Transceiver Optimization for Two-Hop AF MIMO Relay Systems With DFE Receiver and Direct Link

Qiao Su, and Yue Rong, *Senior Member, IEEE*

***Abstract*— In this paper, we consider precoding and receiving matrices optimization for a two-hop amplify-and-forward (AF) multiple-input multiple-output (MIMO) relay system with a decision feedback equalizer (DFE) at the destination node in the presence of the direct source-destination link. By adopting the minimum mean-squared error (MMSE) criterion, we develop two new transceiver design algorithms for such a system. The first one employs an iterative procedure to design the source, relay, feed-forward, and feedback matrices. The second algorithm is a non-iterative suboptimal approach which decomposes the optimization problem into two tractable subproblems and obtains the source and relay precoding matrices by solving the two subproblems sequentially. Simulation results validate the better MSE and bit-error-rate (BER) performance of the proposed algorithms and show that the non-iterative suboptimal method has a negligible performance loss when the ratio of the source node transmission power to the relay node transmission power is small. In addition, the computational complexity analysis suggests that the second algorithm and one iteration of the first algorithm have the same order of complexity. As the first algorithm typically converges within a few iterations, both proposed algorithms exhibit a low complexity order.**

***Index Terms*— Amplify-and-forward, MIMO relay, precoding matrix, decision feedback equalizer (DFE), direct link.**

I. Introduction

SINCE the multiple-input multiple-output (MIMO) relay communication technique combines the advantages of MIMO and relay systems simultaneously which can effectively improve the quality and coverage of wireless communications, it has received much attention in recent years [1]. There are several relay protocols including amplify-and-forward (AF) and decode-and-forward (DF) [2]. For the AF protocol, the relay node only amplifies (including linear precoding) and forwards the received signals, while for the DF protocol, the relay node decodes and re-encodes the received signals before retransmission. Thus, the AF protocol has a much lower complexity than the DF protocol particularly when we consider MIMO relay systems. Based on these advantages, AF MIMO relay systems are promising for improving the performance of next generation wireless communication systems. In the literature, the AF and DF relays are also referred to as the nonregenerative and regenerative relays, respectively [3]–[7].

Most of existing works design AF MIMO relay systems by maximizing the mutual information (MI) between source and destination [3]–[4] or by minimizing the mean-squared error (MSE) of the estimation of the signal waveform [5]–[6]. In [7], a wide class of commonly applied criteria for AF MIMO relay system design have been investigated under a unified framework. In view of the case with mismatch between the true and the estimated channel state information (CSI), robust transceivers have been designed for AF MIMO relay systems in [8]–[9]. AF MIMO relay system design with constrains on the quality-of-service (QoS) has been developed in [10]–[12].

Works such as [5] and [7] ignored the direct link between the source and the destination. However, in some application scenarios, the direct link is non-negligible, which can be exploited to provide valuable spatial diversity. When the direct link is considered, the corresponding MIMO relay system optimization problems become more complicated. Based on the minimum MSE (MMSE) objective, source and relay precoding matrices design with the direct link has been studied in [13], which employs the projected gradient method and the interior-point method to update the source and relay precoding matrices alternately. In [14], a tri-step algorithm and a bi-step algorithm have been proposed for AF MIMO relay systems considering the direct link, where the tri-step algorithm iteratively optimizes the source, relay, and receiver matrices and the bi-step algorithm only alternately updates the source and relay precoding matrices. Transceiver design for AF MIMO relay systems with the direct link and the source node transmits information in both time slots has been investigated in [15]. By introducing two scaling factors multiplied to the received signals at the destination, two AF MIMO relay system design schemes have been derived in [16]. One algorithm in [16] identifies a locally optimal solution by alternately updating the two scaling factors and the source, relay, and receiver matrices in order to further reduce the complexity of the tri-step algorithm in [14]. Without utilizing the two scaling factors, the other algorithm in [16] provides a closed-form suboptimal solution in a non-iterative

Manuscript received June 9, 2021; revised November 13, 2021 and February 9, 2022; accepted April 10, 2022. Date of publication April 18, 2022; date of current version June 16, 2022. This work was supported by the National Natural Science Foundation of China under Grant Nos. 62001513 and 62071485 and the Natural Science Foundation of Jiang Su Province in China under Grant Nos. BK20200579 and BK20201334. The associate editor coordinating the review of this article and approving it for publication was A. El Shafie. *(Corresponding author: Qiao Su.)*

Qiao Su is with the College of Communications Engineering, Army Engineering University of PLA, Nanjing 210007, China (e-mail: qiaosu810@foxmail.com).

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

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TCOMM.2022.3168281.

Digital Object Identifier 10.1109/TCOMM.2022.3168281

manner by decomposing the problem into two simpler subproblems.

The works [3]–[7] and [13]–[16] consider linear receivers. As well known, a receiver with a nonlinear decision feedback equalizer (DFE) usually yields a lower system bit-error-rate (BER) compared with linear receivers. When a DFE decodes a particular data stream, interference caused by data streams that are already decoded is canceled [17]–[18]. Some works have studied the optimization of AF MIMO relay systems with the DFE receiver. In [19], a multi-hop AF MIMO relay system design with an MMSE-DFE receiver has been proposed. To reduce the computational complexity, a non-iterative method has been proposed in [20] to design the source and relay matrices for AF MIMO relay systems equipped with a DFE receiver in the absence of the direct link. By combining the advantages of exploiting the DFE receiver and the direct link, transceiver optimization with the DFE receiver and the direct link has been investigated in [21] with the zero-forcing (ZF) strategy. However, the transceiver with the ZF strategy has poor performance at low signal-to-noise ratio (SNR), due to the effect of noise enhancement of the ZF receiver. We would like to note that the transceiver optimization problem utilizing the MMSE strategy with the DFE receiver and the direct link is more difficult to solve than that of the ZF scheme.

In this article, we study the transceiver design with perfect and imperfect CSI for two-hop AF MIMO relay systems with a DFE receiver and the direct link. The main contributions of this paper over existing works such as [3]–[21] are summarized below.

1) We propose two new transceiver design algorithms for such system based on the MMSE criterion, where both algorithms can be used for the perfect and imperfect CSI cases. The first algorithm optimizes the source, relay, feed-forward, and feedback matrices iteratively by solving convex subproblems. We show that this iterative algorithm converges to a Nash point [22].
2) In the second proposed method, we substitute the optimal feed-forward matrix into the MSE objective function and derive the optimal structure of the relay precoding matrix, which enables the transceiver design problem to be decomposed into two tractable subproblems. By solving the two subproblems sequentially, we obtain suboptimal source, relay, and feedback matrices without iteration. To the best of our knowledge, both algorithms are proposed for the first time in this paper.
3) Compared with [21], which also considered the DFE receiver and the direct link but only works when the number of transmit data streams is equal to the number of antennas at the source node, both proposed algorithms are applicable to more general scenarios.
4) The robust algorithms extend the proposed methods from the ideal case of perfect CSI to practical systems with imperfect CSI knowledge, thus, bringing the proposed methods one step closer to practical deployment.
5) Simulation results demonstrate that the first algorithm has a better MSE and BER performance than the methods in [16], [20], and [21] under various scenarios. Moreover, although the second algorithm is suboptimal, it has a negligible performance loss compared with the first algorithm when the ratio of the source node transmission power (P_s) to the relay node transmission power (P_r) is small (typically less than two), since the sub-optimality of the decomposition decreases with the reduction of P_s/P_r.
6) Computational complexity analysis shows that the second algorithm and one iteration of the first algorithm have the same complexity order. As the first algorithm typically converges within a few iterations, both proposed algorithms exhibit a low complexity order.

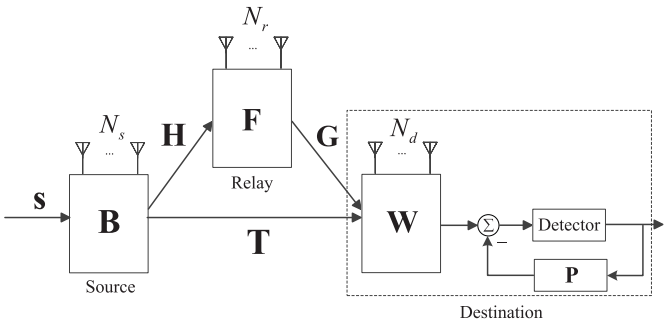


Fig. 1. Block diagram of an AF MIMO relay system with a DFE receiver and direct link.

The remainder of this article is organized as follows. The model of a dual-hop AF MIMO relay system having a DFE receiver and direct link is introduced in Section II. The proposed transceiver optimization algorithms with the exact CSI are presented in Section III. The extension of the proposed transceiver optimization algorithms to the case with imperfect CSI is given in Section IV. Simulation results are provided in Section V to show the performance of the proposed algorithms. Conclusions are drawn in Section VI.

II. System Model

We study a three-node dual-hop AF MIMO relay system as illustrated in Fig. 1, where the information is transmitted from the source node to the destination node with the aid of a relay node. A nonlinear DFE-based receiver is utilized at the destination node. The source node, relay node, and destination node have N_s, N_r, and N_d antennas, respectively. A half-duplex relay is considered in this paper to avoid the loop interference. Therefore, the communication between the source node and the destination node is accomplished in two time intervals.

During the first time interval, an information-bearing vector $\mathbf{s} \in \mathbb{C}^{N_b \times 1}$ ($N_b \leq N_s, N_r, N_d$) is precoded linearly at the source node by a matrix $\mathbf{B} \in \mathbb{C}^{N_s \times N_b}$ as

$$\mathbf{x}_s = \mathbf{B}\mathbf{s} \tag{1}$$

where $E\{\mathbf{s}\mathbf{s}^{\mathrm{H}}\} = \mathbf{I}_{N_b}$, N_b is the number of data streams, and each stream carries independent information bits. Here $(\cdot)^{\mathrm{H}}$ denotes the Hermitian transpose, $E\{\cdot\}$ represents the statistical expectation, and $\mathbf{I}_m$ stands for the $m \times m$ identity matrix. Then $\mathbf{x}_s$ is transmitted from the source node to both the relay node

and the destination node. The received signal vectors at the relay and destination nodes are given respectively by

$$\mathbf{y}_r = \mathbf{H}\mathbf{x}_s + \mathbf{n}_r \tag{2}$$

$$\mathbf{y}_{d_1} = \mathbf{T}\mathbf{x}_s + \mathbf{n}_{d_1} \tag{3}$$

where $\mathbf{H} \in \mathbb{C}^{N_r \times N_s}$ and $\mathbf{T} \in \mathbb{C}^{N_d \times N_s}$ are the channel matrices for the source-relay link and the source-destination link, respectively, $\mathbf{n}_r \in \mathbb{C}^{N_r \times 1}$ is the noise vector at the relay node, and $\mathbf{n}_{d_1} \in \mathbb{C}^{N_d \times 1}$ is the noise vector at the destination node in the first time interval.

In the second time interval, $\mathbf{y}_r$ is linearly precoded by a matrix $\mathbf{F} \in \mathbb{C}^{N_r \times N_r}$ as

$$\mathbf{x}_r = \mathbf{F}\mathbf{y}_r. \tag{4}$$

Then the relay node forwards $\mathbf{x}_r$ to the destination node. The received signal vector at the destination node is written as

$$\mathbf{y}_{d_2} = \mathbf{G}\mathbf{x}_r + \mathbf{n}_{d_2} \tag{5}$$

where $\mathbf{G} \in \mathbb{C}^{N_d \times N_r}$ is the channel matrix for the relay-destination link and $\mathbf{n}_{d_2}$ is the noise vector at the destination node in the second time interval.

According to (1)-(5), the received signal vector at the destination node during two consecutive time intervals can be described by

$$\begin{aligned}\mathbf{y} &= \begin{bmatrix}\mathbf{y}_{d_2}\\ \mathbf{y}_{d_1}\end{bmatrix}\\ &= \begin{bmatrix}\mathbf{GFH}\\ \mathbf{T}\end{bmatrix}\mathbf{B}\mathbf{s} + \begin{bmatrix}\mathbf{GF}\mathbf{n}_r + \mathbf{n}_{d_2}\\ \mathbf{n}_{d_1}\end{bmatrix}\\ &= \mathbf{M}\mathbf{s} + \mathbf{v}\end{aligned} \tag{6}$$

where $\mathbf{M} \in \mathbb{C}^{2N_d \times N_b}$ and $\mathbf{v} \in \mathbb{C}^{2N_d \times 1}$ are defined respectively as

$$\mathbf{M} = \begin{bmatrix}\mathbf{GFHB}\\ \mathbf{TB}\end{bmatrix}, \quad \mathbf{v} = \begin{bmatrix}\mathbf{GF}\mathbf{n}_r + \mathbf{n}_{d_2}\\ \mathbf{n}_{d_1}\end{bmatrix}.$$

We assume that $\mathbf{n}_r$, $\mathbf{n}_{d_1}$, and $\mathbf{n}_{d_2}$ are independent and identically distributed (i.i.d.) Gaussian noise vectors whose entries are zero-mean with unit variance.

At the destination, a nonlinear DFE receiver shown in Fig. 1 is employed to estimate the source signals successively from the N_b-th symbol down to the first symbol. The estimated signal vector can be written as [19]

$$\widehat{\mathbf{s}} = \mathbf{W}^{\mathrm{H}}\mathbf{y} - \mathbf{P}\mathbf{s} \tag{7}$$

where $\mathbf{W} \in \mathbb{C}^{2N_d \times N_b}$ is the feed-forward matrix and $\mathbf{P} \in \mathbb{C}^{N_b \times N_b}$ is a strictly upper triangular feedback matrix with zero main diagonal elements. Based on (6) and (7), at the destination node, the MSE matrix of the signal waveform estimation can be written as

$$\begin{aligned}\mathbf{E} &= E\{(\widehat{\mathbf{s}} - \mathbf{s})(\widehat{\mathbf{s}} - \mathbf{s})^{\mathrm{H}}\}\\ &= (\mathbf{W}^{\mathrm{H}}\mathbf{M} - \mathbf{U})(\mathbf{W}^{\mathrm{H}}\mathbf{M} - \mathbf{U})^{\mathrm{H}} + \mathbf{W}^{\mathrm{H}}\mathbf{C}_v\mathbf{W}\end{aligned} \tag{8}$$

where $\mathbf{U} = \mathbf{P} + \mathbf{I}_{N_b}$ is an upper triangular matrix with unit main diagonal elements and $\mathbf{C}_v = E\{\mathbf{v}\mathbf{v}^{\mathrm{H}}\}$ is the noise covariance matrix and given by

$$\mathbf{C}_v = \begin{bmatrix}\mathbf{GFF}^{\mathrm{H}}\mathbf{G}^{\mathrm{H}} + \mathbf{I}_{N_d} & \mathbf{0}\\ \mathbf{0} & \mathbf{I}_{N_d}\end{bmatrix}.$$

Based on (1) and (4), the transmission power consumption at the source node and the relay node can be computed respectively as

$$E\{\mathrm{tr}(\mathbf{x}_s\mathbf{x}_s^{\mathrm{H}})\} = \mathrm{tr}(\mathbf{B}\mathbf{B}^{\mathrm{H}}) \tag{9}$$

$$E\{\mathrm{tr}(\mathbf{x}_r\mathbf{x}_r^{\mathrm{H}})\} = \mathrm{tr}(\mathbf{F}(\mathbf{H}\mathbf{B}\mathbf{B}^{\mathrm{H}}\mathbf{H}^{\mathrm{H}} + \mathbf{I}_{N_r})\mathbf{F}^{\mathrm{H}}) \tag{10}$$

where $\mathrm{tr}(\cdot)$ is the matrix trace.

Based on (8)-(10), the optimization problem with respect to the source and relay precoding matrices and the feed-forward and feedback matrices at the destination node is formulated as

$$\min_{\mathbf{B},\mathbf{F},\mathbf{W},\mathbf{P}\in\mathcal{U}} \ \mathrm{tr}(\mathbf{E}) \tag{11a}$$

$$\text{s.t. } \mathrm{tr}(\mathbf{B}\mathbf{B}^{\mathrm{H}}) \le P_s \tag{11b}$$

$$\mathrm{tr}(\mathbf{F}(\mathbf{H}\mathbf{B}\mathbf{B}^{\mathrm{H}}\mathbf{H}^{\mathrm{H}} + \mathbf{I}_{N_r})\mathbf{F}^{\mathrm{H}}) \le P_r \tag{11c}$$

where P_s and P_r are the transmission power budget at the source node and the relay node, respectively, and $\mathcal{U}$ denotes the set of $N_b \times N_b$ strictly upper triangular matrices. The problem (11) aims to minimize the MSE of the signal waveform estimation using a DFE receiver, subjecting to the transmission power constraints at the source and relay nodes.

The problem (11) is nonconvex and challenging to solve, particularly when both the direct link and the nonlinear DFE receiver are considered. Note that without the direct link, the structure of the optimal precoding matrices $\mathbf{B}$ and $\mathbf{F}$ are known [19]. However, when the direct link is included, the optimal structure of $\mathbf{B}$ and $\mathbf{F}$ cannot be given in an analytic form. Compared with existing works which consider the direct link without the nonlinear DFE receiver such as [13], [14], and [16], the problem (11) is more complicated since the objective function (11a) contains four matrix variables. Compared with the objective function based on the ZF strategy in [21] which also considers the direct link and the nonlinear DFE receiver, the objective function (11a) with the MMSE strategy is more complex.

III. Proposed Transceiver Design Algorithms

In this section, we present two algorithms for transceiver optimization of AF MIMO relay systems with a DFE receiver and direct link. The first algorithm optimizes the source, relay, feed-forward, and feedback matrices iteratively. While the second algorithm substitutes the optimal feed-forward matrix into the MSE objective function and decomposes the resulting objective function into two parts. Then the second algorithm optimizes the source, relay, and feedback matrices sequentially without iteration.

A. The Iterative Transceiver Design Algorithm

Due to the introduction of $\mathbf{P}$, the problem that iteratively optimizes $\mathbf{B}$, $\mathbf{F}$, $\mathbf{W}$, and $\mathbf{P}$ is more complex compared with that in [14] and [16]. In order to solve the problem (11), we introduce a nonzero real scalar ξ. The received signal vector at the destination node during two consecutive time intervals can be reformulated by $\widehat{\mathbf{y}} = \xi\mathbf{y}$, where ξ can be viewed as a gain control factor at the destination node. In this

case, we rewrite the feed-forward matrix $\widehat{\mathbf{W}} = (1/\xi)\mathbf{W}$. It will be shown later that the introduction of ξ facilitates the optimization of the source precoding matrix $\mathbf{B}$. Note that since $\widehat{\mathbf{W}}^{\mathrm{H}}\widehat{\mathbf{y}} = \mathbf{W}^{\mathrm{H}}\mathbf{y}$, (7) holds unchanged after introducing the nonzero real scalar ξ. Thus, the MSE matrix of the signal waveform estimation in (8) stays the same. We can rewrite the MSE matrix of the signal waveform estimation equivalently by using $\mathbf{W} = \xi\widehat{\mathbf{W}}$, which is given by

$$\widehat{\mathbf{E}}=(\xi\widehat{\mathbf{W}}^{\mathrm{H}}\mathbf{M}-\mathbf{U})(\xi\widehat{\mathbf{W}}^{\mathrm{H}}\mathbf{M}-\mathbf{U})^{\mathrm{H}}+\xi^2\widehat{\mathbf{W}}^{\mathrm{H}}\mathbf{C}_v\widehat{\mathbf{W}}. \quad (12)$$

The sign of ξ can be absorbed by $\widehat{\mathbf{W}}$ according to the equation $\mathbf{W} = \xi\widehat{\mathbf{W}}$. Moreover, our final goal is to obtain $\mathbf{W}$ instead of ξ and $\widehat{\mathbf{W}}$. Thus, the sign of ξ does no affect the transceiver optimization. For the simplicity of presentation, we set $\xi > 0$ in the following text.

The problem (11) can be reformulated as

$$\min_{\mathbf{B},\mathbf{F},\widehat{\mathbf{W}},\xi,\mathbf{P}\in\mathcal{U}} \ \mathrm{tr}(\widehat{\mathbf{E}}) \quad (13a)$$

$$\text{s.t. } \mathrm{tr}(\mathbf{B}\mathbf{B}^{\mathrm{H}}) \le P_s \quad (13b)$$

$$\mathrm{tr}(\mathbf{F}(\mathbf{H}\mathbf{B}\mathbf{B}^{\mathrm{H}}\mathbf{H}^{\mathrm{H}}+\mathbf{I}_{N_r})\mathbf{F}^{\mathrm{H}}) \le P_r. \quad (13c)$$

The proposed algorithm updates $\mathbf{B}$, $\mathbf{F}$, $\widehat{\mathbf{W}}$, ξ, and $\mathbf{P}$ iteratively till convergence. In each iteration, we first optimize $\widehat{\mathbf{W}}$ and $\mathbf{P}$ using given $\mathbf{B}$, ξ, and $\mathbf{F}$. Then, we optimize $\mathbf{F}$ with fixed $\mathbf{B}$, $\widehat{\mathbf{W}}$, ξ, and $\mathbf{P}$. Finally, $\mathbf{B}$ and ξ are optimized with given $\mathbf{F}$, $\widehat{\mathbf{W}}$, and $\mathbf{P}$. This procedure continuous till convergence.

Firstly, when $\mathbf{B}$, ξ, and $\mathbf{F}$ are given, from (13), $\widehat{\mathbf{W}}$ and $\mathbf{P}$ can be obtained from solving the problem of

$$\min_{\widehat{\mathbf{W}},\mathbf{P}\in\mathcal{U}} \ \mathrm{tr}(\widehat{\mathbf{E}}). \quad (14)$$

Let us introduce the QR decomposition [24] of

$$\begin{bmatrix} (\mathbf{C}'_v)^{-\frac{1}{2}}\mathbf{M}' \\ \mathbf{I}_{N_b} \end{bmatrix} = \mathbf{Q}\mathbf{R} \quad (15)$$

where $\mathbf{M}' = \xi\mathbf{M}$, $\mathbf{C}'_v = \xi^2\mathbf{C}_v$, and $(\cdot)^{-1}$ denotes matrix inversion. It can be shown from [19] that the solution to the problem (14) is

$$\widehat{\mathbf{W}} = (\mathbf{C}'_v)^{-\frac{1}{2}}\mathbf{Q}_1\mathbf{D}_r^{-1}, \quad \mathbf{P} = \mathbf{D}_r^{-1}\mathbf{R} - \mathbf{I}_{N_b} \quad (16)$$

where $\mathbf{Q}_1$ contains the first $2N_d$ rows of $\mathbf{Q}$ in (15) and $\mathbf{D}_r$ is a diagonal matrix whose main diagonal elements are taken from the main diagonal elements of $\mathbf{R}$.

Secondly, with given $\mathbf{B}$, $\widehat{\mathbf{W}}$, ξ, and $\mathbf{P}$, from (13), we can obtain the optimal matrix $\mathbf{F}$ via the following problem

$$\min_{\mathbf{F}} \ \mathrm{tr}((\mathbf{H}_2\mathbf{F}\mathbf{H}_1-\mathbf{\Theta})(\mathbf{H}_2\mathbf{F}\mathbf{H}_1-\mathbf{\Theta})^{\mathrm{H}}+\mathbf{H}_2\mathbf{F}\mathbf{F}^{\mathrm{H}}\mathbf{H}_2^{\mathrm{H}}) \quad (17a)$$

$$\text{s.t. } \mathrm{tr}(\mathbf{F}(\mathbf{H}_1\mathbf{H}_1^{\mathrm{H}}+\mathbf{I}_{N_r})\mathbf{F}^{\mathrm{H}}) \le P_r \quad (17b)$$

where $\mathbf{H}_1 = \mathbf{H}\mathbf{B}$, $\mathbf{H}_2 = \xi\widehat{\mathbf{W}}_1^{\mathrm{H}}\mathbf{G}$, $\mathbf{\Theta} = \mathbf{U} - \xi\widehat{\mathbf{W}}_2^{\mathrm{H}}\mathbf{T}\mathbf{B}$, $\widehat{\mathbf{W}}_1$ and $\widehat{\mathbf{W}}_2$ contain the first and last N_d rows of $\widehat{\mathbf{W}}$, respectively. The solution to the problem (17) can be easily found by resorting to the Lagrange multiplier method as

$$\mathbf{F} = \mathbf{H}_2^{\mathrm{H}}(\mathbf{H}_2\mathbf{H}_2^{\mathrm{H}}+\gamma\mathbf{I}_{N_b})^{-1}\mathbf{\Theta}\boldsymbol{H}_1^{\mathrm{H}}(\mathbf{H}_1\mathbf{H}_1^{\mathrm{H}}+\mathbf{I}_{N_r})^{-1} \quad (18)$$

where $\gamma \ge 0$ is the Lagrange multiplier.

When $\gamma = 0$, $\mathbf{F}$ is obtained from (18) as

$$\mathbf{F} = \mathbf{H}_2^{\mathrm{H}}(\mathbf{H}_2\mathbf{H}_2^{\mathrm{H}})^{-1}\mathbf{\Theta}\boldsymbol{H}_1^{\mathrm{H}}(\mathbf{H}_1\mathbf{H}_1^{\mathrm{H}}+\mathbf{I}_{N_r})^{-1}. \quad (19)$$

If $\mathbf{F}$ given by (19) satisfies the constraint (17b), then (19) is the optimal solution to the problem (17). Otherwise, we have $\gamma > 0$ such that $\mathbf{F}$ in (18) satisfies the constraint (17b) with equality as

$$\mathrm{tr}(\mathbf{H}_2^{\mathrm{H}}(\mathbf{H}_2\mathbf{H}_2^{\mathrm{H}}+\gamma\mathbf{I}_{N_b})^{-1}\mathbf{\Upsilon}(\mathbf{H}_2\mathbf{H}_2^{\mathrm{H}}+\gamma\mathbf{I}_{N_b})^{-1}\mathbf{H}_2) = P_r \quad (20)$$

where $\mathbf{\Upsilon} = \mathbf{\Theta}\boldsymbol{H}_1^{\mathrm{H}}(\mathbf{H}_1\mathbf{H}_1^{\mathrm{H}}+\mathbf{I}_{N_r})^{-1}\mathbf{H}_1\mathbf{\Theta}^{\mathrm{H}}$. By using the singular value decomposition (SVD) of $\mathbf{H}_2 = \mathbf{U}_2\mathbf{\Lambda}_2\mathbf{V}_2$ and defining $\mathbf{\Omega} = \mathbf{U}_2^{\mathrm{H}}\mathbf{\Upsilon}\mathbf{U}_2$, (20) can be rewritten as

$$\sum_{i=1}^{N_b} \lambda_{2,i}^2\omega_i(\lambda_{2,i}^2+\gamma)^{-2} = P_r \quad (21)$$

where $\lambda_{2,i}$ and ω_i are the ith main diagonal elements of $\mathbf{\Lambda}_2$ and $\mathbf{\Omega}$, respectively. We can obtain γ from (21) by the bisection method [26], since the left-hand side (LHS) of (21) monotonically decreases with $\gamma > 0$.

Thirdly, with fixed $\widehat{\mathbf{W}}$, $\mathbf{P}$, and $\mathbf{F}$, the problem (13) becomes

$$\min_{\mathbf{B},\xi} \ \mathrm{tr}((\xi\mathbf{D}_1\mathbf{B}-\mathbf{U})(\xi\mathbf{D}_1\mathbf{B}-\mathbf{U})^{\mathrm{H}}+\xi^2\widehat{\mathbf{W}}^{\mathrm{H}}\mathbf{C}_v\widehat{\mathbf{W}}) \quad (22a)$$

$$\text{s.t. } \mathrm{tr}(\mathbf{B}\mathbf{B}^{\mathrm{H}}) \le P_s \quad (22b)$$

$$\mathrm{tr}(\mathbf{B}^{\mathrm{H}}\mathbf{D}_2^{\mathrm{H}}\mathbf{D}_2\mathbf{B}) \le \hat{P}_r \quad (22c)$$

where $\mathbf{D}_1 = \widehat{\mathbf{W}}_1^{\mathrm{H}}\mathbf{G}\mathbf{F}\mathbf{H} + \widehat{\mathbf{W}}_2^{\mathrm{H}}\mathbf{T}$, $\mathbf{D}_2 = \mathbf{F}\mathbf{H}$, and $\hat{P}_r = P_r - \mathrm{tr}(\mathbf{F}\mathbf{F}^{\mathrm{H}})$. Note that optimizing $\mathbf{B}$ is more challenging compared with the optimization of $\mathbf{W}$, $\mathbf{P}$, and $\mathbf{F}$, as $\mathbf{B}$ appears in both constraints (22b) and (22c). The introduced ξ can be used to solve this problem without using any optimization toolbox (which usually has a high complexity order).

The Lagrangian function of the problem (22) can be expressed as

$$\begin{aligned}\mathcal{L}=&\mathrm{tr}((\xi\mathbf{D}_1\mathbf{B}-\mathbf{U})(\xi\mathbf{D}_1\mathbf{B}-\mathbf{U})^{\mathrm{H}}+\xi^2\widehat{\mathbf{W}}^{\mathrm{H}}\mathbf{C}_v\widehat{\mathbf{W}})\\ &+\mu_1(\mathrm{tr}(\mathbf{B}\mathbf{B}^{\mathrm{H}})-P_s)+\mu_2(\mathrm{tr}(\mathbf{B}^{\mathrm{H}}\mathbf{D}_2^{\mathrm{H}}\mathbf{D}_2\mathbf{B})-\hat{P}_r)\end{aligned} \quad (23)$$

where $\mu_1 \ge 0$ and $\mu_2 \ge 0$ are the Lagrange multipliers.

Based on the first-order derivative condition of $\partial\mathcal{L}/\partial\mathbf{B} = 0$, we can obtain from (23) that

$$(\mathbf{D}_1^{\mathrm{H}}\mathbf{D}_1 + \widetilde{\mu}_1\mathbf{I}_{N_s} + \widetilde{\mu}_2\mathbf{D}_2^{\mathrm{H}}\mathbf{D}_2)\mathbf{B} = (1/\xi)\mathbf{D}_1^{\mathrm{H}}\mathbf{U} \quad (24)$$

where $\widetilde{\mu}_1 = \mu_1/\xi^2$ and $\widetilde{\mu}_2 = \mu_2/\xi^2$. By employing the first-order derivative condition of $\partial\mathcal{L}/\partial\xi = 0$, we get

$$\begin{aligned}&\mathrm{tr}(\mathbf{D}_1\mathbf{B}\mathbf{B}^{\mathrm{H}}\mathbf{D}_1^{\mathrm{H}}) + \mathrm{tr}(\widehat{\mathbf{W}}^{\mathrm{H}}\mathbf{C}_v\widehat{\mathbf{W}})\\ &\qquad = (1/2\xi)(\mathrm{tr}(\mathbf{U}\mathbf{B}^{\mathrm{H}}\mathbf{D}_1^{\mathrm{H}}) + \mathrm{tr}(\mathbf{D}_1\mathbf{B}\mathbf{U}^{\mathrm{H}})).\end{aligned} \quad (25)$$

Pre-multiplying (24) by $\mathbf{B}^{\mathrm{H}}$, we have

$$\mathbf{B}^{\mathrm{H}}(\mathbf{D}_1^{\mathrm{H}}\mathbf{D}_1+\widetilde{\mu}_1\mathbf{I}_{N_s}+\widetilde{\mu}_2\mathbf{D}_2^{\mathrm{H}}\mathbf{D}_2)\mathbf{B} = (1/\xi)\mathbf{B}^{\mathrm{H}}\mathbf{D}_1^{\mathrm{H}}\mathbf{U}. \quad (26)$$

From (26), we can see that $\mathbf{B}^{\mathrm{H}}\mathbf{D}_1^{\mathrm{H}}\mathbf{U}$ is a Hermitian Matrix. Thus, it can be obtained easily that $\mathrm{tr}(\mathbf{U}\mathbf{B}^{\mathrm{H}}\mathbf{D}_1^{\mathrm{H}}) = \mathrm{tr}(\mathbf{D}_1\mathbf{B}\mathbf{U}^{\mathrm{H}})$. Then, by combining (25) and (26), we can get

$$\mathrm{tr}(\widehat{\mathbf{W}}^{\mathrm{H}}\mathbf{C}_v\widehat{\mathbf{W}}) = \widetilde{\mu}_1\mathrm{tr}(\mathbf{B}^{\mathrm{H}}\mathbf{B}) + \widetilde{\mu}_2\mathrm{tr}(\mathbf{B}^{\mathrm{H}}\mathbf{D}_2^{\mathrm{H}}\mathbf{D}_2\mathbf{B}). \quad (27)$$

Interestingly, (27) shows the relationship between $\widetilde{\mu}_1$ and $\widetilde{\mu}_2$. From (24), we can compute $\mathbf{B}$ as

$$\mathbf{B} = (1/\xi)(\mathbf{D}_1^{\mathrm{H}}\mathbf{D}_1 + \widetilde{\mu}_1\mathbf{I}_{N_s} + \widetilde{\mu}_2\mathbf{D}_2^{\mathrm{H}}\mathbf{D}_2)^{-1}\mathbf{D}_1^{\mathrm{H}}\mathbf{U}. \quad (28)$$

If $\widetilde{\mu}_1 = \widetilde{\mu}_2 = 0$, we can see from (27) that $\mathrm{tr}(\widehat{\mathbf{W}}^{\mathrm{H}}\mathbf{C}_v\widehat{\mathbf{W}}) = 0$, which cannot be true. Thus, we consider the following three cases: i) $\widetilde{\mu}_1 = 0$ and $\widetilde{\mu}_2 > 0$; ii) $\widetilde{\mu}_1 > 0$ and $\widetilde{\mu}_2 = 0$; iii) $\widetilde{\mu}_1 > 0$ and $\widetilde{\mu}_2 > 0$. Assuming that $\widetilde{\mu}_1 = 0$ and $\widetilde{\mu}_2 > 0$, we have $\mathrm{tr}(\mathbf{B}^{\mathrm{H}}\mathbf{D}_2^{\mathrm{H}}\mathbf{D}_2\mathbf{B}) = \hat{P}_r$ according to the complementary slackness condition. Thus, it can be deduced from (27) and (28) that

$$\mathbf{B} = \frac{1}{\xi}\left(\mathbf{D}_1^{\mathrm{H}}\mathbf{D}_1 + \frac{\mathrm{tr}(\widehat{\mathbf{W}}^{\mathrm{H}}\mathbf{C}_v\widehat{\mathbf{W}})}{\widehat{P}_r}\mathbf{D}_2^{\mathrm{H}}\mathbf{D}_2\right)^{-1}\mathbf{D}_1^{\mathrm{H}}\mathbf{U} \quad (29)$$

where $\xi = \sqrt{\mathrm{tr}(\overline{\mathbf{B}}_1^{\mathrm{H}}\mathbf{D}_2^{\mathrm{H}}\mathbf{D}_2\overline{\mathbf{B}}_1)\Big/\widehat{P}_r}$ with

$$\overline{\mathbf{B}}_1 = \left(\mathbf{D}_1^{\mathrm{H}}\mathbf{D}_1 + \frac{\mathrm{tr}(\widehat{\mathbf{W}}^{\mathrm{H}}\mathbf{C}_v\widehat{\mathbf{W}})}{\widehat{P}_r}\mathbf{D}_2^{\mathrm{H}}\mathbf{D}_2\right)^{-1}\mathbf{D}_1^{\mathrm{H}}\mathbf{U}. \quad (30)$$

If the constraint (22b) is satisfied by $\mathbf{B}$ in (29), then (29) is the optimal solution to the problem (22). Otherwise, $\widetilde{\mu}_1 = 0$ and $\widetilde{\mu}_2 > 0$ do not hold true.

Now we consider the case of $\widetilde{\mu}_1 > 0$ and $\widetilde{\mu}_2 = 0$. According to the complementary slackness condition, we get $\mathrm{tr}(\mathbf{B}\mathbf{B}^{\mathrm{H}}) = P_s$ in this case. Combining (27) and (28), we obtain

$$\mathbf{B} = \frac{1}{\xi}\left(\mathbf{D}_1^{\mathrm{H}}\mathbf{D}_1 + \frac{\mathrm{tr}(\widehat{\mathbf{W}}^{\mathrm{H}}\mathbf{C}_v\widehat{\mathbf{W}})}{P_s}\mathbf{I}_{N_s}\right)^{-1}\mathbf{D}_1^{\mathrm{H}}\mathbf{U} \quad (31)$$

where $\xi = \sqrt{\mathrm{tr}(\overline{\mathbf{B}}_2\overline{\mathbf{B}}_2^{\mathrm{H}})\Big/P_s}$ with

$$\overline{\mathbf{B}}_2 = \left(\mathbf{D}_1^{\mathrm{H}}\mathbf{D}_1 + \frac{\mathrm{tr}(\widehat{\mathbf{W}}^{\mathrm{H}}\mathbf{C}_v\widehat{\mathbf{W}})}{P_s}\mathbf{I}_{N_s}\right)^{-1}\mathbf{D}_1^{\mathrm{H}}\mathbf{U}. \quad (32)$$

If the constraint (22c) is satisfied by $\mathbf{B}$ in (31), then (31) is the optimal solution of the problem (22). Otherwise, $\widetilde{\mu}_1 > 0$ and $\widetilde{\mu}_2 = 0$ is not true.

For the case of $\widetilde{\mu}_1 > 0$ and $\widetilde{\mu}_2 > 0$, we can deduce from (27) and (28) that

$$\mathbf{B} = \frac{1}{\xi}(\mathbf{D}_1^{\mathrm{H}}\mathbf{D}_1 + \widetilde{\mu}_1\mathbf{I}_{N_s} + \frac{\mathrm{tr}(\widehat{\mathbf{W}}^{\mathrm{H}}\mathbf{C}_v\widehat{\mathbf{W}}) - \widetilde{\mu}_1 P_s}{\hat{P}_r}\mathbf{D}_2^{\mathrm{H}}\mathbf{D}_2)^{-1}\mathbf{D}_1^{\mathrm{H}}\mathbf{U} \quad (33)$$

where $\xi = \sqrt{\mathrm{tr}(\overline{\mathbf{B}}_3\overline{\mathbf{B}}_3^{\mathrm{H}})\Big/P_s}$ with

$$\overline{\mathbf{B}}_3 = (\mathbf{D}_1^{\mathrm{H}}\mathbf{D}_1 + \widetilde{\mu}_1\mathbf{I}_{N_s} + \frac{\mathrm{tr}(\widehat{\mathbf{W}}^{\mathrm{H}}\mathbf{C}_v\widehat{\mathbf{W}}) - \widetilde{\mu}_1 P_s}{\hat{P}_r}\mathbf{D}_2^{\mathrm{H}}\mathbf{D}_2)^{-1}\mathbf{D}_1^{\mathrm{H}}\mathbf{U}. \quad (34)$$

It can be seen that $\mathbf{B}$ in (33) is a function of $\widetilde{\mu}_1$. Note that in this case, $\mathbf{B}$ in (33) should satisfy $\mathrm{tr}(\mathbf{B}\mathbf{B}^{\mathrm{H}}) = P_s$ and $\mathrm{tr}(\mathbf{B}^{\mathrm{H}}\mathbf{D}_2^{\mathrm{H}}\mathbf{D}_2\mathbf{B}) = \hat{P}_r$ simultaneously. Thus, we have

$$\frac{\mathrm{tr}(\mathbf{B}\mathbf{B}^{\mathrm{H}})}{\mathrm{tr}(\mathbf{B}^{\mathrm{H}}\mathbf{D}_2^{\mathrm{H}}\mathbf{D}_2\mathbf{B})} = \frac{P_s}{\hat{P}_r}. \quad (35)$$

From (33) and (35), we can get

$$\mathrm{tr}(\mathbf{Q}\mathbf{B}_0^{-1}(\widetilde{\mu}_1)\mathbf{D}_1^{\mathrm{H}}\mathbf{U}\mathbf{U}^{\mathrm{H}}\mathbf{D}_1\mathbf{B}_0^{-1}(\widetilde{\mu}_1)) = 0 \quad (36)$$

where $\mathbf{B}_0(\widetilde{\mu}_1) = \mathbf{D}_1^{\mathrm{H}}\mathbf{D}_1 + \widetilde{\mu}_1\mathbf{I}_{N_s} + \frac{\mathrm{tr}(\widehat{\mathbf{W}}^{\mathrm{H}}\mathbf{C}_v\widehat{\mathbf{W}}) - \widetilde{\mu}_1 P_s}{\hat{P}_r}\mathbf{D}_2^{\mathrm{H}}\mathbf{D}_2$ and $\mathbf{Q} = \mathbf{I}_{N_s} - \frac{P_s}{\hat{P}_r}\mathbf{D}_2^{\mathrm{H}}\mathbf{D}_2$. Let us introduce

$$f(\widetilde{\mu}_1) = \mathrm{tr}(\mathbf{Q}\mathbf{B}_0^{-1}(\widetilde{\mu}_1)\mathbf{D}_1^{\mathrm{H}}\mathbf{U}\mathbf{U}^{\mathrm{H}}\mathbf{D}_1\mathbf{B}_0^{-1}(\widetilde{\mu}_1)).$$

In Appendix, we prove that $f(\widetilde{\mu}_1)$ is a non-increasing function with respect to $\widetilde{\mu}_1$. Therefore, the bisection method can be employed to solve $\widetilde{\mu}_1$ from (36). Then, $\mathbf{B}$ and ξ are obtained from (33) by using the calculated $\widetilde{\mu}_1$.

According to (16), (18), (29), (31), and (33), the steps of using the proposed iterative algorithm to solve the transceiver optimization problem (13) are listed in Algorithm 1, where the superscript (n) stands for variables at the nth iteration and $\varepsilon > 0$ is a small number for which convergence is acceptable. Since the subproblem (14) of optimizing $\widehat{\mathbf{W}}$ and $\mathbf{P}$, the subproblem (17) of updating $\mathbf{F}$, and the subproblem (22) of optimizing $\mathbf{B}$ and ξ are all convex optimization problem, based on [22], Algorithm 1 converges to a Nash point of the problem (13).

Algorithm 1 Iterative Transceiver Optimization Algorithm

Input: $\mathbf{H}$, $\mathbf{G}$, $\mathbf{T}$, P_s, P_r.
Output: $\mathbf{B}$, $\mathbf{F}$, $\widehat{\mathbf{W}}$, ξ, $\mathbf{P}$.
Initialize $\mathbf{B}^{(0)}$, $\xi^{(0)}$ and $\mathbf{F}^{(0)}$, set $n = 0$ and $\mathrm{tr}(\mathbf{E}^{(0)}) = N_b$.
repeat
  Set $n := n + 1$.
  Optimize $\widehat{\mathbf{W}}^{(n)}$ and $\mathbf{P}^{(n)}$ for fixed $\mathbf{B}^{(n-1)}$, $\xi^{(n-1)}$, and $\mathbf{F}^{(n-1)}$ as (16).
  Calculate $\mathbf{F}^{(n)}$ for fixed $\widehat{\mathbf{W}}^{(n)}$, $\mathbf{P}^{(n)}$, $\mathbf{B}^{(n-1)}$, and $\xi^{(n-1)}$ as (18).
  Update $\mathbf{B}^{(n)}$ and $\xi^{(n)}$ for given $\widehat{\mathbf{W}}^{(n)}$, $\mathbf{P}^{(n)}$, and $\mathbf{F}^{(n)}$ according to the following three cases:
  1) If $\mathbf{B}^{(n)}$ in (29) satisfies (22b), then $\mathbf{B}^{(n)}$ and $\xi^{(n)}$ are given by (29);
  2) If $\mathbf{B}^{(n)}$ in (31) satisfies (22c), then $\mathbf{B}^{(n)}$ and $\xi^{(n)}$ are given by (31);
  3) Otherwise, $\mathbf{B}^{(n)}$ and $\xi^{(n)}$ are given by (33)
until $(\mathrm{tr}(\mathbf{E}^{(n-1)}) - \mathrm{tr}(\mathbf{E}^{(n)}))/\mathrm{tr}(\mathbf{E}^{(n-1)}) \leq \varepsilon$

B. The Non-Iterative Transceiver Design Algorithm

In this subsection, we present a non-iterative transceiver design approach, which is suboptimal but has a lower computational complexity compared with the iterative algorithm in Section III-A. We consider the problem (11). Taking the first order derivative of $\mathrm{tr}(\mathbf{E})$ in (8) with respect to $\mathbf{W}$ and equating the result to zero, we obtain the optimal feed-forward matrix $\mathbf{W}$ minimizing $\mathrm{tr}(\mathbf{E})$ in (8) with fixed $\mathbf{B}$, $\mathbf{F}$, and $\mathbf{P}$, which is given by [25]

$$\mathbf{W} = (\mathbf{M}\mathbf{M}^{\mathrm{H}} + \mathbf{C}_v)^{-1}\mathbf{M}(\mathbf{P} + \mathbf{I}_{N_b})^{\mathrm{H}}. \quad (37)$$

By substituting (37) back into (8), we obtain

$$\mathbf{E} = \mathbf{U}[\mathbf{I}_{N_b} + \mathbf{B}^{\mathrm{H}}\mathbf{T}^{\mathrm{H}}\mathbf{T}\mathbf{B} + \mathbf{B}^{\mathrm{H}}\mathbf{H}^{\mathrm{H}}\mathbf{F}^{\mathrm{H}}\mathbf{G}^{\mathrm{H}} \times (\mathbf{G}\mathbf{F}\mathbf{F}^{\mathrm{H}}\mathbf{G}^{\mathrm{H}} + \mathbf{I}_{N_d})^{-1}\mathbf{G}\mathbf{F}\mathbf{H}\mathbf{B}]^{-1}\mathbf{U}^{\mathrm{H}}. \quad (38)$$

Based on [27], it can be shown that the optimal $\mathbf{F}$ that minimizes $\mathrm{tr}(\mathbf{E})$ in (38) can be written as

$$\mathbf{F} = \mathbf{DL} \tag{39}$$

where $\mathbf{L} = (\mathbf{B}^{\mathrm{H}}\mathbf{H}^{\mathrm{H}}\mathbf{H}\mathbf{B} + \mathbf{B}^{\mathrm{H}}\mathbf{T}^{\mathrm{H}}\mathbf{T}\mathbf{B} + \mathbf{I}_{N_b})^{-1}\mathbf{B}^{\mathrm{H}}\mathbf{H}^{\mathrm{H}}$ and $\mathbf{D} \in \mathbb{C}^{N_r \times N_b}$ remains to be optimized. Interestingly, $\mathbf{D}$ can be viewed as the linear transmitter matrix at the relay node, while $\mathbf{L}$ can be treated as the linear receiver matrix at the relay node.

By substituting (39) back into (38), we can decompose (38) into two MSE matrices as

$$\mathbf{E} = \mathbf{E}_1 + \mathbf{E}_2 \tag{40}$$

where

$$\mathbf{E}_1 = \mathbf{U}(\mathbf{B}^{\mathrm{H}}\mathbf{H}^{\mathrm{H}}\mathbf{H}\mathbf{B} + \mathbf{B}^{\mathrm{H}}\mathbf{T}^{\mathrm{H}}\mathbf{T}\mathbf{B} + \mathbf{I}_{N_b})^{-1}\mathbf{U}^{\mathrm{H}} \tag{41}$$

$$\mathbf{E}_2 = \mathbf{U}(\mathbf{D}^{\mathrm{H}}\mathbf{G}^{\mathrm{H}}\mathbf{G}\mathbf{D} + \mathbf{\Omega}^{-1})^{-1}\mathbf{U}^{\mathrm{H}}. \tag{42}$$

Here $\mathbf{\Omega} = \mathbf{L}(\mathbf{H}\mathbf{B}(\mathbf{B}^{\mathrm{H}}\mathbf{T}^{\mathrm{H}}\mathbf{T}\mathbf{B} + \mathbf{I}_{N_b})^{-1}\mathbf{B}^{\mathrm{H}}\mathbf{H}^{\mathrm{H}} + \mathbf{I}_{N_r})\mathbf{L}^{\mathrm{H}}$. Using (40), the problem (11) can be rewritten as

$$\min_{\mathbf{B},\mathbf{D},\mathbf{P}\in\mathcal{U}} \ \mathrm{tr}(\mathbf{E}_1) + \mathrm{tr}(\mathbf{E}_2) \tag{43a}$$

$$\text{s.t. } \mathrm{tr}(\mathbf{B}\mathbf{B}^{\mathrm{H}}) \le P_s \tag{43b}$$

$$\mathrm{tr}(\mathbf{DL}(\mathbf{H}\mathbf{B}\mathbf{B}^{\mathrm{H}}\mathbf{H}^{\mathrm{H}} + \mathbf{I}_{N_r})\mathbf{L}^{\mathrm{H}}\mathbf{D}^{\mathrm{H}}) \le P_r. \tag{43c}$$

Since $\mathbf{D}$ does not appear in $\mathrm{tr}(\mathbf{E}_1)$ and the constraint (43b), we propose to optimize $\mathbf{B}$ and $\mathbf{P}$ through solving the problem below

$$\min_{\mathbf{B},\mathbf{P}\in\mathcal{U}} \ \mathrm{tr}(\mathbf{E}_1) \tag{44a}$$

$$\text{s.t. } \mathrm{tr}(\mathbf{B}\mathbf{B}^{\mathrm{H}}) \le P_s. \tag{44b}$$

Then, using the solution of $\mathbf{B}$ and $\mathbf{P}$ from the problem (44), we can obtain $\mathbf{D}$ by solving the optimization problem of

$$\min_{\mathbf{D}} \ \mathrm{tr}(\mathbf{E}_2) \tag{45a}$$

$$\text{s.t. } \mathrm{tr}(\mathbf{DL}(\mathbf{H}\mathbf{B}\mathbf{B}^{\mathrm{H}}\mathbf{H}^{\mathrm{H}} + \mathbf{I}_{N_r})\mathbf{L}^{\mathrm{H}}\mathbf{D}^{\mathrm{H}}) \le P_r. \tag{45b}$$

Thus, the problem (43) is decomposed into two subproblems, the problem (44) and the problem (45). Since we consider the direct link and the nonlinear DFE receiver, the two MSE matrices $\mathbf{E}_1$ and $\mathbf{E}_2$ are complicated functions of $\mathbf{B}$ and $\mathbf{D}$ because of the introduction of the upper triangular matrix $\mathbf{U}$ and the channel matrix $\mathbf{T}$ for the direct link. Thus, the optimization problems (44) and (45) are challenging to solve compared with existing works. We would like to note that as $\mathbf{B}$ is in (45a) and (45b), and $\mathbf{P}$ appears in the objective function (45a), such decomposition is suboptimal. The performance of the decomposition depends on the value of P_s relative to P_r. When P_s is small compared with P_r, the feasible region specified by (44b) is small. Therefore, the suboptimality of obtaining $\mathbf{B}$ by solving the problem (44) has smaller impact on the feasible region of $\mathbf{D}$ in (45b) with smaller P_s. Thus, with increasing P_r/P_s, the performance of the non-iterative method by solving the two subproblems is getting closer to the iterative algorithm in Section III-A. Moreover, this non-iterative method has a lower computational complexity compared with the iterative algorithm. It will be shown through numerical simulations that the non-iterative method has a negligible performance loss when the source node transmission power is less than the relay node transmission power.

In order to solve the problem (44), we first introduce the arithmetic-geometric mean inequality [24] which shows that for a positive semidefinite matrix $\mathbf{K} \in \mathbb{C}^{n\times n}$, there is $\frac{1}{n}\mathrm{tr}(\mathbf{K}) \ge |\mathbf{K}|^{\frac{1}{n}}$, where $|\cdot|$ denotes matrix determinant. The equality holds if and only if $\mathbf{K} = \alpha\mathbf{I}_n$ with $\alpha \ge 0$. Therefore, we have

$$\begin{aligned}\frac{1}{N_b}\mathrm{tr}(\mathbf{E}_1) &\ge |\mathbf{E}_1|^{\frac{1}{N_b}} \\ &= \left|\mathbf{U}(\mathbf{B}^{\mathrm{H}}\mathbf{H}^{\mathrm{H}}\mathbf{H}\mathbf{B} + \mathbf{B}^{\mathrm{H}}\mathbf{T}^{\mathrm{H}}\mathbf{T}\mathbf{B} + \mathbf{I}_{N_b})^{-1}\mathbf{U}^{\mathrm{H}}\right|^{\frac{1}{N_b}} \\ &= \left|\mathbf{B}^{\mathrm{H}}\mathbf{H}^{\mathrm{H}}\mathbf{H}\mathbf{B} + \mathbf{B}^{\mathrm{H}}\mathbf{T}^{\mathrm{H}}\mathbf{T}\mathbf{B} + \mathbf{I}_{N_b}\right|^{-\frac{1}{N_b}}\end{aligned} \tag{46}$$

where the fact of $|\mathbf{U}| = 1$ is used. Let us define $\mathbf{Z} = \mathbf{H}^{\mathrm{H}}\mathbf{H} + \mathbf{T}^{\mathrm{H}}\mathbf{T}$ and introduce the SVD of $\mathbf{B} = \mathbf{U}_b\mathbf{\Lambda}_b\mathbf{V}_b^{\mathrm{H}}$ and the eigenvalue decomposition (EVD) of $\mathbf{Z} = \mathbf{V}_z\mathbf{\Lambda}_z\mathbf{V}_z^{\mathrm{H}}$, where $\mathbf{\Lambda}_b$ and $\mathbf{\Lambda}_z$ are $N_b \times N_b$ and $N_s \times N_s$ diagonal matrices, respectively, and the main diagonal elements of $\mathbf{\Lambda}_b$ and $\mathbf{\Lambda}_z$ are arranged in descending order. Using the Hadamard's inequality [24], from (46) we obtain

$$\left|\mathbf{B}^{\mathrm{H}}\mathbf{Z}\mathbf{B} + \mathbf{I}_{N_b}\right|^{-\frac{1}{N_b}} \ge \prod_{i=1}^{N_b}(\lambda_{b,i}^2\lambda_{z,i} + 1)^{-\frac{1}{N_b}} \tag{47}$$

where $\lambda_{b,i}$ and $\lambda_{z,i}$ are the ith main diagonal elements of $\mathbf{\Lambda}_b$ and $\mathbf{\Lambda}_z$, respectively. The equality in (47) can be achieved if $\mathbf{U}_b = \mathbf{V}_{z,1}$, where $\mathbf{V}_{z,1}$ contains the first N_b columns of $\mathbf{V}_z$. The right-hand side (RHS) of (47) can be minimized by solving the problem below

$$\min_{\boldsymbol{\lambda}_b} \ \prod_{i=1}^{N_b}(\lambda_{b,i}^2\lambda_{z,i} + 1)^{-\frac{1}{N_b}} \tag{48a}$$

$$\text{s.t. } \sum_{i=1}^{N_b}\lambda_{b,i}^2 \le P_s \tag{48b}$$

where $\boldsymbol{\lambda}_b = [\lambda_{b,1}, \ldots, \lambda_{b,N_b}]^{\mathrm{T}}$, and $(\cdot)^{\mathrm{T}}$ is the matrix and vector transpose. The problem (48) has the water-filling solution given by

$$\lambda_{b,i}^2 = \left(\eta_1 - \frac{1}{\lambda_{z_i}}\right)^+, \qquad i = 1, \cdots, N_b \tag{49}$$

where $(x)^+ = \max(x, 0)$ and $\eta_1 > 0$ is the Lagrange multiplier which is computed to satisfy equality in the constraint (48b).

In (46), $\frac{1}{N_b}\mathrm{tr}(\mathbf{E}_1)$ achieves its lower bound $|\mathbf{E}_1|^{\frac{1}{N_b}}$ if and only if

$$\mathbf{E}_1 = \alpha\mathbf{I}_{N_b} \tag{50}$$

where α is the minimum value of the objective function (48a). By substituting $\mathbf{B} = \mathbf{V}_{z,1}\mathbf{\Lambda}_b\mathbf{V}_b^{\mathrm{H}}$ into (50), we have

$$\mathbf{U}(\mathbf{V}_b\mathbf{\Lambda}_b^2\mathbf{\Lambda}_{z,1}\mathbf{V}_b^{\mathrm{H}} + \mathbf{I}_{N_b})^{-1}\mathbf{U}^{\mathrm{H}} = \alpha\mathbf{I}_{N_b} \tag{51}$$

where $\mathbf{\Lambda}_{z,1}$ is a diagonal matrix containing the largest N_b eigenvalues of $\mathbf{Z}$. Similar to [20], $\mathbf{U}$ and $\mathbf{V}_b$ satisfying (51)

can be obtained from the geometric mean decomposition (GMD) [28] of $\widetilde{\mathbf{\Lambda}} = \alpha(\mathbf{\Lambda}_b^2\mathbf{\Lambda}_{z,1} + \mathbf{I}_{N_b})$ as

$$\widetilde{\mathbf{\Lambda}}^{\frac{1}{2}} = \mathbf{\Psi}^{\mathrm{H}}\mathbf{U}\mathbf{V}_b \tag{52}$$

where $\mathbf{\Psi}$ is an $N_b \times N_b$ unitary matrix. The feedback matrix $\mathbf{P}$ is obtained from $\mathbf{U}$ as $\mathbf{P} = \mathbf{U} - \mathbf{I}_{N_b}$.

Then, we consider solving the problem (45) with $\mathbf{B}$ and $\mathbf{P}$ derived above. We rewrite $\mathbf{E}_2$ as

$$\mathbf{E}_2 = (\overline{\mathbf{D}}^{\mathrm{H}}\mathbf{G}^{\mathrm{H}}\mathbf{G}\overline{\mathbf{D}} + \mathbf{\Phi}^{-1})^{-1} \tag{53}$$

where $\overline{\mathbf{D}} = \mathbf{D}\mathbf{U}^{-1}$ and $\mathbf{\Phi} = \mathbf{U}\mathbf{\Omega}\mathbf{U}^{\mathrm{H}}$. Let us introduce the EVD of $\mathbf{\Phi} = \mathbf{U}_\Phi\mathbf{\Lambda}_\Phi\mathbf{U}_\Phi^{\mathrm{H}}$ and the SVDs of $\overline{\mathbf{D}} = \mathbf{U}_{\overline{d}}\mathbf{\Lambda}_{\overline{d}}\mathbf{V}_{\overline{d}}^{\mathrm{H}}$ and $\mathbf{G} = \mathbf{U}_g\mathbf{\Lambda}_g\mathbf{V}_g^{\mathrm{H}}$, where $\mathbf{\Lambda}_\Phi$, $\mathbf{\Lambda}_{\overline{d}}$, and $\mathbf{\Lambda}_g$ are $N_b \times N_b$, $N_b \times N_b$, and $N_d \times N_r$ diagonal matrices whose main diagonal elements are sorted in descending order. From (53), we can rewrite the objective function (45a) as

$$\mathrm{tr}(\mathbf{E}_2) = \mathrm{tr}((\mathbf{U}_\Phi^{\mathrm{H}}\overline{\mathbf{D}}^{\mathrm{H}}\mathbf{G}^{\mathrm{H}}\mathbf{G}\overline{\mathbf{D}}\mathbf{U}_\Phi + \mathbf{\Lambda}_\Phi^{-1})^{-1}). \tag{54}$$

Let us define $\mathbf{\Gamma} = \mathbf{U}\mathbf{L}(\mathbf{H}\mathbf{B}\mathbf{B}^{\mathrm{H}}\mathbf{H}^{\mathrm{H}} + \mathbf{I}_{N_r})\mathbf{L}^{\mathrm{H}}\mathbf{U}^{\mathrm{H}}$. The constraint (45b) can be rewritten as

$$\mathrm{tr}(\overline{\mathbf{D}}\mathbf{\Gamma}\overline{\mathbf{D}}^{\mathrm{H}}) \le P_r. \tag{55}$$

By investigating (54) and (55), we find that in general there does not exist an optimal $\overline{\mathbf{D}}$ which simultaneously diagonalizes the objective function (54) and the constraint (55). Similar to [27], we propose to apply a suboptimal structure of

$$\overline{\mathbf{D}} = \mathbf{V}_{g,1}\mathbf{\Lambda}_{\overline{d}}\mathbf{U}_\Phi^{\mathrm{H}} \tag{56}$$

where $\mathbf{V}_{g,1}$ contains the first N_b columns of $\mathbf{V}_g$. Based on (56), the problem (45) becomes

$$\min_{\boldsymbol{\lambda}_{\overline{d}}} \ \sum_{i=1}^{N_b} (\lambda_{\overline{d},i}^2\lambda_{g,i}^2 + \lambda_{\Phi,i}^{-1})^{-1} \tag{57a}$$

$$\text{s.t.} \ \sum_{i=1}^{N_b} \lambda_{\overline{d},i}^2\tau_i \le P_r \tag{57b}$$

where $\boldsymbol{\lambda}_{\overline{d}} = [\lambda_{\overline{d},1}, \ldots, \lambda_{\overline{d},N_b}]^{\mathrm{T}}$, $\lambda_{\overline{d},i}$, $\lambda_{g,i}$, $\lambda_{\Phi,i}$, and τ_i are the ith main diagonal elements of $\mathbf{\Lambda}_{\overline{d}}$, $\mathbf{\Lambda}_g$, $\mathbf{\Lambda}_\Phi$, and $\mathbf{U}_\Phi^{\mathrm{H}}\mathbf{\Gamma}\mathbf{U}_\Phi$, respectively. It can be shown that the problem (57) has a water-filling solution given by

$$\lambda_{\overline{d},i}^2 = \frac{1}{\lambda_{g,i}^2\lambda_{\Phi,i}}\left(\sqrt{\frac{\lambda_{g,i}^2\lambda_{\Phi,i}^2}{\eta_2\tau_i}} - 1\right)^+ \tag{58}$$

where $\eta_2 > 0$ is the Lagrange multiplier calculated to satisfy the constraint (57b) with equality.

Finally, $\mathbf{F}$ is obtained by $\mathbf{F} = \overline{\mathbf{D}}\mathbf{U}\mathbf{L}$. The proposed non-iterative transceiver design algorithm is summarized in Algorithm 2.

Algorithm 2 Non-Iterative Transceiver Optimization Algorithm

Input: $\mathbf{H}$, $\mathbf{G}$, $\mathbf{T}$, P_s, P_r.
Output: $\mathbf{B}$, $\mathbf{F}$, $\mathbf{W}$, $\mathbf{P}$.
1: Calculate $\mathbf{\Lambda}_b$ by (49).
2: Obtain $\mathbf{U}$ and $\mathbf{V}_b$ by employing the GMD of $\widetilde{\mathbf{\Lambda}}^{\frac{1}{2}}$ according to (52), and then obtain $\mathbf{B}$ and $\mathbf{P}$ as $\mathbf{B} = \mathbf{V}_{z,1}\mathbf{\Lambda}_b\mathbf{V}_b^{\mathrm{H}}$ and $\mathbf{P} = \mathbf{U} - \mathbf{I}_{N_b}$, respectively.
3: Compute $\mathbf{\Lambda}_{\overline{d}}$ by (58) and calculate $\overline{\mathbf{D}}$ by (56). Obtain $\mathbf{F}$ as $\mathbf{F} = \overline{\mathbf{D}}\mathbf{U}\mathbf{L}$.

IV. Proposed Robust Transceiver Design Algorithms

The discussions above are based on the exact knowledge of CSI. In this section, the practical case of CSI mismatch is considered. We show that the transceiver design problem under imperfect CSI can be converted to the problem with perfect CSI knowledge. We denote $\overline{\mathbf{H}}$, $\overline{\mathbf{G}}$, and $\overline{\mathbf{T}}$ as the estimated channel matrices of $\mathbf{H}$, $\mathbf{G}$, and $\mathbf{T}$, respectively. By introducing the well-known Gaussian-Kronecker model [14], the true channel matrices $\mathbf{H}$, $\mathbf{G}$, and $\mathbf{T}$ can be written as $\mathbf{H} = \overline{\mathbf{H}} + \mathbf{A}_{\mathbf{\Phi_H}}\mathbf{H}_{w_\mathbf{H}}\mathbf{A}_{\mathbf{\Theta_H}}^{\mathrm{H}}$, $\mathbf{G} = \overline{\mathbf{G}} + \mathbf{A}_{\mathbf{\Phi_G}}\mathbf{H}_{w_\mathbf{G}}\mathbf{A}_{\mathbf{\Theta_G}}^{\mathrm{H}}$, and $\mathbf{T} = \overline{\mathbf{T}} + \mathbf{A}_{\mathbf{\Phi_T}}\mathbf{H}_{w_\mathbf{T}}\mathbf{A}_{\mathbf{\Theta_T}}^{\mathrm{H}}$, respectively, where $\mathbf{A}_{\mathbf{\Phi}_i}\mathbf{A}_{\mathbf{\Phi}_i}^{\mathrm{H}} = \mathbf{\Phi}_i$, $\mathbf{A}_{\mathbf{\Theta}_i}\mathbf{A}_{\mathbf{\Theta}_i}^{\mathrm{H}} = \mathbf{\Theta}_i^{\mathrm{T}}$, $\mathbf{\Phi}_i$ and $\mathbf{\Theta}_i$, $i \in \{\mathbf{H}, \mathbf{G}, \mathbf{T}\}$, are the covariance matrix of channel estimation error seen from the receiver side and the transmitter side, respectively, and $\mathbf{H}_{w_i}$ (the unknown part in the CSI mismatch) is a Gaussian random matrix with i.i.d. zero mean and unit variance entries.

We consider the statistical expectation of $\mathbf{E}$ in (8) in this case, which is described by

$$\begin{aligned} E(\mathbf{E}) &= E((\mathbf{W}^{\mathrm{H}}\mathbf{M} - \mathbf{U})(\mathbf{W}^{\mathrm{H}}\mathbf{M} - \mathbf{U})^{\mathrm{H}} + \mathbf{W}^{\mathrm{H}}\mathbf{C}_v\mathbf{W}) \\ &= \mathbf{W}^{\mathrm{H}}E(\mathbf{M}\mathbf{M}^{\mathrm{H}} + \mathbf{C}_v)\mathbf{W} - \mathbf{U}E(\mathbf{M}^{\mathrm{H}})\mathbf{W} \\ &\quad -\mathbf{W}^{\mathrm{H}}E(\mathbf{M})\mathbf{U}^{\mathrm{H}} + \mathbf{U}\mathbf{U}^{\mathrm{H}}. \end{aligned} \tag{59}$$

By defining $\alpha_1 = \mathrm{tr}(\mathbf{B}\mathbf{B}^{\mathrm{H}}\mathbf{\Theta}_\mathbf{H}^{\mathrm{T}})$, $\beta = \mathrm{tr}(\mathbf{F}(\overline{\mathbf{H}}\mathbf{B}\mathbf{B}^{\mathrm{H}}\overline{\mathbf{H}}^{\mathrm{H}} + \alpha_1\mathbf{\Phi_H})\mathbf{F}^{\mathrm{H}}\mathbf{\Theta}_\mathbf{G}^{\mathrm{T}})$, $\alpha_2 = \beta + \mathrm{tr}(\mathbf{F}\mathbf{F}^{\mathrm{H}}\mathbf{\Theta}_\mathbf{G}^{\mathrm{T}})$, and $\alpha_3 = \mathrm{tr}(\mathbf{B}\mathbf{B}^{\mathrm{H}}\mathbf{\Theta}_\mathbf{T}^{\mathrm{T}})$, we can deduce that

$$E(\mathbf{M}\mathbf{M}^{\mathrm{H}} + \mathbf{C}_v) = \overline{\mathbf{M}}\,\overline{\mathbf{M}}^{\mathrm{H}} + \overline{\mathbf{C}}_v + \mathbf{R} \tag{60}$$

where

$$\overline{\mathbf{M}} = \begin{bmatrix} \overline{\mathbf{G}}\mathbf{F}\overline{\mathbf{H}}\mathbf{B} \\ \overline{\mathbf{T}}\mathbf{B} \end{bmatrix}$$

$$\overline{\mathbf{C}}_v = \begin{bmatrix} \overline{\mathbf{G}}\mathbf{F}\mathbf{F}^{\mathrm{H}}\overline{\mathbf{G}}^{\mathrm{H}} + \mathbf{I}_{N_d} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{N_d} \end{bmatrix}$$

$$\mathbf{R} = \begin{bmatrix} \alpha_1\overline{\mathbf{G}}\mathbf{F}\mathbf{\Phi_H}\mathbf{F}^{\mathrm{H}}\overline{\mathbf{G}}^{\mathrm{H}} + \alpha_2\mathbf{\Phi_G} & \mathbf{0} \\ \mathbf{0} & \alpha_3\mathbf{\Phi_T} \end{bmatrix}.$$

Note that (60) is obtained based on the fact [14] that $E(\mathbf{H}\mathbf{X}\mathbf{H}^{\mathrm{H}}) = \overline{\mathbf{H}}\mathbf{X}\overline{\mathbf{H}}^{\mathrm{H}} + \mathrm{tr}(\mathbf{X}\mathbf{\Theta}_\mathbf{H}^{\mathrm{T}})\mathbf{\Phi_H}$, and similarly for equations involving $\mathbf{G}$ and $\mathbf{T}$.

By introducing $\mathbf{P_H} = \alpha_1\mathbf{\Phi_H} + \mathbf{I}_{N_r}$, $\mathbf{P_G} = \alpha_2\mathbf{\Phi_G} + \mathbf{I}_{N_d}$, and $\mathbf{P_T} = \alpha_3\mathbf{\Phi_T} + \mathbf{I}_{N_d}$, (60) can be rewritten as (61), as shown at the bottom of the next page.

Since the true channel matrix $\mathbf{H}$ is unknown, the averaged transmission power at the relay node is considered, which is

$$\begin{aligned} &E(\mathrm{tr}(\mathbf{F}(\mathbf{H}\mathbf{B}\mathbf{B}^{\mathrm{H}}\mathbf{H}^{\mathrm{H}} + \mathbf{I}_{N_r})\mathbf{F}^{\mathrm{H}})) \\ &\qquad = \mathrm{tr}(\mathbf{F}(\overline{\mathbf{H}}\mathbf{B}\mathbf{B}^{\mathrm{H}}\overline{\mathbf{H}}^{\mathrm{H}} + \alpha_1\mathbf{\Phi_H} + \mathbf{I}_{N_r})\mathbf{F}^{\mathrm{H}}). \end{aligned} \tag{62}$$

Let us introduce $\widetilde{\mathbf{H}} = \mathbf{P}_\mathbf{H}^{-\frac{1}{2}}\overline{\mathbf{H}}$, $\widetilde{\mathbf{G}} = \mathbf{P}_\mathbf{G}^{-\frac{1}{2}}\overline{\mathbf{G}}$, $\widetilde{\mathbf{T}} = \mathbf{P}_\mathbf{T}^{-\frac{1}{2}}\overline{\mathbf{T}}$, $\widetilde{\mathbf{F}} = \mathbf{F}\mathbf{P}_\mathbf{H}^{\frac{1}{2}}$, $\widetilde{\mathbf{W}}_1^{\mathrm{H}} = \mathbf{W}_1^{\mathrm{H}}\mathbf{P}_\mathbf{G}^{\frac{1}{2}}$, and $\widetilde{\mathbf{W}}_2^{\mathrm{H}} = \mathbf{W}_2^{\mathrm{H}}\mathbf{P}_\mathbf{T}^{\frac{1}{2}}$, where $\mathbf{W}_1$ and

$\mathbf{W}_2$ contain the first and last N_d rows of $\mathbf{W}$, respectively. Combining (59) and (61), we get

$$E(\mathbf{E}) = (\widetilde{\mathbf{W}}^{\mathrm{H}}\widetilde{\mathbf{M}} - \mathbf{U})(\widetilde{\mathbf{W}}^{\mathrm{H}}\widetilde{\mathbf{M}} - \mathbf{U})^{\mathrm{H}} + \widetilde{\mathbf{W}}^{\mathrm{H}}\widetilde{\mathbf{C}}_v\widetilde{\mathbf{W}} \quad (63)$$

where

$$\widetilde{\mathbf{W}}^{\mathrm{H}} = [\,\widetilde{\mathbf{W}}_1^{\mathrm{H}}, \quad \widetilde{\mathbf{W}}_2^{\mathrm{H}}\,]$$
$$\widetilde{\mathbf{M}} = \begin{bmatrix} \widetilde{\mathbf{G}}\widetilde{\mathbf{F}}\widetilde{\mathbf{H}}\mathbf{B} \\ \widetilde{\mathbf{T}}\mathbf{B} \end{bmatrix}$$
$$\widetilde{\mathbf{C}}_v = \begin{bmatrix} \widetilde{\mathbf{G}}\widetilde{\mathbf{F}}\widetilde{\mathbf{F}}^{\mathrm{H}}\widetilde{\mathbf{G}}^{\mathrm{H}} + \mathbf{I}_{N_d} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{N_d} \end{bmatrix}.$$

The averaged transmission power at the relay node shown in (62) can be rewritten as $\mathrm{tr}(\widetilde{\mathbf{F}}(\widetilde{\mathbf{H}}\mathbf{B}\mathbf{B}^{\mathrm{H}}\widetilde{\mathbf{H}}^{\mathrm{H}} + \mathbf{I}_{N_r})\widetilde{\mathbf{F}}^{\mathrm{H}})$.

Thus, the optimization problem for robust transceiver design is described by

$$\min_{\mathbf{B},\widetilde{\mathbf{F}},\widetilde{\mathbf{W}},\mathbf{P}\in\mathcal{U}} \quad \mathrm{tr}(E(\mathbf{E})) \quad (64a)$$
$$\text{s.t. } \mathrm{tr}(\mathbf{B}\mathbf{B}^{\mathrm{H}}) \le P_s \quad (64b)$$
$$\mathrm{tr}(\widetilde{\mathbf{F}}(\widetilde{\mathbf{H}}\mathbf{B}\mathbf{B}^{\mathrm{H}}\widetilde{\mathbf{H}}^{\mathrm{H}} + \mathbf{I}_{N_r})\widetilde{\mathbf{F}}^{\mathrm{H}}) \le P_r. \quad (64c)$$

Comparing the problem (64) with the problem (11), we can easily find that these two problems are similar. Therefore, the proposed iterative and non-iterative transceiver design algorithms for solving the problem (11) can be directly applied to solve the problem (64) for robust transceiver design where we only need to replace $\mathbf{F}$, $\mathbf{W}$, $\mathbf{H}$, $\mathbf{G}$, and $\mathbf{T}$ with $\widetilde{\mathbf{F}}$, $\widetilde{\mathbf{W}}$, $\widetilde{\mathbf{H}}$, $\widetilde{\mathbf{G}}$, and $\widetilde{\mathbf{T}}$. Then we obtain the iterative and non-iterative algorithms for robust transceiver design.

V. Numerical Examples

We study the performance of the proposed transceiver design algorithms in this section by numerical simulations. In the simulations, the source-relay channel matrix $\mathbf{H}$, the relay-destination channel matrix $\mathbf{G}$, and the source-destination channel matrix $\mathbf{T}$ have complex Gaussian entries with zero mean and variances of σ_s^2/N_s, σ_r^2/N_r, and σ_d^2/N_s, respectively. We define $\mathrm{SNR_s} = \sigma_s^2 P_s/N_s$, $\mathrm{SNR_r} = \sigma_r^2 P_r/N_r$, and $\mathrm{SNR_d} = \sigma_d^2 P_s/N_s$ as the SNRs of the source-relay, relay-destination, and source-destination links, respectively. When the BER performance is studied, signals are modulated by QPSK constellations. The simulation results are averaged through 2000 Monte-Carlo runs.

The performance of the proposed iterative (ITA) and non-iterative (NITA) algorithms for transceiver design with the exact CSI are compared with the following three benchmarking approaches: i) The transceiver design for AF MIMO relay systems with a ZF-based DFE receiver and direct link [21] (denoted as ZF-DFE); ii) Joint MMSE transceiver design for AF MIMO relay systems with a linear receiver and direct link [16] (denoted as JMLD); iii) Joint MMSE transceiver design for AF MIMO relay systems with a DFE receiver where the direct link is ignored [20] (denoted as JMDND). Note that the ZF-DFE algorithm can only work when $N_s = N_b$. For the case of CSI mismatch, we compare the performance of the proposed iterative (ITA-R) and non-iterative (NITA-R) algorithms for robust transceiver design with the joint MMSE robust transceiver design algorithm in [16] (denoted as JMLD-R).

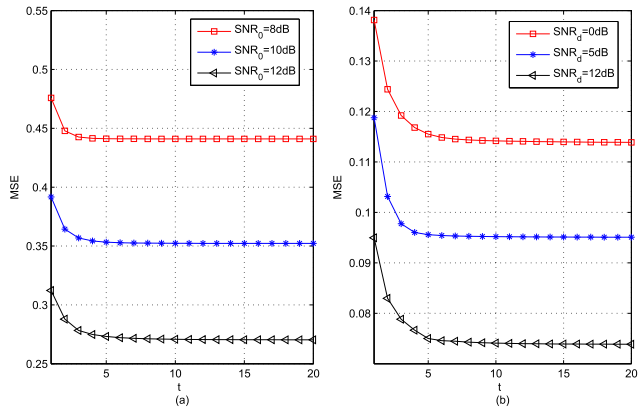


Fig. 2. Example 1: MSE of the ITA algorithm versus the number of iterations under two scenarios, where in (a) $\mathrm{SNR_s} = \mathrm{SNR_r} = \mathrm{SNR_0}$, $\mathrm{SNR_d} = \mathrm{SNR_s} - 10\mathrm{dB}$, $N_s = N_r = N_d = N_b = 3$, and $\mathrm{SNR_0} = 8\mathrm{dB}$, 10dB or 12dB, and in (b) $\mathrm{SNR_s} = 15\mathrm{dB}$, $\mathrm{SNR_r} = 10\mathrm{dB}$, $N_s = 4$, $N_r = 6$, $N_d = 3$, $N_b = 2$, and $\mathrm{SNR_d} = 0\mathrm{dB}$, 5dB or 12dB.

In the first simulation example, we investigate the convergence of the proposed ITA algorithm under two scenarios. We set $\mathrm{SNR_s} = \mathrm{SNR_r} = \mathrm{SNR_0}$, $\mathrm{SNR_d} = \mathrm{SNR_s} - 10\mathrm{dB}$, and $N_s = N_r = N_d = N_b = 3$ in the first scenario. In the second scenario, we set $\mathrm{SNR_s} = 15\mathrm{dB}$, $\mathrm{SNR_r} = 10\mathrm{dB}$, $N_s = 4$, $N_r = 6$, $N_d = 3$, and $N_b = 2$. Fig. 2(a) demonstrates the MSE of the proposed ITA algorithm versus the number of iterations (denoted as t) for $\mathrm{SNR_0} = 8\mathrm{dB}$, $\mathrm{SNR_0} = 10\mathrm{dB}$, and $\mathrm{SNR_0} = 12\mathrm{dB}$ under the first scenario. In Fig. 2(b), the MSE of the proposed ITA algorithm versus the number of iterations for $\mathrm{SNR_d} = 0\mathrm{dB}$, $\mathrm{SNR_d} = 5\mathrm{dB}$, and $\mathrm{SNR_d} = 12\mathrm{dB}$ is shown under the second scenario. It can be clearly seen from both Fig. 2(a) and Fig. 2(b) that the MSE of the proposed ITA algorithm converges in about 5 iterations for various scenarios.

In the second example, we set $\mathrm{SNR_s} = \mathrm{SNR_r} = \mathrm{SNR_0}$, $\mathrm{SNR_d} = \mathrm{SNR_s} - 10\mathrm{dB}$, and $N_s = N_r = N_d = N_b = 4$. The CSI mismatch is considered in this example, where for $i \in \{\mathbf{H}, \mathbf{G}, \mathbf{T}\}$, $\mathbf{\Theta}_i$ and $\mathbf{\Phi}_i$ are set as $\mathbf{\Theta}_i = \sigma_e^2\mathbf{I}_{N_b}$ and

$$\mathbf{\Phi}_i = \begin{bmatrix} 1 & \phi & \phi^2 & \phi^3 \\ \phi & 1 & \phi & \phi^2 \\ \phi^2 & \phi & 1 & \phi \\ \phi^3 & \phi^2 & \phi & 1 \end{bmatrix}.$$

In this example, we select $\phi = 0.45$. Fig. 3 shows the MSE of the proposed algorithms versus $\mathrm{SNR_0}$ with the exact CSI and imperfect CSI. We can see from Fig. 3 that for

$$E(\mathbf{M}\mathbf{M}^{\mathrm{H}} + \mathbf{C}_v) = \begin{bmatrix} \overline{\mathbf{G}}\mathbf{F}\overline{\mathbf{H}}\mathbf{B}\mathbf{B}^{\mathrm{H}}\overline{\mathbf{H}}^{\mathrm{H}}\mathbf{F}^{\mathrm{H}}\overline{\mathbf{G}}^{\mathrm{H}} + \overline{\mathbf{G}}\mathbf{F}\mathbf{P}_{\mathbf{H}}\mathbf{F}^{\mathrm{H}}\overline{\mathbf{G}}^{\mathrm{H}} + \mathbf{P}_{\mathbf{G}} & \overline{\mathbf{G}}\mathbf{F}\overline{\mathbf{H}}\mathbf{B}\mathbf{B}^{\mathrm{H}}\overline{\mathbf{T}}^{\mathrm{H}} \\ \overline{\mathbf{T}}\mathbf{B}\mathbf{B}^{\mathrm{H}}\overline{\mathbf{H}}^{\mathrm{H}}\mathbf{F}^{\mathrm{H}}\overline{\mathbf{G}}^{\mathrm{H}} & \overline{\mathbf{T}}\mathbf{B}\mathbf{B}^{\mathrm{H}}\overline{\mathbf{T}}^{\mathrm{H}} + \mathbf{P}_{\mathbf{T}} \end{bmatrix} \quad (61)$$

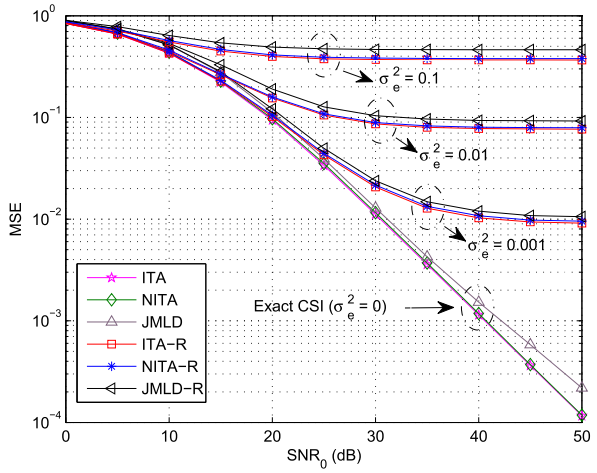


Fig. 3. Example 2: MSE versus SNR_0 with $\mathrm{SNR}_\mathrm{s} = \mathrm{SNR}_\mathrm{r} = \mathrm{SNR}_0$, $\mathrm{SNR}_\mathrm{d} = \mathrm{SNR}_\mathrm{s} - 10\mathrm{dB}$, and $N_s = N_r = N_d = N_b = 4$.

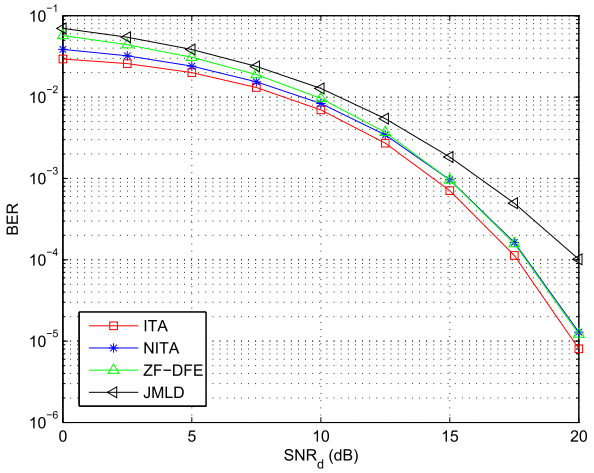


Fig. 5. Example 3: BER versus SNR_d with $\mathrm{SNR}_\mathrm{s} = 18\mathrm{dB}$, $\mathrm{SNR}_\mathrm{r} = 15\mathrm{dB}$, and $N_s = N_r = N_d = N_b = 4$.

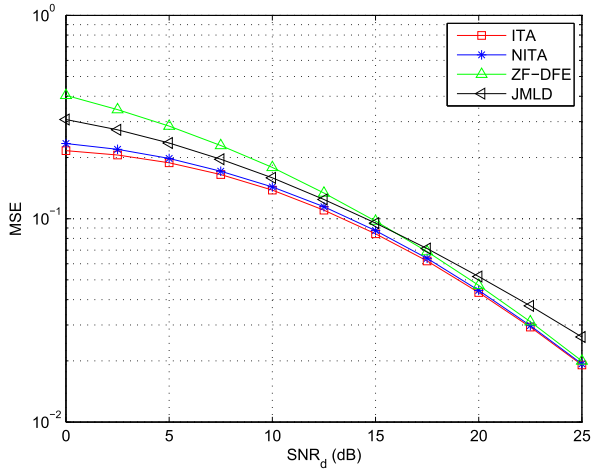


Fig. 4. Example 3: MSE versus SNR_d with $\mathrm{SNR}_\mathrm{s} = 18\mathrm{dB}$, $\mathrm{SNR}_\mathrm{r} = 15\mathrm{dB}$, and $N_s = N_r = N_d = N_b = 4$.

both the exact and imperfect CSI cases, the proposed NITA and ITA algorithms perform almost same and both of them have a better MSE performance than the JMLD algorithm. That is to say, the DFE receiver shows a better performance than the linear receiver. As expected, the MSEs of the three tested algorithms increase with σ_e^2. In the following numerical examples, we focus on the exact CSI case.

In the third example, we study the MSE and BER performance of the proposed algorithms versus SNR_d with fixed SNR_s and SNR_r. Fig. 4 displays the MSE performance of the four tested algorithms versus SNR_d with $\mathrm{SNR}_\mathrm{s} = 18\mathrm{dB}$, $\mathrm{SNR}_\mathrm{r} = 15\mathrm{dB}$, and $N_s = N_r = N_d = N_b = 4$. From Fig. 4, we can see that the MSE performance of the proposed ITA and NITA algorithms is better than that of the ZF-DFE and JMLD algorithms with different SNR_d. Moreover, the NITA algorithm has a slightly higher MSE than the ITA algorithm throughout the range of SNR_d, since the former one is a suboptimal method. In Fig. 5, we show the BER performance of the four tested algorithms versus SNR_d with the same configuration. Similar to Fig. 4, we can see from Fig. 5 that the proposed ITA and NITA algorithms have a better BER performance than the ZF-DFE and JMLD algorithms. From Figs. 4 and 5 we can observe that the performance of the ZF-DFE algorithm is good at high SNR, while it deteriorates at low SNR. The reason is that the ZF-DFE algorithm takes use of the ZF strategy, which has poor performance at low SNR, due to the effect of noise enhancement of the ZF receiver.

In the next example, we fix SNR_s (or SNR_r) and investigate the impact of SNR_r (or SNR_s) on the performance of the proposed algorithms. We set $\mathrm{SNR}_\mathrm{d} = \mathrm{SNR}_\mathrm{s} - 10\mathrm{dB}$ and $N_s = N_r = N_d = N_b = 3$. Fig. 6 illustrates the MSE performance of the five algorithms versus SNR_r with $\mathrm{SNR}_\mathrm{s} = 20\mathrm{dB}$. It can be observed from Fig. 6 that the ITA algorithm yields the lowest MSE among the five approaches over the entire SNR_r range, and the NITA algorithm has a lower MSE than the JMDND and ZF-DFE algorithms. Note that the JMDND algorithm does not exploit the direct link, so it usually performs worse than the other algorithms. When $\mathrm{SNR}_\mathrm{r} > 15\mathrm{dB}$, the MSE of the NITA algorithm is close to that of the ITA algorithm, while the former deteriorates and is even worse than that of the JMLD algorithm when $\mathrm{SNR}_\mathrm{r} < 15\mathrm{dB}$. The reasons for the worse performance of the NITA algorithm at low SNR_r are explained below. As mentioned in Section III-B, the NITA algorithm obtains $\mathbf{B}$ and $\mathbf{P}$ by solving the problem (44). This may result in suboptimal $\mathbf{B}$ and $\mathbf{P}$ since $\mathbf{E}_2$ contains $\mathbf{B}$ and $\mathbf{P}$, and the constraint (45b) is also related to $\mathbf{B}$. The impact of such sub-optimality increases as the value of P_s/P_r grows.

For this simulation example, we display the MSE of the five algorithms versus SNR_s with $\mathrm{SNR}_\mathrm{r} = 12\mathrm{dB}$ in Fig. 7. From Fig. 7, we can see that the ITA algorithm has a lower MSE compared with the other four algorithms throughout the entire range of SNR_s. Additionally, the MSE performance of the NITA algorithm is better than that of the ZF-DFE, JMLD, and JMDND algorithms when $\mathrm{SNR}_\mathrm{s} \leq 15\mathrm{dB}$, and is worse than that of the ZF-DFE and JMLD algorithms when $\mathrm{SNR}_\mathrm{s} \geq 25\mathrm{dB}$, which corroborates the conclusion above in Fig. 6. Moreover, we can find that there is a switching point

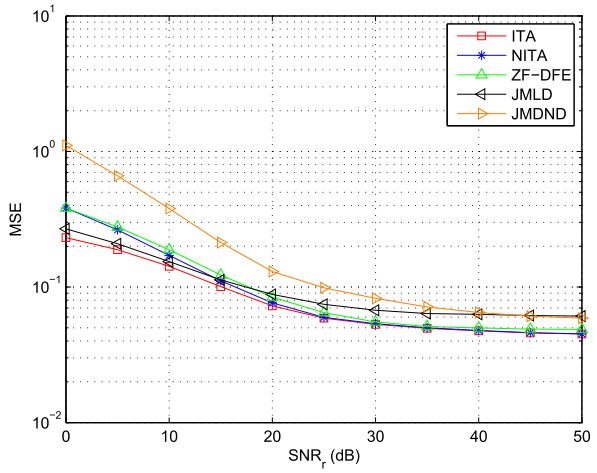


Fig. 6. Example 4: MSE versus $\mathrm{SNR_r}$ with $\mathrm{SNR_s} = 20\mathrm{dB}$, $\mathrm{SNR_d} = 10\mathrm{dB}$, and $N_s = N_r = N_d = N_b = 3$.

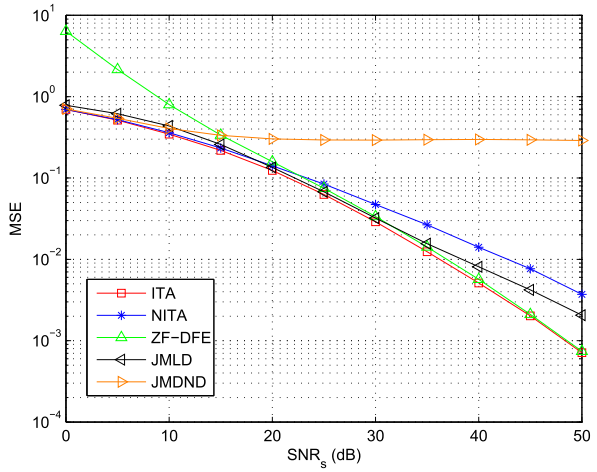


Fig. 7. Example 4: MSE versus $\mathrm{SNR_s}$ with $\mathrm{SNR_r} = 12\mathrm{dB}$, $\mathrm{SNR_d} = \mathrm{SNR_s} - 10\mathrm{dB}$, and $N_s = N_r = N_d = N_b = 3$.

(at $\mathrm{SNR_s} = 15\mathrm{dB}$) between the ZF-DFE and JMDND algorithms. The reason for the occurrence of the switching point is that the ZF-DFE algorithm has poor performance at low SNR, which has been explained above, so it may perform worse than the JMDND algorithm at low SNR. Since the JMDND algorithm does not exploit the direct link, the superiority of the ZF-DFE algorithm which considers the direct link appears at high SNR. Therefore, the switching point occurs and it may appear in the following examples between the ZF-DFE and JMDND algorithms. Fig. 8 shows the BER performance of the five algorithms versus $\mathrm{SNR_s}$ with $\mathrm{SNR_r} = 12\mathrm{dB}$, where we obtain observations similar to those in Fig. 7. Based on Figs. 6-8, we can conclude that when P_s/P_r is small (i.e., $P_s/P_r < 2$), the NITA algorithm has a negligible performance loss compared with the ITA algorithms, and performs better than the other benchmarking algorithms.

In the fifth example, we set $\mathrm{SNR_s} = \mathrm{SNR_r} = \mathrm{SNR_0}$ and $\mathrm{SNR_d} = \mathrm{SNR_s} - 10\mathrm{dB}$. Fig. 9 illustrates the MSE performance of the tested algorithms versus $\mathrm{SNR_0}$ with different number of antennas at three nodes. It can be seen from

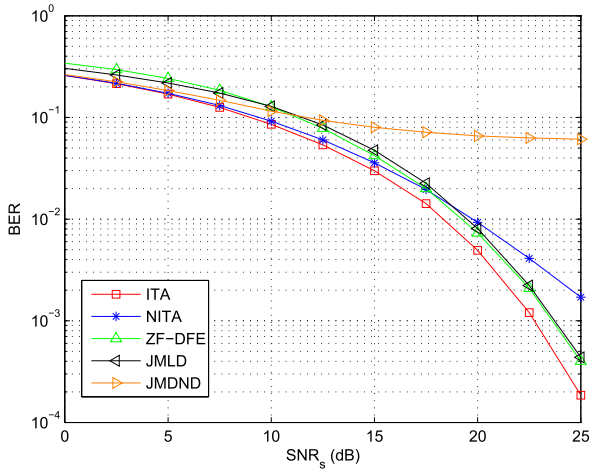


Fig. 8. Example 4: BER versus $\mathrm{SNR_s}$ with $\mathrm{SNR_r} = 12\mathrm{dB}$, $\mathrm{SNR_d} = \mathrm{SNR_s} - 10\mathrm{dB}$, and $N_s = N_r = N_d = N_b = 3$.

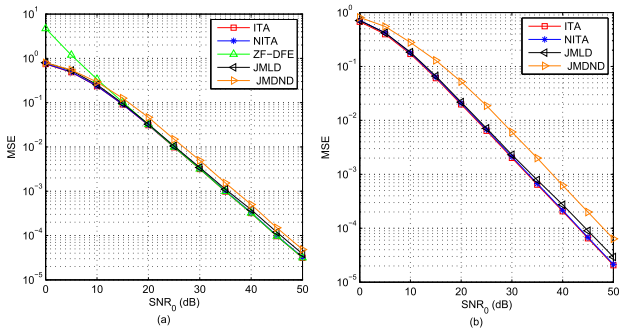


Fig. 9. Example 5: MSE versus $\mathrm{SNR_0}$ with $\mathrm{SNR_s} = \mathrm{SNR_r} = \mathrm{SNR_0}$ and $\mathrm{SNR_d} = \mathrm{SNR_s} - 10\mathrm{dB}$, where in (a) $N_b = 2, N_s = 2, N_r = 4, N_d = 6$, and in (b) $N_b = 2, N_s = 4, N_r = 2, N_d = 6$.

Fig. 9 that the ITA algorithm possesses the best performance in all tested algorithms with various antenna configurations. Moreover, the NITA algorithm has almost the same MSE performance as the ITA algorithm and has better performance than the other benchmarking approaches. The performance of the ZF-DFE algorithm is not shown in Fig. 9(b) since this algorithm does not work when $N_b \neq N_s$.

In the sixth numerical example, we set $N_s = N_r = N_d = N_b = 4$ and study the performance of the proposed algorithms with $\mathrm{SNR_r} = 2\mathrm{SNR_s} = \mathrm{SNR_0}$ and $\mathrm{SNR_d} = \mathrm{SNR_s} - 10\mathrm{dB}$. The MSE and BER performance of the five tested algorithms versus $\mathrm{SNR_0}$ is shown in Figs. 10 and 11, respectively. It can be observed from Figs. 10 and 11 that the proposed ITA and NITA algorithms almost have identical MSE and BER performance, and they both yield lower MSEs and BERs than the ZF-DFE, JMLD, and JMDND algorithms. It is obvious that the results in this example corroborate the conclusion drawn in example four.

Finally, we set $N_s = N_r = N_d = N_b = N$ and compare the computational complexity of the proposed algorithms with the ZF-DFE, JMLD (JMLD-R), and JMDND methods. since the NITA and ZF-DFE algorithms mainly involve matrix SVD/EVD and matrix inversion, their computational complexity order is $\mathcal{O}(N^3)$. For each iteration of the ITA algorithm,

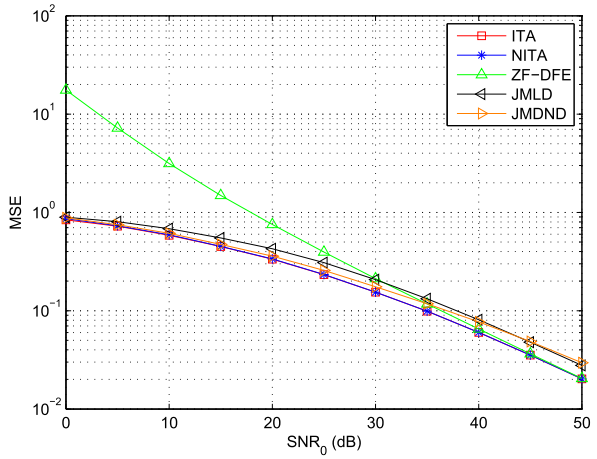


Fig. 10. Example 6: MSE versus SNR_0 with $\mathrm{SNR}_\mathrm{r} = 2\mathrm{SNR}_\mathrm{s} = \mathrm{SNR}_0$, $\mathrm{SNR}_\mathrm{d} = \mathrm{SNR}_\mathrm{s} - 10\mathrm{dB}$, and $N_s = N_r = N_d = N_b = 4$.

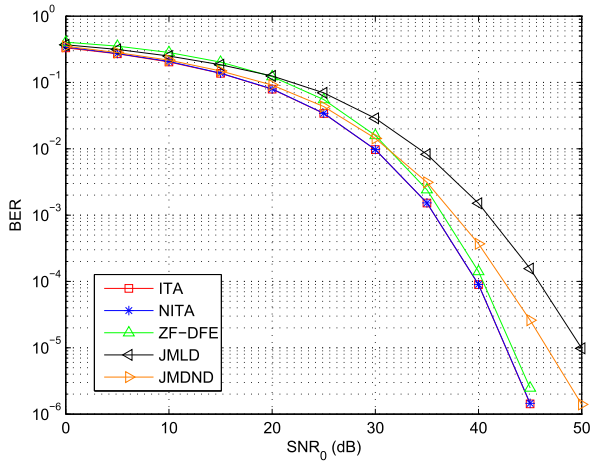


Fig. 11. Example 6: BER versus SNR_0 with $\mathrm{SNR}_\mathrm{r} = 2\mathrm{SNR}_\mathrm{s} = \mathrm{SNR}_0$, $\mathrm{SNR}_\mathrm{d} = \mathrm{SNR}_\mathrm{s} - 10\mathrm{dB}$, and $N_s = N_r = N_d = N_b = 4$.

matrix QR decomposition, matrix SVD and matrix inversion are employed. Thus, the ITA algorithm has an overall complexity order of $\mathcal{O}(tN^3)$, where t is the number of iterations till convergence (usually around five according to example one). From [16] and [20], it can be found that the JMLD (JMLD-R) and JMDND algorithms possess the overall complexity order of $\mathcal{O}(t'N^3)$ and $\mathcal{O}(N^3)$, respectively, where t' refers to the number of iterations and usually equals to three according to [16]. Therefore, the ITA (ITA-R) and JMLD (JMLD-R) algorithms almost have the same complexity and the former is slightly higher than the latter. In addition, the computational complexity of the ITA (ITA-R) algorithm is higher than that of the NITA (NITA-R), ZF-DFE, and JMDND algorithms. Considering that the performance of the ITA approach is better, such performance-complexity tradeoff is interesting in practical AF MIMO relay systems.

VI. Conclusion

We have investigated the optimization of precoding matrices for a dual-hop AF MIMO relay system with a DFE receiver and direct link. An iterative algorithm and a non-iterative suboptimal approach have been developed to design the transceiver matrices with exact and imperfect CSI based on the MMSE criterion. Simulation results have shown that the proposed iterative design has a better performance than existing approaches under various scenarios. When the ratio of the source node transmission power to the relay node transmission power is small (typically less than two), the proposed non-iterative suboptimal algorithm has a negligible performance loss compared with the iterative algorithm.

Appendix

We calculate the first-order derivative of $f(\widetilde{\mu}_1)$ with respect to $\widetilde{\mu}_1$ as

$$\begin{aligned}
&\frac{\partial f(\widetilde{\mu}_1)}{\partial \widetilde{\mu}_1} \\
&= \frac{\partial \mathrm{tr}(\mathbf{Q}\mathbf{B}_0^{-1}(\widetilde{\mu}_1)\mathbf{D}_1^{\mathrm{H}}\mathbf{U}\mathbf{U}^{\mathrm{H}}\mathbf{D}_1\mathbf{B}_0^{-1}(\widetilde{\mu}_1))}{\partial \widetilde{\mu}_1} \\
&= -\mathrm{tr}(\mathbf{Q}\mathbf{B}_0^{-1}(\widetilde{\mu}_1)\mathbf{Q}\mathbf{B}_0^{-1}(\widetilde{\mu}_1)\mathbf{D}_1^{\mathrm{H}}\mathbf{U}\mathbf{U}^{\mathrm{H}}\mathbf{D}_1\mathbf{B}_0^{-1}(\widetilde{\mu}_1)) \\
&\quad -\mathrm{tr}(\mathbf{Q}\mathbf{B}_0^{-1}(\widetilde{\mu}_1)\mathbf{D}_1^{\mathrm{H}}\mathbf{U}\mathbf{U}^{\mathrm{H}}\mathbf{D}_1\mathbf{B}_0^{-1}(\widetilde{\mu}_1)\mathbf{Q}\mathbf{B}_0^{-1}(\widetilde{\mu}_1)) \\
&= -2\mathrm{tr}(\mathbf{Q}\mathbf{B}_0^{-1}(\widetilde{\mu}_1)\mathbf{D}_1^{\mathrm{H}}\mathbf{U}\mathbf{U}^{\mathrm{H}}\mathbf{D}_1\mathbf{B}_0^{-1}(\widetilde{\mu}_1)\mathbf{Q}\mathbf{B}_0^{-1}(\widetilde{\mu}_1)). \quad (65)
\end{aligned}$$

By introducing the eigenvalue decomposition (EVD) of $\mathbf{B}_0^{-1}(\widetilde{\mu}_1) = \mathbf{U}_{\mathbf{B}_0^{-1}}\mathbf{\Lambda}_{\mathbf{B}_0^{-1}}\mathbf{U}_{\mathbf{B}_0^{-1}}^{\mathrm{H}}$, where $\mathbf{\Lambda}_{\mathbf{B}_0^{-1}}$ is an $N_s \times N_s$ diagonal matrix, (65) is rewritten as

$$\begin{aligned}
\frac{\partial f(\widetilde{\mu}_1)}{\partial \widetilde{\mu}_1} &= -2\mathrm{tr}\left(\mathbf{\Gamma}\mathbf{\Gamma}^{\mathrm{H}}\mathbf{U}_{\mathbf{B}_0^{-1}}\mathbf{\Lambda}_{\mathbf{B}_0^{-1}}\mathbf{U}_{\mathbf{B}_0^{-1}}^{\mathrm{H}}\right) \\
&= -2\mathrm{tr}\left(\mathbf{\Lambda}_{\mathbf{B}_0^{-1}}^{-1/2}\mathbf{U}_{\mathbf{B}_0^{-1}}^{\mathrm{H}}\mathbf{\Gamma}\mathbf{\Gamma}^{\mathrm{H}}\mathbf{U}_{\mathbf{B}_0^{-1}}\mathbf{\Lambda}_{\mathbf{B}_0^{-1}}^{-1/2}\right) \quad (66)
\end{aligned}$$

where $\mathbf{\Gamma} = \mathbf{Q}\mathbf{B}_0^{-1}(\widetilde{\mu}_1)\mathbf{D}_1^{\mathrm{H}}\mathbf{U}$. From (66), we have $\frac{\partial f(\widetilde{\mu}_1)}{\partial \widetilde{\mu}_1} \leq 0$. Thus, $f(\widetilde{\mu}_1)$ is a non-increasing function with respect to $\widetilde{\mu}_1$.

Acknowledgment

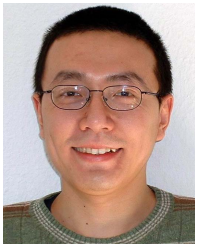
The authors would like to thank an editor and anonymous reviewers for their valuable comments and suggestions that helped improve the quality of the article.

References

[1] L. Sanguinetti, A. A. D'Amico, and Y. Rong, "A tutorial on the optimization of amplify- and-forward MIMO relay systems," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 8, pp. 1331–1346, Sep. 2012.
[2] J. N. Laneman, D. N. C. Tse, and G. W. Wornell, "Cooperative diversity in wireless networks: Efficient protocols and outage behavior," *IEEE Trans. Inf. Theory*, vol. 50, no. 12, pp. 3062–3080, Dec. 2004.
[3] 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.
[4] 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, Jun. 2007.
[5] 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.
[6] Y. Rong, "Linear non-regenerative multicarrier MIMO relay communications based on MMSE criterion," *IEEE Trans. Commun.*, vol. 58, no. 7, pp. 1918–1923, Jul. 2010.
[7] 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.

[8] F. Tseng, M. Chang, and W. Wu, "Robust Tomlinson-Harashima source and linear relay precoders design in amplify-and-forward MIMO relay systems," *IEEE Trans. Commun.*, vol. 60, no. 4, pp. 1124–1137, Apr. 2012.
[9] C. Xing, S. Ma, Z. Fei, Y.-C. Wu, and H. V. Poor, "A general robust linear transceiver design for multi-hop amplify- and-forward MIMO relaying systems," *IEEE Trans. Signal Process.*, vol. 61, no. 5, pp. 1196–1209, Mar. 2013.
[10] L. Sanguinetti, A. A. D'Amico, and Y. Rong, "On the design of amplify-and-forward MIMO-OFDM relay systems with QoS requirements specified as Schur-convex functions of the MSEs," *IEEE Trans. Veh. Technol.*, vol. 62, no. 4, pp. 1871–1877, May 2013.
[11] Y. Fu, L. Yang, W.-P. Zhu, and C. Liu, "Optimum linear design of two-hop MIMO relay networks with QoS requirements," *IEEE Trans. Signal Process.*, vol. 59, no. 5, pp. 2257–2269, May 2011.
[12] L. Sanguinetti and A. A. D'Amico, "Power allocation in two-hop amplify-and-forward MIMO relay systems with QoS requirements," *IEEE Trans. Signal Process.*, vol. 60, no. 5, pp. 2494–2507, May 2012.
[13] Y. Rong, "Optimal joint source and relay beamforming for MIMO relays with direct link," *IEEE Commun. Lett.*, vol. 14, no. 5, pp. 390–392, May 2010.
[14] Z. He, W. Jiang, and Y. Rong, "Robust design for amplify-and-forward MIMO relay systems with direct link and imperfect channel information," *IEEE Trans. Wireless Commun.*, vol. 14, no. 1, pp. 353–363, Jan. 2015.
[15] Z. He, J. Zhang, W. Liu, and Y. Rong, "New results on transceiver design for two-hop amplify- and-forward MIMO relay systems with direct link," *IEEE Trans. Signal Process.*, vol. 64, no. 20, pp. 5232–5241, Oct. 2016.
[16] H.-B. Kong, H. M. Shin, T. Oh, and I. Lee, "Joint MMSE transceiver designs for MIMO AF relaying systems with direct link," *IEEE Trans. Wireless Commun.*, vol. 16, no. 6, pp. 3547–3560, Jun. 2017.
[17] I. Lee and J. M. Cioffi, "A fast computation algorithm for the decision feedback equalizer," *IEEE Trans. Commun.*, vol. 43, no. 11, pp. 2742–2749, Nov. 1995.
[18] G. J. Foschini, G. D. Golden, R. A. Valenzuela, and P. W. Wolniansky, "Simplified processing for high spectral efficiency wireless communication employing multi-element arrays," *IEEE J. Sel. Areas Commun.*, vol. 17, no. 11, pp. 1841–1852, Nov. 1999.
[19] Y. Rong, "Optimal linear non-regenerative multi-hop MIMO relays with MMSE-DFE receiver at the destination," *IEEE Trans. Wireless Commun.*, vol. 9, no. 7, pp. 2268–2279, Jul. 2010.
[20] M. Ahn, H.-B. Kong, T. Kim, C. Song, and I. Lee, "Precoding techniques for MIMO AF relaying systems with decision feedback receiver," *IEEE Trans. Wireless Commun.*, vol. 14, no. 1, pp. 446–455, Jan. 2015.
[21] A. P. Millar, S. Weiss, and R. W. Stewart, "Precoder design for MIMO relay networks with direct link and decision feedback equalisation," *IEEE Commun. Lett.*, vol. 15, no. 10, pp. 1044–1046, Oct. 2011.
[22] Y. Xu and W. Yin, "A block coordinate descent method for regularized multiconvex optimization with applications to nonnegative tensor factorization and completion," *SIAM J. Imag. Sci.*, vol. 6, no. 3, pp. 1758–1789, Jan. 2013.
[23] Y. Rong, M. R. A. Khandaker, and Y. Xiang, "Channel estimation of dual-hop MIMO relay system via parallel factor analysis," *IEEE Trans. Wireless Commun.*, vol. 11, no. 6, pp. 2224–2233, Jun. 2012.
[24] R. A. Horn and C. R. Johson, *Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1985.
[25] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication* Cambridge, U.K.: Cambridge Univ. Press, 2005.
[26] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.
[27] C. Song, K.-J. Lee, and I. Lee, "MMSE-based MIMO cooperative relaying systems: Closed-form designs and outage behavior," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 8, pp. 1390–1401, Sep. 2012.
[28] 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.

Qiao Su received the Ph.D. degree from the College of Communications Engineering, Army Engineering University of PLA, in 2019. Since July 2019, he has been with Army Engineering University of PLA, where he is currently a Lecturer. His research interests include wireless communications, blind signal processing, and statistical and array signal processing.



Yue Rong (Senior Member, IEEE) received the Ph.D. degree *(summa cum laude)* in electrical engineering from the Darmstadt University of Technology, Darmstadt, Germany, in 2005.

He was a Post-Doctoral Researcher with the Department of Electrical Engineering, University of California at Riverside, Riverside, CA, USA, from February 2006 to November 2007. Since December 2007, he has been with Curtin University, Bentley, Australia, where he is currently a Professor. His research interests include signal processing for communications, wireless communications, underwater acoustic communications, underwater optical wireless communications, applications of linear algebra and optimization methods, and statistical and array signal processing. He has published over 190 journals and conference papers in these areas. He was a TPC Member of the IEEE ICC, IEEE GlobalSIP, EUSIPCO, IEEE ICCC, WCSP, IWCMC, and ChinaCom. He was a recipient of the Best Paper Award from the 2011 International Conference on Wireless Communications and Signal Processing, the Best Paper Award from the 2010 Asia–Pacific Conference on Communications, and the Young Researcher from the Year Award of the Faculty of Science and Engineering at Curtin University in 2010. He was an Associate Editor of the IEEE TRANSACTIONS ON SIGNAL PROCESSING from 2014 to 2018, an Editor of the IEEE WIRELESS COMMUNICATIONS LETTERS from 2012 to 2014, and a Guest Editor of the *IEEE Journal on Selected Areas in Communications* special issue on theories and methods for advanced wireless relays. He is a Senior Area Editor of the IEEE TRANSACTIONS ON SIGNAL PROCESSING.