IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received September 26, 2017, accepted November 1, 2017, date of publication November 8, 2017,
date of current version November 28, 2017.

Digital Object Identifier 10.1109/ACCESS.2017.2771305

Transceiver Optimization for Two-Hop MIMO Relay Systems With Direct Link and MSE Constraints

JIAOLONG YANG[1], ZHIQIANG HE[1], (Member, IEEE), AND YUE RONG[2], (Senior Member, IEEE)
[1]Key Laboratory of Universal Wireless Communication, Ministry of Education, Beijing University of Posts and Telecommunications, Beijing 100876, China
[2]Department of Electrical and Computer Engineering, Curtin University, Bentley, WA 6102, Australia

Corresponding author: Yue Rong (y.rong@curtin.edu.au)

This work was supported in part by the National Natural Science Foundation of China under Grant 61671080 and in part by the Australian Research Council's Discovery Projects funding scheme under Grant DP140102131.

ABSTRACT In this paper, we investigate the transceiver design in a two-hop amplify-and-forward (AF) multiple-input multiple-output (MIMO) relay system with the direct link and quality-of-service (QoS) constraint. The QoS criterion is specified by the upper bound of the mean-squared error of the signal waveform estimation at the destination node. To minimize the total system transmission power, we apply a new AF relay protocol where the source node transmits signals during both time slots. Two iterative algorithms are proposed to jointly optimize the source and relay precoding matrices for the general case with multiple concurrent data streams, and the special case where a single data stream is transmitted, respectively. Simulation results show the effectiveness of both proposed algorithms. Interestingly, for the single data stream case, the second algorithm converges faster than the first algorithm. Compared with conventional two-hop MIMO relay systems where the source node is silent at the second time slot, the new protocol reduces the system transmission power. Under a given system transmission power, the proposed algorithms yield a higher system mutual information thanks to a better utilization of the direct link. Simulation results in this paper shed lights on some fundamental questions in relay system design such as when direct communication between source and destination (without any relay nodes) is optimal, and when the new AF relay protocol provides a larger gain compared with the conventional single phase source power allocation.

INDEX TERMS MIMO relay, QoS, MSE, direct link, single data stream.

I. INTRODUCTION

Wireless relay communication is an important technique to improve the coverage and reliability of wireless communication systems [1], [2]. By deploying multiple antennas at the nodes, we have multiple-input multiple-output (MIMO) relay systems. As an emerging and promising technique for future high-speed reliable wireless communication, MIMO relay systems have gained much attention recently [3].

In this paper, we focus on amplify-and-forward (AF) MIMO relay systems due to their low cost and simplicity in implementation. For AF MIMO relay systems, the relay precoding matrix that minimizes the mean-squared error (MSE) of the signal waveform estimation at the destination node has been proposed in [4]. A unified framework has been proposed in [5] to optimize the source and relay precoding matrices with a broad class of objective functions. In [4] and [5], the direct link between the source and destination nodes is ignored. However, in practical systems, the direct link provides valuable spatial diversity, and thus, should be properly considered in the MIMO relay system design.

Transceiver optimization considering the direct link has been addressed in [6] based on the optimal structure of the relay precoding matrix. Robust transceiver design for MIMO relay systems with the direct link has been proposed in [7]. The aim of [4]–[7] is to optimize a given objective function, subjecting to the transmission power constraint at the source and relay nodes. However, the quality-of-service (QoS) constraints are not addressed in [4]–[7]. Note that in practical communication systems, QoS criteria are very important, as they greatly affect the user experience.

QoS constraints have been considered in [8] for power allocation in AF MIMO relay systems. In [9], transceiver design has been proposed where the QoS requirements are specified as Schur-convex functions of the MSEs.

Transceiver optimization for two-way MIMO relay systems with MSE constraints has been studied in [10]. However, the direct link is not considered in [8] and [9]. Moreover, we note that in the second time slot of conventional AF MIMO relay systems [3]–[9], the source node is silent and only the relay node forwards the signal to the destination node. In the single antenna relay system of [11] and [12], the second time slot is utilized by the source node to transmit signals to the destination node through the direct link, which improves the system capacity. However, [11] and [12] do not consider MIMO relay systems and QoS constraints.

In this paper, we investigate the transceiver design in a two-hop AF MIMO relay system with the direct link and QoS constraint. The QoS criterion is specified by the upper bound of the MSE of the signal waveform estimation at the destination node, since MSE is directly related to other commonly used QoS criteria such as the system bit-error-rate (BER) and the source-destination mutual information [13]. Different to [3]–[9], we consider a new AF relay protocol [11], [12] where the source node transmits signals during both time slots. Based on this new protocol, we investigate the transceiver design which minimizes the total network transmission power subjecting to the MSE constraint. As this protocol enables the source node to exploit the diversity of the channel in both time slots, we expect that the system transmission power can be reduced compared with conventional systems in [3]–[9]. We would like to note that this new protocol was first proposed for single-antenna AF relay systems [11], [12] and recently extended to AF MIMO relay systems [14]. However, the QoS issue has not been studied in [11], [12], and [14].

Since the transceiver optimization problem is nonconvex with matrix variables, the globally optimal solution is difficult to obtain. We propose two iterative algorithms to solve the problem. In particular, for the general case with multiple concurrent data streams, we iteratively optimize the source, relay, and receiver matrices till convergence. For the special case where a single data stream is transmitted, we iteratively optimize the source precoding vectors and the relay precoding matrix by exploiting the optimal structure of the relay matrix. In this case, the receiver matrix is obtained after the convergence of the algorithm. Simulation results show that both algorithms have fast convergence rate. Interestingly, for the single data stream case, the second algorithm converges faster than the first algorithm, as it reduces the number of optimization variables during iterations. Compared with conventional two-hop MIMO relay systems where the source node is silent at the second time slot, the new protocol reduces the system transmission power, particularly when the direct link is strong. Under a given system transmission power, the proposed algorithms yield a higher system mutual information (MI) thanks to a better utilization of the direct link.

We would like to note that although both [14] and this paper use two phases for source power allocation, the novelties and contributions of this paper compared with [14] are summarized below.

1) The approaches used to optimize the source and relay precoding matrices are different. For general AF MIMO relay systems, the source precoding matrices are updated through quadratically constrained quadratic programming (QCQP) in [14], while in this paper, the Lagrange multiplier method is applied to optimize the source precoding matrices, which has a lower computational complexity than QCQP.
2) The structure of the relay precoding matrix is different. It is demonstrated in this paper that for both multiple data streams and single data stream cases, the structure of the optimal relay precoding matrix depends on the strength of the direct link and the QoS constraint, which is not shown in [14].
3) Compared with [14], more insights on the optimal power allocation in AF MIMO relay systems are revealed in this paper. We find that using both phases for source power allocation provides a larger gain over the conventional single phase source power allocation when the direct link is not very strong or weak. When the direct link is very strong compared with the relay link, direct communication between source and destination (without any relay nodes) is actually optimal.

The rest of the paper is organized as follows. The model of a two-hop AF MIMO relay system with direct link where the source node transmits signals in both time slots is presented in Section II. The transceiver optimization problem with MSE constraints is also formulated in Section II. Two algorithms are proposed in Section III. Section IV shows numerical results which demonstrate the effectiveness of the proposed algorithms. Conclusions are drawn in Section V.

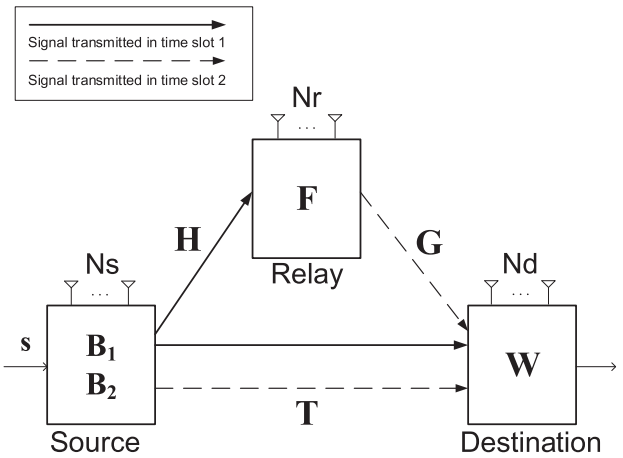


FIGURE 1. Block diagram of a two-hop AF MIMO relay system with direct link where the source node transmits signals in both time slots.

II. SYSTEM MODEL

In this section, we introduce the model of a two-hop AF MIMO relay system with direct link where the source node transmits signals in both time slots. As shown in Fig. 1, the MIMO relay system we consider in this paper has three nodes,

and there are N_s, N_r, and N_d antennas in the source, relay, and destination nodes, respectively. The data transmission from the source node to the destination node is completed in two time slots. At the first time slot, the source node transmits a linearly precoded signal vector $\mathbf{B}_1\mathbf{s}$ to both the relay node and the destination node, where $\mathbf{s}$ is the $N_1 \times 1$ source signal vector and $\mathbf{B}_1$ is the $N_s \times N_1$ source precoding matrix. The signal vectors received at the relay node and the destination node at the first time slot can be written as

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

$$\mathbf{y}_{d,1} = \mathbf{T}\mathbf{B}_1\mathbf{s} + \mathbf{n}_{d,1} \tag{2}$$

where $\mathbf{H}$ is the $N_r \times N_s$ channel matrix between the source node and the relay node, $\mathbf{T}$ is the $N_d \times N_s$ channel matrix between the source node and the destination node, $\mathbf{n}_r$ is the $N_r \times 1$ noise vector at the relay node, and $\mathbf{n}_{d,1}$ is the $N_d \times 1$ noise vector at the destination node at the first time slot. We assume that $E[\mathbf{s}\mathbf{s}^H] = \mathbf{I}_{N_1}$, where $E[\cdot]$ stands for the statistical expectation, $(\cdot)^H$ denotes the Hermitian transpose, and $\mathbf{I}_n$ denotes the $n \times n$ identity matrix.

At the second time slot, $\mathbf{y}_r$ is linearly precoded by an $N_r \times N_r$ matrix $\mathbf{F}$ at the relay node, and the source signal vector $\mathbf{s}$ is linearly precoded by an $N_s \times N_1$ matrix $\mathbf{B}_2$ at the source node. Then the source node and the relay node send the precoded signal vectors to the destination node. From (1), the signal vector received at the destination node is given by

$$\begin{aligned}\mathbf{y}_{d,2} &= \mathbf{G}\mathbf{y}_r + \mathbf{T}\mathbf{B}_2\mathbf{s} + \mathbf{n}_{d,2} \\ &= (\mathbf{G}\mathbf{H}\mathbf{B}_1 + \mathbf{T}\mathbf{B}_2)\mathbf{s} + \mathbf{G}\mathbf{n}_r + \mathbf{n}_{d,2}\end{aligned} \tag{3}$$

where $\mathbf{G}$ is the $N_d \times N_r$ channel matrix between the relay node and the destination node, $\mathbf{n}_{d,2}$ is the $N_d \times 1$ noise vector at the destination node at the second time slot.

From (2) and (3), the received signal vectors at the destination node in two time slots can be written as

$$\begin{aligned}\mathbf{y}_d &= \begin{pmatrix}\mathbf{G}\mathbf{H}\mathbf{B}_1 + \mathbf{T}\mathbf{B}_2 \\ \mathbf{T}\mathbf{B}_1\end{pmatrix}\mathbf{s} + \begin{pmatrix}\mathbf{G}\mathbf{n}_r + \mathbf{n}_{d,2} \\ \mathbf{n}_{d,1}\end{pmatrix} \\ &= \bar{\mathbf{H}}\mathbf{s} + \bar{\mathbf{v}}\end{aligned} \tag{4}$$

where

$$\bar{\mathbf{H}} = \begin{pmatrix}\mathbf{G}\mathbf{H}\mathbf{B}_1 + \mathbf{T}\mathbf{B}_2 \\ \mathbf{T}\mathbf{B}_1\end{pmatrix}, \quad \bar{\mathbf{v}} = \begin{pmatrix}\mathbf{G}\mathbf{n}_r + \mathbf{n}_{d,2} \\ \mathbf{n}_{d,1}\end{pmatrix}.$$

We would like to note that (4) is the most general case for a three-node two-hop AF MIMO relay system. If $\mathbf{B}_2 = \mathbf{0}$, we have a conventional MIMO relay system [3]–[9] where the source node is silent at the second time slot. At the destination node, $\mathbf{y}_d$ is processed by a linear receiver due to its simplicity, and the estimated signal is given by

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

where $\mathbf{W}$ is the $2N_d \times N_1$ weight matrix of the linear receiver.

We assume that all channels are quasi-static, i.e., channel matrices $\mathbf{H}$, $\mathbf{T}$, and $\mathbf{G}$ remain unchanged during one transmission cycle. We assume that all noises are independent and identically distributed additive white Gaussian noise with zero mean and unit variance. We also assume that the destination node knows the channel state information (CSI) of $\mathbf{H}$, $\mathbf{T}$, and $\mathbf{G}$. The destination node performs the transceiver optimization, and then feeds back the optimized $\mathbf{B}_1$ and $\mathbf{B}_2$ to the source node and $\mathbf{F}$ to the relay node.

The transmission power consumed by the source node during two time slots is given by

$$P_s = tr(\mathbf{B}_1\mathbf{B}_1^H + \mathbf{B}_2\mathbf{B}_2^H) \tag{6}$$

where $tr(\cdot)$ denotes the matrix trace. From (1), the transmission power consumed by the relay node is

$$P_r = tr(\mathbf{F}(\mathbf{H}\mathbf{B}_1\mathbf{B}_1^H\mathbf{H}^H + \mathbf{I}_{N_r})\mathbf{F}^H). \tag{7}$$

From (6) and (7), the total system transmission power can be written as

$$P_t = P_r + P_s. \tag{8}$$

From (4) and (5), the MSE of the signal waveform estimation at the destination node is given by

$$\begin{aligned}\mathrm{MSE} &= tr(E[(\hat{\mathbf{s}} - \mathbf{s})(\hat{\mathbf{s}} - \mathbf{s})^H]) \\ &= tr((\mathbf{W}^H\bar{\mathbf{H}} - \mathbf{I}_{N_1})(\mathbf{W}^H\bar{\mathbf{H}} - \mathbf{I}_{N_1})^H + \mathbf{W}^H\mathbf{C}\mathbf{W})\end{aligned} \tag{9}$$

where $\mathbf{C}$ is the equivalent noise covariance matrix given by

$$\mathbf{C} = E[\bar{\mathbf{v}}\bar{\mathbf{v}}^H] = \begin{pmatrix}\mathbf{G}\mathbf{F}\mathbf{F}^H\mathbf{G}^H + \mathbf{I}_{N_d} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{N_d}\end{pmatrix}.$$

Based on (8) and (9), the transceiver optimization problem which minimizes the total system transmission power subjecting to the MSE constraint can be written as

$$\min_{\mathbf{B}_1,\mathbf{B}_2,\mathbf{W},\mathbf{F}} P_t \tag{10}$$

$$\text{s.t. } \mathrm{MSE} \le e \tag{11}$$

where e denotes the upper-bound of MSE. We would like to note that MSE-based transceiver designs have been considered extensively due to their good performance and significantly reduced complexity. It has been shown in [15] that minimal MSE (MMSE) estimation plays an important role in approaching the information-theoretic limits of Gaussian channels. As mentioned in [16], the user-wise MSE can be used to approximate the achievable rate of the users when they jointly decode their streams. In particular, when MMSE receivers are employed, the achievable rate of a user is written as the negative logarithm of the determinant of the MSE error covariance matrix, which is tightly related to the user MSE.

III. PROPOSED TRANSCEIVER OPTIMIZATION ALGORITHMS

The problem (10)-(11) is nonconvex with matrix variables and the globally optimal solution is difficult to obtain. Moreover, compared with the transceiver design for conventional two-hop AF MIMO relay systems [8], [9], the problem (10)-(11) is more challenging to solve as $\mathbf{B}_2$ also needs to be optimized. In this section, we propose two iterative algorithms to solve the problem (10)-(11). The first algorithm

is developed for the general case of multiple concurrent data streams ($N_1 \geq 1$), where we iteratively update $\mathbf{W}$, $(\mathbf{B}_1, \mathbf{B}_2)$, and $\mathbf{F}$ till convergence. The second algorithm is designed for the single data stream case where $N_1 = 1$. In this case, we iteratively optimize $(\mathbf{b}_1, \mathbf{b}_2)$ and $\mathbf{F}$ till convergence, and $\mathbf{W}$ is obtained after the convergence of $(\mathbf{b}_1, \mathbf{b}_2)$ and $\mathbf{F}$.

A. TRANSCEIVER OPTIMIZATION ALGORITHM FOR $N_1 \geq 1$

In this subsection, we present an iterative algorithm to solve the problem (10)-(11) for the general case where multiple concurrent data streams are transmitted from the source node to the destination node.

Firstly, we consider the optimal receiver matrix $\mathbf{W}$. Since the objective function (10) does not contain $\mathbf{W}$, with given $\mathbf{B}_1$, $\mathbf{B}_2$, and $\mathbf{F}$, the optimal $\mathbf{W}$ that minimizes the left-hand side (LHS) of (11) is the well-known Wiener filter [17] given by

$$\mathbf{W} = (\mathbf{C} + \bar{\mathbf{H}}\bar{\mathbf{H}}^H)^{-1}\bar{\mathbf{H}} \tag{12}$$

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

Secondly, we fix $\mathbf{W}$, $\mathbf{B}_1$, and $\mathbf{B}_2$ and optimize the relay precoding matrix $\mathbf{F}$ by solving the following problem

$$\min_{\mathbf{F}} tr(\mathbf{F}(\mathbf{H}\mathbf{B}_1\mathbf{B}_1^H\mathbf{H}^H + \mathbf{I}_{N_r})\mathbf{F}^H) \tag{13}$$

$$\text{s.t. MSE} \leq e. \tag{14}$$

The problem (13)-(14) is a convex optimization problem with a convex quadratic objective function and a convex quadratic constraint. The Lagrangian function associated with the problem (13)-(14) is given by

$$\mathcal{L}(\lambda, \mathbf{F}) = tr(\mathbf{F}(\mathbf{H}\mathbf{B}_1\mathbf{B}_1^H\mathbf{H}^H + \mathbf{I}_{N_r})\mathbf{F}^H) + \lambda(\text{ MSE} - e) \tag{15}$$

where $\lambda \geq 0$ is the Lagrange multiplier. By taking the partial derivative of $\mathcal{L}(\lambda, \mathbf{F})$ with respect to $\mathbf{F}$ and let $\partial\mathcal{L}(\lambda, \mathbf{F})/\partial\mathbf{F} = 0$, we obtain the optimal structure of $\mathbf{F}$ as

$$\mathbf{F} = \mathbf{F}_1\mathbf{F}_2\mathbf{F}_3 \tag{16}$$

where

$$\mathbf{F}_1 = \lambda(\mathbf{I}_{N_r} + \lambda\mathbf{G}^H\mathbf{W}_1\mathbf{W}_1^H\mathbf{G})^{-1}\mathbf{G}^H\mathbf{W}_1 \tag{17}$$

$$\mathbf{F}_2 = \mathbf{I}_{N_1} - \mathbf{W}_2^H\mathbf{T}\mathbf{B}_1 - \mathbf{W}_1^H\mathbf{T}\mathbf{B}_2 \tag{18}$$

$$\mathbf{F}_3 = \mathbf{B}_1^H\mathbf{H}^H(\mathbf{H}\mathbf{B}_1\mathbf{B}_1^H\mathbf{H}^H + \mathbf{I}_{N_r})^{-1} \tag{19}$$

and $\mathbf{W}_1$, $\mathbf{W}_2$ contain the first and the last N_d rows of $\mathbf{W}$, respectively.

The unknown λ can be obtained from the following complementary slackness condition of the problem (13)-(14) as

$$\lambda(\text{ MSE} - e) = 0. \tag{20}$$

If $\lambda = 0$, then (20) holds. Based on (17) and (16), this results in $\mathbf{F} = \mathbf{0}$. If the MSE constraint (14) is satisfied under $\mathbf{F} = \mathbf{0}$ for given $\mathbf{W}$, $\mathbf{B}_1$, and $\mathbf{B}_2$, then $\mathbf{F} = \mathbf{0}$ is the optimal solution to the problem (13)-(14). This may occur when the direct link is strong such that it is optimal to switch off the relay node. Otherwise, there must be $\lambda > 0$ such that MSE $= e$ in (20). The steps of obtaining such λ are presented in Appendix A. Then the optimal $\mathbf{F}$ is obtained from (16). We would like to note that as a contribution of this paper, the dependence of $\mathbf{F}$ on the strength of the direct link and the QoS constraint is not shown in [14].

Thirdly, we fix $\mathbf{W}$ and $\mathbf{F}$ and optimize the source precoding matrices $\mathbf{B}_1$ and $\mathbf{B}_2$. With fixed $\mathbf{F}$, we can ignore the term $tr(\mathbf{F}\mathbf{F}^H)$ in the objective function (10). By introducing

$$\mathbf{L}_1^H = \mathbf{W}_2^H\mathbf{T} + \mathbf{W}_1^H\mathbf{G}\mathbf{F}\mathbf{H}, \quad \mathbf{L}_2^H = \mathbf{W}_1^H\mathbf{T}$$
$$\mathbf{D} = \mathbf{I}_{N_s} + \mathbf{H}^H\mathbf{F}^H\mathbf{F}\mathbf{H}, \quad \hat{e} = tr(\mathbf{W}^H\mathbf{C}\mathbf{W})$$

the problem (10)-(11) can be rewritten as

$$\min_{\mathbf{B}_1,\mathbf{B}_2} tr(\mathbf{B}_1^H\mathbf{D}\mathbf{B}_1 + \mathbf{B}_2^H\mathbf{B}_2) \tag{21}$$

$$\text{s.t. } tr(\bar{\mathbf{B}}\bar{\mathbf{B}}^H) \leq e - \hat{e} \tag{22}$$

where

$$\bar{\mathbf{B}} = \mathbf{I}_{N_1} - \mathbf{L}_1^H\mathbf{B}_1 - \mathbf{L}_2^H\mathbf{B}_2. \tag{23}$$

We apply the Lagrange multiplier method [18] to solve the problem (21)-(22). The Lagrangian function of (21)-(22) is

$$\mathcal{L}(\mathbf{B}_1, \mathbf{B}_2, \mu) = tr(\mathbf{B}_1^H\mathbf{D}\mathbf{B}_1 + \mathbf{B}_2^H\mathbf{B}_2) + \mu(tr(\bar{\mathbf{B}}\bar{\mathbf{B}})^H - e + \hat{e})$$

where $\mu \geq 0$ is the Lagrange multiplier.

By letting $\frac{\partial\mathcal{L}}{\partial\mathbf{B}_1} = \mathbf{0}$ and $\frac{\partial\mathcal{L}}{\partial\mathbf{B}_2} = \mathbf{0}$, where $\frac{\partial\mathcal{L}}{\partial\mathbf{B}_i}$ denotes the partial derivative of $\mathcal{L}$ with respect to $\mathbf{B}_i$, we obtain

$$\mathbf{B}_1 = \mu(\mathbf{D} + \mu\mathbf{L}_1\mathbf{L}_1^H)^{-1}\mathbf{L}_1(\mathbf{I}_{N_1} - \mathbf{L}_2^H\mathbf{B}_2) \tag{24}$$

$$\mathbf{B}_2 = \mu(\mathbf{I}_{N_s} + \mu\mathbf{L}_2\mathbf{L}_2^H)^{-1}\mathbf{L}_2(\mathbf{I}_{N_1} - \mathbf{L}_1^H\mathbf{B}_1). \tag{25}$$

The Lagrange multiplier μ in (24) and (25) can be obtained through the complementary slackness condition of the problem (21)-(22) given by

$$\mu(tr(\bar{\mathbf{B}}\bar{\mathbf{B}}^H) - e + \hat{e}) = 0. \tag{26}$$

Note that if $\mu = 0$, then (26) holds. But this results in both $\mathbf{B}_1 = \mathbf{0}$ and $\mathbf{B}_2 = \mathbf{0}$ according to (24) and (25). Therefore, there must be a $\mu > 0$ such that $\bar{\mathbf{B}}$ in (23) satisfies

$$tr(\bar{\mathbf{B}}\bar{\mathbf{B}}^H) = e - \hat{e}. \tag{27}$$

Lemma 1: By solving $\mathbf{L}_1^H\mathbf{B}_1$ and $\mathbf{L}_2^H\mathbf{B}_2$ from (24) and (25), we obtain the following equation

$$\bar{\mathbf{B}} = (\mu(\mathbf{L}_1\mathbf{D}^{-1}\mathbf{L}_1^H + \mathbf{L}_2\mathbf{L}_2^H) + \mathbf{I}_{N_1})^{-1}. \tag{28}$$

Proof: See Appendix B. ■

By substituting (28) into (27), and introducing the eigenvalue decomposition (EVD) of $\mathbf{L}_1^H\mathbf{D}^{-1}\mathbf{L}_1 + \mathbf{L}_2^H\mathbf{L}_2 = \mathbf{Q}\boldsymbol{\Upsilon}\mathbf{Q}^H$, where $\boldsymbol{\Upsilon}$ is the eigenvalue matrix, (27) can be equivalently rewritten as

$$\sum_{i=1}^{N_1} \frac{1}{(1 + \upsilon_i\mu)^2} = e - \hat{e} \tag{29}$$

where υ_i is the ith diagonal element of $\boldsymbol{\Upsilon}$. It can be seen that as $\upsilon_i \geq 0$, $i = 1, \ldots, N_1$, the LHS of (29) is monotonically decreasing with respect to μ when $\mu > 0$. Thus, we

can efficiently solve μ from (29) for example by bisection method [18]. After obtaining μ, we can obtain $\mathbf{L}_1^H\mathbf{B_1}$ and $\mathbf{L}_2^H\mathbf{B}_2$ according to (70) and (71), respectively, and then obtain the optimal $\mathbf{B}_1$ and $\mathbf{B}_2$ from (24) and (25). Note that in [14], $\mathbf{B}_1$ and $\mathbf{B}_2$ are updated through quadratically constrained quadratic programming (QCQP), which has a higher computational complexity than the Lagrange multiplier method.

TABLE 1. Procedure of the proposed algorithm for $N_1 \geq 1$.

1) Generate feasible $\mathbf{B}_1^{(0)}, \mathbf{B}_2^{(0)}$, and $\mathbf{F}^{(0)}$; set $n = 0$.
2) Use (12) to calculate $\mathbf{W}^{(n)}$ using $\mathbf{B}_1^{(n)}, \mathbf{B}_2^{(n)}$, and $\mathbf{F}^{(n)}$.
3) Set $\mathbf{F}^{(n+1)} = \mathbf{0}$ and check if it satisfies (14). Otherwise, solve λ from (61) and obtain $\mathbf{F}^{(n+1)}$ from (16).
4) Solve μ from (29) and obtain $\mathbf{L}_1^H\mathbf{B}_1^{(n+1)}$ and $\mathbf{L}_2^H\mathbf{B}_2^{(n+1)}$ from (70) and (71). Then obtain the optimal $\mathbf{B}_1^{(n+1)}$ and $\mathbf{B}_2^{(n+1)}$ from (24) and (25).
5) Calculate $P_t^{(n+1)}$ in (8) using $\mathbf{F}^{(n+1)}$, $\mathbf{B}_1^{(n+1)}$, and $\mathbf{B}_2^{(n+1)}$. If $\frac{P_t^{(n)} - P_t^{(n+1)}}{P_t^{(n)}} < \varepsilon$, iteration ends. Otherwise set $n = n + 1$ and go to Step 2).

With the analysis above, we can jointly optimize the source, relay, and receiver matrices as shown in Table 1. This algorithm starts with feasible $\mathbf{F}$, $\mathbf{B}_1$, and $\mathbf{B}_2$, and iteratively optimizes $\mathbf{W}$, $\mathbf{F}$, and $(\mathbf{B}_1, \mathbf{B}_2)$. At each iteration, $\mathbf{W}$ is firstly updated using (12). Then we update $\mathbf{F}$ by solving λ from (61). Finally we calculate μ in (29) and update $\mathbf{B}_1$ and $\mathbf{B}_2$ according to (24) and (25). This iterative process continues until the termination condition is satisfied. We would like to note that since the objective function P_t is non-increasing in every iteration and $P_t \geq 0$, the convergence of this algorithm follows from this observation.

B. TRANSCEIVER OPTIMIZATION ALGORITHM FOR $N_1 = 1$

In the case of a single data stream, the dimension of source signal vector $\mathbf{s}$ is 1. Thus, the source precoding matrices $\mathbf{B}_1$ and $\mathbf{B}_2$ become vectors $\mathbf{b}_1$ and $\mathbf{b}_2$, and the receiver matrices $\mathbf{W}_1$ and $\mathbf{W}_2$ become vectors $\mathbf{w}_1$ and $\mathbf{w}_2$. In this subsection, we develop an iterative algorithm to solve the transceiver optimization problem (10)-(11) by exploiting the rank-1 structure of the relay precoding matrix in the special case of $N_1 = 1$.

By substituting (12) back into (9) and considering that $N_1 = 1$, the MSE becomes

$$\text{MSE} = (1 + \mathbf{b}_1^H\mathbf{T}^H\mathbf{T}\mathbf{b}_1 + \tilde{\mathbf{b}}^H\bar{\mathbf{C}}^{-1}\tilde{\mathbf{b}})^{-1} \tag{30}$$

where $\tilde{\mathbf{b}} = \mathbf{T}\mathbf{b}_2 + \mathbf{G}\mathbf{F}\mathbf{H}\mathbf{b}_1$ and $\bar{\mathbf{C}} = \mathbf{G}\mathbf{F}\mathbf{F}^H\mathbf{G}^H + \mathbf{I}_{N_d}$. The total transmission power in this case becomes

$$P_t = \mathbf{b}_1^H\mathbf{b}_1 + \mathbf{b}_2^H\mathbf{b}_2 + tr(\mathbf{F}(\mathbf{H}_1\mathbf{b}_1\mathbf{b}_1^H\mathbf{H}_1^H + \mathbf{I}_{N_r})\mathbf{F}^H). \tag{31}$$

Based on (30) and (31), the problem (10)-(11) can be rewritten as

$$\min_{\mathbf{b}_1,\mathbf{b}_2,\mathbf{F}} \ \mathbf{b}_1^H\mathbf{b}_1 + \mathbf{b}_2^H\mathbf{b}_2 + tr(\mathbf{F}(\mathbf{H}_1\mathbf{b}_1\mathbf{b}_1^H\mathbf{H}_1^H + \mathbf{I}_{N_r})\mathbf{F}^H) \tag{32}$$

$$\text{s.t. } 1 + \mathbf{b}_1^H\mathbf{T}^H\mathbf{T}\mathbf{b}_1 + \tilde{\mathbf{b}}^H\bar{\mathbf{C}}^{-1}\tilde{\mathbf{b}} \geq 1/e. \tag{33}$$

We first optimize the relay matrix $\mathbf{F}$ with $\mathbf{b}_1$ and $\mathbf{b}_2$ fixed. We have shown that the optimal $\mathbf{F}$ has the structure in (16)-(19). In the single data stream case, (16) becomes

$$\mathbf{F} = \alpha\beta\gamma\mathbf{G}^H\mathbf{w}_1\mathbf{b}_1^H\mathbf{H}^H \tag{34}$$

where $\alpha = \lambda/(1 + \lambda\mathbf{w}_1^H\mathbf{G}\mathbf{G}^H\mathbf{w}_1)$, $\beta = 1/(1 + \mathbf{b}_1^H\mathbf{H}^H\mathbf{H}\mathbf{b}_1)$, and $\gamma = 1 - \mathbf{w}_2^H\mathbf{T}\mathbf{b}_1 - \mathbf{w}_1^H\mathbf{T}\mathbf{b}_2$. Similar to the case of $N_1 \geq 1$, we first check whether $\lambda = 0$ is a valid solution, if not then there is $\lambda > 0$. When $\lambda > 0$, by introducing $\mathbf{q} = \alpha\beta\gamma\mathbf{G}^H\mathbf{w}_1$, (34) can be rewritten as

$$\mathbf{F} = \mathbf{q}\mathbf{b}_1^H\mathbf{H}^H. \tag{35}$$

It can be seen from (35) that the optimal $\mathbf{F}$ in the single data stream case is rank-1. Using (35), $\tilde{\mathbf{b}}$ and $\bar{\mathbf{C}}$ in (30) become

$$\tilde{\mathbf{b}} = \mathbf{T}\mathbf{b}_2 + c\mathbf{G}\mathbf{q}, \quad \bar{\mathbf{C}} = c\mathbf{G}\mathbf{q}\mathbf{q}^H\mathbf{G}^H + \mathbf{I}_{N_d} \tag{36}$$

where $c = \mathbf{b}_1^H\mathbf{H}^H\mathbf{H}\mathbf{b}_1$. Based on (35), (36), and the following identity

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

we obtain

$$\tilde{\mathbf{b}}^H\bar{\mathbf{C}}^{-1}\tilde{\mathbf{b}} = \mathbf{b}_2^H\mathbf{T}^H\mathbf{T}\mathbf{b}_2 + c\left(1 - \frac{|\mathbf{b}_2^H\mathbf{T}^H\mathbf{G}\mathbf{q} - 1|^2}{1 + c\mathbf{q}^H\mathbf{G}^H\mathbf{G}\mathbf{q}}\right). \tag{37}$$

Using (35), the transmission power consumed by the relay node becomes $(c^2 + c)\mathbf{q}^H\mathbf{q}$. Based on (37), it can be seen that with fixed $\mathbf{b}_1$ and $\mathbf{b}_2$, the problem (32)-(33) can be rewritten as the following problem

$$\min_{\mathbf{q}} \ \mathbf{q}^H\mathbf{q} \tag{38}$$

$$\text{s.t. } \frac{|\mathbf{b}_2^H\mathbf{T}^H\mathbf{G}\mathbf{q} - 1|^2}{1 + c\mathbf{q}^H\mathbf{G}^H\mathbf{G}\mathbf{q}} \leq \tilde{e} \tag{39}$$

where $\tilde{e} = \frac{1}{c}\left(1 + \mathbf{b}_1^H(\mathbf{T}^H\mathbf{T} + \mathbf{H}^H\mathbf{H})\mathbf{b}_1 + \mathbf{b}_2^H\mathbf{T}^H\mathbf{T}\mathbf{b}_2 - \frac{1}{e}\right)$. In the following, we show that the problem (38)-(39) can be efficiently solved by the Charnes-Cooper transformation [19] and the semidefinite relaxation (SDR) technique [20].

The constraint (39) can be rewritten as

$$\mathbf{q}^H\mathbf{G}^H\mathbf{T}\mathbf{b}_2\mathbf{b}_2^H\mathbf{T}^H\mathbf{G}\mathbf{q} + a \leq \tilde{e}c(\mathbf{G}\mathbf{q} + \boldsymbol{\eta})^H(\mathbf{G}\mathbf{q} + \boldsymbol{\eta}) \tag{40}$$

where $\boldsymbol{\eta} = \frac{1}{\tilde{e}c}\mathbf{T}\mathbf{b}_1$ and $a = 1 - \tilde{e} + \tilde{e}c\boldsymbol{\eta}^H\boldsymbol{\eta}$. Let us introduce an auxiliary variable t with $|t| = 1$ and $\tilde{\mathbf{f}} = t\mathbf{q}$. By multiplying both sides of (40) with $|t|^2$, the constraint (40) becomes

$$\tilde{\mathbf{f}}^H\mathbf{G}^H\mathbf{T}\mathbf{b}_2\mathbf{b}_2^H\mathbf{T}^H\mathbf{G}\tilde{\mathbf{f}} + a \leq \tilde{e}c(\mathbf{G}\tilde{\mathbf{f}} + \boldsymbol{\eta}t)^H(\mathbf{G}\tilde{\mathbf{f}} + \boldsymbol{\eta}t). \tag{41}$$

By introducing $\bar{\mathbf{f}} = [\tilde{\mathbf{f}}^T, t]^T$ in (41), the problem (38)-(39) can be equivalently rewritten as

$$\min_{\bar{\mathbf{f}}} \ \bar{\mathbf{f}}^H\mathbf{A}_1\bar{\mathbf{f}} \tag{42}$$

$$\text{s.t. } \bar{\mathbf{f}}^H\mathbf{A}_2\bar{\mathbf{f}} + a \leq \tilde{e}c\bar{\mathbf{f}}^H\mathbf{A}_3\bar{\mathbf{f}} \tag{43}$$

$$\bar{\mathbf{f}}^H\mathbf{A}_4\bar{\mathbf{f}} = 1 \tag{44}$$

where

$$\mathbf{A}_1 = \begin{pmatrix} \mathbf{I}_{N_r} & \\ & 0 \end{pmatrix}, \quad \mathbf{A}_2 = \begin{pmatrix} \mathbf{G}^H\mathbf{T}\mathbf{b}_2\mathbf{b}_2^H\mathbf{T}^H\mathbf{G} & \\ & 0 \end{pmatrix}$$
$$\mathbf{A}_3 = \begin{pmatrix} \mathbf{G}^H \\ \boldsymbol{\eta}^H \end{pmatrix}(\mathbf{G} \quad \boldsymbol{\eta}), \quad \mathbf{A}_4 = \begin{pmatrix} \mathbf{0} & \\ & 1 \end{pmatrix}.$$

By using the SDR technique, the problem (42)-(44) can be relaxed to the following semidefinite programming (SDP) problem

$$\min_{\mathbf{R}\succeq 0} \ tr(\mathbf{R}\mathbf{A}_1) \tag{45}$$
$$\text{s.t. } tr(\mathbf{R}\mathbf{A}_2) + a \le \tilde{e}c\, tr(\mathbf{R}\mathbf{A}_3) \tag{46}$$
$$tr(\mathbf{R}\mathbf{A}_4) = 1. \tag{47}$$

The SDP problem (45)-(47) can be efficiently solved by interior-point methods [21].

The optimal $\bar{\mathbf{f}}$ for the problem (42)-(44) can be found by first solving the SDP problem (45)-(47) and then recovering $\bar{\mathbf{f}}$ from $\mathbf{R}$ as follows. If the solution $\mathbf{R}$ is rank-1, i.e., $\mathbf{R} = \mathbf{r}\mathbf{r}^H$, then we let $\bar{\mathbf{f}} = \mathbf{r}$. For the case of higher rank $\mathbf{R}$, according to Theorem 2.3 (Algorithm 3) in [22], we can find a vector $\bar{\mathbf{f}}_0$ in polynomial time such that $tr(\bar{\mathbf{f}}_0^H\mathbf{A}_i\bar{\mathbf{f}}_0) = tr(\mathbf{R}\mathbf{A}_i)$, $i = 1, 2, 3, 4$, which means we can find a rank-1 optimal solution $\bar{\mathbf{f}}_0\bar{\mathbf{f}}_0^H$ to the relaxation problem (45)-(47) and $\bar{\mathbf{f}}_0$ is the optimal solution to the original problem (42)-(44). Then we can obtain $\mathbf{q} = \bar{\mathbf{f}}/t$ and the optimal $\mathbf{F} = \mathbf{q}\mathbf{b}_1^H\mathbf{H}^H$.

After optimizing $\mathbf{F}$, we fix $\mathbf{F}$ in the problem (32)-(33) and optimize the source precoding vectors $\mathbf{b}_1$ and $\mathbf{b}_2$. With fixed $\mathbf{F}$, by introducing a vector $\mathbf{b} = [\mathbf{b}_1^T, \mathbf{b}_2^T]^T$, the constraint (33) can be rewritten as

$$(1 + \mathbf{b}^H\boldsymbol{\Phi}\mathbf{b})^{-1} \le e \tag{48}$$

where

$$\boldsymbol{\Phi} = \begin{pmatrix} \mathbf{H}^H\mathbf{F}^H\mathbf{G}^H \\ \mathbf{T}^H \end{pmatrix} \bar{\mathbf{C}}^{-1} \begin{pmatrix} \mathbf{G}\mathbf{F}\mathbf{H} & \mathbf{T} \end{pmatrix} + \begin{pmatrix} \mathbf{T}^H\mathbf{T} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}.$$

With fixed $\mathbf{F}$, we can ignore the item $tr(\mathbf{F}\mathbf{F}^H)$ in (32). And the objective function can be rewritten as $\mathbf{b}^H\mathbf{P}\mathbf{b}$, where

$$\mathbf{P} = \begin{pmatrix} \mathbf{H}^H\mathbf{F}^H\mathbf{F}\mathbf{H} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} + \mathbf{I}_{2N_s}. \tag{49}$$

Using (48) and (49), the problem (32)-(33) can be rewritten as

$$\min_{\mathbf{b}} \ \mathbf{b}^H\mathbf{P}\mathbf{b} \tag{50}$$
$$\text{s.t. } \mathbf{b}^H\boldsymbol{\Phi}\mathbf{b} \ge \frac{1}{e} - 1. \tag{51}$$

The problem (50)-(51) can be solved by using the SDR technique, where we obtain a relaxation of the problem (50)-(51) as

$$\min_{\tilde{\mathbf{Q}}\succeq 0} \ tr(\mathbf{P}\tilde{\mathbf{Q}}) \tag{52}$$
$$\text{s.t. } tr(\boldsymbol{\Phi}\tilde{\mathbf{Q}}) \ge \frac{1}{e} - 1. \tag{53}$$

After solving the SDP problem (52)-(53), by using the randomization technique in [23], the optimal $\mathbf{b}$ for the problem (50)-(51) can be recovered from $\tilde{\mathbf{Q}}$ by finding a vector $\mathbf{b}_0$ such that $tr(\mathbf{P}\tilde{\mathbf{Q}}) = \mathbf{b}_0^H\mathbf{P}\mathbf{b}_0$ and $tr(\boldsymbol{\Phi}\tilde{\mathbf{Q}}) = \mathbf{b}_0^H\boldsymbol{\Phi}\mathbf{b}_0$. Then the optimal $\mathbf{b}_1$ and $\mathbf{b}_2$ can be obtained from $\mathbf{b}_0$.

TABLE 2. Procedure of the proposed algorithm for $N_1 = 1$.

1) Generate feasible $\mathbf{b}_1^{(0)}$, $\mathbf{b}_2^{(0)}$, and $\mathbf{q}^{(0)}$; set $n = 0$.
2) Solve the SDP problem (45)-(47) to obtain the optimal $\mathbf{R}$. Then use Algorithm 3 in [22] to find a vector $\bar{\mathbf{f}}$ such that $tr(\mathbf{R}\mathbf{A}_i) = \bar{\mathbf{f}}^H\mathbf{A}_i\bar{\mathbf{f}}$, $i = 1, 2, 3, 4$, and obtain $\mathbf{q}$ according to $[t\mathbf{q}^T, t] = \bar{\mathbf{f}}^T$. Let $\mathbf{F}^{(n+1)} = \mathbf{q}(\mathbf{b}_1^{(n)})^H\mathbf{H}^H$.
3) Solve the SDP problem (52)-(53) to obtain the optimal $\tilde{\mathbf{Q}}$. Then find $\mathbf{b}$ using [23] such that $tr(\mathbf{P}\tilde{\mathbf{Q}}) = \mathbf{b}^H\mathbf{P}\mathbf{b}$ and $tr(\boldsymbol{\Phi}\tilde{\mathbf{Q}}) = \mathbf{b}^H\boldsymbol{\Phi}\mathbf{b}$, and obtain $\mathbf{b}_1^{(n+1)}$ and $\mathbf{b}_2^{(n+1)}$ from $[(\mathbf{b}_1^{n+1})^T, (\mathbf{b}_2^{n+1})^T] = \mathbf{b}^T$.
4) Calculate the total transmission power $P_t^{(n+1)}$ using $\mathbf{F}^{(n+1)}$, $\mathbf{b}_1^{(n+1)}$, and $\mathbf{b}_2^{(n+1)}$. If $\frac{P_t^{(n)} - P_t^{(n+1)}}{P_t^{(n)}} < \varepsilon$, iteration ends. Otherwise set $n = n + 1$ and go to Step 2).

Now we can iteratively optimize the source precoding vectors $\mathbf{b}_1$, $\mathbf{b}_2$, and the relay precoding matrix $\mathbf{F}$. In each iteration, we first solve the SDP problem (45)-(47) and obtain the optimal $\mathbf{F}$ as (35) by recovering $\mathbf{q}$ from $\mathbf{R}$. Then we solve the SDP problem (52)-(53) and obtain the optimal $\mathbf{b}_1$ and $\mathbf{b}_2$ by recovering them from $\tilde{\mathbf{Q}}$. The procedure of the proposed algorithm for the case of $N_1 = 1$ is summarized in Table 2. Since the update of $\mathbf{b}$ and $\mathbf{F}$ can only decrease or maintain the objective function (31), the convergence of the proposed algorithm in Table 2 follows from this observation.

IV. NUMERICAL RESULTS

In this section, we study the performance of two proposed algorithms through numerical simulations. In the simulations, the channel matrices $\mathbf{H}$, $\mathbf{G}$, and $\mathbf{T}$ are set to complex Gaussian random matrices whose entries are independent and identically distributed (i.i.d.) Gaussian random variables with zero mean and variance of σ_H, σ_G, and σ_T, respectively. We set $\sigma_H = \sigma_G = 1$ and introduce $\Delta = \sigma_G/\sigma_T$ to measure the gain of the relay link over the direct link. A large Δ indicates a week direct link relative to the relay link. For convenience, we refer to the algorithm for the general case of $N_1 \ge 1$ as Algorithm 1 and the algorithm for the special case of $N_1 = 1$ as Algorithm 2. The SDP problems in Algorithm 2 are solved by the CVX convex optimization toolbox [24]. As comparison, the performance of the conventional AF MIMO relay system where the source node is silent at the second time slot (i.e. $\mathbf{B}_2 = \mathbf{0}$ for $N_1 > 1$ and $\mathbf{b}_2 = \mathbf{0}$ for $N_1 = 1$) is shown. We also include the performance of a two-hop AF MIMO relay system without the direct link ($\sigma_T = 0$) as a benchmark [5].

Let us introduce the singular value decompositions (SVDs) of $\mathbf{H} = \mathbf{U}_h\mathbf{D}_h\mathbf{V}_h^H$, $\mathbf{G} = \mathbf{U}_g\mathbf{D}_g\mathbf{V}_g^H$, and $\mathbf{T} = \mathbf{U}_t\mathbf{D}_t\mathbf{V}_t^H$. The proposed algorithms are initialized by $\mathbf{B}_1 = \zeta\mathbf{V}_{h,1}\mathbf{D}_1$, $\mathbf{F} = \zeta\mathbf{V}_{g,1}\mathbf{D}_f\mathbf{U}_{h,1}^H$, and $\mathbf{B}_2 = \mathbf{0}$, where ζ is a positive value, $\mathbf{V}_{h,1}$, $\mathbf{V}_{g,1}$, and $\mathbf{U}_{h,1}$ contain N_1 columns of $\mathbf{V}_h$, $\mathbf{V}_g$, and $\mathbf{U}_h$, respectively, associated with the corresponding largest N_1

singular values, $\mathbf{D}_1$ and $\mathbf{D}_f$ are $N_1 \times N_1$ diagonal matrices with random positive diagonal elements. Note that such initialization enables a fair comparison with the conventional AF MIMO relay system where the source node is silent at the second time slot. To obtain feasible initial precoding matrices, we start with a small ζ and increase the value of ζ with a given step length until the MSE constraint (11) is satisfied. In the simulations, we have also tried to initialize $\mathbf{B}_2 = \mathbf{V}_{t,1}\mathbf{D}_2$, where $\mathbf{V}_{t,1}$ contains N_1 columns of $\mathbf{V}_t$ associated with the largest N_1 singular values of $\mathbf{T}$, and $\mathbf{D}_2$ is an $N_1 \times N_1$ diagonal matrix with random positive diagonal elements. Simulation results show that these two initialization approaches yield similar results. All simulation results are averaged over 1000 channel realizations.

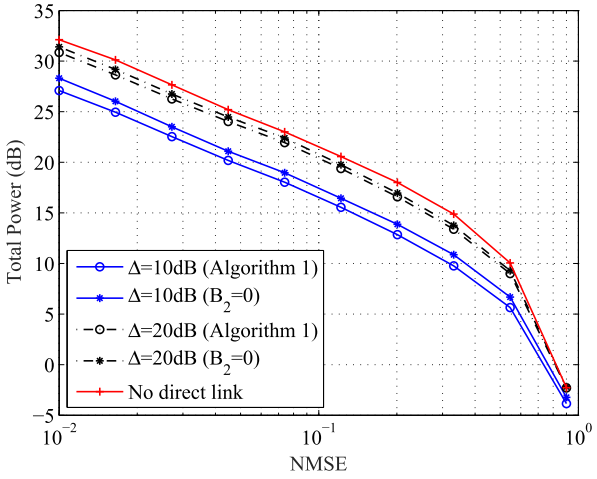


FIGURE 2. Example 1: P_t versus the NMSE at various direct link strength. Algorithm 1, $N_1 = 2$, $N_s = N_r = N_d = 2$.

In the first example, we study the impact of the direct link and $\mathbf{B}_2$ on the system performance. The number of antennas is fixed to $N_s = N_r = N_d = 2$. To test the system performance at various strength of the direct link, we set Δ as 10dB and 20dB. Fig. 2 shows the system total transmission power P_t versus the normalized MSE (NMSE), where $N_1 = 2$ is chosen for Algorithm 1. Here, the NMSE is obtained by dividing the MSE (9) by the number of data streams N_1. It can be seen from Fig. 2 that for all systems, the transmission power required to achieve a certain MSE increases when the MSE constraint becomes stricter (i.e. smaller NMSE). For a given NMSE, a remarkable power saving can be achieved if the direct link is considered, and such power saving is more significant when the MSE constraint is stricter. We can also observe from Fig. 2 that compared with the conventional AF MIMO relay system where $\mathbf{B}_2 = \mathbf{0}$, Algorithm 1 further reduces the system power level, especially in the case of strict MSE constraint and strong direct link (smaller Δ). This can be explained by the fact that when the direct link is strong, it is important to exploit the direct link in the second time slot. While for a weak direct link, more power should be allocated to the relay link. In this case, there is not much difference between $\mathbf{B}_2 = \mathbf{0}$ and $\mathbf{B}_2 \neq \mathbf{0}$.

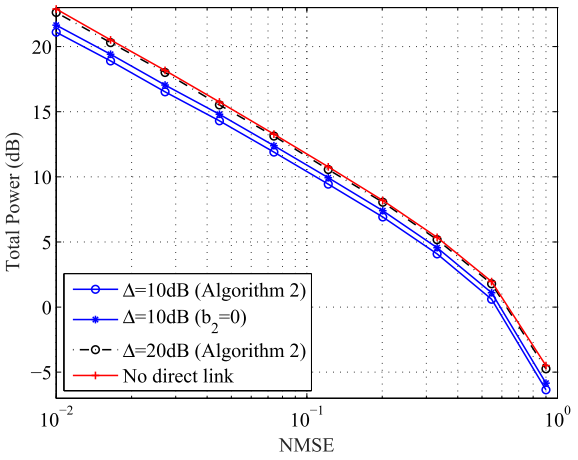


FIGURE 3. Example 1: P_t versus the NMSE at various direct link strength. Algorithm 2, $N_1 = 1$, $N_s = N_r = N_d = 2$.

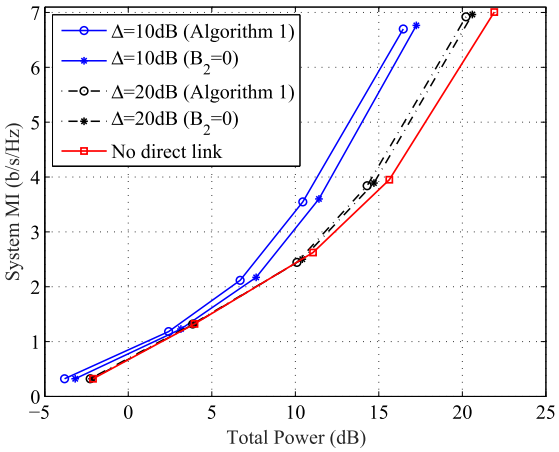


FIGURE 4. Example 2: System MI versus P_t at various direct link strength. Algorithm 1, $N_1 = 2$, $N_s = N_r = N_d = 2$.

Fig. 3 illustrates the system total transmission power versus the NMSE with $N_1 = 1$ using Algorithm 2. Similar to Fig. 2, we can observe from Fig. 3 that the power required to maintain a certain MSE decreases as the direct link becomes stronger. It can also be seen that Algorithm 2 reduces the system power consumption throughout the whole range of NMSE.

In the second example, we show that the performance gain achieved by the two proposed algorithms in terms of saving the system transmission power while ensuring QoS is also reflected in the increase of the system MI under a given power. In this example, the simulation parameters are set identical to those in the first example. The system MI is calculated after $\mathbf{F}$, $\mathbf{B}_1$, and $\mathbf{B}_2$ are obtained using Algorithm 1 in Table 1 as

$$\mathrm{MI} = \log_2 |\mathbf{I}_{N_1} + \bar{\mathbf{H}}^H \mathbf{C}^{-1} \bar{\mathbf{H}}|.$$

Note that we have ignored the constant factor $1/2$ that accounts for the two time slots used. As this factor is ignored for all other schemes, it has no effect on the performance comparison between the proposed algorithms and other schemes. Fig. 4 shows the system MI versus the total transmission

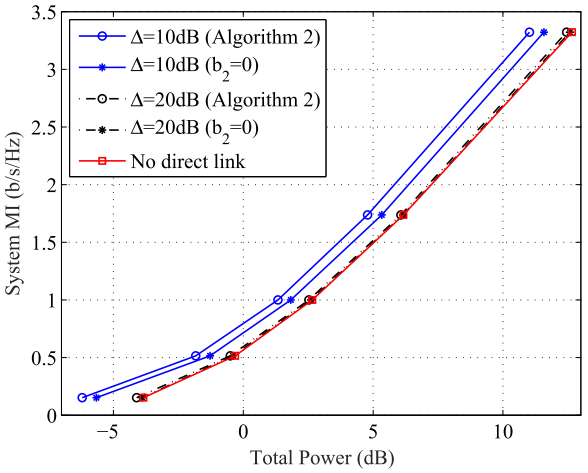


FIGURE 5. Example 2: System MI versus P_t at various direct link strength. Algorithm 2, $N_1 = 1$, $N_s = N_r = N_d = 2$.

power P_t. It can be seen that among the systems tested, Algorithm 1 yields the highest system MI through optimizing $\mathbf{B}_2$. The increase of the MI is more significant at high power levels and when the direct link is stronger.

Fig. 5 illustrates the system MI versus P_t with $N_1 = 1$ using Algorithm 2. Similar to Fig. 4, it can be observed from Fig. 5 that Algorithm 2 has the highest system MI. Moreover, the MI gain is larger at $\Delta = 10$dB than $\Delta = 20$dB. Interestingly, unlike in Fig. 4, the MI gain is almost constant at all power levels. Such difference is mainly due to different number of data streams in two scenarios as will be verified later on.

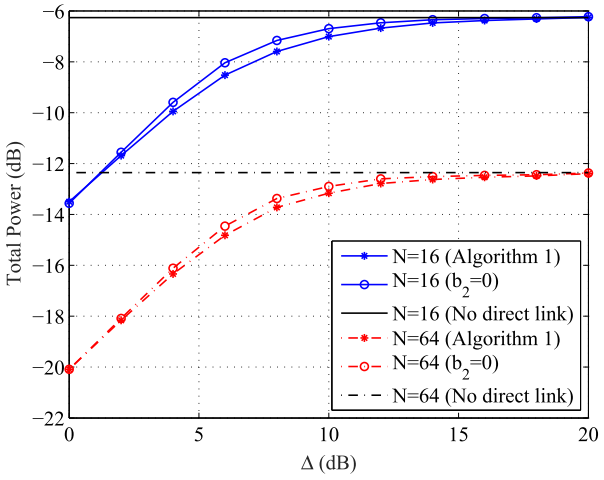


FIGURE 6. Example 3: P_t versus Δ. Algorithm 1, $e = 0.3$, $N_1 = 1$, $N_s = N_r = N_d = N$.

In the third example, we fix the MSE constraint $e = 0.3$ in (11), and vary Δ to provide an insight of Algorithm 1. We set $N_1 = 1$ and $N_s = N_r = N_d = N$. Fig. 6 shows P_t versus Δ. It can be seen that as Δ increases, the system power required approaches to that of the system without the direct link. Interestingly, we observe that the power required under Algorithm 1 is very close to that with $\mathbf{B}_2 = \mathbf{0}$ when Δ is very large or small. This can be explained as follows. When Δ is small, the direct link is strong enough and only little or no power should be transmitted through the relay link. Since the direct link channel $\mathbf{T}$ is assumed to be constant during two time slots, setting $\mathbf{B}_2 = \mathbf{0}$ does not change the system MSE and power consumption. On the other hand, when Δ is large, the direct link is very weak such that all the transmission power should be directed to the relay link. Thus, $\mathbf{B}_2 = \mathbf{0}$ is optimal in this scenario. It can also be observed from Fig. 6 that for both $N = 16$ and $N = 64$, Algorithm 1 saves the system power in the middle range of Δ.

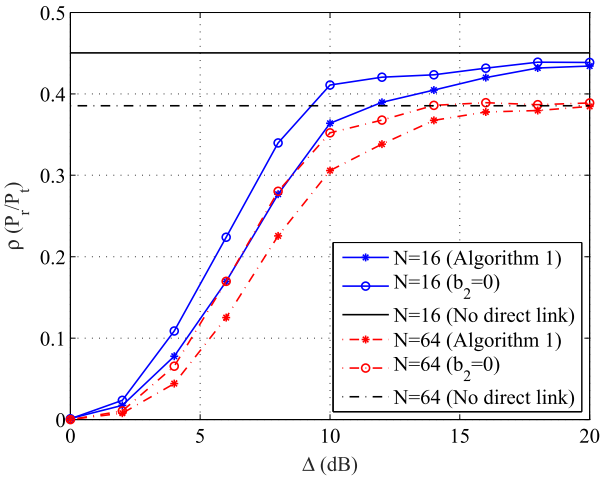


FIGURE 7. Example 3: $\rho = P_r/P_t$ versus Δ. Algorithm 1, $e = 0.3$, $N_1 = 1$, $N_s = N_r = N_d = N$.

Let us introduce $\rho = P_r/P_t$ as the percentage of the transmission power consumed by the relay node out of the total transmission power. Fig. 7 shows ρ versus Δ. We can see that when $\Delta = 0$, the transmission power in the relay link is reduced to 0, which means $\mathbf{F} = \mathbf{0}$ (i.e., (20) holds with $\lambda = 0$) when the direct link is strong. This agrees with the observation in Fig. 6. Moreover, it can be seen from Fig. 7 that ρ increases with Δ and approaches to that of the system without the direct link. Due to a better utilization of the direct link through $\mathbf{B}_2$, Algorithm 1 yields a smaller ρ and converges to the system without the direct link slower compared with setting $\mathbf{B}_2 = \mathbf{0}$. Interestingly, we can also observe from Fig. 7 that the number of antennas affects ρ. For each Δ, a larger N results in a smaller proportion of power transmitted through the relay link. From Figs. 6 and 7, we can see that to achieve a given system MSE, increasing the number of antennas effectively reduces the system transmission power consumption.

In the fourth simulation example, we compare the performance of the two proposed algorithms with $N_1 = 1$, as Algorithm 2 is only applicable to a single data stream system. We set $N_s = N_r = N_d = N$ and $\Delta = 10$dB. Fig. 8 illustrates the system MI versus the total transmission power. It can be seen from Fig. 8 that when N is small ($N = 4$), both algorithms have a very close MI performance.

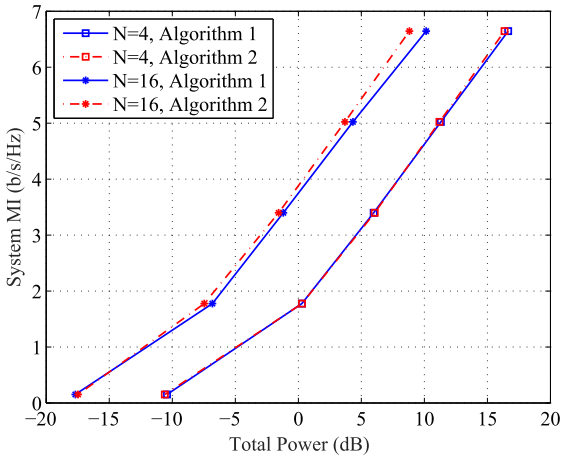


FIGURE 8. Example 4: Comparison between two proposed algorithms. $\Delta = 10$dB, $N_1 = 1$, $N_s = N_r = N_d = N$.

When N is large ($N = 16$), Algorithm 2 has a slightly better MI performance than Algorithm 1. This can be explained by the fact that Algorithm 2 involves less optimization parameters. It can also be observed from Fig. 8 that both algorithms are capable of reducing the transmission power as the number of antennas increases, which agrees with the observation in Figs. 6 and 7.

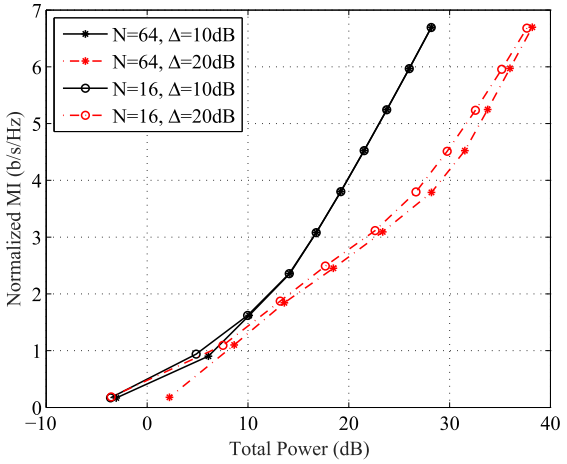


FIGURE 9. Example 5: Normalized MI versus P_t. $N_1 = N_s = N_r = N_d = N$.

In the fifth example, we study the performance of Algorithm 1 with a large number of data streams. Fig. 9 shows the normalized MI (system MI divided by N) versus P_t. We set $N_1 = N_s = N_r = N_d = N$ and chose Δ to be 10dB and 20dB. The MSE constraints are set as logarithmically spaced numbers from $0.01N$ to $0.9N$. It can be seen from Fig. 9 that at $\Delta = 10$dB, the MI per data stream remains almost unchanged as N increases from 16 to 64. For $\Delta = 20$dB, the normalized MI is slightly higher at $N = 16$. Therefore, Algorithm 1 can be applied in relay systems with a large number of antennas such as massive MIMO relay systems.

TABLE 3. Average number of iterations required by two proposed algorithms till convergence.

e	0.01	0.1	0.5
Algorithm 1, $N = 2$	11.8	5.4	4.3
Algorithm 1, $N = 4$	14.8	6.4	6.1
Algorithm 2, $N = 2$	4.5	4.4	4.2
Algorithm 2, $N = 4$	6.5	6.4	5.2

In the last example, we study the convergence behavior of the two proposed algorithms. We choose $\Delta = 10$dB, $N_s = N_r = N_d = N$, and set the convergence criterion $\varepsilon = 10^{-4}$ in Tables 1 and 2. It can be seen from Table 3 that as expected, for both proposed algorithms, the number of iterations increases when e decreases and N increases. We can also see that Algorithm 2 has a faster convergence rate than Algorithm 1.

Finally, we compare the computational complexity of the two proposed algorithms. We set $N_1 = N_s = N_r = N_d = N$ for Algorithm 1 and $N_1 = 1, N_s = N_r = N_d = N$ for Algorithm 2. In one iteration of Algorithm 1, since the bisection method has a low complexity, the computation of Algorithm 1 is mainly on matrix multiplication, inversion, EVD and SVD operations, which have a complexity order of $\mathcal{O}(N^3)$. For one iteration of Algorithm 2, the computational complexity is mainly from the solving the two SDP problems (45)-(47) and (52)-(53) which has a complexity order of $\mathcal{O}(N^7)$ [21]. Thus, Algorithm 2 has a higher per iteration complexity than Algorithm 1. However, Algorithm 2 converges faster (see Table 3) and has a better performance (see Fig. 8) than Algorithm 1 when $N_1 = 1$. Such performance-complexity tradeoff is interesting in practical MIMO relay systems.

V. CONCLUSIONS

We have investigated the transceiver optimization for two-hop AF MIMO relay systems with the direct link and MSE constraints, where the source node transmits signals in both time slots. Two iterative algorithms have been developed to minimize the total system transmission power subjecting to MSE constraints by jointly optimizing the source, relay, and receiver matrices. Algorithm 1 works for the general case of multiple concurrent data streams, while Algorithm 2 is developed for the single data stream case. Simulation results show that both proposed algorithms converge fast and reduce the system transmission power compared with conventional two-hop AF MIMO relay systems where the source node is silent at the second time slot. Under a given system transmission power, the proposed algorithms yield a higher system MI thanks to a better utilization of the direct link. We have found that using both phases for source power allocation provides a larger gain over the conventional single phase source power allocation when the direct link is not very strong or weak. When the direct link is very strong compared with the relay link, direct communication between source

and destination (without any relay nodes) is actually optimal. As future works, transceiver optimization which guarantees QoS for the system in this paper under imperfect CSI will be investigated. Applying the new protocol to AF MIMO relay systems with nonlinear receivers is also an interesting problem for our future research.

APPENDIX A
SOLUTION TO λ IN (20)

By substituting (16) back into (9), we have

$$\begin{aligned}\text{MSE} = tr(&\mathbf{W}_1^H\mathbf{G}\mathbf{F}_1\mathbf{F}_2\mathbf{F}_3(\mathbf{H}\mathbf{B}_1\mathbf{B}_1^H\mathbf{H}^H + \mathbf{I}_{N_1})\mathbf{F}_3^H\mathbf{F}_2^H\mathbf{F}_1^H\\ &\times\mathbf{G}^H\mathbf{W}_1 - \mathbf{F}_2\mathbf{B}_1^H\mathbf{H}^H\mathbf{F}_3^H\mathbf{F}_2^H\mathbf{F}_1^H\mathbf{G}^H\mathbf{W}_1\\ &- \mathbf{W}_1^H\mathbf{G}\mathbf{F}_1\mathbf{F}_2\mathbf{F}_3\mathbf{H}\mathbf{B}_1\mathbf{F}_2^H) + tr(\mathbf{F}_2\mathbf{F}_2^H + \mathbf{W}^H\mathbf{W}).\end{aligned} \tag{54}$$

From (19), we obtain the identity of

$$\mathbf{F}_3(\mathbf{H}\mathbf{B}_1\mathbf{B}_1^H\mathbf{H}^H + \mathbf{I}_{N_1})\mathbf{F}_3^H = \mathbf{F}_3\mathbf{H}\mathbf{B}_1. \tag{55}$$

Denoting

$$\check{\mathbf{F}}_1 = \mathbf{W}_1^H\mathbf{G}\mathbf{F}_1 = \mathbf{F}_1^H\mathbf{G}^H\mathbf{W}_1 \tag{56}$$

$$\check{\mathbf{F}}_3 = \mathbf{F}_3\mathbf{H}\mathbf{B}_1 = \mathbf{B}_1^H\mathbf{H}^H\mathbf{F}_3^H \tag{57}$$

and using (55), we can rewrite (54) as

$$\text{MSE} = tr((\check{\mathbf{F}}_1 - \mathbf{I}_{N_1})\mathbf{F}_2\check{\mathbf{F}}_3\mathbf{F}_2^H(\check{\mathbf{F}}_1 - \mathbf{I}_{N_1})) + \check{e} \tag{58}$$

where $\check{e} = tr(\mathbf{F}_2\mathbf{F}_2^H + \mathbf{W}^H\mathbf{W} - \mathbf{F}_2\check{\mathbf{F}}_3\mathbf{F}_2^H)$. By introducing the SVD of $\mathbf{W}_1^H\mathbf{G} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^H$, from (56) and (17), $\check{\mathbf{F}}_1$ becomes

$$\check{\mathbf{F}}_1 = \lambda\mathbf{U}\Sigma(\mathbf{I}_{N_1} + \lambda\mathbf{\Sigma}^2)^{-1}\mathbf{\Sigma}\mathbf{U}^H. \tag{59}$$

By substituting (59) back into (58), we have

$$\begin{aligned}\text{MSE} = tr((&\lambda\mathbf{\Sigma}(\mathbf{I}_{N_1} + \lambda\mathbf{\Sigma}^2)^{-1}\mathbf{\Sigma} - \mathbf{I}_{N_1})\mathbf{U}^H\mathbf{F}_2\check{\mathbf{F}}_3\mathbf{F}_2^H\mathbf{U}\\ &\times(\lambda\mathbf{\Sigma}(\mathbf{I}_{N_1} + \lambda\mathbf{\Sigma}^2)^{-1}\mathbf{\Sigma} - \mathbf{I}_{N_1})) + \check{e}.\end{aligned} \tag{60}$$

Denoting the ith main diagonal elements of $\mathbf{U}^H\mathbf{F}_2\check{\mathbf{F}}_3\mathbf{F}_2^H\mathbf{U}$ and $\mathbf{\Sigma}$ by ω_i and σ_i, $i = 1, \ldots, N_1$, respectively, MSE $= e$ can be equivalently rewritten from (60) as

$$\sum_{i=1}^{N_1}\frac{\omega_i}{(1+\lambda\sigma_i^2)^2} = e - \check{e}. \tag{61}$$

Since $\mathbf{U}^H\mathbf{F}_2\check{\mathbf{F}}_3\mathbf{F}_2^H\mathbf{U}$ is positive semidefinite, we have $\omega_i \geq 0$, $i = 1, \ldots, N_1$. Thus, the LHS of (61) is monotonically decreasing with respect to $\lambda > 0$. Therefore, we can efficiently solve $\lambda > 0$ from (61) using the bisection method.

APPENDIX B
PROOF OF LEMMA 1

Using the matrix identity of

$$(\mathbf{A}+\mathbf{B}\mathbf{C}\mathbf{D})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{B}(\mathbf{D}\mathbf{A}^{-1}\mathbf{B} + \mathbf{C}^{-1})^{-1}\mathbf{D}\mathbf{A}^{-1} \tag{62}$$

we have

$$\begin{aligned}&\mu\mathbf{L}_1^H(\mathbf{D} + \mu\mathbf{L}_1\mathbf{L}_1^H)^{-1}\mathbf{L}_1\\ &= \mu\mathbf{L}_1^H\mathbf{D}^{-1}\mathbf{L}_1 - \mu\mathbf{L}_1^H\mathbf{D}^{-1}\mathbf{L}_1(\mathbf{L}_1^H\mathbf{D}^{-1}\mathbf{L}_1 + \mu^{-1}\mathbf{I}_{N_1})^{-1}\\ &\quad\times\mathbf{L}_1^H\mathbf{D}^{-1}\mathbf{L}_1\\ &= \mu\mathbf{L}_1^H\mathbf{D}^{-1}\mathbf{L}_1(\mathbf{I}_{N_1} - (\mathbf{L}_1^H\mathbf{D}^{-1}\mathbf{L}_1 + \mu^{-1}\mathbf{I}_{N_1})^{-1}\mathbf{L}_1^H\mathbf{D}^{-1}\mathbf{L}_1)\\ &= \mu\mathbf{L}_1^H\mathbf{D}^{-1}\mathbf{L}_1(\mu\mathbf{L}_1^H\mathbf{D}^{-1}\mathbf{L}_1 + \mathbf{I}_{N_1})^{-1}\\ &= \mathbf{A}_1(\mathbf{A}_1 + \mathbf{I}_{N_1})^{-1}\end{aligned} \tag{63}$$

where $\mathbf{A}_1 = \mu\mathbf{L}_1^H\mathbf{D}^{-1}\mathbf{L}_1$. Similarly, we have

$$\mu\mathbf{L}_2^H(\mathbf{I}_{N_1} + \mu\mathbf{L}_2\mathbf{L}_2^H)^{-1}\mathbf{L}_2 = \mathbf{A}_2(\mathbf{A}_2 + \mathbf{I}_{N_1})^{-1} \tag{64}$$

where $\mathbf{A}_2 = \mu\mathbf{L}_2^H\mathbf{L}_2$.

Left-multiplying both sides of (24) and (25) by $\mathbf{L}_1^H$ and $\mathbf{L}_2^H$, respectively, and using (63) and (64), we have

$$\mathbf{L}_1^H\mathbf{B}_1 = \mathbf{A}_1(\mathbf{I}_{N_1} + \mathbf{A}_1)^{-1}(\mathbf{I}_{N_1} - \mathbf{L}_2^H\mathbf{B}_2) \tag{65}$$

$$\mathbf{L}_2^H\mathbf{B}_2 = \mathbf{A}_2(\mathbf{I}_{N_1} + \mathbf{A}_2)^{-1}(\mathbf{I}_{N_1} - \mathbf{L}_1^H\mathbf{B}_1). \tag{66}$$

By substituting (66) into (65), we obtain

$$\begin{aligned}\mathbf{L}_1^H\mathbf{B}_1 &= \mathbf{A}_1(\mathbf{I}_{N_1} + \mathbf{A}_1)^{-1}(\mathbf{I}_{N_1} - \mathbf{A}_2(\mathbf{I}_{N_1} + \mathbf{A}_2)^{-1}\\ &\quad\times(\mathbf{I}_{N_1} - \mathbf{L}_1^H\mathbf{B}_1))\\ &= \mathbf{A}_1(\mathbf{I}_{N_1} + \mathbf{A}_1)^{-1}(\mathbf{I}_{N_1} + \mathbf{A}_2)^{-1}\\ &\quad+\mathbf{A}_1(\mathbf{I}_{N_1} + \mathbf{A}_1)^{-1}\mathbf{A}_2(\mathbf{I}_{N_1} + \mathbf{A}_2)^{-1}\mathbf{L}_1^H\mathbf{B}_1.\end{aligned} \tag{67}$$

Solving (67) for $\mathbf{L}_1^H\mathbf{B}_1$, we have

$$\begin{aligned}\mathbf{L}_1^H\mathbf{B}_1 = (&\mathbf{I}_{N_1} - \mathbf{A}_1(\mathbf{I}_{N_1} + \mathbf{A}_1)^{-1}\mathbf{A}_2(\mathbf{I}_{N_1} + \mathbf{A}_2)^{-1})^{-1}\\ &\times\mathbf{A}_1(\mathbf{I}_{N_1} + \mathbf{A}_1)^{-1}(\mathbf{I}_{N_1} + \mathbf{A}_2)^{-1}.\end{aligned} \tag{68}$$

Applying the matrix identity (62), we obtain

$$\begin{aligned}&(\mathbf{I}_{N_1} - \mathbf{A}_1(\mathbf{I}_{N_1} + \mathbf{A}_1)^{-1}\mathbf{A}_2(\mathbf{I}_{N_1} + \mathbf{A}_2)^{-1})^{-1}\\ &= ((\mathbf{I}_{N_1} + (\mathbf{I}_{N_1} - \mathbf{A}_1(\mathbf{I}_{N_1} + \mathbf{A}_1)^{-1})\mathbf{A}_2)(\mathbf{I}_{N_1} + \mathbf{A}_2)^{-1})^{-1}\\ &= ((\mathbf{I}_{N_1} + (\mathbf{I}_{N_1} + \mathbf{A}_1)^{-1}\mathbf{A}_2)(\mathbf{I}_{N_1} + \mathbf{A}_2)^{-1})^{-1}\\ &= (\mathbf{I}_{N_1} + \mathbf{A}_2)(\mathbf{I}_{N_1} + (\mathbf{I}_{N_1} + \mathbf{A}_1)^{-1}\mathbf{A}_2)^{-1}\\ &= (\mathbf{I}_{N_1} + \mathbf{A}_2)(\mathbf{I}_{N_1} + \mathbf{A}_1 + \mathbf{A}_2)^{-1}(\mathbf{I}_{N_1} + \mathbf{A}_1).\end{aligned} \tag{69}$$

Substituting (69) back into (68), we have that

$$\begin{aligned}\mathbf{L}_1^H\mathbf{B}_1 &= (\mathbf{I}_{N_1} + \mathbf{A}_2)(\mathbf{I}_{N_1} + \mathbf{A}_1 + \mathbf{A}_2)^{-1}\mathbf{A}_1(\mathbf{I}_{N_1} + \mathbf{A}_2)^{-1}\\ &= (\mathbf{A}_1(\mathbf{I}_{N_1} + \mathbf{A}_2)^{-1} + \mathbf{I}_{N_1})^{-1}\mathbf{A}_1(\mathbf{I}_{N_1} + \mathbf{A}_2)^{-1}\\ &= \mathbf{A}_1(\mathbf{I}_{N_1} + (\mathbf{I}_{N_1} + \mathbf{A}_2)^{-1}\mathbf{A}_1)^{-1}(\mathbf{I}_{N_1} + \mathbf{A}_2)^{-1}\\ &= \mathbf{A}_1(\mathbf{I}_{N_1} + \mathbf{A}_1 + \mathbf{A}_2)^{-1}.\end{aligned} \tag{70}$$

In a similar way, we can obtain

$$\mathbf{L}_2^H\mathbf{B}_2 = \mathbf{A}_2(\mathbf{I}_{N_1} + \mathbf{A}_1 + \mathbf{A}_2)^{-1}. \tag{71}$$

Finally, by substituting (70) and (71) back into (23), we have

$$\begin{aligned}\bar{\mathbf{B}} &= \mathbf{I}_{N_1} - \mathbf{L}_1^H\mathbf{B}_1 - \mathbf{L}_2^H\mathbf{B}_2\\ &= \mathbf{I}_{N_1} - (\mathbf{A}_1 + \mathbf{A}_2)(\mathbf{I}_{N_1} + \mathbf{A}_1 + \mathbf{A}_2)^{-1}\\ &= (\mathbf{I}_{N_1} + \mathbf{A}_1 + \mathbf{A}_2)^{-1}.\end{aligned}$$

REFERENCES

[1] J. N. Laneman, D. N. C. Tse, and G. W. Wornell, "Cooperative diversity in wireless networks: Efficient protocols and outage behavior," *IEEE Trans. Inf. Theory*, vol. 50, no. 12, pp. 3062–3080, Dec. 2004.

[2] *Relay Architectures for E-UTRA (LTE-Advanced)*, document TR 36.806, 3GPP, Mar. 2010.

[3] 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.

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

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

[6] 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.

[7] Z. He, W. Jiang, and Y. Rong, "Robust design for amplify-and-forward MIMO relay systems with direct link and imperfect channel information," *IEEE Trans. Wireless Commun.*, vol. 14, no. 1, pp. 353–363, Jan. 2015.

[8] L. Sanguinetti and A. A. D'Amico, "Power allocation in two-hop amplify-and-forward MIMO relay systems with QoS requirements," *IEEE Trans. Signal Process.*, vol. 60, no. 5, pp. 2494–2507, May 2012.

[9] L. Sanguinetti, A. A. D'Amico, and Y. Rong, "On the design of amplify-and-forward MIMO-OFDM relay systems with QoS requirements specified as Schur-convex functions of the MSEs," *IEEE Trans. Veh. Technol.*, vol. 62, no. 4, pp. 1871–1877, May 2013.

[10] Z. He, Z. Lang, Y. Rong, and S. Qu, "Joint transceiver optimization for two-way MIMO relay systems with MSE constraints," *IEEE Wireless Commun. Lett.*, vol. 3, no. 6, pp. 613–616, Dec. 2014.

[11] Y. Ma and Y. Hua, "Splitting source power for a multicarrier relay system with direct link," in *Proc. Asilomar Conf. Signals, Syst. Comput.*, Pacific Grove, CA, USA, Nov. 2013, pp. 1253–1257.

[12] 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, no. 1, pp. 5–16, Jan. 2014.

[13] D. P. Palomar, M. A. Lagunas, and J. M. Cioffi, "Optimum linear joint transmit-receive processing for MIMO channels with QoS constraints," *IEEE Trans. Signal Process.*, vol. 52, no. 5, pp. 1179–1197, May 2004.

[14] Z. He, J. Zhang, W. Liu, and Y. Rong, "New results on transceiver design for two-hop amplify-and-forward MIMO relay systems with direct link," *IEEE Trans. Signal Process.*, vol. 64, no. 20, pp. 5232–5241, Oct. 2016.

[15] G. D. Forney, Jr., "Shannon meets Wiener II: On MMSE estimation in successive decoding schemes," in *Proc. Allerton Conf. Commun., Control, Comput.*, Sep. 2004, pp. 1–9.

[16] R. Hunger, M. Joham, and W. Utschick, "On the MSE-duality of the broadcast channel and the multiple access channel," *IEEE Trans. Signal Process.*, vol. 57, no. 2, pp. 698–713, Feb. 2009.

[17] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Upper Saddle River, NJ, USA: Prentice-Hall, 1993.

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

[19] A. Charnes and W. W. Cooper, "Programming with linear fractional functionals," *Naval Res. Logistics*, vol. 9, pp. 181–186, Sep./Oct. 1962.

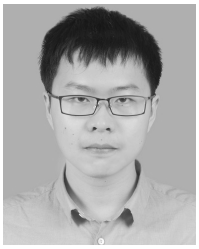
[20] Z.-Q. Luo, W.-K. Ma, A. M.-C. So, Y. Ye, and S. Zhang, "Semidefinite relaxation of quadratic optimization problems," *IEEE Signal Process. Mag.*, vol. 27, no. 3, pp. 20–34, May 2010.

[21] Y. Nesterov and A. Nemirovskii, *Interior-Point Polynomial Algorithms in Convex Programming* (Studies in Applied Mathematics), vol. 13. Philadelphia, PA, USA: SIAM, 1994.

[22] W. Ai, Y. Huang, and S. Zhang, "New results on Hermitian matrix rank-one decomposition," *Math. Program.*, vol. 128, pp. 253–283, Jun. 2009.

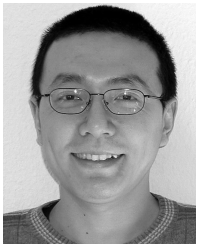
[23] Y. Huang and D. P. Palomar, "Randomized algorithms for optimal solutions of double-sided QCQP with applications in signal processing," *IEEE Trans. Signal Process.*, vol. 62, no. 5, pp. 1093–1108, Mar. 2014.

[24] M. C. Grant and S. P. Boyd. (Oct. 2014). *The CVX Users Guide, Release 2.1*. [Online]. Available: http://web.cvxr.com/cvx/doc/CVX.pdf



JIAOLONG YANG received the B.E. degree in information engineering from the Beijing University of Posts and Telecommunications, Beijing, China, in 2015, where he is currently pursuing the M.Sc. degree in electronics and telecommunication engineering. His research interests are mainly on wireless communications.

ZHIQIANG HE (S'01–M'04) received the B.E. degree and Ph.D. degree (Hons.) in signal and information processing from the Beijing University of Posts and Telecommunications, China, in 1999 and 2004, respectively. Since 2004, he has been with the School of Information and Communication Engineering, Beijing University of Posts and Telecommunications, where he is currently a 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, machine learning, and underwater acoustic 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 at Riverside, Riverside, CA, USA, from 2006 to 2007. Since 2007, he has been with the Department of Electrical and Computer Engineering, Curtin University, Bentley, WA, Australia, where he is currently a Professor. He has authored or co-authored over 150 journal and conference papers in his research areas. His research interests include signal processing for communications, wireless communications, underwater acoustic communications, the applications of linear algebra and optimization methods, and statistical and array signal processing.

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 of the IEEE ICC, WCSP, IWCMC, EUSIPCO, and ChinaCom.