

# A Distributionally Robust Linear Receiver Design for Multi-Access Space-Time Block Coded MIMO Systems

Bin Li, Yue Rong, *Senior Member, IEEE*, Jie Sun, and Kok Lay Teo, *Senior Member, IEEE*

**Abstract**—A receiver design problem for multi-access space-time block coded multiple-input multiple-output (MIMO) systems is considered. To hedge the mismatch between the true and the estimated channel state information (CSI), several robust receivers have been developed in the past decades. Among these receivers, the Gaussian robust receiver has been shown to be superior in performance. This receiver is designed based on the assumption that the CSI mismatch has Gaussian distribution. However, in real world applications, the assumption of Gaussianity might not hold. Motivated by this fact, a more general distributionally robust receiver is proposed in this paper, where only the mean and the variance of the CSI mismatch distribution are required in the receiver design. A tractable semi-definite programming (SDP) reformulation of the robust receiver design is developed. To suppress the self-interferences, a more advanced distributionally robust receiver is proposed. A tight convex approximation is given and the corresponding tractable SDP reformulation is developed. Moreover, for the sake of easy implementation, we present a simplified distributionally robust receiver. Simulations results are provided to show the effectiveness of our design by comparing with some existing well-known receivers.

**Index Terms**—MIMO systems, space-time block codes (STBC-s), distributionally robust optimization, individual chance constraint, joint chance constraint, Gaussian mixture.

## I. INTRODUCTION

### A. Motivation

Space-time coding has been shown to be a convincing approach to exploit the spacial diversity and improve the immunity to fading in multiple-input and multiple-output (MIMO) communication systems [1]. Minimum variance (MV) based receivers have been proposed in [2] to suppress the multi-access interferences (MAI) in multi-access space-time block coded MIMO systems. However, the receivers in [2] are designed based on the assumption that the channel state information (CSI) is perfectly known at the receiver. To combat the mismatch between the true and the estimated CSI, robust linear multiuser MIMO receiver design has been studied in the last decade. For example, the diagonal loading minimum

variance (DLMV) receiver has been developed in [2] and worst-case based robust receivers have been proposed in [3] and [4].

However, the worst-case based robust receiver design [3] and [4] has been shown to be conservative in practice since the actual worst-case may occur with a very low probability in real world applications. Furthermore, it is difficult to identify the physical meaning in practice by measuring the channel mismatch with Frobenius norm. Therefore, a chance-constrained receiver has been developed in [5] by exploring the stochastic characteristics of the channel mismatch. As a Gaussian CSI mismatch is assumed in [5], we refer to this receiver as Gaussian robust receiver in this paper. Although the performance degradation of the Gaussian robust receiver has been shown to be less than that of the worst-case based receiver in most cases, CSI mismatch distribution may not be available in practice. In addition, even the CSI mismatch distribution is available, it may not be subject to a Gaussian distribution.

Motivated by these facts, a more general robust linear receiver, distributionally robust receiver, is proposed in this paper. Our new receiver design provides a more general and practical formulation, in which we do not assume the full knowledge of the CSI mismatch. The term distributionally robust is introduced from the concept in distributionally robust optimization [6], [7], [8], [9]. In our formulation, we consider that all the possible CSI mismatch distributions belong to a set, which is called ‘ambiguity set’ in distributionally robust optimization. This ambiguity set contains all the distributions which have the same mean and the same covariance. The receiver design is formulated as a distributionally robust optimization problem with an individual chance constraint.

Different from the Gaussian robust receiver in [5], the chance constraint is required to be satisfied for all the distributions in the ambiguity set by optimizing the worst-case distribution. Furthermore, we incorporate another chance constraint in our formulation to suppress the self-interferences when the signal-to-noise ratio (SNR) is high. This chance constraint is called joint chance constraint in stochastic optimization, where a group of constraints are required to be satisfied simultaneously with a same probability. However, an optimization problem with such constraints usually cannot find a tractable solution. Hence, our goal in this paper is to provide tractable reformulations for these general distributionally robust receivers.

B. Li is with the School of Information Engineering, Guangdong University of Technology, Guangzhou 510006, China, and also with the Department of Mathematics and Statistics, Curtin University, GPO Box U1987, Perth, WA 6845, Australia, e-mail: bin.li@curtin.edu.au.

J. Sun and K. L. Teo are with the Department of Mathematics and Statistics, Curtin University, GPO Box U1987, Perth, WA 6845, Australia, e-mail: jie.sun@curtin.edu.au, k.l.teo@curtin.edu.au.

Y. Rong is with the Department of Electrical and Computer Engineering, Curtin University, GPO Box U1987, Perth, WA 6845, Australia, e-mail: y.rong@curtin.edu.au.

Manuscript received March 24, 2016; revised September 4, 2016.

## B. Literature Review

It is well-known that in general a chance-constrained optimization problem is non-convex and hence is computationally intractable. Only in some special cases, a chance-constrained optimization problem is convex. For example, under multivariate Gaussian distribution, an individual chance constraint can be represented by a second-order cone. More generally, it has been shown in [7] that an individual chance constraint can be converted into second-order cone constraints when the random parameters are under radial distributions. However, in most cases, chance-constrained problems are computationally intractable.

There are several methods to solve chance-constrained optimization problems, for example, the Monte-Carlo sampling method [10]. However, it may be computational prohibitive for large scale problems or problems under high feasibility requirement. An attractive approach to solve chance-constrained problem is the convex approximation (also known as safe approximation) method which yields a tractable and feasible solution to the original problem. The conditional value-at-risk (CVaR), which was introduced in [11] [12], is known as the tightest convex approximation to chance constraints [13], [14]. In general, CVaR is computationally prohibitive since the evaluation of a multidimensional integration is required. Fortunately, it has been shown in [6] that for certain constraint functions in distributionally robust scenario, CVaR is computationally tractable. Furthermore, for individual chance constraint, the CVaR approximation has been shown to be exact [6]. For the joint chance constraint, although the CVaR approximation is inexact, it is a tight convex approximation. There are some other deterministic approximation methods, for example, the Chebyshev inequality, and the Bernstein inequality [15].

In this paper, based on our distributionally robust problem formulation, we show that our distributionally robust receiver with the individual chance constraint design can be converted into a semi-definite programming (SDP) problem, which is computational tractable and hence can be solved efficiently with standard optimization package, such as CVX. For the receiver with the joint chance constraint, we provide a tight convex approximation. Furthermore, for the purpose of easy implementation, we provide a simplified design by using the Chebyshev inequality.

There are some recent works on distributionally robust optimization based design in wireless communications, for example, [16] and [17]. In [16], two types of distributionally robust beamformers have been proposed for multiple-input single-output (MISO) downlink systems and the corresponding approximate tractable reformulations are developed. In [17], an efficient distributionally robust slow adaptive orthogonal frequency-division multiple access (OFDMA) scheme has been proposed, which aims to increase the capacity gain of adaptive OFDMA. The formulated problem has been solved by converting the original problem to a tractable linear program.

## C. Contributions

We summarize the contributions of this work as follows: 1. Our formulation is more general and practical than the state-of-the-art Gaussian robust receiver, and it is applicable to general linear dispersion (LD) space-time block codes; 2. A tractable SDP reformulation is developed; 3. A tight convex approximation and its corresponding tractable SDP reformulation are provided to suppress the self-interferences; 4. A simple design is provided for the self-interferences suppression receiver.

## D. Structure

The rest of this paper is organized as follows. Background on multi-access space-time block coded (STBC) MIMO systems and linear multiuser receiver algorithms is given in Section II. In Section III, the formulation of the distributionally robust receiver is provided, and we show that this problem can be reformulated as an SDP optimization problem. Section IV presents simulation results that compare the performance of the proposed receivers with the existing techniques. Finally, we conclude this paper in Section V.

## II. BACKGROUND

### A. Multi-Access STBC MIMO Systems

We consider an uplink multiuser MIMO communication system with multiple transmitters and a single receiver. Each transmitter is assumed to have the same number of antennas and to encode information-bearing symbols using the same STBC. The received signal can be written as (see [2], [5])

$$\mathbf{Y} = \sum_{i=1}^I \mathbf{X}_i \mathbf{H}_i + \mathbf{N} \quad (1)$$

where

$$\mathbf{Y} \triangleq [\mathbf{y}^T(1), \mathbf{y}^T(2), \dots, \mathbf{y}^T(T)]^T \quad (2)$$

$$\mathbf{X}_i \triangleq [\mathbf{x}_i^T(1), \mathbf{x}_i^T(2), \dots, \mathbf{x}_i^T(T)]^T \quad (3)$$

$$\mathbf{N} \triangleq [\mathbf{n}^T(1), \mathbf{n}^T(2), \dots, \mathbf{n}^T(T)]^T \quad (4)$$

are the matrices of the received signals, transmitted signals of the  $i$ th transmitter, and noise, respectively,  $\mathbf{H}_i$  is the  $N \times M$  complex channel matrix between the  $i$ th transmitter and the receiver,  $N$  is the number of transmit antennas,  $M$  is the number of receive antennas,  $I$  is the number of transmitters,  $T$  is the block length,  $(\cdot)^T$  denotes the transpose, and for  $t = 1, 2, \dots, T$ ,

$$\mathbf{y}(t) \triangleq [y_1(t), y_2(t), \dots, y_M(t)] \quad (5)$$

$$\mathbf{x}_i(t) \triangleq [x_{i,1}(t), x_{i,2}(t), \dots, x_{i,N}(t)] \quad (6)$$

$$\mathbf{n}(t) \triangleq [n_1(t), n_2(t), \dots, n_M(t)] \quad (7)$$

are the complex row vectors of the received signals, transmitted signals of the  $i$ th user, and noise, respectively.

We denote the complex information-bearing symbols of the  $i$ th transmitter prior to space-time encoding as

$$\mathbf{s}_i \triangleq [s_{i,1}, s_{i,2}, \dots, s_{i,K}]^T \quad (8)$$

where  $K$  is the constellation size.

It can be shown that for any LD code,  $\mathbf{X}(s_i)$  can be written as (see [18], [19])

$$\mathbf{X}(s_i) = \sum_{k=1}^K (\mathbf{C}_k \text{Re}\{s_{i,k}\} + \mathbf{D}_k \text{Im}\{s_{i,k}\}) \quad (9)$$

where  $\mathbf{C}_k \triangleq \mathbf{X}(\mathbf{q}_k)$ ,  $\mathbf{D}_k \triangleq \mathbf{X}(j\mathbf{q}_k)$ ,  $j = \sqrt{-1}$  and  $\mathbf{q}_k$  is the  $K \times 1$  vector having one in its  $k$ th position and zeros elsewhere. Using (9), we can rewrite (1) as (see [2])

$$\underline{\mathbf{Y}} = \sum_{i=1}^I \mathbf{A}(\mathbf{H}_i) \underline{s}_i + \underline{\mathbf{N}} \quad (10)$$

where the ‘underline’ operator for any matrix  $\mathbf{P}$  is defined as

$$\underline{\mathbf{P}} \triangleq \begin{bmatrix} \text{vec}(\text{Re}\{\mathbf{P}\}) \\ \text{vec}(\text{Im}\{\mathbf{P}\}) \end{bmatrix}. \quad (11)$$

Here,  $\text{vec}(\cdot)$  is the vectorization operator stacking all columns of a matrix on top of each other, and the  $2MT \times 2K$  real-valued matrix  $\mathbf{A}(\mathbf{H}_i)$  is given by [19]

$$\begin{aligned} & \mathbf{A}(\mathbf{H}_i) \\ &= [\mathbf{C}_1 \mathbf{H}_i, \dots, \mathbf{C}_K \mathbf{H}_i, \mathbf{D}_1 \mathbf{H}_i, \dots, \mathbf{D}_K \mathbf{H}_i] \\ &\triangleq [\mathbf{a}_1(\mathbf{H}_i), \dots, \mathbf{a}_K(\mathbf{H}_i), \mathbf{a}_{K+1}(\mathbf{H}_i), \dots, \mathbf{a}_{2K}(\mathbf{H}_i)] \end{aligned} \quad (12)$$

### B. Robust MV Receivers

The goal of designing a receiver is to extract the signals received from the user-of-interest, while rejecting the interference and noise components. Without any loss of generality, let us assume that the first user is the user-of-interest. The estimated value of the data vector  $\underline{s}_1$  at the output of a linear receiver can be expressed as

$$\underline{\hat{s}}_1 = \mathbf{W}^T \underline{\mathbf{Y}} \quad (13)$$

where

$$\mathbf{W} = [\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_{2K}] \quad (14)$$

is the  $2MT \times 2K$  matrix of the receiver weight coefficients, and  $\mathbf{w}_k$  is the  $2MT \times 1$  weight vector that is used to decode the  $k$ th entry of  $\underline{s}_1$ . Given the matrix  $\mathbf{W}$ , the estimate of the vector of information symbols of the transmitter-of-interest can be computed as

$$\hat{\mathbf{s}}_1 = [\mathbf{I}_K \ j\mathbf{I}_K] \underline{\hat{s}}_1 \quad (15)$$

where  $\mathbf{I}_K$  is a  $K \times K$  identity matrix.

To suppress MAI, an MV receiver was proposed in [2]. This MV receiver is designed to estimate  $\underline{s}_1$  by minimizing the receiver output power while preserving a unity gain for this particular entry of  $\underline{s}_1$ . The corresponding optimization problem can be written as [2]

$$\min_{\mathbf{w}_k} \mathbf{w}_k^T \hat{\mathbf{R}} \mathbf{w}_k \quad \text{s.t.} \quad \mathbf{a}_k^T(\mathbf{H}_1) \mathbf{w}_k = 1 \quad (16)$$

for all  $k = 1, \dots, 2K$ , where

$$\hat{\mathbf{R}} = \frac{1}{J} \sum_{i=1}^J \underline{\mathbf{Y}}_i \underline{\mathbf{Y}}_i^T \quad (17)$$

is the sample estimate of the full rank  $2MT \times 2MT$  covariance matrix

$$\mathbf{R} \triangleq \mathbb{E}\{\underline{\mathbf{Y}} \underline{\mathbf{Y}}^T\} \quad (18)$$

of the vectorized data,  $\underline{\mathbf{Y}}_i$  is the  $i$ th received data block, and  $\mathbb{E}\{\cdot\}$  denotes the statistical expectation.

To improve the robustness of the receivers (16) against CSI errors, some more advanced robust receivers have been developed. These receivers are designed by considering the following error matrix

$$\Delta_1 \triangleq \mathbf{H}_1 - \hat{\mathbf{H}}_1 \quad (19)$$

where  $\mathbf{H}_1$  and  $\hat{\mathbf{H}}_1$  denote the actual channel matrix of the user-of-interest and its estimated value available at the receiver, respectively, and  $\Delta_1$  denotes the CSI mismatch. In [3], a worst-case based robust receiver has been proposed, where the Frobenius norm of the error matrix is bounded by a known constant  $\eta$

$$\|\Delta_1\|_F \leq \eta. \quad (20)$$

Here  $\|\cdot\|_F$  denotes the matrix Frobenius norm. Then the worst-case based receiver is formulated as

$$\min_{\mathbf{w}_k} \mathbf{w}_k^T \hat{\mathbf{R}} \mathbf{w}_k \quad \text{s.t.} \quad \min_{\|\Delta_1\|_F \leq \eta} \mathbf{w}_k^T \mathbf{a}_k(\hat{\mathbf{H}}_1 + \Delta_1) \geq 1 \quad (21)$$

where  $\mathbf{a}_k(\hat{\mathbf{H}}_1 + \Delta_1)$  is defined in the same manner as that in (12). The main modification of (21) with respect to (16) is that for each  $k$ , instead of requiring fixed distortionless response towards the single mismatched space-time signature  $\mathbf{a}_k(\hat{\mathbf{H}}_1)$ , in (21), such distortionless response is maintained by means of inequality constraints for a continuum of all space-time signatures given by the set

$$\mathcal{A}(\eta) = \left\{ \mathbf{a}_k(\hat{\mathbf{H}}_1 + \Delta_1) \mid \|\Delta_1\|_F \leq \eta \right\}.$$

The inequality constraint in (21) guarantees that the distortionless response is maintained in the worst case, i.e., for a particular vector in  $\mathcal{A}(\eta)$ , which corresponds to the smallest value of  $\mathbf{w}_k^T \mathbf{a}_k(\hat{\mathbf{H}}_1 + \Delta_1)$ .

However, it is difficult to determine  $\eta$  in real world applications because its physical meaning in practice is not obvious. In addition, the worst-case scenario that was considered in [3] may occur with a very low chance in practice. Therefore, a chance constraint based robust receiver was proposed in [5], where the receiver is designed by considering the stochastic property of the CSI mismatch. This chance constraint based robust receiver design is formulated as

$$\min_{\mathbf{w}_k} \mathbf{w}_k^T \hat{\mathbf{R}} \mathbf{w}_k \quad (22)$$

$$\text{s.t.} \quad \Pr_{[\mathbb{G}]} \left[ \mathbf{w}_k^T \left( \mathbf{a}_k(\hat{\mathbf{H}}_1) + \mathbf{e}_k(\Delta_1) \right) \geq 1 \right] \geq 1 - \epsilon \quad (23)$$

where  $\Pr_{[\mathbb{G}]}[\cdot]$  stands for the probability under the Gaussian distribution  $\mathbb{G}$  and

$$\mathbf{e}_k(\Delta_1) \triangleq \mathbf{a}_k(\mathbf{H}_1) - \mathbf{a}_k(\hat{\mathbf{H}}_1). \quad (24)$$

According to [5], the entries of  $\Delta_1$  are assumed to have uncorrelated Gaussian distribution, and we refer to (22)-(23) as the Gaussian robust receiver in our paper. We would like to mention that  $p$  in [5] is replaced by  $1 - \epsilon$  in (23), and  $\epsilon$  is chosen according to quality-of-service (QoS) specifications in practice.

### III. A DISTRIBUTIONALLY ROBUST LINEAR RECEIVER

Although the robust receiver in [5] explores the stochastic information of the channel, the exact distribution of the CSI mismatch is usually not available in practice. In most cases, we may only have partial information on the distribution of the CSI mismatch. Moreover, even the mismatch distribution is available, it may not be Gaussian. Motivated by this fact, in this section, we develop a distributionally robust linear receiver. Here, the concept of distributionally robust is from distributionally robust optimization in the literature. More specifically, the receiver design is based only on the first-order and second-order moments of the mismatch. In this paper, we assume that the mean and the covariance of  $\underline{\Delta}_1$  under the distribution  $\mathbb{P}$  is  $\underline{\mu}$  and  $\underline{\Sigma}$ , respectively. In addition, we define a distribution set  $\mathcal{P}$  as

$$\mathcal{P} = \left\{ \mathbb{P} : \mathbb{E}_{\mathbb{P}} [\underline{\Delta}_1] = \underline{\mu}, \mathbb{E}_{\mathbb{P}} [(\underline{\Delta}_1 - \underline{\mu})(\underline{\Delta}_1 - \underline{\mu})^T] = \underline{\Sigma} \right\}. \quad (25)$$

The robustness is in the sense of finding the worst-case distribution among all the possible distributions in the distribution set  $\mathcal{P}$  such that the chance constraint (23) is satisfied.

Using the notations of model (10), we can write

$$\begin{aligned} e_k(\underline{\Delta}_1) &\triangleq \mathbf{a}_k(\mathbf{H}_1) - \mathbf{a}_k(\hat{\mathbf{H}}_1) \\ &= \underline{\mathbf{F}}_k \underline{\mathbf{H}}_1 - \underline{\mathbf{F}}_k \hat{\mathbf{H}}_1 \\ &= \underline{\mathbf{F}}_k \underline{\Delta}_1, \quad k = 1, \dots, 2K \end{aligned} \quad (26)$$

where

$$\underline{\mathbf{F}}_k \triangleq \begin{cases} \mathbf{C}_k, & k = 1, \dots, K \\ \mathbf{D}_{k-K}, & k = K + 1, \dots, 2K \end{cases} \quad (27)$$

and the last equality in (26) follows from the linearity of the underline operator (11). We note that  $e_k(\underline{\Delta}_1)$  depends linearly on  $\underline{\Delta}_1$ . Indeed, applying the underline operator (11) to (26) and using some properties of the Kronecker matrix product [20], we have

$$\begin{aligned} e_k(\underline{\Delta}_1) &= \begin{bmatrix} \text{vec}(\text{Re}\{\underline{\mathbf{F}}_k \underline{\Delta}_1\}) \\ \text{vec}(\text{Im}\{\underline{\mathbf{F}}_k \underline{\Delta}_1\}) \end{bmatrix} \\ &= \begin{bmatrix} \text{Re}\{(\mathbf{I}_M \otimes \underline{\mathbf{F}}_k) \text{vec}(\underline{\Delta}_1)\} \\ \text{Im}\{(\mathbf{I}_M \otimes \underline{\mathbf{F}}_k) \text{vec}(\underline{\Delta}_1)\} \end{bmatrix} \\ &= \begin{bmatrix} \text{Re}\{\mathbf{I}_M \otimes \underline{\mathbf{F}}_k\} & -\text{Im}\{\mathbf{I}_M \otimes \underline{\mathbf{F}}_k\} \\ \text{Im}\{\mathbf{I}_M \otimes \underline{\mathbf{F}}_k\} & \text{Re}\{\mathbf{I}_M \otimes \underline{\mathbf{F}}_k\} \end{bmatrix} \begin{bmatrix} \text{vec}(\text{Re}\{\underline{\Delta}_1\}) \\ \text{vec}(\text{Im}\{\underline{\Delta}_1\}) \end{bmatrix} \\ &= \underline{\Psi}_k \underline{\Delta}_1 \end{aligned} \quad (28)$$

where

$$\underline{\Psi}_k \triangleq \begin{bmatrix} \text{Re}\{\mathbf{I}_M \otimes \underline{\mathbf{F}}_k\} & -\text{Im}\{\mathbf{I}_M \otimes \underline{\mathbf{F}}_k\} \\ \text{Im}\{\mathbf{I}_M \otimes \underline{\mathbf{F}}_k\} & \text{Re}\{\mathbf{I}_M \otimes \underline{\mathbf{F}}_k\} \end{bmatrix} \quad (29)$$

and  $\otimes$  denotes the matrix Kronecker product.

In view of (23), the single chance constraint is calculated based on the Gaussian distribution. However, in the context of distributionally robust optimization, rather than knowing the exact distribution  $\mathbb{P}$ , there are infinite number of possible distributions which all have the same mean and variance and

belong to a distribution set  $\mathcal{P}$  as defined in (25). By considering the worst-case robust philosophy, the corresponding distributionally robust receiver design can be formulated as

$$\begin{aligned} \min_{\mathbf{w}_k} \quad & \mathbf{w}_k^T \hat{\mathbf{R}} \mathbf{w}_k \quad (30) \\ \text{s.t.} \quad & \inf_{\mathbb{P} \in \mathcal{P}} \Pr_{[\mathbb{P}]} \left[ \mathbf{w}_k^T \left( \mathbf{a}_k(\hat{\mathbf{H}}_1) + \underline{\Psi}_k \underline{\Delta}_1 \right) \geq 1 \right] \geq 1 - \epsilon \quad (31) \end{aligned}$$

It is well-known [14] that chance constraint problem is usually non-convex and hence it is difficult to solve the optimization problem (30)-(31). In the following, we shall derive a tractable reformulation of the problem (30)-(31).

A popular treatment of chance constraints is using convex approximation. Among them, conditional value-at-risk (CVaR) is widely accepted as the tightest convex approximation according to [13]. CVaR is a special class of risk measure introduced in [11] and further discussed in [12] as a tractable alternative for solving value-at-risk (VaR) problems in financial applications. There are different definitions of CVaR. In this paper, we adopt the definition in [12] as follows. For a random variable  $\tilde{\xi}$ , its CVaR under distribution  $\mathbb{P}$  is defined as

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

Here,  $\mathbb{R}$  denotes the set of all real numbers and  $(a)^+$  denotes  $\max(a, 0)$ , respectively, where  $a$  is a real number.

In general, for the chance constraints like (23), CVaR can only provide a ‘best’ convex approximation. However, for the distributionally robust constraint like (31), it has been proved that the ‘worst-case’ CVaR constraint among all possible distributions is equivalent to the distributionally robust constraint on condition that the distributionally robust constraint function is either concave or quadratic in  $\mathbf{w}_k$  [6]. Fortunately, we can justify that the problem (30)-(31) satisfies this condition with the following lemma (Theorem 2.2 in [6]).

LEMMA 1: Let  $L : \mathbb{R}^k \rightarrow \mathbb{R}$  be a continuous loss function that is either concave in  $\boldsymbol{\xi}$  or quadratic in  $\boldsymbol{\xi}$ . Then, the following equivalence holds.

$$\inf_{\mathbb{P} \in \mathcal{P}} \Pr_{[\mathbb{P}]} [L(\boldsymbol{\xi}) \leq 0] \geq 1 - \epsilon \iff \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{P} - \text{CVaR}_{\epsilon} [L(\boldsymbol{\xi})] \leq 0 \quad (32)$$

where  $\mathcal{P}$  is defined in (25).

Although CVaR is convex, it is difficult to calculate because the expectation involves multidimensional integration. Therefore, the evaluation of CVaR is computational prohibitive. However, the worst-case CVaR under the distributionally robust framework can be represented as an SDP from the following lemma (Theorem 21 in [6]).

LEMMA 2: The feasible set

$$\left\{ \mathbf{x} \in \mathbb{R}^n : \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{P} - \text{CVaR}_{\epsilon} [y^0(\mathbf{x}) + \mathbf{y}^T(\mathbf{x})\boldsymbol{\xi}] \leq 0 \right\} \quad (33)$$

can be written as

$$\left\{ \mathbf{x} \in \mathbb{R}^n : \mathbf{M} \succeq 0, \beta + \frac{1}{\epsilon} \text{Tr}(\boldsymbol{\Omega} \mathbf{M}) \leq 0, \mathbf{M} - \begin{bmatrix} \mathbf{0} & \frac{1}{2} \mathbf{y}(\mathbf{x}) \\ \frac{1}{2} \mathbf{y}^T(\mathbf{x}) & y^0(\mathbf{x}) - \beta \end{bmatrix} \succeq 0 \right\} \quad (34)$$

where

$$\boldsymbol{\Omega} = \begin{bmatrix} \underline{\Sigma} + \underline{\mu} \underline{\mu}^T & \underline{\mu} \\ \underline{\mu}^T & 1 \end{bmatrix} \quad (35)$$

$\Sigma$  and  $\mu$  are defined as in (25),  $y^0(\mathbf{x})$  and  $\mathbf{y}(\mathbf{x})$  depend affinely on  $\mathbf{x}$ , and  $\text{Tr}(\cdot)$  denotes the matrix trace.

Based on the results above, we can show that the problem (30)-(31) can be represented as an SDP problem and hence is computationally tractable.

**THEOREM 1:** The problem (30)-(31) can be reformulated as the following conic optimization problem

$$\min_{\tau, \beta, \mathbf{w}_k, \mathbf{M}} \tau \quad (36)$$

$$\text{s.t.} \quad \|\mathbf{U}\mathbf{w}_k\| \leq \tau \quad (37)$$

$$\beta + \frac{1}{\epsilon} \text{Tr}(\mathbf{\Omega}\mathbf{M}) \leq 0 \quad (38)$$

$$\mathbf{M} - \begin{bmatrix} \mathbf{0} & -\frac{1}{2}\mathbf{\Psi}_k^T \mathbf{w}_k \\ -\frac{1}{2}\mathbf{w}_k^T \mathbf{\Psi}_k & 1 - \mathbf{w}_k^T \mathbf{a}_k(\hat{\mathbf{H}}_1) - \beta \end{bmatrix} \succeq \mathbf{0} \quad (39)$$

$$\mathbf{M} \succeq \mathbf{0} \quad (40)$$

where  $\|\cdot\|$  denotes the Euclidean norm,  $\mathbf{\Omega}$  is defined in (35), and  $\mathbf{M} \in \mathbb{S}^{2MN+1}$  means all the  $(2MN+1) \times (2MN+1)$  symmetric matrices.

**PROOF:** In view of (31), we can see that  $L(\underline{\Delta}_1) = 1 - \mathbf{w}_k^T (\mathbf{a}_k(\hat{\mathbf{H}}_1) + \mathbf{\Psi}_k \underline{\Delta}_1)$  depends linearly on  $\underline{\Delta}_1$ , which also implies that  $L(\underline{\Delta}_1)$  is concave in  $\underline{\Delta}_1$ . Thus, from LEMMA 1, we know that (31) is equivalent to

$$\sup_{\mathbb{P} \in \mathcal{P}} \mathbb{P} - \text{CVaR}_\epsilon \left[ 1 - \mathbf{w}_k^T (\mathbf{a}_k(\hat{\mathbf{H}}_1) + \mathbf{\Psi}_k \underline{\Delta}_1) \right] \leq 0 \quad (41)$$

where

$$\begin{aligned} & \mathbb{P} - \text{CVaR}_\epsilon \left[ 1 - \mathbf{w}_k^T (\mathbf{a}_k(\hat{\mathbf{H}}_1) + \mathbf{\Psi}_k \underline{\Delta}_1) \right] \\ &= \inf_{\beta \in \mathbb{R}} \left\{ \beta + \frac{1}{\epsilon} \mathbb{E}_{\mathbb{P}} \left[ \left( 1 - \mathbf{w}_k^T (\mathbf{a}_k(\hat{\mathbf{H}}_1) + \mathbf{\Psi}_k \underline{\Delta}_1) - \beta \right)^+ \right] \right\}. \end{aligned}$$

Let  $y^0(\mathbf{w}_k) = 1 - \mathbf{w}_k^T \mathbf{a}_k(\hat{\mathbf{H}}_1)$  and  $\mathbf{y}(\mathbf{w}_k) = -\mathbf{\Psi}_k^T \mathbf{w}_k$ . Clearly,  $y^0(\mathbf{w}_k)$  and  $\mathbf{y}(\mathbf{w}_k)$  affinely depend on  $\mathbf{w}_k$ . Then, from LEMMA 2, (41) can be rewritten as the constraints (38)-(40). Then, we take the Cholesky factorization of  $\hat{\mathbf{R}}$  in (30), which yields

$$\hat{\mathbf{R}} = \mathbf{U}^T \mathbf{U}. \quad (42)$$

By introducing a new decision variable  $\tau$ , (30) can be rewritten in the epigraph form as that in (36)-(37). This completes the proof.  $\square$

### A. Self-Interferences Suppression

For the distributionally robust receiver (36)-(40), both the self-interferences and MAI are suppressed by minimizing the output power in (36). However, according to [3] and [5], the performance of the distributionally robust receiver in (36)-(40) may degrade dramatically when the SNR is high. This is because when the SNR is high, the power of the self-interferences is high as well.

To suppress the self-interferences, some additional constraints are imposed in the formulation as follows

$$\min_{\mathbf{w}_k, \delta} \mathbf{w}_k^T \hat{\mathbf{R}} \mathbf{w}_k + \|\delta\|^2 \quad (43)$$

$$\text{s.t.} \quad \Pr_{\mathbb{P}} \left[ \mathbf{w}_k^T (\mathbf{a}_k(\hat{\mathbf{H}}_1) + \mathbf{e}_k(\Delta_1)) \geq 1 \right] \geq 1 - \epsilon \quad (44)$$

$$\Pr_{\mathbb{P}} \left[ \sigma_1 \left| \mathbf{w}_k^T (\mathbf{a}_l(\hat{\mathbf{H}}_1) + \mathbf{e}_l(\Delta_1)) \right| \leq \delta_l \right] \geq 1 - \epsilon, \quad l = 1, \dots, 2K, \quad l \neq k \quad (45)$$

where  $\delta = [\delta_1, \dots, \delta_{k-1}, \delta_{k+1}, \dots, \delta_{2K}]^T$  is the  $(2K-1) \times 1$  vector whose entries limit the contribution of self-interferences, and  $\sigma_1$  is the standard deviation of the waveform of the user-of-interest. In fact,  $\|\delta\|^2$  is the power of self-interferences for the user-of-interest.

Similar to the distributionally robust receiver (30)-(31), we can derive the formulation of the distributionally robust version of (43)-(45) as follows

$$\min_{\mathbf{w}_k, \delta} \mathbf{w}_k^T \hat{\mathbf{R}} \mathbf{w}_k + \|\delta\|^2 \quad (46)$$

$$\text{s.t.} \quad \inf_{\mathbb{P} \in \mathcal{P}} \Pr_{\mathbb{P}} \left[ \mathbf{w}_k^T (\mathbf{a}_k(\hat{\mathbf{H}}_1) + \mathbf{e}_k(\Delta_1)) \geq 1 \right] \geq 1 - \epsilon \quad (47)$$

$$\inf_{\mathbb{P} \in \mathcal{P}} \Pr_{\mathbb{P}} \left[ \sigma_1 \left| \mathbf{w}_k^T (\mathbf{a}_l(\hat{\mathbf{H}}_1) + \mathbf{e}_l(\Delta_1)) \right| \leq \delta_l \right] \geq 1 - \epsilon, \quad l = 1, \dots, 2K, \quad l \neq k. \quad (48)$$

Clearly, for each  $l$ , the constraint (45) can be rewritten as

$$\Pr_{\mathbb{P}} \left[ \begin{array}{l} \mathbf{w}_k^T (\mathbf{a}_l(\hat{\mathbf{H}}_1) + \mathbf{e}_l(\Delta_1)) \leq \delta_l / \sigma_1, \\ \mathbf{w}_k^T (\mathbf{a}_l(\hat{\mathbf{H}}_1) + \mathbf{e}_l(\Delta_1)) \geq -\delta_l / \sigma_1 \end{array} \right] \geq 1 - \epsilon. \quad (49)$$

Compared with (44), there are two inequalities in (45) to be satisfied. In fact, (45) is called joint chance constraint in stochastic optimization. Only when the distribution is log-concave, joint chance constraint is convex [14], [15]. Therefore, considering all the possible distributions in  $\mathcal{P}$ , the distributionally robust constraint (48) cannot be convex.

Here, we provide a tight convex approximation of (48) based on the results in [6]. In particular, we approximate (48) by the following constraint

$$\begin{aligned} & \inf_{\mathbb{P} \in \mathcal{P}} \Pr_{\mathbb{P}} \left[ \max \left\{ \alpha_{l,1} \left( \mathbf{w}_k^T (\mathbf{a}_l(\hat{\mathbf{H}}_1) + \mathbf{e}_l(\Delta_1)) - \delta_l / \sigma_1 \right), \right. \right. \\ & \left. \left. \alpha_{l,2} \left( -\mathbf{w}_k^T (\mathbf{a}_l(\hat{\mathbf{H}}_1) + \mathbf{e}_l(\Delta_1)) - \delta_l / \sigma_1 \right) \right\} \leq 0 \right] \geq 1 - \epsilon \quad (50) \end{aligned}$$

where  $\alpha_{l,1}$  and  $\alpha_{l,2}$  are positive numbers. It has been shown by [14] that the feasible set of (50) is a subset of (48).

To proceed further, we need the following result (Theorem 3.3 in [6]).

**LEMMA 3:** For any fixed  $\mathbf{x} \in \mathbb{R}^n$ ,  $\mathbf{y}_i : \mathbb{R}^n \rightarrow \mathbb{R}^k$ , and  $\alpha_i$ , the feasible set of

$$\inf_{\mathbb{P} \in \mathcal{P}} \Pr_{\mathbb{P}} \left[ \max_{i=1,2,\dots,m} \{ \alpha_i (y_i^0(\mathbf{x}) + \mathbf{y}_i^T(\mathbf{x})\boldsymbol{\xi}) \} \leq 0 \right] \geq 1 - \epsilon$$

is equivalent to the following SDP representable set

$$\left\{ \mathbf{x} \in \mathbb{R}^n : \begin{array}{l} \beta_i + \frac{1}{\epsilon} \text{Tr}(\mathbf{\Omega}\mathbf{M}_i) \leq 0, \mathbf{M}_i \succeq \mathbf{0}, \\ \mathbf{M}_i - \begin{bmatrix} \mathbf{0} & \frac{1}{2}\alpha_i \mathbf{y}_i(\mathbf{x}) \\ \frac{1}{2}\alpha_i \mathbf{y}_i^T(\mathbf{x}) & \alpha_i y_i^0(\mathbf{x}) - \beta_i \end{bmatrix} \succeq \mathbf{0}, \\ i = 1, 2, \dots, m, \beta \in \mathbb{R}, \mathbf{M}_i \in \mathbb{S}^{k+1} \end{array} \right\}.$$

Now we give a tight SDP approximation of the problem (46)-(48) with the following theorem.

**THEOREM 2:** The distributionally robust receiver with self-interferences suppression (46)-(48) can be approximated by the SDP optimization problem (51)-(59) shown at the top of the next page, where the decision variables are  $\tau, \beta_0, \beta_{l,1}, \beta_{l,2}, \mathbf{w}_k, \mathbf{M}_0, \mathbf{M}_{l,1}, \mathbf{M}_{l,2}, \alpha_{l,1}, \alpha_{l,2}, \delta$ .

$$\min \quad \tau + \|\boldsymbol{\delta}\|^2 \quad (51)$$

$$\text{s.t.} \quad \|\mathbf{U}\mathbf{w}_k\| \leq \tau \quad (52)$$

$$\beta_0 + \frac{1}{\epsilon} \text{Tr}(\boldsymbol{\Omega}\mathbf{M}_0) \leq 0 \quad (53)$$

$$\beta_{l,1} + \frac{1}{\epsilon} \text{Tr}(\boldsymbol{\Omega}\mathbf{M}_{l,1}) \leq 0, \quad l = 1, \dots, 2K, \quad l \neq k \quad (54)$$

$$\beta_{l,2} + \frac{1}{\epsilon} \text{Tr}(\boldsymbol{\Omega}\mathbf{M}_{l,2}) \leq 0, \quad l = 1, \dots, 2K, \quad l \neq k \quad (55)$$

$$\mathbf{M}_0 - \begin{bmatrix} \mathbf{0} & -\frac{1}{2}\boldsymbol{\Psi}_k^T \mathbf{w}_k \\ -\frac{1}{2}\mathbf{w}_k^T \boldsymbol{\Psi}_k & 1 - \mathbf{w}_k^T \mathbf{a}_k(\hat{\mathbf{H}}_1) - \beta_0 \end{bmatrix} \succeq 0, \quad \mathbf{M}_0 \succeq 0 \quad (56)$$

$$\mathbf{M}_{l,1} - \begin{bmatrix} \mathbf{0} & \frac{1}{2}\alpha_{l,1}\boldsymbol{\Psi}_l^T \mathbf{w}_k \\ \frac{1}{2}\alpha_{l,1}\mathbf{w}_k^T \boldsymbol{\Psi}_l & \alpha_{l,1}(\mathbf{w}_k^T \mathbf{a}_l(\hat{\mathbf{H}}_1) - \delta_l/\sigma_1) - \beta_{l,1} \end{bmatrix} \succeq 0 \quad (57)$$

$$\mathbf{M}_{l,2} - \begin{bmatrix} \mathbf{0} & -\frac{1}{2}\alpha_{l,2}\boldsymbol{\Psi}_l^T \mathbf{w}_k \\ -\frac{1}{2}\alpha_{l,2}\mathbf{w}_k^T \boldsymbol{\Psi}_l & \alpha_{l,2}(-\mathbf{w}_k^T \mathbf{a}_l(\hat{\mathbf{H}}_1) - \delta_l/\sigma_1) - \beta_{l,2} \end{bmatrix} \succeq 0 \quad (58)$$

$$\mathbf{M}_{l,1} \succeq 0, \quad \mathbf{M}_{l,2} \succeq 0, \quad l = 1, \dots, 2K, \quad l \neq k \quad (59)$$

Theorem 2 can be proved by considering Lemma 3 and Theorem 1. Here, we omit the proof.

By observing the problem (51)-(59), we can see that it requires to solve  $(4K - 1) \times 4K$  linear matrix inequalities (LMIs) to design the distributionally robust receiver with the self-interferences suppression function, which is computational prohibitive. Therefore, it is difficult to be implemented in practice. Hence, we shall give a simplified design in the next section.

### B. An Implementable Self-Interferences Suppression Receiver

According to [5], (45) can be simplified by the famous Chebyshev inequality, which states that for any random variable  $\theta$  and any positive real number  $\alpha$ ,

$$\Pr_{[\mathbb{P}]} \{|\theta| \geq \alpha\} \leq \frac{\mathbb{E}\{\theta^2\}}{\alpha^2}. \quad (60)$$

Since all constraints in (45) have the same structure, for the simplicity of presentation, we further discuss only the  $l$ th constraint.

Note that

$$\begin{aligned} & \mathbb{E}_{\mathbb{P}} \left[ \left| \mathbf{w}_k^T \left( \mathbf{a}_l(\hat{\mathbf{H}}_1) + \mathbf{e}_l(\boldsymbol{\Delta}_1) \right) \right|^2 \right] \\ &= \mathbb{E}_{\mathbb{P}} \left[ \mathbf{w}_k^T \left( \mathbf{a}_l(\hat{\mathbf{H}}_1) \mathbf{a}_l^T(\hat{\mathbf{H}}_1) + \mathbf{a}_l(\hat{\mathbf{H}}_1) \mathbf{e}_l^T(\boldsymbol{\Delta}_1) \right. \right. \\ & \quad \left. \left. + \mathbf{e}_l(\boldsymbol{\Delta}_1) \mathbf{a}_l^T(\hat{\mathbf{H}}_1) + \mathbf{e}_l(\boldsymbol{\Delta}_1) \mathbf{e}_l^T(\boldsymbol{\Delta}_1) \right) \mathbf{w}_k \right]. \end{aligned} \quad (61)$$

From (28), we know that

$$\begin{aligned} & \mathbb{E}_{\mathbb{P}} [\mathbf{e}_l(\boldsymbol{\Delta}_1)] = \boldsymbol{\Psi}_l \mathbb{E}_{\mathbb{P}} [\boldsymbol{\Delta}_1] = \boldsymbol{\Psi}_l \boldsymbol{\mu} \quad (62) \\ & \mathbb{E}_{\mathbb{P}} [\mathbf{e}_l(\boldsymbol{\Delta}_1) \mathbf{e}_l^T(\boldsymbol{\Delta}_1)] \\ &= \boldsymbol{\Psi}_l \mathbb{E}_{\mathbb{P}} [\boldsymbol{\Delta}_1 \boldsymbol{\Delta}_1^T] \boldsymbol{\Psi}_l^T = \boldsymbol{\Psi}_l (\boldsymbol{\Sigma} + \boldsymbol{\mu} \boldsymbol{\mu}^T) \boldsymbol{\Psi}_l^T. \end{aligned} \quad (63)$$

From (61)-(63), it follows that

$$\begin{aligned} & \mathbb{E}_{\mathbb{P}} \left[ \left| \mathbf{w}_k^T \left( \mathbf{a}_l(\hat{\mathbf{H}}_1) + \mathbf{e}_l(\boldsymbol{\Delta}_1) \right) \right|^2 \right] \\ &= \mathbf{w}_k^T \left( \mathbf{a}_l(\hat{\mathbf{H}}_1) \mathbf{a}_l^T(\hat{\mathbf{H}}_1) + \mathbf{a}_l(\hat{\mathbf{H}}_1) \boldsymbol{\mu}^T \boldsymbol{\Psi}_l^T \right. \\ & \quad \left. + \boldsymbol{\Psi}_l \boldsymbol{\mu} \mathbf{a}_l^T(\hat{\mathbf{H}}_1) + \boldsymbol{\Psi}_l (\boldsymbol{\Sigma} + \boldsymbol{\mu} \boldsymbol{\mu}^T) \boldsymbol{\Psi}_l^T \right) \mathbf{w}_k. \end{aligned} \quad (64)$$

Using (60) and (64), the left-hand side of the  $l$ th constraint in (45) can be lower bounded as

$$\begin{aligned} & \Pr_{[\mathbb{P}]} \left\{ \sigma_1 \left| \mathbf{w}_k^T \left( \mathbf{a}_l(\hat{\mathbf{H}}_1) + \mathbf{e}_l(\boldsymbol{\Delta}_1) \right) \right| \leq \delta_l \right\} \\ &= 1 - \Pr_{[\mathbb{P}]} \left\{ \sigma_1 \left| \mathbf{w}_k^T \left( \mathbf{a}_l(\hat{\mathbf{H}}_1) + \mathbf{e}_l(\boldsymbol{\Delta}_1) \right) \right| \geq \delta_l \right\} \\ &\geq 1 - \frac{\sigma_1^2}{\delta_l^2} \mathbf{w}_k^T \left( \mathbf{a}_l(\hat{\mathbf{H}}_1) \mathbf{a}_l^T(\hat{\mathbf{H}}_1) + \mathbf{a}_l(\hat{\mathbf{H}}_1) \boldsymbol{\mu}^T \boldsymbol{\Psi}_l^T \right. \\ & \quad \left. + \boldsymbol{\Psi}_l \boldsymbol{\mu} \mathbf{a}_l^T(\hat{\mathbf{H}}_1) + \boldsymbol{\Psi}_l (\boldsymbol{\Sigma} + \boldsymbol{\mu} \boldsymbol{\mu}^T) \boldsymbol{\Psi}_l^T \right) \mathbf{w}_k. \end{aligned} \quad (65)$$

Replacing all the constraints in (45) by their lower bounds (65), we obtain the following set of constraints

$$\begin{aligned} & \frac{\sigma_1^2}{\epsilon} \mathbf{w}_k^T \left( \mathbf{a}_l(\hat{\mathbf{H}}_1) \mathbf{a}_l^T(\hat{\mathbf{H}}_1) + \mathbf{a}_l(\hat{\mathbf{H}}_1) \boldsymbol{\mu}^T \boldsymbol{\Psi}_l^T \right. \\ & \quad \left. + \boldsymbol{\Psi}_l \boldsymbol{\mu} \mathbf{a}_l^T(\hat{\mathbf{H}}_1) + \boldsymbol{\Psi}_l (\boldsymbol{\Sigma} + \boldsymbol{\mu} \boldsymbol{\mu}^T) \boldsymbol{\Psi}_l^T \right) \mathbf{w}_k \leq \delta_l^2, \\ & \quad l = 1, \dots, 2K, \quad l \neq k. \end{aligned} \quad (66)$$

The constraints in (66) are referred to as *safe approximations* of the original constraints in (45), meaning that the constraints in (66) are stricter than those in (45). Therefore, the constraints in (45) always hold true provided that those in (66) are satisfied.

For the sake of simplicity, we further approximate the constraints in (66) by summing them together to obtain a single constraint of the following form

$$\mathbf{w}_k^T \mathbf{Q}_k \mathbf{w}_k \leq \|\boldsymbol{\delta}\|^2 \quad (67)$$

where

$$\begin{aligned} \mathbf{Q}_k &\triangleq \frac{\sigma_1^2}{\epsilon} \sum_{l=1, l \neq k}^{2K} \left[ \mathbf{a}_l(\hat{\mathbf{H}}_1) \mathbf{a}_l^T(\hat{\mathbf{H}}_1) + \mathbf{a}_l(\hat{\mathbf{H}}_1) \boldsymbol{\mu}^T \boldsymbol{\Psi}_l^T \right. \\ & \quad \left. + \boldsymbol{\Psi}_l \boldsymbol{\mu} \mathbf{a}_l^T(\hat{\mathbf{H}}_1) + \boldsymbol{\Psi}_l (\boldsymbol{\Sigma} + \boldsymbol{\mu} \boldsymbol{\mu}^T) \boldsymbol{\Psi}_l^T \right]. \end{aligned} \quad (68)$$

Substituting the left-hand side of (67) into the objective function (43) instead of the term  $\|\boldsymbol{\delta}\|^2$ , we can eliminate the constraint (67) from the final optimization problem. Then, the new objective function can be written as

$$\mathbf{w}_k^T (\hat{\mathbf{R}} + \mathbf{Q}_k) \mathbf{w}_k. \quad (69)$$

TABLE I: Computational Complexity of the Gaussian Robust Receiver [5] and the Distributionally Robust Receiver (70)-(74)

Receiver	[5]	(70)-(74)
Complexity order	$\mathcal{O}(M^3T^3)$	$\mathcal{O}(M^{4.5}T^{4.5})$

The final optimization problem which approximates the original problem (43)-(45) can be written as the following second-order cone programming (SOCP) problem

$$\min_{\tau, \beta, \mathbf{w}_k, M} \tau \quad (70)$$

$$\text{s.t.} \quad \|\mathbf{Z}_k \mathbf{w}_k\| \leq \tau \quad (71)$$

$$\beta + \frac{1}{\epsilon} \text{Tr}(\mathbf{\Omega} \mathbf{M}) \leq 0 \quad (72)$$

$$\mathbf{M} - \begin{bmatrix} \mathbf{0} & -\frac{1}{2} \mathbf{\Psi}_k^T \mathbf{w}_k \\ -\frac{1}{2} \mathbf{w}_k^T \mathbf{\Psi}_k & 1 - \mathbf{w}_k^T \mathbf{a}_k(\hat{\mathbf{H}}_1) - \beta \end{bmatrix} \succeq \mathbf{0} \quad (73)$$

$$\mathbf{M} \succeq \mathbf{0} \quad (74)$$

where

$$\hat{\mathbf{R}} + \mathbf{Q}_k = \mathbf{Z}_k^T \mathbf{Z}_k \quad (75)$$

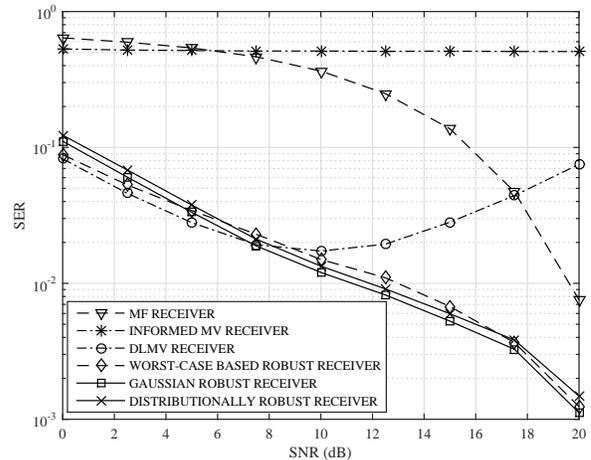
is the Cholesky factorization of  $\hat{\mathbf{R}} + \mathbf{Q}_k$ , and  $\tau$  is a new variable such that  $\|\mathbf{Z}_k \mathbf{w}_k\| \leq \tau$ .

We observe that only  $4K$  LMIs are involved in the computation of designing a distributionally robust self-interferences suppression receiver, which is much less than  $(4K - 1) \times 4K$  in the receiver (51)-(59). Therefore, the receiver (70)-(74) is much easier to be implemented in practice. We compare the computational complexity of the Gaussian robust receiver [5] and the proposed receiver (70)-(74) in Table I according to the results in [21]. Here, for simplicity we assume  $T = N$ . It can be seen from Table I that the proposed receiver has a higher computational complexity than the Gaussian robust receiver. It will be seen in the next section that the proposed receiver has a better symbol-error-rate (SER) performance than the Gaussian robust receiver, such performance-complexity trade-off is interesting for practical multi-access space-time block coded MIMO systems.

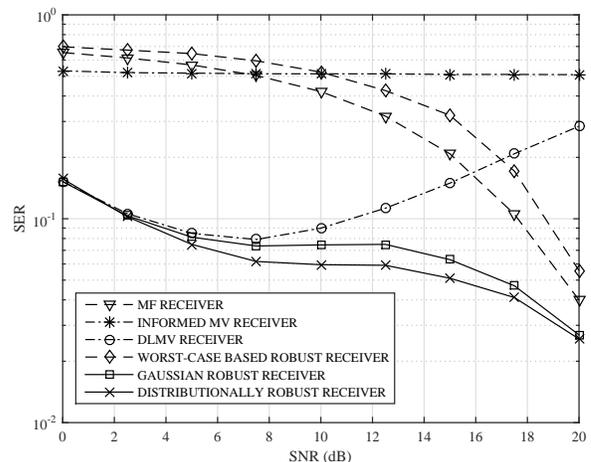
#### IV. SIMULATIONS

We consider an uplink cellular communication system with  $I$  transmitters each equipped with  $N$  antennas and a single receiver equipped with  $M$  antennas. The interfering transmitters use the same STBC as the transmitter-of-interest. The block length is  $T$ . The interference-to-noise ratio (INR) is equal to 20 dB and the QPSK modulation scheme is used. The MIMO channel between the  $i$ th transmitter and the receiver is assumed to be quasi-static Rayleigh flat fading, and  $p = 1 - \epsilon = 0.95$  is taken for the proposed distributionally robust receiver and the Gaussian robust receiver all through our simulations. For each example, a Monte-Carlo simulation of 1000 runs is performed.

The following receivers are compared in terms of SERs: the proposed distributionally robust receiver (70)-(74), the Gaussian robust receiver [5], the worst-case optimization-based robust receiver [3] with the parameter  $\eta = 6\sigma_h$ , the DLMV receiver [5] with the DL factor  $\nu = 10\sigma_v^2$  (where  $\sigma_v^2$  is the noise variance), the MF receiver, and the 'informed' MV



(a)  $\sigma_h^2 = 0.1$ .



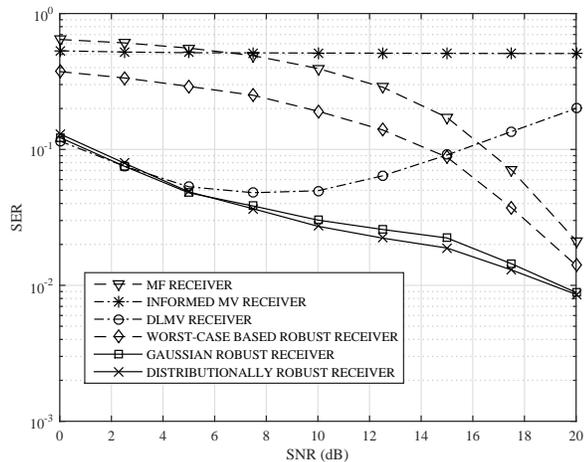
(b)  $\sigma_h^2 = 0.5$ .

Fig. 1: Example One. Alamouti's code, i.i.d. Gaussian CSI mismatch,  $I = 2$ ,  $N = K = T = 2$ ,  $M = 8$ , and  $J = 35$ .

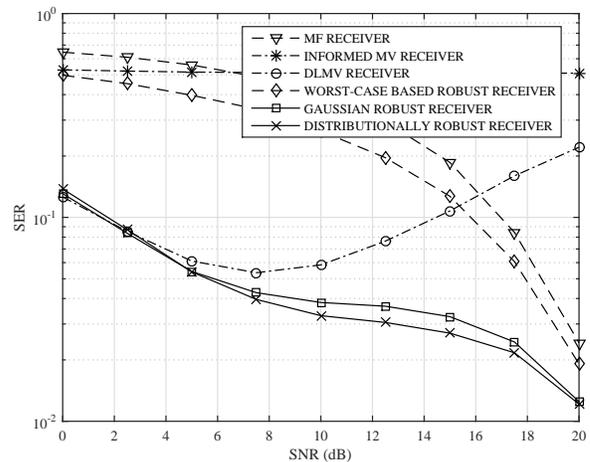
receiver [5]. Note that the latter receiver does not correspond to any practical situation and is included in our simulations for the sake of comparison only. The distributionally robust receiver (36)-(40) and the Gaussian robust receiver (22)-(23) are not considered here since their performance will degrade when the SNR increases according to [5], [3]. This is because the self-interferences are not suppressed as mentioned in Section III-A.

In our first example, we set  $I = 2$ ,  $N = T = 2$ ,  $M = 8$ , and adopt the Alamouti's code [22]. Hence, there is  $K = 2$ . The number of data blocks  $J$  is fixed at 35. The channel mismatch  $\Delta_i$  is independent of  $\mathbf{H}_i$  and has independent and identically distributed (i.i.d.) Gaussian entries with  $[\Delta_i]_{n,m} \sim \mathcal{CN}(0, \sigma_h^2)$  and  $[\mathbf{H}_i]_{n,m} \sim \mathcal{CN}(0, 1)$ .

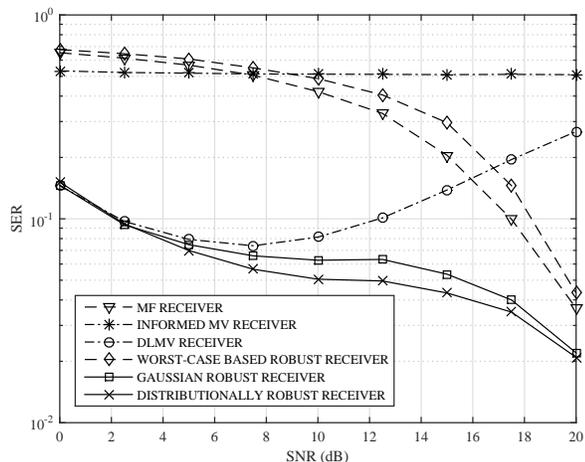
We first set  $\sigma_h^2 = 0.1$  and plot the SERs as a function of the SNR for all the receivers tested in Fig. 1a. Then, we set  $\sigma_h^2 = 0.5$  and plot the SERs versus SNR for all the receivers tested in Fig. 1b. From Fig. 1a, we observe that all the advanced robust receivers have similar performances



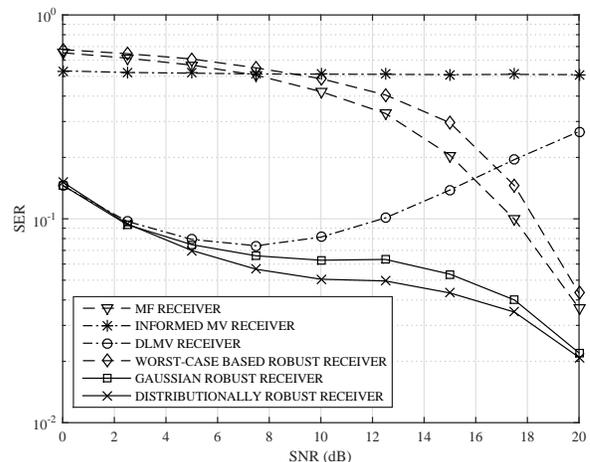
(a)  $\sigma_{h,1}^2 = 0.3, \lambda_1 = 0.9, \sigma_{h,2}^2 = 5, \lambda_2 = 0.1$ .



(b)  $\sigma_{h,1}^2 = 0.3, \lambda_1 = 0.9, \sigma_{h,2}^2 = 10, \lambda_2 = 0.1$ .



(c)  $\sigma_{h,1}^2 = 0.5, \lambda_1 = 0.9, \sigma_{h,2}^2 = 5, \lambda_2 = 0.1$ .



(d)  $\sigma_{h,1}^2 = 0.5, \lambda_1 = 0.9, \sigma_{h,2}^2 = 10, \lambda_2 = 0.1$ .

Fig. 2: Example Two, Alamouti's code, i.i.d. Gaussian mixture CSI mismatch,  $I = 2, N = K = T = 2, M = 8$ , and  $J = 35$ .

and are much superior than the other receivers. Among the advanced receivers, the Gaussian robust receiver performs slightly better than the other two receivers. However, when we increase  $\sigma_h^2$  to 0.5, we observe from Fig. 1b that the proposed distributionally robust receiver performs better than the Gaussian robust receiver. It implies that the proposed receiver appears to be more robust in the case of large CSI mismatch. In addition, we also observe from Fig. 1b that the performance degradation is severe for the worst-case based receiver in such situation.

In the second example, we set  $I = 2, N = K = T = 2, M = 8, J = 35, [\mathbf{H}_i]_{n,m} \sim \mathcal{CN}(0,1)$ , and test all the receivers in non-Gaussian CSI mismatch scenarios. The Alamouti's code is used in this example. Firstly, we consider the Gaussian mixture model, which is widely used to approximate the non-Gaussian noise in communication channels [23]. The

probability density function (pdf) of  $[\Delta_i]_{n,m}$  is given as

$$f([\Delta_i]_{n,m}) = \sum_{l=1}^L \frac{\lambda_l}{\pi \sigma_{h,l}^2} \exp \left\{ -\frac{|[\Delta_i]_{n,m}|^2}{\sigma_{h,l}^2} \right\} \quad (76)$$

where  $\sum_{l=1}^L \lambda_l = 1$ . According to [23], (76) is a spherically symmetric, bivariate pdf for the complex-valued random variable  $[\Delta_i]_{n,m}$ . In particular, for the case of  $L = 2$ , it is a typical model for impulsive noise if  $\sigma_{h,2} \gg \sigma_{h,1}$  and  $\lambda_2 < \lambda_1$ . Fig. 2a shows the receiver SERs versus the SNR for a CSI mismatch scenario with  $\sigma_{h,1}^2 = 0.3, \lambda_1 = 0.9, \sigma_{h,2}^2 = 5$ , and  $\lambda_2 = 0.1$ , while Fig. 2b displays the receiver SERs versus SNR for the CSI mismatch environment where  $\sigma_{h,1}^2 = 0.3, \lambda_1 = 0.9, \sigma_{h,2}^2 = 10$ , and  $\lambda_2 = 0.1$ . Similarly, Fig. 2c and Fig. 2d show the SERs versus the SNR where the Gaussian mixture model has the setting with  $\sigma_{h,1}^2 = 0.5, \lambda_1 = 0.9, \sigma_{h,2}^2 = 5, \lambda_2 = 0.1$  and  $\sigma_{h,1}^2 = 0.5, \lambda_1 = 0.9, \sigma_{h,2}^2 = 10, \lambda_2 = 0.1$ , respectively. In Fig. 2, it shows that the proposed robust receiver has the best performance among all the receivers. We also observe that

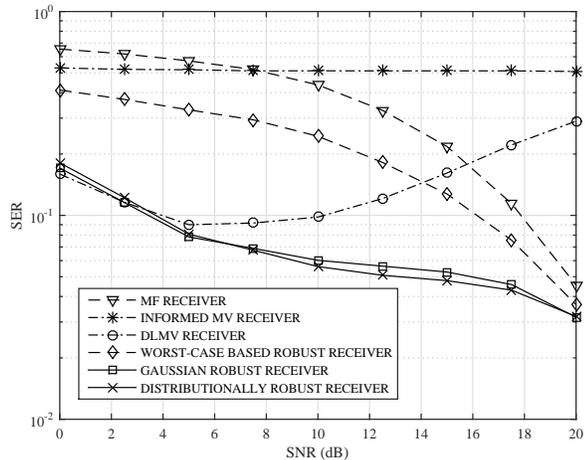
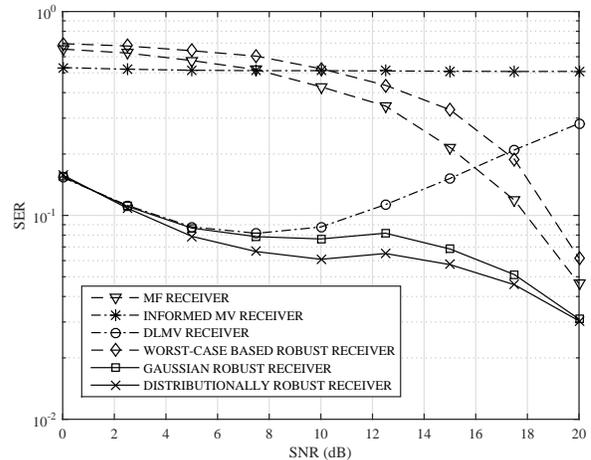


Fig. 3: Example Two. Alamouti's code, i.i.d. Laplacian CSI mismatch,  $I = 2$ ,  $N = K = T = 2$ ,  $M = 8$ , and  $J = 35$ .



(a) Alamouti's code,  $K = 2$ ,  $M = 8$ , and  $J = 35$ .

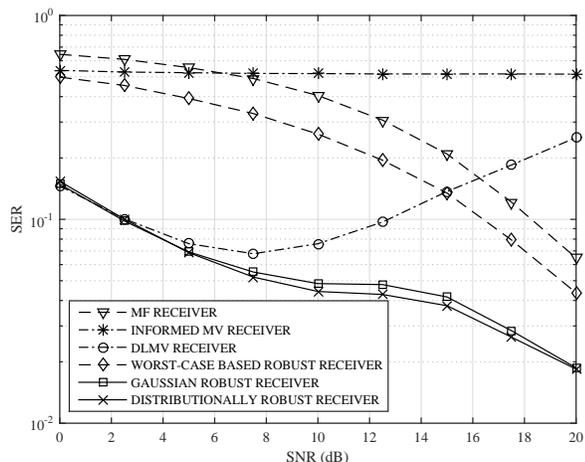
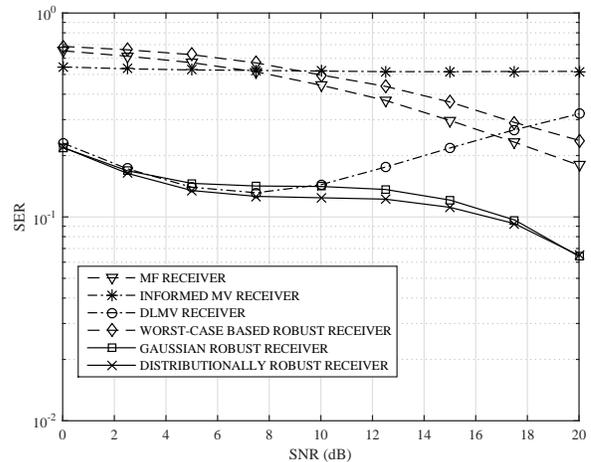


Fig. 4: Example Three. LD code, i.i.d. Gaussian mixture CSI mismatch,  $I = 2$ ,  $N = T = 2$ ,  $K = 4$ ,  $M = 16$ , and  $J = 70$ .



(b) LD code,  $K = 4$ ,  $M = 16$ , and  $J = 70$ .

Fig. 5: Example Four. Correlated Gaussian mixture CSI mismatch,  $I = 2$  and  $N = T = 2$ .

the performance gain of the proposed receiver is more obvious in the case where  $\sigma_{h,2}$  is larger. An interesting phenomenon is that the worst-case based receiver experiences a severe performance degradation when  $\sigma_{h,2}$  increases.

Secondly, a Laplacian CSI mismatch [24] is considered with zero-mean and variance of  $\sigma_h^2$ . The pdf of  $[\Delta_i]_{n,m}$  is given by

$$f([\Delta_i]_{n,m}) = \frac{1}{\sigma_h} \exp \left\{ -\frac{2}{\sigma_h} \left( \left| \text{Re}\{[\Delta_i]_{n,m}\} \right| + \left| \text{Im}\{[\Delta_i]_{n,m}\} \right| \right) \right\}.$$

We set  $\sigma_h^2 = 0.3$  and plot SERs versus the SNR in Fig. 3. The performance of the proposed receiver is better than other receivers as shown in Fig. 3.

In the third example, we use a higher rate LD code in [18]. In particular, we set  $I = 2$ ,  $N = T = 2$ ,  $K = 4$ ,  $M = 16$ , and  $J = 70$ . Thus, the rate of this code  $K/T = 2$  is higher than

that of the Alamouti's code where  $K/T = 1$ . In this code,  $C_k$ ,  $k = 1, 2, 3, 4$ , are chosen as

$$C_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, C_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \\ C_3 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, C_4 = \frac{1}{\sqrt{2}} \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \quad (77)$$

and  $D_k = jC_k$ ,  $k = 1, 2, 3, 4$ . In Fig. 4, we plot SERs versus the SNR with a Gaussian mixture CSI mismatch (76) where  $\sigma_{h,1}^2 = 0.3$ ,  $\lambda_1 = 0.9$ ,  $\sigma_{h,2}^2 = 10$ , and  $\lambda_2 = 0.1$  (the same settings as those in Fig. 2b). Compared with Fig. 2b, we observe that the SERs of all receivers increase in Fig. 4. The can be interpreted by the typical trade-off between the coding rate and SER in space-time block coded MIMO systems. Nevertheless, it can be seen from Fig. 4 that the proposed robust receiver still yields the lowest SER among all receivers tested.

In the fourth example, we consider a scenario where the channel elements are not i.i.d. The spatial correlation of

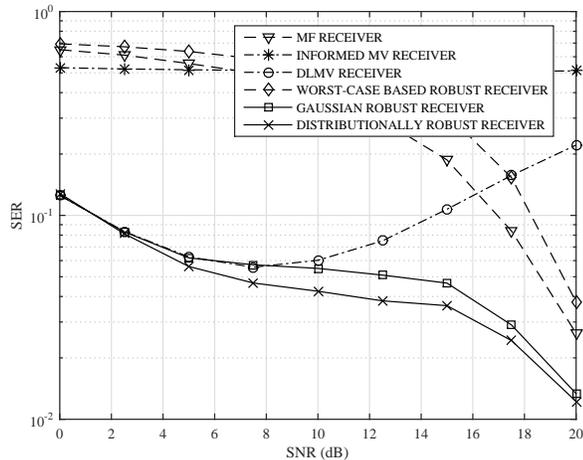


Fig. 6: Example Five. OSTBC, i.i.d. Gaussian mixture CSI mismatch,  $I = 4$ ,  $N = K = 3$ ,  $T = 4$ ,  $M = 12$ , and  $J = 100$ .

the channel at the transmitter side and the receiver side is represented as [25]

$$\mathbf{H}_i = \mathbf{A}_{\Theta_i} \mathbf{H}_{i,w} \mathbf{A}_{\Phi}^H, \quad \Delta_i = \mathbf{A}_{\Theta_i} \Delta_{i,w} \mathbf{A}_{\Phi}^H, \quad i = 1, \dots, I \quad (78)$$

where  $\mathbf{A}_{\Theta_i} \mathbf{A}_{\Theta_i}^H = \Theta_i$  and  $\mathbf{A}_{\Phi} \mathbf{A}_{\Phi}^H = \Phi^T$  stand for the correlation at the transmitter and receiver side of the  $i$ th user, respectively,  $[\mathbf{H}_{i,w}]_{n,m} \sim \mathcal{CN}(0,1)$ , and a Gaussian mixture distribution (76) is used for elements in  $\Delta_{i,w}$  with  $\sigma_{h,1}^2 = 0.5$ ,  $\lambda_1 = 0.9$ ,  $\sigma_{h,2}^2 = 10$ , and  $\lambda_2 = 0.1$  (the same settings as those in Fig. 2d). Here  $(\cdot)^H$  denotes the matrix Hermitian transpose. As all user channels have the same  $\Phi$ , (78) also models the inter-user channel correlation. In this example, similar to [25], we choose  $[\Theta_i]_{m,n} = \theta_i^{|m-n|}$ ,  $m, n = 1, \dots, N$ ,  $i = 1, \dots, I$ ,  $[\Phi]_{m,n} = \phi^{|m-n|}$ ,  $m, n = 1, \dots, M$ , and set  $I = 2$ ,  $N = T = 2$ ,  $\theta_1 = 0.5$ ,  $\theta_2 = 0.4$ , and  $\phi = 0.2$ .

We first consider the Alamouti's code and follow the settings in the second example ( $K = 2$ ,  $M = 8$ , and  $J = 35$ ) and plot SERs versus the SNR in Fig. 5a. Then, we use the higher rate code in Example Three (77) and set  $K = 4$ ,  $M = 16$ , and  $J = 70$ . SERs versus the SNR is plotted in Fig. 5b. It can be seen from Figs. 5a and 5b that similar to the point-to-point case [26], the SERs of all receivers increase compared with those in Figs. 2d and 4, respectively, due to the channel correction. Nevertheless, seen from Figs. 5a and 5b, the distributionally robust receiver still has the lowest SER among all receivers tested. Comparing Fig. 5a with Fig. 5b, the trade-off between coding rate and SER is observed again as expected.

In the last example, we simulate a multi-access MIMO system with  $I = 4$  and  $N = 3$ . In particular, we set  $K = 3$ ,  $T = 4$ ,  $M = 12$ ,  $J = 100$ , and use the  $K/T = 3/4$  rate orthogonal STBC (OSTBC) in [1]. A Gaussian mixture CSI mismatch is used with  $\sigma_{h,1}^2 = 0.3$ ,  $\lambda_1 = 0.9$ ,  $\sigma_{h,2}^2 = 10$ , and  $\lambda_2 = 0.1$  (the same settings as those in Figs. 2b and 4). The SER of all algorithms tested versus the SNR is shown in Fig. 6. It can be clearly seen from Fig. 6 that the performance of the proposed robust receiver is better than the other receivers.

## V. CONCLUSIONS

A distributionally robust receiver design is proposed in this paper. The proposed receiver works better than the existing receivers in most cases in our simulations. The advantage of the proposed receiver is more obvious for the situation where the variance of the CSI mismatch is large. The worst-case based receiver is shown to be sensitive to the CSI mismatch with a large variance.

## REFERENCES

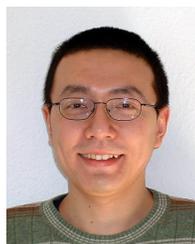
- [1] V. Tarokh, H. Jafarkhani, and A. R. Calderbank, "Space-time block codes from orthogonal designs," *IEEE Trans. Inf. Theory*, vol. 45, pp. 1456-1467, July 1999.
- [2] S. Shahbazpanahi, M. Beheshti, A. B. Gershman, M. Gharavi-Alkhansari, and K. M. Wong, "Minimum variance linear receivers for multi-access MIMO wireless systems with space-time block coding," *IEEE Trans. Signal Process.*, vol. 52, pp. 3306-3313, Dec. 2004.
- [3] Y. Rong, S. Shahbazpanahi, and A. B. Gershman, "Robust linear receivers for space-time block coded multi-access MIMO systems with imperfect channel state information," *IEEE Trans. Signal Process.*, vol. 53, pp. 3081-3090, Aug. 2005.
- [4] Y. Rong, S. Shahbazpanahi, and A. B. Gershman, "Exploiting the structure of OSTBCs to improve the robustness of worst-case optimization based linear multi-user MIMO receivers," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Process.*, Philadelphia, PA, Mar. 2005, vol. 4, pp. 781-784.
- [5] Y. Rong, S. Vorobyov, and A. B. Gershman, "Robust linear receivers for multiaccess space-time block-coded MIMO systems: A probabilistically constrained approach," *IEEE J. Selected Areas Commun.*, vol. 24, no. 8, pp. 1560-1570, Aug. 2006.
- [6] S. Zymler, D. Kuhn, and B. Rustem, "Distributionally robust joint chance constraints with second-order moment information," *Math. Program. Ser. A*, vol. 137, pp. 167-198, 2013.
- [7] G. Calafiore and L. El Ghaoui, "Distributionally robust chance constraints linear programs with applications," *J. Optim. Theory Appl.*, vol. 130, no. 1, pp. 1-22, 2006.
- [8] E. Delage and Y. Ye, "Distributionally robust optimization under moment uncertainty with application to data-driven problems," *Oper. Res.*, vol. 58, no. 3, pp. 595-612, 2010.
- [9] A. M.-C. So, "Moment inequalities for sums of random matrices and their applications in optimization," *Math. Program.*, vol. 130, no. 1, pp. 125-151, 2011.
- [10] A. Ruszczyński, "Probabilistic programming with discrete distributions and precedence constrained knapsack polyhedra," *Math. Program.*, vol. 93, no.2, pp. 195-215, 2002.
- [11] A. Ben-Tal and M. Teboulle, "Expected utility, penalty functions and duality in stochastic nonlinear programming," *Management Sci.*, vol. 32, pp. 1445-1466, 1986.
- [12] R. T. Rockafellar and S. Uryasev, "Optimization of conditional value-at-risk," *J. Risk*, vol. 2, pp. 21-41, 2000.
- [13] A. Nemirovski and A. Shapiro, "Convex approximations of chance constrained programs," *SIAM J. Optim.*, vol. 17, no. 4, pp. 969-996, 2007.
- [14] W. Chen, M. Sim, J. Sun, and C. P. Teo, "From CVaR to uncertainty set: Implications in joint-chance-constrained optimization," *Oper. Res.*, vol. 58, no. 2, pp. 470-485, 2010.
- [15] A. Prékopa, *Stochastic Programming*. Dordrecht, Netherlands: Kluwer Academic Publishers, 1995.
- [16] Q. Li, A. M.-C. So, and W. K. Ma, "Distributionally robust chance-constrained transmit beamforming for multiuser MISO downlink," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Process.*, Florence, Italy, May. 2014, pp. 3479-3483.
- [17] A. M.-C. So, and Y. J. Zhang, "Distributionally robust slow adaptive OFDMA with soft QoS via linear programming," *IEEE J. Selected Areas Commun.*, vol. 31, no. 5, pp. 947-958, May 2013.
- [18] B. Hassibi and B. M. Hochwald, "High-rate codes that are linear in space and time," *IEEE Trans. Inf. Theory*, vol. 48, pp. 1804-1824, July 2002.
- [19] M. Gharavi-Alkhansari and A. B. Gershman, "Constellation space invariance of orthogonal space-time block codes," *IEEE Trans. Inf. Theory*, vol. 51, pp. 331-334, Jan. 2005.
- [20] J. W. Brewer, "Kronecker products and matrix calculus in system theory," *IEEE Trans. Circuits and Systems*, vol. 25, pp. 772-781, Sep. 1978.

- [21] M. S. Lobo, L. Vandenberghe, and S. Boyd, "Applications of second-order cone programming," *Linear Algebra and its Applications*, vol. 284, pp. 193-228, Nov. 1998.
- [22] S. M. Alamouti, "A simple transmitter diversity scheme for wireless communications," *IEEE J. Selected Areas Commun.*, vol. 16, pp. 1451-1458, Oct. 1998.
- [23] R. J. Kozick and B. M. Sadler, "Maximum-likelihood array processing in non-Gaussian noise with Gaussian mixtures," *IEEE Trans. Signal Process.*, vol. 48, pp. 3520-3535, Aug. 2000.
- [24] S. A. Vorobyov, Y. Rong, N. D. Sidiropoulos, and A. B. Gershman, "Robust iterative fitting of multilinear models," *IEEE Trans. Signal Process.*, vol. 53, pp. 2678-2689, Aug. 2005.
- [25] Z. He, W. Jiang, and Y. Rong, "Robust design for amplify-and-forward MIMO relay systems with direct link and imperfect channel information," *IEEE Trans. Wireless Commun.*, vol. 14, pp. 353-363, Jan. 2015.
- [26] B. Clerckx and C. Oestges, "Finite-SNR performance analysis of space-time coding in correlated Ricean MIMO channels," *IEEE Trans. Inf. Theory*, vol. 53, pp. 4761-4777, Dec. 2007.



**Bin Li** received the Bachelor degree in automation and the Master degree in control science and engineering from Harbin Institute of Technology, China, in 2005 and 2008, respectively, and Ph.D. degrees in mathematics and statistics from Curtin University, Australia, in 2011. From 2012 - 2014, he was a Research Associate with the School of Electrical, Electronic and Computer Engineering, the University of Western Australia, Australia. Since 2014, he is a Research Fellow with the Department of Mathematics and Statistics, Curtin University,

Australia. Currently, he is also with the School of Information Engineering, Guangdong University of Technology. His research interests include signal processing, wireless communications, optimization and optimal control.



**Yue Rong** (S'03-M'06-SM'11) received the Ph.D. degree (summa cum laude) in electrical engineering from the Darmstadt University of Technology, Darmstadt, Germany, in 2005.

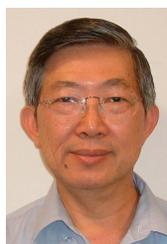
He was a Post-Doctoral Researcher with the Department of Electrical Engineering, University of California, Riverside, from February 2006 to November 2007. Since December 2007, he has been with the Department of Electrical and Computer Engineering, Curtin University, Bentley, Australia, where he is currently a Full Professor. His research

interests include signal processing for communications, wireless communications, underwater acoustic communications, applications of linear algebra and optimization methods, and statistical and array signal processing. He has published over 140 journal and conference paper in these areas.

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



**Jie Sun** (PhD in Applied mathematics, University of Washington, 1986)'s research interest focuses on operations research, in particular in theory and computation of optimization models. His area of expertise includes continuous optimization and decision under uncertainty. Specifically, he has been a frontier researcher in semismooth Newton's methods, interior point methods, and matrix variational inequalities. He was a winner of Outstanding Researcher Award of National University of Singapore and is currently a professor at Curtin University, Western Australia.



**Kok Lay Teo** (M74SM87) received the B.Sc. degree in telecommunications engineering from Ngee Ann Technical College, Singapore, and the M.A.Sc and Ph.D. degrees in electrical engineering from the University of Ottawa, Canada. He was with the Department of Applied Mathematics, University of New South Wales, Australia, the Department of Industrial and Systems Engineering, National University of Singapore, Singapore, the Department of Mathematics, the University of Western Australia, Australia. In 1996, he joined the Department of Mathematics and

Statistics, Curtin University of Technology, Australia, as Professor. He then took up the position of Chair Professor of Applied Mathematics and Head of Department of Applied Mathematics at the Hong Kong Polytechnic University, China, from 1999 to 2004. He is currently John Curtin Distinguished Professor at Curtin University. He has published 5 books and over 350 journal papers. He has a software package, MISER3.3, for solving general constrained optimal control problems. His editorial positions include serving as Editor-in-Chief of the Journal of Industrial and Management Optimization, and Numerical Algebra, Control and Optimization, and as a member of editorial board of Automatica, Journal of Global Optimization, Journal of Optimization Theory and Applications, Optimization and Engineering, Discrete and Continuous Dynamic Systems, Optimization Letters, Differential Equations and Dynamical Systems, and Applied Mathematical Modeling. His research interests include both the theoretical and practical aspects of optimal control and optimization, and their practical applications such as in signal processing in telecommunications, and financial portfolio optimization.