

Robust Linear Receivers for Multiaccess Space–Time Block-Coded MIMO Systems: A Probabilistically Constrained Approach

Yue Rong, *Member, IEEE*, Sergiy A. Vorobyov, *Senior Member, IEEE*, and Alex B. Gershman, *Fellow, IEEE*

Abstract—Traditional multiuser receiver algorithms developed for multiple-input–multiple-output (MIMO) wireless systems are based on the assumption that the channel state information (CSI) is precisely known at the receiver. However, in practical situations, the exact CSI may be unavailable because of channel estimation errors and/or outdated training. In this paper, we address the problem of robustness of multiuser MIMO receivers against imperfect CSI and propose a new linear technique that guarantees the robustness against CSI errors with a certain selected probability.

The proposed receivers are formulated as probabilistically constrained stochastic optimization problems. Provided that the CSI mismatch is Gaussian, each of these problems is shown to be convex and to have a unique solution. The fact that the CSI mismatch is Gaussian also enables to convert the original stochastic problems to a more tractable deterministic form and to solve them using the second-order cone programming approach.

Numerical simulations illustrate an improved robustness of the proposed receivers against CSI errors and validate their better flexibility as compared with the robust multiuser MIMO receivers based on the worst case designs.

Index Terms—Multiple-input–multiple-output (MIMO) systems, multiaccess communications, optimization methods, probability-constrained optimization, robustness.

I. INTRODUCTION

IN UPLINK CELLULAR communications with multiple receive antennas at the base station and transmit antennas at each mobile station, spatial diversity techniques can be employed to increase the capacity [1]–[3] and improve the immunity to fading [4]–[6]. Among space–time codes developed so far, orthogonal space–time block codes (OSTBCs) are very popular [4], [6], mainly because they enable a very simple (linear) maximum-likelihood (ML) decoding in the point-to-point communication case and, at the same time, achieve the full diversity order [6].

Manuscript received September 1, 2005; revised April 15, 2006. The work of A. B. Gershman was supported in part by the Wolfgang Paul Award Program of the Alexander von Humboldt Foundation, Germany. This paper was presented in part at the IEEE Vehicular Technology Conference–Spring, May 2004, Milan, Italy, and in part at the IEEE Workshop on Statistical Signal Processing, July 2005, Bordeaux, France.

Y. Rong was with the Department of Communication Systems, Darmstadt University of Technology, Darmstadt D-64283, Germany. He is now with the Department of Electrical Engineering, University of California, Riverside, CA 92521 USA (e-mail: yue.rong@ieee.org).

S. A. Vorobyov and A. B. Gershman are with the Department of Communication Systems, Darmstadt University of Technology, Merckstr. 25, Darmstadt D-64283, Germany (e-mail: svor@ieee.org; gershman@ieee.org).

Digital Object Identifier 10.1109/JSAC.2006.879379

In the multiaccess MIMO case, the optimal ML receiver becomes prohibitively expensive and, hence, suboptimal linear multiuser receivers have gained recently much interest as computationally attractive alternatives to the ML receiver [7]–[11]. Unfortunately, all the receivers of [7]–[11] assume that the exact knowledge of the channel state information (CSI) of at least the user-of-interest is available. When the exact CSI is unavailable, the performance of these receivers may degrade severely.

Motivated by the latter fact, the problem of robust linear multiuser MIMO receiver design has been recently addressed in [12] and [13], where the worst case optimization approach [14]–[21] has been used to improve the robustness of the minimum variance (MV) receivers of [11] against CSI mismatches. However, the actual worst case may occur in practice with a very low probability. Hence, the worst case approach may be overly pessimistic and, therefore, may lead to unnecessary performance degradation.

In this paper (see also [22] and [23]), we design robust linear multiuser MIMO receivers which are less conservative than the worst case optimization-based receivers of [12] and [13]. Our new receivers guarantee the robustness against CSI errors with a certain selected (high) probability.¹ The mathematical formulation of this receiver design problem is developed in terms of probability-constrained stochastic optimization (also referred to as *chance programming*) [27], [28]. Based on this approach, we formulate robust receiver design problems whose convexity is then proven using the assumption that the CSI mismatch is Gaussian. These problems are further converted into equivalent deterministic forms based on nonlinear programming (NLP) [29] and/or second-order cone programming (SOCP) [30]–[32]. The latter problems can be efficiently solved using modern convex optimization algorithms [32], [33]. An interesting relationship between the proposed techniques and the worst case optimization-based receivers is also established.

The remainder of this paper is organized as follows. Some background on multiaccess space–time block-coded (STBC) MIMO systems and linear multiuser receiver algorithms is given in Section II. In Section III, we formulate new robust linear receivers based on probability-constrained stochastic optimization. Then, we prove the convexity of the obtained stochastic optimization problems and develop computationally efficient techniques to solve them. Section IV presents simulation results that compare the performance of the proposed receivers with the existing techniques. Section V concludes our work.

¹Interestingly, related probabilistic approaches have been recently applied to several other important problems including robust beamforming [24], robust multiuser detection [25], and robust signal parameter estimation [26].

II. BACKGROUND

A. Multiaccess STBC MIMO Systems

Let us consider an uplink multiuser MIMO communication system with multiple transmitters, single receiver, and flat block fading. Each transmitter is assumed to have the same number of antennas and to encode information-bearing symbols using the same STBC.² Then, the received signal can be written as [11]

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

where

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

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

$$\mathbf{V} \triangleq [\mathbf{v}^T(1) \cdots \mathbf{v}^T(T)]^T \quad (4)$$

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

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

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

$$\mathbf{v}(t) \triangleq [v_1(t) \cdots v_M(t)] \quad (7)$$

are the complex row vectors of the received signals, transmitted signals of the i th user, and noise, respectively. In the sequel, we make use of the following assumptions.

- The entries of \mathbf{H}_i , $i = 1, 2, \dots, I$ are independent identically distributed (i.i.d.) complex zero-mean Gaussian, that is, the channel is Rayleigh flat fading;
- The noise is spatially and temporally i.i.d. additive white Gaussian noise (AWGN).

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

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

where K is the constellation size. It can be shown that for any linear dispersion (LD) code, the code matrix $\mathbf{X}(\mathbf{s}_i)$ can be written as [34], [35]

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

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), one can rewrite (1) as [11]

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

²The assumption of the same number of antennas and the same STBC across all transmitters is only needed for notational simplicity and can be relaxed, see [11] for more detail.

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

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

$\text{vec}(\cdot)$ 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 [35]

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

B. Linear Receivers

Without any loss of generality, let us assume that the first user is the user-of-interest. The estimate of the data vector $\hat{\mathbf{s}}_1$ at the output of a linear receiver can be expressed as

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

where

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

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 $\underline{\mathbf{s}}_1$. Given the matrix \mathbf{W} , the estimate of the vector of information symbols of the transmitter-of-interest can be computed as

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

where \mathbf{I}_K denotes an $K \times K$ identity matrix.

The problem of linear receiver design is how to find the matrix \mathbf{W} that extracts the signals received from the user-of-interest, while rejecting the interference and noise components. It has been proposed in [11] to design \mathbf{W} using the MV approach. According to it, each entry of $\underline{\mathbf{s}}_1$ is estimated by minimizing the receiver output power, while preserving a unity gain for this particular entry of $\underline{\mathbf{s}}_1$. The corresponding optimization problem can be written as [11]

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

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.

The solution to (16) is given by

$$\mathbf{w}_{\text{MV},k} = \frac{1}{\mathbf{a}_k^T(\mathbf{H}_1)\hat{\mathbf{R}}^{-1}\mathbf{a}_k(\mathbf{H}_1)}\hat{\mathbf{R}}^{-1}\mathbf{a}_k(\mathbf{H}_1) \quad (19)$$

for $k = 1, \dots, 2K$. The form of the MV receiver in (19) is rather similar to that of the MV distortionless response (MVDR) beamformer [36] and the minimum output energy (MOE) receiver used in single-input–single-output (SISO) multiuser detection [37]. The receiver (19) can be also interpreted in terms of *user-separating* receivers; see [38] where such receivers have been developed for multiuser code-division multiple-access (CDMA) systems.

Although the receiver (19) is able to reject multiaccess interference (MAI), it does not completely cancel self-interference, which, for each \mathbf{w}_k , is caused by the other than the k th entries of $\underline{\mathbf{s}}_1$. Note that the complete cancellation of self-interference is a strongly desirable feature as, otherwise, the symbol-by-symbol detector is far from being optimal [11], [25]. More specifically, the symbol-by-symbol detector uses the assumption that the output of each linear receiver corresponding to any particular symbol is independent of the other symbols. However, this assumption is violated in the presence of even a small amount of noncancelled self-interference.

To incorporate the self-interference cancellation feature into (16), it was proposed in [11] to use additional zero-forcing (ZF) constraints

$$\mathbf{a}_l^T(\mathbf{H}_1)\mathbf{w}_k = 0, \quad \text{for all } l \neq k \quad (20)$$

that guarantee complete cancellation of self-interference. With the additional constraints of (20), the problem (16) can be reformulated as [11]

$$\min_{\mathbf{W}} \text{tr}\{\mathbf{W}^T\hat{\mathbf{R}}\mathbf{W}\} \quad \text{subject to} \quad \mathbf{A}^T(\mathbf{H}_1)\mathbf{W} = \mathbf{I}_{2K}. \quad (21)$$

Using the Lagrange multiplier method, the solution to (21) can be written in the following form [11]:

$$\tilde{\mathbf{W}}_{\text{MV}} = \hat{\mathbf{R}}^{-1}\mathbf{A}(\mathbf{H}_1)\left[\mathbf{A}^T(\mathbf{H}_1)\hat{\mathbf{R}}^{-1}\mathbf{A}(\mathbf{H}_1)\right]^{-1}. \quad (22)$$

Note also that a conceptually similar ZFMV receiver has been earlier used in multiuser detection [25] and adaptive filtering [39, Ch. 3, Prob. 4].

It can be seen from (16) and (21) that the MV receivers require the CSI of the user-of-interest. However, in practice, it is unrealistic to obtain the exact CSI at the receiver because of a limited/outdated training, negative effects of interference and noise, and channel variability. Therefore, the performance of the MV receivers may be subject to a severe degradation due to CSI imperfections, and some robustness against imperfect CSI is required.

To improve the robustness of (19) and (22), it has been proposed in [11] to apply *ad hoc* diagonal loading (DL) by means of using the matrix $\hat{\mathbf{R}} + \nu\mathbf{I}_{2MT}$ instead of $\hat{\mathbf{R}}$, where ν is the DL

factor. The DL-based modifications of the MV receivers (19) and (22) can be written as

$$\mathbf{w}_{\text{DLMV},k} = \frac{1}{\mathbf{a}_k^T(\mathbf{H}_1)(\hat{\mathbf{R}} + \nu\mathbf{I}_{2MT})^{-1}\mathbf{a}_k(\mathbf{H}_1)} \times (\hat{\mathbf{R}} + \nu\mathbf{I}_{2MT})^{-1}\mathbf{a}_k(\mathbf{H}_1) \quad (23)$$

$$\tilde{\mathbf{W}}_{\text{DLMV}} = (\hat{\mathbf{R}} + \nu\mathbf{I}_{2MT})^{-1}\mathbf{A}(\mathbf{H}_1) \times \left[\mathbf{A}^T(\mathbf{H}_1)(\hat{\mathbf{R}} + \nu\mathbf{I}_{2MT})^{-1}\mathbf{A}(\mathbf{H}_1)\right]^{-1} \quad (24)$$

respectively. Unfortunately, the DLMV receivers (23) and (24) chose the DL parameter ν in an *ad hoc* way. As a result, the choice of this parameter may be nonoptimal.

To further improve the robustness of the receivers (23) and (24) against CSI errors, several worst case-based receiver designs have been proposed in [12] and [13]. The latter techniques have more rigorous theoretical background than the DL-based receivers, because they chose the DL parameter that is optimally matched to the worst case. As a result, the techniques of [12] and [13] substantially outperform the receivers (23) and (24) in scenarios with imperfect CSI.

III. PROBABILISTICALLY CONSTRAINED ROBUST LINEAR RECEIVERS

Although the worst case-based receiver designs provide excellent robustness against CSI imperfections, they may be overly pessimistic in practical applications, where the actual worst case may occur with a very low probability. Motivated by this fact, in this section, we develop a more flexible approach based on probabilistic constraints.

Let us consider the CSI error matrix of the user-of-interest. This matrix can be defined as

$$\Delta_1 \triangleq \mathbf{H}_1 - \hat{\mathbf{H}}_1 \quad (25)$$

where \mathbf{H}_1 and $\hat{\mathbf{H}}_1$ denote the actual channel matrix of the user-of-interest and its estimate available at the receiver, respectively. Using the notations of model (10), we can write that

$$\begin{aligned} \mathbf{e}_k(\Delta_1) &\triangleq \mathbf{a}_k(\mathbf{H}_1) - \mathbf{a}_k(\hat{\mathbf{H}}_1) \\ &= \underline{\mathbf{F}}_k\mathbf{H}_1 - \underline{\mathbf{F}}_k\hat{\mathbf{H}}_1 \\ &= \underline{\mathbf{F}}_k\Delta_1, \quad k = 1, \dots, 2K \end{aligned} \quad (26)$$

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).

Let us obtain the receiver weight vector \mathbf{w}_k for the k th entry of $\underline{\mathbf{s}}_1$ as the solution of the following probability-constrained optimization problem

$$\min_{\mathbf{w}_k} \mathbf{w}_k^T\hat{\mathbf{R}}\mathbf{w}_k \quad (28)$$

$$\text{subject to} \quad \Pr\left\{\mathbf{w}_k^T\left(\mathbf{a}_k(\hat{\mathbf{H}}_1) + \mathbf{e}_k(\Delta_1)\right) \geq 1\right\} \geq p. \quad (29)$$

In this problem, the receiver output power is minimized under the constraint that the distortionless response for the k th entry of $\underline{\mathbf{s}}_1$ is maintained with the probability not smaller than p , where

p is selected from the interval $(0,1)$, and $\Pr\{\cdot\}$ denotes the probability operator whose form is assumed to be known. In fact, the value of p in (28) and (29) determines the probability of outage³ that is allowed at the receiver. Therefore, to restrict the outage probability to be small, the typical choice of p should be close to one, that is, $1 - p \ll 1$. It is easy to make the particular choice of p based on quality-of-service (QoS) system specifications. Note that the problems of the type of (28) and (29) are referred in the optimization literature to as *chance-constrained* or *probability-constrained* stochastic programming problems [27], [28].

In what follows, we will assume that the CSI errors are Gaussian distributed. This assumption can be validated by the results of [40] and [41], where it has been shown that in MIMO communications with optimal (orthogonal) training signals and AWGN, the CSI errors are Gaussian.

The following theorem establishes the convexity of the problem (28) and (29).

Theorem 1: If the entries of Δ_1 are uncorrelated and each entry has circular complex Gaussian distribution so that $[\Delta_1]_{n,m} \sim \mathcal{CN}(0, \sigma_h^2)$, then the optimization problem (28) and (29) is convex for $p \in (0.5, 1)$.

Proof: The objective function in (28) is a quadratic function of \mathbf{w}_k , and $\hat{\mathbf{R}}$ is non-negative definite. Therefore, this function is convex.

Now, let us prove that the constraint (29) is also convex under the assumptions of the theorem. First, we note that $\mathbf{e}_k(\Delta_1)$ depends linearly on Δ_1 . Indeed, applying the underline operator (11) to (26) and using some properties of the Kronecker matrix product [42], we have

$$\begin{aligned}
 \mathbf{e}_k(\Delta_1) &= \begin{bmatrix} \text{vec}(\text{Re}\{\mathbf{F}_k \Delta_1\}) \\ \text{vec}(\text{Im}\{\mathbf{F}_k \Delta_1\}) \end{bmatrix} \\
 &= \begin{bmatrix} \text{Re}\{(\mathbf{I}_M \otimes \mathbf{F}_k) \text{vec}(\Delta_1)\} \\ \text{Im}\{(\mathbf{I}_M \otimes \mathbf{F}_k) \text{vec}(\Delta_1)\} \end{bmatrix} \\
 &= \begin{bmatrix} \text{Re}\{\mathbf{I}_M \otimes \mathbf{F}_k\} & -\text{Im}\{\mathbf{I}_M \otimes \mathbf{F}_k\} \\ \text{Im}\{\mathbf{I}_M \otimes \mathbf{F}_k\} & \text{Re}\{\mathbf{I}_M \otimes \mathbf{F}_k\} \end{bmatrix} \\
 &\quad \times \begin{bmatrix} \text{vec}(\text{Re}\{\Delta_1\}) \\ \text{vec}(\text{Im}\{\Delta_1\}) \end{bmatrix} \\
 &= \Psi_k \underline{\Delta}_1
 \end{aligned} \tag{30}$$

where

$$\Psi_k \triangleq \begin{bmatrix} \text{Re}\{\mathbf{I}_M \otimes \mathbf{F}_k\} & -\text{Im}\{\mathbf{I}_M \otimes \mathbf{F}_k\} \\ \text{Im}\{\mathbf{I}_M \otimes \mathbf{F}_k\} & \text{Re}\{\mathbf{I}_M \otimes \mathbf{F}_k\} \end{bmatrix} \tag{31}$$

and \otimes denotes the Kronecker matrix product. It follows from (30) that $\mathbf{e}_k(\Delta_1)$ is a linear combination of the real and imaginary parts of all entries of the CSI error matrix Δ_1 . If the entries of Δ_1 are uncorrelated and have circular complex Gaussian distribution

$$[\Delta_1]_{n,m} \sim \mathcal{CN}(0, \sigma_h^2), \quad n = 1, \dots, N, \quad m = 1, \dots, M \tag{32}$$

then, using (30), we find that $\mathbf{e}_k(\Delta_1)$ has multivariate real Gaussian distribution. Its mean vector and covariance matrix can be calculated, respectively, as

$$\text{E}\{\mathbf{e}_k(\Delta_1)\} = \Psi_k \text{E}\{\underline{\Delta}_1\} = \mathbf{0}_{2MT} \tag{33}$$

$$\begin{aligned}
 \text{E}\{\mathbf{e}_k(\Delta_1) \mathbf{e}_k^T(\Delta_1)\} &= \Psi_k \text{E}\{\underline{\Delta}_1 \underline{\Delta}_1^T\} \Psi_k^T \\
 &= \frac{\sigma_h^2}{2} \Psi_k \Psi_k^T
 \end{aligned} \tag{34}$$

³In the context of our paper, the *receiver outage* is defined as the case when the distortionless response constraint is violated.

where $\mathbf{0}_{2MT}$ denotes a $2MT \times 1$ vector with all zero entries.

Since the only random term in the product $\mathbf{w}_k^T(\mathbf{a}_k(\hat{\mathbf{H}}_1) + \mathbf{e}_k(\Delta_1))$ is $\mathbf{e}_k(\Delta_1)$, and both \mathbf{w}_k and $\mathbf{a}_k(\hat{\mathbf{H}}_1)$ are deterministic, the random variable

$$\gamma = \mathbf{w}_k^T (\mathbf{a}_k(\hat{\mathbf{H}}_1) + \mathbf{e}_k(\Delta_1)) \tag{35}$$

has also Gaussian distribution. Its mean and covariance can be computed, respectively, as

$$\begin{aligned}
 \text{E}\{\gamma\} &= \mathbf{w}_k^T \text{E}\{\mathbf{a}_k(\hat{\mathbf{H}}_1) + \mathbf{e}_k(\Delta_1)\} \\
 &= \mathbf{w}_k^T \mathbf{a}_k(\hat{\mathbf{H}}_1)
 \end{aligned} \tag{36}$$

$$\begin{aligned}
 \text{E}\{\gamma^2\} - \text{E}\{\gamma\}^2 &= \text{E}\{\mathbf{w}_k^T \mathbf{e}_k(\Delta_1) \mathbf{e}_k^T(\Delta_1) \mathbf{w}_k\} \\
 &= \mathbf{w}_k^T \text{E}\{\mathbf{e}_k(\Delta_1) \mathbf{e}_k^T(\Delta_1)\} \mathbf{w}_k \\
 &= \frac{\sigma_h^2}{2} \|\Psi_k^T \mathbf{w}_k\|^2
 \end{aligned} \tag{37}$$

where $\|\cdot\|$ denotes the Euclidean norm of a vector.

Using the error function for any real Gaussian random variable x

$$\text{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt \tag{38}$$

we can find the probability $\Pr\{x \geq 1\}$ as follows [43]:

$$\Pr\{x \geq 1\} = \frac{1}{2} - \frac{1}{2} \text{erf}\left(\frac{1 - \text{E}\{x\}}{\sqrt{2\text{E}\{(x - \text{E}\{x\})^2\}}}\right). \tag{39}$$

For the random variable γ in (35), (39) can be rewritten as

$$\Pr\{\gamma \geq 1\} = \frac{1}{2} - \frac{1}{2} \text{erf}\left(\frac{1 - \mathbf{w}_k^T \mathbf{a}_k(\hat{\mathbf{H}}_1)}{\sigma_h \|\Psi_k^T \mathbf{w}_k\|}\right). \tag{40}$$

Applying (40) to (29), we obtain the following equivalent constraint:

$$\text{erf}\left(\frac{\mathbf{w}_k^T \mathbf{a}_k(\hat{\mathbf{H}}_1) - 1}{\sigma_h \|\Psi_k^T \mathbf{w}_k\|}\right) \geq 2p - 1. \tag{41}$$

The inequality (41) is convex if and only if its left-hand side is positive. The latter is guaranteed if $2p - 1 > 0$ or, equivalently, $p > 0.5$. In this case, (41) can be rewritten as

$$\sigma_h \|\Psi_k^T \mathbf{w}_k\| \leq \frac{\mathbf{w}_k^T \mathbf{a}_k(\hat{\mathbf{H}}_1) - 1}{\text{erf}^{-1}(2p - 1)} \tag{42}$$

where $\text{erf}^{-1}(\cdot)$ denotes the inverse error function. The constraint (42) is called the second-order cone (SOC) constraint and is convex; see also [30, pp. 157–158] for an outline of the conversion of a probabilistic constraint to a SOCP constraint.

Summarizing, both the objective function and the constraint are convex. Therefore, the problem (28) and (29) is convex. \square

Although the problem (28) and (29) is convex, it is nonlinear and does not have closed-form solution. However, using (42) it can be equivalently converted to the following SOCP problem:

$$\min_{\mathbf{w}_k, \chi} \chi \quad (43)$$

$$\text{subject to } \|\mathbf{L}\mathbf{w}_k\| \leq \chi \quad (44)$$

$$\sigma_h \left\| \Psi_k^T \mathbf{w}_k \right\| \leq \frac{\mathbf{w}_k^T \mathbf{a}_k(\hat{\mathbf{H}}_1) - 1}{\text{erf}^{-1}(2p - 1)} \quad (45)$$

where

$$\hat{\mathbf{R}} = \mathbf{L}^T \mathbf{L} \quad (46)$$

is the Cholesky factorization of $\hat{\mathbf{R}}$, and χ is a new optimization variable such that $\|\mathbf{L}\mathbf{w}_k\| \leq \chi$. Using the standard optimization tools [33], the problem (43)–(45) can be solved with the complexity order of $\mathcal{O}(M^3 T^3)$ [45].

Interestingly, if $\mathbf{F}_k \mathbf{F}_k^H = \mathbf{I}_T$, then $\Psi_k \Psi_k^T = \mathbf{I}_{2MT}$ (note that, in particular, this property holds true for the Alamouti's code). Then, the constraint (45) is equivalent to

$$\text{erf}^{-1}(2p - 1) \sigma_h \|\mathbf{w}_k\| \leq \mathbf{w}_k^T \mathbf{a}_k(\hat{\mathbf{H}}_1) - 1. \quad (47)$$

In this specific case, we can establish a connection between the proposed receiver (43)–(45) and the worst case optimization-based receiver of [12]. The latter receiver can be represented as the following optimization problem:

$$\min_{\mathbf{w}_k, \chi} \chi \quad (48)$$

$$\text{subject to } \|\mathbf{L}\mathbf{w}_k\| \leq \chi \quad (49)$$

$$\varepsilon \|\mathbf{w}_k\| \leq \mathbf{w}_k^T \mathbf{a}_k(\hat{\mathbf{H}}_1) - 1 \quad (50)$$

where ε is a constant which bounds the uncertainty region of the CSI mismatch as

$$\|\Delta_1\| \leq \varepsilon. \quad (51)$$

Comparing (47) and (50), we obtain that the problems (43)–(45) and (48)–(50) become identical if

$$\varepsilon = \text{erf}^{-1}(2p - 1) \sigma_h. \quad (52)$$

Equation (52) illustrates an interesting link between the worst case-based and probabilistically constrained robust receiver designs. It provides an explicit relationship between the probability p and the uncertainty parameter ε and, therefore, provides a practical rule for choosing ε based on the QoS requirements.

In (28) and (29), both self-interference and MAI are suppressed by minimizing the objective function (28). In [12], it has been proposed to add separate constraints to the worst case optimization problem to mitigate self-interference in the CSI uncertainty region. With such additional constraints, the problem (48)–(50) was modified in [12] as

$$\min_{\mathbf{w}_k, \chi, \delta_k} \chi + \delta_k \quad (53)$$

$$\text{subject to } \|\mathbf{L}\mathbf{w}_k\| \leq \chi \quad (54)$$

$$\varepsilon \|\mathbf{w}_k\| \leq \mathbf{w}_k^T \mathbf{a}_k(\hat{\mathbf{H}}_1) - 1 \quad (55)$$

$$\left\| \mathbf{B}_k^T \mathbf{w}_k \right\| + \varepsilon \sqrt{2K - 1} \|\mathbf{w}_k\| \leq \delta_k \quad (56)$$

where

$$\mathbf{B}_k \triangleq \left[\mathbf{a}_1(\hat{\mathbf{H}}_1) \cdots \mathbf{a}_{k-1}(\hat{\mathbf{H}}_1) \mathbf{a}_{k+1}(\hat{\mathbf{H}}_1) \cdots \mathbf{a}_{2K}(\hat{\mathbf{H}}_1) \right] \quad (57)$$

and, similar to (48)–(50), χ is a new auxiliary variable. The latter problem is easily convertible to a SOCP form.

It has been demonstrated in [12] that, due to the aforementioned self-interference mitigation constraints, the receiver (53)–(56) significantly outperforms the receiver (48)–(50).

As mentioned before, the worst case receiver designs can be overly conservative as the probability of the actual worst case may be quite low [22]–[25]. Therefore, let us use a more flexible problem formulation with probabilistic constraints which guarantee that self-interference is cancelled with a certain selected probability p . With such additional constraints, the problem (28) and (29) can be modified as

$$\min_{\mathbf{w}_k, \delta} \mathbf{w}_k^T \hat{\mathbf{R}} \mathbf{w}_k + \|\delta\|^2 \quad (58)$$

$$\text{subject to } \Pr \left\{ \mathbf{w}_k^T \left(\mathbf{a}_k(\hat{\mathbf{H}}_1) + \mathbf{e}_k(\Delta_1) \right) \geq 1 \right\} \geq p \quad (59)$$

$$\Pr \left\{ \sigma_1 \left| \mathbf{w}_k^T \left(\mathbf{a}_l(\hat{\mathbf{H}}_1) + \mathbf{e}_l(\Delta_1) \right) \right| \leq \delta_l \right\} \geq p, \\ l = 1, \dots, 2K, l \neq k \quad (60)$$

where $\delta = [\delta_1, \dots, \delta_{k-1}, \delta_{k+1}, \dots, \delta_{2K}]^T$ is the $(2K - 1) \times 1$ vector whose values limit the contribution of self-interference, and σ_1 is the standard deviation of the waveform of the user-of-interest.

Note that the contribution of any particular user to the objective function (58) is proportional to its power. That is why the scaling by σ_1 is used in (60). The power of the user-of-interest is assumed to be approximately known (or estimated) at the receiver.

The following theorem establishes the convexity of the problem (58)–(60).

Theorem 2: If the elements of Δ_1 are uncorrelated and each element has circular complex Gaussian distribution $[\Delta_1]_{n,m} \sim \mathcal{CN}(0, \sigma_k^2)$, then the optimization problem (58)–(60) is convex if $p \in (0.5, 1)$.

Proof: See Appendix A. \square

Since the problem (58)–(60) is convex, a locally optimal solution is also globally optimal. This problem can be converted into its deterministic equivalent form in the following way.

Let us introduce a new vector

$$\zeta \triangleq \left[(\mathbf{L}\mathbf{w}_k)^T, \delta^T \right]^T. \quad (61)$$

Minimizing $\|\zeta\|$ is equivalent to minimizing the objective function of (58). Hence, introducing a new scalar non-negative variable η and a new constraint

$$\|\zeta\| \leq \eta \quad (62)$$

we can define a new objective function minimization as follows:

$$\min_{\mathbf{w}_k, \delta, \eta} \eta. \quad (63)$$

The deterministic equivalent form of the constraint (59) is given by (42). The probability constraints (60) also can be converted into their deterministic equivalents. Using (36), (37), and (38), the left-hand side of (60) can be written as

$$\begin{aligned} & \Pr \left\{ \sigma_1 \left| \mathbf{w}_k^T \left(\mathbf{a}_l(\hat{\mathbf{H}}_1) + \mathbf{e}_l(\Delta_1) \right) \right| \leq \delta_l \right\} \\ &= \Pr \left\{ \sigma_1 \mathbf{w}_k^T \left(\mathbf{a}_l(\hat{\mathbf{H}}_1) + \mathbf{e}_l(\Delta_1) \right) \leq \delta_l \right\} \\ & \quad - \Pr \left\{ \sigma_1 \mathbf{w}_k^T \left(\mathbf{a}_l(\hat{\mathbf{H}}_1) + \mathbf{e}_l(\Delta_1) \right) \leq -\delta_l \right\} \\ &= \frac{1}{2} \operatorname{erf} \left(\frac{\delta_l - \sigma_1 \mathbf{w}_k^T \mathbf{a}_l(\hat{\mathbf{H}}_1)}{\sigma_h \sigma_1 \left\| \Psi_l^T \mathbf{w}_k \right\|} \right) \\ & \quad - \frac{1}{2} \operatorname{erf} \left(\frac{-\delta_l - \sigma_1 \mathbf{w}_k^T \mathbf{a}_l(\hat{\mathbf{H}}_1)}{\sigma_h \sigma_1 \left\| \Psi_l^T \mathbf{w}_k \right\|} \right) \\ & \quad l = 1, \dots, 2K, \quad l \neq k. \end{aligned} \quad (64)$$

Combining (42), (62), (63), and (64) together, we can rewrite the stochastic programming problem (58)–(60) as the following equivalent deterministic problem:

$$\min_{\mathbf{w}_k, \delta, \eta} \eta \quad (65)$$

$$\text{subject to } \|\zeta\| \leq \eta \quad (66)$$

$$\sigma_h \left\| \Psi_k^T \mathbf{w}_k \right\| \leq \frac{\mathbf{w}_k^T \mathbf{a}_k(\hat{\mathbf{H}}_1) - 1}{\operatorname{erf}^{-1}(2p - 1)} \quad (67)$$

$$\begin{aligned} & \operatorname{erf} \left(\frac{\delta_l - \sigma_1 \mathbf{w}_k^T \mathbf{a}_l(\hat{\mathbf{H}}_1)}{\sigma_h \sigma_1 \left\| \Psi_l^T \mathbf{w}_k \right\|} \right) \\ & \quad - \operatorname{erf} \left(\frac{-\delta_l - \sigma_1 \mathbf{w}_k^T \mathbf{a}_l(\hat{\mathbf{H}}_1)}{\sigma_h \sigma_1 \left\| \Psi_l^T \mathbf{w}_k \right\|} \right) \geq 2p, \\ & \quad l = 1, \dots, 2K, \quad l \neq k \end{aligned} \quad (68)$$

where ζ is given by (61). Note that since the problem (65)–(68) is equivalent to the problem (58)–(60), the former problem is also convex.

Although the problems (58)–(60) and (65)–(68) are equivalent when the channel mismatch is Gaussian, the original stochastic programming problem is computationally intractable, whereas the deterministic problem (65)–(68) belongs to the class of NLP problems and can be numerically solved using interior-point algorithms [44] or sequential quadratic programming (SQP) [29] techniques. The latter technique is an iterative procedure in which each search direction is the solution of a particular quadratic programming (QP) subproblem. The computational complexity of solving each QP subproblem using the primal-dual potential reduction method is $\mathcal{O}(M^{4.5} N^{4.5} K^{4.5})$ [45]. Overall complexity of SQP algorithm is determined by the number of iterations, which may depend on problem-specific parameters and the given batch of data. As

a result, the overall computational complexity of solving the problem (65)–(68) may be prohibitively expensive for practical wireless communication systems.

Below we show that the problem (58)–(60) can be approximated to a much simpler SOCP form. The key idea is to approximate the nonlinear constraints (60) by SOC constraints. Let us use Chebyshev inequality, which states that for any random variable ξ and any positive real number α

$$\Pr \{ |\xi| \geq \alpha \} \leq \frac{\mathbb{E}\{\xi^2\}}{\alpha^2}. \quad (69)$$

Since all the constraints in (60) have the same structure, we further discuss only the l th constraint. Under the assumption that $\mathbf{e}_l(\Delta_1)$ has Gaussian distribution and using (36) and (37), we have

$$\begin{aligned} & \mathbb{E} \left\{ \left| \mathbf{w}_k^T \left(\mathbf{a}_l(\hat{\mathbf{H}}_1) + \mathbf{e}_l(\Delta_1) \right) \right|^2 \right\} \\ &= \mathbf{w}_k^T \left(\mathbf{a}_l(\hat{\mathbf{H}}_1) \mathbf{a}_l^T(\hat{\mathbf{H}}_1) + \frac{\sigma_h^2}{2} \Psi_l \Psi_l^T \right) \mathbf{w}_k. \end{aligned} \quad (70)$$

Using (69) and (70), the left-hand side of the l th constraint in (60) can be lower bounded as

$$\begin{aligned} & \Pr \left\{ \sigma_1 \left| \mathbf{w}_k^T \left(\mathbf{a}_l(\hat{\mathbf{H}}_1) + \mathbf{e}_l(\Delta_1) \right) \right| \leq \delta_l \right\} \\ &= 1 - \Pr \left\{ \sigma_1 \left| \mathbf{w}_k^T \left(\mathbf{a}_l(\hat{\mathbf{H}}_1) + \mathbf{e}_l(\Delta_1) \right) \right| \geq \delta_l \right\} \\ &\geq 1 - \frac{\sigma_1^2}{\delta_l^2} \mathbf{w}_k^T \left(\mathbf{a}_l(\hat{\mathbf{H}}_1) \mathbf{a}_l^T(\hat{\mathbf{H}}_1) + \frac{\sigma_h^2}{2} \Psi_l \Psi_l^T \right) \mathbf{w}_k. \end{aligned} \quad (71)$$

Replacing all the constraints (60) by their lower bounds (71), we obtain the following set of constraints:

$$\begin{aligned} & \frac{\sigma_1^2}{1-p} \mathbf{w}_k^T \left(\mathbf{a}_l(\hat{\mathbf{H}}_1) \mathbf{a}_l^T(\hat{\mathbf{H}}_1) + \frac{\sigma_h^2}{2} \Psi_l \Psi_l^T \right) \mathbf{w}_k \leq \delta_l^2, \\ & \quad l = 1, \dots, 2K, \quad l \neq k. \end{aligned} \quad (72)$$

The constraints (72) can be referred to as a *safe approximations* of the original constraints (60), meaning that the constraints in (72) are more strict than that in (60). Therefore, the constraints in (60) always hold true provided that those in (72) are satisfied.

For the sake of simplicity, we further approximate the constraints in (72) by summing them together to obtain a single constraint of the following form:

$$\mathbf{w}_k^T \mathbf{Q}_k \mathbf{w}_k \leq \|\delta\|^2 \quad (73)$$

where

$$\mathbf{Q}_k \triangleq \frac{\sigma_1^2}{1-p} \sum_{l=1, l \neq k}^{2K} \left[\mathbf{a}_l(\hat{\mathbf{H}}_1) \mathbf{a}_l^T(\hat{\mathbf{H}}_1) + \frac{\sigma_h^2}{2} \Psi_l \Psi_l^T \right]. \quad (74)$$

Substituting the left-hand side of (73) into the objective function (58) instead of the term $\|\delta\|^2$, we can eliminate the constraint (73) from the final optimization problem. Then, the new objective function can be written as follows:

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

The final optimization problem which approximates the original problem (58)–(60) can be written as the following SOCP problem:

$$\min_{\mathbf{w}_k, \tau} \tau \quad (76)$$

$$\text{subject to } \|\mathbf{Z}_k \mathbf{w}_k\| \leq \tau \quad (77)$$

$$\sigma_h \left\| \Psi_k^T \mathbf{w}_k \right\| \leq \frac{\mathbf{w}_k^T \mathbf{a}_k(\hat{\mathbf{H}}_1) - 1}{\text{erf}^{-1}(2p - 1)}. \quad (78)$$

where

$$\hat{\mathbf{R}} + \mathbf{Q}_k = \mathbf{Z}_k^T \mathbf{Z}_k \quad (79)$$

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

The problem (76)–(78) has the same structure as the problem (43)–(45). However, an important practical difference between (43)–(45) and (76)–(78) is that in (76)–(78) an additional mitigation of self-interference is used. It should be also stressed that the relaxed SOCP problem (76)–(78) has much lower computational complexity than the original NLP problem (65)–(68). The problem (76)–(78) can be easily solved using modern interior-point methods-based convex optimization software [33]. For example, using the primal-dual potential reduction method, this problem can be solved with the complexity order of $\mathcal{O}(M^3 T^3)$ [45].

Note also that the problem (76)–(78) is computationally simpler than (53)–(56), because it has less constraints and smaller number of optimization variables.

IV. SIMULATIONS

Throughout our simulations, an uplink cellular communication system with $I = 2$ transmitters and a single receiver equipped with $M = 8$ antennas is assumed. The interfering transmitter uses the same OSTBC as the transmitter-of-interest. The interference-to-noise ratio (INR) is equal to 20 dB and the quadrature phase-shift keying (QPSK) modulation scheme is used. All plots are averaged over 1000 independent simulation runs. The MIMO channel between the i th transmitter and the receiver is assumed to be quasi-static Rayleigh flat fading with $[\mathbf{H}_i]_{n,m} \sim \mathcal{CN}(0, 1)$.

The following receivers are compared in terms of symbol error rates (SERs): the proposed SOCP-based receiver (76)–(78), the worst case optimization-based receiver (53)–(56), the DLMV receiver (24) with the DL factor $\nu = 10\sigma_v^2$ (where σ_v^2 is the noise variance), the matched filter (MF) receiver [35]

$$\mathbf{W}_{\text{MF}} = \frac{1}{\|\hat{\mathbf{H}}_1\|^2} \mathbf{A}(\hat{\mathbf{H}}_1) \quad (80)$$

and the “informed” MV receiver, which corresponds to the ideal case when (22) is used with the exactly known \mathbf{H}_1 . 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 receivers (48)–(50) and (43)–(45) are not included in the subsequent plots because in [12] it has been shown that the receiver (53)–(56) has better performance than (48)–(50), and because the receiver (43)–(45) does not provide any performance improvements as compared with the receiver (48)–(50).

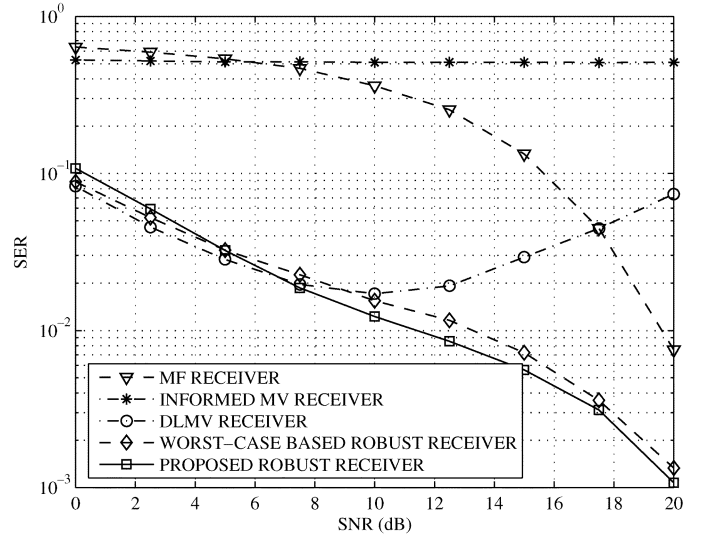


Fig. 1. SERs versus SNR; first example.

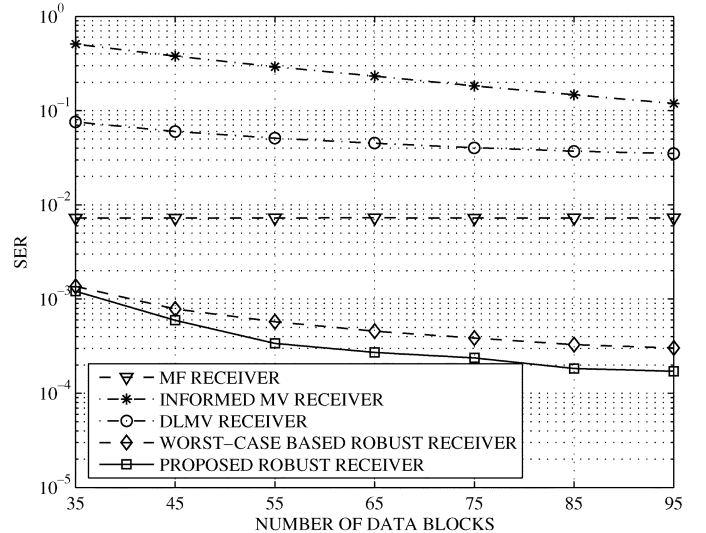


Fig. 2. SERs versus J ; first example.

The latter fact was confirmed through extensive simulations that are not included below for brevity reasons.

In our first and the second examples, the channel mismatch Δ_i is assumed to be independent on \mathbf{H}_i with $[\Delta_i]_{n,m} \sim \mathcal{CN}(0, \sigma_h^2)$ and $\sigma_h^2 = 0.1$ (the latter value of σ_h^2 corresponds to quite a substantial CSI estimation error). Note that to verify the proposed approach, we use in these two examples the same CSI mismatch model as in Theorems 1 and 2. In both these examples, $p = 0.95$ is taken for the proposed robust receiver (76)–(78).

In the first example, we simulate a scenario where each transmitter uses $N = 2$ antennas with the Alamouti’s code [4]. The parameter $\varepsilon = 6\sigma_h$ is used for the worst case-based robust receiver as suggested in [12]. In Fig. 1, SERs of all the receivers tested are displayed versus the SNR for the number of data blocks $J = 35$. Fig. 2 shows the SERs of the same receivers versus the number of data blocks for SNR = 20 dB.

In the second example, we consider the scenario with $N = 3$ antennas per transmitter. Both transmitters use the 3/4-rate

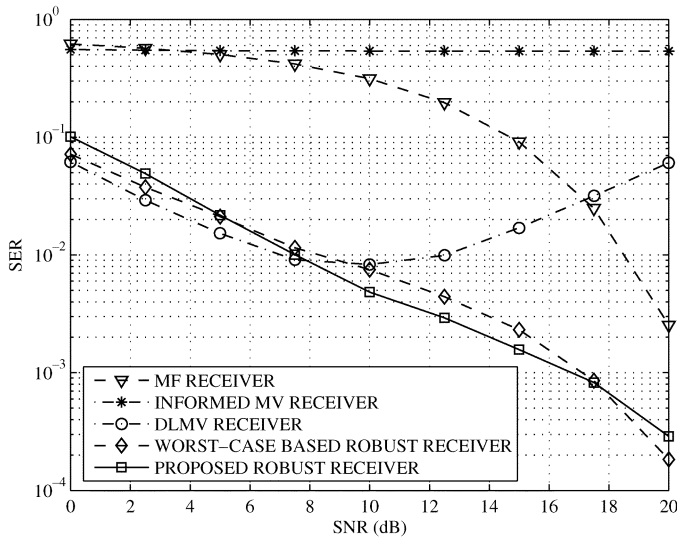


Fig. 3. SERs versus SNR; second example.

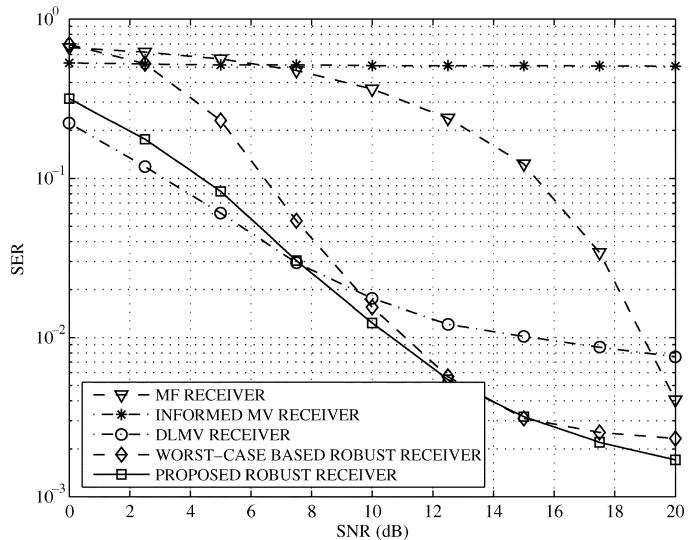


Fig. 5. SERs versus SNR; third example.

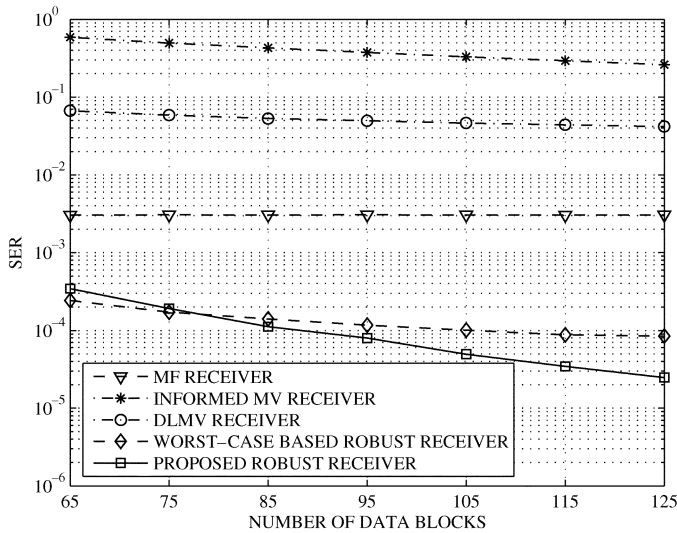


Fig. 4. SERs versus J ; second example.

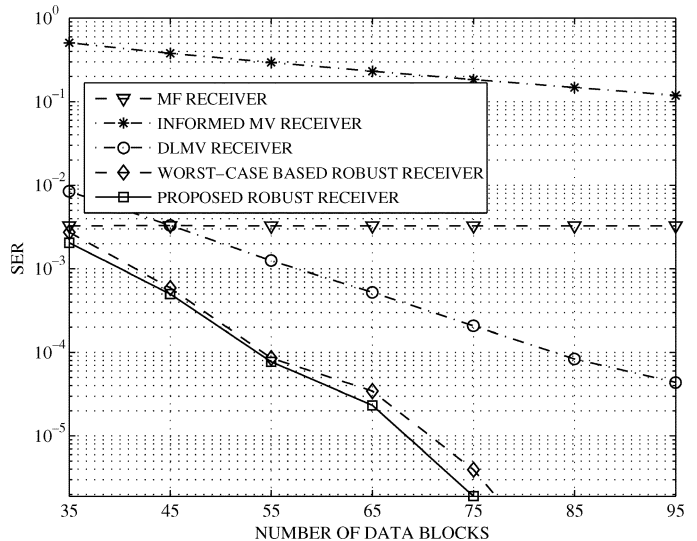


Fig. 6. SERs versus J ; third example.

($K = 3$; $T = 4$) OSTBC from [6]. The parameter $\varepsilon = 7\sigma_h$ is taken for the worst case-based receiver as suggested in [12]. Fig. 3 shows the receiver SERs versus the SNR for $J = 70$, while Fig. 4 displays the receiver SERs versus the number of data blocks for SNR = 20 dB.

In the third and fourth examples, we assume that the channel knowledge at the receiver is obtained by means of training-based channel estimation. In both these examples, the least-squares (LS) technique with the optimal training sequence [46] is used to estimate the CSI based on a single training data block, and $p = 0.98$ is chosen for the proposed receiver. To compute the value of σ_h^2 for the proposed receiver, (12) from [46] is used.

In the third example, we simulate a scenario with the same parameter setup as that of our first example. Fig. 5 displays the receiver SERs versus SNR for $J = 35$. Fig. 6 shows the receiver SERs versus J for SNR = 20 dB. Furthermore, Fig. 7 displays the performance of the proposed receiver versus the mismatch between the actual value of the standard deviation of the channel estimation error and the value of σ_h used in our design. In the latter figure, SNR = 15 dB and $J = 35$.

In the fourth example, we simulate a scenario with the same parameter setup as that of our second example. Fig. 8 shows the receiver SERs versus SNR for $J = 70$, while Fig. 9 displays the receiver SERs versus J for SNR = 20 dB.

From Figs. 1–9, it follows that in all simulation examples, the proposed robust receiver (76)–(78) consistently provides better performance as compared with the nonrobust receivers and the worst case robust receiver. Note that throughout all the figures, the “informed” MV receiver has quite a poor performance which can be explained by its insufficient robustness against finite sample effects in the case when the signal of interest “contaminates” the sample estimate of the interference-plus-noise covariance matrix. This signal “self-nulling” phenomenon is well known in adaptive beamforming and multiuser detection, see [14] and references therein. It can also be seen from Fig. 7 that the proposed receiver is relatively insensitive to a mismatch in σ_h .

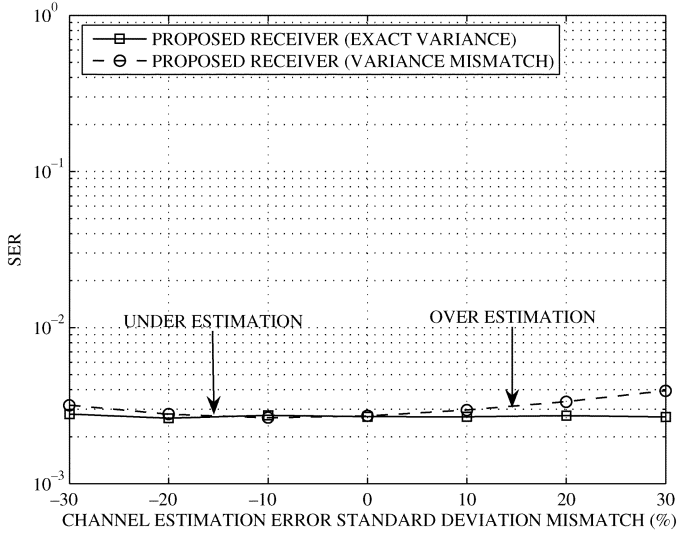


Fig. 7. SERs versus the channel estimation error standard deviation mismatch; third example.

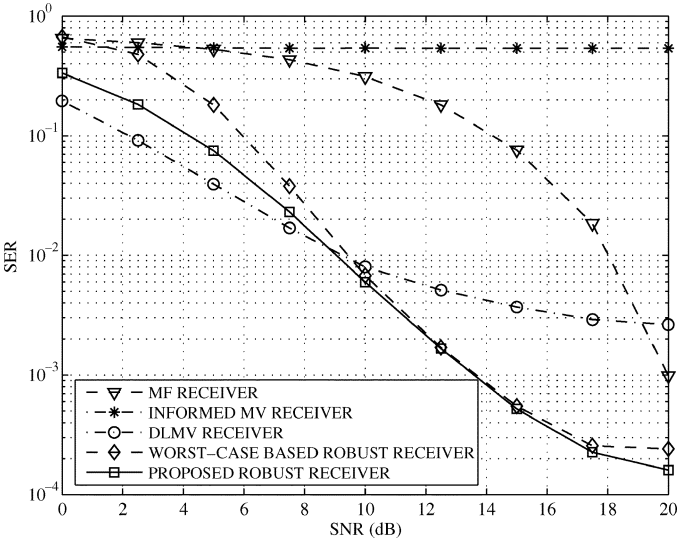


Fig. 8. SERs versus SNR; fourth example.

From Figs. 1, 3, 5, and 8, it can be seen that the DLMV receiver experiences substantial performance degradation at high SNR values. This effect has been earlier observed in [12] and can be explained by the fact that the DLMV receiver uses non-adaptive (fixed) DL factor.

In summary, Figs. 1–9 clearly demonstrate that the probability-based robust designs are, as expected, more flexible than the worst case designs.

V. CONCLUSION

The problem of robustness of multiuser MIMO receivers against imperfect CSI has been addressed, and a new approach that guarantees the robustness against CSI errors with a certain selected probability has been proposed. The new receiver techniques have been formulated as probabilistically constrained stochastic optimization problems. In the case of Gaussian CSI mismatch, these problems have been proved to be convex, and

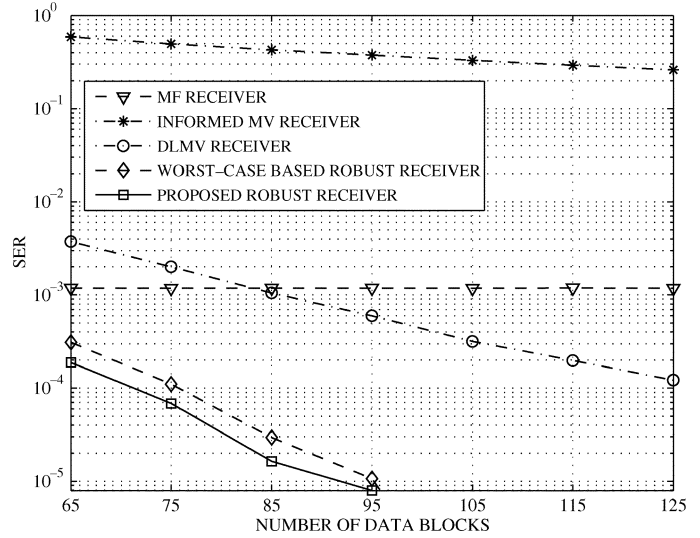


Fig. 9. SERs versus J ; fourth example.

have been converted to a more tractable deterministic SOCP form whose computational complexity is lower than that of the known worst case optimization-based robust receivers.

Simulation results have validated an improved performance of the proposed approach as compared to the existing linear multiuser MIMO receivers.

APPENDIX A
PROOF OF THEOREM 2

Let us first introduce the following lemma.

Lemma 1: Let vectors $\mathbf{v}_1, \dots, \mathbf{v}_n$ have a joint real Gaussian distribution with a covariance matrix \mathbf{B} , so that

$$\mathbb{E} \left\{ (\mathbf{v}_i - \mathbb{E}\{\mathbf{v}_i\}) (\mathbf{v}_l - \mathbb{E}\{\mathbf{v}_l\})^T \right\} = r_{il} \mathbf{B}, \quad \forall i, l, \quad i, l = 1, \dots, n \quad (81)$$

where r_{il} are some constants. Then, the set

$$\mathcal{K}(p) = \{ \mathbf{x} \mid \Pr \{ \mathbf{v}_1^T \mathbf{x} \geq \kappa_1 \wedge \dots \wedge \mathbf{v}_n^T \mathbf{x} \geq \kappa_n \} \geq p \} \quad (82)$$

is convex for $p \geq 0.5$. Here, \wedge denotes the set intersection operation, $0 < p \leq 1$, and κ_i are arbitrary real constants.

Proof: See [28, p. 312]. □

To prove Theorem 2, we first observe that the objective function (58) is a sum of two convex quadratic functions. Thus, it is convex.

It has been proven earlier in this paper [see (42)] that the constraint (59) is convex if $[\Delta_1]_{n,m} \sim \mathcal{CN}(0, \sigma_h^2)$ and $p \in (0.5, 1)$.

The constraints (60) share the same structure. Thus, it is enough to show that at least one of them is convex. Let us rewrite the l th constraint in (60) in the following equivalent form:

$$\Pr \left\{ \left(\mathbf{a}_l(\hat{\mathbf{H}}_1) + \mathbf{e}_l(\Delta_1) \right)^T \mathbf{w}_k + \delta_l \geq 0 \wedge \right. \\ \left. - \left(\mathbf{a}_l(\hat{\mathbf{H}}_1) + \mathbf{e}_l(\Delta_1) \right)^T \mathbf{w}_k + \delta_l \geq 0 \right\} \geq p. \quad (83)$$

The constraint (83) is called *joint chance constraint* in the stochastic programming literature [27], [28].

To be consistent with the notations used in Lemma 1, let us denote

$$\mathbf{v}_1 \triangleq \left[\left(\mathbf{a}_l(\hat{\mathbf{H}}_1) + \mathbf{e}_l(\Delta_1) \right)^T, 1 \right]^T \quad (84)$$

$$\mathbf{v}_2 \triangleq \left[- \left(\mathbf{a}_l(\hat{\mathbf{H}}_1) + \mathbf{e}_l(\Delta_1) \right)^T, 1 \right]^T \quad (85)$$

$$\mathbf{x} \triangleq [\mathbf{w}_k^T, \delta_l]^T \quad (86)$$

$$\kappa_1 \triangleq 0, \quad \kappa_2 \triangleq 0. \quad (87)$$

Then, (83) can be equivalently written as

$$\Pr \{ \mathbf{v}_1^T \mathbf{x} \geq \kappa_1 \wedge \mathbf{v}_2^T \mathbf{x} \geq \kappa_2 \} \geq p. \quad (88)$$

Using (34), we obtain that the vectors \mathbf{v}_1 and \mathbf{v}_2 have joint Gaussian distribution with

$$\begin{aligned} & \mathbb{E} \left\{ (\mathbf{v}_1 - \mathbb{E}\{\mathbf{v}_1\})(\mathbf{v}_1 - \mathbb{E}\{\mathbf{v}_1\})^T \right\} \\ &= \mathbb{E} \left\{ (\mathbf{v}_2 - \mathbb{E}\{\mathbf{v}_2\})(\mathbf{v}_2 - \mathbb{E}\{\mathbf{v}_2\})^T \right\} = \mathbf{B} \end{aligned} \quad (89)$$

$$\mathbb{E} \left\{ (\mathbf{v}_1 - \mathbb{E}\{\mathbf{v}_1\})(\mathbf{v}_2 - \mathbb{E}\{\mathbf{v}_2\})^T \right\} = -\mathbf{B} \quad (90)$$

where

$$\mathbf{B} = \begin{bmatrix} \mathbb{E} \left\{ \mathbf{e}_l(\Delta_1) \mathbf{e}_l^T(\Delta_1) \right\} & \mathbf{0}_{2MT} \\ \mathbf{0}_{2MT}^T & 0 \end{bmatrix} = \begin{bmatrix} \frac{\sigma_l^2}{2} \Psi_l \Psi_l^T & \mathbf{0}_{2MT} \\ \mathbf{0}_{2MT}^T & 0 \end{bmatrix}. \quad (91)$$

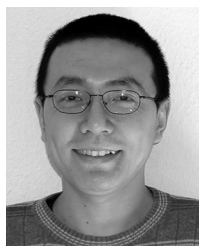
From (89) and (90), it is obvious that Lemma 1 can be directly applied. Thus, the constraints (60) are convex if $p \in [0.5, 1)$.

This completes the proof. \square

REFERENCES

- [1] A. J. Paulraj and C. B. Papadias, "Space-time processing for wireless communications," *IEEE Signal Process. Mag.*, vol. 14, pp. 49–83, Nov. 1997.
- [2] E. Telatar, "Capacity of multi-antenna Gaussian channels," *Eur. Trans. Telecommun.*, vol. 10, pp. 585–595, Nov. 1999.
- [3] G. J. Foschini and M. J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Pers. Commun.*, vol. 6, pp. 311–335, 1998.
- [4] S. M. Alamouti, "A simple transmitter diversity scheme for wireless communications," *IEEE J. Sel. Areas Commun.*, vol. 16, pp. 1451–1458, Oct. 1998.
- [5] V. Tarokh, N. Seshadri, and A. R. Calderbank, "Space-time codes for high data rate wireless communication: Performance criterion and code construction," *IEEE Trans. Inf. Theory*, vol. 44, pp. 744–765, Mar. 1998.
- [6] V. Tarokh, H. Jafarkhani, and A. R. Calderbank, "Space-time block codes from orthogonal designs," *IEEE Trans. Inf. Theory*, vol. 45, pp. 1456–1467, Jul. 1999.
- [7] H. Li, X. Lu, and G. B. Giannakis, "Capon multiuser receiver for CDMA systems with space-time coding," *IEEE Trans. Signal Process.*, vol. 50, pp. 1193–1204, May 2002.
- [8] D. Reynolds, X. Wang, and H. V. Poor, "Blind adaptive space-time multiuser detection with multiple transmitter and receiver antennas," *IEEE Trans. Signal Process.*, vol. 50, pp. 1261–1276, Jun. 2002.
- [9] A. F. Naguib, N. Seshadri, and A. R. Calderbank, "Applications of space-time block codes and interference suppression for high capacity and high data rate wireless systems," in *Proc. IEEE Asilomar Conf. Signals, Syst., Comp.*, Pacific Grove, CA, Nov. 1998, vol. 2, pp. 1803–1810.
- [10] W. M. Younis, A. H. Sayed, and N. Al-Dhahir, "Efficient adaptive receivers for joint equalization and interference cancellation in multi-user space-time block-coded systems," *IEEE Trans. Signal Process.*, vol. 51, pp. 2849–2862, Nov. 2003.
- [11] 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.
- [12] 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.
- [13] —, "Exploiting the structure of OSTBCs to improve the robustness of worst-case optimization based linear multi-user MIMO receivers," in *Proc. IEEE Inter. Conf. Acoust., Speech, Signal Process.*, Philadelphia, PA, Mar. 2005, vol. 4, pp. 781–784.
- [14] S. A. Vorobyov, A. B. Gershman, and Z.-Q. Luo, "Robust adaptive beamforming using worst-case performance optimization: A solution to the signal mismatch problem," *IEEE Trans. Signal Process.*, vol. 51, pp. 313–324, Feb. 2003.
- [15] A. H. Sayed and H. Chen, "A uniqueness result concerning a robust regularized least-squared solution," *Syst. Control Lett.*, vol. 46, pp. 361–369, Jan. 2002.
- [16] A. B. Gershman, "Robustness issues in adaptive beamforming and high-resolution direction finding," in *High-Resolution and Robust Signal Processing*, Y. Hua, A. B. Gershman, and Q. Cheng, Eds. New York: Marcel Dekker, 2003, pp. 63–110.
- [17] S. Shahbazpanahi, A. B. Gershman, Z.-Q. Luo, and K. M. Wong, "Robust adaptive beamforming for general-rank signal models," *IEEE Trans. Signal Process.*, vol. 51, pp. 2257–2269, Sept. 2003.
- [18] R. G. Lorenz and S. P. Boyd, "Robust minimum variance beamforming," *IEEE Trans. Signal Process.*, vol. 53, pp. 1684–1696, May 2005.
- [19] S. A. Vorobyov, A. B. Gershman, Z.-Q. Luo, and N. Ma, "Adaptive beamforming with joint robustness against mismatched signal steering vector and interference nonstationarity," *IEEE Signal Process. Lett.*, vol. 11, pp. 108–111, Feb. 2004.
- [20] D. P. Palomar, A. Pascual-Iserte, J. M. Cioffi, and M. A. Lagunas, "Convex optimization theory applied to joint transmitter-receiver design in MIMO channels," in *Space-Time Processing for MIMO Communications*, A. B. Gershman and N. D. Sidiropoulos, Eds. New York: Wiley, 2005, pp. 269–318.
- [21] Y. C. Eldar, A. Ben-Tal, and A. Nemirovski, "Robust mean-squared error estimation in the presence of model uncertainties," *IEEE Trans. Signal Process.*, vol. 53, pp. 168–181, Jan. 2005.
- [22] Y. Rong, S. A. Vorobyov, and A. B. Gershman, "Robust linear receiver design for multi-access space-time block coded MIMO systems using stochastic optimization," in *Proc. IEEE Workshop Stat. Signal Process.*, Bordeaux, France, Jul. 2005, pp. 65–70.
- [23] —, "A robust linear receiver for multi-access space-time block coded MIMO systems based on probability-constrained optimization," in *Proc. IEEE Veh. Technol. Conf.—Spring*, Milan, Italy, May 2004, vol. 1, pp. 118–122.
- [24] S. A. Vorobyov, Y. Rong, and A. B. Gershman, "Robust adaptive beamforming using probability-constrained optimization," in *Proc. IEEE Workshop Stat. Signal Process.*, Bordeaux, France, Jul. 2005, pp. 934–939.
- [25] R. Wang, H. Li, and T. Li, "Robust multiuser detection for multicarrier CDMA systems," *IEEE J. Sel. Areas Comm.*, vol. 24, pp. 673–683, Mar. 2006.
- [26] S. A. Vorobyov, Y. C. Eldar, A. Nemirovski, and A. B. Gershman, "Probability-constrained approach to estimation of random Gaussian parameters," in *Proc. IEEE Workshop Comp. Advances Multi-Sensor Adaptive Processing*, Puerto Vallarta, Mexico, Dec. 2005, pp. 101–104.
- [27] J. R. Birge and F. Louveaux, *Introduction to Stochastic Programming*. New York: Springer-Verlag, 1997.
- [28] A. Prékopa, *Stochastic Programming*. Dordrecht, The Netherlands: Kluwer, 1995.
- [29] D. P. Bertsekas, *Nonlinear Programming*. Belmont, MA: Athena Scientific, 1995.
- [30] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2003.
- [31] M. Lobo, L. Vandenberghe, S. Boyd, and H. Lebret, "Application of second-order cone programming," *Linear Algebra and Its Applications*, vol. 284, no. 1–3, pp. 193–228, Nov. 1998.
- [32] Z.-Q. Luo, "Applications of convex optimization in signal processing and digital communication," *Math. Program.*, ser. B, vol. 97, no. 1–2, pp. 177–207, Jul. 2003.
- [33] J. F. Sturm, "Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones," *Optim. Meth. Softw.*, vol. 11–12, pp. 625–653, Aug. 1999.
- [34] 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.
- [35] 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.

- [36] H. L. Van Trees, *Optimum Array Processing*. New York: Wiley, 2002.
- [37] S. Verdú, *Multisuser Detection*. Cambridge, U.K.: Cambridge Univ. Press, 1998.
- [38] M. Rupp, F. Tarköy, and J. L. Massey, "User-separating demodulation for code-division multiple-access systems," *IEEE J. Sel. Areas Comm.*, vol. 12, pp. 786–795, Jun. 1994.
- [39] A. H. Sayed, *Fundamentals of Adaptive Filtering*. New York: Wiley, 2003.
- [40] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links?," *IEEE Trans. Inf. Theory*, vol. 49, pp. 951–963, Apr. 2003.
- [41] T. L. Marzetta, "BLAST training: estimating channel characteristics for high-capacity space-time wireless," in *Proc. IEEE Annu. Allerton Conf. Commun., Control, Comput.*, Monticello, IL, Sept. 22–24, 1999, pp. 958–966.
- [42] J. W. Brewer, "Kronecker products and matrix calculus in system theory," *IEEE Trans. Circuits Syst.*, vol. 25, pp. 772–781, Sep. 1978.
- [43] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, 3rd ed. New York: McGraw-Hill, 1991.
- [44] J. Nocedal, A. Wächter, and R. A. Waltz, Adaptive barrier strategies for nonlinear interior methods, IBM T.J. Watson Research Center, Yorktown Heights, NY, Tech. Rep., RC 23563, Mar. 2005.
- [45] Y. Nesterov and A. Nemirovski, *Interior Point Polynomial Algorithms in Convex Programming*. Philadelphia, PA: SIAM, 1994.
- [46] M. Biguesh and A. B. Gershman, "Training-based MIMO channel estimation: A study of estimator tradeoffs and optimal training signals," *IEEE Trans. Signal Process.*, vol. 54, pp. 884–893, Mar. 2006.

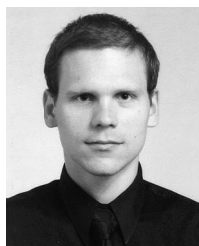


Yue Rong (S'03–M'06) was born in Jiangsu, China, in 1976. He received the B.E. degree from Shanghai Jiao Tong University, Shanghai, China, the M.Sc. degree from the University of Duisburg-Essen, Duisburg, Germany, and the Ph.D. degree (*summa cum laude*) from Darmstadt University of Technology, Darmstadt, Germany, in 1999, 2002, and 2005, respectively, all in electrical engineering.

From April 2001 to April 2002, he was a Student Research Assistant at the Fraunhofer Institute of Microelectronic Circuits and Systems, Duisburg,

Germany. From October 2001 to March 2002, he was with the Application-Specific Integrated Circuit Design Department, Nokia Ltd., Bochum, Germany. Currently, he is a Postdoctoral Researcher with the Department of Electrical Engineering, University of California, Riverside. His research interests include signal processing for communications, MIMO communication systems, multicarrier communications, applications of linear algebra and optimization methods, statistical and array signal processing, and parallel factor analysis.

Dr. Rong received the 2001–2002 Graduate Sponsoring Asia Scholarship of DAAD/ABB (Germany) and the 2004 Chinese Government Award for Outstanding Self-Financed Students Abroad (China).



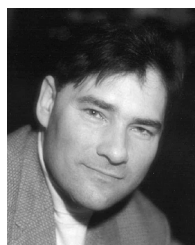
Sergiy A. Vorobyov (M'02–SM'05) was born in Ukraine in 1972. He received the M.S. and Ph.D. degrees in systems and control from the Kharkiv National University of Radioelectronics, Kharkiv, Ukraine, in 1994 and 1997, respectively.

From 1995 to 2000, he was with the Control Systems Research Laboratory, Kharkiv National University of Radioelectronics, where he became a Senior Research Scientist in 1999. From 1999 to 2001, he was with the Brain Science Institute, Institute of Physical and Chemical Research (RIKEN),

Japan, as a Research Scientist. From 2001 to 2003, he was with the Department of Electrical and Computer Engineering, McMaster University, Hamilton, ON, Canada, as a Postdoctoral Fellow. From 2003 to 2004 and from 2005 to 2006, he was a Senior Researcher at the Departments of Communication Systems, University of Duisburg-Essen, Duisburg, Germany, and the Darmstadt University of Technology, Darmstadt, Germany, respectively. From August 2006, he is with the Department of Electrical and Computer Engineering,

University of Alberta, Edmonton, AB, Canada, as an Assistant Professor. He also held short-time visiting appointments at the Institute of Applied Computer Science, Karlsruhe, Germany, in 1996; Gerhard-Mercator University, Duisburg, Germany, in 2002; and Technion—Israel Institute of Technology, Haifa, Israel, in 2005. His research interests include statistical array signal processing, robust adaptive beamforming, applications of linear algebra and optimization methods, estimation and detection theory, wireless and multicarrier communications, blind source separation, and control theory.

Dr. Vorobyov was a recipient of the 2004 IEEE Signal Processing Society Best Paper Award for his paper on robust minimum variance beamforming. He was also a recipient of the 1999 DAAD Fellowship (Germany); the 1996 and 1997 Young Scientist Research Grants from the George Soros Foundation; and the 1996–1998 Young Scientist Fellowship of the Ukrainian Cabinet of Ministers. He is an Associate Editor of the IEEE TRANSACTIONS ON SIGNAL PROCESSING.



Alex B. Gershman (M'97–SM'98–F'06) received the Diploma and Ph.D. degrees in radiophysics from the Nizhny Novgorod State University, Nizhny Novgorod, Russia, in 1984 and 1990, respectively.

From 1984 to 1999, he held several full-time and visiting research appointments in Russia, Switzerland, and Germany. In 1999, he joined the Department of Electrical and Computer Engineering, McMaster University, Hamilton, ON, Canada, where he became a Full Professor in 2002.

From 2003 to 2005, he was a Visiting Professor at

the Department of Communication Systems, University of Duisburg-Essen, Duisburg, Germany. Since April 2005, he has been with Darmstadt University of Technology, Darmstadt, Germany, as a full Professor and Head of the Communication Systems Group. He has coedited *High-Resolution and Robust Signal Processing* (Berlin, Germany: Marcel Dekker, 2003), with Y. Hua and Q. Cheng, and *Space-Time Processing for MIMO Communications* (New York: Wiley, 2005), with N. D. Sidiropoulos. His research interests are in the area of signal processing and communications with primary emphasis on statistical signal and array processing, adaptive beamforming, spatial diversity in MIMO communications, multiuser and multicarrier communications, and parameter estimation and detection. He has (co)authored several book chapters and more than 90 journal and 150 conference papers on these subjects.

Dr. Gershman was a corecipient of the 2004 IEEE Signal Processing Society Best Paper Award for his paper on robust minimum variance beamforming. He coauthored the paper that received the 2005 IEEE Signal Processing Society Young Author Best Paper Award. He was a recipient of the 2002 Young Explorers Prize from the Canadian Institute for Advanced Research; the 2001 Wolfgang Paul Award from the Alexander von Humboldt Foundation, Germany; the 2000 Premier's Research Excellence Award, Ontario, Canada; the 1995–1996 Alexander von Humboldt Fellowship (Germany); the 1994 Swiss Academy of Engineering Science and Branco Weiss Fellowship (Switzerland); the 1994 Outstanding Young Scientist Presidential Fellowship (Russia); and the 1993 International Union of Radio Science (URSI) Young Scientist Award. He is currently serving as Editor-in-Chief of the IEEE SIGNAL PROCESSING LETTERS. He has been on the Editorial Boards of the *EURASIP Journal on Wireless Communications and Networking* and *EURASIP Signal Processing* from 2003 and 2005, respectively. He was Associate Editor of the IEEE TRANSACTIONS ON SIGNAL PROCESSING (1999–2005) and Guest Editor of the *EURASIP Journal of Applied Signal Processing* (2004). He is Vice-Chair of the Sensor Array and Multichannel (SAM) Technical Committee (TC) and a Member of the Signal Processing Theory and Methods (SPTM) TC of the IEEE Signal Processing Society. He was a Technical Co-Chair of the IEEE International Symposium on Signal Processing and Information Technology (ISSPIT), Darmstadt, Germany, December 2003; a Co-Chair of the International ITG/IEEE Workshop on Smart Antennas, Duisburg, Germany, April 2005; and a General Co-Chair of the First IEEE Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP), Puerto-Vallarta, Mexico, December 2005. He was a Technical Co-Chair of the Fourth IEEE Sensor Array and Multichannel Signal Processing Workshop, Waltham, MA, June 2006, and is a Tutorial Chair of EUSIPCO 2006, Florence, Italy, September 2006.