Joint Source and Relay Optimization for Non-Regenerative MIMO Two-Way Relay Systems with Imperfect CSI

Jun Zou, Hanwen Luo, Meixia Tao, *Senior Member, IEEE,* and Rui Wang

Abstract—In this paper, we consider a non-regenerative MIMO two-way relay system with imperfect channel state information (CSI). We employ a stochastic approach to model the channel uncertainties and address the robust joint source and relay optimization problem based on the minimum mean squared error (MMSE) criterion. With imperfect CSI, the self-interference (SI) cannot be completely canceled at destination nodes. Hence, both channel uncertainties and residual self-interference should be considered. We develop an optimization framework that unifies both frequency-division duplex (FDD) and time-division duplex (TDD) systems despite their different channel statistical properties. Two robust algorithms are proposed to minimize the sum mean squared error (MSE) averaged over channel uncertainties. The first algorithm adopts alternating optimization to update the source precoders, relay precoder and destination receivers iteratively with guaranteed convergence. In the second algorithm, only the relay precoder with certain structure is considered. Then the relay precoder design is reduced to the simple power allocation problem. Simulation results show that the proposed algorithms provide robustness against channel uncertainties, especially when the signal-to-noise ratio is high.

Index Terms—Precoder, analog network coding (ANC), two-way, multiple-input multiple-output (MIMO), relay, minimum mean squared error (MMSE).

I. INTRODUCTION

RELAYS can be deployed in wireless networks to extend coverage and improve throughput [1]. A wireless relay retransmits the signal received from the source to the destination in either a regenerative fashion such as decode-and-forward (DF) or a non-regenerative fashion such as amplify-and-forward (AF) [2]. To further improve the spectral efficiency of relay-assisted communications, two-way relaying has recently been proposed by using network coding. Network coding was first introduced in [3] as means to improve network throughput, in which intermediate nodes mixes the data received from different source nodes and retransmits the mixed information to destination nodes. Based on the principle of network coding, the authors in [4] proposed a non-regenerative two-way relaying scheme with analog network

coding (ANC), in which two source nodes simultaneously transmit their information over the same radio resource to the relay node during the first channel use, and the relay node amplifies and forwards the superimposed signal to both destination nodes during the second channel use. Note that in two-way relay systems, both source nodes are also the corresponding destination nodes of each other. Since the transmitted information from each node is known by itself, the destination nodes can cancel self-interference (SI) from the received signal before detecting the required data if perfect channel state information (CSI) is available. Hence, only two orthogonal channel uses are needed in ANC-based two-way relay systems for each round of information exchange.

Meanwhile, multiple-input multiple-output (MIMO) techniques can be incorporated into relay systems by equipping each node with multiple antennas to provide multiplexing and diversity gains [5]. Transceiver designs for traditional one-way MIMO relay systems were proposed in [6]–[10]. The optimal two-way relay strategies were studied in [11] to maximize achievable sum-rate and to achieve optimal diversity-multiplexing tradeoff. For an ANC-based MIMO two-way relay system consisting of a multi-antenna relay node and two single-antenna source nodes, the optimal relay beamforming was studied in [12] to characterize the achievable capacity region, and also in [13] to minimize the mean squared error. In [14] and [15], the authors considered jointly optimizing the source and relay precoders to maximize sum-rate for MIMO two-way relay systems where all nodes are equipped with multiple antennas. The joint optimization problem was also investigated in [16] based on the minimum mean squared error (MMSE) criterion, and an iterative algorithm as well as a heuristic algorithm based on channel parallelization was proposed therein. In [17] and [18], the authors addressed joint source and relay optimization for multiple-relay systems. Gradient descent algorithms were proposed in [17] to optimize source and relay iteratively based on sum-rate and mean squared error criteria, while in [18] an iterative algorithm based on solving convex sub-problems was proposed to minimize sum mean squared error.

All of the aforementioned works assume perfect channel state information (CSI) is available. In practice, channel state information has to be estimated using pilot symbols, and channel estimation errors are inevitable due to noise and the time-varying nature of wireless channels. Hence, it is desirable to provide robustness against channel uncertainties when de-

signing transceivers, i.e., precoders and receivers, for MIMO relay systems. Generally, there are two basic approaches for dealing with channel uncertainties, the stochastic approach and the worst-case approach. Depending on the system design criterion adopted, the first one is suitable when the average performance such as mean squared error is concerned, whereas the second one is preferred when certain quality-of-service requirement needs to be guaranteed. The robust designs for conventional one-way MIMO relay systems were studied in [19]–[21]. The worst-case approach was utilized in [19] to minimize the consumed power at the relay node subject to minimum signal-to-noise-plus-interference ratio (SINR) requirements for each data stream, while in [20] and [21] the stochastic approach was used to minimize the total MSE averaged over channel uncertainties. Unfortunately, the results therein cannot be generalized to two-way relay systems due to the bidirectional channels and self-interference. The effect of imperfect CSI was studied for MIMO two-way relay systems in [22] and [23]. The authors in [22] investigated the achievable rate region with imperfect CSI. In [23], the authors considered joint optimization of source and relay precoders with norm bounded CSI error model, and employed a worst-case approach to minimize the sum mean squared error. However, channel estimation errors are not necessarily bounded, e.g., when channels are estimated using minimum mean squared error (MMSE) method [26], the estimation errors are Gaussian random variables. The solutions in [23] cannot be applied in such scenarios.

In this paper, we address the robust joint source and relay optimization for an ANC-based MIMO two-way relay system to minimize the average sum MSE of detected signals at both destination nodes by using a stochastic approach to take channel uncertainties into account. Note that our earlier work in [24] only considered relay precoder optimization. Since self-interference (SI) cannot be completely canceled due to the imperfect CSI, the residual SI is also considered in our design. We develop a robust design framework that unifies both frequency-division duplex (FDD) and time-division duplex (TDD) systems despite their different channel statistical properties. An iterative algorithm is first proposed to jointly optimize the source and relay precoders as well as the destination receivers. Then to reduce complexity, a constrained structure algorithm is proposed by only considering the relay precoder with source precoders fixed and imposing a specific structure on the relay precoder.

The paper is organized as follows. The system model is described in Section II. The problem formulation for both FDD and TDD systems are developed in Section III. In Section IV and V, the iterative algorithm and the constrained structure algorithm are proposed, respectively. Then in Section VI, we provide simulation results to verify the effectiveness of the proposed robust algorithms. Finally, conclusions are given in Section VII.

Notations: Bold-face lower-case letters and bold-face upper-case letters are used for vectors and matrices, respectively. $(\cdot)^T$, $(\cdot)^H$, $(\cdot)^{-1}$, $(\cdot)^*$, $\|\cdot\|_F$, and $\mathrm{Tr}(\cdot)$ denote transpose, conjugate transpose, inverse, conjugate, Frobenius norm, and trace of a matrix, respectively. $\mathbb{E}\{\cdot\}$, $\mathrm{vec}(\cdot)$, $\mathrm{Re}\{\cdot\}$, and $\otimes$ denote expectation operator, vectorization operator, the real
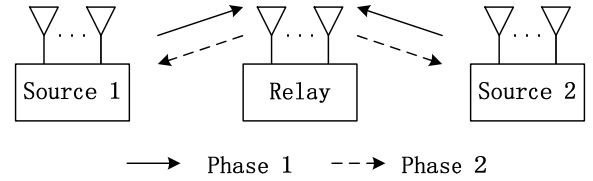

Fig. 1. Illustration of a MIMO two-way relay system.

part, and the Kronecker product, respectively. $\mathbf{A} \succeq \mathbf{0}$ means $\mathbf{A}$ is a positive semi-definite matrix, and $\mathbf{A} \succeq \mathbf{B}$ means $\mathbf{A} - \mathbf{B}$ is positive semi-definite. $\mathbb{C}^{K \times K}$ denotes space of $K \times K$ matrices with complex entries. $\mathbf{I}_K$ stands for a $K \times K$ identity matrix. $\mathrm{diag}\{\mathbf{A}\}$ represents a diagonal matrix formed from entries of the main diagonal of matrix $\mathbf{A}$. $\lambda_1(\mathbf{A})$ denotes the maximum eigenvalue of matrix $\mathbf{A}$. $\mathbf{A}(k, l)$ represents the entry on the kth row and lth column of matrix $\mathbf{A}$.

II. SYSTEM MODEL

The MIMO two-way relay system is shown in Fig. 1. We denote two source nodes as $\mathcal{S}_1$ and $\mathcal{S}_2$, respectively, and denote the relay node as $\mathcal{R}$. Both $\mathcal{S}_1$ and $\mathcal{S}_2$ have M antennas, and $\mathcal{R}$ has N antennas. All nodes are subject to half-duplex constraint. An analog network coding-based two-phase protocol is employed for $\mathcal{S}_1$ and $\mathcal{S}_2$ to exchange information via the aid of $\mathcal{R}$. The direct path between source nodes are not considered for FDD or TDD systems, because both source nodes simultaneously transmit signals to the relay node over the same channel. Each source node transmits M independent data streams, and to support transmission of all data streams the number of antennas at relay node should satisfy $N \geq M$ [16]. All channels are assumed to be flat fading and remain unchanged during the two phases.

A. FDD Mode

In the first phase, $\mathcal{S}_1$ and $\mathcal{S}_2$ simultaneously transmit data to $\mathcal{R}$ over the same radio resource. When the precoder $\mathbf{P}_i \in \mathbb{C}^{M \times M}$ is used by $\mathcal{S}_i$, $\forall i = 1, 2$, the transmitted signal from $\mathcal{S}_i$ is described by $\mathbf{x}_i = \mathbf{P}_i \mathbf{s}_i$, where $\mathbf{s}_i \in \mathbb{C}^{M \times 1}$ is the data vector with covariance matrix $\mathbf{R}_{s_i} = \mathbb{E}\{\mathbf{s}_i \mathbf{s}_i^H\}$. Denote $\mathbf{R}_{x_i} = \mathbb{E}\{\mathbf{x}_i \mathbf{x}_i^H\}$, and the transmit power constraint at $\mathcal{S}_i$ can be expressed as

$$\mathrm{Tr}(\mathbf{R}_{x_i}) = \mathrm{Tr}(\mathbf{P}_i \mathbf{R}_{s_i} \mathbf{P}_i^H) \leq P_{s_i}, \qquad (1)$$

where P_{s_i} is the maximum power which can be used by $\mathcal{S}_i$. The superimposed received signal $\mathbf{r} \in \mathbb{C}^{N \times 1}$ at $\mathcal{R}$ can be written as

$$\mathbf{r} = \mathbf{H}_1 \mathbf{x}_1 + \mathbf{H}_2 \mathbf{x}_2 + \mathbf{n}_r, \qquad (2)$$

where $\mathbf{H}_i \in \mathbb{C}^{N \times M}$ is the backward channel from $\mathcal{S}_i$ to $\mathcal{R}$, and $\mathbf{n}_r \in \mathbb{C}^{N \times 1}$ is the additive Gaussian noise vector with covariance matrix $\mathbf{R}_{n_r}$.

In the second phase, after multiplying the received signal $\mathbf{r}$ with precoding matrix $\mathbf{F} \in \mathbb{C}^{N \times N}$, the relay node $\mathcal{R}$ forwards the signal $\mathbf{F}\mathbf{r}$ to $\mathcal{S}_1$ and $\mathcal{S}_2$ subject to transmit power constraint P_r at $\mathcal{R}$, namely,

$$\mathrm{Tr}\left[\mathbf{F}\left(\mathbf{H}_1 \mathbf{R}_{x_1} \mathbf{H}_1^H + \mathbf{H}_2 \mathbf{R}_{x_2} \mathbf{H}_2^H + \mathbf{R}_{n_r}\right) \mathbf{F}^H\right] \leq P_r. \quad (3)$$

The observed signal at $\mathcal{S}_i$ can be expressed as

$$\mathbf{y}_i = \mathbf{G}_i\mathbf{F}\mathbf{H}_j\mathbf{x}_j + \mathbf{G}_i\mathbf{F}\mathbf{H}_i\mathbf{x}_i + \mathbf{G}_i\mathbf{F}\mathbf{n}_r + \mathbf{n}_i, \qquad (4)$$

where $j = 2$ if $i = 1$ and $j = 1$ if $i = 2$, $\mathbf{G}_i \in \mathbb{C}^{M \times N}$ is the forward channel from $\mathcal{R}$ to $\mathcal{S}_i$, and $\mathbf{n}_i \in \mathbb{C}^{M \times 1}$ is the additive Gaussian noise vector with covariance matrix $\mathbf{R}_{n_i}$.

B. TDD Mode

When the system works in TDD mode, we assume the uplink channel in the first phase and the downlink channel in the second phase are reciprocal for each source node. The transmission protocol for TDD systems is similar to FDD systems. However, in order to formulate a unified framework for both the FDD and TDD systems, we introduce a conjugate operation applied to signals at the TDD relay node $\mathcal{R}$ in the second phase. After receiving signal $\mathbf{r}$, $\mathcal{R}$ forwards the signal $(\mathbf{F}\mathbf{r})^*$ rather than $\mathbf{F}\mathbf{r}$ to $\mathcal{S}_1$ and $\mathcal{S}_2$. This conjugate operation applied at $\mathcal{R}$ helps to reduce the impact of channel uncertainties on the system as will be discussed later in Sec. III-B. Considering channel reciprocity as assumed, the received signal at $\mathcal{S}_i$ is described by

$$\mathbf{y}_i^* = \mathbf{H}_i^T(\mathbf{F}\mathbf{r})^* + \mathbf{n}_i. \qquad (5)$$

$\mathcal{S}_i$ then applies conjugate operation to $\mathbf{y}_i^*$ and get

$$\mathbf{y}_i = \mathbf{H}_i^H\mathbf{F}\mathbf{H}_j\mathbf{x}_j + \mathbf{H}_i^H\mathbf{F}\mathbf{H}_i\mathbf{x}_i + \mathbf{H}_i^H\mathbf{F}\mathbf{n}_r + \mathbf{n}_i^*. \qquad (6)$$

Note that $\mathbf{n}_i^*$ has the same statistical properties as $\mathbf{n}_i$, and thus is replaced with $\mathbf{n}_i$ in the following for notational convenience.

To avoid redundant elaboration on TDD systems, we generally use the same notation $\mathbf{G}_i$ as in FDD systems to refer to $\mathbf{H}_i^H$ through out the paper. Unless specific otherwise, all the following discussions and conclusions apply to both FDD and TDD systems.

C. Imperfect Channel State Information

Since mounting multiple antennas closely on one node introduces channel correlation, the Kronecker model [25] is used to describe $\mathbf{H}_i$ and $\mathbf{G}_i$ as

$$\mathbf{H}_i = \mathbf{R}_{hi}^{\frac{1}{2}}\mathbf{H}_{w_i}\mathbf{T}_{hi}^{\frac{1}{2}}, \quad \mathbf{G}_i = \mathbf{R}_{gi}^{\frac{1}{2}}\mathbf{G}_{w_i}\mathbf{T}_{gi}^{\frac{1}{2}}, \qquad (7)$$

where $\mathbf{R}_{hi} \in \mathbb{C}^{N \times N}$ and $\mathbf{R}_{gi} \in \mathbb{C}^{M \times M}$ are the row correlation matrices, and $\mathbf{T}_{hi} \in \mathbb{C}^{M \times M}$ and $\mathbf{T}_{gi} \in \mathbb{C}^{N \times N}$ are the column correlation matrices, all being positive semi-definite. We assume all correlation matrices are perfectly known, and the entries of $\mathbf{H}_{w_i}$ and $\mathbf{G}_{w_i}$ are independent and identically distributed (i.i.d.) zero-mean circularly symmetric complex Gaussian (ZMCSCG) random variables with unit variance. Note that the uncorrelated channel case is included by assigning identity matrices to all correlation matrices in (7).

$\mathbf{H}_{w_i}$ and $\mathbf{G}_{w_i}$ are estimated using minimum mean squared error (MMSE) method [26], with estimation error given by

$$\mathbf{E}_{hw_i} = \mathbf{H}_{w_i} - \hat{\mathbf{H}}_{w_i}, \quad \mathbf{E}_{gw_i} = \mathbf{G}_{w_i} - \hat{\mathbf{G}}_{w_i}, \qquad (8)$$

where $\hat{\mathbf{H}}_{w_i}$ and $\hat{\mathbf{G}}_{w_i}$ are the estimated channels. According to the MMSE estimation property, estimated channels are uncorrelated with estimation errors. The entries of $\mathbf{E}_{hw_i}$ and

$\mathbf{E}_{gw_i}$ are i.i.d. ZMCSCG with variance $\sigma_{e_i}^2$. We can rewrite $\mathbf{H}_i$ and $\mathbf{G}_i$ as

$$\mathbf{H}_i = \underbrace{\mathbf{R}_{hi}^{\frac{1}{2}}\hat{\mathbf{H}}_{w_i}\mathbf{T}_{hi}^{\frac{1}{2}}}_{\hat{\mathbf{H}}_i} + \underbrace{\mathbf{R}_{hi}^{\frac{1}{2}}\mathbf{E}_{hw_i}\mathbf{T}_{hi}^{\frac{1}{2}}}_{\mathbf{E}_{hi}}, \qquad (9)$$

$$\mathbf{G}_i = \underbrace{\mathbf{R}_{gi}^{\frac{1}{2}}\hat{\mathbf{G}}_{w_i}\mathbf{T}_{gi}^{\frac{1}{2}}}_{\hat{\mathbf{G}}_i} + \underbrace{\mathbf{R}_{gi}^{\frac{1}{2}}\mathbf{E}_{gw_i}\mathbf{T}_{gi}^{\frac{1}{2}}}_{\mathbf{E}_{gi}}. \qquad (10)$$

D. Data Detection

Before detecting the needed data, $\mathcal{S}_i$ first cancels self-interference (SI) with estimated CSI and gets

$$\begin{aligned} \hat{\mathbf{y}}_i &= \mathbf{y}_i - \hat{\mathbf{G}}_i\mathbf{F}\hat{\mathbf{H}}_i\mathbf{x}_i \\ &= \mathbf{G}_i\mathbf{F}\mathbf{H}_j\mathbf{x}_j + \mathbf{G}_i\mathbf{F}\mathbf{n}_r + \mathbf{z}_i + \mathbf{n}_i, \end{aligned} \qquad (11)$$

where $\mathbf{z}_i \triangleq (\mathbf{G}_i\mathbf{F}\mathbf{H}_i - \hat{\mathbf{G}}_i\mathbf{F}\hat{\mathbf{H}}_i)\mathbf{x}_i$ is the residual SI with covariance matrix $\mathbf{R}_{z_i} = \mathbb{E}\{\mathbf{z}_i\mathbf{z}_i^H\}$. Then a linear receiver $\mathbf{W}_i$ is employed by S_i to detect data $\mathbf{s}_j$, and the mean squared error (MSE) can be expressed as

$$J_i(\mathbf{P}_i, \mathbf{P}_j, \mathbf{F}, \mathbf{W}_i) = \mathbb{E}\{\|\mathbf{W}_i\hat{\mathbf{y}}_i - \mathbf{s}_j\|_F^2\}, \qquad (12)$$

where $\mathbb{E}\{\cdot\}$ is performed with respect to $\mathbf{s}_1$, $\mathbf{s}_2$, $\mathbf{n}_i$, and $\mathbf{n}_r$.

III. PROBLEM FORMULATION

We consider the joint design of source precoders $\mathbf{P}_i$, $\forall i = 1, 2$, and the relay precoder $\mathbf{F}$ as well as destination receivers $\mathbf{W}_i$, $\forall i = 1, 2$, to minimize the average sum MSE of the detected data at $\mathcal{S}_1$ and $\mathcal{S}_2$ with imperfect CSI. To take into account channel estimation errors, we take expectation of $J_i(\mathbf{P}_i, \mathbf{P}_j, \mathbf{F}, \mathbf{W}_i)$ with respect to $\mathbf{E}_{h1}$, $\mathbf{E}_{h2}$, $\mathbf{E}_{g1}$ and $\mathbf{E}_{g2}$ as follows

$$\begin{aligned} \hat{J}_i(\mathbf{P}_i, \mathbf{P}_j, \mathbf{F}, \mathbf{W}_i) &= \mathbb{E}\{J_i(\mathbf{P}_i, \mathbf{P}_j, \mathbf{F}, \mathbf{W}_i)\} \\ &= \mathbb{E}_{\mathbf{E}_{gi}, \mathbf{E}_{hj}}\left\{\mathrm{Tr}\left[\mathbf{W}_i\mathbf{G}_i\mathbf{F}\mathbf{H}_j\mathbf{R}_{x_j}(\mathbf{W}_i\mathbf{G}_i\mathbf{F}\mathbf{H}_j)^H\right]\right\} \\ &\quad + \mathbb{E}_{\mathbf{E}_{gi}}\left\{\mathrm{Tr}\left[\mathbf{W}_i\mathbf{G}_i\mathbf{F}\mathbf{R}_{n_r}\mathbf{F}^H\mathbf{G}_i^H\mathbf{W}_i^H\right]\right\} \\ &\quad - \mathbb{E}_{\mathbf{E}_{gi}, \mathbf{E}_{hj}}\left\{2\mathrm{Re}\left\{\mathrm{Tr}\left[\mathbf{W}_i\mathbf{G}_i\mathbf{F}\mathbf{H}_j\mathbf{P}_j\mathbf{R}_{s_j}\right]\right\}\right\} \\ &\quad + \mathbb{E}_{\mathbf{E}_{hi}, \mathbf{E}_{gi}}\left\{\mathrm{Tr}\left[\mathbf{W}_i\mathbf{R}_{z_i}\mathbf{W}_i^H\right]\right\} \\ &\quad + \mathrm{Tr}(\mathbf{R}_{s_j}) + \mathrm{Tr}(\mathbf{W}_i\mathbf{R}_{n_i}\mathbf{W}_i^H). \end{aligned} \qquad (13)$$

Note that $\mathbf{G}_i$ and $\mathbf{H}_i$ are generally independent for FDD systems. However, in TDD systems, due to channel reciprocity, i.e., $\mathbf{G}_i = \mathbf{H}_i^H$, the backward and forward channels are not independent. Therefore, the results of (13) for FDD and TDD systems should be expected to be different. However, we make some efforts to formulate the problems for both FDD and TDD systems into a unified framework under the assumption that the error variance $\sigma_{e_i}^2$ is much smaller than one. With this assumption, we can approximate the mean squared error with some more tractable expressions. Otherwise, the problem would be more complicated. The accuracy of this approximation is examined through simulation in Sec. VI-A.

A. FDD Mode

We first deal with the term involving $\mathbf{R}_{z_i}$,

$$\begin{aligned} \mathbb{E}_{\mathbf{E}_{hi}, \mathbf{E}_{gi}}\left\{\mathrm{Tr}\left[\mathbf{W}_i\mathbf{R}_{z_i}\mathbf{W}_i^H\right]\right\} &= \\ \mathrm{Tr}\left[\mathbf{W}_i\mathbb{E}_{\mathbf{E}_{hi}, \mathbf{E}_{gi}}\{\mathbf{R}_{z_i}\}\mathbf{W}_i^H\right]. \end{aligned} \qquad (14)$$

Since $\mathbf{H}_i$ and $\mathbf{G}_i$ are estimated independently, $\mathbf{E}_{hi}$ and $\mathbf{E}_{gi}$ are independent. Hence, we have

$$
\begin{aligned}
\mathbb{E}_{\mathbf{E}_{hi},\mathbf{E}_{gi}}\{\mathbf{R}_{z_i}\} &= \mathbb{E}_{\mathbf{E}_{hi},\mathbf{E}_{gi}}\{\hat{\mathbf{G}}_i\mathbf{F}\mathbf{E}_{hi}\mathbf{R}_{x_i}\mathbf{E}_{hi}^H\mathbf{F}^H\hat{\mathbf{G}}_i^H\} \\
&\quad + \underbrace{\mathbb{E}_{\mathbf{E}_{hi},\mathbf{E}_{gi}}\{2\mathrm{Re}\{\hat{\mathbf{G}}_i\mathbf{F}\mathbf{E}_{hi}\mathbf{R}_{x_i}\hat{\mathbf{H}}_i^H\mathbf{F}^H\mathbf{E}_{gi}^H\}\}}_{0} \\
&\quad + \underbrace{\mathbb{E}_{\mathbf{E}_{hi},\mathbf{E}_{gi}}\{2\mathrm{Re}\{\hat{\mathbf{G}}_i\mathbf{F}\mathbf{E}_{hi}\mathbf{R}_{x_i}\mathbf{E}_{hi}^H\mathbf{F}^H\mathbf{E}_{gi}^H\}\}}_{0} \\
&\quad + \mathbb{E}_{\mathbf{E}_{hi},\mathbf{E}_{gi}}\{\mathbf{E}_{gi}\mathbf{F}\hat{\mathbf{H}}_i\mathbf{R}_{x_i}\hat{\mathbf{H}}_i^H\mathbf{F}^H\mathbf{E}_{gi}^H\} \\
&\quad + \underbrace{\mathbb{E}_{\mathbf{E}_{hi},\mathbf{E}_{gi}}\{2\mathrm{Re}\{\mathbf{E}_{gi}\mathbf{F}\hat{\mathbf{H}}_i\mathbf{R}_{x_i}\mathbf{E}_{hi}^H\mathbf{F}^H\mathbf{E}_{gi}^H\}\}}_{0} \\
&\quad + \mathbb{E}_{\mathbf{E}_{hi},\mathbf{E}_{gi}}\{\mathbf{E}_{gi}\mathbf{F}\mathbf{E}_{hi}\mathbf{R}_{x_i}\mathbf{E}_{hi}^H\mathbf{F}^H\mathbf{E}_{gi}^H\}. \quad (15)
\end{aligned}
$$

The last term in (15) is a second order term of the estimation error variance $\sigma_{e_i}^2$, and thus can be ignored as $\sigma_{e_i}^2$ is assumed to be much smaller than one. Hence, we have

$$
\begin{aligned}
&\mathbb{E}_{\mathbf{E}_{hi},\mathbf{E}_{gi}}\{\mathbf{R}_{z_i}\} \\
&\approx \hat{\mathbf{G}}_i\mathbf{F}\mathbf{R}_{hi}^{\frac{1}{2}}\mathbb{E}_{\mathbf{E}_{hi}}\left\{\mathbf{E}_{hw_i}\mathbf{T}_{hi}^{\frac{1}{2}}\mathbf{R}_{x_i}\mathbf{T}_{hi}^{\frac{1}{2}}\mathbf{E}_{hw_i}^H\right\}\mathbf{R}_{hi}^{\frac{1}{2}}\mathbf{F}^H\hat{\mathbf{G}}_i^H \\
&\quad + \mathbf{R}_{gi}^{\frac{1}{2}}\mathbb{E}_{\mathbf{E}_{gi}}\left\{\mathbf{E}_{gw_i}\mathbf{T}_{gi}^{\frac{1}{2}}\mathbf{F}\hat{\mathbf{H}}_i\mathbf{R}_{x_i}\hat{\mathbf{H}}_i^H\mathbf{F}^H\mathbf{T}_{gi}^{\frac{1}{2}}\mathbf{E}_{gw_i}^H\right\}\mathbf{R}_{gi}^{\frac{1}{2}} \\
&= \hat{\mathbf{G}}_i\mathbf{F}\mathrm{Tr}(\mathbf{R}_{x_i}\hat{\mathbf{T}}_{hi})\mathbf{R}_{hi}\mathbf{F}^H\hat{\mathbf{G}}_i^H \\
&\quad + \mathrm{Tr}(\mathbf{F}\hat{\mathbf{H}}_i\mathbf{R}_{x_i}\hat{\mathbf{H}}_i^H\mathbf{F}^H\hat{\mathbf{T}}_{gi})\mathbf{R}_{gi}, \quad (16)
\end{aligned}
$$

where for notational convenience, we have denoted $\hat{\mathbf{T}}_{hi} \triangleq \sigma_{e_i}^2\mathbf{T}_{hi}$ and $\hat{\mathbf{T}}_{gi} \triangleq \sigma_{e_i}^2\mathbf{T}_{gi}$.

Similarly, we can calculate the other terms in (13) and obtain the following results

$$
\begin{aligned}
&\mathbb{E}_{\mathbf{E}_{gi},\mathbf{E}_{hj}}\left\{\mathrm{Tr}\left[\mathbf{W}_i\mathbf{G}_i\mathbf{F}\mathbf{H}_j\mathbf{R}_{x_j}(\mathbf{W}_i\mathbf{G}_i\mathbf{F}\mathbf{H}_j)^H\right]\right\} \\
&= \mathrm{Tr}\left[\mathbf{W}_i(\mathrm{Tr}(\mathbf{F}\mathbf{\Phi}_j\mathbf{F}^H\hat{\mathbf{T}}_{hi})\mathbf{R}_{hi} + \hat{\mathbf{G}}_i^H\mathbf{F}\mathbf{\Phi}_j\mathbf{F}^H\hat{\mathbf{G}}_i^H)\mathbf{W}_i^H\right]
\end{aligned}
$$

$$(17)$$

$$
\begin{aligned}
&\mathbb{E}_{\mathbf{E}_{gi}}\left\{\mathrm{Tr}\left[\mathbf{W}_i\mathbf{G}_i\mathbf{F}\mathbf{R}_{n_r}\mathbf{F}^H\mathbf{G}_i^H\mathbf{W}_i^H\right]\right\} \\
&= \mathrm{Tr}\left[\mathbf{W}_i(\mathrm{Tr}(\mathbf{F}\mathbf{R}_{n_r}\mathbf{F}^H\hat{\mathbf{T}}_{gi})\mathbf{R}_{gi} + \hat{\mathbf{G}}_i\mathbf{F}\mathbf{R}_{n_r}\mathbf{F}^H\hat{\mathbf{G}}_i^H)\mathbf{W}_i^H\right]
\end{aligned}
$$

$$(18)$$

$$
\begin{aligned}
&\mathbb{E}_{\mathbf{E}_{hj},\mathbf{E}_{gi}}\left\{2\mathrm{Re}\left\{\mathrm{Tr}(\mathbf{W}_i\mathbf{G}_i\mathbf{F}\mathbf{H}_j\mathbf{P}_j\mathbf{R}_{s_j})\right\}\right\} \\
&= 2\mathrm{Re}\left\{\mathrm{Tr}(\mathbf{W}_i\hat{\mathbf{G}}_i\mathbf{F}\hat{\mathbf{H}}_j\mathbf{P}_j\mathbf{R}_{s_j})\right\} \quad (19)
\end{aligned}
$$

where

$$
\mathbf{\Phi}_j \triangleq \mathbb{E}_{\mathbf{E}_{hj}}\left\{\mathbf{H}_j\mathbf{R}_{x_j}\mathbf{H}_j^H\right\}\mathrm{Tr}(\mathbf{R}_{x_j}\hat{\mathbf{T}}_{hj})\mathbf{R}_{hj} + \hat{\mathbf{H}}_j\mathbf{R}_{x_j}\hat{\mathbf{H}}_j^H. \quad (20)
$$

Substituting the above equations into (13), we finally have

$$
\begin{aligned}
\hat{J}_i(\mathbf{P}_i,\mathbf{P}_j,\mathbf{F},\mathbf{W}_i) &= \mathrm{Tr}\left[\mathbf{W}_i\left(\hat{\mathbf{G}}_i\mathbf{F}\mathbf{\Psi}_i\mathbf{F}^H\hat{\mathbf{G}}_i^H + \mathbf{\Omega}_i\right)\mathbf{W}_i^H\right] \\
&\quad - 2\mathrm{Re}\left\{\mathrm{Tr}(\mathbf{W}_i\hat{\mathbf{G}}_i\mathbf{F}\hat{\mathbf{H}}_j\mathbf{P}_j\mathbf{R}_{s_j})\right\} + \mathrm{Tr}(\mathbf{R}_{s_j}), \quad (21)
\end{aligned}
$$

where $\mathbf{\Psi}_i \triangleq \mathbf{\Phi}_j + \mathrm{Tr}(\mathbf{R}_{x_i}\hat{\mathbf{T}}_{hi})\mathbf{R}_{hi} + \mathbf{R}_{n_r}$ and

$$
\mathbf{\Omega}_i \triangleq \mathrm{Tr}[\mathbf{F}(\mathbf{\Phi}_j + \mathbf{R}_{n_r} + \hat{\mathbf{H}}_i\mathbf{R}_{x_i}\hat{\mathbf{H}}_i^H)\mathbf{F}^H\hat{\mathbf{T}}_{gi}]\mathbf{R}_{gi} + \mathbf{R}_{n_i}. \quad (22)
$$

The problem of robust joint design of $\mathbf{P}_i$, $\mathbf{F}$, and $\mathbf{W}_i$, $\forall i = 1, 2$, to minimize the average sum MSE of $\mathcal{S}_1$ and $\mathcal{S}_2$ can be formulated as

$$
\min_{\mathbf{F},\mathbf{P}_i,\mathbf{W}_i,\forall i=1,2} \quad \hat{J} = \sum_{i=1}^{2}\hat{J}_i(\mathbf{P}_i,\mathbf{P}_j,\mathbf{F},\mathbf{W}_i) \quad (23a)
$$

$$
\mathrm{s.t.} \quad \mathrm{Tr}(\mathbf{P}_i\mathbf{R}_{s_i}\mathbf{P}_i^H) \leq P_{s_i}, \forall i = 1, 2 \quad (23b)
$$

$$
\mathrm{Tr}(\mathbf{F}\mathbf{R}_r\mathbf{F}^H) \leq P_r, \quad (23c)
$$

where

$$
\mathbf{R}_r \triangleq \mathbb{E}_{\mathbf{E}_{h1},\mathbf{E}_{h2}}\left\{\mathbb{E}\{\mathbf{r}\mathbf{r}^H\}\right\} = \mathbf{\Phi}_1 + \mathbf{\Phi}_2 + \mathbf{R}_{n_r}. \quad (24)
$$

Note that since only estimated CSI is available, the average transmit power constraint at $\mathcal{R}$ is considered here.

B. TDD Mode

According to channel reciprocity, $\mathbf{G}_i = \mathbf{H}_i^H$, correspondingly $\hat{\mathbf{G}}_i = \hat{\mathbf{H}}_i^H$ and $\mathbf{E}_{gi} = \mathbf{E}_{hi}^H$. The residual self-interference becomes $\mathbf{z}_i = (\mathbf{H}_i^H\mathbf{F}\mathbf{H}_i - \hat{\mathbf{H}}_i^H\mathbf{F}\hat{\mathbf{H}}_i)\mathbf{x}_i$. A close examination on (13) reveals that after replacing $\mathbf{G}_i$ with $\mathbf{H}_i^H$ only the forth term, $\mathbb{E}_{\mathbf{E}_{hi},\mathbf{E}_{gi}}\left\{\mathrm{Tr}\left[\mathbf{W}_i\mathbf{R}_{z_i}\mathbf{W}_i^H\right]\right\} = \mathbb{E}_{\mathbf{E}_{hi}}\left\{\mathrm{Tr}\left[\mathbf{W}_i\mathbf{R}_{z_i}\mathbf{W}_i^H\right]\right\}$, needs to be recalculated. The other terms are essentially the same as FDD systems since $\mathbf{H}_1$ and $\mathbf{H}_2$ are still independent, and their results can be easily derived from (17) to (19) with the following substitution

$$
\hat{\mathbf{G}}_i = \hat{\mathbf{H}}_i^H, \mathbf{R}_{gi} = \hat{\mathbf{T}}_{hi}, \hat{\mathbf{T}}_{gi} = \mathbf{R}_{hi}. \quad (25)
$$

We next calculate $\mathbb{E}_{\mathbf{E}_{hi}}\{\mathbf{R}_{z_i}\}$ to compute the forth term in (13). Estimation error variance $\sigma_{e_i}^2$ is assumed to be much smaller than one, and thus the terms with order of $\mathbf{E}_{hi}$ higher than two are negligible, i.e.,

$$
\begin{aligned}
\mathbb{E}_{\mathbf{E}_{hi}}\{\mathbf{R}_{z_i}\} &\approx \mathbb{E}_{\mathbf{E}_{hi}}\left\{\hat{\mathbf{H}}_i^H\mathbf{F}\mathbf{E}_{hi}\mathbf{R}_{x_i}\mathbf{E}_{hi}^H\mathbf{F}^H\hat{\mathbf{H}}_i\right\} \\
&\quad + \mathbb{E}_{\mathbf{E}_{hi}}\left\{2\mathrm{Re}\left\{\hat{\mathbf{H}}_i^H\mathbf{F}\mathbf{E}_{hi}\mathbf{R}_{x_i}\hat{\mathbf{H}}_i\mathbf{F}^H\mathbf{E}_{hi}\right\}\right\} \\
&\quad + \mathbb{E}_{\mathbf{E}_{hi}}\left\{\mathbf{E}_{hi}^H\mathbf{F}\hat{\mathbf{H}}_i\mathbf{R}_{x_i}\hat{\mathbf{H}}_i^H\mathbf{F}^H\mathbf{E}_{hi}\right\}. \quad (26)
\end{aligned}
$$

For any ZMCSCG entry $e = a+jb$ of $\mathbf{E}_{hw_i}$, we have $\mathbb{E}\{e^2\} = \mathbb{E}\{a^2\} - \mathbb{E}\{b^2\} + 2j\mathbb{E}\{ab\} = 0$ [27]. It follows that

$$
\begin{aligned}
&\mathbb{E}_{\mathbf{E}_{hi}}\left\{\hat{\mathbf{H}}_i^H\mathbf{F}\mathbf{E}_{hi}\mathbf{R}_{x_i}\hat{\mathbf{H}}_i\mathbf{F}^H\mathbf{E}_{hi}\right\} = \\
&\hat{\mathbf{H}}_i^H\mathbf{F}\mathbf{R}_{hi}^{\frac{1}{2}}\mathbb{E}_{\mathbf{E}_{hi}}\left\{\mathbf{E}_{hw_i}\mathbf{T}_{hi}^{\frac{1}{2}}\mathbf{R}_{x_i}\hat{\mathbf{H}}_i^H\mathbf{F}^H\mathbf{R}_{hi}^{\frac{1}{2}}\mathbf{E}_{hw_i}\right\}\mathbf{T}_{hi}^{\frac{1}{2}} = \mathbf{0}.
\end{aligned}
$$

Hence, the second term in (26) is zero. Note that the result holds here because we have applied the conjugate operation at relay node $\mathcal{R}$ as described in Sec. II-B. If no conjugate operation is applied there, $\mathbf{G}_i = \mathbf{H}_i^T$ and the second term in (26) should be changed as

$$
\begin{aligned}
&\mathbb{E}_{\mathbf{E}_{hi}}\left\{2\mathrm{Re}\left\{\hat{\mathbf{H}}_i^H\mathbf{F}\mathbf{E}_{hi}\mathbf{R}_{x_i}\hat{\mathbf{H}}_i\mathbf{F}^H\mathbf{E}_{hi}\right\}\right\} \rightarrow \\
&\mathbb{E}_{\mathbf{E}_{hi}}\left\{2\mathrm{Re}\left\{\hat{\mathbf{H}}_i^T\mathbf{F}\mathbf{E}_{hi}\mathbf{R}_{x_i}\hat{\mathbf{H}}_i\mathbf{F}^H\mathbf{E}_{hi}^*\right\}\right\},
\end{aligned}
$$

which would not be zero and thus increases the power level of residual self-interference $\mathbf{z}_i$. From (26), we now have

$$
\begin{aligned}
\mathbb{E}_{\mathbf{E}_{hi}}\{\mathbf{R}_{z_i}\} &\approx \hat{\mathbf{H}}_i^H\mathbf{F}\mathrm{Tr}(\mathbf{R}_{x_i}\hat{\mathbf{T}}_{hi})\mathbf{R}_{hi}\mathbf{F}^H\hat{\mathbf{H}}_i \\
&\quad + \mathrm{Tr}(\mathbf{F}\hat{\mathbf{H}}_i\mathbf{R}_{x_i}\hat{\mathbf{H}}_i^H\mathbf{F}^H\mathbf{R}_{hi})\hat{\mathbf{T}}_{hi}. \quad (27)
\end{aligned}
$$

Comparing (27) with (16), we find that (27) can be obtained from (16) using the same substitution as in (25).

Therefore, we have shown that the objective function (23a)

can be generalized to TDD systems following (25) when a conjugate operation is applied to signals at the relay node. Furthermore, the power constraints for TDD systems are the same as FDD systems, and can be exactly represented by (23b) and (23c). Hence, the problem formulation in (23) serves as a unified framework. In the following, we solve (23) and derive solutions that can be applied to both FDD and TDD systems.

IV. ITERATIVE JOINT OPTIMIZATION ALGORITHM

The problem in (23) is a nonconvex problem, and thus is difficult to solve. To make the problem tractable, we first decouple the original problem into three sub-problems where we only solve for part of the variables with others fixed. Since each sub-problem is convex, the optimal solution can be efficiently obtained by using optimization tool package or the derived closed-form solutions. After solving those sub-problems, we propose an iterative algorithm to jointly optimize $\mathbf{P}_i$, $\mathbf{F}$, and $\mathbf{W}_i$, $\forall i = 1, 2$. A brief convergence analysis for the iterative algorithm is also provided.

A. Optimization of $\mathbf{P}_i$

The sub-problem regarding $\mathbf{P}_i$ for fixed $\mathbf{F}$ and $\mathbf{W}_i$, $\forall i = 1, 2$, is given as

$$\begin{aligned} \min_{\mathbf{P}_i, \forall i = 1, 2} \quad & \hat{J} = \sum_{i=1}^2 \hat{J}_i(\mathbf{P}_1, \mathbf{P}_2) \\ \text{s.t.} \quad & \mathrm{Tr}(\mathbf{P}_i \mathbf{R}_{s_i} \mathbf{P}_i^H) \le P_{s_i}, \forall i = 1, 2, \\ & \mathrm{Tr}(\mathbf{F} \mathbf{R}_r \mathbf{F}^H) \le P_r. \end{aligned} \quad (28)$$

From (21), we can rewrite the objective function as follows

$$\begin{aligned} \hat{J} &= \sum_{i=1}^2 \hat{J}_i(\mathbf{P}_1, \mathbf{P}_2) \\ &= \sum_{i=1}^2 \Big\{ \mathrm{Tr}(\mathbf{W}_i \hat{\mathbf{G}}_i \mathbf{F} \mathbf{R}_{hj} \mathbf{F}^H \hat{\mathbf{G}}_i^H \mathbf{W}_i^H) \mathrm{Tr}(\mathbf{P}_j \mathbf{R}_{s_j} \mathbf{P}_j^H \hat{\mathbf{T}}_{hj}) \\ &\quad + \mathrm{Tr}(\mathbf{W}_i \hat{\mathbf{G}}_i \mathbf{F} \mathbf{R}_{hi} \mathbf{F}^H \hat{\mathbf{G}}_i^H \mathbf{W}_i^H) \mathrm{Tr}(\mathbf{P}_i \mathbf{R}_{s_i} \mathbf{P}_i^H \hat{\mathbf{T}}_{hi}) \\ &\quad + \mathrm{Tr}(\mathbf{P}_j \mathbf{R}_{s_j} \mathbf{P}_j^H \hat{\mathbf{H}}_j^H \mathbf{F}^H \hat{\mathbf{G}}_i^H \mathbf{W}_i^H \mathbf{W}_i \hat{\mathbf{G}}_i \mathbf{F} \hat{\mathbf{H}}_j) \\ &\quad + \mathrm{Tr}(\mathbf{F} \mathbf{R}_{hj} \mathbf{F}^H \hat{\mathbf{T}}_{gi}) \mathrm{Tr}(\mathbf{W}_i \mathbf{R}_{gi} \mathbf{W}_i^H) \mathrm{Tr}(\mathbf{P}_j \mathbf{R}_{s_j} \mathbf{P}_j^H \hat{\mathbf{T}}_{hj}) \\ &\quad + \mathrm{Tr}(\mathbf{W}_i \mathbf{R}_{gi} \mathbf{W}_i^H) \mathrm{Tr}(\mathbf{P}_i \mathbf{R}_{s_i} \mathbf{P}_i^H \hat{\mathbf{H}}_i^H \mathbf{F}^H \hat{\mathbf{T}}_{gi}) \\ &\quad + \mathrm{Tr}(\mathbf{W}_i \mathbf{R}_{gi} \mathbf{W}_i^H) \mathrm{Tr}(\mathbf{P}_j \mathbf{R}_{s_j} \mathbf{P}_j^H \hat{\mathbf{H}}_j^H \mathbf{F}^H \hat{\mathbf{T}}_{gi}) \\ &\quad - 2\mathrm{Re}\Big\{ \mathrm{Tr}(\mathbf{W}_i \hat{\mathbf{G}}_i \mathbf{F} \hat{\mathbf{H}}_j \mathbf{P}_j \mathbf{R}_{s_j}) \Big\} \\ &\quad + \mathrm{Tr}(\mathbf{W}_i \hat{\mathbf{G}}_i \mathbf{F} \mathbf{R}_{n_r} \mathbf{F}^H \hat{\mathbf{G}}_i^H \mathbf{W}_i^H) + \mathrm{Tr}(\mathbf{F} \mathbf{R}_{n_r} \mathbf{F}^H \hat{\mathbf{T}}_{gi}) \\ &\quad + \mathrm{Tr}(\mathbf{W}_i \mathbf{R}_{n_i} \mathbf{W}_i^H) + \mathrm{Tr}(\mathbf{R}_{s_j}) \Big\} \quad (29) \end{aligned}$$

We also rewrite the power constraint at relay as

$$\begin{aligned} \mathrm{Tr}(\mathbf{F} \mathbf{R}_r \mathbf{F}^H) &= \sum_{i=1}^2 \Big\{ \mathrm{Tr}(\mathbf{F} \mathbf{R}_{hi} \mathbf{F}^H) \mathrm{Tr}(\mathbf{P}_i \mathbf{R}_{s_i} \mathbf{P}_i^H \hat{\mathbf{T}}_{hi}) \\ &\quad + \mathrm{Tr}(\mathbf{P}_i \mathbf{R}_{s_i} \mathbf{P}_i^H \hat{\mathbf{H}}_i^H \mathbf{F}^H \mathbf{F} \hat{\mathbf{H}}_i) \Big\} \\ &\quad + \mathrm{Tr}(\mathbf{F} \mathbf{R}_{n_r} \mathbf{F}^H) \le P_r. \quad (30) \end{aligned}$$

It is easy to verify that both (29) and (30) are jointly convex in $\mathbf{P}_1$ and $\mathbf{P}_2$. Indeed, the sub-problem (28) can be transformed into a quadratically constrained quadratic programming (QCQP) problem [28]. Hence, the convex optimization package CVX [29] can be used to solve (28).

B. Optimization of $\mathbf{F}$

Given $\mathbf{P}_i$ and $\mathbf{W}_i$, $i = 1, 2$, we formulate the sub-problem for optimization of $\mathbf{F}$ as

$$\min_{\mathbf{F}} \quad \hat{J} = \sum_{i=1}^2 \hat{J}_i(\mathbf{F}) \quad \text{s.t.} \ \ \mathrm{Tr}(\mathbf{F} \mathbf{R}_r \mathbf{F}^H) \le P_r. \quad (31)$$

It can be verified that (31) is convex with respect to $\mathbf{F}$. The Lagrangian function of (31) is given as $L(\mathbf{F}, \lambda) = \hat{J} + \lambda(\mathrm{Tr}(\mathbf{F} \mathbf{R}_r \mathbf{F}^H) - P_r)$, where $\lambda \ge 0$ is the lagrangian multiplier. Its derivative with respect to $\mathbf{F}^*$ is

$$\begin{aligned} \frac{\partial L}{\partial \mathbf{F}^*} &= \sum_{i=1}^2 \hat{\mathbf{G}}_i^H \mathbf{W}_i^H \mathbf{W}_i \hat{\mathbf{G}}_i \mathbf{F} \mathbf{\Psi}_i - \sum_{i=1}^2 \hat{\mathbf{G}}_i^H \mathbf{W}_i^H \mathbf{R}_{s_j} \mathbf{P}_j^H \hat{\mathbf{H}}_j^H \\ &\quad + \sum_{i=1}^2 \mathrm{Tr}(\mathbf{W}_i \mathbf{R}_{gi} \mathbf{W}_i^H) \hat{\mathbf{T}}_{gi} \mathbf{F} \mathbf{\Pi}_j + \lambda \mathbf{F} \mathbf{R}_r, \quad (32) \end{aligned}$$

where $\mathbf{\Pi}_j \triangleq \mathbf{\Phi}_j + \hat{\mathbf{H}}_i \mathbf{R}_{x_i} \hat{\mathbf{H}}_i^H + \mathbf{R}_{n_r}$. By setting $\frac{\partial L}{\partial \mathbf{F}^*} = \mathbf{0}$, we have

$$\mathrm{vec}(\mathbf{F}) = \mathbf{P}^{-1} \mathrm{vec}(\mathbf{Q}), \quad (33)$$

where $\mathbf{Q} \triangleq \sum_{i=1}^2 \hat{\mathbf{G}}_i^H \mathbf{W}_i^H \mathbf{R}_{s_j} \mathbf{P}_j^H \hat{\mathbf{H}}_j^H$ and

$$\begin{aligned} \mathbf{P} \triangleq \sum_{i=1}^2 \Big\{ & \mathbf{\Psi}_i^T \otimes (\hat{\mathbf{G}}_i^H \mathbf{W}_i^H \mathbf{W}_i \hat{\mathbf{G}}_i) \\ & + \mathbf{\Pi}_j^T \otimes (\mathrm{Tr}(\mathbf{W}_i \mathbf{R}_{gi} \mathbf{W}_i^H) \hat{\mathbf{T}}_{gi}) \Big\} + \lambda \mathbf{R}_r^T \otimes \mathbf{I}_N. \quad (34) \end{aligned}$$

The optimal λ should satisfy the following Karush-Kuhn-Tucker (KKT) conditions [28],

$$\lambda(\mathrm{Tr}(\mathbf{F} \mathbf{R}_r \mathbf{F}^H) - P_r) = 0, \quad (35)$$

$$\mathrm{Tr}(\mathbf{F} \mathbf{R}_r \mathbf{F}^H) \le P_r. \quad (36)$$

In Appendix A, we show that the optimal λ can be bounded as follows

$$0 \le \lambda \le \sqrt{\frac{\lambda_1(\mathbf{R}_r) \lambda_1(\mathrm{vec}(\mathbf{Q}) \mathrm{vec}(\mathbf{Q})^H)}{P_r}} \mathrm{Tr}((\mathbf{R}_r^T \otimes \mathbf{I}_N)^{-1}),$$

where $\lambda_1(\mathbf{A})$ denotes the maximum eigenvalue of matrix $\mathbf{A}$. Then we can conveniently search for the optimal λ using bisection method within the bounds on λ.

C. Optimization of $\mathbf{W}_i$

Observing that the objective function of (23) is convex with respect to $\mathbf{W}_i$, $\forall i = 1, 2$, and $\mathbf{W}_i$ is not involved in any constraints, hence, by calculating the derivative of $\hat{J}$ with respect to $\mathbf{W}_i$ and setting it to zero, i.e. $\frac{\partial \hat{J}}{\partial \mathbf{W}_i} = \mathbf{0}$, we can easily obtain the optimal receiver $\mathbf{W}_i$ at S_i as

$$\mathbf{W}_i = \mathbf{R}_{s_j} \mathbf{P}_j^H \hat{\mathbf{H}}_j^H \mathbf{F}^H \hat{\mathbf{G}}_i^H (\hat{\mathbf{G}}_i \mathbf{F} \mathbf{\Psi}_i \mathbf{F}^H \hat{\mathbf{G}}_i^H + \mathbf{\Omega}_i)^{-1}. \quad (37)$$

D. Iterative Optimization

After characterizing and solving the above sub-problems, we iteratively optimize $\mathbf{P}_i$, $\mathbf{F}$, and $\mathbf{W}_i$, $\forall i = 1, 2$, as in the algorithm below to get the final solutions.

Algorithm 1: Iterative Algorithm for Joint Optimization

- **Initialization**: The source precoders $\mathbf{P}_i$, $\forall i = 1, 2$, and relay precoder $\mathbf{F}$ are initialized as $\mathbf{P}_i^{(0)} =$

$\sqrt{P_{s_i}/\mathrm{Tr}(\mathbf{R}_{s_i})}\mathbf{I}_M$ and $\mathbf{F}^{(0)} = \sqrt{P_r/\mathrm{Tr}(\mathbf{R}_r)}\mathbf{I}_N$, respectively. Compute $\mathbf{W}_i^{(0)}$, $\forall i = 1, 2$, using (37). n = 1;

- **Iteration**:
  (a) Solve for $\mathbf{P}_i^{(n)}$ in sub-problem (28) with $\mathbf{F} = \mathbf{F}^{(n-1)}$ and $\mathbf{W}_i = \mathbf{W}_i^{(n-1)}$, $\forall i = 1, 2$;
  (b) Compute $\mathbf{F}^{(n)}$ using (33) with $\mathbf{P}_i = \mathbf{P}_i^{(n)}$ and $\mathbf{W}_i = \mathbf{W}_i^{(n-1)}$, $\forall i = 1, 2$;
  (c) Compute $\mathbf{W}_i^{(n)}$ using (37) with $\mathbf{P}_i = \mathbf{P}_i^{(n)}$ and $\mathbf{F} = \mathbf{F}^{(n)}$, $\forall i = 1, 2$;
  (d) $n = n + 1$;

- **Termination**: The algorithm terminates when $\frac{\|\mathbf{F}^{(n)} - \mathbf{F}^{(n-1)}\|_F^2}{\|\mathbf{F}^{(n-1)}\|_F^2} \leq \delta$ or $n \geq N_{\max}$, where δ is a predefined threshold and $N_{\max}$ is the maximum number of iterations allowed.

In the iteration process, each step updates the concerned variable to a new optimal value by solving a convex sub-problem. Note that all sub-problems have the same objective function and the solution of each sub-problem does not violate the power constraints of the other sub-problems. We conclude that the sum MSE is decreased in each iteration. As the sum MSE is lower bounded by zero, the proposed iterative algorithm converges at least to some local minimum.

The performance of the iterative algorithm can be improved by running the algorithm several times with different random initialization matrices and then choosing the best result among them. However, this would significantly increase the computational complexity. We have shown in Sec. VI-B that using scaled identity initialization matrices provides better average performance than random initialization.

V. CONSTRAINED STRUCTURE ALGORITHM

Since the iterative algorithm could take several iterations to converge, its complexity is high. Hence, in the following we develop a low complexity algorithm by only optimizing the relay precoder with the source precoders $\mathbf{P}_i$, $\forall i = 1, 2$, fixed. Moreover, we propose a structure for the relay precoder and derive a closed-form solution which can be efficiently computed.

Substituting (37) into (21), we have

$$
\begin{aligned}
\hat{J}_i(\mathbf{F}, \mathbf{W}_i) &= \mathrm{Tr}(\mathbf{R}_{s_j}) - \mathrm{Tr}\left[\mathbf{R}_{s_j}\mathbf{P}_j^H\hat{\mathbf{H}}_j^H\mathbf{F}^H\hat{\mathbf{G}}_i^H \right. \\
&\quad \left. (\hat{\mathbf{G}}_i\mathbf{F}\boldsymbol{\Psi}_i\mathbf{F}^H\hat{\mathbf{G}}_i^H + \boldsymbol{\Omega}_i)^{-1}\hat{\mathbf{G}}_i\mathbf{F}\hat{\mathbf{H}}_j\mathbf{P}_j\mathbf{R}_{s_j}\right] \\
&= \mathrm{Tr}(\mathbf{R}_{s_j}) - \mathrm{Tr}(\mathbf{R}_{s_j}\mathbf{P}_j^H\hat{\mathbf{H}}_j^H\boldsymbol{\Psi}_i^{-1}\hat{\mathbf{H}}_j\mathbf{P}_j\mathbf{R}_{s_j}) + \tilde{J}_i(\mathbf{F}),
\end{aligned}
\tag{38}
$$

where

$$
\tilde{J}_i(\mathbf{F}) \triangleq \mathrm{Tr}\left[\mathbf{R}_{s_j}\mathbf{P}_j^H\hat{\mathbf{H}}_j^H\boldsymbol{\Psi}_i^{-\frac{1}{2}}(\boldsymbol{\Psi}_i^{\frac{1}{2}}\mathbf{F}^H\hat{\mathbf{G}}_i^H\boldsymbol{\Omega}_i^{-1}\hat{\mathbf{G}}_i\mathbf{F}\boldsymbol{\Psi}_i^{\frac{1}{2}}\right.
$$
$$
\left. + \mathbf{I}_N)^{-1}\boldsymbol{\Psi}_i^{-\frac{1}{2}}\hat{\mathbf{H}}_j\mathbf{P}_j\mathbf{R}_{s_j}\right]. \tag{39}
$$

Observing that only the term $\tilde{J}_i(\mathbf{F})$ in (38) is related to $\mathbf{F}$, the problem in (23) is equivalent to minimizing $\tilde{J}_1(\mathbf{F}) + \tilde{J}_2(\mathbf{F})$. To further simplify the problem, we next consider minimizing an upper bound on $\tilde{J}_1(\mathbf{F}) + \tilde{J}_2(\mathbf{F})$. In Appendix B, we show

that $\tilde{J}_i(\mathbf{F}) \leq \tilde{J}_i^U(\mathbf{F})$, where

$$
\tilde{J}_i^U(\mathbf{F}) \triangleq \mathrm{Tr}\left[\mathbf{R}_{s_j}\mathbf{P}_j^H\hat{\mathbf{H}}_j^H\boldsymbol{\Psi}_i^{-\frac{1}{2}}(\boldsymbol{\Psi}_i^{\frac{1}{2}}\mathbf{F}^H\hat{\mathbf{G}}_i^H\bar{\boldsymbol{\Omega}}_i^{-1}\hat{\mathbf{G}}_i\mathbf{F}\boldsymbol{\Psi}_i^{\frac{1}{2}}\right.
$$
$$
\left. + \mathbf{I}_N)^{-1}\boldsymbol{\Psi}_i^{-\frac{1}{2}}\hat{\mathbf{H}}_j\mathbf{P}_j\mathbf{R}_{s_j}\right], \tag{40}
$$

and $\bar{\boldsymbol{\Omega}}_i \triangleq \lambda_1(\hat{\mathbf{T}}_{gi})P_r\mathbf{R}_{gi} + \mathbf{R}_{n_i}$. The relaxed problem is described as

$$
\begin{aligned}
\min_{\mathbf{F}} \quad & \tilde{J}_1^U(\mathbf{F}) + \tilde{J}_2^U(\mathbf{F}) \\
\text{s.t.} \quad & \mathrm{Tr}(\mathbf{F}\mathbf{R}_r\mathbf{F}^H) \leq P_r.
\end{aligned}
\tag{41}
$$

For one-way relay systems, it is shown in [8] and [20] that the optimal relay precoder should diagonalize the MSE matrix. However, in our problem, it is generally impossible to find such an $\mathbf{F}$ which could diagonalize the MSE matrices of both $\tilde{J}_1^U(\mathbf{F})$ and $\tilde{J}_2^U(\mathbf{F})$. We thus resort to some sub-optimal solutions with tractable structures. $\tilde{J}_i^U(\mathbf{F})$ can be rewritten as

$$
\tilde{J}_i^U(\mathbf{F}) = \mathrm{Tr}\left[\mathbf{A}_i(\boldsymbol{\Psi}_i^{\frac{1}{2}}\mathbf{F}^H\boldsymbol{\Theta}_i\mathbf{F}\boldsymbol{\Psi}_i^{\frac{1}{2}} + \mathbf{I}_N)^{-1}\right], \tag{42}
$$

where $\boldsymbol{\Theta}_i \triangleq \hat{\mathbf{G}}_i^H\bar{\boldsymbol{\Omega}}_i^{-1}\hat{\mathbf{G}}_i$ and $\mathbf{A}_i \triangleq \boldsymbol{\Psi}_i^{-\frac{1}{2}}\hat{\mathbf{H}}_j\mathbf{P}_j\mathbf{R}_{s_j}^2\mathbf{P}_j^H$ $\hat{\mathbf{H}}_j^H\boldsymbol{\Psi}_i^{-\frac{1}{2}}$. Substituting the eigenvalue decomposition $\mathbf{A}_i = \mathbf{U}_{A_i}\boldsymbol{\Lambda}_{A_i}\mathbf{U}_{A_i}^H$ into (42) gives

$$
\tilde{J}_i^U(\mathbf{F}) = \mathrm{Tr}\left[\boldsymbol{\Lambda}_{A_i}(\mathbf{B}_i^H\mathbf{F}^H\boldsymbol{\Theta}_i^{\frac{H}{2}}\boldsymbol{\Theta}_i^{\frac{1}{2}}\mathbf{F}\mathbf{B}_i + \mathbf{I}_N)^{-1}\right], \tag{43}
$$

where $\mathbf{B}_i \triangleq \boldsymbol{\Psi}_i^{\frac{1}{2}}\mathbf{U}_{A_i}$. We propose the following structure for the relay precoder

$$
\mathbf{F} = \mathbf{U}_\Theta\boldsymbol{\Lambda}_F\mathbf{U}_B^{-1}, \tag{44}
$$

where $\mathbf{U}_\Theta$ is a unitary matrix, $\mathbf{U}_B$ is a non-singular matrix, and $\boldsymbol{\Lambda}_F$ is an $N \times N$ diagonal matrix to be determined in the following. $\mathbf{U}_B$ is obtained by applying generalized singular value decomposition to $\mathbf{B}_1^H$ and $\mathbf{B}_2^H$ as $\mathbf{B}_i^H = \mathbf{V}_{B_i}\boldsymbol{\Lambda}_{B_i}\mathbf{U}_B^H$, where $\mathbf{V}_{B_i}$ is a unitary matrix and $\boldsymbol{\Lambda}_{B_i}$ is a diagonal matrix. $\mathbf{U}_\Theta$ is obtained from the following singular value decomposition (SVD)

$$
\begin{pmatrix} \boldsymbol{\Theta}_1^{\frac{1}{2}} \\ \boldsymbol{\Theta}_2^{\frac{1}{2}} \end{pmatrix} = \begin{pmatrix} \mathbf{V}_{\Theta_1} & \tilde{\mathbf{V}}_{\Theta_1} \\ \mathbf{V}_{\Theta_2} & \tilde{\mathbf{V}}_{\Theta_2} \end{pmatrix} \begin{pmatrix} \boldsymbol{\Lambda}_\Theta \\ \mathbf{0} \end{pmatrix} \mathbf{U}_\Theta^H, \tag{45}
$$

where $\mathbf{V}_{\Theta_i}$ and $\tilde{\mathbf{V}}_{\Theta_i}$ are $N \times N$ complex matrices, and $\boldsymbol{\Lambda}_\Theta$ is a $N \times N$ diagonal matrix. Note that the similar relay precoder structure has been used in [16] to parallelize the bidirectional channels in two-way relay system under perfect CSI. From (45), $\boldsymbol{\Theta}_i^{\frac{1}{2}} = \mathbf{V}_{\Theta_i}\boldsymbol{\Lambda}_\Theta\mathbf{U}_\Theta^H$. Substituting (44) into (43) gives

$$
\begin{aligned}
\tilde{J}_i^U(\mathbf{F}) &= \mathrm{Tr}\left[\boldsymbol{\Lambda}_{A_i}(\mathbf{V}_{B_i}\tilde{\boldsymbol{\Lambda}}\mathbf{V}_{\Theta_i}^H\mathbf{V}_{\Theta_i}\tilde{\boldsymbol{\Lambda}}\mathbf{V}_{B_i}^H + \mathbf{I}_N)^{-1}\right] \\
&= \mathrm{Tr}\left[\mathbf{C}_i(\tilde{\boldsymbol{\Lambda}}\mathbf{V}_{\Theta_i}^H\mathbf{V}_{\Theta_i}\tilde{\boldsymbol{\Lambda}} + \mathbf{I}_N)^{-1}\right]
\end{aligned}
\tag{46}
$$

where $\tilde{\boldsymbol{\Lambda}} \triangleq \boldsymbol{\Lambda}_{B_i}\boldsymbol{\Lambda}_F\boldsymbol{\Lambda}_\Theta$ and $\mathbf{C}_i \triangleq \mathbf{V}_{B_i}^H\boldsymbol{\Lambda}_{A_i}\mathbf{V}_{B_i}$. In order to achieve some simple solutions, we further relax the problem by only considering the main diagonal of $\mathbf{V}_{\Theta_i}^H\mathbf{V}_{\Theta_i}$ in (46). Finally, the problem in (41) is simplified to the following problem

$$
\begin{aligned}
\min_{\boldsymbol{\Lambda}_F} \quad & \sum_{i=1}^2 \mathrm{Tr}\left[\mathbf{C}_i(\mathbf{D}_{\Theta_i}\boldsymbol{\Lambda}_{B_i}^2\boldsymbol{\Lambda}_\Theta^2\boldsymbol{\Lambda}_F^2 + \mathbf{I}_N)^{-1}\right] \\
\text{s.t.} \quad & \mathrm{Tr}(\mathbf{D}\boldsymbol{\Lambda}_F^2) \leq P_r
\end{aligned}
\tag{47}
$$

where $\mathbf{D}_{\Theta_i} \triangleq \mathrm{diag}\{\mathbf{V}_{\Theta_i}^H \mathbf{V}_{\Theta_i}\}$ and $\mathbf{D} \triangleq \mathbf{U}_B^{-1} \mathbf{R}_r \mathbf{U}_B^{-H}$. The optimization problem (47) can be rewritten in scalar form as

$$\min_{p_1,\ldots,p_N} \quad \sum_{i=1}^{2} \sum_{n=1}^{N} \frac{c_{in}}{1 + k_{in} p_n}$$

$$\text{s.t.} \quad \sum_{n=1}^{N} d_n p_n \leq P_r, \tag{48}$$

$$p_n \geq 0, \forall n = 1, \ldots, N$$

where p_n, k_{in}, c_{in} and d_n, $\forall n = 1, \ldots, N$, are the diagonal entries of $\boldsymbol{\Lambda}_F^2$, $(\mathbf{D}_{\Theta_i} \boldsymbol{\Lambda}_{B_i}^2 \boldsymbol{\Lambda}_{\Theta}^2)$, $\mathbf{C}_i$, and $\mathbf{D}$, respectively. It is easy to recognize that (48) is a convex optimization problem. By using the KKT conditions we obtain its optimal solution as $p_n^{opt} \triangleq \max(0, r_n)$, where r_n is the maximum real root of the equation

$$\alpha_4^{(n)} x^4 + \alpha_3^{(n)} x^3 + \alpha_2^{(n)} x^2 + \alpha_1^{(n)} x + \alpha_0^{(n)} = 0 \tag{49}$$

with variable x and coefficients

$$
\begin{aligned}
\alpha_4^{(n)} &= \mu d_n k_{1n}^2 k_{2n}^2, \\
\alpha_3^{(n)} &= 2\mu d_n (k_{1n} k_{2n}^2 + k_{1n}^2 k_{2n}), \\
\alpha_2^{(n)} &= \mu d_n (k_{1n}^2 + k_{2n}^2 + 4 k_{1n} k_{2n}) \\
&\quad - c_{1n} k_{1n} k_{2n}^2 - c_{2n} k_{1n}^2 k_{2n}, \\
\alpha_1^{(n)} &= 2\mu d_n (k_{1n} + k_{2n}) - 2(c_{1n} + c_{2n}) k_{1n} k_{2n}, \\
\alpha_0^{(n)} &= \mu d_n - c_{1n} k_{1n} - c_{2n} k_{2n},
\end{aligned}
$$

where $\mu > 0$ and should be chosen such that $\sum_{n=1}^{N} d_n p_n^{opt} = P_r$. In Appendix C, we show that the optimal μ is upper bounded as

$$\mu \leq \max \left\{ \frac{c_{1n} k_{1n} + c_{2n} k_{2n}}{d_n}, \forall n = 1, \ldots, N \right\}. \tag{50}$$

Hence, we can again use the bisection method to find the optimal μ. After obtaining $\boldsymbol{\Lambda}_F$, we compute $\mathbf{F}$ and $\mathbf{W}_i$, $\forall i = 1, 2$, using (44) and (37).

The algorithm can be deployed at the relay node to compute $\mathbf{F}$ and $\mathbf{W}_i$, $\forall i = 1, 2$, and the relay then feed forward $\mathbf{W}_i$ to $\mathcal{S}_i$. The estimated channels and their correlation matrices should be known at relay. In practice, the estimated backward channel $\hat{\mathbf{H}}_i$ and its correlation matrices $\mathbf{R}_{hi}$ and $\mathbf{T}_{hi}$ can be locally obtained at the relay, whereas for the forward channel $\hat{\mathbf{G}}_i$ and its correlation matrices $\mathbf{R}_{gi}$ and $\mathbf{T}_{gi}$, if the system works in TDD mode, they can be easily obtained through channel reciprocity, otherwise source $\mathcal{S}_i$ should feedback such knowledge to the relay.

VI. SIMULATION RESULTS AND DISCUSSIONS

In this section, we provide simulation results to evaluate the performance of the proposed algorithms. FDD mode is assumed in our simulation. The channel correlation matrices of $\mathbf{H}_i$ and $\mathbf{G}_i$, $\forall i = 1, 2$, are generated using the exponential model [30]. The entries of $\mathbf{R}_{hi}$ and $\mathbf{T}_{hi}$ are given respectively by $\mathbf{R}_{hi}(k, l) = \beta^{|k-l|}$ and $\mathbf{T}_{hi}(p, q) = \alpha^{|p-q|}$, $\forall 1 \leq k, l \leq N$ and $1 \leq p, q \leq M$, where β and α are the row and column correlation coefficients, respectively. $\mathbf{R}_{gi}$ and $\mathbf{T}_{gi}$ are obtained similarly, and the same channel correlation coefficients are assumed for $\mathbf{G}_i$ with $\mathbf{R}_{gi}(p, q) = \beta^{|p-q|}$ and $\mathbf{T}_{gi}(k, l) =$
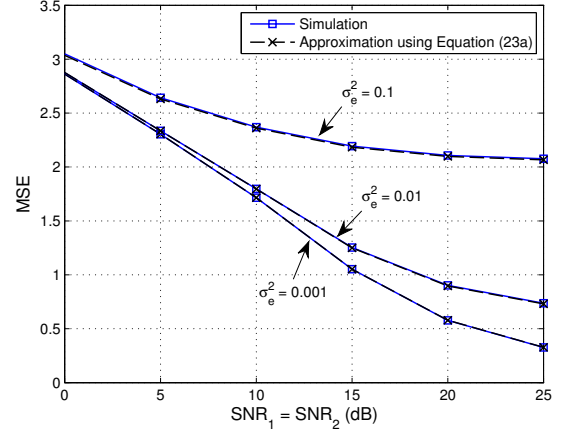

Fig. 2. Approximation of MSE. $N = M = 2$, $\alpha = \beta = 0.5$.

$\alpha^{|k-l|}$. The channel estimation error variances are given by $\sigma_{e_1}^2 = \sigma_{e_2}^2 = \sigma_e^2$.

The transmit data from both source nodes are normalized as $\mathbf{R}_{s_1} = \mathbf{R}_{s_2} = \mathbf{I}_M$. The receive noises at all nodes are complex white Gaussian with $\mathbf{R}_{n_r} = \sigma_r^2 \mathbf{I}_N$ and $\mathbf{R}_{n_1} = \mathbf{R}_{n_2} = \sigma_d^2 \mathbf{I}_M$. The transmit power budgets at both source nodes are assumed to be equal, i.e., $P_{s_1} = P_{s_2} = P_s$. The average *signal-to-noise ratio* (SNR) for the first phase and the second phase of the two-way communication protocol are defined as $\mathrm{SNR}_1 \triangleq \frac{P_s}{M\sigma_r^2}$ and $\mathrm{SNR}_2 \triangleq \frac{P_r}{N\sigma_d^2}$, respectively. All results are averaged over 2000 independent random channel realizations, and 1000 BPSK symbols are transmitted for each data stream per realization.

A. Accuracy of MSE Approximation

The MSE of the system is approximated with (23a). We now examine the accuracy of the approximation in Fig. 2, when $N = M = 2$ and $\alpha = \beta = 0.5$. $\mathbf{P}_i$, $\mathbf{F}$, and $\mathbf{W}_i$, $\forall i = 1, 2$ are computed using the proposed iterative algorithm. It can be seen that the approximation results fit very well with the simulation results even for σ_e^2 as large as 0.1.

B. Proposed Iterative Algorithm

The convergence behavior of the proposed iterative algorithm is shown in Fig. 3, where $N = M = 2$, $\alpha = \beta = 0.5$, and $\sigma_e^2 = 0.005$. It can be seen that identity initialization exhibits better convergence performance than random initialization.

We compare the robust joint MMSE (JMMSE) precoding in terms of *bit error rate* (BER) with the non-robust joint MMSE precoding scheme in [16] which does not take channel estimation errors into account. We also simulate the dual channel matching scheme in [11] as well as the conventional AF relay where the simple scaled identity matrix is used at relay to amplify the received signal. To illustrate the advantage of joint precoding, we also show the results of the robust relay MMSE precoding scheme, in which only the relay precoder $\mathbf{F}$ and destination receiver $\mathbf{W}_i$, $\forall i = 1, 2$, are optimized by using Algorithm 1 and skipping Step (a) in the iteration process, and the source precoders are fixed as $\mathbf{P}_i = \sqrt{P_{s_i}/M} \mathbf{I}_M$,
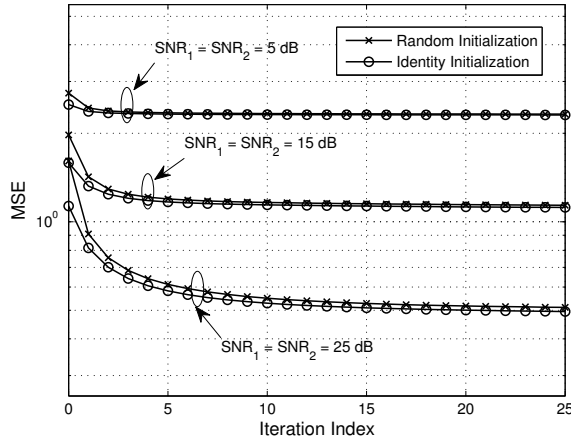
Fig. 3. Convergence behavior of the iterative algorithm. $N = M = 2$, $\alpha = \beta = 0.5$, $\sigma_e^2 = 0.005$.
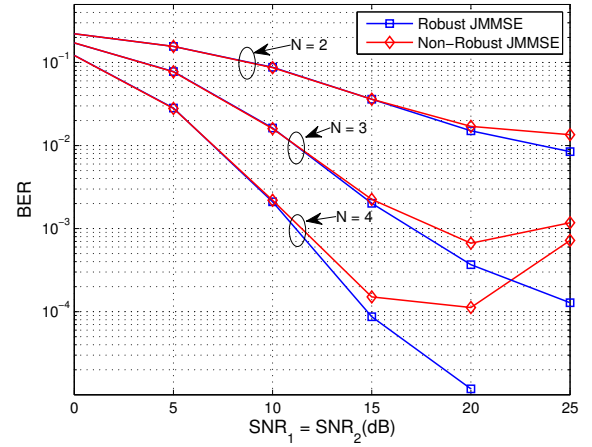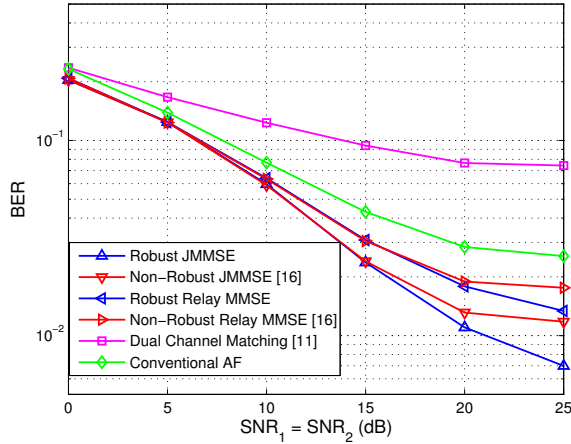

Fig. 4. BER performance comparison with uncorrelated channels. $\sigma_e^2 = 0.005$, $N = M = 2$, $\alpha = \beta = 0$.
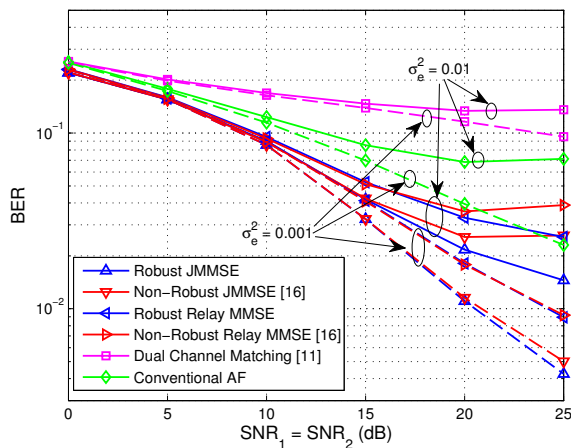

Fig. 5. BER performance comparison with different σ_e^2. $N = M = 2$, $\alpha = \beta = 0.5$.

$\forall i = 1, 2$. The non-robust relay MMSE precoding scheme is similarly derived from [16].

When the channels are uncorrelated, i.e., $\alpha = \beta = 0$, the BER performance is shown in Fig. 4 with $N = M = 2$

Fig. 6. BER performance comparison with different N. $M = 2$, $\sigma_e^2 = 0.005$, $\alpha = \beta = 0.5$.

and $\sigma_e^2 = 0.005$. The robust schemes show better robustness against channel errors than the non-robust schemes especially in median and high SNR regions. In Fig. 5, we take channel correlation into account by setting $\alpha = \beta = 0.5$, and show the results for $\sigma_e^2 = 0.001$ and 0.01. The superiority of the robust schemes becomes more significant when σ_e^2 increases. We can also see that the BER of the non-robust schemes increases with SNR in the high SNR region when the error variance is large. This is because part of transmit power contributes to noise due to channel mismatch and residual self-interference resulted from channel estimation errors, which is ignored in the non-robust schemes. Meanwhile, the robust joint precoding scheme always outperforms the robust relay precoding scheme. However, this performance gain is achieved at the extra cost of solving a QCQP problem with $2M^2$ complex variables in each iteration of Algorithm 1.

In Fig. 6, we further investigate the BER performance when the antenna number N at $\mathcal{R}$ changes from 2 to 4 with $M = 2$, $\sigma_e^2 = 0.005$, and $\alpha = \beta = 0.5$. It can be observed that the system performance significantly improves when N is greater than 2. This is because extra diversity gain can be exploited to enhance the reliability of data transmission. Moreover, the robust joint MMSE precoding scheme consistently overcomes the error floor problem occurrs in the non-robust precoding scheme, and the performance gap widens as N increases from 2 to 4.

C. Proposed Constrained Structure Algorithm

In Fig. 7, we compare the BER performance of the proposed constrained structure algorithm with the iterative algorithm for robust relay MMSE precoding when $M = N = 4$, $\alpha = \beta = 0.5$, and $\sigma_e^2 = 0.001$ or 0.01. In both algorithms, the source precoders are fixed as $\mathbf{P}_i = \sqrt{P_{s_i}/M}\mathbf{I}_M$, $\forall i = 1, 2$. We can see that the constrained structure algorithm achieves very close performance with the iterative algorithm under different σ_e^2. However, we would like to point out that due to the specific structure imposed on the relay precoder, when $M < N$ the performance of the constrained structure algorithm degrades severely, showing apparent diversity loss compared with the iterative algorithm.
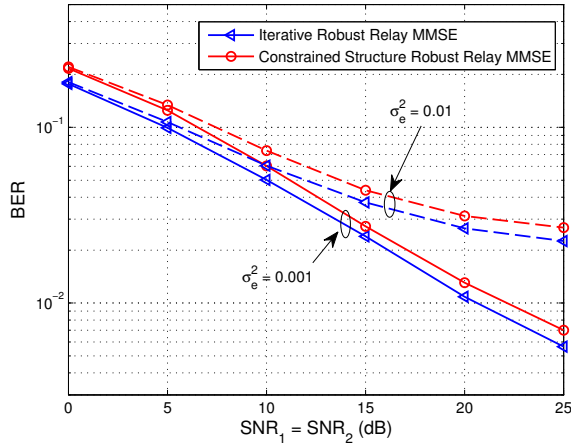
Fig. 7. BER performance comparison with different σ_e^2. $M = N = 4$, $\alpha = \beta = 0.5$.
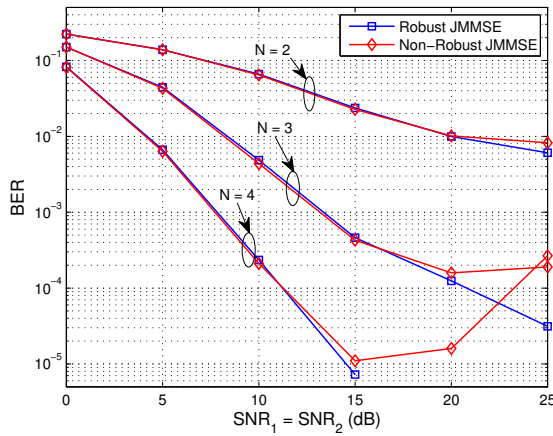

Fig. 8. BER performance comparison when using convolutional code with different N. $M = 2$, $\sigma_e^2 = 0.005$, $\alpha = \beta = 0.5$.

D. Systems with Convolutional Codes

In practical wireless communication systems, the convolutional code as a type of error correction code is usually employed to improve communication reliability. We thus would like to examine whether the robust schemes still have evident advantages over the non-robust schemes when the convolutional code is used. In the following results, the convolutional code with code rate $\frac{1}{2}$ and constrained length 3 is used. We show the BER results in Fig. 8 with the same simulation settings as in Fig. 6. It can be seen that the robust joint MMSE precoding scheme still shows its robustness at high SNR.

VII. Conclusion

This paper addressed the robust joint optimization of source and relay precoders for ANC-based MIMO two-way relay systems with both channel uncertainties and residual SI at destination nodes taken into account. Using a stochastic approach, we calculated the sum MSE averaged over channel uncertainties, and it was shown that the joint optimization problems for both FDD and TDD systems can be solved in

a unified framework. Due to the nonconvexity of the original problem, we first proposed an iterative algorithm by alternately solving three convex sub-problems. Then to further reduce complexity, we proposed a constrained structure algorithm in which the relay precoder has a specific structure and the source precoders are fixed. Simulation results confirmed the superior performance of the proposed robust schemes compared to the non-robust schemes especially in median and high SNR regions, and this advantage is also evident even when the convolutional code is employed. With the source precoders fixed, the constrained structure algorithm achieves close performance with the proposed iterative algorithm when the relay node has the same number of antennas as the source nodes.

Appendix A
Proof of Upper Bound on λ

Proof: If (36) is satisfied with $\lambda = 0$, then the optimal relay precoder is easy to obtain by setting $\lambda = 0$ in (34). Otherwise, $\lambda > 0$, by (35) we have

$$P_r = \text{Tr}(\mathbf{F}\mathbf{R}_r\mathbf{F}^H) = \text{Tr}(\mathbf{R}_r\mathbf{F}^H\mathbf{F}) \leq \lambda_1(\mathbf{R}_r)\text{Tr}(\mathbf{F}^H\mathbf{F}), \quad (51)$$

where the inequality holds because for any $\mathbf{A} \succeq \mathbf{0}$ and $\mathbf{B} \succeq \mathbf{0}$, $\text{Tr}(\mathbf{A}\mathbf{B}) \leq \lambda_1(\mathbf{A})\text{Tr}(\mathbf{B})$. Using (33), we get

$$
\begin{aligned}
\text{Tr}(\mathbf{F}^H\mathbf{F}) &= \text{Tr}(\text{vec}(\mathbf{F})\text{vec}(\mathbf{F})^H) \\
&= \text{Tr}\left[\mathbf{P}^{-1}\text{vec}(\mathbf{Q})\text{vec}(\mathbf{Q})^H\mathbf{P}^{-H}\right] \\
&\leq \lambda_1(\text{vec}(\mathbf{Q})\text{vec}(\mathbf{Q})^H)\text{Tr}(\mathbf{P}^{-1}\mathbf{P}^{-H}). \quad (52)
\end{aligned}
$$

Also, for any two matrices $\mathbf{A} \succeq \mathbf{0}$ and $\mathbf{B} \succeq \mathbf{0}$, $\mathbf{A} \otimes \mathbf{B} \succeq \mathbf{0}$. From (34), we have $\mathbf{P} - \lambda\mathbf{R}_r^T \otimes \mathbf{I}_N \succeq \mathbf{0}$, and hence $\mathbf{P} \succeq \lambda\mathbf{R}_r^T \otimes \mathbf{I}_N$. Moreover, if $\mathbf{A} \succeq \mathbf{B}$, then $\mathbf{A}^{-1} \preceq \mathbf{B}^{-1}$ and $\text{Tr}(\mathbf{A}) \geq \text{Tr}(\mathbf{B})$. Thus, we have

$$\text{Tr}(\mathbf{P}^{-1}) \leq \text{Tr}((\lambda\mathbf{R}_r^T \otimes \mathbf{I}_N)^{-1}). \quad (53)$$

Further, for any $\mathbf{A} \succeq \mathbf{0}$ and $\mathbf{B} \succeq \mathbf{0}$, $\text{Tr}(\mathbf{A}\mathbf{B}) \leq \text{Tr}(\mathbf{A})\text{Tr}(\mathbf{B})$. It follows that

$$
\begin{aligned}
\text{Tr}(\mathbf{P}^{-1}\mathbf{P}^{-H}) &\leq \text{Tr}(\mathbf{P}^{-1})\text{Tr}(\mathbf{P}^{-H}) \\
&\leq (\text{Tr}((\lambda\mathbf{R}_r^T \otimes \mathbf{I}_N)^{-1}))^2 \\
&= \lambda^{-2}(\text{Tr}((\mathbf{R}_r^T \otimes \mathbf{I}_N)^{-1}))^2. \quad (54)
\end{aligned}
$$

Combining (51), (52) and (54), we can obtain the upper bound on λ. ∎

Appendix B
Proof of Upper Bound on $\tilde{J}_i(\mathbf{F})$

Proof: From (20) and (24), we have

$$\mathbf{\Phi}_j + \mathbf{R}_{n_r} + \hat{\mathbf{H}}_i\mathbf{R}_{x_i}\hat{\mathbf{H}}_i^H = \mathbf{R}_r - \text{Tr}(\mathbf{R}_{x_i}\hat{\mathbf{T}}_{hi})\mathbf{R}_{hi}. \quad (55)$$

Substituting (55) into (22) gives

$$
\begin{aligned}
\mathbf{\Omega}_i &= \text{Tr}[\mathbf{F}(\mathbf{R}_r - \text{Tr}(\mathbf{R}_{x_i}\hat{\mathbf{T}}_{hi})\mathbf{R}_{hi})\mathbf{F}^H\hat{\mathbf{T}}_{gi}]\mathbf{R}_{gi} + \mathbf{R}_{n_i} \\
&\stackrel{(a)}{\preceq} \text{Tr}(\mathbf{F}\mathbf{R}_r\mathbf{F}^H\hat{\mathbf{T}}_{gi})\mathbf{R}_{gi} + \mathbf{R}_{n_i} \\
&\preceq \lambda_1(\hat{\mathbf{T}}_{gi})\text{Tr}(\mathbf{F}\mathbf{R}_r\mathbf{F}^H)\mathbf{R}_{gi} + \mathbf{R}_{n_i} \\
&\preceq \lambda_1(\hat{\mathbf{T}}_{gi})P_r\mathbf{R}_{gi} + \mathbf{R}_{n_i}, \quad (56)
\end{aligned}
$$

hence $\bar{\boldsymbol{\Omega}}_i \succeq \boldsymbol{\Omega}_i$. Note that (a) should be tight when $\sigma_{e_i}^2$ is quite small. Using the result in [20],

$$\mathbf{B}^H(\mathbf{A}^H\mathbf{M}^{-1}\mathbf{A}+\mathbf{I})^{-1}\mathbf{B} \succeq \mathbf{B}^H(\mathbf{A}^H\mathbf{N}^{-1}\mathbf{A}+\mathbf{I})^{-1}\mathbf{B},$$

where $\mathbf{A}$ and $\mathbf{B}$ are arbitrary matrices and $\mathbf{M} \succeq \mathbf{N} \succeq \mathbf{0}$, we have $\tilde{J}_i^U(\mathbf{F}) \geq \tilde{J}_i(\mathbf{F})$. ∎

APPENDIX C
PROOF OF UPPER BOUND ON μ

Proof: From (48), we construct the Lagrangian function

$$L(\mathbf{p},\mu) = \sum_{i=1}^{2}\sum_{n=1}^{N}\frac{c_{in}}{1+k_{in}p_n} + \mu(\sum_{n=1}^{N}d_n p_n - P_r), \quad (57)$$

where $\mathbf{p} \triangleq [p_1, p_2, \ldots, p_N]$. By setting $\frac{\partial L(\mathbf{p},\mu)}{\partial p_n} = 0$, we have

$$\sum_{i=1}^{2} -\frac{c_{in}k_{in}}{(1+k_{in}p_n)^2} + \mu d_n = 0. \quad (58)$$

Since $\sum_{n=1}^{N} d_n p_n^{opt} = P_r$, there exits at least one $m \in \{1,\ldots,N\}$ such that $p_m^{opt} > 0$. Hence, from (58) we have

$$\mu = \sum_{i=1}^{2}\frac{c_{im}k_{im}}{d_m(1+k_{im}p_m^{opt})^2}$$
$$\leq \sum_{i=1}^{2}\frac{c_{im}k_{im}}{d_m}$$
$$\leq \max\left\{\sum_{i=1}^{2}\frac{c_{in}k_{in}}{d_n}, \forall n=1,\ldots,N\right\}. \quad (59)$$

Thus the upper bound on μ is derived. ∎

REFERENCES

[1] R. Pabst, B. H. Walke, D. C. Schultz, *et al.*, "Relay based deployment concepts for wireless and mobile broadband radio," *IEEE Commun. Mag.*, vol. 42, pp. 80–89, Sep. 2004.
[2] G. Kramar, M. Gastpar, and P. Gupta, "Cooperative strategies and capacity theorems for relay networks," *IEEE Trans. Inf. Theory*, vol. 51, pp. 3037–3063, Sep. 2005.
[3] R. Ahlswede, N. Cai, S.-Y. R. Li, and R. W. Yeung, "Network information flow," *IEEE Trans. Inf. Theory*, vol. 46, no. 4, pp. 1204–1216, July 2000.
[4] S. Katti, S. Gollakota, and D. Katabi, "Embracing wireless interference: analog network coding," in *Proc. 2007 ACM SIGCOMM*, pp. 397–408.
[5] D. Tse and P. Viswanath, *Fundamentals of Wireless Communications*. Cambridge University Press, 2005.
[6] X. Tang and Y. Hua, "Optimal design of non-regenerative MIMO wireless relays," *IEEE Trans. Wireless Commun.*, vol. 6, no. 4, pp. 1398–1407, Apr. 2007.
[7] O. Munoz-Medina, J. Vidal, and A. Agustin, "Linear transceiver design in nonregenerative relays with channel state information," *IEEE Trans. Signal Process.*, vol. 55, no. 6, pp. 2593–2604, June 2007.
[8] W. Guan and H. Luo, "Joint MMSE transceiver design in non-regenerative MIMO relay systems," *IEEE Commun. Lett.*, vol. 12, pp. 517–519, July 2008.
[9] R. Mo and Y. H. Chew, "MMSE-based joint source and relay precoding design for amplify-and-forward MIMO relay networks," *IEEE Trans. Wireless Commun.*, vol. 8, pp. 4668–4676, Sep. 2009.
[10] C. Song, K.-J. Lee, and I. Lee, "MMSE based transceiver designs in closed-loop non-regenerative MIMO relaying systems," *IEEE Trans. Wireless Commun.*, vol. 9, no. 7, pp. 2310–2319, July 2010.
[11] R. Vaze and R. W. Heath, "On the capacity and diversity-multiplexing tradeoff of the two-way relay channel," *IEEE Trans. Inf. Theory*, vol. 57, no. 7, pp. 4219–4234, July 2011.

[12] R. Zhang, Y.-C. Liang, C. C. Chai, and S. Cui, "Optimal beamforming for two-way multi-antenna relay channel with analogue network coding," *IEEE J. Sel. Areas Commun.*, vol. 27, no. 5, pp. 699–712, June 2009.
[13] G. Li, Y. Wang, and P. Zhang, "Optimal linear MMSE beamforming for two way multi-antenna relay systems," *IEEE Commun. Lett.*, vol. 15, no. 5, pp. 533–535, May 2011.
[14] X. Wang and X.-D. Zhang, "Optimal beamforming in MIMO two-way relay channels," in *Proc. 2010 IEEE Global Commun. Conf.*, pp. 1–5.
[15] S. Xu and Y. Hua, "Optimal design of spatial source-and-relay matrices for a non-regenerative two-way MIMO relay system," *IEEE Trans. Wireless Commun.*, vol. 10, no. 5, pp. 1645–1655, May 2011.
[16] R. Wang and M. Tao, "Joint source and relay precoding design for MIMO two-way relaying based on MSE criterion," *IEEE Trans. Signal Process.*, vol. 6, no. 3, pp. 1352–1365, Mar. 2012.
[17] K.-J. Lee, H. Sung, E. Park, and I. Lee, "Joint optimization for one and two-way MIMO AF multiple-relay systems," *IEEE Trans. Wireless Commun.*, vol. 9, no. 12, pp. 3671–3681, Dec. 2010.
[18] Y. Rong, "Joint source and relay optimization for two-way MIMO multi-relay networks," *IEEE Commun. Lett.*, vol. 15, pp. 1329–1331, Dec. 2011.
[19] B. K. Chalise and L. Vandendorpe, "MIMO relay design for multipoint-to-multipoint communications with imperfect channel state information," *IEEE Trans. Signal Process.*, vol. 57, no. 7, pp. 2785–2796, July 2009.
[20] 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.
[21] 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.
[22] I. Krikidis and J. S. Thompson, "MIMO two-way relay channel with superposition coding and imperfect channel estimation," in *Proc. 2010 IEEE Global Commun. Conf. Workshops*, pp. 84–88.
[23] E. A. Gharavol and E. G. Larsson, "Robust joint optimization of MIMO two-way relay channels with imperfect CSI," in *Proc. 2011 Allerton Conf. Communication, Control, and Computing*, pp. 1657–1664.
[24] J. Zou, W. Liu, M. Ding, H. Luo, and H. Yu, "Transceiver desgin for AF MIMO two-way relay systems with imperfect channel estimation," in *Proc. 2011 IEEE Global Commun. Conf.*, pp. 1–5.
[25] D. Shiu, G. J. Foschini, M. J. Gans, and J. M. Kahn, "Fading correlation and its effect on the capacity of multielement antenna systems," *IEEE Trans. Commun.*, vol. 48, no. 3, pp. 502–513, Mar. 2000.
[26] T. Yoo and A. Goldsmith, "Capacity and power allocation for fading MIMO channels with channel estimation error," *IEEE Trans. Inf. Theory*, vol. 52, no. 5, pp. 2203–2214, May 2006.
[27] E. Telatar, "Capacity of multi-antenna Gaussian channels," *Europ. Trans. Telecommun.*, vol. 10, no. 6, pp. 585–596, Nov. 1999.
[28] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.
[29] M. Grant and S. Boyd, CVX: Matlab software for disciplined convex programming, version 1.21, http://cvxr.com/cvx, Apr. 2011.
[30] A. Paulraj, R. Nabar, and D. Gore, *Introduction to Space-Time Wireless Communications*. Cambridge University Press, 2003.

Jun Zou received his B.S. degree in electronic engineering from Shanghai Jiao Tong University, Shanghai, China, in 2009. Then he joined the dual-master's program between Shanghai Jiao Tong University and Georgia Institute of Technology, and received his M.S. degree in electronic engineering from Shanghai Jiao Tong University and M.S. degree in electrical and computer engineering from Georgia Institute of Technology, both in 2012. He is now pursuing his Ph.D. degree in electrical and computer engineering at the Georgia Institute of Technology, Atlanta, GA. His research interests include MIMO systems and relay communications.
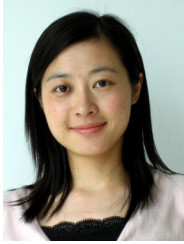
Hanwen Luo is a full Professor with the Department of Electronic Engineering, Shanghai Jiao Tong University, Shanghai, China. He held several positions including Vice Director of Shanghai Institute of Wireless Communications Technology and Vice Director of Institute of Wireless Communication Technology in Shanghai Jiao Tong University. He used to be the leading specialist of China 863 high-tech program on Beyond 3G wireless communication systems and China 973 high-tech program on the researches of military equipment. His research interests include mobile and personal communications.

Meixia Tao (S'00-M'04-SM'10) received the B.S. degree in electronic engineering from Fudan University, Shanghai, China, in 1999, and the Ph.D. degree in electrical and electronic engineering from Hong Kong University of Science and Technology in 2003. She is currently an Associate Professor with the Department of Electronic Engineering, Shanghai Jiao Tong University, China. From August 2003 to August 2004, she was a Member of Professional Staff at Hong Kong Applied Science and Technology Research Institute Co. Ltd. From August 2004 to December 2007, she was with the Department of Electrical and Computer Engineering, National University of Singapore, as an Assistant Professor. Her current research interests include cooperative transmission, physical layer network coding, resource allocation of OFDM networks, and MIMO techniques.

Dr. Tao is an Associate Editor for the IEEE COMMUNICATIONS LETTERS and an Editor for the IEEE WIRELESS COMMUNICATIONS LETTERS. She was on the Editorial Board of the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS from 2007 to 2011. She served as Track/Symposium Co-Chair for WPMC12, APCC09, ChinaCom09, IEEE ICCCN07, and IEEE ICCCAS07. She has also served as Technical Program Committee member for various conferences, including IEEE INFOCOM, IEEE GLOBECOM, IEEE ICC, IEEE WCNC, and IEEE VTC.

Dr. Tao is the recipient of the IEEE ComSoC Asia-Pacific Outstanding Young Researcher Award in 2009.

Rui Wang received the B.S. degree from Anhui Normal University, Wuhu, China, in 2006, and the M.S. degree from Shanghai University, Shanghai, China, in 2009, both in electronic engineering. Currently, he is pursuing his Ph.D. degree at the Institute of Wireless Communication Technology (IWCT) in Shanghai Jiao Tong University. His research interests include digital image processing, cognitive radio and signal processing for wireless cooperative communication.