Robust Transceiver Design for Multi-hop AF MIMO Relay Multicasting from Multiple Sources

Justin Lee Bing, *Student Member, IEEE*, Lenin Gopal, *Member, IEEE*, Yue Rong, *Senior Member, IEEE*
Choo W. R. Chiong, *Member, IEEE*, and Zhuquan Zang, *Member, IEEE*

***Abstract*—In this article, the transceiver design optimization problem is investigated for multi-hop multicasting amplify-and-forward (AF) multiple-input multiple-output (MIMO) relay systems, where multiple source nodes broadcast their message to multiple destination nodes via multiple serial relay nodes. Multiple antennas are installed at the sources, relays, and the destination nodes. In the transceiver design, we consider the mismatch between the true and the estimated channel state information (CSI), where the CSI mismatch follows the Gaussian-Kronecker model. A robust transceiver design algorithm is developed to jointly optimize the source, relay, and destination matrices to minimize the maximal weighted mean-squared error (WMSE) of the received message at all destination nodes. In particular, the WMSE is made statistically robust against the CSI mismatch by averaging through the distributions of the true CSI. Moreover, the WMSE decomposition is exploited to reduce the computational complexity of the transceiver optimization. Numerical simulations show a better performance of the proposed robust transceiver design against the channel mismatch.**

***Index Terms*—Robustness, amplify-and-forward, MIMO relay, multicasting, multi-hop, weighted mean-squared error.**

I. Introduction

With the recent booming of wireless applications such as streaming media and geographic information services, wireless multicasting systems have attracted great research interest as they enable common information to be concurrently transmitted to multiple recipients. This significantly enhances the bandwidth efficiency and reduces the power consumption at the source node side [1]. The wireless channel is undoubtedly the choice for multicasting applications due to its broadcasting nature. Wireless multicasting technology has been included in communication standards such as WiMAX (Worldwide Interoperability for Microwave Access) 802.16m and evolved multimedia broadcast/multicast service standards defined by the Third Generation Partnership Project for Long-Term Evolution (LTE)-advanced networks [2].

Generally, the design of the optimal beamforming vectors for multicasting is difficult due to its non-convex nature. In [3], the transmit beamforming vector for the multicasting problem has been investigated with the full channel state information (CSI) at the source node. In [3], two design criteria have been formulated based on the minimum transmission power and the max-min received power, respectively, and both optimization problems have been shown to be NP-hard. In [4], the channel capacity of a multi-antenna broadcasting system has been investigated. The influence of correlated fading channels on the capacity limits of multi-antenna broadcasting systems has been investigated in [5]. Two broadcasting schemes have been proposed in [6], namely the stochastic beamforming and the Alamouti-assisted rank-two beamforming, where the latter technique is a combination of source beamforming and the Alamouti space-time coding. Precoder design for multi-group multicasting with a common message has been addressed in [7].

The works in [3]-[7] focused on multicasting systems with single-antenna destination nodes, and these works have been extended to multi-antenna destination nodes to improve the system performance. A generalized block diagonalization technique has been developed in [8] to distinguish different multicasting groups. In [9], the designs of linear precoders based on the maximization of the achievable rate and the minimization of the weighted sum delay have been proposed for multicasting systems with multi-antenna transmitter and receivers.

While the works discussed in [3]-[9] considered single-hop multicasting systems, it becomes necessary to deploy relay nodes when the source-destination transmission distance is long or the wireless channel is strongly shadowed [10]. Relay nodes can efficiently mitigate the pathloss and shadowing effects of wireless links due to obstacles such as tall buildings and hills. In [11], a distributed beamforming algorithm has been proposed for multi-group multicasting relay networks to minimize the power consumption of the relay nodes, where each node has a single antenna. In [12], joint optimization of source and relay precoders has been investigated for a dual-hop multicasting multiple-input multiple-output (MIMO) relay system, where all nodes are installed with multiple antennas.

A more general multicasting relay network has been studied in [13] and [14]. In particular, transceiver design for dual-hop multicasting MIMO relay systems with multiple transmitters has been investigated in [13]. In [14], the transceiver design for the system proposed in [13] has been extended to multi-hop MIMO relay systems with multiple transmitters and receivers. Multi-hop relay systems are necessary when dual-hop multicasting relay systems are not able to provide a reliable link between transmitters and receivers. The transceiver designs in [12]-[14] were developed with knowledge of the full CSI.

Copyright (c) 2015 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending a request to pubs-permissions@ieee.org.

J. L. Bing, L. Gopal, C. W. R. Chiong, and Z. Zang are with the Department of Electrical and Computer Engineering, Curtin University, Miri, 98009, Sarawak, Malaysia (e-mails: justin.b.lee@student.curtin.edu.au; lenin@curtin.edu.my; raymond.ccw@curtin.edu.my; zqzang@curtin.edu.my).

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

However, the full CSI is unavailable in practical wireless systems, and there is always mismatch between the exact and the estimated CSI. Such CSI mismatch degrades the performance of the algorithms in [12]-[14], and thus, a more robust transceiver design is needed. The works in [12] and [13] have been extended in [15]-[17] to consider the mismatch between the exact and estimated CSI in the transceiver optimization. This improves the transceiver performance against the channel mismatch. But the transceiver design algorithms in [15]-[18] are for systems with only one source node and one relay node.

In [19], robust transceiver design for amplify-and-forward (AF) multi-hop MIMO relay networks has been developed based on various design criteria, and a general design algorithm has been proposed based on the majorization theory and properties of matrix-monotone functions. This work has been further extended in [20] and [21], where the covariance shaping constraints and per-antenna power constraints have been considered in the transceiver design of multi-hop AF MIMO relay networks. However, the transceiver design algorithms developed in [19]-[21] are for systems with only one source node and one destination node.

In this paper, we explicitly take into account the CSI mismatch to develop a robust transceiver design for a multi-hop AF MIMO relay multicasting system. We consider multiple sources multicast their information to a group of destination nodes through multiple hops, assuming that all nodes have multiple antennas. Different from [14], the impact of the channel estimation error is examined using the Gaussian-Kronecker model. Then the transceiver optimization problem is formulated to minimize the maximal weighted mean-squared error (WMSE) of the estimated signals at all destination nodes, where the WMSE is made statistically robust against the CSI mismatch by averaging through the distributions of the true CSI. Hence, various robust transceiver design problems such as the WMSE minimization and the system capacity maximization can be solved via transferring into a weighted minimum mean-squared error (WMMSE) problem.

The contributions and significance of this paper are listed below:

- The impact of channel estimation error on the performance of multi-hop multicasting AF MIMO relay systems is examined in this paper, under the Gaussian-Kronecker CSI mismatch model.
- The transceiver design proposed in this paper provides robustness against the unavoidable mismatch between the true and the estimated CSI in practical scenarios. The proposed algorithm improves the performance of the existing transceiver design [14].
- This paper extends the existing works in [15]-[18], which consider only a single source node and a single relay node, to systems with multiple source nodes and multiple serial relay nodes. Thus the system model investigated in this paper supports a longer communication range.
- This work extends [19]-[21] (with a single source-destination pair) to systems with multiple source and destination nodes. Thus, a more general system model is considered in this paper where multiple sources transmit signals simultaneously to multiple destinations by exploiting the broadcasting nature of the wireless channel.
- Compared with the existing works [15]-[21], multiple serial relay nodes are considered in this paper which significantly increases the challenges in solving the transceiver optimization problem, as a large number of mutually coupled matrix variables are involved in the optimization problem.
- The optimization problem for transceiver design with the general WMSE objective function in this paper is more challenging to solve than the MSE problems in [14]-[18], particularly when the weight matrix is not an identity matrix, as the subproblems cannot be easily converted to convex problems.

In summary, the main contribution of this paper to the communications society is to propose a robust transmitter design for a general multicasting AF MIMO relay system with multiple source nodes, multiple hops, and multiple destination nodes, under practical CSI mismatch scenarios, and using a general WMSE objective. This contribution is important to extend the range, increase the rate, and improve the reliability of multicasting systems.

To solve the transceiver optimization problem, the optimal structure of the multi-hop relay matrices is derived, which enables the objective function to be decomposed into the sum of L WMSE terms, where L is the number of hops. We show that under high signal-to-noise ratios (SNRs), the original problem can be decomposed into L subproblems. Then by exploiting upper-bounds of the objective functions, we show that the precoding matrices for the first $L-2$ relay nodes have a closed-form water-filling solution. For non-identity weight matrices, the subproblems of optimizing the source matrices and the precoding matrix at the last relay node are nonconvex, and we propose iterative algorithms to solve these challenging subproblems. It is shown through numerical simulations that the proposed robust transceiver design outperforms the algorithm developed in [14].

The rest of this paper is arranged as follows. In Section II, we present the system model of a multi-hop multicasting AF MIMO relay system with multiple sources and CSI mismatch. In Section III, the proposed robust transceiver design algorithm is derived. In Section IV, numerical simulations are shown to validate the performance of the proposed algorithm. The paper is concluded in Section V.

II. System Model

In the proposed system as shown in Fig. 1, a multicasting MIMO relay system with L hops ($L \geq 2$) is considered where K source nodes concurrently multicast their information to M destination nodes via $L-1$ AF relay nodes, where $N_{s,k}$ and $N_{r,l}$ are the number of antennas installed at the kth source node and the lth relay node, respectively. To simplify notations, we assume that each destination node is installed with N_D antennas. However, the algorithm proposed can be easily extended to multicasting systems where destination nodes have different number of antennas. We also assume that direct links do not exist between the source nodes and the destination nodes as well as those between any two non-consecutive relays due to severe pathloss/shadowing.

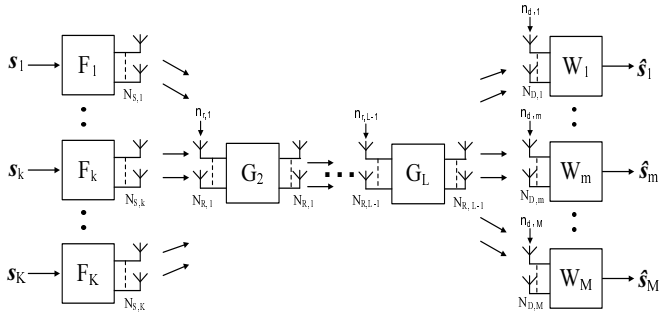


Fig. 1. A multi-hop AF MIMO relay multicasting system with multiple sources.

As the relay nodes are assumed to work in the half-duplex mode to avoid self-interference, the data communication between the source and destination nodes takes place over L time slots. At the first time slot, the kth source node precodes the modulated signal vector $\mathbf{s}_k \in \mathbb{C}^{N_{B,k}\times 1}$ by a linear precoding matrix $\mathbf{F}_k \in \mathbb{C}^{N_{S,k}\times N_{B,k}}$ and transmits the precoded vector $\mathbf{F}_k\mathbf{s}_k$ to the first relay node. The covariance matrix of $\mathbf{s}_k$ is given as $E\{\mathbf{s}_k\mathbf{s}_k^H\} = \mathbf{I}_{N_{B,k}}$ with $N_{B,k} \le N_{S,k}$, where $E\{\cdot\}$ represents statistical expectation with respect to the signal and noise, $\mathbf{I}_n$ represents an identity matrix of order n, and $(\cdot)^H$ denotes the matrix Hermitian transpose. We introduce $N_B = \sum_{k=1}^{K} N_{B,k}$ as the total number of signal streams from all K users, which satisfies $N_B \le \min(N_{R,1}, \cdots, N_{R,L-1}, N_D)$.

The received signal vector at the first relay node is expressed as

$$\begin{aligned}\mathbf{y}_1 &= \sum_{k=1}^{K} \mathbf{H}_{1,k}\mathbf{F}_k\mathbf{s}_k + \mathbf{n}_{r,1} \\ &\triangleq \mathbf{H}_1\mathbf{G}_1\mathbf{s} + \mathbf{n}_{r,1} \\ &\triangleq \mathbf{H}_1\mathbf{x}_1 + \mathbf{n}_{r,1}\end{aligned} \tag{1}$$

where $\mathbf{H}_{1,k} \in \mathbb{C}^{N_{R,1}\times N_{S,k}}$ is the MIMO channel matrix between the kth source node and the first relay node, $\mathbf{n}_{r,1} \in \mathbb{C}^{N_{R,1}\times 1}$ is the additive white Gaussian noise (AWGN) vector at the first relay node, $\mathbf{H}_1 = [\mathbf{H}_{1,1}, \mathbf{H}_{1,2}, \cdots, \mathbf{H}_{1,K}]$, $\mathbf{x}_1 = \mathbf{G}_1\mathbf{s}$, $\mathbf{G}_1 \triangleq \text{bd}(\mathbf{F}_1, \mathbf{F}_2, \cdots, \mathbf{F}_K)$, and $\mathbf{s} \triangleq [\mathbf{s}_1^T, \mathbf{s}_2^T, \cdots, \mathbf{s}_K^T]^T$. Here $(\cdot)^T$ denotes the transposition of a matrix and $\text{bd}(\cdot)$ represents block-diagonalization of matrices.

With the AF protocol, the transmitted signal vector from the lth relay node is expressed as

$$\mathbf{x}_{l+1} = \mathbf{G}_{l+1}\mathbf{y}_l, \quad l = 1, \cdots, L-1 \tag{2}$$

where $\mathbf{G}_{l+1} \in \mathbb{C}^{N_{R,l}\times N_{R,l}}$ is the linear precoding matrix at the lth relay node. Here $\mathbf{y}_l \in \mathbb{C}^{N_{R,l}\times 1}$ is the received signal vector at the lth relay node which is expressed as

$$\mathbf{y}_l = \mathbf{H}_l\mathbf{x}_l + \mathbf{n}_{r,l}, \quad l = 1, \cdots, L-1 \tag{3}$$

where $\mathbf{H}_l \in \mathbb{C}^{N_{R,l}\times N_{R,l-1}}$ is the lth hop MIMO channel matrix ($N_{R,0} = \sum_{i=1}^{K} N_{S,k}$), and $\mathbf{n}_{r,l} \in \mathbb{C}^{N_{R,l}\times 1}$ is the AWGN vector at the lth relay node. In the general scenario (i.e. $L \ge 3$), by applying (1) and (2), the received signal vector (3) at the lth relay node can be equivalently written as

$$\begin{aligned}\mathbf{y}_l = &\prod_{i=l}^{1} (\mathbf{H}_i\mathbf{G}_i)\,\mathbf{s} + \sum_{j=2}^{l}\left(\prod_{i=l}^{j} (\mathbf{H}_i\mathbf{G}_i)\,\mathbf{n}_{r,j-1}\right) + \mathbf{n}_{r,l} \\ &l = 2, \cdots, L-1\end{aligned} \tag{4}$$

where $\prod_{i=l}^{k}\mathbf{X}_i$ represents multiplication over a series of $\mathbf{X}_i$ from the lth to the kth terms.

The received signal vector $\mathbf{x}_L$ at the last relay node is multicast to all M destination nodes. Based on (4), the received signal vector at the mth destination node can be expressed as

$$\begin{aligned}\mathbf{y}_{d,m} = &\mathbf{H}_{L,m}\mathbf{G}_L \prod_{l=L-1}^{1} (\mathbf{H}_l\mathbf{G}_l)\,\mathbf{s} + \mathbf{H}_{L,m}\mathbf{G}_L \\ &\times\left(\sum_{j=2}^{L-1}\left(\prod_{l=L-1}^{j} (\mathbf{H}_l\mathbf{G}_l)\,\mathbf{n}_{r,j-1}\right) + \mathbf{n}_{r,L-1}\right) + \mathbf{n}_{d,m} \\ &\triangleq \bar{\mathbf{A}}_m\mathbf{s} + \mathbf{v}_m, \qquad m = 1, \cdots, M\end{aligned} \tag{5}$$

where $\mathbf{H}_{L,m} \in \mathbb{C}^{N_D\times N_{R,L-1}}$ is the MIMO channel matrix between the last relay node and the mth destination node, $\mathbf{n}_{d,m} \in \mathbb{C}^{N_D\times 1}$ is the AWGN vector at the mth destination node. For $m = 1, \cdots, M$, the equivalent MIMO channel matrix between all source nodes and the mth destination node can be written as

$$\bar{\mathbf{A}}_m = \mathbf{H}_{L,m}\mathbf{G}_L \prod_{l=L-1}^{1} (\mathbf{H}_l\mathbf{G}_l)$$

and the equivalent noise vector at the mth destination node is given by

$$\mathbf{v}_m = \mathbf{H}_{L,m}\mathbf{G}_L\left(\sum_{j=2}^{L-1}\left(\prod_{l=L-1}^{j}(\mathbf{H}_l\mathbf{G}_l)\mathbf{n}_{r,j-1}\right) + \mathbf{n}_{r,L-1}\right) + \mathbf{n}_{d,m}.$$

We would like to note that although the system model (5) is developed for a narrow band system, for orthogonal frequency-division multiplexing (OFDM) based broadband multi-hop AF multicasting MIMO relay systems, the algorithm in this paper can be applied to each subcarrier of the OFDM system.

Usually, the full CSI is needed for the optimal precoder matrices design. However, the exact CSI is unavailable in practice due to channel estimation errors which resulting in mismatch between the true and estimated CSI. Thus, the true channel matrices $\mathbf{H}_{1,k}$, $\mathbf{H}_l$, and $\mathbf{H}_{L,m}$ can be modelled as [18], [19], [23], [24]

$$\mathbf{H}_{1,k} = \widehat{\mathbf{H}}_{1,k} + \mathbf{\Delta}_{1,k}, \qquad k = 1, \cdots, K \tag{6}$$

$$\mathbf{H}_l = \widehat{\mathbf{H}}_l + \mathbf{\Delta}_l, \qquad l = 2, \cdots, L-1, L \ge 3 \tag{7}$$

$$\mathbf{H}_{L,m} = \widehat{\mathbf{H}}_{L,m} + \mathbf{\Delta}_{L,m}, \; m = 1, \cdots, M \tag{8}$$

where $\widehat{\mathbf{H}}_{1,k}$, $\widehat{\mathbf{H}}_l$, and $\widehat{\mathbf{H}}_{L,m}$ are the estimated channel matrices between the kth source node and the first relay node, the $(l-1)$th relay node and the lth relay node, and the last relay node and the mth destination node, respectively, $\mathbf{\Delta}_{1,k}$, $\mathbf{\Delta}_l$, and $\mathbf{\Delta}_{L,m}$ are the corresponding CSI mismatch matrices. We

assume that these CSI mismatch matrices satisfy the Gaussian-Kronecker model [18], [19], [23], [24]

$$\mathbf{\Delta}_{1,k} \sim \mathcal{CN}(\mathbf{0}, \mathbf{\Sigma}_{1,k} \otimes \mathbf{\Psi}_{1,k}^T) \tag{9}$$

$$\mathbf{\Delta}_{l} \sim \mathcal{CN}(\mathbf{0}, \mathbf{\Sigma}_{l} \otimes \mathbf{\Psi}_{l}^T) \tag{10}$$

$$\mathbf{\Delta}_{L,m} \sim \mathcal{CN}(\mathbf{0}, \mathbf{\Sigma}_{L,m} \otimes \mathbf{\Psi}_{L,m}^T) \tag{11}$$

where $\otimes$ denotes the Kronecker product, $\mathbf{\Sigma}_{1,k}$, $\mathbf{\Sigma}_{l}$, and $\mathbf{\Sigma}_{L,m}$ are the row covariance matrices for $\mathbf{\Delta}_{1,k}$, $\mathbf{\Delta}_{l}$, and $\mathbf{\Delta}_{L,m}$ respectively, and $\mathbf{\Psi}_{1,k}$, $\mathbf{\Psi}_{l}$, and $\mathbf{\Psi}_{L,m}$ are the column covariance matrices for $\mathbf{\Delta}_{1,k}$, $\mathbf{\Delta}_{l}$, and $\mathbf{\Delta}_{L,m}$ respectively. These covariance matrices model the correlation among antennas at each node [18].

Based on (6)-(11), the true channel matrices $\mathbf{H}_{1,k}$, $\mathbf{H}_{l}$, and $\mathbf{H}_{L,m}$ are rewritten into [18]

$$\mathbf{H}_{1,k} = \widehat{\mathbf{H}}_{1,k} + \mathbf{\Sigma}_{1,k}^{\frac{1}{2}} \mathbf{H}_{\omega_{1,k}} \mathbf{\Psi}_{1,k}^{\frac{1}{2}}, \qquad k = 1, \cdots, K \tag{12}$$

$$\mathbf{H}_{l} = \widehat{\mathbf{H}}_{l} + \mathbf{\Sigma}_{l}^{\frac{1}{2}} \mathbf{H}_{\omega_{l}} \mathbf{\Psi}_{l}^{\frac{1}{2}}, \qquad l = 2, \cdots, L-1 \tag{13}$$

$$\mathbf{H}_{L,m} = \widehat{\mathbf{H}}_{L,m} + \mathbf{\Sigma}_{L,m}^{\frac{1}{2}} \mathbf{H}_{\omega_{L,m}} \mathbf{\Psi}_{L,m}^{\frac{1}{2}}, \ m = 1, \cdots, M \tag{14}$$

where the elements of $\mathbf{H}_{\omega_{1,k}}$, $\mathbf{H}_{\omega_{l}}$, and $\mathbf{H}_{\omega_{L,m}}$ are drawn from independent and identically distributed (i.i.d.) complex Gaussian distribution with zero mean and unit variance.

Lemma 1: [18], [25] For $\mathbf{H} \sim \mathcal{CN}(\widehat{\mathbf{H}}, \mathbf{\Sigma} \otimes \mathbf{\Psi}^T)$ and any constant matrix $\mathbf{A}$, there is

$$E\{\mathbf{HAH}^H\} = \widehat{\mathbf{H}}\mathbf{A}\widehat{\mathbf{H}}^H + tr\{\mathbf{A}\mathbf{\Psi}\}\mathbf{\Sigma} \tag{15}$$

where $tr\{\cdot\}$ stands for matrix trace.

Due to simplicity, we assume that a linear receiver with a weight matrix $\mathbf{W}_m$ is used at the mth destination node to recover the transmitted signal waveform. Thus, the estimated signal vector at the mth destination node is expressed as

$$\widehat{\mathbf{s}}_m = \mathbf{W}_m \mathbf{y}_{d,m}, \quad m = 1, \cdots, M. \tag{16}$$

It is shown in [19] that various performance metrics for transceiver design such as weighted MSE and system capacity are functions of the MSE matrix of the estimated signal waveform at the destination nodes. Moreover, considering that capacity maximization can be solved via transferring into a WMMSE problem [26], we introduce a Hermitian weight matrix $\mathbf{O}$. Similar to [27], the weight matrix $\mathbf{O}$ is used to convert the capacity maximization problem to the weighted MSE minimization problem. When solving the capacity maximization problem, the weight matrix $\mathbf{O}$ can be updated iteratively following [26]. Thus, by applying (5) and (16), the WMSE of the estimated signal waveform at the destination nodes can be expressed as

$$\begin{aligned} E_m &= E\{tr\{\mathbf{O}(\widehat{\mathbf{s}}_m - \mathbf{s})(\widehat{\mathbf{s}}_m - \mathbf{s})^H\}\} \\ &= tr\big\{\mathbf{O}(\mathbf{W}_m E\{\mathbf{y}_{d,m}\mathbf{y}_{d,m}^H\}\mathbf{W}_m^H - \mathbf{W}_m\bar{\mathbf{A}}_m \\ &\quad -\bar{\mathbf{A}}_m^H\mathbf{W}_m^H + \mathbf{I}_{N_B})\big\}, \quad m = 1, \cdots, M. \end{aligned} \tag{17}$$

The transmission power needed by the kth source node is $tr\{\mathbf{F}_k\mathbf{F}_k^H\}$, $k = 1, \cdots, K$. From (2), the transmission power required by the lth relay node can be derived as

$$\begin{aligned} P(\mathbf{G}_{l+1}) &= tr\big\{E\{\mathbf{x}_{l+1}\mathbf{x}_{l+1}^H\}\big\} \\ &= tr\{\mathbf{G}_{l+1}E\{\mathbf{y}_l\mathbf{y}_l^H\}\mathbf{G}_{l+1}^H\}, \ l = 1, \cdots, L-1. \end{aligned} \tag{18}$$

III. Proposed Robust Transceiver Design Algorithm

From (17), we can see that it is difficult to design $\mathbf{W}_m$ to optimize E_m without the perfect $\mathbf{H}_{1,k}$, $\mathbf{H}_{l}$, and $\mathbf{H}_{L,m}$. If designing $\mathbf{W}_m$, $\mathbf{G}_l$, and $\mathbf{F}_k$ based on the estimated channel matrices, the system performance may degrade significantly due to the channel estimation errors $\mathbf{\Delta}_{1,k}$, $\mathbf{\Delta}_{l}$, and $\mathbf{\Delta}_{L,m}$. Thus, instead of optimizing E_m, we design the transceiver precoding matrices which minimize J_m as

$$J_m \triangleq E_H\{E_m\} \tag{19}$$

where $E_H\{\cdot\}$ denotes the statistical expectation with respect to MIMO channel matrices with distributions in (6)-(11). Using *Lemma 1*, the MSE function in (19) can be written as

$$\begin{aligned} J_m = tr\{&\mathbf{O}(\mathbf{I}_{N_B} + \mathbf{W}_m\mathbf{R}_{\mathbf{y}_{d,m}}\mathbf{W}_m^H - \mathbf{W}_m\widehat{\mathbf{A}}_m \\ &-\widehat{\mathbf{A}}_m^H\mathbf{W}_m^H)\}, \quad m = 1, \cdots, M. \end{aligned} \tag{20}$$

The detailed derivations of (20) can be found in Appendix A. Here $\widehat{\mathbf{A}}_m = \widehat{\mathbf{H}}_{L,m}\mathbf{G}_L\prod_{l=L-1}^{1}(\widehat{\mathbf{H}}_l\mathbf{G}_l)$, $m = 1, \cdots, M$, is the estimated MIMO channel matrix from all source nodes to the mth destination node and $\mathbf{R}_{\mathbf{y}_{d,m}} = E_H\{E\{\mathbf{y}_{d,m}\mathbf{y}_{d,m}^H\}\}$, $m = 1, \cdots, M$, is the expectation of the covariance matrix of $\mathbf{y}_{d,m}$ with respect to the distribution in (12)-(14) as

$$\begin{aligned} \mathbf{R}_{\mathbf{y}_{d,m}} = &\widehat{\mathbf{H}}_{L,m}\mathbf{G}_L\mathbf{R}_{\mathbf{y}_{L-1}}\mathbf{G}_L^H\widehat{\mathbf{H}}_{L,m}^H \\ &+tr\{\mathbf{G}_L\mathbf{R}_{\mathbf{y}_{L-1}}\mathbf{G}_L^H\mathbf{\Psi}_{L,m}\}\mathbf{\Sigma}_{L,m} + \mathbf{I}_{N_D} \end{aligned}$$

where $\mathbf{R}_{\mathbf{y}_l} = E_H\{E\{\mathbf{y}_l\mathbf{y}_l^H\}\}$ is the expectation of the covariance matrix of $\mathbf{y}_l$ with respect to the distribution in (13) and is calculated recursively by

$$\begin{aligned} \mathbf{R}_{\mathbf{y}_l} = &\widehat{\mathbf{H}}_l\mathbf{G}_l\mathbf{R}_{\mathbf{y}_{l-1}}\mathbf{G}_l^H\widehat{\mathbf{H}}_l^H + tr\{\mathbf{G}_l\mathbf{R}_{\mathbf{y}_{l-1}}\mathbf{G}_l^H\mathbf{\Psi}_l\}\mathbf{\Sigma}_l \\ &+\mathbf{I}_{N_{R,l}}, \qquad l = 2, \cdots, L-1 \end{aligned}$$

and for $l = 1$

$$\mathbf{R}_{\mathbf{y}_1} = \widehat{\mathbf{H}}_1\mathbf{G}_1\mathbf{G}_1^H\widehat{\mathbf{H}}_1^H + \sum_{k=1}^{K} tr\{\mathbf{F}_k\mathbf{F}_k^H\mathbf{\Psi}_{1,k}\}\mathbf{\Sigma}_{1,k} + \mathbf{I}_{N_{R,1}}.$$

As the exact MIMO channel matrices are not available, the power constraints at the lth relay node (18) can be expressed as

$$E_H\{P(\mathbf{G}_{l+1})\} = tr\{\mathbf{G}_{l+1}\mathbf{R}_{\mathbf{y}_l}\mathbf{G}_{l+1}^H\}, \quad l = 1, \cdots, L-1. \tag{21}$$

The objective of the proposed robust transceiver design is to minimize the maximum of the MSE in (20) across all destination nodes subjecting to the power constraints at the source and relay nodes. Based on (20) and (21), the optimization problem can be written as

$$\min_{\{\mathbf{F}_k\},\{\mathbf{G}_l\},\{\mathbf{W}_m\}} \ \max_{m} \ J_m \tag{22a}$$

$$s.t. \ tr\{\mathbf{G}_l\mathbf{R}_{\mathbf{y}_{l-1}}\mathbf{G}_l^H\} \le P_{r,l}, \ l = 2, \cdots, L \tag{22b}$$

$$tr\{\mathbf{F}_k\mathbf{F}_k^H\} \le P_{s,k}, \qquad k = 1, \cdots, K \tag{22c}$$

where $\{\mathbf{F}_k\} \triangleq \{\mathbf{F}_k, k = 1, \cdots, K\}$, $\{\mathbf{G}_l\} \triangleq \{\mathbf{G}_l, l = 2, \cdots, L\}$, and $\{\mathbf{W}_m\} \triangleq \{\mathbf{W}_m, m = 1, \cdots, M\}$, (22b) is the power constraint at the $(l-1)$th relay node with a power limit of $P_{r,l} > 0$ and (22c) is the transmission power constraint

at the kth source node with the power budget $P_{s,k} > 0$. Due to the complicated objective function (22a), it is very challenging to determine the solution to the min-max problem (22) with a reasonable computational complexity[1]. Hence, a low computational complexity approach is developed in this paper to solve the problem (22).

A. Optimization of $\mathbf{W}_m$ and $\mathbf{G}_l$

For any given $\{\mathbf{F}_k\}$ and $\{\mathbf{G}_l\}$, the optimal $\mathbf{W}_m$ which minimizes J_m in (20) is the linear minimum MSE (LMMSE) receiver expressed as [28]

$$\mathbf{W}_m = \widehat{\mathbf{A}}_m^H \mathbf{R}_{\mathbf{y}_{d,m}}^{-1}, \quad m = 1, \cdots, M \tag{23}$$

where $(\cdot)^{-1}$ represents the matrix inversion. By substituting (23) to (20), the WMSE function can be rewritten as

$$J_m = tr\big\{\mathbf{O}\big(\mathbf{I}_{N_B} - \widehat{\mathbf{A}}_m^H \mathbf{R}_{\mathbf{y}_{d,m}}^{-1} \widehat{\mathbf{A}}_m\big)\big\}. \tag{24}$$

With the input-output relationship (4), it can be shown that the optimal $\mathbf{G}_l$ has the optimal structure of [26], [29]

$$\mathbf{G}_l = \mathbf{T}_l \mathbf{D}_l, \qquad l = 2, \cdots, L \tag{25}$$

where $\mathbf{T}_l \in \mathbb{C}^{N_{R,l-1} \times N_B}$ is a linear filter which is viewed as the transmitting precoder matrix for the effective lth hop MIMO channel, and $\mathbf{D}_l = \mathbf{O}^{\frac{H}{2}} \mathbf{K}_{l-1}^H \mathbf{R}_{\mathbf{y}_{l-1}}^{-1} \in \mathbb{C}^{N_B \times N_{R,l-1}}$ is the LMMSE receiver weight matrix for the $(l-1)$th relay node received signal. Here $\mathbf{K}_l = \prod_{i=l}^{1} (\widehat{\mathbf{H}}_i \mathbf{G}_i)$, $l = 1, \cdots, L-1$. Using (25), (24) is equivalently expressed as the sum of L MSE functions as

$$\begin{aligned} J_m = & tr\big\{\mathbf{O}(\mathbf{I}_{N_B} + \mathbf{G}_1^H \widehat{\mathbf{H}}_1^H \mathbf{\Xi}_1^{-1} \widehat{\mathbf{H}}_1 \mathbf{G}_1)^{-1}\big\} \\ & + \sum_{l=2}^{L-1} tr\big\{(\mathbf{Z}_l^{-1} + \mathbf{T}_l^H \widehat{\mathbf{H}}_l^H \mathbf{\Xi}_l^{-1} \widehat{\mathbf{H}}_l \mathbf{T}_l)^{-1}\big\} \\ & + tr\big\{(\mathbf{Z}_L^{-1} + \mathbf{T}_L^H \widehat{\mathbf{H}}_{L,m}^H \mathbf{\Xi}_{L,m}^{-1} \widehat{\mathbf{H}}_{L,m} \mathbf{T}_L)^{-1}\big\} \end{aligned} \tag{26}$$

where

$$\mathbf{\Xi}_1 = \sum_{k=1}^{K} tr\{\mathbf{F}_k \mathbf{F}_k^H \mathbf{\Psi}_{1,k}\} \mathbf{\Sigma}_{1,k} + \mathbf{I}_{N_{R,1}} \tag{27}$$

$$\mathbf{Z}_l = \mathbf{O}^{\frac{H}{2}} \mathbf{K}_{l-1}^H \mathbf{R}_{\mathbf{y}_{l-1}}^{-1} \mathbf{K}_{l-1} \mathbf{O}^{\frac{1}{2}}, \qquad l = 2, \cdots, L \tag{28}$$

$$\mathbf{\Xi}_l = tr\{\mathbf{T}_l \mathbf{Z}_l \mathbf{T}_l^H \mathbf{\Psi}_l\} \mathbf{\Sigma}_l + \mathbf{I}_{N_{R,l}}, \quad l = 2, \cdots, L-1 \tag{29}$$

$$\mathbf{\Xi}_{L,m} = tr\{\mathbf{T}_L \mathbf{Z}_L \mathbf{T}_L^H \mathbf{\Psi}_{L,m}\} \mathbf{\Sigma}_{L,m} + \mathbf{I}_{N_D}, m = 1, \cdots, M \tag{30}$$

When the perfect CSI is available, we have $\mathbf{\Xi}_l = \mathbf{I}_{N_{R,l}}$, $l = 1, \cdots, L-1$, and $\mathbf{\Xi}_{L,m} = \mathbf{I}_{N_D}$, $m = 1, \cdots, M$. Then (26) can be viewed as the MSE decomposition based on multi-hop multicasting MIMO relay systems without channel estimation error, i.e., $\widehat{\mathbf{H}}_{1,k} = \mathbf{H}_{1,k}$, $\widehat{\mathbf{H}}_l = \mathbf{H}_l$, and $\widehat{\mathbf{H}}_{L,m} = \mathbf{H}_{L,m}$. Moreover, it can be noticed that the first term in (26) is actually the WMSE of the first-hop signal waveform estimation with the LMMSE receiver $\mathbf{\Gamma} = \mathbf{G}_1^H \widehat{\mathbf{H}}_1^H \mathbf{R}_{\mathbf{y}_1}^{-1}$ at the first relay node which is expressed as

$$\begin{aligned} J_s &= E_H\left\{E\{tr\{\mathbf{O}\big((\mathbf{\Gamma}\mathbf{y}_1 - \mathbf{s})(\mathbf{\Gamma}\mathbf{y}_1 - \mathbf{s})^H\big)\}\}\right\} \\ &= tr\{\mathbf{O}\big(\mathbf{\Gamma}\mathbf{R}_{\mathbf{y}_1}\mathbf{\Gamma}^H - \mathbf{\Gamma}\widehat{\mathbf{H}}_1\mathbf{G}_1 - \mathbf{G}_1^H\widehat{\mathbf{H}}_1^H\mathbf{\Gamma}^H + \mathbf{I}_{N_B}\big)\} \\ &= tr\big\{\mathbf{O}\big(\mathbf{I}_{N_B} + \mathbf{G}_1^H\widehat{\mathbf{H}}_1^H\mathbf{\Xi}_1^{-1}\widehat{\mathbf{H}}_1\mathbf{G}_1\big)^{-1}\big\}. \end{aligned}$$

The subsequent trace terms in (26) can be considered as increments of the MSE caused by the consecutive hops.

From (25) and (28), the transmission power required by the $(l-1)$th relay node is expressed as

$$tr\{\mathbf{G}_l \mathbf{R}_{\mathbf{y}_{l-1}} \mathbf{G}_l^H\} = tr\{\mathbf{T}_l \mathbf{Z}_l \mathbf{T}_l^H\}, \quad l = 2, \cdots, L. \tag{31}$$

Substituting (31) into (22), the optimization problem (22) can be rewritten as

$$\min_{\{\mathbf{F}_k\},\{\mathbf{T}_l\}} \max_m \; J_m \tag{32a}$$

$$s.t.\; tr\{\mathbf{T}_l \mathbf{Z}_l \mathbf{T}_l^H\} \le P_{r,l}, \qquad l = 2, \cdots, L \tag{32b}$$

$$tr\{\mathbf{F}_k \mathbf{F}_k^H\} \le P_{s,k}, \qquad k = 1, \cdots, K \tag{32c}$$

where $\{\mathbf{T}_l\} \triangleq \{\mathbf{T}_l, l = 2, \cdots, L\}$. Based on the matrix inversion lemma [30], (28) can be rewritten as

$$\begin{aligned} \mathbf{Z}_l &= \mathbf{O}^{\frac{H}{2}} \mathbf{K}_{l-1}^H (\mathbf{K}_{l-1}\mathbf{K}_{l-1}^H + \mathbf{\Omega}_{l-1})^{-1} \mathbf{K}_{l-1} \mathbf{O}^{\frac{1}{2}} \\ &= \mathbf{O}^{\frac{H}{2}} \mathbf{K}_{l-1}^H \mathbf{\Omega}_{l-1}^{-1} \mathbf{K}_{l-1} (\mathbf{K}_{l-1}^H \mathbf{\Omega}_{l-1}^{-1} \mathbf{K}_{l-1} + \mathbf{I}_{N_B})^{-1} \mathbf{O}^{\frac{1}{2}} \\ & \qquad l = 2, \cdots, L \end{aligned}$$

where

$$\mathbf{\Omega}_1 = \mathbf{\Xi}_1$$

$$\mathbf{\Omega}_l = \sum_{j=2}^{l} \prod_{i=l}^{j} (\widehat{\mathbf{H}}_i \mathbf{G}_i) \mathbf{\Xi}_{j-1} \prod_{i=j}^{l} (\mathbf{G}_i^H \widehat{\mathbf{H}}_i^H) + \mathbf{\Xi}_l, l = 2, \cdots, L.$$

It is noted that $\mathbf{Z}_l$ can be approximated as $\mathbf{O}$ with high SNRs, i.e., $\mathbf{K}_{l-1}^H \mathbf{\Omega}_{l-1}^{-1} \mathbf{K}_{l-1} \gg \mathbf{I}_{N_B}$. Hence, the problem (32) is expressed into

$$\begin{aligned} \min_{\{\mathbf{F}_k\},\{\mathbf{T}_l\}} \max_m \; & tr\big\{\mathbf{O}(\mathbf{I}_{N_B} + \mathbf{G}_1^H \widehat{\mathbf{H}}_1^H \mathbf{\Xi}_1^{-1} \widehat{\mathbf{H}}_1 \mathbf{G}_1)^{-1}\big\} \\ & + \sum_{l=2}^{L-1} tr\big\{(\mathbf{O}^{-1} + \mathbf{T}_l^H \widehat{\mathbf{H}}_l^H \mathbf{\Upsilon}_l^{-1} \widehat{\mathbf{H}}_l \mathbf{T}_l)^{-1}\big\} \\ & + tr\big\{(\mathbf{O}^{-1} + \mathbf{T}_L^H \widehat{\mathbf{H}}_{L,m}^H \mathbf{\Upsilon}_{L,m}^{-1} \widehat{\mathbf{H}}_{L,m} \mathbf{T}_L)^{-1}\big\} \end{aligned} \tag{33a}$$

$$s.t.\; tr\{\mathbf{T}_l \mathbf{O} \mathbf{T}_l^H\} \le P_{r,l}, \qquad l = 2, \cdots, L \tag{33b}$$

$$tr\{\mathbf{F}_k \mathbf{F}_k^H\} \le P_{s,k}, \qquad k = 1, \cdots, K \tag{33c}$$

where $\{\mathbf{\Upsilon}_l, l = 2, \cdots, L-1\}$ and $\{\mathbf{\Upsilon}_{L,m}, m = 1, \cdots, M\}$ are given by

$$\mathbf{\Upsilon}_l = tr\{\mathbf{T}_l \mathbf{O} \mathbf{T}_l^H \mathbf{\Psi}_l\} \mathbf{\Sigma}_l + \mathbf{I}_{N_{R,l}} \tag{34}$$

$$\mathbf{\Upsilon}_{L,m} = tr\{\mathbf{T}_L \mathbf{O} \mathbf{T}_L^H \mathbf{\Psi}_{L,m}\} \mathbf{\Sigma}_{L,m} + \mathbf{I}_{N_D}. \tag{35}$$

When the exact CSI is available, there are $\mathbf{\Xi}_1 = \mathbf{I}_{N_{R,1}}$, $\mathbf{\Upsilon}_l = \mathbf{I}_{N_{R,l}}$, $l = 2, \cdots, L-1$, and $\mathbf{\Upsilon}_{L,m} = \mathbf{I}_{N_D}$, $m = 1, \cdots, M$, and the problem (33) can be viewed as a non-robust transceiver design problem. In other words, the proposed robust transceiver design extends the non-robust transceiver design algorithm to the general practical multicasting MIMO system with CSI mismatch and the WMSE objective function.

[1]For systems with perfect CSI, the problem (22) can be solved by an iterative algorithm similar to [14] based on the conditional quadratic structure of J_m [22] with respect to the variables $\{\mathbf{F}_k\}, \{\mathbf{G}_l\}, \{\mathbf{W}_m\}$. For practical systems with CSI mismatch, with some approximations, the problem (22) is able to be solved by the iterative approach in [14] with a high computational complexity. It can be shown similar to [14] that the MSE gap between the simplified algorithm and the iterative algorithm reduces with the increase of the system SNR.

It can be observed that each trace term in (33a) can be optimized independently. Thus, the problem (33) can be decomposed into L subproblems. In particular, the problem of optimizing the source precoding matrices can be written as

$$\min_{\{\mathbf{F}_k\}} \ tr\big\{\mathbf{O}(\mathbf{I}_{N_B}+\mathbf{G}_1^H\widehat{\mathbf{H}}_1^H\mathbf{\Xi}_1^{-1}\widehat{\mathbf{H}}_1\mathbf{G}_1)^{-1}\big\} \tag{36a}$$

$$s.t.\ tr\{\mathbf{F}_k\mathbf{F}_k^H\}\le P_{s,k},\qquad k=1,\cdots,K. \tag{36b}$$

For $L\ge 3$, the problem of optimizing the $(l-1)$th relay node precoding matrix, $l=2,\cdots,L-1$, is given by

$$\min_{\mathbf{T}_l} \ tr\big\{(\mathbf{O}^{-1}+\mathbf{T}_l^H\widehat{\mathbf{H}}_l^H\mathbf{\Upsilon}_l^{-1}\widehat{\mathbf{H}}_l\mathbf{T}_l)^{-1}\big\} \tag{37a}$$

$$s.t.\ tr\{\mathbf{T}_l\mathbf{O}\mathbf{T}_l^H\}\le P_{r,l}. \tag{37b}$$

Finally, the precoding matrix at the last relay node is optimized by solving

$$\min_{\tilde{\mathbf{T}}_L}\ \max_m \ tr\big\{\mathbf{O}(\mathbf{I}_{N_B}+\tilde{\mathbf{T}}_L^H\widehat{\mathbf{H}}_{L,m}^H\mathbf{\Upsilon}_{L,m}^{-1}\widehat{\mathbf{H}}_{L,m}\tilde{\mathbf{T}}_L)^{-1}\big\} \tag{38a}$$

$$s.t.\ tr\{\tilde{\mathbf{T}}_L\tilde{\mathbf{T}}_L^H\}\le P_{r,L} \tag{38b}$$

where $\tilde{\mathbf{T}}_L=\mathbf{T}_L\mathbf{O}^{\frac{1}{2}}$.

B. Optimization of $\{\mathbf{F}_k\}$

When $\mathbf{O}=\mathbf{I}_{N_B}$, the problem (36) can be modified to a convex semidefinite programming (SDP) problem. However in general cases ($\mathbf{O}\neq\mathbf{I}_{N_B}$), the problem (36) is difficult to be formulated into a convex optimization problem. Thus, we show in Appendix B that (36a) can be reformulated as follows

$$\begin{aligned}&tr\big\{\mathbf{O}(\mathbf{I}_{N_B}+\mathbf{G}_1^H\widehat{\mathbf{H}}_1^H\mathbf{\Xi}_1^{-1}\widehat{\mathbf{H}}_1\mathbf{G}_1)^{-1}\big\}\\&=\min_{\mathbf{L}} E_H\{tr(\mathbf{O}E\{(\mathbf{L}^H\mathbf{y}_1-\mathbf{s})(\mathbf{L}^H\mathbf{y}_1-\mathbf{s})^H\})\}\end{aligned} \tag{39}$$

where $\mathbf{L}$ is the weight matrix of a linear receiver for the MIMO system in (1). By utilizing (39), the problem (36) can be solved through the following problem

$$\begin{aligned}\min_{\{\mathbf{F}_k\},\mathbf{L}}\ tr\Big\{&(\mathbf{O}^{\frac{H}{2}}\mathbf{L}^H\widehat{\mathbf{H}}_1\mathbf{G}_1-\mathbf{O}^{\frac{H}{2}})(\mathbf{O}^{\frac{H}{2}}\mathbf{L}^H\widehat{\mathbf{H}}_1\mathbf{G}_1-\mathbf{O}^{\frac{H}{2}})^H\\&+\mathbf{O}^{\frac{H}{2}}\mathbf{L}^H\mathbf{\Xi}_1\mathbf{L}\mathbf{O}^{\frac{1}{2}}\Big\}\end{aligned} \tag{40a}$$

$$s.t.\ tr\{\mathbf{F}_k\mathbf{F}_k^H\}\le P_{s,k},\qquad k=1,\cdots,K. \tag{40b}$$

In general cases of $\mathbf{\Psi}_{1,k}\neq\sigma_e^2\mathbf{I}_{N_{S,k}}$, the problem (40a) is difficult to solve as a result of $\mathbf{\Xi}_1$ being a function of $\{\mathbf{F}_k\}$. Thus, the following inequality [31] is applied to overcome the challenge

$$tr\left\{\mathbf{A}\mathbf{B}\right\}\le tr\left\{\mathbf{A}\right\}\lambda_m\left(\mathbf{B}\right) \tag{41}$$

where $\lambda_m(\cdot)$ is defined as the maximal eigenvalue of a matrix. By applying the inequality (41), an upper-bound of $\mathbf{\Xi}_1$ is given by

$$\mathbf{\Xi}_1\le\sum_{k=1}^{K}P_{s,k}\lambda_m(\mathbf{\Psi}_{1,k})\mathbf{\Sigma}_{1,k}+\mathbf{I}_{N_{R,1}}\triangleq\mathbf{\Phi}_1. \tag{42}$$

Using (42), the problem (40) can be written as

$$\begin{aligned}\min_{\{\mathbf{F}_k\},\mathbf{L}}\ tr\Big\{&(\mathbf{O}^{\frac{H}{2}}\mathbf{L}^H\widehat{\mathbf{H}}_1\mathbf{G}_1-\mathbf{O}^{\frac{H}{2}})(\mathbf{O}^{\frac{H}{2}}\mathbf{L}^H\widehat{\mathbf{H}}_1\mathbf{G}_1-\mathbf{O}^{\frac{H}{2}})^H\\&+\mathbf{O}^{\frac{H}{2}}\mathbf{L}^H\mathbf{\Phi}_1\mathbf{L}\mathbf{O}^{\frac{1}{2}}\Big\}\end{aligned} \tag{43a}$$

$$s.t.\ tr\{\mathbf{F}_k\mathbf{F}_k^H\}\le P_{s,k},\qquad k=1,\cdots,K. \tag{43b}$$

In the following, an iterative algorithm is proposed to solve the problem (43). In this iterative algorithm, $\mathbf{L}$ is optimized by (63) in Appendix B based on $\{\mathbf{F}_k\}$ from the previous iteration. Then using the $\mathbf{L}$ obtained from the current iteration, $\{\mathbf{F}_k\}$ are optimized by solving the following problem

$$\min_{\{\mathbf{F}_k\}}\ tr\Big\{(\mathbf{P}\mathbf{G}_1-\mathbf{O}^{\frac{H}{2}})(\mathbf{P}\mathbf{G}_1-\mathbf{O}^{\frac{H}{2}})^H\Big\} \tag{44a}$$

$$s.t.\ tr\{\mathbf{F}_k\mathbf{F}_k^H\}\le P_{s,k},\qquad k=1,\cdots,K \tag{44b}$$

where $\mathbf{P}=\mathbf{O}^{\frac{H}{2}}\mathbf{L}^H\widehat{\mathbf{H}}_1$. We update $\{\mathbf{F}_k\}$ and $\mathbf{L}$ alternatingly till convergence. By introducing $\mathbf{P}_k$ and $\mathbf{O}_k$, which contain the $\sum_{j=0}^{k-1}N_{S,j}+1$ to $\sum_{j=0}^{k}N_{S,j}$ columns of $\mathbf{P}$ and the $\sum_{j=0}^{k-1}N_{B,j}+1$ to $\sum_{j=0}^{k}N_{B,j}$ columns of $\mathbf{O}^{\frac{H}{2}}$ respectively, $k=1,\cdots,K$, with $N_{S,0}=N_{B,0}=0$, the objective function (44a) can be rewritten as

$$\sum_{k=1}^{K}tr\Big\{(\mathbf{P}_k\mathbf{F}_k-\mathbf{O}_k)(\mathbf{P}_k\mathbf{F}_k-\mathbf{O}_k)^H\Big\}. \tag{45}$$

It can be seen from (44b) and (45) that the problem (44) can be decomposed into K subproblems. Each $\mathbf{F}_k$ can be obtained by solving the following optimization problem

$$\min_{\mathbf{F}_k}\ tr\Big\{(\mathbf{P}_k\mathbf{F}_k-\mathbf{O}_k)(\mathbf{P}_k\mathbf{F}_k-\mathbf{O}_k)^H\Big\} \tag{46a}$$

$$s.t.\ tr\{\mathbf{F}_k\mathbf{F}_k^H\}\le P_{s,k}. \tag{46b}$$

The Lagrange multiplier method [32] can be used to solve the optimization problem (46) and the solution is

$$\mathbf{F}_k=(\mathbf{P}_k^H\mathbf{P}_k+\lambda_k\mathbf{I}_{N_{S,k}})^{-1}\mathbf{P}_k^H\mathbf{O}_k \tag{47}$$

where $\lambda_k\ge 0$ is the Lagrangian multiplier. Based on (63) and (47), the computational complexity order of optimizing $\{\mathbf{F}_k\}$ at the first relay node can be estimated as $\mathcal{O}(c_1(KN_{S,k}^3+N_{R,1}^3))$, where c_1 is the number of iterations till the convergence of $\{\mathbf{F}_k\}$. Since the update of $\mathbf{L}$ and $\{\mathbf{F}_k\}$ in each iteration reduces or maintains but can not increase the value of the objective function (43a), and (43a) is bounded below by 0, a monotonic convergence of the iterative approach in solving the problem (43) can be observed.

C. Optimization of $\mathbf{T}_l$

In general cases of $\mathbf{\Psi}_l\neq\sigma_e^2\mathbf{I}_{N_{R,l}}$, the problem (37) is difficult to solve as $\mathbf{\Upsilon}_l$ is a function of $\mathbf{T}_l$. Thus, the inequality (41) is applied to overcome the challenge. By applying the inequality (41), an upper-bound of (34) is given by

$$\mathbf{\Upsilon}_l\le P_{r,l}\lambda_m(\mathbf{\Psi}_l)\mathbf{\Sigma}_l+\mathbf{I}_{N_{R,l}}\triangleq\mathbf{\Phi}_l. \tag{48}$$

Using (48), an upper-bound of (37a) is given by

$$tr\big\{(\mathbf{O}^{-1}+\mathbf{T}_l^H\mathbf{B}_l\mathbf{T}_l)^{-1}\big\} \tag{49}$$

where $\mathbf{B}_l\triangleq\widehat{\mathbf{H}}_l^H\mathbf{\Phi}_l^{-1}\widehat{\mathbf{H}}_l$. From (49), the problem (37) can be written as

$$\min_{\mathbf{T}_l}\ tr\big\{(\mathbf{O}^{-1}+\mathbf{T}_l^H\mathbf{B}_l\mathbf{T}_l)^{-1}\big\} \tag{50a}$$

$$s.t.\ tr\{\mathbf{T}_l\mathbf{O}\mathbf{T}_l^H\}\le P_{r,l}. \tag{50b}$$

The eigenvalue decomposition (EVD) of $\mathbf{O}$ and $\mathbf{B}_l$ is introduced as $\mathbf{O}=\mathbf{U}_O\mathbf{\Sigma}_O\mathbf{U}_O^H$ and $\mathbf{B}_l=\mathbf{V}_l\mathbf{\Lambda}_{\mathbf{B}_l}\mathbf{V}_l^H$ respectively,

where the diagonal elements of $\mathbf{\Sigma}_O$ and $\mathbf{\Lambda}_{\mathbf{B}_l}$ are sorted in descending sequence. The optimization problem (50) can be solved by using *Lemma 2* in [29], and the solution is given by

$$\mathbf{T}_l = \mathbf{V}_{l,1}\mathbf{\Lambda}_{\mathbf{T}_l}^{\frac{1}{2}}\mathbf{U}_O^H \tag{51}$$

where $\mathbf{V}_{l,1}$ contains the leftmost N_B columns of $\mathbf{V}_l$, and $\mathbf{\Lambda}_{\mathbf{T}_l}$ is a diagonal matrix with nonnegative diagonal entries. By applying (51), the problem (50) can be converted to the following problem with scalar variables

$$\min_{\{\lambda_{\mathbf{T}_l,i}\}} \sum_{i=1}^{N_B} \frac{1}{\sigma_{O,i}^{-1} + \lambda_{\mathbf{T}_l,i}\lambda_{\mathbf{B}_l,i}} \tag{52a}$$

$$s.t. \sum_{i=1}^{N_B} \lambda_{\mathbf{T}_l,i}\sigma_{O,i} \le P_{r,l} \tag{52b}$$

$$\lambda_{\mathbf{T}_l,i} \ge 0, \qquad i = 1, \cdots, N_B \tag{52c}$$

where $\{\lambda_{\mathbf{T}_l,i}\} \triangleq \{\lambda_{\mathbf{T}_l,i}, i = 1, \cdots, N_B\}$, $\lambda_{\mathbf{T}_l,i}$, $\sigma_{O,i}$, and $\lambda_{\mathbf{B}_l,i}$ are the ith diagonal elements of $\mathbf{\Lambda}_{\mathbf{T}_l}$, $\mathbf{\Sigma}_O$, and $\mathbf{\Lambda}_{\mathbf{B}_l}$, respectively. By applying the Karush-Kuhn-Tucker (KKT) optimality conditions, the optimal solution for the problem (52) is obtained as

$$\lambda_{\mathbf{T}_l,i} = \frac{1}{\lambda_{\mathbf{B}_l,i}}\left(\sqrt{\frac{\lambda_{\mathbf{B}_l,i}}{\mu_l\sigma_{O,i}}} - \frac{1}{\sigma_{O,i}}\right)^+, \quad i = 1, \cdots, N_B \tag{53}$$

where $(x)^+ = \max(x, 0)$ and $\mu_l > 0$ is the solution to the non-linear equation of $\sum_{i=1}^{N_B} \frac{\sigma_{O,i}}{\lambda_{\mathbf{B}_l,i}}\left(\sqrt{\frac{\lambda_{\mathbf{B}_l,i}}{\mu_l\sigma_{O,i}}} - \frac{1}{\sigma_{O,i}}\right)^+ = P_{r,l}$. This solution is also known as the water-filling solution. As the most computationally intensive operation of optimizing $\mathbf{T}_l$ is calculating the EVD of $\mathbf{O}$, the complexity order of optimizing $\mathbf{T}_l$ is $\mathcal{O}(N_B^3)$.

D. Optimization of $\tilde{\mathbf{T}}_L$

Similar to the problem (36), the problem (38) can be converted to a convex SDP problem if $\mathbf{O} = \mathbf{I}_{N_B}$. But for $\mathbf{O} \neq \mathbf{I}_{N_B}$, the problem (38) is difficult to be formulated into a convex optimization problem. Let us consider a single-hop MIMO channel with

$$\tilde{\mathbf{y}}_{d,m} = \mathbf{H}_{L,m}\tilde{\mathbf{T}}_L\mathbf{s} + \mathbf{n}_{d,m}. \tag{54}$$

Using (54), we show in Appendix C that the function in (38a) can be reformulated as

$$\begin{aligned} &tr\{\mathbf{O}(\mathbf{I}_{N_B} + \tilde{\mathbf{T}}_L^H\widehat{\mathbf{H}}_{L,m}^H\mathbf{\Upsilon}_{L,m}^{-1}\widehat{\mathbf{H}}_{L,m}\tilde{\mathbf{T}}_L)^{-1}\} \\ &= \min_{\mathbf{\Theta}_m} E_H\{tr(\mathbf{O}E\{(\mathbf{\Theta}_m^H\tilde{\mathbf{y}}_{d,m} - \mathbf{s})(\mathbf{\Theta}_m^H\tilde{\mathbf{y}}_{d,m} - \mathbf{s})^H\})\} \end{aligned} \tag{55}$$

where $\mathbf{\Theta}_m$ is the weight matrix of the linear receiver for the MIMO system in (54). By utilizing (55), the problem (38) can be solved through the following problem

$$\min_{\tilde{\mathbf{T}}_L,\{\mathbf{\Theta}_m\}} \max_m tr\Big\{(\mathbf{O}^{\frac{H}{2}}\mathbf{\Theta}_m^H\widehat{\mathbf{H}}_{L,m}\tilde{\mathbf{T}}_L - \mathbf{O}^{\frac{H}{2}})(\mathbf{O}^{\frac{H}{2}}\mathbf{\Theta}_m^H\widehat{\mathbf{H}}_{L,m}\tilde{\mathbf{T}}_L - \mathbf{O}^{\frac{H}{2}})^H + \mathbf{O}^{\frac{H}{2}}\mathbf{\Theta}_m^H\mathbf{\Upsilon}_{L,m}\mathbf{\Theta}_m\mathbf{O}^{\frac{1}{2}}\Big\} \tag{56a}$$

$$s.t.\ tr\{\tilde{\mathbf{T}}_L\tilde{\mathbf{T}}_L^H\} \le P_{r,L} \tag{56b}$$

where $\{\mathbf{\Theta}_m\} \triangleq \{\mathbf{\Theta}_m, m = 1, \cdots, M\}$. Based on the technique used for optimizing $\{\mathbf{F}_k\}$ and $\{\mathbf{T}_l\}$, an upper-bound of $\mathbf{\Upsilon}_{L,m}$ is given by

$$\mathbf{\Upsilon}_{L,m} \le P_{r,L}\lambda_m(\mathbf{\Psi}_{L,m})\mathbf{\Sigma}_{L,m} + \mathbf{I}_{N_D} \triangleq \mathbf{\Phi}_{L,m}. \tag{57}$$

Using (57), the problem (56) can be written as

$$\min_{\tilde{\mathbf{T}}_L,\{\mathbf{\Theta}_m\}} \max_m tr\Big\{(\mathbf{O}^{\frac{H}{2}}\mathbf{\Theta}_m^H\widehat{\mathbf{H}}_{L,m}\tilde{\mathbf{T}}_L - \mathbf{O}^{\frac{H}{2}})(\mathbf{O}^{\frac{H}{2}}\mathbf{\Theta}_m^H\widehat{\mathbf{H}}_{L,m}\tilde{\mathbf{T}}_L - \mathbf{O}^{\frac{H}{2}})^H + \mathbf{O}^{\frac{H}{2}}\mathbf{\Theta}_m^H\mathbf{\Phi}_{L,m}\mathbf{\Theta}_m\mathbf{O}^{\frac{1}{2}}\Big\} \tag{58a}$$

$$s.t.\ tr\{\tilde{\mathbf{T}}_L\tilde{\mathbf{T}}_L^H\} \le P_{r,L}. \tag{58b}$$

Because of the highly complex objective function (58a), the min-max problem (58) is difficult to solve. We propose an iterative algorithm to obtain the solution of the problem (58). In this algorithm, $\mathbf{\Theta}_m$ is optimized by (65) in Appendix C based on $\tilde{\mathbf{T}}_L$ from the previous iteration. Then using $\{\mathbf{\Theta}_m\}$ obtained from the current iteration, $\tilde{\mathbf{T}}_L$ is obtained by optimizing the following problem

$$\min_{\tilde{\mathbf{T}}_L} \max_m tr\Big\{(\mathbf{O}^{\frac{H}{2}}\mathbf{\Theta}_m^H\widehat{\mathbf{H}}_{L,m}\tilde{\mathbf{T}}_L - \mathbf{O}^{\frac{H}{2}}) \times (\mathbf{O}^{\frac{H}{2}}\mathbf{\Theta}_m^H\widehat{\mathbf{H}}_{L,m}\tilde{\mathbf{T}}_L - \mathbf{O}^{\frac{H}{2}})^H\Big\} \tag{59a}$$

$$s.t.\ tr\{\tilde{\mathbf{T}}_L\tilde{\mathbf{T}}_L^H\} \le P_{r,L}. \tag{59b}$$

We update $\{\mathbf{\Theta}_m\}$ and $\tilde{\mathbf{T}}_L$ alternatingly till convergence.

Let us introduce a real-valued slack variable γ and positive semidefinite (PSD) matrices $\mathbf{C}_m$, $m = 1, \cdots, M$, with $\mathbf{C}_m \ge (\mathbf{O}^{\frac{H}{2}}\mathbf{\Theta}_m^H\widehat{\mathbf{H}}_{L,m}\tilde{\mathbf{T}}_L - \mathbf{O}^{\frac{H}{2}})(\mathbf{O}^{\frac{H}{2}}\mathbf{\Theta}_m^H\widehat{\mathbf{H}}_{L,m}\tilde{\mathbf{T}}_L - \mathbf{O}^{\frac{H}{2}})^H$ and $\mathbf{Q} \ge \tilde{\mathbf{T}}_L\tilde{\mathbf{T}}_L^H$. By applying the Schur complement, the optimization problem (59) can be expressed as

$$\min_{\gamma,\mathbf{Q},\{\mathbf{C}_m\},\tilde{\mathbf{T}}_L} \gamma \tag{60a}$$

$$s.t.\ tr\{\mathbf{C}_m\} \le \gamma, \qquad m = 1, \cdots, M \tag{60b}$$

$$\begin{pmatrix} \mathbf{C}_m & \bar{\mathbf{K}}_m \\ \bar{\mathbf{K}}_m^H & \mathbf{I}_{N_B} \end{pmatrix} \ge 0,\ m = 1, \cdots, M \tag{60c}$$

$$tr\{\mathbf{Q}\} \le P_{r,L}, \qquad \mathbf{Q} \ge 0 \tag{60d}$$

$$\begin{pmatrix} \mathbf{Q} & \tilde{\mathbf{T}}_L \\ \tilde{\mathbf{T}}_L^H & \mathbf{I}_{N_B} \end{pmatrix} \ge 0 \tag{60e}$$

where $\bar{\mathbf{K}}_m = \mathbf{O}^{\frac{H}{2}}\mathbf{\Theta}_m^H\widehat{\mathbf{H}}_{L,m}\tilde{\mathbf{T}}_L - \mathbf{O}^{\frac{H}{2}}$, $m = 1, \cdots, M$ and $\{\mathbf{C}_m\} = \{\mathbf{C}_m, m = 1, \cdots, M\}$. The problem (60) is a convex SDP problem which can be solved by the convex programming toolbox CVX [33]. Since the conditional update of $\tilde{\mathbf{T}}_L$ and $\{\mathbf{\Theta}_m\}$ in each iteration reduces or maintains but can not increase the value of the objective function (58a), and (58a) is bounded below by 0, a monotonic convergence of the iterative approach in solving the problem (58) can be observed. Since the complexity of solving the problem (60) is $\mathcal{O}((N_{R,L-1}^2 + (M+1)N_B^2)^2(M+1)^{1.5}N_B^{2.5})$, the computational complexity of solving the problem (58) has an order of $\mathcal{O}(c_2(N_D^3 + (N_{R,L-1}^2 + (M+1)N_B^2)^2(M+1)^{1.5}N_B^{2.5}))$, where c_2 is the number of iterations till convergence. By adding up this complexity order with the complexity of optimizing $\{\mathbf{F}_k\}$ in Section III.B and optimizing $\mathbf{T}_l$ in Section III.C, the overall complexity of the proposed robust transceiver design algorithm

is given by $\mathcal{O}(c_1(KN_{S,k}^3+N_{R,1}^3)+(L-2)N_B^3+c_2(N_D^3+(N_{R,L-1}^2+(M+1)N_B^2)^2(M+1)^{1.5}N_B^{2.5}))$. Note that this complexity order is for the general case where the weight matrix $\mathbf{O}\neq\mathbf{I}_{N_B}$. For the special case of $\mathbf{O}=\mathbf{I}_{N_B}$, the robust transceiver design subproblems (36), (37), and (38) can be solved with non-iterative approaches at the same complexity order as the corresponding subproblems in Section III.B of [14]. Note that the full CSI is required in [14], while the CSI mismatch is considered in the proposed algorithm.

IV. Simulation Results

In this section, the performance of the proposed robust transceiver design algorithm for multi-hop multicasting MIMO relay systems is investigated through numerical simulations. We simulate multi-hop AF MIMO relay multicasting systems with $K=2$ source nodes and $N_{S,1}=N_{S,2}=N_S=2$. Except for the last simulation example, a three-hop ($L=3$) system is simulated with $N_{R,1}=8$, $N_{R,2}=4$, and $N_D=8$. The information-carrying symbols are modulated by quadrature phase-shift keying (QPSK) constellations. We set $P_{s,1}=P_{s,2}=P_s$, and the SNRs of the first hop and the other two hops are set as $\mathrm{SNR}_1=P_s/N_S$ and $\mathrm{SNR}_2=P_{r,1}/N_{R_1}=P_{r,2}/N_{R_2}$, respectively. In each channel realization, each source node sends 1000 randomly generated bits. The simulation results are averaged through 1000 independent channel realizations.

We model the correlation matrices of channel estimation errors as [34]

$$
\begin{aligned}
[\mathbf{\Psi}_{1,k}]_{p,q}&=\alpha^{|p-q|},\ p,q=1,\cdots,N_{S,k},\ k=1,\cdots,K\\
[\mathbf{\Psi}_{l}]_{p,q}&=\alpha^{|p-q|},\ p,q=1,\cdots,N_{R,l-1},\ l=2,\cdots,L-1\\
[\mathbf{\Psi}_{L,m}]_{p,q}&=\alpha^{|p-q|},\ p,q=1,\cdots,N_{L},\ m=1,\cdots,M\\
[\mathbf{\Sigma}_{1,k}]_{p,q}&=\sigma_e^2\beta^{|p-q|},\ p,q=1,\cdots,N_{R_1},\ k=1,\cdots,K\\
[\mathbf{\Sigma}_{l}]_{p,q}&=\sigma_e^2\beta^{|p-q|},\ p,q=1,\cdots,N_{R,l},\ l=2,\cdots,L-1\\
[\mathbf{\Sigma}_{L,m}]_{p,q}&=\sigma_e^2\beta^{|p-q|},\ p,q=1,\cdots,N_{D},\ m=1,\cdots,M
\end{aligned}
$$

where $\alpha\in[0,1]$ and $\beta\in[0,1]$ are the correlation coefficients, and σ_e^2 represents the variance of the estimation error. We set $\alpha=0.1$ and $\beta=0.01$ in the simulations. The estimated channel matrices $\widehat{\mathbf{H}}_{1,k}$, $\widehat{\mathbf{H}}_{l}$, and $\widehat{\mathbf{H}}_{2,m}$ are modelled through the distributions given by [18]

$$
\begin{aligned}
\widehat{\mathbf{H}}_{1,k}&\sim\mathcal{CN}\Big(\mathbf{0},\frac{1-\sigma_e^2}{\sigma_e^2}\mathbf{\Sigma}_{1,k}\otimes\mathbf{\Psi}_{1,k}^T\Big),\quad k=1,\cdots,K\\
\widehat{\mathbf{H}}_{l}&\sim\mathcal{CN}\Big(\mathbf{0},\frac{1-\sigma_e^2}{\sigma_e^2}\mathbf{\Sigma}_{l}\otimes\mathbf{\Psi}_{l}^T\Big),\quad l=2,\cdots,L-1\\
\widehat{\mathbf{H}}_{L,m}&\sim\mathcal{CN}\Big(\mathbf{0},\frac{1-\sigma_e^2}{\sigma_e^2}\mathbf{\Sigma}_{L,m}\otimes\mathbf{\Psi}_{L,m}^T\Big),\ m=1,\cdots,M.
\end{aligned}
$$

The system performance of the proposed robust algorithm (Robust) is compared with the algorithm given in [14] using only the estimated CSI $\widehat{\mathbf{H}}_{1,k}$, $k=1,\cdots,K$, $\widehat{\mathbf{H}}_{l}$, $l=2,\cdots,L-1$, and $\widehat{\mathbf{H}}_{L,m}$, $m=1,\cdots,M$ (denoted the Non-Robust algorithm). The performance of the algorithm in [14] with the full CSI (FCSI) of $\mathbf{H}_{1,k}$, $k=1,\cdots,K$, $\mathbf{H}_{l}$, $l=2,\cdots,L-1$, and $\mathbf{H}_{L,m}$, $m=1,\cdots,M$ (denoted the FCSI algorithm) is also shown and included as a benchmark.

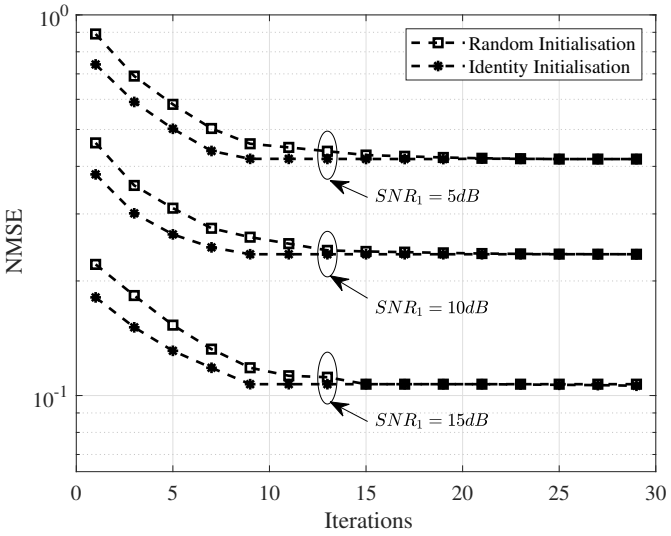


Fig. 2. Example 1: NMSE versus the number of iterations at different SNR_1. $L=3$, $M=4$, and $\sigma_e^2=0.0001$.

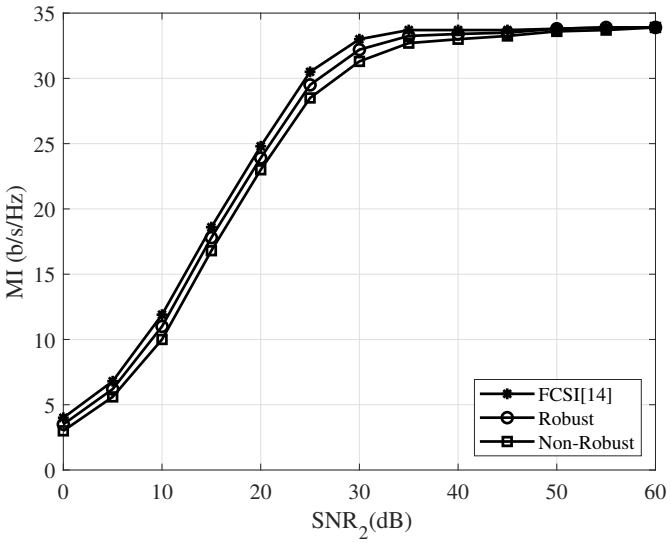


Fig. 3. Example 2: MI versus SNR_2. $L=3$, $M=4$, $\sigma_e^2=0.0001$, and $\mathrm{SNR}_1=25$ dB.

In the first numerical example, the convergence rate of the iterative algorithm in Section III.B to optimize $\{\mathbf{F}_k\}$ is investigated. We set $L=3$, $M=4$, and $\sigma_e^2=0.0001$. The simulation is carried out at three different SNR_1, i.e., $\mathrm{SNR}_1=5$ dB, $\mathrm{SNR}_1=10$ dB, and $\mathrm{SNR}_1=15$ dB, while $\{\mathbf{F}_k\}$ are initialized either by scaled identity matrices or square matrices with randomized entries. Fig. 2 illustrates the number of iterations required by the algorithm in Section III.B to achieve convergence in the normalized MSE (NMSE) (with $\mathbf{O}=\mathbf{I}_{N_B}$) of the signal waveform estimation. It can be noted from Fig. 2 that the NMSE monotonically decreases with the number of iterations towards convergence, which verifies the convergence analysis in Section III.B. It is also observed that the convergence rate of the algorithm in Section III.B increases with SNR_1. Based on the comparison between the two randomization methods, it can be seen that the initialization

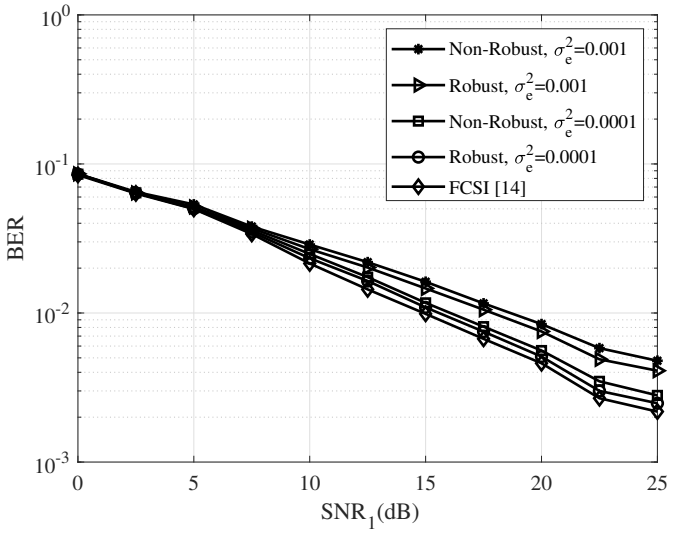


Fig. 4. Example 3: BER versus SNR_1 at various σ_e^2. $L = 3$, $M = 4$, and $\mathrm{SNR}_2 = 25$ dB.

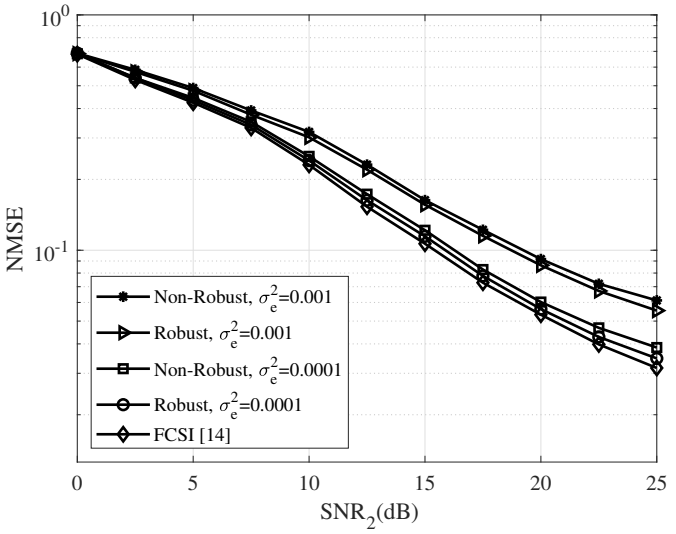


Fig. 5. Example 4: NMSE versus SNR_2 at different σ_e^2. $L = 3$, $M = 4$, and $\mathrm{SNR}_1 = 25$ dB.

with scaled identity matrices requires less iterations to achieve convergence. Thus, in the remaining numerical examples, the algorithm in Section III.B is initialized with scaled identity matrices for a faster convergence rate.

In the second numerical example, we investigate the mutual information (MI) between the source nodes and the destination nodes using the proposed robust transceiver design algorithm. To maximize the system MI, we apply the iterative approach in [26], [27] which turns the MI maximization problem into the WMSE minimization problem. The weight matrix $\mathbf{O} \neq \mathbf{I}_{N_B}$ is updated iteratively, and after the first iteration, $\mathbf{O}$ is equal to the inverse of the MSE matrix in each iteration [26]. Fig. 3 demonstrates the system MI versus SNR_2 at $L = 3$, $M = 4$, $\sigma_e^2 = 0.0001$, and $\mathrm{SNR}_1 = 25$ dB. It can be noticed that the FCSI system provides an upper-bound of the system MI, while the proposed robust algorithm outperforms the existing non-robust approach across all SNR_2. We would like to mention that the FCSI system is an ideal case where the full CSI is known for the transceiver design, which is unavailable in practice. Through this numerical example, it is shown that the proposed algorithm works efficiently for the general case of $\mathbf{O} \neq \mathbf{I}_{N_B}$. In the remaining numerical examples, we choose the weight matrix to be an identity matrix as we focus on the NMSE and bit-error-rate (BER) performance of the proposed robust system.

In the third simulation example, the BER performance of the proposed robust transceiver design algorithm is studied. Fig. 4 illustrates the BER performance of the three algorithms versus SNR_1 at $L = 3$, $M = 4$, and $\mathrm{SNR}_2 = 25$ dB with $\sigma_e^2 = 0.001$, and $\sigma_e^2 = 0.0001$. It is noted from Fig. 4 that the proposed robust algorithm consistently outperforms the non-robust algorithm over the entire SNR_1 region. Moreover, the results in Fig. 4 demonstrate that the FCSI algorithm provides a lower bound of the system BER. When CSI mismatch is small, the performance of the proposed robust optimization algorithm is close to that of the FCSI scheme.

In the fourth numerical example, the performance of the proposed algorithm at $L = 3$, $M = 4$, and $\mathrm{SNR}_1 = 25$ dB with $\sigma_e^2 = 0.001$ and $\sigma_e^2 = 0.0001$ is studied. The NMSE performance is illustrated in Fig. 5 and the BER performance is shown in Fig. 6 for the three algorithms versus SNR_2. It is noticed from Figs. 5 and 6 that the system performances of the proposed robust algorithm in term of NMSE and BER are better than that of the non-robust algorithm over the entire SNR_2 range. When σ_e^2 is reduced, the system NMSE and BER decrease for both the proposed robust and non-robust algorithms. When $\sigma_e^2 = 0.0001$, the performance of the proposed algorithm is very close to that of the FCSI scheme.

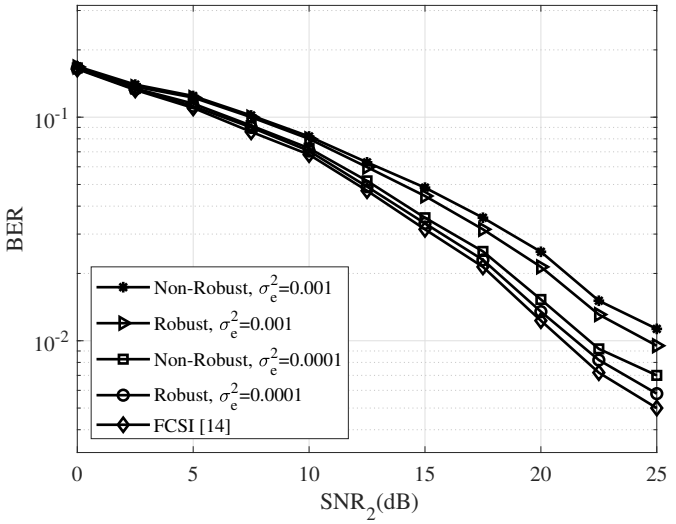


Fig. 6. Example 4: BER versus SNR_2 at different σ_e^2. $L = 3$, $M = 4$, and $\mathrm{SNR}_1 = 25$ dB.

In the fifth example, the BER performance of the proposed robust algorithm with various number of destination nodes is investigated. In this simulation, we set $L = 3$, $\mathrm{SNR}_1 = 25$ dB, and $\sigma_e^2 = 0.0001$. Fig. 7 displays the BER performance of the proposed algorithm and the FCSI scheme versus SNR_2 for $M = 2$, $M = 4$, and $M = 8$. It can be seen from

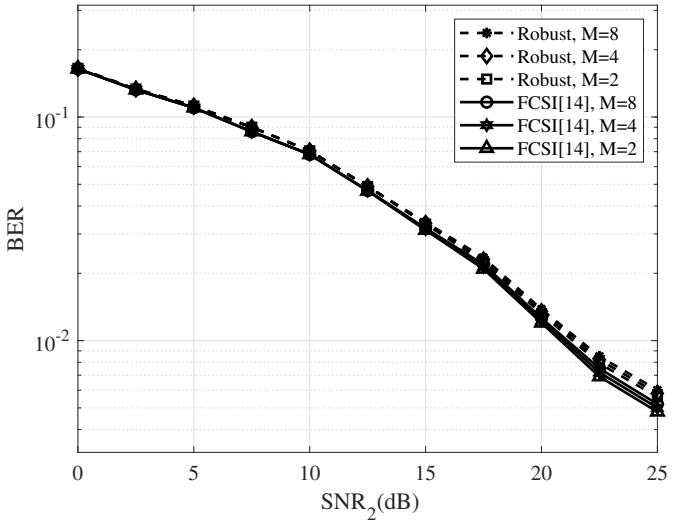


Fig. 7. Example 5: BER versus SNR_2 at different M. $L = 3$, $\sigma_e^2 = 0.0001$, and $\mathrm{SNR}_1 = 25$ dB.

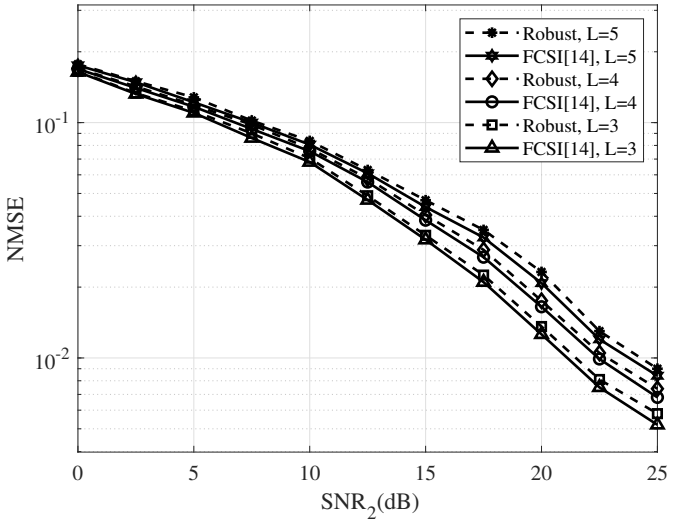


Fig. 8. Example 6: NMSE versus SNR_2 at different L. $M = 4$, $\sigma_e^2 = 0.0001$, and $\mathrm{SNR}_1 = 25$ dB.

Fig. 7 that as the number of destination nodes increases, the BER keeps increasing which is analogous to the results obtained in [15]. This is reasonable because it is more likely to find a worse relay-destination channel among an increased number of destinations and we choose the worst-user MSE as the objective function. For a various number of destination nodes, the average BER of the users is very similar up to $\mathrm{SNR}_2 = 15$ dB.

In the last simulation example, we investigate the NMSE performance of the proposed robust algorithm for multicasting AF MIMO relay communication systems with various number of hops. We set $M = 4$, $\mathrm{SNR}_1 = 25$dB, and $\sigma_e^2 = 0.0001$ for this simulation. Fig. 8 illustrates the NMSE performance of the proposed robust algorithm and the FCSI algorithm versus SNR_2 for $L = 3$, 4, and 5. We set $N_{R,3} = 4$ for $L = 4$ and $N_{R,3} = N_{R,4} = 4$ for $L = 5$. It is noticeable that the NMSE increases with the number of hops. This is because with a larger number of hops, the communication between the source nodes and the destination nodes occurs over a longer distance. It can be seen from (26) that with increasing number of hops, there are more trace terms in the WMSE function. Hence, the system NMSE increases with the number of hops. We also observe from Fig. 8 that the NMSE performance of the robust algorithm is close to that of the FCSI system for $L = 3$, 4, and 5.

V. Conclusion

In this article, we have investigated the problem of robust transceiver design for multi-hop multicasting AF MIMO relay systems from multiple sources when there is mismatch between the exact and estimated CSI. Here, we assumed that multiple source nodes broadcast their message to a group of destination nodes through multiple serial relay nodes. In the proposed transceiver design, the Gaussian-Kronecker model was adopted for the CSI mismatch. Furthermore, a robust transceiver design algorithm was developed to jointly optimize the source, relay, and destination matrices to minimize the maximal WMSE of the received signal at all destination nodes. It can be noticed that the WMSE is made statistically robust against the CSI mismatch by averaging through the distributions of the exact CSI. Moreover, the WMSE decomposition was exploited in the proposed transceiver design to reduce the computational complexity of the transceiver optimization problem. Numerical examples demonstrated the improved performance of the proposed transceiver optimization algorithm against the channel mismatch.

Appendix A
Derivation of (20)

In this appendix, the derivation of (20) is presented. By substituting (17) into the WMSE function (19), we obtain

$$\begin{aligned} J_m \triangleq\ & tr\{\mathbf{O}(\mathbf{W}_m E_H\{E\{\mathbf{y}_{d,m}\mathbf{y}_{d,m}^H\}\}\mathbf{W}_m^H - \mathbf{W}_m E_H\{\bar{\mathbf{A}}_m\} \\ & - E_H\{\bar{\mathbf{A}}_m^H\}\mathbf{W}_m^H + \mathbf{I}_{N_B})\} \\ =\ & tr\{\mathbf{O}(\mathbf{I}_{N_B} + \mathbf{W}_m\mathbf{R}_{\mathbf{y}_{d,m}}\mathbf{W}_m^H - \mathbf{W}_m\widehat{\mathbf{A}}_m \\ & - \widehat{\mathbf{A}}_m^H\mathbf{W}_m^H)\}, \quad m = 1, \cdots, M \end{aligned} \tag{61}$$

where $\mathbf{R}_{\mathbf{y}_{d,m}}$ and $\widehat{\mathbf{A}}_m$ are derived as the following

$$\begin{aligned} \mathbf{R}_{\mathbf{y}_{d,m}} &= E_H\{E\{\mathbf{y}_{d,m}\mathbf{y}_{d,m}^H\}\} \\ &= E_H\{\mathbf{H}_{L,m}\mathbf{G}_L\mathbf{R}_{\mathbf{y}_{L-1}}\mathbf{G}_L^H\mathbf{H}_L^H\} + \mathbf{I}_{N_D} \\ &= \widehat{\mathbf{H}}_{L,m}\mathbf{G}_L\mathbf{R}_{\mathbf{y}_{L-1}}\mathbf{G}_L^H\widehat{\mathbf{H}}_L^H \\ &\quad + tr\{\mathbf{G}_L\mathbf{R}_{\mathbf{y}_{L-1}}\mathbf{G}_L^H\mathbf{\Psi}_{L,m}\}\mathbf{\Sigma}_{L,m} + \mathbf{I}_{N_D} \\ \widehat{\mathbf{A}}_m &= E_H\{\bar{\mathbf{A}}_m\} \\ &= \widehat{\mathbf{H}}_{L,m}\mathbf{G}_L\prod_{l=L-1}^{1}\left(\widehat{\mathbf{H}}_l\mathbf{G}_l\right). \end{aligned}$$

Here *Lemma 1* is used to obtain the third equation of $\mathbf{R}_{\mathbf{y}_{d,m}}$.

APPENDIX B
DERIVATION OF (39)

In this appendix, the derivation of (39) is shown. By taking the expectation on the right-hand side of (39) and using *Lemma 1*, we obtain

$$\begin{aligned}
&E_H\{tr(\mathbf{O}E\{(\mathbf{L}^H\mathbf{y}_1-\mathbf{s})(\mathbf{L}^H\mathbf{y}_1-\mathbf{s})^H\})\}\\
&=E_H\{tr\{\mathbf{O}[\mathbf{L}^H(\mathbf{H}_1\mathbf{G}_1\mathbf{G}_1^H\mathbf{H}_1^H+\mathbf{I}_{N_R})\mathbf{L}\\
&\quad-\mathbf{L}^H\mathbf{H}_1\mathbf{G}_1-\mathbf{G}_1^H\mathbf{H}_1^H\mathbf{L}+\mathbf{I}_{N_B}]\}\}\\
&=tr\{\mathbf{O}[\mathbf{L}^H(\widehat{\mathbf{H}}_1\mathbf{G}_1\mathbf{G}_1^H\widehat{\mathbf{H}}_1^H+\boldsymbol{\Xi}_1)\mathbf{L}\\
&\quad-\mathbf{L}^H\widehat{\mathbf{H}}_1\mathbf{G}_1-\mathbf{G}_1^H\widehat{\mathbf{H}}_1^H\mathbf{L}+\mathbf{I}_{N_B}]\}. \qquad (62)
\end{aligned}$$

Obviously, the optimal $\mathbf{L}$ minimizing (62) is the Wiener filter [28] given by

$$\mathbf{L}=(\widehat{\mathbf{H}}_1\mathbf{G}_1\mathbf{G}_1^H\widehat{\mathbf{H}}_1^H+\boldsymbol{\Xi}_1)^{-1}\widehat{\mathbf{H}}_1\mathbf{G}_1. \qquad (63)$$

After substituting (63) back into (62), the left-hand side of (39) can be obtained.

APPENDIX C
DERIVATION OF (55)

In this appendix, we show the derivation of (55). By calculating the statistical expectation on the right-hand side of (55) using *Lemma 1*, we obtain

$$\begin{aligned}
&E_H\{tr\{\mathbf{O}E\{(\boldsymbol{\Theta}_m^H\tilde{\mathbf{y}}_{d,m}-\mathbf{s})(\boldsymbol{\Theta}_m^H\tilde{\mathbf{y}}_{d,m}-\mathbf{s})^H\}\}\}\\
&=E_H\{tr\{\mathbf{O}[\boldsymbol{\Theta}_m^H(\mathbf{H}_{L,m}\tilde{\mathbf{T}}_L\tilde{\mathbf{T}}_L^H\mathbf{H}_{L,m}^H+\mathbf{I}_{N_D})\boldsymbol{\Theta}_m\\
&\quad-\boldsymbol{\Theta}_m^H\mathbf{H}_{L,m}\tilde{\mathbf{T}}_L-\tilde{\mathbf{T}}_L^H\mathbf{H}_{L,m}^H\boldsymbol{\Theta}_m+\mathbf{I}_{N_B}]\}\}\\
&=tr\{\mathbf{O}[\boldsymbol{\Theta}_m^H(\widehat{\mathbf{H}}_{L,m}\tilde{\mathbf{T}}_L\tilde{\mathbf{T}}_L^H\widehat{\mathbf{H}}_{L,m}^H+\boldsymbol{\Upsilon}_{L,m})\boldsymbol{\Theta}_m\\
&\quad-\boldsymbol{\Theta}_m^H\widehat{\mathbf{H}}_{L,m}\tilde{\mathbf{T}}_L-\tilde{\mathbf{T}}_L^H\widehat{\mathbf{H}}_{L,m}^H\boldsymbol{\Theta}_m+\mathbf{I}_{N_B}]\}. \qquad (64)
\end{aligned}$$

Obviously, the optimal $\boldsymbol{\Theta}_m$ minimizing (64) is the Wiener filter [28] given by

$$\boldsymbol{\Theta}_m=\left(\widehat{\mathbf{H}}_{L,m}\tilde{\mathbf{T}}_L\tilde{\mathbf{T}}_L^H\widehat{\mathbf{H}}_{L,m}^H+\boldsymbol{\Upsilon}_{L,m}\right)^{-1}\widehat{\mathbf{H}}_{L,m}\tilde{\mathbf{T}}_L. \qquad (65)$$

After substituting (65) back into (64), the left-hand side of (55) can be obtained.

REFERENCES

[1] R. O. Afolabi, A. Dadlani, and K. Kim, "Multicast scheduling and resource allocation algorithms for OFDMA-based systems: A survey," *IEEE Commun. Surverys Tut.*, vol. 15, no. 1, pp. 240-254, First Quarter 2013.

[2] D. Lecompte and F. Gabin, "Evolved multimedia broadcast/multicast service (eMBMS) in LTE-advanced: Overview and Rel-11 enhancements," *IEEE Commun. Mag.*, vol. 50, no. 11, pp. 68-74, Nov. 2012.

[3] N. D. Sidiropoulos, T. N. Davidson, and Z.-Q. Luo, "Transmit beamforming for physical-layer multicasting," *IEEE Trans. Signal Process.*, vol. 54, pp. 2239-2251, Jun. 2006.

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

[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, pp. 2002-2013, Jul. 2010.

[6] S. X. Wu, W.-K. Ma, and A. M.-C. So, "Physical-layer multicasting by stochastic transmit beamforming and Alamouti space-time coding," *IEEE Trans. Signal Process.*, vol. 61, no. 17, pp. 4230-4245, Sep. 2013.

[7] A. Z. Yalcin and M. Yuksel, "Precoder design for multi-group multicasting with a common message," *IEEE Trans. Commun.*, vol. 67, pp. 7302-7315, Oct. 2019.

[8] D. Senaratne and C. Tellambura, "Beamforming for physical layer multicasting," in *Proc. IEEE WCNC*, Cancun, Mexico, pp. 1776-1781, Mar. 2011.

[9] H. Zhu, N. Prasad, and S. Rangarajan, "Precoder design for physical layer multicasting," *IEEE Trans. Signal Process.*, vol. 60, no. 11, pp. 5932-5947, Nov. 2012.

[10] R. Duan, J. Wang, H. Zhang, Y. Ren, and L. Hanzo, "Joint multicast beamforming and relay design for maritime communication systems," *IEEE Trans. Green Commun. Network.*, vol. 4, no. 1, pp. 139-151, Mar. 2020.

[11] N. Bornhorst, M. Pesavento, and A. B. Gershman, "Distributed beamforming for multi-group multicasting relay networks," *IEEE Trans. Signal Process.*, vol. 60, no. 1, pp. 221-232, Jan. 2012.

[12] M. R. A. Khandaker and Y. Rong, "Precoding design for MIMO relay multicasting," *IEEE Trans. Wireless Commun.*, vol. 12, pp. 3544-3555, Jul. 2013.

[13] M. R. A. Khandaker and Y. Rong, "Simplified MIMO relay design for multicasting from multiple sources," in *Proc. IEEE ICASSP*, Florence, Italy, pp. 1-5, May 2014.

[14] M. R. A. Khandaker and Y. Rong, "Transceiver optimization for multi-hop MIMO relay multicasting from multiple sources," *IEEE Trans. Wireless Commun.*, vol. 13, no. 9, pp. 5162-5172, Sep. 2014.

[15] L. Gopal, Y. Rong, and Z. Zang, "Robust MMSE transceiver design for nonregenerative multicasting MIMO relay systems," *IEEE Trans. Veh. Technol.*, vol. 66, no. 10, pp. 8979-8989, Oct. 2017.

[16] 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.

[17] 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.

[18] 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.

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

[20] C. Xing, F. Gao, and Y. Zhou, "A framework for transceiver designs for multi-hop communications with covariance shaping constraints," *IEEE Trans. Signal Process.*, vol. 63, no. 15, pp. 3930-3945, Aug. 2015.

[21] C. Xing, Y. Ma, Y. Zhou, and F. Gao, "Transceiver optimization for multi-hop communications with per-antenna power constraints," *IEEE Trans. Signal Process.*, vol. 64, no. 6, pp. 1519-1534, Mar. 2016.

[22] 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, no. 12, pp. 1231-1242, Mar. 2013.

[23] Y. Rong, "Robust design for linear non-regenerative MIMO relays with imperfect channel state information," *IEEE Trans. Signal Process.*, vol. 59, pp. 2455-2460, May 2011.

[24] 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.

[25] A. K. Gupta and D. K. Nagar, *Matrix Variate Distributions*, vol. 104. United States: Chapman & Hall/CRC, 1999.

[26] Z. He, S. Guo, Y. Ou, and Y. Rong, "Multiuser multihop MIMO relay system design based on mutual information maximization," *IEEE Trans. Signal Processing*, vol. 62, pp. 5725-5733, Nov. 2014.

[27] Q. Shi, M. Razaviyayn, Z. Luo, and C. He, "An iteratively weighted MMSE approach to distributed sum-utility maximization for a MIMO interfering broadcast channel," *IEEE Trans. Signal Process.*, vol. 59, no. 9, pp. 4331-4340, Sep. 2011.

[28] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. United States: Prentice-Hall, 1993.

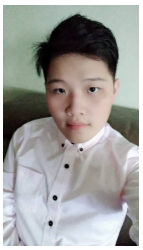
[29] Y. Rong, "Simplified algorithms for optimizing multiuser multi-hop MIMO relay systems," *IEEE Trans. Commun.*, vol. 59, no. 10, pp. 2896-2904, Oct. 2011.

[30] D. Bernstein, *Matrix Mathematics: Theory, Facts, and Formulas*. Princeton University Press, 2011.

[31] A. W. Marshall and I. Olkin, *Inequalities: Theory of Majorization and Its Applications*. New York: Academic Press, 1980.

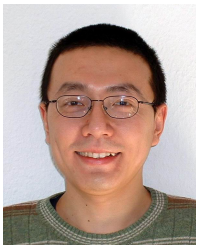
[32] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge, UK: Cambridge University Press, 2004.

[33] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming (web page and software)," Apr. 2010. Available: http://cvxr.com/cvx.
[34] C. Xing, Z. Fei, S. Ma, J. Kuang, and Y.-C. Wu, "Robust linear transceiver design for multi-hop non-regenerative MIMO relaying systems," in *Proc. Int. Conf. Wireless Commun. Signal Process.*, Nanjing, China, Nov. 9-11, 2011,



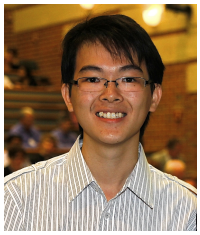
Justin Lee Bing (S'18) received the B.E. (Hons.) degree in electronic and communication engineering from Curtin University, Australia, in 2016. He is currently working toward the Ph.D. degree at the Department of Electrical and Computer Engineering, Curtin University. His research interests include wireless communications, signal processing for communications, simultaneous wireless information and power transfer and energy harvesting.

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 an Associate Professor 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, Power Line Communications, and the Internet of Things (IoT).



Yue Rong (S'03–M'06–SM'11) received the Ph.D. degree (*summa cum laude*) in electrical engineering from the Darmstadt University of Technology, Darmstadt, Germany, in 2005. He was a Post-Doctoral Researcher with the Department of Electrical Engineering, University of California, Riverside, from February 2006 to November 2007. Since December 2007, he has been with Curtin University, Bentley, Australia, where he is currently a Professor. His research interests include signal processing for communications, wireless communications, underwater acoustic communications, underwater optical wireless communications, applications of linear algebra and optimization methods, and statistical and array signal processing. He has published over 180 journal and conference papers in these areas.

Dr. Rong was a recipient of the Best Paper Award at the 2011 International Conference on Wireless Communications and Signal Processing, the Best Paper Award at the 2010 Asia-Pacific Conference on Communications, and the Young Researcher of the Year Award of the Faculty of Science and Engineering at Curtin University in 2010. He is a Senior Area Editor of the IEEE Transactions on Signal Processing, and was an Associate Editor of the IEEE Transactions on Signal Processing from 2014 to 2018, an Editor of the IEEE Wireless Communications Letters from 2012 to 2014, and a Guest Editor of the IEEE Journal on Selected Areas in Communications special issue on theories and methods for advanced wireless relays. He was also a TPC Member for the IEEE ICC, IEEE GlobalSIP, EUSIPCO, IEEE ICCC, WCSP, IWCMC, and ChinaCom.



Choo W.R. Chiong received the B.E. (Hons.) and Ph.D. degrees in electrical engineering from Curtin University, Bentley, WA, Australia, in 2010 and 2015, respectively. He is currently a Senior Lecturer with the Department of Electrical and Computer Engineering, Curtin University, Malaysia Campus. His research interests include signal processing for communications and power system optimization.

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 worked at 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 worked at the Australian Telecommunications Research Institute, Curtin University in Perth, Australia, first as a Research Fellow and then as a Senior Research Fellow. From 2002 to 2005, he worked 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 interest is currently in 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 non-regenerative MIMO wireless communication systems.