Robust MMSE Transceiver Design for Nonregenerative Multicasting MIMO Relay Systems

Lenin Gopal, *Member, IEEE*, Yue Rong, *Senior Member, IEEE*, and Zhuquan Zang, *Member, IEEE*

***Abstract*—In this paper, we investigate the transceiver design for nonregenerative multicasting multiple-input multiple-output (MIMO) relay systems, where one transmitter broadcasts common message to multiple receivers with the aid of a relay node. The transmitter, relay, and receivers are all equipped with multiple antennas. We assume that the true (unknown) channel matrices have Gaussian distribution, with the estimated channels as the mean value, and the channel estimation errors follow the well-known Kronecker model. We first develop an iterative robust algorithm to jointly design the transmitter, relay, and receiver matrices to minimize the maximal mean-squared error (MSE) of the signal waveform estimation among all receivers. Then, we derive the optimal structure of the relay precoding matrix and show that the MSE at each receiver can be decomposed into the sum of the MSEs of the first-hop and second-hop channels. Based on this MSE decomposition, we develop a simplified transceiver design algorithm with a low computational complexity. Numerical simulations demonstrate the improved robustness of the proposed transceiver design algorithms against the mismatch between the true and estimated channels. Interestingly, compared with the iterative algorithm, the simplified transceiver design has only negligible performance loss with significantly reduced computational complexity.**

***Index Terms*—Minimum mean-squared error (MMSE), multicasting, nonregenerative MIMO relay, robustness.**

I. Introduction

In MANY practical communication systems, one source node transmits common information to multiple receivers simultaneously. These systems are referred to as multicast broadcasting or multicasting systems. Recently, multicasting systems have attracted much research interest, due to the increasing demand for mobile applications such as location based video broadcasting and streaming media.

The wireless channel has the multicast broadcasting nature, making it suitable for multicasting applications. However, the wireless system performance may be degraded due to the channel fading and shadowing effects. By deploying multi-antenna and beamforming techniques at the transmitter and receivers, the channel shadowing effect can be mitigated [2]. Next generation wireless standards such as WiMAX 802.16m and 3GPP LTE-Advanced have already included technologies which enable better multicasting solutions based on multi-antenna and beamforming techniques [3].

Due to the nonconvex nature of the problem, designing the optimal transmit beamforming vector for multicasting is difficult in general. Capacity limits of multi-antenna multicasting channel have been studied in [4], and the channel spatial correlation effect on the channel capacity has been investigated in [5]. In [6], transmit beamforming vectors for physical layer multicasting have been designed with the assumption that the channel state information (CSI) is available at the transmitter. In the multicasting systems [3]–[8], single-antenna has been assumed at the receivers. Recently multicasting systems with multi-antenna receivers have been investigated in [9]–[11].

In the case of long distance between the transmitter and receivers, relay node is necessary to efficiently mitigate the pathloss of wireless channel. In [12], a cooperative protocol for multicasting systems with multiple transmit antennas is proposed with the assumption that the users are equipped with single antenna. A two-hop multiple-input multiple-output (MIMO) relay multicasting system has been proposed in [13] where one transmitter multicasts common message to multiple receivers with the aid of a relay node, and the transmitter, relay, and receivers are all equipped with multiple antennas. It is also assumed in [13] that the true CSI of all channels is available at the relay node. However, in practical communication systems, the exact CSI is not available, and therefore, has to be estimated. There is always mismatch between the true and estimated CSI. Hence, the performance of the algorithm in [13] will degrade due to such CSI mismatch.

In this paper, we propose a transceiver design algorithm for multicasting MIMO relay systems which is robust against the CSI mismatch. Similar to [13], the transmitter, relay, and receivers in the system are all equipped with multiple antennas. However, different to [13], the true channel matrices have Gaussian distribution, with the estimated channels as the mean value, and the channel estimation errors follow the well-known Kronecker model [14], [15]. We first propose an iterative robust algorithm to jointly design the transmitter, relay, and receiver matrices to minimize the maximal mean-squared error (MSE) of the signal waveform estimation among all receivers.

Note that recently some transceiver optimization problems in MIMO systems have been unified into a framework of matrix-monotonic optimization in [16]. A general robust linear transceiver design for multi-hop amplify-and-forward MIMO relay systems has been proposed in [17]. Quadratic matrix programming has been applied in [18] for transceiver optimization in MIMO relay systems. However, the problem of robust transceiver design for multicasting MIMO relay systems is not

Manuscript received March 16, 2017; accepted May 5, 2017. Date of publication May 11, 2017; date of current version October 13, 2017. This work was supported by the Australian Research Council's Discovery Projects funding scheme (DP140102131). The review of this paper was coordinated by Dr. T. Taniguchi. This paper was presented in part at the 22nd International Conference on Telecommunications, Sydney, NSW, Australia, Apr. 27–29, 2015. *(Corresponding author: Yue Rong.)*

L. Gopal and Z. Zhang are with the Department of Electrical and Computer Engineering, Curtin University, Miri 98009, Malaysia (e-mail: lenin@curtin.edu.my; zqzang@curtin.edu.my).

Y. Rong is with the Department of Electrical and Computer Engineering, Curtin University, Bentley, WA 6102, Australia (e-mail: y.rong@curtin.edu.au).

Digital Object Identifier 10.1109/TVT.2017.2703641

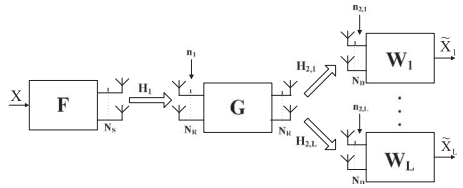


Fig. 1. Block diagram of a two-hop nonregenerative multicasting MIMO relay system.

covered in [16]–[18]. We would like to mention that robust transceiver design for single-user MIMO relay systems [14], [15], [19]–[21], and multiuser MIMO relay systems [22] can not be simply extended to multicasting MIMO relay systems.

Due to the high computational complexity of the proposed iterative algorithm, we develop a simplified robust transceiver design scheme in which we show that the MSE at each receiver can be decomposed into the sum of the MSEs of the first-hop and second-hop channels. This extends the result of MSE decomposition [13] and [23] from MIMO relay systems with perfect CSI to practical MIMO relay systems with imperfect CSI. Based on this MSE decomposition, we develop a transceiver design algorithm with low computational complexity. We show that under some mild conditions, the transmitter and relay precoding matrices can be optimized separately. In particular, the transmitter precoding matrix optimization problem has a closed-form solution, while the relay precoding matrix can be optimized through solving a convex semidefinite programming (SDP) problem [24].

Numerical simulations demonstrate the improved robustness of both proposed iterative and simplified algorithms against the CSI mismatch. Interestingly, compared with the iterative algorithm, the simplified transceiver design has only negligible performance loss with significantly reduced computational complexity.

The rest of the paper is organized as follows. The system model is presented in Section II. In Section III, the proposed robust transceiver design algorithms are developed. Numerical examples are shown in Section IV to verify the performance of the proposed algorithms, and conclusions are drawn in Section V.

II. System Model

We consider a two-hop nonregenerative MIMO relay multicasting system with L receivers as shown in Fig. 1, where the transmitter and relay have N_S and N_R antennas, respectively. For simplicity, we assume that each receiver has N_D antennas. It is assumed that due to severe pathloss, there are no direct links between the transmitter and receivers. The data transmission takes place over two time slots. The received signal at the relay during the first time slot is given by

$$\mathbf{y}_r = \mathbf{H}_1\mathbf{F}\mathbf{x} + \mathbf{n}_1 \tag{1}$$

where $\mathbf{x} \in \mathbb{C}^{N_B \times 1}$ is the source signal vector satisfying $E\{\mathbf{x}\mathbf{x}^H\} = \mathbf{I}_{N_B}$, N_B is chosen to satisfy $N_B \leq \min(N_S, N_R, N_D)$, $\mathbf{H}_1 \in \mathbb{C}^{N_R \times N_S}$ is the MIMO channel matrix between the transmitter and the relay node, $\mathbf{F} \in \mathbb{C}^{N_S \times N_B}$ is the transmitter precoding matrix, and $\mathbf{n}_1 \in \mathbb{C}^{N_R \times 1}$ is the additive noise vector at the relay node. Here $(\cdot)^H$ denotes matrix Hermitian transpose, $E\{\cdot\}$ stands for statistical expectation, and $\mathbf{I}_n$ represents the $n \times n$ identity matrix.

At the second time slot, the relay node linearly precodes $\mathbf{y}_r$ with the relay precoding matrix $\mathbf{G} \in \mathbb{C}^{N_R \times N_R}$, and broadcasts the linearly precoded signal vector

$$\mathbf{x}_r = \mathbf{G}\mathbf{y}_r \tag{2}$$

to all receivers. The received signal vector at the ith receiver can be written as

$$\mathbf{y}_{d,i} = \mathbf{H}_{2,i}\mathbf{G}\mathbf{H}_1\mathbf{F}\mathbf{x} + \mathbf{H}_{2,i}\mathbf{G}\mathbf{n}_1 + \mathbf{n}_{2,i}, \quad i = 1, \cdots, L \tag{3}$$

where $\mathbf{H}_{2,i} \in \mathbb{C}^{N_D \times N_R}$ is the MIMO channel matrix between the relay node and the ith receiver, $\mathbf{n}_{2,i} \in \mathbb{C}^{N_D \times 1}$ is the additive noise vector at the ith receiver. We assume that all noises are independent and identically distributed (i.i.d.) with zero mean and unit variance. In practice, channel $\mathbf{H}_1$ can be estimated at the relay node through channel training using pilot signals sent by the transmitter. Channel $\mathbf{H}_{2,i}$ is estimated at the ith receiver using a training sequence sent from the relay node, and the estimated $\mathbf{H}_{2,i}$ is then sent from the ith receiver to the relay node. The optimal transmitter precoding matrix $\mathbf{F}$ and relay precoding matrix $\mathbf{G}$ are calculated at the relay node. Then the relay node forwards $\mathbf{F}$ to the transmitter and broadcasts $\mathbf{F}$, $\mathbf{G}$, and $\mathbf{H}_1$ to all receivers.

In general, the instantaneous CSI is required for the optimal design of the precoding matrices $\mathbf{F}$ and $\mathbf{G}$. However, in practice, the exact CSI is not available due to channel estimation errors. In fact, the exact CSI of $\mathbf{H}_1$ and $\mathbf{H}_{2,i}$ can be modeled as [15]

$$\mathbf{H}_1 = \widehat{\mathbf{H}}_1 + \mathbf{\Delta}_1 \tag{4}$$

$$\mathbf{H}_{2,i} = \widehat{\mathbf{H}}_{2,i} + \mathbf{\Delta}_{2,i}, \quad i = 1, \cdots, L \tag{5}$$

where $\widehat{\mathbf{H}}_1$ and $\widehat{\mathbf{H}}_{2,i}$ are the estimated transmitter-relay and relay-ith receiver channel matrices, respectively, $\mathbf{\Delta}_1$ and $\mathbf{\Delta}_{2,i}$ are the corresponding channel estimation error matrices. We assume that $\mathbf{\Delta}_1$ and $\mathbf{\Delta}_{2,i}$ satisfy the Gaussian-Kronecker model as [15]

$$\mathbf{\Delta}_1 \sim \mathcal{CN}\left(\mathbf{0}, \mathbf{\Sigma}_1 \otimes \mathbf{\Psi}_1^T\right) \tag{6}$$

$$\mathbf{\Delta}_{2,i} \sim \mathcal{CN}\left(\mathbf{0}, \mathbf{\Sigma}_{2,i} \otimes \mathbf{\Psi}_{2,i}^T\right), \quad i = 1, \cdots, L \tag{7}$$

where $\otimes$ denotes the matrix Kronecker product, $(\cdot)^T$ stands for the matrix transpose, $\mathbf{\Sigma}_1$ and $\mathbf{\Psi}_1^T$ are the row and column covariance matrices of $\mathbf{\Delta}_1$, respectively, and similarly, $\mathbf{\Sigma}_{2,i}$ and $\mathbf{\Psi}_{2,i}^T$ are the row and column covariance matrices of $\mathbf{\Delta}_{2,i}$, respectively. We would like to note that the Gaussian-Kronecker model is well-accepted to model the channel estimation error matrices. As an example, it has been shown in [15] that the error matrices using channel estimation algorithms in [25] and [26] have the Gaussian-Kronecker structure.

From (4)–(7), the channel matrices $\mathbf{H}_1$ and $\mathbf{H}_{2,i}$ can be modeled as [15]

$$\mathbf{H}_1 = \widehat{\mathbf{H}}_1 + \mathbf{\Sigma}_1^{\frac{1}{2}}\mathbf{H}_{\omega_1}\mathbf{\Psi}_1^{\frac{1}{2}} \tag{8}$$

$$\mathbf{H}_{2,i} = \widehat{\mathbf{H}}_{2,i} + \mathbf{\Sigma}_{2,i}^{\frac{1}{2}}\mathbf{H}_{\omega_{2,i}}\mathbf{\Psi}_{2,i}^{\frac{1}{2}}, \quad i = 1, \cdots, L \tag{9}$$

where $\mathbf{H}_{\omega_1}$ and $\mathbf{H}_{\omega_{2,i}}$ are complex Gaussian random matrices whose entries are i.i.d. with zero mean and unit variance.

Lemma 1: [15], [27] For $\mathbf{H} \sim \mathcal{CN}(\widehat{\mathbf{H}}, \mathbf{\Sigma} \otimes \mathbf{\Psi}^T)$ and any constant matrix $\mathbf{A}$, there is

$$E_H\left\{\mathbf{H}\mathbf{A}\mathbf{H}^H\right\} = \widehat{\mathbf{H}}\mathbf{A}\widehat{\mathbf{H}}^H + tr\left\{\mathbf{A}\mathbf{\Psi}\right\}\mathbf{\Sigma} \tag{10}$$

where $tr\{\cdot\}$ denotes the matrix trace.

At the ith receiver, a linear receiver with the weight matrix $\mathbf{W}_i$ is applied to retrieve the source signal vector $\mathbf{x}$. Hence, the estimated signal at the ith receiver can be expressed as

$$\tilde{\mathbf{x}}_i = \mathbf{W}_i\,\mathbf{y}_{d,i}, \quad i = 1, \cdots, L. \tag{11}$$

Using (3) and (11), the MSE of the signal waveform estimation at the ith receiver is given by

$$\begin{aligned} & M_i(\mathbf{W}_i, \mathbf{G}, \mathbf{F}) \\ = \; & E\big\{tr\big\{(\mathbf{W}_i\mathbf{y}_{d,i} - \mathbf{x})(\mathbf{W}_i\mathbf{y}_{d,i} - \mathbf{x})^H\big\}\big\} \\ = \; & tr\big\{\big(\mathbf{W}_i\mathbf{H}_{2,i}\mathbf{G}\mathbf{H}_1\mathbf{F} - \mathbf{I}_{N_B}\big) \\ & \times \big(\mathbf{W}_i\mathbf{H}_{2,i}\mathbf{G}\mathbf{H}_1\mathbf{F} - \mathbf{I}_{N_B}\big)^H + \mathbf{W}_i\mathbf{R}_{n,i}\mathbf{W}_i^H\big\} \end{aligned} \tag{12}$$

where $\mathbf{R}_{n,i}$ is the equivalent noise covariance matrix given by

$$\mathbf{R}_{n,i} = \mathbf{H}_{2,i}\mathbf{G}\mathbf{G}^H\mathbf{H}_{2,i}^H + \mathbf{I}_{N_D}.$$

From (2), the transmission power consumed by the relay node can be written as

$$\begin{aligned} P(\mathbf{G}) = \; & E\{tr\{\mathbf{x}_r\mathbf{x}_r^H\}\} \\ = \; & tr\{\mathbf{G}(\mathbf{H}_1\mathbf{F}\mathbf{F}^H\mathbf{H}_1^H + \mathbf{I}_{N_R})\mathbf{G}^H\}. \end{aligned} \tag{13}$$

III. Proposed Robust Transceiver Design Algorithms

It can be seen from (12) that if the exact $\mathbf{H}_1$ and $\mathbf{H}_{2,i}$ are unavailable at the receivers, it is impossible to design $\mathbf{W}_i$ that optimizes $M_i(\mathbf{W}_i, \mathbf{G}, \mathbf{F})$ in (12). If we design $\mathbf{W}_i$, $\mathbf{F}$, and $\mathbf{G}$ based only on $\widehat{\mathbf{H}}_1$ and $\widehat{\mathbf{H}}_{2,i}$, there can be a great performance degradation due to the mismatch between $\mathbf{H}_1$, $\mathbf{H}_{2,i}$ and $\widehat{\mathbf{H}}_1$, $\widehat{\mathbf{H}}_{2,i}$. Instead of optimizing (12), we design $\mathbf{W}_i$, $\mathbf{F}$, and $\mathbf{G}$ to minimize

$$J_i(\mathbf{W}_i, \mathbf{G}, \mathbf{F}) \triangleq E_{H_1, H_{2,i}}\{M_i(\mathbf{W}_i, \mathbf{G}, \mathbf{F})\}$$

where the statistical expectation is carried out with respect to $\mathbf{H}_1$ and $\mathbf{H}_{2,i}$, with the distribution given in (8) and (9). By applying Lemma 1 in (10) to (12), we obtain

$$\begin{aligned} & J_i(\mathbf{W}_i, \mathbf{G}, \mathbf{F}) \\ = \; & tr\big\{\mathbf{I}_{N_B} + \mathbf{W}_i\big(\widehat{\mathbf{H}}_{2,i}\mathbf{G}\boldsymbol{\Xi}\mathbf{G}^H\widehat{\mathbf{H}}_{2,i}^H + tr\big\{\mathbf{G}\boldsymbol{\Xi}\mathbf{G}^H\boldsymbol{\Psi}_{2,i}\big\}\boldsymbol{\Sigma}_{2,i} \\ & + \mathbf{I}_{N_D}\big)\mathbf{W}_i^H - \mathbf{W}_i\widehat{\mathbf{H}}_{2,i}\mathbf{G}\widehat{\mathbf{H}}_1\mathbf{F} - \mathbf{F}^H\widehat{\mathbf{H}}_1^H\mathbf{G}^H\widehat{\mathbf{H}}_{2,i}^H\mathbf{W}_i^H\big\} \end{aligned} \tag{14}$$

where

$$\boldsymbol{\Xi} = \widehat{\mathbf{H}}_1\mathbf{F}\mathbf{F}^H\widehat{\mathbf{H}}_1^H + tr\{\mathbf{F}\mathbf{F}^H\boldsymbol{\Psi}_1\}\boldsymbol{\Sigma}_1 + \mathbf{I}_{N_R}. \tag{15}$$

The detailed steps leading to (14) are presented in Appendix A.

Since the true $\mathbf{H}_1$ is unknown, we consider the averaged transmission power at the relay node, which can be calculated from (13) as

$$\begin{aligned} E_{H_1}\{P(\mathbf{G})\} = \; & E_{H_1}\{tr\{\mathbf{G}(\mathbf{H}_1\mathbf{F}\mathbf{F}^H\mathbf{H}_1^H + \mathbf{I}_{N_R})\mathbf{G}^H\}\} \\ = \; & tr\{\mathbf{G}\boldsymbol{\Xi}\mathbf{G}^H\}. \end{aligned} \tag{16}$$

In the proposed transceiver design, we aim to minimize the maximum of (14) among all receivers subjecting to the transmitter and relay power constraints, which can be written as the following optimization problem

$$\min_{\mathbf{F}, \mathbf{G}, \{\mathbf{W}_i\}} \quad \max_i \; J_i(\mathbf{W}_i, \mathbf{G}, \mathbf{F}) \tag{17}$$

$$\text{s.t.} \quad tr\big\{\mathbf{G}\boldsymbol{\Xi}\mathbf{G}^H\big\} \le P_r \tag{18}$$

$$tr\{\mathbf{F}\mathbf{F}^H\} \le P_s \tag{19}$$

where $\{\mathbf{W}_i\} = \{\mathbf{W}_i, i = 1, \cdots, L\}$, (18) and (19) are the transmission power constraints at the relay node and the transmitter, respectively, and $P_r > 0$, $P_s > 0$ are the corresponding power budgets.

A. Iterative Transceiver Design Algorithm

The min-max problem (17)–(19) is very hard to solve due to the complicated objective function (17). In the following, we develop an iterative algorithm[1] to solve the problem (17)–(19). Firstly, for any given $\mathbf{F}$ and $\mathbf{G}$, the optimal $\mathbf{W}_i$ that minimizes the MSE function (14) is the well known MMSE receiver (Wiener filter) which is given by [28]

$$\begin{aligned} \mathbf{W}_i = \; & \mathbf{F}^H\widehat{\mathbf{H}}_1^H\mathbf{G}^H\widehat{\mathbf{H}}_{2,i}^H\big(\widehat{\mathbf{H}}_{2,i}\mathbf{G}\boldsymbol{\Xi}\mathbf{G}^H\widehat{\mathbf{H}}_{2,i}^H \\ & + tr\{\mathbf{G}\boldsymbol{\Xi}\mathbf{G}^H\boldsymbol{\Psi}_{2,i}\}\boldsymbol{\Sigma}_{2,i} + \mathbf{I}_{N_D}\big)^{-1}, i = 1, \cdots, L \end{aligned} \tag{20}$$

where $(\cdot)^{-1}$ denotes the matrix inverse.

Secondly, with given $\mathbf{F}$ and $\{\mathbf{W}_i\}$, the problem of optimizing $\mathbf{G}$ can be written as

$$\min_{\mathbf{G}, t_1} \quad t_1 \tag{21}$$

$$\text{s.t.} \quad J_i(\mathbf{G}) \le t_1, \quad i = 1, \cdots, L \tag{22}$$

$$tr\{\mathbf{G}\boldsymbol{\Xi}\mathbf{G}^H\} \le P_r \tag{23}$$

where $J_i(\mathbf{G})$ is (14) written as a function of $\mathbf{G}$ with fixed $\mathbf{F}$ and $\{\mathbf{W}_i\}$ as

$$\begin{aligned} J_i(\mathbf{G}) = \; & tr\{\mathbf{W}_i\mathbf{W}_i^H\} + tr\{\mathbf{W}_i\widehat{\mathbf{H}}_{2,i}\mathbf{G}\boldsymbol{\Xi}\mathbf{G}^H\widehat{\mathbf{H}}_{2,i}^H\mathbf{W}_i^H\} \\ & + tr\{\mathbf{G}\boldsymbol{\Xi}\mathbf{G}^H\boldsymbol{\Psi}_{2,i}\}tr\{\mathbf{W}_i\boldsymbol{\Sigma}_{2,i}\mathbf{W}_i^H\} + N_B \\ & - tr\{\mathbf{W}_i\widehat{\mathbf{H}}_{2,i}\mathbf{G}\widehat{\mathbf{H}}_1\mathbf{F} + \mathbf{F}^H\widehat{\mathbf{H}}_1^H\mathbf{G}^H\widehat{\mathbf{H}}_{2,i}^H\mathbf{W}_i^H\}. \end{aligned} \tag{24}$$

Let us introduce for $i = 1, \cdots, L$

$$\mathbf{D}_i^{\frac{1}{2}}\mathbf{D}_i^{\frac{H}{2}} = \widehat{\mathbf{H}}_{2,i}^H\mathbf{W}_i^H\mathbf{W}_i\widehat{\mathbf{H}}_{2,i} + tr\{\mathbf{W}_i\boldsymbol{\Sigma}_{2,i}\mathbf{W}_i^H\}\boldsymbol{\Psi}_{2,i}.$$

Then (24) can be rewritten as

$$\begin{aligned} J_i(\mathbf{G}) = \; & tr\{\mathbf{W}_i\mathbf{W}_i^H\} + N_B + tr\{\mathbf{D}_i^{\frac{H}{2}}\mathbf{G}\boldsymbol{\Xi}\mathbf{G}^H\mathbf{D}_i^{\frac{1}{2}} \\ & - \mathbf{W}_i\widehat{\mathbf{H}}_{2,i}\mathbf{G}\widehat{\mathbf{H}}_1\mathbf{F} - \mathbf{F}^H\widehat{\mathbf{H}}_1^H\mathbf{G}^H\widehat{\mathbf{H}}_{2,i}^H\mathbf{W}_i^H\}. \end{aligned} \tag{25}$$

Let us introduce $\mathbf{G}\boldsymbol{\Xi}\mathbf{G}^H \preceq \boldsymbol{\Phi}$, where $\mathbf{A} \preceq \mathbf{B}$ means that $\mathbf{B} - \mathbf{A} \succeq 0$ is positive semidefinite (PSD). Using (25), the problem (21)–(23) can be equivalently rewritten as the following

[1]Note that the iterative method provides a system MSE and bit-error-rate (BER) lower-bound which is useful to judge the performance of the simplified algorithm developed in the next subsection.

SDP problem

$$\min_{\mathbf{G},\mathbf{\Phi},t_1} \quad t_1 \tag{26}$$

$$\text{s.t.} \quad J_i(\mathbf{G},\mathbf{\Phi}) \le t_1, \;\; i=1,\cdots,L \tag{27}$$

$$tr\{\mathbf{\Phi}\} \le P_r \tag{28}$$

$$\begin{pmatrix} \mathbf{\Phi} & \mathbf{G} \\ \mathbf{G}^H & \mathbf{\Xi}^{-1} \end{pmatrix} \succeq 0 \tag{29}$$

where

$$J_i(\mathbf{G},\mathbf{\Phi}) = tr\Big\{\mathbf{W}_i\mathbf{W}_i^H\Big\} + N_B + tr\Big\{\mathbf{D}_i^{\frac{H}{2}}\mathbf{\Phi}\mathbf{D}_i^{\frac{1}{2}} - \mathbf{W}_i\widehat{\mathbf{H}}_{2,i}\mathbf{G}\widehat{\mathbf{H}}_1\mathbf{F} - \mathbf{F}^H\widehat{\mathbf{H}}_1^H\mathbf{G}^H\widehat{\mathbf{H}}_{2,i}^H\mathbf{W}_i^H\Big\}.$$

Thirdly, with given $\mathbf{G}$ and $\{\mathbf{W}_i\}$, the problem of optimizing $\mathbf{F}$ can be written as

$$\min_{\mathbf{F},t_2} \quad t_2 \tag{30}$$

$$\text{s.t.} \quad J_i(\mathbf{F}) \le t_2, \quad i=1,\cdots,L \tag{31}$$

$$tr\{\mathbf{G}\mathbf{\Xi}\mathbf{G}^H\} \le P_r \tag{32}$$

$$tr\{\mathbf{F}\mathbf{F}^H\} \le P_s \tag{33}$$

where $J_i(\mathbf{F})$ is (14) written as a function of $\mathbf{F}$ with fixed $\mathbf{G}$ and $\{\mathbf{W}_i\}$ as

$$J_i(\mathbf{F}) = tr\Big\{\mathbf{\Theta}_i^{\frac{H}{2}}\mathbf{F}\mathbf{F}^H\mathbf{\Theta}_i^{\frac{1}{2}}\Big\} + S_i - tr\Big\{\mathbf{W}_i\widehat{\mathbf{H}}_{2,i}\mathbf{G}\widehat{\mathbf{H}}_1\mathbf{F} + \mathbf{F}^H\widehat{\mathbf{H}}_1^H\mathbf{G}^H\widehat{\mathbf{H}}_{2,i}^H\mathbf{W}_i^H\Big\}. \tag{34}$$

Here $\mathbf{\Theta}_i$ and S_i are given in (86) and (87) in Appendix B, respectively. The detailed steps leading to (34) are presented in Appendix B. The left-hand side of the constraint (32) can be rewritten as

$$\begin{aligned} tr\{\mathbf{G}\mathbf{\Xi}\mathbf{G}^H\} &= tr\{\widehat{\mathbf{H}}_1^H\mathbf{G}^H\mathbf{G}\widehat{\mathbf{H}}_1\mathbf{F}\mathbf{F}^H\} \\ &\quad + tr\{\mathbf{G}\mathbf{\Sigma}_1\mathbf{G}^H\}tr\{\mathbf{F}\mathbf{F}^H\mathbf{\Psi}_1\} + tr\{\mathbf{G}\mathbf{G}^H\} \\ &= tr\{\mathbf{O}^{\frac{H}{2}}\mathbf{F}\mathbf{F}^H\mathbf{O}^{\frac{1}{2}}\} + tr\{\mathbf{G}\mathbf{G}^H\} \end{aligned} \tag{35}$$

where $\widehat{\mathbf{H}}_1^H\mathbf{G}^H\mathbf{G}\widehat{\mathbf{H}}_1 + tr\{\mathbf{G}\mathbf{\Sigma}_1\mathbf{G}^H\}\mathbf{\Psi}_1 = \mathbf{O}^{\frac{1}{2}}\mathbf{O}^{\frac{H}{2}}$.

Let us introduce $\mathbf{O}^{\frac{H}{2}}\mathbf{F}\mathbf{F}^H\mathbf{O}^{\frac{1}{2}} \preceq \mathbf{X}$, $\mathbf{F}\mathbf{F}^H \preceq \mathbf{Y}$. Using (34) and (35), the problem (30)–(33) can be rewritten as the following SDP problem

$$\min_{\mathbf{F},\mathbf{X},\mathbf{Y},t_2} \quad t_2 \tag{36}$$

$$\text{s.t.} \quad J_i(\mathbf{F},\mathbf{Y}) \le t_2, \quad i=1,\cdots,L \tag{37}$$

$$tr\{\mathbf{X}\} + tr\{\mathbf{G}\mathbf{G}^H\} \le P_r \tag{38}$$

$$\begin{pmatrix} \mathbf{X} & \mathbf{O}^{\frac{H}{2}}\mathbf{F} \\ \mathbf{F}^H\mathbf{O}^{\frac{1}{2}} & \mathbf{I}_{N_B} \end{pmatrix} \succeq 0 \tag{39}$$

$$tr\{\mathbf{Y}\} \le P_s \tag{40}$$

$$\begin{pmatrix} \mathbf{Y} & \mathbf{F} \\ \mathbf{F}^H & \mathbf{I}_{N_B} \end{pmatrix} \succeq 0 \tag{41}$$

TABLE I
PROCEDURE OF THE PROPOSED ITERATIVE TRANSCEIVER DESIGN ALGORITHM

1) Initialize the algorithm with $\mathbf{F}^{(0)} = \sqrt{P_s/N_B}[\mathbf{I}_{N_B}, \mathbf{0}]^T$ and $\mathbf{G}^{(0)} = \sqrt{P_r/tr\{\mathbf{\Xi}^{(0)}\}}\mathbf{I}_{N_R}$; Set $n=0$.
2) Update $\{\mathbf{W}^{(n)}\}$ using $\mathbf{F}^{(n)}$ and $\mathbf{G}^{(n)}$ as (20).
3) Update $\mathbf{G}^{(n+1)}$ using $\{\mathbf{W}^{(n)}\}$ and $\mathbf{F}^{(n)}$ by solving the problem (26)–(29).
4) Update $\mathbf{F}^{(n+1)}$ using $\{\mathbf{W}^{(n)}\}$ and $\mathbf{G}^{(n+1)}$ by solving the problem (36)–(41).
5) If $\big(\max_i J_i^{(n)} - \max_i J_i^{(n+1)}\big)/\max_i J_i^{(n)} < \varepsilon$, iteration ends; otherwise go to step (2).

where

$$J_i(\mathbf{F},\mathbf{Y}) = tr\Big\{\mathbf{\Theta}_i^{\frac{H}{2}}\mathbf{Y}\mathbf{\Theta}_i^{\frac{1}{2}}\Big\} + S_i - tr\Big\{\mathbf{W}_i\widehat{\mathbf{H}}_{2,i}\mathbf{G}\widehat{\mathbf{H}}_1\mathbf{F} + \mathbf{F}^H\widehat{\mathbf{H}}_1^H\mathbf{G}^H\widehat{\mathbf{H}}_{2,i}^H\mathbf{W}_i^H\Big\}.$$

The procedure of applying the iterative algorithm to solve the problem (17)–(19) is listed in Table I, where the superscript (n) denotes the number of iterations, ε is a small positive number close to zero, and $\max_i J_i^{(n)}$ stands for the value of (17) at the nth iteration. At each iteration, the complexity of solving the SDP problem (26)–(29) is $\mathcal{O}(N_R^4(L+1+4N_R^2)(L+1+2N_R)^{0.5})$, and the complexity order of solving the problem (36)–(41) is $\mathcal{O}((N_SN_B+2N_S^2)^2(L+2+2(N_S+N_B)^2)(L+2+2(N_S+N_B))^{0.5})$. Thus, it can be seen from Table I that the iterative algorithm has a high computational complexity, as its overall complexity is the per iteration complexity multiply with the number of iterations required till convergence. In the following, we develop a simplified algorithm for the minmax MSE problem (17)–(19) such that nearly optimal transmit and relay matrices can be designed with a significantly reduced computational complexity.

B. Simplified Transceiver Design Algorithm

In the following, we propose a low computational complexity approach to optimize $\mathbf{F}$ and $\mathbf{G}$. By substituting (20) back into (14), we have

$$\begin{aligned} &J_i(\mathbf{G},\mathbf{F}) \\ &= tr\big\{\mathbf{I}_{N_B} - \mathbf{F}^H\widehat{\mathbf{H}}_1^H\mathbf{G}^H\widehat{\mathbf{H}}_{2,i}^H\big(\widehat{\mathbf{H}}_{2,i}\mathbf{G}\mathbf{\Xi}\mathbf{G}^H\widehat{\mathbf{H}}_{2,i}^H \\ &\quad + tr\{\mathbf{G}\mathbf{\Xi}\mathbf{G}^H\mathbf{\Psi}_{2,i}\}\mathbf{\Sigma}_{2,i} + \mathbf{I}_{N_D}\big)^{-1}\widehat{\mathbf{H}}_{2,i}\mathbf{G}\widehat{\mathbf{H}}_1\mathbf{F}\big\}. \end{aligned} \tag{42}$$

Then the problem (17)–(19) can be written as the following optimization problem

$$\min_{\mathbf{F},\mathbf{G}} \quad \max_i \; J_i(\mathbf{G},\mathbf{F}) \tag{43}$$

$$\text{s.t.} \quad tr\big\{\mathbf{G}\mathbf{\Xi}\mathbf{G}^H\big\} \le P_r \tag{44}$$

$$tr\{\mathbf{F}\mathbf{F}^H\} \le P_s. \tag{45}$$

The following theorem shows the optimal structure of $\mathbf{G}$ as the solution to the problem (43)–(45).

Theorem 1: The optimal relay precoding matrix $\mathbf{G}$ for each transmitter-relay-receiver link can be expressed as

$$\mathbf{G} = \mathbf{T}\mathbf{W}_r = \mathbf{T}\mathbf{F}^H\widehat{\mathbf{H}}_1^H\mathbf{\Xi}^{-1} \tag{46}$$

where $\mathbf{W}_r \triangleq \mathbf{F}^H \widehat{\mathbf{H}}_1^H \boldsymbol{\Xi}^{-1}$ can be viewed as the linear MMSE receiver at the relay node, and $\mathbf{T}$ is unknown and can be viewed as the precoding matrix at the transmit side of the second-hop MIMO multicasting channel. Using $\mathbf{G}$ in (46), the MSE of the estimated signal at the ith receiver (42) can be reformulated as the sum of two MSE functions

$$\begin{aligned} & J_i(\mathbf{T},\mathbf{F}) \\ =\ & tr\big\{(\mathbf{I}_{N_B}+\mathbf{F}^H\widehat{\mathbf{H}}_1^H\boldsymbol{\Upsilon}^{-1}\widehat{\mathbf{H}}_1\mathbf{F})^{-1}\big\}+tr\Big\{\Big[\mathbf{R}^{-1}+\mathbf{T}^H\widehat{\mathbf{H}}_{2,i}^H \\ & \times\big(tr\big\{\mathbf{T}\mathbf{R}\mathbf{T}^H\boldsymbol{\Psi}_{2,i}\big\}\boldsymbol{\Sigma}_{2,i}+\mathbf{I}_{N_D}\big)^{-1}\widehat{\mathbf{H}}_{2,i}\mathbf{T}\Big]^{-1}\Big\} \end{aligned} \tag{47}$$

where

$$\boldsymbol{\Upsilon} = tr\{\mathbf{F}\mathbf{F}^H\boldsymbol{\Psi}_1\}\boldsymbol{\Sigma}_1+\mathbf{I}_{N_R} \tag{48}$$

$$\mathbf{R} = \mathbf{F}^H\widehat{\mathbf{H}}_1^H\big(\widehat{\mathbf{H}}_1\mathbf{F}\mathbf{F}^H\widehat{\mathbf{H}}_1^H+\boldsymbol{\Upsilon}\big)^{-1}\widehat{\mathbf{H}}_1\mathbf{F}. \tag{49}$$

Proof: See Appendix C. ■

Interestingly, Theorem 1 extends the MSE matrix decomposition from MIMO relay systems with perfect CSI to two-hop MIMO relay systems with imperfect CSI. In fact, the first term in (47) is the MSE of the first-hop signal waveform estimation at the relay node which is given by

$$\begin{aligned} & E_{H_1}\{E\{tr\{(\mathbf{W}_r\mathbf{y}_r-\mathbf{x})(\mathbf{W}_r\mathbf{y}_r-\mathbf{x})^H\}\}\} \\ =\ & E_{H_1}\{tr\{(\mathbf{W}_r\mathbf{H}_1\mathbf{F}-\mathbf{I}_{N_B})(\mathbf{W}_r\mathbf{H}_1\mathbf{F}-\mathbf{I}_{N_B})^H+\mathbf{W}_r\mathbf{W}_r^H\}\} \\ =\ & tr\{\mathbf{W}_r\boldsymbol{\Xi}\mathbf{W}_r^H-\mathbf{W}_r\widehat{\mathbf{H}}_1\mathbf{F}-\mathbf{F}^H\widehat{\mathbf{H}}_1^H\mathbf{W}_r^H+\mathbf{I}_{N_B}\} \\ =\ & tr\big\{(\mathbf{I}_{N_B}+\mathbf{F}^H\widehat{\mathbf{H}}_1^H\boldsymbol{\Upsilon}^{-1}\widehat{\mathbf{H}}_1\mathbf{F})^{-1}\big\} \end{aligned}$$

while the second term in (47) can be viewed as the increment of the MSE introduced by the second-hop.

Using (46), the power consumption at the relay node (16) can be rewritten as $tr(\mathbf{T}\mathbf{R}\mathbf{T}^H)$. Hence, the problem (43)–(45) can be equivalently rewritten as

$$\min_{\mathbf{F},\mathbf{T}} \quad \max_i\ J_i(\mathbf{T},\mathbf{F}) \tag{50}$$

$$\text{s.t.} \quad tr\{\mathbf{T}\mathbf{R}\mathbf{T}^H\} \le P_r \tag{51}$$

$$tr\{\mathbf{F}\mathbf{F}^H\} \le P_s. \tag{52}$$

Using the matrix inversion lemma [29]

$$(\mathbf{A}+\mathbf{B}\mathbf{C}\mathbf{D})^{-1} = \mathbf{A}^{-1}-\mathbf{A}^{-1}\mathbf{B}(\mathbf{D}\mathbf{A}^{-1}\mathbf{B}+\mathbf{C}^{-1})^{-1}\mathbf{D}\mathbf{A}^{-1} \tag{53}$$

matrix $\mathbf{R}$ in (49) can be expressed as

$$\begin{aligned} \mathbf{R} =\ & \mathbf{F}^H\widehat{\mathbf{H}}_1^H\big(\boldsymbol{\Upsilon}^{-1}-\boldsymbol{\Upsilon}^{-1}\widehat{\mathbf{H}}_1\mathbf{F} \\ & \times\big(\mathbf{F}^H\widehat{\mathbf{H}}_1^H\boldsymbol{\Upsilon}^{-1}\widehat{\mathbf{H}}_1\mathbf{F}+\mathbf{I}_{N_B}\big)^{-1}\mathbf{F}^H\widehat{\mathbf{H}}_1^H\boldsymbol{\Upsilon}^{-1}\big)\widehat{\mathbf{H}}_1\mathbf{F} \\ =\ & \mathbf{F}^H\widehat{\mathbf{H}}_1^H\boldsymbol{\Upsilon}^{-1}\widehat{\mathbf{H}}_1\mathbf{F}\big(\mathbf{F}^H\widehat{\mathbf{H}}_1^H\boldsymbol{\Upsilon}^{-1}\widehat{\mathbf{H}}_1\mathbf{F}+\mathbf{I}_{N_B}\big)^{-1}. \end{aligned} \tag{54}$$

In the case of small CSI mismatch, for example, $\boldsymbol{\Psi}_1=\sigma_e^2\mathbf{I}_{N_S}$ and $\boldsymbol{\Sigma}_1=\sigma_e^2\mathbf{I}_{N_R}$ with $\sigma_e^2 \ll 1$, where σ_e^2 denotes the variance of channel mismatch, $tr\{\mathbf{F}\mathbf{F}^H\boldsymbol{\Psi}_1\}\boldsymbol{\Sigma}_1$ is much smaller compared with $\mathbf{I}_{N_R}$. Then $\boldsymbol{\Upsilon}$ can be approximated as $\mathbf{I}_{N_R}$. Consequently it can be seen from (54) that in this case, $\mathbf{R}$ can be approximated as $\mathbf{I}_{N_B}$, if $\mathbf{F}^H\widehat{\mathbf{H}}_1^H\widehat{\mathbf{H}}_1\mathbf{F}$ is much greater than $\mathbf{I}_{N_B}$, which can be satisfied with a large P_s [23]. Therefore, the problem (50)–(52) can be approximated as

$$\begin{aligned} \min_{\mathbf{F},\mathbf{T}} \quad & \max_i\ tr\{(\mathbf{I}_{N_B}+\mathbf{F}^H\widehat{\mathbf{H}}_1^H\boldsymbol{\Upsilon}^{-1}\widehat{\mathbf{H}}_1\mathbf{F})^{-1}\} \\ & +tr\{(\mathbf{I}_{N_B}+\mathbf{T}^H\widehat{\mathbf{H}}_{2,i}^H\mathbf{K}_i^{-1}\widehat{\mathbf{H}}_{2,i}\mathbf{T})^{-1}\} \end{aligned} \tag{55}$$

$$\text{s.t.} \quad tr\{\mathbf{T}\mathbf{T}^H\} \le P_r \tag{56}$$

$$tr\{\mathbf{F}\mathbf{F}^H\} \le P_s \tag{57}$$

where

$$\mathbf{K}_i = tr\big\{\mathbf{T}\mathbf{T}^H\boldsymbol{\Psi}_{2,i}\big\}\boldsymbol{\Sigma}_{2,i}+\mathbf{I}_{N_D}, \quad i=1,\cdots,L. \tag{58}$$

We observe from the problem (55)–(57) that the first trace term in the objective function (55) does not depend on $\mathbf{T}$, while the value of the second trace term in (55) is not affected by $\mathbf{F}$. Therefore, the problem (55)–(57) can be decomposed into the problem of optimizing $\mathbf{F}$ as

$$\min_{\mathbf{F}} \quad tr\{(\mathbf{I}_{N_B}+\mathbf{F}^H\widehat{\mathbf{H}}_1^H\boldsymbol{\Upsilon}^{-1}\widehat{\mathbf{H}}_1\mathbf{F})^{-1}\} \tag{59}$$

$$\text{s.t.} \quad tr\{\mathbf{F}\mathbf{F}^H\} \le P_s \tag{60}$$

and the problem which optimizes $\mathbf{T}$ as

$$\min_{\mathbf{T}} \quad \max_i\ tr\{(\mathbf{I}_{N_B}+\mathbf{T}^H\widehat{\mathbf{H}}_{2,i}^H\mathbf{K}_i^{-1}\widehat{\mathbf{H}}_{2,i}\mathbf{T})^{-1}\} \tag{61}$$

$$\text{s.t.} \quad tr\{\mathbf{T}\mathbf{T}^H\} \le P_r. \tag{62}$$

Interestingly, we observed during simulations that keeping $\boldsymbol{\Upsilon}$ in the optimization of $\mathbf{F}$ achieves a better system performance than treating $\boldsymbol{\Upsilon}=\mathbf{I}_{N_R}$, without increasing the complexity of optimizing $\mathbf{F}$. Moreover, it can be seen from (48) and (58) that for multicasting MIMO relay systems with the exact CSI knowledge, we have $\boldsymbol{\Upsilon}=\mathbf{I}_{N_R}$ and $\mathbf{K}_i=\mathbf{I}_{N_D}$, $i=1,\cdots,L$, Thus, in this case, the problems (59)–(60) and (61)–(62) become non-robust transceiver design problems in [13]. Therefore, the proposed robust transceiver design extends the algorithm in [13] to the general practical multicasting MIMO systems with CSI mismatch.

1) Optimization of $\mathbf{F}$*:* When $\boldsymbol{\Psi}_1=\sigma_e^2\mathbf{I}_{N_S}$, i.e., the columns of $\boldsymbol{\Delta}_1$ are uncorrelated, from (48) we have $\boldsymbol{\Upsilon}=\sigma_e^2 tr\{\mathbf{F}\mathbf{F}^H\}\boldsymbol{\Sigma}_1+\mathbf{I}_{N_R}$. It can be easily shown that the optimal solution of the problem (59)–(60) must meet equality at the constraint (60), i.e., the optimal $\mathbf{F}$ should satisfy $tr\{\mathbf{F}\mathbf{F}^H\}=P_s$. In this case, $\boldsymbol{\Upsilon}=\sigma_e^2P_s\boldsymbol{\Sigma}_1+\mathbf{I}_{N_R}$ does not depend on $\mathbf{F}$, and the problem (59)–(60) has a closed-form solution as shown later.

However, for the general case of $\boldsymbol{\Psi}_1 \neq \sigma_e^2\mathbf{I}_{N_S}$, $\boldsymbol{\Upsilon}$ is a function of $\mathbf{F}$, which makes the problem (59)–(60) difficult to solve. To overcome this challenge, we apply the following inequality [15]

$$tr\big\{\mathbf{F}\mathbf{F}^H\boldsymbol{\Psi}_1\big\} \le P_s\lambda_{\mathrm{M}}(\boldsymbol{\Psi}_1) \tag{63}$$

where $\lambda_{\mathrm{M}}(\cdot)$ stands for the maximal eigenvalue of a matrix. From (63), an upper-bound of (59) is given by

$$\begin{aligned} & tr\{(\mathbf{I}_{N_B}+\mathbf{F}^H\widehat{\mathbf{H}}_1^H\boldsymbol{\Upsilon}^{-1}\widehat{\mathbf{H}}_1\mathbf{F})^{-1}\} \\ & \le tr\{(\mathbf{I}_{N_B}+\mathbf{F}^H\widehat{\mathbf{H}}_1^H(P_s\lambda_{\mathrm{M}}(\boldsymbol{\Psi}_1)\boldsymbol{\Sigma}_1+\mathbf{I}_{N_R})^{-1}\widehat{\mathbf{H}}_1\mathbf{F})^{-1}\}. \end{aligned} \tag{64}$$

Interestingly, the equality in (64) holds when $\boldsymbol{\Psi}_1=\sigma_e^2\mathbf{I}_{N_S}$, as in this case $\lambda_{\mathrm{M}}(\boldsymbol{\Psi}_1)=\sigma_e^2$.

Based on the discussion above, let us introduce

$$\mathbf{A} \triangleq \widehat{\mathbf{H}}_1^H(P_s\lambda_{\mathrm{M}}(\boldsymbol{\Psi}_1)\boldsymbol{\Sigma}_1+\mathbf{I}_{N_R})^{-1}\widehat{\mathbf{H}}_1.$$

The problem (59)–(60) is modified to the following transmitter precoding matrix optimization problem

$$\min_{\mathbf{F}} \quad tr\{(\mathbf{I}_{N_B} + \mathbf{F}^H \mathbf{A}\mathbf{F})^{-1}\} \tag{65}$$

$$\text{s.t.} \quad tr\{\mathbf{F}\mathbf{F}^H\} \le P_s. \tag{66}$$

Let us introduce the eigenvalue decomposition (EVD) of $\mathbf{A}$ as $\mathbf{A} = \mathbf{U}_A \mathbf{\Lambda}_A \mathbf{U}_A^H$, where the diagonal elements of $\mathbf{\Lambda}_A$ are sorted in a decreasing order. It can be shown that the solution to the problem (65)–(66) is given by

$$\mathbf{F} = \mathbf{U}_{A,1} \mathbf{\Lambda}_F^{\frac{1}{2}} \tag{67}$$

where $\mathbf{U}_{A,1}$ contains the leftmost N_B columns of $\mathbf{U}_A$ associated with the largest N_B eigenvalues and $\mathbf{\Lambda}_F$ is a diagonal matrix. Using (67), the problem (65)–(66) can be written as the following constrained optimization problem with scalar variables

$$\min_{\{\lambda_{F,i}\}} \quad \sum_{i=1}^{N_B} \frac{1}{1 + \lambda_{F,i}\lambda_{A,i}} \tag{68}$$

$$\text{s.t.} \quad \sum_{i=1}^{N_B} \lambda_{F,i} \le P_s \tag{69}$$

$$\lambda_{F,i} \ge 0, \qquad i = 1, \cdots, N_B \tag{70}$$

where $\lambda_{F,i}$ and $\lambda_{A,i}$, $i = 1, \cdots, N_B$, are the ith diagonal elements of $\mathbf{\Lambda}_F$ and $\mathbf{\Lambda}_A$, respectively, and $\{\lambda_{F,i}\} = \{\lambda_{F,1}, \cdots, \lambda_{F,N_B}\}$. The problem (68)–(70) has the well-known water-filling solution as

$$\lambda_{F,i} = \frac{1}{\lambda_{A,i}} \left(\sqrt{\frac{\lambda_{A,i}}{\mu}} - 1 \right)^+, \qquad i = 1, \cdots, N_B$$

where $(x)^+ = \max(x, 0)$, and $\mu > 0$ satisfies the nonlinear equation of $\sum_{i=1}^{N_B} \frac{1}{\lambda_{A,i}} \left(\sqrt{\frac{\lambda_{A,i}}{\mu}} - 1 \right)^+ = P_s$.

2) Optimization of $\mathbf{T}$*:* Similar to the technique used in optimizing $\mathbf{F}$, we have $tr\{\mathbf{T}\mathbf{T}^H \mathbf{\Psi}_{2,i}\} \le P_r \lambda_{\mathrm{M}}(\mathbf{\Psi}_{2,i})$. Let us introduce

$$\mathbf{B}_i \triangleq \widehat{\mathbf{H}}_{2,i}^H (P_r \lambda_{\mathrm{M}}(\mathbf{\Psi}_{2,i}) \mathbf{\Sigma}_{2,i} + \mathbf{I}_{N_D})^{-1} \widehat{\mathbf{H}}_{2,i}, \quad i = 1, \cdots, L.$$

The problem (61)–(62) can be modified to the following problem

$$\min_{\mathbf{T}} \quad \max_i \ tr\{(\mathbf{I}_{N_B} + \mathbf{T}^H \mathbf{B}_i \mathbf{T})^{-1}\} \tag{71}$$

$$\text{s.t.} \quad tr\{\mathbf{T}\mathbf{T}^H\} \le P_r. \tag{72}$$

Using the matrix identity of $tr\{(\mathbf{I}_m + \mathbf{A}_{m\times n}\mathbf{B}_{n\times m})^{-1}\} = tr\{(\mathbf{I}_n + \mathbf{B}_{n\times m}\mathbf{A}_{m\times n})^{-1}\} + m - n$, the min-max problem (71)–(72) can be written as

$$\min_{\mathbf{Q}} \quad \max_i \ tr\left\{ \left(\mathbf{I}_{N_R} + \mathbf{B}_i^{\frac{1}{2}} \mathbf{Q} \mathbf{B}_i^{\frac{H}{2}} \right)^{-1} \right\} + N_B - N_R \tag{73}$$

$$\text{s.t.} \quad tr(\mathbf{Q}) \le P_r \tag{74}$$

$$\mathbf{Q} \succeq 0 \tag{75}$$

where $\mathbf{Q} = \mathbf{T}\mathbf{T}^H$. Note that for the case of $N_R > N_B$ (when $\mathbf{T}$ is a tall matrix), to facilitate solving the problem (73)–(75), the constraint of $\mathrm{rank}(\mathbf{Q}) = N_B$ is removed in the problem (73)–(75). Nevertheless, $\mathbf{T}$ can be obtained from $\mathbf{Q}$ as $\mathbf{T} = c\mathbf{U}_{q,1}\mathbf{\Lambda}_{q,1}^{\frac{1}{2}}$, where $\mathbf{U}_q \mathbf{\Lambda}_q \mathbf{U}_q^H$ is the EVD of $\mathbf{Q}$, $\mathbf{\Lambda}_{q,1}$ contains the N_B largest eigenvalues of $\mathbf{Q}$, $\mathbf{U}_{q,1}$ contains the eigenvectors associated with the largest N_B eigenvalues, and $c = \sqrt{P_r / tr\{\mathbf{U}_{q,1}\mathbf{\Lambda}_{q,1}\mathbf{U}_{q,1}^H\}}$ is a scalar to meet equality in the power constraint (74).

Let us introduce a real valued slack variable ρ and PSD matrices $\mathbf{Z}_i$ with $\left(\mathbf{I}_{N_R} + \mathbf{B}_i^{\frac{1}{2}} \mathbf{Q} \mathbf{B}_i^{\frac{H}{2}}\right)^{-1} \preceq \mathbf{Z}_i, i = 1, \cdots, L$. By using the Schur complement [24], the problem (73)–(75) can be equivalently rewritten as

$$\min_{\rho, \mathbf{Q}, \{\mathbf{Z}_i\}} \quad \rho \tag{76}$$

$$\text{s.t.} \quad tr(\mathbf{Z}_i) \le \rho, \quad i = 1, \cdots, L \tag{77}$$

$$tr(\mathbf{Q}) \le P_r \tag{78}$$

$$\begin{pmatrix} \mathbf{Z}_i & \mathbf{I}_{N_R} \\ \mathbf{I}_{N_R} & \mathbf{I}_{N_R} + \mathbf{B}_i^{\frac{1}{2}} \mathbf{Q} \mathbf{B}_i^{\frac{H}{2}} \end{pmatrix} \succeq 0, \quad i = 1, \cdots, L \tag{79}$$

$$\mathbf{Q} \succeq 0 \tag{80}$$

where $\{\mathbf{Z}_i\} = \{\mathbf{Z}_1, \cdots, \mathbf{Z}_L\}$. The problem (76)–(80) is a convex SDP problem and can be solved by the convex programming toolbox CVX [30].

In the proposed simplified transceiver design, the original transmitter and relay matrices optimization problem (43)–(45) is decomposed into two MSE minimization problems. The computational complexity of optimizing the relay precoding matrix is governed by solving the SDP problem (76)–(80), which has an order of $\mathcal{O}((L+1)^{3.5} N_R^{6.5})$. The computation of optimizing the transmitter precoding matrix involves performing matrix EVD and calculating the power loading parameters, whose complexity are negligible compared with that of solving the problem (76)–(80). Thus, the overall complexity order of the proposed simplified design is $\mathcal{O}((L+1)^{3.5} N_R^{6.5})$, which is the same as that of the non-robust transceiver design in [13], and much lower than that of the iterative transceiver design algorithm in Section III-A.

IV. Simulation Results

In this section, we investigate the performance of the proposed robust transceiver design algorithm for multicasting MIMO relay systems through numerical simulations. We simulate a two-hop nonregenerative MIMO relay multicasting system. The information-carrying symbols are modulated using the QPSK constellations. The signal-to-noise ratios (SNRs) of the first-hop and second-hop channels are defined as $\mathrm{SNR}_1 = P_s/N_S$ and $\mathrm{SNR}_2 = P_r/N_R$, respectively. Unless explicitly mentioned, we set $\mathrm{SNR}_1 = 30$ dB and $N_B = N_S = N_R = N_D = 4$. In the simulations, the correlation matrices of the channel estimation errors are modeled as [15]

$$[\mathbf{\Psi}_1]_{m,n} = \alpha^{|m-n|}, \quad m, n = 1, \cdots, N_s$$

$$[\mathbf{\Psi}_{2,i}]_{m,n} = \alpha^{|m-n|}, \quad m, n = 1, \cdots, N_R, \ \ i = 1, \cdots, L$$

$$[\mathbf{\Sigma}_1]_{m,n} = \sigma_e^2 \beta^{|m-n|}, \quad m, n = 1, \cdots, N_R$$

$$[\mathbf{\Sigma}_{2,i}]_{m,n} = \sigma_e^2 \beta^{|m-n|}, \quad m, n = 1, \cdots, N_D, \ \ i = 1, \cdots, L$$

where $0 \le \alpha, \beta \le 1$ are correlation coefficients, and σ_e^2 denotes the variance of the estimation error.

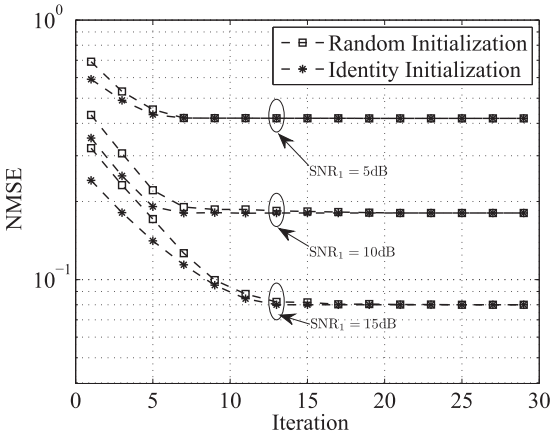


Fig. 2. Example 1: Convergence rate of the proposed robust-ITE algorithm at various SNR_1. $\mathrm{SNR}_2 = 20$ dB, $\sigma_e^2 = 0.02$, $L = 2$, and $\alpha = \beta = 0$.

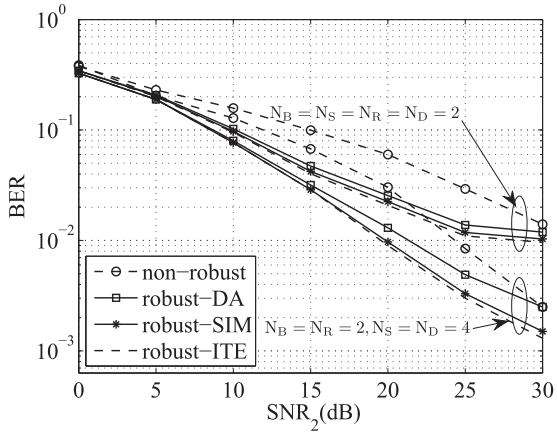


Fig. 3. Example 2: BER versus SNR_2 with different number of antennas. $\sigma_e^2 = 0.02$, $L = 2$, and $\alpha = \beta = 0$.

The estimated channel matrices $\widehat{\mathbf{H}}_1$ and $\widehat{\mathbf{H}}_{2,i}$ are generated based on the following distributions

$$\widehat{\mathbf{H}}_1 \sim \mathcal{CN}\Big(\mathbf{0}, \frac{1-\sigma_e^2}{\sigma_e^2}\boldsymbol{\Sigma}_1 \otimes \boldsymbol{\Psi}_1^T\Big)$$

$$\widehat{\mathbf{H}}_{2,i} \sim \mathcal{CN}\Big(\mathbf{0}, \frac{1-\sigma_e^2}{\sigma_e^2}\boldsymbol{\Sigma}_{2,i} \otimes \boldsymbol{\Psi}_{2,i}^T\Big), \quad i = 1, \cdots, L.$$

We compare the performance of the proposed robust iterative algorithm (robust-ITE), the proposed simplified robust transceiver design (robust-SIM) with

1) The algorithm in [13] using only the estimated CSI $\widehat{\mathbf{H}}_1$, $\widehat{\mathbf{H}}_{2,i}, i = 1, \cdots, L$ (denoted as the non-robust algorithm).
2) The algorithm in [13] with the true CSI $\mathbf{H}_1$, $\mathbf{H}_{2,i}$, $i = 1, \cdots, L$ (denoted as the TCSI algorithm), which provides a lower bound of the system performance.
3) The robust decomposition approach (robust-DA) proposed in [22].

In the first simulation example, we study the convergence rate of the proposed iterative algorithm at various levels of SNR_1. We set $\mathrm{SNR}_2 = 20$ dB, $L = 2$, $\sigma_e^2 = 0.02$, $\alpha = \beta = 0$, and the robust-ITE algorithm is initialized with either scaled identity or random matrices. Fig. 2 illustrates the normalized MSE (NMSE) of the proposed robust-ITE algorithm versus the number of iterations. It can be seen from Fig. 2 that the NMSE always decreases and reduces to fixed values after a few iterations. Interestingly, the number of iterations required till convergence increases SNR_1. Moreover, it can be observed from Fig. 2 that the convergence speed of the robust-ITE algorithm with the scaled identity matrix initialization is faster than that of random matrix initialization. Hence, for the rest of the simulations, the scaled identity matrix is chosen to initialize the robust-ITE algorithm.

In the second simulation example, we study the BER performance of the proposed algorithms with different number of antennas at the transmitter, relay and receivers. In this example, we set $L = 2$, $\sigma_e^2 = 0.02$, and $\alpha = \beta = 0$. Fig. 3 shows the BER of four algorithms versus SNR_2. It can be seen from Fig. 3 that with $N_S = N_D = 4$, the system BER of the proposed robust algorithms approaches 1×10^{-3}. Hence, with a fixed N_B, the BER performance of the proposed robust algorithms improves

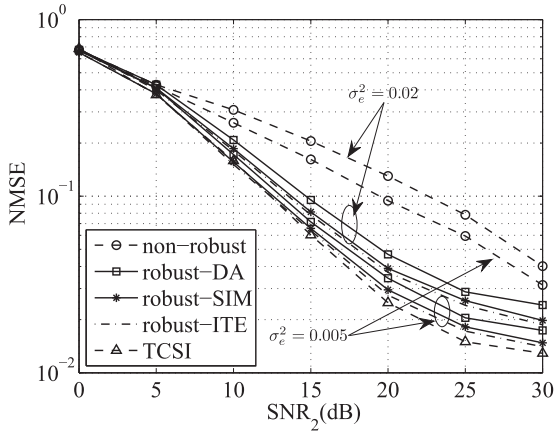


Fig. 4. Example 3: NMSE versus SNR_2 at various σ_e^2. $L = 2$ and $\alpha = \beta = 0$.

when N_S and N_D increase, as the system spatial diversity gain increases with the number of antennas. Moreover, it can be seen from Fig. 3 that the BER improvement of both proposed algorithms over the robust-DA method is larger with $N_S = N_D = 4$ compared with $N_S = N_D = 2$.

In the third simulation example, we study the performance of the proposed algorithms at various values of estimation error variance σ_e^2. In this simulation, the number of receivers is set as $L = 2$ and the correlation coefficients are set as $\alpha = \beta = 0$ (i.e., $\boldsymbol{\Psi}_1 = \mathbf{I}_{N_S}$, $\boldsymbol{\Psi}_{2,i} = \mathbf{I}_{N_R}$, $\boldsymbol{\Sigma}_1 = \sigma_e^2\mathbf{I}_{N_R}$ and $\boldsymbol{\Sigma}_{2,i} = \sigma_e^2\mathbf{I}_{N_D}$). Fig. 4 shows the NMSE of five algorithms versus SNR_2. It can be observed from Fig. 4 that over the whole range of SNR_2, the proposed robust algorithms significantly outperform the non-robust algorithm. Both proposed algorithms have a better MSE performance than the robust-DA algorithm. It can also be seen from Fig. 4 that for all three robust and non-robust algorithms, the system MSE decreases when σ_e^2 is reduced.

For this example, the BER yielded by the five algorithms tested versus SNR_2 is shown in Fig. 5. It can be clearly seen from Fig. 5 that the three robust algorithms yield much lower BER compared with the non-robust algorithm. As σ_e^2 decreases, the system BER of all four algorithms decreases. We also observe

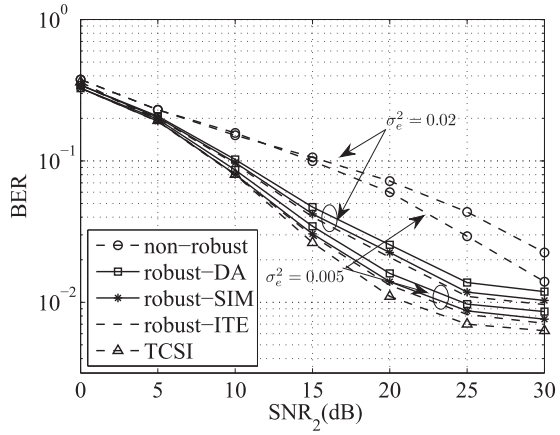


Fig. 5. Example 3: BER versus SNR_2 at various σ_e^2. $L = 2$ and $\alpha = \beta = 0$.

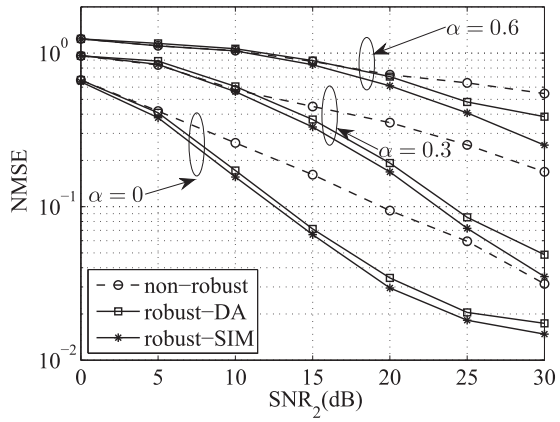


Fig. 6. Example 4: NMSE versus SNR_2 at various α. $L = 2$, $\sigma_e^2 = 0.005$, and $\beta = 0$.

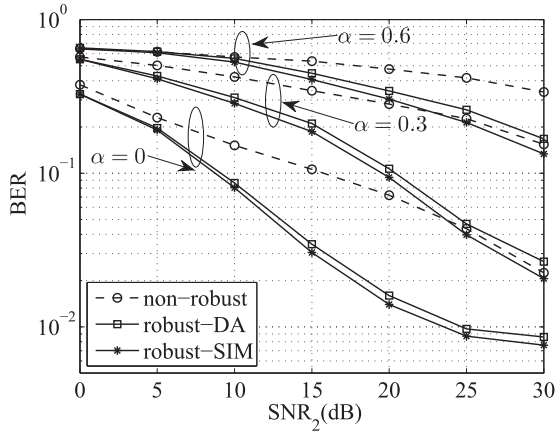


Fig. 7. Example 4: BER versus SNR_2 at various α. $L = 2$, $\sigma_e^2 = 0.005$, and $\beta = 0$.

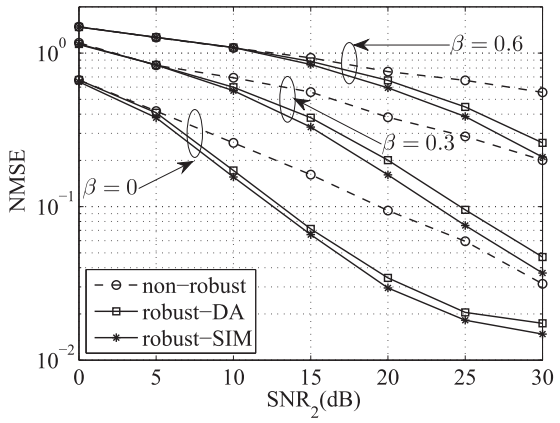


Fig. 8. Example 5: NMSE versus SNR_2 at various β. $L = 2$, $\sigma_e^2 = 0.005$, and $\alpha = 0$.

from Fig. 5 that both proposed transceiver design algorithms have a better BER performance than the robust-DA algorithm.

It can be observed from Figs. 4 and 5 that the proposed robust-ITE and robust-SIM algorithms achieve very close NMSE and BER performance to that of the approach in [13] with true CSI. It can also be seen from Figs. 4 and 5 that the MSE and BER yielded by the proposed simplified design is very close to that of the proposed iterative algorithm. In the following simulation examples, we focus on the simplified transceiver design, as it has negligible performance loss but with significantly reduced computational complexity, which makes it more attractive for practical multicasting MIMO relay systems.

In the fourth example, we investigate the performance of the proposed simplified algorithm at various values of the correlation coefficient α with $\beta = 0$. In this example, the number of receivers is set as $L = 2$ and the estimation error variance is chosen as $\sigma_e^2 = 0.005$. Figs. 6 and 7 demonstrate the MSE and BER performance of the three algorithms tested versus SNR_2, respectively. It can be seen from Figs. 6 and 7 that the proposed robust-SIM algorithm provides better MSE and BER performances than the non-robust and robust-DA algorithms for all the α tested. Both robust algorithms outperform the non-robust algorithm. Moreover, the MSE and BER yielded by all three algorithms increase with α. This is due to the fact that as α increases, the correlation among the elements of channel matrices increase, leading to the loss of spatial diversity.

In the fifth example, we study the performance of the proposed robust-SIM algorithm at various values of the correlation coefficient β with $\alpha = 0$, $L = 2$, and $\sigma_e^2 = 0.005$. The MSE and BER performances of all three algorithms are shown in Figs. 8 and 9, respectively. Similar to Figs. 6 and 7, it can be seen from Figs. 8 and 9 that the proposed robust-SIM algorithm outperforms the non-robust and robust-DA algorithms. It can also be observed that the MSE and BER of the three algorithms increase with the increasing correlation coefficient β.

In the sixth simulation example, we study the performance of the proposed robust-SIM algorithm at various number of receivers L. In Fig. 10, we compare the MSE performance of the proposed robust-SIM algorithm with the TCSI and robust-DA algorithms at various L as a function of SNR_2 with $\sigma_e^2 = 0.005$ and $\alpha = \beta = 0$. It can be noted from Fig. 10 that as we increase the number of receivers, the MSE of the three algorithms increases. This is reasonable since it is more likely to find a worse relay-receiver channel among the increased number of users and we choose the worst-user MSE as the objective function. Moreover, it can be seen from Fig. 10 that the proposed robust-SIM algorithm outperforms the robust-DA algorithm at both $L = 2$ and $L = 6$.

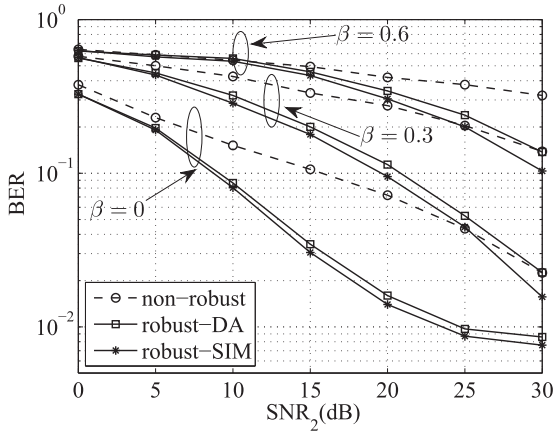


Fig. 9. Example 5: BER versus SNR_2 at various β. $L = 2$, $\sigma_e^2 = 0.005$, and $\alpha = 0$.

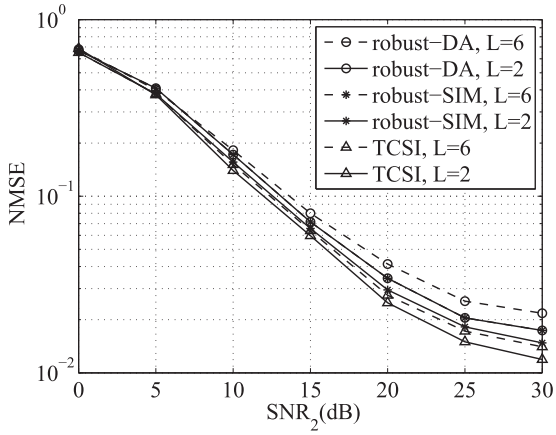


Fig. 10. Example 6: NMSE versus SNR_2 at various L. $\sigma_e^2 = 0.005$ and $\alpha = \beta = 0$.

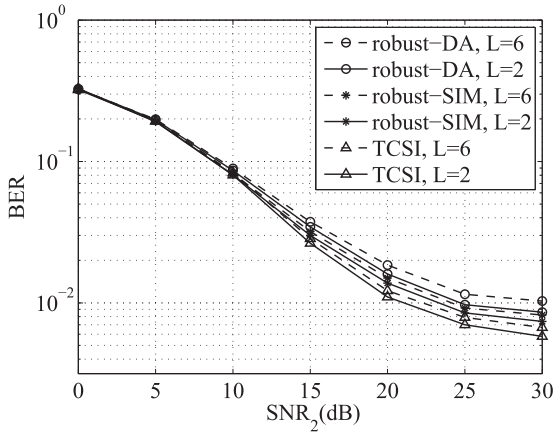


Fig. 11. Example 6: BER versus SNR_2 at various L. $\sigma_e^2 = 0.005$ and $\alpha = \beta = 0$.

For this example, the BER performance of the TCSI, robust-SIM, and robust-DA algorithms is demonstrated in Fig. 11. It can be noted from Fig. 11 that similar to Fig. 10, the BER of the three algorithms increase with the number of receivers and it can be noticed from Fig. 11 that the proposed robust-SIM algorithm has a lower BER compared with the robust-DA algorithm for both $L = 2$ and $L = 6$. From Figs. 10 and 11 we can see that the proposed robust-SIM algorithm has a very closed MSE and BER performance to the TCSI algorithm for both $L = 2$ and $L = 6$.

V. Conclusion

In this paper, we have addressed the challenging issue of robust transceiver optimization for two-hop multicasting MIMO relay systems when there is mismatch between the true and estimated channel matrices. The true channel matrices are assumed as Gaussian random matrices with the estimated channels as the mean value, and estimation error follows the well-known Kronecker model. In the proposed iterative algorithm, the transmitter, relay, and receiver matrices are designed to minimize the maximal MSE of the signal waveform estimation among all receivers. Due to the high computational complexity of the proposed iterative algorithm, we develop a simplified transceiver design in which we show that the MSE at each receiver can be decomposed into the sum of the MSEs of the first-hop and second-hop channels. Based on the proposed MSE decomposition, a transceiver optimization algorithm with low computational complexity has been developed. Numerical simulations demonstrate that the proposed transceiver design algorithms have an improved robustness against the mismatch between the true and estimated channels than existing algorithms.

Appendix A
Derivation of (14)

Taking the statistical expectation of (12) with respect to $\mathbf{H}_1$ and $\mathbf{H}_{2,i}$, we have

$$\begin{aligned} & E_{H_1,H_{2,i}}\{M_i(\mathbf{W}_i,\mathbf{G},\mathbf{F})\} \\ = \; & tr\{\mathbf{W}_i E_{H_{2,i}}\{\mathbf{H}_{2,i}\mathbf{G}E_{H_1}\{\mathbf{H}_1\mathbf{F}\mathbf{F}^H\mathbf{H}_1^H\}\mathbf{G}^H\mathbf{H}_{2,i}^H\}\mathbf{W}_i^H \\ & - \mathbf{W}_i E_{H_1,H_{2,i}}\{\mathbf{H}_{2,i}\mathbf{G}\mathbf{H}_1\}\mathbf{F} \\ & - \mathbf{F}^H E_{H_1,H_{2,i}}\{\mathbf{H}_1^H\mathbf{G}^H\mathbf{H}_{2,i}^H\}\mathbf{W}_i^H + \mathbf{I}_{N_B} \\ & + \mathbf{W}_i\big(E_{H_{2,i}}\{\mathbf{H}_{2,i}\mathbf{G}\mathbf{G}^H\mathbf{H}_{2,i}^H\} + \mathbf{I}_{N_D}\big)\mathbf{W}_i^H\}. \end{aligned} \tag{81}$$

Using the distribution in (4), (6) and applying Lemma 1, we obtain

$$E_{H_1}\{\mathbf{H}_1\mathbf{F}\mathbf{F}^H\mathbf{H}_1^H\} = \widehat{\mathbf{H}}_1\mathbf{F}\mathbf{F}^H\widehat{\mathbf{H}}_1^H + tr\{\mathbf{F}\mathbf{F}^H\mathbf{\Psi}_1\}\mathbf{\Sigma}_1. \tag{82}$$

By substituting (82) back into (81), we have

$$\begin{aligned} & E_{H_1,H_{2,i}}\{M_i(\mathbf{W}_i,\mathbf{G},\mathbf{F})\} \\ = \; & tr\{\mathbf{W}_i\big(E_{H_{2,i}}\{\mathbf{H}_{2,i}\mathbf{G}\mathbf{\Xi}\mathbf{G}^H\mathbf{H}_{2,i}^H\} + \mathbf{I}_{N_D}\big)\mathbf{W}_i^H \\ & - \mathbf{W}_i\widehat{\mathbf{H}}_{2,i}\mathbf{G}\widehat{\mathbf{H}}_1\mathbf{F} - \mathbf{F}^H\widehat{\mathbf{H}}_1^H\mathbf{G}^H\widehat{\mathbf{H}}_{2,i}^H\mathbf{W}_i^H + \mathbf{I}_{N_B}\}. \end{aligned} \tag{83}$$

Based on the distribution in (5), (7) and applying Lemma 1, we obtain

$$\begin{aligned} & E_{H_{2,i}}\{\mathbf{H}_{2,i}\mathbf{G}\mathbf{\Xi}\mathbf{G}^H\mathbf{H}_{2,i}^H\} \\ & = \widehat{\mathbf{H}}_{2,i}\mathbf{G}\mathbf{\Xi}\mathbf{G}^H\widehat{\mathbf{H}}_{2,i}^H + tr\{\mathbf{G}\mathbf{\Xi}\mathbf{G}^H\mathbf{\Psi}_{2,i}\}\mathbf{\Sigma}_{2,i}. \end{aligned} \tag{84}$$

By substituting (84) back into (83), we obtain (14).

APPENDIX B
DERIVATION OF (34)

By substituting (15) back into (14) and using the identity of $tr\{\mathbf{AB}\} = tr\{\mathbf{BA}\}$, (14) can be rewritten as

$$\begin{aligned} J_i(\mathbf{F}) = \; & tr\{\mathbf{F}\mathbf{F}^H\widehat{\mathbf{H}}_1^H\mathbf{G}^H\widehat{\mathbf{H}}_{2,i}^H\mathbf{W}_i^H\mathbf{W}_i\widehat{\mathbf{H}}_{2,i}\mathbf{G}\widehat{\mathbf{H}}_1\} \\ & + tr\{\mathbf{W}_i\widehat{\mathbf{H}}_{2,i}\mathbf{G}\boldsymbol{\Sigma}_1\mathbf{G}^H\widehat{\mathbf{H}}_{2,i}^H\mathbf{W}_i^H\}tr\{\mathbf{F}\mathbf{F}^H\boldsymbol{\Psi}_1\} \\ & + tr\{\mathbf{W}_i\boldsymbol{\Sigma}_{2,i}\mathbf{W}_i^H\}(tr\{\mathbf{F}\mathbf{F}^H\widehat{\mathbf{H}}_1^H\mathbf{G}^H\boldsymbol{\Psi}_{2,i}\mathbf{G}\widehat{\mathbf{H}}_1\} \\ & + tr\{\mathbf{F}\mathbf{F}^H\boldsymbol{\Psi}_1\}tr\{\mathbf{G}\boldsymbol{\Sigma}_1\mathbf{G}^H\boldsymbol{\Psi}_{2,i}\} + tr\{\mathbf{G}\mathbf{G}^H\boldsymbol{\Psi}_{2,i}\}) \\ & + tr\{\mathbf{W}_i\mathbf{W}_i^H\} + N_B + tr\{\mathbf{W}_i\widehat{\mathbf{H}}_{2,i}\mathbf{G}\mathbf{G}^H\widehat{\mathbf{H}}_{2,i}^H\mathbf{W}_i^H\} \\ & - tr\{\mathbf{W}_i\widehat{\mathbf{H}}_{2,i}\mathbf{G}\widehat{\mathbf{H}}_1\mathbf{F} + \mathbf{F}^H\widehat{\mathbf{H}}_1^H\mathbf{G}^H\widehat{\mathbf{H}}_{2,i}^H\mathbf{W}_i^H\}. \quad (85) \end{aligned}$$

By introducing for $i = 1, \cdots, L$

$$\begin{aligned} \boldsymbol{\Theta}_i^{\frac{1}{2}}\boldsymbol{\Theta}_i^{\frac{H}{2}} = \; & \widehat{\mathbf{H}}_1^H\mathbf{G}^H\widehat{\mathbf{H}}_{2,i}^H\mathbf{W}_i^H\mathbf{W}_i\widehat{\mathbf{H}}_{2,i}\mathbf{G}\widehat{\mathbf{H}}_1 \\ & + tr\{\mathbf{W}_i\widehat{\mathbf{H}}_{2,i}\mathbf{G}\boldsymbol{\Sigma}_1\mathbf{G}^H\widehat{\mathbf{H}}_{2,i}^H\mathbf{W}_i^H\}\boldsymbol{\Psi}_1 \\ & + tr\{\mathbf{W}_i\boldsymbol{\Sigma}_{2,i}\mathbf{W}_i^H\}(\widehat{\mathbf{H}}_1^H\mathbf{G}^H\boldsymbol{\Psi}_{2,i}\mathbf{G}\widehat{\mathbf{H}}_1 \\ & + tr\{\mathbf{G}\boldsymbol{\Sigma}_1\mathbf{G}^H\boldsymbol{\Psi}_{2,i}\}\boldsymbol{\Psi}_1) \quad (86) \end{aligned}$$

and

$$\begin{aligned} S_i = \; & tr\{\mathbf{W}_i\boldsymbol{\Sigma}_{2,i}\mathbf{W}_i^H\}tr\{\mathbf{G}\mathbf{G}^H\boldsymbol{\Psi}_{2,i}\} + tr\{\mathbf{W}_i\mathbf{W}_i^H\} \\ & + tr\{\mathbf{W}_i\widehat{\mathbf{H}}_{2,i}\mathbf{G}\mathbf{G}^H\widehat{\mathbf{H}}_{2,i}^H\mathbf{W}_i^H\} + N_B \quad (87) \end{aligned}$$

(85) can be rewritten as (34).

APPENDIX C
PROOF OF THEOREM 1

Let us introduce

$$\mathbf{P}_i \triangleq tr\big\{\mathbf{G}\boldsymbol{\Xi}\mathbf{G}^H\boldsymbol{\Psi}_{2,i}\big\}\boldsymbol{\Sigma}_{2,i} + \mathbf{I}_{N_D}, \qquad i = 1, \cdots, L. \quad (88)$$

The MSE function in (42) can be rewritten as

$$\begin{aligned} & J_i(\mathbf{G}, \mathbf{F}) \\ = \; & tr\{\mathbf{I}_{N_B} - \mathbf{F}^H\widehat{\mathbf{H}}_1^H\mathbf{G}^H\widehat{\mathbf{H}}_{2,i}^H\big(\widehat{\mathbf{H}}_{2,i}\mathbf{G}(\widehat{\mathbf{H}}_1\mathbf{F}\mathbf{F}^H\widehat{\mathbf{H}}_1^H + \boldsymbol{\Upsilon}) \\ & \times\mathbf{G}^H\widehat{\mathbf{H}}_{2,i}^H + \mathbf{P}_i\big)^{-1}\widehat{\mathbf{H}}_{2,i}\mathbf{G}\widehat{\mathbf{H}}_1\mathbf{F}\}. \quad (89) \end{aligned}$$

By introducing

$$\tilde{\mathbf{G}} \triangleq \mathbf{G}\boldsymbol{\Upsilon}^{\frac{1}{2}}, \quad \tilde{\mathbf{H}}_1 \triangleq \boldsymbol{\Upsilon}^{-\frac{1}{2}}\widehat{\mathbf{H}}_1, \quad \tilde{\mathbf{H}}_{2,i} \triangleq \mathbf{P}_i^{-\frac{1}{2}}\widehat{\mathbf{H}}_{2,i} \quad (90)$$

we can rewrite (89) as

$$\begin{aligned} & J_i(\tilde{\mathbf{G}}, \mathbf{F}) \\ = \; & tr\{\mathbf{I}_{N_B} - \mathbf{F}^H\tilde{\mathbf{H}}_1^H\tilde{\mathbf{G}}^H\tilde{\mathbf{H}}_{2,i}^H\big(\tilde{\mathbf{H}}_{2,i}\tilde{\mathbf{G}}(\tilde{\mathbf{H}}_1\mathbf{F}\mathbf{F}^H\tilde{\mathbf{H}}_1^H + \mathbf{I}_{N_R}) \\ & \times\tilde{\mathbf{G}}^H\tilde{\mathbf{H}}_{2,i}^H + \mathbf{I}_{N_D}\big)^{-1}\tilde{\mathbf{H}}_{2,i}\tilde{\mathbf{G}}\tilde{\mathbf{H}}_1\mathbf{F}\} \\ = \; & tr\big\{\big[\mathbf{I}_{N_B} + \mathbf{F}^H\tilde{\mathbf{H}}_1^H\tilde{\mathbf{G}}^H\tilde{\mathbf{H}}_{2,i}^H\big(\tilde{\mathbf{H}}_{2,i}\tilde{\mathbf{G}}\tilde{\mathbf{G}}^H\tilde{\mathbf{H}}_{2,i}^H + \mathbf{I}_{N_D}\big)^{-1} \\ & \times\tilde{\mathbf{H}}_{2,i}\tilde{\mathbf{G}}\tilde{\mathbf{H}}_1\mathbf{F}\big]^{-1}\big\} \quad (91) \end{aligned}$$

where the matrix inversion lemma (53) has been applied to obtain (91). Using (90), the constraint (44) can be rewritten as

$$tr\big\{\tilde{\mathbf{G}}(\tilde{\mathbf{H}}_1\mathbf{F}\mathbf{F}\tilde{\mathbf{H}}_1^H + \mathbf{I}_{N_R})\tilde{\mathbf{G}}^H\big\} \le P_r. \quad (92)$$

Based on (45), (91), and (92), the MMSE-based transceiver optimization problem for the ith receiver can be written as

$$\begin{aligned} \min_{\mathbf{F},\tilde{\mathbf{G}}} \quad & J_i(\tilde{\mathbf{G}}, \mathbf{F}) \quad (93) \\ \text{s.t.} \quad & tr\big\{\tilde{\mathbf{G}}(\tilde{\mathbf{H}}_1\mathbf{F}\mathbf{F}\tilde{\mathbf{H}}_1^H + \mathbf{I}_{N_R})\tilde{\mathbf{G}}^H\big\} \le P_r \quad (94) \\ & tr\{\mathbf{F}\mathbf{F}^H\} \le P_s. \quad (95) \end{aligned}$$

It can be shown similar to [23] that the optimal $\tilde{\mathbf{G}}$ as the solution to the problem (93)–(95) can be written as

$$\tilde{\mathbf{G}} = \mathbf{T}\mathbf{F}^H\tilde{\mathbf{H}}_1^H(\tilde{\mathbf{H}}_1\mathbf{F}\mathbf{F}^H\tilde{\mathbf{H}}_1^H + \mathbf{I}_{N_R})^{-1} \quad (96)$$

and the objective function (93) can be decomposed into two MSE terms as

$$\begin{aligned} J_i(\mathbf{T}, \mathbf{F}) = \; & tr\{(\mathbf{I}_{N_B} + \mathbf{F}^H\tilde{\mathbf{H}}_1^H\tilde{\mathbf{H}}_1\mathbf{F})^{-1}\} \\ & + tr\{(\tilde{\mathbf{R}}^{-1} + \mathbf{T}^H\tilde{\mathbf{H}}_{2,i}^H\tilde{\mathbf{H}}_{2,i}\mathbf{T})^{-1}\} \quad (97) \end{aligned}$$

where

$$\tilde{\mathbf{R}} = \mathbf{F}^H\tilde{\mathbf{H}}_1^H(\tilde{\mathbf{H}}_1\mathbf{F}\mathbf{F}^H\tilde{\mathbf{H}}_1^H + \mathbf{I}_{N_R})^{-1}\tilde{\mathbf{H}}_1\mathbf{F}. \quad (98)$$

Substituting $\tilde{\mathbf{G}}$ and $\tilde{\mathbf{H}}_1$ in (90) and $\boldsymbol{\Upsilon}$ in (48) into (96), we have $\mathbf{G} = \mathbf{T}\mathbf{F}^H\widehat{\mathbf{H}}_1^H\boldsymbol{\Xi}^{-1}$, which proves (46). By Substituting $\tilde{\mathbf{H}}_1$ in (90) into (98), we obtain $\tilde{\mathbf{R}} = \mathbf{F}^H\widehat{\mathbf{H}}_1^H(\widehat{\mathbf{H}}_1\mathbf{F}\mathbf{F}^H\widehat{\mathbf{H}}_1^H + \boldsymbol{\Upsilon})^{-1}\widehat{\mathbf{H}}_1\mathbf{F} = \mathbf{R}$ in (49). Moreover, by substituting (46) into (88), we have

$$\begin{aligned} \mathbf{P}_i = \; & tr\big\{\mathbf{T}\mathbf{F}^H\widehat{\mathbf{H}}_1^H\boldsymbol{\Xi}^{-1}\widehat{\mathbf{H}}_1\mathbf{F}\mathbf{T}^H\boldsymbol{\Psi}_{2,i}\big\}\boldsymbol{\Sigma}_{2,i} + \mathbf{I}_{N_D} \\ = \; & tr\big\{\mathbf{T}\mathbf{R}\mathbf{T}^H\boldsymbol{\Psi}_{2,i}\big\}\boldsymbol{\Sigma}_{2,i} + \mathbf{I}_{N_D}. \quad (99) \end{aligned}$$

Thus, from (90) and (97) we have

$$\begin{aligned} J_i(\mathbf{T}, \mathbf{F}) = \; & tr\{(\mathbf{I}_{N_B} + \mathbf{F}^H\widehat{\mathbf{H}}_1^H\boldsymbol{\Upsilon}^{-1}\widehat{\mathbf{H}}_1\mathbf{F})^{-1}\} \\ & + tr\{(\mathbf{R}^{-1} + \mathbf{T}^H\widehat{\mathbf{H}}_{2,i}^H\mathbf{P}_i^{-1}\widehat{\mathbf{H}}_{2,i}\mathbf{T})^{-1}\}. \quad (100) \end{aligned}$$

Finally, by substituting (99) into (100), we prove (47).

REFERENCES

[1] L. Gopal, Y. Rong, and Z. Zang, "Simplified robust design for nonregenerative multicasting MIMO relay systems," in *Proc. 22nd Int. Conf. Commun.*, Sydney, Australia, Apr. 27–29, 2015, pp. 289–293.

[2] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge, U.K.: Cambridge Univ. Press, 2005.

[3] M. A. Khojastepour, A. Salehi-Golsefidi, and S. Rangarajan, "Towards an optimal beamforming algorithm for physical layer multicasting," in *Proc. IEEE Inf. Theory Workshop*, 2011, pp. 395–399.

[4] N. Jindal and Z.-Q. Luo, "Capacity limits of multiple antenna multicast," in *Proc. IEEE ISIT*, Seatle, WA, USA, Jul. 09–14, 2006, pp. 1841–1845.

[5] S. Y. Park, D. J. Love, and D. H. Kim, "Capacity limits of multi-antenna multicasting under correlated fading channels," *IEEE Trans. Commun.*, vol. 58, no. 7, pp. 2002–2013, Jul. 2010.

[6] N. D. Sidiropoulos, T. N. Davidson, and Z.-Q. Luo, "Transmit beamforming for physical-layer multicasting," *IEEE Trans. Signal Process.*, vol. 54, no. 6, pp. 2239–2251, Jun. 2006.

[7] E. Chiu and V. Lau, "Precoding design for multi-antenna multicast broadcast services with limited feedback," *IEEE Syst. J.*, vol. 4, no. 4, pp. 550–560, Dec. 2010.

[8] E. Matskani, N. D. Sidiropoulos, Z.-Q. Luo, and L. Tassiulas, "Efficient batch and adaptive approximation algorithms for joint multicast beamforming and admission control," *IEEE Trans. Signal Process.*, vol. 57, no. 12, pp. 4882–4894, Dec. 2009.

[9] M. A. Khojastepour, A. Khajehnejad, K. Sundaresan, and S. Rangarajan, "Adaptive beamforming algorithms for wireless link layer multicasting," in *Proc. 22nd Int. Symp. Personal, Indoor Mobile Radio Commun.*, Toronto, Canada, Sep. 11–14, 2011, pp. 1994–1998.

[10] M. Kaliszan, E. Pollakis, and S. Stańczak, "Efficient beamforming algorithms for MIMO multicast with application-layer coding," in *Proc. IEEE Int. Symp. Inf. Theory*, St. Petersburg, Russia, Jul. 31–Aug. 5, 2011, pp. 928–932.

[11] S. Shi, M. Schubert, and H. Boche, "Physical layer multicasting with linear MIMO transceivers," in *Proc. 42nd Annu. Conf. Inf. Sci. Syst.*, Princeton, NJ, USA, Mar. 2008, pp. 884–889.

[12] H. Zhang, J. Jin, X. You, G. Wu, and H. Wang, "Cooperative multi-antenna multicasting for wireless networks," in *Proc. IEEE GLOBECOM*, Miami, FL, USA, Dec. 6–10, 2010.

[13] M. R. A. Khandaker and Y. Rong, "Precoding design for MIMO relay multicasting," *IEEE Trans. Wireless Commun.*, vol. 12, no. 7, pp. 3544–3555, Jul. 2013.

[14] Y. Rong, "Robust design for linear non-regenerative MIMO relays with imperfect channel state information," *IEEE Trans. Signal Process.*, vol. 59, no. 5, pp. 2455–2460, May 2011.

[15] C. Xing, S. Ma, and Y.-C. Wu, "Robust joint design of linear relay precoder and destination equalizer for dual-hop amplify-and-forward MIMO relay systems," *IEEE Trans. Signal Process.*, vol. 58, no. 4, pp. 2273–2283, Apr. 2010.

[16] C. Xing, S. Ma, and Y. Zhou, "Matrix-monotonic optimization for MIMO systems," *IEEE Trans. Signal Process.*, vol. 63, no. 2, pp. 334–348, Jan. 2015.

[17] C. Xing, S. Ma, Z. Fei, Y. C. Wu, and H. V. Poor, "A general robust linear transceiver design for multi-hop amplify-and-forward MIMO relaying systems," *IEEE Trans. Signal Process.*, vol. 61, no. 5, pp. 1196–1209, Mar. 2013.

[18] C. Xing, S. Li, Z. Fei, and J. Kuang, "How to understand linear minimum mean-square-error transceiver design for multiple-input-multiple-output systems from quadratic matrix programming," *IET Commun.*, vol. 7, pp. 1231–1242, Aug. 2013.

[19] B. K. Chalise and L. Vandendorpe, "Joint linear processing for an amplify-and-forward MIMO relay channel with imperfect channel state information," *EURASIP J. Adv. Signal Process.*, vol. 2010, 2010, Art. no. 640186.

[20] F.-S. Tseng, M.-Y. Chang, and W.-R. Wu, "Joint Tomlinson-Harashima source and linear relay precoder design in amplify-and-forward MIMO relay systems via MMSE criterion," *IEEE Trans. Veh. Technol.*, vol. 60, no. 4, pp. 1687–1698, May 2011.

[21] 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, no. 1, pp. 353–363, Jan. 2015.

[22] J. Liu, F. Gao, and Z. Qiu, "Robust transceiver design for multi-user multiple-input multiple-output amplify-and-forward relay systems," *IET Commun.*, vol. 8, pp. 2162–2170, Mar. 2014.

[23] Y. Rong, "Simplified algorithms for optimizing multiuser multi-hop MIMO relay systems," *IEEE Trans. Commun.*, vol. 59, no. 10, pp. 2896–2904, Oct. 2011.

[24] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[25] L. Musavian, M. R. Nakhi, M. Dohler, and A. H. Aghvami, "Effect of channel uncertainty on the mutual information of MIMO fading channels," *IEEE Trans. Veh. Technol.*, vol. 56, no. 5, pp. 2798–2806, Sep. 2007.

[26] M. Ding and S. D. Blostein, "MIMO minimum total MSE transceiver design with imperfect CSI at both ends," *IEEE Trans. Signal Process.*, vol. 57, no. 3, pp. 1141–1150, Mar. 2009.

[27] A. Gupta and D. Nagar, *Matrix Variate Distributions*. London, U.K.: Chapman & Hall, 2000.

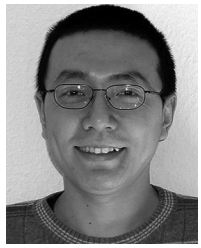
[28] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Englewood Cilffs, NJ, USA: Prentice-Hall, 1993.

[29] D. Bernstein, *Matrix Mathematics: Theory, Facts, and Formulas*. Princeton, NJ, USA: Princeton Univ. Press, 2011.

[30] M. Grant and S. Boyd, "The CVX Users Guide," Release 2.1, Oct. 2014. [Online]. Available: http://web.cvxr.com/cvx/doc/CVX.pdf

Lenin Gopal (M'06) received the B.Eng. degree in electrical and electronics engineering from Madurai Kamaraj University, Madurai, India, in 1996, the M.Eng. degree in telecommunications engineering from Multimedia University, Cyberjeya, Malaysia, in 2006, and the Ph.D. degree from Curtin University, Bentley, WA, Australia, in 2015.

He is currently a Senior Lecture with the Department of Electrical and Computer Engineering, Faculty of Engineering and Science, Curtin University, Miri, Malaysia. His research interests include signal processing for communications, wireless communications, and FPGA applications for wireless communications.



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 Postdoctoral Researcher with the Department of Electrical Engineering, University of California, Riverside, CA, USA, from February 2006 to November 2007. Since December 2007, he has been with the Department of Electrical and Computer Engineering, Curtin University, Bentley, WA, 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 more than 140 journal and conference papers in these areas.

Dr. Rong received 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, EUSIPCO, and ChinaCom.

Zhuquan Zang (M'93) received the B.Sc. degree from Shandong Normal University, Jinan, China, the M.Sc. degree from Shandong University, Jinan, China, and the Ph.D. degree in systems engineering from the Australian National University, Canberra, Australia, in 1993. From 1993 to 1994, he was with the University of Western Australia, Perth, Australia, as a Research Associate in the area of optimization, optimal control, and system identification. From 1994 to 2002, he was at the Australian Telecommunications Research Institute, Curtin University, Perth, Australia, first as a Research Fellow and then as a Senior Research Fellow. From 2002 to 2005, he was at the Western Australian Telecommunications Research Institute (A joint venture between Curtin University and the University of Western Australia), Perth, Australia. During this period, he was also affiliated with the Australian Telecommunications Cooperative Research Centre. Since early 2006, he has been with the Faculty of Engineering and Science, Curtin University, Malaysia. His main research interests include the areas of constrained filter set design for bandwidth-efficient wireless multiuser communications, wideband waveform design for radar and sonar, and computationally efficient optimization methods and their application to effective multiscale control scheme design, PAPR reduction in MIMO OFDM systems, and robust transceiver design for nonregenerative MIMO wireless communication systems.