Robust Tomlinson-Harashima Source and Linear Relay Precoders Design in Amplify-and-Forward MIMO Relay Systems

Fan-Shuo Tseng, *Member, IEEE,* Min-Yao Chang, and Wen-Rong Wu, *Member, IEEE*

***Abstract*—Existing transceiver designs in amplify-and-forward (AF) multiple-input-multiple-output (MIMO) relay systems often assume the availability of perfect channel state informations (CSIs). Robust designs for imperfect CSI have less been considered. In this paper, we propose a robust nonlinear transceiver design for the system with a Tomlinson-Harashima precoder (THP), a linear relay precoder, and a minimum-mean-squared-error (MMSE) receiver. Since two precoders and imperfect CSIs are involved, the robust transceiver design is difficult. To overcome the difficulty, we first propose cascading an additional unitary precoder after the THP. The unitary precoder can not only simplify the optimization but also improve the performance of the MMSE receiver. We then adopt the primal decomposition dividing the original optimization problem into a subproblem and a master problem. With our formulation, the subproblem can be solved and the two-precoder problem can be transferred to a single relay precoder problem. The master problem, however, is not solvable. We then propose a lower bound for the objective function and transfer the master problem into a convex optimization problem. A closed-form solution can then be obtained by the Karush-Kuhn-Tucker (KKT) conditions. Simulations show that the proposed transceiver can significantly outperform existing linear transceivers with perfect or imperfect CSIs.**

***Index Terms*—Amplify-and-forward (AF), multiple-input multiple-output (MIMO), channel state information (CSI), joint source/relay precoders, robust transceiver design, Tomlinson Harashima precoding (THP), minimum-mean-squared-error (MMSE), primal decomposition approach, Karuch-Kuhn-Tucker (KKT) conditions.**

I. Introduction

RECENTLY, the multiple-input multiple-output (MIMO) technique was introduced to cooperative systems [1]-[10]. When multiple antennas are deployed at each relay node, the cooperative system is referred to as a MIMO relay system. Similar to the conventional MIMO systems [11]-[24], [34], MIMO relays can provide additional degrees of freedom for increasing spectral efficiency and/or transmission reliability. The capacity bound for a three-node MIMO relay system was first studied in [1]. Also similar to MIMO systems, the precoding technique can be applied in MIMO relay systems to further improve the performance. A relay precoder was first designed to enhance the capacity of a three-node amplify-and-forward (AF) MIMO relay system [2], [3]. It was soon realized that the capacity can be further enhanced if the direct link is further taken into account [3]. Apart from the capacity, link quality improvement is an alternative criterion that has been considered [4]-[10]. In [4]-[5], a relay precoder was designed for the minimum-mean-squared-error (MMSE) receiver. The same design criterion for multiple relays was later developed [5]. More recently, joint source and relay precoders designs were studied [6]-[10] for the MMSE [6], [7], QR successive-interference-cancellation (SIC) [8], and MMSE-SIC receivers [9], respectively. Note that the aforementioned designs all address the linear relay and/or linear source precoders [2]-[9]. Nonlinear precoders design for AF MIMO relay systems was first discussed in [10] in which a Tomlinson-Harashima (TH) source precoder and a linear relay precoder are jointly optimized for a MMSE receiver. As that in conventional MIMO systems [18], [19], [34], nonlinear precoded MIMO relay systems can yield better performance.

As known, channel state informations (CSIs) are required in the transceiver design. Most designs often assume that perfect CSIs are available [2]-[10]. However, in real-world applications, channel responses are usually estimated at the receiver and fed back to the transmitter. As a result, estimation and quantization errors always exist. The performance of a transceiver designed with imperfect CSIs can be seriously degraded. In some cases, it may even be poorer than that of un-precoded systems [6]. To overcome the problem, robust transceiver designs were then developed for conventional point-to-point MIMO [21]-[23] and MIMO relay systems [25], [33]. Note that in [25] and [33], linear transceivers were considered. In this paper, we extend the work in [10] taking the problem of imperfect CSIs into consideration. We develop a robust nonlinear transceiver for a three-node MIMO relay system in which a Tomlinson-Harashima precoder (THP) is used at the source, a linear relay precoder at the relay, and the MMSE receiver at the destination.

As typical precoder design, the problem can be easily formulated as an optimization problem. However, in our design, the optimization involves two precoders, two coupled power constraint, and three channels (source-to-destination, source-to-relay and relay-to-destination). Even with numerical methods [28], the optimum solution is difficult to obtain. To

Paper approved by D. J. Love, the Editor for MIMO and Adaptive Techniques of the IEEE Communications Society. Manuscript received September 4, 2010; revised July 9 and October 13, 2011.

F.-S. Tseng is with the Institute of Communications Engineering, National Sun Yat-sen University, Kaohsiung, Taiwan (e-mail: fs.tseng@mail.nsysu.edu.tw).

M.-Y. Chang is with MediaTek Inc., Hsinchu Science Park, Taiwan (e-mail: emifyao33@gmail.com).

W.-R. Wu is with the Institute of Communication Engineering, National Chiao-Tung University, Hsinchu, Taiwan (e-mail: wrwu@faculty.nctu.edu.tw).

Digital Object Identifier 10.1109/TCOMM.2012.031712.100514

alleviate the problem, we first propose cascading a unitary precoder after the THP. In this manner, as we will see, not only the mean-squared-error (MSE) can be reduced, but also the optimization can be simplified. We then propose using the primal decomposition transfering the optimization problem into a subproblem and a master problem. In our approach, the source precoder is optimized in the subproblem and subsequently the relay precoder in the master problem. With our formualtion, the optimal source precoder can be solved as a function of the relay precoder in the subproblem. However, the relay precoder in the master problem is not solvable. To complete the design, we further propose a method that can translate the master problem to a standard scalar-valued concave optimization problem. The key idea is to use some relaxation for the objective function in the master problem. In some scenarios, the relaxed objective function is equal to the original objective function. Finally, we can obtain a closed-form solution for the relay and source precoders via Karash-Kuhn-Tucker (KKT) conditions.

This paper is organized as follows. In Section II, we first describe a three-node AF MIMO relay system with the TH source precoder, a linear relay precoder, and an MMSE receiver. Then, we formulate the optimization problem for the precoders design with imperfect CSIs. In Section III, we propose an efficient method solving the optimization problem. The performance of the proposed transceiver is then evaluated in Section IV. Finally, we draw conclusions in Section V.

II. System Model and Problem Formulation

A. MMSE Receiver With TH Source and Linear Relay Precoders

We consider a three-node AF MIMO relay precoding system in which N, R, and M antennas are placed at the source, the relay and the destination, respectively, as shown in Fig. 1. From the figure, we see that the system consists of a TH source precoder, a linear relay precoder, and a linear MMSE receiver. Here, we consider the general two-phase AF protocol [2]-[10]. In the first phase, the source signal vector $\mathbf{s} \in \mathbb{C}^{N\times 1}$ is first fed into the THP. The precoder conducts a successive cancellation operation characterized by a backward squared matrix $\mathbf{B}$ and a modulo operation $\text{MOD}_m(\cdot)$. Each element of the transmit vector, $\mathbf{s} = [s_1, \ldots, s_N]^T$, is a m-QAM modulated signal and takes its real and imaginary values from the set $\{\pm 1, \ldots, \pm(\sqrt{m}-1)\}$. The matrix $\mathbf{B}$ has a lower triangular structure and its diagonal elements are all equal to zeros. The modulo operation, conducted over the real and imaginary parts of the inputs, can be expressed as follows:

$$\text{MOD}_m(x) = x - 2\sqrt{m}\cdot \lfloor \frac{x+\sqrt{m}}{2\sqrt{m}} \rfloor. \tag{1}$$

Let the signal after the modulo operation be expressed as $\mathbf{x}$. It is clear that each element of $\mathbf{x}$ is bounded between $-\sqrt{m}$ and $\sqrt{m}$ (real and imaginary parts). With $\mathbf{B}$ and the operation in (1), the elements of $\mathbf{x}$ can be expressed as [19]

$$x_k = s_k - \sum_{l=1}^{k-1} \mathbf{B}(k,l)x_l + e_k \tag{2}$$

where x_k denotes the kth element of the vector $\mathbf{x}$, $\mathbf{B}(k,l)$ the (k,l) element of the matrix $\mathbf{B}$, and e_k is the error yielded by the modulo operation (the difference between the input and the output). Let $\mathbf{e} = [e_1, \ldots, e_N]^T$ be the error vector. From (2), we can define the transmitted signal $\mathbf{x}$ using the following matrix expression:

$$\mathbf{x} = \mathbf{C}^{-1}\mathbf{v} \tag{3}$$

where $\mathbf{C} = \mathbf{B} + \mathbf{I}_N$ is a lower triangular matrix with ones in its diagonal, and $\mathbf{v} = \mathbf{s} + \mathbf{e}$. The TH precoded signal vector $\mathbf{x}$ is further passed through a unitary precoder matrix $\mathbf{F}_S$ and subsequently sent to the relay and the destination. The unitary precoder, as we will see, can greatly simplify the design and improve the bit-error-rate (BER) performance.

In the second phase, the received signal vector at the relay is multiplied by the relay precoding matrix, and then transmitted to the destination. Therefore, the signal received at the destination (after the two consecutive phases) can be combined into a single vector as [6]-[10]

$$\mathbf{y}_D := \mathbf{H}\mathbf{F}_S\mathbf{x} + \mathbf{w}, \tag{4}$$

where

$$\mathbf{H} = \begin{bmatrix} \mathbf{H}_{SD} \\ \mathbf{H}_{RD}\mathbf{F}_R\mathbf{H}_{SR} \end{bmatrix} \text{ and } \mathbf{w} = \begin{bmatrix} \mathbf{n}_{D,1} \\ \mathbf{H}_{RD}\mathbf{F}_R\mathbf{n}_R + \mathbf{n}_{D,2} \end{bmatrix} \tag{5}$$

denote the equivalent channel matrix and the equivalent noise vector, respectively. In (4), $\mathbf{x} \in \mathbb{C}^{N\times 1}$ is the TH precoded signal vector defined in (3); $\mathbf{y}_D \in \mathbb{C}^{2M\times 1}$ is the received signal vector at the destination; $\mathbf{H}_{SR} \in \mathbb{C}^{R\times N}$, $\mathbf{H}_{SD} \in \mathbb{C}^{M\times N}$, and $\mathbf{H}_{RD} \in \mathbb{C}^{M\times R}$ are the channel matrices of the source-to-relay, the source-to-destination, and the relay-to-destination links, respectively; $\mathbf{n}_{D,1} \in \mathbb{C}^{M\times 1}$, $\mathbf{n}_R \in \mathbb{C}^{R\times 1}$ and $\mathbf{n}_{D,2} \in \mathbb{C}^{M\times 1}$ are the received noise vector at the destination in the first-phase, that at the relay in the first-phase, and that at the destination in the second-phase. Before proceeding to our design, we first consider the error model of channel estimation. Let $\mathbf{H}_{MO}$ denote the channel matrix of a $N\times M$ point-to-point MIMO system. A common model for the generation of $\mathbf{H}_{MO}$ is given by [20]-[22]

$$\mathbf{H}_{MO} = \mathbf{R}_{R,MO}^{1/2}\mathbf{H}_{i.i.d.}\mathbf{R}_{T,MO}^{1/2}, \tag{6}$$

where $\mathbf{H}_{i.i.d.}$ is a spatially white matrix whose entries are independent and identically distributed (i.i.d.), and $\mathbf{R}_{R,MO}$ and $\mathbf{R}_{T,MO}$ are normalized receive and transmit correlation matrices (unit diagonal entries). Also, each element of $\mathbf{H}_{i.i.d.}$ has a zero-mean and unit variance Gaussian distribution. When a linear channel estimation method is adopted, the relationship of the true and estimated channel matrices can be expressed as [20], [21], [25]

$$\mathbf{H}_{MO} = \widehat{\mathbf{H}}_{MO} + \Delta\mathbf{H}_{MO} \tag{7}$$

where $\widehat{\mathbf{H}}_{MO}$ is the estimate of $\mathbf{H}_{MO}$ and $\Delta\mathbf{H}_{MO}$ is the estimation error. It has been shown that [20], [21], [25]

$$\Delta\mathbf{H}_{MO} = \mathbf{\Sigma}_{MO}^{1/2}\Delta\mathbf{H}_{i.i.d.}\mathbf{\Psi}_{MO}^{1/2}, \tag{8}$$

where $\mathbf{\Sigma}_{MO}$ and $\mathbf{\Psi}_{MO}$ are two covariance matrices associated with $\Delta\mathbf{H}_{MO}$ and $\Delta\mathbf{H}_{i.i.d.} = \mathbf{H}_{i.i.d.} - \widehat{\mathbf{H}}_{i.i.d.}$. Here, $\widehat{\mathbf{H}}_{i.i.d.}$ is an estimate of $\mathbf{H}_{i.i.d.}$. The actual values of $\mathbf{\Sigma}_{MO}$ and $\mathbf{\Psi}_{MO}$

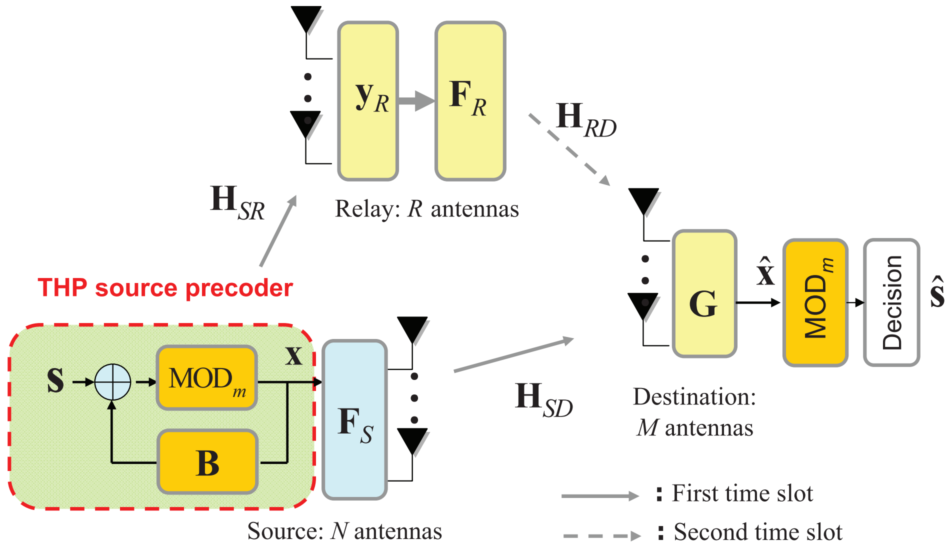


Fig. 1. TH source and linear relay precoded AF MIMO relay system with MMSE receiver.

depend on the channel estimation method we use. For example, if we use the estimation algorithm proposed in [20], then $\mathbf{\Psi}_{MO} = \mathbf{R}_{T,MO}$ and $\mathbf{\Sigma}_{MO} = \sigma^2_{e,MO}\mathbf{R}_{R,MO}$ where $\sigma^2_{e,MO} = E\left[|\mathbf{H}_{i.i.d.}(i,j) - \widehat{\mathbf{H}}_{i.i.d.}(i,j)|^2\right]$. Note that $\mathbf{H}_{i.i.d.}$, instead of $\mathbf{H}_{MO}$, is estimated. As another example, if we use the channel estimation method proposed in [21], we then have $\mathbf{\Psi}_{MO} = \mathbf{R}_{T,MO}$ and $\mathbf{\Sigma}_{MO} = \sigma^2_{e,MO}\left(\mathbf{I}_{N,M} + \sigma^2_{e,MO}\mathbf{R}^{-1}_{R,MO}\right)^{-1}$. With the expression in (8), it is clear that

$$\text{vec}\left(\Delta\mathbf{H}_{MO}\right) \sim \mathcal{CN}\left(\mathbf{0}_{NM\times 1}, \mathbf{\Sigma}_{MO} \otimes \mathbf{\Psi}^T_{MO}\right), \quad (9)$$

where vec($\bullet$) denotes the operation of stacking the columns of a matrix into a vector, $\mathcal{CN}(\mathbf{m}, \mathbf{C})$ denotes a complex Gaussian random vector with a mean vector of $\mathbf{m}$ and a covariance matrix of $\mathbf{C}$, and $\otimes$ represents the operation of the Kronecker product. Thus, we can have the probability density function (PDF) of $\Delta\mathbf{H}_{MO}$ as

$$p\left(\Delta\mathbf{H}_{MO}\right) = \frac{\exp\left(-\text{Tr}\left(\Delta\mathbf{H}^H_{MO}\mathbf{\Sigma}^{-1}_{MO}\Delta\mathbf{H}_{MO}\mathbf{\Psi}^{-1}_{MO}\right)\right)}{\pi^{NM}\left(\det\left(\mathbf{\Sigma}_{MO}\right)\right)^N\left(\det\left(\mathbf{\Psi}_{MO}\right)\right)^M}. \quad (10)$$

We then use (8) as our model for channel estimation error and have

$$\mathbf{H}_{SR} = \widehat{\mathbf{H}}_{SR} + \Delta\mathbf{H}_{SR} \quad (11)$$
$$\mathbf{H}_{RD} = \widehat{\mathbf{H}}_{RD} + \Delta\mathbf{H}_{RD} \quad (12)$$
$$\mathbf{H}_{SD} = \widehat{\mathbf{H}}_{SD} + \Delta\mathbf{H}_{SD} \quad (13)$$

where $\widehat{\mathbf{H}}_{SR}$, $\widehat{\mathbf{H}}_{RD}$, and $\widehat{\mathbf{H}}_{SD}$ are the estimated channel matrices of $\mathbf{H}_{SR}$, $\mathbf{H}_{RD}$, and $\mathbf{H}_{SD}$, respectively; $\Delta\mathbf{H}_{SR}$, $\Delta\mathbf{H}_{RD}$, and $\Delta\mathbf{H}_{SD}$ are the corresponding estimation error matrices. From (9), we can then obtain the PDFs of $\Delta\mathbf{H}_{SR}$, $\Delta\mathbf{H}_{RD}$ and $\Delta\mathbf{H}_{SD}$ as

$$\text{vec}\left(\Delta\mathbf{H}_{SR}\right) \sim \mathcal{CN}\left(\mathbf{0}_{NR\times 1}, \mathbf{\Sigma}_{SR} \otimes \mathbf{\Psi}^T_{SR}\right), \quad (14)$$
$$\text{vec}\left(\Delta\mathbf{H}_{RD}\right) \sim \mathcal{CN}\left(\mathbf{0}_{RM\times 1}, \mathbf{\Sigma}_{RD} \otimes \mathbf{\Psi}^T_{RD}\right), \quad (15)$$
$$\text{vec}\left(\Delta\mathbf{H}_{SD}\right) \sim \mathcal{CN}\left(\mathbf{0}_{NM\times 1}, \mathbf{\Sigma}_{SD} \otimes \mathbf{\Psi}^T_{SD}\right). \quad (16)$$

Here, we assume that all the channels are time-invariant and all the second-order channel statistics, $\mathbf{\Sigma}_{SR}$, $\mathbf{\Sigma}_{RD}$, $\mathbf{\Sigma}_{SD}$, $\mathbf{\Psi}_{SR}$, $\mathbf{\Psi}_{RD}$, and $\mathbf{\Psi}_{SD}$ are known a priori. Note that if $\mathbf{v}$ can be estimated, $\mathbf{s}$ can be recovered by the modulo operation in (1). Since we take both noise and channel estimation error into the MMSE receiver design, we can define the MSE as

$$\begin{aligned}&\text{MSE}\left(\mathbf{C}, \mathbf{F}_S, \mathbf{F}_R, \mathbf{G}\right) = E_{\Delta,w}\left\{\|\mathbf{G}\mathbf{y}_D - \mathbf{v}\|^2\right\}\\ &= E_{\Delta,w}\left\{\text{Tr}\left\{\left(\left(\mathbf{G}\mathbf{H}\mathbf{F}_S - \mathbf{C}\right)\mathbf{x} + \mathbf{G}\mathbf{w}\right) \times\right.\right.\\ &\left.\left.\left(\left(\mathbf{G}\mathbf{H}\mathbf{F}_S - \mathbf{C}\right)\mathbf{x} + \mathbf{G}\mathbf{w}\right)^H\right\}\right\}, \quad (17)\end{aligned}$$

where $\mathbf{G}$ represents the MMSE equalization matrix and the subscripts Δ and w denote that the expectation is taken over channel estimation error and noise, respectively. Here, we assume that x_k's are statistically independent and they have a zero-mean and a same variance. Let the variance of each element in $\mathbf{s}$ be denoted as σ^2_s. We then have $E[\mathbf{x}\mathbf{x}^H] = \sigma^2_s\mathbf{I}_N$ and $E[\mathbf{v}\mathbf{v}^H] = \sigma^2_s\mathbf{C}\mathbf{C}^H$. Note that the independent assumption is valid only for large QAM size (e.g., $m \geq 16$) [18], [19]. Therefore, we can rewrite the MSE in (17) as

$$\begin{aligned}&\text{MSE}\left(\mathbf{C}, \mathbf{F}_S, \mathbf{F}_R, \mathbf{G}\right) = E_\Delta\left\{\text{Tr}\left\{\sigma^2_s\mathbf{G}\mathbf{H}\mathbf{F}_S\mathbf{F}^H_S\mathbf{H}^H\mathbf{G}^H\right\}\right\}\\ &-E_\Delta\left\{\text{Tr}\left\{\sigma^2_s\mathbf{G}\mathbf{H}\mathbf{F}_S\mathbf{C}^H\right\}\right\} - E_\Delta\left\{\text{Tr}\left\{\sigma^2_s\mathbf{C}\mathbf{F}^H_S\mathbf{H}^H\mathbf{G}^H\right\}\right\}\\ &+\text{Tr}\left\{\sigma^2_s\mathbf{C}\mathbf{C}^H\right\} + E_\Delta\left\{\text{Tr}\left\{\mathbf{G}\mathbf{R}_w\mathbf{G}^H\right\}\right\}\\ &= \text{Tr}\left\{E_\Delta\left\{\sigma^2_s\mathbf{G}\mathbf{H}\mathbf{F}_S\mathbf{F}^H_S\mathbf{H}^H\mathbf{G}^H\right\}\right\}\\ &-\text{Tr}\left\{E_\Delta\left\{\sigma^2_s\mathbf{G}\mathbf{H}\mathbf{F}_S\mathbf{C}^H\right\}\right\} - \text{Tr}\left\{E_\Delta\left\{\sigma^2_s\mathbf{C}\mathbf{F}^H_S\mathbf{H}^H\mathbf{G}^H\right\}\right\}\\ &+\text{Tr}\left\{\sigma^2_s\mathbf{C}\mathbf{C}^H\right\} + \text{Tr}\left\{E_\Delta\left\{\mathbf{G}\mathbf{R}_w\mathbf{G}^H\right\}\right\}, \quad (18)\end{aligned}$$

where $\mathbf{R}_w = E\left[\mathbf{w}\mathbf{w}^H\right]$. Using the error models in (11)-(13), we can then rewrite the MSE in (18) as

$$\begin{aligned}&\text{MSE}\left(\mathbf{C}, \mathbf{F}_S, \mathbf{F}_R, \mathbf{G}\right)\\ &= \text{Tr}\left\{\mathbf{G}\left(\sigma^2_s\widehat{\mathbf{H}}\mathbf{F}_S\mathbf{F}^H_S\widehat{\mathbf{H}}^H + \widehat{\mathbf{R}}_w + \Delta\text{err}\right)\mathbf{G}^H\right\} -\\ &\text{Tr}\left\{\sigma^2_s\mathbf{G}\widehat{\mathbf{H}}\mathbf{F}_S\mathbf{C}^H\right\} - \text{Tr}\left\{\sigma^2_s\mathbf{C}\mathbf{F}^H_S\widehat{\mathbf{H}}^H\mathbf{G}^H\right\} +\\ &\text{Tr}\left\{\sigma^2_s\mathbf{C}\mathbf{C}^H\right\}, \quad (19)\end{aligned}$$

where

$$\widehat{\mathbf{H}} = \begin{bmatrix} \widehat{\mathbf{H}}_{SD} \\ \widehat{\mathbf{H}}_{RD}\mathbf{F}_R\widehat{\mathbf{H}}_{SR} \end{bmatrix}, \tag{20}$$

$$\begin{aligned}
&\Delta\text{err} = \begin{bmatrix} \Delta\text{err}_A & \mathbf{0} \\ \mathbf{0} & \Delta\text{err}_B \end{bmatrix}, \\
&\Delta\text{err}_A = \sigma_s^2 \text{Tr}\left(\mathbf{F}_S\mathbf{F}_S^H\mathbf{\Psi}_{SD}\right)\mathbf{\Sigma}_{SD} \\
&\Delta\text{err}_B = \text{Tr}\left(\mathbf{F}_R\left(\sigma_s^2\mathbf{T}_{SR} + \sigma_{n,r}^2\mathbf{I}_R\right)\mathbf{F}_R^H\mathbf{\Psi}_{RD}\right)\mathbf{\Sigma}_{RD} + \\
&\qquad \sigma_s^2\text{Tr}\left(\mathbf{F}_S\mathbf{F}_S^H\mathbf{\Psi}_{SR}\right)\widehat{\mathbf{H}}_{RD}\mathbf{F}_R\mathbf{\Sigma}_{SR}\mathbf{F}_R^H\widehat{\mathbf{H}}_{RD}^H
\end{aligned} \tag{21}$$

$$\widehat{\mathbf{R}}_w = E\left\{\begin{bmatrix} \mathbf{n}_{D,1} \\ \widehat{\mathbf{H}}_{RD}\mathbf{F}_R\mathbf{n}_R + \mathbf{n}_{D,2} \end{bmatrix} \times \begin{bmatrix} \mathbf{n}_{D,1}^H & \mathbf{n}_R^H\mathbf{F}_R^H\widehat{\mathbf{H}}_{RD}^H + \mathbf{n}_{D,2}^H \end{bmatrix}\right\}, \tag{22}$$

$$\mathbf{T}_{SD} = \text{Tr}\left(\mathbf{F}_S\mathbf{F}_S^H\mathbf{\Psi}_{SD}\right)\mathbf{\Sigma}_{SD} + \widehat{\mathbf{H}}_{SD}\mathbf{F}_S\mathbf{F}_S^H\widehat{\mathbf{H}}_{SD}^H, \tag{23}$$

$$\mathbf{T}_{SR} = \text{Tr}\left(\mathbf{F}_S\mathbf{F}_S^H\mathbf{\Psi}_{SR}\right)\mathbf{\Sigma}_{SR} + \widehat{\mathbf{H}}_{SR}\mathbf{F}_S\mathbf{F}_S^H\widehat{\mathbf{H}}_{SR}^H. \tag{24}$$

The detailed derivation of (19) is provided in Appendix A. Since MSE in (19) is convex, the optimum $\mathbf{G}$ minimizing MSE, denoted as $\mathbf{G}_{opt}$, can be found by

$$\frac{\partial \text{MSE}\left(\mathbf{C}, \mathbf{F}_S, \mathbf{F}_R, \mathbf{G}\right)}{\partial \mathbf{G}^H} = \mathbf{0}. \tag{25}$$

If we assume that $\mathbf{F}_S$, $\mathbf{F}_R$, and $\mathbf{C}$ are known, $\mathbf{G}_{opt}$ can be expressed as

$$\mathbf{G}_{opt} = \sigma_s^2\mathbf{C}\mathbf{F}_S^H\widehat{\mathbf{H}}^H\left(\sigma_s^2\widehat{\mathbf{H}}\mathbf{F}_S\mathbf{F}_S^H\widehat{\mathbf{H}}^H + \widehat{\mathbf{R}}_w + \Delta\text{err}\right)^{-1}. \tag{26}$$

Substituting (26) into (19), we then have the minimum MSE as

$$\begin{aligned}
&\text{MSE}\left(\mathbf{C}, \mathbf{F}_S, \mathbf{F}_R, \mathbf{G}_{opt}\right) := \text{MSE}_{min}\left(\mathbf{C}, \mathbf{F}_S, \mathbf{F}_R\right) \\
&= \sigma_s^2\text{Tr}\left\{\mathbf{C}\mathbf{C}^H\right\} - \sigma_s^2\text{Tr}\left\{\mathbf{C}\mathbf{F}_S^H\widehat{\mathbf{H}}^H\left(\widehat{\mathbf{H}}\mathbf{F}_S\mathbf{F}_S^H\widehat{\mathbf{H}}^H\right.\right. \\
&\left.\left. +\sigma_s^{-2}\left(\widehat{\mathbf{R}}_w + \Delta\text{err}\right)\right)^{-1}\widehat{\mathbf{H}}\mathbf{F}_S\mathbf{C}^H\right\}.
\end{aligned} \tag{27}$$

Using the matrix inversion lemma [27], we can further express (27) as

$$\begin{aligned}
&\text{MSE}_{min}\left(\mathbf{C}, \mathbf{F}_S, \mathbf{F}_R\right) \\
&= \sigma_s^2\text{Tr}\left\{\mathbf{C}\left(\mathbf{F}_S^H\widehat{\mathbf{H}}^H\mathbf{R}_\Delta^{-1}\widehat{\mathbf{H}}\mathbf{F}_S + \mathbf{I}_N\right)^{-1}\mathbf{C}^H\right\},
\end{aligned} \tag{28}$$

where

$$\begin{aligned}
\mathbf{R}_\Delta &= \sigma_s^{-2}\left(\widehat{\mathbf{R}}_w + \Delta\text{err}\right) \\
&= \sigma_s^{-2}\begin{bmatrix} \mathbf{R}_{\Delta,1,1} & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_{\Delta,2,2} \end{bmatrix}.
\end{aligned} \tag{29}$$

Here,

$$\mathbf{R}_{\Delta,1,1} = \sigma_{n,d}^2\mathbf{I}_M + \sigma_s^2\text{Tr}\left(\mathbf{F}_S\mathbf{F}_S^H\mathbf{\Psi}_{SD}\right)\mathbf{\Sigma}_{SD} \tag{30}$$

and

$$\begin{aligned}
&\mathbf{R}_{\Delta,2,2} = \text{Tr}\left(\mathbf{F}_R\left(\sigma_s^2\left(\text{Tr}\left(\mathbf{F}_S\mathbf{F}_S^H\mathbf{\Psi}_{SR}\right)\mathbf{\Sigma}_{SR} +\right.\right.\right. \\
&\left.\left.\left.\widehat{\mathbf{H}}_{SR}\mathbf{F}_S\mathbf{F}_S^H\widehat{\mathbf{H}}_{SR}^H\right) + \sigma_{n,r}^2\mathbf{I}_R\right)\mathbf{F}_R^H\mathbf{\Psi}_{RD}\right)\mathbf{\Sigma}_{RD} + \\
&\sigma_s^2\text{Tr}\left(\mathbf{F}_S\mathbf{F}_S^H\mathbf{\Psi}_{SR}\right)\widehat{\mathbf{H}}_{RD}\mathbf{F}_R\mathbf{\Sigma}_{SR}\mathbf{F}_R^H\widehat{\mathbf{H}}_{RD}^H + \\
&\sigma_{n,r}^2\widehat{\mathbf{H}}_{RD}\mathbf{F}_R\mathbf{F}_R^H\widehat{\mathbf{H}}_{RD}^H + \sigma_{n,d}^2\mathbf{I}_M.
\end{aligned} \tag{31}$$

From (29)-(31), we can see that the MSE in (28) is a complicated nonlinear function of $\mathbf{F}_S$ and $\mathbf{F}_R$.

B. Problem Formulation

Our task is to find $\mathbf{C}$, $\mathbf{F}_S$, and $\mathbf{F}_R$ so that the MSE in (28) is minimized. The joint precoder designs problem can now be formulated as:

$$\begin{aligned}
&\min_{\mathbf{C},\mathbf{F}_S,\mathbf{F}_R} \sigma_s^2\text{Tr}\left\{\mathbf{C}\left(\mathbf{I}_N + \mathbf{F}_S^H\widehat{\mathbf{H}}^H\mathbf{R}_\Delta^{-1}\widehat{\mathbf{H}}\mathbf{F}_S\right)^{-1}\mathbf{C}^H\right\} \\
&s.t.\ \ \sigma_s^2\text{Tr}\left\{\mathbf{F}_S\mathbf{F}_S^H\right\} \le P_{S,T}, \\
&E\left[\text{Tr}\left\{\mathbf{F}_R\left(\sigma_{n,r}^2\mathbf{I}_R + \sigma_s^2\mathbf{H}_{SR}\mathbf{F}_S\mathbf{F}_S^H\mathbf{H}_{SR}^H\right)\mathbf{F}_R^H\right\}\right] = \\
&\sigma_{n,r}^2\text{Tr}\left\{\mathbf{F}_R\mathbf{F}_R^H\right\} + \sigma_s^2\text{Tr}\left\{\mathbf{F}_R\left(\text{Tr}\left(\mathbf{F}_S\mathbf{F}_S^H\mathbf{\Psi}_{SR}\right)\mathbf{\Sigma}_{SR}\right.\right. \\
&\left.\left.+\widehat{\mathbf{H}}_{SR}\mathbf{F}_S\mathbf{F}_S^H\widehat{\mathbf{H}}_{SR}^H\right)\mathbf{F}_R^H\right\} \le P_{R,T},
\end{aligned} \tag{32}$$

where the inequalities in (32) are due to the transmission power constraints at the source and the relay (the maximum available powers are $P_{S,T}$ and $P_{R,T}$, respectively). Taking a close look at (32), we can observe that in addition to the cost function, the constraints are also complicated nonlinear functions of $\mathbf{F}_S$ and $\mathbf{F}_R$. Even worse, they are mutually coupled through $\mathbf{F}_S$. It is simple to check that (32) is not a convex optimization problem. Since imperfect CSIs of all links are involved, the problem is much more difficult than that in [10]. In the next section, we propose a new approach to solve the problem.

III. Proposed Robust Joint Source and Relay Precoders Design

A. Primal Decomposition

We resort to the primal decomposition method [28] where the original optimization can be transferred into a subproblem and a master problem. The method is first to split unknown variables into two groups, and the variables in the first group are treated as known constants. Then, the variables in the second group are solved as the functions of the variables in the first group (the subproblem), and the cost function is reduced to a function of the variables in the first group. Finally, the variables in the first group can be solved (the master problem). For our problem, we let the subproblem be the optimization of $\mathbf{C}$ and $\mathbf{F}_S$ ($\mathbf{F}_R$ is treated as a known matrix), and the master problem be that of $\mathbf{F}_R$. However, since the two power constraints are mutually coupled, the primal decomposition cannot be applied directly. We then propose using a unitary structure for $\mathbf{F}_S$. To proceed, we reformulate (32) as

$$\begin{aligned}
&\min_{\mathbf{C},\mathbf{F}_S,\mathbf{F}_R} \text{Tr}\left\{\mathbf{E}\right\} = \min_{\mathbf{F}_R}\min_{\mathbf{C},\mathbf{F}_S} \text{Tr}\left\{\mathbf{E}\right\} \\
&s.t.\quad \mathbf{E} = \sigma_s^2\mathbf{C}\left(\mathbf{I}_N + \mathbf{F}_S^H\widehat{\mathbf{H}}^H\mathbf{R}_\Delta^{-1}\widehat{\mathbf{H}}\mathbf{F}_S\right)^{-1}\mathbf{C}^H, \\
&\mathbf{F}_S = \alpha_S\mathbf{U}_S, \quad \sigma_s^2\alpha_S^2 N \le P_{S,T}, \\
&\text{Tr}\left\{\mathbf{F}_R\left(\sigma_{n,r}^2\mathbf{I}_R + \sigma_s^2\alpha_S^2\widehat{\mathbf{H}}'_{SR}\widehat{\mathbf{H}}'^H_{SR}\right)\mathbf{F}_R^H\right\} \le P_{R,T}.
\end{aligned} \tag{33}$$

where $\widehat{\mathbf{H}}'_{SR} = \left(\widehat{\mathbf{H}}_{SR}\widehat{\mathbf{H}}_{SR}^H + \text{Tr}\left(\mathbf{\Psi}_{SR}\right)\mathbf{\Sigma}_{SR}\right)^{1/2}$, α_S is a scalar and $\mathbf{U}_S \in \mathbb{C}^{N\times N}$ is a unitary matrix to be further specified.

B. Subproblem Optimization

In the subproblem, the optimum $\mathbf{C}$ and $\mathbf{F}_S$ are first derived as a function of $\mathbf{F}_R$. Since we let the additional precoder $\mathbf{F}_S$ have a unitary structure, the cost function and the constraints can then be optimized with respect to $\mathbf{C}$, α_S, and $\mathbf{U}_S$. The subproblem optimization can then be written as:

$$\begin{aligned}&\min_{\mathbf{C}(\mathbf{F}_R),\alpha_S(\mathbf{F}_R),\mathbf{U}_S(\mathbf{F}_R)} \text{Tr}\{\mathbf{E}\}\\ &s.t.\quad \mathbf{E}=\sigma_s^2\mathbf{C}\left(\mathbf{I}_N+\alpha_S^2\mathbf{U}_S^H\widetilde{\mathbf{H}}^H\widetilde{\mathbf{H}}\mathbf{U}_S\right)^{-1}\mathbf{C}^H\\ &N\sigma_s^2\alpha_S^2\le P_{S,T}\\ &tr\left\{\mathbf{F}_R\left(\sigma_{n,r}^2\mathbf{I}_R+\alpha_S^2\sigma_s^2\widehat{\mathbf{H}}'_{SR}\widehat{\mathbf{H}}'^H_{SR}\right)\mathbf{F}_R^H\right\}\le P_{R,T},\end{aligned} \tag{34}$$

where

$$\begin{aligned}&\widetilde{\mathbf{H}}^H\widetilde{\mathbf{H}}:=\widehat{\mathbf{H}}^H\mathbf{R}_\Delta^{-1}\widehat{\mathbf{H}}\\ &=\sigma_s^2\left[\begin{array}{cc}\widehat{\mathbf{H}}_{SD}^H & \widehat{\mathbf{H}}_{SR}^H\mathbf{F}_R^H\widehat{\mathbf{H}}_{RD}^H\end{array}\right]\times\\ &\left[\begin{array}{cc}\left(\sigma_{n,d}^2\mathbf{I}_M+\sigma_s^2\alpha_S^2\text{Tr}\left(\boldsymbol{\Psi}_{SD}\right)\boldsymbol{\Sigma}_{SD}\right)^{-1} & \mathbf{0}\\ \mathbf{0} & (\Delta\mathbf{A}+\mathbf{A})^{-1}\end{array}\right]\times\\ &\left[\begin{array}{c}\widehat{\mathbf{H}}_{SD}\\ \widehat{\mathbf{H}}_{RD}\mathbf{F}_R\widehat{\mathbf{H}}_{SR}\end{array}\right]\\ &=\sigma_s^2\left(\sigma_{n,d}^{-2}\widehat{\mathbf{H}}_{SD}^H\left(\sigma_{n,d}^2\mathbf{I}_M+\sigma_s^2\alpha_S^2\text{Tr}\left(\boldsymbol{\Psi}_{SD}\right)\boldsymbol{\Sigma}_{SD}\right)^{-1}\widehat{\mathbf{H}}_{SD}+\right.\\ &\left.\widehat{\mathbf{H}}_{SR}^H\mathbf{F}_R^H\widehat{\mathbf{H}}_{RD}^H\left(\Delta\mathbf{A}+\mathbf{A}\right)^{-1}\widehat{\mathbf{H}}_{RD}\mathbf{F}_R\widehat{\mathbf{H}}_{SR}\right)\end{aligned} \tag{35}$$

with

$$\begin{aligned}\Delta\mathbf{A} &= \text{Tr}\left(\mathbf{F}_R\left(\sigma_s^2\alpha_S^2\widehat{\mathbf{H}}'_{SR}\widehat{\mathbf{H}}'^H_{SR}+\sigma_{n,r}^2\mathbf{I}_R\right)\mathbf{F}_R^H\boldsymbol{\Psi}_{RD}\right)\boldsymbol{\Sigma}_{RD}\\ &\quad+\sigma_s^2\alpha_S^2\text{Tr}\left(\boldsymbol{\Psi}_{SR}\right)\widehat{\mathbf{H}}_{RD}\mathbf{F}_R\boldsymbol{\Sigma}_{SR}\mathbf{F}_R^H\widehat{\mathbf{H}}_{RD}^H\end{aligned} \tag{36}$$

and

$$\mathbf{A}=\sigma_{n,r}^H\widehat{\mathbf{H}}_{RD}\mathbf{F}_R\mathbf{F}_R^H\widehat{\mathbf{H}}_{RD}^H+\sigma_{n,d}^2\mathbf{I}_M. \tag{37}$$

Note that primal decomposition can be further applied in the subproblem. Treating $\mathbf{C}$, $\mathbf{F}_R$ and $\mathbf{U}_S$ as known entities, we first optimize α_S. To do that, let's consider a maximum power property, stating that the cost function in (34) is monotonically decreasing in α_S. The proof of this property is provided in Appendix B. Using the property and denoting the optimum α_S as $\alpha_{S,opt}$, we can have

$$\alpha_{opt}=\sqrt{\frac{P_{S,T}}{N\sigma_s^2}}. \tag{38}$$

This is an intuitively appealing property since the noise power at the receiver can be reduced if the source transmit power is increased.

Substituting (38) into (34), we see that the relay power constraint is just a function of the relay precoder, and two power constraints are decoupled. Thus, the power constraint of the precoder can be moved to the master problem. If we treat $\mathbf{U}_S$ as a known matrix, the subproblem will be degenerated to the design in conventional point-to-point THP MIMO system, which can be written as

$$\min_{\mathbf{C}(\mathbf{F}_R)}\sigma_s^2\text{Tr}\left\{\mathbf{C}\left(\mathbf{I}_N+\frac{P_{S,T}}{N\sigma_s^2}\mathbf{U}_S^H\widetilde{\mathbf{H}}^H\widetilde{\mathbf{H}}\mathbf{U}_S\right)^{-1}\mathbf{C}^H\right\}. \tag{39}$$

The optimum solution $\mathbf{C}$ of (39), denoted as $\mathbf{C}_{opt}$, has been solved in [18] as:

$$\mathbf{C}_{opt}=\mathbf{D}\mathbf{L}^{-1}, \tag{40}$$

where

$$\mathbf{L}\mathbf{L}^H=\sigma_s^2\left(\mathbf{I}_N+\frac{P_{S,T}}{N\sigma_s^2}\mathbf{U}_S^H\widetilde{\mathbf{H}}^H\widetilde{\mathbf{H}}\mathbf{U}_S\right)^{-1} \tag{41}$$

is the Cholesky factorization of $\sigma_s^2\left(\mathbf{I}_N+\frac{P_{S,T}}{N\sigma_s^2}\mathbf{U}_S^H\widetilde{\mathbf{H}}^H\widetilde{\mathbf{H}}\mathbf{U}_S\right)^{-1}$ and $\mathbf{D}$ is a diagonal matrix scaling the diagonal elements of $\mathbf{C}_{opt}$ to unity. Substituting (40) into (39), we then have the cost function as

$$\text{Tr}\left\{\mathbf{C}\left(\mathbf{I}_N+\frac{P_{S,T}}{N\sigma_s^2}\mathbf{U}_S^H\widetilde{\mathbf{H}}^H\widetilde{\mathbf{H}}\mathbf{U}_S\right)^{-1}\mathbf{C}^H\right\}=\sum_{k=1}^N\mathbf{L}(k,k)^2 \tag{42}$$

which is a function of $\mathbf{U}_S$. We now can find $\mathbf{U}_S$ so that (42) is minimized. To start with, we decompose $\mathbf{U}_S$ as

$$\mathbf{U}_S=\mathbf{V}_{\widetilde{\mathbf{H}}}\mathbf{U}'_S, \tag{43}$$

where $\mathbf{V}_{\widetilde{\mathbf{H}}}\in\mathbb{C}^{N\times N}$ is the left singular matrices of $\widetilde{\mathbf{H}}$ and $\mathbf{U}'_S\in\mathbb{C}^{N\times N}$ is another unitary matrix. Note that this decomposition can always be conducted for a unitary matrix. Substituting (43) into (41), we then have

$$\begin{aligned}\mathbf{L}\mathbf{L}^H &= \sigma_s^2\left(\mathbf{I}_N+\frac{P_{S,T}}{N\sigma_s^2}\left(\mathbf{V}_{\widetilde{\mathbf{H}}}\mathbf{U}'_S\right)^H\widetilde{\mathbf{H}}^H\widetilde{\mathbf{H}}\mathbf{V}_{\widetilde{\mathbf{H}}}\mathbf{U}'_S\right)^{-1}\\ &= \mathbf{U}'^H_S\underbrace{\sigma_s^2\left(\mathbf{I}_N+\frac{P_{S,T}}{N\sigma_s^2}\boldsymbol{\Lambda}\right)^{-1}}_{:=\widetilde{\mathbf{D}}}\mathbf{U}'_S\end{aligned} \tag{44}$$

where $\boldsymbol{\Lambda}=\text{diag}\left\{\lambda_{\widetilde{\mathbf{H}},1},\ldots,\lambda_{\widetilde{\mathbf{H}},N}\right\}$ is a diagonal matrix consisted of the eigenvalues of $\widetilde{\mathbf{H}}^H\widetilde{\mathbf{H}}$. It is simple to see that $\widetilde{\mathbf{D}}$ is a diagonal matrix. Applying arithmetic-geometric inequality (AGI), we see that when $\mathbf{L}(i,i)=\mathbf{L}(j,j)$, $i\neq j$, (42) is minimized.

The remaining work is to find a proper $\mathbf{U}'_S$ such that $\mathbf{L}(i,i)=\mathbf{L}(j,j)$. This problem has been solved in [19] and the result is restated as follows. Consider the following decomposition:

$$\widetilde{\mathbf{D}}=\widetilde{\mathbf{D}}^{1/2}\widetilde{\mathbf{D}}^{1/2}, \tag{45}$$

where $\widetilde{\mathbf{D}}^{1/2}$ is the square-root matrix of $\widetilde{\mathbf{D}}$. Applying geometric mean decomposition (GMD) [16] on $\widetilde{\mathbf{D}}^{1/2}$, we can have

$$\widetilde{\mathbf{D}}^{1/2}=\mathbf{Q}\mathbf{R}\mathbf{P}^H, \tag{46}$$

where $\mathbf{Q}$ and $\mathbf{P}$ are some unitary matrices, and $\mathbf{R}$ is an upper triangular matrix with equal diagonal elements. Substituting (46) in (44), we then have

$$\mathbf{L}\mathbf{L}^H=\mathbf{U}'^H_S\widetilde{\mathbf{D}}\mathbf{U}'_S=\mathbf{U}'^H_S\mathbf{P}\mathbf{R}^H\mathbf{R}\mathbf{P}^H\mathbf{U}'_S. \tag{47}$$

Let $\mathbf{U}'_S=\mathbf{P}$. Equation (47) can be written as

$$\mathbf{L}\mathbf{L}^H=\mathbf{R}^H\mathbf{R}. \tag{48}$$

From (48), it is clear that if $\mathbf{L}=\mathbf{R}^H$, the diagonal elements of $\mathbf{L}$ will be all equal. Therefore, the optimal $\mathbf{F}_S$, denoted as $\mathbf{F}_{S,opt}$, can then be expressed as

$$\mathbf{F}_{S,opt}=\sqrt{\frac{P_{S,T}}{N\sigma_s^2}}\mathbf{V}_{\widetilde{\mathbf{H}}}\mathbf{P}. \tag{49}$$

From (42), the resultant MSE can be expressed as

$$\begin{aligned} J_{\min} &= \sum_{k=1}^{N} \mathbf{L}(k,k)^2 = \sum_{k=1}^{N} \mathbf{R}(k,k)^2 \\ &= N\sigma_s^2 \prod_{k=1}^{N} \left(\frac{P_{S,T}}{N\sigma_s^2} \lambda_{\widetilde{\mathbf{H}},k} + 1 \right)^{-1/N}. \end{aligned} \tag{50}$$

Now, the problem becomes the minimization of (50) which is the master problem. It is noteworthy that if the unitary $\mathbf{F}_S$ is not included in the design, the cost function in (42) can be reformulated with $\mathbf{F}_S = \mathbf{I}_N$. In this scenario, the related minimum MSE will be larger than that in (50). Also, the optimization problem is much more cumbersome to deal with. We note here that the original THP in [18] does not include the unitary matrix $\mathbf{F}_S$[1].

C. Master Problem Optimization

To solve the master problem, let's first consider the following equivalence:

$$\begin{aligned} &\min_{\mathbf{F}_R} N\sigma_s^2 \prod_{k=1}^{N} \left(\frac{P_{S,T}}{N\sigma_s^2} \lambda_{\widetilde{\mathbf{H}},k} + 1 \right)^{-1/N} \\ &= \max_{\mathbf{F}_R} \det \left(\left(\frac{N\sigma_s^2}{P_{S,T}} \mathbf{I}_N + \widetilde{\mathbf{H}}^H \widetilde{\mathbf{H}} \right) \right). \end{aligned} \tag{51}$$

The result in (51) can be easily obtained since

$$\det \left(\mathbf{I}_N + \frac{P_{S,T}}{N\sigma_s^2} \widetilde{\mathbf{H}}^H \widetilde{\mathbf{H}} \right) = \prod_{k=1}^{N} \left(\frac{P_{S,T}}{N\sigma_s^2} \lambda_{\widetilde{\mathbf{H}},k} + 1 \right). \tag{52}$$

Using (51) and (35), we can then reformulate the master problem as

$$\begin{aligned} &\max_{\mathbf{F}_R} \det \Bigg(\mathbf{I}_N + \frac{P_{S,T}}{N\sigma_{n,d}^2} \widehat{\mathbf{H}}_{SD}^H \times \\ &\left(\sigma_{n,d}^2 \mathbf{I}_M + \frac{P_{S,T}}{N} \mathrm{Tr}\left(\mathbf{\Psi}_{SD}\right) \mathbf{\Sigma}_{SD} \right)^{-1} \widehat{\mathbf{H}}_{SD} + \\ &\frac{P_{S,T}}{N} \widehat{\mathbf{H}}_{SR}^H \mathbf{F}_R^H \widehat{\mathbf{H}}_{RD}^H \left(\Delta\mathbf{A} + \mathbf{A}\right)^{-1} \widehat{\mathbf{H}}_{RD} \mathbf{F}_R \widehat{\mathbf{H}}_{SR} \Bigg) \\ &s.t. \\ &\mathrm{Tr}\left\{ \mathbf{F}_R \left(\sigma_{n,r}^2 \mathbf{I}_R + \frac{P_{S,T}}{N} \widehat{\mathbf{H}}'_{SR} \widehat{\mathbf{H}}'^H_{SR} \right) \mathbf{F}_R^H \right\} \le P_{R,T}. \end{aligned} \tag{53}$$

Take a look at (33) and (53) and we can readily find that although the number of the unknowns are reduced, the utility[2] function is a complicated function of $\mathbf{F}_R$ and yet the optimization is not convex. As a result, the problem in (53) is difficult to solve, even with a numerical method [28]. To provide a solution, instead of the original utility function, we propose using a lower bound of the function. By the lower bound, we can reformulate the master optimization as

$$\begin{aligned} &\max_{\mathbf{F}_R} \det \Bigg(\mathbf{I}_N + \frac{P_{S,T}}{N\sigma_{n,d}^2} \widehat{\mathbf{H}}_{SD}^H \times \\ &\left(\sigma_{n,d}^2 \mathbf{I}_M + \frac{P_{S,T}}{N} \mathrm{Tr}\left(\mathbf{\Psi}_{SD}\right) \mathbf{\Sigma}_{SD} \right)^{-1} \widehat{\mathbf{H}}_{SD} + \\ &\frac{P_{S,T}}{N} \widehat{\mathbf{H}}_{SR}^H \mathbf{F}_R^H \widehat{\mathbf{H}}_{RD}^H \left(\Delta\mathbf{A}' + \mathbf{A}\right)^{-1} \widehat{\mathbf{H}}_{RD} \mathbf{F}_R \widehat{\mathbf{H}}_{SR} \Bigg) \\ &s.t. \\ &\mathrm{Tr}\left\{ \mathbf{F}_R \left(\sigma_{n,r}^2 \mathbf{I}_R + \frac{P_{S,T}}{N} \widehat{\mathbf{H}}'_{SR} \widehat{\mathbf{H}}'^H_{SR} \right) \mathbf{F}_R^H \right\} \le P_{R,T}, \end{aligned} \tag{54}$$

where

$$\begin{aligned} \Delta\mathbf{A}' = {} & P_{R,T} \lambda_{\max}(\mathbf{\Psi}_{RD}) \lambda_{\max}(\mathbf{\Sigma}_{RD}) \mathbf{I}_M + \\ & \frac{P_{S,T}}{N} \mathrm{Tr}(\mathbf{\Psi}_{SR}) \lambda_{\max}(\mathbf{\Sigma}_{SR}) \widehat{\mathbf{H}}_{RD} \mathbf{F}_R \mathbf{F}_R^H \widehat{\mathbf{H}}_{RD}^H \end{aligned} \tag{55}$$

Particularly, the lower bound of the utility function is equal to the actual one when the transmit or receive transmit pairs are uncorrelated. The detailed derivation is given in Appendix C.

As we will see in latter development, $\Delta\mathbf{A}'$ in (54) is easier to deal with, compared to $\Delta\mathbf{A}$ in (53). The optimization in (54) remains unsolvable since the utility function is still a complicated function of $\mathbf{F}_R$. In addition, the problem is still not convex. To overcome the problem, we use the Hardamard inequality, described in the following Lemma.

Lemma 1 [27]: Let $\mathbf{M} \in \mathbb{C}^{N\times N}$ be a positive definite matrix, then

$$\det(\mathbf{M}) \le \prod_{i=1}^{N} \mathbf{M}(i,i), \tag{56}$$

where $\mathbf{M}(i,i)$ denotes the ith diagonal element of $\mathbf{M}$. The equality in (56) holds when $\mathbf{M}$ is a diagonal matrix. This suggests a diagonalization of the matrix in the utility function of (54), achieving the maximum of (54). To conduct the diagonalization, we need another lemma described below.

Lemma 2 [27]: Let $\mathbf{P} \in \mathbb{C}^{N\times N}$ be a positive definite matrix and $\mathbf{J} \in \mathbb{C}^{N\times N}$, then

$$\det\left(\mathbf{P} + \mathbf{J}\right) = \det\left(\mathbf{P}\right) \det\left(\mathbf{I}_N + \mathbf{P}^{-1/2} \mathbf{J} \mathbf{P}^{-1/2} \right). \tag{57}$$

We let

$$\begin{aligned} \mathbf{J} &= \frac{P_{S,T}}{N} \widehat{\mathbf{H}}_{SR}^H \mathbf{F}_R^H \widehat{\mathbf{H}}_{RD}^H \left(\Delta\mathbf{A}' + \mathbf{A}\right)^{-1} \widehat{\mathbf{H}}_{RD} \mathbf{F}_R \widehat{\mathbf{H}}_{SR}, \\ \mathbf{P} &= \Bigg(\mathbf{I}_N + \frac{P_{S,T}}{N\sigma_{n,d}^2} \widehat{\mathbf{H}}_{SD}^H \times \\ &\quad \left(\sigma_{n,d}^2 \mathbf{I}_M + \frac{P_{S,T}}{N} \mathrm{Tr}\left(\mathbf{\Psi}_{SD}\right) \mathbf{\Sigma}_{SD} \right)^{-1} \widehat{\mathbf{H}}_{SD} \Bigg) \end{aligned} \tag{58}$$

in (57). Then, we have the following equivalence:

$$\begin{aligned} &\arg\max_{\mathbf{F}_R} \det\left(\mathbf{P} + \mathbf{J}\right) = \\ &\arg\max_{\mathbf{F}_R} \det \Big(\mathbf{I}_N + \widehat{\mathbf{H}}''^H_{SR} \mathbf{F}_R^H \widehat{\mathbf{H}}_{RD}^H \times \\ &\qquad \left(\Delta\mathbf{A}' + \mathbf{A}\right)^{-1} \widehat{\mathbf{H}}_{RD} \mathbf{F}_R \widehat{\mathbf{H}}''_{SR} \Big), \end{aligned} \tag{59}$$

[1]It is noteworthy here that the THP design in [19] does cascade a precoder. However, the precoder is not restricted to a unitary matrix and the purpose is different from ours.

[2]In general, when a problem is to be minimized, the objective function is referred to as a cost function. In contrast, when the problem is to be maximized, the objective function is referred to as a utility function.

where

$$\widehat{\mathbf{H}}''_{SR} = \sqrt{\frac{P_{S,T}}{N}}\widehat{\mathbf{H}}_{SR}\left(\mathbf{I}_N + \frac{P_{S,T}}{N\sigma^2_{n,d}}\widehat{\mathbf{H}}^H_{SD}\times \left(\sigma^2_{n,d}\mathbf{I}_M + \frac{P_{S,T}}{N}\text{Tr}(\mathbf{\Psi}_{SD})\mathbf{\Sigma}_{SD}\right)^{-1}\widehat{\mathbf{H}}_{SD}\right)^{-1/2} \quad (60)$$

Note here that $\det(\mathbf{P})$ is ignored since it is not a function of $\mathbf{F}_R$. Using (59) and the relay power constraint in (54), we then have the master optimization as follows:

$$\max_{\mathbf{F}_R}\det\left(\mathbf{I}_N + \widehat{\mathbf{H}}''^H_{SR}\mathbf{F}^H_R\widehat{\mathbf{H}}^H_{RD}\times \left(\alpha\mathbf{I}_M + \beta\widehat{\mathbf{H}}_{RD}\mathbf{F}_R\mathbf{F}^H_R\widehat{\mathbf{H}}^H_{RD}\right)^{-1}\widehat{\mathbf{H}}_{RD}\mathbf{F}_R\widehat{\mathbf{H}}''_{SR}\right)$$
$$s.t.\quad \text{Tr}\left\{\mathbf{F}_R\left(\sigma^2_{n,r}\mathbf{I}_R + \frac{P_{S,T}}{N}\widehat{\mathbf{H}}'_{SR}\widehat{\mathbf{H}}'^H_{SR}\right)\mathbf{F}^H_R\right\} \le P_{R,T}, \quad (61)$$

where

$$\alpha = P_{R,T}\lambda\max(\mathbf{\Psi}_{RD})\lambda\max(\mathbf{\Sigma}_{RD}) + \sigma^2_{n,d} \quad \text{and}$$
$$\beta = \frac{P_{R,T}}{N}\text{Tr}(\mathbf{\Psi}_{SR})\lambda\max(\mathbf{\Sigma}_{SR}) + \sigma^2_{n,r}. \quad (62)$$

From Lemma 1, we see that if $\left(\mathbf{I}_N + \widehat{\mathbf{H}}''^H_{SR}\mathbf{F}^H_R\widehat{\mathbf{H}}^H_{RD}\left(\Delta\mathbf{A}' + \mathbf{A}\right)^{-1}\widehat{\mathbf{H}}_{RD}\mathbf{F}_R\widehat{\mathbf{H}}''_{SR}\right)$ is diagonalized, the utility function in (61) is maximized. Note that the optimality may not be held when the power constraint in (61) is considered. However, we still use the diagonalization operation here, facilitating the derivation of a solution. To do that, we can let the relay precoder have a certain structure. Considering following singular value decomposition (SVD):

$$\widehat{\mathbf{H}}_{RD} = \mathbf{U}_{rd}\mathbf{\Sigma}_{rd}\mathbf{V}^H_{rd}; \quad (63)$$

$$\widehat{\mathbf{H}}''_{SR} = \mathbf{U}''_{sr}\mathbf{\Sigma}''_{sr}\mathbf{V}''^H_{sr} \quad (64)$$

where $\mathbf{U}_{rd} \in \mathbb{C}^{M\times M}$ and $\mathbf{U}''_{sr} \in \mathbb{C}^{R\times R}$ are left singular matrices of $\widehat{\mathbf{H}}_{RD}$ and $\widehat{\mathbf{H}}''_{SR}$, respectively; $\mathbf{\Sigma}_{rd} \in \mathbb{R}^{M\times R}$ and $\mathbf{\Sigma}''_{sr} \in \mathbb{R}^{R\times N}$ are the diagonal singular value matrices of $\widehat{\mathbf{H}}_{RD}$ and $\widehat{\mathbf{H}}''_{SR}$, respectively; $\mathbf{V}^H_{rd} \in \mathbb{C}^{R\times R}$ and $\mathbf{V}''_{sr} \in \mathbb{C}^{N\times N}$ are the right singular matrices of $\widehat{\mathbf{H}}_{RD}$ and $\widehat{\mathbf{H}}''_{SR}$, respectively. Substituting (63) and (64) into (61), we can have the utility function as

$$\det\left(\mathbf{I}_N + \widehat{\mathbf{H}}''^H_{SR}\mathbf{F}^H_R\widehat{\mathbf{H}}^H_{RD}\left(\Delta\mathbf{A}' + \mathbf{A}\right)^{-1}\widehat{\mathbf{H}}_{RD}\mathbf{F}_R\widehat{\mathbf{H}}''_{SR}\right)$$
$$= \det\left(\mathbf{I}_N + \mathbf{\Sigma}''^H_{sr}\mathbf{U}''^H_{sr}\mathbf{F}^H_R\mathbf{V}_{rd}\mathbf{\Sigma}^H_{rd}\left(\alpha\mathbf{I}_M + \beta\mathbf{\Sigma}_{rd}\mathbf{V}^H_{rd}\times \mathbf{F}_R\mathbf{F}^H_R\mathbf{V}_{rd}\mathbf{\Sigma}^H_{rd}\right)^{-1}\mathbf{\Sigma}_{rd}\mathbf{V}^H_{rd}\mathbf{F}_R\mathbf{U}''_{sr}\mathbf{\Sigma}''_{sr}\right). \quad (65)$$

It turns out that if the $\mathbf{F}_R$ have the following structure, a full diagonalization of the utility matrix in (65) can be achieved:

$$\mathbf{F}_R = \mathbf{V}_{rd}\mathbf{\Sigma}_r\mathbf{U}''^H_{sr}, \quad (66)$$

where $\mathbf{\Sigma}_r$ is a diagonal matrix with the ith diagonal element of $\sigma_{r,i}$, $i = 1\cdots\kappa$ (κ=min$\{N, R\}$). Let $\sigma_{rd,i}$ and $\sigma''_{sr,i}$ be the ith diagonal element of $\mathbf{\Sigma}_{rd}$ and $\mathbf{\Sigma}''_{sr}$, respectively. Using (63), (64) and (66) in (61) and taking the ln operation to the utility function, we can then rewrite (61) as:

$$\max_{p_{r,i},\ 1\le i\le\kappa}\sum_{i=1}^{\kappa}\ln\left(1 + \frac{p_{r,i}\sigma^2_{rd,i}\sigma''_{sr,i}}{\alpha + \beta p_{r,i}\sigma^2_{rd,i}}\right)$$
$$s.t.\quad \sum_{i=1}^{\kappa}p_{r,i}\left(\mathbf{D}_{sr}(i,i) + \sigma^2_{n,r}\right) \le P_{R,T},\quad p_{r,i}\ge 0,\ \ \forall\, i, \quad (67)$$

where $\mathbf{D}_{sr}(i,i)$ is the ith diagonal element of $\mathbf{D}_{sr} = \frac{P_{S,T}}{N}\mathbf{U}''^H_{sr}\left(\widehat{\mathbf{H}}_{SR}\widehat{\mathbf{H}}^H_{SR} + \text{Tr}\left(\mathbf{\Psi}_{SR}\right)\mathbf{\Sigma}_{SR}\right)\mathbf{U}''_{sr}$, and $p_{r,i} = \sigma^2_{r,i}$. The utility function now is simplified to a function of scalar variables. Since the utility function and the inequalities are all concave for $p_{r,i} \ge 0$ [28], (67) is a standard concave optimization problem. As a result, the optimum solution of $p_{r,i}$, $i = 1,\ldots,\kappa$, can be solved by means of KKT conditions given as (68) at the top of the next page, where μ is chosen to satisfy the power constraint in (67). The derivation of (68) is summarized in Appendix D.

The computational complexity of the proposed robust TH precoders mainly involve SVD, GMD, and matrix inversion operations. The overall computational complexity, measured in terms of FLOPs, are summarized as Table I.

IV. Simulation Results

We consider a three-node AF MIMO relay system with $N = R = M = 4$, and model the channel estimation errors with the covariance matrices as [20], [25]:

$$\mathbf{\Psi}_{SR} = \mathbf{\Psi}_{RD} = \mathbf{\Psi}_{SD} = \begin{bmatrix} 1 & \delta & \delta^2 & \delta^3 \\ \delta & 1 & \delta & \delta^2 \\ \delta^2 & \delta & 1 & \delta \\ \delta^3 & \delta^2 & \delta & 1 \end{bmatrix} \quad (69)$$

and

$$\mathbf{\Sigma}_{SR} = \mathbf{\Sigma}_{RD} = \mathbf{\Sigma}_{SD} = \sigma^2_e\begin{bmatrix} 1 & \gamma & \gamma^2 & \gamma^3 \\ \gamma & 1 & \gamma & \gamma^2 \\ \gamma^2 & \gamma & 1 & \gamma \\ \gamma^3 & \gamma^2 & \gamma & 1 \end{bmatrix} \quad (70)$$

where δ and γ denote the correlation coefficients, and σ^2_e the estimation error variance. The covariance matrices in (69) and (70) can be obtained if we use the channel estimation method proposed in [20]. The estimated channels, $\widehat{\mathbf{H}}_{SR}$, $\widehat{\mathbf{H}}_{RD}$, and $\widehat{\mathbf{H}}_{SD}$, are generated by the following distributions:

$$\text{vec}\left(\widehat{\mathbf{H}}_{SR}\right) \sim \mathcal{CN}\left(\mathbf{0}_{NR\times 1}, \frac{1-\sigma^2_e}{\sigma^2_e}\mathbf{\Sigma}_{SR}\otimes\mathbf{\Psi}^T_{SR}\right) \quad (71)$$

$$\text{vec}\left(\widehat{\mathbf{H}}_{RD}\right) \sim \mathcal{CN}\left(\mathbf{0}_{MR\times 1}, \frac{1-\sigma^2_e}{\sigma^2_e}\mathbf{\Sigma}_{RD}\otimes\mathbf{\Psi}^T_{RD}\right) \quad (72)$$

$$\text{vec}\left(\widehat{\mathbf{H}}_{SD}\right) \sim \mathcal{CN}\left(\mathbf{0}_{MN\times 1}, \frac{1-\sigma^2_e}{\sigma^2_e}\mathbf{\Sigma}_{SD}\otimes\mathbf{\Psi}^T_{SD}\right) \quad (73)$$

As mentioned, the relationships of the actual and estimated channels can then be expressed as $\mathbf{H}_{SR} = \widehat{\mathbf{H}}_{SR} + \Delta\mathbf{H}_{SR}$, $\mathbf{H}_{RD} = \widehat{\mathbf{H}}_{RD} + \Delta\mathbf{H}_{RD}$, and $\mathbf{H}_{SD} = \widehat{\mathbf{H}}_{SD} + \Delta\mathbf{H}_{SD}$. Let SNR$_{sr}$, SNR$_{sd}$ and SNR$_{rd}$ denote the received signal-to-noise ratio (SNR) at each relay antenna in the first phase, that at each destination antenna in the first transmission phase, and that at

$$p_{r,i} = \left[\sqrt{\frac{\mu}{\sigma_{rd,i}^2 \frac{\beta}{\alpha}\left(\frac{\beta}{\sigma_{sr,i}''^2}+1\right)\left(\sigma_{n,r}^2+\mathbf{D}_{sr}(i,i)\right)} + \frac{\frac{1}{4}}{\sigma_{rd,i}^4 \frac{\beta^2}{\alpha^2}\left(\frac{\beta^2}{\sigma_{sr,i}''^2}+1\right)^2}} - \frac{\alpha + \frac{\alpha\sigma_{sr,i}''^2}{2\beta}}{\sigma_{rd,i}^2\left(\beta+\sigma_{sr,i}''^2\right)}\right]^+ \quad (68)$$

TABLE I
COMPLEXITY OF PROPOSED TH SOURCE AND RELAY PRECODERS (MMSE RECEIVER).

Step	Operation	FLOPs
1	$\widehat{\mathbf{H}}''_{SR}$, (60)	$O\left(N^2(N+R+M)\right)$
2	SVD $\widehat{\mathbf{H}}_{RD} = \mathbf{U}_{rd}\boldsymbol{\Sigma}_{rd}\mathbf{V}_{rd}^H$, (63)	$O\left(MR^2+R^3\right)$
3	SVD $\widehat{\mathbf{H}}''_{SR} = \mathbf{U}''_{sr}\boldsymbol{\Sigma}''_{sr}\mathbf{V}''^H_{sr}$, (64)	$O\left(RN^2+N^3\right)$
4	$\boldsymbol{\Sigma}_r$, (68)	$O(\kappa I_r)$
5	$\mathbf{F}_R$, (66)	$O(R^3)$
6	$\widetilde{\mathbf{H}} = \mathbf{R}_{\Delta}^{-1/2}\widehat{\mathbf{H}}$, (36)	$O\left(M^2(M+N)\right)$
7	SVD $\widetilde{\mathbf{H}}$	$O\left(N^2(M+N)\right)$
8	GMD $\widetilde{\mathbf{D}}^{1/2} = \mathbf{QRP}^H$ (46) $\mathbf{L} = \mathbf{R}^H$ (48)	$O(N^3)$
9	$\mathbf{C}_{opt} = \mathbf{DL}^{-1}$, (40)	$O(N^3)$
10	$\mathbf{F}_{S,opt}$, (49)	$O(N^3)$
I_r: denotes the iteration number of the water-filling process for for computing $\sigma_{r,i}$ (68)		

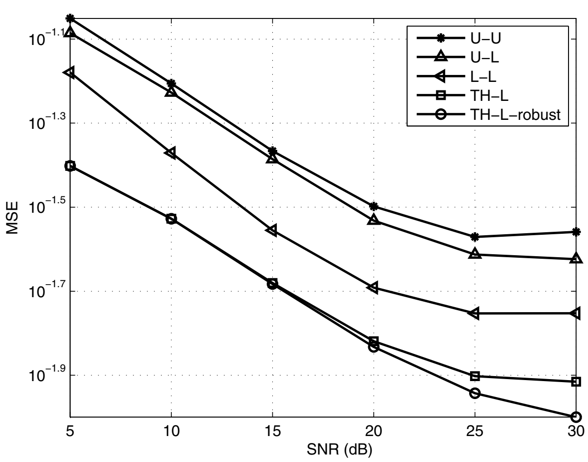


Fig. 2. MSE performance comparison for existing precoded systems and proposed TH source and linear relay precoded system ($\delta = \gamma = 0, \sigma_e^2 = 0.003$).

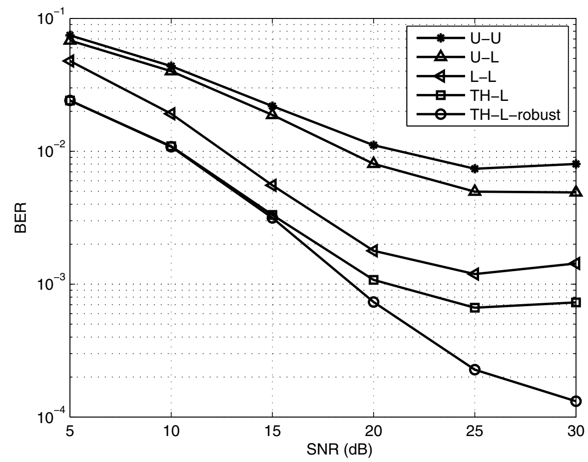


Fig. 3. BER performance comparison for existing precoded systems and proposed TH source and linear relay precoded system ($\delta = \gamma = 0, \sigma_e^2 = 0.003$).

each destination antenna in the second transmission phase, respectively. We use 16-QAM as the modulation scheme.

In the first set of simulations, we let SNR$_{sr}$ = 30, SNR$_{sd}$=15 dB, and SNR$_{rd}$ be varied. We also let $\delta = \gamma = 0$ and $\sigma_e^2 = 0.003$. Fig. 2 and Fig. 3 show the MSE and BER performances, respectively, for (a) an un-precoded system (U-U), (b) the relay precoded system (U-L) [4], (c) the linear source and relay precoded system (L-L) [7], (d) the TH source and linear relay precoded system (TH-L) [10], and (e) the proposed robust TH source and linear relay precoded system (TH-L-robust). All the considered systems, (a)-(e), use the MMSE receiver. Note that the system in [4] only considers the relay link. For fair comparison, we include the direct link in the MMSE receiver. From the figures, we can observe that the un-precoded system is inferior to precoded systems. The nonlinear source precoded systems are superior to the linear ones. Since TH-L-robust takes the CSI uncertainty into consideration, it performs better than TH-L. Interestingly, we can find that the performance of non-robust systems slightly degrade at the high SNR region. This is because noise can somehow offset the CSI uncertainty. Since σ_e is fixed at all SNR range, the robustness at the high SNR region is reduced. A similar result is also observed in conventional MIMO systems [22].

In the second set of simulations, we evaluate the precoding performance in a two-hop system. Fig. 4 and Fig. 5 show the MSE and BER performances for the cases that SNR$_{sr}$=35 dB, SNR$_{rd}$ is varied, $\delta = \gamma = 0$ and $\sigma_e^2 = 0$ ($\sigma_e^2 = 0.003$), respectively. Here, we further incorporate the system with the robust relay precoder in [25] (U-L-robust) for comparison. As we can see, both U-L-robust and TH-L-robust are degenerated

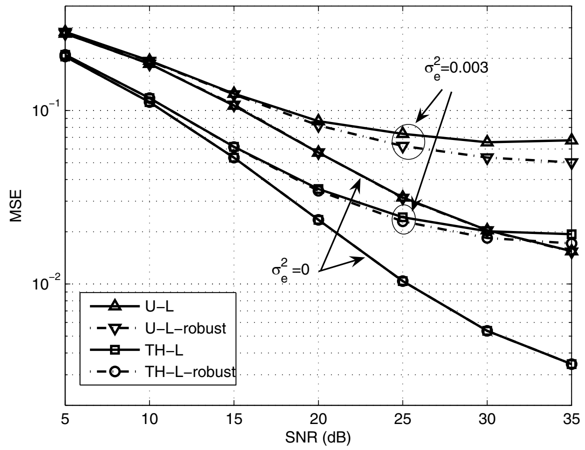


Fig. 4. MSE performance comparison for existing relay precoded systems and proposed TH source and linear relay precoded system (SNR$_{sr}$=35 dB, $\delta = \gamma = 0, \sigma_e^2 = 0/\sigma_e^2 = 0.003$).

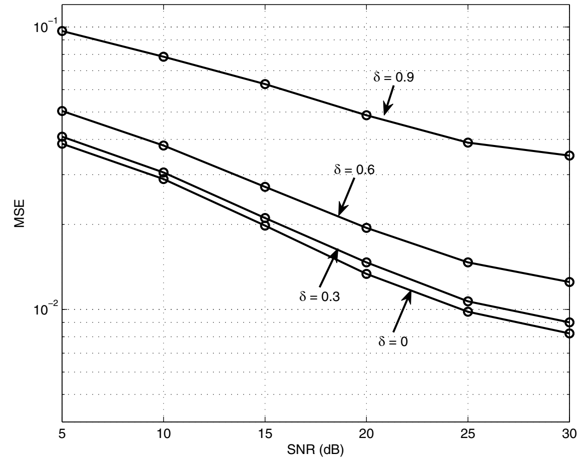


Fig. 6. MSE performance comparison for proposed precoded system with different δ ($\gamma = 0, \sigma_e^2 = 0.002$).

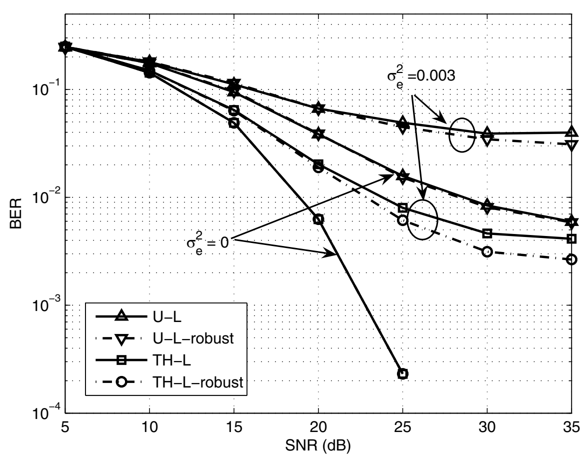


Fig. 5. BER performance comparison for existing relay precoded system and proposed TH source and linear relay precoded system (SNR$_{sr}$=35 dB, $\delta = \gamma = 0, \sigma_e^2 = 0/\sigma_e^2 = 0.003$).

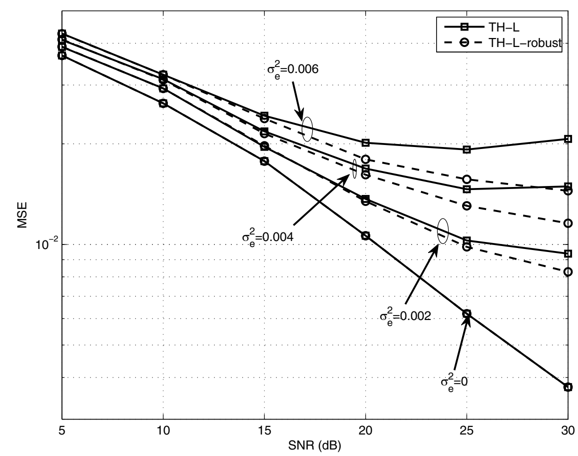


Fig. 7. MSE performance comparison for proposed precoded system with different σ_e^2 ($\delta = \gamma = 0$).

to U-L and TH-L, respectively if CSIs are perfect. Since U-L-robust only considers a relay precoder, its performance is inferior to proposed TH-L-robust no matter CSIs are perfectly available or not.

In the third set of simulations, we evaluate the MSE performance of the proposed method for some settings. Here, let $\gamma = 0$, $\sigma_e^2 = 0.002$ and δ be varied. Also let SNR$_{sr}$=30, SNR$_{sd}$=15 dB and SNR$_{rd}$ be varied. Fig. 6 shows the simulation results. From this figure, we can see that the performance of the proposed method is improved along with the decrease of δ. This is because as δ becomes smaller, $\mathbf{\Sigma}_{SR}$, $\mathbf{\Sigma}_{RD}$, and $\mathbf{\Sigma}_{SD}$ approach to $\sigma_e^2\mathbf{I}$. As a result, the actual utility function is closer to the lower bound in (94).

In the fourth set of simulations, we evaluate the MSE performance of the proposed method under the scenario that $\delta = \gamma = 0$ and σ_e^2 is varied. The SNR of each link is set as that in the previous case, and Fig. 7 shows the simulation results. As we can see, the performance of TH-L and TH-L-robust is significantly degraded when σ_e^2 is large. The performance gap between the TH-L and TH-L-robust is increased when the CSI uncertainty is increased. The performance of the TH-L system degrades at the high SNR region. The phenomenon is similar to that in Fig. 2.

In the fifth set of simulations, we evaluate the impact of imperfect second-order statistics. We assume that there exists errors in the transmit and receive correlation matrices, and use the error model proposed in [31]. Using $\mathbf{R}_{T,SR}$ as an example, we describe the model as follows. Let $\widehat{\mathbf{R}}_{T,SR}$ be an estimate of $\mathbf{R}_{T,SR}$. Then,

$$\widehat{\mathbf{R}}_{T,SR} = (1 - \epsilon)\, \mathbf{R}_{T,SR}, \tag{74}$$

where ϵ indicates the level of the estimation error. With the channel estimation algorithm in [20], $\mathbf{\Psi}_{SR} = \mathbf{R}_{T,SR}$. Thus, we can model the error in $\mathbf{\Psi}_{SR}$ as

$$\widehat{\mathbf{\Psi}}_{SR} = (1 - \epsilon)\, \mathbf{\Psi}_{SR} \tag{75}$$

where $\widehat{\mathbf{\Psi}}_{SR}$ is the estimate of $\mathbf{\Psi}_{SR}$. Similar modeling can be also applied to $\widehat{\mathbf{\Psi}}_{RD}$, $\widehat{\mathbf{\Psi}}_{SD}$, $\widehat{\mathbf{\Sigma}}_{SR}$, $\widehat{\mathbf{\Sigma}}_{RD}$, and $\widehat{\mathbf{\Sigma}}_{SD}$. In

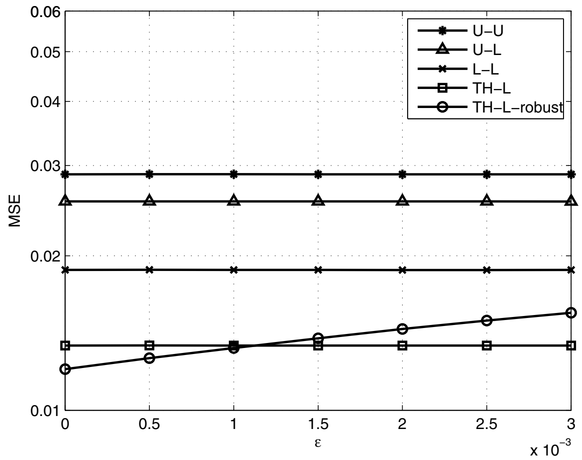


Fig. 8. MSE performance comparison for proposed precoded system with different ϵ (σ_e^2 = 0.003, $\delta = 0.3$, $\gamma = 0$, SNR$_{sr}$ = 30, SNR$_{rd}$ = 25, SNR$_{sd}$ = 15).

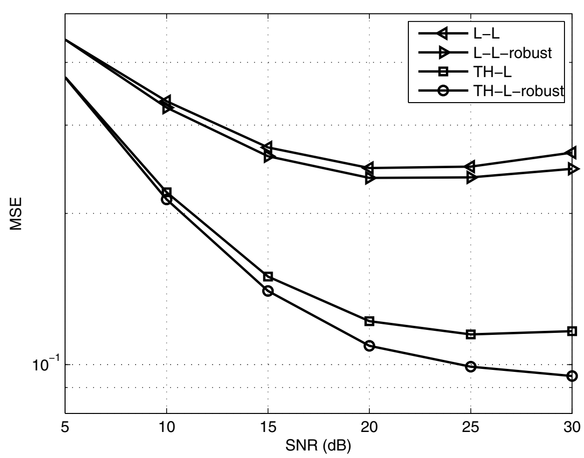


Fig. 9. MSE performance comparison for existing joint precoded systems and proposed TH source and linear relay precoded system (σ_e^2 = 0.02, SNR$_{rd}$ = 20).

the simulations, we let $\sigma_e^2 = 0.003$, $\delta = 0.3$, $\gamma = 0$, SNR$_{sr}$=30, SNR$_{rd}$=25, SNR$_{sd}$ = 15 dB, $\widehat{\boldsymbol{\Sigma}}_{SR} = \widehat{\boldsymbol{\Sigma}}_{RD} = \widehat{\boldsymbol{\Sigma}}_{SD}$, and $\widehat{\boldsymbol{\Psi}}_{SR} = \widehat{\boldsymbol{\Psi}}_{RD} = \widehat{\boldsymbol{\Psi}}_{SD}$. Fig. 8 shows the simulated MSE performance versus ϵ. From the figure, we can see that as the error increases, the MSE of TH-L-robust increases. Since σ_e is fixed, the performance curves of other non-robust algorithms are not changed by ϵ. Up to some level, the MSE of TH-L-robust will exceed that of TH-L.

In the last set of simulations, we compare the current robust joint precoders [32] with our proposed method. Since the precoders proposed in [32] considers the dual-hop scenario, we only consider the relay link in Fig. 9. In this simulation, we set $N = R = M = 4$, $\delta = \gamma = 0$, SNR$_{rd}$ = 20 dB, $\sigma_e^2 = 0.02$, and vary SNR$_{sr}$. Four joint precoders designs are compared, L-L, the robust design in [32] (L-L-robust) , TH-L, and TH-L-robust. As we can see, the robust designs, either linear or nonlinear design, are better than non-robust designs. Also, TH-L-robust are superior to L-L-robust due to the fact that the nonlinear TH precoder is adopted at the source.

V. Conclusion

In this paper, we consider a robust transceiver design for AF MIMO relay system with imperfect CSIs. The transceiver consists of a THP at the source, a linear precoder at the relay, and a MMSE receiver at the destination. The design can be easily formulated as an optimization problem. However, the problem is difficult to solve due to the complicated objective function and constraints. We then propose using the primal decomposition technique transferring the problem into a subproblem and a master problem. With the aid of a lower bound of the objective function, the closed-form solution can then be derived by the KKT conditions. Simulations show that the proposed robust design outperforms the existing non-robust linear or nonlinear designs. For non-robust designs, the effect of the mismatched CSIs is more severe in high SNR regions in which the performance improvement of the proposed design is more significant. The performance gain comes from the fact that the proposed design takes the second order statistics of the channel estimation error into account. In real-world applications, the second order statistics can be derived from the history of the channel estimates. In mobility environments, the channel estimation error may become larger due to the time lag in the feedback channel. An effective robust design will rely on the model of the channel estimation error. How to build the model and complete the design can serve as an interesting subject for further research.

Appendix A
Derivation of the MSE in (19)

Expanding the MSE in (18) as that in (19), we consider the first term in (18) expressed as

$$\begin{aligned}
&\mathrm{Tr}\left\{E_\Delta\left\{\sigma_s^2\mathbf{G}\mathbf{H}\mathbf{F}_S\mathbf{F}_S^H\mathbf{H}^H\mathbf{G}^H\right\}\right\}\\
&=\mathrm{Tr}\left\{\sigma_s^2\mathbf{G}E_\Delta\left\{\mathbf{H}\mathbf{F}_S\mathbf{F}_S^H\mathbf{H}^H\right\}\mathbf{G}^H\right\}\\
&=\sigma_s^2\mathbf{G}\begin{bmatrix}\mathbf{M}_{1,1} & \mathbf{M}_{1,2}\\ \mathbf{M}_{2,1} & \mathbf{M}_{2,2}\end{bmatrix}\mathbf{G}^H,\\
&\text{where}\\
&\mathbf{M}_{1,1}=E_\Delta\left\{\mathbf{H}_{SD}\mathbf{F}_S\mathbf{F}_S^H\mathbf{H}_{SD}^H\right\}\\
&\mathbf{M}_{1,2}=E_\Delta\left\{\mathbf{H}_{SD}\mathbf{F}_S\mathbf{F}_S^H\mathbf{H}_{SR}^H\mathbf{F}_R^H\mathbf{H}_{RD}^H\right\}\\
&\mathbf{M}_{2,1}=E_\Delta\left\{\mathbf{H}_{RD}\mathbf{F}_R\mathbf{H}_{SR}\mathbf{F}_S\mathbf{F}_S^H\mathbf{H}_{SD}^H\right\}\\
&\mathbf{M}_{2,2}=E_\Delta\left\{\mathbf{H}_{RD}\mathbf{F}_R\mathbf{H}_{SR}\mathbf{F}_S\mathbf{F}_S^H\mathbf{H}_{SR}^H\mathbf{F}_R^H\mathbf{H}_{RD}^H\right\}.
\end{aligned}\tag{76}$$

Since $\Delta\mathbf{H}_{SD}$ is multivariate complex Gaussian distributed (with zero mean), we can have the first diagonal matrix in (76) as [30]

$$\begin{aligned}
&E_\Delta\left\{\mathbf{H}_{SD}\mathbf{F}_S\mathbf{F}_S^H\mathbf{H}_{SD}^H\right\}\\
&=E_\Delta\left\{\left(\widehat{\mathbf{H}}_{SD}+\Delta\mathbf{H}_{SD}\right)\mathbf{F}_S\mathbf{F}_S^H\left(\widehat{\mathbf{H}}_{SD}+\Delta\mathbf{H}_{SD}\right)^H\right\}\\
&=\mathrm{Tr}\left\{\mathbf{F}_S\mathbf{F}_S^H\boldsymbol{\Psi}_{SD}\right\}\boldsymbol{\Sigma}_{SD}+\widehat{\mathbf{H}}_{SD}\mathbf{F}_S\mathbf{F}_S^H\widehat{\mathbf{H}}_{SD}^H\\
&:=\mathbf{T}_{SD}
\end{aligned}\tag{77}$$

For the second diagonal matrix, we have

$$\begin{aligned} & E_\Delta \left\{ \mathbf{H}_{RD}\mathbf{F}_R\mathbf{H}_{SR}\mathbf{F}_S\mathbf{F}_S^H\mathbf{H}_{SR}^H\mathbf{F}_R^H\mathbf{H}_{RD}^H \right\} \\ &= E_\Delta \left\{ \mathbf{H}_{RD}\mathbf{F}_R E_\Delta \left\{ \mathbf{H}_{SR}\mathbf{F}_S\mathbf{F}_S^H\mathbf{H}_{SR}^H \right\} \mathbf{F}_R^H\mathbf{H}_{RD}^H \right\} \quad (78) \\ &= E_\Delta \left\{ \mathbf{H}_{RD}\mathbf{F}_R\mathbf{T}_{SR}\mathbf{F}_R^H\mathbf{H}_{RD}^H \right\} \quad (79) \\ &= \mathrm{Tr}\left\{ \mathbf{\Psi}_{RD} \right\} \mathbf{\Sigma}_{RD} + \widehat{\mathbf{H}}_{RD}\mathbf{F}_R\mathbf{T}_{SR}\mathbf{F}_R^H\widehat{\mathbf{H}}_{RD}^H \quad (80) \end{aligned}$$

where the equality in (78) is due to the fact that $\Delta\mathbf{H}_{SR}$ and $\Delta\mathbf{H}_{RD}$ are assumed to be independent. Here, $\mathbf{T}_{SR}$ is defined similarly as that in (77):

$$\begin{aligned} \mathbf{T}_{SR} &:= E_\Delta \left\{ \mathbf{H}_{SR}\mathbf{F}_S\mathbf{F}_S^H\mathbf{H}_{SR}^H \right\} \\ &= \mathrm{Tr}\left\{ \mathbf{F}_S\mathbf{F}_S^H\mathbf{\Psi}_{SR} \right\} \mathbf{\Sigma}_{SR} + \widehat{\mathbf{H}}_{SR}\mathbf{F}_S\mathbf{F}_S^H\widehat{\mathbf{H}}_{SR}^H. \end{aligned} \quad (81)$$

Equation (80) is obtained from (79) also as that in (77). For the off-diagonal matrices, we have

$$\begin{aligned} & E_\Delta \left\{ \mathbf{H}_{SD}\mathbf{F}_S\mathbf{F}_S^H\mathbf{H}_{SR}^H\mathbf{F}_R^H\mathbf{H}_{RD}^H \right\} \\ &= \widehat{\mathbf{H}}_{SD}\mathbf{F}_S\mathbf{F}_S^H\widehat{\mathbf{H}}_{SR}^H\mathbf{F}_R^H\widehat{\mathbf{H}}_{RD}^H \end{aligned} \quad (82)$$

and

$$\begin{aligned} & E_\Delta \left\{ \mathbf{H}_{RD}\mathbf{F}_R\mathbf{H}_{SR}\mathbf{F}_S\mathbf{F}_S^H\mathbf{H}_{SD}^H \right\} \\ &= \widehat{\mathbf{H}}_{RD}\mathbf{F}_R\widehat{\mathbf{H}}_{SR}\mathbf{F}_S\mathbf{F}_S^H\widehat{\mathbf{H}}_{SD}^H. \end{aligned} \quad (83)$$

For the second and third terms in (18), we have

$$\mathrm{Tr}\left\{ E_\Delta \left\{ \sigma_s^2 \mathbf{G}\mathbf{H}\mathbf{F}_S\mathbf{C}^H \right\} \right\} = \mathrm{Tr}\left\{ \sigma_s^2 \mathbf{G}\widehat{\mathbf{H}}\mathbf{F}_S\mathbf{C}^H \right\} \quad (84)$$

and

$$\mathrm{Tr}\left\{ E_\Delta \left\{ \sigma_s^2 \mathbf{C}\mathbf{F}_S^H\mathbf{H}^H\mathbf{G}^H \right\} \right\} = \mathrm{Tr}\left\{ \sigma_s^2 \mathbf{C}\mathbf{F}_S^H\widehat{\mathbf{H}}^H\mathbf{G}^H \right\}. \quad (85)$$

For the last term in (18), we have $\mathbf{R}_w$ as

$$\begin{aligned} \mathbf{R}_w &= E_\Delta \left\{ \left[\begin{array}{c} \mathbf{n}_{D,1} \\ \left(\widehat{\mathbf{H}}_{RD} + \Delta\mathbf{H}_{RD}\right)\mathbf{F}_R\mathbf{n}_R + \mathbf{n}_{D,2} \end{array} \right] \times \right. \\ & \left. \left[\begin{array}{cc} \mathbf{n}_{D,1}^H & \left(\left(\widehat{\mathbf{H}}_{RD} + \Delta\mathbf{H}_{RD}\right)\mathbf{F}_R\mathbf{n}_R + \mathbf{n}_{D,2}\right)^H \end{array} \right] \right\} \\ &= \left[\begin{array}{cc} \sigma_{n,d}^2\mathbf{I}_M & 0 \\ 0 & \mathbf{R}_{w,2,2} \end{array} \right] \end{aligned}$$

with

$$\begin{aligned} \mathbf{R}_{w,2,2} &= \sigma_{n,r}^2 \mathrm{Tr}\left\{ \mathbf{F}_R\mathbf{F}_R^H\mathbf{\Psi}_{RD} \right\} \mathbf{\Sigma}_{RD} + \\ & \sigma_{n,r}^2 \widehat{\mathbf{H}}_{RD}\mathbf{F}_R\mathbf{F}_R^H\widehat{\mathbf{H}}_{RD}^H + \sigma_{n,d}^2\mathbf{I}_M. \end{aligned} \quad (86)$$

Finally, substituting (76), (84)-(86) into (18) and simplifying the result, we can then obtain (19).

Appendix B
Proof of Maximum Power Property

Let's rewrite the MSE matrix as the function of α_S in (33) as

$$\begin{aligned} \mathbf{E}(\alpha_S) &= \sigma_s^2 \mathbf{C} \left(\alpha_S^2 \mathbf{U}_S^H \widetilde{\mathbf{H}}^H \widetilde{\mathbf{H}} \mathbf{U}_S + \mathbf{I}_N \right)^{-1} \mathbf{C}^H \\ &= \sigma_s^2 \mathbf{C} \left(\sigma_s^2 \mathbf{U}_S^H \left(\sigma_{n,d}^{-2} \widehat{\mathbf{H}}_{SD}^H \left(\alpha_S^{-2}\sigma_{n,d}^2 \mathbf{I}_M + \right. \right. \right. \\ & \sigma_s^2 \mathrm{Tr}\left(\mathbf{\Psi}_{SD}\right) \mathbf{\Sigma}_{SD} \right)^{-1} \widehat{\mathbf{H}}_{SD} + \\ & \widehat{\mathbf{H}}_{SR}^H \mathbf{F}_R^H \widehat{\mathbf{H}}_{RD}^H \left(\alpha_S^{-2}\Delta\mathbf{A} + \alpha_S^{-2}\mathbf{A} \right)^{-1} \times \\ & \left. \left. \widehat{\mathbf{H}}_{RD}\mathbf{F}_R\widehat{\mathbf{H}}_{SR} \right) \mathbf{U}_S + \mathbf{I}_N \right)^{-1} \mathbf{C}^H, \end{aligned} \quad (87)$$

where $\Delta\mathbf{A}$ and $\mathbf{A}$ are, respectively defined in (36) and (37). Note that if $\mathbf{E}(\alpha_{S,1}) \preceq \mathbf{E}(\alpha_{S,2})$ for any $\alpha_{S,1} \geq \alpha_{S,2}$, $\mathrm{Tr}\{\mathbf{E}(\alpha_S)\}$ is monotonically decreasing on α_S. So, we have to check if $\mathbf{E}(\alpha_{S,1}) \preceq \mathbf{E}(\alpha_{S,2})$ for any $\alpha_{S,1} \geq \alpha_{S,2}$. To start with, we consider the following lemmas:

Lemma 3 [29]: For any two Hermitian matrices, $\mathbf{P}_1$ and $\mathbf{P}_2$, if $\mathbf{P}_1 \succeq \mathbf{P}_2$, $\mathbf{J}^H\mathbf{P}_1\mathbf{J} \succeq \mathbf{J}^H\mathbf{P}_2\mathbf{J}$ for an arbitrary matrix $\mathbf{J}$. Here, $\mathbf{P}_1 \succeq \mathbf{P}_2$ indicates that $\mathbf{P}_1 - \mathbf{P}_2$ is a positive semidefinite matrix.

Lemma 4 [29]: For any two Hermitian matrices, $\mathbf{P}_1$ and $\mathbf{P}_2$, $\mathbf{P}_1 \succeq \mathbf{P}_2$ if and only if $\mathbf{P}_1^{-1} \preceq \mathbf{P}_2^{-1}$.

Considering $\alpha_S^{-2}\Delta\mathbf{A}$ in (87), we have:

$$\begin{aligned} \alpha_S^{-2}\Delta\mathbf{A} = \; & \mathrm{Tr}\left(\mathbf{F}_R \left(\sigma_s^2 \underbrace{\left(\mathrm{Tr}\left(\mathbf{\Psi}_{SR}\right)\mathbf{\Sigma}_{SR} + \widehat{\mathbf{H}}_{SR}\widehat{\mathbf{H}}_{SR}^H \right)}_{:=\widehat{\mathbf{H}}'_{SR}\widehat{\mathbf{H}}'^H_{SR}} + \right. \right. \\ & \left. \left. \alpha_S^{-2}\sigma_{n,r}^2\mathbf{I}_R \right) \mathbf{F}_R^H \mathbf{\Psi}_{RD} \right) \mathbf{\Sigma}_{RD} \\ & + \sigma_s^2 \mathrm{Tr}\left(\mathbf{\Psi}_{SR}\right) \widehat{\mathbf{H}}_{RD}\mathbf{F}_R\mathbf{\Sigma}_{SR}\mathbf{F}_R^H\widehat{\mathbf{H}}_{RD}^H. \end{aligned} \quad (88)$$

Since $\Delta\mathbf{A}$ and $\mathbf{A}$ are Hermitian matrices and $\Delta\mathbf{A}$ is a function of α_S. By (36), it is easy to find that

$$\alpha_{S,1}^{-2}\Delta\mathbf{A}(\alpha_{S,1}) \preceq \alpha_{S,2}^{-2}\Delta\mathbf{A}(\alpha_{S,2}), \quad \text{if } \alpha_{S,1} \geq \alpha_{S,2}. \quad (89)$$

So, by Lemma 4, we can have

$$\begin{aligned} & \left(\alpha_{S,1}^{-2}\Delta\mathbf{A} + \mathbf{A} \right)^{-1} (\alpha_{S,1}) \succeq \left(\alpha_{S,2}^{-2}\Delta\mathbf{A} + \mathbf{A} \right)^{-1} (\alpha_{S,2}), \\ & \text{if } \alpha_{S,1} \geq \alpha_{S,2}. \end{aligned} \quad (90)$$

By Lemma 3, we have

$$\begin{aligned} & \mathbf{M}(\alpha_{S,1}) \succeq \mathbf{M}(\alpha_{S,2}), \quad \text{if} \quad \alpha_{S,1} \geq \alpha_{S,2}, \quad \text{and} \\ & \mathbf{M} = \left(\widehat{\mathbf{H}}_{SR}^H\mathbf{F}_R^H\widehat{\mathbf{H}}_{RD}^H \left(\alpha_{S,2}^{-2}\Delta\mathbf{A} + \mathbf{A} \right)^{-1} \times \right. \\ & \left. \widehat{\mathbf{H}}_{RD}\mathbf{F}_R\widehat{\mathbf{H}}_{SR} \right). \end{aligned} \quad (91)$$

Similarly,

$$\begin{aligned} & \mathbf{N}(\alpha_{S,1}) \succeq \mathbf{N}(\alpha_{S,2}), \quad \text{if } \alpha_{S,1} \geq \alpha_{S,2}, \quad \text{and} \\ & \mathbf{N} = \left(\sigma_{n,d}^{-2}\widehat{\mathbf{H}}_{SD}^H \times \right. \\ & \left. \left(\alpha_{S,1}^{-2}\sigma_{n,d}^2\mathbf{I}_M + \sigma_s^2\mathrm{Tr}\left(\mathbf{\Psi}_{SD}\right)\mathbf{\Sigma}_{SD} \right)^{-1} \widehat{\mathbf{H}}_{SD} \right). \end{aligned} \quad (92)$$

From (91), (92), Lemma 3 and Lemma 4, we have

$$\begin{aligned} & \left(\sigma_s^2 \mathbf{C} \left(\sigma_s^2 \mathbf{U}_S^H \left(\mathbf{M} + \mathbf{N}\right) \mathbf{U}_S + \mathbf{I}_N \right)^{-1} \mathbf{C}^H \right) (\alpha_{S,1}) \\ & \preceq \left(\sigma_s^2 \mathbf{C} \left(\sigma_s^2 \mathbf{U}_S^H \left(\mathbf{M} + \mathbf{N}\right) \mathbf{U}_S + \mathbf{I}_N \right)^{-1} \mathbf{C}^H \right) (\alpha_{S,2}), \\ & \text{if } \alpha_{S,1} \geq \alpha_{S,2}. \end{aligned} \quad (93)$$

Therefore, we have $\mathbf{E}(\alpha_{S,1}) \preceq \mathbf{E}(\alpha_{S,2})$ for any $\alpha_{S,1} \geq \alpha_{S,2}$ which implies that $\mathrm{Tr}\{\mathbf{E}(\alpha_{S,1})\} \leq \mathrm{Tr}\{\mathbf{E}(\alpha_{S,2})\}$ for any $\alpha_{S,1} \geq \alpha_{S,2}$. Thus, $\mathrm{Tr}\{\mathbf{E}(\alpha_S)\}$ is monotonically decreasing on α_S.

APPENDIX C
DERIVATION OF (54)

We first consider a lower bound of the utility function in (53) given by

$$\begin{aligned}&\det\Bigg(\mathbf{I}_N+\frac{P_{S,T}}{N\sigma_{n,d}^2}\widehat{\mathbf{H}}_{SD}^H\times\\&\left(\sigma_{n,d}^2\mathbf{I}_M+\frac{P_{S,T}}{N}\mathrm{Tr}\left(\mathbf{\Psi}_{SD}\right)\mathbf{\Sigma}_{SD}\right)^{-1}\widehat{\mathbf{H}}_{SD}+\\&\frac{P_{S,T}}{N}\widehat{\mathbf{H}}_{SR}^H\mathbf{F}_R^H\widehat{\mathbf{H}}_{RD}^H\left(\Delta\mathbf{A}+\mathbf{A}\right)^{-1}\widehat{\mathbf{H}}_{RD}\mathbf{F}_R\widehat{\mathbf{H}}_{SR}\Bigg)\geq\\&\det\Bigg(\mathbf{I}_N+\frac{P_{S,T}}{N\sigma_{n,d}^2}\widehat{\mathbf{H}}_{SD}^H\times\\&\left(\sigma_{n,d}^2\mathbf{I}_M+\frac{P_{S,T}}{N}\mathrm{Tr}\left(\mathbf{\Psi}_{SD}\right)\mathbf{\Sigma}_{SD}\right)^{-1}\widehat{\mathbf{H}}_{SD}+\\&\frac{P_{S,T}}{N}\widehat{\mathbf{H}}_{SR}^H\mathbf{F}_R^H\widehat{\mathbf{H}}_{RD}^H\left(\Delta\mathbf{A}'+\mathbf{A}\right)^{-1}\widehat{\mathbf{H}}_{RD}\mathbf{F}_R\widehat{\mathbf{H}}_{SR}\Bigg).\end{aligned}\tag{94}$$

The equality holds when $\mathbf{\Psi}_{RD}=\beta_{RD}\mathbf{I}_R$, $\mathbf{\Sigma}_{RD}=\gamma_{RD}\mathbf{I}_M$, $\mathbf{\Psi}_{SR}=\beta_{SR}\mathbf{I}_N$ and $\mathbf{\Sigma}_{SR}=\gamma_{SR}\mathbf{I}_R$ with some scalars β_{RD}, β_{SR}, γ_{RD} and γ_{SR}. As we will see, optimization with the lower bound is much easier to work with. To prove (94), we first consider following lemmas:

Lemma 5 [27]: For any two positive semidefinite matrices, $\mathbf{P}_1$ and $\mathbf{P}_2$, $\mathrm{Tr}(\mathbf{P}_1)\mathbf{P}_2\leq\mathrm{Tr}(\mathbf{P}_1)\lambda_{\max}(\mathbf{P}_2)$ where $\lambda_{\max}(\mathbf{P}_2)$ indicates the maximum eigenvalue of $\mathbf{P}_2$. The equality holds when $(\mathbf{P}_2)=\lambda_{\max}\mathbf{I}$.

Lemma 6 [27]: For any positive semidefinite matrix, $\mathbf{P}$, $\lambda_{\max}(\mathbf{P})\mathbf{I}\succeq\mathbf{P}$.

Lemma 7: For any two positive matrices $\mathbf{P}_1$, $\mathbf{P}_2$, a positive semidefinite matrix, $\mathbf{K}$, and an arbitrary matrix, $\mathbf{J}$, if $\mathbf{P}_1\succeq\mathbf{P}_2$, then

$$\det\left(\mathbf{K}+\mathbf{J}\mathbf{P}_1^{-1}\mathbf{J}^H\right)\leq\det\left(\mathbf{K}+\mathbf{J}\mathbf{P}_2^{-1}\mathbf{J}^H\right).\tag{95}$$

Lemma 7 can be easily proved by using Lemma 3 and Lemma 4. Since $\mathbf{P}_1\succeq\mathbf{P}_2$ and $\left(\mathbf{K}+\mathbf{J}\mathbf{P}_1^{-1}\mathbf{J}^H\right)\preceq\left(\mathbf{K}+\mathbf{J}\mathbf{P}_2^{-1}\mathbf{J}^H\right)$, (95) results. We now use these lemmas to prove (94). First, consider $\Delta\mathbf{A}$ in (94). By Lemma 5, we have

$$\begin{aligned}\Delta\mathbf{A}=&\,\mathrm{Tr}\left(\mathbf{F}_R\left(\frac{P_{S,T}}{N}\widehat{\mathbf{H}}'_{SR}\widehat{\mathbf{H}}'^H_{SR}+\sigma_{n,r}^2\mathbf{I}_R\right)\times\right.\\&\left.\mathbf{F}_R^H\mathbf{\Psi}_{RD}\right)\mathbf{\Sigma}_{RD}\\&+\sigma_s^2\alpha_S^2\mathrm{Tr}\left(\mathbf{\Psi}_{SR}\right)\widehat{\mathbf{H}}_{RD}\mathbf{F}_R\mathbf{\Sigma}_{SR}\mathbf{F}_R^H\widehat{\mathbf{H}}_{RD}^H\\\preceq&\underbrace{\mathrm{Tr}\left(\mathbf{F}_R\left(\frac{P_{S,T}}{N}\widehat{\mathbf{H}}'_{SR}\widehat{\mathbf{H}}'^H_{SR}+\sigma_{n,r}^2\mathbf{I}_R\right)\mathbf{F}_R^H\right)}_{\leq P_{R,T}}\times\\&\lambda_{\max}(\mathbf{\Psi}_{RD})\lambda_{\max}(\mathbf{\Sigma}_{RD})\mathbf{I}_M+\\&\frac{P_{S,T}}{N}\mathrm{Tr}(\mathbf{\Psi}_{SR})\lambda_{\max}(\mathbf{\Sigma}_{SR})\widehat{\mathbf{H}}_{RD}\mathbf{F}_R\mathbf{F}_R^H\widehat{\mathbf{H}}_{RD}^H\\\preceq&\,P_{R,T}\lambda_{\max}(\mathbf{\Psi}_{RD})\lambda_{\max}(\mathbf{\Sigma}_{RD})\mathbf{I}_M+\\&\frac{P_{S,T}}{N}\mathrm{Tr}(\mathbf{\Psi}_{SR})\lambda_{\max}(\mathbf{\Sigma}_{SR})\widehat{\mathbf{H}}_{RD}\mathbf{F}_R\mathbf{F}_R^H\widehat{\mathbf{H}}_{RD}^H.\end{aligned}\tag{96}$$

The last inequality in (96) is due to the relay power constraint. The equality holds if $\mathbf{\Psi}_{RD}=\beta_{RD}\mathbf{I}_R$, $\mathbf{\Sigma}_{RD}=\gamma_{RD}\mathbf{I}_M$, $\mathbf{\Psi}_{SR}=\beta_{SR}\mathbf{I}_N$ and $\mathbf{\Sigma}_{SR}=\gamma_{SR}\mathbf{I}_R$. This means that the channels corresponding to any two antenna pairs are uncorrelated (the transmit or receive correlation matrix is a scaled identity matrix). Then, from (96) and (55) we have

$$\left(\Delta\mathbf{A}+\mathbf{A}\right)\preceq\left(\Delta\mathbf{A}'+\mathbf{A}\right).\tag{97}$$

Now, let $\mathbf{K}=\mathbf{I}_N+\frac{P_{S,T}}{N\sigma_{n,d}^2}\widehat{\mathbf{H}}_{SD}^H\left(\sigma_{n,d}^2\mathbf{I}_M+\frac{P_{S,T}}{N}\mathrm{Tr}(\mathbf{\Psi}_{SD})\mathbf{\Sigma}_{SD}\right)^{-1}\widehat{\mathbf{H}}_{SD}$, $\mathbf{J}=\sqrt{\frac{P_{S,T}}{N}}\widehat{\mathbf{H}}_{SR}^H\mathbf{F}_R^H\widehat{\mathbf{H}}_{RD}^H$, $\mathbf{P}_1=\Delta\mathbf{A}'+\mathbf{A}$, and $\mathbf{P}_2=\Delta\mathbf{A}+\mathbf{A}$. Using Lemma 7, we then obtain the lower bound in (94) and thus have the optimization (54).

APPENDIX D
DERIVATION OF THE OPTIMAL SOLUTION IN (68)

To solve the optimization problem in (67), we first consider the corresponding Lagrangian function:

$$\begin{aligned}L\quad&=\sum_{i=1}^{\kappa}\ln\left(1+\frac{p_{r,i}{\sigma''}^2_{sr,i}\sigma_{rd,i}^2}{\alpha+\beta p_{r,i}\sigma_{rd,i}^2}\right)+\\&\lambda\left[\sum_{i=1}^{\kappa}p_{r,i}\left(\sigma_{n,r}^2+\mathbf{D}_{sr}(i,i)\right)-P_{R,T}\right]\\&-\sum_{i=1}^{\kappa}\upsilon_{r,i}p_{r,i}\end{aligned}\tag{98}$$

where $\lambda\geq0,\upsilon_{r,i}\geq0$ with $i=1,\cdots,\kappa$. By the KKT conditions, for all i, we have

$$\frac{\partial L}{\partial p_k}=-\frac{\frac{\alpha{\sigma''}^2_{sr,i}\sigma_{rd,i}^2}{\left(\alpha+\beta p_{r,i}\sigma_{rd,i}^2\right)^2}}{1+\frac{p_{r,i}{\sigma''}^2_{sr,i}\sigma_{rd,i}^2}{\alpha+\beta p_{r,i}\sigma_{rd,i}^2}}+\lambda\left[\sigma_{n,r}^2\mathbf{D}_{sr}(i,i)\right]-\upsilon_{r,i}=0\tag{99}$$

$$\upsilon_{r,i}p_{r,i}=0\tag{100}$$

$$\lambda\left[\textstyle\sum_{i=1}^{\kappa}p_{r,i}\left(\sigma_{n,r}^2+\mathbf{D}_{sr}(i,i)\right)-P_{R,T}\right]=0\tag{101}$$

$$\lambda,\upsilon_{r,i},p_{r,i}\geq0\tag{102}$$

Substituting (99) into (100) and considering that $p_{r,i}\geq0$, we have $\upsilon_{r,i}=0$. Thus, we have equations (103)-(106), as shown at the top of the next page.

After simplification, the optimum $p_{r,i}$ can be expressed as equation (107) shown on the next page, where $\mu:=\frac{1}{\lambda}$ is chosen to satisfy the power constraint in (67).

ACKNOWLEDGMENT

The authors would like to thank the Editor and anonymous reviewers for their helpful comments. This work was supported by the National Science Council under grant NSC-98-2221-E-009-046-MY3 and by MediaTek Research Center at National Chiao Tung University, Taiwan.

REFERENCES

[1] B. Wang, J. Zhang, and A. Host-Madsen, "On the capacity of MIMO relay channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 1, pp. 29–43, Jan. 2005.

[2] 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.

[3] 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.

$$\frac{\frac{\alpha\sigma''^2_{sr,i}\sigma^2_{rd,i}}{\left(\alpha+\beta p_{r,i}\sigma^2_{rd,i}\right)^2}}{1+\frac{p_{r,i}\sigma''^2_{sr,i}\sigma^2_{rd,i}}{\alpha+\beta p_{r,i}\sigma^2_{rd,i}}} = \frac{1}{p^2_{r,i}\underbrace{\sigma^2_{rd,i}\frac{\beta}{\alpha}\left(\frac{\beta}{\sigma''^2_{sr,i}}+1\right)}_{:=A_i}+2p_{r,i}\underbrace{\left(\frac{\beta}{\sigma''^2_{sr,i}}+\frac{1}{2}\right)}_{:=B_i}+\underbrace{\frac{\alpha}{\sigma''^2_{sr,i}\sigma^2_{rd,i}}}_{:=C_i}}$$
$$= \lambda\left[\sigma^2_{n,r}+\mathbf{D}_{sr}(i,i)\right] \tag{103}$$

$$p_{r,i} = \sqrt{\frac{1}{\lambda\left(\sigma^2_{n,r}+\mathbf{D}_{sr}(i,i)\right)A_i}+\frac{B_i^2}{A_i^2}-\frac{C_i}{A_i}}-\frac{B_i}{A_i} \tag{104}$$

$$\frac{B_i^2-A_iC_i}{A_i^2} = \frac{1/4}{\sigma^4_{rd,i}\frac{\beta^2}{\alpha^2}\left(\frac{\beta}{\sigma''^2_{sr,i}}+1\right)^2} \tag{105}$$

$$\frac{B_i}{A_i} = \frac{\alpha+\frac{\alpha\sigma''^2_{sr,i}}{2\beta}}{\sigma^2_{rd,i}\left(\beta+\sigma''^2_{sr,i}\right)} \tag{106}$$

$$p_{r,i} = \left[\sqrt{\frac{\mu}{\left(\sigma^2_{n,r}+\mathbf{D}_{sr}(i,i)\right)\sigma^2_{rd,i}\cdot\frac{\beta}{\alpha}\left(\frac{\beta}{\sigma''^2_{sr,i}}+1\right)}+\frac{\frac{1}{4}}{\sigma^4_{rd,i}\frac{\beta^2}{\alpha^2}\left(\frac{\beta}{\sigma''^2_{sr,i}}+1\right)^2}}-\frac{\alpha+\frac{\alpha\sigma''^2_{sr,i}}{2\beta}}{\sigma^2_{rd,i}\left(\beta+\sigma''^2_{sr,i}\right)}\right]^+ \tag{107}$$

[4] W. Guan and H. Luo, "Joint MMSE transceiver design in nonregenerative MIMO relay systems," *IEEE Trans. Commun. Lett.*, vol. 12, no. 7, pp. 517–519, July 2008.

[5] A. S. Behbahani, R. Merched, and A. M. Eltawil, "Optimization of a MIMO relay network," *IEEE Trans. Signal Process.*, vol. 56, no. 10, pp. 5062–5073, Oct. 2008.

[6] F. S. Tseng, W. R. Wu, and J. Y. Wu, "Joint source/relay precoder design in nonregenerative cooperative systems using an MMSE criterion," *IEEE Trans. Wireless Commun.*, vol. 8, no. 10, pp. 4928–4933, Oct. 2009.

[7] F. S. Tseng and W. R. Wu, "Linear MMSE transceiver design in amplify-and-forward MIMO relay systems," *IEEE Trans. Veh. Technol.*, vol. 59, no. 2, pp. 754–765, Feb. 2010.

[8] F. S. Tseng and W. R. Wu, "Joint source/relay precoders design in amplify-and-forward relay systems: a geometric mean decomposition approach," in *Proc. 2009 IEEE ICASSP*.

[9] F. S. Tseng and W. R. Wu, "MMSE-SIC transceiver design in amplify-and-forward MIMO relay systems," in *Proc. 2010 IEEE VTC – Spring*.

[10] F. S. Tseng, M. Y. Chang, and W. R. Wu, "Joint MMSE transceiver design in amplify-and-forward MIMO relay systems with Tomlinson-Harashima source precoding," in *Proc. 2010 IEEE PIMRC*.

[11] A. Goldsmith, S. A. Jafar, N. Jindal, and S. Vishwanath, "Capacity limits of MIMO channels," *IEEE J. Sel. Areas Commun.*, vol. 21, pp. 684–702, June 2003.

[12] M. Joham, W. Utschick, and J. A. Nossek, "Linear transmit processing in MIMO communications systems," *IEEE Trans. Signal Process.*, vol. 53, no. 8, pp. 2700–2712, Aug. 2005.

[13] H. Sampath, P. Stoica, and A. Paulraj, "Generalized linear precoder and decoder design for MIMO channels using the weighted MMSE criterion," *IEEE Trans. Commun.*, vol. 49, no. 12, pp. 2198–2206, Dec. 2001.

[14] D. P. Palomar, J. M. Cioffi, and M. A. Laguna, "Joint Tx-Rx beamforming design for multicarrier MIMO channels: a unified framework for convex optimization," *IEEE Trans. Signal Process.*, vol. 51, no. 9, pp. 2381–2401, Sep. 2003.

[15] D. P. Palomar and S. Barbarossa, "Designing MIMO communication systems: constellation choice and linear transceiver design," *IEEE Trans. Signal Process.*, vol. 53, no. 10, pp. 3804–3818, Oct. 2005.

[16] Y. Jiang, J. Li, and W. W. Hager, "Joint transceiver design for MIMO communications using geometric mean decomposition," *IEEE Trans. Signal Process.*, vol. 53, no. 10, pp. 3791–3803, Oct. 2005.

[17] Y. Jiang, J. Li, and W. W. Hager, "Uniform channel decomposition for MIMO communications," *IEEE Trans. Signal Process.*, vol. 53, no. 11, pp. 4283–4294, Nov. 2005.

[18] R. F. H. Fischer, C. Windpassinger, A. Lampe, and J. H. Huber, "Space time transmission using Tomlinson-Harashima precoding," in *Proc. 2002 ITC Conf. Source Channel Coding*.

[19] A. A. D'Amico and M. Morelli, "Joint Tx-Rx MMSE design for MIMO multicarrier systems with Tomlinson-Harashima precoding," *IEEE Trans. Wireless Commun.*, vol. 7, no. 8, pp. 3118–3127, Aug. 2008.

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

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

[22] J. Duplicy and L. Vandendorpe, "Robust MMSE Precoding for the MIMO complex Gaussian broadcast channel," in *Proc. 2007 IEEE ICASSP*.

[23] Y. Guo and B. C. Levy, "Worst-case MSE precoder design for imperfectly known MIMO communications channels," *IEEE Trans. Signal Process.*, vol. 53, no. 8, pp. 2918–2903, Aug. 2005.

[24] E. G. Larsson and P. Stoica, *Space-Time Block Coding for Wireless Communications*. Cambridge University Press, 2003

[25] 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.

[26] A. T. James, "Distributions of matrix variants and latent roots derived from normal samples," *Ann. Math. Statisitcs*, vol. 35, pp. 475–501, 1964.

[27] D. S. Bernstein, *Matrix Mathematics*. Princeton University Press, 2005.

[28] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.

[29] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge University Press, 1985.

[30] A. K. Gupta and D. K. Nagar, *Matrix Variate Distributions*. Chapman & Hall/CRC 2000.

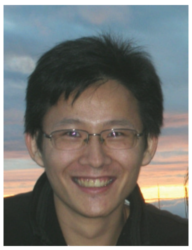
[31] H. E. Cho, H. J. Park, and Y. H. Lee, "Effect of imperfect transmit correlation on statistical beamforming in multi-user cellular systems,"

IEEE Trans. Wireless Commun., vol. 9, no. 11, pp. 3335–3340, Nov. 2010.

[32] Y. Rong, "Robust design for linear non-regenerative MIMO relays," in *Proc. 2010 Australian Commun. Theory Workshop*, pp. 87–92.

[33] 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.

[34] A. A. D'Amico, "Tomlinson-Harashima precoding in MIMO systems: a unified approach to transceiver optimization based on multiplicative Schur-convexity," *IEEE Trans. Signal Process.*, vol. 56, no. 8, pp. 3662–3677, Aug. 2008.



Fan-Shuo Tseng (S'08) received the B.S. degree from Fu Jen Catholic University, Hsinchung, Taiwan, in 2002, the M.S. degree from National Central University, Jhongli, Taiwan, in 2004, and the Ph.D. degree from National Chiao Tung University, Hsinchu, Taiwan, in December 2009. From January 2010 to July 2010, he was a Postdoctoral Research Fellow with the Institute of Communications Engineering, National Chiao Tung University. Since August 2011, he has been an Assistant Professor at National Sun Yat-sen University, Kaohsiung, Taiwan. His research interests include transceiver design in cooperative communication and statistical signal processing in digital communication.

Min-Yao Chang received the B.S. degree in the Department of Communications Engineering in 2007 from National Central University, Jhongli, Taiwan, and the M.S. degree in the Department of Communications Engineering in 2009 from National Chiao Tung University, Hsinchu, Taiwan. He has served as an engineer for MediaTek Inc., Hsinchu Science Park, Taiwan, since 2011. His research interests include digital signal processing in cooperative system and filter design.

Wen-Rong Wu (M'89) received a B.S. degree in mechanical engineering from Tatung Institute of Technology, Taipei, Taiwan, in 1980, and a M.S. degree in mechanical engineering, a M.S. degree in electrical engineering, and a Ph.D. degree in electrical engineering from the State University of New York, Buffalo, in 1985, 1986, and 1989, respectively.

Since August 1989, he has been a faculty member with the Department of Electrical Engineering, National Chiao Tung University, Hsinchu, Taiwan. His research interests include statistical signal processing and digital communications.