New Results on Transceiver Design for Two-hop Amplify-and-forward MIMO Relay Systems with Direct Link

Zhiqiang He, *Member, IEEE*, Jinnian Zhang, Wanning Liu, and Yue Rong, *Senior Member, IEEE*

Abstract—Conventional amplify-and-forward (AF) protocol for half-duplex two-hop multiple-input multiple-output (MIMO) relay systems assumes that the source node transmits signal only at the first time slot. While making the source node silent at the second time slot simplifies the system design, it is strictly suboptimal. To improve the system performance, in this paper, we consider that the source node transmits signals during both time slots. We develop two novel iterative algorithms to optimize the source, relay, and receiver matrices in this new AF MIMO relay system. Both algorithms are based on the minimum mean-squared error (MMSE) criterion. In particular, the first algorithm is applicable for general MIMO relay systems with multiple concurrent data streams, where the source, relay, and receiver matrices are optimized in an alternating fashion till convergence. The second algorithm is developed for MIMO relay systems with a single data stream, where the source precoding vectors and the relay precoding matrix are optimized iteratively and the receiver matrix is obtained after the convergence of the source vectors and the relay matrix. Simulation results show that compared with conventional AF MIMO relay systems, the proposed system provides better bit-error-rate performance for both multiple data streams and single data stream cases.

Index Terms—Amplify-and-forward, direct link, MIMO relay, MMSE, two-hop.

I. INTRODUCTION

Cooperative relay communication has attracted much interest in recent years from both academia and industry due to its potential in increasing the coverage, throughput, and capacity of wireless communication systems [1]-[4]. Relay nodes in a cooperative communication system can be nonregenerative or regenerative [2]. In a nonregenerative relay system, the relay node simply amplifies (including a linear transformation) the received signals and forwards the amplified signals to the

Z. He, J. Zhang, and W. Liu are with the Key Laboratory of Universal Wireless Communication, Ministry of Education, Beijing University of Posts and Telecommunications, Beijing 100876, China (e-mails: hezq@bupt.edu.cn; zhangjinnian@bupt.edu.cn; wanninglu@126.com).

Y. Rong is with the Department of Electrical and Computer Engineering, Curtin University, Bentley, WA 6102, Australia (e-mail: y.rong@curtin.edu.au).

destination node. Therefore, the complexity of the amplify-and-forward (AF) protocol is much lower than that of the regenerative relay schemes, particularly when multiple-input multiple-output (MIMO) relay systems are considered.

Capacity bounds of AF MIMO relay systems have been derived in [5] and [6]. The relay precoding matrix maximizing the source-destination mutual information (MI) of a two-hop AF MIMO relay system has been developed in [7] and [8]. In [9]-[11], the relay precoding matrix that minimizes the mean-squared error (MSE) of the signal waveform estimation was proposed. In [12], a unified framework has been established to optimize the source and relay precoding matrices of linear AF MIMO relay systems with a broad class of commonly used objective functions. A recent tutorial on the transceiver design for AF MIMO relay communication systems can be found in [13].

In the transceiver design works [9]-[12], the direct link between the source and destination nodes is not considered. However, in practical systems, the direct link provides valuable spatial diversity, and thus, should be properly considered in the MIMO relay system design. In [7] and [8], suboptimal structures of the relay precoding matrix have been derived considering the direct link. The optimal structure of the relay precoding matrix for AF MIMO relay systems in the presence of the direct link has been derived in [14]. Joint source and relay precoding matrices optimization with the direct link based on the MSE matrix diagonalization has been proposed in [15] and [16]. Tomlinson-Harashima source precoding matrix and linear relay precoding matrix design has been studied in [17]. A closed-form design of the relay precoding matrix has been proposed in [18]. Recently, a semi-closed form solution to the source beamformer optimization problem with direct link has been presented in [19]. Joint source and relay matrices optimization for multiuser AF MIMO relay networks with direct links has been investigated in [20]. Taking into account the mismatch between the true and the estimated channel state information (CSI), robust transceiver design algorithms for AF MIMO relay systems with the direct link have been proposed in [21]-[23].

It is notable that in the conventional AF relay protocol on half-duplex two-hop MIMO relay systems adopted in [7]-[23], the source node transmits signal only at the first time slot. While making the source node silent at the second time slot simplifies the system design, it is strictly suboptimal [24]. Theoretically, additional diversity gains can be achieved by exploiting the signals transmitted by the source node through

the direct link at the second time slot.

In this paper, we consider two-hop AF MIMO relay systems with the direct source-destination link. Different from the conventional AF relay protocol [7]-[23], the source node transmits signals during both time slots. Thus, valuable time diversity provided by the direct link can be utilized to improve the system performance. We investigate the joint source, relay, and receiver matrices design of this system to minimize the MSE of the signal waveform estimation at the destination node. Note that this new AF relay protocol was first proposed in [24] for single-input single-output relay systems. However, the matrices optimization problem in this paper is more difficult to solve than the problem with scalar variables in [24]. Compared with existing works [7]-[23], the transceiver optimization problem in this paper is more challenging as we need to optimize the source precoding matrices at two time slots, rather than only the first time slot. Since the joint source, relay, and receiver optimization problem is non-convex with matrix variables, the globally optimal solution is intractable to obtain. To solve this problem, we present two iterative algorithms.

The first algorithm is applicable for general AF MIMO relay systems with multiple concurrent data streams. In this algorithm, the source, relay, and receiver matrices are optimized in an alternating fashion till convergence. In particular, we show that with given source and receiver matrices, the optimal relay precoding matrix has a closed-form solution. While with fixed relay and receiver matrices, the two source precoding matrices can be optimized through solving a quadratically constrained quadratic programming (QCQP) problem.

The second algorithm is developed for AF MIMO relay systems with a single data stream, where the source precoding vectors and the relay precoding matrix are optimized iteratively and the receiver matrix is obtained after the convergence of the source vectors and the relay matrix. Interestingly, we show that in this case, the relay precoding matrix can be optimized through the semidefinite relaxation (SDR) technique together with the Charnes-Cooper transformation, and the two source precoding vectors can be optimized through the SDR technique. Simulation results show that compared with conventional AF MIMO relay systems, the proposed system provides better bit-error-rate (BER) performance for both multiple data streams and single data stream cases.

The rest of this paper is organized as follows. In Section II, the model of a general two-hop linear AF MIMO relay system with the direct link is presented, where the source node transmits signals at both time slots. The joint source, relay, and receiver optimization problem is also formulated in this section. In Section III, we present an iterative transceiver design algorithm for general AF MIMO relay systems with multiple concurrent data streams. The transceiver optimization algorithm for the particular case of a single data stream system is developed in Section IV. Simulation results are shown in Section V to verify the benefit of the proposed algorithms. Finally, conclusions are drawn in Section VI.
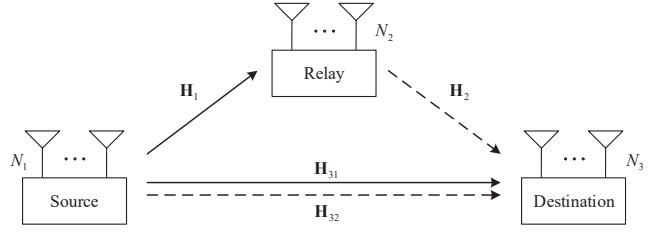

Fig. 1. A two-hop MIMO relay communication system with the direct source-destination link.

II. SYSTEM MODEL

We consider a three-node two-hop MIMO relay communication system as shown in Fig. 1, where the source node (node 1) transmits information to the destination node (node 3) with the aid of a relay node (node 2). We assume that the three nodes are equipped with $N_i, i = 1, 2, 3$ antennas, respectively. With a practical half-duplex relay node, the data transmission from the source node to the destination node is completed in two time slots. At the first time slot, the information-carrying source symbol vector $\mathbf{s} = [s_1, s_2, \cdots, s_{N_b}]^T$ is linearly precoded by a matrix $\mathbf{F}_1 \in \mathbb{C}^{N_1 \times N_b}$ and then transmitted to both the relay node and the destination node, where $(\cdot)^T$ stands for the matrix transpose and N_b denotes the number of independent data streams transmitted simultaneously at the source node.

The signal vectors received at the relay and destination nodes can be written respectively as

$$\mathbf{y}_r = \mathbf{H}_1 \mathbf{F}_1 \mathbf{s} + \mathbf{n}_r \tag{1}$$

$$\mathbf{y}_{d1} = \mathbf{H}_{31} \mathbf{F}_1 \mathbf{s} + \mathbf{n}_{d1} \tag{2}$$

where $\mathbf{H}_1 \in \mathbb{C}^{N_2 \times N_1}$ and $\mathbf{H}_{31} \in \mathbb{C}^{N_3 \times N_1}$ are the MIMO channel matrices of the source-relay link and source-destination link at the first time slot, respectively, $\mathbf{n}_r \in \mathbb{C}^{N_2 \times 1}$ and $\mathbf{n}_{d1} \in \mathbb{C}^{N_3 \times 1}$ are the additive noise vectors at the relay node and the destination node at the first time slot, respectively.

During the second time slot, the relay node linearly precodes $\mathbf{y}_r$ with a matrix $\mathbf{F}_2 \in \mathbb{C}^{N_2 \times N_2}$ and then forwards the precoded signal vector to the destination node. Meanwhile, the source node transmits $\mathbf{F}_3 \mathbf{s}$ to the destination node, where $\mathbf{F}_3 \in \mathbb{C}^{N_1 \times N_b}$ is the source precoding matrix at the second time slot. Note that this new AF relay protocol is different to that used in two-hop AF MIMO relay communication systems [7]-[23], where the source node is silent at the second time slot. The signal vector $\mathbf{y}_{d2}$ received at the destination node at the second time slot is given by

$$\mathbf{y}_{d2} = \mathbf{H}_2 \mathbf{F}_2 \mathbf{y}_r + \mathbf{H}_{32} \mathbf{F}_3 \mathbf{s} + \mathbf{n}_{d2} \tag{3}$$

where $\mathbf{H}_2 \in \mathbb{C}^{N_3 \times N_2}$ and $\mathbf{H}_{32} \in \mathbb{C}^{N_3 \times N_1}$ are the MIMO channel matrices of the relay-destination link and the source-destination link at the second time slot, respectively, and $\mathbf{n}_{d2} \in \mathbb{C}^{N_3 \times 1}$ is the additive noise vector at the destination node at the second time slot. We assume that $\mathbf{n}_r$, $\mathbf{n}_{d1}$, and $\mathbf{n}_{d2}$ are independent and identically distributed (i.i.d.) Gaussian noise vectors with zero-mean and unit variance entries.

From (1)-(3), the signal vector received at the destination

node over two time slots can be written as

$$
\begin{aligned}
\mathbf{y} &= \begin{bmatrix} \mathbf{y}_{d2} \\ \mathbf{y}_{d1} \end{bmatrix} \\
&= \begin{bmatrix} \mathbf{H}_2\mathbf{F}_2\mathbf{H}_1\mathbf{F}_1 + \mathbf{H}_{32}\mathbf{F}_3 \\ \mathbf{H}_{31}\mathbf{F}_1 \end{bmatrix} \mathbf{s} + \begin{bmatrix} \mathbf{H}_2\mathbf{F}_2\mathbf{n}_r + \mathbf{n}_{d2} \\ \mathbf{n}_{d1} \end{bmatrix} \\
&= \mathbf{G}\mathbf{s} + \mathbf{v} \tag{4}
\end{aligned}
$$

where $\mathbf{G} = \begin{bmatrix} \mathbf{H}_2\mathbf{F}_2\mathbf{H}_1\mathbf{F}_1 + \mathbf{H}_{32}\mathbf{F}_3 \\ \mathbf{H}_{31}\mathbf{F}_1 \end{bmatrix}$ is the equivalent source-destination MIMO channel matrix and $\mathbf{v} = \begin{bmatrix} \mathbf{H}_2\mathbf{F}_2\mathbf{n}_r + \mathbf{n}_{d2} \\ \mathbf{n}_{d1} \end{bmatrix}$ is the equivalent noise vector at the destination node. It is worth noting that (4) represents the most general case for a half-duplex three-node two-hop AF MIMO relay system.

- If $\mathbf{F}_3 = \mathbf{0}$, we have a conventional AF MIMO relay system with the direct link [14]-[23], where the source node is silent at the second time slot.
- AF MIMO relay systems where the direct link is neglected [9]-[12] correspond to (4) with $\mathbf{H}_{31} = \mathbf{H}_{32} = \mathbf{0}$ and $\mathbf{n}_{d1} = \mathbf{0}$.
- For slow-fading MIMO relay channels, there is $\mathbf{H}_{31} = \mathbf{H}_{32}$, while for the fast-fading environment, there is $\mathbf{H}_{31} \neq \mathbf{H}_{32}$.
- A suboptimal scheme is to set $\mathbf{F}_1 = \mathbf{F}_3$, i.e., the source node transmits signals in both time slots using the same precoding matrix. This may reduce the computational complexity of transceiver optimization as only $\mathbf{F}_1$ and $\mathbf{F}_2$ need to be optimized. However, setting $\mathbf{F}_1 = \mathbf{F}_3$ is strictly suboptimal, because even for slow fading MIMO relay channels (i.e., $\mathbf{H}_{31} = \mathbf{H}_{32}$), the channels experienced by the signals sent from the source node are different over two time slots ($\begin{pmatrix} \mathbf{H}_2\mathbf{F}_2\mathbf{H}_1 \\ \mathbf{H}_{31} \end{pmatrix}$ and $\mathbf{H}_{31}$ during the first and second time slot, respectively). Therefore, different source precoding matrices should be used over two time slots to optimize the system performance.

For simplicity, we assume a perfect knowledge of the CSI of $\mathbf{H}_1$, $\mathbf{H}_2$, $\mathbf{H}_{31}$, and $\mathbf{H}_{32}$. In practice, the destination node can obtain the knowledge of $\mathbf{H}_2$, $\mathbf{H}_{31}$ and $\mathbf{H}_{32}$ through channel training. The CSI of $\mathbf{H}_1$, which is obtained at the relay node by channel training, can also be transmitted to the destination node. Therefore, by using all CSI obtained, the destination node can perform the transceiver optimization, and then feed back the optimized $\mathbf{F}_1$ and $\mathbf{F}_3$ to the source node and $\mathbf{F}_2$ to the relay node. In the presence of mismatch between the true and the estimated CSI, similar to [21]-[23], robust transceiver design can be carried out based on the transceiver optimization approaches using the exact CSI presented in the following sections.

Due to its simplicity, a linear receiver is used at the destination node to retrieve the source signal vector. Thus, the estimated source signal vector can be written as

$$
\hat{\mathbf{s}} = \mathbf{W}^H\mathbf{y} \tag{5}
$$

where $\mathbf{W} \in \mathbb{C}^{2N_3 \times N_b}$ is the weight matrix of the linear receiver and $(\cdot)^H$ stands for the matrix Hermitian transpose.

From (5), the MSE matrix of the signal waveform estimation is given by

$$
\begin{aligned}
&\mathbf{E}(\mathbf{W}, \mathbf{F}_1, \mathbf{F}_2, \mathbf{F}_3) \\
&= E[(\mathbf{W}^H\mathbf{y} - \mathbf{s})(\mathbf{W}^H\mathbf{y} - \mathbf{s})^H] \\
&= (\mathbf{W}^H\mathbf{G} - \mathbf{I}_{N_b})(\mathbf{W}^H\mathbf{G} - \mathbf{I}_{N_b})^H + \mathbf{W}^H\mathbf{C}_v\mathbf{W} \tag{6}
\end{aligned}
$$

where $\mathbf{I}_n$ stands for an $n \times n$ identity matrix, $E[\cdot]$ is the statistical expectation with respect to signal and noise, and

$$
\mathbf{C}_v = E[\mathbf{v}\mathbf{v}^H] = \begin{bmatrix} \mathbf{H}_2\mathbf{F}_2\mathbf{F}_2^H\mathbf{H}_2^H + \mathbf{I}_{N_3} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{N_3} \end{bmatrix}
$$

is the noise covariance matrix. Here we assume $E[\mathbf{s}\mathbf{s}^H] = \mathbf{I}_{N_b}$.

From (1), the power of the signal transmitted at the relay node is given by

$$
tr(E[\mathbf{y}_r\mathbf{y}_r^H]) = tr(\mathbf{F}_2(\mathbf{H}_1\mathbf{F}_1\mathbf{F}_1^H\mathbf{H}_1^H + \mathbf{I}_{N_2})\mathbf{F}_2^H) \tag{7}
$$

where $tr(\cdot)$ stands for the matrix trace. The total power consumption at the source node over two time slots is

$$
tr(\mathbf{F}_1\mathbf{F}_1^H + \mathbf{F}_3\mathbf{F}_3^H). \tag{8}
$$

The aim of the transceiver design is to minimize the MSE of the signal waveform estimation while satisfying the transmission power constraints at the source and the relay nodes. From (6)-(8), the source, relay, and receiver matrices optimization problem can be written as

$$
\min_{\mathbf{W}, \mathbf{F}_1, \mathbf{F}_2, \mathbf{F}_3} tr(\mathbf{E}(\mathbf{W}, \mathbf{F}_1, \mathbf{F}_2, \mathbf{F}_3)) \tag{9}
$$

$$
\text{s.t. } tr(\mathbf{F}_1\mathbf{F}_1^H + \mathbf{F}_3\mathbf{F}_3^H) \leq P_s \tag{10}
$$

$$
tr(\mathbf{F}_2(\mathbf{H}_1\mathbf{F}_1\mathbf{F}_1^H\mathbf{H}_1^H + \mathbf{I}_{N_2})\mathbf{F}_2^H) \leq P_r \tag{11}
$$

where P_s and P_r denote the transmission power available at the source node and the relay node, respectively. Note that (10) guarantees that the source node in the new AF MIMO relay system consumes the same amount of power as that of the conventional system where the source node is silent at the second time slot. Compared with existing works [7]-[23], the introduction of the source precoding matrix $\mathbf{F}_3$ at the second time slot makes the problem (9)-(11) much more challenging to solve. Note that the problem (9)-(11) is nonconvex with matrix variables, and the globally optimal solution is intractable to obtain. In the next section, we develop an iterative algorithm to solve the problem (9)-(11).

III. Transceiver Design Algorithm for General AF MIMO Relay Systems

In this section, we develop a novel algorithm to optimize $\mathbf{W}$ and $\mathbf{F}_i$, $i = 1, 2, 3$ alternatingly till convergence as follows. As is well known, for any given precoding matrices $\mathbf{F}_i$, $i = 1, 2, 3$, the optimal $\mathbf{W}$ minimizing (9) is the Wiener filter [25] given by

$$
\mathbf{W} = \left(\mathbf{G}\mathbf{G}^H + \mathbf{C}_v\right)^{-1}\mathbf{G} \tag{12}
$$

where $(\cdot)^{-1}$ denotes matrix inversion.

To optimize the relay precoding matrix, let us first rewrite (9) as a function of $\mathbf{F}_2$ as

$$tr(\mathbf{E}(\mathbf{F}_2)) =$$
$$tr((\mathbf{W}_1^H \mathbf{H}_2 \mathbf{F}_2 \mathbf{H}_1 \mathbf{F}_1 + \mathbf{W}_1^H \mathbf{H}_{32} \mathbf{F}_3 + \mathbf{W}_2^H \mathbf{H}_{31} \mathbf{F}_1 - \mathbf{I}_{N_b})$$
$$\times (\mathbf{W}_1^H \mathbf{H}_2 \mathbf{F}_2 \mathbf{H}_1 \mathbf{F}_1 + \mathbf{W}_1^H \mathbf{H}_{32} \mathbf{F}_3 + \mathbf{W}_2^H \mathbf{H}_{31} \mathbf{F}_1 - \mathbf{I}_{N_b})^H$$
$$+ \mathbf{W}_1^H \mathbf{H}_2 \mathbf{F}_2 \mathbf{F}_2^H \mathbf{H}_2^H \mathbf{W}_1 + \mathbf{W}^H \mathbf{W}) \qquad (13)$$

where $\mathbf{W}_1$ and $\mathbf{W}_2$ contain the first and last N_3 rows of $\mathbf{W}$, respectively, i.e., $\mathbf{W} = \begin{bmatrix} \mathbf{W}_1^T, \mathbf{W}_2^T \end{bmatrix}^T$. It can be seen from (13) that with given $\mathbf{W}$, $\mathbf{F}_1$, and $\mathbf{F}_3$, the problem (9)-(11) is converted to the following problem of optimizing $\mathbf{F}_2$

$$\min_{\mathbf{F}_2} \; tr((\breve{\mathbf{H}}_2 \mathbf{F}_2 \breve{\mathbf{H}}_1 - \boldsymbol{\Pi})(\breve{\mathbf{H}}_2 \mathbf{F}_2 \breve{\mathbf{H}}_1 - \boldsymbol{\Pi})^H + \breve{\mathbf{H}}_2 \mathbf{F}_2 \mathbf{F}_2^H \breve{\mathbf{H}}_2^H) \quad (14)$$

$$\text{s.t.} \; tr(\mathbf{F}_2(\breve{\mathbf{H}}_1 \breve{\mathbf{H}}_1^H + \mathbf{I}_{N_2})\mathbf{F}_2^H) \le P_r \qquad (15)$$

where $\breve{\mathbf{H}}_1 = \mathbf{H}_1 \mathbf{F}_1$, $\breve{\mathbf{H}}_2 = \mathbf{W}_1^H \mathbf{H}_2$, $\boldsymbol{\Pi} = \mathbf{I}_{N_b} - \mathbf{W}_1^H \mathbf{H}_{32} \mathbf{F}_3 - \mathbf{W}_2^H \mathbf{H}_{31} \mathbf{F}_1$. The problem (14)-(15) is convex and can be solved by the Lagrange multiplier method as shown in Appendix. The optimal $\mathbf{F}_2$ can be expressed as

$$\mathbf{F}_2 = (\breve{\mathbf{H}}_2^H \breve{\mathbf{H}}_2 + \lambda \mathbf{I}_{N_2})^{-1} \breve{\mathbf{H}}_2^H \boldsymbol{\Pi} \breve{\mathbf{H}}_1^H (\breve{\mathbf{H}}_1 \breve{\mathbf{H}}_1^H + \mathbf{I}_{N_2})^{-1}. \quad (16)$$

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

It is clear that λ can be determined through the power constraint (15) and the following complementary slackness condition associated with the problem (14)-(15)

$$\lambda(tr(\mathbf{F}_2(\breve{\mathbf{H}}_1 \breve{\mathbf{H}}_1^H + \mathbf{I}_{N_2})\mathbf{F}_2^H) - P_r) = 0. \qquad (17)$$

It can be seen from (17) that if $\mathbf{F}_2$ in (16) with $\lambda = 0$ satisfies the constraint (15), then $\mathbf{F}_2 = (\breve{\mathbf{H}}_2^H \breve{\mathbf{H}}_2)^{-1} \breve{\mathbf{H}}_2^H \boldsymbol{\Pi} \breve{\mathbf{H}}_1^H (\breve{\mathbf{H}}_1 \breve{\mathbf{H}}_1^H + \mathbf{I}_{N_2})^{-1}$ is the optimal solution to the problem (14)-(15). Otherwise, there must exist $\lambda > 0$ such that

$$tr(\mathbf{F}_2(\breve{\mathbf{H}}_1 \breve{\mathbf{H}}_1^H + \mathbf{I}_{N_2})\mathbf{F}_2^H) = P_r. \qquad (18)$$

Let us introduce $\breve{\mathbf{H}}_2 = \mathbf{U}_2 \boldsymbol{\Lambda}_2 \mathbf{V}_2^H$ as the singular value decomposition (SVD) of $\breve{\mathbf{H}}_2$. By substituting (16) into (18), we obtain

$$tr(\boldsymbol{\Lambda}_2(\boldsymbol{\Lambda}_2^2 + \lambda \mathbf{I}_{N_b})^{-1} \boldsymbol{\Gamma}(\boldsymbol{\Lambda}_2^2 + \lambda \mathbf{I}_{N_b})^{-1} \boldsymbol{\Lambda}_2) = P_r \qquad (19)$$

where $\boldsymbol{\Gamma} = \mathbf{U}_2^H \boldsymbol{\Pi} \breve{\mathbf{H}}_1^H (\breve{\mathbf{H}}_1 \breve{\mathbf{H}}_1^H + \mathbf{I}_{N_2})^{-1} \breve{\mathbf{H}}_1 \boldsymbol{\Pi}^H \mathbf{U}_2$. By denoting μ_i and γ_i as the ith diagonal elements of $\boldsymbol{\Lambda}_2$ and $\boldsymbol{\Gamma}$ respectively, (19) can be rewritten as

$$\sum_{i=1}^{N_b} \frac{\mu_i^2 \gamma_i}{(\mu_i^2 + \lambda)^2} = P_r. \qquad (20)$$

As the left-hand side (LHS) of (20) is a monotonically decreasing function of λ, the bisection method [26] can be applied to solve (20) to obtain λ.

With given $\mathbf{W}$ and $\mathbf{F}_2$, the MSE function (13) can be rewritten as the following function of $\mathbf{F}_1$ and $\mathbf{F}_3$

$$tr(\mathbf{E}(\mathbf{F}_1, \mathbf{F}_3)) = tr((\mathbf{DB} - \mathbf{I}_{N_b})(\mathbf{DB} - \mathbf{I}_{N_b})^H + \mathbf{W}^H \mathbf{C}_v \mathbf{W})$$
$$\qquad (21)$$

where $\mathbf{D} = \begin{bmatrix} \mathbf{W}_1^H \mathbf{H}_2 \mathbf{F}_2 \mathbf{H}_1 + \mathbf{W}_2^H \mathbf{H}_{31}, \mathbf{W}_1^H \mathbf{H}_{32} \end{bmatrix}$, $\mathbf{B} = [\mathbf{F}_1^T, \mathbf{F}_3^T]^T$. Moreover, the LHS of (10) becomes $tr(\mathbf{B}^H \mathbf{B})$, while the LHS of (11) becomes $tr(\mathbf{B}^H \mathbf{A}_0 \mathbf{B} + \mathbf{F}_2 \mathbf{F}_2^H))$, where

1) Initialize the algorithm with $\mathbf{F}_1^{(0)} = \mathbf{F}_3^{(0)} = \sqrt{P_s/2N_b}[\mathbf{I}_{N_b}, \mathbf{0}]^T$ and $\mathbf{F}_2^{(0)} = \sqrt{P_r/tr(\mathbf{H}_1 \mathbf{F}_1^{(0)}(\mathbf{F}_1^{(0)})^H \mathbf{H}_1^H + \mathbf{I}_{N_2})}\mathbf{I}_{N_2}$, and set $n = 0$.
2) Update $\mathbf{W}^{(n)}$ using $\mathbf{F}_1^{(n)}$, $\mathbf{F}_2^{(n)}$, and $\mathbf{F}_3^{(n)}$ as (12).
3) Update $\mathbf{F}_2^{(n+1)}$ using $\mathbf{W}^{(n)}$, $\mathbf{F}_1^{(n)}$, and $\mathbf{F}_3^{(n)}$ as (16) and (20).
4) Update $\mathbf{F}_1^{(n+1)}$ and $\mathbf{F}_3^{(n+1)}$ using $\mathbf{W}^{(n)}$ and $\mathbf{F}_2^{(n+1)}$ by solving the problem (22)-(24).
5) If $(\text{mse}_1^{(n)} - \text{mse}_1^{(n+1)})/\text{mse}_1^{(n)} \le \varepsilon$, iteration ends; otherwise go to step (2).

$\mathbf{A}_0 = \begin{bmatrix} \mathbf{H}_1^H \mathbf{F}_2^H \mathbf{F}_2 \mathbf{H}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$. Thus, the problem of optimizing $\mathbf{F}_1$ and $\mathbf{F}_3$ can be written as

$$\min_{\mathbf{B}} \; tr((\mathbf{DB} - \mathbf{I}_{N_b})(\mathbf{DB} - \mathbf{I}_{N_b})^H) \qquad (22)$$

$$\text{s.t.} \; tr(\mathbf{B}^H \mathbf{B}) \le P_s \qquad (23)$$

$$tr(\mathbf{B}^H \mathbf{A}_0 \mathbf{B}) \le P_r - tr(\mathbf{F}_2 \mathbf{F}_2^H). \qquad (24)$$

The problem (22)-(24) is a QCQP problem, which can be efficiently solved by the disciplined convex programming toolbox CVX [27]. We would like to note that the conditional updates of $\mathbf{W}$ and $\mathbf{F}_i$, $i = 1, 2, 3$, may either decrease or maintain but cannot increase the objective function (9). Monotonic convergence of $\mathbf{W}$ and $\mathbf{F}_i$, $i = 1, 2, 3$ towards (at least) a stationary point follows directly from this observation.

The procedure of the proposed iterative algorithm is summarized in Table I, where the superscript (n) denotes variables at the nth iteration, ε is a small positive number for which convergence is acceptable, and mse_1 is the MSE calculated from (9). With the decrease of ε, the system MSE performance improves, whereas the computational complexity increases.

IV. TRANSCEIVER OPTIMIZATION FOR THE SPECIAL CASE OF A SINGLE DATA STREAM

The source, relay, and receiver matrices optimization algorithm developed in the last section is applicable for the general case of multiple data streams. In this section, we develop a novel algorithm for the MSE minimization problem for the special case of single data stream two-hop AF MIMO relay systems.

In the special case of a single data stream, the precoding matrices $\mathbf{F}_1$ and $\mathbf{F}_3$ at the source node and the weight matrices $\mathbf{W}_1$ and $\mathbf{W}_2$ at the destination node degenerate to vectors, which are denoted as $\mathbf{f}_1$, $\mathbf{f}_3$, $\mathbf{w}_1$, and $\mathbf{w}_2$, respectively. From (16) we have $\breve{\mathbf{h}}_1 = \mathbf{H}_1 \mathbf{f}_1$, $\breve{\mathbf{h}}_2 = \mathbf{w}_1^H \mathbf{H}_2$ (a row vector), and

$$\mathbf{F}_2 = \eta(\breve{\mathbf{h}}_2^H \breve{\mathbf{h}}_2 + \lambda \mathbf{I}_{N_2})^{-1} \breve{\mathbf{h}}_2^H \breve{\mathbf{h}}_1^H (\breve{\mathbf{h}}_1 \breve{\mathbf{h}}_1^H + \mathbf{I}_{N_2})^{-1} \quad (25)$$

$$= \alpha \beta \eta \breve{\mathbf{h}}_2^H \breve{\mathbf{h}}_1^H \qquad (26)$$

$$= \mathbf{b} \mathbf{f}_1^H \mathbf{H}_1^H \qquad (27)$$

where $\alpha = 1/(\lambda + \breve{\mathbf{h}}_2 \breve{\mathbf{h}}_2^H)$, $\beta = 1/(1 + \breve{\mathbf{h}}_1^H \breve{\mathbf{h}}_1)$, $\eta = 1 - \mathbf{w}_2^H \mathbf{H}_{31} \mathbf{f}_1 - \mathbf{w}_1^H \mathbf{H}_{32} \mathbf{f}_3$, and $\mathbf{b} = \alpha \beta \eta \breve{\mathbf{h}}_2^H$. From (25) to (26), we applied the identity that for a vector $\mathbf{a}$, there is

$$(\mathbf{a}\mathbf{a}^H + \lambda \mathbf{I})^{-1} \mathbf{a} = \frac{1}{\mathbf{a}^H \mathbf{a} + \lambda} \mathbf{a} \qquad (28)$$

which can be easily derived using (160) in [28]. It can be seen from (27) that the optimal $\mathbf{F}_2$ has a rank-1 structure.

By substituting (12) back into (6), the MSE of the signal waveform estimation using an MMSE receiver is given by

$$tr(\mathbf{E}(\mathbf{F}_1, \mathbf{F}_2, \mathbf{F}_3)) = tr((\mathbf{I}_{N_b} + \mathbf{G}^H \mathbf{C}_v^{-1} \mathbf{G})^{-1}). \quad (29)$$

For the single data stream case, using (27), (29) can be equivalently rewritten as

$$\begin{aligned}
\text{MSE} &= \left[1 + \mathbf{f}_1^H \mathbf{H}_{31}^H \mathbf{H}_{31} \mathbf{f}_1 + (\mathbf{H}_{32} \mathbf{f}_3 + \mathbf{H}_2 \mathbf{F}_2 \mathbf{H}_1 \mathbf{f}_1)^H \right. \\
&\quad \left. \times (\mathbf{H}_2 \mathbf{F}_2 \mathbf{F}_2^H \mathbf{H}_2^H + \mathbf{I}_{N_3})^{-1}(\mathbf{H}_{32} \mathbf{f}_3 + \mathbf{H}_2 \mathbf{F}_2 \mathbf{H}_1 \mathbf{f}_1)\right]^{-1} \\
&= \left[1 + \mathbf{f}_1^H \mathbf{H}_{31}^H \mathbf{H}_{31} \mathbf{f}_1 + (\mathbf{H}_{32} \mathbf{f}_3 + c\mathbf{H}_2 \mathbf{b})^H \right. \\
&\quad \left. \times (c\mathbf{H}_2 \mathbf{b}\mathbf{b}^H \mathbf{H}_2^H + \mathbf{I}_{N_3})^{-1}(\mathbf{H}_{32} \mathbf{f}_3 + c\mathbf{H}_2 \mathbf{b})\right]^{-1} \quad (30)
\end{aligned}$$

where $c = \mathbf{f}_1^H \mathbf{H}_1^H \mathbf{H}_1 \mathbf{f}_1$. By substituting (27) back into (11), the power constraint at the relay node can be rewritten as $(c^2 + c)\mathbf{b}^H \mathbf{b} \leq P_r$. Therefore, with fixed $\mathbf{f}_1$ and $\mathbf{f}_3$, optimizing $\mathbf{F}_2$ can be converted to the optimization of $\mathbf{b}$, which is given by the following problem

$$\max_{\mathbf{b}} \ (\mathbf{H}_{32} \mathbf{f}_3 + c\mathbf{H}_2 \mathbf{b})^H (c\mathbf{H}_2 \mathbf{b}\mathbf{b}^H \mathbf{H}_2^H + \mathbf{I}_{N_3})^{-1}$$
$$\times (\mathbf{H}_{32} \mathbf{f}_3 + c\mathbf{H}_2 \mathbf{b}) \quad (31)$$
$$\text{s.t. } \mathbf{b}^H \mathbf{b} \leq \bar{P}_r \quad (32)$$

where $\bar{P}_r = P_r/(c^2 + c)$.

Using the identity (28), it can be shown that

$$(\mathbf{H}_{32} \mathbf{f}_3 + c\mathbf{H}_2 \mathbf{b})^H (c\mathbf{H}_2 \mathbf{b}\mathbf{b}^H \mathbf{H}_2^H + \mathbf{I}_{N_3})^{-1}(\mathbf{H}_{32} \mathbf{f}_3 + c\mathbf{H}_2 \mathbf{b})$$
$$= \mathbf{f}_3^H \mathbf{H}_{32}^H \mathbf{H}_{32} \mathbf{f}_3 + c\left(1 - \frac{|\mathbf{f}_3^H \mathbf{H}_{32}^H \mathbf{H}_2 \mathbf{b} - 1|^2}{1 + c\mathbf{b}^H \mathbf{H}_2^H \mathbf{H}_2 \mathbf{b}}\right).$$

Then, the problem (31)-(32) can be equivalently rewritten as

$$\min_{\mathbf{b}} \ \frac{|\mathbf{f}_3^H \mathbf{H}_{32}^H \mathbf{H}_2 \mathbf{b} - 1|^2}{1 + c\mathbf{b}^H \mathbf{H}_2^H \mathbf{H}_2 \mathbf{b}} \quad (33)$$
$$\text{s.t. } \mathbf{b}^H \mathbf{b} \leq \bar{P}_r. \quad (34)$$

The problem (33)-(34) can be solved by the SDR technique [29] and the Charnes-Cooper transformation [30] as shown below. By introducing $\mathbf{a} = \mathbf{H}_2^H \mathbf{H}_{32} \mathbf{f}_3$, $\mathbf{A} = c\mathbf{H}_2^H \mathbf{H}_2$, $|t|^2 = 1$, and $\tilde{\mathbf{b}} = \mathbf{b}t$, (33) and (34) can be equivalently rewritten as

$$\frac{|t|^2 |\mathbf{f}_3^H \mathbf{H}_{32}^H \mathbf{H}_2 \mathbf{b} - 1|^2}{|t|^2 (1 + c\mathbf{b}^H \mathbf{H}_2^H \mathbf{H}_2 \mathbf{b})} = \frac{(\mathbf{a}^H \tilde{\mathbf{b}} - t)(\tilde{\mathbf{b}}^H \mathbf{a} - t^*)}{1 + \tilde{\mathbf{b}}^H \mathbf{A} \tilde{\mathbf{b}}} \quad (35)$$
$$|t|^2 \mathbf{b}^H \mathbf{b} = \tilde{\mathbf{b}}^H \tilde{\mathbf{b}}. \quad (36)$$

Using (35) and (36), the problem (33)-(34) can be equivalently rewritten as

$$\min_{\tilde{\mathbf{b}}, t} \ \frac{(\mathbf{a}^H \tilde{\mathbf{b}} - t)(\tilde{\mathbf{b}}^H \mathbf{a} - t^*)}{1 + \tilde{\mathbf{b}}^H \mathbf{A} \tilde{\mathbf{b}}} \quad (37)$$
$$\text{s.t. } \tilde{\mathbf{b}}^H \tilde{\mathbf{b}} \leq \bar{P}_r, \qquad |t|^2 = 1. \quad (38)$$

By introducing $\bar{\mathbf{a}} = [\mathbf{a}^T, -1]^T$ and $\bar{\mathbf{b}} = [\tilde{\mathbf{b}}^T, t]^T$, we have

$$(\mathbf{a}^H \tilde{\mathbf{b}} - t)(\tilde{\mathbf{b}}^H \mathbf{a} - t^*) = \bar{\mathbf{a}}^H \bar{\mathbf{b}}\bar{\mathbf{b}}^H \bar{\mathbf{a}} = tr(\mathbf{A}_2 \mathbf{X}) \quad (39)$$
$$\tilde{\mathbf{b}}^H \mathbf{A} \tilde{\mathbf{b}} = \bar{\mathbf{b}}^H \mathbf{A}_1 \bar{\mathbf{b}} = tr(\mathbf{A}_1 \mathbf{X}) \quad (40)$$
$$\tilde{\mathbf{b}}^H \tilde{\mathbf{b}} = \bar{\mathbf{b}}^H \mathbf{B}_1 \bar{\mathbf{b}} = tr(\mathbf{B}_1 \mathbf{X}) \quad (41)$$
$$|t|^2 = \bar{\mathbf{b}}^H \mathbf{B}_2 \bar{\mathbf{b}} = tr(\mathbf{B}_2 \mathbf{X}) \quad (42)$$

where $\mathbf{X} = \bar{\mathbf{b}}\bar{\mathbf{b}}^H$, $\mathbf{A}_2 = \bar{\mathbf{a}}\bar{\mathbf{a}}^H$, $\mathbf{A}_1 = \begin{bmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{0} & 0 \end{bmatrix}$, $\mathbf{B}_1 = \begin{bmatrix} \mathbf{I}_{N_2} & \mathbf{0} \\ \mathbf{0} & 0 \end{bmatrix}$, $\mathbf{B}_2 = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & 1 \end{bmatrix}$. Using (39)-(42), the problem (37)-(38) can be equivalently rewritten as

$$\max_{\mathbf{X}} \ \frac{1 + tr(\mathbf{A}_1 \mathbf{X})}{tr(\mathbf{A}_2 \mathbf{X})} \quad (43)$$
$$\text{s.t. } tr(\mathbf{B}_1 \mathbf{X}) \leq \bar{P}_r \quad (44)$$
$$tr(\mathbf{B}_2 \mathbf{X}) = 1 \quad (45)$$
$$rank(\mathbf{X}) = 1, \quad \mathbf{X} \succeq \mathbf{0} \quad (46)$$

where $rank(\cdot)$ denotes the matrix rank and $\mathbf{A} \succeq \mathbf{0}$ means that matrix $\mathbf{A}$ is positive semidefinite (PSD).

By applying the SDR technique, we drop the rank-1 constraint in (46), then we have the following relaxed problem

$$\max_{\mathbf{X}} \ \frac{1 + tr(\mathbf{A}_1 \mathbf{X})}{tr(\mathbf{A}_2 \mathbf{X})} \quad (47)$$
$$\text{s.t. } tr(\mathbf{B}_1 \mathbf{X}) \leq \bar{P}_r \quad (48)$$
$$tr(\mathbf{B}_2 \mathbf{X}) = 1 \quad (49)$$
$$\mathbf{X} \succeq \mathbf{0}. \quad (50)$$

The problem (47)-(50) is a linear fraction programming problem, which can be solved through the Charnes-Cooper transformation [30]. Let $\mathbf{X} = \mathbf{Z}/\delta$ with $\delta > 0$. The problem (47)-(50) can be equivalently rewritten as

$$\max_{\mathbf{Z}, \delta} \ tr(\mathbf{A}_1 \mathbf{Z}) + \delta \quad (51)$$
$$\text{s.t. } tr(\mathbf{A}_2 \mathbf{Z}) = 1 \quad (52)$$
$$tr(\mathbf{B}_1 \mathbf{Z}) \leq \delta \bar{P}_r \quad (53)$$
$$tr(\mathbf{B}_2 \mathbf{Z}) = \delta \quad (54)$$
$$\mathbf{Z} \succeq \mathbf{0}, \qquad \delta > 0. \quad (55)$$

The problem (51)-(55) is a semidefinite programming (SDP) problem and can be efficiently solved by the disciplined convex programming toolbox CVX [27]. After getting $\mathbf{Z}$ and δ, we obtain $\mathbf{X} = \mathbf{Z}/\delta$. Then we can apply Theorem 2.3 in [31] to find the optimal $\bar{\mathbf{b}}$ such that

$$tr(\mathbf{A}_i \mathbf{X}) = \bar{\mathbf{b}}^H \mathbf{A}_i \bar{\mathbf{b}}, \qquad tr(\mathbf{B}_i \mathbf{X}) = \bar{\mathbf{b}}^H \mathbf{B}_i \bar{\mathbf{b}}, \quad i = 1, 2. \quad (56)$$

We would like to note that the construction of $\bar{\mathbf{b}}$ from $\mathbf{X}$ using (56) does not lose any optimality [31], i.e., the SDR from (43)-(46) to (47)-(50) is in fact tight. In the next step, we extract $\mathbf{b}$ from $\bar{\mathbf{b}}$ as $\mathbf{b} = \bar{\mathbf{b}}_{1:N_2}/\bar{b}_{N_2+1}$, where $\bar{\mathbf{b}}_{1:N_2}$ stands for the first N_2 elements of $\bar{\mathbf{b}}$. Finally, we obtain $\mathbf{F}_2$ from (27).

With fixed $\mathbf{F}_2$, the problem of optimizing $\mathbf{f}_1$ and $\mathbf{f}_3$ is given by

$$\max_{\mathbf{f}_1, \mathbf{f}_3} \ \mathbf{f}_1^H \mathbf{H}_{31}^H \mathbf{H}_{31} \mathbf{f}_1 + (\mathbf{H}_{32} \mathbf{f}_3 + \mathbf{H}_2 \mathbf{F}_2 \mathbf{H}_1 \mathbf{f}_1)^H$$
$$\times (\mathbf{H}_2 \mathbf{F}_2 \mathbf{F}_2^H \mathbf{H}_2^H + \mathbf{I}_{N_3})^{-1}(\mathbf{H}_{32} \mathbf{f}_3 + \mathbf{H}_2 \mathbf{F}_2 \mathbf{H}_1 \mathbf{f}_1) \quad (57)$$
$$\text{s.t. } \mathbf{f}_1^H \mathbf{H}_1^H \mathbf{F}_2^H \mathbf{F}_2 \mathbf{H}_1 \mathbf{f}_1 \leq P_r - tr(\mathbf{F}_2 \mathbf{F}_2^H) \quad (58)$$
$$\mathbf{f}_1^H \mathbf{f}_1 + \mathbf{f}_3^H \mathbf{f}_3 \leq P_s. \quad (59)$$

Let us introduce $\tilde{P}_r = P_r - tr(\mathbf{F}_2\mathbf{F}_2^H)$, $\mathbf{f} = [\mathbf{f}_1^T, \mathbf{f}_3^T]^T$, and

$$\mathbf{D}_1 = \begin{bmatrix} (\mathbf{H}_2\mathbf{F}_2\mathbf{H}_1)^H \\ \mathbf{H}_{32}^H \end{bmatrix} (\mathbf{H}_2\mathbf{F}_2\mathbf{F}_2^H\mathbf{H}_2^H + \mathbf{I}_{N_3})^{-1}$$

$$\times [\mathbf{H}_2\mathbf{F}_2\mathbf{H}_1 \quad \mathbf{H}_{32}] + \begin{bmatrix} \mathbf{H}_{31}^H\mathbf{H}_{31} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$$

$$\mathbf{D}_2 = \begin{bmatrix} \mathbf{H}_1^H\mathbf{F}_2^H\mathbf{F}_2\mathbf{H}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}.$$

The problem (57)-(59) can be rewritten as

$$\min_{\mathbf{f}} \; -\mathbf{f}^H\mathbf{D}_1\mathbf{f} \tag{60}$$

$$\text{s.t. } \mathbf{f}^H\mathbf{D}_2\mathbf{f} \le \tilde{P}_r \tag{61}$$

$$\mathbf{f}^H\mathbf{f} \le P_s. \tag{62}$$

By introducing $\mathbf{Q} = \mathbf{f}\mathbf{f}^H$, the problem (60)-(62) can rewritten as

$$\min_{\mathbf{Q}} \; -tr(\mathbf{D}_1\mathbf{Q}) \tag{63}$$

$$\text{s.t. } tr(\mathbf{D}_2\mathbf{Q}) \le \tilde{P}_r \tag{64}$$

$$tr(\mathbf{Q}) \le P_s, \quad rank(\mathbf{Q}) = 1, \quad \mathbf{Q} \succeq \mathbf{0}. \tag{65}$$

Using the SDR technique to drop the rank-1 constraint in (65), the relaxed problem can be written as

$$\min_{\mathbf{Q}} \; -tr(\mathbf{D}_1\mathbf{Q}) \tag{66}$$

$$\text{s.t. } tr(\mathbf{D}_2\mathbf{Q}) \le \tilde{P}_r \tag{67}$$

$$tr(\mathbf{Q}) \le P_s, \quad \mathbf{Q} \succeq \mathbf{0}. \tag{68}$$

The problem (66)-(68) is an SDP problem which can be solved by CVX. After obtaining $\mathbf{Q}$, we can get the optimal $\mathbf{f}$ from $\mathbf{Q}$ by using the randomization procedure in [29] such that the SDR from (63)-(65) to (66)-(68) is tight.

We are ready to develop an iterative transceiver optimization algorithm for AF MIMO relay systems with a single data stream. In each iteration, with fixed $\mathbf{f}_1$ and $\mathbf{f}_3$, we update $\mathbf{F}_2$ following the steps in (51)-(56). Then with given $\mathbf{F}_2$, we optimize $\mathbf{f}_1$ and $\mathbf{f}_3$ by solving the problem (66)-(68). Note that the conditional updates of $\mathbf{F}_2$, $\mathbf{f}_1$, and $\mathbf{f}_3$ may either decrease or maintain but cannot increase the objective function (30). Monotonic convergence of $\mathbf{F}_2$, $\mathbf{f}_1$, and $\mathbf{f}_3$ towards (at least) a stationary point follows directly from this observation. After the convergence of the iterations, $\mathbf{W}$ is obtained as an MMSE receiver in (12). Table II illustrates the procedure of the proposed iterative algorithm, where mse_2 is the MSE calculated by (30).

V. NUMERICAL EXAMPLES

In this section, we study the performance of two proposed source, relay, and receiver matrices optimization algorithms through numerical simulations. In the simulations, a two-hop AF MIMO relay system with 4 antennas at each node (i.e., $N_1 = N_2 = N_3 = 4$) is considered. The channel matrices $\mathbf{H}_1$, $\mathbf{H}_2$, $\mathbf{H}_{31}$, and $\mathbf{H}_{32}$ have i.i.d. complex Gaussian entries with zero mean and variances of σ_1^2, σ_2^2, σ_3^2, and σ_3^2, respectively. Note that we assume the same statistics for $\mathbf{H}_{31}$ and $\mathbf{H}_{32}$, as usually it remains unchanged over two consecutive time slots.

TABLE II
PROCEDURE OF THE TRANSCEIVER OPTIMIZATION ALGORITHM FOR AF MIMO RELAY SYSTEMS WITH A SINGLE DATA STREAM

1) Initialize the algorithm with $\mathbf{f}_1^{(0)} = \mathbf{f}_3^{(0)} = \sqrt{P_s/2N_b}[1,\mathbf{0}]^T$ and $\mathbf{F}_2^{(0)} = \sqrt{P_r/tr(\mathbf{H}_1\mathbf{f}_1^{(0)}(\mathbf{f}_1^{(0)})^H\mathbf{H}_1^H + \mathbf{I}_{N_2})}\mathbf{I}_{N_2}$, and set $n = 0$.
2) Update $\mathbf{F}_2^{(n+1)}$ using $\mathbf{f}_1^{(n)}$ and $\mathbf{f}_3^{(n)}$ as (51)-(56).
3) Update $\mathbf{f}_1^{(n+1)}$ and $\mathbf{f}_3^{(n+1)}$ using $\mathbf{F}_2^{(n+1)}$ by solving the problem (66)-(68).
4) If $(\text{mse}_2^{(n)} - \text{mse}_2^{(n+1)})/\text{mse}_2^{(n)} \le \varepsilon$, iteration ends; otherwise go to step (2).
5) Obtain $\mathbf{W}$ using (12).

Based on the assumption that the noise has unit power, we define $\text{SNR}_1 = \sigma_1^2 P_s/N_2$, $\text{SNR}_2 = \sigma_2^2 P_r/N_3$, and $\text{SNR}_3 = \sigma_3^2 P_s/N_3$ as the signal-to-noise ratio (SNR) for the source-relay, relay-destination, and source-destination links respectively. Following [23], we choose $\text{SNR}_3 = \text{SNR}_1 - \Delta_{\text{SNR}}$(dB), where Δ_{SNR} stands for the attenuation of the direct link relative to the first-hop channel. In the first two numerical examples, we set $\text{SNR}_1 = \text{SNR}_2 = \text{SNR}$ and in the last four simulation examples, the SNRs are set as $\text{SNR}_1 = 0$dB and $\text{SNR}_2 = \text{SNR}$. Quadrature phase-shift keying (QPSK) constellations are used to modulate the source symbols. All simulation results are averaged over 1000 independent channel realizations.

We compare the proposed algorithms with the algorithm in [16] and the Tri-step and Bi-step algorithms developed in [23] using the exact CSI. We set $\varepsilon = 0.001$ for the proposed schemes and both algorithms in [23]. For the proposed AF MIMO relay system, we consider two cases. In Case 1, $\mathbf{H}_{32} = \mathbf{H}_{31}$, i.e., the exact CSI of the direct link remains unchanged during two transmission slots, corresponding to a slow-fading environment. While in Case 2, we set $\mathbf{H}_{32} \ne \mathbf{H}_{31}$, which simulates a fast-fading environment.

In the first numerical example, we consider the case where $N_b = 4$ independent data streams are simultaneously transmitted and set $\Delta_{\text{SNR}} = 20$dB. The procedure in Table I is carried out for the proposed algorithm. Fig. 2 shows the system BER versus SNR of the four algorithms tested. It can be seen from Fig. 2 that the proposed algorithm has a better BER performance than those in [16] and [23], which confirms that additional gain can be achieved by making the source node transmit signals at the second time slot. The poor performance of the algorithm in [16] is caused by the channel diagonalization constraint, which is strictly suboptimal when $N_b > 1$. We would like to note that for the algorithms in [16] and [23], as the source node is silent at the second time slot, the BER of these three algorithms does not depend on $\mathbf{H}_{32}$.

In the second example, we simulate a MIMO relay system with $N_b = 4$ and $\Delta_{\text{SNR}} = 10$dB. The system BER and MSE yielded by four algorithms tested are shown in Figs. 3 and 4, respectively. Similar to Fig. 2, it can be observed from Figs. 3 and 4 that the proposed algorithm yields a lower BER and MSE than the algorithms in [16] and [23]. Moreover, it can be seen from Figs. 2 and 3 that in a fast-fading channel environment, the BER performance of the proposed AF protocol is further improved due to the valuable time
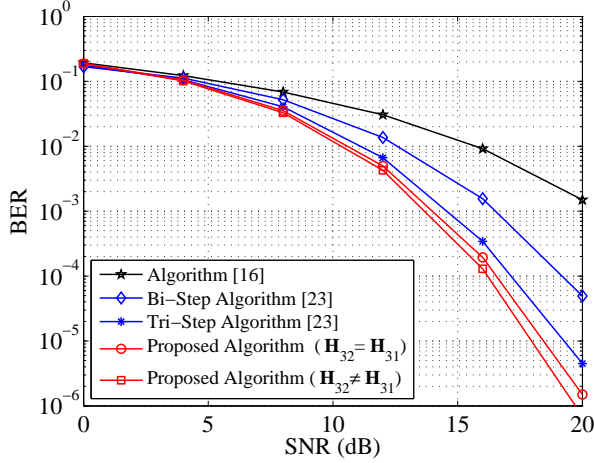
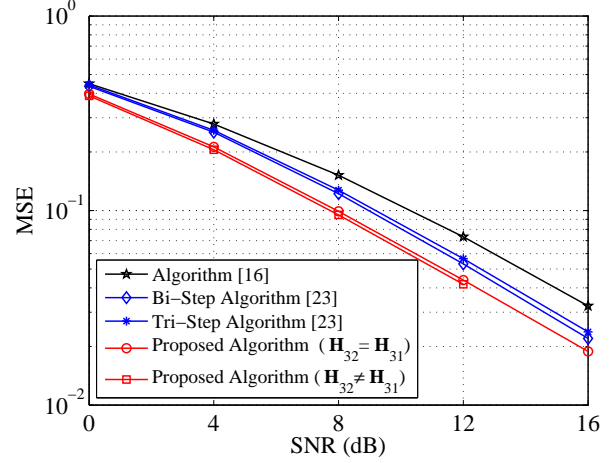Fig. 2. Example 1. BER versus SNR, $N_b = 4$, $\Delta_{\text{SNR}} = 20$dB.

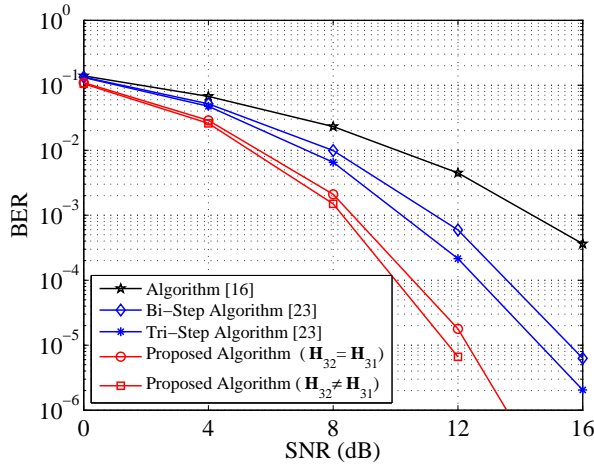Fig. 4. Example 2. MSE versus SNR, $N_b = 4$, $\Delta_{\text{SNR}} = 10$dB.

Fig. 3. Example 2. BER versus SNR, $N_b = 4$, $\Delta_{\text{SNR}} = 10$dB.
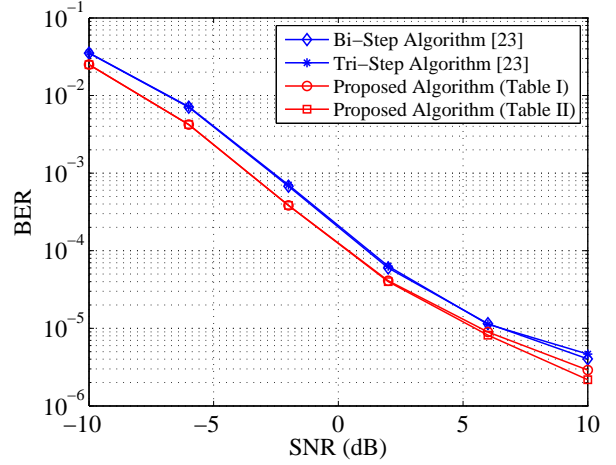

Fig. 5. Example 3. BER versus SNR, $N_b = 1$, $\Delta_{\text{SNR}} = 20$dB, $\mathbf{H}_{32} = \mathbf{H}_{31}$.

diversity as the proposed algorithm jointly optimizes $\mathbf{F}_1$ and $\mathbf{F}_3$ considering both $\mathbf{H}_{31}$ and $\mathbf{H}_{32}$. It is worth noting that the gap between the BER of the proposed algorithm and that of the algorithms in [16] and [23] increases with the SNR, and such performance gap increases as the gain of the direct link is increased from $\Delta_{\text{SNR}} = 20$dB to $\Delta_{\text{SNR}} = 10$dB. This is expected as explained below. When $\text{SNR}_1 = \text{SNR}_2 = \text{SNR}$ is low (and/or the direct link is weak), the benefit of distributing the transmission power at the source node over two time slots is less noticeable as the effective power for the direct link is limited. As the SNR increases (and/or the direct link becomes stronger), more effective power is available at the source node so that it can allocate the power more flexibly over two time slots to reduce the system BER.

In the third simulation example, an AF MIMO relay system with $N_b = 1$, $\Delta_{\text{SNR}} = 20$dB, and $\mathbf{H}_{32} = \mathbf{H}_{31}$ is simulated. We applied both the algorithms in Table I and Table II for the proposed AF protocol. The BER performance of four algorithms is shown in Fig. 5. It can be seen that similar to the general case of multiple data streams, for the case of a

single data stream, the system BER is reduced by letting the source node transmit signals at both time slots. Moreover, we can also observe from Fig. 5 that at high SNR, the algorithm in Table II has a better BER performance than the algorithm in Table I.

We simulate in the next example a MIMO relay system with $N_b = 1$, $\Delta_{\text{SNR}} = 20$dB, and $\mathbf{H}_{32} \neq \mathbf{H}_{31}$. It can be seen from Fig. 6 that the proposed algorithms yield smaller system BER than the algorithms in [23]. Comparing Fig. 6 with Fig. 5, we observe that the time diversity in a fast-fading channel environment helps to reduce the system BER.

In the last two examples, we simulate an AF MIMO relay system with $N_b = 1$ and $\Delta_{\text{SNR}} = 10$dB. Fig. 7 shows the system BER versus SNR of all five algorithms tested when $\mathbf{H}_{32} = \mathbf{H}_{31}$, while Figs. 8 and 9 demonstrate the system BER and MSE performance, respectively, when $\mathbf{H}_{32} \neq \mathbf{H}_{31}$. Compared with Figs. 5 and 6, it can be seen from Figs. 7 and 8 that as the gain of the direct link increases (Δ_{SNR} from 20 dB to 10dB), the BER performance of all algorithms improves. Interestingly, it can be seen from Figs. 7-9 that the

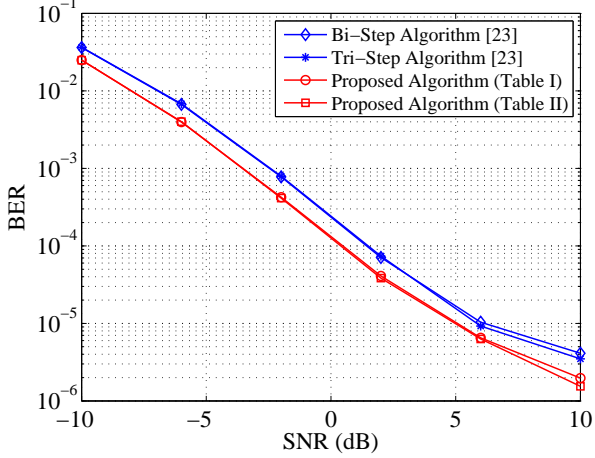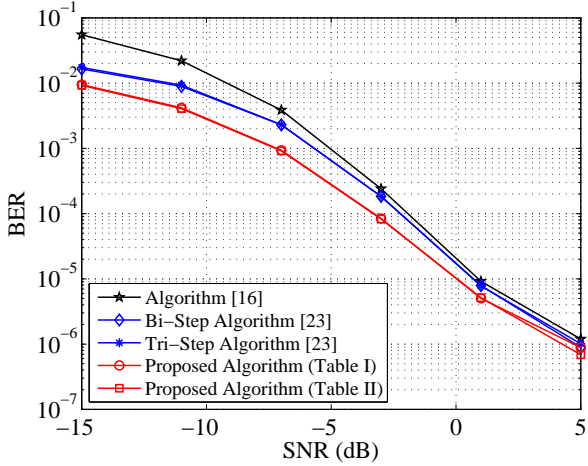Fig. 6. Example 4. BER versus SNR, $N_b = 1$, $\Delta_{\text{SNR}} = 20$dB, $\mathbf{H}_{32} \neq \mathbf{H}_{31}$.

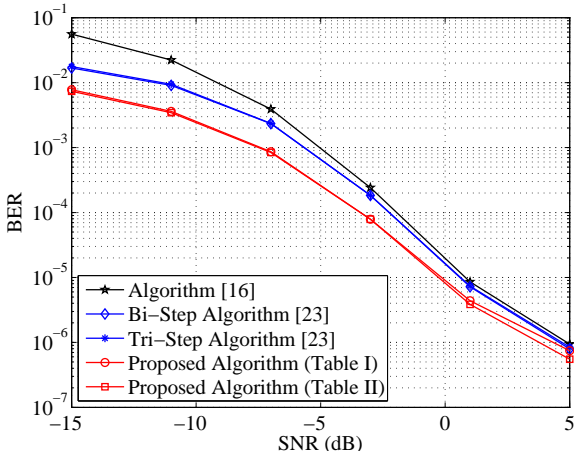Fig. 7. Example 5. BER versus SNR, $N_b = 1$, $\Delta_{\text{SNR}} = 10$dB, $\mathbf{H}_{32} = \mathbf{H}_{31}$.

Fig. 8. Example 6. BER versus SNR, $N_b = 1$, $\Delta_{\text{SNR}} = 10$dB, $\mathbf{H}_{32} \neq \mathbf{H}_{31}$.
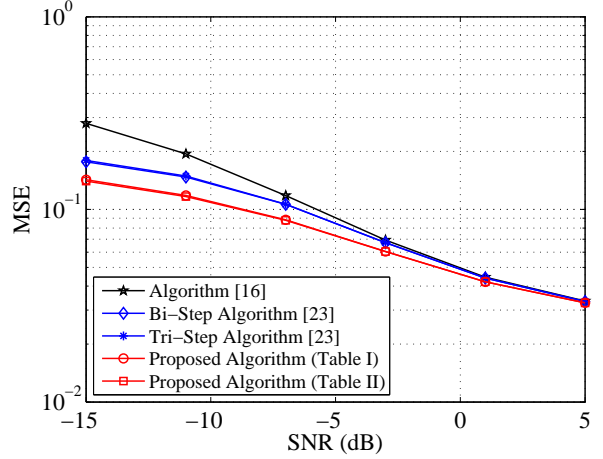

Fig. 9. Example 6. MSE versus SNR, $N_b = 1$, $\Delta_{\text{SNR}} = 10$dB, $\mathbf{H}_{32} \neq \mathbf{H}_{31}$.

gap between the BERs yielded by the proposed algorithms and the approaches in [16] and [23] decreases with the increase of SNR, which is expected as explained below. Remember that in Figs. 7-9, we set $\text{SNR}_1 = 0$dB, $\text{SNR}_3 = -10$dB, and $\text{SNR}_2 = \text{SNR}$. At low SNRs, as the gain of the direct link is larger than that of the relay-destination link, the source node allocates more power to the second time slot to exploit $\mathbf{H}_{32}$. Thus, the proposed algorithms have a better performance than the other three approaches at low SNRs. As SNR increases, the relay-destination link becomes stronger. The source node tends to allocate more power to the first time slot. When the SNR is sufficiently large, almost all power of the source node is allocated at the first time slot. As a result, the proposed algorithms have a similar performance to the other three algorithms at high SNRs.

Finally, we compare the computational complexity of the five algorithms tested. For the sake of notational simplicity, we assume $N_i = N, i = 1, 2, 3$ and $N_b = N$ for the general case. As the algorithm in [16] involves matrix inversion and matrix SVD, it has a complexity order of $O(N^3)$ [16]. It is shown in [23] that the per iteration complexity of the Bi-step and Tri-step algorithms is $O(N^7)$ and $O(N^6)$, respectively. For the proposed algorithm in the general case, it can be seen from Table I that most of computations are spent on solving the QCQP problem (22)-(24). Thus, the per iteration complexity of the proposed algorithm is in the order of $O(N^6)$ [32], which is comparable to that of the Tri-step algorithm.

For the special case of a single data stream, we assume $N_i = N, i = 1, 2, 3$ and $N_b = 1$. The complexity order of the algorithm in [16] and the Bi-step algorithm remains the same as in the general case, while the per iteration complexity of the Tri-step algorithm and the proposed algorithm in Table I is reduced to $O(N^3)$ [32]. For the proposed algorithm in Table II, we need to solve two SDP problems in each iteration, which has a complexity order of $O(N^7)$. From the above analysis, the algorithm in Table II has a higher per iteration complexity than that in Table I.

The overall computational complexity of the four iterative algorithms tested (except for the approach in [16]) depends

TABLE III
AVERAGE NUMBER OF ITERATIONS REQUIRED BY FOUR ITERATIVE
ALGORITHMS TILL CONVERGENCE FOR GENERAL SYSTEMS

SNR (dB)	0	4	8	12	16
Tri-step Algorithm	8	10	15	27	49
Bi-step Algorithm	4	4	5	5	5
Algorithm in Table I ($\mathbf{H}_{32} = \mathbf{H}_{31}$)	14	21	31	42	59
Algorithm in Table I ($\mathbf{H}_{32} \neq \mathbf{H}_{31}$)	13	19	27	37	-

TABLE IV
AVERAGE NUMBER OF ITERATIONS REQUIRED BY FOUR ITERATIVE
ALGORITHMS TILL CONVERGENCE FOR SINGLE DATA STREAM SYSTEMS

SNR (dB)	-15	-11	-7	-3	1	5
Tri-step Algorithm	8	8	7	5	4	4
Bi-step Algorithm	4	4	6	5	3	3
Algorithm in Table I	9	7	6	5	4	4
Algorithm in Table II	3	3	3	3	3	3

also on the number of iterations. Tables III and IV show the average number of iterations required by the four iterative algorithms till convergence in the examples corresponding to Figs. 3 and 8, respectively. It can be seen that for general systems, the number of iterations required by the proposed algorithm in Table I increases with the SNR, and is larger than the other two algorithms. Interestingly, for single data stream systems, it can be seen from Table IV that the number of iterations required by the proposed algorithm in Table II is smaller than the other three algorithms and does not change with the SNR.

VI. CONCLUSIONS

We have studied a new AF relay protocol where the source node transmits signals at both time slots in half-duplex MIMO relay systems with the direct link. Two novel iterative algorithms have been developed to optimize the source, relay, and receiver matrices in this new AF MIMO relay system. Simulation results show that the proposed algorithms have better BER performance compared with conventional AF MIMO relay systems, and more gains can be achieved with the improvement of the direct-link channel quality. When the optimal receiver $\mathbf{W}$ in (12) is substituted into (9), the objective function becomes a very complicated function of three precoding matrices $\mathbf{F}_1$, $\mathbf{F}_2$, and $\mathbf{F}_3$. How to solve this optimization problem efficiently for general AF MIMO relay systems is an interesting future topic.

APPENDIX
PROOF OF (16)

The Lagrangian function associated with the problem (14)-(15) is given by

$$\mathcal{L} = f(\mathbf{F}_2) + \lambda\, g(\mathbf{F}_2)$$

where $f(\mathbf{F}_2) = tr((\breve{\mathbf{H}}_2\mathbf{F}_2\breve{\mathbf{H}}_1 - \mathbf{\Pi})(\breve{\mathbf{H}}_2\mathbf{F}_2\breve{\mathbf{H}}_1 - \mathbf{\Pi})^H + \breve{\mathbf{H}}_2\mathbf{F}_2\mathbf{F}_2^H\breve{\mathbf{H}}_2^H)$, and $g(\mathbf{F}_2) = tr(\mathbf{F}_2(\breve{\mathbf{H}}_1\breve{\mathbf{H}}_1^H + \mathbf{I}_{N_2})\mathbf{F}_2^H) - P_r$. At the optimal $\mathbf{F}_2$, there is [26]

$$\frac{\partial \mathcal{L}}{\partial \mathbf{F}_2} = \frac{\partial f(\mathbf{F}_2)}{\partial \mathbf{F}_2} + \lambda\frac{\partial g(\mathbf{F}_2)}{\partial \mathbf{F}_2} = 0. \tag{69}$$

Using the derivatives of matrices in [28], we have

$$\frac{\partial f(\mathbf{F}_2)}{\partial \mathbf{F}_2} = \breve{\mathbf{H}}_2^T[\breve{\mathbf{H}}_1(\breve{\mathbf{H}}_2\mathbf{F}_2\breve{\mathbf{H}}_1 - \mathbf{\Pi})^H]^T + \breve{\mathbf{H}}_2^T(\mathbf{F}_2^H\breve{\mathbf{H}}_2^H)^T \tag{70}$$

$$\frac{\partial g(\mathbf{F}_2)}{\partial \mathbf{F}_2} = [(\breve{\mathbf{H}}_1\breve{\mathbf{H}}_1^H + \mathbf{I}_{N_2})\mathbf{F}_2^H]^T. \tag{71}$$

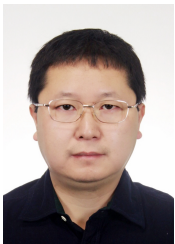By substituting (70) and (71) back into (69) and solving for $\mathbf{F}_2$, we obtain (16).

REFERENCES

[1] J. Boyer, D. D. Falconer, and H. Yanikomeroglu, "Multihop diversity in wireless relaying channels," *IEEE Trans. Commun.*, vol. 52, pp. 1820-1830, Oct. 2004.

[2] J. N. Laneman, D. N. C. Tse, and G. W. Wornell, "Cooperative diversity in wireless networks: Efficient protocols and outage behavior," *IEEE Trans. Inf. Theory*, vol. 50, pp. 3062-3080, Dec. 2004.

[3] 3GPP, "Relay architectures for E-UTRA (LTE-Advanced)," TR 36.806, Mar. 2010.

[4] S. W. Peters and R. W. Heath, Jr., "The future of WiMAX: Multihop relaying with IEEE 802.16j," *IEEE Commun. Magazine*, vol. 47, pp. 104-111, Jan. 2009.

[5] B. Wang, J. Zhang, and A. Høst-Madsen, "On the capacity of MIMO relay channels," *IEEE Trans. Inf. Theory*, vol. 51, pp. 29-43, Jan. 2005.

[6] S. Jin, M. R. McKay, C. Zhong, and K. K. Wong, "Ergodic capacity analysis of amplify-and-forward MIMO dual-hop systems," *IEEE Trans. Inf. Theory*, vol. 56, pp. 2204-2224, May. 2010.

[7] X. Tang and Y. Hua, "Optimal design of non-regenerative MIMO wireless relays," *IEEE Trans. Wireless Commun.*, vol. 6, pp. 1398-1407, Apr. 2007.

[8] O. Muñoz-Medina, J. Vidal, and A. Agustín, "Linear transceiver design in nonregenerative relays with channel state information," *IEEE Trans. Signal Process.*, vol. 55, pp. 2593-2604, June 2007.

[9] W. Guan and H. Luo, "Joint MMSE transceiver design in non-regenerative MIMO relay systems," *IEEE Commun. Lett.*, vol. 12, pp. 517-519, July 2008.

[10] D. Jiang, H. Zhang, and D. Yuan, "Linear precoding with limited feedback in multiuser MIMO relay networks," in *Proc. Int. Symposium Commun. and Inf. Technol.*, Icheon, South Korea, Sep. 2009, pp. 337-342.

[11] B. Zhang, X. Wang, K. Niu, and Z. He, "Joint linear transceiver design for non-regenerative MIMO relay systems," *Electronics Lett.*, vol. 45, pp. 1254-1256, Nov. 2009.

[12] Y. Rong, X. Tang, and Y. Hua, "A unified framework for optimizing linear non-regenerative multicarrier MIMO relay communication systems," *IEEE Trans. Signal Process.*, vol. 57, pp. 4837-4851, Dec. 2009.

[13] L. Sanguinetti, A. A. D'Amico, and Y. Rong, "A tutorial on the optimization of amplify-and-forward MIMO relay systems," *IEEE J. Selet. Areas Commun.*, vol. 30, pp. 1331-1346, Sep. 2012.

[14] Y. Rong, "Optimal joint source and relay beamforming for MIMO relays with direct link," *IEEE Commun. Lett.*, vol.14, pp. 390-392, May 2010.

[15] F.-S. Tseng and W.-R. Wu, "Linear MMSE transceiver design in amplify-and-forward MIMO relay systems," *IEEE Trans. Veh. Technol.*, vol. 59, pp. 754-765, Feb. 2010.

[16] F.-S. Tseng, W.-R. Wu, and J.-Y. Wu, "Joint source/relay precoder design in nonregenerative cooperative systems using an MMSE criterion," *IEEE Trans. Wireless Commun.*, vol. 8, pp. 4928-4933, Oct. 2009.

[17] F.-S. Tseng, M.-Y. Chang, and W.-R. Wu, "Joint Tomlinson-Harashima source and linear relay precoder design in amplify-and-forward MIMO relay systems via MMSE criterion," *IEEE Trans. Veh. Technol.*, vol. 60, pp. 1687-1698, May 2011.

[18] C. Song, K.-J. Lee, and I. Lee, "MMSE-based MIMO cooperative relaying systems: Closed-form designs and outage behavior," *IEEE J. Select. Areas Commun.*, vol. 30, pp. 1390-1401, Sep. 2012.

[19] 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.

[20] H. Wan and W. Chen, "Joint source and relay design for multiuser MIMO nonregenerative relay networks with direct links," *IEEE Trans. Veh. Technol.*, vol. 61, pp. 2871-2876, July 2012.

[21] F.-S. Tseng, M.-Y. Chang, and W.-R. Wu, "Robust Tomlinson-Harashima source and linear relay precoders design in amplify-and-forward MIMO relay systems," *IEEE Trans. Commun.*, vol. 60, pp. 1124-1137, Apr. 2012.

[22] H. Shen, W. Xu, and C. Zhao, "Robust transceiver for AF MIMO relaying with direct link: A globally optimal solution," *IEEE Signal Process. Lett.*, vol. 21, pp. 947-951, Aug. 2014.

[23] Z. He, W. Jiang, and Y. Rong, "Robust design for amplify-and-forward MIMO relay systems with direct link and imperfect channel information," *IEEE Trans. Wireless Commun.*, vol. 14, pp. 353-363, Jan. 2015.

[24] Y. Ma, A. Liu, and Y. Hua, "A dual-phase power allocation scheme for multicarrier relay system with direct link," *IEEE Trans. Signal Process.*, vol. 62, pp. 5-16, Jan. 2014.

[25] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Prentice Hall, 1993.

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

[27] M. Grant and S. Boyd, "The CVX Users Guide," Release 2.1, Oct. 2014. Available: http://web.cvxr.com/cvx/doc/CVX.pdf.

[28] K. B. Petersen and M. S. Petersen, The Matrix Cookbook [Online]. Available: https://www.math.uwaterloo.ca/~hwolkowi/matrixcookbook.pdf, ver. Nov. 15, 2012.

[29] Y. Huang and D. P. Palomar, "Randomized algorithms for optimal solutions of double-sided QCQP with applications in signal processing," *IEEE Trans. Signal Process.*, vol. 62, pp. 1093-1108, Mar. 2014.

[30] A. Charnes and W. W. Cooper, "Programming with linear fractional functionals," *Naval Research Logistics Quarterly*, vol. 9, pp. 181-186, 1962.

[31] W. Ai, Y. Huang, and S. Zhang, "New results on Hermitian matrix rank-one decomposition," *Mathematical programming*, vol. 128, pp. 253-283, Aug. 2009.

[32] Y. Nesterov and A. Nemirovsky, "Interior-point polynomial methods in convex programming," in *Volume 13 of Studies in Applied Mathematics*. Philadelphia, PA, USA: SIAM, 1994.

Wanning Liu received the B.E. degree in communication engineering from North China Electric Power University, Beijing, China. She is currently pursuing the M.Sc. degree in information and telecommunication engineering at Beijing University of Posts and Telecommunications, Beijing, China. Her research interests include data mining, machine learning, and wireless communications.
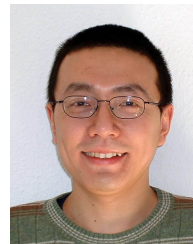
Yue Rong (S'03-M'06-SM'11) received the Ph.D. degree (summa cum laude) in electrical engineering from the Darmstadt University of Technology, Darmstadt, Germany, in 2005.

He was a Post-Doctoral Researcher with the Department of Electrical Engineering, University of California, Riverside, from February 2006 to November 2007. Since December 2007, he has been with the Department of Electrical and Computer Engineering, Curtin University, Bentley, Australia, where he is currently a Full Professor. His research interests include signal processing for communications, wireless communications, underwater acoustic communications, applications of linear algebra and optimization methods, and statistical and array signal processing. He has published over 130 journal and conference paper in these areas.

Dr. Rong 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 at Curtin University in 2010. He is an Associate Editor of the IEEE TRANSACTIONS ON SIGNAL PROCESSING. He was an Editor of the IEEE WIRELESS COMMUNICATIONS LETTERS from 2012 to 2014, a Guest Editor of the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS special issue on theories and methods for advanced wireless relays, and was a TPC Member for the IEEE ICC, WCSP, IWCMC, and ChinaCom.

Zhiqiang He (S'01-M'04) received the B.E. degree and Ph.D. degree (with distinction) from Beijing University of Posts and Telecommunications, China, all in signal and information processing, in 1999 and 2004, respectively. Since July 2004, He has been with the School of Information and Communication Engineering, Beijing University of Posts and Telecommunications, where he is currently an Associate Professor and the director of the Center of Information Theory and Technology.

His research interests include signal and information processing in wireless communications, networking architecture and protocol design, and underwater acoustic communications.

Jinnian Zhang received the B.E. degree in information engineering from Beijing University of Posts and Telecommunications, Beijing, China, in 2014. He is currently pursuing the M.Sc. degree in information and telecommunication engineering at Beijing University of Posts and Telecommunications, Beijing, China. His research interests include wireless communications, synchronization, interference cancellation, and channel estimation.