IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received September 21, 2017, accepted November 12, 2017, date of publication November 20, 2017, date of current version December 22, 2017.

Digital Object Identifier 10.1109/ACCESS.2017.2775202

Channel Estimation of MIMO Relay Systems With Multiple Relay Nodes

ZHIQIANG HE[1,2], (Member, IEEE), JIAOLONG YANG[1], XIAODAN WANG[1], YANG LIU[1], AND YUE RONG[3], (Senior Member, IEEE)

[1]Key Laboratory of Universal Wireless Communication, Ministry of Education, Beijing University of Posts and Telecommunications, Beijing 100876, China
[2] Key Laboratory of Underwater Acoustic Communication and Marine Information Technology, Ministry of Education, Xiamen University, Xiamen 361005, China
[3]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 consider the channel estimation problem for multiple-input multiple-output wireless relay communication systems with multiple relay nodes. In particular, all individual channel matrices of the first-hop and second-hop links are estimated at the destination node by applying the superimposed channel training algorithm, where training sequences are superimposed at the relay nodes to assist the estimation of the relays-destination channel matrices. To improve the performance of channel estimation, we consider the estimation error inherited from the second-hop channel estimation and develop a new minimal mean-squared error-based algorithm to estimate the first-hop channel matrices. Furthermore, we derive the optimal power allocation and training sequences at the source and relay nodes. Numerical examples demonstrate a better performance of the proposed superimposed channel training algorithm.

INDEX TERMS Channel estimation, MIMO relay, MMSE, multiple relay nodes, power allocation, superimposed training.

I. INTRODUCTION

Multiple-input multiple-output (MIMO) relay communication systems have attracted much research interest recently due to their capability in improving the reliability and coverage of wireless systems [1], [2]. Many studies have been carried out on the transceiver optimization of amplify-and-forward (AF) MIMO relay systems. The optimal relay precoding matrix which maximizes the source-destination mutual information (MI) of a two-hop MIMO relay system was developed in [3] and [4]. In [5] and [6], the relay precoding matrix was designed to minimize the mean-squared error (MSE) of the signal waveform estimation. A unified framework was developed in [7] and [8] to jointly optimize the source and relay precoding matrices for a broad class of objective functions.

The works in [3]–[8] focused on MIMO relay systems with a single relay node at each hop. Systems with multiple parallel relay nodes have also attracted great interests [9], [10]. Behbahani *et al.* [9] developed the optimal relay precoding matrices with multiple relay nodes. In [10], joint source and relay precoding matrices optimization has been investigated with power constraint at the output of the second-hop channel considering both linear and nonlinear receivers.

It can be seen that for MIMO relay systems discussed in [3]–[10], the knowledge of the instantaneous channel state information (CSI) is required for designing the optimal linear receiver at the destination node and optimizing the system through precoding matrices design. Nevertheless, in real relay communication systems, the instantaneous CSI is unknown at the destination node, and therefore, needs to be estimated. Lioliou and Viberg [11] and Lioliou *et al.* [12] developed a least-squares (LS) based channel estimation algorithm for MIMO relay systems and further improved its performance by the weighted LS (WLS) approach. A two-stage channel training method was developed in [13], where the optimal training sequence at the source and relay nodes was derived. In [14], a parallel factor (PARAFAC) analysis based MIMO relay channel estimation algorithm was proposed. Superimposed channel training algorithms were developed in [15] and [16] for two-way MIMO relay systems, and in [17] for time-varying MIMO relay systems, where a training sequence is superimposed at the relay node to estimate the CSI of first-hop and second-hop channel at the destination nodes.

The channel estimation algorithms in [11]–[17] were developed for MIMO relay systems with a single relay node at each hop [3]–[8]. Obviously, channel estimation becomes more challenging for systems with multiple parallel relay nodes as more unknowns need to be estimated. In [18], LS and maximum likelihood (ML) based channel estimators were derived based on the parameterization of the estimation problem. The ML estimation technique was used to estimate the basis expansion model (BEM) weighting coefficients in [19]. However, the optimization of training sequence, which may further improve the performance of the estimators, was not discussed in [18] and [19].

In this paper, we apply the principle of superimposed channel training to estimate all individual first-hop and second-hop channel matrices of MIMO relay systems with multiple parallel relay nodes. Note that the knowledge of both the source-relay and relay-destination channels is required at the destination node for developing the optimal receiver [9], [10]. In the proposed algorithm, the channel estimation process is completed in two time blocks. At the first time block, the source node broadcasts training sequence to all relay nodes, while at the second time block, each relay node transmits linearly amplified received signals as well as its own training sequence. Then all individual channel matrices of the first-hop and second-hop links can be efficiently estimated at the destination node.

We derive the optimal structure of the training sequences that minimize the sum MSE of channel estimation and optimize the power allocation between the training sequences at the source and relay nodes. We consider the estimation error inherited from the estimation of the second-hop channel matrices and develop a new minimal MSE (MMSE)-based algorithm to estimate the first-hop channel matrices, for which no efficient method was proposed in [15]. Moreover, compared with [15] and [16], the optimization problems in this paper are more challenging to solve, as more channel training sequences need to be optimized. Numerical examples demonstrate a better performance of the proposed superimposed channel training algorithm compared with the conventional two-stage channel estimation approach for MIMO relay systems with multiple relay nodes.

The rest of this paper is organized as follows. The system model of a two-hop MIMO relay communication system with multiple relay nodes is presented in Section II. The superimposed channel training algorithm is developed in Section III, where the optimal training sequences and power allocation at the source and relay nodes are derived. In Section IV, we present a new MMSE-based algorithm to estimate the first-hop channel matrices considering the error of the second-hop channel estimation. Section V shows numerical simulation results to demonstrate the performance of the proposed algorithms. Finally, conclusions are drawn in Section VI.

The following notations are used throughout the paper. Vectors and matrices are represented by lower case and upper case bold letters, respectively; Superscripts $(\cdot)^T$, $(\cdot)^H$, and $(\cdot)^{-1}$ denote transpose, Hermitian transpose, and matrix inversion, respectively; $\otimes$ denotes the matrix Kronecker product [20]; $tr(\cdot)$ stands for the matrix trace; $vec(\cdot)$ denotes the vectorization operator which stacks all column vectors of a matrix on top of each other; $\mathbf{I}_n$ denotes an $n \times n$ identity matrix; $\mathbf{1}$ is a vector with all one entries; bdiag$[\cdot]$ and diag$(\cdot)$ stand for a block diagonal and diagonal matrix, respectively; and E$[\cdot]$ denotes the statistical expectation.

II. SYSTEM MODEL

We consider a two-hop MIMO relay communication system which consists of one source node, K parallel relay nodes, and one destination node as illustrated in Fig. 1. The source and the destination nodes have N_s and N_d antennas, respectively, and the ith relay node has N_i antennas. For $i = 1, \ldots, K$, $\mathbf{H}_i$ denotes the $N_i \times N_s$ first-hop channel matrix from the source node to the ith relay node and $\mathbf{G}_i$ is the $N_d \times N_i$ second-hop channel matrix from the ith relay node to the destination node. Due to its merit of simplicity, we consider the AF relay scheme at each relay.

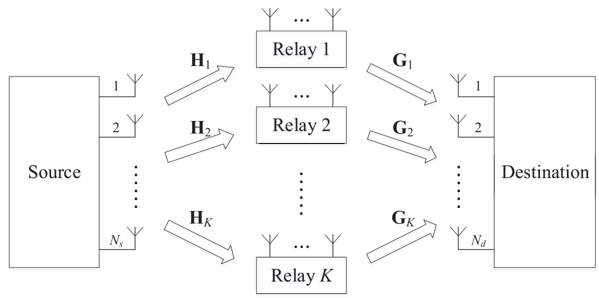


FIGURE 1. Block diagram of a two-hop MIMO relay communication system with multiple parallel relay nodes.

The channel training process is implemented in two time blocks. At the first time block, the source node sends an $N_s \times T$ channel training matrix $\mathbf{S}$ to all relay nodes, where T is the length of the training sequence. The signals received at the ith relay node can be written as

$$\mathbf{Y}_i = \mathbf{H}_i\mathbf{S} + \mathbf{V}_i, \quad i = 1, \ldots, K \tag{1}$$

where $\mathbf{Y}_i$ and $\mathbf{V}_i$ are the $N_i \times T$ received signal matrix and noise matrix at the ith relay node, respectively.

At the second time block, the source node is silent, and the ith relay node amplifies its received signal with an amplifying factor $\alpha > 0$ and superimposes its own training matrix $\mathbf{T}_i$. Therefore, the signal matrix transmitted by the ith relay node is given by

$$\mathbf{X}_i = \sqrt{\alpha}\mathbf{Y}_i + \mathbf{T}_i, \quad i = 1, \ldots, K. \tag{2}$$

The received signal at the destination node can be written as

$$\mathbf{Y}_d = \sum_{i=1}^{K} \mathbf{G}_i\mathbf{X}_i + \mathbf{V}_d \tag{3}$$

where $\mathbf{V}_d$ is the $N_d \times T$ additive Gaussian noise matrix at the destination node.

We assume that the channel matrices $\mathbf{H}_i$ and $\mathbf{G}_i$ satisfy the well-known Gaussian-Kronecker model [21], where $\mathbf{H}_i$ and $\mathbf{G}_i$ are complex Gaussian random matrices with

$$\mathbf{H}_i \sim \mathcal{CN}(\mathbf{0}, \mathbf{T}_s \otimes \mathbf{R}_i), \quad \mathbf{G}_i \sim \mathcal{CN}(\mathbf{0}, \mathbf{C}_i \otimes \mathbf{R}_d) \tag{4}$$

where $\mathbf{T}_s$ and $\mathbf{R}_i$ denote the $N_s \times N_s$ and $N_i \times N_i$ covariance matrix at the transmit and receive side of $\mathbf{H}_i$, respectively, while $\mathbf{C}_i$ and $\mathbf{R}_d$ denote the $N_i \times N_i$ and $N_d \times N_d$ covariance matrix at the transmit and receive side of $\mathbf{G}_i$, respectively. In other words, from (4) we have

$$\mathbf{H}_i = \mathbf{A}_i\mathbf{H}_{i,w}\mathbf{B}_s^H, \quad \mathbf{G}_i = \mathbf{A}_d\mathbf{G}_{i,w}\mathbf{K}_i^H \tag{5}$$

where $\mathbf{A}_i\mathbf{A}_i^H = \mathbf{R}_i, \mathbf{B}_s\mathbf{B}_s^H = \mathbf{T}_s^T, \mathbf{A}_d\mathbf{A}_d^H = \mathbf{R}_d, \mathbf{K}_i\mathbf{K}_i^H = \mathbf{C}_i^T$, $i = 1, \ldots, K$, $\mathbf{H}_{i,w}$ and $\mathbf{G}_{i,w}$ are $N_i \times N_s$ and $N_d \times N_i$ complex Gaussian random matrices with independent and identically distributed (i.i.d.) zero mean and unit variance entries. We assume that $\mathbf{H}_{i,w}$ and $\mathbf{G}_{i,w}$, $i = 1, \ldots, K$, are statistically independent of each other. We also assume that the knowledge on the channel covariance matrices is available at the destination node.

By substituting (1) and (2) into (3), we can rewrite the received signal at the destination node as

$$\begin{aligned}\mathbf{Y}_d &= \sum_{i=1}^{K}\mathbf{G}_i[\sqrt{\alpha}(\mathbf{H}_i\mathbf{S} + \mathbf{V}_i) + \mathbf{T}_i] + \mathbf{V}_d \\ &= \sqrt{\alpha}\mathbf{M}\mathbf{S} + \sum_{i=1}^{K}\mathbf{G}_i\mathbf{T}_i + \mathbf{V}\end{aligned} \tag{6}$$

where $\mathbf{V} = \sum_{i=1}^{K}\sqrt{\alpha}\mathbf{G}_i\mathbf{V}_i + \mathbf{V}_d$ is the total noise matrix at the destination node and $\mathbf{M} = \sum_{i=1}^{K}\mathbf{G}_i\mathbf{H}_i$ can be viewed as the compound source-destination channel matrix. It can be seen from (6) that the first-hop channel $\mathbf{H}_i$ is multiplied with the second-hop channel $\mathbf{G}_i$, which makes it difficult to estimate both $\mathbf{H}_i$ and $\mathbf{G}_i$ in one step. Thus, we propose an efficient two-step channel estimation algorithm to estimate the second-hop channel $\mathbf{G}_i$ first based on $\mathbf{T}_i$, and then estimate the first-hop channels.

III. SUPERIMPOSED CHANNEL TRAINING FOR THE SECOND-HOP CHANNEL ESTIMATION

In this section, the superimposed channel training principle is applied to estimate the channel matrices $\mathbf{M}$ and $\mathbf{G}_i$, $i = 1, \ldots, K$. Moreover, we design the optimal training matrices $\mathbf{S}$ and $\mathbf{T}_i$, $i = 1, \ldots, K$, and the relay amplifying factor α to minimize the sum MSE of the channel estimation.

The main idea of the superimposed channel training algorithm is to use $\mathbf{T}_i$ to estimate the second-hop channel matrices $\mathbf{G}_i$, $i = 1, \ldots, K$. By introducing the eigenvalue decompositions (EVDs) of

$$\mathbf{T}_s^T = \mathbf{U}_s\mathbf{\Lambda}_s\mathbf{U}_s^H, \quad \mathbf{C}_i^T = \mathbf{U}_i\mathbf{\Lambda}_i\mathbf{U}_i^H, \quad i = 1, \ldots, K$$

we have

$$\mathbf{B}_s^H = \mathbf{\Pi}_s\mathbf{\Lambda}_s^{\frac{1}{2}}\mathbf{U}_s^H, \quad \mathbf{K}_i^H = \mathbf{\Pi}_i\mathbf{\Lambda}_i^{\frac{1}{2}}\mathbf{U}_i^H, \quad i = 1, \ldots, K \tag{7}$$

where $\mathbf{\Pi}_s$ and $\mathbf{\Pi}_i$ are arbitrary $N_s \times N_s$ and $N_i \times N_i$ unitary matrix, respectively. Based on (5), the received signal at the destination node (6) can be rewritten as

$$\mathbf{Y}_d = \sqrt{\alpha}\tilde{\mathbf{M}}\tilde{\mathbf{S}} + \sum_{i=1}^{K}\tilde{\mathbf{G}}_i\tilde{\mathbf{T}}_i + \mathbf{V} \tag{8}$$

where

$$\begin{aligned}\tilde{\mathbf{M}} &\triangleq \sum_{i=1}^{K}\mathbf{G}_i\tilde{\mathbf{H}}_i \\ \tilde{\mathbf{S}} &\triangleq \mathbf{U}_s^H\mathbf{S}, \quad \tilde{\mathbf{H}}_i \triangleq \mathbf{H}_i\mathbf{U}_s, \quad i = 1, \ldots, K \\ \tilde{\mathbf{G}}_i &\triangleq \mathbf{G}_i\mathbf{U}_i, \quad \tilde{\mathbf{T}}_i \triangleq \mathbf{U}_i^H\mathbf{T}_i, \quad i = 1, \ldots, K.\end{aligned} \tag{9}$$

In the following, we develop an algorithm to estimate $\tilde{\mathbf{M}}$ and $\tilde{\mathbf{G}}_i$ in (8). In Section IV, we will present the algorithm to estimate the first-hop channels $\tilde{\mathbf{H}}_i$ from (8). It will be shown in Section IV that the MSE of estimating $\tilde{\mathbf{H}}_i$ increases with the MSE of estimating $\tilde{\mathbf{M}}$ and $\tilde{\mathbf{G}}_i$. Thus, it is important to optimize the training matrices $\mathbf{S}$ and $\mathbf{T}_i$, $i = 1, \ldots, K$, to minimize the MSE of estimating $\tilde{\mathbf{H}}_i$.

Applying the identity of $vec(\mathbf{ABC}) = (\mathbf{C}^T \otimes \mathbf{A})vec(\mathbf{B})$ [20], we can rewrite (8) in vector form as

$$\begin{aligned}\mathbf{y}_d &= \left[\sqrt{\alpha}\tilde{\mathbf{S}}^T \otimes \mathbf{I}_{N_d}, \tilde{\mathbf{T}}_1^T \otimes \mathbf{I}_{N_d}, \ldots, \tilde{\mathbf{T}}_K^T \otimes \mathbf{I}_{N_d}\right] \\ &\quad \times \left[\tilde{\mathbf{m}}^T, \tilde{\mathbf{g}}_1^T, \ldots, \tilde{\mathbf{g}}_K^T\right]^T + \mathbf{v} \\ &= \mathbf{L}\boldsymbol{\gamma} + \mathbf{v}\end{aligned} \tag{10}$$

where

$$\begin{aligned}\mathbf{L} &\triangleq \left[\sqrt{\alpha}\tilde{\mathbf{S}}^T \otimes \mathbf{I}_{N_d}, \tilde{\mathbf{T}}_1^T \otimes \mathbf{I}_{N_d}, \ldots, \tilde{\mathbf{T}}_K^T \otimes \mathbf{I}_{N_d}\right] \\ \boldsymbol{\gamma} &\triangleq \left[\tilde{\mathbf{m}}^T, \tilde{\mathbf{g}}_1^T, \ldots, \tilde{\mathbf{g}}_K^T\right]^T \\ \mathbf{y}_d &\triangleq vec(\mathbf{Y}_d), \quad \tilde{\mathbf{m}} \triangleq vec(\tilde{\mathbf{M}}) \\ \mathbf{v} &\triangleq vec(\mathbf{V}), \quad \tilde{\mathbf{g}}_i \triangleq vec(\tilde{\mathbf{G}}_i), \quad i = 1, \ldots, K.\end{aligned}$$

In (10), $\boldsymbol{\gamma}$ is the vector of unknown with a dimension of $Q \triangleq N_d(N_s + \sum_{i=1}^{K}N_i)$, $\mathbf{L}$ has a dimension of $TN_d \times Q$, and there is $T \geq N_s + \sum_{i=1}^{K}N_i$.

Thanks to its computational simplicity, a linear estimator is applied at the destination node to estimate $\boldsymbol{\gamma}$ as

$$\hat{\boldsymbol{\gamma}} = \mathbf{W}^H\mathbf{y}_d \tag{11}$$

where $\hat{\boldsymbol{\gamma}}$ denotes the estimation of $\boldsymbol{\gamma}$ and $\mathbf{W}$ is the weight matrix of the linear estimator. Using (11), the MSE of channel estimation is given by

$$\begin{aligned}\text{MSE}_1 &= \text{E}[tr((\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma})(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma})^H)] \\ &= tr((\mathbf{W}^H\mathbf{L} - \mathbf{I}_Q)\mathbf{R}_{\boldsymbol{\gamma}}(\mathbf{W}^H\mathbf{L} - \mathbf{I}_Q)^H + \mathbf{W}^H\mathbf{R}_{\mathbf{v}}\mathbf{W})\end{aligned} \tag{12}$$

where $\mathbf{R}_{\boldsymbol{\gamma}} = \text{E}[\boldsymbol{\gamma}\boldsymbol{\gamma}^H]$ and $\mathbf{R}_{\mathbf{v}} = \text{E}[\mathbf{v}\mathbf{v}^H]$ are the covariance matrix of $\boldsymbol{\gamma}$ and $\mathbf{v}$, respectively.

Lemma 1: $\mathbf{R}_\gamma$ and $\mathbf{R}_v$ are given by

$$\mathbf{R}_v = \mathbf{I}_T \otimes \left(\sum_{i=1}^{K} \alpha\, tr(\mathbf{C}_i^T)\mathbf{R}_d + \mathbf{I}_{N_d}\right) \quad (13)$$

$$\mathbf{R}_\gamma = \text{bdiag}\,[\mathbf{\Lambda}_s \otimes b\mathbf{R}_d, \mathbf{\Lambda}_1 \otimes \mathbf{R}_d, \ldots, \mathbf{\Lambda}_K \otimes \mathbf{R}_d] \quad (14)$$

where $b \triangleq \sum_{i=1}^{K} tr(\mathbf{R}_i\mathbf{C}_i^T)$.

Proof: See Appendix A. □

It is well-known that (12) is minimized by the linear MMSE estimator [22] given by

$$\mathbf{W} = \left(\mathbf{L}\mathbf{R}_\gamma\mathbf{L}^H + \mathbf{R}_v\right)^{-1}\mathbf{L}\mathbf{R}_\gamma. \quad (15)$$

By substituting (15) back into (12) and applying the matrix inversion lemma of

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

we can obtain the MSE of estimating $\boldsymbol{\gamma}$ as

$$\text{MSE}_1 = tr\left((\mathbf{R}_\gamma^{-1} + \mathbf{L}^H\mathbf{R}_v^{-1}\mathbf{L})^{-1}\right). \quad (16)$$

The transmission power consumed by the source node is given by

$$tr(\mathbf{S}\mathbf{S}^H) = tr(\tilde{\mathbf{S}}\tilde{\mathbf{S}}^H). \quad (17)$$

Using (2) and Lemma 3 in Appendix A, the transmission power of the ith relay node can be calculated as

$$\alpha \text{E}[tr(\mathbf{H}_i\mathbf{S}\mathbf{S}^H\mathbf{H}_i^H + \mathbf{I}_{N_i})] + tr(\mathbf{T}_i\mathbf{T}_i^H) = \alpha N_i + \alpha\, tr(\mathbf{\Lambda}_s\tilde{\mathbf{S}}\tilde{\mathbf{S}}^H)\, tr(\mathbf{R}_i) + tr(\tilde{\mathbf{T}}_i\tilde{\mathbf{T}}_i^H). \quad (18)$$

From (16)-(18), the optimal amplifying factor α and the optimal training matrices can be designed through solving the following optimization problem

$$\min_{\alpha,\tilde{\mathbf{S}},\tilde{\mathbf{T}}_1,\ldots,\tilde{\mathbf{T}}_K} tr\left((\mathbf{R}_\gamma^{-1} + \mathbf{L}^H\mathbf{R}_v^{-1}\mathbf{L})^{-1}\right) \quad (19)$$

$$\text{s.t. } tr(\tilde{\mathbf{S}}\tilde{\mathbf{S}}^H) \le p_s, \quad \alpha > 0 \quad (20)$$

$$\alpha N_i + \alpha\, tr(\mathbf{\Lambda}_s\tilde{\mathbf{S}}\tilde{\mathbf{S}}^H)\, tr(\mathbf{R}_i) + tr(\tilde{\mathbf{T}}_i\tilde{\mathbf{T}}_i^H) \le p_i, \quad i = 1,\ldots,K \quad (21)$$

where p_s is the transmission power available at the source node and p_i is the transmission power at the ith relay node.

Using [15, Th. 1], the optimal $\tilde{\mathbf{S}}$ and $\tilde{\mathbf{T}}_i$ as the solution to the problem (19)–(21) satisfy the equations below for $i, j = 1, \ldots, K$

$$\tilde{\mathbf{S}}\tilde{\mathbf{S}}^H = \mathbf{\Sigma}_s, \quad \tilde{\mathbf{T}}_i\tilde{\mathbf{T}}_i^H = \mathbf{\Sigma}_i \quad (22)$$

$$\tilde{\mathbf{S}}\tilde{\mathbf{T}}_i^H = \mathbf{0}, \quad \tilde{\mathbf{T}}_i\tilde{\mathbf{T}}_j^H = \mathbf{0}, \quad i \neq j \quad (23)$$

where $\mathbf{\Sigma}_s$ and $\mathbf{\Sigma}_i$ are $N_s \times N_s$ and $N_i \times N_i$ diagonal matrix, respectively, with non-negative diagonal elements. Based on (22) and (23), the optimal structure of the training sequences is given by

$$\mathbf{S} = \mathbf{U}_s\mathbf{\Sigma}_s^{\frac{1}{2}}\mathbf{\Phi}_s, \quad \mathbf{T}_i = \mathbf{U}_i\mathbf{\Sigma}_i^{\frac{1}{2}}\mathbf{\Phi}_i, \quad i = 1,\ldots,K \quad (24)$$

where $\mathbf{\Phi}_i$, $i = s, 1, \ldots, K$, is an $N_i \times T$ semi-unitary matrix satisfying the following equations

$$\mathbf{\Phi}_i\mathbf{\Phi}_i^H = \mathbf{I}_{N_i}, \quad \mathbf{\Phi}_i\mathbf{\Phi}_j^H = \mathbf{0}, i, j = s, 1, \ldots, K, i \neq j. \quad (25)$$

We would like to note that semi-unitary $\mathbf{\Phi}_i$ satisfying (25) can be easily obtained, for instance, from the normalized discrete Fourier transform (DFT) matrices.

Based on (22) and (23), it can be seen that the MSE matrix $(\mathbf{R}_\gamma^{-1} + \mathbf{L}^H\mathbf{R}_v^{-1}\mathbf{L})^{-1}$ in (16) becomes block diagonal under the optimal $\tilde{\mathbf{S}}$ and $\tilde{\mathbf{T}}_i$, indicating that the estimation errors of $\tilde{\mathbf{m}}^T$, $\tilde{\mathbf{g}}_i^T$, $i = 1, \ldots, K$, are uncorrelated. By introducing the EVD of $\mathbf{R}_d = \mathbf{U}_d\mathbf{\Lambda}_d\mathbf{U}_d^H$, the problem (19)-(21) can be rewritten as

$$\min_{\alpha,\{\mathbf{\Sigma}_i\}} tr\left((\mathbf{D}_{ds} + \alpha\mathbf{\Sigma}_s \otimes \mathbf{D}_v)^{-1} + \sum_{i=1}^{K}(\mathbf{D}_{di} + \mathbf{\Sigma}_i \otimes \mathbf{D}_v)^{-1}\right) \quad (26)$$

$$\text{s.t. } tr(\mathbf{\Sigma}_s) \le p_s \quad (27)$$

$$\alpha > 0, \mathbf{\Sigma}_i \ge 0, \quad i = s, 1, \ldots, K \quad (28)$$

$$\alpha N_i + \alpha tr(\mathbf{\Lambda}_s\mathbf{\Sigma}_s)tr(\mathbf{R}_i) \quad (29)$$

$$+\, tr(\mathbf{\Sigma}_i) \le p_i, \quad i = 1, \ldots, K \quad (30)$$

where $\{\mathbf{\Sigma}_i\} = \{\mathbf{\Sigma}_i, i = s, 1, \ldots, K\}$ and

$$\mathbf{D}_{ds} = \mathbf{\Lambda}_s^{-1} \otimes (b\mathbf{\Lambda}_d)^{-1} \quad (31)$$

$$\mathbf{D}_{di} = \mathbf{\Lambda}_i^{-1} \otimes \mathbf{\Lambda}_d^{-1}, \quad i = 1, \ldots, K \quad (32)$$

$$\mathbf{D}_v = \left(\sum_{i=1}^{K} \alpha tr(\mathbf{C}_i^T)\mathbf{\Lambda}_d + \mathbf{I}_{N_d}\right)^{-1}. \quad (33)$$

Using (31)-(33), the problem (26)-(29) can be equivalently converted to the following optimization problem with scalar variables

$$\min_{\alpha,\{\sigma_{i,m}\}} \sum_{m=1}^{N_s}\sum_{n=1}^{N_d}\left(\frac{1}{b\lambda_{s,m}\lambda_{d,n}} + \alpha\sigma_{s,m}d_{v,n}\right)^{-1} + \sum_{i=1}^{K}\sum_{m=1}^{N_i}\sum_{n=1}^{N_d}\left(\frac{1}{\lambda_{i,m}\lambda_{d,n}} + \sigma_{i,m}d_{v,n}\right)^{-1} \quad (34)$$

$$\text{s.t. } \sum_{m=1}^{N_s}\sigma_{s,m} \le p_s \quad (35)$$

$$\alpha\left(\sum_{m=1}^{N_s}\lambda_{s,m}\sigma_{s,m}tr(\mathbf{R}_i) + N_i\right) + \sum_{m=1}^{N_i}\sigma_{i,m} \le p_i, \quad i = 1, \ldots, K \quad (36)$$

$$\alpha > 0, \quad \sigma_{i,m} \ge 0, \ m = 1, \ldots, N_i, \ i = s, 1, \ldots, K \quad (37)$$

where $\{\sigma_{i,m}\} = \{\sigma_{i,m}, m = 1, \ldots, N_i, i = s, 1, \ldots, K\}$, $\lambda_{d,n}$ is the nth diagonal element of $\mathbf{\Lambda}_d$, $\lambda_{i,m}, \sigma_{i,m}, i = s, 1, \ldots, K$, are the mth diagonal element of $\mathbf{\Lambda}_i$ and $\mathbf{\Sigma}_i$, respectively, and

$$d_{v,n} \triangleq \left(\sum_{i=1}^{K}\alpha tr(\mathbf{C}_i^T)\lambda_{d,n} + 1\right)^{-1}. \quad (38)$$

For a given α, the problem (34)–(37) is a convex optimization problem with respect to $\{\sigma_{i,m}\}$. This is because when α is fixed, b, $\lambda_{s,m}$, $\lambda_{d,n}$, $d_{v,n}$ and $\lambda_{i,m}$ are known variables, and (34) is monotonically decreasing and convex with respect to $\{\sigma_{i,m}\}$. Moreover, the constraints in (35) and (36) are linear inequality constraints when α is fixed. Thus, with fixed α, the problem (34)–(37) with respect to $\{\sigma_{i,m}\}$ is a convex optimization problem, which can be efficiently solved by the Lagrange multiplier method.

The Karush-Kuhn-Tucker (KKT) optimality conditions [23] of the problem (34)–(37) are given below. Firstly, the gradient conditions are given by

$$\sum_{n=1}^{N_d} \frac{\alpha d_{v,n}}{\left((b\lambda_{s,m}\lambda_{d,n})^{-1} + \alpha\sigma_{s,m}d_{v,n}\right)^2} = \mu_s + \sum_{i=1}^{K} \mu_i e_{i,m} \quad (39)$$
$$m = 1, \ldots, N_s$$

$$\sum_{n=1}^{N_d} \frac{d_{v,n}}{\left((\lambda_{i,m}\lambda_{d,n})^{-1} + \sigma_{i,m}d_{v,n}\right)^2} = \mu_i$$
$$m = 1, \ldots, N_i, \qquad i = 1, \ldots, K \quad (40)$$

where $e_{i,m} = \alpha\lambda_{s,m}tr(\mathbf{R}_i)$, and $\mu_i \geq 0$, $i = s, 1, \ldots, K$, are the Lagrange multipliers. Secondly, the complementary slackness conditions are given by

$$\mu_s\left(p_s - \sum_{m=1}^{N_s}\sigma_{s,m}\right) = 0$$

$$\mu_i\left(p_i - \alpha\left(\sum_{m=1}^{N_s}\lambda_{s,m}\sigma_{s,m}tr(\mathbf{R}_i) + N_i\right) - \sum_{m=1}^{N_i}\sigma_{i,m}\right) = 0$$
$$i = 1, \ldots, K.$$

When α and μ_i, $i = s, 1, \ldots, K$, are fixed, the non-negative $\{\sigma_{s,m}\}$ can be obtained through the bi-section search, as the left-hand side (LHS) of (39) and (40) are monotonically decreasing functions of $\sigma_{s,m}$ and $\sigma_{i,m}$, respectively. To obtain μ_s and μ_i, we can apply an outer bi-section search loop since the LHS of (35) is a decreasing function of $\sigma_{s,m}$, and the LHS of (36) is an increasing function of $\sigma_{s,m}$ and $\sigma_{i,m}$. Meanwhile, in (39), $\sigma_{s,m}$ monotonically decreases with respect to μ_s and μ_i, and $\sigma_{i,m}$ is a monotonically decreasing function of μ_i in (40).

When α also becomes a variable in the optimization, the problem (34)-(37) is not a convex optimization problem. But in this case, the following lemma is in order.

Lemma 2: The object function (34) subjecting to constraints (35)-(37) is a unimodal function with respect to α.

Proof: See Appendix B. □

For a unimodal function, its minimal value can be efficiently obtained by using the golden section search (GSS) [24]. The procedure of applying the GSS technique to find the optimal α is described in Table I, where $\phi > 0$ is the reduction factor, $|\cdot|$ denotes the absolute value, and ε is a positive constant close to 0. It is shown in [24] that the optimal ϕ is 1.618, also known as the golden ratio.

It can be observed from Table I that at each iteration, the GSS method reduces the interval containing the optimal α to 0.618 times of interval at the preceding iteration. Thus, the length of the interval containing the optimal α after the nth iteration is $\Gamma_n = 0.618^n\Gamma_0$ [24], where $\Gamma_0 = \alpha_u - \alpha_l$ is the length of the initial feasible interval . Therefore, the number of iterations required to achieve the desired accuracy ε in Table 1 is given by $N = \lceil\frac{\ln\varepsilon - \ln\Gamma_0}{\ln 0.618}\rceil$, where $\lceil\cdot\rceil$ denotes the ceiling operator. The value of ε can be determined by considering the complexity-performance tradeoff. In the simulations, we choose $\varepsilon = 10^{-3}$ and $[a_l, a_u] = [0, 1]$. Thus, $\Gamma_0 = 1$ and $N = 15$. The unimodality of (34) with respect to α subjecting to (35)-(37) within bounds $[a_l, a_u] = [0, 1]$ is illustrated in Fig. 7 in Appendix B.

TABLE 1. Algorithm I: Procedure of finding the optimal α in the Problem (34)–(37).

1) Choose a feasible bound $[a_l, a_u]$ on α.
2) Define $c_1 = (\phi - 1)\, a_l + (2 - \phi)\, a_u$
   $c_2 = (2 - \phi)\, a_l + (\phi - 1)\, a_u$.
3) Solve the problem (34)-(37) for $\alpha = c_1$;
   Calculate the MSE value defined in (34), $f_{\mathrm{MSE}}(c_1)$ for $\alpha = c_1$.
4) Repeat Step 3 for $\alpha = c_2$.
5) If $f_{\mathrm{MSE}}(c_1) < f_{MSE}(c_2)$, then assign $a_u = c_2$.
   Otherwise, assign $a_l = c_1$.
6) If $|a_u - a_l| \leq \varepsilon$, then end.
   Otherwise, go to step 2.

IV. MMSE-BASED FIRST-HOP CHANNEL ESTIMATION

In this section, we propose an MMSE-based algorithm to estimate the first-hop channel matrices considering the error of the second-hop channel estimation in Section III.

After estimating the second-hop channels $\tilde{\mathbf{G}}_i$, $i = 1, \ldots, K$, $\hat{\tilde{\mathbf{G}}}_i\tilde{\mathbf{T}}_i$ can be subtracted from (8) as

$$\begin{aligned}\bar{\mathbf{Y}}_d &= \mathbf{Y}_d - \sum_{i=1}^{K}\hat{\tilde{\mathbf{G}}}_i\tilde{\mathbf{T}}_i \\ &= \sqrt{\alpha}\mathbf{G}\tilde{\mathbf{H}}\tilde{\mathbf{S}} + \sum_{i=1}^{K}(\tilde{\mathbf{G}}_i - \hat{\tilde{\mathbf{G}}}_i)\tilde{\mathbf{T}}_i + \mathbf{V} \\ &= \sqrt{\alpha}\hat{\mathbf{G}}\tilde{\mathbf{H}}\tilde{\mathbf{S}} + \bar{\mathbf{V}} \end{aligned} \quad (41)$$

where $\tilde{\mathbf{H}} = [\tilde{\mathbf{H}}_1^T, \ldots, \tilde{\mathbf{H}}_K^T]^T$ and

$$\bar{\mathbf{V}} = \sqrt{\alpha}(\mathbf{G} - \hat{\mathbf{G}})\tilde{\mathbf{H}}\tilde{\mathbf{S}} + \sum_{i=1}^{K}(\tilde{\mathbf{G}}_i - \hat{\tilde{\mathbf{G}}}_i)\tilde{\mathbf{T}}_i + \mathbf{V} \quad (42)$$

is the total noise matrix including the channel estimation errors. By applying the vectorization operation to both sides of (41), we have

$$\bar{\mathbf{y}}_d = (\sqrt{\alpha}\tilde{\mathbf{S}}^T \otimes \hat{\mathbf{G}})\tilde{\mathbf{h}} + \bar{\mathbf{v}} = \mathbf{L}_2\tilde{\mathbf{h}} + \bar{\mathbf{v}} \quad (43)$$

where $\bar{\mathbf{y}}_d = vec(\bar{\mathbf{Y}}_d)$, $\tilde{\mathbf{h}} = vec(\tilde{\mathbf{H}})$, $\bar{\mathbf{v}} = vec(\bar{\mathbf{V}})$, and $\mathbf{L}_2 = \sqrt{\alpha}\tilde{\mathbf{S}}^T \otimes \hat{\mathbf{G}}$.

From (43), an MMSE estimate of $\tilde{\mathbf{h}}$ can be obtained by

$$\hat{\tilde{\mathbf{h}}} = \mathbf{W}_2^H\bar{\mathbf{y}}_d \quad (44)$$

where $\mathbf{W}_2 = (\mathbf{L}_2\mathbf{R}_{\tilde{\mathbf{h}}}\mathbf{L}_2^H + \mathbf{R}_{\bar{\mathbf{v}}})^{-1}\mathbf{L}_2\mathbf{R}_{\tilde{\mathbf{h}}}$, $\mathbf{R}_{\tilde{\mathbf{h}}} = \mathrm{E}[\tilde{\mathbf{h}}\tilde{\mathbf{h}}^H]$, and $\mathbf{R}_{\bar{\mathbf{v}}} = \mathrm{E}[\bar{\mathbf{v}}\bar{\mathbf{v}}^H]$. From (44), the MSE of estimating $\tilde{\mathbf{h}}$ is given

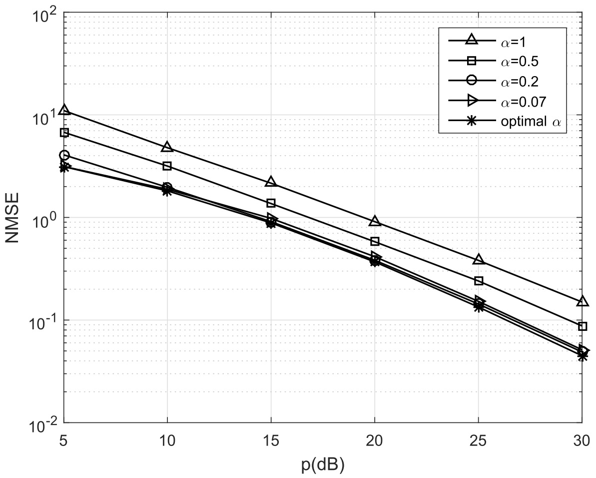


FIGURE 2. Example 1: NMSE versus p at various α with $N_d = N = 4$ and $\rho = 0.8$.

by

$$\text{MSE}_3 = tr\left(\left(\mathbf{R}_{\tilde{\mathbf{h}}}^{-1} + \mathbf{L}_2^H \mathbf{R}_{\tilde{\mathbf{v}}}^{-1} \mathbf{L}_2\right)^{-1}\right) \tag{45}$$

where the derivation of $\mathbf{R}_{\tilde{\mathbf{h}}}$ and $\mathbf{R}_{\tilde{\mathbf{v}}}$ is shown in Appendix C.

Note that in (45), we take into account the equivalent estimation error inherited from the second-hop channel estimation, i.e., $\mathbf{R}_{\tilde{\mathbf{v}}}$. Therefore, the total MSE of the first-hop channel matrices estimation depends on the accuracy of the second-hop channel matrices estimation. In other words, the total MSE in (45) increases with the increasing of the MSE of the second-hop channel estimation.

V. NUMERICAL EXAMPLES

In this section, we study the performance of the proposed superimposed channel training algorithm for MIMO relay systems with multiple relay nodes through numerical simulations. We simulate a MIMO relay system with $K = 2$ relay nodes which have the same number of antennas as the source node, i.e., $N_s = N_i = N$, $i = 1, 2$, and we choose the smallest T possible, i.e., $T = N_s + \sum_{i=1}^{K} N_i = 3N$.

The optimal training matrices for the superimposed channel training method are generated according to (24). In particular, the semi-unitary matrices in (25) are generated based on the normalized DFT matrices as $[\mathbf{\Phi}_s]_{m,n} = \frac{1}{\sqrt{3N}} e^{-j\frac{2\pi mn}{3N}}$, $[\mathbf{\Phi}_1]_{m,n} = \frac{1}{\sqrt{3N}} e^{-j\frac{2\pi(m+N)n}{3N}}$, $[\mathbf{\Phi}_2]_{m,n} = \frac{1}{\sqrt{3N}} e^{-j\frac{2\pi(m+2N)n}{3N}}$, $m = 1, \ldots, N$, $n = 1, \ldots, 3N$. The channel covariance matrices have the widely used exponential Toeplitz structure [21] such that $\mathbf{T}_s = \mathbf{R}_i = \mathbf{C}_i = \rho^{|m-n|}$, $i = 1, 2$, $m, n = 1, \ldots, N$, and $\mathbf{R}_d = \rho^{|m-n|}$, $m, n = 1, \ldots, N_d$, where ρ is the correlation coefficient with magnitude $|\rho| < 1$. Without loss of generality, we consider $\rho = 0.8$ and $\rho = 0.2$ for the high and low channel correlation scenarios, respectively. We assume that all nodes have the same amount of transmission power as $p_s = p_1 = p_2 = p$. In all simulation examples, the normalized MSE (NMSE) of channel estimation at the destination node is computed.

In the first example, we compare the NMSE performance of the superimposed channel training algorithm at various α. Fig. 2 shows the NMSE of the proposed algorithm versus p at various α with $N_d = N = 4$ and $\rho = 0.8$. The curve of the optimal α is obtained by applying the GSS method in the proposed superimposed channel training algorithm to find the optimal α for each value of p. It can be observed from Fig. 2 that the GSS technique is an efficient method to find the optimal α since the optimal α curve constantly has the lowest MSE level for all values of p. Furthermore, we observe from Fig. 2 that the optimal α varies with respect to p, which means that using a fixed α is strictly suboptimal. In particular, for p between 10dB and 30dB, the NMSE with $\alpha = 0.2$ is close to the NMSE using the optimal α. However, the curve associated with $\alpha = 0.07$ has a lower NMSE than the one associated with $\alpha = 0.2$ when $p = 5$dB. In addition, for other simulation examples (e.g. different N and ρ), the NMSE with $\alpha = 0.2$ might not be close to the NMSE using the optimal α.

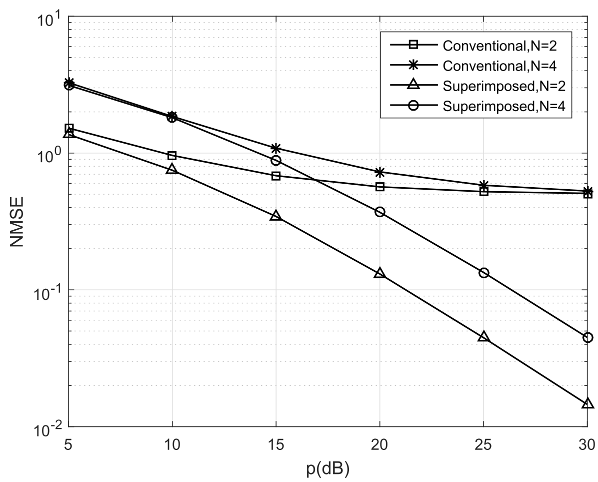


FIGURE 3. Example 2: NMSE versus p at various N with $\rho = 0.8$ and $N_d = N$.

In the second example, we set $N_d = N$ and investigate the NMSE performance of the proposed superimposed channel training algorithm using the optimal α for various simulation scenarios. A comparison of the proposed algorithm has been made with the conventional two-stage MMSE-based channel training algorithm, where random orthogonal pilot sequences are adopted to estimate channel matrices with equally distributed transmission power at the relay node between two stages [15]. Fig. 3 shows the NMSE performance of both methods with $\rho = 0.8$ for $N = 2$ and $N = 4$, while Fig. 4 demonstrates the performance of two methods with $\rho = 0.2$. It can be observed from Figs. 3 and 4 that the two algorithms have similar performance when $p < 10$dB, while at a high power level, the superimposed channel training algorithm yields much smaller estimation error than the conventional two-stage channel estimation method. Moreover, the relationship between the curves almost stays the same when the correlation coefficient ρ changes, indicating that the

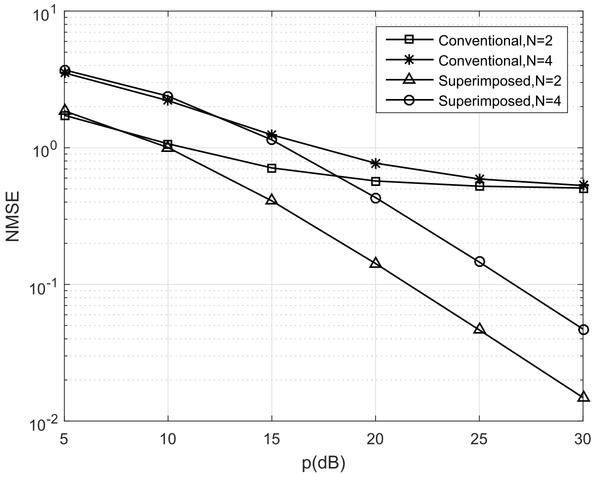


FIGURE 4. Example 2: NMSE versus p at various N with $\rho = 0.2$ and $N_d = N$.

proposed algorithm is efficient in both scenarios of high and low channel correlation. We also observe from Figs. 3 and 4 that for both algorithms the NMSE increases with N, as more unknowns need to be estimated with a larger number of antennas.

In the third example, we study the NMSE performance of the algorithm proposed in Section IV which retrieves the individual CSI of the first-hop channels. Since the first-hop channel estimation is based on the second-hop one, and the error in second-hop estimation is propagated to the first-hop channel estimation, we apply the proposed superimposed channel training algorithm in Section III to estimate the second-hop channels in this example, which has a better NMSE performance than the conventional two-stage algorithm. Fig. 5 shows the NMSE performance of this method with $\rho = 0.2$ for various N and N_d. It can be seen that with fixed N_d, the NMSE increases with N, as more unknowns need to be estimated. As expected, for a given N, the NMSE decreases with N_d since more observations are available.

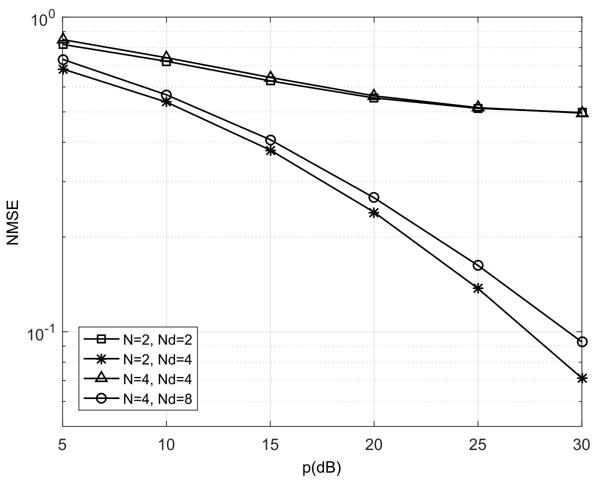


FIGURE 5. Example 3: NMSE of the first-hop channel estimation versus p at various N_d and N with $\rho = 0.2$.

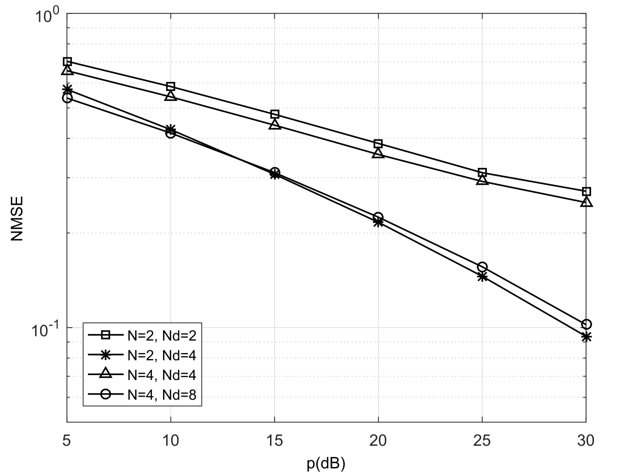


FIGURE 6. Example 3: NMSE of the first-hop channel estimation versus p at various N_d and N with $\rho = 0.8$.

Fig. 6 demonstrates the performance of this approach with $\rho = 0.8$. Similar to Fig. 5, it can be seen from Fig. 6 that a better MSE performance can be achieved by setting $N_d = N_1 + N_2$ instead of $N_d = N_1 = N_2 = N$.

VI. CONCLUSIONS

In this paper, we have applied the superimposed channel training approach to MIMO relay communication systems with multiple relay nodes. The channel matrices of both the first-hop and the second-hop links can be efficiently estimated using the proposed algorithms. The optimal training sequences and power allocation at the source and relay nodes are derived. Furthermore, a new MMSE-based estimator has been developed to retrieve the first-hop channel information. A better performance of the proposed superimposed channel training algorithm has been shown through numerical examples. The channel training algorithm proposed in this paper can be readily extended to two-way MIMO relay networks with multiple relay nodes.

APPENDIX A
PROOF OF LEMMA 1

In order to derive $\mathbf{R_v}$ and $\mathbf{R_\gamma}$, here we introduce the following lemma.

Lemma 3 [25]: For $\mathbf{H} \sim \mathcal{CN}(\mathbf{0}, \mathbf{\Theta} \otimes \mathbf{\Phi})$, there is $\mathrm{E}[\mathbf{H}\mathbf{A}\mathbf{H}^H] = tr(\mathbf{A}\mathbf{\Theta}^T)\mathbf{\Phi}$ and $\mathrm{E}[\mathbf{H}^H\mathbf{A}\mathbf{H}] = tr(\mathbf{\Phi}\mathbf{A})\mathbf{\Theta}^T$.

To derive $\mathbf{R_v}$, we first calculate the covariance matrix of $\mathbf{v}_m$ using (4) and Lemma 3, where $\mathbf{v}_m$ denotes the mth column of $\mathbf{V}$

$$\begin{aligned}\mathrm{E}\left[\mathbf{v}_m\mathbf{v}_m^H\right] &= \mathrm{E}\left[\sum_{i=1}^{K}\alpha\mathbf{G}_i\mathbf{v}_{i,m}\mathbf{v}_{i,m}^H\mathbf{G}_i^H + \mathbf{v}_{d,m}\mathbf{v}_{d,m}^H\right] \\ &= \mathrm{E}\left[\sum_{i=1}^{K}\alpha\mathbf{G}_i\mathbf{G}_i^H\right] + \mathbf{I}_{N_d} \\ &= \sum_{i=1}^{K}\alpha tr(\mathbf{C}_i^T)\mathbf{R}_d + \mathbf{I}_{N_d} \qquad (46)\end{aligned}$$

where $\mathbf{v}_{i,m}$ and $\mathbf{v}_{d,m}$ are the mth columns of $\mathbf{V}_i$ and $\mathbf{V}_d$, respectively. As $\mathbf{v}_m$, $m = 1, \ldots, T$, are independent, $\mathbf{R}_\mathbf{v}$ can be written as

$$\mathbf{R}_\mathbf{v} = \mathbf{I}_T \otimes \left(\sum_{i=1}^{K} \alpha tr(\mathbf{C}_i^T)\mathbf{R}_d + \mathbf{I}_{N_d} \right).$$

Now let us calculate $\mathbf{R}_\gamma$. From (5), (7), and (9), the mth column of $\tilde{\mathbf{M}}$ and $\tilde{\mathbf{G}}_i$ can be written as

$$\tilde{\mathbf{m}}_m = \lambda_{s,m}^{\frac{1}{2}} \sum_{i=1}^{K} \mathbf{G}_i\mathbf{A}_i\mathbf{H}_{i,w}\boldsymbol{\pi}_{s,m}, \quad \tilde{\mathbf{g}}_{i,m} = \lambda_{i,m}^{\frac{1}{2}}\mathbf{A}_d\mathbf{G}_{i,w}\boldsymbol{\pi}_{i,m}$$

where $\boldsymbol{\pi}_{s,m}$ and $\boldsymbol{\pi}_{i,m}$ are the mth column of $\mathbf{\Pi}_s$ and $\mathbf{\Pi}_i$, respectively, $\lambda_{s,m}$ and $\lambda_{i,m}$ are the mth diagonal element of $\mathbf{\Lambda}_s$ and $\mathbf{\Lambda}_i$, respectively. Then similar to (46), the covariance matrix of $\tilde{\mathbf{m}}_m$ and $\tilde{\mathbf{g}}_{i,m}$ are given by

$$\begin{aligned} \mathrm{E}[\tilde{\mathbf{m}}_m\tilde{\mathbf{m}}_m^H] &= \lambda_{s,m}\mathrm{E}\left[\sum_{i=1}^{K}\mathbf{G}_i\mathbf{A}_i\mathbf{H}_{i,w}\boldsymbol{\pi}_{s,m}\boldsymbol{\pi}_{s,m}^H\mathbf{H}_{i,w}^H\mathbf{A}_i^H\mathbf{G}_i^H\right] \\ &= \lambda_{s,m}\sum_{i=1}^{K} tr(\mathbf{R}_i\mathbf{C}_i^T)\mathbf{R}_d \\ &= \lambda_{s,m}b\mathbf{R}_d \\ \mathrm{E}[\tilde{\mathbf{g}}_{i,m}\tilde{\mathbf{g}}_{i,m}^H] &= \lambda_{i,m}\mathrm{E}\left[\mathbf{A}_d\mathbf{G}_{i,w}\boldsymbol{\pi}_{i,m}\boldsymbol{\pi}_{i,m}^H\mathbf{G}_{i,w}^H\mathbf{A}_d^H\right] \\ &= \lambda_{i,m}\mathbf{R}_d. \end{aligned}$$

As $\tilde{\mathbf{m}}_m$, $m = 1, \ldots, N_s$ and $\tilde{\mathbf{g}}_{i,m}$, $m = 1, \ldots, N_i$, $i = 1, \ldots, K$, are independent, we have

$$\mathbf{R}_\gamma = \mathrm{bdiag}[\mathbf{\Lambda}_s \otimes b\mathbf{R}_d, \mathbf{\Lambda}_1 \otimes \mathbf{R}_d, \ldots, \mathbf{\Lambda}_K \otimes \mathbf{R}_d].$$

APPENDIX B
PROOF OF LEMMA 2

Denoting $a_{m,n} \triangleq 1/\left(b\lambda_{s,m}\lambda_{d,n}\right)$, $c_m \triangleq \alpha\sigma_{s,m}$, and $g_{i,m,n} \triangleq 1/\left(\lambda_{i,m}\lambda_{d,n}\right)$, we can rewrite the problem (34)–(37) as

$$\min_{\alpha,\mathbf{c},\{\boldsymbol{\sigma}_i\}} \sum_{m=1}^{N_s}\sum_{n=1}^{N_d}\frac{1}{a_{m,n}+c_m d_{v,n}} + \sum_{i=1}^{K}\sum_{m=1}^{N_i}\sum_{n=1}^{N_d}\frac{1}{g_{i,m,n}+\sigma_{i,m}d_{v,n}} \quad (47)$$

$$s.t.\ \mathbf{1}^T\mathbf{c} \le \alpha p_s \quad (48)$$

$$\mathbf{z}_i^T\mathbf{c} + \mathbf{1}^T\boldsymbol{\sigma}_i \le p_i - \alpha N_i, \quad i = 1, \ldots, K \quad (49)$$

$$\alpha > 0, \quad c_m \ge 0, \quad m = 1, \ldots, N_s \quad (50)$$

$$\sigma_{i,m} \ge 0, \quad m = 1, \ldots, N_i, \quad i = 1, \ldots, K \quad (51)$$

where $\mathbf{c} \triangleq \left[c_1, \ldots, c_{N_s}\right]^T$, $\mathbf{z}_i \triangleq tr(\mathbf{R}_i)\left[\lambda_{s,1}, \ldots, \lambda_{s,N_s}\right]^T$, $\boldsymbol{\sigma}_i \triangleq \left[\sigma_{i,1}, \ldots, \sigma_{i,N_i}\right]^T$, and $\{\boldsymbol{\sigma}_i\} = \{\boldsymbol{\sigma}_i, i = 1, \ldots, K\}$.

For a very small α, the value of (47) is mainly governed by the constraint in (48), since the constraints in (49) is not active compared with the constraint (48) when α is small. As α increases from a small value, the feasible region determined by (48) expands, which contributes to the decrease in the value of (47). On the other hand, when α is large, the value of (47) is mainly affected by the constraints in (49), as the constraint (48) is not active compared with those in (49) for a large value of α. As α decreases from a large value, the feasible region specified by (49) expands, resulting in a decreasing of (47).

Now let us consider the effect of α on $d_{v,n}$. Note that $d_{v,n}$ in (38) is monotonically decreasing when α increases, and the value of (47) increases with decreasing $d_{v,n}$. Taking the two effects above into account, we can draw the following conclusion regarding the value of (47) with respect to α. When α increases from a very small positive number, the value of (47) starts to decrease as the potential decrease of (47) caused by the relaxed feasible region (48) dominates the potential increase of (47) due to the decreasing $d_{v,n}$. The value of (47) keeps decreasing as α increases till a 'turning point' where the decreasing of $d_{v,n}$ starts to dominate the effect of relaxed feasible region (48). After this turning point, the value of (47) will monotonically increase with an increasing α. Therefore, the objective function (34) subjecting to (35)-(37) is proved to be a unimodal function with respect to α.

To validate the analysis above, we plot the MSE value (34) versus α in Fig. 7 at various levels of p_s. Here we set $p_1 = p_2 = 20$dB and the other simulation setups are the same as those in Fig. 2. It can be clearly seen from Fig. 7 that the objective function (34) subjecting to the constraints (35)–(37) is a unimodal function of α.

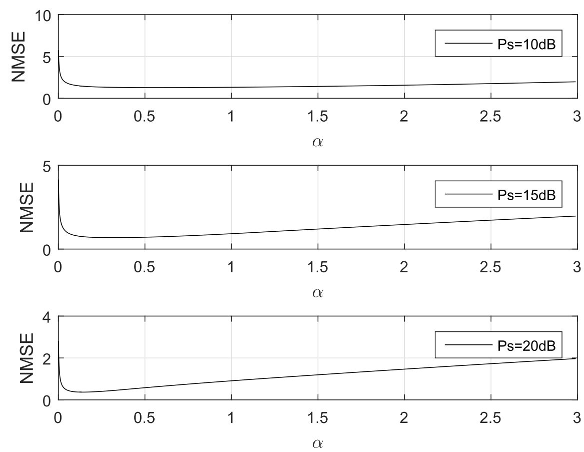


FIGURE 7. NMSE versus α at various p_s with $p_1 = p_2 = 20$dB.

APPENDIX C
CALCULATION OF $\mathbf{R}_{\tilde{\mathbf{h}}}$ AND $\mathbf{R}_{\tilde{\mathbf{v}}}$

From (5) and (7), we have

$$\tilde{\mathbf{H}}_i = \mathbf{H}_i\mathbf{U}_s = \mathbf{A}_i\mathbf{H}_{i,w}\mathbf{\Pi}_s\mathbf{\Lambda}_s^{\frac{1}{2}}. \quad (52)$$

Using (52), the covariance matrix of the pth column of $\tilde{\mathbf{H}}_i$, denoted as $\tilde{\mathbf{h}}_{i,p}$, is given by

$$\begin{aligned} \mathrm{E}[\tilde{\mathbf{h}}_{i,p}\tilde{\mathbf{h}}_{i,p}^H] &= \lambda_{s,p}\mathrm{E}\left[\mathbf{A}_i\mathbf{H}_{i,w}\boldsymbol{\pi}_{s,p}\boldsymbol{\pi}_{s,p}^H\mathbf{H}_{i,w}^H\mathbf{A}_i^H\right] \\ &= \lambda_{s,p}\mathbf{R}_i. \end{aligned} \quad (53)$$

Then the covariance matrix of $\tilde{\mathbf{h}}_p = [\tilde{\mathbf{h}}_{1,p}^T, \ldots, \tilde{\mathbf{h}}_{K,p}^T]^T$, the pth column of $\tilde{\mathbf{H}}$, can be obtained based on (53) as

$$\mathbf{R}_{\tilde{\mathbf{h}}_p} = \mathrm{bdiag}\left[\lambda_{s,p}\mathbf{R}_1, \lambda_{s,p}\mathbf{R}_2, \ldots, \lambda_{s,p}\mathbf{R}_K\right]. \quad (54)$$

As $\tilde{\mathbf{h}} = [\tilde{\mathbf{h}}_1^T, \cdots, \tilde{\mathbf{h}}_{N_s}^T]^T$, we have

$$\mathbf{R}_{\tilde{\mathbf{h}}} = \mathrm{E}[\tilde{\mathbf{h}}\tilde{\mathbf{h}}^H] = \mathrm{bdiag}\left[\mathbf{R}_{\tilde{\mathbf{h}}_1}, \mathbf{R}_{\tilde{\mathbf{h}}_2}, \ldots, \mathbf{R}_{\tilde{\mathbf{h}}_{N_s}}\right].$$

Now we start to calculate $\mathbf{R}_{\bar{\mathbf{v}}} = \mathrm{E}[\bar{\mathbf{v}}\bar{\mathbf{v}}^H]$. From (42), the correlation matrix of the mth and nth columns of $\bar{\mathbf{V}}$, $m, n = 1, \ldots, T$, is given by

$$\mathrm{E}[\bar{\mathbf{v}}_m\bar{\mathbf{v}}_n^H] = \alpha \mathrm{E}[\boldsymbol{\Delta}\tilde{\mathbf{H}}\tilde{\mathbf{s}}_m\tilde{\mathbf{s}}_n^H\tilde{\mathbf{H}}^H\boldsymbol{\Delta}^H] + \mathrm{E}\left[\sum_{i=1}^{K}\boldsymbol{\Psi}_i\tilde{\mathbf{t}}_{i,m}\tilde{\mathbf{t}}_{i,n}^H\boldsymbol{\Psi}_i^H\right] + \mathrm{E}[\mathbf{v}_m\mathbf{v}_n^H] \quad (55)$$

where $\boldsymbol{\Delta} = (\boldsymbol{\Delta}_1, \ldots, \boldsymbol{\Delta}_K)$, $\boldsymbol{\Delta}_i = \boldsymbol{\Psi}_i\mathbf{U}_i^H$, $\boldsymbol{\Psi}_i = \tilde{\mathbf{G}}_i - \hat{\tilde{\mathbf{G}}}_i$, $i = 1, \ldots, K$, $\hat{\tilde{\mathbf{G}}}_i$ is the estimate of $\tilde{\mathbf{G}}_i$, $\tilde{\mathbf{s}}_m$, $\tilde{\mathbf{t}}_{i,m}$, and $\mathbf{v}_m$ are the mth columns of $\tilde{\mathbf{S}}$, $\tilde{\mathbf{T}}_i$, and $\mathbf{V}$, respectively. From (46), we obtain

$$\mathrm{E}[\mathbf{v}_m\mathbf{v}_n^H] = \begin{cases} \sum_{i=1}^{K}\alpha tr(\mathbf{C}_i^T)\mathbf{R}_d + \mathbf{I}_{N_d}, & m = n; \\ \mathbf{0}, & m \neq n. \end{cases}$$

With $\tilde{\mathbf{t}}_{i,m} = \boldsymbol{\Sigma}_i^{\frac{1}{2}}\boldsymbol{\phi}_{i,m}$, where $\boldsymbol{\phi}_{i,m}$ is the mth column of $\boldsymbol{\Phi}_i$, the (e,f)-th entry of the second term in (55) can be calculated as

$$\begin{aligned}\left[\sum_{i=1}^{K}\mathrm{E}[\boldsymbol{\Psi}_i\tilde{\mathbf{t}}_{i,m}\tilde{\mathbf{t}}_{i,n}^H\boldsymbol{\Psi}_i^H]\right]_{e,f} &= \sum_{i=1}^{K}\mathrm{E}\Big[\left(\delta_{i,1,e}, \ldots, \delta_{i,N_i,e}\right)\boldsymbol{\Sigma}_i^{\frac{1}{2}}\boldsymbol{\phi}_{i,m}\boldsymbol{\phi}_{i,n}^H \\ &\quad \times \boldsymbol{\Sigma}_i^{\frac{1}{2}}\left(\delta_{i,1,f}, \ldots, \delta_{i,N_i,f}\right)^H\Big] \\ &= \sum_{i=1}^{K} tr(\boldsymbol{\Sigma}_i^{\frac{1}{2}}\boldsymbol{\phi}_{i,m}\boldsymbol{\phi}_{i,n}^H\boldsymbol{\Sigma}_i^{\frac{1}{2}}\mathbf{D}_{i,e,f}) \end{aligned} \quad (56)$$

where $\delta_{i,p,q}$ is the qth entry of $\boldsymbol{\delta}_{i,p} = \tilde{\mathbf{g}}_{i,p} - \hat{\tilde{\mathbf{g}}}_{i,p}, p = 1, \ldots, N_i$, $q = 1, \ldots, N_d$, $\tilde{\mathbf{g}}_{i,p}$ and $\hat{\tilde{\mathbf{g}}}_{i,p}$ denote the pth columns of $\tilde{\mathbf{G}}_i$ and $\hat{\tilde{\mathbf{G}}}_i$, respectively, and $\mathbf{D}_{i,e,f}$ is given by

$$\begin{aligned}\mathbf{D}_{i,e,f} &= \mathrm{E}\left[\left(\delta_{i,1,f}, \ldots, \delta_{i,N_i,f}\right)^H\left(\delta_{i,1,e}, \ldots, \delta_{i,N_i,e}\right)\right] \\ &= \mathrm{diag}(\mathrm{E}[\delta_{i,1,e}\delta_{i,1,f}^*], \ldots, \mathrm{E}[\delta_{i,N_i,e}\delta_{i,N_i,f}^*]) \\ &= \mathrm{diag}\left(\left[\boldsymbol{\Omega}_{i,1}\right]_{e,f}, \ldots, \left[\boldsymbol{\Omega}_{i,N_i}\right]_{e,f}\right). \end{aligned} \quad (57)$$

Here

$$\begin{aligned}\boldsymbol{\Omega}_{i,p} &= \mathrm{E}\left[\boldsymbol{\delta}_{i,p}\boldsymbol{\delta}_{i,p}^H\right] \\ &= \mathrm{E}\left[\left(\tilde{\mathbf{g}}_{i,p} - \hat{\tilde{\mathbf{g}}}_{i,p}\right)\left(\tilde{\mathbf{g}}_{i,p} - \hat{\tilde{\mathbf{g}}}_{i,p}\right)^H\right] \\ &= \left[(\lambda_{i,p}\mathbf{R}_d)^{-1} + \sigma_{i,p}\left(\sum_{i=1}^{K}\alpha tr(\mathbf{C}_i^T)\mathbf{R}_d + \mathbf{I}_{N_d}\right)^{-1}\right]^{-1}. \end{aligned} \quad (58)$$

In other words, $\mathbf{D}_{i,e,f}$ in (57) is a diagonal matrix whose pth diagonal entry, $p = 1, \ldots, N_i$, is the (e,f)-th entry of $\boldsymbol{\Omega}_{i,p}$ in (58).

Finally, the first term in (55) can be calculated as

$$\begin{aligned}\mathrm{E}[\boldsymbol{\Delta}\tilde{\mathbf{H}}\tilde{\mathbf{s}}_m\tilde{\mathbf{s}}_n^H\tilde{\mathbf{H}}^H\boldsymbol{\Delta}^H] &= \mathrm{E}\left[\sum_{i=1}^{K}\boldsymbol{\Delta}_i\tilde{\mathbf{H}}_i\tilde{\mathbf{s}}_m\tilde{\mathbf{s}}_n^H\sum_{i=1}^{K}\tilde{\mathbf{H}}_i^H\boldsymbol{\Delta}_i^H\right] \\ &= \sum_{i=1}^{K}\mathrm{E}[\boldsymbol{\Delta}_i\tilde{\mathbf{H}}_i\tilde{\mathbf{s}}_m\tilde{\mathbf{s}}_n^H\tilde{\mathbf{H}}_i^H\boldsymbol{\Delta}_i^H] \\ &= \tilde{\mathbf{s}}_n^H\boldsymbol{\Lambda}_s\tilde{\mathbf{s}}_m\sum_{i=1}^{K}\mathrm{E}[\boldsymbol{\Delta}_i\mathbf{R}_i\boldsymbol{\Delta}_i^H] \end{aligned} \quad (59)$$

where we used the fact that

$$\begin{aligned}\mathrm{E}[\tilde{\mathbf{H}}_i\tilde{\mathbf{s}}_m\tilde{\mathbf{s}}_n^H\tilde{\mathbf{H}}_i^H] &= \mathrm{E}[\mathbf{A}_i\mathbf{H}_{i,w}\boldsymbol{\Pi}_s\boldsymbol{\Lambda}_s^{\frac{1}{2}}\tilde{\mathbf{s}}_m\tilde{\mathbf{s}}_n^H\boldsymbol{\Lambda}_s^{\frac{1}{2}}\boldsymbol{\Pi}_s^H\mathbf{H}_{i,w}^H\mathbf{A}_i^H] \\ &= \tilde{\mathbf{s}}_n^H\boldsymbol{\Lambda}_s\tilde{\mathbf{s}}_m\mathbf{R}_i. \end{aligned}$$

To derive $\mathrm{E}[\boldsymbol{\Delta}_i\mathbf{R}_i\boldsymbol{\Delta}_i^H]$, its (e,f)-th entry, $e, f = 1, \ldots, N_d$, can be calculated as

$$\begin{aligned}&\left[\mathrm{E}[\boldsymbol{\Delta}_i\mathbf{R}_i\boldsymbol{\Delta}_i^H]\right]_{e,f} \\ &= \mathrm{E}\left[\left(\delta_{i,1,e}, \ldots, \delta_{i,N_i,e}\right)\mathbf{U}_i^H\mathbf{R}_i\mathbf{U}_i\left(\delta_{i,1,f}, \ldots, \delta_{i,N_i,f}\right)^H\right] \\ &= tr(\mathbf{U}_i^H\mathbf{R}_i\mathbf{U}_i\mathbf{D}_{i,e,f}). \end{aligned} \quad (60)$$

REFERENCES

[1] Y. Fan and J. Thompson, "MIMO configurations for relay channels: Theory and practice," *IEEE Trans. Wireless Commun.*, vol. 6, no. 5, pp. 1774–1786, May 2007.

[2] 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.

[3] X. Tang and Y. Hua, "Optimal design of non-regenerative MIMO wireless relays," *IEEE Trans. Wireless Commun.*, vol. 6, no. 4, pp. 1398–1407, Apr. 2007.

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

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

[6] Y. Rong, "Linear non-regenerative multicarrier MIMO relay communications based on MMSE criterion," *IEEE Trans. Commun.*, vol. 58, no. 7, pp. 1918–1923, Jul. 2010.

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

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

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

[10] A. Toding, M. R. A. Khandaker, and Y. Rong, "Joint source and relay optimization for parallel MIMO relay networks," *EURASIP J. Adv. Signal Process.*, vol. 2012, p. 174, Aug. 2012.

[11] P. Lioliou and M. Viberg, "Least-squares based channel estimation for MIMO relays," in *Proc. Int. ITG Workshop Smart Antennas*, Feb. 2008, pp. 90–95.

[12] P. Lioliou, M. Viberg, and M. Coldrey, "Efficient channel estimation techniques for amplify and forward relaying systems," *IEEE Trans. Commun.*, vol. 60, no. 11, pp. 3150–3155, Nov. 2012.

[13] T. Kong and Y. Hua, "Optimal design of source and relay pilots for MIMO relay channel estimation," *IEEE Trans. Signal Process.*, vol. 59, no. 9, pp. 4438–4446, Sep. 2011.

[14] Y. Rong, M. R. A. Khandaker, and Y. Xiang, "Channel estimation of dual-hop MIMO relay system via parallel factor analysis," *IEEE Trans. Wireless Commun.*, vol. 11, no. 6, pp. 2224–2233, Jun. 2012.

[15] C. W. R. Chiong, Y. Rong, and Y. Xiang, "Channel training algorithms for two-way MIMO relay systems," *IEEE Trans. Signal Process.*, vol. 61, no. 16, pp. 3988–3998, Aug. 2013.

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

[17] C. W. R. Chiong, Y. Rong, and Y. Xiang, "Channel estimation for time-varying MIMO relay systems," *IEEE Trans. Wireless Commun.*, vol. 14, no. 12, pp. 6752–6762, Dec. 2015.

[18] H. Mehrpouyan, S. D. Blostein, and B. Ottersten, "Simultaneous estimation of multi-relay MIMO channels," in *Proc. IEEE Global Commun. Conf.*, Atlanta, GA, USA, Dec. 2013, pp. 4269–4274.

[19] L. Canonne-Velasquez, H. Nguyen-Le, and T. Le-Ngoc, "Cascaded doubly-selective channel estimation in multi-relay AF OFDM transmissions," in *Proc. IEEE Int. Conf. Commun.*, Ottawa, ON, Canada, Jun. 2012, pp. 3836–3840.

[20] J. W. Brewer, "Kronecker products and matrix calculus in system theory," *IEEE Trans. Circuits Syst.*, vol. CAS-25, no. 9, pp. 772–781, Sep. 1978.

[21] D. S. Shiu, G. Foschini, M. Gans, and J. Kahn, "Fading correlation and its effect on the capacity of multielement antenna systems," *IEEE Trans. Commun.*, vol. 48, no. 3, pp. 503–513, Mar. 2000.

[22] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*, vol. 1. Englewood Cilffs, NJ, USA: Prentice-Hall, 1993.

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

[24] A. Antoniou and W.-S. Lu, *Practical Optimization: Algorithms and Engineering Applications*. Spring Street, NY, USA: Springer, 2007.

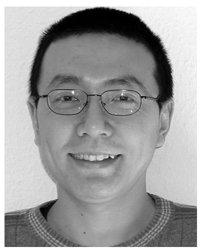
[25] A. Gupta and D. Nagar, *Matrix Variate Distributions.* London, U.K.: Chapman & Hall, 2000.

ZHIQIANG HE (S'01–M'04) received the B.E. degree in signal and information processing 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.

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.

XIAODAN WANG received the B.E. and M.Sc. degrees in information and communication engineering from the Beijing University of Posts and Telecommunications, China, in 2014 and 2017, respectively. She majors in new physical layer technology applied in 5G, with an emphasis on channel estimation and different schemes of channel coding and nonorthogonal multiple access.

YANG LIU received the M.Sc. degree in information and communication engineering from the Beijing University of Posts and Telecommunications, Beijing, China, in 2017. His research focuses on frequency-domain equalization algorithms, and the results have been published at IC-NIDC.



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 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. 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 150 journal and conference papers 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 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 a TPC Member for the IEEE ICC, WCSP, IWCMC, EUSIPCO, and ChinaCom. He is an Associate Editor of the IEEE Transactions on Signal Processing.