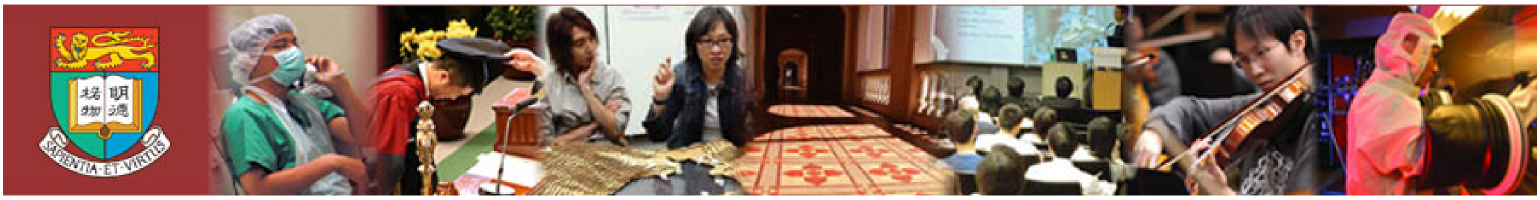

Title	Probabilistic QoS Constrained Robust Downlink Multiuser MIMO Transceiver Design with Arbitrarily Distributed Channel Uncertainty
Author(s)	He, X; Wu, YC
Citation	IEEE Transactions on Wireless Communications, 2013, v. 12 n. 12, p. 6292-6302
Issued Date	2013
URL	http://hdl.handle.net/10722/199094
Rights	IEEE Transactions on Wireless Communications. Copyright © IEEE.

Probabilistic QoS Constrained Robust Downlink Multiuser MIMO Transceiver Design with Arbitrarily Distributed Channel Uncertainty

Xin He and Yik-Chung Wu

***Abstract*—We study the robust transceiver optimization in downlink multiuser multiple-input multiple-output (MU-MIMO) systems aiming at minimizing transmit power under probabilistic quality-of-service (QoS) requirements. Owing to the unknown distributed interference, the channel estimation error obtained from the linear minimum mean square error (LMMSE) estimator can be arbitrarily distributed. Under this situation, the QoS requirements should account for the worst-case channel estimation error distribution. While directly finding the worst-case distribution is challenging, two methods are proposed to solve the robust transceiver design problem. One is based on the Markov's inequality, while the other is based on a novel duality method. Two convergence-guaranteed iterative algorithms are proposed to solve the transceiver design problems. Furthermore, for the special case of MU multiple-input single-output (MISO) systems, the corresponding robust transceiver design problems are shown to be convex. Simulation results show that, compared to the non-robust method, the QoS requirement is satisfied by both proposed algorithms. Among the two proposed methods, the duality method shows a superior performance in transmit power, while the Markov method demonstrates a lower computational complexity. Furthermore, the proposed duality method results in less conservative QoS performance than the Gaussian approximated probabilistic robust method and bounded robust method.**

***Index Terms*—LMMSE channel estimation, QoS, robust MU-MIMO transceiver design, arbitrarily distributed uncertainty.**

I. INTRODUCTION

IN downlink MU-MIMO systems, there are two commonly used transceiver design methodologies, aiming at different goals. The first one is to maximize or minimize a performance metric, such as capacity, mutual information or sum data mean square error (MSE) subject to a power constraint [1]- [4]. However, this framework does not take fairness into account in multiuser system. The other methodology is QoS and fairness based design: minimizing total transmit power subject to different QoS constraints [5]- [7]. Fairness among users is automatically introduced by defining different QoS requirements.

For QoS constraint, it can be in terms of MSE or bit error rate (BER). If the noise is Gaussian, BER is strongly related to the signal-to-interference-plus-noise ratio (SINR) through the Gaussian $\mathcal{Q}$ function, and the SINR in each data stream is related to the inverse of MSE [8], [9]. Therefore, the BER requirement on the substream can be transformed into the MSE requirement on the substream. However, in practice, owing to the interference from unintended users (e.g., co-channel interference in a cellular system or interference from secondary user in cognitive radio systems), or even hostile jamming signals, the distribution of the interference plus noise might be unknown. In this case, the relation between BER and MSE cannot be expressed explicitly, and directly using the BER as the QoS constraint is intractable.

In general, channel state information (CSI) is required for transceiver design [1]- [7]. In practice, CSI has to to be estimated, and estimation error is unavoidable. Therefore, robust transceiver design, which takes the channel estimation uncertainty into consideration, is important. For existing works, the estimation errors are usually assumed to be bounded or Gaussian distributed. In [10]- [12], bounded channel estimation error is introduced in the QoS constraints in the multiuser transceiver design problem. Nevertheless, in general, the distribution of estimation error of a random variable is unbounded [13]. To tackle this problem, several approximation based probabilistic robust transceiver designs with Gaussian distributed channel uncertainty were proposed. In uplink SIMO system, [14] aims to minimize total users' transmit power with individual probabilistic SINR constraints. However, the method is only suitable for users with single antenna. Furthermore, owing to the difference between uplink and downlink inter-user interference management, the method in [14] cannot be generalized to downlink system. For cognitive radio systems when communication is carried out between single antenna secondary users, a robust power control is proposed in [15] aiming at maximizing secondary users' capacity subject to probabilistically constrained interference to the primary user. The geometric programming (GP) method was used to solve the power allocation problem. However, the GP method cannot be extended to transceiver matrices design in MU-MIMO system. In single-user MIMO system, a robust transmitter was designed in [16] aiming at maximizing average SINR while total SINR was probabilistically constrained. However, the framework of [16] cannot handle individual SINR constraint on each data stream. Only in a recent paper [17], a robust transmitter design in downlink MU-MISO system is proposed to minimize total transmit power subject to individual probabilistic SINR constraint for different users. However, there is no corresponding receiver design in [17].

Manuscript received February 23, 2013; revised July 4 and September 12, 2013; accepted September 23, 2013. The associate editor coordinating the review of this paper and approving it for publication was Y.-C. Ko.

The authors are with the Department of Electrical and Electronic Engineering, The University of Hong Kong, Pokfulam Road, Hong Kong (e-mail: {hexin, ycwu}@eee.hku.hk).

Digital Object Identifier 10.1109/TWC.2013.102413.130343

In this paper, we consider a general scenario that the distribution of the interference plus noise is not known, and the distribution of the channel estimation error under interference cannot be modeled in prior. To the best of our knowledge, this is the first work designing probabilistic QoS based robust MU-MIMO transceiver when the channel uncertainty distribution is not known. In particular, we formulate the probabilistic QoS constraints under arbitrarily distributed channel estimation error into worst-case probabilistic constraints, and two methods are proposed to tackle this problem. One is based on the Markov's inequality, which provides an upper bound for the worst-case probability. The other is based on a novel duality method, in which the worst-case probability problem is transformed into a deterministic finite constrained problem by using the Lagrange duality and S-Lemma, with strong duality guaranteed. For both proposed methods, the resulting nonconvex problems are solved by convergence-guaranteed iterative algorithms between two convex subproblems. Furthermore, for the special case of MU-MISO systems, the robust transceiver design problems are shown to be convex, and thus global optimal solutions can be guaranteed. Simulation results show that the duality method has an excellent performance on the transmit power with QoS guaranteed, while the Markov method exhibits low computational complexity. Furthermore, comparisons with Gaussian approximated robust method [17] and bounded robust method [11] show superior performance of the proposed duality method.

The rest of this paper is organized as follows. In Section II, the system model and the statistical information of the CSI error modeling is investigated under unknown distributed interference. In Section III, two different methods are proposed to solve the robust transceiver design problem under the probabilistic QoS requirements. Simulation results are presented in Section IV, and conclusions are drawn in Section V.

Notation: In this paper, $\mathbb{E}(\cdot)$, $(\cdot)^*$, $(\cdot)^T$, and $(\cdot)^H$ denote statistical expectation, conjugation, transposition and Hermitian, respectively, while $\|\cdot\|_2$ denotes the norm of a vector. In addition, $\mathrm{Tr}(\cdot)$ and $\|\cdot\|_F$ refer to the trace and Frobenius norm of a matrix, respectively. The notations $\mathrm{vec}(\cdot)$, $\circ$ and $\otimes$ stands for the vectorization, Hadamard and Kronecker product, respectively. Symbol $\mathrm{diag}(\mathbf{x})$ denotes a diagonal matrix with vector $\mathbf{x}$ on its diagonal. Finally, $\mathbf{I}_K$ is a $K \times K$ identity matrix.

II. System Model

A. Downlink MU-MIMO System

The downlink MU-MIMO system under consideration consists of one base station (BS) equipped with N transmit antennas and K active users, with the k^{th} user equipped with M_k antennas ($\sum_{k=1}^{K} M_k = M$). It is assumed that L_k independent data streams are transmitted to the k^{th} user and $\sum_{k=1}^{K} L_k = L$. In order to guarantee data recovery at the users, it is required that $L_k \le M_k$ and $L \le \min\{M, N\}$. Let $\mathbf{s}_k$ be the $L_k \times 1$ data vector transmitted to the k^{th} user, we have $\mathbb{E}(\mathbf{s}_k\mathbf{s}_k^H) = \mathbf{I}_{L_k}$.

Let $\mathbf{G}$ be the $N \times L$ precoding matrix at the BS, and $\mathbf{s} = [\mathbf{s}_1^T \cdots \mathbf{s}_K^T]^T$ is the data vector for all active users, then the received $M_k \times 1$ signal at the k^{th} user is

$$\mathbf{y}_k = \mathbf{H}_k\mathbf{G}\mathbf{s} + \mathbf{n}_k, \tag{1}$$

where $\mathbf{H}_k$ and $\mathbf{n}_k$ are the $M_k \times N$ channel matrix and the received $M_k \times 1$ interference plus noise vector of the k^{th} user, respectively. It is assumed that the interference plus noise can be arbitrarily distributed with its first two moments being known. Without loss of generality, we assume $\mathbb{E}(\mathbf{n}_k) = \mathbf{0}$ and $\mathbb{E}(\mathbf{n}_k\mathbf{n}_k^H) = \mathbf{R}_k$, and we write $\mathbf{n}_k \sim \mathcal{A}(\mathbf{0}, \mathbf{R}_k)$, where $\mathcal{A}$ denotes an arbitrary distribution.

At the receiver, an $L_k \times M_k$ equalizer $\mathbf{F}_k$ is used to process the received signal. The recovered $L_k \times 1$ data vector at the k^{th} user is

$$\hat{\mathbf{s}}_k = \mathbf{F}_k\mathbf{H}_k\mathbf{G}\mathbf{s} + \mathbf{F}_k\mathbf{n}_k. \tag{2}$$

Since the transmitted data are independent with the interference and noise, the total MSE of the k^{th} user's data can be calculated as

$$\mathrm{MSE}_k = \mathbb{E}_{\mathbf{s}_k,\mathbf{n}_k}\left[\mathrm{Tr}\{(\mathbf{D}_k\mathbf{s} - \hat{\mathbf{s}}_k)(\mathbf{D}_k\mathbf{s} - \hat{\mathbf{s}}_k)^H\}\right] \tag{3}$$

$$= \|\mathbf{F}_k\mathbf{H}_k\mathbf{G} - \mathbf{D}_k\|_F^2 + \mathrm{Tr}(\mathbf{F}_k\mathbf{R}_k\mathbf{F}_k^H), \tag{4}$$

where the matrix $\mathbf{D}_k = [\mathbf{0}_{L_k \times \sum_{k=1}^{k-1} L_k} \quad \mathbf{I}_{L_k} \quad \mathbf{0}_{L_k \times \sum_{k=k+1}^{K} L_k}]$ is used to select the k^{th} user's data streams. The basic QoS based MU-MIMO transceiver design problem is [7]

$$\begin{aligned} \min_{\mathbf{G},\mathbf{F}_1,\ldots,\mathbf{F}_K} \quad & \mathrm{Tr}(\mathbf{G}\mathbf{G}^H) \\ \text{s.t.} \quad & \mathrm{MSE}_k(\mathbf{G}, \mathbf{F}_k) \le \varepsilon_k, \quad k = 1, \cdots, K, \end{aligned}$$

where ε_k is the required QoS target in terms of MSE. However, as shown in (4), in addition to the precoder $\mathbf{G}$ and equalizer $\mathbf{F}_k$, the MSE also depends on the channel realization $\mathbf{H}_k$, which has to be estimated in practice.

B. Channel Estimation Under Arbitrarily Distributed Interference

During channel estimation, an $N \times L_t$ training matrix $\mathbf{S}$ is transmitted from the BS, where L_t is the training length with $L_t \ge N$. Let $\mathbf{G} = \mathbf{I}_N$ during training, the received signal at the k^{th} receiver with observation window length L_t is

$$\mathbf{Y}_k = \mathbf{H}_k\mathbf{S} + \mathbf{N}_k^t, \tag{5}$$

where $\mathbf{N}_k^t$ is an $M_k \times L_t$ matrix with arbitrarily distributed elements and represent interference plus noise at the training stage. In this paper, the well-known Kronecker model is used for the MIMO channel [18], i.e., $\mathbf{H}_k = \mathbf{R}_{rk}^{\frac{1}{2}}\mathbf{H}_{wk}\mathbf{R}_t^{\frac{1}{2}}$, where $\mathbf{R}_{rk}$ and $\mathbf{R}_t$ are the correlation matrix at the k^{th} receiver and the BS, respectively, and $\mathbf{H}_{wk}$ has independent and identically distributed (i.i.d.) complex Gaussian elements distributed as $\mathcal{CN}(0, 1)$. After taking vectorization on both sides of (5), we have $\mathrm{vec}(\mathbf{Y}_k) = (\mathbf{S}^T \otimes \mathbf{I}_{M_k})\mathrm{vec}(\mathbf{H}_k) + \mathrm{vec}(\mathbf{N}_k^t)$. With $\mathbf{y}_k^t \triangleq \mathrm{vec}(\mathbf{Y}_k)$, $\mathbf{h}_k \triangleq \mathrm{vec}(\mathbf{H}_k)$ and $\mathbf{n}_k^t \triangleq \mathrm{vec}(\mathbf{N}_k^t)$, we get

$$\mathbf{y}_k^t = (\mathbf{S}^T \otimes \mathbf{I}_{M_k})\mathbf{h}_k + \mathbf{n}_k^t, \tag{6}$$

where the first two moments of $\mathbf{n}_k^t$ is assumed to be known, i.e., $\mathbf{n}_k^t \sim \mathcal{A}(\mathbf{0}, \mathbf{\Phi}_k)$. According to the Kronecker channel model, we have $\mathbf{h}_k \sim \mathcal{CN}(\mathbf{0}, \mathbf{R}_t^T \otimes \mathbf{R}_{rk})$.

In order to utilize the statistical information of the channel and the interference plus noise, we employ the linear minimum mean squared error (LMMSE) estimator, which takes the expression $\hat{\mathbf{h}}_k = \left((\mathbf{R}_t^{-T} \otimes \mathbf{R}_{rk}^{-1}) + (\mathbf{S}^* \otimes \mathbf{I}_{M_k})\boldsymbol{\Phi}_k^{-1}(\mathbf{S}^T \otimes \mathbf{I}_{M_k})\right)^{-1} \cdot (\mathbf{S}^* \otimes \mathbf{I}_{M_k})\boldsymbol{\Phi}_k^{-1} \cdot \mathbf{y}_k$. It is easy to prove that the mean of the channel estimation error is zero, i.e., $\mathbb{E}(\mathbf{h}_k - \hat{\mathbf{h}}_k) = \mathbf{0}$, and the covariance of the error is [13]

$$\boldsymbol{\Sigma}_k = \left((\mathbf{R}_t^{-T} \otimes \mathbf{R}_{rk}^{-1}) + (\mathbf{S}^* \otimes \mathbf{I}_{M_k})\boldsymbol{\Phi}_k^{-1}(\mathbf{S}^T \otimes \mathbf{I}_{M_k})\right)^{-1}. \quad (7)$$

Therefore, the channel estimation error can be modeled as

$$\mathbf{H}_k = \hat{\mathbf{H}}_k + \boldsymbol{\Delta}_k, \quad (8)$$

where $\mathrm{vec}(\hat{\mathbf{H}}_k) = \hat{\mathbf{h}}_k$. Since the distribution of the interference plus noise is unknown, we can only model the channel estimation uncertainty $\boldsymbol{\Delta}_k$ as arbitrarily distributed with its first two moments known, i.e., $\mathrm{vec}(\boldsymbol{\Delta}_k) \sim \mathcal{A}(\mathbf{0}, \boldsymbol{\Sigma}_k)$.

After substituting (8) into (4), the total MSE of the k^{th} user is

$$\mathrm{MSE}_k(\boldsymbol{\Delta}_k) = \|\mathbf{F}_k(\hat{\mathbf{H}}_k + \boldsymbol{\Delta}_k)\mathbf{G} - \mathbf{D}_k\|_F^2 + \mathrm{Tr}(\mathbf{F}_k\mathbf{R}_k\mathbf{F}_k^H). \quad (9)$$

Note that the distribution of the MSE depends on that of the $\boldsymbol{\Delta}_k$. Furthermore, $\mathbf{G}$ and $\mathbf{F}_k$ are unknown, and in general depends on the statistical information of $\boldsymbol{\Delta}_k$. Therefore, the distribution of the MSE cannot be obtained or approximated in advance.

Remark 1. Besides the estimation error, the quantization error is also unavoidable in frequency-division duplex (FDD) systems (e.g., channel is estimated at the users and then quantized and fed back to the BS.). Channel directional information is usually quantized by Grassmannian codebook or random vector quantization (RVQ) [19], [20], while the channel quality information is quantized by entropy-maximizing method [21]. In general, accurate analytical statistical model of quantization error is intractable as it depends on the distribution of the estimated channel $\hat{\mathbf{H}}_k$. For example, for the simple case of scalar quantization [12], only an error bound can be derived. For RVQ technique with i.i.d. Rayleigh distributed channel, only the error angle is modeled. Fortunately, empirical mean and variance of the quantization error can be obtained once the codebook is fixed. Since the channel estimation error is independent with the quantization error, the combined channel uncertainty can also be modeled using (8) with updated mean and variance. As the bias of the quantization error can be shifted to the estimated channel, the channel uncertainty can still be modeled as $\mathrm{vec}(\boldsymbol{\Delta}_k) \sim \mathcal{A}(\mathbf{0}, \boldsymbol{\Sigma}_k)$.

III. Robust MU-MIMO Transceiver Design

In multiuser system, different users may have different QoS requirements, which can be formulated as $\mathrm{Prob}\{\mathrm{MSE}_k(\boldsymbol{\Delta}_k) \geq \varepsilon_k\} \leq p_k$, where ε_k and p_k are the required QoS target and guaranteed probability at the k^{th} receiver, respectively. After the QoS requirements are satisfied, it is crucial to save transmission power at the BS. Therefore, the probabilistic QoS constrained transceiver design problem can be formulated as

$$\begin{aligned} &\min_{\mathbf{G},\mathbf{F}_1,\ldots,\mathbf{F}_K} \mathrm{Tr}(\mathbf{G}\mathbf{G}^H) \\ &\text{s.t.} \sup_{\mathrm{vec}(\boldsymbol{\Delta}_k) \sim \mathcal{A}(\mathbf{0},\boldsymbol{\Sigma}_k)} \mathrm{Prob}\{\mathrm{MSE}_k(\boldsymbol{\Delta}_k) \geq \varepsilon_k\} \leq p_k, \quad \forall k. \end{aligned} \quad (\mathbb{P}0)$$

In the QoS constraints, due to the arbitrarily distributed channel estimation errors, the supremum of the outage probability is used to guarantee the QoS performance is satisfied even at the worst case.

Note that the problem ($\mathbb{P}0$) is a bilevel optimization problem, and the lower level problem involves finding the supremum of the outage probability. Owing to the unknown distribution of $\boldsymbol{\Delta}_k$, it is difficult to get an analytical solution for the lower level problem. Below, we consider two methods to solve this problem. The first one is to use the Markov's inequality to eliminate the supremum. Another method is the proposed duality method, in which the lower level problem is transformed from a stochastic problem into a solvable deterministic problem.

A. The Markov's Inequality Based Method

Since the distribution of the MSE is unknown, the Markov's inequality [22] can be used to get an upper bound of the outage probability

$$\mathrm{Prob}\{\mathrm{MSE}_k(\boldsymbol{\Delta}_k) \geq \varepsilon_k\} \leq \frac{\mathbb{E}_{\boldsymbol{\Delta}_k}\{\mathrm{MSE}_k(\boldsymbol{\Delta}_k)\}}{\varepsilon_k}, \quad (10)$$

where $\mathbb{E}_{\boldsymbol{\Delta}_k}\{\mathrm{MSE}_k(\boldsymbol{\Delta}_k)\}$ can be derived as

$$\begin{aligned} \mathbb{E}_{\boldsymbol{\Delta}_k}\{\mathrm{MSE}_k\} &= \mathbb{E}_{\boldsymbol{\Delta}_k}\{\|\mathbf{F}_k\boldsymbol{\Delta}_k\mathbf{G}\|_F^2\} + \|\mathbf{F}_k\hat{\mathbf{H}}_k\mathbf{G} - \mathbf{D}_k\|_F^2 \\ &\quad + \|\mathbf{R}_k^{1/2}\mathbf{F}_k^H\|_F^2 & (11) \\ &= \mathbb{E}_{\boldsymbol{\Delta}_k}\{\|(\mathbf{G}^T \otimes \mathbf{F}_k)\mathrm{vec}(\boldsymbol{\Delta}_k)\|_2^2\} \\ &\quad + \|\mathbf{F}_k\hat{\mathbf{H}}_k\mathbf{G} - \mathbf{D}_k\|_F^2 + \|\mathbf{R}_k^{1/2}\mathbf{F}_k^H\|_F^2 & (12) \\ &= \|[\mathrm{vec}\left((\mathbf{G}^T \otimes \mathbf{F}_k)\boldsymbol{\Sigma}_k^{\frac{1}{2}}\right)^T \\ &\quad \mathrm{vec}(\mathbf{F}_k\hat{\mathbf{H}}_k\mathbf{G} - \mathbf{D}_k)^T \mathrm{vec}(\mathbf{R}_k^{1/2}\mathbf{F}_k^H)^T]\|_2^2. & (13) \end{aligned}$$

Since the right hand side of (10) is independent of the exact distribution, the problem ($\mathbb{P}0$) can be approximated as

$$\begin{aligned} &\min_{\mathbf{G},\mathbf{F}_1,\ldots,\mathbf{F}_K} \mathrm{Tr}(\mathbf{G}\mathbf{G}^H) \\ &\text{s.t.} \ \|[\mathrm{vec}\left((\mathbf{G}^T \otimes \mathbf{F}_k)\boldsymbol{\Sigma}_k^{\frac{1}{2}}\right)^T \mathrm{vec}(\mathbf{F}_k\hat{\mathbf{H}}_k\mathbf{G} - \mathbf{D}_k)^T \\ &\qquad \mathrm{vec}(\mathbf{R}_k^{1/2}\mathbf{F}_k^H)^T]\|_2^2 \leq p_k\varepsilon_k, \quad \forall k. \end{aligned} \quad (14)$$

From (14), it is obvious that the probabilistic robust approach becomes a statistically average approach [23], [24]. Although this problem is still nonconvex, it is observed that it becomes a convex problem of $\mathbf{G}$ when all equalizer $\mathbf{F}_k$ are fixed. Furthermore, when the precoder $\mathbf{G}$ is fixed, the objective function is not related to the equalizer $\mathbf{F}_k$. In order to provide a larger feasible space for the next round precoder design, the equalizer $\mathbf{F}_k$ can be used to minimize the left side of the QoS constraints in (14). Therefore, the problem (14) can be solved by iterations between two convex subproblems as follows.

In the first subproblem, with the equalizer $\mathbf{F}_k$ fixed, the optimum precoder $\mathbf{G}$ can be obtained from the second-order cone programming (SOCP) problem

$$\begin{aligned} &\min_{\mathbf{G},P} P \\ &\text{s.t.} \ \|\mathrm{vec}(\mathbf{G})\|_2 \leq P \\ &\quad \|[\mathrm{vec}\left((\mathbf{G}^T \otimes \mathbf{F}_k)\boldsymbol{\Sigma}_k^{\frac{1}{2}}\right)^T \mathrm{vec}(\mathbf{F}_k\hat{\mathbf{H}}_k\mathbf{G} - \mathbf{D}_k)^T \\ &\qquad \mathrm{vec}(\mathbf{R}_k^{1/2}\mathbf{F}_k^H)^T]\|_2 \leq \sqrt{p_k\varepsilon_k}, \quad \forall k, \end{aligned} \quad (15)$$

where P is a slack variable.

In the second subproblem, with the precoder $\mathbf{G}$ fixed, the equalizer $\mathbf{F}_k$ can be used to minimize the left side of the QoS constraint. Expressing the left side of the QoS constraint in (14) with the Frobenius norm, the equalizer $\mathbf{F}_k$ can be updated from the following problem

$$\min_{\mathbf{F}_k} \|(\mathbf{G}^T \otimes \mathbf{F}_k)\boldsymbol{\Sigma}_k^{\frac{1}{2}}\|_F^2 + \|\mathbf{F}_k\hat{\mathbf{H}}_k\mathbf{G} - \mathbf{D}_k\|_F^2 + \|\mathbf{R}_k^{1/2}\mathbf{F}_k^H\|_F^2. \quad (16)$$

Furthermore, writing the first term of the cost function (16) as a quadratic form,

$$\|(\mathbf{G}^T \otimes \mathbf{F}_k)\boldsymbol{\Sigma}_k^{\frac{1}{2}}\|_F^2 = \mathrm{Tr}\left(\boldsymbol{\Sigma}_k\left((\mathbf{G}^*\mathbf{G}^T) \otimes (\mathbf{F}_k^H\mathbf{F}_k)\right)\right) \quad (17)$$

$$= \mathrm{Tr}(\sum_{j=1}^{N}\sum_{i=1}^{N} g_{ij}\boldsymbol{\Sigma}_k^{ji}\mathbf{F}_k^H\mathbf{F}_k), \quad (18)$$

where g_{ij} is the $(i,j)^{\text{th}}$ element of the matrix $\mathbf{G}^*\mathbf{G}^T$, $\boldsymbol{\Sigma}_k^{ji}$ is the $(j,i)^{\text{th}}$ $M_k \times M_k$ subblock of the matrix $\boldsymbol{\Sigma}_k$. Putting (18) into (16), and setting the derivative of the cost function (16) with respect to $\mathbf{F}_k^*$ to zero, the optimum equalizer $\mathbf{F}_k$ is

$$\mathbf{F}_k = (\hat{\mathbf{H}}_k\mathbf{G}\mathbf{D}_k^H)^H(\hat{\mathbf{H}}_k\mathbf{G}\mathbf{G}^H\hat{\mathbf{H}}_k^H + \mathbf{R}_k + \sum_{j=1}^{N}\sum_{i=1}^{N} g_{ij}\boldsymbol{\Sigma}_k^{ji})^{-1}. \quad (19)$$

It is observed that the obtained equalizer is a conventional MMSE equalizer with additional regularization from the weighted uncertainty $\sum_{j=1}^{N}\sum_{i=1}^{N} g_{ij}\boldsymbol{\Sigma}_k^{ji}$.

The iterative algorithm between the two subproblems is summarized at Table I. With a feasible initialization, the transmit power in each iteration obtained by the iterative algorithm decreases monotonically and converges. Similar proofs have been provided in [11].

With regard to the initialization, it is a common problem for the QoS based MU-MIMO transceiver design since a feasible initial transceiver pair is required [7] [11]. Conventionally, the receiver $\mathbf{F}_k$ is initialized with an identity or a randomly generated matrix. However, these initializations are not guaranteed to satisfy the QoS constraints, and the precoder $\mathbf{G}$ design in (15) may not exist. It is observed in (15) that if the elements of $\mathbf{F}_k$ are small, we have a better chance of satisfying the QoS constraints. Based on this observation, it is suggested in [11] that scaling factors $1/a_k$ are introduced into the initial chosen equalizer in (15), and the scaling factors can be obtained from the following SOCP problem

$$\begin{aligned} &\min_{\mathbf{G},P,\gamma_1,\ldots,\gamma_K} P \\ &\text{s.t. } \|\mathrm{vec}(\mathbf{G})\|_2 \le P \\ &\quad \|[\mathrm{vec}\Big((\mathbf{G}^T \otimes \mathbf{F}_{ko})\boldsymbol{\Sigma}_k^{\frac{1}{2}}\Big)^T \mathrm{vec}(\mathbf{F}_{ko}\hat{\mathbf{H}}_k\mathbf{G} - a_k\mathbf{D}_k)^T \\ &\quad \mathrm{vec}(\mathbf{R}_k^{1/2}\mathbf{F}_{ko}^H)^T]\|_2 \le a_k\sqrt{p_k\varepsilon_k},\ \forall k, \end{aligned} \quad (20)$$

where $\mathbf{F}_{ko}$ is the initial chosen equalizer. Compared to the precoder design with fixed equalizers in (15), the joint precoder and scalable equalizers $(\mathbf{G}, \frac{1}{a_1}\mathbf{F}_{1o}, \cdots, \frac{1}{a_K}\mathbf{F}_{Ko})$ in (20) has more degree of freedom and thus has a higher chance of satisfying the QoS constraints. If problem (20) is feasible, the result $(\mathbf{G}, \frac{1}{a_1}\mathbf{F}_{1o}, \cdots, \frac{1}{a_K}\mathbf{F}_{Ko})$ is used as the initial starting point for the iterative algorithm. Otherwise, another $\mathbf{F}_{ko}$ with a different beamforming direction may be chosen and (20) is solved again. In practice, the feasibility problem can also be mitigated by utilizing it as the user admission criterion if the number of active user is large [25]. However, the corresponding cross-layer design problem is beyond the scope of this paper.

TABLE I
MARKOV'S INEQUALITY BASED ROBUST TRANSCEIVER DESIGN

1. Set iteration number $j = 1$, initialize with a feasible transceiver set $(\mathbf{G}[0], \mathbf{F}_1[0], \cdots, \mathbf{F}_K[0])$, define $P[0] = \mathrm{Tr}(\mathbf{G}[0]^H\mathbf{G}[0])$
2. Update $\mathbf{G}[j]$ using the solution of (15), calculate $P[j]=\mathrm{Tr}(\mathbf{G}[j]^H\mathbf{G}[j])$
3. Update $\mathbf{F}_k[j]$ using (19)
4. If $P[j-1] - P[j] \le \epsilon$ (ϵ is a pre-defined threshold) then stop, otherwise increment j and go to step 2

B. The Duality Based Method

In the MSE expression (9), the random variable $\boldsymbol{\Delta}_k$ is weighted by unknown $\mathbf{F}_k$ and $\mathbf{G}$, whose values in general depend on $\boldsymbol{\Delta}_k$. Therefore, the MSE is a sum of correlated elements. According to the generalized weak-convergence theorem [26], a sum of many random variables with dependence will tend to be distributed according to one of a small set of stable distributions. This means that although the channel estimation uncertainty $\boldsymbol{\Delta}_k$ is arbitrarily distributed, the MSE in (9) is in fact not arbitrarily distributed. Therefore, the Markov's inequality in (10) is quite loose [22], and the QoS and power saving performance of the Markov method is expected to be conservative. In this subsection, the exact solution of the lower level problem is derived by the proposed duality method.

Since the MSE is a function of the channel estimation uncertainty, let $\psi_k(\mathbf{x}_k) = \mathrm{MSE}_k(\boldsymbol{\Delta}_k)$, where $\mathbf{x}_k \triangleq \mathrm{vec}(\boldsymbol{\Delta}_k)$. The lower level problem $\sup_{\mathrm{vec}(\boldsymbol{\Delta}_k)\sim\mathcal{A}(\mathbf{0},\boldsymbol{\Sigma}_k)} \mathrm{Prob}\{\mathrm{MSE}_k(\boldsymbol{\Delta}_k) \ge \varepsilon_k\}$ can be reformulated as

$$\begin{aligned} \sup_{f(\mathbf{x}_k)} \quad & \mathrm{Prob}\{\psi_k(\mathbf{x}_k) \ge \varepsilon_k\} \\ \text{s.t.} \quad & \int_{\mathbf{x}_k\in\mathbb{C}^{NM_k}} f(\mathbf{x}_k)d\mathbf{x}_k = 1 \\ & \mathbb{E}(\mathbf{x}_k) = \mathbf{0} \\ & \mathbb{E}(\mathbf{x}_k\mathbf{x}_k^H) = \boldsymbol{\Sigma}_k, \end{aligned} \quad (21)$$

where $f(\mathbf{x}_k)$ is the probability density function (PDF) of the vectorized uncertainty $\mathbf{x}_k$.

Transforming problem (21) *into its dual problem*

The Lagrangian of this problem is presented as

$$\begin{aligned} &\mathcal{L}_k\left(f(\mathbf{x}_k), \alpha_k, \boldsymbol{\eta}_k, \boldsymbol{\Xi}_k\right) \\ &= \mathrm{Prob}\{\psi_k(\mathbf{x}_k) \ge \varepsilon_k\} + \alpha_k\left(1 - \int_{\mathbf{x}_k\in\mathbb{C}^{NM_k}} f(\mathbf{x}_k)d\mathbf{x}_k\right) \\ &\quad + \boldsymbol{\eta}_k^H\left(\mathbf{0} - \mathbb{E}(\mathbf{x}_k)\right) + \mathrm{Tr}\left(\boldsymbol{\Xi}_k^H(\boldsymbol{\Sigma}_k - \mathbb{E}(\mathbf{x}_k\mathbf{x}_k^H))\right) \quad (22) \\ &= \alpha_k + \mathrm{Tr}(\boldsymbol{\Xi}_k^H\boldsymbol{\Sigma}_k) + \int_{\psi_k(\mathbf{x}_k)\ge\varepsilon_k} f(\mathbf{x}_k)d\mathbf{x}_k \\ &\quad - \int_{\mathbf{x}_k\in\mathbb{C}^{NM_k}} \left(\alpha_k + \boldsymbol{\eta}_k^H\mathbf{x}_k + \mathrm{Tr}(\boldsymbol{\Xi}_k^H\mathbf{x}_k\mathbf{x}_k^H)\right)\cdot f(\mathbf{x}_k)d\mathbf{x}_k \quad (23) \end{aligned}$$

$$
\begin{aligned}
&= \alpha_k + \mathrm{Tr}(\boldsymbol{\Xi}_k^H \boldsymbol{\Sigma}_k) \\
&\quad + \int_{\psi_k(\mathbf{x}_k)\geq\varepsilon_k} \left(1 - \alpha_k - \boldsymbol{\eta}_k^H \mathbf{x}_k - \mathrm{Tr}(\boldsymbol{\Xi}_k^H \mathbf{x}_k \mathbf{x}_k^H)\right) \cdot f(\mathbf{x}_k) d\mathbf{x}_k \\
&\quad + \int_{\psi_k(\mathbf{x}_k)<\varepsilon_k} \left(0 - \alpha_k - \boldsymbol{\eta}_k^H \mathbf{x}_k - \mathrm{Tr}(\boldsymbol{\Xi}_k^H \mathbf{x}_k \mathbf{x}_k^H)\right) \cdot f(\mathbf{x}_k) d\mathbf{x}_k. \quad (24)
\end{aligned}
$$

where $\alpha_k, \boldsymbol{\eta}_k, \boldsymbol{\Xi}_k$ are the Lagrangian multipliers, and $\boldsymbol{\Xi}_k = \boldsymbol{\Xi}_k^H$. With the implicit PDF constraint $f(\mathbf{x}_k) \geq 0$, the Lagrange dual function of the problem (21) is

$$
g_k(\alpha_k, \boldsymbol{\eta}_k, \boldsymbol{\Xi}_k) = \sup_{f(\mathbf{x}_k)\geq 0} \mathcal{L}_k(f(\mathbf{x}_k), \alpha_k, \boldsymbol{\eta}_k, \boldsymbol{\Xi}_k). \quad (25)
$$

Note that the supremum of the first integral in (24) with the nonnegative PDF constraint is zero if $\alpha_k + \boldsymbol{\eta}_k^H \mathbf{x}_k + \mathrm{Tr}(\boldsymbol{\Xi}_k^H \mathbf{x}_k \mathbf{x}_k^H) > 1$, otherwise the supremum is infinity. Similarly, the supremum of the second integral with the nonnegative PDF constraint is also zero if $\alpha_k + \boldsymbol{\eta}_k^H \mathbf{x}_k + \mathrm{Tr}(\boldsymbol{\Xi}_k^H \mathbf{x}_k \mathbf{x}_k^H) \geq 0$, otherwise the supremum is infinity. Therefore, the Lagrange dual function $g_k(\alpha_k, \boldsymbol{\eta}_k, \boldsymbol{\Xi}_k)$ is

$$
g_k(\alpha_k, \boldsymbol{\eta}_k, \boldsymbol{\Xi}_k) = \begin{cases} \alpha_k + \mathrm{Tr}(\boldsymbol{\Xi}_k^H \boldsymbol{\Sigma}_k) & \text{if} \begin{array}{c} \alpha_k + \boldsymbol{\eta}_k^H \mathbf{x}_k + \mathrm{Tr}(\boldsymbol{\Xi}_k^H \mathbf{x}_k \mathbf{x}_k^H) \geq 0, \\ \forall \mathbf{x}_k : \psi_k(\mathbf{x}_k) < \varepsilon_k; \text{ and} \\ \alpha_k + \boldsymbol{\eta}_k^H \mathbf{x}_k + \mathrm{Tr}(\boldsymbol{\Xi}_k^H \mathbf{x}_k \mathbf{x}_k^H) > 1, \\ \forall \mathbf{x}_k : \psi_k(\mathbf{x}_k) \geq \varepsilon_k \end{array} \\ +\infty & \text{otherwise.} \end{cases} \quad (26)
$$

Combining the two conditions that make $g_k(\alpha_k, \boldsymbol{\eta}_k, \boldsymbol{\Xi}_k) = \alpha_k + \mathrm{Tr}(\boldsymbol{\Xi}_k^H \boldsymbol{\Sigma}_k)$, the first condition can be replaced by $\alpha_k + \boldsymbol{\eta}_k^H \mathbf{x}_k + \mathrm{Tr}(\boldsymbol{\Xi}_k^H \mathbf{x}_k \mathbf{x}_k^H) \geq 0$ for all $\mathbf{x}_k \in \mathbb{C}^{NM_k}$. Therefore, the dual of the problem (21) can be formulated as

$$
\begin{aligned}
\min_{\alpha_k, \boldsymbol{\eta}_k, \boldsymbol{\Xi}_k} \quad & \alpha_k + \mathrm{Tr}(\boldsymbol{\Xi}_k^H \boldsymbol{\Sigma}_k) \\
\text{s.t.} \quad & \alpha_k + \boldsymbol{\eta}_k^H \mathbf{x}_k + \mathrm{Tr}(\boldsymbol{\Xi}_k^H \mathbf{x}_k \mathbf{x}_k^H) \geq 0, \quad \forall \mathbf{x}_k : \mathbf{x}_k \in \mathbb{C}^{NM_k} \\
& \alpha_k + \boldsymbol{\eta}_k^H \mathbf{x}_k + \mathrm{Tr}(\boldsymbol{\Xi}_k^H \mathbf{x}_k \mathbf{x}_k^H) > 1, \quad \forall \mathbf{x}_\mathbf{k} : \psi_k(\mathbf{x}_k) \geq \varepsilon_k \\
& \boldsymbol{\Xi}_k = \boldsymbol{\Xi}_k^H.
\end{aligned} \quad (27)
$$

Remark 2. It is recognized that the problem (21) is known as the moment problem [27]. Since only the first two moments of the random vector $\mathbf{x}_k$ are used in (21), the feasible moment vector set of $\mathbf{x}_k$ is $\mathcal{M} = \{(\mathbf{0}, \boldsymbol{\Sigma}_k) | \boldsymbol{\Sigma}_k \succeq 0\}$. According to the general theory on the moment problem [27, p. 812], strong duality holds between the primal moment problem (21) and its dual problem (27) when the moment vector of $\mathbf{x}_k$ is an interior point of $\mathcal{M}$, i.e., $\boldsymbol{\Sigma}_k \succ 0$.

Remark 3. From (7), the condition $\boldsymbol{\Sigma}_k \succ 0$ is equivalent to requiring that either the covariance matrix of the channel $\mathbf{R}_t^T \otimes \mathbf{R}_{rk}$ or the interference plus noise covariance matrix $\boldsymbol{\Phi}_k$ is positive definite. Furthermore, when strong duality holds, the optimal value of (21) is equivalent to that of (27).

In order to get a compact form of (27), we define $\mathbf{Z}_k \triangleq \begin{bmatrix} \boldsymbol{\Xi}_k^H & \frac{1}{2}\boldsymbol{\eta}_k \\ \frac{1}{2}\boldsymbol{\eta}_k^H & \alpha_k \end{bmatrix}$, $\tilde{\boldsymbol{\Sigma}}_k \triangleq \begin{bmatrix} \boldsymbol{\Sigma}_k & \mathbf{0} \\ \mathbf{0} & 1 \end{bmatrix}$ and $\mathbf{u}_k \triangleq \begin{bmatrix} \mathbf{x}_k^T & 1 \end{bmatrix}^T$, then we get

$$
\begin{aligned}
\min_{\mathbf{Z}_k} \quad & \mathrm{Tr}(\mathbf{Z}_k \tilde{\boldsymbol{\Sigma}}_k) \\
\text{s.t.} \quad & \mathbf{u}_k^H \mathbf{Z}_k \mathbf{u}_k \geq 0, \quad \forall \mathbf{x}_k : \mathbf{x}_k \in \mathbb{C}^{NM_k} \\
& \mathbf{u}_k^H \mathbf{Z}_k \mathbf{u}_k - 1 > 0, \quad \forall \mathbf{x}_k : \psi_k(\mathbf{x}_k) \geq \varepsilon_k \\
& \mathbf{Z}_k = \mathbf{Z}_k^H.
\end{aligned} \quad (28)
$$

Note that the first and the third constraints of (28) can be combined into $\mathbf{Z}_k \succeq 0$. Furthermore, after replacing the MSE term $\psi_k(\mathbf{x}_k)$ with $\mathrm{MSE}_k(\boldsymbol{\Delta}_k)$ in (9), the problem (28) becomes

$$
\begin{aligned}
\min_{\mathbf{Z}_k} \quad & \mathrm{Tr}(\mathbf{Z}_k \tilde{\boldsymbol{\Sigma}}_k) \\
\text{s.t.} \quad & \mathbf{Z}_k \succeq 0 \\
& \mathbf{u}_k^H \mathbf{Z}_k \mathbf{u}_k - 1 > 0, \forall \boldsymbol{\Delta}_k : \|\mathbf{F}_k(\hat{\mathbf{H}}_k + \boldsymbol{\Delta}_k)\mathbf{G} - \mathbf{D}_k\|_F^2 \\
& \qquad + \mathrm{Tr}(\mathbf{F}_k \mathbf{R}_k \mathbf{F}_k^H) \geq \varepsilon_k.
\end{aligned} \quad (29)
$$

Since (29) does not contain probabilistic constraint, it is a deterministic optimization problem. But it is still an infinite constrained problem.

Transforming problem (29) into finite constrained problem

We first reformulate the Frobenius norm as a spectral norm as follows

$$
\begin{aligned}
&\|\mathbf{F}_k(\hat{\mathbf{H}}_k + \boldsymbol{\Delta}_k)\mathbf{G} - \mathbf{D}_k\|_F^2 \\
&= \|\mathrm{vec}(\mathbf{F}_k(\hat{\mathbf{H}}_k + \boldsymbol{\Delta}_k)\mathbf{G} - \mathbf{D}_k)\|_2^2 \quad (30) \\
&= \|\mathrm{vec}(\mathbf{F}_k \boldsymbol{\Delta}_k \mathbf{G}) + \mathrm{vec}(\mathbf{F}_k \hat{\mathbf{H}}_k \mathbf{G} - \mathbf{D}_k)\|_2^2 \quad (31) \\
&= \left\| \left[\mathbf{G}^T \otimes \mathbf{F}_k \ \ \mathrm{vec}(\mathbf{F}_k \hat{\mathbf{H}}_k \mathbf{G} - \mathbf{D}_k) \right] \begin{bmatrix} \mathrm{vec}(\boldsymbol{\Delta}_k) \\ 1 \end{bmatrix} \right\|_2^2 \quad (32) \\
&= \mathbf{u}_k^H \mathbf{Q}_k^H \mathbf{Q}_k \mathbf{u}_k, \quad (33)
\end{aligned}
$$

where $\mathrm{vec}(\mathbf{F}_k \boldsymbol{\Delta}_k \mathbf{G}) = (\mathbf{G}^T \otimes \mathbf{F}_k)\mathrm{vec}(\boldsymbol{\Delta}_k)$ is used from (31) to (32) [28], and $\mathbf{Q}_k \triangleq [\mathbf{G}^T \otimes \mathbf{F}_k \ \ \mathrm{vec}(\mathbf{F}_k \hat{\mathbf{H}}_k \mathbf{G} - \mathbf{D}_k)]$. Putting (33) into the condition of the second constraint of (29), we get the quadratic form

$$
\begin{aligned}
&\|\mathbf{F}_k(\hat{\mathbf{H}}_k + \boldsymbol{\Delta}_k)\mathbf{G} - \mathbf{D}_k\|_F^2 + \mathrm{Tr}(\mathbf{F}_k \mathbf{R}_k \mathbf{F}_k^H) - \varepsilon_k \\
&= \mathbf{u}_k^H \left(\mathbf{Q}_k^H \mathbf{Q}_k - \begin{bmatrix} \mathbf{0}_{NM_k} & \mathbf{0}_{NM_k \times 1} \\ \mathbf{0}_{1 \times NM_k} & \varepsilon_k - \mathrm{Tr}(\mathbf{F}_k \mathbf{R}_k \mathbf{F}_k^H) \end{bmatrix} \right) \mathbf{u}_k \quad (34) \\
&= \mathbf{u}_k^H (\mathbf{Q}_k^H \mathbf{Q}_k - \mathrm{diag}([\mathbf{0}, \varepsilon_k - \mathrm{Tr}(\mathbf{F}_k \mathbf{R}_k \mathbf{F}_k^H)])) \mathbf{u}_k. \quad (35)
\end{aligned}
$$

Therefore, the second constraint of the problem (29) can be transformed as

$$
\begin{aligned}
\mathbf{u}_k^H (\mathbf{Z}_k - \mathrm{diag}([\mathbf{0}, 1])) \mathbf{u}_k > 0, \quad \forall \mathbf{x}_k : \mathbf{u}_k^H \big(\mathbf{Q}_k^H \mathbf{Q}_k \\
- \mathrm{diag}([\mathbf{0}, \varepsilon_k - \mathrm{Tr}(\mathbf{F}_k \mathbf{R}_k \mathbf{F}_k^H)]) \big) \mathbf{u}_k \geq 0. \quad (36)
\end{aligned}
$$

By using the S-Lemma in control theory [29, p. 23], (36) can be equivalently formulated as

$$
\begin{aligned}
\exists \lambda_k \geq 0 : \ \mathbf{Z}_k - \mathrm{diag}([\mathbf{0}, 1]) - \lambda_k \big(\mathbf{Q}_k^H \mathbf{Q}_k \\
- \mathrm{diag}([\mathbf{0}, \varepsilon_k - \mathrm{Tr}(\mathbf{F}_k \mathbf{R}_k \mathbf{F}_k^H)]) \big) \succeq 0. \quad (37)
\end{aligned}
$$

It is shown in Appendix A that the problem $(\mathbb{P}0)$ will be infeasible if $\lambda_k = 0$ for any $k \in \{1, 2, \cdots, K\}$. Therefore, we can omit $\lambda_k = 0$ in (37). Letting $\beta_k \triangleq 1/\lambda_k > 0$, the constraint (37) can be reformulated as

$$
\exists \beta_k > 0 : \ \beta_k \mathbf{Z}_k + \mathrm{diag}([\mathbf{0}, \varepsilon_k - \mathrm{Tr}(\mathbf{F}_k \mathbf{R}_k \mathbf{F}_k^H) - \beta_k]) - \mathbf{Q}_k^H \mathbf{Q}_k \succeq 0. \quad (38)
$$

According to the Schur's Complement [30], (38) can be transformed into an linear matrix inequality (LMI) form of β_k as

$$
\exists \beta_k > 0 : \begin{bmatrix} \beta_k \mathbf{Z}_k + \mathrm{diag}([\mathbf{0}, \varepsilon_k - \mathrm{Tr}(\mathbf{F}_k \mathbf{R}_k \mathbf{F}_k^H) - \beta_k]) & \mathbf{Q}_k^H \\ \mathbf{Q}_k & \mathbf{I}_{LL_k} \end{bmatrix} \succeq 0. \quad (39)
$$

After replacing the second constraint of the problem (29) with its equivalent form (39), the problem (29) becomes

$$\begin{aligned}&\min_{\beta_k,\mathbf{Z}_k} \operatorname{Tr}(\mathbf{Z}_k\tilde{\boldsymbol{\Sigma}}_k)\\ &\text{s.t. } \mathbf{Z}_k \succeq 0, \quad \beta_k > 0\\ &\quad \begin{bmatrix}\beta_k\mathbf{Z}_k+\operatorname{diag}([\mathbf{0},\varepsilon_k-\operatorname{Tr}(\mathbf{F}_k\mathbf{R}_k\mathbf{F}_k^H)-\beta_k]) & \mathbf{Q}_k^H\\ \mathbf{Q}_k & \mathbf{I}_{LL_k}\end{bmatrix}\succeq 0.\end{aligned} \tag{40}$$

Defining a new variable $\tilde{\mathbf{Z}}_k \triangleq \beta_k\mathbf{Z}_k$, then the problem (40) further becomes an LMI problem

$$\begin{aligned}&\min_{\beta_k,\tilde{\mathbf{Z}}_k} \operatorname{Tr}(\tilde{\mathbf{Z}}_k\tilde{\boldsymbol{\Sigma}}_k)/\beta_k\\ &\text{s.t. } \tilde{\mathbf{Z}}_k \succeq 0, \quad \beta_k > 0\\ &\quad \begin{bmatrix}\tilde{\mathbf{Z}}_k+\operatorname{diag}([\mathbf{0},\varepsilon_k-\operatorname{Tr}(\mathbf{F}_k\mathbf{R}_k\mathbf{F}_k^H)-\beta_k]) & \mathbf{Q}_k^H\\ \mathbf{Q}_k & \mathbf{I}_{LL_k}\end{bmatrix}\succeq 0.\end{aligned} \tag{41}$$

Therefore, the stochastic problem (21) is transformed into the deterministic finite constrained problem (41).

Transforming ($\mathbb{P}0$) into a solvable single-level problem

Since the strong duality holds between the primal and dual problem, after replacing the lower level problem of ($\mathbb{P}0$) with (41), we get the following results.

Proposition 1. *The original bilevel optimization problem ($\mathbb{P}0$) is equivalent to the following single-level problem*

$$\begin{aligned}&\min_{\substack{\mathbf{G},\mathbf{F}_1,\ldots,\mathbf{F}_K\\ \beta_1,\ldots,\beta_K\\ \tilde{\mathbf{Z}}_1,\ldots,\tilde{\mathbf{Z}}_K}} \operatorname{Tr}(\mathbf{G}\mathbf{G}^H)\\ &\textit{s.t. } \operatorname{Tr}(\tilde{\mathbf{Z}}_k\tilde{\boldsymbol{\Sigma}}_k)/\beta_k \le p_k, \quad \beta_k>0, \quad \tilde{\mathbf{Z}}_k\succeq 0, \ \forall k\\ &\quad \begin{bmatrix}\tilde{\mathbf{Z}}_k+\operatorname{diag}([\mathbf{0},\varepsilon_k-\operatorname{Tr}(\mathbf{F}_k\mathbf{R}_k\mathbf{F}_k^H)-\beta_k]) & \mathbf{Q}_k^H\\ \mathbf{Q}_k & \mathbf{I}_{LL_k}\end{bmatrix}\succeq 0, \forall k.\end{aligned} \tag{$\mathbb{P}1$}$$

Proof. See Appendix B.

Since the term $\mathbf{Q}_k$ in problem $\mathbb{P}1$ contains the product of $\mathbf{G}$ and $\mathbf{F}_k$, it is a nonconvex problem of $\mathbf{G}$ and $\mathbf{F}_k$. However, it can be solved by an iteration between two subproblems as follows.

In the first subproblem, when all equalizers $\mathbf{F}_k$ are fixed, the subproblem becomes a convex problem of $\mathbf{G}$, and the precoder design problem is

$$\begin{aligned}&\min_{\substack{\mathbf{G},\beta_1,\ldots,\beta_K\\ \tilde{\mathbf{Z}}_1,\ldots,\tilde{\mathbf{Z}}_K}} \operatorname{Tr}(\mathbf{G}\mathbf{G}^H)\\ &\text{s.t. } \operatorname{Tr}(\tilde{\mathbf{Z}}_k\tilde{\boldsymbol{\Sigma}}_k) \le p_k\cdot\beta_k, \quad \beta_k>0, \quad \tilde{\mathbf{Z}}_k\succeq 0, \quad \forall k\\ &\quad \begin{bmatrix}\tilde{\mathbf{Z}}_k+\operatorname{diag}([\mathbf{0},\varepsilon_k-\operatorname{Tr}(\mathbf{F}_k\mathbf{R}_k\mathbf{F}_k^H)-\beta_k]) & \mathbf{Q}_k^H\\ \mathbf{Q}_k & \mathbf{I}_{LL_k}\end{bmatrix}\succeq 0, \forall k.\end{aligned} \tag{42}$$

This convex problem can be efficiently solved by the interior point method in [31, p. 561].

For the second subproblem, similar to the idea in the Markov method, the equalizer $\mathbf{F}_k$ is chosen to minimize the guaranteed data MSE $\tilde{\varepsilon}_k$ to create a larger feasible space for the next round precoder design. Therefore, $\mathbf{F}_k$ is updated using the following problem

$$\begin{aligned}&\min_{\mathbf{F}_k,\beta_k,\tilde{\mathbf{Z}}_k,\tilde{\varepsilon}_k} \tilde{\varepsilon}_k\\ &\text{s.t. } \operatorname{Tr}(\tilde{\mathbf{Z}}_k\tilde{\boldsymbol{\Sigma}}_k) \le p_k\cdot\beta_k, \quad \beta_k > 0, \quad \tilde{\mathbf{Z}}_k \succeq 0\\ &\quad \begin{bmatrix}\tilde{\mathbf{Z}}_k+\operatorname{diag}([\mathbf{0},\tilde{\varepsilon}_k-\operatorname{Tr}(\mathbf{R}_k\mathbf{F}_k^H\mathbf{F}_k)-\beta_k]) & \mathbf{Q}_k^H\\ \mathbf{Q}_k & \mathbf{I}_{LL_k}\end{bmatrix}\succeq 0.\end{aligned} \tag{43}$$

In contrast to the problem (16), we cannot get a closed-form solution from (43). Furthermore, owing to the nonlinear term $\operatorname{Tr}(\mathbf{R}_k\mathbf{F}_k^H\mathbf{F}_k)$ in the positive semidefinite constraint, this subproblem seems to be a nonconvex problem on $\mathbf{F}_k$. But since $\operatorname{Tr}(\mathbf{R}_k\mathbf{F}_k^H\mathbf{F}_k)$ is a scalar and lies on the diagonal of the matrix $\begin{bmatrix}\tilde{\mathbf{Z}}_k+\operatorname{diag}([\mathbf{0},\tilde{\varepsilon}_k-\operatorname{Tr}(\mathbf{R}_k\mathbf{F}_k^H\mathbf{F}_k)-\beta_k]) & \mathbf{Q}_k^H\\ \mathbf{Q}_k & \mathbf{I}_{LL_k}\end{bmatrix}$, by introducing a slack variable t_k, the problem (43) is transformed into

$$\begin{aligned}&\min_{\substack{\mathbf{F}_k,t_k\\ \beta_k,\tilde{\mathbf{Z}}_k,\tilde{\varepsilon}_k}} \tilde{\varepsilon}_k\\ &\text{s.t. } \operatorname{Tr}(\tilde{\mathbf{Z}}_k\tilde{\boldsymbol{\Sigma}}_k) \le p_k\cdot\beta_k, \quad \beta_k > 0, \quad \tilde{\mathbf{Z}}_k \succeq 0\\ &\quad \begin{bmatrix}\tilde{\mathbf{Z}}_k+\operatorname{diag}([\mathbf{0},\tilde{\varepsilon}_k-t_k-\beta_k]) & \mathbf{Q}_k^H\\ \mathbf{Q}_k & \mathbf{I}_{LL_k}\end{bmatrix}\succeq 0\\ &\quad \operatorname{Tr}(\mathbf{R}_k\mathbf{F}_k^H\mathbf{F}_k) \le t_k,\end{aligned} \tag{44}$$

which is a convex problem as $\mathbf{R}_k \succeq 0$. The problem (44) is equivalent with (43), because the last inequality constraint in (44) becomes active at the optimal solution. This can be proved by contradiction. If the last constraint is not active, then $\tilde{\varepsilon}_k$ and t_k can always be reduced by the same amount, which does not change the positive semidefinite constraint, until the last inequality becomes active.

The iterative algorithm for the duality method can also be represented as in Table I, except that the two subproblems (15) and (19) are replaced by (42) and (44), respectively. The proof for its convergence is similar to that in [11], and is omitted in this paper.

Again for initialization, the conventional identity or randomly generated equalizer may not be feasible. Below, we introduce a feasibility enhancement initialization analogous to that used in the Markov based method. Let $\mathbf{F}_k \triangleq \frac{1}{a_k}\mathbf{F}_{ko}$ in ($\mathbb{P}1$), where $\mathbf{F}_{ko}$ is an initial chosen equalizer, and $a_k > 0$ is the amplitude factor. Multiplying both sides of the last K inequalities by a_k, ($\mathbb{P}1$) becomes

$$\begin{aligned}&\min_{\substack{\mathbf{G},a_1,\ldots,a_K\\ \check{\beta}_1,\ldots,\check{\beta}_K\\ \check{\mathbf{Z}}_1,\ldots,\check{\mathbf{Z}}_K}} \operatorname{Tr}(\mathbf{G}\mathbf{G}^H)\\ &\text{s.t. } \operatorname{Tr}(\check{\mathbf{Z}}_k\tilde{\boldsymbol{\Sigma}}_k) \le p_k\check{\beta}_k, \quad \check{\beta}_k > 0, \quad \check{\mathbf{Z}}_k \succeq 0, \quad a_k > 0, \ \forall k\\ &\quad \begin{bmatrix}\check{\mathbf{Z}}_k+\operatorname{diag}([\mathbf{0},a_k\varepsilon_k-\frac{1}{a_k}\operatorname{Tr}(\mathbf{R}_k\mathbf{F}_{ko}^H\mathbf{F}_{ko})-\check{\beta}_k]) & \check{\mathbf{Q}}_k^H\\ \check{\mathbf{Q}}_k & a_k\mathbf{I}\end{bmatrix}\succeq 0, \forall k,\end{aligned} \tag{45}$$

where $\check{\mathbf{Q}}_k \triangleq [\mathbf{G}^T\otimes\mathbf{F}_{ko} \ \operatorname{vec}(\mathbf{F}_{ko}\hat{\mathbf{H}}_k\mathbf{G}-a_k\mathbf{D}_k)]$, $\check{\beta}_k \triangleq a_k\beta_k$ and $\check{\mathbf{Z}}_k \triangleq a_k\tilde{\mathbf{Z}}_k$. Further introducing a slack variable c_k, the problem (45) is equivalent to

$$\begin{aligned}&\min_{\substack{\mathbf{G},a_1,\ldots,a_K,c_1,\ldots,c_K\\ \check{\beta}_1,\ldots,\check{\beta}_K\\ \check{\mathbf{Z}}_1,\ldots,\check{\mathbf{Z}}_K}} \operatorname{Tr}(\mathbf{G}\mathbf{G}^H)\\ &\text{s.t. } \operatorname{Tr}(\check{\mathbf{Z}}_k\tilde{\boldsymbol{\Sigma}}_k) \le p_k\check{\beta}_k, \quad \check{\beta}_k > 0, \quad \check{\mathbf{Z}}_k \succeq 0, \quad \forall k\\ &\quad \begin{bmatrix}\check{\mathbf{Z}}_k+\operatorname{diag}([\mathbf{0},a_k\varepsilon_k-c_k\operatorname{Tr}(\mathbf{R}_k\mathbf{F}_{ko}^H\mathbf{F}_{ko})-\check{\beta}_k]) & \check{\mathbf{Q}}_k^H\\ \check{\mathbf{Q}}_k & a_k\mathbf{I}\end{bmatrix}\succeq 0, \forall k\\ &\quad a_kc_k \ge 1, \quad a_k > 0, \quad \forall k,\end{aligned} \tag{46}$$

where the positive amplitude constraint $a_k > 0$ regularizes the set $\{(a_k, c_k)|a_k c_k \geq 1\}$ into a convex set, which can be represented as a LMI $\left[\begin{smallmatrix} a_k & 1 \\ 1 & c_k \end{smallmatrix}\right] \succeq 0$. Therefore, the initial transceiver pair $(\mathbf{G}, \frac{1}{a_1}\mathbf{F}_{1o}, \cdots, \frac{1}{a_K}\mathbf{F}_{Ko})$ can be efficiently obtained from the convex problem (46).

C. Convex Robust MU-MISO Transceiver Design

In MU-MIMO systems, the proposed iterative algorithms converge, but in general it is not known whether the converged solution is the global optimal or not [1]- [7]. However, for the special case of MU-MISO systems, we will show that the robust probabilistic QoS constrained transceiver design problems are convex, thus global optimal solution is guaranteed.

1) Markov's Inequality Based Method: Since each user only equipped with one antenna in the MU-MISO system, $L_k = M_k = 1, L = K$ and the k^{th} user's equalizer $\mathbf{F}_k$ in (14) should be replaced by $1/a_k{\cdot}e^{j\theta_k}$. After multiplying a_k^2 on both sides of all inequalities in problem (14), it becomes

$$\begin{aligned} \min_{\mathbf{G}, a_1,\ldots,a_K, \theta_1,\ldots,\theta_K} \quad & \mathrm{Tr}(\mathbf{G}\mathbf{G}^H) \\ \text{s.t.} \quad & \|\mathrm{vec}\left(\mathbf{G}^T \boldsymbol{\Sigma}_k^{\frac{1}{2}} e^{j\theta_k}\right)\|_2^2 + \|\hat{\mathbf{H}}_k \mathbf{G} e^{j\theta_k} - a_k \mathbf{D}_k\|_2^2 \\ & + \|\mathbf{R}_k^{1/2} e^{j\theta_k}\|_2^2 \leq a_k^2 p_k \varepsilon_k, \quad \forall k, \end{aligned} \tag{47}$$

where $\mathbf{D}_k = [0, \cdots, 1, \cdots, 0]$, with 1 appears at the k^{th} position, is a $1 \times K$ basis vector in MU-MISO system. Now, define $\bar{\mathbf{G}} \triangleq \mathbf{G} \circ [e^{j\theta_1}\mathbf{1}, \cdots, e^{j\theta_K}\mathbf{1}]$, we get $\mathrm{Tr}(\mathbf{G}\mathbf{G}^H) = \|\mathbf{G}\|_F^2 = \mathrm{Tr}(\bar{\mathbf{G}}\bar{\mathbf{G}}^H)$ and $\|\mathrm{vec}\left(\mathbf{G}^T \boldsymbol{\Sigma}_k^{\frac{1}{2}} e^{j\theta_k}\right)\|_2^2 = \|\mathbf{G}^T \boldsymbol{\Sigma}_k^{\frac{1}{2}}\|_F^2 = \|\bar{\mathbf{G}}^T \boldsymbol{\Sigma}_k^{\frac{1}{2}}\|_F^2$. It is also observed that the k^{th} elements of vectors $[\hat{\mathbf{H}}_k \mathbf{G} e^{j\theta_k} - a_k \mathbf{D}_k]$ and $[\hat{\mathbf{H}}_k \bar{\mathbf{G}} - a_k \mathbf{D}_k]$ are the same, and other elements in them only differ by a phase rotation because the corresponding elements in $\mathbf{D}_k$ are zeros. Therefore, (47) is equivalent to the following SOCP problem,

$$\begin{aligned} \min_{\bar{\mathbf{G}}, a_1,\ldots,a_K} \quad & \mathrm{Tr}(\bar{\mathbf{G}}\bar{\mathbf{G}}^H) \\ \text{s.t.} \quad & \|[\mathrm{vec}\left(\bar{\mathbf{G}}^T \boldsymbol{\Sigma}_k^{\frac{1}{2}}\right)^T, \hat{\mathbf{H}}_k \bar{\mathbf{G}} - a_k \mathbf{D}_k, \mathbf{R}_k^{1/2}]\|_2 \leq a_k \sqrt{p_k \varepsilon_k}, \forall k, \end{aligned} \tag{48}$$

and the global optimal transceiver $(\bar{\mathbf{G}}, \frac{1}{a_1}, \cdots, \frac{1}{a_K})$ for the Markov's inequality based MU-MISO system can be obtained.

2) Duality Based Method: Similar to the above Markov method, letting $\mathbf{F}_k \triangleq \frac{1}{a_k} e^{j\theta_k}$ and introducing $\bar{\mathbf{G}}$ to (30), we get $\|\frac{1}{a_k} e^{j\theta_k}(\hat{\mathbf{H}}_k + \boldsymbol{\Delta}_k)\mathbf{G} - \mathbf{D}_k\|_F^2 = \frac{1}{a_k^2} \mathbf{u}_k^H \bar{\mathbf{Q}}_k^H \bar{\mathbf{Q}}_k \mathbf{u}_k$, where $\bar{\mathbf{Q}}_k \triangleq [\bar{\mathbf{G}}^T \ \mathrm{vec}(\hat{\mathbf{H}}_k \bar{\mathbf{G}} - a_k \mathbf{D}_k)]$. Following the same derivations in the MU-MIMO case, the MU-MISO transceiver design problem corresponds to $(\mathbb{P}1)$ is

$$\begin{aligned} \min_{\substack{\bar{\mathbf{G}}, a_1,\ldots,a_K, c_1,\ldots,c_K \\ \check{\beta}_1,\ldots,\check{\beta}_K \\ \check{\mathbf{Z}}_1,\ldots,\check{\mathbf{Z}}_K}} \quad & \mathrm{Tr}(\bar{\mathbf{G}}\bar{\mathbf{G}}^H) \\ \text{s.t.} \quad & \mathrm{Tr}(\check{\mathbf{Z}}_k \tilde{\boldsymbol{\Sigma}}_k) \leq p_k \check{\beta}_k, \quad \check{\beta}_k > 0, \quad \check{\mathbf{Z}}_k \succeq 0, \quad \forall k \\ & \begin{bmatrix} \check{\mathbf{Z}}_k + \mathrm{diag}([\mathbf{0}, a_k \varepsilon_k - c_k \mathrm{Tr}(\mathbf{R}_k) - \check{\beta}_k]) & \bar{\mathbf{Q}}_k^H \\ \bar{\mathbf{Q}}_k & a_k \mathbf{I}_L \end{bmatrix} \succeq 0, \forall k \\ & \begin{bmatrix} a_k & 1 \\ 1 & c_k \end{bmatrix} \succeq 0, \quad \forall k. \end{aligned} \tag{49}$$

Obviously, (49) is convex and is similar to (46).

D. Implementation Consideration and Complexity Analysis

In practical systems, the proposed Markov and duality method can be used when the estimated downlink channels and the statistical information of the estimation error are obtained at the BS. In particular, since the downlink channel is the same as the uplink channel in the time-division duplex (TDD) system, the BS can estimate the uplink channels and then infer the downlink channels. Once these information are obtained, the BS can calculate the optimal precoder $\mathbf{G}$ and all equalizers, and send the equalizer $\mathbf{F}_k$ to the corresponding user.

For computational complexity, since the first subproblem of the Markov method is a SOCP problem, the complexity in each iteration is $O\left(n^2 \cdot \sum_{k=1}^{K} d_k\right)$ [32], where $n = NL$ is the number of unknown variables and $d_k = NLM_kL_k + L_k^2 + M_k^2$ is the vector length in the k^{th} second order cone constraint. The complexity of the second subproblem is $O(M_k^3)$ owing to the dominant matrix inversion operation.

For the computational complexity of the duality method, since all subproblems are SDP problems, the complexity in each iteration is $O\left(n^2 \cdot \sum_{k=1}^{K} d_k^2\right)$ [32] where n is the number of unknown variables and d_k is the matrix dimension in the k^{th} positive semidefinite constraint. For the first subproblem, $n = NL + \sum_{k=1}^{K} \frac{(NM_k+1)^2}{2} + K$ and $d_k = NM_k + 1 + LL_k$. Although the matrix dimension of the second subproblem is the same as that of the first subproblem, the variable number is reduced to $n = L_kM_k + \frac{(NM_k+1)^2}{2} + 3$ owing to the fact that all equalizers can be computed in parallel. Therefore, the computational complexity of the Markov method is lower than that of the duality method.

IV. Simulation Results and Discussions

In this section, the performance of the two proposed algorithms is illustrated by simulations. Unlike most existing works, the channel estimation is implemented first, and the estimated channel and statistical information of the estimation errors are taken into consideration in the transceiver design. The MIMO channel in the simulations has complex Gaussian entries with zero mean and unit variance. The channel correlation matrices are taken as the exponential model, i.e., $[\mathbf{R}_t]_{ij} = \rho_t^{|i-j|}, [\mathbf{R}_{rk}]_{ij} = \rho_r^{|i-j|}$, with the correlation coefficient $\rho_t = 0.2$ and $\rho_r = 0.5$. We assume the system is interference dominated, and the received interference at every antenna is i.i.d. with zero mean and variance 0.01. Since interference in future networks (e.g., Cognitive Radio [33] and Heterogeneous Network [34]) is often modeled by heavy-tailed distributions, it is assumed in this section that the distribution of interference is Laplace unless stated otherwise. The BS is equipped with four antennas and there are two active users, and each is equipped with two antennas, i.e., $N = 4, K = 2, M_k = 2$. Spatial multiplexing is implemented for the two users, i.e., $L_k = 2$. At the channel estimation stage, orthogonal Walsh sequences with length $L_t = 2N$ are employed in different transmit antennas. At the data detection stage, the required outage probability is $p_1 = p_2 = 0.15$, the QoS target for the

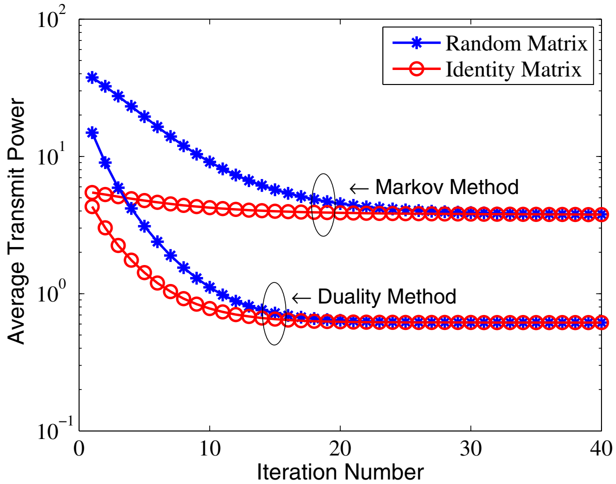

Fig. 1. The convergence performance of the proposed iterative algorithms.

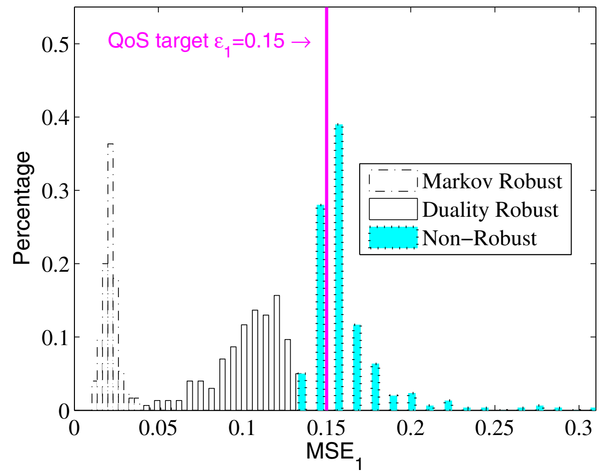

Fig. 2. The residual MSE of the first user with QoS requirement $p_1 = p_2 = 0.15, \varepsilon_1 = 0.15, \varepsilon_2 = 0.3$.

second user is $\varepsilon_2 = 0.3$ while that for the first user is varied between 0.05 and 0.3, and the threshold to terminate the iteration is $\epsilon = 5 \times 10^{-4}$. All simulation results are averaged over 10^3 randomly chosen channel realizations. For fair comparison in the transmit power performance, the randomly chosen channel realizations are feasible for all algorithms under consideration.

In Fig. 1, the convergence performance of the proposed algorithms with $\mathbf{F}_{ko}$ being identity matrix and random matrix as initialization is compared. The QoS target for the first user is $\varepsilon_1 = 0.3$. It can be seen that for any particular proposed method, both initializations lead to the same transmit power level after convergence, but the method with identity matrix initialization converges slightly faster than the random one. Therefore, $\mathbf{F}_{ko} = \mathbf{I}$ is used for equalizer initialization in the subsequent simulations. Besides the convergence speed, from Fig. 1, it is further noted that the duality method achieves a lower transmission power level than the Markov method, as the Markov method involves a loose upper bound, and therefore is more conservative.

In Fig. 2, the histograms of the recovered first user's data MSEs of the two proposed methods and the non-robust method using the basic formulation after (4) [7] is compared when the QoS target for the first user is $\varepsilon_1 = 0.15$. The non-robust method takes the estimated channel as the true channel, and the channel estimation error is ignored[1]. It is observed that the residual MSEs of the two proposed robust methods are smaller than the prefixed QoS target, while that of the non-robust method exceeds the QoS target for about 70% of the channel realizations. Furthermore, it is also noted that the residual MSEs of the duality method are very close to the QoS target, while that of the Markov method are much smaller than the target. This shows the conservative nature of the Markov's inequality based design.

The required transmit power of the two proposed algorithms to guarantee QoS performance is illustrated in Fig. 3. It is observed that over a wide range of QoS target for the first

[1]Note that for the non-robust method, if the channel is perfectly known, the QoS performance will always coincide with the "QoS target ε_1" in Fig. 2.

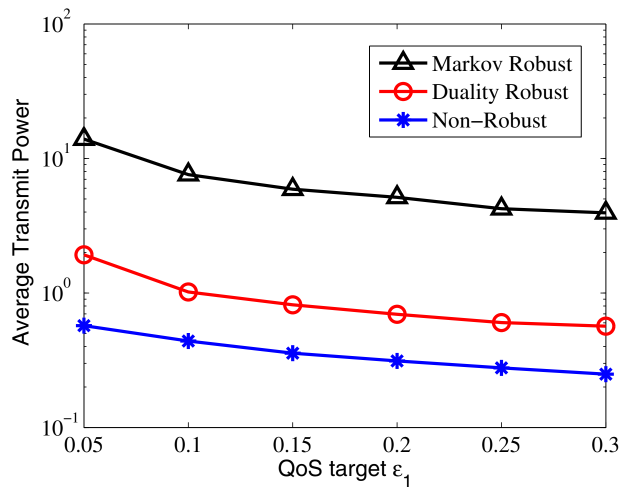

Fig. 3. The transmission power with different QoS requirement.

user, the duality method consistently requires a significantly lower transmission power than that of the Markov method, saving more than 8 dB in transmission power. This is owing to the loose Markov's inequality which makes the approximated QoS requirements in (14) over conservative. The additional required transmit power of the duality method compared to the non-robust method is about 3 dB, which is the cost of the guaranteed QoS performance against the channel uncertainty. At Fig. 4, the required transmit power as a function of outage probability p_1 (with $\varepsilon_1 = 0.15$) is shown. It can be seen that similar conclusions can be drawn as in Fig. 3.

From Figs. 2, 3 and 4, it can be concluded that the proposed duality method guarantees the QoS requirement, while maintaining a low transmission power. While the Markov method also guarantee the QoS constraints, the duality method achieves a good balance between the QoS requirement and the power saving requirement. Notice that although the non-robust method seems to indicate a low transmit power, it does not represent a valid solution because the QoS constraints are not satisfied as shown in Fig. 2. Similar conclusions can be drawn

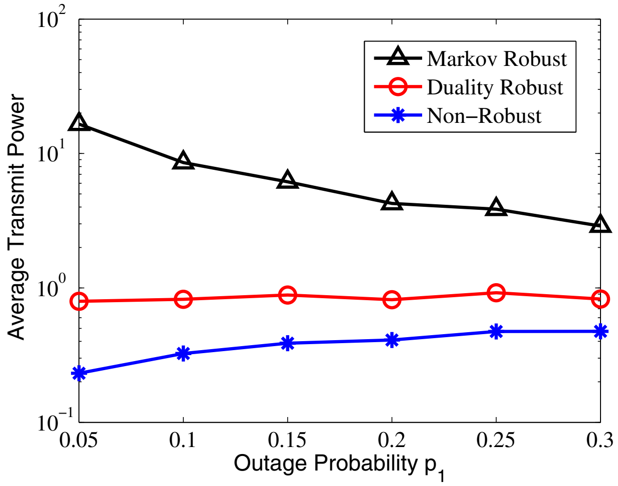

Fig. 4. The transmission power with different outage probability requirement.

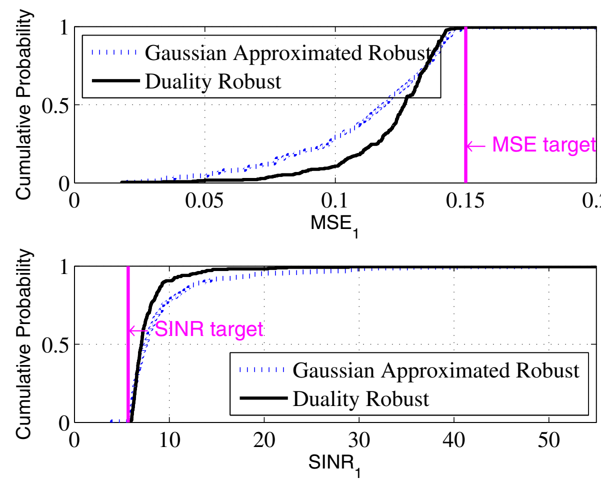

Fig. 5. The CDF of residual MSE and SINR for the proposed duality method and Gaussian approximated robust method.

in other simulation settings with different interference distributions, e.g., Gaussian and uniform distributed interference. The corresponding figures are not repeated in this paper.

One may argue that in order to have fair comparison, the transmit power of the non-robust method should be increased to the same level as the robust method. However, this is possible only after we executed the proposed robust transceiver design algorithm (i.e., after we obtained the required transmit power to guarantee the QoS requirements). Without the proposed algorithms, the only way to guarantee the QoS requirements for the non-robust method is to perform extensive simulations on the MSE with increasing transmit power level, until the QoS requirements are satisfied. This would takes prohibitively significant amount of time, as the minimum required power level would also depend on numerous other factors, such as pilot pattern, pilot transmit power, total number of users, antenna number in BS and different users, number of data stream of each user, signal-to-noise ratio, and most importantly the interference plus noise distribution. Simulations have to be re-run if any one of the above parameters is changed. Therefore, this approach is almost impossible in practice.

The performance comparison between the Gaussian approximated robust method in [17] and the duality method is illustrated in Fig. 5. In the simulation, there are two users, both equipped with single antenna, and one BS equipped with two antennas. This corresponds to the MU-MISO case. The MSE target is 0.15 for both users (the corresponding SINR target can be calculated to be 5.66 [9]), and the outage probability is 0.15. As shown in Fig. 5, the residual MSE and SINR of the method in [17] is more spread out (conservative) than the proposed duality method, the reason is that the Bernstein-type inequality in [17] is a conservative approximation for linear and quadratic formulated problems with probabilistic constraint over a set of ambiguous uncertainties [35, p.91, p.118], while the proposed duality method is a tight solution for the specific problem. Also note that the method in [17] can only be applied to the MU-MISO system, while our proposed methods are applicable to the more general MU-MIMO system.

In Fig. 6, the cumulative distribution function (CDF) of the residual first user's data MSE for the proposed duality method and the bounded robust method in [11] are compared. The interference distribution is assumed to be Gaussian, but not known to the design procedure. For the bounded method, since the interference distribution is unknown, the bound is heuristically selected as $3b$ and $0.3b$, where $b = \|\mathbf{\Sigma}_k^{\frac{1}{2}}\|_F$. It can be seen that the QoS performance of the bounded method with bound $3b$ is too conservative, while the QoS requirement is not satisfied for the bound $0.3b$. Therefore, without the interference distribution information, there is no systematic way to choose an accurate bound to realize the probabilistic QoS target. On the other hand, if we assume the interference distribution is known to be Gaussian, the bound can be set as $1.44b$ to make this bound to cover 85% of the Gaussian channel estimation error. Only under this strong assumption, the bounded method has a comparable QoS performance to the proposed duality method, which assumes no knowledge on the interference distribution. However, even with the accurate bound $1.44b$, the distribution of the MSEs of the bounded method is more spread out than that of the duality method. This is because the bounded robust method only exploited the deterministic bound information of channel uncertainty, while the duality method fully utilized the first and second-order moment information. It can be concluded that the bounded robust method in [11] is not a practical way to realize the probabilistic QoS constrained transceiver design without distribution information.

In practical wireless systems, the transmission power is limited. Therefore, besides the feasibility problem at the initialization, the problem is also considered as infeasible if the required transmission power is larger than the peak power, resulting in two factors that affect the feasibility of the whole transceiver design problem. With the QoS target for the first user $\varepsilon_1 = 0.3$, the impact of the peak power constraint on the feasible rate of different methods is illustrated in Fig. 7. On the left hand side of Fig. 7, the rapidly increasing feasible rate with increasing peak power reveals that the peak power is the main limiting factor. On the other hand, on the right

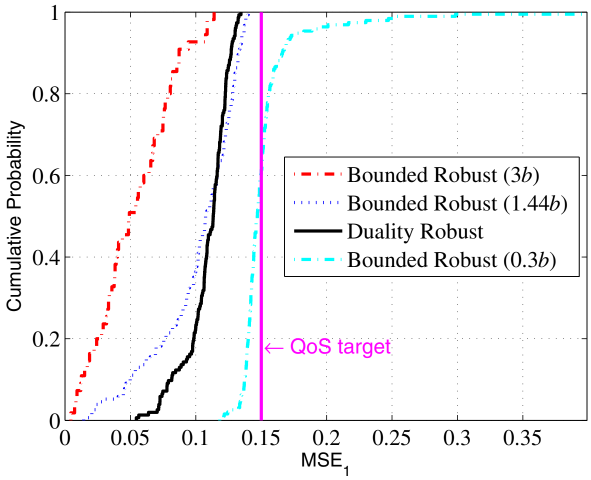

Fig. 6. The CDF of residual MSE for the proposed duality method and the bounded robust method ($p_1 = p_2 = 0.15, \varepsilon_1 = 0.15, \varepsilon_2 = 0.3, b = \|\mathbf{\Sigma}_k^{\frac{1}{2}}\|_F$).

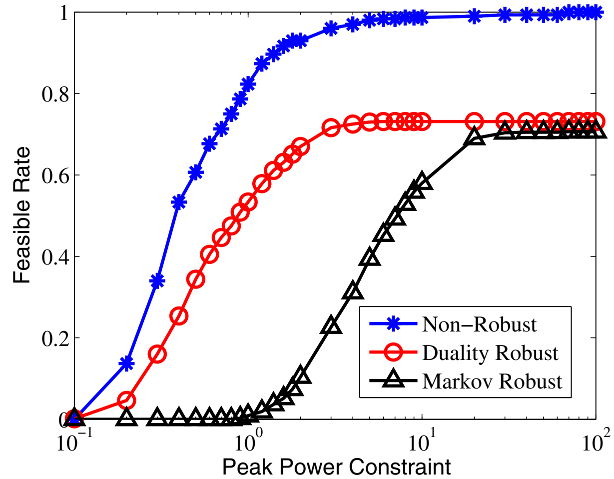

Fig. 7. The feasible rate of different methods with different peak power constraints.

hand side of Fig. 7, the flat feasible rate with increasing peak power reveals that the peak power is not the limiting factor, but the feasibility rate of initialization would become the limiting factor. For practical peak power around 1 to 2 (corresponds to SINR 20 to 23 dB), we observe that the feasible rate of the proposed duality method is close to that of the non-robust method, while that of the Markov method is close to zero owing to its high transmit power requirement.

V. Conclusions

In this paper, probabilistic QoS constrained robust downlink MU-MIMO transceiver design was investigated. The objective of the proposed design is to minimize the transmit power, while still guarantees a probabilistic QoS requirement under arbitrarily distributed channel estimation error. Markov's inequality based method and a novel duality method were proposed to solve the problem. The convergence of both iterative algorithms and the tightness of the duality method were guaranteed, and the convexity of robust transceiver designs for the MU-MISO scenario was proved. Simulation results showed that the QoS requirement is guaranteed for both proposed methods. For the minimized transmit power, the duality method showed superior performance than the Markov method, due to its tight reformulation. Furthermore, compared with Gaussian approximated probabilistic robust method and bounded robust method, the duality method has less conservative QoS performance.

Appendix A

Replacing the second constraint of problem (29) with its equivalent form (37), (29) becomes

$$\begin{aligned}&\min_{\mathbf{Z}_k,\lambda_k} \mathrm{Tr}(\mathbf{Z}_k\tilde{\mathbf{\Sigma}}_k)\\ &\text{s.t. } \mathbf{Z}_k \succeq 0, \lambda_k \geq 0\\ &\mathbf{Z}_k - \mathrm{diag}([\mathbf{0},1]) - \lambda_k(\mathbf{Q}_k^H\mathbf{Q}_k - \mathrm{diag}([\mathbf{0},\varepsilon_k - \mathrm{Tr}(\mathbf{F}_k\mathbf{R}_k\mathbf{F}_k^H)])) \succeq 0.\end{aligned} \tag{50}$$

If $\lambda_k = 0$, then the constraints of (50) reduces to $\mathbf{Z}_k - \mathrm{diag}([\mathbf{0},1]) \succeq 0$. Since $\tilde{\mathbf{\Sigma}}_k \succeq 0$, the optimal $\mathbf{Z}_k$ occurs on the boundary of the constraint and the optimal value of problem (50) becomes $\mathrm{Tr}(\mathrm{diag}([\mathbf{0},1])\tilde{\mathbf{\Sigma}}_k) = 1$. Putting this optimal value into ($\mathbb{P}0$), the resulted constraint in ($\mathbb{P}0$) becomes $1 \leq p_k$, which is infeasible because the outage probability has to be smaller than one.

Appendix B
Proof of the Proposition 1

Owing to the strong duality between the supremum of the outage probability in ($\mathbb{P}0$) and its dual problem (41), the problem ($\mathbb{P}0$) is represented as

$$\begin{aligned}&\min_{\mathbf{G},\mathbf{F}_1,\ldots,\mathbf{F}_K} \mathrm{Tr}(\mathbf{G}\mathbf{G}^H)\\ &\text{s.t.} \min_{\substack{\beta_k>0\\ \tilde{\mathbf{Z}}_k\succeq 0\\ \mathbf{B}_k(\mathbf{G},\mathbf{F}_k,\beta_k,\tilde{\mathbf{Z}}_k)\succeq 0}} \mathrm{Tr}(\tilde{\mathbf{Z}}_k\tilde{\mathbf{\Sigma}}_k)/\beta_k \leq p_k, \quad \forall k.\end{aligned} \tag{51}$$

where $\mathbf{B}_k(\mathbf{G},\mathbf{F}_k,\beta_k,\tilde{\mathbf{Z}}_k)$ represents the matrix in the last constraint of problem (41). Let $(\mathbf{G}^0,\mathbf{F}_1^0,\cdots,\mathbf{F}_K^0)$ be any transceiver, then the constraints in problem (51) is

$$\begin{aligned}&\beta_k > 0,\ \tilde{\mathbf{Z}}_k \succeq 0,\ \mathbf{B}_k(\mathbf{G}^0,\mathbf{F}_k^0,\beta_k,\tilde{\mathbf{Z}}_k)\succeq 0,\\ &\min \mathrm{Tr}(\tilde{\mathbf{Z}}_k\tilde{\mathbf{\Sigma}}_k)/\beta_k \leq p_k, \forall k.\end{aligned} \tag{52}$$

If there exist $\beta_k, \tilde{\mathbf{Z}}_k$ such that

$$\begin{aligned}&\beta_k > 0,\quad \tilde{\mathbf{Z}}_k \succeq 0,\quad \mathbf{B}_k(\mathbf{G}^0,\mathbf{F}_k^0,\beta_k,\tilde{\mathbf{Z}}_k)\succeq 0,\\ &\mathrm{Tr}(\tilde{\mathbf{Z}}_k\tilde{\mathbf{\Sigma}}_k)/\beta_k \leq p_k, \forall k\end{aligned} \tag{53}$$

is satisfied, then (52) is also satisfied. On the other hand, if (52) is satisfied, then there exist $\beta_k, \tilde{\mathbf{Z}}_k$ such that (53) is satisfied. Therefore, the constraint (52) and (53) are equivalent, and the proposition is proved.

References

[1] S. Serbetli and A. Yener, "Transceiver optimization for multiuser MIMO systems," *IEEE Trans. Signal Process.*, vol. 52, no. 1, pp. 214–226, 2004.

[2] M. Schubert, S. Shi, E. Jorswieck, and H. Boche, "Downlink sum-MSE transceiver optimization for linear multi-user MIMO systems," in *Proc. 2005 Asilomar CSSC*.

[3] J. Zhang, Y. Wu, S. Zhou, and J. Wang, "Joint linear transmitter and receiver design for the downlink of multiuser MIMO systems," *IEEE Commun. Lett.*, vol. 9, no. 11, pp. 991–993, 2005.

[4] S. Abraham, D. C. Popescu, and O. A. Dobre, "Joint beamforming and power control in downlink multiuser MIMO systems," *Wiley J. Commun. Comput.*, Mar. 2013.

[5] A. Mezghani, M. Joham, R. Hunger, and W. Utschick, "Transceiver design for mutli-user MIMO systems," in *Proc. 2006 Int. ITG/WSA Conference*.

[6] S. Shi, M. Schubert, and H. Boche, "Downlink MMSE transceiver optimization for multiuser MIMO systems: duality and sum-MSE minimization," *IEEE Trans. Signal Process.*, vol. 13, no. 11, pp. 5436–5446, 2007.

[7] S. Shi, M. Schubert, and H. Boche, "Downlink MMSE transceiver optimization for multiuser MIMO systems: MMSE balancing," *IEEE Trans. Signal Process.*, vol. 56, no. 8, pp. 3702–3712, 2008.

[8] L. G. Ordóñez, D. P. Palomar, A. Pagès-Zamora, and J. R. Fonollosa, "Minimum BER linear MIMO transceivers with adaptive number of substreams," *IEEE Trans. Signal Process.*, vol. 57, no. 6, pp. 2336–2353, 2009.

[9] D. P. Palomar, J. M. Cioffi, and M. Á. 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, 2003.

[10] N. Vučić and H. Boche, "Robust QoS-constrained optimization of downlink multiuser MISO systems," *IEEE Trans. Signal Process.*, vol. 57, no. 2, pp. 714–725, 2009.

[11] N. Vučić, H. Boche, and S. Shi, "Robust transceiver optimization in downlink multiuser MIMO systems," *IEEE Trans. Signal Process.*, vol. 57, no. 9, pp. 3576–3587, 2009.

[12] M. Payaró, A. Pascual-Iserte, and M. Á. Lagunas, "Robust power allocation designs for multiuser and multiantenna downlink communication systems through convex optimization," *IEEE J. Sel. Areas Commun.*, vol. 25 no. 7, pp. 1390–1401, 2007.

[13] S. M. Kay, *Fundamentals of Statistical Signal Processing, Volume I: Estimation Theory*. Prentice Hall, 1993.

[14] G. Zheng, K. K. Wong, and T. S. Ng, "Energy-efficient multiuser SIMO: achieving probabilistic robustness with Gaussian channel uncertainty," *IEEE Trans. Commun.*, vol. 57, no. 6, pp. 1866–1878, 2009.

[15] E. D. Anese, S.-J. Kim, G. B. Giannakis, and S. Pupolin, "Power control for cognitive radio networks under channel uncertainty," *IEEE Trans. Wireless Commun.*, vol. 10, no. 10, pp. 3541–3551, 2011.

[16] P.-J. Chung, H. Du, and J. Gondzio, "A probabilistic constraint approach for robust transmit beamforming with imperfect channel information," *IEEE Trans. Signal Process.*, vol. 59, no. 6, pp. 2773–2782, 2011.

[17] K.-Y. Wang, A. M.-C. So, T.-H. Chang, W.-K. Ma, and C.-Y. Chi, "Outage constrained robust transmit optimization for multiuser MISO downlinks: tractable approximations by conic optimization," submitted to *IEEE Trans. Signal Process.*, Aug. 2011.

[18] Y. Liu, T. F. Wong, and W. W. Hager, "Training signal design for estimation of correlated MIMO channels with colored interference," *IEEE Trans. Signal Process.*, vol. 55, no. 4, pp. 1486–1497, 2007.

[19] N. Jindal, "MIMO broadcast channels with finite-rate feedback," *IEEE Trans. Inf. Theory*, vol. 52, pp. 5045–5060, 2006.

[20] C. K. Au-Yeung and D. J. Love, "On the performance of random vector quantization limited feedback beamforming in a MISO system," *IEEE Trans. Wireless Commun.*, vol. 6, no. 2, pp. 458–462, 2007.

[21] E. Björnson, D. Hammarwall, and B. Ottersten, "Exploiting quantized channel norm feedback through conditional statistics in arbitrarily correlated MIMO system," *IEEE Trans. Signal Process.*, no. 57, pp. 4027–4041, 2009.

[22] H. Stark and J. W. Woods, *Probability and Random Processes with Applications to Signal Processing*, 3rd ed. Prentice Hall, 2001.

[23] M. B. Shenouda and T. N. Davidson, "Tomlinson-Harashima precoding for broadcast channels with uncertainty," *IEEE J. Sel. Areas Commun.*, vol. 25, no. 7, pp. 1380–1389, 2007.

[24] X. Zhang, D. P. Palomar, and B. Ottersten, "Statistically robust design of linear MIMO transceivers," *IEEE Trans. Signal Process.*, vol. 56, no. 8, pp. 3678–3689, 2008.

[25] E. Matskani, N. D. Sidiropoulos, Z.-Q. Luo, and L. Tassiulas, "Convex approximation techniques for joint multiuser downlink beamforming and admission control," *IEEE Trans. Wireless Commun.*, vol. 7, no. 7, pp. 2682–2693, 2008.

[26] S. Umarov, C. Tsallis, and S. Steinberg, "On a q-central limit theorem consistent with nonextensive statistical mechanics," *Milan J. Math.*, vol. 76, pp. 307–328, 2008.

[27] J. Smith, "Generalized Chebychev inequalities: theory and applications in decision analysis," *Oper. Res.*, vol. 43, pp. 807–825, 1995.

[28] T. K. Moon and W. C. Stirling, *Mathematical Methods and Algorithms for Signal Processing*. Prentlce Hall, 2000.

[29] S. Boyd, L. El Ghaoui, E. Feron, and V. Balakrishnan, *Linear Matrix Inequalities in System and Control Theory*. SIAM Studies in Applied Mathematics, 1994.

[30] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge University Press, 1990.

[31] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.

[32] M. Lobo, L. Vandenberghe, S. Boyd, and H. Lebret, "Applications of second-order cone programming," *Linear Algebra and its Applications*, no. 284, pp. 193–228, 1998.

[33] A. Rabbachin, T. Q. S. Quek, H. Shin, and M. Z. Win, "Cognitive network interference," *IEEE J. Sel. Areas Commun.*, vol. 29, no. 2, pp. 480–493, 2011.

[34] R. W. Heath Jr., M. Kountouris, and T. Bai, "Modeling heterogeneous network interference using Poisson point processes," submitted to *IEEE Trans. Signal Process.*, July 2012.

[35] A. Ben-Tal, L. El Ghaoui, and A. Nemirovski, *Robust Optimization*. Princeton Series in Applied Mathematics, Princeton University Press, 2009.

Xin He received the B.Eng. degree in Electronic Information from SiChuan University, Chengdu, China, in 2007, the M.Eng. degree in Information and Communication from Beijing University of Posts and Telecommunications, in 2010, and the M.Phil. (EEE) degree in 2012 from the University of Hong Kong (HKU). He is currently working toward the Ph.D. degree in HKU. His research interests are optimization and statistics for probabilistic inference and beamforming.

Yik-Chung Wu received the B.Eng. (EEE) degree in 1998 and the M.Phil. degree in 2001 from the University of Hong Kong (HKU). He received the Croucher Foundation scholarship in 2002 to study Ph.D. degree at Texas A&M University, College Station, and graduated in 2005. From August 2005 to August 2006, he was with the Thomson Corporate Research, Princeton, NJ, as a Member of Technical Staff. Since September 2006, he has been with HKU, currently as an Associate Professor. He was a visiting scholar at Princeton University, in summer 2011. His research interests are in general area of signal processing and communication systems, and in particular distributed signal processing and communications; optimization theories for communication systems; estimation and detection theories in transceiver designs; and smart grid. Dr. Wu served as an Editor for IEEE COMMUNICATIONS LETTERS, is currently an Editor for IEEE TRANSACTIONS ON COMMUNICATIONS and the *Journal of Communications and Networks*.