Title

Robust Optimization for Amplify-and-Forward MIMO Relaying from a Worst-Case Perspective

Journal
IEEE Transactions on Signal Processing, 61(21)

ISSN
1053-587X

Authors
Shen, Hong
Wang, Jiaheng
Levy, Bernard C
et al.

Publication Date
2013-11-01

DOI
10.1109/tsp.2013.2278819

Peer reviewed

Robust Optimization for Amplify-and-Forward MIMO Relaying from a Worst-Case Perspective

Hong Shen, *Student Member, IEEE,* Jiaheng Wang, *Member, IEEE,* Bernard C. Levy, *Fellow, IEEE,* and Chunming Zhao, *Member, IEEE*

***Abstract*— In this paper, we consider robust optimization of amplify-and-forward (AF) multiple-input multiple-output (MIMO) relay precoders in presence of deterministic imperfect channel state information (CSI), when the CSI uncertainty lies in a norm bounded region. Two widely used performance metrics, mutual information (MI) and mean square error (MSE), are adopted as design objectives. According to the philosophy of worst-case robustness, the robust optimization problems with respect to maximizing the worst-case MI and minimizing the worst-case MSE are formulated as maximin and minimax problems, respectively. Due to the fact that these two problems do not have a concave-convex or convex-concave structure, we cannot rely on the conventional saddle point theory to find the robust solutions. Nevertheless, by exploiting majorization theory, we show that the formulated maximin and minimax problems both admit saddle points. We further analytically characterize the saddle points, and provide closed-form solutions to robust relay precoder designs. Interestingly, we find that, under both MI and MSE metrics, the robust relay optimization leads to a channel-diagonalizing structure, meaning that eigenmode transmission is optimal from the worst-case robustness perspective. The proposed robust designs can improve the spectral efficiency and reliability of AF MIMO relaying against CSI uncertainties at the similar cost of computational complexity as the existing non-robust schemes.**

***Index Terms*— Multiple-input multiple-output (MIMO) cooperative transmission, mutual information (MI), mean square error (MSE), worst-case robust design, relay precoding, imperfect channel state information (CSI).**

I. Introduction

Cooperative relay networking has emerged as a promising technique for enhancing communication reliability and expanding coverage of next generation wireless systems [1]. Meanwhile, multiple-input multiple-output (MIMO) techniques which are widely used to improve the capacity and/or reliability of wireless channels, have been introduced into relay systems, namely MIMO relaying, in order to gain further performance enhancement [2] [3]. Among various existing relay protocols, amplify-and-forward (AF) MIMO relaying is a very promising strategy well known for its simplicity, as the relay only performs linear transformation (or precoding) on the received signal and re-transmits it to the destination.

H. Shen, J. Wang, and C. Zhao are with National Mobile Communications Research Laboratory, Southeast University, Nanjing 210096, China. H. Shen is also with the Department of Electrical and Computer Engineering, University of California, Davis, CA 95616, USA (e-mail: shenhongseu@gmail.com, jhwangee@gmail.com, cmzhao@seu.edu.cn).

B. Levy is with the Department of Electrical and Computer Engineering, University of California, Davis, CA 95616, USA (e-mail: bclevy@ucdavis.edu).

Transceiver optimization for AF MIMO relays is an important research topic that has received significant attention [4]–[8]. To be more specific, the authors in [4] and [5] optimized the relay precoder to maximize the mutual information (MI) between the source and destination. In [6] and [7], an alternative relay design based on the mean square error (MSE), was considered for transceiver optimization. The authors in [8] proposed a unified optimization framework for AF MIMO relaying by means of majorization theory. Interestingly, a common conclusion of the aforementioned works is that the optimal transceiver design has a simple channel-diagonalizing structure, implying that eigenmode transmission is still optimal as in the case of point-to-point MIMO communications [9]. This attractive feature greatly simplifies the original matrix-variable optimization problem by converting it to a simpler scalar-based power allocation problem.

As the performance of a MIMO relay system is sensitive to the accuracy of available channel state information (CSI), ignoring the deviation between the true CSI and the estimated CSI as in [4]–[8] may lead to severe performance degradation in practical systems, where CSI errors generally exist as a result of inaccurate channel estimation, quantization and delayed feedback. Accordingly, in order to mitigate the degradation caused by CSI imperfection, it is necessary to perform a robust design taking CSI mismatch into consideration. Thus far, two types of robustness have been considered in existing works, namely statistical and worst-case robustness. In particular, statistically robust optimization [10]–[16] assumes that statistical information about the CSI such as its mean and/or covariance can be acquired. When CSI statistics are available, the statistically robust design usually seeks to enhance the average or outage performance. In contrast with the philosophy of statistical robustness, the worst-case robust designs [17]–[25] assume no other CSI knowledge except that the actual channel belongs to a bounded uncertainty set centered by a nominal channel. In this context, optimizing the worst-case performance becomes a meaningful approach to achieve robustness.

The statistically robust transceiver design for AF MIMO relaying has been examined in prior works such as [13]–[15], where the channel-diagonalizing structure is shown to be optimal, hence being consistent with the results under the perfect CSI assumption [4]–[8]. In light of these existing results, it is natural to ask whether similar conclusions also apply to AF MIMO relay systems subject to deterministic CSI uncertainties. Although there have been a few works [22]–[25] on the worst-case robust optimization of AF MIMO relay

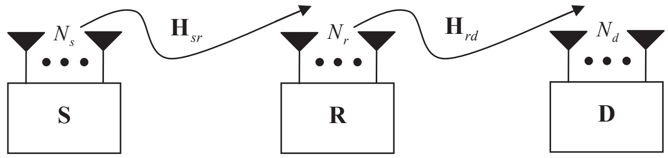

Fig. 1. A two-hop AF MIMO relay system.

precoding, to the best of our knowledge, whether the optimal robust design diagonalizes the MIMO channel is still unknown yet. Our goal is to address this open problem.

In this paper, we investigate robust optimization for a two-hop AF MIMO relay system from the worst-case CSI uncertainty perspective and give a positive answer to the above proposed question. Specifically, we address two robust design problems with respect to the MI and MSE objectives, aiming at enhancing the transmission efficiency and reliability of MIMO relay systems in presence of bounded CSI errors. The two robust optimization problems are formulated as maximin and minimax problems, respectively. The common way to address these problems is to use saddle point theory, e.g. Von Neumann's theorem. However, the considered maximin and minimax problems do not have a classical concave-convex (resp., convex-concave) structure and hence there is even no guarantee of the existence of a saddle point. Nevertheless, by applying majorization theory, we show that there indeed exist saddle point solutions for both problems, and moreover, they can be obtained in explicit analytical forms. An interesting implication arising from the solution is that eigenmode transmission is still the optimal strategy even for the case of bounded CSI uncertainties, which is consistent with the conclusions for the perfect and stochastic CSI cases. As verified by numerical results, our proposed robust designs achieve a noticeable performance gain over the non-robust schemes of [4] [6].

The manuscript is organized as follows. A system model for AF MIMO relaying is introduced in Section II. In Section III, we obtain the optimal AF relay precoder for maximizing the worst-case MI of MIMO relaying. We then derive the optimal transceiver for minimizing the worst-case MSE of AF MIMO relay systems in Section IV. Simulation results are described in Section V and conclusions are presented in Section VI.

Notation: We use uppercase and lowercase boldface letters to denote matrices and vectors, respectively. Notations $\mathbf{A}^{-1}$, $\mathbf{A}^T$ and $\mathbf{A}^H$ represent the inverse, transpose and conjugate transpose of matrix $\mathbf{A}$, respectively. The determinant, trace and rank of $\mathbf{A}$ are denoted by $|\mathbf{A}|$, $\text{tr}(\mathbf{A})$ and $\text{rank}(\mathbf{A})$, respectively. Notations $\mathbf{x} \prec \mathbf{y}$ and $\mathbf{x} \prec_w \mathbf{y}$ indicate that $\mathbf{x}$ is majorized by $\mathbf{y}$ and $\mathbf{x}$ is weakly majorized by $\mathbf{y}$, respectively (see Appendix I for a brief introduction of majorization theory). $(\mathbf{A})_{i,j}$ represents the (ith, jth) element of $\mathbf{A}$. The spectral and Frobenius norms of $\mathbf{A}$ are denoted by $\|\mathbf{A}\|_2$ and $\|\mathbf{A}\|_F$, respectively. $\mathbf{0}_{M\times N}$ represents a zero matrix of size $M \times N$. $\text{diag}\{\mathbf{a}\}$ denotes a diagonal matrix whose diagonal elements are the entries of $\mathbf{a}$. $(x)_+$ denotes $\max\{x, 0\}$. Finally, $\mathbb{C}^{n\times m}$ and $\mathbb{R}^{n\times m}$ denote the ensemble of all $n \times m$ complex and real matrices, respectively, and $\mathbb{E}\{\cdot\}$ represents the expectation operation.

II. System Model and Problem Formulation

A. Signal Model Description

We consider a dual-hop AF MIMO relay system as shown in Fig. 1, where the source, relay and destination are equipped with N_s, N_r and N_d antennas, respectively. The direct link between the source and destination is assumed to be sufficiently weak so that it can be ignored. At the first hop, the source transmits a symbol vector $\mathbf{s} \in \mathbb{C}^{N_s}$ to the relay node, where $\mathbb{E}\{\mathbf{s}\mathbf{s}^H\} = \frac{P_s}{N_s}\mathbf{I}$ with P_s being the source transmit power. The received signal $\mathbf{y}_r \in \mathbb{C}^{N_r}$ at the relay takes the form

$$\mathbf{y}_r = \mathbf{H}_{sr}\mathbf{s} + \mathbf{n}_r, \tag{1}$$

where $\mathbf{H}_{sr} \in \mathbb{C}^{N_r\times N_s}$ represents the source-relay channel and $\mathbf{n}_r \in \mathbb{C}^{N_r}$ is the additive white Gaussian noise (AWGN) vector at the relay with zero mean and covariance matrix $\mathbf{R}_{n_r} = \sigma_r^2\mathbf{I}$. At the second hop, the relay multiplies the received signal $\mathbf{y}_r$ by a precoding matrix $\mathbf{F}_r \in \mathbb{C}^{N_r\times N_r}$, which results in $\mathbf{x}_r = \mathbf{F}_r\mathbf{y}_r$. Generally, the relay imposes a power constraint on the precoder $\mathbf{F}_r$ as

$$\text{tr}\left(\mathbf{F}_r\left(\frac{P_s}{N_s}\mathbf{H}_{sr}\mathbf{H}_{sr}^H + \sigma_r^2\mathbf{I}\right)\mathbf{F}_r^H\right) \leq P_r, \tag{2}$$

where P_r is the maximum transmit power of the relay. After the relay forwards $\mathbf{x}_r$ to the destination, the received signal $\mathbf{y}_d \in \mathbb{C}^{N_d}$ at the destination is given by

$$\mathbf{y}_d = \mathbf{H}_{rd}\mathbf{F}_r\mathbf{H}_{sr}\mathbf{s} + \mathbf{H}_{rd}\mathbf{F}_r\mathbf{n}_r + \mathbf{n}_d, \tag{3}$$

where $\mathbf{H}_{rd} \in \mathbb{C}^{N_d\times N_r}$ denotes the relay-destination channel and $\mathbf{n}_d \in \mathbb{C}^{N_d}$ is the AWGN vector at the destination with zero mean and covariance matrix $\mathbf{R}_{n_d} = \sigma_d^2\mathbf{I}$.

In this paper, we consider two widely used performance metrics: the mutual information (MI) and the mean square error (MSE), which are often used to characterize the transmission efficiency and reliability of a communication system, respectively. Assuming that the source uses a Gaussian code, the MI of the above MIMO relay system is given by [4]

$$\text{MI} = \frac{1}{2}\log_2\left|\mathbf{I} + \frac{P_s}{N_s}\tilde{\mathbf{H}}^H\mathbf{R}^{-1}\tilde{\mathbf{H}}\right|, \tag{4}$$

where $\tilde{\mathbf{H}} = \mathbf{H}_{rd}\mathbf{F}_r\mathbf{H}_{sr}$, $\mathbf{R} = \sigma_r^2\mathbf{H}_{rd}\mathbf{F}_r\mathbf{F}_r^H\mathbf{H}_{rd}^H + \sigma_d^2\mathbf{I}$ and the coefficient $1/2$ accounts for the half-duplex loss. In practice, to reduce the implementation complexity at the receiver, a linear decoder matrix $\mathbf{G} \in \mathbb{C}^{N_s\times N_d}$ is usually applied on the received signal $\mathbf{y}_d$ to obtain the estimate $\hat{\mathbf{s}} = \mathbf{G}\mathbf{y}_d$ of transmit signal $\mathbf{s}$, and the corresponding MSE takes the form

$$\begin{aligned}\text{MSE} &= \mathbb{E}\left\{(\hat{\mathbf{s}} - \mathbf{s})^H(\hat{\mathbf{s}} - \mathbf{s})\right\} \\ &= \mathbb{E}\left\{(\mathbf{G}\mathbf{y}_d - \mathbf{s})^H(\mathbf{G}\mathbf{y}_d - \mathbf{s})\right\}.\end{aligned} \tag{5}$$

The optimal $\mathbf{G}$ that minimizes the MSE is known as the Wiener filter and given by [26]

$$\mathbf{G} = \frac{P_s}{N_s}\left(\frac{P_s}{N_s}\tilde{\mathbf{H}}\tilde{\mathbf{H}}^H + \mathbf{R}\right)^{-1}\tilde{\mathbf{H}}^H. \tag{6}$$

$$\begin{aligned}\max_{\mathbf{F}_r}\ \min_{\mathbf{\Delta}_{rd}}\quad & \frac{1}{2}\log_2\left|\mathbf{I}+\frac{P_s}{N_s}\mathbf{H}_{sr}^H\mathbf{F}_r^H(\hat{\mathbf{H}}_{rd}+\mathbf{\Delta}_{rd})^H\left(\sigma_r^2(\hat{\mathbf{H}}_{rd}+\mathbf{\Delta}_{rd})\mathbf{F}_r\mathbf{F}_r^H(\hat{\mathbf{H}}_{rd}+\mathbf{\Delta}_{rd})^H+\sigma_d^2\mathbf{I}\right)^{-1}\right.\\ & \left.\times(\hat{\mathbf{H}}_{rd}+\mathbf{\Delta}_{rd})\mathbf{F}_r\mathbf{H}_{sr}\right|\\ \text{subject to}\quad & \mathrm{tr}\left(\mathbf{F}_r\left(\frac{P_s}{N_s}\mathbf{H}_{sr}\mathbf{H}_{sr}^H+\sigma_{n_r}^2\mathbf{I}\right)\mathbf{F}_r^H\right)\le P_r,\\ & \forall\mathbf{\Delta}_{rd}:\ \|\mathbf{\Delta}_{rd}\|_2\le\epsilon_{rd}\end{aligned}\tag{10}$$

$$\begin{aligned}\min_{\mathbf{F}_r}\ \max_{\mathbf{\Delta}_{rd}}\quad & \frac{P_s}{N_s}\mathrm{tr}\Bigg(\Bigg(\mathbf{I}+\frac{P_s}{N_s}\mathbf{H}_{sr}^H\mathbf{F}_r^H(\hat{\mathbf{H}}_{rd}+\mathbf{\Delta}_{rd})^H\left(\sigma_r^2(\hat{\mathbf{H}}_{rd}+\mathbf{\Delta}_{rd})\mathbf{F}_r\mathbf{F}_r^H(\hat{\mathbf{H}}_{rd}+\mathbf{\Delta}_{rd})^H+\sigma_d^2\mathbf{I}\right)^{-1}\\ & \times(\hat{\mathbf{H}}_{rd}+\mathbf{\Delta}_{rd})\mathbf{F}_r\mathbf{H}_{sr}\Bigg)^{-1}\Bigg)\\ \text{subject to}\quad & \mathrm{tr}\left(\mathbf{F}_r\left(\frac{P_s}{N_s}\mathbf{H}_{sr}\mathbf{H}_{sr}^H+\sigma_{n_r}^2\mathbf{I}\right)\mathbf{F}_r^H\right)\le P_r,\\ & \forall\mathbf{\Delta}_{rd}:\ \|\mathbf{\Delta}_{rd}\|_2\le\epsilon_{rd}\end{aligned}\tag{11}$$

By substituting the above expression inside (5), we have

$$\mathrm{MSE}=\frac{P_s}{N_s}\mathrm{tr}\left(\left(\mathbf{I}+\frac{P_s}{N_s}\tilde{\mathbf{H}}^H\mathbf{R}^{-1}\tilde{\mathbf{H}}\right)^{-1}\right),\tag{7}$$

While the relay precoder design under the maximum MI and minimum MSE criteria has been well investigated in [4] [6] with perfect CSI, in practice CSI is seldom perfect, which thus calls for a robust design taking into account the imperfection of CSI.

B. Imperfect CSI Model

In general, it is reasonable to assume that perfect CSI at the receiver (CSIR) is available since CSIR is relatively easy to acquire with the aid of training sequences, whereas it is much harder to obtain accurate CSI at the transmitter (CSIT) due to practical factors such as quantization, delays or feedback errors. Thereby, imperfect CSIT has to be considered in the system design. For the AF MIMO relay system, we assume that the relay knows the perfect source-relay channel $\mathbf{H}_{sr}$ and the destination knows perfectly the equivalent channels $\tilde{\mathbf{H}}$ and $\mathbf{H}_{rd}\mathbf{F}_r$, while the relay can only acquire imperfect information about the relay-destination channel $\mathbf{H}_{rd}$. Note that these assumptions have also been widely adopted in previous studies such as [15] [22] [25] [27].

To characterize the mismatched relay-destination CSI, we adopt a common deterministic imperfect CSI model as in [19]–[21]. Specifically, the actual $\mathbf{H}_{rd}$ takes the form

$$\mathbf{H}_{rd}=\hat{\mathbf{H}}_{rd}+\mathbf{\Delta}_{rd}\tag{8}$$

and

$$\mathbf{\Delta}_{rd}\in\mathcal{U}\triangleq\{\mathbf{\Delta}_{rd}:\|\mathbf{\Delta}_{rd}\|_2\le\epsilon_{rd}\},\tag{9}$$

where $\hat{\mathbf{H}}_{rd}$ represents the mismatched channel obtained by means of channel estimation or quantization, $\mathbf{\Delta}_{rd}$ denotes the channel uncertainty lying in a spectral norm bounded uncertainty region $\mathcal{U}$ with a given radius ϵ_{rd}. In practice, there are several ways to determine ϵ_{rd}. One approach consists of using a channel emulator to simulate wireless channels based on the specified parameters, and record the generated channel matrices $\mathbf{H}_{rd}$. The mismatched channel matrices $\hat{\mathbf{H}}_{rd}$ are found by estimating or quantizing the channels generated by the channel emulator. The channel error matrix $\mathbf{\Delta}_{rd}$ is obtained as the difference between the true and mismatched channels. Then ϵ_{rd} corresponds to the upper bound of the spectral norms of all channel error matrices. One can also determine ϵ_{rd} in a stochastic way when the distribution of the channel error is known, and satisfies for example a Gaussian distribution. In this case, one can apply numerical methods to evaluate ϵ_{rd} to ensure that the norms of possible errors are upper bounded with a certain probability. It is necessary to point out that we can carry out the aforementioned offline computation for different channel parameters and make a look-up table storing all calculated ϵ_{rd}. In this way, we update the value of ϵ_{rd} when the channel conditions change.

The norm bounded CSI error model has been extensively used in robust precoder designs for point-to-point MIMO channels [17] [18], downlink multiuser channels [28]–[30], cognitive radio systems [31] [32] and so on. As pointed out in [18] [28], the shape of the uncertainty region corresponding to channel quantization is a bounded polytope $\mathcal{P}$ with the quantized channel at its center. Hence, from a geometric perspective, one can always find a value of ϵ_{rd} such that the uncertainty region $\|\mathbf{\Delta}_{rd}\|_2\le\epsilon_{rd}$ covers the entire polytope $\mathcal{P}$. In contrast with the quantization error which naturally belongs to a bounded region, the CSI imperfection caused by feedback delay or error is usually assumed to be Gaussian distributed [33] [34] (similar to the estimation error) and hence lies inside the norm bounded uncertainty region with a certain probability P_{in}, i.e., $\Pr(\|\mathbf{\Delta}_{rd}\|_2\le\epsilon_{rd})=P_{\text{in}}<1$.

In other words, given any probability P_{in}, it is always possible to choose a proper radius ϵ_{rd} and let the norm bounded uncertainty region cover the channel errors with probability P_{in}. In particular, we would like to note that using the spectral norm in defining the uncertainty region has a number of advantages [20]. Specifically, it is a unitary-invariant matrix norm which indicates that the uncertainties in the spectral norm bounded region are statistically equal in all directions. Furthermore, the spectral norm serves as a lower bound of any unitary-invariant matrix norm, meaning that given the same error radius, it models the largest uncertainty set among all unitary-invariant matrix norms. Besides, it is also an indicator of the strongest eigenmode of the uncertainty. It is important to note that, since the inequality $\|\mathbf{\Delta}_{rd}\|_2 \leq \|\mathbf{\Delta}_{rd}\|_F \leq \sqrt{\min\{N_d, N_r\}}\|\mathbf{\Delta}_{rd}\|_2$ always holds, studying spectral norm bounded uncertainties provides some insight on the case where Frobenius norm bounded errors are considered.

C. Problem Formulation

According to the philosophy of worst-case robustness [17]–[25], [28]–[32], a robust design is obtained by ensuring the best possible level of MI (or MSE) performance for all channel realizations within the uncertainty region given by (9), which is equivalent to optimizing the worst-case MI (or MSE) performance in our system. To be more exact, the robust design problem with respect to optimizing the worst-case MI can be expressed as (10). Similarly, the worst-case MSE optimization can be formulated as (11).

In the context of AF MIMO relaying, the MI maximization and MSE minimization problems with perfect CSI have been solved in [4] and [6], respectively, where the optimal relay precoder was obtained in closed form. Different from [4] and [6], we here consider the relay precoder optimization problems from the worst-case robust perspective, which, to the best of our knowledge, have not been studied by others before. It can be observed that problems (10) and (11) are quite difficult to solve due to the intricate form of maximin (minimax) optimization and semi-infinite constraint with respect to $\mathbf{\Delta}_{rd}$. Moreover, the above two problems do not have a classical concave-convex (resp., convex-concave) structure and hence there may not exist a saddle point. Despite these difficulties, in the rest of this paper, we will show that the optimal solutions to both problems can be achieved in analytical forms. Due to their complicated structure, problems (10) and (11) are examined separately in Sections III and IV.

III. Worst-Case MI Maximization for AF MIMO Relaying

In this section, we focus on the worst-case MI maximization, i.e., the maximin problem (10). As pointed out above, (10) is difficult to solve, because there is even no guarantee of the existence of a saddle point as the solution to (10). Nevertheless, we will prove that the maximin problem (10) indeed admits a saddle point, and therefore can be optimally solved. More importantly, we analytically characterize such a saddle point and provide a closed-form solution to (10). Our result implies that eigenmode transmission is the optimal transmit strategy in presence of deterministic CSI errors as in the cases of perfect CSI [4]–[8] and stochastic CSI [13]–[15].

Since the optimization variables $\mathbf{F}_r$ and $\mathbf{\Delta}_{rd}$ appear both inside and outside the matrix inversion of the objective function, we first apply the matrix inversion lemma [35] and perform some manipulations to rewrite it into the following form:

$$\begin{aligned}&\frac{1}{2}\log_2\Big|\mathbf{I}+\gamma_r\mathbf{H}_{sr}\mathbf{H}_{sr}^H-\gamma_r\mathbf{H}_{sr}\mathbf{H}_{sr}^H\Big(\mathbf{I}+\tilde{\mathbf{F}}_r^H\\&\quad\times(\hat{\mathbf{H}}_{rd}+\mathbf{\Delta}_{rd})^H(\hat{\mathbf{H}}_{rd}+\mathbf{\Delta}_{rd})\tilde{\mathbf{F}}_r\Big)^{-1}\Big|\\&\triangleq\text{MI}(\tilde{\mathbf{F}}_r,\mathbf{\Delta}_{rd}),\end{aligned}$$

where $\tilde{\mathbf{F}}_r \triangleq \frac{\sigma_r}{\sigma_d}\mathbf{F}_r$ and $\gamma_r \triangleq \frac{P_s}{N_s\sigma_r^2}$ denotes the signal to noise ratio (SNR) at the relay. Then, the problem (10) becomes

$$\begin{aligned}\max_{\tilde{\mathbf{F}}_r}\min_{\mathbf{\Delta}_{rd}}\quad&\text{MI}(\tilde{\mathbf{F}}_r,\mathbf{\Delta}_{rd})\\\text{subject to}\quad&\text{tr}(\tilde{\mathbf{F}}_r(\gamma_r\mathbf{H}_{sr}\mathbf{H}_{sr}^H+\mathbf{I})\tilde{\mathbf{F}}_r^H)\leq\gamma_dN_r,\\&\forall\mathbf{\Delta}_{rd}:\ \|\mathbf{\Delta}_{rd}\|_2\leq\epsilon_{rd},\end{aligned}\tag{12}$$

where $\gamma_d \triangleq \frac{P_r}{N_r\sigma_d^2}$ represents the destination SNR. At this point, it is worth pointing out that the main difficulty in solving the maximin problem (12) is that although the scaled precoder $\tilde{\mathbf{F}}_r$ and channel perturbation $\mathbf{\Delta}_{rd}$ belong to convex sets, the objective function $\text{MI}(\tilde{\mathbf{F}}_r,\mathbf{\Delta}_{rd})$ is not concave in $\tilde{\mathbf{F}}_r$ and convex in $\mathbf{\Delta}_{rd}$ so that classical saddle point theory results, such as Von Neumann's theorem [36, Section 6.2, Theorem 8] is not applicable. We would also like to point out that, although the S-lemma [37] is a powerful tool for handling worst-case robust design problems [30] [32], it can not be applied to our problem to obtain the optimal robust solution.

To proceed, let us consider the following minimax problem as a counterpart to the maximin problem (12)

$$\begin{aligned}\min_{\mathbf{\Delta}_{rd}}\max_{\tilde{\mathbf{F}}_r}\quad&\text{MI}(\tilde{\mathbf{F}}_r,\mathbf{\Delta}_{rd})\\\text{subject to}\quad&\text{tr}(\tilde{\mathbf{F}}_r(\gamma_r\mathbf{H}_{sr}\mathbf{H}_{sr}^H+\mathbf{I})\tilde{\mathbf{F}}_r^H)\leq\gamma_dN_r,\\&\forall\mathbf{\Delta}_{rd}:\ \|\mathbf{\Delta}_{rd}\|_2\leq\epsilon_{rd}.\end{aligned}\tag{13}$$

As will be shown later, studying this problem indeed paves the way to solving the original maximin problem (12). Let us introduce eigenvalue decomposition (EVD) $\mathbf{H}_{sr}\mathbf{H}_{sr}^H = \mathbf{U}_{sr}\mathbf{\Sigma}_{sr}\mathbf{U}_{sr}^H$ and singular value decomposition (SVD) $\hat{\mathbf{H}}_{rd} = \hat{\mathbf{U}}_{rd}\hat{\mathbf{\Lambda}}_{rd}\hat{\mathbf{V}}_{rd}^H$, where $\mathbf{\Sigma}_{sr} = \text{diag}\{\sigma_{sr,1}^2,\cdots,\sigma_{sr,N_s}^2,0,\cdots,0\}$ and $\hat{\mathbf{\Lambda}}_{rd} = \text{diag}\{\hat{\sigma}_{rd,1},\cdots,\hat{\sigma}_{rd,N_p},0,\cdots,0\}$ with $N_p = \min\{N_r,N_d\}$. Then, the minimax problem (13) admits a closed-form solution that can be described as follows.

Proposition 1: The worst-case CSI error $\mathbf{\Delta}_{rd}^w$ and the optimal relay precoder $\tilde{\mathbf{F}}_r^{opt}$ that together solve the minimax problem (13) are given respectively by (14) and

$$\tilde{\mathbf{F}}_r^{opt} = \hat{\mathbf{V}}_{rd}\mathbf{\Lambda}_{\tilde{f}_r}^{opt}(\mathbf{I}+\gamma_r\mathbf{\Sigma}_{sr})^{-1/2}\mathbf{U}_{sr}^H,\tag{15}$$

where $\mathbf{\Lambda}_{\tilde{f}_r}^{opt} = \text{diag}\{\tilde{f}_{r,1}^{opt},\cdots,\tilde{f}_{r,N_s}^{opt},0,\cdots,0\}$ and its ith diagonal entry takes the form of (16), where $\mu > 0$ is chosen such that $\sum_{i=1}^{N_s}(\tilde{f}_{r,i}^{opt})^2 = \gamma_dN_r$ is satisfied. In addition, the

$$\mathbf{\Delta}_{rd}^{w}=\begin{cases}-\hat{\mathbf{U}}_{rd}[\text{diag}\left\{\min\left\{\hat{\sigma}_{rd,1},\epsilon_{rd}\right\},\cdots,\min\left\{\hat{\sigma}_{rd,N_d},\epsilon_{rd}\right\}\right\}\ \ \mathbf{0}_{N_d\times(N_r-N_d)}]\hat{\mathbf{V}}_{rd}^{H}, & N_r>N_d\\ -\hat{\mathbf{U}}_{rd}[\text{diag}\left\{\min\left\{\hat{\sigma}_{rd,1},\epsilon_{rd}\right\},\cdots,\min\left\{\hat{\sigma}_{rd,N_r},\epsilon_{rd}\right\}\right\}\ \ \mathbf{0}_{N_r\times(N_d-N_r)}]^{T}\hat{\mathbf{V}}_{rd}^{H}, & N_d\geq N_r\end{cases} \tag{14}$$

$$\tilde{f}_{r,i}^{opt}=\begin{cases}\sqrt{\frac{\left(\sqrt{\gamma_r^2\sigma_{sr,i}^4+4\mu\gamma_r\sigma_{sr,i}^2(\hat{\sigma}_{rd,i}-\epsilon_{rd})^2}-\gamma_r\sigma_{sr,i}^2-2\right)_+}{2(\hat{\sigma}_{rd,i}-\epsilon_{rd})^2}}, & \hat{\sigma}_{rd,i}>\epsilon_{rd}\\ 0, & \hat{\sigma}_{rd,i}\leq\epsilon_{rd}\end{cases} \tag{16}$$

optimal value of the problem is

$$\begin{aligned}&\text{MI}_{\text{min,max}}\\ &=\frac{1}{2}\sum_{i=1}^{N_s}\log_2\left(1+\frac{\gamma_r\sigma_{sr,i}^2(\tilde{f}_{r,i}^{opt})^2(\hat{\sigma}_{rd,i}-\epsilon_{rd})_+^2}{1+\gamma_r\sigma_{sr,i}^2+(\tilde{f}_{r,i}^{opt})^2(\hat{\sigma}_{rd,i}-\epsilon_{rd})_+^2}\right).\end{aligned} \tag{17}$$

Proof: See Appendix II. ■

In general, the optimal solution of the minimax problem is not necessarily the same as that of the maximin problem. Nevertheless, from [36, Section 6.2, Proposition 1] (see also [38, Corollary 9.16]), we know that if there exists a saddle point in $\text{MI}(\tilde{\mathbf{F}}_r,\mathbf{\Delta}_{rd})$, then it is optimal for both problems at the same time. In the sequel, we prove that $(\tilde{\mathbf{F}}_r^{opt},\mathbf{\Delta}_{rd}^{w})$ obtained in *Proposition* 1 is indeed a saddle point. Before presenting this main result, we first introduce two lemmas that will be used.

Lemma 1: Given a Hermitian matrix $\mathbf{W}\in\mathbb{C}^{N\times N}$, suppose that its ith diagonal element $d_i(\mathbf{W})$ is lower bounded by $d_i^{'}(\mathbf{W})$, i.e., $d_i(\mathbf{W})\geq d_i^{'}(\mathbf{W}), i=1,\cdots,N$. Denote $\mathbf{d}^{'}(\mathbf{W})=[d_{i_1}^{'}(\mathbf{W}),\cdots,d_{i_N}^{'}(\mathbf{W})]^T$ whose elements are arranged in decreasing order. Let $\boldsymbol{\lambda}(\mathbf{W})=[\lambda_1(\mathbf{W}),\cdots,\lambda_N(\mathbf{W})]^T$ be a vector formed by arranging all the eigenvalues of $\mathbf{W}$ in descending order. Then, $\boldsymbol{\lambda}(\mathbf{W})\succ_w\mathbf{d}^{'}(\mathbf{W})$.

Proof: See Appendix III. ■

Lemma 2 ([39, Chapter 3, Theorem A.8]): A real-valued function ϕ defined on a set $\mathcal{A}\subset\mathbb{R}^n$ satisfies

$$\mathbf{x}\prec_w\mathbf{y}\text{ on }\mathcal{A}\Rightarrow\phi(\mathbf{x})\geq\phi(\mathbf{y}), \tag{18}$$

if and only if ϕ is decreasing and Schur-convex on $\mathcal{A}$.

Now we show the main conclusion of this section in the following theorem.

Theorem 1: The inequality $\text{MI}(\tilde{\mathbf{F}}_r,\mathbf{\Delta}_{rd}^{w})\leq\text{MI}(\tilde{\mathbf{F}}_r^{opt},\mathbf{\Delta}_{rd}^{w})\leq\text{MI}(\tilde{\mathbf{F}}_r^{opt},\mathbf{\Delta}_{rd})$ holds for any admissible $\tilde{\mathbf{F}}_r$ and $\mathbf{\Delta}_{rd}$, meaning that $(\tilde{\mathbf{F}}_r^{opt},\mathbf{\Delta}_{rd}^{w})$ is a saddle point of $\text{MI}(\tilde{\mathbf{F}}_r,\mathbf{\Delta}_{rd})$ and is hence optimal for both maximin problem (12) and minimax problem (13).

Proof: By fixing $\mathbf{\Delta}_{rd}=\mathbf{\Delta}_{rd}^{w}$ in $\text{MI}(\tilde{\mathbf{F}}_r,\mathbf{\Delta}_{rd})$, it is easy to verify that $\text{MI}(\tilde{\mathbf{F}}_r,\mathbf{\Delta}_{rd}^{w})\leq\text{MI}(\tilde{\mathbf{F}}_r^{opt},\mathbf{\Delta}_{rd}^{w})$ holds based on the results in [4]. However, showing $\text{MI}(\tilde{\mathbf{F}}_r^{opt},\mathbf{\Delta}_{rd})\geq\text{MI}(\tilde{\mathbf{F}}_r^{opt},\mathbf{\Delta}_{rd}^{w})$ is more intricate. We first express $\text{MI}(\tilde{\mathbf{F}}_r^{opt},\mathbf{\Delta}_{rd})$ as

$$\begin{aligned}&\text{MI}(\tilde{\mathbf{F}}_r^{opt},\mathbf{\Delta}_{rd})=\log_2|\mathbf{I}+\gamma_r\mathbf{\Sigma}_{sr}|\\ &+\log_2\frac{\left|\mathbf{I}+\mathbf{\Lambda}_{\tilde{f}_r}^{opt}(\hat{\mathbf{\Lambda}}_{rd}+\tilde{\mathbf{\Delta}}_{rd})^H(\hat{\mathbf{\Lambda}}_{rd}+\tilde{\mathbf{\Delta}}_{rd})\mathbf{\Lambda}_{\tilde{f}_r}^{opt}\right|}{\left|\mathbf{I}+\gamma_r\mathbf{\Sigma}_{sr}+\mathbf{\Lambda}_{\tilde{f}_r}^{opt}(\hat{\mathbf{\Lambda}}_{rd}+\tilde{\mathbf{\Delta}}_{rd})^H(\hat{\mathbf{\Lambda}}_{rd}+\tilde{\mathbf{\Delta}}_{rd})\mathbf{\Lambda}_{\tilde{f}_r}^{opt}\right|},\end{aligned} \tag{19}$$

where $\tilde{\mathbf{\Delta}}_{rd}=\hat{\mathbf{U}}_{rd}^H\mathbf{\Delta}_{rd}\hat{\mathbf{V}}_{rd}$. Then, proving $\text{MI}(\tilde{\mathbf{F}}_r^{opt},\mathbf{\Delta}_{rd})\geq\text{MI}(\tilde{\mathbf{F}}_r^{opt},\mathbf{\Delta}_{rd}^{w})$ is equivalent to showing

$$\begin{aligned}&\frac{\left|\mathbf{I}+\mathbf{\Lambda}_{\tilde{f}_r}^{opt}(\hat{\mathbf{\Lambda}}_{rd}+\tilde{\mathbf{\Delta}}_{rd})^H(\hat{\mathbf{\Lambda}}_{rd}+\tilde{\mathbf{\Delta}}_{rd})\mathbf{\Lambda}_{\tilde{f}_r}^{opt}\right|}{\left|\mathbf{I}+\gamma_r\mathbf{\Sigma}_{sr}+\mathbf{\Lambda}_{\tilde{f}_r}^{opt}(\hat{\mathbf{\Lambda}}_{rd}+\tilde{\mathbf{\Delta}}_{rd})^H(\hat{\mathbf{\Lambda}}_{rd}+\tilde{\mathbf{\Delta}}_{rd})\mathbf{\Lambda}_{\tilde{f}_r}^{opt}\right|}\\ &\geq\prod_{i=1}^{N_s}\frac{1+(\tilde{f}_{r,i}^{opt})^2(\hat{\sigma}_{rd,i}-\epsilon_{rd})_+^2}{1+\gamma_r\sigma_{sr,i}^2+(\tilde{f}_{r,i}^{opt})^2(\hat{\sigma}_{rd,i}-\epsilon_{rd})_+^2}.\end{aligned} \tag{20}$$

Note that the matrix $\tilde{\mathbf{\Delta}}_{rd}$ appears in both the nominator and denominator of the left-hand side of (20), which complicates the proof.

Let us consider the following inequality from taking the reciprocal of both sides of (20)

$$\begin{aligned}&\frac{\left|\mathbf{I}+\gamma_r\mathbf{\Sigma}_{sr}+\mathbf{\Lambda}_{\tilde{f}_r}^{opt}(\hat{\mathbf{\Lambda}}_{rd}+\tilde{\mathbf{\Delta}}_{rd})^H(\hat{\mathbf{\Lambda}}_{rd}+\tilde{\mathbf{\Delta}}_{rd})\mathbf{\Lambda}_{\tilde{f}_r}^{opt}\right|}{\left|\mathbf{I}+\mathbf{\Lambda}_{\tilde{f}_r}^{opt}(\hat{\mathbf{\Lambda}}_{rd}+\tilde{\mathbf{\Delta}}_{rd})^H(\hat{\mathbf{\Lambda}}_{rd}+\tilde{\mathbf{\Delta}}_{rd})\mathbf{\Lambda}_{\tilde{f}_r}^{opt}\right|}\\ &\leq\prod_{i=1}^{N_s}\frac{1+\gamma_r\sigma_{sr,i}^2+(\tilde{f}_{r,i}^{opt})^2(\hat{\sigma}_{rd,i}-\epsilon_{rd})_+^2}{1+(\tilde{f}_{r,i}^{opt})^2(\hat{\sigma}_{rd,i}-\epsilon_{rd})_+^2}.\end{aligned} \tag{21}$$

Without loss of generality, we assume that the first N_l diagonal elements of matrix $\mathbf{\Sigma}_{sr}$ are non-zero, i.e., $\mathbf{\Sigma}_{sr}=\text{diag}\{\sigma_{sr,1}^2,\cdots,\sigma_{sr,N_l}^2,0,\cdots,0\}$ with $\sigma_{sr,i}^2>0, i=1,\cdots,N_l$. Then, based on (16), $\tilde{f}_{r,i}^{opt}=0,\ i=N_l+1,\cdots,N_r$. Moreover, the (N_l+1)th to N_rth rows and columns of matrix $\mathbf{\Lambda}_{\tilde{f}_r}^{opt}(\hat{\mathbf{\Lambda}}_{rd}+\tilde{\mathbf{\Delta}}_{rd})^H(\hat{\mathbf{\Lambda}}_{rd}+\tilde{\mathbf{\Delta}}_{rd})\mathbf{\Lambda}_{\tilde{f}_r}^{opt}$ all become zero. We denote the first N_l rows and columns of matrix $\gamma_r\mathbf{\Sigma}_{sr}$ with $\tilde{\mathbf{\Sigma}}_{sr}=\text{diag}\{\tilde{\sigma}_{sr,1}^2,\cdots,\tilde{\sigma}_{sr,N_l}^2\}$, let $\mathbf{\Omega}$ be the first N_l rows and columns of matrix $\mathbf{\Lambda}_{\tilde{f}_r}^{opt}(\hat{\mathbf{\Lambda}}_{rd}+\tilde{\mathbf{\Delta}}_{rd})^H(\hat{\mathbf{\Lambda}}_{rd}+\tilde{\mathbf{\Delta}}_{rd})\mathbf{\Lambda}_{\tilde{f}_r}^{opt}$ and also denote $\mathbf{\Psi}=\tilde{\mathbf{\Sigma}}_{sr}^{-1}+\tilde{\mathbf{\Sigma}}_{sr}^{-1/2}\mathbf{\Omega}\tilde{\mathbf{\Sigma}}_{sr}^{-1/2}$. Accordingly, the left-hand side of (21) can be rewritten as

$$\frac{\left|\mathbf{I}+\tilde{\mathbf{\Sigma}}_{sr}+\mathbf{\Omega}\right|}{|\mathbf{I}+\mathbf{\Omega}|}=\left|\mathbf{I}+\tilde{\mathbf{\Sigma}}_{sr}(\mathbf{I}+\mathbf{\Omega})^{-1}\right|$$
$$=\left|\mathbf{I}+\tilde{\mathbf{\Sigma}}_{sr}^{1/2}(\mathbf{I}+\mathbf{\Omega})^{-1}\tilde{\mathbf{\Sigma}}_{sr}^{1/2}\right|$$
$$=\left|\mathbf{I}+(\tilde{\mathbf{\Sigma}}_{sr}^{-1}+\tilde{\mathbf{\Sigma}}_{sr}^{-1/2}\mathbf{\Omega}\tilde{\mathbf{\Sigma}}_{sr}^{-1/2})^{-1}\right|$$
$$=\prod_{i=1}^{N_l}\left(1+\lambda_i^{-1}(\mathbf{\Psi})\right). \quad (22)$$

We consider the ith diagonal element of matrix $\mathbf{\Psi}$, which is given by

$$d_i(\mathbf{\Psi})$$
$$=\tilde{\sigma}_{sr,i}^{-2}+\tilde{\sigma}_{sr,i}^{-2}(\tilde{f}_{r,i}^{opt})^2\left(\sum_{j\neq i}\left|(\tilde{\mathbf{\Delta}}_{rd})_{i,j}\right|^2+\left|\hat{\sigma}_{rd,i}+(\tilde{\mathbf{\Delta}}_{rd})_{i,i}\right|^2\right)$$
$$\geq\tilde{\sigma}_{sr,i}^{-2}+\tilde{\sigma}_{sr,i}^{-2}(\tilde{f}_{r,i}^{opt})^2\left|\hat{\sigma}_{rd,i}+(\tilde{\mathbf{\Delta}}_{rd})_{i,i}\right|^2$$
$$\geq\tilde{\sigma}_{sr,i}^{-2}+\tilde{\sigma}_{sr,i}^{-2}(\tilde{f}_{r,i}^{opt})^2\left(\hat{\sigma}_{rd,i}-\left|(\tilde{\mathbf{\Delta}}_{rd})_{i,i}\right|\right)_+^2. \quad (23)$$

Since $\left|(\tilde{\mathbf{\Delta}}_{rd})_{i,i}\right|\leq\|\tilde{\mathbf{\Delta}}_{rd}\|_2$ [39, Chapter 9, Theorem D.1] and $\|\tilde{\mathbf{\Delta}}_{rd}\|_2=\|\mathbf{\Delta}_{rd}\|_2\leq\epsilon_{rd}$, $d_i(\mathbf{\Psi})$ is lower bounded by

$$d_i(\mathbf{\Psi})\geq\tilde{\sigma}_{sr,i}^{-2}+\tilde{\sigma}_{sr,i}^{-2}(\tilde{f}_{r,i}^{opt})^2\left(\hat{\sigma}_{rd,i}-\epsilon_{rd}\right)_+^2\triangleq d_i^{'}(\mathbf{\Psi}). \quad (24)$$

Then, by applying *Lemma* 1, we have $\boldsymbol{\lambda}(\mathbf{\Psi})\succ_w \mathbf{d}^{'}(\mathbf{\Psi})$.

Note that, given $x_i>0$, $f(\mathbf{x})=\prod_{i=1}^{N_l}\left(1+\frac{1}{x_i}\right)$ is decreasing in each x_i. In addition, as the function $\log\left(1+\frac{1}{x_i}\right)$ is convex, $f(\mathbf{x})$ is a Schur-convex function (according to *Lemma* 4 in Appendix I). By using *Lemma* 2, we immediately find

$$\prod_{i=1}^{N_l}(1+\lambda_i^{-1}(\mathbf{\Psi}))\leq\prod_{i=1}^{N_l}(1+(d_i^{'}(\mathbf{\Psi}))^{-1})$$
$$=\prod_{i=1}^{N_l}\frac{1+\gamma_r\sigma_{sr,i}^2+(\tilde{f}_{r,i}^{opt})^2(\hat{\sigma}_{rd,i}-\epsilon_{rd})_+^2}{1+(\tilde{f}_{r,i}^{opt})^2(\hat{\sigma}_{rd,i}-\epsilon_{rd})_+^2},$$

which is equivalent to (21). Therefore, we have proved that the inequality $\mathrm{MI}(\tilde{\mathbf{F}}_r,\mathbf{\Delta}_{rd}^w)\leq\mathrm{MI}(\tilde{\mathbf{F}}_r^{opt},\mathbf{\Delta}_{rd}^w)\leq\mathrm{MI}(\tilde{\mathbf{F}}_r^{opt},\mathbf{\Delta}_{rd})$ holds. Then, according to [36, Section 6.2, Proposition 1], $(\tilde{\mathbf{F}}_r^{opt},\mathbf{\Delta}_{rd}^w)$ is a saddle point in $\mathrm{MI}(\tilde{\mathbf{F}}_r^{opt},\mathbf{\Delta}_{rd})$ and hence is the optimal solution to both the maximin problem (12) and minimax problem (13). ∎

The importance of such a result lies in:

- *Theorem* 1 provides some interesting insights into the worst-case MI maximization for MIMO relaying. It can be found from (14) that, the worst-case CSI uncertainty $\mathbf{\Delta}_{rd}^w$ has the similar SVD structure as the nominal channel $\hat{\mathbf{H}}_{rd}$ and it decreases the ith singular value of $\hat{\mathbf{H}}_{rd}$ by $\min\{\hat{\sigma}_{rd,i},\epsilon_{rd}\}$, which, intuitively, can be explained by the fact that the worst-case CSI perturbation shall attempt to degrade the nominal channel as much as possible. From (15), we observe that the optimal relay precoder has a channel-diagonalizing structure, where the unitary matrices $\mathbf{U}_{sr}^H$ and $\hat{\mathbf{V}}_{rd}$ match the eigen-directions of the source-relay channel $\mathbf{H}_{sr}$ and the nominal relay-destination channel $\hat{\mathbf{H}}_{rd}$, respectively. The power allocation matrix $\mathbf{\Lambda}_{\tilde{f}_r}^{opt}$ is similar to the one obtained under the perfect CSI assumption [4], but with the original singular value $\hat{\sigma}_{rd,i}$ replaced by a degraded version $(\hat{\sigma}_{rd,i}-\epsilon_{rd})_+$. Therefore, our proposed robust design requires nearly the same computational complexity as the non-robust scheme in [4].
- In prior works [20] and [21], the worst-case MI maximization for point-to-point MIMO systems with spectral norm bounded CSI errors was studied. The latter work proved that the optimal solution is actually a saddle point to the original maximin problem and that eigenmode transmission is the optimum transmit strategy. The main technical challenge of [21] lies in the proof of inequality (19) in that paper. As a byproduct, it is easy to verify that the inequality used in [21] can be alternatively derived by following the proof of *Theorem* 1 with mild modifications.

IV. Worst-Case MSE Minimization for AF MIMO Relaying

In the previous section, we have examined the problem of mutual information maximization against worst-case channel uncertainties for improving the spectral efficiency of AF MIMO relaying. In this section, we consider the worst-case MSE minimization problem (11) to enhance the reliability of the MIMO relay system subject to norm-bounded CSI uncertainties. Different from the worst-case MI maximization, the MSE minimization problem contains joint optimization of the transceiver, where the optimal receiver is given by the linear MMSE decoder in (6).

Note that the objective function of problem (11) is more involved than that of problem (10) due to the multiple levels of matrix inversions. To overcome this difficulty, we use variables $\tilde{\mathbf{F}}_r$, γ_r and γ_d defined in (12) and perform some matrix manipulations to express the MSE as (25), where (a) is obtained from the matrix inversion lemma and (b) holds since $\mathrm{tr}(\mathbf{AB})=\mathrm{tr}(\mathbf{BA})$. Then, the problem (11) becomes

$$\begin{aligned}\min_{\tilde{\mathbf{F}}_r}\max_{\mathbf{\Delta}_{rd}}\quad & \mathrm{MSE}(\tilde{\mathbf{F}}_r,\mathbf{\Delta}_{rd})\\ \text{subject to}\quad & \mathrm{tr}(\tilde{\mathbf{F}}_r(\gamma_r\mathbf{H}_{sr}\mathbf{H}_{sr}^H+\mathbf{I})\tilde{\mathbf{F}}_r^H)\leq\gamma_d N_r,\\ & \forall\mathbf{\Delta}_{rd}:\ \|\mathbf{\Delta}_{rd}\|_2\leq\epsilon_{rd}.\end{aligned} \quad (26)$$

Like the MI objective function considered in the previous section, the function $\mathrm{MSE}(\tilde{\mathbf{F}}_r,\mathbf{\Delta}_{rd})$ does not have a convex-concave structure which makes searching the solution of problem (26) rather challenging. As before, instead of directly solving the above problem, we first consider its counterpart maximin problem, whose solution is given by the following proposition.

Proposition 2: The maximin problem

$$\begin{aligned}\max_{\mathbf{\Delta}_{rd}}\min_{\tilde{\mathbf{F}}_r}\quad & \mathrm{MSE}(\tilde{\mathbf{F}}_r,\mathbf{\Delta}_{rd})\\ \text{subject to}\quad & \mathrm{tr}(\tilde{\mathbf{F}}_r(\gamma_r\mathbf{H}_{sr}\mathbf{H}_{sr}^H+\mathbf{I})\tilde{\mathbf{F}}_r^H)\leq\gamma_d N_r,\\ & \forall\mathbf{\Delta}_{rd}:\ \|\mathbf{\Delta}_{rd}\|_2\leq\epsilon_{rd}\end{aligned} \quad (27)$$

$$
\begin{aligned}
&\frac{P_s}{N_s}\mathrm{tr}\left(\left(\mathbf{I}+\gamma_r\mathbf{H}_{sr}^H\tilde{\mathbf{F}}_r^H(\hat{\mathbf{H}}_{rd}+\boldsymbol{\Delta}_{rd})^H\left(\mathbf{I}+(\hat{\mathbf{H}}_{rd}+\boldsymbol{\Delta}_{rd})\tilde{\mathbf{F}}_r\tilde{\mathbf{F}}_r^H(\hat{\mathbf{H}}_{rd}+\boldsymbol{\Delta}_{rd})^H\right)^{-1}(\hat{\mathbf{H}}_{rd}+\boldsymbol{\Delta}_{rd})\tilde{\mathbf{F}}_r\mathbf{H}_{sr}\right)^{-1}\right)\\
&\overset{(a)}{=}\frac{P_s}{N_s}\mathrm{tr}\Big(\mathbf{I}-\gamma_r\mathbf{H}_{sr}^H\tilde{\mathbf{F}}_r^H(\hat{\mathbf{H}}_{rd}+\boldsymbol{\Delta}_{rd})^H\Big(\mathbf{I}+(\hat{\mathbf{H}}_{rd}+\boldsymbol{\Delta}_{rd})\tilde{\mathbf{F}}_r\tilde{\mathbf{F}}_r^H(\hat{\mathbf{H}}_{rd}+\boldsymbol{\Delta}_{rd})^H+\gamma_r(\hat{\mathbf{H}}_{rd}+\boldsymbol{\Delta}_{rd})\tilde{\mathbf{F}}_r\mathbf{H}_{sr}\\
&\quad\times\mathbf{H}_{sr}^H\tilde{\mathbf{F}}_r^H(\hat{\mathbf{H}}_{rd}+\boldsymbol{\Delta}_{rd})^H\Big)^{-1}(\hat{\mathbf{H}}_{rd}+\boldsymbol{\Delta}_{rd})\tilde{\mathbf{F}}_r\mathbf{H}_{sr}\Big)\\
&\overset{(b)}{=}\frac{P_s}{N_s}\mathrm{tr}\Big(\mathbf{I}-\gamma_r(\hat{\mathbf{H}}_{rd}+\boldsymbol{\Delta}_{rd})\tilde{\mathbf{F}}_r\mathbf{H}_{sr}\mathbf{H}_{sr}^H\tilde{\mathbf{F}}_r^H(\hat{\mathbf{H}}_{rd}+\boldsymbol{\Delta}_{rd})^H\Big(\mathbf{I}+(\hat{\mathbf{H}}_{rd}+\boldsymbol{\Delta}_{rd})\tilde{\mathbf{F}}_r\tilde{\mathbf{F}}_r^H(\hat{\mathbf{H}}_{rd}+\boldsymbol{\Delta}_{rd})^H\\
&\quad+\gamma_r(\hat{\mathbf{H}}_{rd}+\boldsymbol{\Delta}_{rd})\tilde{\mathbf{F}}_r\mathbf{H}_{sr}\mathbf{H}_{sr}^H\tilde{\mathbf{F}}_r^H(\hat{\mathbf{H}}_{rd}+\boldsymbol{\Delta}_{rd})^H\Big)^{-1}\Big)\\
&=\frac{P_s}{N_s}\mathrm{tr}\left(\left(\mathbf{I}+(\hat{\mathbf{H}}_{rd}+\boldsymbol{\Delta}_{rd})\tilde{\mathbf{F}}_r(\mathbf{I}+\gamma_r\mathbf{H}_{sr}\mathbf{H}_{sr}^H)\tilde{\mathbf{F}}_r^H(\hat{\mathbf{H}}_{rd}+\boldsymbol{\Delta}_{rd})^H\right)^{-1}\left(\mathbf{I}+(\hat{\mathbf{H}}_{rd}+\boldsymbol{\Delta}_{rd})\tilde{\mathbf{F}}_r\tilde{\mathbf{F}}_r^H(\hat{\mathbf{H}}_{rd}+\boldsymbol{\Delta}_{rd})^H\right)\right)\\
&\triangleq\mathrm{MSE}(\tilde{\mathbf{F}}_r,\boldsymbol{\Delta}_{rd})
\end{aligned}
\tag{25}
$$

$$
\boldsymbol{\Delta}_{rd}^w=\begin{cases}-\hat{\mathbf{U}}_{rd}[\mathrm{diag}\left\{\min\left\{\hat{\sigma}_{rd,1},\epsilon_{rd}\right\},\cdots,\min\left\{\hat{\sigma}_{rd,N_d},\epsilon_{rd}\right\}\right\}\ \ \mathbf{0}_{N_d\times(N_r-N_d)}]\hat{\mathbf{V}}_{rd}^H, & N_r>N_d\\ -\hat{\mathbf{U}}_{rd}[\mathrm{diag}\left\{\min\left\{\hat{\sigma}_{rd,1},\epsilon_{rd}\right\},\cdots,\min\left\{\hat{\sigma}_{rd,N_r},\epsilon_{rd}\right\}\right\}\ \ \mathbf{0}_{N_r\times(N_d-N_r)}]^T\hat{\mathbf{V}}_{rd}^H, & N_d\geq N_r\end{cases}
\tag{28}
$$

---

has a closed-form solution where the worst-case CSI uncertainty is (28) and the optimal relay precoder is given by

$$
\tilde{\mathbf{F}}_r^{opt}=\hat{\mathbf{V}}_{rd}\boldsymbol{\Lambda}_{\tilde{f}_r}^{opt}(\mathbf{I}+\gamma_r\boldsymbol{\Sigma}_{sr})^{-1/2}\mathbf{U}_{sr}^H, \tag{29}
$$

with $\boldsymbol{\Lambda}_{\tilde{f}_r}^{opt}=\mathrm{diag}\{\tilde{f}_{r,1}^{opt},\cdots,\tilde{f}_{r,N_s}^{opt},0,\cdots,0\}$, whose ith diagonal element is

$$
\tilde{f}_{r,i}^{opt}=\begin{cases}\sqrt{\dfrac{\left(\sqrt{\frac{\gamma_r\sigma_{sr,i}^2(\hat{\sigma}_{rd,i}-\epsilon_{rd})^2}{\nu(1+\gamma_r\sigma_{sr,i}^2)}}-1\right)_+}{(\hat{\sigma}_{rd,i}-\epsilon_{rd})^2}}, & \hat{\sigma}_{rd,i}>\epsilon_{rd}\\ 0, & \hat{\sigma}_{rd,i}\leq\epsilon_{rd},\end{cases} \tag{30}
$$

where $\nu>0$ is chosen such that $\sum_{i=1}^{N_s}(\tilde{f}_{r,i}^{opt})^2=\gamma_d N_r$ is satisfied. Moreover, the optimal value of this problem is

$$
\begin{aligned}
&\mathrm{MSE}_{\max,\min}\\
&=\frac{P_s}{N_s}\sum_{i=1}^{N_s}\frac{1+\gamma_r\sigma_{sr,i}^2+(\tilde{f}_{r,i}^{opt})^2(\hat{\sigma}_{rd,i}-\epsilon_{rd})_+^2}{(1+\gamma_r\sigma_{sr,i}^2)(1+(\tilde{f}_{r,i}^{opt})^2(\hat{\sigma}_{rd,i}-\epsilon_{rd})_+^2)}.
\end{aligned} \tag{31}
$$

Proof: Please refer to Appendix IV. ■

Like the worst-case MI optimization problem studied in the previous section, the worst-case MSE minimization problem also admits a saddle point, which, according to the following theorem, is given exactly by the solution in *Proposition* 2.

Theorem 2: The inequality $\mathrm{MSE}(\tilde{\mathbf{F}}_r^{opt},\boldsymbol{\Delta}_{rd})\leq\mathrm{MSE}(\tilde{\mathbf{F}}_r^{opt},\boldsymbol{\Delta}_{rd}^w)\leq\mathrm{MSE}(\tilde{\mathbf{F}}_r,\boldsymbol{\Delta}_{rd}^w)$ holds for any admissible $\tilde{\mathbf{F}}_r$ and $\boldsymbol{\Delta}_{rd}$. Hence, $(\tilde{\mathbf{F}}_r^{opt},\boldsymbol{\Delta}_{rd}^w)$ is a saddle point of $\mathrm{MSE}(\tilde{\mathbf{F}}_r,\boldsymbol{\Delta}_{rd})$ and hence optimal for both minimax problem (26) and maximin problem (27).

Proof: See Appendix V. ■

The results derived up to this point lead to the following observations:

- Interestingly, by comparing the above theorem and *Theorem* 1, we find that their conclusions are consistent in that they both imply the optimality of eigenmode transmission. The major difference between them lies in the power allocation matrices due to the fact that the design objective functions are different. At this moment, we are able to conclude that the optimal robust transceiver that minimizes the worst-case MSE of AF MIMO relaying consists of the relay precoder in (29) and the Wiener filter in (6).
- In both Sections III and IV, we consider the robust optimization under the additive channel uncertainty model. Note that [20] studied another type of multiplicative uncertainty model that can be used to characterize calibration errors. Under this channel model, the actual relay-destination channel is given by $\mathbf{H}_{rd}=(\mathbf{I}+\mathbf{E}_{rd})\hat{\mathbf{H}}_{rd}$, where the multiplicative CSI uncertainty $\mathbf{E}_{rd}$ satisfies $\|\mathbf{E}_{rd}\|_2\leq\kappa_{rd}<1$. For the multiplicative uncertainty model, we can also obtain the closed-form solution to the corresponding robust optimization problem using the analogous techniques as in the additive uncertainty case and we omit the detailed steps for brevity. We find that the solution to the worst-case MI maximization problem with multiplicative uncertainty differs from that of the additive case in two aspects: 1) $(\hat{\sigma}_{rd,i}-\epsilon_{rd})_+$ in (16) is replaced by $(1-\kappa_{rd})\hat{\sigma}_{rd,i}$; 2) the worst-case CSI perturbation is $\mathbf{E}_{rd}^w=-\kappa_{rd}\mathbf{I}$. This finding also applies to the solution of the worst-case MSE minimization problem with multiplicative uncertainty.

V. Simulation Results

In this section, we present some numerical results for our proposed robust designs concerning a two-hop AF MIMO relay system. The channels on both hops are assumed to have independent and identically distributed (i.i.d.) complex

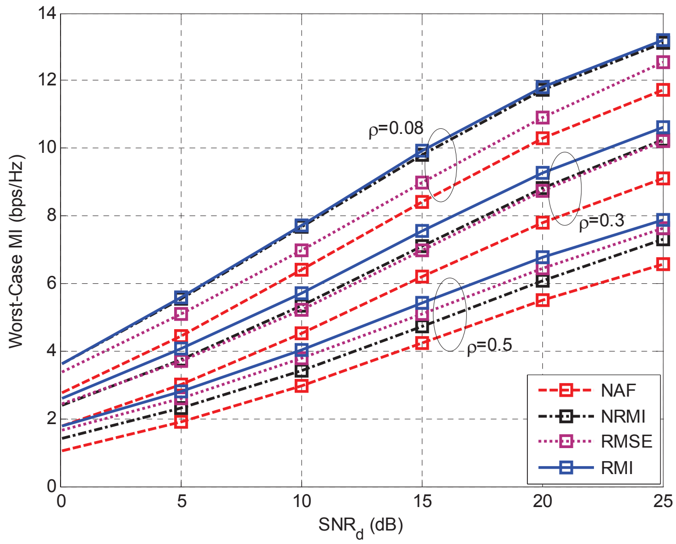

Fig. 2. Worst-case MI performance versus SNR_d with different ρ ($N_s = N_r = N_d = 4$, SNR_r = 20 dB).

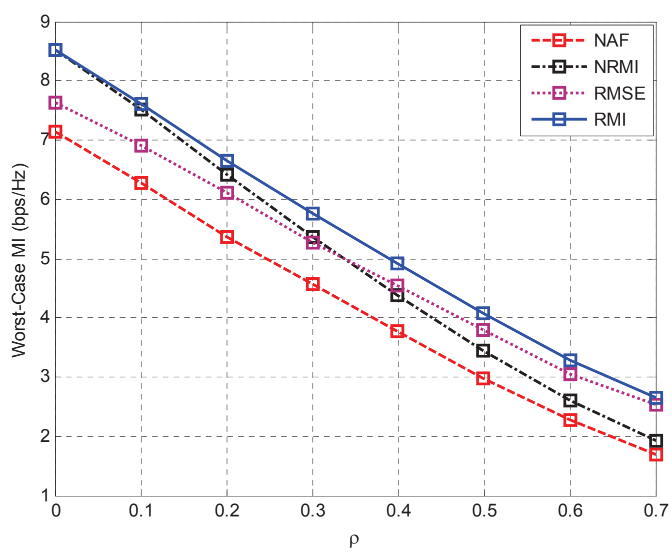

Fig. 3. Worst-case MI performance versus ρ ($N_s = N_r = N_d = 4$, SNR_r = 20 dB, SNR_d = 10 dB).

Gaussian entries with zero mean and unit variance. We define the relay and destination SNR by $\text{SNR}_r = \frac{P_s}{N_s \sigma_r^2}$ and $\text{SNR}_d = \frac{P_r}{N_r \sigma_d^2}$, respectively. We also use the parameter $\rho = \epsilon_{rd}/\|\hat{\mathbf{H}}_{rd}\|_2$ to denote the normalized radius of the norm bounded channel uncertainty region which reflects the quality of CSI, i.e., the larger ρ is, the poorer CSI quality will be. In our simulations, we fix ρ and hence ϵ_{rd} is adapted to each mismatched channel realization. We investigate the performance of the following schemes for AF MIMO relaying:

- Non-robust MI maximization in [4] (NRMI)
- Robust MI maximization in *Theorem* 1 (RMI)
- Non-robust MSE minimization in [6] (NRMSE)
- Robust MSE minimization in *Theorem* 2 (RMSE)
- Naive AF transmit strategy (NAF)

where in NAF, the relay utilizes no special precoding and simply scales the received signal with the constant $\frac{\sigma_d}{\sigma_r}\sqrt{\frac{\gamma_d N_r}{\text{tr}(\mathbf{I}+\gamma_r \mathbf{H}_{sr}\mathbf{H}_{sr}^H)}}$ according to the power constraint (2). For the numerical tests below, we generate estimated relay-destination channels with random matrices whose entries are independent and identically distributed (i.i.d.) complex Gaussian variables with zero mean and unit variance. Moreover, for the above schemes, the worst-case channel error changes with $\hat{\mathbf{H}}_{rd}$ according to (14) which can be readily verified from the proof of *Theorem* 1.

A. Worst-Case Performance Evaluation

Fig. 2 shows the worst-case MI performance of the NAF, NRMI, RMSE and RMI schemes as a function of SNR_d, the destination SNR. The SNR at the relay SNR_r is fixed at 20 dB. It can be observed that when the uncertainty size is relatively large, the performance gain of the robust scheme over the non-robust one is evident under various SNR_d. The NAF scheme exhibits the worst performance among all of three strategies and the performance of the RMSE scheme is inferior to that of the RMI scheme as expected, since it aims to minimize the worst-case MSE instead of maximizing the worst-case MI. Interestingly, the RMSE method still outperforms NRMI and NAF strategies when ρ is large, which is due to the fact that it takes into account the CSI mismatch when performing the precoder optimization. Also, as shown in Fig. 3, if we fix SNR_d, the gap between the RMI and NRMI schemes increases as ρ becomes larger. Therefore, our proposed RMI design technique indeed improves the spectral efficiency of AF MIMO relaying with bounded CSI uncertainties.

Figs. 4 and 5 compare the worst-case MSE performance of the NAF, NRMSE, RMI and RMSE schemes. We observe that RMSE can achieve a noticeable performance gain over NRMSE, especially for a large uncertainty size or high SNR. The RMI scheme performs worse than RMSE because its optimization goal is worst-case MI. The RMI scheme outperforms NRMSE and NAF methods for large ρ since it considers CSI uncertainties when optimizing the relay precoder, however, it performs worse than RMSE because its optimization goal is worst-case MI. Accordingly, our proposed RMSE design can provide robustness against CSI errors in terms of the MSE metric compared to other strategies.

Finally, we evaluate the worst-case bit error rate (BER) performance of our proposed RMSE scheme in Figs. 6 and 7, where we adopt binary phase shift keying (BPSK) modulation. It can be found from Fig. 6 that the robust scheme RMSE outperforms other methods especially in the low and medium SNR regimes. The RMI scheme performs worse than RMSE since it aims to improve the system efficiency instead of reliability. However, since RMI incorporates CSI errors into the precoder optimization, it exhibits better performance than NRMSE and NAF schemes when ρ is large. Fig. 7 shows that RMSE has the lowest worst-case BER under different ρ, and its superiority over the NRMSE scheme becomes more pronounced as ρ increases. Thereby, we conclude that RMSE can enhance the reliability of AF MIMO relay systems with bounded CSI errors.

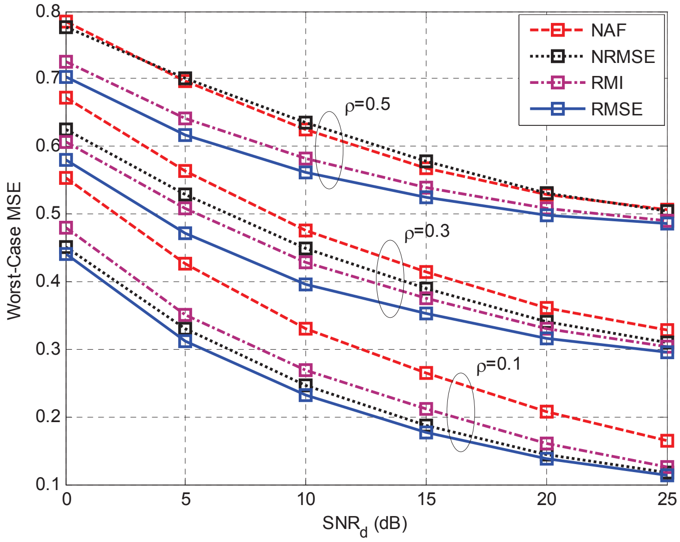

Fig. 4. Worst-case MSE performance versus SNR_d with different ρ ($N_s = N_r = N_d = 4$, SNR_r = 20 dB).

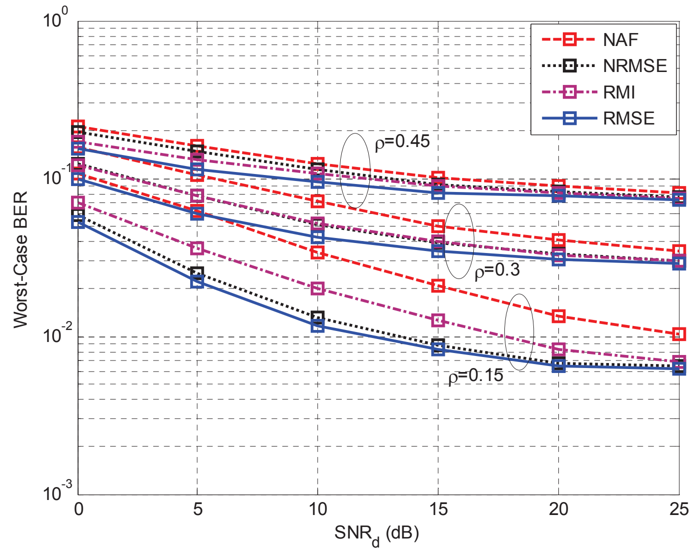

Fig. 6. Worst-case BER performance versus SNR_d with different ρ ($N_s = N_d = 4$, $N_r = 5$, SNR_r = 20 dB)

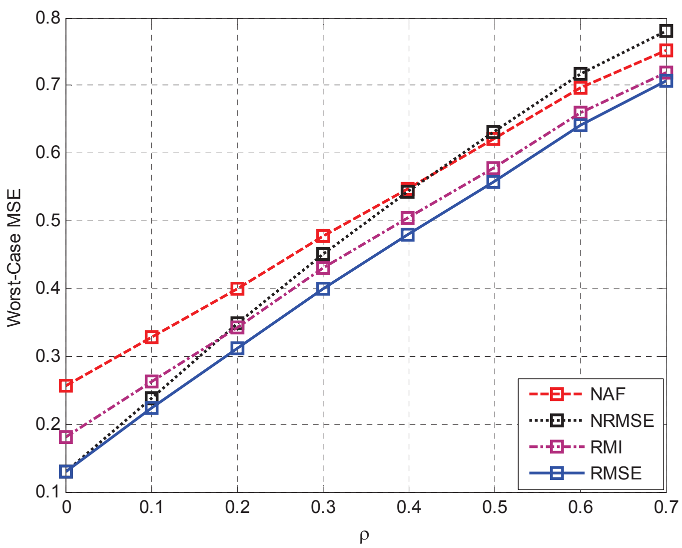

Fig. 5. Worst-case MSE performance versus ρ ($N_s = N_r = N_d = 4$, SNR_r = 20 dB, SNR_d = 10 dB).

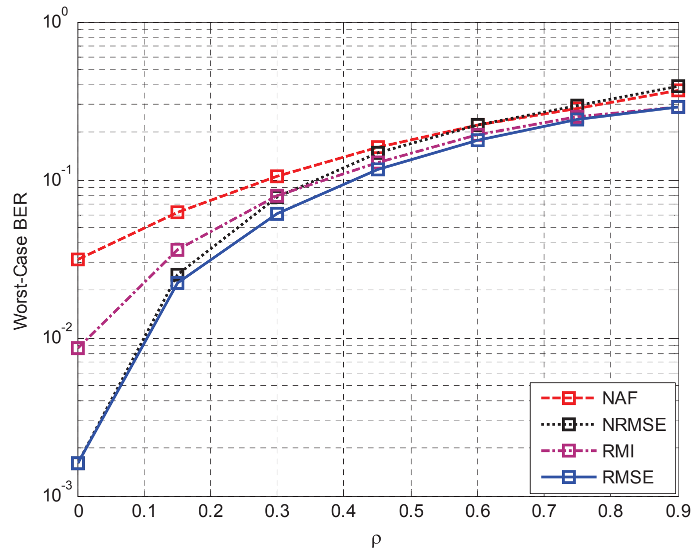

Fig. 7. Worst-case BER performance versus ρ ($N_s = N_d = 4$, $N_r = 5$, SNR_r = 20 dB, SNR_d = 5 dB).

B. QoS Consideration

The robust design can not only improve the worst-case performance of an AF MIMO relay system, but also effectively enhance its QoS performance with presence of deterministic CSI errors. Specifically, we can formulate the QoS problems with respect to MI and MSE metrics as

$$\begin{aligned} \underset{\tilde{\mathbf{F}}_r, p}{\text{minimize}} \quad & p \\ \text{subject to} \quad & \sigma_d^2 \text{tr}(\tilde{\mathbf{F}}_r(\gamma_r \mathbf{H}_{sr}\mathbf{H}_{sr}^H + \mathbf{I})\tilde{\mathbf{F}}_r^H) \le p, \\ & \text{MI}(\tilde{\mathbf{F}}_r, \mathbf{\Delta}_{rd}) \ge \tau_{\text{MI}}, \\ & \forall \mathbf{\Delta}_{rd} : \|\mathbf{\Delta}_{rd}\|_2 \le \epsilon_{rd}, \end{aligned} \tag{32}$$

and

$$\begin{aligned} \underset{\tilde{\mathbf{F}}_r, p}{\text{minimize}} \quad & p \\ \text{subject to} \quad & \sigma_d^2 \text{tr}(\tilde{\mathbf{F}}_r(\gamma_r \mathbf{H}_{sr}\mathbf{H}_{sr}^H + \mathbf{I})\tilde{\mathbf{F}}_r^H) \le p, \\ & \text{MSE}(\tilde{\mathbf{F}}_r, \mathbf{\Delta}_{rd}) \le \tau_{\text{MSE}}, \\ & \forall \mathbf{\Delta}_{rd} : \|\mathbf{\Delta}_{rd}\|_2 \le \epsilon_{rd}. \end{aligned} \tag{33}$$

The above two problems are complementary to problems (12) and (26), respectively. For instance, the problem (32) can be solved by searching the relay transmit power P_r such that the optimal value of the problem (12) equals to the threshold τ_{MI}. This can be implemented via a simple bi-section search over P_r and the same approach also applies to the problem (33).

We compare the minimum relay transmit power required to satisfy different QoS thresholds with respect to MI in Fig. 8. It can be found that the robust scheme RMI requires less

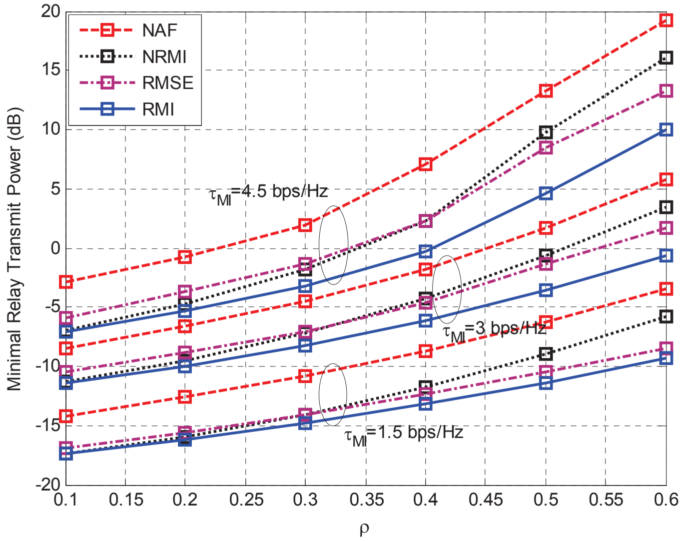

Fig. 8. Minimal relay transmit power versus ρ ($N_s = N_r = N_d = 4$, σ_r^2 = 0.0025, σ_d^2 = 0.025).

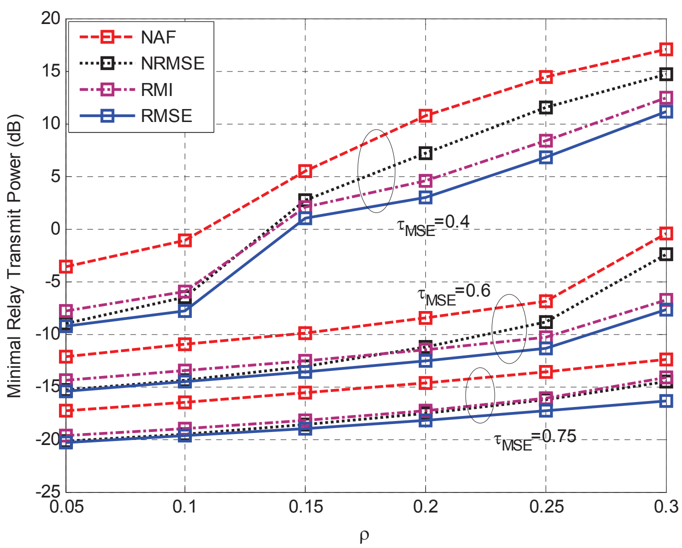

Fig. 9. Minimal relay transmit power versus ρ ($N_s = N_r = N_d = 4$, σ_r^2 = 0.0025, σ_d^2 = 0.025).

power than other approaches under various MI thresholds and uncertainty sizes. In addition, the RMSE method outperforms NRMI and NAF schemes when ρ gets large since channel uncertainties are considered in this strategy. Similar observations can be made in Fig. 9 where the QoS metric is the MSE. Therefore, our proposed robust designs can provide better power efficiency with a prescribed QoS constraint involving bounded CSI errors.

C. Performance Evaluation with Random CSI Errors

In this part, we present numerical results for random CSI errors instead of the worst-case errors considered before. In our test, random CSI error matrices are generated according to an i.i.d. Gaussian distribution and restricted to lying in a spectral norm bounded region. We compare robust and non-

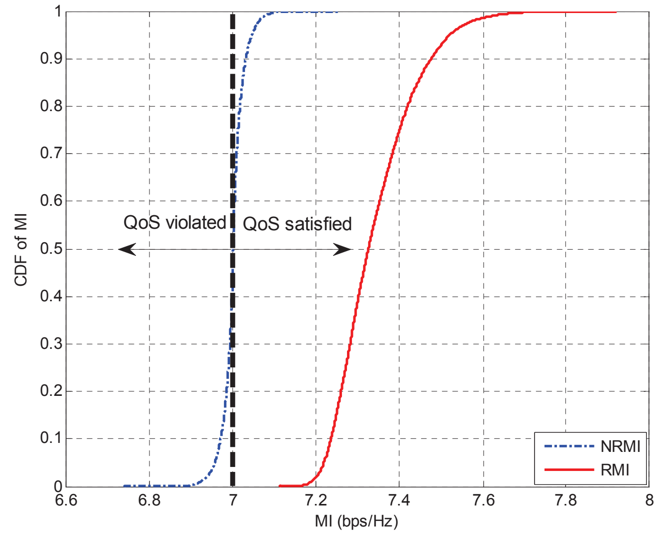

Fig. 10. Comparison of non-robust and robust methods for MIMO relaying with an MI QoS constraint (MI threshold = 7 bps/Hz, $N_s = N_r = N_d = 4$, σ_r^2 = 0.0079, σ_d^2 = 0.025, $\rho = 0.05$).

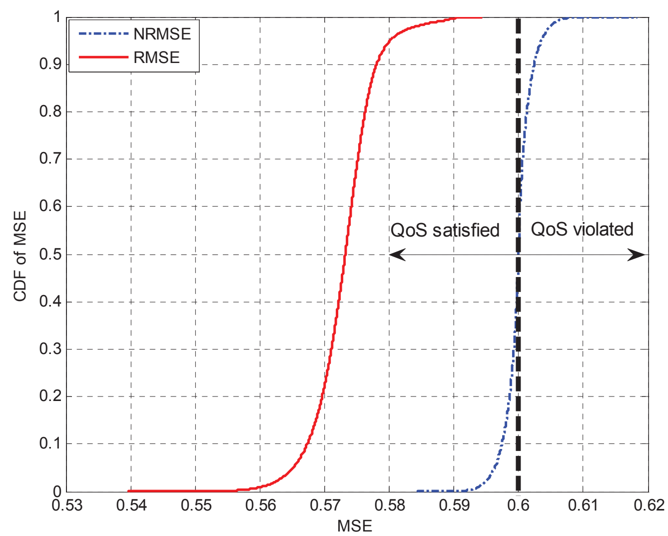

Fig. 11. Comparison of non-robust and robust methods for MIMO relaying with an MSE QoS constraint (MSE threshold = 0.6, $N_s = N_r = N_d = 4$, σ_r^2 = 0.0079, σ_d^2 = 0.025, $\rho = 0.05$).

robust designs in MIMO relay systems with an MI or MSE QoS constraint. Fig. 10 shows the cumulative distribution of MI using robust and non-robust design methods, where we can find that for the non-robust design, the MI QoS constraint cannot be satisfied for almost 50% channel realizations. This is due to the fact that the non-robust design does not incorporate channel uncertainties in the precoder optimization. On the other hand, the robust design can guarantee that all MI values are beyond the threshold. Note that a similar conclusion also applies to the case with an MSE QoS constraint, as shown in Fig. 11. Therefore, in presence of randomly generated CSI errors, the robust design outperforms the non-robust one in terms of the feasibility of QoS constraint.

VI. Conclusions

We studied optimal worst-case robust designs for AF MIMO relay systems adopting either MI or MSE as the design objective. Although the formulated optimization problems do not have a conventional concave-convex or convex-concave structure, we have derived their optimal solutions in closed form. Specifically, we proved that eigenmode transmission is the best strategy even with deterministic but bounded CSI errors, which is consistent with the results under perfect and stochastic CSI assumptions. Thus in practice, the same eigenmode architecture can be utilized regardless of CSI knowledge. The available information about CSI errors affects the optimal precoder only through the diagonal power allocation design. The proposed robust designs can enhance the transmission efficiency and reliability of AF MIMO relaying with norm-bounded CSI uncertainties without increasing computational requirements beyond those needed for conventional non-robust schemes.

The work presented in this paper can be extended in several directions. It would be of interest to study worst-case robust relay transceiver optimization with a direct line of sight, which is a non-trivial extension of the current work. In addition, whereas we assumed here that perfect CSI knowledge is available at the receiver, it will be both meaningful and challenging to analyze the effect of imperfect CSIR in the robust design.

Appendix I
A Brief Introduction to Majorization Theory

We herein introduce some basic definitions and results of majorization theory that are necessary for this paper. Interested readers are referred to [39] for a comprehensive presentation of this subject.

Definition 1 ([39, Chapter 1, Definition A.1]): For $\mathbf{x},\mathbf{y}\in\mathbb{R}^N$, $\mathbf{x}$ is said to be majorized by $\mathbf{y}$ (denoted as $\mathbf{x}\prec\mathbf{y}$) if

$$\begin{aligned}\sum_{i=1}^{k} x_{[i]} &\le \sum_{i=1}^{k} y_{[i]},\ k=1,\cdots,N-1,\\ \sum_{i=1}^{N} x_{[i]} &= \sum_{i=1}^{N} y_{[i]},\end{aligned}\tag{34}$$

where $x_{[1]},\cdots,x_{[N]}$ (and $y_{[1]},\cdots,y_{[N]}$) are the components of $\mathbf{x}$ (and $\mathbf{y}$) arranged in decreasing order.

Definition 2 ([39, Chapter 1, Definition A.2]): For $\mathbf{x},\mathbf{y}\in\mathbb{R}^N$, $\mathbf{x}$ is said to be weakly majorized by $\mathbf{y}$ (denoted as $\mathbf{x}\prec_w\mathbf{y}$) if

$$\sum_{i=1}^{k} x_{[i]} \le \sum_{i=1}^{k} y_{[i]},\ k=1,\cdots,N.\tag{35}$$

Definition 3 ([39, Chapter 3, Definition A.1]): A real-valued function ϕ defined on a set $\mathcal{A}\subset\mathbb{R}^n$ is said to be Schur-convex on $\mathcal{A}$ if

$$\mathbf{x}\prec\mathbf{y}\text{ on }\mathcal{A}\Rightarrow\phi(\mathbf{x})\le\phi(\mathbf{y}).\tag{36}$$

In the following lemmas, we present two kinds of Schur-convex functions that will be used in the paper.

Lemma 3 ([39, Chapter 3, Proposition C.1]): If $I\subset\mathbb{R}$ is an interval and $g:I\to\mathbb{R}$ is convex, then

$$\phi(\mathbf{x})=\sum_{i=1}^{n} g(x_i),\ \mathbf{x}\in I^n\tag{37}$$

is Schur-convex on I^n.

Lemma 4 ([39, Chapter 3, Proposition E.1]): Let g be a continuous nonnegative function defined on an interval $I\subset\mathbb{R}$. Then

$$\phi(\mathbf{x})=\prod_{i=1}^{n} g(x_i),\ \mathbf{x}\in I^n\tag{38}$$

is Schur-convex on I^n if and only if $\log g$ is convex on I.

Appendix II
Proof of Proposition 1

Denote the EVD of $(\hat{\mathbf{H}}_{rd}+\mathbf{\Delta}_{rd})^H(\hat{\mathbf{H}}_{rd}+\mathbf{\Delta}_{rd})$ as $(\hat{\mathbf{H}}_{rd}+\mathbf{\Delta}_{rd})^H(\hat{\mathbf{H}}_{rd}+\mathbf{\Delta}_{rd})=\mathbf{V}_{rd}\mathbf{\Sigma}_{rd}\mathbf{V}_{rd}^H$, where $\mathbf{\Sigma}_{rd}=\text{diag}\{\sigma_{rd,1}^2,\cdots,\sigma_{rd,N_p}^2,0,\cdots,0\}$ with $N_p=\min\{N_r,N_d\}$. Then, the optimal solution to the inner maximization problem is [4]

$$\tilde{\mathbf{F}}_r^*=\mathbf{V}_{rd}\mathbf{\Lambda}_{\tilde{f}_r^*}(\mathbf{I}+\gamma_r\mathbf{\Sigma}_{sr})^{-1/2}\mathbf{U}_{sr}^H,\tag{39}$$

where $\mathbf{\Lambda}_{\tilde{f}_r^*}=\text{diag}\{\tilde{f}_{r,1}^*,\cdots,\tilde{f}_{r,N_s}^*,0,\cdots,0\}$ and its ith diagonal element $\tilde{f}_{r,i}^*$ is determined by

$$\tilde{f}_{r,i}^*=\begin{cases}\sqrt{\frac{\left(\sqrt{\gamma_r^2\sigma_{sr,i}^4+4\mu\gamma_r\sigma_{sr,i}^2\sigma_{rd,i}^2}-\gamma_r\sigma_{sr,i}^2-2\right)_+}{2\sigma_{rd,i}^2}}, & \sigma_{rd,i}>0\\ 0, & \sigma_{rd,i}=0,\end{cases}\tag{40}$$

where $\mu>0$ is chosen such that $\sum_{i=1}^{N_s}(\tilde{f}_{r,i}^*)^2=\gamma_d N_r$ is satisfied. In addition, the maximum value of the inner problem is obtained as

$$\frac{1}{2}\log_2|\mathbf{I}+\gamma_r\mathbf{\Sigma}_{sr}|+\frac{1}{2}\log_2\frac{\left|\mathbf{I}+\mathbf{\Lambda}_{\tilde{f}_r^*}\mathbf{\Sigma}_{rd}\mathbf{\Lambda}_{\tilde{f}_r^*}\right|}{\left|\mathbf{I}+\gamma_r\mathbf{\Sigma}_{sr}+\mathbf{\Lambda}_{\tilde{f}_r^*}\mathbf{\Sigma}_{rd}\mathbf{\Lambda}_{\tilde{f}_r^*}\right|}.\tag{41}$$

Note that the first term of (41) is a constant and the second term can be expressed as (42). The inequality holds due to the facts that $\sigma_i(\mathbf{A}+\mathbf{B})\ge(\sigma_i(\mathbf{A})-\sigma_1(\mathbf{B}))_+$ and $\sigma_1(\mathbf{\Delta}_{rd})\le\epsilon_{rd}$, where $\sigma_1(\cdot)$ denotes the largest singular value of a matrix. We can achieve the lower bound in the above expression by selecting the worst-case CSI error $\mathbf{\Delta}_{rd}^w$ as indicated in (14) and letting the relay precoder be $\tilde{\mathbf{F}}_r^{opt}$ in (15). Therefore, we conclude that $(\mathbf{\Delta}_{rd}^w,\tilde{\mathbf{F}}_r^{opt})$ is the optimal solution to the minimax problem (13) and its optimal value is (17).

Appendix III
Proof of Lemma 1

We first construct a column vector $\mathbf{d}(\mathbf{W})$ whose entries are the diagonal elements of matrix $\mathbf{W}$ arranged in decreasing order, i.e., $d_1(\mathbf{W})\ge\cdots\ge d_N(\mathbf{W})$. Concerning the first entry of $\mathbf{d}(\mathbf{W})$, i.e., $d_1(\mathbf{W})$, we have $d_1(\mathbf{W})\ge d_{i_1}(\mathbf{W})\ge d'_{i_1}(\mathbf{W})$ since $i_1\ge1$ and $d_i(\mathbf{W})\ge d'_i(\mathbf{W})$. Similarly, for the last entry of $\mathbf{d}(\mathbf{W})$, i.e., $d_N(\mathbf{W})$, it can be readily shown that

$$
\begin{aligned}
&\frac{1}{2}\log_2 \frac{\left|\mathbf{I}+\mathbf{\Lambda}_{\tilde{f}_r^*}\mathbf{\Sigma}_{rd}\mathbf{\Lambda}_{\tilde{f}_r^*}\right|}{\left|\mathbf{I}+\gamma_r\mathbf{\Sigma}_{sr}+\mathbf{\Lambda}_{\tilde{f}_r^*}\mathbf{\Sigma}_{rd}\mathbf{\Lambda}_{\tilde{f}_r^*}\right|}\\
=&\frac{1}{2}\log_2 \frac{\prod_{i=1}^{N_s}\left(1+(\tilde{f}_{r,i}^*)^2\sigma_{rd,i}^2\right)}{\prod_{i=1}^{N_s}\left(1+\gamma_r\sigma_{sr,i}^2+(\tilde{f}_{r,i}^*)^2\sigma_{rd,i}^2\right)}\\
=&\frac{1}{2}\sum_{i=1}^{N_s}\log_2\left(1-\frac{\gamma_r\sigma_{sr,i}^2}{1+\gamma_r\sigma_{sr,i}^2+(\tilde{f}_{r,i}^*)^2\sigma_{rd,i}^2}\right)\\
=&\frac{1}{2}\sum_{i=1}^{N_s}\log_2\left(1-\frac{\gamma_r\sigma_{sr,i}^2}{1+\gamma_r\sigma_{sr,i}^2+\frac{1}{2}\left(\sqrt{\gamma_r^2\sigma_{sr,i}^4+4\mu\gamma_r\sigma_{sr,i}^2\sigma_{rd,i}^2}-\gamma_r\sigma_{sr,i}^2-2\right)_+}\right)\\
\geq&\frac{1}{2}\sum_{i=1}^{N_s}\log_2\left(1-\frac{\gamma_r\sigma_{sr,i}^2}{1+\gamma_r\sigma_{sr,i}^2+\frac{1}{2}\left(\sqrt{\gamma_r^2\sigma_{sr,i}^4+4\mu\gamma_r\sigma_{sr,i}^2(\hat{\sigma}_{rd,i}-\epsilon_{rd})_+^2}-\gamma_r\sigma_{sr,i}^2-2\right)_+}\right)
\end{aligned}
\tag{42}
$$

$d_N(\mathbf{W}) \geq d_N^{'}(\mathbf{W}) \geq d_{i_N}^{'}(\mathbf{W})$. Now let us compare $d_p(\mathbf{W})$ with $d_{i_p}^{'}(\mathbf{W})$ where $1 < p < N$. We consider the following three cases:

a) $i_p > p$. For this case, we can readily verify that $d_p(\mathbf{W}) \geq d_{i_p}(\mathbf{W}) \geq d_{i_p}^{'}(\mathbf{W})$ holds.

b) $i_p \leq p$ and $d_{i_p}^{'}(\mathbf{W}) \leq d_p^{'}(\mathbf{W})$. In this context, it is easy to show that $d_p(\mathbf{W}) \geq d_p^{'}(\mathbf{W}) \geq d_{i_p}^{'}(\mathbf{W})$.

c) $i_p \leq p$ and $d_{i_p}^{'}(\mathbf{W}) > d_p^{'}(\mathbf{W})$. For this case, there must exist a $t > 0$ such that $d_{p+t}^{'}(\mathbf{W}) \geq d_{i_p}^{'}(\mathbf{W})$, and hence we have $d_p(\mathbf{W}) \geq d_{p+t}(\mathbf{W}) \geq d_{p+t}^{'}(\mathbf{W}) \geq d_{i_p}^{'}(\mathbf{W})$.

Therefore, we conclude that $d_k(\mathbf{W}) \geq d_{i_k}^{'}(\mathbf{W}), \forall 1 \leq k \leq N$. Considering that $\boldsymbol{\lambda}(\mathbf{W}) \succ \mathbf{d}(\mathbf{W})$ holds for any Hermitian matrix $\mathbf{W}$ [39, Chapter 9, Theorem B.1], we eventually arrive at the conclusion that $\boldsymbol{\lambda}(\mathbf{W}) \succ_w \mathbf{d}^{'}(\mathbf{W})$.

Appendix IV
Proof of Proposition 2

As shown in [6], the optimal solution to the inner minimization problem is given by

$$\tilde{\mathbf{F}}_r^{'} = \mathbf{V}_{rd}\mathbf{\Lambda}_{\tilde{f}_r^{'}}(\mathbf{I}+\gamma_r\mathbf{\Sigma}_{sr})^{-1/2}\mathbf{U}_{sr}^H, \tag{43}$$

where the diagonal elements of $\mathbf{\Lambda}_{\tilde{f}_r^{'}} = \text{diag}\{\tilde{f}_{r,1}^{'}, \cdots, \tilde{f}_{r,N_s}^{'}, 0, \cdots, 0\}$ can be expressed as

$$\tilde{f}_{r,i}^{'} = \begin{cases} \sqrt{\frac{\left(\sqrt{\frac{\gamma_r\sigma_{sr,i}^2\sigma_{rd,i}^2}{\nu(1+\gamma_r\sigma_{sr,i}^2)}}-1\right)_+}{\sigma_{rd,i}^2}}, & \sigma_{rd,i} > 0\\ 0, & \sigma_{rd,i} = 0, \end{cases} \tag{44}$$

with $\nu > 0$ chosen such that $\sum_{i=1}^{N_s}(\tilde{f}_{r,i}^{'})^2 = \gamma_d N_r$ holds. In addition, the optimal value takes the form

$$\frac{P_s}{N_s}\sum_{i=1}^{N_s}\frac{1+\gamma_r\sigma_{sr,i}^2+(\tilde{f}_{r,i}^{'})^2\sigma_{rd,i}^2}{(1+\gamma_r\sigma_{sr,i}^2)(1+(\tilde{f}_{r,i}^{'})^2\sigma_{rd,i}^2)}. \tag{45}$$

With techniques similar to those used to derive (42), we can show that (45) is upper bounded by

$$\frac{P_s}{N_s}\sum_{i=1}^{N_s}\frac{1+\gamma_r\sigma_{sr,i}^2+\left(\sqrt{\frac{\gamma_r\sigma_{sr,i}^2(\hat{\sigma}_{rd,i}-\epsilon_{rd})_+^2}{\nu(1+\gamma_r\sigma_{sr,i}^2)}}-1\right)_+}{(1+\gamma_r\sigma_{sr,i}^2)\left(1+\left(\sqrt{\frac{\gamma_r\sigma_{sr,i}^2(\hat{\sigma}_{rd,i}-\epsilon_{rd})_+^2}{\nu(1+\gamma_r\sigma_{sr,i}^2)}}-1\right)_+\right)}, \tag{46}$$

which can be achieved by selecting the worst-case CSI perturbation as indicated in (28) and letting the relay precoder be (29). Therefore, we arrive at the conclusion that $(\mathbf{\Delta}_{rd}^w, \tilde{\mathbf{F}}_r^{opt})$ is the optimal solution to the maximin problem (27) whose optimal value is given by (31).

Appendix V
Proof of Theorem 2

By fixing $\mathbf{\Delta}_{rd}$ in $\text{MSE}(\tilde{\mathbf{F}}_r^{opt}, \mathbf{\Delta}_{rd})$ with $\mathbf{\Delta}_{rd}^w$, it is easy to verify that $\text{MSE}(\tilde{\mathbf{F}}_r^{opt}, \mathbf{\Delta}_{rd}^w) \leq \text{MSE}(\tilde{\mathbf{F}}_r, \mathbf{\Delta}_{rd}^w)$ holds based on the results in [6]. However, verifying $\text{MSE}(\tilde{\mathbf{F}}_r^{opt}, \mathbf{\Delta}_{rd}^w) \geq \text{MSE}(\tilde{\mathbf{F}}_r^{opt}, \mathbf{\Delta}_{rd})$ requires proving the following trace inequality

$$
\begin{aligned}
&\text{tr}\Bigg(\left(\mathbf{I}+(\hat{\mathbf{\Lambda}}_{rd}+\tilde{\mathbf{\Delta}}_{rd})(\mathbf{\Lambda}_{\tilde{f}_r}^{opt})^2(\hat{\mathbf{\Lambda}}_{rd}+\tilde{\mathbf{\Delta}}_{rd})^H\right)^{-1}\\
&\times\left(\mathbf{I}+(\hat{\mathbf{\Lambda}}_{rd}+\tilde{\mathbf{\Delta}}_{rd})(\mathbf{\Lambda}_{\tilde{f}_r}^{opt})^2(\mathbf{I}+\gamma_r\mathbf{\Sigma}_{sr})^{-1}\big(\hat{\mathbf{\Lambda}}_{rd}\right.\\
&\left.+\tilde{\mathbf{\Delta}}_{rd}\big)^H\right)\Bigg)\\
\leq&\sum_{i=1}^{N_s}\frac{1+\gamma_r\sigma_{sr,i}^2+(\tilde{f}_{r,i}^{opt})^2(\hat{\sigma}_{rd,i}-\epsilon_{rd})_+^2}{(1+\gamma_r\sigma_{sr,i}^2)(1+(\tilde{f}_{r,i}^{opt})^2(\hat{\sigma}_{rd,i}-\epsilon_{rd})_+^2)}.
\end{aligned}
\tag{47}
$$

Note that the above inequality is quite different from (21) in the proof of *Theorem* 1. Before proceeding, we introduce the following lemma that will be used to simplify the left-hand side of (47).

$$
\begin{aligned}
&N_d - N_r + \text{tr}\left(\left(\mathbf{I} + \boldsymbol{\Lambda}_{\tilde{f}_r}^{opt}(\hat{\boldsymbol{\Lambda}}_{rd} + \tilde{\boldsymbol{\Delta}}_{rd})^H(\hat{\boldsymbol{\Lambda}}_{rd} + \tilde{\boldsymbol{\Delta}}_{rd})\boldsymbol{\Lambda}_{\tilde{f}_r}^{opt}\right)^{-1}\right) + \text{tr}\bigg((\hat{\boldsymbol{\Lambda}}_{rd} + \tilde{\boldsymbol{\Delta}}_{rd})\boldsymbol{\Lambda}_{\tilde{f}_r}^{opt} \\
&\quad \times \left(\mathbf{I} + \boldsymbol{\Lambda}_{\tilde{f}_r}^{opt}(\hat{\boldsymbol{\Lambda}}_{rd} + \tilde{\boldsymbol{\Delta}}_{rd})^H(\hat{\boldsymbol{\Lambda}}_{rd} + \tilde{\boldsymbol{\Delta}}_{rd})\boldsymbol{\Lambda}_{\tilde{f}_r}^{opt}\right)^{-1}\boldsymbol{\Lambda}_{\tilde{f}_r}^{opt}(\mathbf{I} + \gamma_r\boldsymbol{\Sigma}_{sr})^{-1}(\hat{\boldsymbol{\Lambda}}_{rd} + \tilde{\boldsymbol{\Delta}}_{rd})^H\bigg) \\
=&N_d - N_r + \text{tr}\left(\left(\mathbf{I} + \boldsymbol{\Lambda}_{\tilde{f}_r}^{opt}(\hat{\boldsymbol{\Lambda}}_{rd} + \tilde{\boldsymbol{\Delta}}_{rd})^H(\hat{\boldsymbol{\Lambda}}_{rd} + \tilde{\boldsymbol{\Delta}}_{rd})\boldsymbol{\Lambda}_{\tilde{f}_r}^{opt}\right)^{-1}\right) + \text{tr}\bigg((\mathbf{I} + \gamma_r\boldsymbol{\Sigma}_{sr})^{-1}\boldsymbol{\Lambda}_{\tilde{f}_r}^{opt} \\
&\quad \times (\hat{\boldsymbol{\Lambda}}_{rd} + \tilde{\boldsymbol{\Delta}}_{rd})^H(\hat{\boldsymbol{\Lambda}}_{rd} + \tilde{\boldsymbol{\Delta}}_{rd})\boldsymbol{\Lambda}_{\tilde{f}_r}^{opt}\left(\mathbf{I} + \boldsymbol{\Lambda}_{\tilde{f}_r}^{opt}(\hat{\boldsymbol{\Lambda}}_{rd} + \tilde{\boldsymbol{\Delta}}_{rd})^H(\hat{\boldsymbol{\Lambda}}_{rd} + \tilde{\boldsymbol{\Delta}}_{rd})\boldsymbol{\Lambda}_{\tilde{f}_r}^{opt}\right)^{-1}\bigg) \\
=&N_d - N_r + \text{tr}\bigg(\left(\mathbf{I} + (\mathbf{I} + \gamma_r\boldsymbol{\Sigma}_{sr})^{-1}\boldsymbol{\Lambda}_{\tilde{f}_r}^{opt}(\hat{\boldsymbol{\Lambda}}_{rd} + \tilde{\boldsymbol{\Delta}}_{rd})^H(\hat{\boldsymbol{\Lambda}}_{rd} + \tilde{\boldsymbol{\Delta}}_{rd})\boldsymbol{\Lambda}_{\tilde{f}_r}^{opt}\right)\bigg(\mathbf{I} + \boldsymbol{\Lambda}_{\tilde{f}_r}^{opt} \\
&\quad \times (\hat{\boldsymbol{\Lambda}}_{rd} + \tilde{\boldsymbol{\Delta}}_{rd})^H(\hat{\boldsymbol{\Lambda}}_{rd} + \tilde{\boldsymbol{\Delta}}_{rd})\boldsymbol{\Lambda}_{\tilde{f}_r}^{opt}\bigg)^{-1}\bigg) \\
=&N_d - N_r + \text{tr}\left((\mathbf{I} + \gamma_r\boldsymbol{\Sigma}_{sr})^{-1}\left(\mathbf{I} + \gamma_r\boldsymbol{\Sigma}_{sr}\left(\mathbf{I} + \boldsymbol{\Lambda}_{\tilde{f}_r}^{opt}(\hat{\boldsymbol{\Lambda}}_{rd} + \tilde{\boldsymbol{\Delta}}_{rd})^H(\hat{\boldsymbol{\Lambda}}_{rd} + \tilde{\boldsymbol{\Delta}}_{rd})\boldsymbol{\Lambda}_{\tilde{f}_r}^{opt}\right)^{-1}\right)\right) \\
=&C + \gamma_r\text{tr}\left(\boldsymbol{\Sigma}_{sr}(\mathbf{I} + \gamma_r\boldsymbol{\Sigma}_{sr})^{-1/2}\left(\mathbf{I} + \boldsymbol{\Lambda}_{\tilde{f}_r}^{opt}(\hat{\boldsymbol{\Lambda}}_{rd} + \tilde{\boldsymbol{\Delta}}_{rd})^H(\hat{\boldsymbol{\Lambda}}_{rd} + \tilde{\boldsymbol{\Delta}}_{rd})\boldsymbol{\Lambda}_{\tilde{f}_r}^{opt}\right)^{-1}(\mathbf{I} + \gamma_r\boldsymbol{\Sigma}_{sr})^{-1/2}\right) \\
=&C + \gamma_r\text{tr}\left(\boldsymbol{\Sigma}_{sr}\left((\mathbf{I} + \gamma_r\boldsymbol{\Sigma}_{sr}) + (\mathbf{I} + \gamma_r\boldsymbol{\Sigma}_{sr})^{1/2}\boldsymbol{\Lambda}_{\tilde{f}_r}^{opt}(\hat{\boldsymbol{\Lambda}}_{rd} + \tilde{\boldsymbol{\Delta}}_{rd})^H(\hat{\boldsymbol{\Lambda}}_{rd} + \tilde{\boldsymbol{\Delta}}_{rd})\boldsymbol{\Lambda}_{\tilde{f}_r}^{opt}(\mathbf{I} + \gamma_r\boldsymbol{\Sigma}_{sr})^{1/2}\right)^{-1}\right) \quad (48)
\end{aligned}
$$

Lemma 5: For any matrix $\mathbf{P} \in \mathbb{C}^{M\times N}$, the trace equality $\text{tr}((\mathbf{I}+\mathbf{P}^H\mathbf{P})^{-1}) = \text{tr}((\mathbf{I}+\mathbf{P}\mathbf{P}^H)^{-1}) + N - M$ holds.

Proof: Denote the SVD of $\mathbf{P}$ with $\mathbf{P} = \mathbf{U}_p\boldsymbol{\Sigma}_p\mathbf{V}_p^H$, then we have $\text{tr}((\mathbf{I}+\mathbf{P}^H\mathbf{P})^{-1}) = \text{tr}((\mathbf{I}+\mathbf{V}_p\boldsymbol{\Sigma}_p^H\boldsymbol{\Sigma}_p\mathbf{V}_p^H)^{-1}) = \text{tr}((\mathbf{I}+\boldsymbol{\Sigma}_p^H\boldsymbol{\Sigma}_p)^{-1})$ and $\text{tr}((\mathbf{I}+\mathbf{P}\mathbf{P}^H)^{-1}) = \text{tr}((\mathbf{I}+\boldsymbol{\Sigma}_p\boldsymbol{\Sigma}_p^H)^{-1})$. Let us assume $M < N$ temporarily, then $\boldsymbol{\Sigma}_p$ can be expressed as $\boldsymbol{\Sigma}_p = [\boldsymbol{\Sigma}_p^{'}\ \mathbf{0}_{M\times(N-M)}]$, where $\boldsymbol{\Sigma}_p^{'}$ is an $M\times M$ real diagonal matrix. Accordingly, $\text{tr}((\mathbf{I}+\mathbf{P}^H\mathbf{P})^{-1}) = \text{tr}((\mathbf{I}+(\boldsymbol{\Sigma}_p^{'})^2)^{-1}) + N - M$ and $\text{tr}((\mathbf{I}+\mathbf{P}\mathbf{P}^H)^{-1}) = \text{tr}((\mathbf{I}+(\boldsymbol{\Sigma}_p^{'})^2)^{-1})$. Therefore, it is true that $\text{tr}((\mathbf{I}+\mathbf{P}^H\mathbf{P})^{-1}) = \text{tr}((\mathbf{I}+\mathbf{P}\mathbf{P}^H)^{-1}) + N - M$. The proof for the case of $M \geq N$ is similar and hence is omitted for brevity. ■

By using *Lemma* 5 and the equality $(\mathbf{P}^H\mathbf{P}+\mathbf{I})^{-1}\mathbf{P}^H = \mathbf{P}^H(\mathbf{P}\mathbf{P}^H+\mathbf{I})^{-1}$, we rewrite the left-hand side of (47) by (48) on the top of next page, where $C = \text{tr}\,(\mathbf{I}+\gamma_r\boldsymbol{\Sigma}_{sr})^{-1} + N_d - N_r = \sum_{i=1}^{N_s}\frac{1}{1+\gamma_r\sigma_{sr,i}^2} + N_d - N_s$.

Let us assume that the first N_l diagonal elements of matrix $\boldsymbol{\Sigma}_{sr}$ are non-zero. Then, in accordance with (30), $\tilde{f}_{r,i}^{opt} = 0, i = N_l+1, \cdots, N_r$. So we can rewrite (48) as

$$
\begin{aligned}
&C + \text{tr}\bigg(\left(\mathbf{I} + \tilde{\boldsymbol{\Sigma}}_{sr}^{-1} + \tilde{\boldsymbol{\Sigma}}_{sr}^{-1/2}\left(\mathbf{I} + \tilde{\boldsymbol{\Sigma}}_{sr}\right)^{1/2}\boldsymbol{\Upsilon}\left(\mathbf{I} + \tilde{\boldsymbol{\Sigma}}_{sr}\right)^{1/2} \\
&\times \tilde{\boldsymbol{\Sigma}}_{sr}^{-1/2}\right)^{-1}\bigg) = C + \sum_{i=1}^{N_l}\lambda_i^{-1}(\boldsymbol{\Xi}),
\end{aligned}
$$

where $\tilde{\boldsymbol{\Sigma}}_{sr} = \text{diag}\{\tilde{\sigma}_{sr,1}^2, \cdots, \tilde{\sigma}_{sr,N_l}^2\}$ denotes the first N_l rows and columns of matrix $\gamma_r\boldsymbol{\Sigma}_{sr}$, $\boldsymbol{\Upsilon}$ denotes the first N_l rows and columns of matrix $\boldsymbol{\Lambda}_{\tilde{f}_r}^{opt}(\hat{\boldsymbol{\Lambda}}_{rd} + \tilde{\boldsymbol{\Delta}}_{rd})^H(\hat{\boldsymbol{\Lambda}}_{rd} + \tilde{\boldsymbol{\Delta}}_{rd})\boldsymbol{\Lambda}_{\tilde{f}_r}^{opt}$ and $\boldsymbol{\Xi} = \mathbf{I} + \tilde{\boldsymbol{\Sigma}}_{sr}^{-1} + \tilde{\boldsymbol{\Sigma}}_{sr}^{-1/2}\left(\mathbf{I}+\tilde{\boldsymbol{\Sigma}}_{sr}\right)^{1/2}\boldsymbol{\Upsilon}\left(\mathbf{I}+\tilde{\boldsymbol{\Sigma}}_{sr}\right)^{1/2}\tilde{\boldsymbol{\Sigma}}_{sr}^{-1/2}$. As in (24), we can show $d_i(\boldsymbol{\Xi}) \geq 1 + \tilde{\sigma}_{sr,i}^{-2} + (1+\tilde{\sigma}_{sr,i}^{-2})(\tilde{f}_{r,i}^{opt})^2\left(\hat{\sigma}_{rd,i} - \epsilon_{rd}\right)_+^2 \triangleq d_i^{'}(\boldsymbol{\Xi})$. Then, by using *Lemma* 1, we obtain that $\boldsymbol{\lambda}(\boldsymbol{\Xi}) \succ_w \mathbf{d}^{'}(\boldsymbol{\Xi})$. Moreover, as $g(\mathbf{x}) = \sum_{i=1}^{N_l}\frac{1}{x_i},\ x_i > 0$ is a Schur-convex function (according to *Lemma* 3 in Appendix I), we can readily verify (47) using *Lemma* 2. So at this point, we have proved that $(\tilde{\mathbf{F}}_r^{opt}, \boldsymbol{\Delta}_{rd}^w)$ is a saddle point of $\text{MSE}(\tilde{\mathbf{F}}_r, \boldsymbol{\Delta}_{rd})$ and hence optimal for both minimax problem (26) and maximin problem (27).

ACKNOWLEDGEMENT

The authors would like to thank Prof. Zhi Ding for very helpful discussions on the proof in this paper. The authors also appreciate Dr. Wei Xu and Xiao Liang for their insightful views and suggestions.

REFERENCES

[1] S. W. Peters, A. Y. Panah, K. T. Truong, and R. W. Heath, Jr., "Relay architectures for 3GPP LTE-advanced," *EURASIP J. Wireless Commun. Netw.*, vol. 2009, Article ID 618787, 14 pages, May 2009.

[2] 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.

[3] S. Jin, M. R. McKay, C. Zhong, and K.-K. Wong, "Ergodic capacity analysis of amplify-and-forward MIMO dual-hop systems," *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2204–2224, May 2010.

[4] 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.

[5] O. Muñoz-Medina, J. Vidal, and A. Agustín, "Linear transceiver design in nonregenerative relays with channel state information," *IEEE Trans. Signal Process.*, vol. 55, no. 6, pp. 2593–2604, Jun. 2007.

[6] W. Guan and H. Luo, "Joint MMSE transceiver design in non-regenerative MIMO relay systems," *IEEE Commun. Lett.*, vol. 12, no. 7, pp. 517–519, Jul. 2008.

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

[8] Y. Rong, X. Tang, and Y. Hua, "A unified framework for optimizing linear nonregenerative multicarrier MIMO relay communication systems," *IEEE Trans. Signal Process.*, vol. 57, no. 12, pp. 4837–4851, Dec. 2009.

[9] D. P. Palomar, J. M. Cioffi, and M. Á. Lagunas, "Joint Tx-Rx beamforming design for multicarrier MIMO channels: A unified framework for convex optimization," *IEEE Trans. Signal Process.*, vol. 51, pp. 2381–2401, Sep. 2003.

[10] T. Yoo and A. Goldsmith, "Capacity and power allocation for fading MIMO channels with channel estimation error," *IEEE Trans. Inform. Theory*, vol. 52, no. 5, pp. 2203–2214, May 2006.

[11] X. Zhang, D. P. Palomar, and B. Ottersten, "Statistically robust design of linear MIMO transceivers," *IEEE Trans. Signal Process.*, vol. 56, no. 8, pp. 3678–3689, Aug. 2008.

[12] 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.

[13] C. Xing, S. Ma, Y.-C. Wu, and T.-S. Ng, "Transceiver design for dual-hop non-regenerative MIMO-OFDM relay systems under channel uncertainties," *IEEE Trans. Signal Process.*, vol. 58, no. 12, pp. 6325–6339, Dec. 2010.

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

[15] C. Jeong, B. Seo, S. R. Lee, H.-M. Kim, and I.-M. Kim, "Relay precoding for non-regenerative MIMO relay systems with partial CSI feedback," *IEEE Trans. Wireless Commun.*, vol. 11, no. 5, pp. 1698–1711, May 2012.

[16] W. Xu and X. Dong, "Optimized one-way relaying strategy with outdated CSI quantization for spatial multiplexing," *IEEE Trans. Signal Process.*, vol. 60, no. 8, pp. 4458–4464, Aug. 2012.

[17] 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. 2918–2930, Aug. 2005.

[18] A. Pascual-Iserte, D. P. Palomar, A. I. Pérez-Neira, and M. Á. Lagunas, "A robust maximin approach for MIMO communications with imperfect channel state information based on convex optimization," *IEEE Trans. Signal Process.*, vol. 54, no. 1, pp. 346–360, Jan. 2006.

[19] J. Wang and M. Payaró, "On the robustness of transmit beamforming," *IEEE Trans. Signal Process.*, vol. 58, no. 11, pp. 5933–5938, Nov. 2010.

[20] S. Loyka and C. D. Charalambous, "On the compound capacity of a class of MIMO channels subject to normed uncertainty," *IEEE Trans. Inf. Theory*, vol. 58, no. 4, pp. 2048–2063, Apr. 2012.

[21] Y. Sun, C. E. Koksal, and N. B. Shroff, "Capacity of compound MIMO channels with additive uncertainty," arXiv:1208.4656v3 [cs.IT], Aug. 2012.

[22] G. Zheng, K.-K. Wong, A. Paulraj, and B. Ottersten, "Robust collaborative-relay beamforming," *IEEE Trans. Signal Process.*, vol. 57, no. 8, pp. 3130–3143, Aug. 2009.

[23] B. K. Chalise and L. Vandendorpe, "Optimization of MIMO relays for multipoint-to-multipoint communications: Nonrobust and robust designs," *IEEE Trans. Signal Process.*, vol. 58, no. 12, pp. 6355–6368, Dec. 2010.

[24] P. Ubaidulla and A. Chockalingam, "Relay precoder optimization in MIMO relay networks with imperfect CSI," *IEEE Trans. Signal Process.*, vol. 59, no. 11, pp. 5473–5484, Nov. 2011.

[25] M. Tao and R. Wang, "Robust relay beamforming for two-way relay networks," *IEEE Commun. Lett.*, vol. 16, no. 7, pp. 1052–1055, Jul. 2012.

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

[27] Z. Wang, W. Chen, F. Gao, and J. Li, "Capacity performance of relay beamformings for MIMO multirelay networks with imperfect R-D CSI at relays," *IEEE Trans. Veh. Technol.*, vol. 60, no. 6, pp. 2608–2619, Jul. 2011.

[28] M. Payaró, A. Pascual-Iserte, and M. Á. Lagunas, "Robust power allocation designs for multiuser and multiantenna downlink communication systems through convex optimization," *IEEE J. Sel. Areas Commun.*, vol. 25, no. 7, pp. 1390–1401, Sep. 2007.

[29] M. B. Shenouda and T. N. Davidson, "On the design of linear transceivers for multiuser systems with channel uncertainty," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 6, pp. 1015–1024, Aug. 2008.

[30] N. Vučić, H. Boche, and S. Shi, "Robust transceiver optimization in downlink multiuser MIMO systems," *IEEE Trans. Signal Process.*, vol. 57, no. 9, pp. 3576–3587, Sep. 2009.

[31] G. Zheng, K.-K. Wong, and B. Ottersten, "Robust cognitive beamforming with bounded channel uncertainties," *IEEE Trans. Signal Process.*, vol. 57, no. 12, pp. 4871–4881, Dec. 2009.

[32] Y. Huang, Q. Li, W.-K. Ma, and S. Zhang, "Robust multicast beamforming for spectrum sharing-based cognitive radios," *IEEE Trans. Signal Process.*, vol. 60, no. 1, pp. 527–533, Jan. 2012.

[33] J. Zhang, R. W. Heath, Jr., M. Kountouris, and J. G. Andrews, "Mode switching for the multi-antenna broadcast channel based on delay and channel quantization," *EURASIP J. Advances Signal Process.*, vol. 2009, Article ID 802548, 15 pages, Feb. 2009.

[34] J. Zhang, M. Kountouris, J. G. Andrews, and R. W. Heath, Jr., "Multi-mode transmission for the MIMO broadcast channel with imperfect channel state information," *IEEE Trans. Commun.*, vol. 59, no. 3, pp. 803–814, Mar. 2011.

[35] R. A. Horn and C. R. Johnson, *Matrix Analysis*. New York: Cambridge University Press, 1985.

[36] J.-P. Aubin and I. Ekeland, *Applied Nonlinear Analysis*. New York: Dover Publications, 2006.

[37] Y. C. Eldar and N. Merhav, "A competitive minimax approach to robust estimation of random parameters," *IEEE Trans. Signal Process.*, vol. 52, no. 7, pp. 1931–1946, Jul. 2004.

[38] E. Zeidler, *Nonlinear Functional Analysis and Its Applications*. New York: Springer-Verlag, 1986, vol. I, Fixed-Point Theorems.

[39] A. W. Marshall and I. Olkin, *Inequalities: Theory of Majorization and Its Applications*. New York: Academic Press, 1979.