Channel Estimation for Two-Way MIMO Relay Systems in Frequency-Selective Fading Environments

Choo W. R. Chiong, Yue Rong, *Senior Member, IEEE*, and Yong Xiang, *Senior Member, IEEE*

Abstract—In this paper, we investigate the channel estimation problem for two-way multiple-input multiple-output (MIMO) relay communication systems in frequency-selective fading environments. We apply the method of superimposed channel training to estimate the individual channel state information (CSI) of the first-hop and second-hop links for two-way MIMO relay systems with frequency-selective fading channels. In this algorithm, a relay training sequence is superimposed on the received signals at the relay node to assist the estimation of the second-hop channel matrices. The optimal structure of the source and relay training sequences is derived to minimize the mean-squared error (MSE) of channel estimation. Moreover, the optimal power allocation between the source and relay training sequences is derived to improve the performance of channel estimation. Numerical examples are shown to demonstrate the performance of the proposed superimposed channel training algorithm for two-way MIMO relay systems in frequency-selective fading environments.

Index Terms—Channel estimation, MIMO relay, frequency-selective fading, superimposed training, MMSE, two-way relay, power allocation.

I. Introduction

RECENTLY, multiple-input multiple-output (MIMO) relay communication systems have attracted much attention as the demand for high rate and reliable wireless communications keeps increasing [1], [2]. In [3] and [4], the optimal relay precoding matrix for a three-node two-hop MIMO relay communication system has been developed to maximize the mutual information between the source and destination nodes. A unified framework has been developed in [5] to optimize the source and relay precoding matrices of two-hop MIMO relay systems based on a broad class of commonly used objective functions. In [6], the joint transmit and relay precoding design problems were investigated for two-hop multicasting MIMO relay systems.

The MIMO relay communication systems discussed in [2]–[5] are one-way relay systems, where the signals are transmitted from a source node to a destination node through relay node(s). In a two-way relay system, two source nodes exchange their information through relay node(s). Initially studied by Shannon in [7], two-way relay communications have attracted great interests recently as they can provide higher spectral efficiency compared with one-way relay systems. The joint source and relay optimization for two-way MIMO relay systems has been studied in [8]–[10].

For the MIMO relay systems discussed in [2]–[10], the knowledge of the instantaneous channel state information (CSI) is essential for extracting the source signals at the destination node and the optimization of MIMO relay systems through precoding matrices design and power allocation. However, the instantaneous CSI is unknown in practical wireless relay communication systems, and therefore, needs to be estimated at the destination node. In [11], a least-squares (LS) based channel estimation algorithm was developed for MIMO relay systems. The performance of [11] was further analyzed and improved in [12] by using the weighted least-squares (WLS) fitting method. A two-stage channel estimation algorithm based on the linear minimum mean-squared error (LMMSE) criterion was introduced in [13], while in [14], a parallel factor (PARAFAC) analysis based MIMO relay channel estimation algorithm has been developed. A superimposed channel training algorithm has been proposed in [15] for orthogonal frequency division multiplexing (OFDM) modulated one-way relay systems.

The channel estimation algorithms in [11]–[15] were developed for one-way MIMO relay systems. Due to a larger number of unknowns, channel estimation problems are generally more challenging in two-way relay systems than those in one-way relay systems. In [16], two-way relay channel estimation algorithms based on the maximum likelihood (ML) and linear maximum signal-to-noise ratio (SNR) criteria have been proposed. However, the algorithms in [16] were designed for single antenna relay systems, and the extension to MIMO relay systems is not straightforward. Two methods were presented in [17] for two-way MIMO relay systems, namely, cascaded channel estimation and individual channel estimation. In the first algorithm, the cascaded channel matrices are estimated at two source nodes. However, this approach cannot estimate the individual second-hop channel matrices, which are essential for the optimization of MIMO relay networks [10].

This problem has been addressed by the superimposed channel training algorithm for two-way MIMO relay systems in

[18], where a training sequence is superimposed at the relay node. The purpose of superimposing a training matrix at the relay node is to estimate the CSI of *individual* first-hop and second-hop channel matrices at the destination nodes, which cannot be achieved by simply multiplying the received signals at the relay node with a relay precoding matrix. Individual CSI can also be obtained by first estimating the first-hop channel matrices at the relay node and then forwarding the estimated channel matrices to the destination nodes, as the individual channel estimation algorithm in [17]. Obviously, the approach in [17] increases the cost and complexity of the relay node.

The relay systems in [16]–[18] are assumed to have frequency-flat fading channels, which is only valid for narrow-band communication systems. In this paper, we consider a more general situation where two-way MIMO relay systems are operating in frequency-selective fading environments, i.e., there are multiple paths between each transmit-receive antenna pair. We apply the method of superimposed channel training to estimate the individual channel matrices of the first-hop and second-hop links for two-way MIMO relay systems in frequency-selective fading environments. In particular, the channel training is completed in two time blocks. In the first time block, both source nodes transmit their training sequences simultaneously to the relay node. The relay then amplifies the received signals and superimposes its own training sequences before broadcasting the superimposed signals to the destination nodes. The channel estimation processes are implemented at the destination nodes to minimize the amount of signal processing at the relay node.

Since the superimposed channel training approach does not require the relay node to be capable of performing the advanced signal processing of channel estimation, and hence, provides an easy and cost-effective implementation of two-way relay communication systems. Such advantage of the superimposed channel training approach is particularly important under frequency-selective channels, as the complexity at the relay node increases significantly compared with the frequency-flat fading environment when the approach in [17] is used. Thus, the superimposed channel estimation method is preferred from practical point of view.

We derive the optimal source and relay training sequences by minimizing the sum MSE of channel estimation. We also optimize the power allocation between the source and relay training sequences at the relay node. The algorithm developed in this paper generalizes the results in [18] from frequency-flat fading channel to frequency-selective fading channels. We would like to note that such extension is non-trivial as the optimization problem for channel estimation in frequency-selective two-way MIMO relay systems is much more complicated than that of frequency-flat relay systems. Moreover, we develop a new MMSE-based algorithm to retrieve the first-hop channel matrices, which takes into account the estimation error inherited from the estimation of the second-hop channel matrices.

The rest of the paper is organized as follows. The system model of a two-way MIMO relay system in frequency-selective fading environments 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. Section IV shows
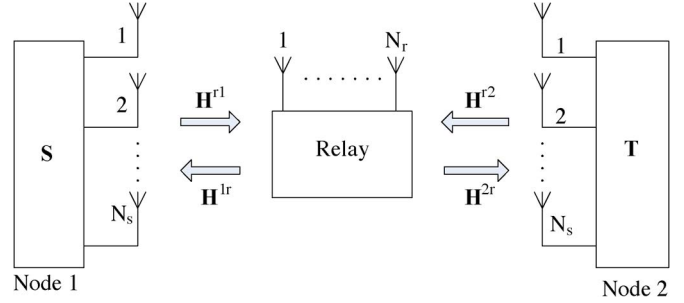

Fig. 1. Block diagram of a two-way MIMO relay communication system.

numerical simulations to demonstrate the performance of the proposed algorithm. Conclusions are drawn in Section V.

II. SYSTEM MODEL

We consider a three-node two-way MIMO relay communication system operating in a frequency-selective fading environment, where two source nodes, node 1 and node 2, exchange information through a relay node as shown in Fig. 1. The source nodes and relay node are equipped with N_s and N_r antennas, respectively. In this paper, we assume that the practical half-duplex mode is used at all nodes, i.e., each node is not able to send and receive signals at the same time. With this assumption, there is no direct link between two source nodes as both source nodes are transmitting signals at the first time block and cannot receive signals from each other. The implementation of half-duplex mode is common in two-way relay communications.

Let us denote $\mathbf{h}_{n,m}^{ri} = [h_{n,m,1}^{ri}, \ldots, h_{n,m,Q}^{ri}]^T$ as the $Q \times 1$ first-hop multipath channel vector from the mth antenna at node i to the nth antenna at the relay node, $i = 1, 2$, $m = 1, \ldots, N_s$, and $n = 1, \ldots, N_r$, where $(\cdot)^T$ denotes the matrix (vector) transpose and we assume that all channels have the same number of taps Q. The extension to systems with different number of channel taps between each transmit and receive antenna pair is straightforward. In a similar way, $\mathbf{h}_{n,m}^{ir} = [h_{n,m,1}^{ir}, \ldots, h_{n,m,Q}^{ir}]^T$ is used to denote the $Q \times 1$ second-hop multipath channel vector from the mth antenna at the relay node to the nth antenna at node i.

The channel estimation process is completed in two time blocks. In the first time block, source node 1 transmits an $N_s \times L$ training signal matrix $\mathbf{S} = [\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_{N_s}]^T$ and node 2 transmits an $N_s \times L$ training matrix $\mathbf{T} = [\mathbf{t}_1, \mathbf{t}_2, \ldots, \mathbf{t}_{N_s}]^T$, respectively, where $L > Q$ is the length of the training sequence and will be determined later. Cyclic prefixes of length $L_{cp} \geq Q$ are inserted at $\mathbf{s}_m$ and $\mathbf{t}_n$, $m, n = 1, \ldots, N_s$, to prevent the inter-block interference at the relay node [15]. The received signal vectors at the relay node over L time slots after removing the cyclic prefix can be written as

$$\mathbf{y}_{r,n} = \sum_{m=1}^{N_s} \mathbf{H}_{n,m}^{r1}\mathbf{s}_m + \sum_{m=1}^{N_s} \mathbf{H}_{n,m}^{r2}\mathbf{t}_m + \mathbf{v}_{r,n}$$

$$= \sum_{m=1}^{N_s} \mathbf{C}_Q(\mathbf{s}_m)\mathbf{h}_{n,m}^{r1} + \sum_{m=1}^{N_s} \mathbf{C}_Q(\mathbf{t}_m)\mathbf{h}_{n,m}^{r2} + \mathbf{v}_{r,n},$$

$$n = 1, \ldots, N_r \tag{1}$$

where $\mathbf{y}_{r,n}$ and $\mathbf{v}_{r,n}$ are the $L \times 1$ received signal vector and noise vector at the nth antenna of the relay node, respectively, $\mathbf{H}_{n,m}^{r1}$ and $\mathbf{H}_{n,m}^{r2}$ are $L \times L$ circulant channel matrices whose first columns are given by $[(\mathbf{h}_{n,m}^{r1})^T, \mathbf{0}_{1 \times (L-Q)}]^T$ and $[(\mathbf{h}_{n,m}^{r2})^T, \mathbf{0}_{1 \times (L-Q)}]^T$, respectively, and $\mathbf{C}_Q(\mathbf{s})$ represents an $L \times Q$ column-wise circulant matrix taking $\mathbf{s}$ as the first column.

In the second time block, the relay node amplifies $\mathbf{y}_{r,n}$, $n = 1, \ldots, N_r$, and superimposes its own training matrix $\mathbf{R} = [\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_{N_r}]^T$. Thus, the signal vector transmitted by the nth antenna of the relay node is given by

$$\mathbf{x}_{r,n} = \sqrt{\alpha}\mathbf{y}_{r,n} + \mathbf{r}_n, \quad n = 1, \ldots, N_r \tag{2}$$

where $\alpha > 0$ is the relay amplifying factor. Similarly, a cyclic prefix is inserted at $\mathbf{x}_{r,n}$ prior to the transmission. The received signal vectors at the source node i, $i = 1, 2$, after removing the cyclic prefix are given by[1]

$$\mathbf{y}_{i,k} = \sum_{n=1}^{N_r} \mathbf{H}_{k,n}^{ir}\mathbf{x}_{r,n} + \mathbf{v}_{i,k}, \quad k = 1, \ldots, N_s \tag{3}$$

where $\mathbf{y}_{i,k}$ and $\mathbf{v}_{i,k}$ are the $L \times 1$ received signal vector and noise vector at the kth antenna of node i, respectively, $\mathbf{H}_{k,n}^{ir}$ is an $L \times L$ circulant channel matrix whose first column is $[(\mathbf{h}_{k,n}^{ir})^T, \mathbf{0}_{1 \times (L-Q)}]^T$.

The main idea of the superimposed channel training algorithm is to exploit $\mathbf{R}$ to estimate the second-hop channels $\{\mathbf{h}_{k,n}^{ir}\} \triangleq \{\mathbf{h}_{k,n}^{ir}, i = 1, 2, k = 1, \ldots, N_s, n = 1, \ldots, N_r\}$, and then estimate the first-hop channels $\{\mathbf{h}_{n,m}^{ri}\} \triangleq \{\mathbf{h}_{n,m}^{ri}, i = 1, 2, n = 1, \ldots, N_r, m = 1, \ldots, N_s\}$ using $\mathbf{S}$, $\mathbf{T}$, and the estimated $\{\mathbf{h}_{k,n}^{ir}\}$. In this paper, we assume that

1) All channel taps are zero-mean circularly symmetric complex Gaussian (CSCG) random variables.
2) Channel taps associated with the same transmit-receive antenna pair, as well as different transmit-receive antenna pairs are independent from each other.
3) Channels are assumed to be quasi-static, i.e., channels do not change within one cycle of transmission.
4) All noises are independent and identically distributed (i.i.d.) additive white Gaussian noise (AWGN) with zero mean and unit variance.

III. MMSE-Based Optimal Training Matrices

In this section, we design the optimal training matrices $\mathbf{S}$, $\mathbf{T}$, $\mathbf{R}$, and the relay amplifying factor α to minimize the MSE of channel estimation. By substituting (1) and (2) into (3), we obtain

$$\mathbf{y}_{i,k} = \sum_{n=1}^{N_r} \mathbf{H}_{k,n}^{ir}\left(\sqrt{\alpha}\sum_{m=1}^{N_s}\mathbf{H}_{n,m}^{r1}\mathbf{s}_m + \sqrt{\alpha}\sum_{m=1}^{N_s}\mathbf{H}_{n,m}^{r2}\mathbf{t}_m + \mathbf{r}_n + \sqrt{\alpha}\mathbf{v}_{r,n}\right) + \mathbf{v}_{i,k}$$

$$= \sqrt{\alpha}\sum_{m=1}^{N_s}\sum_{n=1}^{N_r}\mathbf{H}_{k,n}^{ir}\mathbf{H}_{n,m}^{r1}\mathbf{s}_m + \sqrt{\alpha}\sum_{m=1}^{N_s}\sum_{n=1}^{N_r}\mathbf{H}_{k,n}^{ir}\mathbf{H}_{n,m}^{r2}\mathbf{t}_m$$

$$+ \sum_{n=1}^{N_r}\mathbf{H}_{k,n}^{ir}\mathbf{r}_n + \bar{\mathbf{v}}_{i,k}, \quad k = 1, \ldots, N_s \tag{4}$$

where

$$\bar{\mathbf{v}}_{i,k} \triangleq \sqrt{\alpha}\sum_{n=1}^{N_r}\mathbf{H}_{k,n}^{ir}\mathbf{v}_{r,n} + \mathbf{v}_{i,k}, \quad k = 1, \ldots, N_s \tag{5}$$

is the equivalent noise vector at the kth antenna of node i. Since both $\mathbf{H}_{k,n}^{ir}$ and $\mathbf{H}_{n,m}^{ri}$ are circulant matrices, (4) can be rewritten by exploiting the property of circulant matrix as

$$\mathbf{y}_{i,k} = \sqrt{\alpha}\sum_{m=1}^{N_s}\left[\mathbf{C}_{2Q-1}(\mathbf{s}_m)\sum_{n=1}^{N_r}\mathbf{h}_{k,n}^{ir} * \mathbf{h}_{n,m}^{r1}\right]$$

$$+ \sqrt{\alpha}\sum_{m=1}^{N_s}\left[\mathbf{C}_{2Q-1}(\mathbf{t}_m)\sum_{n=1}^{N_r}\mathbf{h}_{k,n}^{ir} * \mathbf{h}_{n,m}^{r2}\right]$$

$$+ \sum_{n=1}^{N_r}\mathbf{C}_Q(\mathbf{r}_n)\mathbf{h}_{k,n}^{ir} + \bar{\mathbf{v}}_{i,k}$$

$$= \sqrt{\alpha}\mathbf{\Phi}(\mathbf{s})\mathbf{d}_k^{i1} + \sqrt{\alpha}\mathbf{\Phi}(\mathbf{t})\mathbf{d}_k^{i2} + \mathbf{\Phi}(\mathbf{r})\mathbf{d}_k^{ir} + \bar{\mathbf{v}}_{i,k},$$

$$k = 1, \ldots, N_s \tag{6}$$

where $\mathbf{a} * \mathbf{b}$ denotes the linear convolution between vectors $\mathbf{a}$ and $\mathbf{b}$, and

$$\mathbf{\Phi}(\mathbf{s}) \triangleq [\mathbf{C}_{2Q-1}(\mathbf{s}_1), \mathbf{C}_{2Q-1}(\mathbf{s}_2), \ldots,$$
$$\mathbf{C}_{2Q-1}(\mathbf{s}_{N_s})] \in \mathcal{C}^{L \times (2Q-1)N_s} \tag{7}$$

$$\mathbf{\Phi}(\mathbf{t}) \triangleq [\mathbf{C}_{2Q-1}(\mathbf{t}_1), \mathbf{C}_{2Q-1}(\mathbf{t}_2), \ldots,$$
$$\mathbf{C}_{2Q-1}(\mathbf{t}_{N_s})] \in \mathcal{C}^{L \times (2Q-1)N_s} \tag{8}$$

$$\mathbf{\Phi}(\mathbf{r}) \triangleq [\mathbf{C}_Q(\mathbf{r}_1), \mathbf{C}_Q(\mathbf{r}_2), \ldots, \mathbf{C}_Q(\mathbf{r}_{N_r})] \in \mathcal{C}^{L \times QN_r} \tag{9}$$

$$\mathbf{d}_k^{i1} \triangleq \left[\left(\sum_{n=1}^{N_r}\mathbf{h}_{k,n}^{ir} * \mathbf{h}_{n,1}^{r1}\right)^T, \left(\sum_{n=1}^{N_r}\mathbf{h}_{k,n}^{ir} * \mathbf{h}_{n,2}^{r1}\right)^T, \ldots,\right.$$
$$\left.\left(\sum_{n=1}^{N_r}\mathbf{h}_{k,n}^{ir} * \mathbf{h}_{n,N_s}^{r1}\right)^T\right]^T \tag{10}$$

$$\mathbf{d}_k^{i2} \triangleq \left[\left(\sum_{n=1}^{N_r}\mathbf{h}_{k,n}^{ir} * \mathbf{h}_{n,1}^{r2}\right)^T, \left(\sum_{n=1}^{N_r}\mathbf{h}_{k,n}^{ir} * \mathbf{h}_{n,2}^{r2}\right)^T, \ldots,\right.$$
$$\left.\left(\sum_{n=1}^{N_r}\mathbf{h}_{k,n}^{ir} * \mathbf{h}_{n,N_s}^{r2}\right)^T\right]^T \tag{11}$$

$$\mathbf{d}_k^{ir} \triangleq \left[\left(\mathbf{h}_{k,1}^{ir}\right)^T, \left(\mathbf{h}_{k,2}^{ir}\right)^T, \ldots, \left(\mathbf{h}_{k,N_r}^{ir}\right)^T\right]^T. \tag{12}$$

Here $\mathbf{d}_k^{i1}$ in (10) and $\mathbf{d}_k^{i2}$ in (11) can be viewed as the compound channel from all antennas of node 1 and node 2 to the kth antenna at node i, respectively, and $\mathbf{d}_k^{ir}$ in (12) is the channel from all antennas of the relay node to the kth antennas at node i.

---

[1]In this paper, cyclic prefix is removed to facilitate the superimposed channel training algorithm. It is an interesting topic to combine the proposed approach and the channel estimation using the cyclic prefix, which may improve the accuracy of channel estimation.

By introducing

$$\mathbf{A} \triangleq \left[\sqrt{\alpha}\boldsymbol{\Phi}(\mathbf{s}), \sqrt{\alpha}\boldsymbol{\Phi}(\mathbf{t}), \boldsymbol{\Phi}(\mathbf{r})\right] \in \mathcal{C}^{L \times ((4Q-2)N_s + QN_r)} \quad (13)$$

$$\boldsymbol{\theta}_{i,k} \triangleq \left[\left(\mathbf{d}_k^{i1}\right)^T, \left(\mathbf{d}_k^{i2}\right)^T, \left(\mathbf{d}_k^{ir}\right)^T\right]^T, \quad k=1,\ldots,N_s \quad (14)$$

we can rewrite (6) as

$$\mathbf{y}_{i,k} = \mathbf{A}\boldsymbol{\theta}_{i,k} + \bar{\mathbf{v}}_{i,k}, \quad k = 1,\ldots,N_s. \quad (15)$$

Here $\boldsymbol{\theta}_{i,k}$ in (14) is the vector of unknowns that need to be estimated at node i.

Due to its simplicity, a linear estimator is applied at node i to estimate $\boldsymbol{\theta}_{i,k}$ as

$$\hat{\boldsymbol{\theta}}_{i,k} = \mathbf{W}_{i,k}^H \mathbf{y}_{i,k}, \quad k = 1,\ldots,N_s, \quad i = 1, 2 \quad (16)$$

where $\hat{\boldsymbol{\theta}}_{i,k}$ denotes an estimation of $\boldsymbol{\theta}_{i,k}$, $\mathbf{W}_{i,k}$ is the weight matrix of the linear receiver, and $(\cdot)^H$ denotes the matrix (vector) Hermitian transpose. As a linear estimator is used, we can see from (13) that the length of the training sequences should satisfy $L \geq (4Q - 2)N_s + QN_r$. Based on (15) and (16), the sum MSE of channel estimation at two nodes can be written as

$$\text{MSE} = \sum_{i=1}^{2}\sum_{k=1}^{N_s} \text{tr}\left(\text{E}\left[(\hat{\boldsymbol{\theta}}_{i,k} - \boldsymbol{\theta}_{i,k})(\hat{\boldsymbol{\theta}}_{i,k} - \boldsymbol{\theta}_{i,k})^H\right]\right)$$

$$= \sum_{i=1}^{2}\sum_{k=1}^{N_s} \text{tr}\left(\text{E}\left[\left(\mathbf{W}_{i,k}^H\mathbf{A} - \mathbf{I}_B\right)\mathbf{C}_x^{i,k}\left(\mathbf{W}_{i,k}^H\mathbf{A} - \mathbf{I}_B\right)^H\right.\right.$$

$$\left.\left. + \mathbf{W}_{i,k}^H\mathbf{C}_{\bar{v}}^{i,k}\mathbf{W}_{i,k}\right]\right) \quad (17)$$

where $\text{tr}(\cdot)$ denotes the matrix trace, $\mathbf{I}_n$ stands for the $n \times n$ identity matrix, $B \triangleq (4Q - 2)N_s + QN_r$, $\mathbf{C}_x^{i,k} = \text{E}[\boldsymbol{\theta}_{i,k}\boldsymbol{\theta}_{i,k}^H]$ is the covariance matrix of $\boldsymbol{\theta}_{i,k}$, and $\mathbf{C}_{\bar{v}}^{i,k} = \text{E}[\bar{\mathbf{v}}_{i,k}\bar{\mathbf{v}}_{i,k}^H]$ is the noise covariance matrix. Here $\text{E}[\cdot]$ stands for the statistical expectation.

From (5), we have

$$\mathbf{C}_{\bar{v}}^{i,k} = \left(\alpha\sum_{n=1}^{N_r}\sum_{j=1}^{Q}\sigma_{k,n,j}^{ir} + 1\right)\mathbf{I}_L, \quad i = 1, 2, \quad k = 1,\ldots,N_s$$

where $\sigma_{k,n,j}^{ir} = \text{E}[h_{k,n,j}^{ir}(h_{k,n,j}^{ir})^*]$ is the variance of $h_{k,n,j}^{ir}$, $j = 1,\ldots,Q$, and $(\cdot)^*$ denotes complex conjugate. Based on (10)–(12) and (14), we obtain that $\mathbf{C}_x^{i,k} = \text{bd}[\mathbf{C}_{i1}^k, \mathbf{C}_{i2}^k, \mathbf{C}_{ir}^k]$, where $\text{bd}[\cdot]$ represents a block diagonal matrix and

$$\mathbf{C}_{ij}^k = \text{E}\left[\mathbf{d}_k^{ij}\left(\mathbf{d}_k^{ij}\right)^H\right] = \text{bd}\left[\mathbf{C}_{k,1}^{ij},\ldots,\mathbf{C}_{k,N_s}^{ij}\right], \quad j = 1, 2 \quad (18)$$

$$\mathbf{C}_{ir}^k = \text{E}\left[\mathbf{d}_k^{ir}\left(\mathbf{d}_k^{ir}\right)^H\right] = \text{bd}\left[\mathbf{C}_{k,1}^{ir},\ldots,\mathbf{C}_{k,N_r}^{ir}\right]. \quad (19)$$

By introducing $\boldsymbol{\sigma}_{k,n}^{ir} = [\sigma_{k,n,1}^{ir},\ldots,\sigma_{k,n,Q}^{ir}]^T$ and $\boldsymbol{\sigma}_{n,m}^{rj} = [\sigma_{n,m,1}^{rj},\ldots,\sigma_{n,m,Q}^{rj}]^T$, where $\sigma_{n,m,p}^{rj} = \text{E}[h_{n,m,p}^{rj}(h_{n,m,p}^{rj})^*]$ is

the variance of $h_{n,m,p}^{rj}$, $j = 1, 2$, $p = 1,\ldots,Q$, we obtain that

$$\mathbf{C}_{k,m}^{ij} = \text{E}\left[\left(\sum_{n=1}^{N_r}\mathbf{h}_{k,n}^{ir} * \mathbf{h}_{n,m}^{rj}\right)\left(\sum_{n=1}^{N_r}\mathbf{h}_{k,n}^{ir} * \mathbf{h}_{n,m}^{rj}\right)^H\right]$$

$$= \sum_{n=1}^{N_r}\text{diag}\left[\boldsymbol{\sigma}_{k,n}^{ir} * \boldsymbol{\sigma}_{n,m}^{rj}\right], \quad j = 1, 2, \quad m = 1,\ldots,N_s$$

$$\mathbf{C}_{k,n}^{ir} = \text{E}\left[\mathbf{h}_{k,n}^{ir}\left(\mathbf{h}_{k,n}^{ir}\right)^H\right]$$

$$= \text{diag}\left[\sigma_{k,n,1}^{ir},\ldots,\sigma_{k,n,Q}^{ir}\right], \quad n = 1,\ldots,N_r.$$

Here $\text{diag}[\mathbf{x}]$ stands for a diagonal matrix taking $\mathbf{x}$ as the diagonal elements.

A. Structure of Optimal Training Sequences

The matrices $\mathbf{W}_{i,k}, i = 1, 2, k = 1,\ldots,N_s$ that minimize MSE in (17) are given by

$$\mathbf{W}_{i,k} = \left(\mathbf{A}\mathbf{C}_x^{i,k}\mathbf{A}^H + \mathbf{C}_{\bar{v}}^{i,k}\right)^{-1}\mathbf{A}\mathbf{C}_x^{i,k}, \quad i = 1, 2, \quad k = 1,\ldots,N_s \quad (20)$$

where $(\cdot)^{-1}$ denotes the matrix inversion. Substituting (20) back into (17), the MSE of channel estimation at both source nodes can be written as

$$\text{MSE} = \sum_{i=1}^{2}\sum_{k=1}^{N_s}\text{tr}\left(\left[\left(\mathbf{C}_x^{i,k}\right)^{-1} + \mathbf{A}^H\left(\mathbf{C}_{\bar{v}}^{i,k}\right)^{-1}\mathbf{A}\right]^{-1}\right). \quad (21)$$

The transmission power constraints at the source nodes are given by

$$\sum_{m=1}^{N_s}\mathbf{s}_m^H\mathbf{s}_m \leq P_1, \quad \sum_{m=1}^{N_s}\mathbf{t}_m^H\mathbf{t}_m \leq P_2 \quad (22)$$

where P_1 and P_2 are the transmission power available at source nodes 1 and 2, respectively. From (1) and (2), the transmission power constraint at the relay node is given by

$$\sum_{n=1}^{N_r}\text{E}\left[\text{tr}\left(\mathbf{x}_{r,n}\mathbf{x}_{r,n}^H\right)\right]$$

$$= \sum_{n=1}^{N_r}\left(\alpha\text{tr}\left(\sum_{m=1}^{N_s}\left(\mathbf{C}_Q(\mathbf{s}_m)\mathbf{D}_{n,m}^{r1}\mathbf{C}_Q^H(\mathbf{s}_m)\right.\right.\right.$$

$$\left.\left.\left. + \mathbf{C}_Q(\mathbf{t}_m)\mathbf{D}_{n,m}^{r2}\mathbf{C}_Q^H(\mathbf{t}_m)\right) + \mathbf{I}_L\right) + \mathbf{r}_n^H\mathbf{r}_n\right) \leq P_r \quad (23)$$

where $\mathbf{D}_{n,m}^{ri} \triangleq \text{diag}[\sigma_{n,m,1}^{ri},\ldots,\sigma_{n,m,Q}^{ri}]$, $i = 1, 2$, and P_r is the transmission power available at the relay node. It can be seen from (23) that the feasible region of α depends on P_r as $0 < \alpha < (P_r - \sum_{n=1}^{N_r}\mathbf{r}_n^H\mathbf{r}_n)/\Omega$, where

$$\Omega \triangleq \sum_{n=1}^{N_r}\text{tr}\left(\sum_{m=1}^{N_s}\left(\mathbf{C}_Q(\mathbf{s}_m)\mathbf{D}_{n,m}^{r1}\mathbf{C}_Q^H(\mathbf{s}_m)\right.\right.$$

$$\left.\left. + \mathbf{C}_Q(\mathbf{t}_m)\mathbf{D}_{n,m}^{r2}\mathbf{C}_Q^H(\mathbf{t}_m)\right) + \mathbf{I}_L\right).$$

From (21)–(23), the optimal training sequences and the optimal α design problem can be written as

$$\min_{\mathbf{S},\mathbf{T},\mathbf{R},\alpha>0} \sum_{i=1}^{2} \sum_{k=1}^{N_s} \mathrm{tr}\left(\left[\left(\mathbf{C}_x^{i,k}\right)^{-1} + \mathbf{A}^H \left(\mathbf{C}_{\tilde{v}}^{i,k}\right)^{-1} \mathbf{A}\right]^{-1}\right)$$

(24)

$$\text{s.t. } \sum_{m=1}^{N_s} \mathbf{s}_m^H \mathbf{s}_m \leq P_1$$

(25)

$$\sum_{m=1}^{N_s} \mathbf{t}_m^H \mathbf{t}_m \leq P_2$$

(26)

$$\sum_{n=1}^{N_r}\left(\alpha\mathrm{tr}\left(\sum_{m=1}^{N_s}\left(\mathbf{C}_Q(\mathbf{s}_m)\mathbf{D}_{n,m}^{r1}\mathbf{C}_Q^H(\mathbf{s}_m)\right.\right.\right.$$
$$\left.\left.\left.+\mathbf{C}_Q(\mathbf{t}_m)\mathbf{D}_{n,m}^{r2}\mathbf{C}_Q^H(\mathbf{t}_m)\right)+\mathbf{I}_L\right)+\mathbf{r}_n^H\mathbf{r}_n\right)\leq P_r.$$

(27)

The following theorem establishes the optimal structure of $\mathbf{S}$, $\mathbf{T}$, and $\mathbf{R}$ as the solution to the problem (24)–(27).

Theorem 1: The optimal training matrices $\mathbf{S}$, $\mathbf{T}$, and $\mathbf{R}$ satisfy the following equations for all $m, n = 1, \ldots, N_s$, and $p = 1, \ldots, N_r$

$$\mathbf{C}_{2Q-1}^H(\mathbf{s}_m)\mathbf{C}_{2Q-1}(\mathbf{s}_m) = \beta_m\mathbf{I}_{2Q-1},$$
$$\mathbf{C}_{2Q-1}^H(\mathbf{t}_n)\mathbf{C}_{2Q-1}(\mathbf{t}_n) = \gamma_n\mathbf{I}_{2Q-1}, \quad \mathbf{C}_Q^H(\mathbf{r}_p)\mathbf{C}_Q(\mathbf{r}_p) = \delta_p\mathbf{I}_Q$$

(28)

$$\mathbf{C}_{2Q-1}^H(\mathbf{s}_m)\mathbf{C}_{2Q-1}(\mathbf{t}_n) = \mathbf{0},$$
$$\mathbf{C}_{2Q-1}^H(\mathbf{s}_m)\mathbf{C}_Q(\mathbf{r}_p) = \mathbf{0}, \quad \mathbf{C}_{2Q-1}^H(\mathbf{t}_n)\mathbf{C}_Q(\mathbf{r}_p) = \mathbf{0}$$

(29)

where $\beta_m = \mathbf{s}_m^H\mathbf{s}_m$, $\gamma_n = \mathbf{t}_n^H\mathbf{t}_n$, and $\delta_p = \mathbf{r}_p^H\mathbf{r}_p$.

Proof: See Appendix A. □

It is worth noting that the training matrices $\mathbf{S}$, $\mathbf{T}$, and $\mathbf{R}$ satisfying (28) and (29) are not unique in general. Indeed, we are not particularly interested in a unique solution of the problem. The minimum MSE of channel estimation is achieved as long as the training matrices satisfy (28) and (29). One example of achieving (28) and (29) is given below

$$\mathbf{s}_1 = \mathbf{F}\tilde{\mathbf{s}}_1, \ |\tilde{s}_{1,i}| = \sqrt{\frac{\beta_1}{L}}, \quad i = 1, \ldots, L$$

$$\mathbf{s}_m = \mathbf{F}\tilde{\mathbf{s}}_m, \ \tilde{s}_{m,i} = \sqrt{\frac{\beta_m}{\beta_1}}\tilde{s}_{1,i}e^{\frac{j2\pi(i-1)(2Q-1)(m-1)}{L}},$$
$$i = 1, \ldots, L, \quad m = 2, \ldots, N_s$$

$$\mathbf{t}_m = \mathbf{F}\tilde{\mathbf{t}}_m, \ \tilde{t}_{m,i} = \sqrt{\frac{\gamma_m}{\beta_1}}\tilde{s}_{1,i}e^{\frac{j2\pi(i-1)(2Q-1)(N_s-1+m)}{L}},$$
$$i = 1, \ldots, L, \quad m = 1, \ldots, N_s$$

$$\mathbf{r}_n = \mathbf{F}\tilde{\mathbf{r}}_n, \ \tilde{r}_{n,i} = \sqrt{\frac{\delta_n}{\beta_1}}\tilde{s}_{1,i}e^{\frac{j2\pi(i-1)[(2Q-1)(2N_s-1)+Q(n-1)]}{L}},$$
$$i = 1, \ldots, L, \quad n = 1, \ldots, N_r$$

where $|\cdot|$ denotes the modulus of a complex number, $j = \sqrt{-1}$, and $\mathbf{F}$ is an $L \times L$ normalized FFT matrix with $[\mathbf{F}]_{m,n} = (1/\sqrt{L})e^{-j2\pi(m-1)(n-1)/L}$. The training matrices shown above have the advantage that they are easy to implement as the elements of $\tilde{\mathbf{s}}_m$ (also $\tilde{\mathbf{t}}_m$ and $\tilde{\mathbf{r}}_n$) have a constant magnitude.

B. Optimal Power Loading

Applying Theorem 1, the MSE function in (21) can be written as

$$\text{MSE} = \sum_{i=1}^{2} \sum_{k=1}^{N_s} \mathrm{tr}\left(\sum_{m=1}^{N_s}\left[\left(\mathbf{C}_{k,m}^{i1}\right)^{-1} + \alpha\beta_m\eta_{i,k}\mathbf{I}_{2Q-1}\right]^{-1}\right.$$
$$+ \sum_{m=1}^{N_s}\left[\left(\mathbf{C}_{k,m}^{i2}\right)^{-1} + \alpha\gamma_m\eta_{i,k}\mathbf{I}_{2Q-1}\right]^{-1}$$
$$\left.+ \sum_{n=1}^{N_r}\left[\left(\mathbf{C}_{k,n}^{ir}\right)^{-1} + \delta_n\eta_{i,k}\mathbf{I}_Q\right]^{-1}\right) \quad (30)$$

where $\eta_{i,k}$ is defined in (56). Let us denote $c_{k,m,q}^{ij} \triangleq [(\mathbf{C}_{k,m}^{ij})^{-1}]_{q,q}$, $c_{k,n,p}^{ir} \triangleq [(\mathbf{C}_{k,n}^{ir})^{-1}]_{p,p}$, and $\kappa_{i,m} \triangleq \sum_{n=1}^{N_r}\sum_{q=1}^{Q}\sigma_{n,m,q}^{ri}$, $i = 1, 2$. The problem (24)–(27) with matrix variables can be equivalently rewritten as the following problem in scalar variables

$$\min_{\beta,\gamma,\delta,\alpha} \sum_{m=1}^{N_s}\sum_{k=1}^{N_s}\sum_{q=1}^{2Q-1}\sum_{i=1}^{2}\left(\frac{1}{c_{k,m,q}^{i1}+\alpha\beta_m\eta_{i,k}} + \frac{1}{c_{k,m,q}^{i2}+\alpha\gamma_m\eta_{i,k}}\right)$$
$$+ \sum_{n=1}^{N_r}\sum_{k=1}^{N_s}\sum_{p=1}^{Q}\sum_{i=1}^{2}\frac{1}{c_{k,n,p}^{ir} + \delta_n\eta_{i,k}} \quad (31)$$

$$\text{s.t. } \sum_{m=1}^{N_s} \beta_m \leq P_1$$

(32)

$$\sum_{m=1}^{N_s} \gamma_m \leq P_2$$

(33)

$$\alpha\left(\sum_{m=1}^{N_s}\kappa_{1,m}\beta_m + \sum_{m=1}^{N_s}\kappa_{2,m}\gamma_m\right) + \sum_{n=1}^{N_r}\delta_n + \alpha LN_r \leq P_r$$

(34)

$$\alpha > 0, \beta_m \geq 0, \gamma_m \geq 0, m = 1, \ldots, N_s, \delta_n \geq 0, n = 1, \ldots, N_r$$

(35)

where $\boldsymbol{\beta} \triangleq [\beta_1, \ldots, \beta_{N_s}]^T$, $\boldsymbol{\gamma} \triangleq [\gamma_1, \ldots, \gamma_{N_s}]^T$, and $\boldsymbol{\delta} \triangleq [\delta_1, \ldots, \delta_{N_r}]^T$.

Given that $c_{k,m,q}^{i1}$, $c_{k,m,q}^{i2}$, $c_{k,m,q}^{ir}$, and $\eta_{i,k}$ are known variables with fixed α, it can be observed that the fractions in the objective function (31) are monotonically decreasing and convex functions with respect to β_m, γ_m, and δ_n. Moreover, when α is fixed, the constraints in (32)–(35) are linear inequality constraints. Therefore, with fixed α, the problem (31)–(35) with

respect to β_m, γ_m, and δ_n is a convex optimization problem where the optimal β_m, γ_m, and δ_n can be efficiently obtained through the Karush-Kuhn-Tucker (KKT) optimality conditions [20] of the problem (31)–(35). The gradient conditions are given by

$$\sum_{k=1}^{N_s} \sum_{q=1}^{2Q-1} \sum_{i=1}^{2} \frac{\alpha \eta_{i,k}}{\left(c_{k,m,q}^{i1} + \alpha \beta_m \eta_{i,k} \right)^2}$$
$$= \mu_1 + \mu_3 \alpha \kappa_{1,m}, \ m = 1, \ldots, N_s \quad (36)$$

$$\sum_{k=1}^{N_s} \sum_{q=1}^{2Q-1} \sum_{i=1}^{2} \frac{\alpha \eta_{i,k}}{\left(c_{k,m,q}^{i2} + \alpha \gamma_m \eta_{i,k} \right)^2}$$
$$= \mu_2 + \mu_3 \alpha \kappa_{2,m}, \ m = 1, \ldots, N_s \quad (37)$$

$$\sum_{k=1}^{N_s} \sum_{p=1}^{Q} \sum_{i=1}^{2} \frac{\eta_{i,k}}{\left(c_{k,n,p}^{ir} + \delta_n \eta_{i,k} \right)^2}$$
$$= \mu_3, \ n = 1, \ldots, N_r \quad (38)$$

where $\mu_i \geq 0$, $i = 1, 2, 3$, are Lagrange multipliers such that the complementary slackness conditions given by

$$\mu_1 \left(P_1 - \sum_{m=1}^{N_s} \beta_m \right) = 0 \quad (39)$$

$$\mu_2 \left(P_2 - \sum_{m=1}^{N_s} \gamma_m \right) = 0 \quad (40)$$

$$\mu_3 \left(P_r - \alpha N L - \sum_{n=1}^{N_r} \delta_n - \alpha \sum_{m=1}^{N_s} \kappa_{1,m} \beta_m - \alpha \sum_{m=1}^{N_s} \kappa_{2,m} \gamma_m \right) = 0 \quad (41)$$

are satisfied.

When α and μ_i, $i = 1, 2, 3$, are fixed, the non-negative β_m, γ_m, $m = 1, \ldots, N_s$, and δ_n, $n = 1, \ldots, N_r$, can be found by using the bi-section search, as the left-hand-side (LHS) of (36), (37), and (38) are monotonically decreasing functions of β_m, γ_m, and δ_n, respectively. An outer bi-section search is applied to find the optimal μ_i, $i = 1, 2, 3$, since the LHS of (32) and (33) are increasing functions of β_m and γ_m, respectively, and the LHS of (34) is an increasing function of β_m, γ_m, and δ_n. Moreover, in (36), β_m is a monotonically decreasing function of μ_1 and μ_3, γ_m is monotonically decreasing with respect to μ_2 and μ_3 in (37), while in (38), δ_n is a monotonically decreasing function of μ_3.

When α is not fixed, i.e., α is an optimization variable, the problem (31)–(35) as a whole is not a convex optimization problem. However, the following theorem states that (31) is a unimodal function of α.

Theorem 2: The objective function (31) subjecting to (32)–(35) is a unimodal (quasi-convex) function with respect to α.
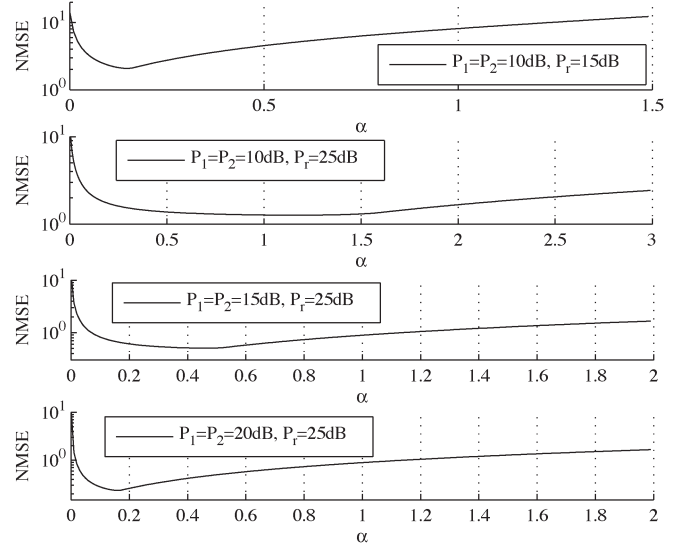

Fig. 2. NMSE versus α for different $P_1 = P_2$ and P_r with $N = 2$ and $Q = 4$.

TABLE I
PROCEDURE OF APPLYING THE GOLDEN SECTION SEARCH (GSS) TO FIND THE OPTIMAL α IN THE PROBLEM (31)–(35)

1) Set a feasible bound $[a, b]$ on α.
2) Define $c_1 = (\phi - 1)a + (2 - \phi)b$ and $c_2 = (2 - \phi)a + (\phi - 1)b$.
3) Solve the problem (31)-(35) for $\alpha = c_1$;
   Compute the MSE value defined in (31), $f_{MSE}(c_1)$ for $\alpha = c_1$.
4) Repeat Step 3 for $\alpha = c_2$.
5) If $f_{MSE}(c_1) < f_{MSE}(c_2)$, then assign $b = c_2$.
   Otherwise, assign $a = c_1$.
6) If $|b - a| \leq \varepsilon$, then end.
   Otherwise, go to step 2.

Proof: See Appendix B. □

To verify Theorem 2, a plot of the MSE value over a range of feasible values of α is shown in Fig. 2. We consider the case where all nodes have the same number of antennas, i.e., $N_s = N_r = N = 2$, and $Q = 4$. Fig. 2 shows the normalized MSE (NMSE) versus α for different $P_1 = P_2$ and P_r. Note that the NMSE is obtained by dividing (31) with $L = (5Q - 2)N$. It can be seen from Fig. 2 that (31) is a unimodal function of α. For a unimodal function, the minimum value can be efficiently found by the golden section search (GSS) [21] technique. Hence, the optimal α for the problem (31)–(35) can be found by applying the GSS technique as described in Table I, where ε is a positive constant close to 0, and $\phi > 0$ is the reduction factor. It is shown in [21] that the optimal $\phi = 1.618$, also known as the golden ratio. It can be seen from Fig. 2 that the optimal value of α varies with P_1, P_2, and P_r. For fixed P_1 and P_2, the optimal α has a larger value when P_r increases. For a given P_r, the optimal value of α decreases when P_1 and P_2 increases.

Since at each iteration, the GSS method reduces the interval containing the optimal α to 0.618 times of the interval at the preceding iteration, the length of the interval of uncertainty after the nth iteration is $\Gamma_n = (0.618)^n \Gamma_0$, where Γ_0 is the length of the initial feasible interval [21]. Therefore, the complexity of the GSS method depends on the number of iterations, which is determined by the desired accuracy.

C. Retrieving the Multipath Channel Vectors

Based on (12) and (14), the second-hop channels $\mathbf{h}_{k,n}^{ir}$, $i = 1, 2$ can be directly obtained from $\hat{\boldsymbol{\theta}}_{i,k}$. The first-hop channels $\mathbf{h}_{n,m}^{r1}$ and $\mathbf{h}_{n,m}^{r2}$ can be estimated based on $\hat{\boldsymbol{\theta}}_{i,k}$ as follows. Since $\mathbf{h}_{k,n}^{ir} * \mathbf{h}_{n,m}^{r1} = \mathbf{T}(\mathbf{h}_{k,n}^{ir})\mathbf{h}_{n,m}^{r1}$, where $\mathbf{T}(\mathbf{h})$ stands for a $(2Q-1) \times Q$ circulant matrix taking $[\mathbf{h}^T, \mathbf{0}_{1 \times (Q-1)}]^T$ as its first column, we have

$$\sum_{n=1}^{N_r} \mathbf{h}_{k,n}^{ir} * \mathbf{h}_{n,m}^{r1} = \sum_{n=1}^{N_r} \mathbf{T}(\mathbf{h}_{k,n}^{ir})\mathbf{h}_{n,m}^{r1} = \mathbf{d}_{k,m}^{i1}, \quad k, m = 1, \ldots, N_s \tag{42}$$

$$\sum_{n=1}^{N_r} \mathbf{h}_{k,n}^{ir} * \mathbf{h}_{n,m}^{r2} = \sum_{n=1}^{N_r} \mathbf{T}(\mathbf{h}_{k,n}^{ir})\mathbf{h}_{n,m}^{r2} = \mathbf{d}_{k,m}^{i2}, \quad k, m = 1, \ldots, N_s. \tag{43}$$

Equations (42) and (43) can be represented in matrix form as

$$\boldsymbol{\Psi}_{ir}\mathbf{h}_m^{r1} = \mathbf{e}_m^{i1}, \quad m = 1, \ldots, N_s \tag{44}$$
$$\boldsymbol{\Psi}_{ir}\mathbf{h}_m^{r2} = \mathbf{e}_m^{i2}, \quad m = 1, \ldots, N_s \tag{45}$$

where

$$\boldsymbol{\Psi}_{ir} \triangleq \begin{bmatrix} \mathbf{T}(\mathbf{h}_{1,1}^{ir}), & \cdots & , \mathbf{T}(\mathbf{h}_{1,N_r}^{ir}) \\ \vdots & \ddots & \vdots \\ \mathbf{T}(\mathbf{h}_{N_s,1}^{ir}), & \cdots & , \mathbf{T}(\mathbf{h}_{N_s,N_r}^{ir}) \end{bmatrix},$$
$$\mathbf{h}_m^{ri} \triangleq \begin{bmatrix} \mathbf{h}_{1,m}^{ri} \\ \vdots \\ \mathbf{h}_{N_r,m}^{ri} \end{bmatrix}, \quad i = 1, 2,$$
$$\mathbf{e}_m^{ij} \triangleq \begin{bmatrix} \mathbf{d}_{1,m}^{ij} \\ \vdots \\ \mathbf{d}_{N_s,m}^{ij} \end{bmatrix}, \quad j = 1, 2. \tag{46}$$

In the following, we develop an LMMSE estimator to retrieve the first-hop multipath channel vectors $\{\mathbf{h}_{n,m}^{ri}\}$. Taking into account the estimation errors in $\mathbf{h}_{k,m}^{ir}$ and $\mathbf{d}_{n,m}^{ij}$, we have

$$\mathbf{h}_{k,m}^{ir} = \hat{\mathbf{h}}_{k,m}^{ir} + \boldsymbol{\kappa}_{k,m}^{ir}, \quad \mathbf{d}_{n,m}^{ij} = \hat{\mathbf{d}}_{n,m}^{ij} + \boldsymbol{\iota}_{n,m}^{ij} \tag{47}$$

where $\hat{\mathbf{h}}_{k,m}^{ir}$ and $\hat{\mathbf{d}}_{n,m}^{ij}$ are the estimates of $\mathbf{h}_{k,m}^{ir}$ and $\mathbf{d}_{n,m}^{ij}$, respectively, obtained from $\hat{\boldsymbol{\theta}}_{i,k}$, and $\boldsymbol{\kappa}_{k,m}^{ir}$ and $\boldsymbol{\iota}_{n,m}^{ij}$ are the estimation error vectors. Substituting (47) back into (44) and (45), we have

$$(\hat{\boldsymbol{\Psi}}_{ir} + \boldsymbol{\Delta}_i)\mathbf{h}_m^{rj} = \hat{\mathbf{e}}_m^{ij} + \mathbf{g}_m^{ij}, \quad m = 1, \ldots, N_s, \; j = 1, 2 \tag{48}$$

where

$$\hat{\boldsymbol{\Psi}}_{ir} \triangleq \begin{bmatrix} \mathbf{T}(\hat{\mathbf{h}}_{1,1}^{ir}), & \cdots & , \mathbf{T}(\hat{\mathbf{h}}_{1,N_r}^{ir}) \\ \vdots & \ddots & \vdots \\ \mathbf{T}(\hat{\mathbf{h}}_{N_s,1}^{ir}), & \cdots & , \mathbf{T}(\hat{\mathbf{h}}_{N_s,N_r}^{ir}) \end{bmatrix}, \; \hat{\mathbf{e}}_m^{ij} \triangleq \begin{bmatrix} \hat{\mathbf{d}}_{1,m}^{ij} \\ \vdots \\ \hat{\mathbf{d}}_{N_s,m}^{ij} \end{bmatrix}$$
$$\boldsymbol{\Delta}_i \triangleq \begin{bmatrix} \mathbf{T}(\boldsymbol{\kappa}_{1,1}^{ir}), & \cdots & , \mathbf{T}(\boldsymbol{\kappa}_{1,N_r}^{ir}) \\ \vdots & \ddots & \vdots \\ \mathbf{T}(\boldsymbol{\kappa}_{N_s,1}^{ir}), & \cdots & , \mathbf{T}(\boldsymbol{\kappa}_{N_s,N_r}^{ir}) \end{bmatrix}, \; \mathbf{g}_m^{ij} \triangleq \begin{bmatrix} \boldsymbol{\iota}_{1,m}^{ij} \\ \vdots \\ \boldsymbol{\iota}_{N_s,m}^{ij} \end{bmatrix}.$$

We can rewrite (48) as

$$\hat{\mathbf{e}}_m^{ij} = \hat{\boldsymbol{\Psi}}_{ir}\mathbf{h}_m^{rj} + \boldsymbol{\varepsilon}_m^{ij}, \quad m = 1, \ldots, N_s, \; j = 1, 2 \tag{49}$$

where $\boldsymbol{\varepsilon}_m^{ij}$ is the equivalent estimation error vector given by

$$\boldsymbol{\varepsilon}_m^{ij} \triangleq \boldsymbol{\Delta}_i\mathbf{h}_m^{rj} - \mathbf{g}_m^{ij}. \tag{50}$$

Using a linear estimator to estimate $\mathbf{h}_m^{rj}$ at node i, we have

$$\hat{\mathbf{h}}_m^{rj} = \mathbf{V}_{ijm}^H \hat{\mathbf{e}}_m^{ij}, \quad m = 1, \ldots, N_s, \; j = 1, 2 \tag{51}$$

where $\mathbf{V}_{ijm}$ is the weight matrix of the LMMSE estimator at node i. From (49) and (51), the sum MSE of the first-hop channel estimation at node i is given by

$$\mathrm{MSE}_i = \sum_{j=1}^{2} \sum_{m=1}^{N_s} \mathrm{tr}\left(\mathrm{E}\left[\left(\hat{\mathbf{h}}_m^{rj} - \mathbf{h}_m^{rj}\right)\left(\hat{\mathbf{h}}_m^{rj} - \mathbf{h}_m^{rj}\right)^H \right] \right)$$

$$= \sum_{j=1}^{2} \sum_{m=1}^{N_s} \mathrm{tr}\left(\mathrm{E}\left[\left(\mathbf{V}_{ijm}^H \boldsymbol{\Psi}_{ir} - \mathbf{I}_{N_r Q}\right) \mathbf{R}_{\mathbf{h}_m^{rj}} \right.\right.$$
$$\left.\left. \times \left(\mathbf{V}_{ijm}^H \boldsymbol{\Psi}_{ir} - \mathbf{I}_{N_r Q}\right)^H + \mathbf{V}_{ijm}^H \mathbf{R}_{\boldsymbol{\varepsilon}_m^{ij}} \mathbf{V}_{ijm} \right]\right) \tag{52}$$

where $\mathbf{R}_{\mathbf{h}_m^{rj}} \triangleq \mathrm{E}[\mathbf{h}_m^{rj}(\mathbf{h}_m^{rj})^H]$ is the covariance matrix of $\mathbf{h}_m^{rj}$ and $\mathbf{R}_{\boldsymbol{\varepsilon}_m^{ij}} \triangleq \mathrm{E}[\boldsymbol{\varepsilon}_m^{ij}(\boldsymbol{\varepsilon}_m^{ij})^H]$ is the estimation error covariance matrix. From (46), we have

$$\mathbf{R}_{\mathbf{h}_m^{rj}} = \mathrm{bd}\left[\mathbf{D}_{1,m}^{rj}, \ldots, \mathbf{D}_{N_r,m}^{rj}\right], \quad m = 1, \ldots, N_s, \; j = 1, 2$$

where $\mathbf{D}_{n,m}^{rj} \triangleq \mathrm{diag}[\sigma_{n,m,1}^{rj}, \ldots, \sigma_{n,m,Q}^{rj}]$.

The estimation error covariance matrix $\mathbf{R}_{\boldsymbol{\varepsilon}_m^{ij}}$ is obtained from (50) as

$$\mathbf{R}_{\boldsymbol{\varepsilon}_m^{ij}} = \mathrm{E}\left[\left(\boldsymbol{\Delta}_i\mathbf{h}_m^{rj} - \mathbf{g}_m^{ij}\right)\left(\boldsymbol{\Delta}_i\mathbf{h}_m^{rj} - \mathbf{g}_m^{ij}\right)^H \right]$$

$$= \mathrm{E}\left[\boldsymbol{\Delta}_i\mathbf{R}_{\mathbf{h}_m^{rj}}\boldsymbol{\Delta}_i^H\right] + \mathrm{E}\left[\mathbf{g}_m^{ij}\left(\mathbf{g}_m^{ij}\right)^H\right]$$

$$m = 1, \ldots, N_s, \; j = 1, 2.$$

Due to the circulant structure of $\boldsymbol{\Delta}_i$ and the fact that $\mathbf{R}_{\mathbf{h}_m^{rj}}$ is a diagonal matrix, we have

$$\mathbf{R}_{\boldsymbol{\varepsilon}_m^{ij}} = \mathrm{bd}\left[\mathbf{R}_{i,j}^{1,m}, \ldots, \mathbf{R}_{i,j}^{N_s,m}\right] + \mathbf{R}_{\mathbf{g}_m^{ij}}$$

$$m = 1, \ldots, N_s, \; j = 1, 2$$

where $\mathbf{R}_{\mathbf{g}_m^{ij}} = \mathrm{E}[\mathbf{g}_m^{ij}(\mathbf{g}_m^{ij})^H]$ can be obtained from (30) and $\mathbf{R}_{i,j}^{k,m} = \sum_{n=1}^{Nr} \mathrm{diag}(\boldsymbol{\sigma}_{n,m}^{rj} * \mathbf{d}_{\boldsymbol{\kappa}_{k,n}^i})$. Here $\boldsymbol{\sigma}_{n,m}^{rj}$ is defined in the line after (19) and $\mathbf{d}_{\boldsymbol{\kappa}_{k,n}^i}$ contains the diagonal elements of $\mathbf{R}_{\boldsymbol{\kappa}_{k,n}^i} \triangleq \mathrm{E}[\boldsymbol{\kappa}_{k,n}^{ir}(\boldsymbol{\kappa}_{k,n}^{ir})^H]$, which can be obtained from the MSE expression (30).

The weight matrices $\mathbf{V}_{ijm}, j = 1, 2, m = 1, \ldots, N_s$, that minimize MSE_i in (52) are given by

$$\mathbf{V}_{ijm} = \left(\boldsymbol{\Psi}_{ir}\mathbf{R}_{\mathbf{h}_m^{rj}}\boldsymbol{\Psi}_{ir}^H + \mathbf{R}_{\boldsymbol{\varepsilon}_m^{ij}}\right)^{-1}\boldsymbol{\Psi}_{ir}\mathbf{R}_{\mathbf{h}_m^{rj}}. \tag{53}$$
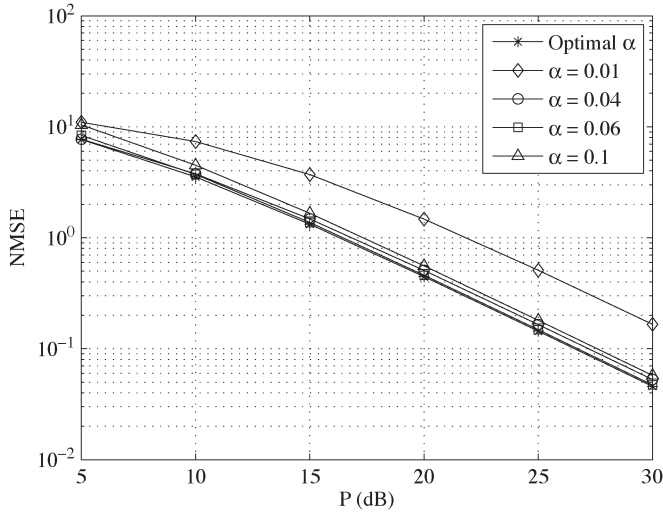
Fig. 3.   Example 1: NMSE versus P for different α with $N = 2$ and $Q = 4$.

Fig. 4.   Example 2: NMSE versus P for different N with $Q = 6$.

Substituting (53) back into (52), we obtain the MSE of channel estimation at node i as

$$\text{MSE}_i = \sum_{j=1}^{2} \sum_{m=1}^{N_s} \text{tr}\left(\left[\left(\mathbf{R}_{\mathbf{h}_m^{rj}}\right)^{-1} + \mathbf{\Psi}_{ir}^H \left(\mathbf{R}_{\boldsymbol{\varepsilon}_m^{ij}}\right)^{-1} \mathbf{\Psi}_{ir}\right]^{-1}\right). \tag{54}$$

It can be seen from (54) that the MSE of the first-hop channel estimation depends on the covariance matrix of the second-hop channel estimation error $\mathbf{R}_{\boldsymbol{\varepsilon}_m^{ij}}$. When the MSE of the second-hop channel estimation increases, the MSE of the first-hop channel estimation also increases.

IV. NUMERICAL EXAMPLES

In this section, we study the performance of the proposed superimposed channel training algorithm for two-way MIMO relay systems operating in frequency-selective fading environments through numerical simulations. We consider a three-node two-way MIMO relay system where all nodes are equipped with the same number of antennas, i.e., $N_s = N_r = N$. For simplicity, we assume that all channel taps have unit variances. We use the shortest length of training sequence possible with $L = (5Q - 2)N$. For all scenarios, the normalized MSE (NMSE) of channel estimation at nodes 1 and 2 are computed.

For the first three simulation examples, we assume that all nodes have the same transmission power $P_i = P$, $i = 1, 2, r$. In the first example, we investigate the performance of the superimposed channel training algorithm for different α. Fig. 3 shows the NMSE of the proposed algorithm versus P with different α when $N = 2$ and $Q = 4$. The optimal α curve is obtained by applying the GSS technique to the proposed superimposed channel training algorithm to obtain the optimal α for different P. It can be observed from Fig. 3 that the optimal α curve consistently has the lowest MSE level for all P. This proves that the GSS technique is able to obtain the optimal α at different P efficiently.

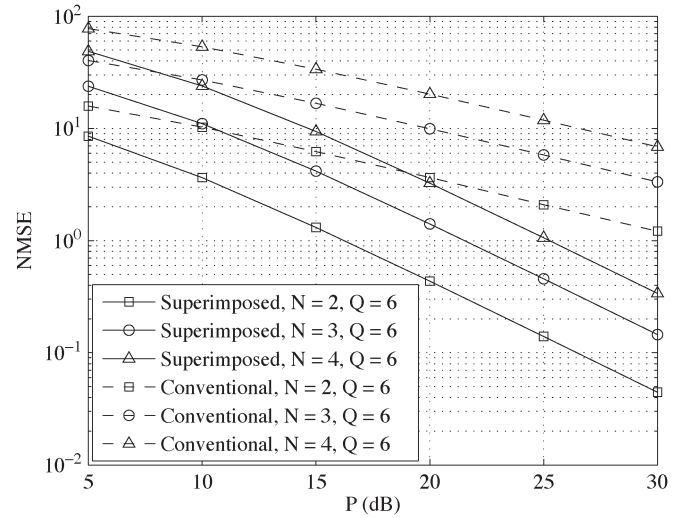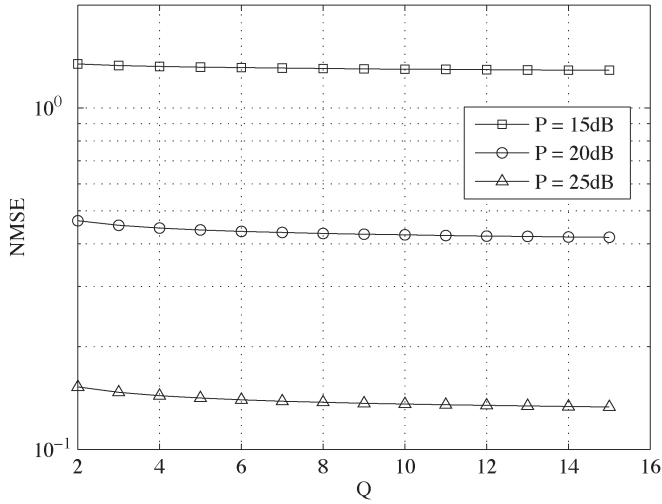Interestingly, we notice from Fig. 3 that the optimal α varies with respect to P, indicating that using a constant α is strictly

suboptimal. Although the NMSE with $\alpha = 0.06$ is close to the NMSE using the optimal α for P between 10 dB and 30 dB, $\alpha = 0.06$ yields a higher NMSE than $\alpha = 0.04$ at $P = 5$ dB. Moreover, for other simulation examples (e.g. different N and Q), the NMSE with $\alpha = 0.06$ might not be close to the NMSE using the optimal α. In practical systems, a table containing the value of the optimal α at different P, N, and Q can be constructed for reference.

In the second example, we study the performance of the proposed superimposed channel training algorithm when the optimal α is used under different simulation parameters. We compare the proposed algorithm with the conventional two-stage MMSE channel estimation algorithm, where the second-hop channel matrices are estimated at the first stage by using the training sequence sent from the relay node, and the first-hop channel matrices are estimated at the second stage by exploiting the training signals sent from the source nodes [13]. Fig. 4 demonstrates the NMSE performance of both methods versus P for different N and $Q = 6$. As expected, when the number of antennas increases, the NMSE of channel estimation at both sides also increases as there are more unknowns to be estimated. It can also be seen from Fig. 4 that the performance of the proposed algorithm is always better than the conventional two-stage channel estimation method, especially at high power levels.

Fig. 5 demonstrates the NMSE performance of the algorithm proposed in Section III-C which retrieves the individual CSI $\{\mathbf{h}_{k,n}^{ir}\}$ and $\{\mathbf{h}_{n,m}^{ri}\}$. It can be observed that the NMSE performance for the estimation of $\{\mathbf{h}_{k,n}^{ir}\}$ is always better than that for the estimation of $\{\mathbf{h}_{n,m}^{ri}\}$, as the estimation of $\{\mathbf{h}_{n,m}^{ri}\}$ depends on the estimation of $\{\mathbf{h}_{k,n}^{ir}\}$.

In the third example, the effect of the number of multipath Q on the performance of the proposed superimposed channel training algorithm is investigated. The results are shown in Fig. 6 for the case of $N = 2$. It can be seen that the NMSE performance of channel estimation improves when Q increases, as all channel taps are set to have unit variance. It can also be seen from Fig. 6 that such improvement diminishes when Q becomes larger.
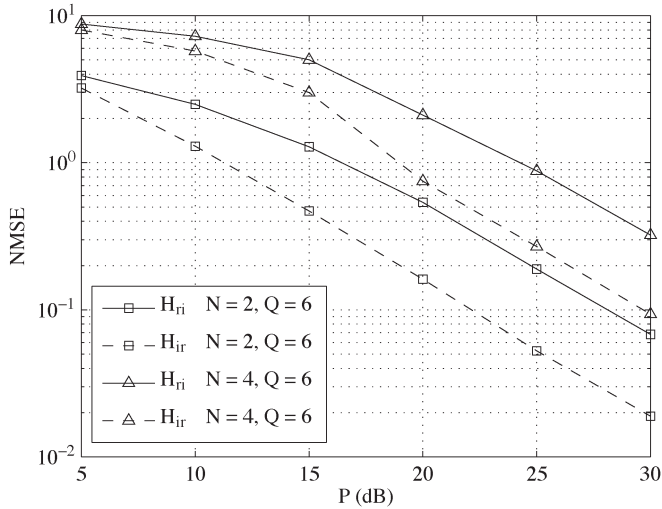
Fig. 5. Example 2: Individual channel NMSE versus P for different N with $Q = 6$.

Fig. 7. Example 4: NMSE versus P_s for different P_r with $N = 2$ and $Q = 6$.

APPENDIX A
PROOF OF THEOREM 1

The MSE in (24) can be rewritten as

$$
\text{MSE} = \sum_{i=1}^{2} \sum_{k=1}^{N_s} \text{tr} \left(\left[\begin{pmatrix} \mathbf{C}_{i1}^{k} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_{i2}^{k} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{C}_{ir}^{k} \end{pmatrix}^{-1} \right. \right.
$$
$$
\left. \left. + \eta_{i,k} \begin{pmatrix} \sqrt{\alpha} \mathbf{\Phi}^{H}(\mathbf{s}) \\ \sqrt{\alpha} \mathbf{\Phi}^{H}(\mathbf{t}) \\ \mathbf{\Phi}^{H}(\mathbf{r}) \end{pmatrix} \left(\sqrt{\alpha} \mathbf{\Phi}(\mathbf{s}), \sqrt{\alpha} \mathbf{\Phi}(\mathbf{t}), \mathbf{\Phi}(\mathbf{r}) \right) \right]^{-1} \right)
$$

(55)

where

$$
\eta_{i,k} \triangleq \left(\alpha \sum_{n=1}^{N_r} \sum_{j=1}^{Q} \sigma_{k,n,j}^{ir} + 1 \right)^{-1}, \quad i = 1, 2, \quad k = 1, \dots, N_s.
$$

(56)

It can be seen that (55) is minimized only if all off-diagonal matrices of the second term are zero, i.e.,

$$
\mathbf{\Phi}^{H}(\mathbf{s}) \mathbf{\Phi}(\mathbf{t}) = \mathbf{0} \quad \mathbf{\Phi}^{H}(\mathbf{s}) \mathbf{\Phi}(\mathbf{r}) = \mathbf{0} \quad \mathbf{\Phi}^{H}(\mathbf{r}) \mathbf{\Phi}(\mathbf{t}) = \mathbf{0}. \quad (57)
$$

Based on (7)–(9) and (57), we have that for $m, n = 1, \dots, N_s$, $p = 1, \dots, N_r$

$$
\mathbf{C}_{2Q-1}^{H}(\mathbf{s}_m) \mathbf{C}_{2Q-1}(\mathbf{t}_n) = \mathbf{0}, \quad \mathbf{C}_{2Q-1}^{H}(\mathbf{s}_m) \mathbf{C}_{Q}(\mathbf{r}_p) = \mathbf{0},
$$
$$
\mathbf{C}_{2Q-1}^{H}(\mathbf{t}_n) \mathbf{C}_{Q}(\mathbf{r}_p) = \mathbf{0}. \quad (58)
$$

Using (57), MSE in (55) can be written as

$$
\text{MSE} = \sum_{i=1}^{2} \sum_{k=1}^{N_s} \text{tr} \left(\left[\left(\mathbf{C}_{i1}^{k} \right)^{-1} + \alpha \eta_{i,k} \mathbf{\Phi}^{H}(\mathbf{s}) \mathbf{\Phi}(\mathbf{s}) \right]^{-1} \right.
$$
$$
+ \left[\left(\mathbf{C}_{i2}^{k} \right)^{-1} + \alpha \eta_{i,k} \mathbf{\Phi}^{H}(\mathbf{t}) \mathbf{\Phi}(\mathbf{t}) \right]^{-1}
$$
$$
\left. + \left[\left(\mathbf{C}_{ir}^{k} \right)^{-1} + \eta_{i,k} \mathbf{\Phi}^{H}(\mathbf{r}) \mathbf{\Phi}(\mathbf{r}) \right]^{-1} \right). \quad (59)
$$

Fig. 6. Example 3: NMSE versus Q for different P and $N = 2$.

The fourth simulation example studies the scenario where the power constraints at the source nodes and relay node are different. Fig. 7 shows the NMSE of the proposed algorithm versus $P_1 = P_2 = P_s$ for different fixed P_r when $N = 2$ and $Q = 6$. It can be seen that as expected, the proposed algorithm has a better NMSE performance when the power at the source/relay node is increased.

V. CONCLUSION

We have applied the method of superimposed channel training to two-way MIMO relay communication systems in frequency-selective fading environments. The proposed algorithm can efficiently estimate the individual CSI for two-way MIMO relay systems with frequency-selective fading channels. We also derived the optimal structure of the training sequences that minimize the MSE of the channel estimation and optimize the power allocation between the source and relay training sequences.

Since from (18) and (19), $\mathbf{C}_{i1}^k$, $\mathbf{C}_{i2}^k$, and $\mathbf{C}_{ir}^k$ are all diagonal, to minimize (59), $\mathbf{\Phi}^H(\mathbf{s})\mathbf{\Phi}(\mathbf{s})$, $\mathbf{\Phi}^H(\mathbf{t})\mathbf{\Phi}(\mathbf{t})$, and $\mathbf{\Phi}^H(\mathbf{r})\mathbf{\Phi}(\mathbf{r})$ must be diagonal, and together with (7)–(9), we have

$$\mathbf{C}_{2Q-1}^H(\mathbf{s}_m)\mathbf{C}_{2Q-1}(\mathbf{s}_m)=\mathbf{D}_{s,m}, \quad \mathbf{C}_{2Q-1}^H(\mathbf{t}_n)\mathbf{C}_{2Q-1}(\mathbf{t}_n)=\mathbf{D}_{t,n},$$

$$\mathbf{C}_Q^H(\mathbf{r}_p)\mathbf{C}_Q(\mathbf{r}_p)=\mathbf{D}_{r,p} \tag{60}$$

where $\mathbf{D}_{s,m}$ and $\mathbf{D}_{t,n}$ are $(2Q-1) \times (2Q-1)$ diagonal matrices, while $\mathbf{D}_{r,p}$ is a $Q \times Q$ diagonal matrix.

It is worth noting that (58) and (60) do not change the value of $\mathbf{s}_m^H\mathbf{s}_m$, $\mathbf{t}_m^H\mathbf{t}_m$, and $\mathbf{r}_n^H\mathbf{r}_n$ in the constraints (25)–(27). Moreover, $\mathrm{tr}(\mathbf{C}_Q(\mathbf{s}_m)\mathbf{D}_{n,m}^{r1}\mathbf{C}_Q^H(\mathbf{s}_m))$ in the constraint (27) is minimized if $\mathbf{C}_Q^H(\mathbf{s}_m)\mathbf{C}_Q(\mathbf{s}_m)$ is diagonal and its diagonal elements are in the inverse order to that of $\mathbf{D}_{n,m}^{r1}$[22]. Similarly, the term of $\mathrm{tr}(\mathbf{C}_Q(\mathbf{t}_m)\mathbf{D}_{n,m}^{r2}\mathbf{C}_Q^H(\mathbf{t}_m))$ in (27) is minimized if $\mathbf{C}_Q^H(\mathbf{t}_m)\mathbf{C}_Q(\mathbf{t}_m)$ is diagonal and its diagonal elements are in the inverse order to that of $\mathbf{D}_{n,m}^{r2}$. Obviously, these two requirements are satisfied by (60).

Considering (58), (60), and the circulant structure of $\mathbf{C}_{2Q-1}(\mathbf{s}_m)$, $\mathbf{C}_{2Q-1}(\mathbf{t}_n)$, and $\mathbf{C}_Q(\mathbf{r}_p)$, we have

$$\mathbf{D}_{s,m} = \beta_m\mathbf{I}_{2Q-1}, \quad \mathbf{D}_{t,n} = \gamma_n\mathbf{I}_{2Q-1}, \quad \mathbf{D}_{r,p} = \delta_p\mathbf{I}_Q$$

where $\mathbf{s}_m^H\mathbf{s}_m = \beta_m$, $\mathbf{t}_n^H\mathbf{t}_n = \gamma_n$, and $\mathbf{r}_p^H\mathbf{r}_p = \delta_p$. $\qquad\square$

APPENDIX B
PROOF OF THEOREM 2

By introducing $\xi_{1,m} \triangleq \alpha\beta_m$, $\xi_{2,m} \triangleq \alpha\gamma_m$, $m = 1, \ldots, N_s$, the problem (31)–(35) can be equivalently rewritten as

$$\min_{\boldsymbol{\xi}_1,\boldsymbol{\xi}_2,\boldsymbol{\delta},\alpha} \sum_{m=1}^{N_s}\sum_{k=1}^{N_s}\sum_{q=1}^{2Q-1}\sum_{i=1}^{2}\left(\frac{1}{c_{k,m,q}^{i1}+\xi_{1,m}\eta_{i,k}} + \frac{1}{c_{k,m,q}^{i2}+\xi_{2,m}\eta_{i,k}}\right)$$

$$+ \sum_{n=1}^{N_r}\sum_{k=1}^{N_s}\sum_{p=1}^{Q}\sum_{i=1}^{2}\frac{1}{c_{k,n,p}^{ir}+\delta_n\eta_{i,k}} \tag{61}$$

$$\text{s.t.} \quad \mathbf{1}^T\boldsymbol{\xi}_i \leq \alpha P_i, \quad i = 1, 2 \tag{62}$$

$$\boldsymbol{\kappa}_1^T\boldsymbol{\xi}_1 + \boldsymbol{\kappa}_2^T\boldsymbol{\xi}_2 + \mathbf{1}^T\boldsymbol{\delta} \leq P_r - \alpha LN_r \tag{63}$$

$$\alpha > 0, \ \xi_{i,m} \geq 0, \ i = 1, 2,$$

$$m = 1, \ldots, N_s, \ \delta_n \geq 0, \ n = 1, \ldots, N_r \tag{64}$$

where $\boldsymbol{\kappa}_i \triangleq [\kappa_{i,1}, \ldots, \kappa_{i,N_s}]^T$, $\boldsymbol{\xi}_i \triangleq [\xi_{i,1}, \ldots, \xi_{i,N_s}]^T$, $i = 1$, 2, and $\mathbf{1}$ is a column vector of all ones with a commensurate dimension.

Let us first ignore the effect of α on all $\eta_{i,k}$ by treating them as known variables. Then the problem (61)–(64) becomes a convex optimization problem, as (61) is a convex function of $\boldsymbol{\xi}_1$, $\boldsymbol{\xi}_2$, $\boldsymbol{\delta}$, and (62)–(64) are linear inequality constraints. When α has a sufficiently small value, the value of (61) is strongly governed by the constraints in (62), since the constraint (63) is inactive compared with the constraints in (62) for small value of α. Once α increases from a small value, the feasible region specified by (62) expands, and thus, the value of (61) decreases.

On the other hand, when α is large (close to $P_r/(LN_r)$), the value of (61) is strongly governed by the constraint (63), as the constraints in (62) are inactive compared with that of (63) when α is large. Once α decreases from a large value, the feasible region specified by (63) expands, leading to the decreasing of (61).

Now we consider the effect of α on $\eta_{i,k}$. It can be seen from (56) that $\eta_{i,k}$ monotonically decreases with increasing α, and (61) increases when $\eta_{i,k}$ decreases. From the analysis above, it can be deduced that when α increases from a significantly small positive number, the objective function (61) starts to decrease since the potential decrease of (61) due to the expanded feasible region of (62) dominates the potential increase of (61) caused by the decreasing $\eta_{i,k}$. The value of (61) keeps decreasing till a 'turning point' where the decreasing of $\eta_{i,k}$ starts to dominate the effect of relaxed feasible region in (62). After such turning point, the value of (61) is monotonically increasing with an increasing α. Therefore, the objective function (31) subjecting to (32)–(35) is a unimodal function with respect to α. $\qquad\square$

REFERENCES

[1] L. Sanguinetti, A. A. D'Amico, and Y. Rong, "A tutorial on transceiver design for amplify-and-forward MIMO relay systems," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 8, pp. 1331–1346, Sep. 2012.

[2] 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.

[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] I. Hammerström and A. Wittneben, "Power allocation schemes for amplify-and-forward MIMO-OFDM relay links," *IEEE Trans. Wireless Commun.*, vol. 6, no. 8, pp. 2798–2802, Aug. 2007.

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

[6] M. R. A. Kandaker and Y. Rong, "Precoding design for MIMO relay multicasting," *IEEE Trans. Wireless Commun.*, vol. 12, no. 7, pp. 3544–3555, Jul. 2013.

[7] C. E. Shannon, "Two-way communication channels," in *Proc. 4th Berkeley Symp. Probability Statist.*, Berkeley, CA, USA, 1961, vol. 1, pp. 611–644.

[8] S. Xu and Y. Hua, "Optimal design of spatial source-and-relay matrices for a non-regenerative two-way MIMO relay system," *IEEE Trans. Wireless Commun.*, vol. 10, no. 5, pp. 1645–1655, May 2011.

[9] R. Wang, M. Tao, and Y. Huang, "Linear precoding designs for amplify-and-forward multiuser two-way relay systems," *IEEE Trans. Wireless Commun.*, vol. 11, no. 12, pp. 4457–4469, Dec. 2012.

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

[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 systems via parallel factor analysis," *IEEE Trans. Wireless Commun.*, vol. 11, no. 6, pp. 2224–2233, Jun. 2012.

[15] F. Gao, B. Jiang, X. Gao, and X. Zhang, "Superimposed training based channel estimation for OFDM modulated amplify-and-forward relay networks," *IEEE Trans. Commun.*, vol. 59, no. 7, pp. 2029–2039, Jul. 2011.

[16] F. Gao, R. Zhang, and Y.-C. Liang, "Optimal channel estimation and training design for two-way relay networks," *IEEE Trans. Commun.*, vol. 57, no. 10, pp. 3024–3033, Oct. 2009.

[17] Z. Fang, J. Shi, and H. Shan, "Comparison of channel estimation schemes for MIMO two-way relaying systems," in *Proc. CSQRWC*, Jul. 2011, vol. 1, pp. 719–722.

[18] 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.

[19] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1993.

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

[21] A. Antoniou and W.-S. Lu, *Practical Optimization: Algorithms and Engineering Applications*. New York, NY, USA: Springer Science+Business Media, 2007.

[22] A. W. Marshall, I. Olkin, and B. C. Arnold, *Inequalities: Theory of Majorization and Its Applications*, 2nd ed. New York, NY, USA: Springer-Verlag, 2009.

Yue Rong (S'03–M'06–SM'11) received the Ph.D. degree (*summa cum laude*) in electrical engineering from the Darmstadt University of Technology, Darmstadt, Germany, in 2005.

He was a Post-Doctoral Researcher with the Department of Electrical Engineering, University of California, Riverside, from February 2006 to November 2007. Since December 2007, he has been with the Department of Electrical and Computer Engineering, Curtin University, Bentley, Australia, where he is currently an Associate 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.

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 Editor of the IEEE WIRELESS COMMUNICATIONS LETTERS, a Guest Editor of the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS special issue on theories and methods for advanced wireless relays, and was a TPC Member for the IEEE ICC, WCSP, IWCMC, and ChinaCom.

Choo W. R. Chiong received the B.E. (Hons.) degree in electrical power engineering from Curtin University, Australia, in 2010. He is currently working toward the Ph.D. degree at the Department of Electrical and Computer Engineering, Curtin University, Australia.

In 2009, he was with Curtin University Sarawak Campus, Malaysia, as a Research Assistant, and he mainly worked on the development of Smart Drive Build in AC Motors for Optimal Operations. His research interests include MIMO channel estimation and blind source separation.

Mr. Chiong was awarded the Digby Leach Medal by Engineers Australia in 2010, and the Curtin International Postgraduate Research Scholarship (CIPRS) for his Ph.D. study in 2011.

Yong Xiang (SM'12) received the Ph.D. degree in electrical and electronic engineering from The University of Melbourne, Australia. He is an Associate Professor and the Director of the Artificial Intelligence and Image Processing Research Cluster, School of Information Technology, Deakin University, Australia. His research interests include signal and system estimation, information and network security, multimedia (speech/image/video) processing, and wireless sensor networks. He has served as Program Chair, TPC Chair, Symposium Chair, and Session Chair for a number of international conferences. He currently serves as Associate Editor of IEEE Access.