Two-Way AF MIMO Multi-Relay System Design Using MMSE-DFE Techniques

Yang Lv, *Graduate Student Member, IEEE*, Zhiqiang He, *Member, IEEE*, and Yue Rong, *Senior Member, IEEE*

***Abstract*— Targeting at a better design of the analogue network coding (ANC)-assisted two-way amplify-and-forward (AF) multiple-input multiple-output (MIMO) multi-relay communication systems, we bring in the nonlinear minimal mean-squared error (MMSE)-decision feedback equalization (DFE) receiving technique to jointly optimize the source precoding, relay amplifying, feed-forward and feedback matrices. Under the transmission power constraints at both source nodes and each relay node, the two-way sum mean-squared error (MSE) of the signal waveform estimation of all data streams is minimized. To solve the complicated nonconvex optimization problem with four groups of system parameters, this paper develops an iterative block coordinate descent (BCD) algorithm, which converges to at least a Nash point. On the basis of it, for mitigating the error propagation in MMSE-DFE receivers, a group of permutation matrix variables, determining the detection orders of all data streams, are further introduced in our system optimization. Moreover, in case there is no sufficiently precise channel state information (CSI), we also make an extension of the developed algorithms, yielding a robust design scheme, to handle the channel uncertainties. Numerical simulation results show that, compared with the existing linear MMSE receiving-based algorithm, our proposed nonlinear ones provide improved MSE and bit-error-rate (BER) performance as well as good robustness against the imperfect CSI, indicating a promising application prospect of this research.**

***Index Terms*— Two-way, ANC, MIMO relay, AF, multi-relay, MMSE, DFE, error propagation, imperfect CSI, robust design.**

I. Introduction

OVER the past decade, the multiple-input multiple-output (MIMO) relay communications have been widely studied and valued for their advantages of being able to expand the system coverage, improve the link reliability as well as reduce the power consumption [1]–[3]. Certainly, they will have a significant impact on the future 5G wireless networks, especially on Internet of Things (IoT) and device-to-device (D2D) communications [4], [5]. In general, there are three common relay protocols, including amplify-and-forward (AF), decode-and-forward (DF) and compress-and-forward (CF) [6], [7], among which the AF protocol attracts more extensive attention due to its simplicity and high speed [3].

For a three-terminal MIMO relay system, the upper and lower bounds on its ergodic capacity over Rayleigh fading channels were studied in [8]. When half-duplex (HD) strategy was adopted, [9] designed the optimal AF MIMO relay matrix to maximize the system capacity between source and destination nodes in the absence of a direct link. Considering the multi-hop scenarios, [10] established the optimal structures of source and relay matrices, which jointly diagonalize the MIMO relay channel into a set of parallel single-input single-output (SISO) ones, and the generalized results for multicarrier transmission were given in [11]. With imperfect channel state information (CSI) taken into account, [12] explored the robust joint optimization of the linear relay precoder and destination equalizer in a dual-hop MIMO relay system. References [13] and [14] investigated the robust transceiver design for, respectively, interference and multicasting relay communications. Reference [15] proposed robust algorithms to optimize multiuser relay systems with direct source-destination links. Also considering direct links in three-terminal AF MIMO relay systems, [16] developed an iterative optimization algorithm on the basis of the optimal beamforming structures of source and relay matrices, [17] derived the optimal source beamformer as a semi-closed form solution with the known results of the optimal receiver and relay precoder by using the semidefinite relaxation approach.

Applying the principle of analogue network coding (ANC), a two-way HD relay communication system allows two source nodes to simultaneously exchange information through assistant relay node(s) in just two time slots, which makes a more efficient use of the limited spectrum [18], [19]. For two-way systems with a single AF MIMO relay node, [20] investigated the joint source and relay optimization problem under a unified framework which contains a broad class of frequently adopted design criteria. Considering imperfect CSI, [21] developed a robust system design on the basis of the sum mean-squared error (MSE) minimization criterion. In regard to multi-relay scenarios, [7] employed the CF protocol for parallel multiple relay nodes, aiming to maximize the sum-rate of two-way

Manuscript received November 7, 2019; revised May 12, 2020 and August 30, 2020; accepted September 11, 2020. Date of publication September 25, 2020; date of current version January 8, 2021. This work was supported by the National Key Research and Development Program of China under Grant 2018YFB1801101. The associate editor coordinating the review of this article and approving it for publication was W. P. Tay. *(Corresponding author: Yue Rong.)*

Yang Lv and Zhiqiang He are with the Key Laboratory of Universal Wireless Communications, Ministry of Education, Beijing University of Posts and Telecommunications, Beijing 100876, China (e-mail: lvyangcn@foxmail.com; hezq@bupt.edu.cn).

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/TWC.2020.3024898.

Digital Object Identifier 10.1109/TWC.2020.3024898

communications. Reference [22] utilized the gradient descent method to jointly optimize the source and relay matrices, and the iterative algorithm proposed in [23] further enhanced the MSE and system bit-error-rate (BER) performance.

For the purpose of improving the quality of signal reception without incurring a high complexity as the optimal maximum likelihood (ML) detection, the well-known decision feedback equalization (DFE) technique [24], also known as the vertical Bell laboratories layered space-time (BLAST) technique [25], is introduced in recent one-way AF MIMO relay researches. Naturally evolving from the traditional linear minimal mean-squared error (MMSE) receivers, the nonlinear MMSE-DFE receivers are more generally adopted and perform better than the zero-forcing (ZF)-DFE receivers [26], [27]. By using the MMSE-DFE technique, [28] explored the problem of multi-hop relay system optimization with Schur-convex and Schur-concave composite objective functions. Reference [29] developed a joint source and relay power loading algorithm for two-hop systems with multiple relay nodes. Towards a three-terminal model, [30] proposed two well-performing closed-form precoding schemes which, compared to iterative algorithms, had much reduced complexity. For multiuser multi-hop scenarios, [3] designed two distributed transceiver optimization algorithms with no need for centralized processing and better performance than linear receiving-based algorithms.

Noteworthily, [31] studied the cyclic prefixed single-carrier (CP-SC) transmission in a two-way AF MIMO single-relay system using the DFE with noise prediction (DFE-NP) receivers. Under the assumption that the number of antennas at the relay node is more than twice that at both source nodes, [31] proposed an iterative algorithm to jointly optimize the source and relay matrices. Note that the DFE-NP technique, or called the frequency domain equalization with time domain noise prediction (FDE-NP) technique [32], though able to alleviate the error propagation, has relatively high complexity, thus is out of our consideration here.

This paper focuses on the design of an ANC-assisted two-way AF MIMO multi-relay system with MMSE-DFE receivers under both perfect and imperfect CSI scenarios. As far as we know, it is the first time to investigate the utilization of the nonlinear MMSE-DFE receiving technique in ANC-assisted two-way AF MIMO relay communications. Moreover, we consider a general system with multiple parallel relay nodes. The algorithms proposed here can also be applied to single-relay, one-way or linear receiving-based relay systems. Besides, if a system has multiple user-pairs, they can separately utilize our algorithms to communicate over orthogonal channels in time or frequency domain [33]. Specifically, under the transmission power constraints at both source nodes and each relay node, we adopt the sum MSE minimization design criterion to jointly optimize four groups of system parameters, i.e., the precoding matrices at source nodes, the amplifying matrices at relay nodes, as well as the feed-forward and the feedback matrices within the MMSE-DFE receivers. For solving the intractable nonconvex system optimization problem, the block coordinate descent (BCD) method [34, Sec. 2.7], [35] is employed to develop an iterative algorithm which is guaranteed to converge towards, at least, a Nash point [36]. To further improve the performance via mitigating the error propagation in MMSE-DFE receivers, we introduce a group of permutation matrix variables, whose optimization determines the detection orders of all data streams according to their signal-to-interference-and-noise ratios (SINRs). Given that a practical system does not always have sufficiently precise CSI, this paper also extends the proposed algorithms into a robust design scheme for tackling the channel uncertainties. By comparison with the linear receiving-based algorithm developed in [23], our proposed nonlinear receiving-based ones exhibit their superior MSE and BER performance and good robustness against the imperfect CSI in numerical simulations.

The rest of this paper is organized as follows. Section II describes the system model and formulates a nonconvex optimization problem. Section III develops an iterative BCD algorithm to jointly optimize the system parameters. Following that, Section IV gives some additional comments. Section V addresses the issue of optimizing the detection orders. Section VI conducts a robust system design with imperfect CSI. The analyses of simulation results are presented in Section VII. At last, Section VIII draws a conclusion.

The following notations and operators are used throughout this paper: $\triangleq$ stands for the phrase "(is) defined as"; $\mathbb{C}^n$, $\mathbb{C}^{m\times n}$ denote the complex column vector space and matrix space with their dimensions being n and $m\times n$, respectively; the reciprocal, complex conjugate, and modulus of scalar x are expressed as x^{-1}, x^*, and $|x|$; $\|\boldsymbol{x}\|$ denotes the Euclidean norm of vector $\boldsymbol{x}$; $(\cdot)^T$, $(\cdot)^H$ represent the transpose and Hermitian transpose of a vector or matrix; the inverse, pseudo-inverse, rank, trace, and Frobenius norm of matrix $\boldsymbol{X}$ are denoted by $\boldsymbol{X}^{-1}$, $\boldsymbol{X}^{\dagger}$, $\mathrm{rank}(\boldsymbol{X})$, $\mathrm{tr}(\boldsymbol{X})$, and $\|\boldsymbol{X}\|_F$; $\boldsymbol{X}\otimes\boldsymbol{Y}$ represents the Kronecker product of matrices $\boldsymbol{X}$ and $\boldsymbol{Y}$; $[\boldsymbol{X}]_n$, $[\boldsymbol{X}]_{m,m}$, and $[\boldsymbol{X}]_{m,n}$ indicate the nth column vector, the mth diagonal element, and the element at the mth row and the nth column of matrix $\boldsymbol{X}$; $[\boldsymbol{X}]_{1:n}$, $[\boldsymbol{X}]_{1:m,1:m}$ stand for submatrices of matrix $\boldsymbol{X}$, containing its leftmost n columns, and its first m rows and first m columns, respectively; $\mathcal{U}[\boldsymbol{X}]$ denotes a square matrix whose lower triangular part is filled by zeros and strictly upper triangular part is the same as that of matrix $\boldsymbol{X}$; $\mathrm{diag}(\cdot)$, $\mathrm{bd}(\cdot)$ represent a diagonal and a block diagonal matrix respectively, whose diagonal entries are given in parentheses; $\mathrm{vec}(\boldsymbol{X})$ rearranges all the columns in matrix $\boldsymbol{X}$, from left to right, into a single column vector, from top to bottom; for a Hermitian matrix $\boldsymbol{X}\in\mathbb{C}^{n\times n}$, all its eigenvalues are real and denoted in nonincreasing order by $\lambda_i(\boldsymbol{X})$, $i=1,\ldots,n$; $\boldsymbol{I}_n$ is an nth-order identity matrix and $\boldsymbol{0}_{m\times n}$ stands for an $m\times n$ zero matrix; $\mathrm{E}[\cdot]$ denotes the statistical expectation with respect to signal and noise and $\mathrm{E}_H[\cdot]$ denotes that with respect to the channel uncertainties.

II. System Model and Problem Formulation

In the absence of a direct link due to the propagation path loss, a two-way HD relay communication system enables a pair of users to concurrently send information to each other within only two time slots as shown in Fig. 1, where we introduce K assistant relay nodes to amplify their received signals from both source nodes. For the conciseness in

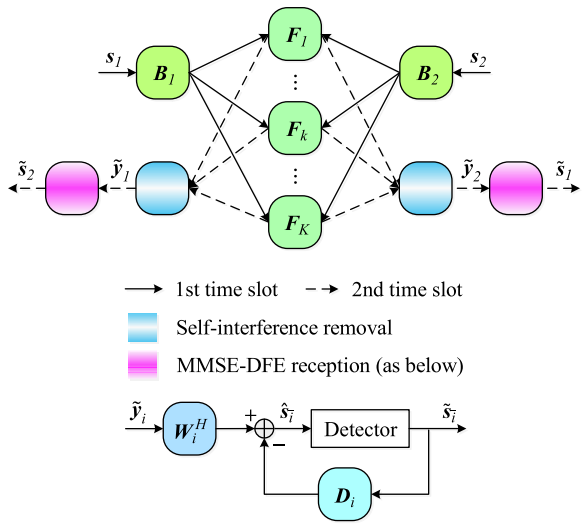


Fig. 1. System model for two-way AF MIMO multi-relay communications with MMSE-DFE receivers.

mathematical derivations, both source nodes are set to transmit M_s independent data streams via M antennas, and each relay node is equipped with N antennas. The generalization of derivations to a system with different numbers of antennas at different nodes is straightforward.

During the first time slot, for $i = 1, 2$, the ith source node at first linearly precodes its modulated signal vector $\boldsymbol{s}_i \in \mathbb{C}^{M_s}$ with the source precoding matrix $\boldsymbol{B}_i \in \mathbb{C}^{M \times M_s}$, then transmits the precoded signal vector $\boldsymbol{x}_i = \boldsymbol{B}_i \boldsymbol{s}_i$ through its wireless interface towards every relay node.

For $k = 1, \ldots, K$, the received signal vector at the kth relay node is given by

$$\boldsymbol{y}_{r,k} = \sum_{i=1}^{2} \boldsymbol{H}_{ri,k} \boldsymbol{x}_i + \boldsymbol{v}_{r,k} \tag{1}$$

where $\boldsymbol{H}_{ri,k} \in \mathbb{C}^{N \times M}$ is the MIMO channel matrix between the ith source node and the kth relay node, and $\boldsymbol{v}_{r,k} \in \mathbb{C}^N$ is the independent and identically distributed (i.i.d.) additive white Gaussian noise (AWGN) vector at the kth relay node.

During the second time slot, under the AF protocol, the kth relay node at first amplifies $\boldsymbol{y}_{r,k}$ in a linear non-regenerative manner with the relay amplifying matrix $\boldsymbol{F}_k \in \mathbb{C}^{N \times N}$, then broadcasts the amplified signal vector $\boldsymbol{x}_{r,k} = \boldsymbol{F}_k \boldsymbol{y}_{r,k}$ towards both source nodes.

Thus, the received signal vector at the ith source node is given by

$$\boldsymbol{y}_i = \sum_{k=1}^{K} \boldsymbol{H}_{ir,k} \boldsymbol{x}_{r,k} + \boldsymbol{v}_i \tag{2}$$

where $\boldsymbol{H}_{ir,k} \in \mathbb{C}^{M \times N}$ is the MIMO channel matrix between the kth relay node and the ith source node, and $\boldsymbol{v}_i \in \mathbb{C}^M$ is the i.i.d. AWGN vector at the ith source node.

Here, as is commonly done, we assume that the two users are perfectly symbol-synchronous and all the symbols in $\boldsymbol{s}_i$ are independent to each other, having zero means and normalized variances, i.e., $\mathrm{E}\left[\boldsymbol{s}_i \boldsymbol{s}_i^H\right] = \boldsymbol{I}_{M_s}$, besides, all the noise vectors have the circular symmetry property [37, Sec. A.1.3] with zero mean vectors and identity covariance matrices.

Substituting (1) and expressions for $\boldsymbol{x}_{r,k}$ and $\boldsymbol{x}_i$ into (2), we obtain

$$\boldsymbol{y}_i = \sum_{k=1}^{K} \boldsymbol{H}_{ir,k} \boldsymbol{F}_k \left(\boldsymbol{H}_{ri,k} \boldsymbol{B}_i \boldsymbol{s}_i + \boldsymbol{H}_{r\bar{i},k} \boldsymbol{B}_{\bar{i}} \boldsymbol{s}_{\bar{i}} \right) + \tilde{\boldsymbol{v}}_i \tag{3}$$

where $\bar{i} = 2$ for $i = 1$ and vice versa, $\tilde{\boldsymbol{v}}_i \triangleq \sum_{k=1}^{K} \boldsymbol{H}_{ir,k} \boldsymbol{F}_k \boldsymbol{v}_{r,k} + \boldsymbol{v}_i$ is the equivalent additive noise vector with its covariance matrix given by $\boldsymbol{C}_i \triangleq \mathrm{E}\left[\tilde{\boldsymbol{v}}_i \tilde{\boldsymbol{v}}_i^H\right] = \sum_{k=1}^{K} \boldsymbol{H}_{ir,k} \boldsymbol{F}_k \boldsymbol{F}_k^H \boldsymbol{H}_{ir,k}^H + \boldsymbol{I}_M$.

In this paper, at first, we assume $\boldsymbol{H}_{ri,k}$ and $\boldsymbol{H}_{ir,k}$ are quasi-static block fading channels and high-precision CSI can be acquired with negligible mismatch against the reality. This corresponds to the situation where all nodes in the system have relatively low mobility or even be static. Here not only traditional channel estimation means [38] but also those specific to two-way MIMO relay systems [39], [40] are available. Then, in Section VI, our research will be extended to a robust design scheme, which utilizes a common analytical approach for channel uncertainties as in [12]–[15], [21], and [41].

To accomplish the self-interference removal operation, the ith source node for $i = 1, 2$ needs record in advance its original modulated signal $\boldsymbol{s}_i$ as well as the precoding matrix $\boldsymbol{B}_i$, and obtain the channel matrices $\boldsymbol{H}_{ri,k}$ and $\boldsymbol{H}_{ir,k}$ as well as the relay amplifying matrices $\boldsymbol{F}_k$ for $k = 1, \ldots, K$. Then, the self-interference term $\sum_{k=1}^{K} \boldsymbol{H}_{ir,k} \boldsymbol{F}_k \boldsymbol{H}_{ri,k} \boldsymbol{B}_i \boldsymbol{s}_i$ can be removed from the received signal (3), resulting in

$$\tilde{\boldsymbol{y}}_i = \sum_{k=1}^{K} \boldsymbol{H}_{ir,k} \boldsymbol{F}_k \boldsymbol{H}_{r\bar{i},k} \boldsymbol{B}_{\bar{i}} \boldsymbol{s}_{\bar{i}} + \tilde{\boldsymbol{v}}_i = \tilde{\boldsymbol{H}}_i \boldsymbol{s}_{\bar{i}} + \tilde{\boldsymbol{v}}_i \tag{4}$$

where $\tilde{\boldsymbol{H}}_i \triangleq \sum_{k=1}^{K} \boldsymbol{H}_{ir,k} \boldsymbol{F}_k \boldsymbol{H}_{r\bar{i},k} \boldsymbol{B}_{\bar{i}} = \boldsymbol{H}_{ir} \boldsymbol{F} \boldsymbol{H}_{r\bar{i}} \boldsymbol{B}_{\bar{i}}$ is the equivalent MIMO channel matrix with $\boldsymbol{H}_{r\bar{i}} \triangleq \left[\boldsymbol{H}_{r\bar{i},1}^T, \ldots, \boldsymbol{H}_{r\bar{i},K}^T\right]^T$, $\boldsymbol{F} \triangleq \mathrm{bd}(\boldsymbol{F}_1, \ldots, \boldsymbol{F}_K)$ and $\boldsymbol{H}_{ir} \triangleq \left[\boldsymbol{H}_{ir,1}, \ldots, \boldsymbol{H}_{ir,K}\right]$.

Similar to the single-hop MIMO system in [42] as well as the multi-hop MIMO relay system in [3], for supporting the transmission of M_s independent data streams from the $\bar{i}$th to the ith source node with acceptable performance, we typically require $M_s \leqslant \min\{\mathrm{rank}(\boldsymbol{H}_{r\bar{i}}), \mathrm{rank}(\boldsymbol{H}_{ir})\}$, from which, as the rank of a matrix is not greater than its dimensions, we have $M_s \leqslant \min\{M, K\ N\}$.

To recover the $\bar{i}$th user's information from signal $\tilde{\boldsymbol{y}}_i$ obtained at the ith source node, the nonlinear MMSE-DFE technique is applied to successively detect all symbols in $\boldsymbol{s}_{\bar{i}}$ with the M_sth symbol detected first and the first symbol detected last.

Specifically, via the M_sth feed-forward vector $\boldsymbol{w}_{i;M_s} \in \mathbb{C}^M$, the M_sth symbol is estimated as $\hat{s}_{\bar{i};M_s} = \boldsymbol{w}_{i;M_s}^H \tilde{\boldsymbol{y}}_i$, then detected as $\tilde{s}_{\bar{i};M_s}$. Following that, the mth symbol is estimated as

$$\hat{s}_{\bar{i};m} = \boldsymbol{w}_{i;m}^H \tilde{\boldsymbol{y}}_i - \sum_{l=m+1}^{M_s} d_{i;m,l} \tilde{s}_{\bar{i};l}, \quad m = 1, \ldots, M_s - 1 \tag{5}$$

where $\boldsymbol{w}_{i;m} \in \mathbb{C}^M$ is the mth feed-forward vector, and $d_{i;m,l}$ for $l = m+1, \ldots, M_s$ is the (m,l)th feedback coefficient

used to eliminate the inter-symbol interference (ISI) produced by the lth previously detected symbol $\tilde{s}_{\bar{i};l}$ from the detection process of the mth symbol.

Here, it is a straightforward and usual way to specify that the mth detected symbol belongs to the mth data stream. However, their corresponding relationships, i.e., the detection order of all the M_s data streams, can be altered and optimized, which will be discussed later in Section V.

Reformulating (5) in matrix form, we have

$$\hat{\boldsymbol{s}}_{\bar{i}} = \boldsymbol{W}_i^H \tilde{\boldsymbol{y}}_i - \boldsymbol{D}_i \tilde{\boldsymbol{s}}_{\bar{i}} \tag{6}$$

where $\hat{\boldsymbol{s}}_{\bar{i}} \triangleq \left[\hat{s}_{\bar{i};1}, \ldots, \hat{s}_{\bar{i};M_s}\right]^T$ and $\tilde{\boldsymbol{s}}_{\bar{i}} \triangleq \left[\tilde{s}_{\bar{i};1}, \ldots, \tilde{s}_{\bar{i};M_s}\right]^T$ denote respectively the estimated and the detected signal vector at the ith source node, $\boldsymbol{W}_i \triangleq [\boldsymbol{w}_{i;1}, \ldots, \boldsymbol{w}_{i;M_s}] \in \mathbb{C}^{M \times M_s}$ stands for the feed-forward matrix, and $\boldsymbol{D}_i \in \mathbb{C}^{M_s \times M_s}$ represents the decision feedback matrix, which is filled by elements $d_{i;m,l}$, thus has strictly upper triangular property.

Moreover, for mathematical derivations, it's a common practice, as in [3], [24]–[30], to assume that no error propagation appears in MMSE-DFE receivers, i.e., we have $\tilde{\boldsymbol{s}}_{\bar{i}} = \boldsymbol{s}_{\bar{i}}$, so (6) becomes

$$\hat{\boldsymbol{s}}_{\bar{i}} = \boldsymbol{W}_i^H \tilde{\boldsymbol{y}}_i - \boldsymbol{D}_i \boldsymbol{s}_{\bar{i}} = \left(\boldsymbol{W}_i^H \tilde{\boldsymbol{H}}_i - \boldsymbol{D}_i\right) \boldsymbol{s}_{\bar{i}} + \boldsymbol{W}_i^H \tilde{\boldsymbol{v}}_i. \tag{7}$$

At the ith source node for $i = 1, 2$, the MSE of the signal waveform estimation for the mth data stream is given by

$$\begin{aligned} E_{i;m} &\triangleq \mathrm{E}\left[\left|\hat{s}_{\bar{i};m} - s_{\bar{i};m}\right|^2\right] \\ &= \sum_{l=1}^{m-1} \left|\boldsymbol{w}_{i;m}^H \left[\tilde{\boldsymbol{H}}_i\right]_l\right|^2 + \left|\boldsymbol{w}_{i;m}^H \left[\tilde{\boldsymbol{H}}_i\right]_m - 1\right|^2 \\ &\quad + \sum_{l=m+1}^{M_s} \left|\boldsymbol{w}_{i;m}^H \left[\tilde{\boldsymbol{H}}_i\right]_l - d_{i;m,l}\right|^2 + \boldsymbol{w}_{i;m}^H \boldsymbol{C}_i \boldsymbol{w}_{i;m}, \\ &\qquad\qquad m = 1, \ldots, M_s \end{aligned} \tag{8}$$

where the third term in (8) will vanish when $m = M_s$.

Accordingly, for all data streams at both source nodes, the two-way sum MSE of the signal waveform estimation can be written as

$$E_s \triangleq \sum_{i=1}^{2} \sum_{m=1}^{M_s} E_{i;m} = \sum_{i=1}^{2} \mathrm{tr}\left\{\mathrm{E}\left[\left(\hat{\boldsymbol{s}}_{\bar{i}} - \boldsymbol{s}_{\bar{i}}\right)\left(\hat{\boldsymbol{s}}_{\bar{i}} - \boldsymbol{s}_{\bar{i}}\right)^H\right]\right\} \tag{9}$$

$$= \sum_{i=1}^{2} \mathrm{tr}\left[\left(\boldsymbol{W}_i^H \tilde{\boldsymbol{H}}_i - \boldsymbol{U}_i\right)\left(\boldsymbol{W}_i^H \tilde{\boldsymbol{H}}_i - \boldsymbol{U}_i\right)^H + \boldsymbol{W}_i^H \boldsymbol{C}_i \boldsymbol{W}_i\right] \tag{10}$$

where $\boldsymbol{U}_i \triangleq \boldsymbol{I}_{M_s} + \boldsymbol{D}_i$ for $i = 1, 2$, defined to facilitate analysis, is an upper triangular matrix with all diagonal elements equal to 1 and is also called the feedback matrix.

Besides, the transmission power consumed at the ith source node for $i = 1, 2$ and the kth relay node for $k = 1, \ldots, K$ is respectively given by $Q_i \triangleq \mathrm{tr}\{\mathrm{E}[\boldsymbol{x}_i \boldsymbol{x}_i^H]\} = \mathrm{tr}\left(\boldsymbol{B}_i \boldsymbol{B}_i^H\right)$ and $P_k \triangleq \mathrm{tr}\{\mathrm{E}[\boldsymbol{x}_{r,k} \boldsymbol{x}_{r,k}^H]\} = \mathrm{tr}[\boldsymbol{F}_k(\sum_{i=1}^{2} \boldsymbol{H}_{ri,k} \boldsymbol{B}_i \boldsymbol{B}_i^H \boldsymbol{H}_{ri,k}^H + \boldsymbol{I}_N)\boldsymbol{F}_k^H]$.

So the problem of the joint optimization of the source precoding, relay amplifying, feed-forward and feedback matrices to minimize E_s under transmission power constraints can be written as

$$\begin{aligned} \min_{\{\boldsymbol{B}_i\},\{\boldsymbol{F}_k\},\{\boldsymbol{W}_i\},\{\boldsymbol{U}_i\}} & \sum_{i=1}^{2} \mathrm{tr}\left[\left(\boldsymbol{W}_i^H \tilde{\boldsymbol{H}}_i - \boldsymbol{U}_i\right) \right. \\ & \left. \times \left(\boldsymbol{W}_i^H \tilde{\boldsymbol{H}}_i - \boldsymbol{U}_i\right)^H + \boldsymbol{W}_i^H \boldsymbol{C}_i \boldsymbol{W}_i\right] \end{aligned} \tag{11}$$

$$\text{s.t. } Q_i = \mathrm{tr}\left(\boldsymbol{B}_i \boldsymbol{B}_i^H\right) \leqslant q_i, \quad i = 1, 2, \tag{12}$$

$$P_k = \mathrm{tr}\left[\boldsymbol{F}_k \left(\sum_{i=1}^{2} \boldsymbol{H}_{ri,k} \boldsymbol{B}_i \boldsymbol{B}_i^H \boldsymbol{H}_{ri,k}^H + \boldsymbol{I}_N\right) \boldsymbol{F}_k^H\right] \leqslant p_k, \quad k = 1, \ldots, K, \tag{13}$$

$$[\boldsymbol{U}_i]_{m,n} = \begin{cases} 0, & m > n, \\ 1, & m = n, \end{cases} \quad i = 1, 2. \tag{14}$$

where variable sets $\{\boldsymbol{B}_i\} \triangleq \{\boldsymbol{B}_1, \boldsymbol{B}_2\}$, $\{\boldsymbol{F}_k\} \triangleq \{\boldsymbol{F}_1, \ldots, \boldsymbol{F}_K\}$, $\{\boldsymbol{W}_i\} \triangleq \{\boldsymbol{W}_1, \boldsymbol{W}_2\}$, $\{\boldsymbol{U}_i\} \triangleq \{\boldsymbol{U}_1, \boldsymbol{U}_2\}$ involve four groups of system parameters remaining to be optimized, and constants q_i and p_k are the power budgets at the ith source node and the kth relay node, respectively.

It can be seen that the problem (11)–(14) is an intractable nonconvex optimization problem, whose globally optimal solution is almost impossible to be found out unless we resort to the exhaustive searching, which is currently not realistic to be accomplished. So we take a step back and try to look for its suboptimal solutions. Here, via utilizing the BCD method, we shall develop an iterative algorithm with guaranteed convergence in Section III, where all the optimization variables are divided into $2 + K$ separate blocks to take turns being updated with other blocks fixed.

Note that the objective function (11) and constraints (12)–(14) bear a certain resemblance to the expressions (18) and (28)–(30) in [3], but nevertheless there are significant differences between them. What [3] optimized is a one-way multi-hop system with multiple cascading relay nodes, thus $\boldsymbol{A}_L$ in (18) of [3], which represents the equivalent MIMO channel matrix between the source modulated signals and the destination node, has a multiplicative form. Only under the assumption of (moderately) high signal-to-noise ratio (SNR) environment, the equivalent target problem (35)–(38) in [3], with a cascading construction, is able to be decomposed, leading to its distributed parameter optimization. Unlike [3], the two-way dual-hop system in this paper allows multiple parallel relay nodes, thus $\tilde{\boldsymbol{H}}_i$ in (11) has an additive form.

To the best of our knowledge, it is the first work to optimize an MMSE-DFE receiving-based two-way AF MIMO multi-relay system. Compared with the problems solved in previous works as listed in Table I, which optimized linear receiving-based two-way AF MIMO (multi-)relay systems, the problem (11)–(14) is more complicated due to the introduction of not only the nonlinear receiving structure made up of decision feed-forward and feedback matrices but also multiple relay amplifying matrices. Besides, optimizing the detection orders within MMSE-DFE receivers in Section V

TABLE I
PREVIOUS WORKS AND THEIR CONTRIBUTIONS

Paper	Main Contributions
K.-J. Lee *et al.* [22]	For both one-way and two-way AF MIMO multi-relay systems with linear MMSE receivers and perfect CSI, based on the sum-rate maximization and sum MSE minimization design criteria with transmission power budgets integrated into the objective functions, the gradient descent method was adopted to jointly optimize the source and relay matrices.
Y. Rong [23]	For two-way AF MIMO multi-relay systems with linear MMSE receivers and perfect CSI, based on the sum MSE minimization design criterion with transmission power constraints, an iterative algorithm was proposed to jointly optimize the source, relay and receiving matrices, which outperforms the gradient algorithm proposed in [22].
J. Zou *et al.* [21]	For two-way AF MIMO relay systems under both frequency-division and time-division duplex modes with linear MMSE receivers and imperfect CSI, based on the sum MSE minimization design criterion with transmission power constraints, by ignoring the second-order terms of the variances of channel estimation errors in mathematical derivations, an iterative algorithm was proposed to jointly optimize the source, relay and receiving matrices. Besides, a constrained structure algorithm was also proposed to optimize the relay matrix with a closed-form solution.
Y. Rong [20]	For two-way AF MIMO relay systems with linear MMSE receivers and perfect CSI, based on a broad class of frequently used design criteria with Schur-convex and Schur-concave objective functions and transmission power constraints, an iterative algorithm was proposed to jointly optimize the source and relay matrices via deriving their optimal beamforming structures.
C.-C. Hu *et al.* [43]	For two-way AF MIMO multi-relay systems with linear MMSE receivers and perfect CSI in correlated fading channels, based on the sum MSE minimization design criterion with transmission power budgets integrated into the objective function, the nonlinear matrix-form conjugate gradient search method was adopted to jointly optimize the source and relay matrices.
Z. Fei *et al.* [44]	For two-way AF MIMO multi-relay systems with linear MMSE receivers and perfect CSI, based on a pure signal processing perspective and the weighted MSE minimization design criterion with constraints of the transmission power at relay nodes, an element-wise semidefinite relaxation method was proposed to jointly optimize the relay and receiving matrices.
M. Zhang *et al.* [45]	For two-way AF MIMO multi-relay systems with multiple user-pairs, linear MMSE receivers and both the perfect and imperfect CSI, based on the design criteria of minimizing all users' sum MSE and the maximal single user's MSE with transmission power constraints, by ignoring the second-order terms of the variances of channel estimation errors in mathematical derivations, iterative algorithms were proposed to jointly optimize the source, relay and receiving matrices.
C.-C. Hu *et al.* [46]	For two-way AF MIMO multi-relay systems with linear MMSE receivers and perfect CSI in correlated fading channels, based on the sum MSE minimization design criterion with transmission power constraints, the singular value decomposition technique was adopted to jointly optimize the source and relay matrices along with the subset selection of relay nodes and their antennas.

and extending the developed algorithms to their robust design counterparts with imperfect CSI in Section VI, where we consider both the first-order and second-order terms of the variances of channel estimation errors in mathematical derivations, further enhance the contributions of this paper.

III. ALGORITHM DESIGN

The proposed iterative BCD algorithm has three major steps in each iteration, corresponding to the optimization of $\{\boldsymbol{B}_i\}$, $\{\boldsymbol{F}_k\}$, as well as $\{\boldsymbol{W}_i\}$ and $\{\boldsymbol{U}_i\}$, respectively.

To start with, we initialize $\{\boldsymbol{B}_i\}$ and $\{\boldsymbol{F}_k\}$ as some feasible values satisfying constraints (12) and (13), e.g.,

$$\boldsymbol{B}_i = \left[\sqrt{q_i/M_s}\boldsymbol{I}_{M_s}, \boldsymbol{0}^T_{(M-M_s)\times M_s}\right]^T, \quad i = 1, 2 \quad (15)$$

and

$$\boldsymbol{F}_k = \sqrt{p_k/N}\left(\sum_{i=1}^{2}\boldsymbol{H}_{ri,k}\boldsymbol{B}_i\boldsymbol{B}_i^H\boldsymbol{H}_{ri,k}^H + \boldsymbol{I}_N\right)^{-1/2}, \quad k = 1, \ldots, K \quad (16)$$

which make full use of the power budgets q_i and p_k. Note that the definition of the squareroot $\boldsymbol{A}^{1/2}$ for a Hermitian positive semidefinite (PSD) matrix $\boldsymbol{A}$ is a natural generalization of that for a symmetric PSD matrix in [47, Sec. A.5.2], which enables $\boldsymbol{A}^{1/2}\boldsymbol{A}^{1/2} = \boldsymbol{A}$ with $\boldsymbol{A}^{1/2}$ being the unique eligible Hermitian PSD matrix as proved in [48, Theorem 7.2.6]. Theoretically, a nonconvex problem like (11)–(14) can have multiple suboptimal solutions. When its variables are iteratively optimized via, e.g., the BCD method, different initialization operations may lead to different optimization results. The reason of choosing (15) and (16) as initial point, rather than random initialization, is that they provide a stable performance.

A. Joint Feed-Forward and Feedback Matrices Optimization

Now, for the first step, based on fixed $\{\boldsymbol{B}_i\}$ and $\{\boldsymbol{F}_k\}$, the optimal $\{\boldsymbol{W}_i\}$ and $\{\boldsymbol{U}_i\}$ can be derived as below.

Let's bring in the following QR factorization [49, Sec. II.7]:

$$\begin{bmatrix}\boldsymbol{C}_i^{-1/2}\tilde{\boldsymbol{H}}_i\\ \boldsymbol{I}_{M_s}\end{bmatrix} = \boldsymbol{Q}_i\boldsymbol{R}_i = \begin{bmatrix}\bar{\boldsymbol{Q}}_i\\ \ddot{\boldsymbol{Q}}_i\end{bmatrix}\boldsymbol{R}_i, \quad i = 1, 2 \quad (17)$$

where $\boldsymbol{Q}_i \in \mathbb{C}^{(M+M_s)\times M_s}$ is a semi-unitary matrix satisfying

$$\boldsymbol{Q}_i^H\boldsymbol{Q}_i = \bar{\boldsymbol{Q}}_i^H\bar{\boldsymbol{Q}}_i + \ddot{\boldsymbol{Q}}_i^H\ddot{\boldsymbol{Q}}_i = \boldsymbol{I}_{M_s}, \quad (18)$$

$\bar{\boldsymbol{Q}}_i \in \mathbb{C}^{M\times M_s}$ and $\ddot{\boldsymbol{Q}}_i \in \mathbb{C}^{M_s\times M_s}$ are made up of the first M and the last M_s rows of $\boldsymbol{Q}_i$, respectively, and $\boldsymbol{R}_i \in \mathbb{C}^{M_s\times M_s}$ is a nonsingular upper triangular matrix with all its diagonal elements being nonzero, or even being positive if the Gram-Schmidt algorithm [49, Sec. II.8] is applied. Then, according to (17), we have

$$\boldsymbol{C}_i^{-1/2}\tilde{\boldsymbol{H}}_i = \bar{\boldsymbol{Q}}_i\boldsymbol{R}_i, \quad \ddot{\boldsymbol{Q}}_i = \boldsymbol{R}_i^{-1}. \quad (19)$$

By utilizing (17)–(19), the theorem below gives the optimal $\{\boldsymbol{W}_i\}$ and $\{\boldsymbol{U}_i\}$.

Theorem 1: With the QR factorization (17)–(19), the optimal feed-forward and feedback matrices are respectively given by

$$\boldsymbol{W}_i = \boldsymbol{C}_i^{-1/2}\bar{\boldsymbol{Q}}_i\boldsymbol{D}_{\boldsymbol{R}_i}^{-H}, \quad \boldsymbol{U}_i = \boldsymbol{D}_{\boldsymbol{R}_i}^{-1}\boldsymbol{R}_i, \quad i=1,2 \tag{20}$$

where $\boldsymbol{D}_{\boldsymbol{R}_i} \triangleq \mathrm{diag}\Big([\boldsymbol{R}_i]_{1,1},\ldots,[\boldsymbol{R}_i]_{M_s,M_s}\Big) \in \mathbb{C}^{M_s\times M_s}$ denotes a diagonal matrix composed of all diagonal elements in $\boldsymbol{R}_i$.

Proof: See Appendix A. □

The derivation of Theorem 1 is inspired by Theorem 2 in [3] and Theorem 1 in [28] which optimized the MMSE-DFE receivers in respectively a multiuser single-hop MIMO system and a single-user multi-hop MIMO relay system, while in this paper, Theorem 1 targets at a two-way dual-hop MIMO multi-relay system instead.

B. Source Precoding Matrices Optimization

Next, for the second step, with given $\{\boldsymbol{F}_k\}$, $\{\boldsymbol{W}_i\}$ and $\{\boldsymbol{U}_i\}$, this subsection aims to optimize $\{\boldsymbol{B}_i\}$ via solving the problem as below.

$$\min_{\{\boldsymbol{B}_i\}} \sum_{i=1}^{2}\mathrm{tr}\Big[\Big(\tilde{\boldsymbol{W}}_i\boldsymbol{B}_{\bar{i}}-\boldsymbol{U}_i\Big)\Big(\tilde{\boldsymbol{W}}_i\boldsymbol{B}_{\bar{i}}-\boldsymbol{U}_i\Big)^H\Big] \tag{21}$$

$$\text{s.t. } \mathrm{tr}\Big(\boldsymbol{B}_i\boldsymbol{B}_i^H\Big)\leqslant q_i, \quad i=1,2, \tag{22}$$

$$\sum_{i=1}^{2}\mathrm{tr}\Big(\boldsymbol{F}_k\boldsymbol{H}_{ri,k}\boldsymbol{B}_i\boldsymbol{B}_i^H\boldsymbol{H}_{ri,k}^H\boldsymbol{F}_k^H\Big)\leqslant\tilde{p}_k, \quad k=1,\ldots,K \tag{23}$$

where $\tilde{\boldsymbol{W}}_i \triangleq \boldsymbol{W}_i^H\sum_{k=1}^{K}\boldsymbol{H}_{ir,k}\boldsymbol{F}_k\boldsymbol{H}_{r\bar{i},k}$ for $i=1,2$ and $\tilde{p}_k \triangleq p_k - \mathrm{tr}\Big(\boldsymbol{F}_k\boldsymbol{F}_k^H\Big)$ for $k=1,\ldots,K$.

Here we present the following theorem at first.

Theorem 2: The function $f(\boldsymbol{X}) = \mathrm{tr}\Big(\boldsymbol{A}\boldsymbol{X}\boldsymbol{B}\boldsymbol{X}^H\Big)$ with matrix variable $\boldsymbol{X}\in\mathbb{C}^{m\times n}$ as well as arbitrary Hermitian PSD matrix constants $\boldsymbol{A}\in\mathbb{C}^{m\times m}$ and $\boldsymbol{B}\in\mathbb{C}^{n\times n}$ is convex.

Proof: See Appendix B. □

Based on Theorem 2, the quadratically constrained quadratic programming (QCQP) problem (21)–(23) can be readily verified as a convex optimization problem, whose unique optimal solution can be efficiently obtained through the well-known interior-point method [47, Ch. 11] or some other approaches. Note that there are several powerful convex optimization tools, e.g., CVX [50], available to solve the problem (21)–(23).

C. Relay Amplifying Matrices Optimization

Finally, for the third step, with given $\{\boldsymbol{F}_j\,|\,j>k\}$ obtained in the last iteration as well as fixed $\{\boldsymbol{W}_i\}$, $\{\boldsymbol{U}_i\}$, $\{\boldsymbol{B}_i\}$ and $\{\boldsymbol{F}_j\,|\,j<k\}$ just updated in the current iteration, we shall get the optimal $\boldsymbol{F}_k$ for $k=1,\ldots,K$, one after another, by solving the following problem.

$$\begin{aligned}\min_{\boldsymbol{F}_k} f_k(\boldsymbol{F}_k) = \sum_{i=1}^{2}\mathrm{tr}\Big[&\big(\boldsymbol{H}_{i,k}\boldsymbol{F}_k\boldsymbol{G}_{\bar{i},k}-\boldsymbol{A}_{i,k}\big)\\ &\times\big(\boldsymbol{H}_{i,k}\boldsymbol{F}_k\boldsymbol{G}_{\bar{i},k}-\boldsymbol{A}_{i,k}\big)^H+\boldsymbol{H}_{i,k}\boldsymbol{F}_k\boldsymbol{F}_k^H\boldsymbol{H}_{i,k}^H\Big]\end{aligned} \tag{24}$$

$$\text{s.t.} g_k(\boldsymbol{F}_k)=\mathrm{tr}\left[\boldsymbol{F}_k\left(\sum_{i=1}^{2}\boldsymbol{G}_{i,k}\boldsymbol{G}_{i,k}^H+\boldsymbol{I}_N\right)\boldsymbol{F}_k^H\right]-p_k\leqslant 0 \tag{25}$$

where for $i=1,2$, $\boldsymbol{H}_{i,k}\triangleq\boldsymbol{W}_i^H\boldsymbol{H}_{ir,k}$, $\boldsymbol{G}_{i,k}\triangleq\boldsymbol{H}_{ri,k}\boldsymbol{B}_i$ and $\boldsymbol{A}_{i,k}\triangleq\boldsymbol{U}_i-\boldsymbol{W}_i^H\sum_{j=1,j\neq k}^{K}\boldsymbol{H}_{ir,j}\boldsymbol{F}_j\boldsymbol{H}_{r\bar{i},j}\boldsymbol{B}_{\bar{i}}$.

According to Theorem 2, the problem (24)–(25) is a convex QCQP problem with a tractable quadratic structure. Thus, the unique optimal $\boldsymbol{F}_k$ can be analytically obtained from the necessary and sufficient Karush-Kuhn-Tucker (KKT) optimality conditions [47, Sec. 5.5.3] written here as

$$\begin{aligned}&g_k(\boldsymbol{F}_k)\leqslant 0, \quad \mu_k\,g_k(\boldsymbol{F}_k)=0,\\ &\mu_k\geqslant 0, \quad \nabla_{\boldsymbol{F}_k}L_k(\boldsymbol{F}_k,\mu_k)=\boldsymbol{0}_{N\times N}\end{aligned} \tag{26}$$

where μ_k is the Lagrange multiplier and $L_k(\boldsymbol{F}_k,\mu_k)\triangleq f_k(\boldsymbol{F}_k)+\mu_k\,g_k(\boldsymbol{F}_k)$ is the Lagrangian.

The gradient of $L_k(\boldsymbol{F}_k,\mu_k)$ with respect to $\boldsymbol{F}_k$ is given by

$$\begin{aligned}&\nabla_{\boldsymbol{F}_k}L_k(\boldsymbol{F}_k,\mu_k)\\ &=2\sum_{i=1}^{2}\Big[\boldsymbol{H}_{i,k}^H\left(\boldsymbol{H}_{i,k}\boldsymbol{F}_k\boldsymbol{G}_{\bar{i},k}-\boldsymbol{A}_{i,k}\right)\boldsymbol{G}_{\bar{i},k}^H+\boldsymbol{H}_{i,k}^H\boldsymbol{H}_{i,k}\boldsymbol{F}_k\Big]\\ &\quad+2\mu_k\boldsymbol{F}_k\left(\sum_{i=1}^{2}\boldsymbol{G}_{i,k}\boldsymbol{G}_{i,k}^H+\boldsymbol{I}_N\right)\end{aligned} \tag{27}$$

which should be equal to zero, resulting in

$$\begin{aligned}&\sum_{i=1}^{2}\boldsymbol{H}_{i,k}^H\boldsymbol{H}_{i,k}\boldsymbol{F}_k\left(\boldsymbol{G}_{\bar{i},k}\boldsymbol{G}_{\bar{i},k}^H+\boldsymbol{I}_N\right)\\ &\qquad+\mu_k\boldsymbol{F}_k\left(\sum_{i=1}^{2}\boldsymbol{G}_{i,k}\boldsymbol{G}_{i,k}^H+\boldsymbol{I}_N\right)\\ &\qquad=\sum_{i=1}^{2}\boldsymbol{H}_{i,k}^H\boldsymbol{A}_{i,k}\boldsymbol{G}_{\bar{i},k}^H.\end{aligned} \tag{28}$$

Through vectorizing both sides of (28), we have

$$\begin{aligned}\mathrm{vec}(\boldsymbol{F}_k)=&\left[\sum_{i=1}^{2}\left(\boldsymbol{G}_{\bar{i},k}\boldsymbol{G}_{\bar{i},k}^H+\boldsymbol{I}_N\right)^T\otimes\left(\boldsymbol{H}_{i,k}^H\boldsymbol{H}_{i,k}\right)\right.\\ &\left.+\mu_k\left(\sum_{i=1}^{2}\boldsymbol{G}_{i,k}\boldsymbol{G}_{i,k}^H+\boldsymbol{I}_N\right)^T\otimes\boldsymbol{I}_N\right]^{\dagger}\\ &\mathrm{vec}\left(\sum_{i=1}^{2}\boldsymbol{H}_{i,k}^H\boldsymbol{A}_{i,k}\boldsymbol{G}_{\bar{i},k}^H\right).\end{aligned} \tag{29}$$

For solving the problem (24)–(25), it can be observed from the KKT conditions (26) that there exist two possibilities:

1) One is for $\mu_k=0$, in which case the second term of the operand of the pseudo-inverse operator in (29) vanishes. Then, if the $\boldsymbol{F}_k$ in (29) indeed satisfies $g_k(\boldsymbol{F}_k)\leqslant 0$, it is the optimal solution. Otherwise, we resort to possibility 2).

2) The other is for $\mu_k>0$, in which case the pseudo-inverse operator in (29) becomes the inverse one as its operand matrix becomes a Hermitian positive definite (PD) matrix according to [51, Corollary 4.2.13], and $g_k(\boldsymbol{F}_k)$ should be equal to zero. Consequently, the optimal $\boldsymbol{F}_k$ can be obtained

by substituting (29) back into $g_k(\boldsymbol{F}_k) = 0$ and seeking the eligible μ_k. To this end, we first derive the following theorem.

Theorem 3: For the optimal $\boldsymbol{F}_k$ satisfying $g_k(\boldsymbol{F}_k) = 0$, there exists an upper bound of its corresponding μ_k in (29), i.e.,

$$\mu_k \leqslant B_{\text{upper}} \triangleq \sqrt{\frac{\lambda_1(\boldsymbol{G}_k)}{p_k}} \frac{\|\boldsymbol{A}_k\|_F}{\lambda_N(\boldsymbol{G}_k)} \tag{30}$$

where $\boldsymbol{G}_k \triangleq \sum_{i=1}^{2} \boldsymbol{G}_{i,k} \boldsymbol{G}_{i,k}^H + \boldsymbol{I}_N$ and $\boldsymbol{A}_k \triangleq \sum_{i=1}^{2} \boldsymbol{H}_{i,k}^H \boldsymbol{A}_{i,k} \boldsymbol{G}_{i,k}^H$.

Proof: See Appendix C. □

At this point, since g_k is a monotonically decreasing function with respect to the positive variable μ_k, which is also bounded above by B_{upper} as demonstrated in Theorem 3, we can readily apply the bisection method [52] to find the unique μ_k which satisfies $g_k(\boldsymbol{F}_k) = 0$, thereby getting the optimal $\boldsymbol{F}_k$ in (29).

D. Convergence, Complexity and Summary

So far we have accomplished one iteration of the proposed iterative BCD algorithm, which is scheduled to be repeated for several times until the algorithm converges. Note that, during the above iterative process, each conditional update, with other variable blocks fixed, of respective $\{\boldsymbol{B}_i\}$, $\boldsymbol{F}_k$ for $k = 1, \ldots, K$ as well as $\{\boldsymbol{W}_i\}$ and $\{\boldsymbol{U}_i\}$ will monotonically decrease the value of the objective function (11), which is also bounded below by at least zero. Therefore, according to [34, Proposition A.3], the algorithm indeed converges. Moreover, since in each update, the optimal solution of the corresponding degenerate optimization problem is uniquely obtained, every limit point achieved by this iterative BCD algorithm will satisfy the Nash equilibrium, i.e., there will be no further improvement of the two-way sum MSE E_s if we individually change any of the $2+K$ variable blocks. This type of point is called a "Nash point" in [36], which, in general, is not necessarily a stationary point. Although here we cannot make a further judgement on whether this Nash point is a stationary (or even optimal) point, the subsequent simulation results in Section VII will exhibit an outstanding MSE and BER performance of our algorithm, which, for convenience, is hereafter called "the MMSE-DFE algorithm".

Now we will analyze the computational complexity for one iteration of the algorithm described within Subsections III-A–III-C. Firstly, optimizing $\{\boldsymbol{W}_i\}$ and $\{\boldsymbol{U}_i\}$ mainly involves the calculations of $\tilde{\boldsymbol{H}}_i$ and $\boldsymbol{C}_i^{-1/2}$ for $i = 1, 2$, the QR factorization (17) as well as several matrix multiplications, whose computational overheads add up to $\mathcal{O}\big(K\left(M^3 + M^2\ N + MN^2\right)\big)$, where $M_s \leqslant M$ is taken into consideration. Secondly, optimizing $\{\boldsymbol{B}_i\}$ mainly involves the calculations of $\tilde{\boldsymbol{W}}_i$ for $i = 1, 2$ and $\tilde{p}_k$ for $k = 1, \ldots, K$, which cost $\mathcal{O}\big(M^3 + K\left(M^2\ N + MN^2 + N^3\right)\big)$, as well as solving the convex QCQP problem (21)–(23) via, e.g., the barrier-generated path-following (BGPF) interior-point method [53, Sec. 3.2]. Specifically, by utilizing the formulae $\mathrm{tr}\Big(\boldsymbol{X}\boldsymbol{X}^H\Big) = \mathrm{vec}(\boldsymbol{X})^H \mathrm{vec}(\boldsymbol{X})$ and $\mathrm{vec}(\boldsymbol{X}\boldsymbol{Y}) = (\boldsymbol{I}_n \otimes \boldsymbol{X})\,\mathrm{vec}(\boldsymbol{Y})$ for $\boldsymbol{X} \in \mathbb{C}^{m\times l}$, $\boldsymbol{Y} \in \mathbb{C}^{l\times n}$, which can be derived from Theorem 1.2.22(i)–(ii) in [54], we are able to convert the problem (21)–(23) into the form similar to that of the problem (34)–(36) in [13] as follows:

$$\begin{aligned} &\min_{\boldsymbol{b}} \boldsymbol{b}^H \boldsymbol{\Psi}_0\, \boldsymbol{b} - \boldsymbol{b}^H \boldsymbol{\psi} - \boldsymbol{\psi}^H \boldsymbol{b} \\ &\text{s.t. } \boldsymbol{b}^H \tilde{\boldsymbol{I}}_i\, \boldsymbol{b} \leqslant q_i, \quad i = 1, 2, \\ &\qquad \boldsymbol{b}^H \boldsymbol{\Psi}_k\, \boldsymbol{b} \leqslant \tilde{p}_k, \quad k = 1, \ldots, K \end{aligned} \tag{31}$$

where we have the vector variable $\boldsymbol{b} \triangleq \big[\mathrm{vec}\big(\boldsymbol{B}_1\big)^T, \mathrm{vec}\big(\boldsymbol{B}_2\big)^T\big]^T$ and the constants $\boldsymbol{\Psi}_0 \triangleq \mathrm{bd}\big(\boldsymbol{I}_{M_s} \otimes \big(\tilde{\boldsymbol{W}}_2^H \tilde{\boldsymbol{W}}_2\big), \boldsymbol{I}_{M_s} \otimes \big(\tilde{\boldsymbol{W}}_1^H \tilde{\boldsymbol{W}}_1\big)\big)$, $\boldsymbol{\psi} \triangleq \big[\mathrm{vec}\big(\boldsymbol{U}_2\big)^H \big(\boldsymbol{I}_{M_s} \otimes \tilde{\boldsymbol{W}}_2\big), \mathrm{vec}\big(\boldsymbol{U}_1\big)^H \big(\boldsymbol{I}_{M_s} \otimes \tilde{\boldsymbol{W}}_1\big)\big]^H$, and $\tilde{\boldsymbol{I}}_i \triangleq \mathrm{bd}\big(\tilde{\boldsymbol{I}}_{i,1}, \tilde{\boldsymbol{I}}_{i,2}\big)$ with $\tilde{\boldsymbol{I}}_{i,i} \triangleq \boldsymbol{I}_{M_s M}$ and $\tilde{\boldsymbol{I}}_{i,\bar{i}} \triangleq \boldsymbol{0}_{\left(M_s M\right)\times\left(M_s M\right)}$ for $i = 1, 2$ as well as $\boldsymbol{\Psi}_k \triangleq \mathrm{bd}\Big(\boldsymbol{I}_{M_s} \otimes \big(\boldsymbol{H}_{r1,k}^H \boldsymbol{F}_k^H \boldsymbol{F}_k \boldsymbol{H}_{r1,k}\big), \boldsymbol{I}_{M_s} \otimes \big(\boldsymbol{H}_{r2,k}^H \boldsymbol{F}_k^H \boldsymbol{F}_k \boldsymbol{H}_{r2,k}\big)\Big)$ for $k = 1, \ldots, K$. Note that here calculating $\boldsymbol{\Psi}_0$, $\boldsymbol{\psi}$ and all $\boldsymbol{\Psi}_k$ costs $\mathcal{O}\big(K\big(M^4 + M^2\ N + MN^2\big)\big)$. Then according to the analysis of solving convex QCQP problems by the BGPF interior-point method in [53, Sec. 6.2.1], the target problem (31), with the number of scalar variables being $a \triangleq 2\ M_s M$, can be solved through $\mathcal{O}\big(\sqrt{K+2}\ln\big[c(K+2)\big]\big)$ iterative steps, where c is a constant greater than 2, and in each step, it costs $\mathcal{O}\big(Ka^2 + a^3\big)$ to form and solve a Newton system. Hence the overheads for optimizing $\big\{\boldsymbol{B}_i\big\}$ add up to $\mathcal{O}\big(\sqrt{K+2}\big(KM^4 + M^6\big)\ln\big[c\big(K+2\big)\big] + K\big(M^2\ N + MN^2 + N^3\big)\big)$. Thirdly, within the optimization of $\boldsymbol{F}_k$ for $k = 1, \ldots, K$, computing, mainly, $\boldsymbol{H}_{i,k}$, $\boldsymbol{G}_{i,k}$ and $\boldsymbol{A}_{i,k}$ for $i = 1, 2$ as well as $\boldsymbol{G}_k$, $\boldsymbol{A}_k$ and B_{upper} costs $\mathcal{O}\big(K\big(M^3 + M^2\ N + MN^2\big) + N^3\big)$. In addition, for the bisection searching of μ_k, the number of its iterative steps is a constant which depends on B_{upper} and the required precision, and in each step, the matrix (pseudo-)inverse in (29) costs the most with complexity $\mathcal{O}\big(N^6\big)$. So the total overheads for optimizing $\big\{\boldsymbol{F}_k\big\}$ are $\mathcal{O}\big(K^2\big(M^3 + M^2\ N + MN^2\big) + KN^6\big)$. All in all, the computational complexity for one iteration of the proposed algorithm is $\mathcal{O}\big(\sqrt{K+2}\big(KM^4 + M^6\big)\ln\big[c\big(K+2\big)\big] + K^2\big(M^3 + M^2\ N + MN^2\big) + KN^6\big)$.

The complete procedures of the MMSE-DFE algorithm are summarized in Table II, where the variables with superscript (n) are the output of the nth iteration. Note that the termination condition adopted at Step 5) is $n \geqslant 10$, meaning that the iteration from Step 2) to Step 4) of Table II is going to be executed for just ten times, which are already enough since we have found in numerical simulations that the performance gain is always nearly negligible after the 10th iteration. This also indicates that our algorithm has a fast convergence rate.

IV. Additional Comments

In this paper, we compare the MMSE-DFE algorithm with the one proposed in [23], which utilized the linear MMSE receivers and, thus, is hereafter called "the LMMSE algorithm". Note that the author in [23] chose 10 as a suitable amount of iterations, which is also followed by us here. Therefore, the comparisons between the MMSE-DFE algorithm and the LMMSE algorithm are fair and reasonable. Besides, both of the two algorithms need to be implemented in

TABLE II
PROCEDURES OF THE MMSE-DFE ALGORITHM

1) Initialize $\left\{\boldsymbol{B}_i^{(0)}\right\}$, $\left\{\boldsymbol{F}_k^{(0)}\right\}$ as (15)–(16). Set counter $n=0$.
2) Compute $\left\{\boldsymbol{W}_i^{(n+1)}\right\}$, $\left\{\boldsymbol{U}_i^{(n+1)}\right\}$ as (20) based on fixed $\left\{\boldsymbol{B}_i^{(n)}\right\}$, $\left\{\boldsymbol{F}_k^{(n)}\right\}$.
3) Solve the problem (21)–(23) as described in Subsection III-B to obtain $\left\{\boldsymbol{B}_i^{(n+1)}\right\}$ with given $\left\{\boldsymbol{F}_k^{(n)}\right\}$, $\left\{\boldsymbol{W}_i^{(n+1)}\right\}$ and $\left\{\boldsymbol{U}_i^{(n+1)}\right\}$.
4) For $k=1,\cdots,K$: Solve the problem (24)–(25) as described in Subsection III-C to obtain $\boldsymbol{F}_k^{(n+1)}$ with given $\left\{\boldsymbol{F}_j^{(n)}\middle|j>k\right\}$, $\left\{\boldsymbol{W}_i^{(n+1)}\right\}$, $\left\{\boldsymbol{U}_i^{(n+1)}\right\}$, $\left\{\boldsymbol{B}_i^{(n+1)}\right\}$ and $\left\{\boldsymbol{F}_j^{(n+1)}\middle|j<k\right\}$.
5) Let $n=n+1$. If $n\geqslant 10$, end the algorithm; Otherwise, continue to Step 2).

a centralized manner. To achieve this, optional schemes may include: one relay node performs the algorithms and delivers the optimized system parameters to the source and other relay nodes, or both source nodes simultaneously perform the algorithms and one source node is responsible for delivering the optimized amplifying matrices to the corresponding relay nodes along with its own transmitted information data.

The newly developed MMSE-DFE algorithm has extensive applicability. Despite targeting at a two-way multi-relay system, the algorithm can also be directly employed in one-way dual-hop relay and two-way single-relay systems. Although, for the sake of notational convenience, this paper focuses on single-carrier transmission, the generalized results concerning multicarrier can be derived without much difficulty, following either subcarrier-independent or subcarrier-cooperative mode as elaborated in [11]. Besides, orthogonal channels in time or frequency domain can be allocated to different user-pairs for supporting their respective communications by using the proposed algorithm. Moreover, the theorem below demonstrates that our algorithm is compatible with those linear receiving-based systems as well.

Theorem 4: When a linear receiver is adopted at both source nodes, i.e., we have $\boldsymbol{U}_i=\boldsymbol{I}_{M_s}$ for $i=1,2$, the optimal feed-forward matrix $\boldsymbol{W}_i$, given in (20), becomes

$$\bar{\boldsymbol{W}}_i=\boldsymbol{C}_i^{-1/2}\bar{\boldsymbol{Q}}_i\boldsymbol{R}_{\bar{i}}^{-H} \tag{32}$$

$$=\left(\tilde{\boldsymbol{H}}_i\tilde{\boldsymbol{H}}_i^H+\boldsymbol{C}_i\right)^{-1}\tilde{\boldsymbol{H}}_i,\quad i=1,2. \tag{33}$$

Proof: See Appendix D. □

Note that (33) is exactly the optimal linear MMSE receiving matrix which has already been derived in [23]. That is to say, in theory, those two-way relay systems with linear receivers can also utilize our proposed algorithm by simply optimizing their linear receiving matrices as (32), though, conventionally, (33) is preferred in practice. It is also worth mentioning that, despite having no calculations of matrix squareroot and QR factorization, which are required by the nonlinear receivers in (20), the linear receivers in (33) still need to compute $\tilde{\boldsymbol{H}}_i$, $\boldsymbol{C}_i$ as well as matrix inverses and multiplications. So the computational complexity of (33) does not decrease by comparison with that of (20). Besides, within the optimization of $\{\boldsymbol{B}_i\}$, $\{\boldsymbol{F}_k\}$ in the LMMSE algorithm, the degeneration from $\boldsymbol{U}_i$ in (20) to $\boldsymbol{I}_{M_s}$ for $i=1,2$ has no influence on the overall computational complexity as well. Thus, our newly developed MMSE-DFE algorithm just has the same order of computational complexity as the LMMSE algorithm. All in all, we believe that the MMSE-DFE algorithm is capable of contributing to the development of 5G wireless networks.

V. OPTIMIZING THE DETECTION ORDERS

As pointed out in [24], the phenomenon of error propagation in MMSE-DFE receivers can lead to non-negligible performance degradation of our proposed algorithm in practice. With respect to the issue of mitigating the error propagation, [24] recommended the automatic repeat request (ARQ) approach for SISO systems. However, if it is adopted in MIMO relay systems, there will be excessive signalling overheads. Reference [3] implemented an unequal error protection (UEP) scheme, which, despite being conceptually simple, needs the assistance of channel coding. Reference [25] verified the effectiveness of several algorithms used to optimize the detection orders in ZF/MMSE-DFE receivers for single-hop MIMO systems, which inspires our following research in this section. It is also worth mentioning that, for the vertical BLAST receiver of a single-hop two-input multiple-output system, [55] optimized its detection order according to the energy of received symbols and analysed its symbol error probabilities (SEPs) for both of the two input data streams. Note that when there are an arbitrary number of input data streams, such analyses of SEPs will become intractable. Hence, for MIMO relay systems, the statistical modeling of error propagation, which requires the information of SEPs, is quite difficult and there are no previous relevant studies as far as we know.

In this paper, for the BCD-based joint parameter optimization of a two-way dual-hop multi-relay system, we introduce a new group of variables, the permutation matrices $\{\boldsymbol{P}_i\}\triangleq\{\boldsymbol{P}_1,\boldsymbol{P}_2\}$, to optimize the detection orders of all data streams.

Specifically, at the ith source node for $i=1,2$, $\boldsymbol{P}_{\bar{i}}\in\mathbb{C}^{M_s\times M_s}$ is used to change the detection order of the target signal vector $\boldsymbol{s}_{\bar{i}}$ sent by the the $\bar{i}$th source node, resulting in

$$\boldsymbol{s}_{\bar{i}}'\triangleq\boldsymbol{P}_{\bar{i}}\boldsymbol{s}_{\bar{i}}=\boldsymbol{P}_{\bar{i}}\left[s_{\bar{i};1},\ldots,s_{\bar{i};M_s}\right]^T=\left[s_{\bar{i};\xi_{\bar{i};1}},\ldots,s_{\bar{i};\xi_{\bar{i};M_s}}\right]^T \tag{34}$$

where $\left\{\xi_{\bar{i};1},\ldots,\xi_{\bar{i};M_s}\right\}$ is a reordering of $\{1,\ldots,M_s\}$, denoting that for $m=1,\ldots,M_s$, the mth symbol in $\boldsymbol{s}_{\bar{i}}'$ is just the $\xi_{\bar{i};m}$th symbol in $\boldsymbol{s}_{\bar{i}}$, i.e., belonging to the $\xi_{\bar{i};m}$th data stream.

Here the optimization of $\{\boldsymbol{P}_i\}$ is carried out before that of $\{\boldsymbol{W}_i\}$, $\{\boldsymbol{U}_i\}$ in Step 2) of Table II. Since (4) can be reformulated to be $\tilde{\boldsymbol{y}}_i=\sum_{m=1}^{M_s}\left[\tilde{\boldsymbol{H}}_i\right]_m s_{\bar{i};m}+\tilde{\boldsymbol{v}}_i$, based on fixed $\{\boldsymbol{B}_i\}$, $\{\boldsymbol{F}_k\}$, the SINR for $s_{\bar{i};m}$, i.e., for the mth data stream, before the feed-forward filtering can be derived as

$$\begin{aligned}\mathrm{SINR}_{\bar{i};m}&\triangleq\mathrm{E}\left[\left\|\left[\tilde{\boldsymbol{H}}_i\right]_m s_{\bar{i};m}\right\|^2\right]\Big/\mathrm{E}\left[\left\|\boldsymbol{\rho}_{\bar{i};m}+\tilde{\boldsymbol{v}}_i\right\|^2\right]\\&=\left\|\left[\tilde{\boldsymbol{H}}_i\right]_m\right\|^2\Big/\left[\sum_{j=1,j\neq m}^{M_s}\left\|\left[\tilde{\boldsymbol{H}}_i\right]_j\right\|^2+\mathrm{tr}(\boldsymbol{C}_i)\right]\end{aligned} \tag{35}$$

where $\boldsymbol{\rho}_{\bar{i};m} \triangleq \sum_{j=1,j\neq m}^{M_s} \left[\tilde{\boldsymbol{H}}_i\right]_j s_{\bar{i};j}$ denotes the interference from other data streams. In order to mitigate the error propagation from earlier detected data streams to later detected ones in the MMSE-DFE receiver at the ith source node, we optimize its detection order under the principle of preferentially detecting those data streams with higher SINRs. Thus, $\left\{\xi_{\bar{i};1},\ldots,\xi_{\bar{i};M_s}\right\}$ satisfies

$$\mathrm{SINR}_{\bar{i};\xi_{\bar{i};1}} \leqslant \ldots \leqslant \mathrm{SINR}_{\bar{i};\xi_{\bar{i};M_s}} \tag{36}$$

from which, the permutation matrix $\boldsymbol{P}_{\bar{i}}$ in (34) is determined.

Now, with the optimized detection orders (i.e., the optimized $\{\boldsymbol{P}_i\}$), (7) turns into $\hat{\boldsymbol{s}}'_{\bar{i}} = \boldsymbol{W}_i^H \tilde{\boldsymbol{y}}_i - \boldsymbol{D}_i \boldsymbol{s}'_{\bar{i}}$ for $i = 1, 2$, where $\hat{\boldsymbol{s}}'_{\bar{i}} \triangleq \boldsymbol{P}_{\bar{i}} \hat{\boldsymbol{s}}_{\bar{i}}$ and $\tilde{\boldsymbol{y}}_i$, given by (4), can be rewritten as

$$\tilde{\boldsymbol{y}}_i = \tilde{\boldsymbol{H}}_i \boldsymbol{P}_{\bar{i}}^T \boldsymbol{s}'_{\bar{i}} + \tilde{\boldsymbol{v}}_i = \boldsymbol{H}_{ir} \boldsymbol{F} \boldsymbol{H}_{r\bar{i}} \boldsymbol{B}'_{\bar{i}} \boldsymbol{s}'_{\bar{i}} + \tilde{\boldsymbol{v}}_i \tag{37}$$

with $\boldsymbol{B}'_{\bar{i}} \triangleq \boldsymbol{B}_{\bar{i}} \boldsymbol{P}_{\bar{i}}^T$, called the permuted source precoding matrix. Since we have $\mathrm{E}\left[\boldsymbol{s}'_{\bar{i}} \boldsymbol{s}'^H_{\bar{i}}\right] = \boldsymbol{P}_{\bar{i}} \boldsymbol{P}_{\bar{i}}^T = \boldsymbol{I}_{M_s} = \mathrm{E}\left[\boldsymbol{s}_{\bar{i}} \boldsymbol{s}_{\bar{i}}^H\right]$, as long as $\boldsymbol{B}_{\bar{i}}$ is substituted by $\boldsymbol{B}'_{\bar{i}}$ for $i = 1, 2$, those derivations within Subsections III-A–III-C remain unaffected, so can Steps 2)–4) in Table II, and we are able to recover $\boldsymbol{B}_{\bar{i}}$ just by $\boldsymbol{B}_{\bar{i}} = \boldsymbol{B}'_{\bar{i}} \boldsymbol{P}_{\bar{i}}$. Therefore, an extended version of the MMSE-DFE algorithm, involving the optimization of detection orders, is proposed and we name it "the MMSE-DFE-O algorithm".

Note that it only costs $\mathcal{O}(M^3)$ to compute and sort $\mathrm{SINR}_{\bar{i};m}$, determine $\boldsymbol{P}_{\bar{i}}$ as well as convert between $\boldsymbol{B}_{\bar{i}}$ and $\boldsymbol{B}'_{\bar{i}}$ for $i = 1, 2$ and $m = 1, \ldots, M_s$, hence the evolution from the MMSE-DFE algorithm to the MMSE-DFE-O algorithm does not change the order of computational complexity. Here the optimization of detection orders not only can mitigate the error propagation, as the earlier detected data streams have higher SINRs and thus produce less detection errors, but also enables those data streams with lower SINRs to be detected last and gain more feedbacks, thus achieves a better use of the nonlinear receiving structure. However, under poor communication conditions, these feedbacks may have the opposite effect due to the severe error propagation. Besides, the MMSE-DFE-O algorithm has more variables to be optimized, which may decrease its convergence rate. Nevertheless, numerical simulations in Section VII will show that, after 10 iterations, compared with the MMSE-DFE algorithm, the MMSE-DFE-O algorithm can always perform better under (moderately) good communication conditions, and even in the worst cases, its performance degradation is negligible, which confirms that the MMSE-DFE-O algorithm has a fast convergence rate. It should be pointed out that, since E_s is derived under the assumption of no error propagation, the improvement of MSE performance obtained by optimizing the detection orders is not as obvious as that of practical BER performance, as will be illustrated later in Section VII.

VI. Robust System Design With Imperfect CSI

Here we shall extend the previously developed algorithms to conduct a robust system design with imperfect CSI. As in [12]–[15], [21], and [41], the well-known Gaussian-Kronecker channel model is adopted, i.e., for $i = 1, 2$ and $k = 1, \ldots, K$, $\boldsymbol{H}_{ri,k}$, $\boldsymbol{H}_{ir,k}$ have the matrix variate complex Gaussian distributions [12], [54, Ch. 2] as follows:

$$\begin{aligned} \boldsymbol{H}_{ri,k} &\sim \mathcal{CN}_{N,M}\left(\hat{\boldsymbol{H}}_{ri,k},\, \sigma^2_{e;\,ri,k} \boldsymbol{\Theta}_{ri,k} \otimes \boldsymbol{\Phi}^T_{ri,k}\right), \\ \boldsymbol{H}_{ir,k} &\sim \mathcal{CN}_{M,N}\left(\hat{\boldsymbol{H}}_{ir,k},\, \sigma^2_{e;\,ir,k} \boldsymbol{\Theta}_{ir,k} \otimes \boldsymbol{\Phi}^T_{ir,k}\right) \end{aligned} \tag{38}$$

where $\hat{\boldsymbol{H}}_{ri,k} \in \mathbb{C}^{N\times M}$, $\hat{\boldsymbol{H}}_{ir,k} \in \mathbb{C}^{M\times N}$ are the estimated CSI; $\sigma^2_{e;\,ri,k}$, $\sigma^2_{e;\,ir,k}$ denote the channel estimation error variances; $\boldsymbol{\Phi}_{ri,k} \in \mathbb{C}^{M\times M}$, $\boldsymbol{\Phi}_{ir,k} \in \mathbb{C}^{N\times N}$ stand for the column correlation matrices of channel estimation error, corresponding to the side of transmitting antennas; $\boldsymbol{\Theta}_{ri,k} \in \mathbb{C}^{N\times N}$, $\boldsymbol{\Theta}_{ir,k} \in \mathbb{C}^{M\times M}$ stand for the row correlation matrices of channel estimation error, corresponding to the side of receiving antennas. Here, $\boldsymbol{\Phi}_{ri,k}$, $\boldsymbol{\Phi}_{ir,k}$, $\boldsymbol{\Theta}_{ri,k}$, $\boldsymbol{\Theta}_{ir,k}$ are all Hermitian PD matrices.

In the following, we give a robust design extension of the MMSE-DFE algorithm to make it capable of handling the channel uncertainties (38).

First of all, it is realized that with imperfect CSI, the receiver at the ith source node for $i = 1, 2$ suffers from residual self-interference (RSI), therefore $\tilde{\boldsymbol{y}}_i$ in (4) becomes

$$\hat{\boldsymbol{y}}_i = \tilde{\boldsymbol{H}}_i \boldsymbol{s}_{\bar{i}} + \tilde{\boldsymbol{v}}_i + \boldsymbol{t}_i \tag{39}$$

where $\boldsymbol{t}_i \triangleq \sum_{k=1}^{K} \left(\boldsymbol{H}_{ir,k} \boldsymbol{F}_k \boldsymbol{H}_{ri,k} - \hat{\boldsymbol{H}}_{ir,k} \boldsymbol{F}_k \hat{\boldsymbol{H}}_{ri,k}\right) \boldsymbol{B}_i \boldsymbol{s}_i$ denotes the RSI vector with its covariance matrix being $\boldsymbol{T}_i \triangleq \mathrm{E}\left[\boldsymbol{t}_i \boldsymbol{t}_i^H\right]$. Consequently, E_s in (10) changes to

$$\begin{aligned} \hat{E}_s = \sum_{i=1}^{2} \mathrm{tr}\Big[&\left(\boldsymbol{W}_i^H \tilde{\boldsymbol{H}}_i - \boldsymbol{U}_i\right) \\ &\times \left(\boldsymbol{W}_i^H \tilde{\boldsymbol{H}}_i - \boldsymbol{U}_i\right)^H + \boldsymbol{W}_i^H \left(\boldsymbol{C}_i + \boldsymbol{T}_i\right) \boldsymbol{W}_i \Big]. \end{aligned} \tag{40}$$

Since $\boldsymbol{H}_{ri,k}$, $\boldsymbol{H}_{ir,k}$ in (40) are not perfectly known, we shall derive $\mathrm{E}_H\left[\hat{E}_s\right]$ via using the theorem below with $\hat{\boldsymbol{\Theta}}_{ri,k} \triangleq \sigma^2_{e;\,ri,k} \boldsymbol{\Theta}_{ri,k}$ and $\hat{\boldsymbol{\Theta}}_{ir,k} \triangleq \sigma^2_{e;\,ir,k} \boldsymbol{\Theta}_{ir,k}$.

Theorem 5: For $\eta = ri,\ ir$ and any complex square matrix constant $\boldsymbol{X}$ with matched dimensions,

$$\mathrm{E}_H\left[\boldsymbol{H}_{\eta,k} \boldsymbol{X} \boldsymbol{H}_{\eta,k}^H\right] = \hat{\boldsymbol{H}}_{\eta,k} \boldsymbol{X} \hat{\boldsymbol{H}}_{\eta,k}^H + \mathrm{tr}(\boldsymbol{X} \boldsymbol{\Phi}_{\eta,k}) \hat{\boldsymbol{\Theta}}_{\eta,k}, \tag{41}$$

$$\mathrm{E}_H\left[\boldsymbol{H}_{\eta,k}^H \boldsymbol{X} \boldsymbol{H}_{\eta,k}\right] = \hat{\boldsymbol{H}}_{\eta,k}^H \boldsymbol{X} \hat{\boldsymbol{H}}_{\eta,k} + \mathrm{tr}\left(\boldsymbol{X} \hat{\boldsymbol{\Theta}}_{\eta,k}^H\right) \boldsymbol{\Phi}_{\eta,k}^H. \tag{42}$$

Proof: This theorem generalizes Theorem 2.3.5(i)–(ii) in [54] from the real matrix space to the complex one, so their proofs are similar and not covered again here. □

On the basis of Theorem 5, $\mathrm{E}_H\left[\tilde{\boldsymbol{H}}_i \tilde{\boldsymbol{H}}_i^H\right]$ and $\mathrm{E}_H[\boldsymbol{C}_i + \boldsymbol{T}_i]$ for $i = 1, 2$ are respectively given by

$$\begin{aligned} &\mathrm{E}_H\left[\tilde{\boldsymbol{H}}_i \tilde{\boldsymbol{H}}_i^H\right] \\ &= \hat{\boldsymbol{H}}_i \hat{\boldsymbol{H}}_i^H + \sum_{k=1}^{K} \Big\{ \hat{\boldsymbol{H}}_{ir,k} \boldsymbol{F}_k \hat{\boldsymbol{\Theta}}_{\bar{i},k} \boldsymbol{F}_k^H \hat{\boldsymbol{H}}_{ir,k}^H \\ &\quad + \mathrm{tr}\Big[\boldsymbol{F}_k \left(\hat{\boldsymbol{H}}_{r\bar{i},k} \boldsymbol{B}_{\bar{i}} \boldsymbol{B}_{\bar{i}}^H \hat{\boldsymbol{H}}_{r\bar{i},k}^H + \hat{\boldsymbol{\Theta}}_{\bar{i},k}\right) \boldsymbol{F}_k^H \boldsymbol{\Phi}_{ir,k}\Big] \hat{\boldsymbol{\Theta}}_{ir,k} \Big\}, \end{aligned}$$

$$\begin{aligned}&\mathrm{E}_H[\boldsymbol{C}_i+\boldsymbol{T}_i]\\&=\sum_{k=1}^{K}\Big\{\hat{\boldsymbol{H}}_{ir,k}\boldsymbol{F}_k\left(\hat{\boldsymbol{\Theta}}_{i,k}+\boldsymbol{I}_N\right)\boldsymbol{F}_k^H\hat{\boldsymbol{H}}_{ir,k}^H\\&\quad+\mathrm{tr}\left[\boldsymbol{F}_k\left(\hat{\boldsymbol{H}}_{ri,k}\boldsymbol{B}_i\boldsymbol{B}_i^H\hat{\boldsymbol{H}}_{ri,k}^H+\hat{\boldsymbol{\Theta}}_{i,k}+\boldsymbol{I}_N\right)\right.\\&\quad\left.\boldsymbol{F}_k^H\boldsymbol{\Phi}_{ir,k}\right]\hat{\boldsymbol{\Theta}}_{ir,k}\Big\}+\boldsymbol{I}_M\end{aligned}\tag{43}$$

where $\hat{\boldsymbol{H}}_i\triangleq\sum_{k=1}^{K}\hat{\boldsymbol{H}}_{ir,k}\boldsymbol{F}_k\hat{\boldsymbol{H}}_{r\bar{i},k}\boldsymbol{B}_{\bar{i}}$ and, for $k=1,\ldots,K$, $\hat{\boldsymbol{\Theta}}_{i,k}\triangleq\mathrm{tr}\Big(\boldsymbol{B}_i\boldsymbol{B}_i^H\boldsymbol{\Phi}_{ri,k}\Big)\hat{\boldsymbol{\Theta}}_{ri,k}$. Hence,

$$\begin{aligned}&\mathrm{E}_H\Big[\hat{E}_s\Big]\\&=\sum_{i=1}^{2}\mathrm{tr}\Bigg\{\Big(\boldsymbol{W}_i^H\hat{\boldsymbol{H}}_i-\boldsymbol{U}_i\Big)\Big(\boldsymbol{W}_i^H\hat{\boldsymbol{H}}_i-\boldsymbol{U}_i\Big)^H+\boldsymbol{W}_i^H\\&\quad\times\Bigg\{\sum_{k=1}^{K}\Big[\hat{\boldsymbol{H}}_{ir,k}\boldsymbol{F}_k\hat{\boldsymbol{\Theta}}_k\boldsymbol{F}_k^H\hat{\boldsymbol{H}}_{ir,k}^H\\&\quad+\mathrm{tr}\Big(\boldsymbol{F}_k\boldsymbol{S}_k\boldsymbol{F}_k^H\boldsymbol{\Phi}_{ir,k}\Big)\hat{\boldsymbol{\Theta}}_{ir,k}\Big]+\boldsymbol{I}_M\Bigg\}\\&\quad\times\boldsymbol{W}_i\Bigg\}=\sum_{i=1}^{2}\mathrm{tr}\Bigg\{\Big(\boldsymbol{W}_i^H\hat{\boldsymbol{H}}_i-\boldsymbol{U}_i\Big)\Big(\boldsymbol{W}_i^H\hat{\boldsymbol{H}}_i-\boldsymbol{U}_i\Big)^H\\&\quad+\boldsymbol{W}_i^H\boldsymbol{V}_i\boldsymbol{W}_i\Bigg\}\end{aligned}\tag{44}$$

where $\hat{\boldsymbol{\Theta}}_k\triangleq\sum_{l=1}^{2}\hat{\boldsymbol{\Theta}}_{l,k}+\boldsymbol{I}_N$, $\boldsymbol{S}_k\triangleq\sum_{l=1}^{2}\hat{\boldsymbol{H}}_{rl,k}\boldsymbol{B}_l\boldsymbol{B}_l^H\hat{\boldsymbol{H}}_{rl,k}^H+\hat{\boldsymbol{\Theta}}_k$, $\boldsymbol{V}_i\triangleq\sum_{k=1}^{K}\Big[\hat{\boldsymbol{H}}_{ir,k}\boldsymbol{F}_k\hat{\boldsymbol{\Theta}}_k\boldsymbol{F}_k^H\hat{\boldsymbol{H}}_{ir,k}^H+\mathrm{tr}\Big(\boldsymbol{F}_k\boldsymbol{S}_k\boldsymbol{F}_k^H\boldsymbol{\Phi}_{ir,k}\Big)\hat{\boldsymbol{\Theta}}_{ir,k}\Big]+\boldsymbol{I}_M$. Likewise, we can get $\mathrm{E}_H[P_k]=\mathrm{tr}\Big(\boldsymbol{F}_k\boldsymbol{S}_k\boldsymbol{F}_k^H\Big)$ for $k=1,\ldots,K$. Thus, evolving from the original problem (11)–(14), the problem of robust system design is written as

$$\begin{aligned}&\min_{\{\boldsymbol{B}_i\},\{\boldsymbol{F}_k\},\{\boldsymbol{W}_i\},\{\boldsymbol{U}_i\}}\mathrm{E}_H\Big[\hat{E}_s\Big]\\&\quad\text{s.t. (12), (14) and }\mathrm{E}_H[P_k]\\&\qquad\leqslant p_k\text{ for }\quad k=1,\ldots,K.\end{aligned}\tag{45}$$

In a similar way as summarized in Table II, via using the BCD method, the matrix variables of the above problem can be iteratively optimized. To start with, we initialize $\{\boldsymbol{B}_i\}$ as (15) and $\boldsymbol{F}_k=\sqrt{p_k/N}\,\boldsymbol{S}_k^{-1/2}$ for $k=1,\ldots,K$. Then, based on fixed $\{\boldsymbol{B}_i\}$, $\{\boldsymbol{F}_k\}$, so long as we substitute $\hat{\boldsymbol{H}}_i$ for $\tilde{\boldsymbol{H}}_i$ and $\boldsymbol{V}_i$ for $\boldsymbol{C}_i$, the optimization results of $\{\boldsymbol{W}_i\}$, $\{\boldsymbol{U}_i\}$ in Subsection III-A can be directly applied. Next, with given $\{\boldsymbol{F}_k\}$, $\{\boldsymbol{W}_i\}$, $\{\boldsymbol{U}_i\}$ as well as $\hat{\boldsymbol{W}}_i\triangleq\boldsymbol{W}_i^H\sum_{k=1}^{K}\hat{\boldsymbol{H}}_{ir,k}\boldsymbol{F}_k\hat{\boldsymbol{H}}_{r\bar{i},k}$, $\alpha_{i,k}\triangleq\mathrm{tr}\big(\boldsymbol{W}_i^H\hat{\boldsymbol{\Theta}}_{ir,k}\boldsymbol{W}_i\big)$, $\beta_{l,k}\triangleq\mathrm{tr}\big[\boldsymbol{F}_k\hat{\boldsymbol{\Theta}}_{rl,k}\boldsymbol{F}_k^H\sum_{i=1}^{2}\big(\hat{\boldsymbol{H}}_{ir,k}^H\boldsymbol{W}_i\boldsymbol{W}_i^H\hat{\boldsymbol{H}}_{ir,k}+\alpha_{i,k}\boldsymbol{\Phi}_{ir,k}\big)\big]$ and $\gamma_{i,k}\triangleq\mathrm{tr}\big(\boldsymbol{F}_k\hat{\boldsymbol{\Theta}}_{ri,k}\boldsymbol{F}_k^H\big)$ for $i=1,2$, $l=1,2$ and $k=1,\ldots,K$, the optimization problem with respect to $\{\boldsymbol{B}_i\}$ is written as

$$\begin{aligned}&\min_{\{\boldsymbol{B}_i\}}\sum_{i=1}^{2}\mathrm{tr}\Big[\Big(\hat{\boldsymbol{W}}_i\boldsymbol{B}_{\bar{i}}-\boldsymbol{U}_i\Big)\Big(\hat{\boldsymbol{W}}_i\boldsymbol{B}_{\bar{i}}-\boldsymbol{U}_i\Big)^H\Big]\\&\qquad+\sum_{l=1}^{2}\mathrm{tr}\Big(\boldsymbol{B}_l\boldsymbol{B}_l^H\boldsymbol{\Gamma}_l\Big)\\&\text{s.t. }\mathrm{tr}\Big(\boldsymbol{B}_i\boldsymbol{B}_i^H\Big)\leqslant q_i,\quad i=1,2,\\&\qquad\sum_{i=1}^{2}\mathrm{tr}\Big[\boldsymbol{B}_i\boldsymbol{B}_i^H\Big(\hat{\boldsymbol{H}}_{ri,k}^H\boldsymbol{F}_k^H\boldsymbol{F}_k\hat{\boldsymbol{H}}_{ri,k}+\gamma_{i,k}\boldsymbol{\Phi}_{ri,k}\Big)\Big]\leqslant\tilde{p}_k,\\&\qquad\qquad k=1,\ldots,K\end{aligned}\tag{46}$$

where matrix $\boldsymbol{\Gamma}_l\triangleq\sum_{k=1}^{K}\Big[\hat{\boldsymbol{H}}_{rl,k}^H\boldsymbol{F}_k^H\Big(\sum_{i=1}^{2}\alpha_{i,k}\boldsymbol{\Phi}_{ir,k}\Big)\boldsymbol{F}_k\hat{\boldsymbol{H}}_{rl,k}+\beta_{l,k}\boldsymbol{\Phi}_{rl,k}\Big]$. This problem, like the one in (21)–(23), is a convex QCQP problem, thus can also be solved by the interior-point method. Finally, with other variables fixed, the optimization problem of $\boldsymbol{F}_k$ for $k=1,\ldots,K$ is given by

$$\begin{aligned}&\min_{\boldsymbol{F}_k}\hat{f}_k(\boldsymbol{F}_k)\\&=\sum_{i=1}^{2}\mathrm{tr}\Big[\Big(\hat{\boldsymbol{H}}_{i,k}\boldsymbol{F}_k\hat{\boldsymbol{G}}_{\bar{i},k}-\hat{\boldsymbol{A}}_{i,k}\Big)\Big(\hat{\boldsymbol{H}}_{i,k}\boldsymbol{F}_k\hat{\boldsymbol{G}}_{\bar{i},k}-\hat{\boldsymbol{A}}_{i,k}\Big)^H\\&\qquad+\boldsymbol{F}_k\hat{\boldsymbol{\Theta}}_k\boldsymbol{F}_k^H\hat{\boldsymbol{H}}_{i,k}^H\hat{\boldsymbol{H}}_{i,k}+\alpha_{i,k}\boldsymbol{F}_k\boldsymbol{S}_k\boldsymbol{F}_k^H\boldsymbol{\Phi}_{ir,k}\Big]\\&\text{s.t. }\hat{g}_k(\boldsymbol{F}_k)=\mathrm{tr}\Big(\boldsymbol{F}_k\boldsymbol{S}_k\boldsymbol{F}_k^H\Big)-p_k\leqslant0\end{aligned}\tag{47}$$

where for $i=1,2$, $\hat{\boldsymbol{H}}_{i,k}\triangleq\boldsymbol{W}_i^H\hat{\boldsymbol{H}}_{ir,k}$, $\hat{\boldsymbol{G}}_{i,k}\triangleq\hat{\boldsymbol{H}}_{ri,k}\boldsymbol{B}_i$ and $\hat{\boldsymbol{A}}_{i,k}\triangleq\boldsymbol{U}_i-\boldsymbol{W}_i^H\sum_{j=1,j\neq k}^{K}\hat{\boldsymbol{H}}_{ir,j}\boldsymbol{F}_j\hat{\boldsymbol{H}}_{r\bar{i},j}\boldsymbol{B}_{\bar{i}}$. As in Subsection III-C, the KKT conditions can be utilized to solve this problem. Here, the solution becomes $\mathrm{vec}(\boldsymbol{F}_k)=\Big[\sum_{i=1}^{2}\Big(\hat{\boldsymbol{G}}_{\bar{i},k}\hat{\boldsymbol{G}}_{\bar{i},k}^H+\hat{\boldsymbol{\Theta}}_k\Big)^T\otimes\Big(\hat{\boldsymbol{H}}_{i,k}^H\hat{\boldsymbol{H}}_{i,k}\Big)+\sum_{i=1}^{2}\boldsymbol{S}_k^T\otimes\left(\alpha_{i,k}\boldsymbol{\Phi}_{ir,k}\right)+\mu_k\boldsymbol{S}_k^T\otimes\boldsymbol{I}_N\Big]^{\dagger}\mathrm{vec}\Big(\hat{\boldsymbol{A}}_k\Big)$ with $\hat{\boldsymbol{A}}_k\triangleq\sum_{i=1}^{2}\hat{\boldsymbol{H}}_{i,k}^H\hat{\boldsymbol{A}}_{i,k}\hat{\boldsymbol{G}}_{\bar{i},k}^H$, and B_{upper} in (30) turns into $\hat{B}_{\mathrm{upper}}\triangleq\sqrt{\lambda_1(\boldsymbol{S}_k)/p_k}\Big(\Big\|\hat{\boldsymbol{A}}_k\Big\|_F\Big/\lambda_N(\boldsymbol{S}_k)\Big)$.

Note that the LMMSE algorithm proposed in [23] can also be extended to achieve the robust design of a linear receiving-based two-way relay system. To be specific, we optimize its linear receivers as those in Theorem 4 with $\hat{\boldsymbol{H}}_i$ and $\boldsymbol{V}_i$ substituted for $\tilde{\boldsymbol{H}}_i$ and $\boldsymbol{C}_i$, respectively, and the optimization of $\{\boldsymbol{B}_i\}$, $\{\boldsymbol{F}_k\}$ can follow the same procedures as described above for the MMSE-DFE algorithm. Regarding the robust design extension of the MMSE-DFE-O algorithm, as pointed out in Section V, after $\{\boldsymbol{P}_i\}$ is determined, we can obtain the permuted source precoding matrices, with which the optimization procedures of the MMSE-DFE algorithm can be directly adopted. Now, for optimizing $\{\boldsymbol{P}_i\}$, we change $\mathrm{SINR}_{\bar{i};m}$ in (35) into $\widehat{\mathrm{SINR}}_{\bar{i};m}\triangleq\mathrm{E}_H\Big[\mathrm{E}\Big[\Big\|\Big[\tilde{\boldsymbol{H}}_i\Big]_m s_{\bar{i};m}\Big\|^2\Big]\Big]\Big/\mathrm{E}_H\Big[\mathrm{E}\Big[\Big\|\boldsymbol{\rho}_{\bar{i};m}+\tilde{\boldsymbol{v}}_i+\boldsymbol{t}_i\Big\|^2\Big]\Big]$ for $i=1,2$ and $m=1,\ldots,M_s$. Since $\Big\|\Big[\tilde{\boldsymbol{H}}_i\Big]_m\Big\|^2=\Big[\tilde{\boldsymbol{H}}_i^H\tilde{\boldsymbol{H}}_i\Big]_{m,m}$ and

$$\begin{aligned}&\mathrm{E}_H\Big[\tilde{\boldsymbol{H}}_i^H\tilde{\boldsymbol{H}}_i\Big]\\&=\hat{\boldsymbol{H}}_i^H\hat{\boldsymbol{H}}_i+\boldsymbol{B}_{\bar{i}}^H\sum_{k=1}^{K}\Big\{\hat{\boldsymbol{H}}_{r\bar{i},k}^H\boldsymbol{F}_k^H\mathrm{tr}\Big(\hat{\boldsymbol{\Theta}}_{ir,k}^H\Big)\boldsymbol{\Phi}_{ir,k}^H\boldsymbol{F}_k\hat{\boldsymbol{H}}_{r\bar{i},k}\\&\quad+\mathrm{tr}\Big\{\boldsymbol{F}_k^H\Big[\hat{\boldsymbol{H}}_{ir,k}^H\hat{\boldsymbol{H}}_{ir,k}+\mathrm{tr}\Big(\hat{\boldsymbol{\Theta}}_{ir,k}^H\Big)\boldsymbol{\Phi}_{ir,k}^H\Big]\boldsymbol{F}_k\hat{\boldsymbol{\Theta}}_{r\bar{i},k}^H\Big\}\boldsymbol{\Phi}_{r\bar{i},k}^H\Big\}\boldsymbol{B}_{\bar{i}}\end{aligned}\tag{48}$$

along with the mutual independence of $\boldsymbol{\rho}_{\bar{i};m}$, $\tilde{\boldsymbol{v}}_i$ and $\boldsymbol{t}_i$, $\widehat{\mathrm{SINR}}_{\bar{i};m}$ is derived as

$$\widehat{\mathrm{SINR}}_{\bar{i};m} = \frac{\left[\mathrm{E}_H\left[\tilde{\boldsymbol{H}}_i^H \tilde{\boldsymbol{H}}_i\right]\right]_{m,m}}{\sum_{j=1,j\neq m}^{M_s}\left[\mathrm{E}_H\left[\tilde{\boldsymbol{H}}_i^H \tilde{\boldsymbol{H}}_i\right]\right]_{j,j} + \mathrm{tr}(\mathrm{E}_H[\boldsymbol{C}_i + \boldsymbol{T}_i])}. \tag{49}$$

So far we have completed all the robust design extensions of the LMMSE, MMSE-DFE and MMSE-DFE-O algorithms with the order of computational complexity remaining unchanged, since the increased calculations are mainly matrix multiplications. Hereafter, for the sake of brevity, the original three algorithms as described in Sections III–V, which use the perfect CSI, are denoted together by "the (L)MMSE(-DFE)(-O) algorithms". Their robust design extensions as described in this section, using the imperfect CSI, i.e., the channel matrices with distributions (38), are called "the R-(L)MMSE(-DFE)(-O) algorithms". Besides, for comparison purposes, the algorithms in Sections III–V can also be executed with only the estimated CSI, which provide a non-robust system design under channel uncertainties, thus called "the NR-(L)MMSE(-DFE)(-O) algorithms".

VII. Numerical Simulations

This section presents the MSE and BER performance of all the proposed algorithms through numerical simulations. Here, MATLAB R2019b running in 64-bit Windows Server 2019 operating system is adopted as the simulation platform, whose hardware is constructed via Baidu Cloud Compute, and we utilize the MATLAB-based software tool CVX to solve those convex QCQP problems with respect to $\{\boldsymbol{B}_i\}$. Besides, all the following Monte Carlo simulation results are obtained from the average of 1500 independent channel realizations.

Here the performance validations and comparisons are made under two scenarios: one is for the systems with perfect CSI and the other is for those with imperfect CSI. In the perfect CSI scenario, we assume an i.i.d. Rayleigh fading channel model, where all the entries in every channel matrix are i.i.d. circularly symmetric complex Gaussian (CSCG) random variables with zero means and variances being $1/M$ for $\boldsymbol{H}_{ri,k}$ or $1/N$ for $\boldsymbol{H}_{ir,k}$, $i = 1, 2$, $k = 1, \ldots, K$. Note that the intention of setting the variances like these is to normalize the influence of the numbers of transmitting antennas. In the imperfect CSI scenario, for the Gaussian-Kronecker channel model (38), we employ the exponential model [12]–[15], [21] to represent the spatial correlation, i.e., for $1 \leqslant l, j \leqslant M$ and $1 \leqslant s, t \leqslant N$, $[\boldsymbol{\Phi}_{ri,k}]_{l,j} = \phi_{ri,k}^{|l-j|}$, $[\boldsymbol{\Theta}_{ri,k}]_{s,t} = \theta_{ri,k}^{|s-t|}$, $[\boldsymbol{\Phi}_{ir,k}]_{s,t} = \phi_{ir,k}^{|s-t|}$, $[\boldsymbol{\Theta}_{ir,k}]_{l,j} = \theta_{ir,k}^{|l-j|}$, where $\phi_{ri,k}$, $\theta_{ri,k}$, $\phi_{ir,k}$, $\theta_{ir,k}$ are the correlation coefficients and we set them as τ. Besides, we also set $\sigma_{e;\,ri,k}^2 = \sigma_e^2/M$, $\sigma_{e;\,ir,k}^2 = \sigma_e^2/N$, and as in [12], [14], the estimated CSI is generated by

$$\begin{aligned}\hat{\boldsymbol{H}}_{ri,k} &\sim \mathcal{CN}_{N,M}\left(\boldsymbol{0}_{N\times M},\, \hat{\sigma}_{ri,k}^2 \boldsymbol{\Theta}_{ri,k} \otimes \boldsymbol{\Phi}_{ri,k}^T\right),\\ \hat{\boldsymbol{H}}_{ir,k} &\sim \mathcal{CN}_{M,N}\left(\boldsymbol{0}_{M\times N},\, \hat{\sigma}_{ir,k}^2 \boldsymbol{\Theta}_{ir,k} \otimes \boldsymbol{\Phi}_{ir,k}^T\right)\end{aligned} \tag{50}$$

TABLE III
Examples of System Settings

Ex.	1	2	3	4	5	6
M_s	3	4	4	4	8	6
M	4	4	4	4	8	8
Q (dB)	25	25	20	20	30	25
N	4	4	4	4	8	4
K	1	1	1	2	1	2

with $\hat{\sigma}_{ri,k}^2 = \left(1-\sigma_e^2\right)/M$, $\hat{\sigma}_{ir,k}^2 = \left(1-\sigma_e^2\right)/N$, where σ_e^2, given for the sake of simplicity, denotes the variance of all channel estimation errors. Here, from (38) and (50), we can obtain $\boldsymbol{H}_{ri,k} \sim \mathcal{CN}_{N,M}\big(\boldsymbol{0}_{N\times M},\, \big(1/M\big)\boldsymbol{\Theta}_{ri,k} \otimes \boldsymbol{\Phi}_{ri,k}^T\big)$, $\boldsymbol{H}_{ir,k} \sim \mathcal{CN}_{M,N}\big(\boldsymbol{0}_{M\times N},\, \big(1/N\big)\boldsymbol{\Theta}_{ir,k} \otimes \boldsymbol{\Phi}_{ir,k}^T\big)$. Note that once $\tau = 0$, $\boldsymbol{H}_{ri,k}$, $\boldsymbol{H}_{ir,k}$ become the i.i.d. Rayleigh fading channel matrices as used in the perfect CSI scenario.

Meanwhile, the distributions of $\boldsymbol{v}_{r,k}$ for $k = 1, \ldots, K$ and $\boldsymbol{v}_i$ for $i = 1, 2$ are subject to $\mathcal{CN}(\boldsymbol{0}, \boldsymbol{I}_N)$ and $\mathcal{CN}(\boldsymbol{0}, \boldsymbol{I}_M)$, respectively. That is, we set all these noise vectors to be zero mean CSCG random vectors with identity covariance matrices. Furthermore, without loss of generality, a unified transmission power budget is assumed for both source nodes as well as for all relay nodes, i.e., $q_i = Q$ for $i = 1, 2$ and $p_k = P$ for $k = 1, \ldots, K$. Noteworthily, the signal propagation path loss is implicitly considered within Q and P, and so are the noise powers as we normalize them to be unity. All the simulations below are carried out with P varying from 0 dB to 40 dB under six different examples of system settings regarding M_s, M, Q, N and K as listed in Table III, where "Ex" is the abbreviation of "Example".

A. Performance Comparisons With Perfect CSI

1) MSE Performance: Figs. 2–3 illustrate the superiority of the MMSE-DFE(-O) algorithms over the LMMSE algorithm in term of the MSE performance with perfect CSI. Here what we exhibit is the arithmetic average of the MSEs for all data streams, i.e., $E_s/\left(2\ M_s\right)$. In order to clearly present the iterative process of the MMSE-DFE algorithm, we also show its MSE simulation results after the 2nd and the 5th iteration (shortened to "itr. 2" and "itr. 5"). It can be observed that, for Exs. 1–4, the MMSE-DFE(-O) algorithms always outperform the LMMSE algorithm after they go through 10 iterations. In particular, as regards Exs. 2–3, even 2 iterations are sufficient for the MMSE-DFE algorithm to obtain a much better MSE performance than the LMMSE algorithm with 10 iterations. Besides, since E_s is derived with no error propagation, the MSE performance of the MMSE-DFE-O algorithm is just a little better than that of the MMSE-DFE algorithm, and the improvement is even invisible for Ex. 4. However, as shown later, the practical BER performance of these two algorithms is much more distinguishable.

Compared with Ex. 2, Ex. 1 has a better MSE performance due to the reduced number of concurrently transmitted data streams, from which the spatial diversity order is increased. Ex. 3 cuts down the power budget at both source nodes,

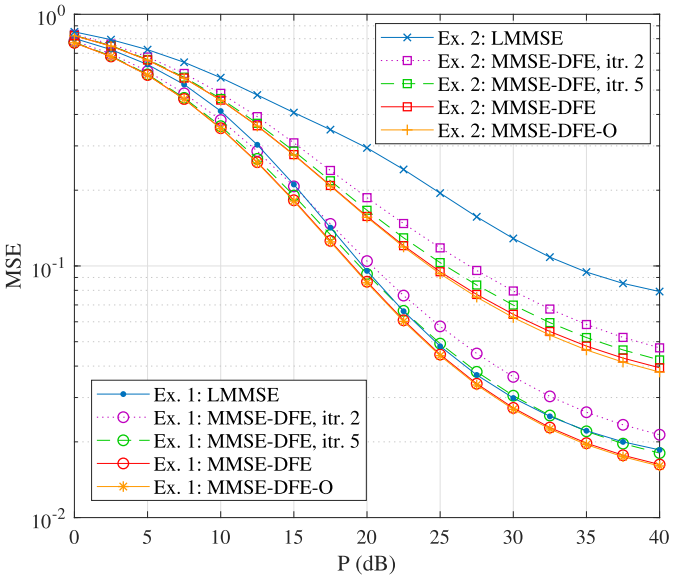


Fig. 2. MSE versus P comparisons for Exs. 1–2 with perfect CSI.

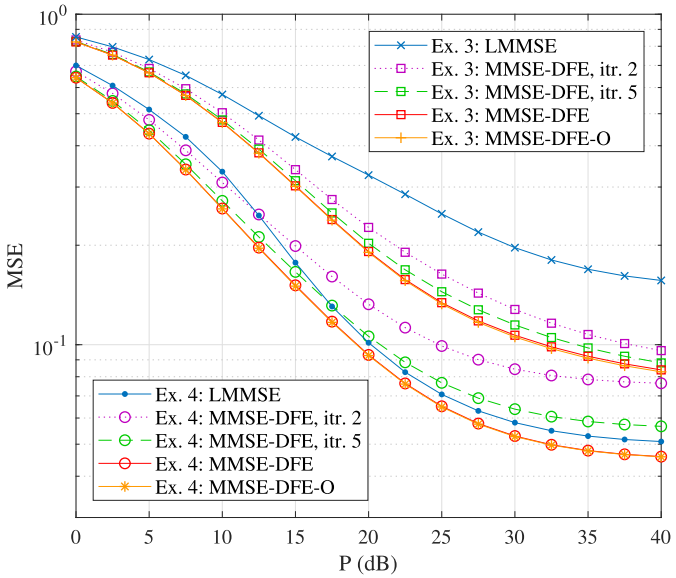


Fig. 3. MSE versus P comparisons for Exs. 3–4 with perfect CSI.

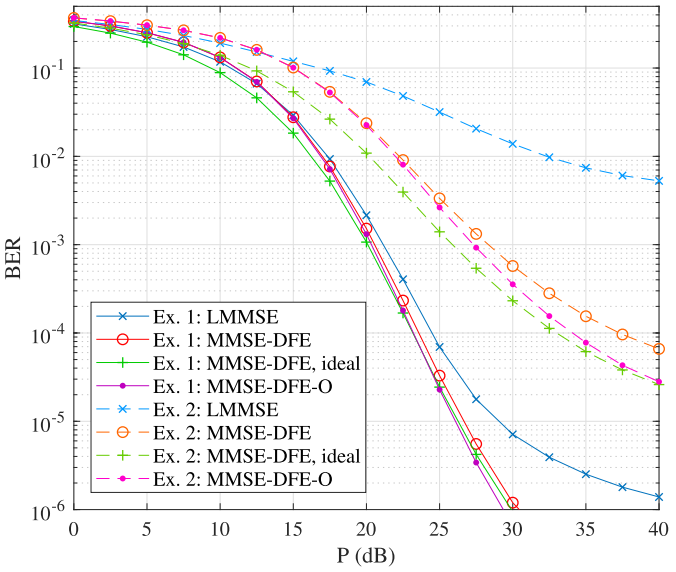


Fig. 4. BER versus P comparisons for Exs. 1–2 with perfect CSI.

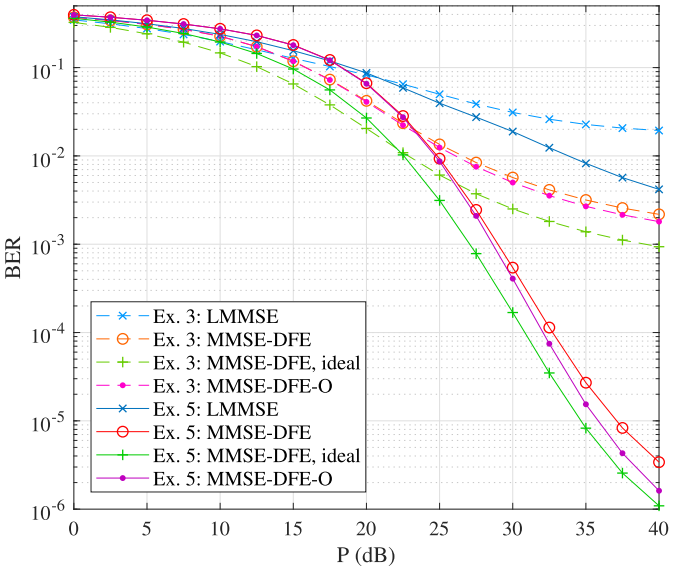


Fig. 5. BER versus P comparisons for Exs. 3, 5 with perfect CSI.

so its MSE curves always lie above the corresponding ones of Ex. 2. By introducing more parallel relay nodes, Ex. 4 has more power available in the system, hence outperforms Ex. 3. Besides, arising from the invariance of Q, which hinders the performance from being further improved when P becomes large enough, the saturation effect is obvious in Ex. 4 after P exceeds 30 dB.

2) BER Performance: Figs. 4–6 indicate the better BER performance of the MMSE-DFE(-O) algorithms by comparison with the LMMSE algorithm. To obtain these simulation results, for each channel realization, we transmit half a million information bits per data stream at both source nodes to pass through the two-way relay system with the QPSK modulation mode. Here, for the MMSE-DFE algorithm, we show not only its practical BER performance curves but also the ideal ones as performance benchmarks. The formers are obtained in the presence of the error propagation within receivers where the symbols fed back are regenerated from previously detected information bits, while the latters follow the assumption of no error propagation within receivers where the symbols fed back are entirely correct. It can be seen that the relationships among Exs. 1–4 in terms of BER performance are generally similar to those regarding MSE performance, which further verifies the effectiveness of the MSE minimization design criterion.

For all examples, under ideal circumstances, the MMSE-DFE algorithm always outperforms the LMMSE algorithm. Meanwhile, the practical BER performance of the MMSE-DFE algorithm is still excellent on the whole. Specifically, within the low range of P, i.e., for about $P < 15$ dB, the practical performance curves of the

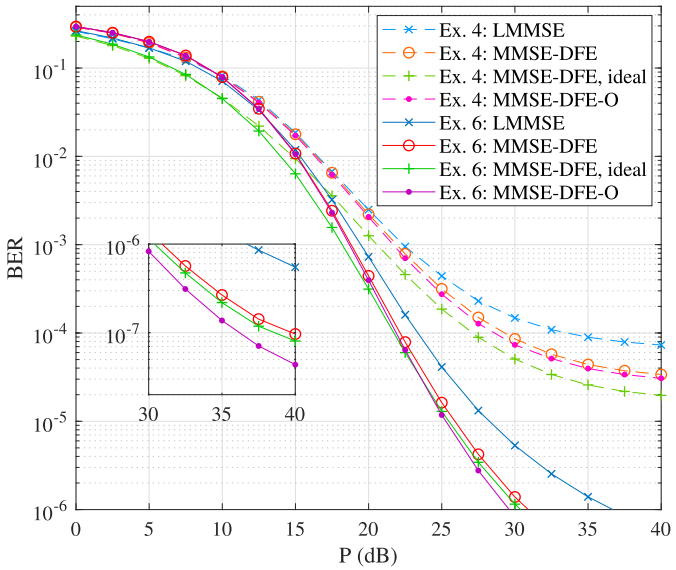


Fig. 6. BER versus P comparisons for Exs. 4, 6 with perfect CSI.

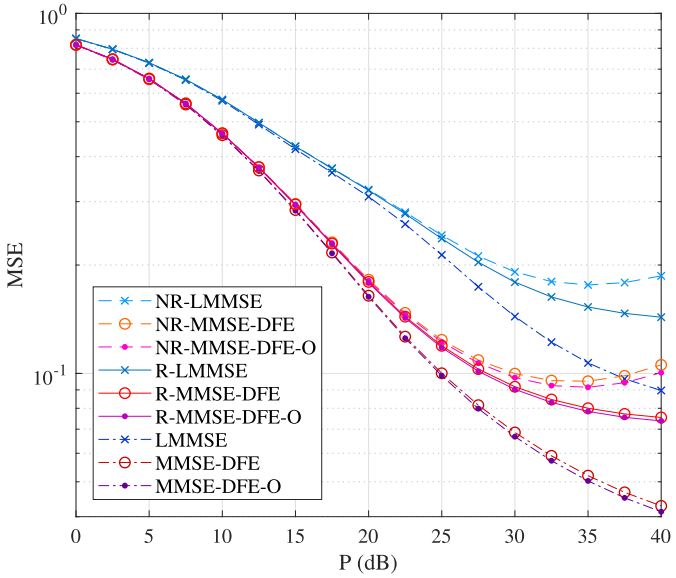


Fig. 7. MSE versus P comparisons for Ex. 2 with $\tau = 0.25$ and $\sigma_e^2 = 0.003$.

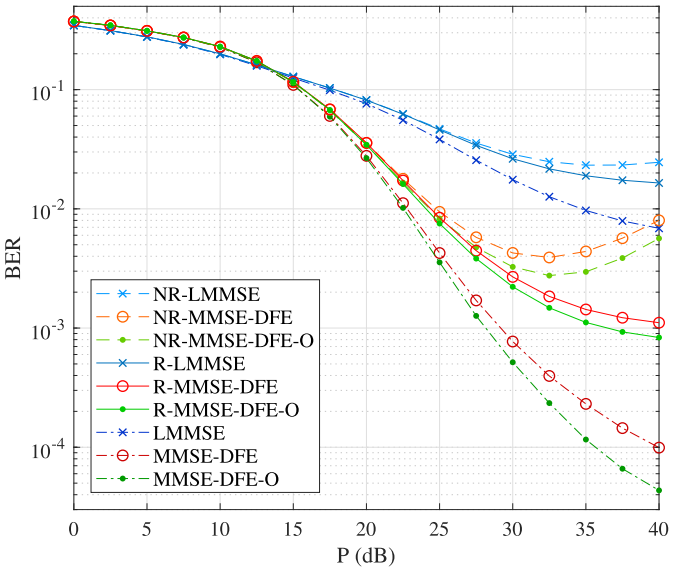


Fig. 8. BER versus P comparisons for Ex. 2 with $\tau = 0.25$ and $\sigma_e^2 = 0.003$.

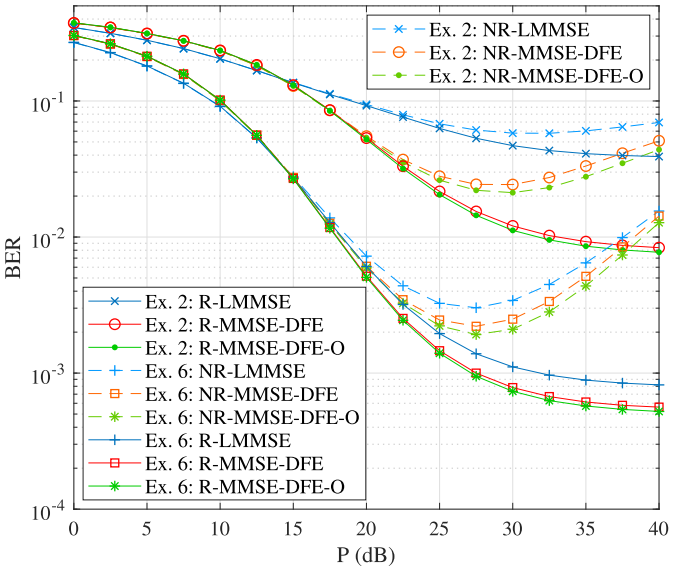


Fig. 9. BER versus P comparisons for Exs. 2, 6 with $\tau = 0.25$ and $\sigma_e^2 = 0.01$.

MMSE-DFE algorithm are slightly inferior to the curves of the LMMSE algorithm. This is because the worse the communication conditions, the more possible the occurrence of detection errors and error propagation in MMSE-DFE receivers. Nevertheless, once P exceeds 15 dB, the practical performance of the MMSE-DFE algorithm gradually becomes better than that of the LMMSE algorithm. The outstanding practical BER performance of the MMSE-DFE algorithm is especially evident as $P > 25$ dB, e.g., there is nearly three orders of magnitude improvement in Ex. 5 at $P = 40$ dB.

Here, we can also observe that the practical BER performance of the MMSE-DFE-O algorithm is superior to that of the MMSE-DFE algorithm. Although they are almost the same within the low range of P, there is a distinct improvement when P is (moderately) high. Besides, as shown in the close-up view of Fig. 6, for Ex. 6, when $P > 30$ dB, the practical BER performance of the MMSE-DFE-O algorithm is even better than the ideal BER performance of the MMSE-DFE algorithm. This is because under good communication conditions, there is nearly no error propagation in MMSE-DFE receivers, and the optimization of detection orders can let those data streams with lower SINRs obtain more beneficial feedbacks, making better use of the nonlinear receiving structure.

B. Performance Comparisons With Imperfect CSI

In the following simulations, the imperfect CSI is taken into account with τ set as 0.25. For Ex. 2 and $\sigma_e^2 = 0.003$, Fig. 7 and Fig. 8 respectively show the MSE and BER performance of the NR-(L)MMSE(-DFE)(-O), R-(L)MMSE(-DFE)(-O) and (L)MMSE(-DFE)(-O) algorithms, where the

algorithms based on perfect CSI show the performance limits and the BER performance of the nonlinear receiving-based algorithms is the practical one involving the error propagation. It can be seen that, since only the estimated CSI is utilized, the non-robust design algorithms perform much worse than the algorithms with perfect CSI, and as P becomes higher, due to the increasingly severe impact of channel estimation errors and RSI, the non-robust design performance curves even gradually go upward. This undesired situation is effectively relieved by the robust design algorithms, which confirms their validity and practicability. Besides, for Exs. 2, 6 and $\sigma_e^2 = 0.01$, Fig. 9 shows the BER performance of the non-robust design and the robust design algorithms. Here we can observe that, for Ex. 2, due to the increase of σ_e^2 from 0.003 to 0.01, there is a significant performance degradation, while the robust design algorithms still provide a considerable performance improvement by comparison with the non-robust design algorithms. For Ex. 6, such improvement is more evident, e.g., the performance is improved by more than an order of magnitude at $P = 40\,\mathrm{dB}$.

VIII. Conclusion

For a two-way AF MIMO multi-relay system with MMSE-DFE receivers, this paper designs an iterative BCD algorithm to solve the joint precoding, amplifying, feed-forward and feedback matrices optimization problem. To mitigate the error propagation within receivers, we further bring in the optimization of the detection orders of data streams. Moreover, a robust design extension is conducted to deal with the imperfect CSI. Through Monte Carlo simulations, the proposed algorithms are shown to not only have better MSE and BER performance than the existing linear receiving-based algorithm, but also provide good robustness against the channel uncertainties.

Appendix A
Proof of Theorem 1

Here we begin by minimizing $E_{i;m}$, i.e., the MSE of the signal waveform estimation for a single data stream. Obviously, the optimal feedback coefficients $d_{i;m,l}$ for (8) ought to satisfy $\boldsymbol{w}_{i;m}^H \left[\tilde{\boldsymbol{H}}_i\right]_l - d_{i;m,l} = 0$ with $l = m+1, \ldots, M_s$ and $m = 1, \ldots, M_s - 1$, or expressed in matrix form,

$$\boldsymbol{D}_i = \mathcal{U}\left[\boldsymbol{W}_i^H \tilde{\boldsymbol{H}}_i\right], \quad i = 1, 2. \tag{51}$$

Thus, via substituting such $d_{i;m,l}$ back into (8), we obtain

$$E_{i;m} = \sum_{l=1}^{m-1} \left|\boldsymbol{w}_{i;m}^H \left[\tilde{\boldsymbol{H}}_i\right]_l\right|^2 + \left|\boldsymbol{w}_{i;m}^H \left[\tilde{\boldsymbol{H}}_i\right]_m - 1\right|^2 + \boldsymbol{w}_{i;m}^H \boldsymbol{C}_i \boldsymbol{w}_{i;m}, \quad m = 1, \ldots, M_s \tag{52}$$

whose gradient with respect to $\boldsymbol{w}_{i;m}$ [56], [57] is given by

$$\nabla_{\boldsymbol{w}_{i;m}} E_{i;m} = 2\left(\sum_{l=1}^{m} \left[\tilde{\boldsymbol{H}}_i\right]_l \left[\tilde{\boldsymbol{H}}_i\right]_l^H + \boldsymbol{C}_i\right) \boldsymbol{w}_{i;m} - 2\left[\tilde{\boldsymbol{H}}_i\right]_m. \tag{53}$$

According to the second-order convexity condition [47, Sec. 3.1.4], (52) can be confirmed as a convex function due to the positive definiteness of its Hessian matrix, i.e., $\nabla^2_{\boldsymbol{w}_{i;m}} E_{i;m} = 2\left(\sum_{l=1}^{m} \left[\tilde{\boldsymbol{H}}_i\right]_l \left[\tilde{\boldsymbol{H}}_i\right]_l^H + \boldsymbol{C}_i\right)$. Then, from the optimality condition (4.22) in [47], minimizing (52) is equivalent to making (53) equal to zero, hence we have

$$\begin{aligned} \boldsymbol{w}_{i;m} &= \left(\sum_{l=1}^{m} \left[\tilde{\boldsymbol{H}}_i\right]_l \left[\tilde{\boldsymbol{H}}_i\right]_l^H + \boldsymbol{C}_i\right)^{-1} \left[\tilde{\boldsymbol{H}}_i\right]_m \\ &= \left[\left(\left[\tilde{\boldsymbol{H}}_i\right]_{1:m} \left[\tilde{\boldsymbol{H}}_i\right]_{1:m}^H + \boldsymbol{C}_i\right)^{-1} \left[\tilde{\boldsymbol{H}}_i\right]_{1:m}\right]_m \end{aligned} \tag{54}$$

for $m = 1, \ldots, M_s$. By using the matrix inversion lemma [58]:

$$(\boldsymbol{A} + \boldsymbol{BCD})^{-1} = \boldsymbol{A}^{-1} - \boldsymbol{A}^{-1}\boldsymbol{B}\left(\boldsymbol{C}^{-1} + \boldsymbol{DA}^{-1}\boldsymbol{B}\right)^{-1}\boldsymbol{DA}^{-1} \tag{55}$$

with nonsingular matrices $\boldsymbol{A}$ and $\boldsymbol{C}$, (54) can be rewritten as

$$\begin{aligned} &\boldsymbol{w}_{i;m} \\ &= \left[\left[\boldsymbol{C}_i^{-1} - \boldsymbol{C}_i^{-1}\left[\tilde{\boldsymbol{H}}_i\right]_{1:m}\left(\boldsymbol{I}_m + \left[\tilde{\boldsymbol{H}}_i\right]_{1:m}^H \boldsymbol{C}_i^{-1}\left[\tilde{\boldsymbol{H}}_i\right]_{1:m}\right)^{-1}\right.\right. \\ &\quad \left.\left. \times \left[\tilde{\boldsymbol{H}}_i\right]_{1:m}^H \boldsymbol{C}_i^{-1}\right]\left[\tilde{\boldsymbol{H}}_i\right]_{1:m}\right]_m \\ &= \left[\boldsymbol{C}_i^{-1}\left[\tilde{\boldsymbol{H}}_i\right]_{1:m}\left(\boldsymbol{I}_m + \left[\tilde{\boldsymbol{H}}_i\right]_{1:m}^H \boldsymbol{C}_i^{-1}\left[\tilde{\boldsymbol{H}}_i\right]_{1:m}\right)^{-1}\right]_m. \end{aligned} \tag{56}$$

At this point, via the QR factorization (17)–(19), we have $\boldsymbol{C}_i^{-1/2}\left[\tilde{\boldsymbol{H}}_i\right]_{1:m} = \left[\bar{\boldsymbol{Q}}_i\right]_{1:m}[\boldsymbol{R}_i]_{1:m,1:m}$. Besides, $\boldsymbol{I}_m$ can also be decomposed as

$$\begin{aligned} \boldsymbol{I}_m &= [\boldsymbol{I}_{M_s}]_{1:m}^H [\boldsymbol{I}_{M_s}]_{1:m} = \left[\ddot{\boldsymbol{Q}}_i \boldsymbol{R}_i\right]_{1:m}^H \left[\ddot{\boldsymbol{Q}}_i \boldsymbol{R}_i\right]_{1:m} \\ &= [\boldsymbol{R}_i]_{1:m,1:m}^H \left[\ddot{\boldsymbol{Q}}_i\right]_{1:m}^H \left[\ddot{\boldsymbol{Q}}_i\right]_{1:m} [\boldsymbol{R}_i]_{1:m,1:m}. \end{aligned} \tag{57}$$

Hence, substituting them back into (56) further makes for

$$\begin{aligned} &\boldsymbol{w}_{i;m} \\ &= \left[\boldsymbol{C}_i^{-1/2}\left[\bar{\boldsymbol{Q}}_i\right]_{1:m}[\boldsymbol{R}_i]_{1:m,1:m}\left[[\boldsymbol{R}_i]_{1:m,1:m}^H\left(\left[\ddot{\boldsymbol{Q}}_i\right]_{1:m}^H\right.\right.\right. \\ &\quad \left.\left.\left. \times \left[\ddot{\boldsymbol{Q}}_i\right]_{1:m} + \left[\bar{\boldsymbol{Q}}_i\right]_{1:m}^H\left[\bar{\boldsymbol{Q}}_i\right]_{1:m}\right)[\boldsymbol{R}_i]_{1:m,1:m}\right]^{-1}\right]_m \\ &= \left[\boldsymbol{C}_i^{-1/2}\left[\bar{\boldsymbol{Q}}_i\right]_{1:m}[\boldsymbol{R}_i]_{1:m,1:m}\left([\boldsymbol{R}_i]_{1:m,1:m}^H[\boldsymbol{Q}_i]_{1:m}^H[\boldsymbol{Q}_i]_{1:m}\right.\right. \\ &\quad \left.\left. \times [\boldsymbol{R}_i]_{1:m,1:m}\right)^{-1}\right]_m = \left[\boldsymbol{C}_i^{-1/2}\left[\bar{\boldsymbol{Q}}_i\right]_{1:m}[\boldsymbol{R}_i]_{1:m,1:m}^{-H}\right]_m \\ &= \boldsymbol{C}_i^{-1/2}\left[\bar{\boldsymbol{Q}}_i\right]_m [\boldsymbol{R}_i]_{m,m}^{-*}, \quad m = 1, \ldots, M_s. \end{aligned} \tag{58}$$

Accordingly, via rewriting (58) in matrix form, the optimal $\{\boldsymbol{W}_i\}$ is given by $\boldsymbol{W}_i = \boldsymbol{C}_i^{-1/2}\bar{\boldsymbol{Q}}_i\boldsymbol{D}_{\boldsymbol{R}_i}^{-H}$, $i = 1, 2$.

On the basis of it, utilizing (19) and (18) in turn, we have

$$\begin{aligned}&\boldsymbol{W}_i^H\tilde{\boldsymbol{H}}_i\\&=\boldsymbol{D}_{\boldsymbol{R}_i}^{-1}\bar{\boldsymbol{Q}}_i^H\boldsymbol{C}_i^{-1/2}\tilde{\boldsymbol{H}}_i=\boldsymbol{D}_{\boldsymbol{R}_i}^{-1}\bar{\boldsymbol{Q}}_i^H\bar{\boldsymbol{Q}}_i\boldsymbol{R}_i\\&=\boldsymbol{D}_{\boldsymbol{R}_i}^{-1}\left(\boldsymbol{I}_{M_s}-\ddot{\boldsymbol{Q}}_i^H\ddot{\boldsymbol{Q}}_i\right)\boldsymbol{R}_i=\boldsymbol{D}_{\boldsymbol{R}_i}^{-1}\boldsymbol{R}_i-\boldsymbol{D}_{\boldsymbol{R}_i}^{-1}\boldsymbol{R}_i^{-H}.\end{aligned}\quad(59)$$

Then, with (59) substituted back into (51), the optimal $\boldsymbol{D}_i$ for $i=1,2$ is given by $\boldsymbol{D}_i=\boldsymbol{D}_{\boldsymbol{R}_i}^{-1}\boldsymbol{R}_i-\boldsymbol{I}_{M_s}$, from which, we can obtain the optimal $\{\boldsymbol{U}_i\}$ as $\boldsymbol{U}_i=\boldsymbol{D}_{\boldsymbol{R}_i}^{-1}\boldsymbol{R}_i$, $i=1,2$.

APPENDIX B
PROOF OF THEOREM 2

To proof Theorem 2, the definition of convex functions in [47, Sec. 3.1.1] can be utilized.

Specifically, for all $\boldsymbol{X}_1,\boldsymbol{X}_2\in\mathbb{C}^{m\times n}$ and any real constant θ with $0\leqslant\theta\leqslant 1$, we have

$$\begin{aligned}&\theta f(\boldsymbol{X}_1)+(1-\theta)f(\boldsymbol{X}_2)-f[\theta\boldsymbol{X}_1+(1-\theta)\boldsymbol{X}_2]\\&=\theta\operatorname{tr}\left(\boldsymbol{A}\boldsymbol{X}_1\boldsymbol{B}\boldsymbol{X}_1^H\right)+(1-\theta)\operatorname{tr}\left(\boldsymbol{A}\boldsymbol{X}_2\boldsymbol{B}\boldsymbol{X}_2^H\right)\\&\quad-\operatorname{tr}\left\{\boldsymbol{A}\left[\theta\boldsymbol{X}_1+(1-\theta)\boldsymbol{X}_2\right]\boldsymbol{B}\left[\theta\boldsymbol{X}_1+(1-\theta)\boldsymbol{X}_2\right]^H\right\}\\&=\theta(1-\theta)\operatorname{tr}\left[\boldsymbol{A}\left(\boldsymbol{X}_1-\boldsymbol{X}_2\right)\boldsymbol{B}\left(\boldsymbol{X}_1-\boldsymbol{X}_2\right)^H\right].\end{aligned}\quad(60)$$

Here let us introduce the eigenvalue decompositions (EVDs) of $\boldsymbol{A}$ and $\boldsymbol{B}$ as $\boldsymbol{A}=\boldsymbol{V}_a\boldsymbol{\Lambda}_a\boldsymbol{V}_a^H$, $\boldsymbol{B}=\boldsymbol{V}_b\boldsymbol{\Lambda}_b\boldsymbol{V}_b^H$ where $\boldsymbol{V}_a\in\mathbb{C}^{m\times m}$, $\boldsymbol{V}_b\in\mathbb{C}^{n\times n}$ are unitary matrices and $\boldsymbol{\Lambda}_a\in\mathbb{C}^{m\times m}$, $\boldsymbol{\Lambda}_b\in\mathbb{C}^{n\times n}$ are diagonal matrices with all their diagonal elements being nonnegative. Then, substituting them back into (60) results in

$$\begin{aligned}&\theta(1-\theta)\operatorname{tr}\left\{\boldsymbol{\Lambda}_a\left[\boldsymbol{V}_a^H\left(\boldsymbol{X}_1-\boldsymbol{X}_2\right)\boldsymbol{V}_b\right]\right.\\&\qquad\left.\boldsymbol{\Lambda}_b\left[\boldsymbol{V}_a^H\left(\boldsymbol{X}_1-\boldsymbol{X}_2\right)\boldsymbol{V}_b\right]^H\right\}\\&=\theta(1-\theta)\operatorname{tr}\left\{\left[\boldsymbol{\Lambda}_a^{1/2}\boldsymbol{V}_a^H\left(\boldsymbol{X}_1-\boldsymbol{X}_2\right)\boldsymbol{V}_b\boldsymbol{\Lambda}_b^{1/2}\right]\right.\\&\qquad\left.\times\left[\boldsymbol{\Lambda}_a^{1/2}\boldsymbol{V}_a^H\left(\boldsymbol{X}_1-\boldsymbol{X}_2\right)\boldsymbol{V}_b\boldsymbol{\Lambda}_b^{1/2}\right]^H\right\}.\end{aligned}\quad(61)$$

Clearly, there appears a Hermitian PSD matrix within the braces of the trace operator in (61), which, together with $\theta(1-\theta)\geqslant 0$, makes (61) not less than zero, leading to

$$f[\theta\boldsymbol{X}_1+(1-\theta)\boldsymbol{X}_2]\leqslant\theta f(\boldsymbol{X}_1)+(1-\theta)f(\boldsymbol{X}_2).\quad(62)$$

So we can confirm that $f(\boldsymbol{X})$ is a convex function.

APPENDIX C
PROOF OF THEOREM 3

To proof Theorem 3, we need the following two lemmas.

Lemma 1: For Hermitian matrices $\boldsymbol{X},\boldsymbol{Y}\in\mathbb{C}^{n\times n}$, $\operatorname{tr}(\boldsymbol{X}\boldsymbol{Y})\leqslant\sum_{i=1}^n\lambda_i(\boldsymbol{X})\lambda_i(\boldsymbol{Y})\leqslant\lambda_1(\boldsymbol{X})\operatorname{tr}(\boldsymbol{Y})$.

Proof: This lemma can be deduced directly from Theorem 4.3.53 in [48]. □

Here, to prepare for Lemma 2, we bring in Definition 7.7.1 in [48]: for $\boldsymbol{X},\boldsymbol{Y}\in\mathbb{C}^{n\times n}$, we write $\boldsymbol{X}\succeq\boldsymbol{Y}$ ($\boldsymbol{X}\succ\boldsymbol{Y}$) if $\boldsymbol{X},\boldsymbol{Y}$ are Hermitian matrices and $\boldsymbol{X}-\boldsymbol{Y}$ is a PSD (PD) matrix; further, we write $\boldsymbol{X}\succeq\boldsymbol{0}$ ($\boldsymbol{X}\succ\boldsymbol{0}$) if $\boldsymbol{X}$ is a Hermitian PSD (PD) matrix.

Lemma 2: If matrices $\boldsymbol{A},\boldsymbol{B}\in\mathbb{C}^{n\times n}$ satisfy $\boldsymbol{A}\succeq\boldsymbol{B}$ ($\boldsymbol{A}\succ\boldsymbol{B}$), then $\lambda_i(\boldsymbol{A})\geqslant\lambda_i(\boldsymbol{B})$ ($\lambda_i(\boldsymbol{A})>\lambda_i(\boldsymbol{B})$) for $i=1,\ldots,n$.

Proof: This lemma just restates Corollary 4.3.12 in [48]. □

Now, from $g_k(\boldsymbol{F}_k)=0$ and Lemma 1, we can derive

$$\begin{aligned}p_k&=\operatorname{tr}\left(\boldsymbol{F}_k\boldsymbol{G}_k\boldsymbol{F}_k^H\right)\\&=\operatorname{tr}\left(\boldsymbol{G}_k\boldsymbol{F}_k^H\boldsymbol{F}_k\right)\leqslant\lambda_1(\boldsymbol{G}_k)\operatorname{tr}\left(\boldsymbol{F}_k^H\boldsymbol{F}_k\right)\end{aligned}\quad(63)$$

where $\boldsymbol{G}_k$ is essentially the covariance matrix of $\boldsymbol{y}_{r,k}$, namely, $\boldsymbol{G}_k=\mathrm{E}\left[\boldsymbol{y}_{r,k}\boldsymbol{y}_{r,k}^H\right]$.

Then substituting (29) back into $\operatorname{tr}\left(\boldsymbol{F}_k^H\boldsymbol{F}_k\right)$ leads to

$$\begin{aligned}\operatorname{tr}\left(\boldsymbol{F}_k^H\boldsymbol{F}_k\right)&=\operatorname{vec}(\boldsymbol{F}_k)^H\operatorname{vec}(\boldsymbol{F}_k)\\&=\operatorname{vec}(\boldsymbol{A}_k)^H\boldsymbol{Z}_k^{-H}\boldsymbol{Z}_k^{-1}\operatorname{vec}(\boldsymbol{A}_k)\end{aligned}\quad(64)$$

where $\boldsymbol{Z}_k\triangleq\sum_{i=1}^2\left(\boldsymbol{G}_{\bar{i},k}\boldsymbol{G}_{\bar{i},k}^H+\boldsymbol{I}_N\right)^T\otimes\left(\boldsymbol{H}_{i,k}^H\boldsymbol{H}_{i,k}\right)+\mu_k\boldsymbol{G}_k^T\otimes\boldsymbol{I}_N$ is a Hermitian PD matrix as mentioned before. Via utilizing Lemma 1 again, we obtain

$$\begin{aligned}\operatorname{tr}\left(\boldsymbol{F}_k^H\boldsymbol{F}_k\right)&=\operatorname{tr}\left[\operatorname{vec}(\boldsymbol{A}_k)\operatorname{vec}(\boldsymbol{A}_k)^H\boldsymbol{Z}_k^{-H}\boldsymbol{Z}_k^{-1}\right]\\&\leqslant\sum_{i=1}^{N^2}\lambda_i\left[\operatorname{vec}(\boldsymbol{A}_k)\operatorname{vec}(\boldsymbol{A}_k)^H\right]\lambda_i\left(\boldsymbol{Z}_k^{-1}\boldsymbol{Z}_k^{-1}\right).\end{aligned}\quad(65)$$

For matrix $\operatorname{vec}(\boldsymbol{A}_k)\operatorname{vec}(\boldsymbol{A}_k)^H$, since its rank equals 1, all but one of its eigenvalues are zeros, i.e.,

$$\begin{aligned}\lambda_1\left[\operatorname{vec}(\boldsymbol{A}_k)\operatorname{vec}(\boldsymbol{A}_k)^H\right]&=\operatorname{tr}\left[\operatorname{vec}(\boldsymbol{A}_k)\operatorname{vec}(\boldsymbol{A}_k)^H\right]\\&=\|\boldsymbol{A}_k\|_F^2,\end{aligned}\quad(66)$$

and $\lambda_i\left[\operatorname{vec}(\boldsymbol{A}_k)\operatorname{vec}(\boldsymbol{A}_k)^H\right]=0$ for $i=2,\ldots,N^2$. So we have

$$\begin{aligned}\operatorname{tr}\left(\boldsymbol{F}_k^H\boldsymbol{F}_k\right)&\leqslant\lambda_1\left[\operatorname{vec}(\boldsymbol{A}_k)\operatorname{vec}(\boldsymbol{A}_k)^H\right]\lambda_1\left(\boldsymbol{Z}_k^{-1}\right)^2\\&=\|\boldsymbol{A}_k\|_F^2\,\lambda_{N^2}(\boldsymbol{Z}_k)^{-2}.\end{aligned}\quad(67)$$

On the basis of Corollary 4.2.13 in [51], it is not hard to get $\boldsymbol{Z}_k\succeq\mu_k\boldsymbol{G}_k^T\otimes\boldsymbol{I}_N\succ\boldsymbol{0}$, which, due to Lemma 2 as well as [51, Theorem 4.2.12], results in

$$\begin{aligned}\lambda_{N^2}(\boldsymbol{Z}_k)&\geqslant\lambda_{N^2}\left(\mu_k\boldsymbol{G}_k^T\otimes\boldsymbol{I}_N\right)\\&=\mu_k\lambda_N\left(\boldsymbol{G}_k^T\right)\lambda_N(\boldsymbol{I}_N)=\mu_k\lambda_N(\boldsymbol{G}_k).\end{aligned}\quad(68)$$

Combining (63), (67) and (68), we finally obtain

$$p_k\leqslant\lambda_1(\boldsymbol{G}_k)\,\|\boldsymbol{A}_k\|_F^2\,\mu_k^{-2}\lambda_N(\boldsymbol{G}_k)^{-2}\quad(69)$$

which leads to (30).

APPENDIX D
PROOF OF THEOREM 4

For nonlinear MMSE-DFE receivers, based on Theorem 1, the optimal feedback matrix $\boldsymbol{U}_i$ satisfies $\boldsymbol{D}_{\boldsymbol{R}_i}^{-1} = \boldsymbol{U}_i \boldsymbol{R}_i^{-1}$ for $i = 1, 2$, from which, the optimal feed-forward matrix $\boldsymbol{W}_i$ can be rewritten as

$$\boldsymbol{W}_i = \boldsymbol{C}_i^{-1/2} \bar{\boldsymbol{Q}}_i \boldsymbol{R}_i^{-H} \boldsymbol{U}_i^H, \quad i = 1, 2. \tag{70}$$

When we switch to linear receivers with $\boldsymbol{U}_i = \boldsymbol{I}_{M_s}$, (70) becomes (32). Since we have $\bar{\boldsymbol{Q}}_i = \boldsymbol{C}_i^{-1/2} \tilde{\boldsymbol{H}}_i \boldsymbol{R}_i^{-1}$ from (19), (32) is equivalent to

$$\bar{\boldsymbol{W}}_i = \boldsymbol{C}_i^{-1} \tilde{\boldsymbol{H}}_i \boldsymbol{R}_i^{-1} \boldsymbol{R}_i^{-H} = \boldsymbol{C}_i^{-1} \tilde{\boldsymbol{H}}_i \left(\boldsymbol{R}_i^H \boldsymbol{R}_i \right)^{-1}. \tag{71}$$

Here, by exploiting (17)–(18), we can obtain

$$\begin{aligned} &\boldsymbol{R}_i^H \boldsymbol{R}_i \\ &= (\boldsymbol{Q}_i \boldsymbol{R}_i)^H (\boldsymbol{Q}_i \boldsymbol{R}_i) = \begin{bmatrix} \boldsymbol{C}_i^{-1/2} \tilde{\boldsymbol{H}}_i \\ \boldsymbol{I}_{M_s} \end{bmatrix}^H \begin{bmatrix} \boldsymbol{C}_i^{-1/2} \tilde{\boldsymbol{H}}_i \\ \boldsymbol{I}_{M_s} \end{bmatrix} \\ &= \left(\boldsymbol{C}_i^{-1/2} \tilde{\boldsymbol{H}}_i \right)^H \left(\boldsymbol{C}_i^{-1/2} \tilde{\boldsymbol{H}}_i \right) + \boldsymbol{I}_{M_s} = \tilde{\boldsymbol{H}}_i^H \boldsymbol{C}_i^{-1} \tilde{\boldsymbol{H}}_i + \boldsymbol{I}_{M_s}. \end{aligned} \tag{72}$$

Then, substituting (72) back into (71) results in

$$\begin{aligned} &\bar{\boldsymbol{W}}_i \\ &= \boldsymbol{C}_i^{-1} \tilde{\boldsymbol{H}}_i \left(\tilde{\boldsymbol{H}}_i^H \boldsymbol{C}_i^{-1} \tilde{\boldsymbol{H}}_i + \boldsymbol{I}_{M_s} \right)^{-1} \\ &= \boldsymbol{C}_i^{-1} \tilde{\boldsymbol{H}}_i \left[\boldsymbol{I}_{M_s} - \tilde{\boldsymbol{H}}_i^H \left(\boldsymbol{C}_i + \tilde{\boldsymbol{H}}_i \tilde{\boldsymbol{H}}_i^H \right)^{-1} \tilde{\boldsymbol{H}}_i \right] \qquad (73) \\ &= \left[\boldsymbol{C}_i^{-1} - \boldsymbol{C}_i^{-1} \tilde{\boldsymbol{H}}_i \tilde{\boldsymbol{H}}_i^H \left(\boldsymbol{I}_M + \boldsymbol{C}_i^{-1} \tilde{\boldsymbol{H}}_i \tilde{\boldsymbol{H}}_i^H \right)^{-1} \boldsymbol{C}_i^{-1} \right] \tilde{\boldsymbol{H}}_i \\ &= \left(\tilde{\boldsymbol{H}}_i \tilde{\boldsymbol{H}}_i^H + \boldsymbol{C}_i \right)^{-1} \tilde{\boldsymbol{H}}_i, \quad i = 1, 2 \qquad (74) \end{aligned}$$

where we obtain (73) and (74) via making use of the matrix inversion lemma (55).

ACKNOWLEDGMENT

The authors would like to thank Prof. Weiling Wu for his beneficial guidance as well as Prof. Jiaru Lin, Prof. Kai Niu, Prof. Wenjun Xu, and Prof. Wenbo Xu for helpful discussions.

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] G. Liu, F. R. Yu, H. Ji, V. C. M. Leung, and X. Li, "In-band full-duplex relaying: A survey, research issues and challenges," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 2, pp. 500–524, 2nd Quart., 2015.

[3] Y. Lv, Z. He, and Y. Rong, "Multiuser multi-hop AF MIMO relay system design based on MMSE-DFE receiver," *IEEE Access*, vol. 7, pp. 42518–42535, Apr. 2019.

[4] A. Gupta and R. K. Jha, "A survey of 5G network: Architecture and emerging technologies," *IEEE Access*, vol. 3, pp. 1206–1232, Aug. 2015.

[5] G. Chen, J. P. Coon, A. Mondal, B. Allen, and J. A. Chambers, "Performance analysis for multihop full-duplex IoT networks subject to Poisson distributed interferers," *IEEE Internet Things J.*, vol. 6, no. 2, pp. 3467–3479, Apr. 2019.

[6] A. Nosratinia, T. E. Hunter, and A. Hedayat, "Cooperative communication in wireless networks," *IEEE Commun. Mag.*, vol. 42, no. 10, pp. 74–80, Oct. 2004.

[7] Y. Rong, "Two-way Compress-and-forward relaying with multiple MIMO relay nodes," *IEEE Commun. Lett.*, vol. 18, no. 8, pp. 1387–1390, Aug. 2014.

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

[9] 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.

[10] Y. Rong and Y. Hua, "Optimality of diagonalization of multi-hop MIMO relays," *IEEE Trans. Wireless Commun.*, vol. 8, no. 12, pp. 6068–6077, Dec. 2009.

[11] Y. Rong and Y. Hua, "Optimality of diagonalization of multicarrier multi-hop linear non-regenerative MIMO relays," in *Proc. IEEE Wireless Commun. Netw. Conf.*, Sydney, NSW, Australia, Apr. 2010, pp. 1–6.

[12] C. Xing, S. Ma, and Y.-C. Wu, "Robust joint design of linear relay precoder and destination equalizer for dual-hop amplify-and-forward MIMO relay systems," *IEEE Trans. Signal Process.*, vol. 58, no. 4, pp. 2273–2283, Apr. 2010.

[13] K. X. Nguyen, Y. Rong, and S. Nordholm, "MMSE-based transceiver design algorithms for interference MIMO relay systems," *IEEE Trans. Wireless Commun.*, vol. 14, no. 11, pp. 6414–6424, Nov. 2015.

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

[15] Z. He, X. Zhang, Y. Bi, W. Jiang, and Y. Rong, "Optimal source and relay design for multiuser MIMO AF relay communication systems with direct links and imperfect channel information," *IEEE Trans. Wireless Commun.*, vol. 15, no. 3, pp. 2025–2038, Mar. 2016.

[16] 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.

[17] H. Shen, W. Xu, and C. Zhao, "A semi-closed form solution to MIMO relaying optimization with source-destination link," *IEEE Signal Process. Lett.*, vol. 23, no. 2, pp. 247–251, Feb. 2016.

[18] B. Rankov and A. Wittneben, "Spectral efficient protocols for half-duplex fading relay channels," *IEEE J. Sel. Areas Commun.*, vol. 25, no. 2, pp. 379–389, Feb. 2007.

[19] R. Zhang, Y.-C. Liang, C. C. Chai, and S. Cui, "Optimal beamforming for two-way multi-antenna relay channel with analogue network coding," *IEEE J. Sel. Areas Commun.*, vol. 27, no. 5, pp. 699–712, Jun. 2009.

[20] Y. Rong, "Joint source and relay optimization for two-way linear non-regenerative MIMO relay communications," *IEEE Trans. Signal Process.*, vol. 60, no. 12, pp. 6533–6546, Dec. 2012.

[21] J. Zou, H. Luo, M. Tao, and R. Wang, "Joint source and relay optimization for non-regenerative MIMO two-way relay systems with imperfect CSI," *IEEE Trans. Wireless Commun.*, vol. 11, no. 9, pp. 3305–3315, Sep. 2012.

[22] K.-J. Lee, H. Sung, E. Park, and I. Lee, "Joint optimization for one and two-way MIMO AF multiple-relay systems," *IEEE Trans. Wireless Commun.*, vol. 9, no. 12, pp. 3671–3681, Dec. 2010.

[23] Y. Rong, "Joint source and relay optimization for two-way MIMO multi-relay networks," *IEEE Commun. Lett.*, vol. 15, no. 12, pp. 1329–1331, Dec. 2011.

[24] C. A. Belfiore and J. H. Park, Jr., "Decision feedback equalization," *Proc. IEEE*, vol. 67, no. 8, pp. 1143–1156, Aug. 1979.

[25] E. Biglieri, G. Taricco, and A. Tulino, "Decoding space-time codes with BLAST architectures," *IEEE Trans. Signal Process.*, vol. 50, no. 10, pp. 2547–2552, Oct. 2002.

[26] J. M. Cioffi, G. P. Dudevoir, M. V. Eyuboglu, and G. D. Forney, Jr., "MMSE decision-feedback equalizers and coding—Part I: Equalization results," *IEEE Trans. Commun.*, vol. 43, no. 10, pp. 2582–2594, Oct. 1995.

[27] J. M. Cioffi, G. P. Dudevoir, M. V. Eyuboglu, and G. D. Forney, Jr., "MMSE decision-feedback equalizers and coding—Part II: Coding results," *IEEE Trans. Commun.*, vol. 43, no. 10, pp. 2595–2604, Oct. 1995.

[28] 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.

[29] A. Toding, M. R. A. Khandaker, and Y. Rong, "Joint source and relay optimization for parallel MIMO relays using MMSE-DFE receiver," in *Proc. 16th Asia–Pacific Conf. Commun. (APCC)*, Auckland, New Zealand, Oct. 2010, pp. 12–16.

[30] 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.

[31] H. Liu, Z. Bai, and K. Sup Kwak, "Precoding design for two-way MIMO relay networks based on CP-SC transmission," in *Proc. 14th Int. Symp. Commun. Inf. Technol. (ISCIT)*, Incheon, South Korea, Sep. 2014, pp. 6–10.

[32] Y. Zhu and K. Letaief, "Single carrier frequency domain equalization with time domain noise prediction for wideband wireless communications," *IEEE Trans. Wireless Commun.*, vol. 5, no. 12, pp. 3548–3557, Dec. 2006.

[33] A. Goldsmith, *Wireless Communications*. Cambridge, U.K.: Cambridge Univ. Press, Aug. 2005, sec. 14.2.

[34] D. P. Bertsekas, *Nonlinear Programming*, 2nd ed. Belmont, MA, USA: Athena Scientific, 1999.

[35] Z. Wen, D. Goldfarb, and K. Scheinberg, "Block coordinate descent methods for semidefinite programming," in *Handbook on Semidefinite, Conic and Polynomial Optimization* (International Series in Operations Research & Management Science), vol. 166, M. F. Anjos and J. B. Lasserre, Eds. New York, NY, USA: Springer, 2012, sec. 19.1.1.

[36] 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.

[37] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge, U.K.: Cambridge Univ. Press, 2005.

[38] H. Meyr, M. Moeneclaey, and S. A. Fechtel, *Digital Communication Receivers: Synchronization, Channel Estimation, and Signal Processing*. New York, NY, USA: Wiley, 1997, chs. 14–15.

[39] C. W. R. Chiong, Y. Rong, and Y. Xiang, "Channel estimation for two-way MIMO relay systems in frequency-selective fading environments," *IEEE Trans. Wireless Commun.*, vol. 14, no. 1, pp. 399–409, Jan. 2015.

[40] J.-M. Kang and H.-M. Kim, "Training designs for estimation of spatially correlated fading channels in MIMO amplify-and-forward two-way multi-relay networks," *IEEE Commun. Lett.*, vol. 20, no. 4, pp. 772–775, Apr. 2016.

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

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

[43] C.-C. Hu and Y.-F. Chou, "Precoding design of MIMO AF two-way multiple-relay systems," *IEEE Signal Process. Lett.*, vol. 20, no. 6, pp. 623–626, Jun. 2013.

[44] F. Zesong, L. Chen, X. Chengwen, and K. Jingming, "Joint linear processing design for distributed two-way amplify-and-forward MIMO relaying networks," *China Commun.*, vol. 10, no. 7, pp. 126–133, Jul. 2013.

[45] M. Zhang, H. Yi, H. Yu, H. Luo, and W. Chen, "Joint optimization in bidirectional multi-user multi-relay MIMO systems: Non-robust and robust cases," *IEEE Trans. Veh. Technol.*, vol. 62, no. 7, pp. 3228–3244, Sep. 2013.

[46] C.-C. Hu, G.-F. Liu, and B.-H. Chen, "Joint relay/antenna selection and precoding design for two-way MIMO amplify-and-forward relaying systems," *IEEE Trans. Veh. Technol.*, vol. 65, no. 7, pp. 4854–4864, Jul. 2016.

[47] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[48] R. A. Horn and C. R. Johnson, *Matrix Analysis*, 2nd ed. New York, USA: Cambridge Univ. Press, 2013.

[49] L. N. Trefethen and D. Bau, III, *Numerical Linear Algebra*. Philadelphia, PA, USA: SIAM, 1997.

[50] M. C. Grant and S. P. Boyd. The CVX Users' Guide. Release 2.1, CVX Research, Dec. 2018. [Online]. Available: http://web.cvxr.com/cvx/doc/CVX.pdf

[51] R. A. Horn and C. R. Johnson, *Topics in Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1991.

[52] R. L. Burden, J. D. Faires, and A. M. Burden, *Numerical Analysis*, 10th ed. Boston, MA, USA: Cengage Learning, 2015, sec. 2.1, pp. 48–54.

[53] Y. Nesterov and A. Nemirovskii, *Interior-Point Polynomial Algorithms in Convex Programming*. Philadelphia, PA, USA: SIAM, 1994, secs. 6.1–6.2, pp. 217–229.

[54] A. Gupta and D. Nagar, *Matrix Variate Distributions*. Boca Raton, Florida, USA: CRC Press, 2000.

[55] W. Peng, S. Ma, T.-S. Ng, J. Wang, and F. Adachi, "An analytical approach to V-BLAST detection with optimal ordering for two input multiple output systems," in *Proc. IEEE 68th Veh. Technol. Conf. (VTC)*, Calgary, BC, Canada, Sep. 2008, pp. 1–5.

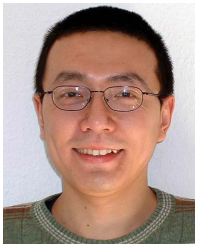
[56] A. Hjorungnes and D. Gesbert, "Complex-valued matrix differentiation: Techniques and key results," *IEEE Trans. Signal Process.*, vol. 55, no. 6, pp. 2740–2746, Jun. 2007.

[57] K. B. Petersen and M. S. Pedersen, *The Matrix Cookbook*. Nov. 2012, ch. 4, pp. 24–26. [Online]. Available: http://www2.imm.dtu.dk/pubdb/edoc/imm3274.pdf

[58] H. V. Henderson and S. R. Searle, "On deriving the inverse of a sum of matrices," *SIAM Rev.*, vol. 23, no. 1, pp. 53–60, Jan. 1981.

Yang Lv (Graduate Student Member, IEEE) was born in Shenyang, China, in 1993. He received the B.E. degree in information engineering from the Beijing University of Posts and Telecommunications (BUPT), Beijing, China, in 2016, where he is currently pursuing the Ph.D. degree in information and communication engineering. His research interests include signal and information processing for wireless communications, applications of matrix, probability and optimization theories, underwater acoustic communications, the Internet of Things and device-to-device communications, mobile edge computing, and wireless networks empowered by reconfigurable intelligent surfaces.

Zhiqiang He (Member, IEEE) received the B.E. degree in signal and information processing and the Ph.D. degree (Hons.) in signal and information processing from the Beijing University of Posts and Telecommunications (BUPT), Beijing, China, in 1999 and 2004, respectively. Since 2004, he has been with the School of Information and Communication Engineering, BUPT, where he is currently a Professor with the Center of Information Theory and Technology. His research interests include signal and information processing in wireless communications, networking architecture and protocol design, machine learning, and underwater acoustic communications.



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, 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 170 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 at the 2011 International Conference on Wireless Communications and Signal Processing, the Best Paper Award at the 2010 Asia-Pacific Conference on Communications, and the Young Researcher of the Year Award of the Faculty of Science and Engineering, Curtin University, in 2010. He is a Senior Area Editor of the IEEE TRANSACTIONS ON SIGNAL PROCESSING, and was an Associate Editor of the IEEE TRANSACTIONS ON SIGNAL PROCESSING from 2014 to 2018, and an Editor of the IEEE WIRELESS COMMUNICATIONS LETTERS from 2012 to 2014. He was also a Guest Editor of the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS Special Issue on "Theories and Methods for Advanced Wireless Relays."