



<b>Title</b>	<b>Tight Probabilistic SINR Constrained Beamforming Under Channel Uncertainties</b>
<b>Author(s)</b>	<b>HE, X; Wu, YC</b>
<b>Citation</b>	<b>IEEE Transactions on Signal Processing, 2015, v. 63, p. 3490-3505</b>
<b>Issued Date</b>	<b>2015</b>
<b>URL</b>	<b><a href="http://hdl.handle.net/10722/214164">http://hdl.handle.net/10722/214164</a></b>
<b>Rights</b>	<b>Creative Commons: Attribution 3.0 Hong Kong License</b>

# Tight Probabilistic SINR Constrained Beamforming Under Channel Uncertainties

Xin He and Yik-Chung Wu

**Abstract**—In downlink multi-user beamforming, a single base station is serving a number of users simultaneously. However, energy intended for one user may leak to other unintended users, causing interference. With signal-to-interference-plus-noise ratio (SINR) being one of the most crucial quality metrics to users, beamforming design with SINR guarantee has always been an important research topic. However, when the channel state information is not accurate, the SINR requirements become probabilistic constraints, which unfortunately are not tractable analytically for general uncertainty distribution. Therefore, existing probabilistic beamforming methods focus on the relatively simple Gaussian and uniform channel uncertainties, and mainly rely on probability inequality based approximated solutions, resulting in conservative SINR outage realizations. In this paper, based on the local structure of the feasible set in the probabilistic beamforming problem, a systematic method is proposed to realize tight SINR outage control for a large class of channel uncertainty distributions. With channel estimation and quantization errors as examples, simulation results show that the SINR outage can be realized tightly, which results in reduced transmit power compared to the existing inequality based probabilistic beamformers.

**Index Terms**—Probabilistic SINR constrained beamforming, tight probabilistic control, channel uncertainty.

## I. INTRODUCTION

**D**UE to diverse nature of data (e.g., video call, VoIP, online game, instant message, etc) simultaneously transmitting through modern wireless systems, different quality of services (QoS) are needed from different users. Exact QoS control is a desirable property for future heterogeneous networks with dense small cell deployments [1], since over-satisfied QoS inevitably leads to interference leakage to unintended users. While bit error rate (BER) is undoubtedly one of the most important QoS criteria in a communication system, it is a highly non-linear function of the beamformer, and various approximations are needed in the beamformer optimization [2]. Therefore, commonly used surrogate QoS criteria in beamforming design include mean square error (MSE) of data, signal-to-interference-plus-noise ratio (SINR) and channel capacity [3], [4]. Among the above criteria, SINR is a compelling QoS criterion, due

to the direct relationship between SINR and BER through the Gaussian  $Q$ -function [4], while other criteria have only indirect connections to BER.

However, the ideal case of exact SINR control in a multiuser system is hindered by channel uncertainties [3], [8]. By modeling the channel uncertainties lie in a bounded region, SINR constrained robust beamforming and transceiver design are proposed to tackle the worse-case error in [5], [6], [9] and [10]. Unfortunately, the bounded robust optimization is generally conservative owing to its worst-case criterion [13]. On the other hand, probabilistic SINR constrained beamforming provides a soft SINR control if the probability density function (PDF) of the channel uncertainty is known. Previous probabilistic beamforming schemes mainly consider the Gaussian channel uncertainty, and only approximation solutions are available by using different probability inequalities, e.g., triangle inequality for the array beamforming [11], Vysochanskii-Petunin inequality for power allocation [12], [13], Bernstein-type inequality [14], [15] and Bernstein approximation [16] for probabilistic SINR constrained beamforming. However, owing to the restricted feasible set in those safe approximations, the SINR requirement of these designs are over-satisfied, which leads to unnecessarily high transmit powers. Although a bisection calibration method is proposed in [17] to mitigate the high transmit power problem under independent Gaussian and uniform channel uncertainties, the bisection range might not exist and thus it is not guaranteed to be implementable.

In this paper, a tight probabilistic SINR control is achieved in multiuser beamforming under a large class of bounded or unbounded channel uncertainties with known PDF. Facing the challenge of intractable probabilistic constraints, a successive method is proposed to reconstruct the feasible set. In particular, we first find a feasible subset based on the moment and support information of the channel uncertainty. Then, a joint feasible subsets refinement and sequential optimization is proposed to analyze the unexplored feasible subsets. In contrast to the oscillating convergence behavior of the bisection calibration [17], the proposed iterative method ensures the transmit power decrease monotonically and achieves tight outage control quickly. Simulation results under channel estimation and quantization errors show that the probabilistic SINR requirements are fulfilled tightly, which leads to improved performance on transmit power compared to existing approximation based probabilistic beamforming.

The rest of this paper is organized as follows. In Section II, the probabilistic beamforming problem is formulated and a systematic way of finding a feasible subset is introduced. Joint feasible subsets refinement and optimization is described in Section III.

Manuscript received September 16, 2014; revised January 26, 2015 and March 31, 2015; accepted April 10, 2015. Date of publication April 28, 2015; date of current version June 03, 2015. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Yongming Huang.

The authors are with the Department of Electrical and Electronic Engineering, The University of Hong Kong, Hong Kong, 60010, Hong Kong (e-mail: hexin@eee.hku.hk; ycwu@eee.hku.hk).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TSP.2015.2425806

The computation details of the iterative procedure is presented in Section IV. Simulation results are presented in Section V, and conclusions are drawn in Section VI.

*Notation:* In this paper,  $\mathbb{E}(\cdot)$ ,  $(\cdot)^T$ , and  $(\cdot)^H$  denote statistical expectation, transposition and Hermitian, respectively. In addition,  $\text{Tr}(\cdot)$  and  $\|\cdot\|_F$  refer to the trace and Frobenius norm of a matrix, respectively, while  $\|\cdot\|_2$  denotes the norm of a vector.  $\text{Re}\{\cdot\}$  and  $\text{Im}\{\cdot\}$  extract real and imaginary parts of the argument, respectively. Symbol  $\text{Diag}(\mathbf{x})$  denotes a diagonal matrix with vector  $\mathbf{x}$  on its diagonal, and  $\mathbf{I}_K$  is a  $K \times K$  identity matrix.

## II. PROBLEM FORMULATION AND FINDING A FEASIBLE SUBSET

The downlink multiuser beamforming system under consideration consists of one base station (BS) equipped with  $N$  transmit antennas and  $K$  single-antenna active users, where  $K \leq N$ . The  $k$ th user's channel and beamformer are represented as  $\mathbf{h}_k$  and  $\mathbf{w}_k$ , respectively, and the noise at the  $k$ th user is distributed as  $\mathcal{CN}(0, \delta_k^2)$ . The beamforming design aiming at minimizing transmit power at BS with guaranteed probabilistic SINR requirements can be formulated as

$$\begin{aligned} \min_{\mathbf{W}} \quad & \|\mathbf{W}\|_F^2 \\ \text{s.t.} \quad & \Pr\{\text{SINR}_k(\mathbf{W}, \mathbf{h}_k) \geq \gamma_k\} \geq 1 - p_k, \quad \forall k \in \mathcal{K}, \end{aligned} \quad (1)$$

where  $\text{SINR}_k(\mathbf{W}, \mathbf{h}_k) = \frac{\|\mathbf{h}_k^H \mathbf{w}_k\|_2^2}{\sum_{l=1, l \neq k}^K \|\mathbf{h}_k^H \mathbf{w}_l\|_2^2 + \delta_k^2}$ ,  $\mathbf{W} = [\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_K]$ ,  $\mathcal{K} = \{1, 2, \dots, K\}$  and outage probability  $p_k \in [0, 1]$ . To model channel uncertainty, it is noted that  $\mathbf{h}_k = \hat{\mathbf{h}}_k + \mathbf{x}_k$ , where  $\hat{\mathbf{h}}_k$  is the obtained channel information (e.g., through estimation or quantization), and  $\mathbf{x}_k$  is the channel uncertainty with continuous PDF  $f(\mathbf{x}_k)$ . It is assumed that the mean vector  $\mathbb{E}(\mathbf{x}_k) = \mathbf{0}$ <sup>1</sup>. Furthermore, the covariance matrix  $\Sigma_k \succeq 0$  and the support  $\mathcal{X}_k = \{\mathbf{x}_k | \mathbf{x}_k^H \mathbf{A}_k \mathbf{x}_k \leq r_k^2, \mathbf{A}_k^H = \mathbf{A}_k\}$  of  $f(\mathbf{x}_k)$  are assumed to be available. Notice that the quadratic support  $\mathcal{X}_k$  covers the unbounded support ( $\mathbf{A}_k = \mathbf{I}_N, r_k \rightarrow +\infty$ ) from many common distributions, e.g., Gaussian, Laplace and  $t$ -distribution, and ellipsoid bounded support ( $\mathbf{A}_k \succeq 0$  and  $r_k$  is bounded).

Due to the unknown beamforming matrix  $\mathbf{W}$ , and the non-linear SINR expression, the distribution of  $\text{SINR}_k(\mathbf{W}, \mathbf{h}_k)$  is difficult to be determined for a general PDF  $f(\mathbf{x}_k)$ . Therefore,

<sup>1</sup>Since the nonzero mean of the channel uncertainty  $\mathbf{x}_k$  can be shifted to the estimated channel  $\hat{\mathbf{h}}_k$ , without loss of generality,  $\mathbb{E}(\mathbf{x}_k) = \mathbf{0}$ .

the closed-form expression for the constraints of (1), and subsequently the feasible set of problem (1)  $\mathcal{W}_0$  is not directly available. Even worse is that the location of the feasible set of (1), which is the premise for solving the optimization problem, is not known. An usual way to tackle the problem is to find a tractable lower bound function of  $\Pr\{\text{SINR}_k(\mathbf{W}, \mathbf{h}_k) \geq \gamma_k\}$ . For Gaussian uncertainty in  $\mathbf{h}_k$ , Bernstein-type inequality provides a lower bound function of  $\Pr\{\text{SINR}_k(\mathbf{W}, \mathbf{h}_k) \geq \gamma_k\}$  [14]. However, it is not guaranteed to be a lower bound for channel uncertainties with other distributions.

On the other hand, with the moment and support information of the channel uncertainty, the inequality

$$\begin{aligned} \Pr\{\text{SINR}_k(\mathbf{W}, \mathbf{h}_k) \geq \gamma_k\} \\ \geq \inf_{\substack{\mathbf{x}_k \in \mathcal{X}_k \\ \mathbb{E}(\mathbf{x}_k) = \mathbf{0}, \mathbb{E}(\mathbf{x}_k \mathbf{x}_k^H) = \Sigma_k}} \Pr\{\text{SINR}_k(\mathbf{W}, \mathbf{h}_k) \geq \gamma_k\} \end{aligned} \quad (2)$$

holds for all channel uncertainty distributions. The rationale of using this lower bound is that by only keeping the moment and support information, an analytic expression can be obtained. To see this, the lower bound is first reformulated as

$$\begin{aligned} \inf_{f_0(\mathbf{x}_k)} \quad & \Pr\{\text{SINR}_k(\mathbf{W}, \mathbf{h}_k) \geq \gamma_k\} \\ \text{s.t.} \quad & \int_{\mathbf{x}_k \in \mathcal{X}_k} f_0(\mathbf{x}_k) d\mathbf{x}_k = 1 \\ & \mathbb{E}(\mathbf{x}_k) = \mathbf{0} \\ & \mathbb{E}(\mathbf{x}_k \mathbf{x}_k^H) = \Sigma_k, \end{aligned} \quad (3)$$

where  $f_0(\mathbf{x}_k)$  is a PDF satisfying the moment and support constraint of  $\mathbf{x}_k$ . With real Lagrangian multiplier  $\nu_k$  and complex Lagrangian multipliers  $\boldsymbol{\eta}_k, \Xi_k$  (with  $\Xi_k = \Xi_k^H$ ), it is shown in *Appendix A* that the dual problem of (3) can be expressed as

$$\begin{aligned} \max_{\mathbf{Z}_k} \quad & \text{Tr}(\mathbf{Z}_k \bar{\Sigma}_k) \\ \text{s.t.} \quad & \mathbf{u}_k^H \mathbf{Z}_k \mathbf{u}_k \leq 0, \quad \forall \mathbf{u}_k : \mathbf{u}_k^H \begin{bmatrix} \mathbf{A}_k & \mathbf{0} \\ \mathbf{0}^T & -r_k^2 \end{bmatrix} \mathbf{u}_k \leq 0, \quad \mathbf{u}_k^H \mathbf{L}(\mathbf{Q}_k) \mathbf{u}_k \leq 0 \\ & \mathbf{u}_k^H \mathbf{Z}_k \mathbf{u}_k \leq 1, \quad \forall \mathbf{u}_k : \mathbf{u}_k^H \begin{bmatrix} \mathbf{A}_k & \mathbf{0} \\ \mathbf{0}^T & -r_k^2 \end{bmatrix} \mathbf{u}_k \leq 0 \\ & \mathbf{Z}_k = \mathbf{Z}_k^H, \end{aligned} \quad (4)$$

where [see the equations at the bottom of the page], and  $\mathbf{Q}_k := \mathbf{w}_k \mathbf{w}_k^H$ . Furthermore, with S-Lemma in the complex domain

$$\begin{aligned} \mathbf{u}_k &:= [\mathbf{x}_k^T \quad 1]^T, \quad \mathbf{Z}_k := \begin{bmatrix} \Xi_k & \frac{1}{2}\boldsymbol{\eta}_k \\ \frac{1}{2}\boldsymbol{\eta}_k^H & \nu_k \end{bmatrix}, \quad \bar{\Sigma}_k := \begin{bmatrix} \Sigma_k & \mathbf{0} \\ \mathbf{0} & 1 \end{bmatrix}, \\ \mathbf{L}(\mathbf{Q}_k) &:= \begin{bmatrix} \mathbf{Q}_k - \gamma_k \sum_{l=1, l \neq k}^K \mathbf{Q}_l & \left( \mathbf{Q}_k - \gamma_k \sum_{l=1, l \neq k}^K \mathbf{Q}_l \right) \hat{\mathbf{h}}_k \\ \hat{\mathbf{h}}_k^H \left( \mathbf{Q}_k - \gamma_k \sum_{l=1, l \neq k}^K \mathbf{Q}_l \right) & \hat{\mathbf{h}}_k^H \left( \mathbf{Q}_k - \gamma_k \sum_{l=1, l \neq k}^K \mathbf{Q}_l \right) \hat{\mathbf{h}}_k - \gamma_k \delta_k^2 \end{bmatrix} \end{aligned}$$

[27], the dual problem (4) is equivalent to

$$\begin{aligned} \max_{\mathbf{Z}_k, \alpha_k, \beta_k, \kappa_k} \quad & \text{Tr}(\mathbf{Z}_k \bar{\Sigma}_k) \\ \text{s.t.} \quad & \beta_k \begin{bmatrix} \mathbf{A}_k & \mathbf{0} \\ \mathbf{0}^T & -r_k^2 \end{bmatrix} + \alpha_k \mathbf{L}(\mathbf{Q}_k) - \mathbf{Z}_k \succeq 0 \\ & \kappa_k \begin{bmatrix} \mathbf{A}_k & \mathbf{0} \\ \mathbf{0}^T & -r_k^2 \end{bmatrix} - \mathbf{Z}_k + \text{Diag}([0, 1]) \succeq 0 \\ & \mathbf{Z}_k = \mathbf{Z}_k^H, \alpha_k \geq 0, \beta_k \geq 0, \kappa_k \geq 0. \end{aligned} \quad (5)$$

By replacing the intractable function  $\Pr\{\text{SINR}_k(\mathbf{W}, \mathbf{h}_k) \geq \gamma_k\}$  in (1) with the dual problem (5), we have the beamforming design problem shown in (6) at the bottom of the page. Obviously, the constraint functions in (6) are in analytic expressions.

However, since the bilinear formulation of  $\alpha_k$  and  $\mathbf{Q}_k$  in (6) hinders further analysis, simplification of (6) is needed by using the following property.

*Property 1:* In problem (6),  $\alpha_k$  does not occur at zero for any  $k \in \mathcal{K}$ .

*Proof:* If any  $\alpha_k$  equals zero in problem (6), the first two constraints of the  $k$ th user become

$$\text{Tr}(\mathbf{Z}_k \bar{\Sigma}_k) \geq 1 - p_k, \quad (7)$$

$$\beta_k \begin{bmatrix} \mathbf{A}_k & \mathbf{0} \\ \mathbf{0}^T & -r_k^2 \end{bmatrix} - \mathbf{Z}_k \succeq 0. \quad (8)$$

Since  $\Sigma_k \succeq 0$ ,  $\bar{\Sigma}_k = \begin{bmatrix} \Sigma_k & \mathbf{0} \\ \mathbf{0}^T & 1 \end{bmatrix} \succeq 0$  is obtained. Together with the positive semidefinite property in (8), we

have  $(\bar{\Sigma}_k^{\frac{1}{2}})^H (\beta_k \begin{bmatrix} \mathbf{A}_k & \mathbf{0} \\ \mathbf{0}^T & -r_k^2 \end{bmatrix} - \mathbf{Z}_k) \bar{\Sigma}_k^{\frac{1}{2}} \succeq 0$ , which makes  $\text{Tr}((\beta_k \begin{bmatrix} \mathbf{A}_k & \mathbf{0} \\ \mathbf{0}^T & -r_k^2 \end{bmatrix} - \mathbf{Z}_k) \bar{\Sigma}_k) \geq 0$ . Furthermore, notice that  $\text{Tr}(\beta_k \begin{bmatrix} \mathbf{A}_k & \mathbf{0} \\ \mathbf{0}^T & -r_k^2 \end{bmatrix} \bar{\Sigma}_k) = \beta_k (\text{Tr}(\mathbf{A}_k \Sigma_k) - r_k^2)$ , we have

$$\beta_k (\text{Tr}(\mathbf{A}_k \Sigma_k) - r_k^2) \geq \text{Tr}(\mathbf{Z}_k \bar{\Sigma}_k). \quad (9)$$

Owing to the Jensen's inequality, we have

$$\text{Tr}(\mathbf{A}_k \Sigma_k) = \text{Tr}(\mathbf{A}_k \mathbb{E}(\mathbf{x}_k \mathbf{x}_k^H)) \quad (10)$$

$$\leq \mathbb{E}(\text{Tr}(\mathbf{A}_k \mathbf{x}_k \mathbf{x}_k^H)) \quad (11)$$

$$= \mathbb{E}(\mathbf{x}_k^H \mathbf{A}_k \mathbf{x}_k) \quad (12)$$

$$\leq r_k^2, \quad (13)$$

where the last inequality is due to the support set definition  $\{\mathbf{x}_k | \mathbf{x}_k^H \mathbf{A}_k \mathbf{x}_k \leq r_k^2, \mathbf{A}_k^H = \mathbf{A}_k\}$ . Taking the result  $\text{Tr}(\mathbf{A}_k \Sigma_k) \leq r_k^2$  and the constraint  $\beta_k \geq 0$  into (9) shows that  $\text{Tr}(\mathbf{Z}_k \bar{\Sigma}_k) \leq 0$ . Therefore, the constraint in (7) becomes  $0 \geq 1 - p_k$ , which is infeasible since the outage target  $p_k \in [0, 1)$ . ■

Therefore, the equality part of  $\{\alpha_k \geq 0\}_{k=1}^K$  in (6) is redundant, and the constraint can be reduced to  $\{\alpha_k > 0\}_{k=1}^K$ . The nonlinear formulation of  $\alpha_k$  and  $\mathbf{Q}_k$  in (6) is then removed by letting  $\tilde{\alpha}_k := 1/\alpha_k > 0$  and multiplying  $\tilde{\alpha}_k$  to the first three constraints of the  $k$ th user, and (6) becomes (14), shown at the bottom of the page with new variables  $\tilde{\mathbf{Z}}_k := \tilde{\alpha}_k \mathbf{Z}_k$ ,  $\tilde{\beta}_k := \tilde{\alpha}_k \beta_k$  and  $\tilde{\kappa}_k := \tilde{\alpha}_k \kappa_k$ .

$$\begin{aligned} \min_{\{\mathbf{w}_k, \mathbf{Q}_k, \mathbf{Z}_k, \alpha_k, \beta_k, \kappa_k\}_{k=1}^K} \quad & \sum_{k=1}^K \text{Tr}(\mathbf{Q}_k) \\ \text{s.t.} \quad & \text{Tr}(\mathbf{Z}_k \bar{\Sigma}_k) \geq 1 - p_k, \quad \forall k \in \mathcal{K} \\ & \beta_k \begin{bmatrix} \mathbf{A}_k & \mathbf{0} \\ \mathbf{0}^T & -r_k^2 \end{bmatrix} - \mathbf{Z}_k + \alpha_k \mathbf{L}(\mathbf{Q}_k) \succeq 0, \quad \forall k \in \mathcal{K} \\ & \kappa_k \begin{bmatrix} \mathbf{A}_k & \mathbf{0} \\ \mathbf{0}^T & -r_k^2 \end{bmatrix} - \mathbf{Z}_k + \text{Diag}([0, 1]) \succeq 0, \quad \forall k \in \mathcal{K} \\ & \mathbf{Z}_k = \mathbf{Z}_k^H, \quad \alpha_k \geq 0, \quad \beta_k \geq 0, \quad \kappa_k \geq 0, \quad \mathbf{Q}_k = \mathbf{w}_k \mathbf{w}_k^H, \quad \forall k \in \mathcal{K}. \end{aligned} \quad (6)$$

$$\begin{aligned} \min_{\{\mathbf{w}_k, \mathbf{Q}_k, \tilde{\mathbf{Z}}_k, \tilde{\alpha}_k, \tilde{\beta}_k, \tilde{\kappa}_k\}_{k=1}^K} \quad & \sum_{k=1}^K \text{Tr}(\mathbf{Q}_k) \\ \text{s.t.} \quad & \text{Tr}(\tilde{\mathbf{Z}}_k \bar{\Sigma}_k) \geq (1 - p_k) \tilde{\alpha}_k, \quad \forall k \in \mathcal{K} \\ & \tilde{\beta}_k \begin{bmatrix} \mathbf{A}_k & \mathbf{0} \\ \mathbf{0}^T & -r_k^2 \end{bmatrix} - \tilde{\mathbf{Z}}_k + \mathbf{L}(\mathbf{Q}_k) \succeq 0, \quad \forall k \in \mathcal{K} \\ & \tilde{\kappa}_k \begin{bmatrix} \mathbf{A}_k & \mathbf{0} \\ \mathbf{0}^T & -r_k^2 \end{bmatrix} - \tilde{\mathbf{Z}}_k + \text{Diag}([0, \tilde{\alpha}_k]) \succeq 0, \quad \forall k \in \mathcal{K} \\ & \tilde{\mathbf{Z}}_k = \tilde{\mathbf{Z}}_k^H, \quad \tilde{\alpha}_k > 0, \quad \tilde{\beta}_k \geq 0, \quad \tilde{\kappa}_k \geq 0, \quad \mathbf{Q}_k = \mathbf{w}_k \mathbf{w}_k^H, \quad \forall k \in \mathcal{K}, \end{aligned} \quad (14)$$

**Theorem 1:** Any feasible solution  $[\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_K]$  in (14) is a feasible solution of problem (1).

*Proof:* Since the objective and constraint functionals of the primal problem (3) are linear functionals of the PDF  $f_0(\mathbf{x}_k)$ , the primal problem (3) is a convex problem. Furthermore, since the moment and support information belongs to the known PDF  $f(\mathbf{x}_k)$ , the primal problem (3) is feasible. Therefore, feasible primal problem (3) guarantees the weak duality holds between (3) and its dual (5), i.e., (5) is a lower bound of (3). Furthermore, (5) is also a lower bound of  $\Pr\{\text{SINR}_k(\mathbf{W}, \mathbf{h}_k) \geq \gamma_k\}$ , which leads to the feasible set of (14) must be a feasible subset of (1). ■

Owing to the nonconvex quadratic constraint  $\mathbf{Q}_k = \mathbf{w}_k \mathbf{w}_k^H$ , solution for (14) is difficult to be obtained. A popular method to solve (14) is semidefinite relaxation (SDR), i.e., replacing the nonconvex constraint with  $\mathbf{Q}_k \succeq 0$  and deleting the rank-one constraint. Although the SDR converts (14) into an efficiently solvable problem, and it is known that Gaussian randomization procedure [19] can be used to mitigate the rank-one issue, there is still a chance that rank-one feasible solution of (1) cannot be obtained. Fortunately, owing to the special relationship between SINR and MSE, a convex problem can be proposed to find the feasible subset of (1) as follows.

**Theorem 2:** Any feasible solution  $\mathbf{W}$  in the convex problem (15) is a feasible solution of problem (1). In (15),  $\tilde{\mathbf{L}}(\mathbf{W}, b_k) := [\mathbf{W}^H, \mathbf{W}^H \hat{\mathbf{h}}_k - b_k \mathbf{e}_k]$  and the  $K \times 1$  vector  $\mathbf{e}_k = [0, \dots, 1, 0, \dots, 0]^T$  with the 1 appears at the  $k$ th position.

*Proof:* See Appendix B. ■

The tradeoff between SINR criterion (14) and MSE criterion (15) can be analyzed from the conservativeness and complexity perspectives. From the conservativeness aspect, since the MSE constraint is only a sufficient condition to guarantee the SINR constraint, the MSE criterion (15) is generally more conservative than the SINR criterion (14). From the computational complexity aspect, the convex problem (15) enables efficient method to obtain a feasible solution, while the SDR based (14) needs an extra Gaussian randomization process to obtain a rank-one solution if a higher rank solution is obtained. Therefore, the complexity of obtaining a guaranteed rank-one solution from (14) is generally higher than that of (15).

The connections of problems (14) and (15) with existing robust optimization problems are revealed as follows.

**Property 2:** When  $p_k = 0, \Sigma_k \succ 0, \forall k \in \mathcal{K}$ , (14) is degenerated to the robust beamforming problem in [5]

$$\begin{aligned} & \min_{\{\mathbf{w}_k, \mathbf{Q}_k, \tilde{\beta}_k\}_{k=1}^K} \sum_{k=1}^K \text{Tr}(\mathbf{Q}_k) \\ & \text{s.t.}, \tilde{\beta}_k \begin{bmatrix} \mathbf{A}_k & \mathbf{0} \\ \mathbf{0}^T & -r_k^2 \end{bmatrix} + \mathbf{L}(\mathbf{Q}_k) \succeq 0, \\ & \tilde{\beta}_k \geq 0, \quad \mathbf{Q}_k = \mathbf{w}_k \mathbf{w}_k^H, \quad \forall k \in \mathcal{K}. \end{aligned} \quad (16)$$

*Proof:* When  $p_k = 0$ , the  $k$ th user's constraints in (14) become

$$\text{Tr}(\tilde{\mathbf{Z}}_k \tilde{\Sigma}_k) \geq \tilde{\alpha}_k, \quad (17)$$

$$\tilde{\beta}_k \begin{bmatrix} \mathbf{A}_k & \mathbf{0} \\ \mathbf{0}^T & -r_k^2 \end{bmatrix} + \mathbf{L}(\mathbf{Q}_k) \succeq \tilde{\mathbf{Z}}_k, \quad (18)$$

$$\tilde{\kappa}_k \begin{bmatrix} \mathbf{A}_k & \mathbf{0} \\ \mathbf{0}^T & -r_k^2 \end{bmatrix} + \text{Diag}([0, \tilde{\alpha}_k]) \succeq \tilde{\mathbf{Z}}_k, \quad (19)$$

$$\tilde{\mathbf{Z}}_k = \tilde{\mathbf{Z}}_k^H, \tilde{\alpha}_k > 0, \tilde{\beta}_k \geq 0, \tilde{\kappa}_k \geq 0, \mathbf{Q}_k = \mathbf{w}_k \mathbf{w}_k^H. \quad (20)$$

Since  $\tilde{\Sigma}_k = \begin{bmatrix} \Sigma_k & \mathbf{0} \\ \mathbf{0}^T & 1 \end{bmatrix} \succeq 0$ , (19) implies  $\text{Tr}((\tilde{\kappa}_k \begin{bmatrix} \mathbf{A}_k & \mathbf{0} \\ \mathbf{0}^T & -r_k^2 \end{bmatrix} + \text{Diag}([0, \tilde{\alpha}_k]) \begin{bmatrix} \Sigma_k & \mathbf{0} \\ \mathbf{0}^T & 1 \end{bmatrix})) = \tilde{\kappa}_k (\text{Tr}(\mathbf{A}_k \Sigma_k) - r_k^2) + \tilde{\alpha}_k \geq \text{Tr}(\tilde{\mathbf{Z}}_k \tilde{\Sigma}_k)$ . Combining this result with (17), we have

$$\tilde{\kappa}_k (\text{Tr}(\mathbf{A}_k \Sigma_k) - r_k^2) + \tilde{\alpha}_k \geq \text{Tr}(\tilde{\mathbf{Z}}_k \tilde{\Sigma}_k) \geq \tilde{\alpha}_k. \quad (21)$$

Taking the two ends of the inequalities in (21), we have  $\tilde{\kappa}_k (\text{Tr}(\mathbf{A}_k \Sigma_k) - r_k^2) \geq 0$ . Since  $\text{Tr}(\mathbf{A}_k \Sigma_k) \leq r_k^2$  was proved in (10)–(13) and  $\tilde{\kappa}_k \geq 0$  is a constraint from (20),  $\tilde{\kappa}_k = 0$  is obtained.

$\tilde{\kappa}_k = 0$  has two consequences. Firstly,  $\tilde{\kappa}_k = 0$  makes (19) become  $\text{Diag}([0, \tilde{\alpha}_k]) - \tilde{\mathbf{Z}}_k \succeq 0$ . Making use of the definition  $\tilde{\mathbf{Z}}_k$ , we have  $\tilde{\alpha}_k \begin{bmatrix} -\Xi_k & -\frac{1}{2}\boldsymbol{\eta}_k \\ -\frac{1}{2}\boldsymbol{\eta}_k^H & 1 - \nu_k \end{bmatrix} \succeq 0$ , which implies

$$\Xi_k \preceq 0 \quad \text{and} \quad \nu_k \leq 1. \quad (22)$$

Secondly,  $\tilde{\kappa}_k = 0$  makes (21) become  $\text{Tr}(\tilde{\mathbf{Z}}_k \tilde{\Sigma}_k) = \tilde{\alpha}_k$ . Expanding this condition gives  $\text{Tr}(\tilde{\mathbf{Z}}_k \tilde{\Sigma}_k) = \tilde{\alpha}_k \text{Tr}(\begin{bmatrix} \Xi_k & \frac{1}{2}\boldsymbol{\eta}_k \\ \frac{1}{2}\boldsymbol{\eta}_k^H & \nu_k \end{bmatrix} \begin{bmatrix} \Sigma_k & \mathbf{0} \\ \mathbf{0}^T & 1 \end{bmatrix}) = \tilde{\alpha}_k (\text{Tr}(\Xi_k \Sigma_k) + \nu_k) = \tilde{\alpha}_k$ , which leads to

$$\text{Tr}(\Xi_k \Sigma_k) + \nu_k = 1. \quad (23)$$

$$\begin{aligned} & \min \|\mathbf{W}\|_F^2 \\ & \mathbf{W}, \{b_k, c_k, \tilde{\mathbf{Z}}_k, \tilde{\alpha}_k, \tilde{\beta}_k, \tilde{\kappa}_k\}_{k=1}^K \\ & \text{s.t.}, \text{Tr}(\tilde{\mathbf{Z}}_k \tilde{\Sigma}_k) \leq p_k \tilde{\alpha}_k, \quad \forall k \in \mathcal{K} \\ & \begin{bmatrix} \tilde{\mathbf{Z}}_k + \text{Diag}\left(\left[0, \frac{b_k}{\gamma_k+1} - c_k \delta_k^2 - \tilde{\alpha}_k\right]\right) + \tilde{\beta}_k \begin{bmatrix} \mathbf{A}_k & \mathbf{0} \\ \mathbf{0}^T & -r_k^2 \end{bmatrix}, & \tilde{\mathbf{L}}^H(\mathbf{W}, b_k) \\ \tilde{\mathbf{L}}(\mathbf{W}, b_k) & b_k \mathbf{I}_K \end{bmatrix} \succeq 0, \forall k \in \mathcal{K} \\ & \tilde{\mathbf{Z}}_k + \tilde{\kappa}_k \begin{bmatrix} \mathbf{A}_k & \mathbf{0} \\ \mathbf{0}^T & -r_k^2 \end{bmatrix} \succeq 0, \quad \|[2, b_k - c_k]\|_2 \leq b_k + c_k, \quad \forall k \in \mathcal{K} \\ & \tilde{\mathbf{Z}}_k = \tilde{\mathbf{Z}}_k^H, \tilde{\alpha}_k > 0, \tilde{\beta}_k \geq 0, \tilde{\kappa}_k \geq 0, b_k > 0, \quad \forall k \in \mathcal{K}, \end{aligned} \quad (15)$$

Since  $\Sigma_k \succ 0$  and  $\Xi_k \preceq 0$  in (22), we have  $\text{Tr}(\Xi_k \Sigma_k) \leq 0$ . Together with  $\nu_k \leq 1$ , it is obvious that only  $\Xi_k = \mathbf{0}$  and  $\nu_k = 1$  satisfy (23). Taking  $\Xi_k = \mathbf{0}$  and  $\nu_k = 1$  into  $\tilde{\alpha}_k \begin{bmatrix} -\Xi_k & -\frac{1}{2}\eta_k \\ -\frac{1}{2}\eta_k^H & 1 - \nu_k \end{bmatrix} \succeq 0$  reveals  $\eta_k = \mathbf{0}$ , which eventually leads to  $\tilde{\mathbf{Z}}_k = \text{Diag}([0, \tilde{\alpha}_k])$ . Putting  $\tilde{\kappa}_k = 0$  and  $\tilde{\mathbf{Z}}_k = \text{Diag}([0, \tilde{\alpha}_k])$  into (17)–(20), it can be easily shown that (17) and (19) can be eliminated and the simplified constraints become

$$\tilde{\beta}_k \begin{bmatrix} \mathbf{A}_k & \mathbf{0} \\ \mathbf{0}^T & -r_k^2 \end{bmatrix} + \mathbf{L}(\mathbf{Q}_k) \succeq \text{Diag}([0, \tilde{\alpha}_k]),$$

$$\tilde{\alpha}_k > 0, \tilde{\beta}_k \geq 0, \mathbf{Q}_k = \mathbf{w}_k \mathbf{w}_k^H. \quad (24)$$

Since  $\tilde{\alpha}_k$  is not involved in the objective function of (14), the largest feasible set of  $(\mathbf{w}_k, \mathbf{Q}_k, \tilde{\beta}_k)$  in (24) occurs at the limit condition  $\tilde{\alpha}_k \rightarrow 0^+$ . Therefore,  $p_k = 0, \forall k \in \mathcal{K}$  make the problem (14) degenerate to the robust optimization problem (16). ■

Similarly,  $p_k = 0, \Sigma_k \succ 0, \forall k \in \mathcal{K}$  in (15) also leads to  $\tilde{\kappa}_k = 0, \tilde{\mathbf{Z}}_k = \mathbf{0}$  and  $\tilde{\alpha}_k \rightarrow 0^+, \forall k \in \mathcal{K}$ , and (15) degenerates into the robust transceiver design problem in [8].

*Property 3:* When the channel uncertainty is unbounded, i.e.,  $r_k \rightarrow +\infty, \forall k \in \mathcal{K}$ , we have  $\tilde{\kappa}_k = 0$  and  $\tilde{\beta}_k = 0, \forall k \in \mathcal{K}$  in (14).

*Proof:* Since (14) is equivalent to (6), we start from (6) instead. The first and the third constraints of the  $k$ th user in (6) are

$$\text{Tr} \left( \underbrace{\begin{bmatrix} \Xi_k & \frac{1}{2}\eta_k \\ \frac{1}{2}\eta_k^H & \nu_k \end{bmatrix}}_{\mathbf{Z}_k} \underbrace{\begin{bmatrix} \Sigma_k & \mathbf{0} \\ \mathbf{0}^T & 1 \end{bmatrix}}_{\tilde{\Sigma}_k} \right) \geq 1 - p_k, \quad (25)$$

$$\kappa_k \begin{bmatrix} \mathbf{A}_k & \mathbf{0} \\ \mathbf{0}^T & -r_k^2 \end{bmatrix} - \begin{bmatrix} \Xi_k & \frac{1}{2}\eta_k \\ \frac{1}{2}\eta_k^H & \nu_k \end{bmatrix} + \text{Diag}([0, 1]) \succeq 0. \quad (26)$$

First, (25) is simplified to  $\nu_k + \text{Tr}(\Xi_k \Sigma_k) \geq 1 - p_k > 0$  owing to the outage probability  $p_k \in [0, 1]$ . Furthermore, since the diagonal elements and the principle submatrices of a positive semidefinite matrix must be nonnegative and positive semidefinite respectively, (26) implies

$$-\nu_k \geq \kappa_k r_k^2 - 1, \quad (27)$$

$$\lambda_{\max}(\kappa_k \mathbf{A}_k) \geq \lambda_{\max}(\Xi_k), \quad (28)$$

where  $\lambda_{\max}(\cdot)$  denotes the largest eigenvalue of a Hermitian matrix, and (28) is due to the fact that for  $\kappa_k \mathbf{A}_k \succeq \Xi_k$ , we have  $\lambda_{\max}(\kappa_k \mathbf{A}_k) \geq \lambda_{\max}(\Xi_k)$  [21, p. 162]. Putting  $-\nu_k \geq \kappa_k r_k^2 - 1$  from (27) into  $\nu_k + \text{Tr}(\Xi_k \Sigma_k) > 0$ , we have

$$\text{Tr}(\Xi_k \Sigma_k) > \kappa_k r_k^2 - 1. \quad (29)$$

With the covariance matrix  $\Sigma_k \succeq 0$ , we have  $\lambda_{\max}(\Xi_k) \text{Tr}(\Sigma_k) \geq \text{Tr}(\Xi_k \Sigma_k)$  [20]. Putting this result into (29), we obtain

$$\lambda_{\max}(\Xi_k) \text{Tr}(\Sigma_k) > \kappa_k r_k^2 - 1. \quad (30)$$

Combining (28) with (30), we have

$$\kappa_k \lambda_{\max}(\mathbf{A}_k) - (\kappa_k r_k^2 - 1) / \text{Tr}(\Sigma_k) > 0. \quad (31)$$

Since  $\kappa_k \geq 0$  from the constraint of (6), in the following, we divide the discussion into  $\kappa_k > 0$  and  $\kappa_k = 0$ . If  $\kappa_k > 0$ , (31) is equivalent to  $\kappa_k (\lambda_{\max}(\mathbf{A}_k) - (r_k^2 - \frac{1}{\kappa_k}) / \text{Tr}(\Sigma_k)) > 0$ , which is infeasible when  $r_k \rightarrow +\infty$  as  $\text{Tr}(\Sigma_k) > 0$  and  $\mathbf{A}_k$  is a matrix with bounded elements. On the other hand,  $\kappa_k = 0$  satisfies the constraint in (31). Therefore,  $\kappa_k \geq 0$  in (6) becomes  $\kappa_k = 0$ , i.e.,  $\tilde{\kappa}_k = \tilde{\alpha}_k \kappa_k = 0$  in (14). Similar contradiction proofs can be used to prove  $\tilde{\beta}_k = 0$  in (14) when  $r_k \rightarrow +\infty$ . ■

Similarly,  $r_k \rightarrow +\infty, \forall k \in \mathcal{K}$  in (15) leads to  $\tilde{\beta}_k = 0$  and  $\tilde{\kappa}_k = 0, \forall k \in \mathcal{K}$ , and (15) degenerates into the moment-constrained transceiver design problem in [18].

*Remark 1:* In practice, the parameters of the support set  $\mathcal{X}_k$  can be determined as follows. For channel estimation error, since the noise is generally modelled as a Gaussian random variable, the estimation error is unbounded. Therefore, the parameters of the support set are  $\mathbf{A}_k = \mathbf{I}_N$  and  $r_k \rightarrow +\infty$ . On the other hand, a finite bound is more appropriate for quantization error. In particular, in such case, it is a common practice to choose the covariance matrix  $\Sigma_k$  as  $\mathbf{A}_k$ . Furthermore, with  $N_s$  independent error sample  $\{\mathbf{x}_k[i]\}_{i=1}^{N_s}$ , an estimator for  $r_k$  is  $\hat{r}_k = \max(\{\sqrt{\mathbf{x}_k[i]^H \Sigma_k \mathbf{x}_k[i]}\}_{i=1}^{N_s})$ . According to the one-sided Chernoff bound in [22, p. 115, (8.16)],  $N_s = \frac{1}{2\epsilon^2} \ln \frac{1}{\delta}$  independent error samples guarantee the estimated support set cover  $100(1 - \epsilon)$  percent channel uncertainty with reliability  $1 - \delta$ .

### III. JOINT FEASIBLE SUBSETS REFINEMENT AND OPTIMIZATION

Since the lower bound function in (2) is only a conservative bound, the obtained feasible set is a conservative feasible subset of (1). Although other lower bound functions might exist, conservativeness is still the key problem. In this section, the local structure of a given feasible solution is utilized systematically to explore other feasible subset of (1). For any given feasible solution  $\bar{\mathbf{W}}$ , we first define  $\mathcal{H}_k(\bar{\mathbf{W}}) := \{\mathbf{h}_k | \mathbf{h}_k \in \{\hat{\mathbf{h}}_k + \mathcal{X}_k\}, \text{SINR}_k(\bar{\mathbf{W}}, \mathbf{h}_k) \geq \gamma_k\}$ , where the Minkowski sum in the set  $\{\hat{\mathbf{h}}_k + \mathcal{X}_k\}$  means  $\{\mathbf{h}_k | (\mathbf{h}_k - \hat{\mathbf{h}}_k)^H \mathbf{A}_k (\mathbf{h}_k - \hat{\mathbf{h}}_k) \leq r_k^2\}$ . Obviously,  $\mathcal{H}_k(\bar{\mathbf{W}})$  is a support subset of  $\mathbf{h}_k$  satisfying the SINR requirement with the given  $\bar{\mathbf{W}}$ . Then, we define another set.

*Definition 1:*  $\mathcal{W}(\bar{\mathbf{W}}) := \{\mathbf{W} | \{\text{SINR}_k(\mathbf{W}, \mathbf{h}_k) \geq \gamma_k, \forall \mathbf{h}_k \in \mathcal{H}_k(\bar{\mathbf{W}})\}_{k=1}^K\}$ .

An important property of  $\mathcal{W}(\bar{\mathbf{W}})$  can be obtained as follow.

*Property 4:*  $\bar{\mathbf{W}} \in \mathcal{W}(\bar{\mathbf{W}}) \subseteq \mathcal{W}_0$

*Proof:* See Appendix C. ■

Therefore, each feasible solution  $\bar{\mathbf{W}}$  of (1) can generate a feasible subset  $\mathcal{W}(\bar{\mathbf{W}})$  which contains  $\bar{\mathbf{W}}$  itself. Although optimization over  $\mathcal{W}(\bar{\mathbf{W}})$  may find better solution than  $\bar{\mathbf{W}}$ , obviously, an even larger feasible subset than  $\mathcal{W}(\bar{\mathbf{W}})$  is highly desirable.

More specifically, from the coupling effect between the support subset  $\mathcal{H}_k(\bar{\mathbf{W}})$  and the feasible subset  $\mathcal{W}(\bar{\mathbf{W}})$  in *Definition 1*, it can be seen that reducing the number of elements in the support subset  $\mathcal{H}_k(\bar{\mathbf{W}})$  may enlarge the feasible subset  $\mathcal{W}(\bar{\mathbf{W}})$ . Therefore, we consider a squeezed support subset  $\mathcal{H}_k(\bar{\mathbf{W}})$  as

$$\mathcal{H}_k(\bar{\mathbf{W}}, q_k) := \{\mathbf{h}_k | \mathbf{h}_k \in \{\hat{\mathbf{h}}_k + \mathcal{X}_k\}, \text{SINR}_k(\bar{\mathbf{W}}, \mathbf{h}_k) \geq q_k\}, \quad (32)$$

where  $q_k \geq \gamma_k$ , and we have

$$\mathcal{H}_k(\bar{\mathbf{W}}, q_k) \subseteq \mathcal{H}_k(\bar{\mathbf{W}}). \quad (33)$$

Then the corresponding set generated from  $\mathcal{H}_k(\bar{\mathbf{W}}, q_k)$  is

$$\mathcal{W}(\bar{\mathbf{W}}, \mathbf{q}) := \{\mathbf{W} | \{\text{SINR}_k(\mathbf{W}, \mathbf{h}_k) \geq \gamma_k, \forall \mathbf{h}_k \in \mathcal{H}_k(\bar{\mathbf{W}}, q_k)\}_{k=1}^K\}, \quad (34)$$

where  $\mathbf{q} = [q_1, q_2, \dots, q_K]^T$ . In order to make  $\mathcal{W}(\bar{\mathbf{W}}, \mathbf{q})$  a feasible subset of  $\mathcal{W}_0$ , the parameters  $\{q_k\}_{k=1}^K$  should be chosen such that for any  $\mathbf{W} \in \mathcal{W}(\bar{\mathbf{W}}, \mathbf{q})$ , it must satisfy the constraints in (1), i.e.,  $\Pr\{\text{SINR}_k(\mathbf{W}, \mathbf{h}_k) \geq \gamma_k\} \geq 1 - p_k$ . With similar derivations to (60) of *Appendix C*, it can be easily established that  $\Pr\{\text{SINR}_k(\mathbf{W}, \mathbf{h}_k) \geq \gamma_k\} = \Pr\{\mathbf{h}_k \in \mathcal{H}_k(\bar{\mathbf{W}}, q_k)\} + c$ , where  $c$  is always nonnegative. Since increasing  $q_k$  would decrease  $\Pr\{\mathbf{h}_k \in \mathcal{H}_k(\bar{\mathbf{W}}, q_k)\}$ , in order to guarantee  $\Pr\{\text{SINR}_k(\mathbf{W}, \mathbf{h}_k) \geq \gamma_k\} \geq 1 - p_k$ , the maximum  $q_k$  is chosen to satisfy  $\Pr\{\mathbf{h}_k \in \mathcal{H}_k(\bar{\mathbf{W}}, q_k)\} = 1 - p_k$ . Furthermore, since  $\Pr\{\mathbf{h}_k \in \mathcal{H}_k(\bar{\mathbf{W}}, q_k)\} = \Pr\{\mathbf{h}_k \in \{\hat{\mathbf{h}}_k + \mathcal{X}_k\}, \text{SINR}_k(\bar{\mathbf{W}}, \mathbf{h}_k) \geq q_k\}$  and  $\Pr\{\mathbf{h}_k \in \{\hat{\mathbf{h}}_k + \mathcal{X}_k\}\} = 1$ , we have  $q_k$  be selected such that  $\Pr\{\text{SINR}_k(\bar{\mathbf{W}}, \mathbf{h}_k) \geq q_k\} = 1 - p_k$ .

To reveal the inter-relationship between  $\mathcal{W}(\bar{\mathbf{W}})$  and  $\mathcal{W}(\bar{\mathbf{W}}, \mathbf{q})$ , we consider

$$\begin{aligned} \mathcal{W}(\bar{\mathbf{W}}, \mathbf{q}) \cap \mathcal{W}(\bar{\mathbf{W}}) &= \left\{ \mathbf{W} | \{\text{SINR}_k(\mathbf{W}, \mathbf{h}_k) \geq \gamma_k, \forall \mathbf{h}_k \in \mathcal{H}_k(\bar{\mathbf{W}}, q_k)\}_{k=1}^K, \right. \\ &\quad \left. \{\text{SINR}_k(\mathbf{W}, \mathbf{h}_k) \geq \gamma_k, \forall \mathbf{h}_k \in \mathcal{H}_k(\bar{\mathbf{W}})\}_{k=1}^K \right\} \quad (35) \end{aligned}$$

$$= \left\{ \mathbf{W} | \{\text{SINR}_k(\mathbf{W}, \mathbf{h}_k) \geq \gamma_k, \forall \mathbf{h}_k \in \mathcal{H}_k(\bar{\mathbf{W}})\}_{k=1}^K \right\} \quad (36)$$

$$= \mathcal{W}(\bar{\mathbf{W}}), \quad (37)$$

where the second equality comes from the inclusive relationship in (33) and the final equality comes from the *Definition 1*. Therefore, an important property of those constructed feasible subsets is

$$\bar{\mathbf{W}} \in \mathcal{W}(\bar{\mathbf{W}}) \subseteq \mathcal{W}(\bar{\mathbf{W}}, \mathbf{q}) \subseteq \mathcal{W}_0. \quad (38)$$

That is, the squeezed support subsets  $\{\mathcal{H}_k(\bar{\mathbf{W}}, q_k)\}_{k=1}^K$  in (32) enlarge the corresponding feasible subset  $\mathcal{W}(\bar{\mathbf{W}}, \mathbf{q})$  in (34).

With the largest feasible subset  $\mathcal{W}(\bar{\mathbf{W}}, \mathbf{q})$  tuned by  $\mathbf{q}$ , owing to  $\bar{\mathbf{W}} \in \mathcal{W}(\bar{\mathbf{W}}, \mathbf{q})$ , better feasible solution than  $\bar{\mathbf{W}}$  can be found via  $\min\{\|\mathbf{W}\|_F^2 | \mathbf{W} \in \mathcal{W}(\bar{\mathbf{W}}, \mathbf{q})\}$ . With the obtained new solution, we can construct another feasible subset of  $\mathcal{W}_0$  and perform another optimization, and so on. That makes iterative improvement of the objective function becomes possible. The proposed iterative procedure begins with finding a feasible solution  $\mathbf{W}^{[0]} = \bar{\mathbf{W}}$  from (14) or (15) (or the solution in [13], [14], [16] under Gaussian channel uncertainty), followed by iterations between the following two steps until convergence.

- **P-step:** Finding  $q_k^{[i]} \in [\gamma_k, +\infty)$  such that  $\Pr\{\text{SINR}_k(\mathbf{W}^{[i]}, \mathbf{h}_k) \geq q_k^{[i]}\} = 1 - p_k$ .
- **O-step:** Solving the  $i$ th subproblem  $\min\{\|\mathbf{W}\|_F^2 | \{\text{SINR}_k(\mathbf{W}, \mathbf{h}_k) \geq \gamma_k, \forall \mathbf{h}_k \in \mathcal{H}_k(\mathbf{W}^{[i]}, q_k^{[i]})\}_{k=1}^K\}$ , denoting the solution as  $\mathbf{W}^{[i+1]}$ . Increment  $i$  by one.

*Lemma 1:* If  $\mathbf{W}^{[i]}$  generated from the  $(i-1)$ th O-step does not activate the  $k$ th inequality constraint in the original problem (1), then  $\mathbf{W}^{[i]}$  does not activate the  $k$ th inequality constraint of the  $i$ th O-step subproblem.

*Proof:* If the  $(i-1)$ th O-step solution  $\mathbf{W}^{[i]}$  does not activate the  $k$ th constraint in (1), i.e.,  $\Pr\{\text{SINR}_k(\mathbf{W}^{[i]}, \mathbf{h}_k) \geq \gamma_k\} > 1 - p_k$ , the parameter  $q_k^{[i]} > \gamma_k$  is needed to make  $\Pr\{\text{SINR}_k(\mathbf{W}^{[i]}, \mathbf{h}_k) \geq q_k^{[i]}\} = 1 - p_k$  at P-step. Together with the definition  $\mathcal{H}_k(\mathbf{W}^{[i]}, q_k^{[i]}) = \{\mathbf{h}_k | \mathbf{h}_k \in \{\hat{\mathbf{h}}_k + \mathcal{X}_k\}, \text{SINR}_k(\mathbf{W}^{[i]}, \mathbf{h}_k) \geq q_k^{[i]}\}$ , we have  $\text{SINR}_k(\mathbf{W}^{[i]}, \mathbf{h}_k) \geq q_k^{[i]} > \gamma_k$  is satisfied for all  $\mathbf{h}_k \in \mathcal{H}_k(\mathbf{W}^{[i]}, q_k^{[i]})$ . Therefore,  $\mathbf{W}^{[i]}$  does not activate the  $k$ th constraint  $\text{SINR}_k(\mathbf{W}^{[i]}, \mathbf{h}_k) \geq \gamma_k$  in the  $i$ th O-step subproblem. ■

*Lemma 1* reveals the connection between the original problem (1) and the O-step subproblem, which facilitates the convergence analysis of the iterative procedure presented as follows.

*Proposition 1:* If every O-step generates descent solution, the iterative procedure converges. With strictly descent solution in every O-step, the limit solution activates all  $K$  users' constraints in problem (1).

*Proof:* First, since  $\mathbf{W}^{[i]} \in \mathcal{W}(\mathbf{W}^{[i]}, \mathbf{q})$  is established in (38), a descent solution with  $\|\mathbf{W}^{[i+1]}\|_F^2 \leq \|\mathbf{W}^{[i]}\|_F^2$  is possible. With the monotonic decreasing property of  $\|\mathbf{W}\|_F^2$ , and the transmit power is bounded below by zero, the convergence of iterative procedure is guaranteed.

Second, if  $\mathbf{W}^{[i]}$  does not activate the  $k$ th constraint in the original problem (1), according to *Lemma 1*,  $\mathbf{W}^{[i]}$  does not activate the  $k$ th constraint  $\frac{\|\mathbf{h}_k^H \mathbf{w}_k\|_2^2}{\sum_{l=1, l \neq k}^K \|\mathbf{h}_k^H \mathbf{w}_l\|_2^2 + \delta_k^2} \geq \gamma_k$  in  $i$ th O-step subproblem. This implies directly scaling down the  $k$ th beamformer  $\mathbf{w}_k$  (the  $k$ th column of  $\mathbf{W}^{[i]}$ ) until  $\frac{\|\mathbf{h}_k^H \mathbf{w}_k\|_2^2}{\sum_{l=1, l \neq k}^K \|\mathbf{h}_k^H \mathbf{w}_l\|_2^2 + \delta_k^2} = \gamma_k$  would reduce transmit power strictly, hence  $\|\mathbf{W}^{[i+1]}\|_F^2 < \|\mathbf{W}^{[i]}\|_F^2$  becomes possible. Furthermore, scaling down  $\mathbf{w}_k$  reduces interference leakage to other users, and other SINR constraints would remain valid. Therefore, the iterative procedure with strictly decreasing transmit power in successive O-steps would not stop, as long as any of the user's constraint in (1) is not active. That is, the limit solution activates all constraints in (1). ■

#### IV. COMPUTATION DETAILS OF THE ITERATIVE PROCEDURE

In the previous section, the framework of sequential optimization is established. The details of P-step and O-step are derived in this section.

##### A. P-Step

The P-step is to find the quantile  $q_k^{[i]}$  such that  $\Pr\{\text{SINR}_k(\mathbf{W}^{[i]}, \mathbf{h}_k) \geq q_k^{[i]}\} = 1 - p_k$ . Owing to the monotonic increasing property of the continuous cumulative distribution function (CDF) of  $\text{SINR}_k(\mathbf{W}^{[i]}, \mathbf{h}_k)$ , the parameter  $q_k^{[i]}$  can be found in  $[\gamma_k, +\infty)$  by bisection. However, even if the PDF of channel uncertainty  $\mathbf{x}_k$  is known, due to the complicated dependence of SINR with respect to  $\mathbf{x}_k$ , the exact CDF of SINR in general is difficult to be obtained. One straightforward way is to use Monte Carlo methods, in which

samples  $\{\mathbf{x}_k[j]\}_{j=1}^{N_s}$  are generated, and fed into the SINR expression to obtain its empirical distribution. Then  $q_k^{[i]}$  can be found by bisection. To make the realized outage probability  $\hat{p}_k := 1 - \Pr\{\text{SINR}_k(\mathbf{W}^{[i]}, \mathbf{h}_k) \geq q_k^{[i]}\}$  approach the target outage probability  $p_k$  with accuracy  $\epsilon$  and reliability  $1 - \delta$ , i.e.,  $\Pr(|\hat{p}_k - p_k| \leq \epsilon) \geq 1 - \delta$ , the number of independent samples needed is  $N_s = \frac{1}{2\epsilon^2} \ln \frac{2}{\delta}$  [22, p. 114]. Since the Monte Carlo method involves vector multiplications and a sorting process, the complexity is  $\mathcal{O}(N_s(NK + \log N_s))$ .

Although Monte Carlo methods are general and easily be implemented, the computational complexity is high, e.g., a mild accuracy requirement  $\epsilon = 1\%$  makes  $N_s$  on the order of  $10^4$ . On the other hand, notice that the SINR outage probability can be equivalently written as

$$\Pr\{\mathbf{h}_k^H (\mathbf{Q}_k^{[i]} - q_k^{[i]} \sum_{l=1, l \neq k}^K \mathbf{Q}_l^{[i]}) \mathbf{h}_k \geq q_k^{[i]} \delta_k^2\} = 1 - p_k, \quad (39)$$

where  $\mathbf{Q}_k^{[i]} = \mathbf{w}_k^{[i]} (\mathbf{w}_k^{[i]})^H$ . If the cumulant-generating function (CGF) of  $d_k$  is known, saddlepoint approximation [23] provides accurate and efficient way to evaluate the left hand side of (39). In particular, with CGF of  $d_k$  denoted by  $g(t)$ , the second-order saddlepoint approximation of the probability in (39) is [24, p. 53]

$$\Pr(d_k \geq q_k^{[i]} \delta_k^2) \simeq 1 - \Phi(u) - \phi(u) \left\{ \frac{1}{u} - \frac{1}{v} - \frac{1}{v} \left( \frac{O_4}{8} - \frac{5}{24} (O_3)^2 \right) + \frac{1}{v^3} + \frac{O_3}{2v^2} - \frac{1}{u^3} \right\}, \quad (40)$$

where  $\Phi(\cdot)$  and  $\phi(\cdot)$  are the CDF and PDF of the standard normal distribution,  $O_n = g^{(n)}(t_0)/\{g''(t_0)\}^{n/2}$  with  $n = \{3, 4\}$ ,  $u = \text{sgn}(t_0) \sqrt{2(t_0 \cdot q_k^{[i]} \delta_k^2 - g(t_0))}$ ,  $v = t_0 \sqrt{g''(t_0)}$ , and the saddlepoint  $t_0$  is determined from

$$g'(t_0) = q_k^{[i]} \delta_k^2. \quad (41)$$

For Gaussian channel uncertainty, it is shown in *Appendix D* that the CGF of  $d_k$  is  $g(t) = \sum_{j=1}^N (\frac{|m_j|^2 \lambda_j t}{1 - 2\lambda_j t} - \ln(1 - 2\lambda_j t))$  with domain  $(1/(2\lambda_N), 1/(2\lambda_1))$ , where  $m_j$  being the  $j$ th element of vector  $\mathbf{V}_k^H (\mathbf{\Sigma}_k/2)^{-\frac{1}{2}} \hat{\mathbf{h}}_k$ , and  $\mathbf{V}_k, \{\lambda_j\}_{j=1}^N$  come from the eigendecomposition  $((\mathbf{\Sigma}_k/2)^{\frac{1}{2}})^H (\mathbf{Q}_k^{[i]} - q_k^{[i]} \sum_{l=1, l \neq k}^K \mathbf{Q}_l^{[i]}) (\mathbf{\Sigma}_k/2)^{\frac{1}{2}} = \mathbf{V}_k \text{Diag}([\lambda_1, \dots, \lambda_N]) \mathbf{V}_k^H$  with descending eigenvalue order. Furthermore, the uniqueness of the saddlepoint  $t_0$  in (41) is guaranteed by the fact that  $g''(t) > 0$  in its domain. The relative error  $\Pr(d_k \geq q_k^{[i]} \delta_k^2)/(1 - p_k)$  by using (40) can be calculated in [25], then pre-distorting the relative error in the outage target ensures tight outage probability control even under saddlepoint

approximation error. Note that the saddlepoint method involves one eigenvalue decomposition in each bisection iteration, the overall complexity is  $\mathcal{O}(\log_2(R)N^3)$ , where  $R$  is the ratio of bisection range divided by required accuracy. Since  $\log_2(R)$  represents the bisection iteration number and is generally smaller than 100, the complexity of the saddlepoint method is generally smaller than that of the Monte Carlo method.

### B. O-Step

With  $q_k^{[i]}$ , the  $i$ th O-step subproblem is

$$\begin{aligned} \min_{\mathbf{W}} \quad & \|\mathbf{W}\|_F^2 \\ \text{s.t.} \quad & \text{SINR}_k(\mathbf{W}, \mathbf{h}_k) \geq \gamma_k, \forall \mathbf{h}_k \in \{\hat{\mathbf{h}}_k + \mathcal{X}_k\}, \text{ and } \text{SINR}_k(\mathbf{W}^{[i]}, \mathbf{h}_k) \geq q_k^{[i]}, \forall k \in \mathcal{K}. \end{aligned} \quad (42)$$

With  $\mathbf{Q}_k^{[i]} = \mathbf{w}_k^{[i]} (\mathbf{w}_k^{[i]})^H$ , the  $k$ th constraint of (42) is reformulated as (43) at the bottom of the page. After applying the S-Lemma in complex domain [27], (43) is equivalent to

$$\begin{aligned} & \begin{bmatrix} \mathbf{Q}_k - \gamma_k \sum_{l=1, l \neq k}^K \mathbf{Q}_l & \mathbf{0} \\ \mathbf{0}^T & -\gamma_k \delta_k^2 \end{bmatrix} \\ & - \alpha_k \begin{bmatrix} \mathbf{Q}_k^{[i]} - q_k^{[i]} \sum_{l=1, l \neq k}^K \mathbf{Q}_l^{[i]} & \mathbf{0} \\ \mathbf{0}^T & -q_k^{[i]} \delta_k^2 \end{bmatrix} \\ & + \beta_k \begin{bmatrix} \mathbf{A}_k & -\mathbf{A}_k \hat{\mathbf{h}}_k \\ -\hat{\mathbf{h}}_k^H \mathbf{A}_k & \hat{\mathbf{h}}_k^H \mathbf{A}_k \hat{\mathbf{h}}_k - r_k^2 \end{bmatrix} \succeq 0, \alpha_k \geq 0, \quad \beta_k \geq 0. \end{aligned} \quad (44)$$

Therefore, the O-step subproblem (42) can be transformed to

$$\begin{aligned} \min_{\{\mathbf{Q}_k, \alpha_k, \beta_k\}_{k=1}^K} \quad & \sum_{k=1}^K \text{Tr}(\mathbf{Q}_k) \\ \text{s.t.} \quad & \begin{bmatrix} \mathbf{Q}_k - \gamma_k \sum_{l=1, l \neq k}^K \mathbf{Q}_l & \mathbf{0} \\ \mathbf{0}^T & -\gamma_k \delta_k^2 \end{bmatrix} \\ & - \alpha_k \begin{bmatrix} \mathbf{Q}_k^{[i]} - q_k^{[i]} \sum_{l=1, l \neq k}^K \mathbf{Q}_l^{[i]} & \mathbf{0} \\ \mathbf{0}^T & -q_k^{[i]} \delta_k^2 \end{bmatrix} \\ & + \beta_k \begin{bmatrix} \mathbf{A}_k & -\mathbf{A}_k \hat{\mathbf{h}}_k \\ -\hat{\mathbf{h}}_k^H \mathbf{A}_k & \hat{\mathbf{h}}_k^H \mathbf{A}_k \hat{\mathbf{h}}_k - r_k^2 \end{bmatrix} \succeq 0, \quad \forall k \in \mathcal{K} \\ & \alpha_k \geq 0, \quad \beta_k \geq 0, \quad \mathbf{Q}_k \succeq 0, \quad \text{Rank}(\mathbf{Q}_k) = 1, \quad \forall k \in \mathcal{K}. \end{aligned} \quad (45)$$

A popular approach to solve the nonconvex rank-one constrained problem (45) is SDR, i.e., deleting the rank-one constraints in (45), and the property of the SDR solution of (45) is described as follows.

$$\mathbf{h}_k^H \left( \mathbf{Q}_k - \gamma_k \sum_{l=1, l \neq k}^K \mathbf{Q}_l \right) \mathbf{h}_k \geq \gamma_k \delta_k^2, \forall \mathbf{h}_k \in \{\hat{\mathbf{h}}_k + \mathcal{X}_k\}, \text{ and } \mathbf{h}_k^H \left( \mathbf{Q}_k^{[i]} - q_k^{[i]} \sum_{l=1, l \neq k}^K \mathbf{Q}_l^{[i]} \right) \mathbf{h}_k \geq q_k^{[i]} \delta_k^2 \quad (43)$$



**Theorem 3:** If  $\gamma_k \gamma_l \geq 1$  with  $k \neq l$ , the SDR solution of (45) guarantees to be low rank as  $\text{rank}(\mathbf{Q}_k), \text{rank}(\mathbf{Q}_l) \in [1, N-1]$  and  $\text{Rank}(\mathbf{Q}_k) + \text{Rank}(\mathbf{Q}_l) \leq N$ .

*Proof:* See Appendix E. ■

**Theorem 3** reveals that when the number of BS antenna  $N = 2$ , the SDR solution is equivalent to the solution of the non-convex problem (45) provided that the product of SINR targets is larger than one (0 dB). When  $N \geq 3$ , although the SDR solutions are still forced to be low ranks, the rank-one property in general cannot be guaranteed. Fortunately, according to *Proposition 1*, we only need a descent solution for (45) in order to guarantee the proposed procedure to converge. By observing the special structure of (45), a rank-one descent solution at O-step can be obtained for any number of BS antenna  $N$  as follow.

**Property 5:** Let  $\mathbf{Q}_k^{[0]} = \mathbf{w}_k \mathbf{w}_k^H$ , where  $\{\mathbf{w}_k\}_{k=1}^K$  is a feasible solution from (1). A descent solution with respect to  $\mathbf{Q}_k^{[i]}$  in (45) is  $\{(\mathbf{Q}_k, \alpha_k, \beta_k) = (\frac{\gamma_k}{q_k^{[i]}} \mathbf{Q}_k^{[i]}, \frac{\gamma_k}{q_k^{[i]}} \mathbf{Q}_k^{[i]}, 0)\}_{k=1}^K$ .

*Proof:* First, it can be checked that  $\mathbf{Q}_k = \frac{\gamma_k}{q_k^{[i]}} \mathbf{Q}_k^{[i]}, \alpha_k = \frac{\gamma_k}{q_k^{[i]}}$  and  $\beta_k = 0$  satisfy the positive semidefinite constraint in (45). Second, since the initial solution  $\mathbf{Q}_k^{[0]}$  (required by (14) or obtained from (15)) is rank-one, later iterative solutions  $\frac{\gamma_k}{q_k^{[i]}} \mathbf{Q}_k^{[i]}$  are also rank-one. Finally, the P-step requirement  $q_k^{[i]} \geq \gamma_k$  ensures the objective value of (45) decrease monotonically. ■

To solve the O-step subproblem (45), the complexity of the SDR method is  $\mathcal{O}(N^6 K^3/4)$  in each interior-point iteration [26], while that of the descent method in *Property 5* is only  $\mathcal{O}(NK)$ . Therefore, the descent method is simpler than the SDR method in terms of complexity. However, if the SDR is tight, the SDR method gives the optimal solution in (45) and is better than the descent method.

**Remark 2:** Let us consider the case when the channel uncertainty is unbounded, i.e.,  $r_k \rightarrow +\infty$  in (45). From (44), the positive semidefinite constraint implies

$$\alpha_k \geq \left( \gamma_k \delta_k^2 + \beta_k \left( r_k^2 - \hat{\mathbf{h}}_k^H \mathbf{A}_k \hat{\mathbf{h}}_k \right) \right) / \left( q_k^{[i]} \delta_k^2 \right), \quad (46)$$

$$\mathbf{Q}_k - \gamma_k \sum_{l=1, l \neq k}^K \mathbf{Q}_l - \alpha_k \left( \mathbf{Q}_k^{[i]} - q_k^{[i]} \sum_{l=1, l \neq k}^K \mathbf{Q}_l^{[i]} \right) + \beta_k \mathbf{A}_k \succeq 0. \quad (47)$$

If  $\beta_k > 0$ , the constraint (46) would result in  $\alpha_k \rightarrow +\infty$  when  $r_k \rightarrow +\infty$ . However, putting  $\alpha_k \rightarrow +\infty$  into (47) leads to infinite power in  $\mathbf{Q}_k$ , which cannot be the optimal solution owing to the existence of finite power solution in *Property 5*. Therefore,  $\beta_k$  must be equal to zero and (45) is degenerated to the following problem

$$\begin{aligned} & \min_{\{\mathbf{Q}_k, \alpha_k\}_{k=1}^K} \sum_{k=1}^K \text{Tr}(\mathbf{Q}_k) \\ & \text{s.t. } \mathbf{Q}_k - \gamma_k \sum_{l=1, l \neq k}^K \mathbf{Q}_l - \alpha_k \left( \mathbf{Q}_k^{[i]} - q_k^{[i]} \sum_{l=1, l \neq k}^K \mathbf{Q}_l^{[i]} \right) \succeq 0, \quad \forall k \in \mathcal{K} \\ & \alpha_k q_k^{[i]} \geq \gamma_k, \quad \mathbf{Q}_k \succeq 0, \quad \text{Rank}(\mathbf{Q}_k) = 1, \quad \forall k \in \mathcal{K}. \end{aligned} \quad (48)$$

Since (48) is the limiting case of (45), problem (48) naturally inherits the properties in *Theorem 3* and *Property 5*.

### C. Summary and Convergence Properties

The proposed iterative procedure for the probabilistic beamforming problem starts with any feasible solution of (1), and follows iterations between (40) for P-step and (45) for O-step until the difference between successive transmit power is smaller than a pre-defined threshold.

Suppose the O-step is solved using the descent solution from *Property 5*. At the  $i$ th iteration, if the O-step solution does not activate the  $k$ th constraint of (1), the  $(i+1)$ th P-step will enforce  $q_k^{[i]} > \gamma_k$ , which ensures the  $(i+1)$ th O-step solution from *Property 5* being a strictly descent solution. According to *Proposition 1*, the iterative procedure converges, and the limit solution activates all constraints of (1). Therefore, with any feasible solution of (1) as initialization, the iterative procedure with descent solution at O-step guarantees monotonic transmit power improvement and tight outage control under any continuous channel uncertainty.

**Remark 3:** The comparisons between the proposed iterative procedure and the bisection calibration [17] are given as follows. From the theoretical perspective, with a feasible solution as initialization, the proposed iterative procedure is applicable to any continuous channel uncertainty with guaranteed tight outage realization. However, since the parameter  $\lambda_k$  in [17] needs to be obtained from a nonconvex optimization or a nonlinear equation, the uniqueness or existence of  $\lambda_k$  is not guaranteed and the bisection in  $[0, \lambda_k]$  is not guaranteed to be implementable. From the computational complexity perspective, the subproblem in [17] is a semidefinite programming problem (for uniform channel uncertainty) with  $K$  positive semidefinite constraints of  $(N+1)(K+2)$  dimension or a second order cone programming problem (for Gaussian channel uncertainty). The corresponding complexity orders in each interior-point iteration are  $\mathcal{O}(N^4 K^5)$  and  $\mathcal{O}(N^3 K^4)$  [26], respectively. On the other hand, for the proposed method, the complexity is  $\mathcal{O}(N^6 K^3/4)$  in each interior-point iteration if SDR is used in the O-step, and  $\mathcal{O}(NK)$  if the descent method is used in the O-step. Therefore, the proposed method with descent solution in the O-step has the least complexity, while the complexity of the proposed method with SDR O-step lies between the two bisection calibration methods.

## V. SIMULATION RESULTS AND DISCUSSIONS

In this section, the performance of the iterative procedure is illustrated under different channel uncertainties. The downlink channel for each user is modeled as  $\mathbf{h}_k = \mathbf{R}_t^{\frac{1}{2}} \mathbf{h}_w$ , where the elements of  $\mathbf{h}_w$  are standard complex Gaussian variables, and the channel correlation matrix is  $[\mathbf{R}_t]_{ij} = \rho_t^{|i-j|}$  with correlation coefficient  $\rho_t = 0.2$ . The BS is equipped with four antennas, and there are two active users unless stated otherwise. The variance of the complex Gaussian noise at every user is  $\delta_k^2 = 0.01$ . The SINR requirement for the second user is fixed as  $10 \log_{10}(\gamma_2) = 5$  (dB) and  $p_2 = 10\%$ , while that for the first user is specified in the figures presented below. Except for Fig. 1, each point in the figures is an average performance under 500 independent feasible channel realizations. The proposed iterative procedure is initialized from the convex

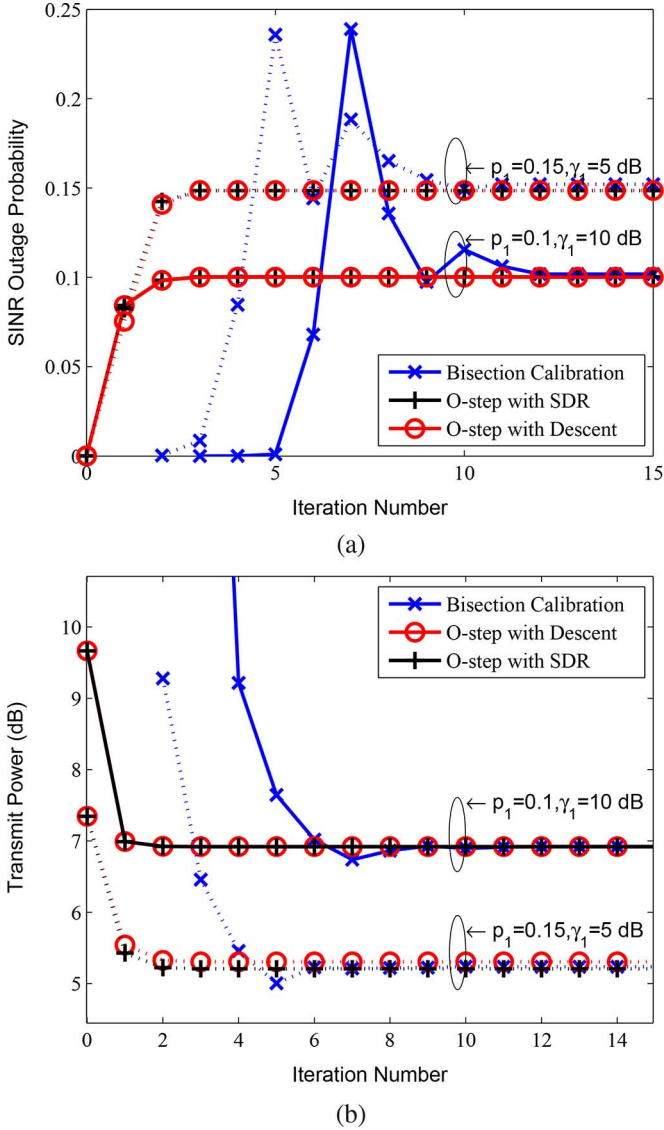


Fig. 1. The convergence performance of the proposed iterative procedure and bisection calibration in a single channel realization.

problem (15) unless stated otherwise, and relative power difference  $\|\mathbf{W}^{[i]}\|_F^2 - \|\mathbf{W}^{[i+1]}\|_F^2 / \|\mathbf{W}^{[i]}\|_F^2 \leq 10^{-2}$  is used to terminate the iterative procedure. The bisection accuracy in finding  $q_k^{[i]}$  is  $10^{-4}$ , and the bisection accuracy in finding the saddle-point is  $10^{-8}$ . For fair comparison with other existing beamforming schemes, under a particular QoS setting, the generated channel realizations should be feasible for all methods under consideration.

#### A. Gaussian Estimation Error

With the training sequence from the  $k$ th user being  $[s_1, \dots, s_L]$ , the received signal at BS is  $\mathbf{y}_k = \mathbf{S}\mathbf{h}_k + \mathbf{n}_k$  with  $\mathbf{S} = [s_1\mathbf{I}_N, \dots, s_L\mathbf{I}_N]^T$ . By using the linear minimum mean square error channel estimator  $\mathbf{E}_k \cdot \mathbf{y}_k = (\mathbf{R}_t^{-T} + \frac{1}{\delta_k^2} \mathbf{S}^H \mathbf{S})^{-1} \mathbf{S}^H / \delta_k^2 \cdot \mathbf{y}_k$ , it is easy to prove that the channel estimation error is zero mean with covariance matrix  $\mathbf{\Sigma}_k = (\mathbf{R}_t^{-T} + \frac{1}{\delta_k^2} \mathbf{S}^H \mathbf{S})^{-1}$  [29]. In the following, the uplink training-to-noise ratio  $\sum_{l=1}^L |s_l|^2 / \delta_k^2$  is set to be 20 dB.

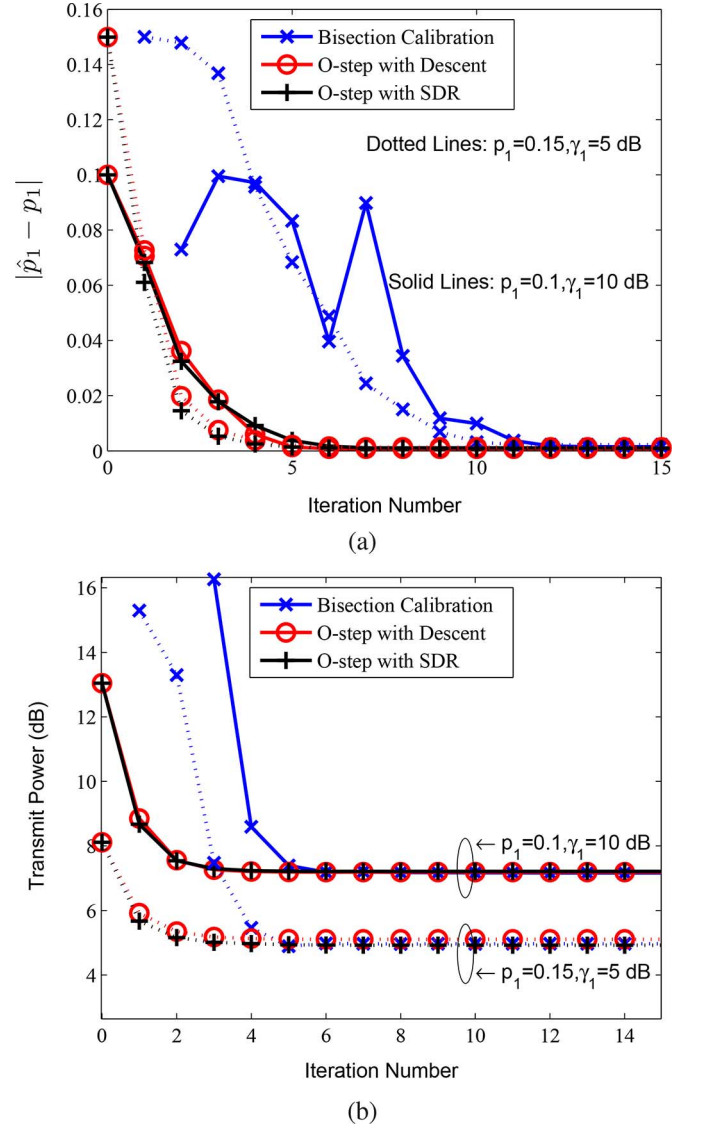


Fig. 2. The convergence performance of the proposed iterative procedure and bisection calibration in multiple channel realizations.

The convergence performances of the proposed iterative procedure and the bisection calibration [17] are compared with a single channel realization in Fig. 1. From Fig. 1(a), it is clear that the realized outage probability of the proposed iterative procedure monotonically increases toward the outage target, while that of the bisection calibration oscillates around the outage target. The same convergence behaviors in transmit power can also be observed in Fig. 1(b). The reason is that the proposed iterative procedure finds the largest feasible subset at P-step and descends at O-step, which makes the transmit power monotonically decrease and the realized outage probability monotonically increase toward the outage target. On the other hand, the bisection calibration only calibrates the outage probability toward the outage target, which causes the oscillation phenomenon in the realized outage and transmit power.

In order to compare the convergence rate of the proposed method and the bisection calibration method, Fig. 2(a) shows the  $|\hat{p}_1 - p_1|$  averaged over 500 feasible channel realizations. From Fig. 2(a), it can be seen that the proposed method needs 5

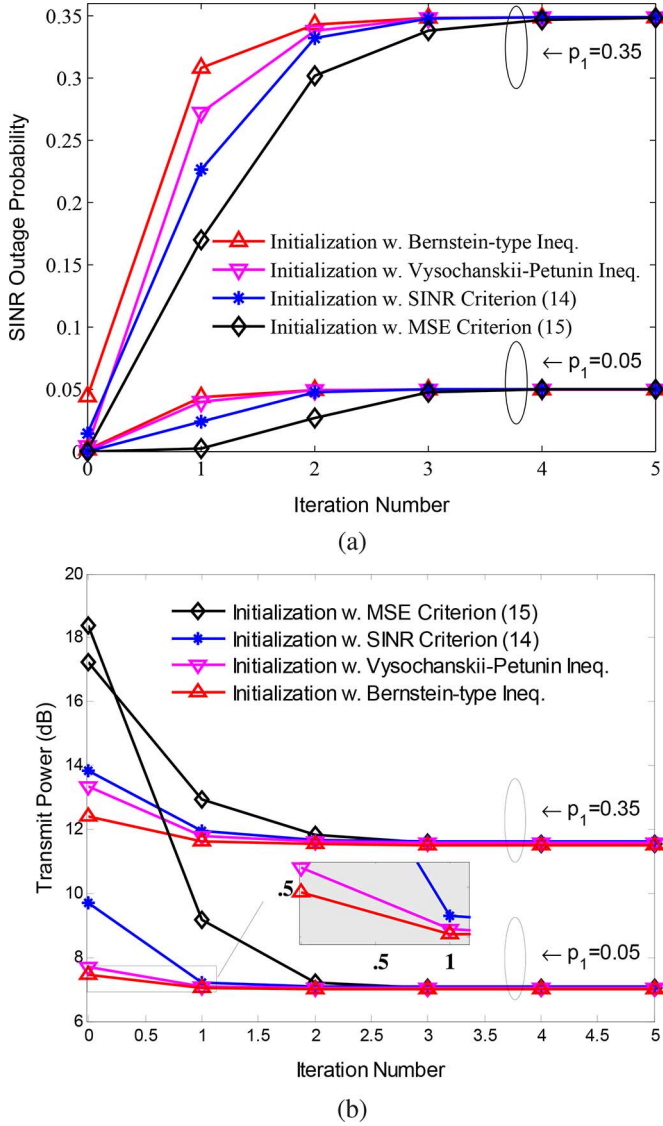


Fig. 3. The convergence performance of the proposed iterative procedure with different safe approximations as initialization.

iterations to converge, while the bisection calibration needs 10 iterations due to the oscillating behavior during convergence. Furthermore, Fig. 2(b) shows the convergence of transmit power for the two methods. It is clear that the proposed iterative procedure converges faster than the bisection calibration method.

From Figs. 1 and 2, it can be noticed that the proposed iterative procedure with SDR and descent solution (*Property 5*) in O-step exhibit similar transmit power performance, which can be explained as follows. Although the safe approximation (15) is a restricted beamforming problem, the beamformers align well with their own channel information and avoid most of the interference. Once the feasible solution is found, the beamforming directions are relatively good directions, and subsequently in every O-step, the solution is significantly affected by the power allocation rather than the beamforming direction adjustment. Therefore, the final transmit powers of SDR and descent solution based O-step are similar. Due to its simplicity and good

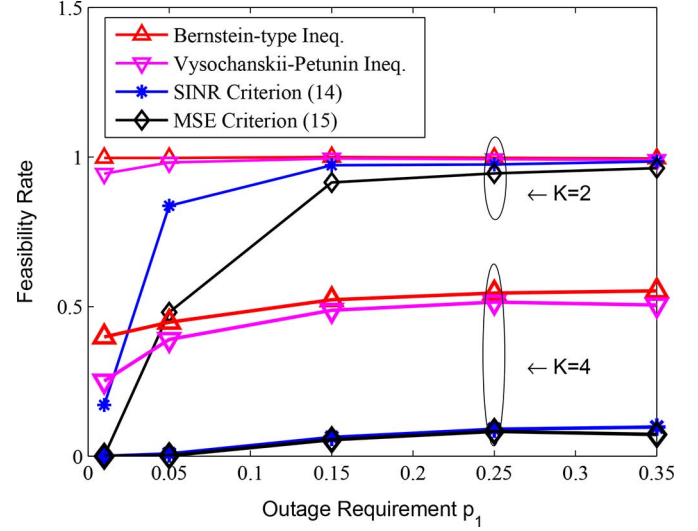


Fig. 4. The feasible rate of different safe approximations under Gaussian channel uncertainty.

performance, only descent solution in O-step is implemented for the rest of the paper.

Next, the performance of the iterative procedure with the proposed initializations, Bernstein-type inequality initialization [14], and the Vysochanskii-Petunin inequality initialization [13] are compared in Fig. 3 under different outage probability targets with  $10 \log_{10}(\gamma_1) = 15$  (dB). It can be seen from Fig. 3(a) that converged solutions with different initializations all realize the outage probability targets tightly. As a result of the tight outage probability control, the proposed iterative procedure reduces transmit power for all initializations as shown in Fig. 3(b). In particular, 0.5 to 1 dB transmit power is saved from the Bernstein-type inequality initialization, and 0.6 to 2 dB transmit power is saved from the Vysochanskii-Petunin inequality initialization. Notice that since the feasible channels are determined by the worst performing MSE criterion (15), the feasible channels for the setting  $p_1 = 0.05$  have higher average channel gain than that of  $p_1 = 0.35$ . Thus, the former setting in Fig. 3(b) shows less transmit power than the latter setting.

In Fig. 4, the feasibility rates of different safe approximation schemes under outage range  $p_1 \in [1\%, 35\%]$  and  $10 \log_{10}(\gamma_1) = 10$  (dB) are shown. The scenario with  $K = 4$  is also included in Fig. 4, and the SINR requirements of the third and fourth users are the same as that of the second user. Without utilizing the Gaussian density information, the proposed safe approximation in (14) and (15) do not perform as well as other existing schemes. However, (14) and (15) are applicable to general uncertainty distributions, which have no alternative solution at this moment. Therefore, the proposed safe approximations (14) and (15) are valuable in terms of generality, and they become indispensable in the next subsection, where mixed estimation and quantization error is considered.

#### B. Mixed Estimation and Quantization Error

For downlink beamforming in frequency-division duplexing (FDD) system, the channel estimation has to be done in the

downlink direction, and then feedback to the BS. Since the feedback cannot afford infinite precision, in addition to the channel estimation error, quantization error also exists. In particular, for channel estimation, the estimator is still  $\mathbf{E}_k \mathbf{y}_k$  but now  $\mathbf{y}_k$  is a  $L \times 1$  vector received at the  $k$ th user, and  $\mathbf{S}$  is a  $L \times N$  matrix with each column represents a training sequence from one BS antenna. For simplicity, we consider orthogonal training with  $\mathbf{S} = \mathbf{I}_N$ . Furthermore, for simple illustration, the scalar quantization [30] in every dimension of the estimated channel is used in this subsection. More specifically, for a random variable distributed as  $\mathcal{N}(0, 1)$ , the maximum entropy quantization criterion [31] leads to the codebook  $\mathcal{C} = \{\Phi^{-1}(1/2^{\xi+1} + i/2^{\xi}) | i = 0, 1, \dots, 2^{\xi} - 1\}$ , where  $\xi$  is the number of quantization bits and  $\Phi^{-1}(\cdot)$  is the inverse CDF of the standard normal distribution. Since the estimated channel  $\mathbf{E}_k \cdot \mathbf{y}_k$  is distributed as  $\mathcal{CN}(\mathbf{0}, \Psi_k)$  with  $\Psi_k = \mathbf{E}_k(\mathbf{S} \mathbf{r}_t \mathbf{S}^H + \delta_k^2 \mathbf{I}_N) \mathbf{E}_k^H$ , the mixed estimation and quantization error is

$$\mathbf{x}_k = \mathbf{h}_k - (\Psi_k/2)^{\frac{1}{2}} Q \left( (\Psi_k/2)^{-\frac{1}{2}} \mathbf{E}_k \cdot \mathbf{y}_k, \mathcal{C} \right), \quad (49)$$

where  $Q(\cdot, \mathcal{C})$  quantizes the real and imaginary parts of each element of the input vector, and  $(\Psi_k/2)^{-\frac{1}{2}}$  is for whitening the estimated channel such that its covariance matrix becomes  $2\mathbf{I}_N$ . For a given channel correlation and noise statistics, samples of  $\{\mathbf{x}_k[i]\}_{i=1}^{N_s}$  can be generated offline, and an empirical distribution can be obtained. In order to guarantee probability evaluation accuracy within 1% from the outage target with reliability 99.999%,  $N_s = 2^{16}$  samples of  $\mathbf{x}_k[i]$  are generated. Notice that the support of the mixed error is theoretically unbounded, so we have  $\mathbf{A}_k = \mathbf{I}_N$  and  $r_k \rightarrow +\infty$ . Since the convergence behavior of the proposed algorithm under mixed error is similar to that of Gaussian case, the convergence figures are not repeated here. In the following, only the performance of the initialization using (15) and that of the converged iterative solution are reported.

In Fig. 5, the performance of the proposed iterative procedure is illustrated under different quantization bits, with the first user's SINR requirement being  $10 \log_{10}(\gamma_1) = 5$  and  $p_1 = 10\%$ . Fig. 5(a) shows that the realized outage probability of the non-robust method ([3], (18.29)) is about 50% owing to the ignorance of the channel uncertainty. Furthermore, although the initialization using (15) is conservative, the iterative procedure always approaches the outage target. Correspondingly, Fig. 5(b) shows that, while the transmit power of the initialization is large, the proposed iterative procedure significantly reduces the transmit power, with the converged power becomes very close to that of the non-robust method. It is interesting to see in Fig. 5(b) that the power gap between the non-robust method and the iterative procedure is almost constant. This is due to the almost constant gap in the SINR outage realizations in Fig. 5(a). Furthermore, for the initialization, owing to its worst case design criterion, its transmit power performance is sensitive to the channel error. This explains why the required transmit power decreases when  $\xi$  increases. On the other hand, notice that the feasible channels are determined by the initialization, and the feasible rate increases as  $\xi$  increases. This leads to a decrease in the average channel gain of the feasible channels,

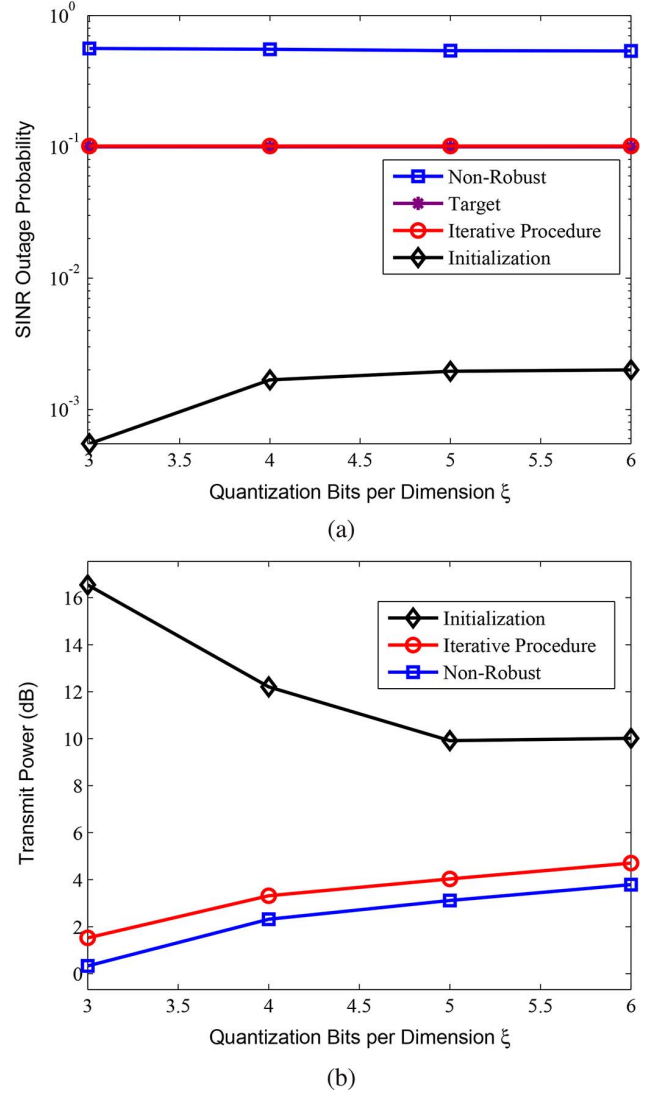


Fig. 5. Performance of the proposed algorithm under different quantization bits.

and thus more transmit power for the non-robust method and the iterative procedure, as  $\xi$  increases. Finally, Fig. 6 shows the performance of the proposed algorithm under different SINR requirements with quantization bits  $\xi = 6$ . It can be seen that, similar to the conclusions of Fig. 5, although the initialization is conservative, the iterative procedure always realizes the outage target tightly, resulting in much reduced transmit power.

## VI. CONCLUSION

In this paper, probabilistic SINR constrained beamforming under channel uncertainties was studied, and a novel method was proposed to achieve tight outage probability control under a large class of bounded or unbounded channel uncertainties. In particular, a systematic method for finding a feasible subset of the probabilistic beamforming problem was proposed based on the moment and support of channel uncertainties. Then, with an iterative procedure, the local structure of the obtained feasible-subset is utilized systematically to explore other feasible subsets of the original problem, leading to tight outage control tailored for the specific channel uncertainty distribution. Simula-



tion results showed that, as a result of tight SINR outage control, transmit power was saved compared to existing approximation based probabilistic beamformers.

#### APPENDIX A

With real Lagrangian multiplier  $\nu_k$  and complex Lagrangian multipliers  $\boldsymbol{\eta}_k, \Xi_k$  (with  $\Xi_k = \Xi_k^H$ ), the Lagrangian of problem (3) in the complex domain is

$$\begin{aligned} \mathcal{L}(f_0(\mathbf{x}_k), \nu_k, \boldsymbol{\eta}_k, \Xi_k) &= \Pr\{\text{SINR}_k(\mathbf{W}, \mathbf{h}_k) \geq \gamma_k\} + \nu_k \left(1 - \int_{\mathbf{x}_k \in \mathcal{X}_k} f_0(\mathbf{x}_k) d\mathbf{x}_k\right) \\ &\quad + \text{Re}\{\boldsymbol{\eta}_k^T\} \cdot \text{Re}\{\mathbf{0} - \mathbb{E}(\mathbf{x}_k)\} \\ &\quad + \text{Im}\{\boldsymbol{\eta}_k^T\} \cdot \text{Im}\{\mathbf{0} - \mathbb{E}(\mathbf{x}_k)\} \\ &\quad + \text{Tr}(\Xi_k^H (\boldsymbol{\Sigma}_k - \mathbb{E}(\mathbf{x}_k \mathbf{x}_k^H))) \end{aligned} \quad (50a)$$

$$\begin{aligned} &= \nu_k + \text{Tr}(\Xi_k^H \boldsymbol{\Sigma}_k) \\ &\quad + \int_{\mathbf{x}_k \in \mathcal{X}_k, \text{SINR}_k(\mathbf{W}, \mathbf{h}_k) \geq \gamma_k} (1 - \nu_k - \text{Re}\{\boldsymbol{\eta}_k^H \mathbf{x}_k\} \\ &\quad - \text{Tr}(\Xi_k^H \mathbf{x}_k \mathbf{x}_k^H)) \cdot f_0(\mathbf{x}_k) d\mathbf{x}_k \\ &\quad + \int_{\mathbf{x}_k \in \mathcal{X}_k, \text{SINR}_k(\mathbf{W}, \mathbf{h}_k) \leq \gamma_k} (0 - \nu_k - \text{Re}\{\boldsymbol{\eta}_k^H \mathbf{x}_k\} \\ &\quad - \text{Tr}(\Xi_k^H \mathbf{x}_k \mathbf{x}_k^H)) \cdot f_0(\mathbf{x}_k) d\mathbf{x}_k. \end{aligned} \quad (50b)$$

With the implicit PDF constraint  $f_0(\mathbf{x}_k) \geq 0$ , the Lagrange dual function of the problem (3) is (51) at the bottom of the page. Comparing the two conditions in (51) that make the dual function equal  $\nu_k + \text{Tr}(\Xi_k^H \boldsymbol{\Sigma}_k)$ , the second condition can be replaced by  $\nu_k + \text{Re}\{\boldsymbol{\eta}_k^H \mathbf{x}_k\} + \text{Tr}(\Xi_k^H \mathbf{x}_k \mathbf{x}_k^H) \leq 1, \forall \mathbf{x}_k \in \mathcal{X}_k$ . Therefore, the dual of problem (3) can be formulated as

$$\begin{aligned} \max_{\nu_k, \boldsymbol{\eta}_k, \Xi_k} \quad & \nu_k + \text{Tr}(\Xi_k^H \boldsymbol{\Sigma}_k) \\ \text{s.t.} \quad & \nu_k + \text{Re}\{\boldsymbol{\eta}_k^H \mathbf{x}_k\} + \text{Tr}(\Xi_k^H \mathbf{x}_k \mathbf{x}_k^H) \leq 0, \\ & \forall \mathbf{x}_k : \begin{cases} \mathbf{x}_k \in \mathcal{X}_k, \text{ and} \\ \text{SINR}_k(\mathbf{W}, \mathbf{h}_k) \leq \gamma_k \end{cases} \\ & \nu_k + \text{Re}\{\boldsymbol{\eta}_k^H \mathbf{x}_k\} \\ & \quad + \text{Tr}(\Xi_k^H \mathbf{x}_k \mathbf{x}_k^H) \leq 1, \quad \forall \mathbf{x}_k \in \mathcal{X}_k \\ & \Xi_k = \Xi_k^H. \end{aligned} \quad (52)$$

By using the expression of  $\mathcal{X}_k = \{\mathbf{x}_k | \mathbf{x}_k^H \mathbf{A}_k \mathbf{x}_k \leq r_k^2\}$  and a quadratic reformulation of  $\text{SINR}_k(\mathbf{W}, \mathbf{h}_k + \mathbf{x}_k) \leq \gamma_k$ , i.e., [see

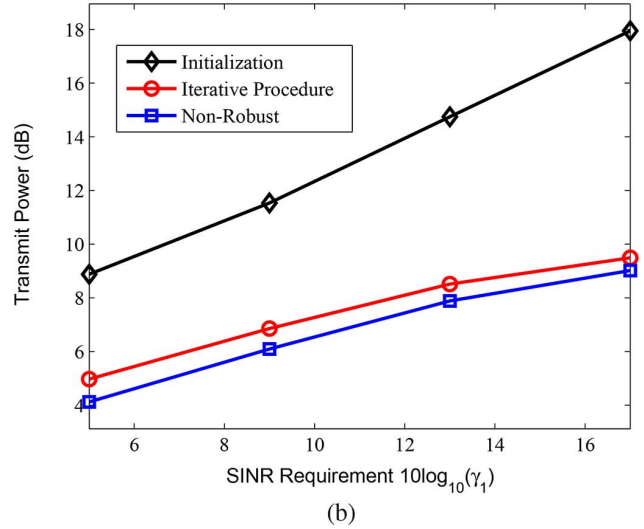
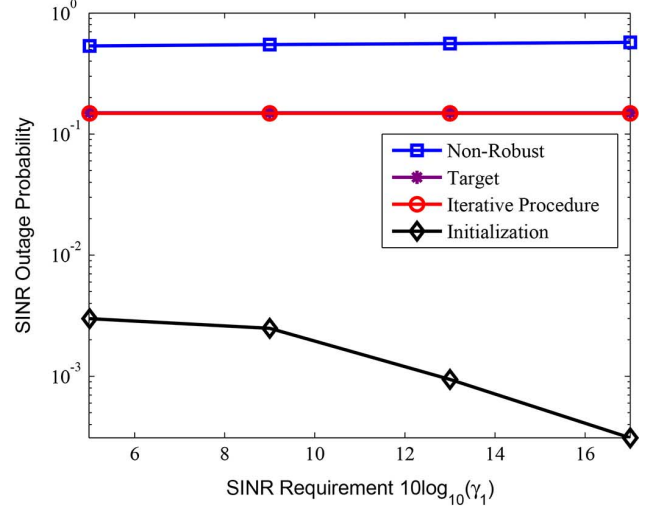


Fig. 6. Performance of the proposed algorithm under different SINR requirements.

the (53) at the bottom of the next page], with  $\mathbf{Q}_k := \mathbf{w}_k \mathbf{w}_k^H$ , (52) can be transformed to the compact quadratic form in (4).

#### APPENDIX B PROOF OF THEOREM 2

According to Theorem 4 in [8], the feasible transceiver solution  $(\mathbf{W}, \{g_k\}_{k=1}^K)$  in  $\text{MSE}_k(\mathbf{W}, g_k, \mathbf{h}_k) \leq 1/(\gamma_k + 1)$  guarantees  $\text{SINR}_k(\mathbf{W}, \mathbf{h}_k) \geq \gamma_k$  with fixed channel realization  $\mathbf{h}_k$ . Therefore, feasible solution  $\mathbf{W}$  in  $\Pr\{\text{MSE}_k(\mathbf{W}, g_k, \mathbf{h}_k) \leq 1/(\gamma_k + 1)\} \geq 1 - p_k$  guarantees  $\Pr\{\text{SINR}_k(\mathbf{W}, \mathbf{h}_k) \geq \gamma_k\} \geq$

$$\begin{aligned} \inf_{f_0(\mathbf{x}_k) \geq 0} \mathcal{L}(f_0(\mathbf{x}_k), \nu_k, \boldsymbol{\eta}_k, \Xi_k) &= \begin{cases} \nu_k + \text{Tr}(\Xi_k^H \boldsymbol{\Sigma}_k) & \text{if } \begin{cases} \nu_k + \text{Re}\{\boldsymbol{\eta}_k^H \mathbf{x}_k\} + \text{Tr}(\Xi_k^H \mathbf{x}_k \mathbf{x}_k^H) \leq 0, \\ \forall \mathbf{x}_k \in \mathcal{X}_k : \text{SINR}_k(\mathbf{W}, \mathbf{h}_k) \leq \gamma_k; \text{ and} \\ \nu_k + \text{Re}\{\boldsymbol{\eta}_k^H \mathbf{x}_k\} + \text{Tr}(\Xi_k^H \mathbf{x}_k \mathbf{x}_k^H) \leq 1, \\ \forall \mathbf{x}_k \in \mathcal{X}_k : \text{SINR}_k(\mathbf{W}, \mathbf{h}_k) \geq \gamma_k \end{cases} \\ -\infty & \text{otherwise.} \end{cases} \end{aligned} \quad (51)$$

$1 - p_k$ , and therefore a feasible subset of problem (1) can be obtained from the feasible set of the following problem

$$\begin{aligned} & \min_{\mathbf{W}, \{\mathbf{g}_k\}_{k=1}^K} \|\mathbf{W}\|_F^2 \\ \text{s.t. } & \Pr \left\{ \text{MSE}_k(\mathbf{W}, g_k, \mathbf{h}_k) \geq \frac{1}{\gamma_k + 1} \right\} \leq p_k, \quad \forall k \in \mathcal{K}. \end{aligned} \quad (54)$$

In order to obtain a tractable feasible subset of (54) for all possible channel uncertainty distribution, by using the moment and support information, an upper bound of  $\Pr\{\text{MSE}_k(\mathbf{W}, g_k, \mathbf{h}_k) \geq 1/(\gamma_k + 1)\}$  is formulated as

$$\begin{aligned} & \sup_{f_0(\mathbf{x}_k)} \Pr \left\{ \text{MSE}_k(\mathbf{W}, g_k, \mathbf{h}_k) \geq \frac{1}{\gamma_k + 1} \right\} \\ \text{s.t. } & \int_{\mathbf{x}_k \in \mathcal{X}_k} f_0(\mathbf{x}_k) d\mathbf{x}_k = 1 \\ & \mathbb{E}(\mathbf{x}_k) = \mathbf{0} \\ & \mathbb{E}(\mathbf{x}_k \mathbf{x}_k^H) = \Sigma_k. \end{aligned} \quad (55)$$

Using a factorized equalizer  $g_k = \frac{1}{b_k} e^{j\theta_k}$  with  $b_k > 0$ , the MSE requirement  $\text{MSE}_k(\mathbf{W}, g_k, \mathbf{h}_k) = \|\frac{1}{b_k} e^{j\theta_k} \mathbf{h}_k^H \mathbf{W} - \mathbf{e}_k^T\|_2^2 + \frac{1}{b_k^2} \delta_k^2 \geq 1/(\gamma_k + 1)$  can be reformulated as  $\mathbf{u}_k^H (\frac{1}{b_k} \ddot{\mathbf{L}}^H(\ddot{\mathbf{W}}, b_k) \ddot{\mathbf{L}}(\ddot{\mathbf{W}}, b_k) - \text{Diag}([0, b_k/(\gamma_k + 1) - \frac{1}{b_k} \delta_k^2])) \mathbf{u}_k \geq 0$ , where  $\mathbf{u}_k := [\mathbf{x}_k^T, 1]^T$ ,  $\ddot{\mathbf{L}}(\ddot{\mathbf{W}}, b_k) := [\ddot{\mathbf{W}}^H, \ddot{\mathbf{W}}^H \hat{\mathbf{h}}_k - b_k \mathbf{e}_k]$ ,  $\ddot{\mathbf{W}} := \mathbf{W} \circ [e^{j\theta_1} \mathbf{1}, \dots, e^{j\theta_K} \mathbf{1}]$  with  $\circ$  stands for the Hadamard product, and the  $K \times 1$  vector  $\mathbf{e}_k = [0, \dots, 1, 0, \dots, 0]^T$  with the element 1 located at the  $k$ th position. With similar derivations in *Appendix A*, the dual of problem (55) is [see (56) at the bottom of the page], where  $\bar{\Sigma}_k := \begin{bmatrix} \Sigma_k & \mathbf{0} \\ \mathbf{0}^T & 1 \end{bmatrix}$  was first defined after (4). Since the primal problem (55) is a concave maximizing problem and is feasible, (55) is upper bounded by its dual problem (56). Therefore, putting (56) into (54) to replace  $\Pr\{\text{MSE}_k(\mathbf{W}, g_k, \mathbf{h}_k) \geq \frac{1}{\gamma_k + 1}\}$ , and noticing that  $\alpha_k \neq 0$  can be proved similar to that in *Property 1*, a feasible subset of (54) can be obtained from the feasible set of the problem in (57) at the bottom of the page, with new variables  $\tilde{\alpha}_k := 1/\alpha_k$ ,  $\tilde{\mathbf{Z}}_k := \tilde{\alpha}_k \mathbf{Z}_k$ ,  $\tilde{\beta}_k := \tilde{\alpha}_k \beta_k$  and  $\tilde{\kappa}_k := \tilde{\alpha}_k \kappa_k$ .

$$[\mathbf{x}_k^H, 1] \begin{bmatrix} \mathbf{Q}_k - \gamma_k \sum_{l=1, l \neq k}^K \mathbf{Q}_l & \left( \mathbf{Q}_k - \gamma_k \sum_{l=1, l \neq k}^K \mathbf{Q}_l \right) \hat{\mathbf{h}}_k \\ \hat{\mathbf{h}}_k^H \left( \mathbf{Q}_k - \gamma_k \sum_{l=1, l \neq k}^K \mathbf{Q}_l \right) & \hat{\mathbf{h}}_k^H \left( \mathbf{Q}_k - \gamma_k \sum_{l=1, l \neq k}^K \mathbf{Q}_l \right) \hat{\mathbf{h}}_k - \gamma_k \delta_k^2 \end{bmatrix} \begin{bmatrix} \mathbf{x}_k \\ 1 \end{bmatrix} \leq 0 \quad (53)$$

$$\begin{aligned} & \min_{b_k, \mathbf{Z}_k, \alpha_k, \beta_k, \kappa_k} \text{Tr}(\mathbf{Z}_k \bar{\Sigma}_k) \\ \text{s.t. } & \mathbf{Z}_k + \text{Diag}([0, -1]) + \beta_k \begin{bmatrix} \mathbf{A}_k & \mathbf{0} \\ \mathbf{0}^T & -r_k^2 \end{bmatrix} - \alpha_k \left( \frac{1}{b_k} \ddot{\mathbf{L}}^H(\ddot{\mathbf{W}}, b_k) \ddot{\mathbf{L}}(\ddot{\mathbf{W}}, b_k) - \text{Diag} \left( \begin{bmatrix} 0, \frac{b_k}{\gamma_k + 1} - \frac{1}{b_k} \delta_k^2 \end{bmatrix} \right) \right) \succeq 0, \\ & \mathbf{Z}_k + \kappa_k \begin{bmatrix} \mathbf{A}_k & \mathbf{0} \\ \mathbf{0}^T & -r_k^2 \end{bmatrix} \succeq 0, \\ & \mathbf{Z}_k = \mathbf{Z}_k^H, \alpha_k \geq 0, \beta_k \geq 0, \kappa_k \geq 0, b_k > 0, \end{aligned} \quad (56)$$

$$\begin{aligned} & \min_{\ddot{\mathbf{W}}, \{\tilde{\mathbf{Z}}_k, \tilde{\alpha}_k, \tilde{\beta}_k, \tilde{\kappa}_k\}_{k=1}^K} \|\ddot{\mathbf{W}}\|_F^2 \\ \text{s.t. } & \text{Tr}(\tilde{\mathbf{Z}}_k \bar{\Sigma}_k) \leq p_k \tilde{\alpha}_k, \quad \forall k \in \mathcal{K} \\ & \tilde{\mathbf{Z}}_k + \text{Diag} \left( \begin{bmatrix} 0, \frac{b_k}{\gamma_k + 1} - \frac{1}{b_k} \delta_k^2 - \tilde{\alpha}_k \end{bmatrix} \right) + \tilde{\beta}_k \begin{bmatrix} \mathbf{A}_k & \mathbf{0} \\ \mathbf{0}^T & -r_k^2 \end{bmatrix} \\ & \quad - \frac{1}{b_k} \ddot{\mathbf{L}}^H(\ddot{\mathbf{W}}, b_k) \ddot{\mathbf{L}}(\ddot{\mathbf{W}}, b_k) \succeq 0, \quad \forall k \in \mathcal{K} \\ & \tilde{\mathbf{Z}}_k + \tilde{\kappa}_k \begin{bmatrix} \mathbf{A}_k & \mathbf{0} \\ \mathbf{0}^T & -r_k^2 \end{bmatrix} \succeq 0, \quad \forall k \in \mathcal{K} \\ & \tilde{\mathbf{Z}}_k = \tilde{\mathbf{Z}}_k^H, \quad \tilde{\alpha}_k > 0, \quad \tilde{\beta}_k \geq 0, \quad \tilde{\kappa}_k \geq 0, \quad b_k > 0, \quad \forall k \in \mathcal{K}, \end{aligned} \quad (57)$$

Introducing a slack variable  $c_k$  with  $c_k \geq \frac{1}{b_k}$  and using Schur complement, (57) becomes (58), shown at the bottom of the page. Since the constraints  $c_k \geq \frac{1}{b_k}$  and  $b_k > 0$  are equivalent to the second order cone programming form  $\|[2, b_k - c_k]\|_2 \leq b_k + c_k$  and  $b_k > 0$ , problem (58) can be transformed to the convex problem (15).

#### APPENDIX C

##### PROOF OF PROPERTY 4

According to the definition  $\mathcal{H}_k(\bar{\mathbf{W}}) := \{\mathbf{h}_k | \mathbf{h}_k \in \{\hat{\mathbf{h}}_k + \mathcal{X}_k\}, \text{SINR}_k(\bar{\mathbf{W}}, \mathbf{h}_k) \geq \gamma_k\}$ , the constraint  $\text{SINR}_k(\bar{\mathbf{W}}, \mathbf{h}_k) \geq \gamma_k$  is automatically satisfied for all  $\mathbf{h}_k \in \mathcal{H}_k(\bar{\mathbf{W}})$ . Combining with the *Definition 1*  $\mathcal{W}(\bar{\mathbf{W}}) := \{\mathbf{W} | \{\text{SINR}_k(\mathbf{W}, \mathbf{h}_k) \geq \gamma_k, \forall \mathbf{h}_k \in \mathcal{H}_k(\bar{\mathbf{W}})\}_{k=1}^K\}$ , then we directly have  $\bar{\mathbf{W}} \in \mathcal{W}(\bar{\mathbf{W}})$ . Furthermore, any  $\mathbf{W}$  in  $\mathcal{W}(\bar{\mathbf{W}})$  satisfies the condition in (59)–(61) at the bottom of the page, where  $\mathbb{I}(\cdot)$  is an indicator function. Therefore, any  $\mathbf{W}$  in  $\mathcal{W}(\bar{\mathbf{W}})$  is a feasible solution of (1), i.e.,  $\mathcal{W}(\bar{\mathbf{W}}) \subseteq \mathcal{W}_0$ .

#### APPENDIX D

With  $\mathbf{h}_k \sim \mathcal{CN}(\hat{\mathbf{h}}_k, \Sigma_k)$ , we first write  $d_k$  as (62) at the bottom of the page, where the normalized

channel  $\ddot{\mathbf{h}}_k \sim \mathcal{CN}((\Sigma_k/2)^{-\frac{1}{2}}\hat{\mathbf{h}}_k, 2\mathbf{I}_N)$ . Since  $\mathbf{B}_k$  is a Hermitian matrix, the eigenvalue decomposition  $\mathbf{B}_k = \mathbf{V}_k \text{Diag}([\lambda_1, \dots, \lambda_N])\mathbf{V}_k^H$  with the descending real eigenvalues  $\{\lambda_j\}_{j=1}^N$  and orthogonal eigenvectors in  $\mathbf{V}_k$  leads to  $\mathbf{h}_k^H(\mathbf{Q}_k^{[i]} - q_k^{[i]} \sum_{l=1, l \neq k}^K \mathbf{Q}_l^{[i]})\mathbf{h}_k = \ddot{\mathbf{h}}_k^H \mathbf{V}_k \text{Diag}([\lambda_1, \dots, \lambda_N])\mathbf{V}_k^H \ddot{\mathbf{h}}_k$ . Since  $\mathbf{V}_k^H \ddot{\mathbf{h}}_k \sim \mathcal{CN}(\mathbf{V}_k^H (\Sigma_k/2)^{-\frac{1}{2}}\hat{\mathbf{h}}_k, 2\mathbf{I}_N)$ , the statistical representation of (62) is a weighted sums of independent noncentral Chi-squared distributions with two degrees of freedom, i.e.,

$$\mathbf{h}_k^H \left( \mathbf{Q}_k^{[i]} - q_k^{[i]} \sum_{l=1, l \neq k}^K \mathbf{Q}_l^{[i]} \right) \mathbf{h}_k \sim \sum_{j=1}^N \lambda_j \chi_{(|m_j|^2, 2)}, \quad (63)$$

where  $m_j$  is the  $j$ th element of the complex vector  $\mathbf{V}_k^H (\Sigma_k/2)^{-\frac{1}{2}}\hat{\mathbf{h}}_k$ . With the moment-generating function of  $\chi_{(|m_j|^2, 2)}$  being  $\exp(|m_j|^2 t / (1 - 2t)) / (1 - 2t)$  with domain  $2t < 1$ , the CGF of (63) is

$$g(t) = \sum_{j=1}^N \left( \frac{|m_j|^2 \lambda_j t}{1 - 2\lambda_j t} - \ln(1 - 2\lambda_j t) \right), \quad (64)$$

---


$$\begin{aligned} & \min_{\check{\mathbf{W}}, \{b_k, c_k, \tilde{\mathbf{Z}}_k, \tilde{\alpha}_k, \tilde{\beta}_k, \tilde{\kappa}_k\}_{k=1}^K} \|\check{\mathbf{W}}\|_F^2 \\ & \text{s.t. } \text{Tr}(\tilde{\mathbf{Z}}_k \bar{\Sigma}_k) \leq p_k \tilde{\alpha}_k, \quad \forall k \in \mathcal{K} \\ & \begin{bmatrix} \tilde{\mathbf{Z}}_k + \text{Diag}([0, \frac{b_k}{\gamma_k + 1} - c_k \delta_k^2 - \tilde{\alpha}_k]) + \tilde{\beta}_k \begin{bmatrix} \mathbf{A}_k & \mathbf{0} \\ \mathbf{0}^T & -r_k^2 \end{bmatrix}, & \tilde{\mathbf{L}}^H(\check{\mathbf{W}}, b_k) \\ \tilde{\mathbf{L}}(\check{\mathbf{W}}, b_k) & b_k \mathbf{I}_K \end{bmatrix} \succeq 0, \forall k \in \mathcal{K} \\ & \tilde{\mathbf{Z}}_k + \tilde{\kappa}_k \begin{bmatrix} \mathbf{A}_k & \mathbf{0} \\ \mathbf{0}^T & -r_k^2 \end{bmatrix} \succeq 0, \quad c_k \geq \frac{1}{b_k}, \quad \forall k \in \mathcal{K} \\ & \tilde{\mathbf{Z}}_k = \tilde{\mathbf{Z}}_k^H, \tilde{\alpha}_k > 0, \tilde{\beta}_k \geq 0, \tilde{\kappa}_k \geq 0, b_k > 0, \quad \forall k \in \mathcal{K}. \end{aligned} \quad (58)$$


---

$$\begin{aligned} & \Pr\{\text{SINR}_k(\mathbf{W}, \mathbf{h}_k) \geq \gamma_k\} \\ & = \int_{\mathbf{h}_k \in \mathcal{H}_k(\bar{\mathbf{W}})} \mathbb{I}(\text{SINR}_k(\mathbf{W}, \mathbf{h}_k) \geq \gamma_k) f(\mathbf{h}_k) d\mathbf{h}_k \\ & \quad + \int_{\mathbf{h}_k \in \{\{\hat{\mathbf{h}}_k + \mathcal{X}_k\} \setminus \mathcal{H}_k(\bar{\mathbf{W}})\}} \mathbb{I}(\text{SINR}_k(\mathbf{W}, \mathbf{h}_k) \geq \gamma_k) f(\mathbf{h}_k) d\mathbf{h}_k \end{aligned} \quad (59)$$

$$= \underbrace{\int_{\mathbf{h}_k \in \mathcal{H}_k(\bar{\mathbf{W}})} f(\mathbf{h}_k) d\mathbf{h}_k}_{=\Pr\{\text{SINR}_k(\bar{\mathbf{W}}, \mathbf{h}_k) \geq \gamma_k\} \geq 1 - p_k} + \underbrace{\int_{\mathbf{h}_k \in \{\{\hat{\mathbf{h}}_k + \mathcal{X}_k\} \setminus \mathcal{H}_k(\bar{\mathbf{W}})\}} \mathbb{I}(\text{SINR}_k(\mathbf{W}, \mathbf{h}_k) \geq \gamma_k) f(\mathbf{h}_k) d\mathbf{h}_k}_{\geq 0} \quad (60)$$

$$\geq 1 - p_k, \quad \forall k \in \mathcal{K}, \quad (61)$$


---

$$\mathbf{h}_k^H (\mathbf{Q}_k^{[i]} - q_k^{[i]} \sum_{l=1, l \neq k}^K \mathbf{Q}_l^{[i]}) \mathbf{h}_k = \underbrace{\ddot{\mathbf{h}}_k^H \left( (\Sigma_k/2)^{\frac{1}{2}} \right)^H \left( \mathbf{Q}_k^{[i]} - q_k^{[i]} \sum_{l=1, l \neq k}^K \mathbf{Q}_l^{[i]} \right) (\Sigma_k/2)^{\frac{1}{2}} \ddot{\mathbf{h}}_k}_{\mathbf{B}_k}, \quad (62)$$

with the domain of  $g(t)$  satisfying  $2\lambda_j t < 1, \forall j \in \{1, \dots, N\}$ . Since  $K$  independent data streams are transmitted simultaneously,  $\mathbf{w}_1^{[i]}, \dots, \mathbf{w}_K^{[i]}$  have to be linearly independent. According to the Sylvester's law of inertia, the largest and smallest eigenvalue of  $\mathbf{B}_k = ((\Sigma_k/2)^{\frac{1}{2}})^H [\mathbf{w}_1^{[i]}, \dots, \mathbf{w}_K^{[i]}] \text{Diag}([1, -q_k^{[i]}, \dots, -q_k^{[i]}]) [\mathbf{w}_1^{[i]}, \dots, \mathbf{w}_K^{[i]}]^H (\Sigma_k/2)^{\frac{1}{2}}$  are positive and negative, respectively. Therefore, the results  $\lambda_1 > 0$  and  $\lambda_N < 0$  make the domain of  $g(t)$  being  $(1/(2\lambda_N), 1/(2\lambda_1))$ .

#### APPENDIX E PROOF OF THEOREM 3

First, the SDR formulation of (45) is

$$\begin{aligned} \min_{\{\mathbf{Q}_k, \alpha_k, \beta_k\}_{k=1}^K} & \sum_{k=1}^K \text{Tr}(\mathbf{Q}_k) \\ \text{s.t. } & \mathbf{Q}_k \succeq 0, \begin{bmatrix} \mathbf{Q}_k - \gamma_k \sum_{l=1, l \neq k}^K \mathbf{Q}_l & \mathbf{0} \\ \mathbf{0}^T & -\gamma_k \delta_k^2 \end{bmatrix} \\ & - \alpha_k \begin{bmatrix} \mathbf{Q}_k^{[i]} - q_k^{[i]} \sum_{l=1, l \neq k}^K \mathbf{Q}_l^{[i]} & \mathbf{0} \\ \mathbf{0}^T & -q_k^{[i]} \delta_k^2 \end{bmatrix} \\ & + \beta_k \begin{bmatrix} \mathbf{A}_k & -\mathbf{A}_k \hat{\mathbf{h}}_k \\ -\hat{\mathbf{h}}_k^H \mathbf{A}_k & \hat{\mathbf{h}}_k^H \mathbf{A}_k \hat{\mathbf{h}}_k - r_k^2 \end{bmatrix} \succeq 0, \quad \forall k \in \mathcal{K} \\ & \alpha_k \geq 0, \beta_k \geq 0, \quad \forall k \in \mathcal{K}. \end{aligned} \quad (65)$$

With Lagrange multipliers  $\mathbf{P}_k \succeq 0, \begin{bmatrix} \mathbf{Z}_k & \mathbf{u}_k \\ \mathbf{u}_k^H & v_k \end{bmatrix} \succeq 0, \mu_k \geq 0, \eta_k \geq 0$ , the dual of problem (65) is

$$\begin{aligned} \max_{\{\mathbf{Z}_k, \mathbf{u}_k, v_k\}_{k=1}^K} & \sum_{k=1}^K v_k \gamma_k \delta_k^2 \\ \text{s.t. } & \mathbf{P}_k = \mathbf{I} + \left( \sum_{j=1, j \neq k}^K \gamma_j \mathbf{Z}_j - \mathbf{Z}_k \right) \succeq 0, \quad \forall k \in \mathcal{K} \\ & \begin{bmatrix} \mathbf{Z}_k & \mathbf{u}_k \\ \mathbf{u}_k^H & v_k \end{bmatrix} \succeq 0, \quad \forall k \in \mathcal{K} \\ & \mu_k = \text{Tr} \left( \mathbf{Z}_k \left( \mathbf{Q}_k^{[i]} - q_k^{[i]} \sum_{j=1, j \neq k}^K \mathbf{Q}_j^{[i]} \right) \right) - v_k q_k^{[i]} \delta_k^2 \geq 0, \quad \forall k \in \mathcal{K} \\ & \eta_k = 2\text{Re}\{\mathbf{u}_k^H \mathbf{A}_k \hat{\mathbf{h}}_k\} - \text{Tr}(\mathbf{A}_k \mathbf{Z}_k) \\ & \quad - v_k (\hat{\mathbf{h}}_k^H \mathbf{A}_k \hat{\mathbf{h}}_k - r_k^2) \geq 0, \quad \forall k \in \mathcal{K}. \end{aligned} \quad (66)$$

We can establish two facts about (66).

- 1) Notice that the primal problem (65) is always feasible with solution  $\{\mathbf{Q}_k = \mathbf{Q}_k^{[i]}, \alpha_k = 1, \beta_k = 0\}_{k=1}^K$ . Furthermore, simple substitution verifies that  $\left\{ \begin{bmatrix} \mathbf{Z}_k & \mathbf{u}_k \\ \mathbf{u}_k^H & v_k \end{bmatrix} = \mathbf{0} \right\}_{k=1}^K$  is a feasible solution for the dual problem (66). Therefore, weak duality holds between (65) and (66), and the lower bounded primal problem (65) makes the dual problem (66) be upper bounded by a positive number. Thus the nonnegative  $\{v_k\}_{k=1}^K$  in the objective function of (66) are upper and lower bounded. Since the bounded  $v_k$  comes from the restriction  $v_k \leq \text{Tr} \left( \mathbf{Z}_k (\mathbf{Q}_k^{[i]} - q_k^{[i]} \sum_{j=1, j \neq k}^K \mathbf{Q}_j^{[i]}) \right) / (q_k^{[i]} \delta_k^2)$  or

$$v_k \leq (2\text{Re}\{\mathbf{u}_k^H \mathbf{A}_k \hat{\mathbf{h}}_k\} - \text{Tr}(\mathbf{A}_k \mathbf{Z}_k)) / (\hat{\mathbf{h}}_k^H \mathbf{A}_k \hat{\mathbf{h}}_k - r_k^2) \leq (2\text{Re}\{\mathbf{u}_k^H \mathbf{A}_k \hat{\mathbf{h}}_k\} - \mathbf{u}_k^H \mathbf{A}_k \mathbf{u}_k / v_k) / (\hat{\mathbf{h}}_k^H \mathbf{A}_k \hat{\mathbf{h}}_k - r_k^2),$$

together with  $\begin{bmatrix} \mathbf{Z}_k & \mathbf{u}_k \\ \mathbf{u}_k^H & v_k \end{bmatrix} \succeq 0$ ,  $\mathbf{Z}_k$  and  $\mathbf{u}_k$  are bounded.

Therefore, the compact feasible set of (66) makes its optimal solution attainable.

- 2) In the dual problem (66), the zero solution  $\{\mathbf{Z}_k = \mathbf{0}, \mathbf{u}_k = \mathbf{0}, v_k = 0\}_{k=1}^K$  can be perturbed such that  $\mu_k > 0, \eta_k > 0, \forall k \in \mathcal{K}$  and make the positive semidefinite constraints strictly feasible owing to the identity matrix.

Combining the two facts above makes the strong duality hold between convex problems (65) and (66).

Owing to the strong duality between (65) and (66), the complementary slackness makes  $\text{rank}(\mathbf{P}_k) + \text{rank}(\mathbf{Q}_k) \leq N, \forall k \in \mathcal{K}$ . Furthermore, since the zero matrices are not the optimal solutions in (65) and (66), the optimal primal solutions satisfy  $\text{rank}(\mathbf{P}_k) \in [1, N-1]$ . Let  $\mathbf{C}_k := \sum_{j=1, j \neq k}^K \gamma_j \mathbf{Z}_j - \mathbf{Z}_k$ , and put  $\mathbf{Z}_k = \sum_{j=1, j \neq k}^K \gamma_j \mathbf{Z}_j - \mathbf{C}_k$  into  $\mathbf{C}_l = \sum_{j=1, j \neq l, j \neq k}^K \gamma_j \mathbf{Z}_j + \gamma_k \mathbf{Z}_k - \mathbf{Z}_l$  with  $l \neq k$ , we have

$$\mathbf{C}_l = (\gamma_k \gamma_l - 1) \mathbf{Z}_l + (1 + \gamma_k) \sum_{j=1, j \neq k, j \neq l}^K \gamma_j \mathbf{Z}_j - \gamma_k \mathbf{C}_k. \quad (67)$$

Since  $\gamma_k > 0, \mathbf{Z}_k \succeq 0, \forall k \in \mathcal{K}$  and  $\gamma_k \gamma_l \geq 1$ , the first two terms of (67) are positive semidefinite. Without loss of generality, let  $\mathbf{C}_k$  has  $s_k$  positive eigenvalues and  $N - s_k$  nonpositive eigenvalues. According to the Weyl's inequality [28, p. 274],  $\mathbf{C}_l$  has at least  $N - s_k$  nonnegative eigenvalues. Therefore,  $\text{rank}(\mathbf{P}_k) = \text{rank}(\mathbf{I} + \mathbf{C}_k) \geq s_k$  and  $\text{rank}(\mathbf{P}_l) = \text{rank}(\mathbf{I} + \mathbf{C}_l) \geq N - s_k$ . Owing to  $\text{rank}(\mathbf{P}_k) + \text{rank}(\mathbf{Q}_k) \leq N, \forall k \in \mathcal{K}$ , the optimal primal solutions  $\mathbf{Q}_k$  and  $\mathbf{Q}_l$  satisfy  $\text{rank}(\mathbf{Q}_k) \leq N - s_k$  and  $\text{rank}(\mathbf{Q}_l) \leq s_k$ , which directly leads to  $\text{rank}(\mathbf{Q}_k) + \text{rank}(\mathbf{Q}_l) \leq N$ .

#### ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for their constructive comments that helped to improve the paper.

#### REFERENCES

- [1] I. Hwang, B. Song, and S. S. Soliman, "A holistic view on hyper-dense heterogeneous and small cell networks," *IEEE Commun. Mag.*, vol. 51, no. 6, pp. 20–27, 2013.
- [2] 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.
- [3] M. Bengtsson and B. Ottersten, "Optimal and suboptimal transmit beamforming," in *Handbook of Antennas in Wireless Communications*, L. C. Godara, Ed. Boca Raton, FL, USA: CRC Press, 2001, ch. 18.
- [4] K. Cho and D. Yoon, "On the general BER expression of one- and two-dimensional amplitude modulations," *IEEE Trans. Commun.*, vol. 50, no. 7, pp. 1074–1080, 2002.
- [5] E. Song, Q. Shi, M. Sanjabi, R. Sun, and Z.-Q. Luo, "Robust SINR-constrained MISO downlink beamforming: When is semidefinite programming relaxation tight?," *EURASIP J. Wireless Commun. Netw.*, vol. 1, pp. 1–11, 2012.
- [6] M. B. Shenoouda and T. N. Davidson, "Convex conic formulations of robust downlink precoder designs with quality of service constraints," *IEEE J. Sel. Topics Signal Process.*, vol. 1, pp. 714–724, 2007.
- [7] 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.



- [8] 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.
- [9] 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.
- [10] G. Zheng, K.-K. Wong, and B. Ottersten, "Robust cognitive beamforming with bounded channel uncertainties," *IEEE Trans. Signal Process.*, vol. 57, no. 12, pp. 4871–4881, 2009.
- [11] S. A. Vorobyov, H. Chen, and A. B. Gershman, "On the relationship between robust minimum variance beamformers with probabilistic and worst-case distortionless response constraints," *IEEE Trans. Signal Process.*, vol. 56, no. 11, pp. 5719–5724, 2008.
- [12] K.-L. Hsiung and R. Chen, "Power allocation with outage probability constraints in lognormal fading wireless channels: A relaxation approach," in *Proc. 39th Southeast. Symp. Syst. Theory*, 2007, pp. 173–175.
- [13] N. Vučić and H. Boche, "A tractable method for chance-constrained power control in downlink multiuser MISO systems with channel uncertainty," *IEEE Signal Process. Lett.*, vol. 16, no. 5, pp. 346–349, 2009.
- [14] K.-Y. Wang, T. H. Chang, W.-K. Ma, A. M.-C. So, and C.-Y. Chi, "Probabilistic SINR constrained robust transmit beamforming: A Bernstein-type inequality based conservative approach," in *Proc. IEEE ICASSP*, 2011, pp. 3080–3083.
- [15] K.-Y. Wang, N. Jacklin, Z. Ding, and C.-Y. Chi, "Robust MISO transmit optimization under outage-based QoS constraints in two-tier heterogeneous networks," *IEEE Trans. Wireless Commun.*, vol. 12, no. 4, pp. 1883–1897, 2013.
- [16] W. Xu, A. Tajer, X. Wang, and S. Alshomrani, "Power allocation in MISO interference channels with stochastic CSIT," *IEEE Trans. Wireless Commun.*, vol. 13, no. 3, pp. 1716–1727, 2014.
- [17] M. B. Shenoouda and T. N. Davidson, "Probabilistically-constrained approaches to the design of the multiple antenna downlink," in *Proc. Asilomar Conf.*, 2008, pp. 1120–1124.
- [18] X. He and Y. C. Wu, "Probabilistic QoS constrained robust downlink multiuser MIMO transceiver design with arbitrarily distributed channel uncertainty," *IEEE Trans. Wireless Commun.*, vol. 12, no. 12, pp. 6292–6302, 2013.
- [19] A. d'Aspremont and S. Boyd, "Relaxations and randomized methods for nonconvex QCQPs," Stanford Univ., EE392o, 2003.
- [20] Y. Fang, K. A. Loparo, and X. Feng, "Inequalities for the trace of matrix product," *IEEE Trans. Autom. Control*, vol. 39, no. 12, pp. 2489–2490, 1994.
- [21] P. D. Lax, *Linear Algebra and Its Applications*. New York, NY, USA: Wiley, 2007.
- [22] R. Tempo, G. Calafiore, and F. Dabbene, *Randomized Algorithms for Analysis and Control of Uncertain Systems*. London, U.K.: Springer-Verlag, 2013.
- [23] J. L. Jensen, *Saddlepoint Approximations*. Oxford, U.K.: Oxford Univ. Press., 1995.
- [24] R. W. Butler, *Saddlepoint Approximations With Applications*. Cambridge, U.K.: Cambridge Univ. Press, 2007.
- [25] R. W. Butler and M. S. Paoletta, "Uniform saddlepoint approximations for ratios of quadratic forms," *Bernoulli*, vol. 14, no. 1, pp. 140–154, 2008.
- [26] M. Lobo, L. Vandenberghe, S. Boyd, and H. Lebret, "Applications of second-order cone programming," *Linear Algebra Its Appl.*, no. 284, pp. 193–228, 1998.
- [27] A. L. Fradkov and V. A. Yakubovich, "The s-procedure and a duality relations in nonconvex problems of quadratic programming," *Vestnik Leningrad Univ. Math.*, vol. 5, no. 1, pp. 101–109, 1979.
- [28] F. Zhang, *Matrix Theory: Basic Results and Techniques*, 2nd ed. New York, NY, USA: Springer, 2011.
- [29] S. M. Kay, *Fundamentals of Statistical Signal Processing, Volume I: Estimation Theory*. : Prentice Hall, 1993.
- [30] Q. Wang, H. Feng, L. J. Cimini, L. J. Greenstein, D. S. Chan, and A. Hedayat, "Comparison of quantization techniques for downlink multiuser MIMO channels with limited feedback," *IEEE Wireless Commun. Lett.*, vol. 3, no. 2, pp. 165–168, 2014.
- [31] E. Björnson, K. Ntontin, and B. Ottersten, "Channel quantization design in multiuser MIMO systems: Asymptotic versus practical conclusions," in *Proc. IEEE ICASSP*, 2011, pp. 3072–3075.



**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 Journal of Communications and Networks.