Channel Estimation for Frequency-Selective Two-Way MIMO Relay Systems

Choo W. R. Chiong
Dept. Electrical and Computer Engineering
Curtin University
Bentley, WA, 6102, Australia
choowee.chiong@student.curtin.edu.au

Yue Rong
Dept. Electrical and Computer Engineering
Curtin University
Bentley, WA, 6102, Australia
y.rong@curtin.edu.au

Yong Xiang
School of Information Technology
Deakin University
Melbourne, VIC 3125, Australia
yong.xiang@deakin.edu.au

***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 propose a superimposed channel training algorithm 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. We also derive the optimal power allocation between the source and relay training sequences. Numerical examples are shown to demonstrate the performance of the proposed algorithm.**

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

I. INTRODUCTION

Two-way multiple-input multiple-output (MIMO) relay communications have attracted great interests recently as they can provide higher spectral efficiency compared with one-way relay systems [1]-[4]. In a two-way relay system, two source nodes exchange their information through relay node(s). The joint source and relay optimization for two-way MIMO relay systems has been studied in [2]-[4].

For the two-way MIMO relay systems discussed in [2]-[4], the knowledge of the instantaneous channel state information (CSI) is essential for extracting the source signals at the destination nodes 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 nodes. In [5], two-way relay channel estimation based on the maximum likelihood (ML) and linear maximum signal-to-noise ratio (SNR) have been proposed. A superimposed training-based channel estimation for two-way relay systems has been developed in [6], and has been extended to two-way MIMO relay systems in [7].

The relay system in [5]-[7] is assumed to have frequency-flat fading channels, which is only valid for narrowband 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. The estimation of frequency-selective fading channels has been discussed in [8] for single antenna one-way relay systems. We develop a superimposed channel training algorithm 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.

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 [7] 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.

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 for all nodes, i.e., each node is not able to transmit and receive signals at the same time.

Let us denote $\mathbf{h}^{ri}_{n,m} = \left[h^{ri}_{n,m,1}, \cdots, h^{ri}_{n,m,Q}\right]^T$ 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, $i = 1, 2$, $m = 1, \cdots, N_s$, and $n = 1, \cdots, N_r$, 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,

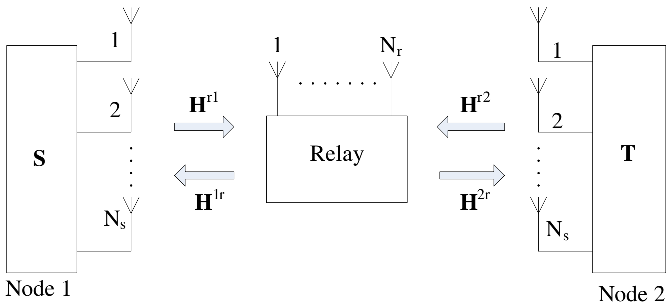

Fig. 1. Block diagram of a two-way MIMO relay communication system.

$\mathbf{h}_{m,n}^{ir} = \left[h_{m,n,1}^{ir}, \cdots, h_{m,n,Q}^{ir}\right]^T$ is used to denote the $Q \times 1$ second-hop multipath channel vector from the nth antenna at the relay node to the mth 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 $\mathbf{S} = [\mathbf{s}_1, \mathbf{s}_2, \cdots, \mathbf{s}_{N_s}]^T$ and source node 2 transmits an $N_s \times L$ training matrix $\mathbf{T} = [\mathbf{t}_1, \mathbf{t}_2, \cdots, \mathbf{t}_{N_s}]^T$, 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 $\mathbf{t}_n$, $m, n = 1, \cdots, N_s$, to prevent the inter-block interference at the relay node [9]. The received signal vectors at the relay node over L time slots after removing the cyclic prefix can be written as

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

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 $\left[(\mathbf{h}_{n,m}^{r1})^T, \mathbf{0}_{1\times(L-Q)}\right]^T$ and $\left[(\mathbf{h}_{n,m}^{r2})^T, \mathbf{0}_{1\times(L-Q)}\right]^T$, respectively.

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

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

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

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

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 $\left[(\mathbf{h}_{k,n}^{ir})^T, \mathbf{0}_{1\times(L-Q)}\right]^T$.

The main idea of the superimposed channel training algorithm is to exploit $\mathbf{R}$ to estimate the second-hop channels $\{\mathbf{h}_{k,n}^{ir}\} \triangleq \{\mathbf{h}_{k,n}^{ir}, i = 1, 2, k = 1, \cdots, N_s, n = 1, \cdots, N_r\}$, and then estimate the first-hop channels $\{\mathbf{h}_{n,m}^{ri}\} \triangleq \{\mathbf{h}_{n,m}^{ri}, i = 1, 2, n = 1, \cdots, N_r, m = 1, \cdots, N_s\}$ 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., the 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

$$\mathbf{y}_{i,k} = \sqrt{\alpha}\sum_{m=1}^{N_s}\sum_{n=1}^{N_r} \mathbf{H}_{k,n}^{ir}\mathbf{H}_{n,m}^{r1}\mathbf{s}_m + \sqrt{\alpha}\sum_{m=1}^{N_s}\sum_{n=1}^{N_r} \mathbf{H}_{k,n}^{ir}\mathbf{H}_{n,m}^{r2}\mathbf{t}_m + \sum_{n=1}^{N_r} \mathbf{H}_{k,n}^{ir}\mathbf{r}_n + \bar{\mathbf{v}}_{i,k}, \quad k = 1, \cdots, N_s \tag{4}$$

where

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

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

$$\begin{aligned}
\mathbf{y}_{i,k} &= \sqrt{\alpha}\sum_{m=1}^{N_s}\left[\mathbf{C}_{2Q-1}(\mathbf{s}_m)\sum_{n=1}^{N_r}\mathbf{h}_{k,n}^{ir} * \mathbf{h}_{n,m}^{r1}\right] \\
&\quad + \sqrt{\alpha}\sum_{m=1}^{N_s}\left[\mathbf{C}_{2Q-1}(\mathbf{t}_m)\sum_{n=1}^{N_r}\mathbf{h}_{k,n}^{ir} * \mathbf{h}_{n,m}^{r2}\right] \\
&\quad + \sum_{n=1}^{N_r}\mathbf{C}_Q(\mathbf{r}_n)\mathbf{h}_{k,n}^{ir} + \bar{\mathbf{v}}_{i,k} \\
&= \sqrt{\alpha}\boldsymbol{\Phi}(\mathbf{s})\mathbf{d}_k^{i1} + \sqrt{\alpha}\boldsymbol{\Phi}(\mathbf{t})\mathbf{d}_k^{i2} + \boldsymbol{\Phi}(\mathbf{r})\mathbf{d}_k^{ir} + \bar{\mathbf{v}}_{i,k} \\
&\qquad k = 1, \cdots, N_s
\end{aligned} \tag{6}$$

where $\mathbf{C}_Q(\mathbf{s})$ represents an $L \times Q$ column-wise circulant matrix taking $\mathbf{s}$ as the first column, $\mathbf{a} * \mathbf{b}$ denotes the linear convolution between vectors $\mathbf{a}$ and $\mathbf{b}$, and

$$\boldsymbol{\Phi}(\mathbf{s}) \triangleq [\mathbf{C}_{2Q-1}(\mathbf{s}_1), \mathbf{C}_{2Q-1}(\mathbf{s}_2), \cdots, \mathbf{C}_{2Q-1}(\mathbf{s}_{N_s})] \tag{7}$$

$$\boldsymbol{\Phi}(\mathbf{t}) \triangleq [\mathbf{C}_{2Q-1}(\mathbf{t}_1), \mathbf{C}_{2Q-1}(\mathbf{t}_2), \cdots, \mathbf{C}_{2Q-1}(\mathbf{t}_{N_s})] \tag{8}$$

$$\boldsymbol{\Phi}(\mathbf{r}) \triangleq [\mathbf{C}_Q(\mathbf{r}_1), \mathbf{C}_Q(\mathbf{r}_2), \cdots, \mathbf{C}_Q(\mathbf{r}_{N_r})] \tag{9}$$

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

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

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

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.

By introducing

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

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

we can rewrite (6) as

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

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

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

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

where $\hat{\mathbf{x}}_{i,k}$ denotes an estimation of $\mathbf{x}_{i,k}$, $(\cdot)^H$ denotes the matrix (vector) Hermitian transpose, and $\mathbf{W}_{i,k}$ is the weight matrix given by

$$\mathbf{W}_{i,k} = \left(\mathbf{A}\mathbf{C}_x^{i,k}\mathbf{A}^H + \mathbf{C}_{\bar{v}}^{i,k}\right)^{-1}\mathbf{A}\mathbf{C}_x^{i,k}$$
$$i = 1, 2, \quad k = 1, \cdots, N_s. \quad (17)$$

Here, $(\cdot)^{-1}$ denotes the matrix inversion. As a linear estimator is used, we can see from (13) that the length of the training sequences should satisfy $L \geq (4Q-2)N_s + QN_r$.

A. Structure of Optimal Training Sequences

Based on (15)-(17), the sum MSE of channel estimation at two nodes can be written as

$$\mathrm{MSE} = \sum_{i=1}^{2}\sum_{k=1}^{N_s} \mathrm{tr}\left(\mathrm{E}\left[(\hat{\mathbf{x}}_{i,k} - \mathbf{x}_{i,k})(\hat{\mathbf{x}}_{i,k} - \mathbf{x}_{i,k})^H\right]\right)$$
$$= \sum_{i=1}^{2}\sum_{k=1}^{N_s} \mathrm{tr}\left(\left[\left(\mathbf{C}_x^{i,k}\right)^{-1} + \mathbf{A}^H\left(\mathbf{C}_{\bar{v}}^{i,k}\right)^{-1}\mathbf{A}\right]^{-1}\right) \quad (18)$$

where $\mathrm{tr}(\cdot)$ denotes the matrix trace, $\mathbf{C}_x^{i,k} = \mathrm{E}\left[\mathbf{x}_{i,k}\mathbf{x}_{i,k}^H\right]$ is the covariance matrix of $\mathbf{x}_{i,k}$, and $\mathbf{C}_{\bar{v}}^{i,k} = \mathrm{E}\left[\bar{\mathbf{v}}_{i,k}\bar{\mathbf{v}}_{i,k}^H\right]$ is the noise covariance matrix. Here $\mathrm{E}[\cdot]$ stands for the statistical expectation.

From (5), we have

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

where $\sigma_{k,n,j}^{ir} = \mathrm{E}\left[h_{k,n,j}^{ir}\left(h_{k,n,j}^{ir}\right)^*\right]$ is the variance of $h_{k,n,j}^{ir}$, $j = 1, \cdots, Q$, $(\cdot)^*$ denotes complex conjugate, and $\mathbf{I}_n$ stands for the $n \times n$ identity matrix. Based on (10)-(12) and (14), we obtain that $\mathbf{C}_x^{i,k} = \mathrm{bd}\left[\mathbf{C}_{i1}^k, \mathbf{C}_{i2}^k, \mathbf{C}_{ir}^k\right]$, where $\mathrm{bd}[\cdot]$ represents a block diagonal matrix and

$$\mathbf{C}_{ij}^k = \mathrm{E}\left[\mathbf{d}_k^{ij}(\mathbf{d}_k^{ij})^H\right] = \mathrm{bd}\left[\mathbf{C}_{k,1}^{ij}, \cdots, \mathbf{C}_{k,N_s}^{ij}\right], \quad j = 1, 2 \quad (19)$$

$$\mathbf{C}_{ir}^k = \mathrm{E}\left[\mathbf{d}_k^{ir}(\mathbf{d}_k^{ir})^H\right] = \mathrm{bd}\left[\mathbf{C}_{k,1}^{ir}, \cdots, \mathbf{C}_{k,N_r}^{ir}\right]. \quad (20)$$

By introducing $\boldsymbol{\sigma}_{k,n}^{ir} = \left[\sigma_{k,n,1}^{ir}, \cdots, \sigma_{k,n,Q}^{ir}\right]^T$ and $\boldsymbol{\sigma}_{n,m}^{rj} = \left[\sigma_{n,m,1}^{rj}, \cdots, \sigma_{n,m,Q}^{rj}\right]^T$, where $\sigma_{n,m,p}^{rj} = \mathrm{E}\left[h_{n,m,p}^{rj}\left(h_{n,m,p}^{rj}\right)^*\right]$ is the variance of $h_{n,m,p}^{rj}$, $j = 1, 2$, $p = 1, \cdots, Q$, we obtain

$$\mathbf{C}_{k,m}^{ij} = \mathrm{E}\left[\left(\sum_{n=1}^{N_r} \mathbf{h}_{k,n}^{ir} * \mathbf{h}_{n,m}^{rj}\right)\left(\sum_{n=1}^{N_r} \mathbf{h}_{k,n}^{ir} * \mathbf{h}_{n,m}^{rj}\right)^H\right]$$
$$= \sum_{n=1}^{N_r} \mathrm{diag}\left[\boldsymbol{\sigma}_{k,n}^{ir} * \boldsymbol{\sigma}_{n,m}^{rj}\right], \quad j = 1, 2, \quad m = 1, \cdots, N_s$$

$$\mathbf{C}_{k,n}^{ir} = \mathrm{E}\left[\mathbf{h}_{k,n}^{ir}(\mathbf{h}_{k,n}^{ir})^H\right]$$
$$= \mathrm{diag}\left[\sigma_{k,n,1}^{ir}, \cdots, \sigma_{k,n,Q}^{ir}\right], \quad n = 1, \cdots, N_r.$$

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

The transmission power constraints at the source nodes are given by

$$\sum_{m=1}^{N_s} \mathbf{s}_m^H\mathbf{s}_m \leq P_1, \quad \sum_{m=1}^{N_s} \mathbf{t}_m^H\mathbf{t}_m \leq P_2 \quad (21)$$

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

$$\sum_{n=1}^{N_r} \mathrm{E}\left[\mathrm{tr}\left(\mathbf{x}_{r,n}\mathbf{x}_{r,n}^H\right)\right]$$
$$= \sum_{n=1}^{N_r}\left(\alpha\mathrm{tr}\left(\sum_{m=1}^{N_s}\left(\mathbf{C}_Q(\mathbf{s}_m)\mathbf{D}_{n,m}^{r1}\mathbf{C}_Q^H(\mathbf{s}_m)\right.\right.\right.$$
$$\left.\left.\left.+\mathbf{C}_Q(\mathbf{t}_m)\mathbf{D}_{n,m}^{r2}\mathbf{C}_Q^H(\mathbf{t}_m)\right) + \mathbf{I}_L\right) + \mathbf{r}_n^H\mathbf{r}_n\right) \leq P_r \quad (22)$$

where $\mathbf{D}_{n,m}^{ri} \triangleq \mathrm{diag}\left[\sigma_{n,m,1}^{ri}, \cdots, \sigma_{n,m,Q}^{ri}\right]$, $i = 1, 2$, and P_r is the transmission power available at the relay node.

The following theorem establishes the optimal structure of $\mathbf{S}$, $\mathbf{T}$, and $\mathbf{R}$ that minimize (18) subjected to the power constraints in (21) and (22).

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

$$\mathbf{C}_{2Q-1}^H(\mathbf{s}_m)\mathbf{C}_{2Q-1}(\mathbf{s}_m) = \beta_m\mathbf{I}_{2Q-1}$$
$$\mathbf{C}_{2Q-1}^H(\mathbf{t}_n)\mathbf{C}_{2Q-1}(\mathbf{t}_n) = \gamma_n\mathbf{I}_{2Q-1}$$
$$\mathbf{C}_Q^H(\mathbf{r}_p)\mathbf{C}_Q(\mathbf{r}_p) = \delta_p\mathbf{I}_Q \quad (23)$$
$$\mathbf{C}_{2Q-1}^H(\mathbf{s}_m)\mathbf{C}_{2Q-1}(\mathbf{t}_n) = \mathbf{0}$$
$$\mathbf{C}_{2Q-1}^H(\mathbf{s}_m)\mathbf{C}_Q(\mathbf{r}_p) = \mathbf{0}$$
$$\mathbf{C}_{2Q-1}^H(\mathbf{t}_n)\mathbf{C}_Q(\mathbf{r}_p) = \mathbf{0} \quad (24)$$

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: The MSE in (18) can be rewritten as

$$\mathrm{MSE} = \sum_{i=1}^{2}\sum_{k=1}^{N_s}\mathrm{tr}\left(\left[\begin{pmatrix}\mathbf{C}_{i1}^{k} & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathbf{C}_{i2}^{k} & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \mathbf{C}_{ir}^{k}\end{pmatrix}^{-1} + \eta_{i,k}\begin{pmatrix}\sqrt{\alpha}\boldsymbol{\Phi}^H(\mathbf{s})\\ \sqrt{\alpha}\boldsymbol{\Phi}^H(\mathbf{t})\\ \boldsymbol{\Phi}^H(\mathbf{r})\end{pmatrix} \times\left(\sqrt{\alpha}\boldsymbol{\Phi}(\mathbf{s}),\ \sqrt{\alpha}\boldsymbol{\Phi}(\mathbf{t}),\ \boldsymbol{\Phi}(\mathbf{r})\right)\right]^{-1}\right) \tag{25}$$

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

It can be seen that (25) is minimized only if all off-diagonal matrices of the second term are zero, i.e.,

$$\boldsymbol{\Phi}^H(\mathbf{s})\boldsymbol{\Phi}(\mathbf{t}) = \mathbf{0} \quad \boldsymbol{\Phi}^H(\mathbf{s})\boldsymbol{\Phi}(\mathbf{r}) = \mathbf{0} \quad \boldsymbol{\Phi}^H(\mathbf{r})\boldsymbol{\Phi}(\mathbf{t}) = \mathbf{0}. \tag{26}$$

Based on (7)-(9) and (26), we have that for $m, n = 1, \cdots, N_s$, $p = 1, \cdots, N_r$

$$\begin{aligned}\mathbf{C}_{2Q-1}^H(\mathbf{s}_m)\mathbf{C}_{2Q-1}(\mathbf{t}_n) &= \mathbf{0}\\ \mathbf{C}_{2Q-1}^H(\mathbf{s}_m)\mathbf{C}_{Q}(\mathbf{r}_p) &= \mathbf{0}\\ \mathbf{C}_{2Q-1}^H(\mathbf{t}_n)\mathbf{C}_{Q}(\mathbf{r}_p) &= \mathbf{0}.\end{aligned}$$

Using (26), MSE in (25) can be written as

$$\begin{aligned}\mathrm{MSE} = \sum_{i=1}^{2}\sum_{k=1}^{N_s}\mathrm{tr}\Big(&\left[\left(\mathbf{C}_{i1}^{k}\right)^{-1} + \alpha\eta_{i,k}\boldsymbol{\Phi}^H(\mathbf{s})\boldsymbol{\Phi}(\mathbf{s})\right]^{-1}\\ &+\left[\left(\mathbf{C}_{i2}^{k}\right)^{-1} + \alpha\eta_{i,k}\boldsymbol{\Phi}^H(\mathbf{t})\boldsymbol{\Phi}(\mathbf{t})\right]^{-1}\\ &+\left[\left(\mathbf{C}_{ir}^{k}\right)^{-1} + \eta_{i,k}\boldsymbol{\Phi}^H(\mathbf{r})\boldsymbol{\Phi}(\mathbf{r})\right]^{-1}\Big).\end{aligned} \tag{27}$$

Since from (19) and (20), $\mathbf{C}_{i1}^{k}$, $\mathbf{C}_{i2}^{k}$, and $\mathbf{C}_{ir}^{k}$ are all diagonal, to minimize (27), $\boldsymbol{\Phi}^H(\mathbf{s})\boldsymbol{\Phi}(\mathbf{s})$, $\boldsymbol{\Phi}^H(\mathbf{t})\boldsymbol{\Phi}(\mathbf{t})$, and $\boldsymbol{\Phi}^H(\mathbf{r})\boldsymbol{\Phi}(\mathbf{r})$ must be diagonal, and together with (7)-(9), we have

$$\begin{aligned}\mathbf{C}_{2Q-1}^H(\mathbf{s}_m)\mathbf{C}_{2Q-1}(\mathbf{s}_m) &= \mathbf{D}_{s,m}\\ \mathbf{C}_{2Q-1}^H(\mathbf{t}_n)\mathbf{C}_{2Q-1}(\mathbf{t}_n) &= \mathbf{D}_{t,n}\\ \mathbf{C}_{Q}^H(\mathbf{r}_p)\mathbf{C}_{Q}(\mathbf{r}_p) &= \mathbf{D}_{r,p}\end{aligned}$$

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 $\mathrm{tr}\left(\mathbf{C}_Q(\mathbf{s}_m)\mathbf{D}_{n,m}^{r1}\mathbf{C}_Q^H(\mathbf{s}_m)\right)$ in the constraint (22) 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 $\mathbf{D}_{n,m}^{r1}$ [10]. Similarly, the term of $\mathrm{tr}\left(\mathbf{C}_Q(\mathbf{t}_m)\mathbf{D}_{n,m}^{r2}\mathbf{C}_Q^H(\mathbf{t}_m)\right)$ in (22) 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 $\mathbf{D}_{n,m}^{r2}$.

Considering 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 $\mathbf{r}_p^H\mathbf{r}_p = \delta_p$. □

B. Optimal Power Loading

Applying Theorem 1, the MSE function in (18) can be written as

$$\begin{aligned}\mathrm{MSE} = \sum_{i=1}^{2}\sum_{k=1}^{N_s}\mathrm{tr}\Bigg(&\sum_{m=1}^{N_s}\left[\left(\mathbf{C}_{k,m}^{i1}\right)^{-1} + \alpha\beta_m\eta_{i,k}\mathbf{I}_{2Q-1}\right]^{-1}\\ &+\sum_{m=1}^{N_s}\left[\left(\mathbf{C}_{k,m}^{i2}\right)^{-1} + \alpha\gamma_m\eta_{i,k}\mathbf{I}_{2Q-1}\right]^{-1}\\ &+\sum_{n=1}^{N_r}\left[\left(\mathbf{C}_{k,n}^{ir}\right)^{-1} + \delta_n\eta_{i,k}\mathbf{I}_{Q}\right]^{-1}\Bigg).\end{aligned} \tag{28}$$

Let us denote $c_{k,m,q}^{ij} \triangleq \left[\left(\mathbf{C}_{k,m}^{ij}\right)^{-1}\right]_{q,q}$, $c_{k,n,p}^{ir} \triangleq \left[\left(\mathbf{C}_{k,n}^{ir}\right)^{-1}\right]_{p,p}$, and $\kappa_{i,m} \triangleq \sum_{n=1}^{N_r}\sum_{q=1}^{Q}\sigma_{n,m,q}^{ri}$, $i = 1, 2$, the optimization problem can be equivalently rewritten as the following problem in scalar variables

$$\begin{aligned}\min_{\boldsymbol{\beta},\boldsymbol{\gamma},\boldsymbol{\delta},\alpha} &\sum_{m=1}^{N_s}\sum_{k=1}^{N_s}\sum_{q=1}^{2Q-1}\sum_{i=1}^{2}\left(\frac{1}{c_{k,m,q}^{i1}+\alpha\beta_m\eta_{i,k}} + \frac{1}{c_{k,m,q}^{i2}+\alpha\gamma_m\eta_{i,k}}\right)\\ &+\sum_{n=1}^{N_r}\sum_{k=1}^{N_s}\sum_{p=1}^{Q}\sum_{i=1}^{2}\frac{1}{c_{k,n,p}^{ir}+\delta_n\eta_{i,k}}\end{aligned} \tag{29}$$

$$\text{s.t.} \sum_{m=1}^{N_s}\beta_m \le P_1 \tag{30}$$

$$\sum_{m=1}^{N_s}\gamma_m \le P_2 \tag{31}$$

$$\alpha\left(\sum_{m=1}^{N_s}\kappa_{1,m}\beta_m + \sum_{m=1}^{N_s}\kappa_{2,m}\gamma_m\right) + \sum_{n=1}^{N_r}\delta_n + \alpha L N_r \le P_r \tag{32}$$

$$\alpha > 0, \quad \beta_m \ge 0, \gamma_m \ge 0, \quad m = 1, \cdots, N_s \tag{33}$$

$$\delta_n \ge 0, \quad n = 1, \cdots, N_r \tag{34}$$

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

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 (29) are monotonically decreasing and convex function with respect to β_m, γ_m, and δ_n. Moreover, when α is fixed, the constraints in (30)-(34) are linear inequality constraints. Therefore, with fixed α, the problem (29)-(34) 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 [11].

When α is not fixed, i.e., α is an optimization variable, the problem (29)-(34) as a whole is not a convex optimization problem. However, it can be shown that (29) is a unimodal function of α. Due to space limit, the proof is omitted. For a unimodal function, the minimum value can be efficiently found by the golden section search (GSS) [12] technique.

IV. NUMERICAL EXAMPLES

In this section, we study the performance of the proposed superimposed channel training algorithm for two-way

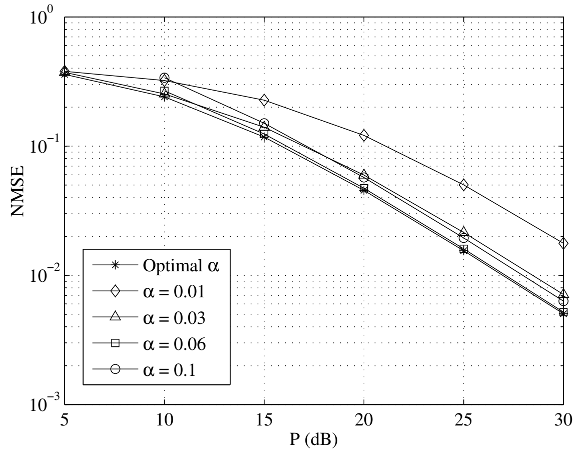

Fig. 2. NMSE versus P for different α with $N = 2$ and $Q = 4$.

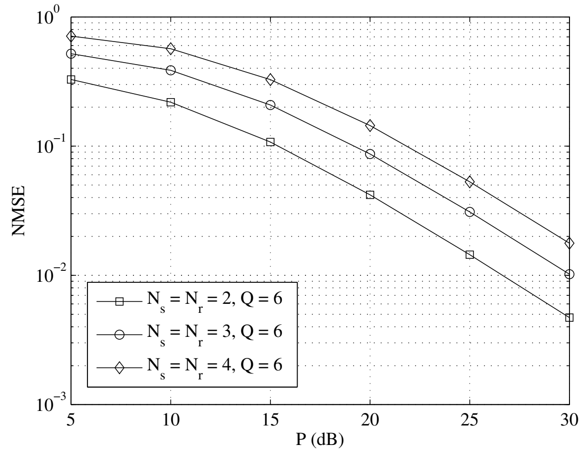

Fig. 3. NMSE versus P for different N with $Q = 6$.

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 nodes have the same transmission power $P_i = P$, $i = 1, 2, r$, and 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.

In the first example, we investigate the performance of the superimposed channel training algorithm for different α. Fig. 2 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 on the proposed superimposed channel training algorithm to obtain the optimal α for different P. It can be observed from Fig. 2 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. Note that the starting point for the curves associated with $\alpha = 0.06$ and $\alpha = 0.1$ is at $P = 10$dB, as these values of α have exceeded the upper bound limit of the α for $P = 5$dB.

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. Fig. 3 demonstrates the NMSE performance of the proposed algorithm 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.

V. Conclusions

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

Acknowledgment

This research was supported under the Australian Research Council's Discovery Projects funding scheme (project numbers DP140102131, DP110102076).

References

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

[2] 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, pp. 1645-1655, May 2011.

[3] 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, pp. 4457-4469, Dec. 2012.

[4] Y. Rong, "Joint source and relay optimization for two-way linear non-regenerative MIMO relay communications," *IEEE Trans. Signal Process.*, vol. 60, pp. 6533-6546, Dec. 2012.

[5] F. Gao, R. Zhang, and Y.-C. Liang, "Optimal channel estimation and training design for two-way relay networks," *IEEE Trans. Commun.*, vol. 57, pp. 3024-3033, Oct. 2009.

[6] G. Wang, F. Gao, X. Zhang and C. Tellambura, "Superimposed training-based joint CFO and channel estimation for CP-OFDM modulated two-way relay networks," *EURASIP J. Wireless Commun. Netw.*, vol. 2010, Jul. 2010.

[7] C. W. R. Chiong, Y. Rong, and Y. Xiang, "Channel training algorithms for two-way MIMO relay systems," *IEEE Trans. Signal Process.*, vol. 61, pp. 3988-3998, Aug. 2013.

[8] H. Zhang, S.-H. Leung, S. Gao, F. Gao, D. Pan, and X. Dai, "Doubly selective channel estimation for OFDM modulated amplify-and-forward relay networks using superimposed training," *EURASIP J. Wireless Commun. Netw.*, Aug. 2012.

[9] 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, pp. 2029-2039, Jul. 2011.

[10] A. W. Marshall, I. Olkin, and B. C. Arnold, *Inequalities: Theory of Majorization and Its Applications*. 2nd Ed., Springer, 2009.

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

[12] A. Antoniou and W.-S. Lu, *Practical Optimization: Algorithms and Engineering Applications*. Spring Street, NY: Springer Science+Business Media, LCC, 2007.