Channel Training Algorithms for Two-Way MIMO Relay Systems

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

Abstract—Two-way relay systems are known to be capable of providing higher spectral efficiency compared with conventional one-way relay systems. However, the channel estimation problem for two-way relay systems is more complicated than that of one-way relay systems. In this paper, we propose and compare two channel estimation algorithms, namely the superimposed channel training scheme and the two-stage channel estimation algorithm, for two-way multiple-input multiple-output (MIMO) relay communication systems, where the individual channel state information (CSI) for the first-hop and second-hop links is estimated. For both algorithms, we derive the optimal structure of the source and relay training sequences which minimize the mean-squared error (MSE) of channel estimation. In the superimposed channel training scheme, the power allocation between the source and relay training sequences is optimized. For the two-stage channel estimation algorithm, we optimize the power allocation at the relay node between two stages to improve the performance of the algorithm. Numerical examples are shown to demonstrate and compare the performance of the proposed channel training algorithms.

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

I. INTRODUCTION

In recent years, multiple-input multiple-output (MIMO) relay communication systems have attracted many research interests due to significant growth in the demand for fast and reliable wireless communications [1]-[5]. In [2] and [3], the optimal relay precoding matrix is derived to maximize the mutual information between the source and destination nodes for a three-node two-hop MIMO relay communication system. In [4], a unified framework has been developed to optimize the source and relay precoding matrices for two-hop MIMO relay systems with a broad class of commonly used objective functions. A recent survey on transceiver design for amplify-and-forward MIMO relay systems is presented in [5].

It can be seen from [1]-[5] that in a MIMO relay system, the knowledge of the instantaneous channel state information (CSI) is required at the destination node to retrieve the information transmitted by the source node. However, in practical relay communication systems, the instantaneous CSI is unknown, and therefore, has to be estimated. A least-squares (LS) fitting-based channel estimation algorithm is

Choo W. R. Chiong and Y. Rong (corresponding author) are with the Department of Electrical and Computer Engineering, Curtin University, Bentley, WA 6102, Australia. (e-mail: choowee.chiong@student.curtin.edu.au; y.rong@curtin.edu.au).

Y. Xiang is with the School of Information Technology, Deakin University, Melbourne, VIC 3125, Australia. (e-mail: yong.xiang@deakin.edu.au).

proposed in [6] for MIMO relay systems. A two-stage channel training algorithm is developed in [7], where the optimal training sequence at the source and relay nodes is derived. In [8], a parallel factor (PARAFAC) analysis based algorithm is proposed which can estimate MIMO relay channels up to a scaling ambiguity that is inherent in the PARAFAC model.

The channel estimation algorithms in [6] and [7] are developed for one-way relay systems [1]-[4], where a source node sends signals to a destination node through relay node(s). In two-way relay systems, two source nodes exchange their information through assisting relay node(s). Initially studied by Shannon in [9], two-way relay systems are getting more attention recently as they have higher spectral efficiency compared with one-way relay systems. For two-way MIMO relay systems, the joint source and relay optimization is recently investigated in [10] assuming the channel matrices are known. Channel estimation issue is not discussed in [10].

The channel estimation problem becomes more complicated in two-way relay systems and several algorithms have been proposed in [11]-[13]. Maximum likelihood (ML) and linear maximum signal-to-noise ratio (SNR) channel estimation techniques have been introduced in [11], while block-based training and pilot-tone based training algorithms are presented in [12]. However, the algorithms in [11] and [12] are based on the assumption that each node is equipped with single antenna only, and extension to MIMO systems is not straightforward.

For two-way MIMO relay systems, cascaded channel estimation and individual channel estimation algorithms have been proposed in [13]. The cascaded channel estimation is easy to implement but does not provide the second-hop CSI, which is necessary for system optimization [10]. In the individual channel estimation algorithm, the first-hop CSI is first estimated at the relay node and then fed-forward to the receive nodes. However, this algorithm requires the relay node to be capable of performing advanced signal processing, and therefore, increases the cost and complexity at the relay node.

The major challenge in channel estimation for two-way MIMO relay systems is to obtain the instantaneous CSI of both the first-hop and second-hop links with a minimal amount of signal processing at the relay node. In this paper, we address this challenge by proposing two algorithms: the superimposed channel training scheme and the two-stage channel estimation algorithm. In the superimposed channel training algorithm, both source nodes transmit their training sequence simultaneously to the relay node in the first time block. The relay node then amplifies the received signals and superimposes its own training sequence, before transmitting the superimposed signals to both receive nodes. By exploiting the training

sequences from the source and relay nodes, the individual CSI of the first-hop and second-hop links can be successfully estimated.

In the two-stage channel estimation algorithm, both source nodes are silent at the first stage, while the relay node broadcasts a pilot matrix to both receive nodes for the estimation of the channel matrices from the relay node to the receive nodes (second-hop links). During the second stage, both source nodes transmit their training sequence simultaneously to the relay node, and the relay node amplifies the received signals and forwards them to the receive nodes. Then, the channel matrices from the source nodes to the relay node (first-hop links) are estimated by exploiting the second-hop channel matrices estimated at the first stage. We would like to mention that although the estimation of the second-hop channels at the first stage is similar to the problem in [14] and [15], an efficient estimation of the first-hop channels is a non-conventional problem.

For both algorithms, we derive the structure of the optimal training sequences that minimize the sum mean-squared error (MSE) of channel estimation. In particular, we show that the optimal training matrix for each hop matches the eigenvector matrix of the correlation matrix of the MIMO channel at that hop. Moreover, in the superimposed channel training scheme, the power allocation between the source and relay training sequences is optimized. For the two-stage channel estimation algorithm, we optimize the power allocation at the relay node between two stages to minimize the MSE of channel estimation. The performance of the superimposed channel training scheme and the two-stage channel estimation algorithm are demonstrated and compared through numerical examples.

The rest of this paper is organized as follows. The system model of a two-way MIMO relay system is presented in Section II. The superimposed channel training algorithm is developed in Section III. In Section IV, we introduce the two-stage channel estimation algorithm and derive the optimal training sequences and power allocation at the relay node. Section V shows numerical examples to demonstrate the performance of the proposed algorithms. Finally, conclusions are drawn in Section VI.

II. SYSTEM MODEL

We consider a three-node two-way MIMO communication system where node 1 and node 2 exchange information through a relay node as shown in Fig. 1. Nodes 1 and 2 are equipped with N_1 and N_2 antennas, respectively, while the relay node has N_r antennas. For $i = 1, 2$, $\mathbf{H}_{ir}$ is the $N_i \times N_r$ channel matrix from the relay node to node i, while $\mathbf{H}_{ri}$ denotes the $N_r \times N_i$ channel matrix from node i to the relay node. In this paper, we consider that all nodes are operating in the half-duplex mode, i.e., one node cannot transmit and receive at the same time. Since in a two-way relay system, both source nodes transmit signals to the relay node at the first time slot, they cannot receive signals from each other. Therefore, there is no direct link between two source nodes. The half-duplex mode has been widely used in two-way relay communications [10]-[12].
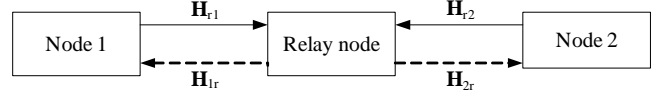

Fig. 1. Block diagram of a two-way MIMO relay communication system.

In this paper, we assume that the channel matrices $\mathbf{H}_{ri}$ and $\mathbf{H}_{ir}$ satisfy the well-known Gaussian-Kronecker model [16], where $\mathbf{H}_{ri}$ and $\mathbf{H}_{ir}$ are complex-valued Gaussian random matrices with

$$\mathbf{H}_{ri} \sim \mathcal{CN}(\mathbf{0}, \mathbf{T}_{ri} \otimes \mathbf{R}_{ri}),$$
$$\mathbf{H}_{ir} \sim \mathcal{CN}(\mathbf{0}, \mathbf{C}_r \otimes \mathbf{R}_{ir}), \qquad i = 1, 2. \qquad (1)$$

Here $\otimes$ stands for the matrix Kronecker product [17], $\mathbf{T}_{ri}$ and $\mathbf{R}_{ri}$ denote the $N_i \times N_i$ and $N_r \times N_r$ covariance matrix at the transmit and receive side of $\mathbf{H}_{ri}$, respectively, while $\mathbf{C}_r$ and $\mathbf{R}_{ir}$ stand for the $N_r \times N_r$ and $N_i \times N_i$ covariance matrix at the transmit and receive side of $\mathbf{H}_{ir}$, respectively. In other words, from (1) we have

$$\mathbf{H}_{ri} = \mathbf{A}_{ri}\mathbf{H}_{ri,\mathrm{w}}\mathbf{B}_{ri}^H,$$
$$\mathbf{H}_{ir} = \mathbf{A}_{ir}\mathbf{H}_{ir,\mathrm{w}}\mathbf{K}_r^H, \qquad i = 1, 2 \qquad (2)$$

where $\mathbf{A}_{ri}\mathbf{A}_{ri}^H = \mathbf{R}_{ri}$, $\mathbf{B}_{ri}\mathbf{B}_{ri}^H = \mathbf{T}_{ri}^T$, $\mathbf{A}_{ir}\mathbf{A}_{ir}^H = \mathbf{R}_{ir}$, $\mathbf{K}_r\mathbf{K}_r^H = \mathbf{C}_r^T$, $i = 1, 2$, $\mathbf{H}_{ri,\mathrm{w}}$ and $\mathbf{H}_{ir,\mathrm{w}}$ are $N_r \times N_i$ and $N_i \times N_r$ Gaussian random matrices with independent and identically distributed (i.i.d.) zero mean and unit variance entries. Here $(\cdot)^T$ and $(\cdot)^H$ denote matrix (vector) transpose and Hermitian transpose, respectively. We assume that $\mathbf{H}_{ri,\mathrm{w}}$ and $\mathbf{H}_{ir,\mathrm{w}}$, $i = 1, 2$, are statistically independent of each other. The following lemma is useful in this paper.

LEMMA 1 [18]: For $\mathbf{H} \sim \mathcal{CN}(\mathbf{0}, \mathbf{\Theta} \otimes \mathbf{\Phi})$, there is $\mathrm{E}[\mathbf{H}\mathbf{A}\mathbf{H}^H] = \mathrm{tr}(\mathbf{A}\mathbf{\Theta}^T)\mathbf{\Phi}$, and $\mathrm{E}[\mathbf{H}^H\mathbf{A}\mathbf{H}] = \mathrm{tr}(\mathbf{\Phi}\mathbf{A})\mathbf{\Theta}^T$. Here $\mathrm{E}[\cdot]$ stands for statistical expectation, and $\mathrm{tr}(\cdot)$ denotes matrix trace.

III. SUPERIMPOSED CHANNEL TRAINING ALGORITHM

In this section, we develop a superimposed channel training algorithm to estimate $\mathbf{H}_{ri}$ and $\mathbf{H}_{ir}$, $i = 1, 2$. This channel estimation scheme is completed in two time blocks. In the first time block, the source node i transmits an $N_i \times T$ training signal matrix $\mathbf{S}_i$, where T is the length of the training sequence. The $N_r \times T$ received signal matrix $\mathbf{Y}_r$ at the relay node is given by

$$\mathbf{Y}_r = \sum_{i=1}^{2} \mathbf{H}_{ri}\mathbf{S}_i + \mathbf{V}_r \qquad (3)$$

where $\mathbf{V}_r$ is an $N_r \times T$ noise matrix at the relay node.

In the second time block, the relay node amplifies $\mathbf{Y}_r$ and superimposes its own training matrix $\mathbf{S}_r$. Thus, the $N_r \times T$ signal matrix transmitted by the relay node can be written as

$$\mathbf{X}_r = \sqrt{\alpha}\mathbf{Y}_r + \mathbf{S}_r \qquad (4)$$

where $\alpha > 0$ is the relay amplifying factor. From (3) and (4), the $N_i \times T$ received signal matrix at node i is given by

$$\mathbf{Y}_i = \mathbf{H}_{ir}\mathbf{X}_r + \mathbf{V}_i$$
$$= \sqrt{\alpha}\mathbf{H}_{ir}\mathbf{H}_{ri}\mathbf{S}_i + \sqrt{\alpha}\mathbf{H}_{ir}\mathbf{H}_{r\bar{i}}\mathbf{S}_{\bar{i}} + \mathbf{H}_{ir}\mathbf{S}_r$$
$$+ \sqrt{\alpha}\mathbf{H}_{ir}\mathbf{V}_r + \mathbf{V}_i, \qquad i = 1, 2 \qquad (5)$$

where $\mathbf{V}_i$ is an $N_i \times T$ noise matrix at node i. Here, $\bar{i} = 2$ for $i = 1$, and $\bar{i} = 1$ for $i = 2$. The main idea of the superimposed channel training algorithm is to use $\mathbf{S}_r$ to estimate the second-hop channels $\mathbf{H}_{ir}$. Then the first-hop channels $\mathbf{H}_{rj}$, $j = i, \bar{i}$, can be estimated by exploiting $\mathbf{S}_j$ and the estimated $\mathbf{H}_{ir}$.

Let us introduce the eigenvalue decomposition (EVD) of $\mathbf{T}_{ri}^T$ as $\mathbf{U}_i\mathbf{\Lambda}_i\mathbf{U}_i^H$, $i = 1, 2$, and the EVD of $\mathbf{C}_r^T$ as $\mathbf{U}_r\mathbf{\Lambda}_r\mathbf{U}_r^H$. Then we have $\mathbf{B}_{ri}^H = \mathbf{\Pi}_i\mathbf{\Lambda}_i^{\frac{1}{2}}\mathbf{U}_i^H$, $i = 1, 2$, and $\mathbf{K}_r^H = \mathbf{\Pi}_r\mathbf{\Lambda}_r^{\frac{1}{2}}\mathbf{U}_r^H$, where $\mathbf{\Pi}_i$ and $\mathbf{\Pi}_r$ are arbitrary $N_i \times N_i$ and $N_r \times N_r$ unitary matrix, respectively. Using (2), we can rewrite (5) as

$$\mathbf{Y}_i = \sqrt{\alpha}\mathbf{G}_{ii}\tilde{\mathbf{S}}_i + \sqrt{\alpha}\mathbf{G}_{i\bar{i}}\tilde{\mathbf{S}}_{\bar{i}} + \tilde{\mathbf{H}}_{ir}\tilde{\mathbf{S}}_r + \bar{\mathbf{V}}_i, \quad i = 1, 2 \quad (6)$$

where $\tilde{\mathbf{S}}_r \triangleq \mathbf{U}_r^H\mathbf{S}_r$,

$$\mathbf{G}_{ij} \triangleq \mathbf{H}_{ir}\tilde{\mathbf{H}}_{rj}, \quad \tilde{\mathbf{S}}_j \triangleq \mathbf{U}_j^H\mathbf{S}_j, \quad \tilde{\mathbf{H}}_{rj} = \mathbf{H}_{rj}\mathbf{U}_j, \quad j = i, \bar{i},$$
$$\tilde{\mathbf{H}}_{ir} \triangleq \mathbf{H}_{ir}\mathbf{U}_r, \quad i = 1, 2 \qquad (7)$$

and

$$\bar{\mathbf{V}}_i \triangleq \sqrt{\alpha}\mathbf{H}_{ir}\mathbf{V}_r + \mathbf{V}_i, \qquad i = 1, 2 \qquad (8)$$

is the equivalent noise matrix at node i. In the following, we develop an algorithm to estimate $\tilde{\mathbf{H}}_{ir}$ and $\mathbf{G}_{ij}$ in (6). Then an estimate of $\mathbf{H}_{ir}$ and $\mathbf{H}_{rj}$ can be obtained from (7) as $\hat{\mathbf{H}}_{ir} = \breve{\mathbf{H}}_{ir}\mathbf{U}_r^H$ and $\hat{\mathbf{H}}_{rj} = \hat{\mathbf{H}}_{ir}^\dagger\breve{\mathbf{G}}_{ij}\mathbf{U}_j^H$, $j = i, \bar{i}$, where $(\cdot)^\dagger$ stands for matrix pseudo-inverse, $\breve{\mathbf{H}}_{ir}$ and $\breve{\mathbf{G}}_{ij}$ are the estimates of $\tilde{\mathbf{H}}_{ir}$ and $\mathbf{G}_{ij}$, respectively.

By vectorizing both sides of (6), we obtain

$$\mathbf{y}_i = \left[\sqrt{\alpha}\tilde{\mathbf{S}}_i^T\otimes\mathbf{I}_{N_i}, \; \sqrt{\alpha}\tilde{\mathbf{S}}_{\bar{i}}^T\otimes\mathbf{I}_{N_i}, \; \tilde{\mathbf{S}}_r^T\otimes\mathbf{I}_{N_i}\right]\left[\mathbf{g}_{ii}^T, \; \mathbf{g}_{i\bar{i}}^T, \; \tilde{\mathbf{h}}_{ir}^T\right]^T$$
$$+ \bar{\mathbf{v}}_i \qquad (9)$$
$$\triangleq \mathbf{M}_i\boldsymbol{\gamma}_i + \bar{\mathbf{v}}_i, \qquad i = 1, 2 \qquad (10)$$

where for $i = 1, 2$, $\mathbf{y}_i \triangleq \text{vec}(\mathbf{Y}_i)$, $\mathbf{g}_{ij} \triangleq \text{vec}(\mathbf{G}_{ij})$, $j = i, \bar{i}$, $\tilde{\mathbf{h}}_{ir} \triangleq \text{vec}(\tilde{\mathbf{H}}_{ir})$, $\bar{\mathbf{v}}_i \triangleq \text{vec}(\bar{\mathbf{V}}_i)$, $\mathbf{I}_n$ denotes an $n \times n$ identity matrix. Here $\text{vec}(\cdot)$ denotes the vectorization operator which stacks all column vectors of a matrix on top of each other, and the identity of $\text{vec}(\mathbf{ABC}) = (\mathbf{C}^T \otimes \mathbf{A})\text{vec}(\mathbf{B})$ [17] has been used to obtain (9) from (6). In (10), $\boldsymbol{\gamma}_i \triangleq \left[\mathbf{g}_{ii}^T, \; \mathbf{g}_{i\bar{i}}^T, \; \tilde{\mathbf{h}}_{ir}^T\right]^T$ is the vector of unknown variables at node i with a dimension of $Q_i \triangleq N_i(N_i + N_{\bar{i}} + N_r)$, and $\mathbf{M}_i \triangleq \left[\sqrt{\alpha}\tilde{\mathbf{S}}_i^T\otimes\mathbf{I}_{N_i}, \; \sqrt{\alpha}\tilde{\mathbf{S}}_{\bar{i}}^T\otimes\mathbf{I}_{N_i}, \; \tilde{\mathbf{S}}_r^T\otimes\mathbf{I}_{N_i}\right]$ has a dimension of $TN_i \times Q_i$.

Due to its simplicity, a linear MMSE estimator [19] is applied at node i to estimate $\boldsymbol{\gamma}_i$. We have

$$\hat{\boldsymbol{\gamma}}_i = \mathbf{W}_i^H\mathbf{y}_i, \qquad i = 1, 2 \qquad (11)$$

where $\hat{\boldsymbol{\gamma}}_i$ stands for an estimation of $\boldsymbol{\gamma}_i$ and $\mathbf{W}_i$ is the weight matrix of the MMSE estimator. It can be seen from (11) that

since a linear estimator is used, there is $T \geq N_1 + N_2 + N_r$, and the MSE of estimating $\boldsymbol{\gamma}_i$ can be written as

$$\text{MSE}_i = \text{E}\left[\text{tr}\left((\hat{\boldsymbol{\gamma}}_i - \boldsymbol{\gamma}_i)(\hat{\boldsymbol{\gamma}}_i - \boldsymbol{\gamma}_i)^H\right)\right]$$
$$= \text{tr}\left(\left(\mathbf{W}_i^H\mathbf{M}_i - \mathbf{I}_{Q_i}\right)\mathbf{R}_{\gamma_i}\left(\mathbf{W}_i^H\mathbf{M}_i - \mathbf{I}_{Q_i}\right)^H\right.$$
$$\left. + \mathbf{W}_i^H\mathbf{R}_{\bar{v}_i}\mathbf{W}_i\right), \quad i = 1, 2 \qquad (12)$$

where $\mathbf{R}_{\gamma_i} \triangleq \text{E}[\boldsymbol{\gamma}_i\boldsymbol{\gamma}_i^H]$ is the covariance matrix of $\boldsymbol{\gamma}_i$ and $\mathbf{R}_{\bar{v}_i} \triangleq \text{E}[\bar{\mathbf{v}}_i\bar{\mathbf{v}}_i^H]$ is the noise covariance matrix. Using (2), (8), and Lemma 1, we obtain that

$$\mathbf{R}_{\bar{v}_i} = \mathbf{I}_T\otimes\left(\alpha\text{tr}(\mathbf{K}_r^H\mathbf{K}_r)\mathbf{A}_{ir}\mathbf{A}_{ir}^H + \mathbf{I}_{N_i}\right)$$
$$= \mathbf{I}_T\otimes\left(\alpha\text{tr}(\mathbf{C}_r^T)\mathbf{R}_{ir} + \mathbf{I}_{N_i}\right), \qquad i = 1, 2. \quad (13)$$

Using Lemma 1, $\mathbf{R}_{\gamma_i}$ can be calculated as follows. First, the mth column of $\mathbf{G}_{ij}$ is given by $[\mathbf{G}_{ij}]_m = \lambda_{j,m}^{\frac{1}{2}}\mathbf{A}_{ir}\mathbf{H}_{ir,\text{w}}\mathbf{K}_r^H\mathbf{A}_{rj}\mathbf{H}_{rj,\text{w}}[\mathbf{\Pi}_j]_m$, $m = 1, \cdots, N_j$, where $\lambda_{j,m}$ is the mth diagonal element of $\mathbf{\Lambda}_j$, and $[\mathbf{\Pi}_j]_m$ is the mth column of $\mathbf{\Pi}_j$. Since $\mathbf{H}_{ir,\text{w}}$ and $\mathbf{H}_{rj,\text{w}}$ are independent, the covariance matrix of $[\mathbf{G}_{ij}]_m$ can be calculated as

$$\text{E}\left[[\mathbf{G}_{ij}]_m[\mathbf{G}_{ij}]_m^H\right] = \lambda_{j,m}\text{tr}(\mathbf{K}_r^H\mathbf{A}_{rj}\mathbf{A}_{rj}^H\mathbf{K}_r)\mathbf{A}_{ir}\mathbf{A}_{ir}^H$$
$$= \lambda_{j,m}b_j\mathbf{R}_{ir}, \quad m = 1, \cdots, N_j, \; j = i, \bar{i} \quad (14)$$

where $b_j \triangleq \text{tr}(\mathbf{R}_{rj}\mathbf{C}_r^T)$. Second, the covariance matrix of the mth column of $\tilde{\mathbf{H}}_{ir}$, denoted as $[\tilde{\mathbf{H}}_{ir}]_m$, is given by

$$\text{E}\left[[\tilde{\mathbf{H}}_{ir}]_m[\tilde{\mathbf{H}}_{ir}]_m^H\right] = \lambda_{r,m}\mathbf{R}_{ir}, \qquad m = 1, \cdots, N_r \quad (15)$$

where $\lambda_{r,m}$ is the mth diagonal element of $\mathbf{\Lambda}_r$. From (14) and (15), $\mathbf{R}_{\gamma_i}$ can be written as

$$\mathbf{R}_{\gamma_i} = \text{Bdiag}\left[\mathbf{\Lambda}_i\otimes b_i\mathbf{R}_{ir}, \; \mathbf{\Lambda}_{\bar{i}}\otimes b_{\bar{i}}\mathbf{R}_{ir}, \; \mathbf{\Lambda}_r\otimes\mathbf{R}_{ir}\right], \quad i = 1, 2 \qquad (16)$$

where $\text{Bdiag}[\cdot]$ denotes a block diagonal matrix.

The matrix $\mathbf{W}_i$ minimizing MSE_i in (12) is given by

$$\mathbf{W}_i = \left(\mathbf{M}_i\mathbf{R}_{\gamma_i}\mathbf{M}_i^H + \mathbf{R}_{\bar{v}_i}\right)^{-1}\mathbf{M}_i\mathbf{R}_{\gamma_i}, \quad i = 1, 2 \qquad (17)$$

where $(\cdot)^{-1}$ denotes matrix inversion. Substituting (17) back into (12), and using the matrix inversion lemma of $(\mathbf{A} + \mathbf{BCD})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{B}(\mathbf{D}\mathbf{A}^{-1}\mathbf{B} + \mathbf{C}^{-1})^{-1}\mathbf{D}\mathbf{A}^{-1}$, the MSE of estimating $\boldsymbol{\gamma}_i$ can be obtained as

$$\text{MSE}_i = \text{tr}\left(\left[\mathbf{R}_{\gamma_i}^{-1} + \mathbf{M}_i^H\mathbf{R}_{\bar{v}_i}^{-1}\mathbf{M}_i\right]^{-1}\right), \quad i = 1, 2. \quad (18)$$

The transmission power consumed at nodes 1 and 2 is

$$\text{tr}(\mathbf{S}_i\mathbf{S}_i^H) = \text{tr}(\tilde{\mathbf{S}}_i\tilde{\mathbf{S}}_i^H), \qquad i = 1, 2. \qquad (19)$$

From (4), the power consumed at the relay node is given by

$$\alpha\text{E}\left[\text{tr}\left(\sum_{i=1}^{2}\mathbf{H}_{ri}\mathbf{S}_i\mathbf{S}_i^H\mathbf{H}_{ri}^H + \mathbf{I}_{N_r}\right)\right] + \text{tr}(\mathbf{S}_r\mathbf{S}_r^H)$$
$$= \alpha N_r + \alpha\sum_{i=1}^{2}\text{tr}(\mathbf{\Lambda}_i\tilde{\mathbf{S}}_i\tilde{\mathbf{S}}_i^H)\text{tr}(\mathbf{R}_{ri}) + \text{tr}(\tilde{\mathbf{S}}_r\tilde{\mathbf{S}}_r^H). \quad (20)$$

From (18)-(20), the optimal training matrices and the optimal α can be designed by solving the following optimization

problem

$$\min_{\alpha, \tilde{\mathbf{S}}_1, \tilde{\mathbf{S}}_2, \tilde{\mathbf{S}}_r} \sum_{i=1}^{2} \text{tr}\left(\left[\mathbf{R}_{\gamma_i}^{-1} + \mathbf{M}_i^H \mathbf{R}_{\bar{v}_i}^{-1} \mathbf{M}_i \right]^{-1} \right) \tag{21}$$

$$\text{s.t. } \text{tr}(\tilde{\mathbf{S}}_i \tilde{\mathbf{S}}_i^H) \leq p_i, \qquad i = 1, 2 \tag{22}$$

$$\alpha \left[N_r + \sum_{i=1}^{2} \text{tr}(\mathbf{\Lambda}_i \tilde{\mathbf{S}}_i \tilde{\mathbf{S}}_i^H) \text{tr}(\mathbf{R}_{ri}) \right] + \text{tr}(\tilde{\mathbf{S}}_r \tilde{\mathbf{S}}_r^H) \leq p_r \tag{23}$$

where p_i is the transmission power available at node i, $i = 1, 2$, and p_r is the transmission power available at the relay node. The following theorem establishes the optimal structure of $\mathbf{S}_1$, $\mathbf{S}_2$, and $\mathbf{S}_r$ as solution to the problem (21)-(23).

THEOREM 1: The optimal training sequences $\mathbf{S}_1$, $\mathbf{S}_2$, and $\mathbf{S}_r$ satisfy $\mathbf{S}_i \mathbf{S}_i^H = \mathbf{U}_i \mathbf{\Sigma}_i \mathbf{U}_i^H$, $i = 1, 2, r$, and $\mathbf{S}_i \mathbf{S}_j^H = \mathbf{0}$, $i, j = 1, 2, r$, $i \neq j$, where $\mathbf{\Sigma}_i$, $i = 1, 2, r$, is an $N_i \times N_i$ diagonal matrices.

PROOF: See Appendix A. $\qquad\square$

The optimal structure of $\mathbf{S}_i$, $i = 1, 2, r$, can be obtained from Theorem 1 as $\mathbf{S}_i = \mathbf{U}_i \mathbf{\Sigma}_i^{\frac{1}{2}} \mathbf{\Omega}_i$, where $\mathbf{\Omega}_i$ is an $N_i \times T$ semi-unitary matrix satisfying $\mathbf{\Omega}_i \mathbf{\Omega}_i^H = \mathbf{I}_{N_i}$, $i = 1, 2, r$, and $\mathbf{\Omega}_i \mathbf{\Omega}_j^H = \mathbf{0}$, $i, j = 1, 2, r$, $i \neq j$. Such $\mathbf{\Omega}_r$, $\mathbf{\Omega}_1$, and $\mathbf{\Omega}_2$ can be easily constructed, for example, from the normalized discrete Fourier transform (DFT) matrix when $T \geq N_1 + N_2 + N_r$.

Interestingly, it can be seen that the optimal training matrix at node i matches the eigenvector matrix of the transmitter correlation matrix of $\mathbf{H}_{ri}$, and the optimal training matrix at the relay node matches the eigenvector matrix of $\mathbf{C}_r^T$. Using Theorem 1 and (81) in Appendix A, the problem (21)-(23) is equivalently converted to the following problem

$$\min_{\alpha, \mathbf{\Sigma}_1, \mathbf{\Sigma}_2, \mathbf{\Sigma}_r} \sum_{i=1}^{2} \text{tr}\Bigg(\sum_{j=1}^{2} \left[\mathbf{D}_{ij} + \alpha \mathbf{\Sigma}_j \otimes \mathbf{D}_{ri} \right]^{-1}$$
$$+ \left[\mathbf{D}_{si} + \mathbf{\Sigma}_r \otimes \mathbf{D}_{ri} \right]^{-1} \Bigg) \tag{24}$$

$$\text{s.t. } \text{tr}(\mathbf{\Sigma}_i) \leq p_i, \qquad i = 1, 2 \tag{25}$$

$$\alpha N_r + \alpha \sum_{i=1}^{2} \text{tr}(\mathbf{\Lambda}_i \mathbf{\Sigma}_i) \text{tr}(\mathbf{R}_{ri}) + \text{tr}(\mathbf{\Sigma}_r) \leq p_r \tag{26}$$

$$\alpha > 0, \qquad \mathbf{\Sigma}_i \geq 0, \quad i = 1, 2, r \tag{27}$$

where for a matrix $\mathbf{A}$, $\mathbf{A} \geq 0$ means that $\mathbf{A}$ is a positive semi-definite (PSD) matrix. Using the definition of $\mathbf{D}_{ij}$, $\mathbf{D}_{si}$, and $\mathbf{D}_{ri}$ in (78) in Appendix A, the problem (24)-(27) can be equivalently rewritten as the following problem with scalar variables

$$\min_{\alpha, \boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2, \boldsymbol{\sigma}_r} \sum_{i=1}^{2} \Bigg(\sum_{j=1}^{2} \sum_{m=1}^{N_j} \sum_{n=1}^{N_i} \left[\frac{1}{b_j \lambda_{j,m} \delta_{i,n}} + \alpha d_{i,n} \sigma_{j,m} \right]^{-1}$$
$$+ \sum_{m=1}^{N_r} \sum_{n=1}^{N_i} \left[\frac{1}{\lambda_{r,m} \delta_{i,n}} + d_{i,n} \sigma_{r,m} \right]^{-1} \Bigg) \tag{28}$$

$$\text{s.t. } \sum_{m=1}^{N_i} \sigma_{i,m} \leq p_i, \quad i = 1, 2 \tag{29}$$

$$\alpha N_r + \alpha \sum_{i=1}^{2} \left(\sum_{m=1}^{N_i} \lambda_{i,m} \sigma_{i,m} \text{tr}(\mathbf{R}_{ri}) \right)$$
$$+ \sum_{m=1}^{N_r} \sigma_{r,m} \leq p_r \tag{30}$$

$$\alpha > 0, \quad \sigma_{i,m} \geq 0, \quad m = 1, \cdots, N_i, \quad i = 1, 2, r \tag{31}$$

where $d_{i,n} \triangleq 1/(\alpha \text{tr}(\mathbf{C}_r^T) \delta_{i,n} + 1)$, $n = 1, \cdots, N_i$, $i = 1, 2$, $\boldsymbol{\sigma}_i \triangleq \left[\sigma_{i,1}, \cdots, \sigma_{i,N_i} \right]^T$, $i = 1, 2, r$, $\delta_{i,m}$, $i = 1, 2$, is the mth diagonal element of $\mathbf{\Delta}_i$, and $\lambda_{i,m}$, $\sigma_{i,m}$, $i = 1, 2, r$, are the mth diagonal element of $\mathbf{\Lambda}_i$ and $\mathbf{\Sigma}_i$, respectively.

Given that $b_j, \lambda_{j,m}, \delta_{i,n}, d_{i,n}$, and $\lambda_{r,m}$ are known variables with fixed α, the objective function (28) can be rewritten as

$$\min_{\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2, \boldsymbol{\sigma}_r} \sum_{i=1}^{2} \Bigg(\sum_{j=1}^{2} \sum_{m=1}^{N_j} \sum_{n=1}^{N_i} \frac{1}{a_{i,j,m,n} + c_{i,n} \sigma_{j,m}}$$
$$+ \sum_{m=1}^{N_r} \sum_{n=1}^{N_i} \frac{1}{g_{i,m,n} + d_{i,n} \sigma_{r,m}} \Bigg)$$

where $a_{i,j,m,n} \triangleq 1/(b_j \lambda_{j,m} \delta_{i,n})$, $c_{i,n} \triangleq \alpha d_{i,n}$, and $g_{i,m,n} \triangleq 1/(\lambda_{r,m} \delta_{i,n})$ are known variables. It can be seen from the above equation that the triple summation terms and the double summation terms are monotonically decreasing and convex with respect to $\sigma_{j,m}$ and $\sigma_{r,m}$, respectively. Moreover, with fixed α, the constraints in (29) and (30) are linear inequality constraints which can be rewritten as $\mathbf{1}^T \boldsymbol{\sigma}_i \leq p_i$, $i = 1, 2$, and $\bar{\mathbf{z}}_1^T \boldsymbol{\sigma}_1 + \bar{\mathbf{z}}_2^T \boldsymbol{\sigma}_2 + \mathbf{1}^T \boldsymbol{\sigma}_r \leq p_r - \alpha N_r$, respectively, where $\bar{\mathbf{z}}_i \triangleq \alpha \text{tr}(\mathbf{R}_{ri})[\lambda_{i,1}, \cdots, \lambda_{i,N_i}]^T$, $i = 1, 2$, and $\mathbf{1}$ is a vector of all ones with a commensurate dimension. Therefore, the problem (28)-(31) with respect to $\boldsymbol{\sigma}_1$, $\boldsymbol{\sigma}_2$, and $\boldsymbol{\sigma}_r$ is a convex optimization problem when α is fixed, where the optimal $\boldsymbol{\sigma}_1$, $\boldsymbol{\sigma}_2$, and $\boldsymbol{\sigma}_r$ can be efficiently obtained through the Karush-Kuhn-Tucker (KKT) optimality conditions of the problem (28)-(31). In particular, the gradient conditions are given by

$$\sum_{j=1}^{2} \sum_{n=1}^{N_j} \frac{\alpha d_{j,n}}{\left[(b_i \lambda_{i,m} \delta_{j,n})^{-1} + \alpha d_{j,n} \sigma_{i,m} \right]^2} = \mu_i + \mu_3 e_{i,m},$$
$$m = 1, \cdots, N_i, \quad i = 1, 2 \tag{32}$$

$$\sum_{i=1}^{2} \sum_{n=1}^{N_i} \frac{d_{i,n}}{\left[(\lambda_{r,m} \delta_{i,n})^{-1} + d_{i,n} \sigma_{r,m} \right]^2} = \mu_3,$$
$$m = 1, \cdots, N_r \tag{33}$$

where $e_{i,m} \triangleq \alpha \text{tr}(\mathbf{R}_{ri}) \lambda_{i,m}$, $i = 1, 2$, and $\mu_i \geq 0$, $i = 1, 2, 3$, are Lagrange multipliers such that the complementary slackness conditions [20] given by

$$\mu_i \left(p_i - \sum_{m=1}^{N_i} \sigma_{i,m} \right) = 0, \qquad i = 1, 2 \tag{34}$$

$$\mu_3 \left(p_r - \alpha N_r - \alpha \sum_{i=1}^{2} \sum_{m=1}^{N_i} \lambda_{i,m} \sigma_{i,m} \text{tr}(\mathbf{R}_{ri}) - \sum_{m=1}^{N_r} \sigma_{r,m} \right) = 0 \tag{35}$$

are satisfied.

With fixed α and μ_i, $i = 1, 2, 3$, for each m, the non-negative $\sigma_{1,m}$, $\sigma_{2,m}$, and $\sigma_{r,m}$ can be found by using the bi-section search, since the left-hand-side (LHS) of (32) and (33) are monotonically decreasing function of $\sigma_{i,m}$ and $\sigma_{r,m}$,

respectively. To find the optimal μ_i, $i = 1, 2, 3$, an outer bi-section search loop is used as the LHS of (29) is an increasing function of $\sigma_{i,m}$, and the LHS of (30) is an increasing function of $\sigma_{1,m}$, $\sigma_{2,m}$, and $\sigma_{r,m}$, while in (32), $\sigma_{i,m}$ is a monotonically decreasing function of μ_i and μ_3, and $\sigma_{r,m}$ is a monotonically decreasing function of μ_3 in (33).

When α is an optimization variable (not fixed), the problem (28)-(31) as a whole is not a convex optimization problem. However, we can show that (28) subjecting to (29)-(31) is a unimodal (quasi-convex) function with respect to α. Let us introduce $\chi_{i,m} \triangleq \alpha\sigma_{i,m}$, $m = 1, \cdots, N_i, i = 1, 2$, the problem (28)-(31) can be rewritten as

$$\min_{\alpha, \boldsymbol{\chi}_1, \boldsymbol{\chi}_2, \boldsymbol{\sigma}_r} \sum_{i=1}^{2} \left(\sum_{j=1}^{2} \sum_{m=1}^{N_j} \sum_{n=1}^{N_i} \frac{1}{a_{i,j,m,n} + d_{i,n}\chi_{j,m}} \right.$$
$$\left. + \sum_{m=1}^{N_r} \sum_{n=1}^{N_i} \frac{1}{g_{i,m,n} + d_{i,n}\sigma_{r,m}} \right) \quad (36)$$

s.t. $\mathbf{1}^T \boldsymbol{\chi}_i \leq \alpha p_i, \qquad i = 1, 2$ (37)

$\mathbf{z}_1^T \boldsymbol{\chi}_1 + \mathbf{z}_2^T \boldsymbol{\chi}_2 + \mathbf{1}^T \boldsymbol{\sigma}_r \leq p_r - \alpha N_r$ (38)

$\alpha > 0, \ \sigma_{r,m} \geq 0, \ \chi_{i,m} \geq 0, \ m = 1, \cdots, N_i, \ i = 1, 2$ (39)

where $\boldsymbol{\chi}_i \triangleq [\chi_{i,1}, \cdots, \chi_{i,N_i}]^T$, and $\mathbf{z}_i \triangleq \mathrm{tr}(\mathbf{R}_{ri})[\lambda_{i,1}, \cdots, \lambda_{i,N_i}]^T$, $i = 1, 2$.

Let us first ignore the effect of all $d_{i,n}$ by treating them as known variables. Then the problem (36)-(39) is a convex optimization problem, since (36) is a convex function of $\boldsymbol{\chi}_1$, $\boldsymbol{\chi}_2$, $\boldsymbol{\sigma}_r$, and (37)-(39) are linear inequality constraints. In particular, with increasing α, the value of (36) first decreases and then increases based on the following reasons. For a significantly small α, the value of (36) is strongly governed by the constraints in (37), since constraint (38) is inactive compared with those in (37) when α is small. Once α increases from a small value, the feasible region specified by (37) expands, and thus, the value of (36) decreases. On the other hand, when α is large, the value of (36) is strongly governed by the constraint in (38), since constraints in (37) are inactive compared with that in (38) when α is large. Once α decreases from a large value, the feasible region specified by (38) expands, resulting in a deceasing of the value of (36). Now we consider the effect of $d_{i,n}$. Since $d_{i,n} = 1/(\alpha\mathrm{tr}(\mathbf{C}_r^T)\delta_{i,n}+1)$, $d_{i,n}$ monotonically decreases with increasing α, and (36) increases when $d_{i,n}$ decreases.

Considering the two effects above, we can draw the following conclusion regarding the value of (36) with respect to α. When α increases from a significantly small positive number, the value of (36) starts to decrease since the potential decrease of (36) due to the expanded feasible region (37) dominates the potential increase of (36) caused by the decreasing $d_{i,n}$. The value of (36) keeps decreasing as α increases till a 'turning point' where the decreasing of $d_{i,n}$ starts to dominate the effect of relaxed feasible region (37). After such turning point, the value of (36) will monotonically increase with an increasing α.

To validate the analysis above, a plot of the MSE value (28) over a range of feasible values of α is generated in Fig. 2 for the case where all nodes have the same number of antennas,

i.e., $N_i = N = 4$, $i = 1, 2, r$, and the channel matrices have i.i.d. entries, i.e., $\mathbf{T}_{ri} = \mathbf{R}_{ri} = \mathbf{R}_{ir} = \mathbf{C}_r = \mathbf{I}_N$, $i = 1, 2$. Fig. 2 shows the normalized MSE (NMSE), which is (28) divided by $6N^2$, versus α for different $p_1 = p_2$, and p_r is set to be 20dB. It can be observed from Fig. 2 that (28) is a unimodal (quasi-convex) function of α. Thus, the optimal α for the problem (28)-(31) can be efficiently found by applying the golden section search (GSS) technique described in Table I, where $|\cdot|$ denotes the absolute value, 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. The GSS method can guarantee that the minimum of a unimodal function to be found by bracketing the minimum to an interval of 0.618 times the size of the preceding interval. Unlike the Fibonacci search, the GSS method is able to perform up to the desired accuracy and does not require the number of iterations as input. However, the GSS method may need more iterations compared with the Fibonacci search.
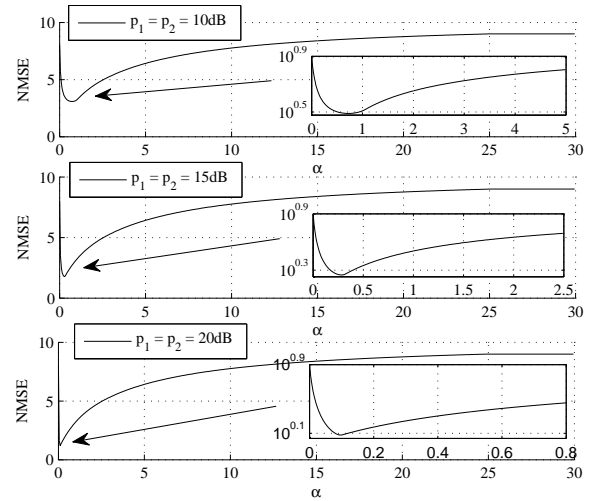

Fig. 2. Superimposed channel training: NMSE versus α for different $p_1 = p_2$ with $N = 4$ and $p_r = 20$dB.

TABLE I
PROCEDURE OF APPLYING THE GOLDEN SECTION SEARCH (GSS) TO FIND THE OPTIMAL α IN THE PROBLEM (28)-(31).

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 (28)-(31) for $\alpha = c_1$;
   Compute the MSE value defined in (28), $f_{MSE}(c_1)$ 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.

The complexity of the superimposed channel training algorithm can be estimated as $\mathcal{O}(c_\alpha c_\mu c_\sigma (N_1 + N_2)(N_1 + N_2 + N_r))$, where c_α is the number of GSS iterations required to obtain the optimal α, c_μ stands for the number of iterations in the outer bi-section loop to obtain the optimal μ_1, μ_2, and μ_3, and c_σ represents the number of bi-section operations required to obtain the optimal $\boldsymbol{\sigma}_1$, $\boldsymbol{\sigma}_2$, and $\boldsymbol{\sigma}_r$.

IV. TWO-STAGE CHANNEL ESTIMATION ALGORITHM

There are two stages in this channel estimation scheme. In particular, the channel matrices $\mathbf{H}_{ir}$, $i = 1, 2$, from the relay node to the receive nodes are estimated in the first stage, while the channel matrices $\mathbf{H}_{ri}$, $i = 1, 2$, from the source nodes to the relay node are estimated in the second stage. The first stage requires one time block while the second stage requires two time blocks.

A. Stage One

At the first stage, the relay node transmits an $N_r \times T_1$ training signal matrix $\mathbf{P}_r$ to both receive nodes, where T_1 is the length of the training sequence and will be determined later. The $N_i \times T_1$ received signal matrix $\mathbf{Y}_{i,1}$ at node i is given by

$$\mathbf{Y}_{i,1} = \mathbf{H}_{ir}\mathbf{P}_r + \mathbf{V}_{i,1}, \qquad i = 1, 2 \tag{40}$$

where $\mathbf{V}_{i,1}$ is an $N_i \times T_1$ noise matrix at node i in stage one. By vectorizing both sides of (40), we obtain

$$\mathbf{y}_{i,1} = (\mathbf{P}_r^T \otimes \mathbf{I}_{N_i})\mathbf{h}_{ir} + \mathbf{v}_{i,1}, \qquad i = 1, 2 \tag{41}$$

where $\mathbf{y}_{i,1} \triangleq \text{vec}(\mathbf{Y}_{i,1})$, $\mathbf{h}_{ir} \triangleq \text{vec}(\mathbf{H}_{ir})$, and $\mathbf{v}_{i,1} \triangleq \text{vec}(\mathbf{V}_{i,1})$.

Using a linear MMSE estimator at node i to estimate $\mathbf{h}_{ir}$, we obtain

$$\hat{\mathbf{h}}_{ir} = \mathbf{W}_{i,1}^H \mathbf{y}_{i,1}, \qquad i = 1, 2 \tag{42}$$

where $\hat{\mathbf{h}}_{ir}$ denotes an estimation of $\mathbf{h}_{ir}$ and $\mathbf{W}_{i,1}$ is the weight matrix of the MMSE estimator given by

$$\mathbf{W}_{i,1} = \mathbf{R}_{i,1}^{-1}\mathbf{C}_{i,1}, \qquad i = 1, 2. \tag{43}$$

Here $\mathbf{R}_{i,1} \triangleq \text{E}\big[\mathbf{y}_{i,1}\mathbf{y}_{i,1}^H\big] = (\mathbf{P}_r^T\mathbf{C}_r\mathbf{P}_r^*)\otimes\mathbf{R}_{ir}+\mathbf{I}_{N_iT_1}$, $\mathbf{C}_{i,1} \triangleq \text{E}\big[\mathbf{y}_{i,1}\mathbf{h}_{ir}^H\big] = (\mathbf{P}_r^T\mathbf{C}_r) \otimes \mathbf{R}_{ir}$, $i = 1, 2$, and $(\cdot)^*$ stands for complex conjugate. From (41) and (42), we find that since a linear estimator is used, there is $T_1 \geq N_r$. Using (41)-(43), the MSE of estimating $\mathbf{h}_{ir}$ can be written as

$$\begin{aligned}\text{MSE}_{i,1} &= \text{E}\big[\text{tr}\big((\hat{\mathbf{h}}_{ir} - \mathbf{h}_{ir})(\hat{\mathbf{h}}_{ir} - \mathbf{h}_{ir})^H\big)\big] \\ &= \text{tr}\big([(\mathbf{C}_r\otimes\mathbf{R}_{ir})^{-1}+(\mathbf{P}_r^*\mathbf{P}_r^T)\otimes\mathbf{I}_{N_i}]^{-1}\big), \quad i = 1, 2.\end{aligned}$$

Since the transmission power consumed by the relay node at stage one is $\text{tr}(\mathbf{P}_r\mathbf{P}_r^H)$, the optimal $\mathbf{P}_r$ can be derived by solving the following optimization problem

$$\min_{\mathbf{P}_r} \sum_{i=1}^2 \text{tr}\Big([(\mathbf{C}_r\otimes\mathbf{R}_{ir})^{-1}+(\mathbf{P}_r^*\mathbf{P}_r^T) \otimes \mathbf{I}_{N_i}]^{-1}\Big) \tag{44}$$

$$\text{s.t. } \text{tr}(\mathbf{P}_r\mathbf{P}_r^H) \leq q_{r,1} \tag{45}$$

where $q_{r,1}$ is the power allocation at the relay node at the first stage. The following theorem establishes the optimal structure of $\mathbf{P}_r$ as the solution to the problem (44)-(45).

THEOREM 2: The optimal training sequence $\mathbf{P}_r$ satisfies $\mathbf{P}_r\mathbf{P}_r^H = \mathbf{U}_r\boldsymbol{\Xi}_r\mathbf{U}_r^H$, where $\boldsymbol{\Xi}_r$ is an $N_r \times N_r$ diagonal matrix.

PROOF: Similar to the proof of Theorem 1. □

The optimal structure of $\mathbf{P}_r$ can be obtained from Theorem 2 as $\mathbf{P}_r = \mathbf{U}_r\boldsymbol{\Xi}_r^{\frac{1}{2}}\boldsymbol{\Omega}_{r,1}$, where $\boldsymbol{\Omega}_{r,1}$ is an $N_r \times T_1$ semi-unitary matrix satisfying $\boldsymbol{\Omega}_{r,1}\boldsymbol{\Omega}_{r,1}^H = \mathbf{I}_{N_r}$ and can be easily constructed from the normalized DFT matrix when $T_1 \geq N_r$.

Using Theorem 2, the problem (44)-(45) is equivalently converted to the following problem

$$\min_{\boldsymbol{\Xi}_r} \sum_{i=1}^2 \text{tr}\Big([(\boldsymbol{\Lambda}_r^{-1} \otimes \boldsymbol{\Lambda}_{ir}^{-1}) + (\boldsymbol{\Xi}_r \otimes \mathbf{I}_{N_i})]^{-1}\Big) \tag{46}$$

$$\text{s.t. } \text{tr}(\boldsymbol{\Xi}_r) \leq q_{r,1}, \qquad \boldsymbol{\Xi}_r \geq 0. \tag{47}$$

The problem (46)-(47) can be equivalently rewritten as the following problem with scalar variables

$$\min_{\boldsymbol{\xi}_r} \sum_{i=1}^2 \sum_{m=1}^{N_r} \sum_{n=1}^{N_i} \frac{1}{\lambda_{r,m}^{-1}\lambda_{ir,n}^{-1} + \xi_{r,m}} \tag{48}$$

$$\text{s.t. } \sum_{m=1}^{N_r} \xi_{r,m} \leq q_{r,1}, \qquad \xi_{r,m} \geq 0, \quad m = 1, \cdots, N_r \tag{49}$$

where $\boldsymbol{\xi}_r \triangleq \big[\xi_{r,1}, \cdots, \xi_{r,N_r}\big]^T$ and $\xi_{r,m}$ is the mth diagonal element of $\boldsymbol{\Xi}_r$.

The problem (48)-(49) is convex and thus can be efficiently solved through the KKT optimality conditions. The gradient condition is given by

$$\sum_{i=1}^2 \sum_{n=1}^{N_i} \frac{1}{\big[\lambda_{r,m}^{-1}\lambda_{ir,n}^{-1} + \xi_{r,m}\big]^2} = \mu, \qquad m = 1, \cdots, N_r \tag{50}$$

where $\mu \geq 0$ is the Lagrange multiplier such that the complementary slackness condition $\mu\big(q_{r,1}-\sum_{m=1}^{N_r} \xi_{r,m}\big) = 0$ is satisfied. For each m, with fixed μ, the non-negative $\xi_{r,m}$ can be found using the bi-section search, since the LHS of (50) is a monotonically decreasing function of $\xi_{r,m}$. To find the optimal μ, an outer bi-section search is used as the LHS of (49) is an increasing function of $\xi_{r,m}$, while in (50), $\xi_{r,m}$ is a monotonically decreasing function of μ.

B. Stage Two

At the second stage, the source node i transmits an $N_i \times T_2$ training signal matrix $\mathbf{P}_i$ to the relay node. The $N_r \times T_2$ received signal matrix $\mathbf{Y}_{r,2}$ at the relay node is given by

$$\mathbf{Y}_{r,2} = \sum_{i=1}^2 \mathbf{H}_{ri}\mathbf{P}_i + \mathbf{V}_{r,2}$$

where $\mathbf{V}_{r,2}$ is an $N_r \times T_2$ noise matrix at the relay node. Then the relay node amplifies $\mathbf{Y}_{r,2}$ and retransmits $\mathbf{X}_{r,2} = \sqrt{\eta}\mathbf{Y}_{r,2}$, where $\eta > 0$ is the relay amplifying factor. The $N_i \times T_2$ received signal matrix at node i is given by

$$\begin{aligned}\mathbf{Y}_{i,2} &= \mathbf{H}_{ir}\mathbf{X}_{r,2} + \mathbf{V}_{i,2} \\ &= \sqrt{\eta}\mathbf{H}_{ir}\mathbf{H}_{ri}\mathbf{P}_i + \sqrt{\eta}\mathbf{H}_{ir}\mathbf{H}_{r\bar{i}}\mathbf{P}_{\bar{i}} + \sqrt{\eta}\mathbf{H}_{ir}\mathbf{V}_{r,2} \\ &\quad + \mathbf{V}_{i,2}, \qquad i = 1, 2\end{aligned} \tag{51}$$

where $\mathbf{V}_{i,2}$ is an $N_i \times T_2$ noise matrix at node i.

Introducing $\tilde{\mathbf{P}}_j \triangleq \mathbf{U}_j^H\mathbf{P}_j$, $j = i, \bar{i}$, $\bar{\mathbf{V}}_{i,2} \triangleq \sqrt{\eta}\mathbf{H}_{ir}\mathbf{V}_{r,2} + \mathbf{V}_{i,2}$, $i = 1, 2$, we can rewrite (51) as

$$\mathbf{Y}_{i,2} = \sqrt{\eta}\mathbf{G}_{ii}\tilde{\mathbf{P}}_i + \sqrt{\eta}\mathbf{G}_{i\bar{i}}\tilde{\mathbf{P}}_{\bar{i}} + \bar{\mathbf{V}}_{i,2}, \qquad i = 1, 2 \tag{52}$$

where $\mathbf{G}_{ij}$ is defined in (7). Similar to Section III, we first estimate $\mathbf{G}_{ij}$. Then an estimation of $\mathbf{H}_{rj}$ is obtained as $\breve{\mathbf{H}}_{rj} = \breve{\mathbf{H}}_{ir}^\dagger\breve{\mathbf{G}}_{ij}\mathbf{U}_j^H$, $j = i, \bar{i}$, where $\breve{\mathbf{H}}_{ir}$ is the estimation of $\mathbf{H}_{ir}$

obtained from stage one and $\breve{\mathbf{G}}_{ij}$ is the estimation of $\mathbf{G}_{ij}$. By vectorizing both sides of (52), we obtain

$$\begin{aligned}
\mathbf{y}_{i,2} &= \left[\sqrt{\eta}\tilde{\mathbf{P}}_i^T \otimes \mathbf{I}_{N_i}, \ \sqrt{\eta}\tilde{\mathbf{P}}_{\bar{i}}^T \otimes \mathbf{I}_{N_i}\right]\left[\mathbf{g}_{ii}^T, \ \mathbf{g}_{i\bar{i}}^T\right]^T + \bar{\mathbf{v}}_{i,2} \\
&\triangleq \mathbf{N}_i\boldsymbol{\theta}_i + \bar{\mathbf{v}}_{i,2}, \qquad i = 1, 2
\end{aligned} \quad (53)$$

where $\mathbf{y}_{i,2} \triangleq \mathrm{vec}(\mathbf{Y}_{i,2})$, $\bar{\mathbf{v}}_{i,2} \triangleq \mathrm{vec}(\bar{\mathbf{V}}_{i,2})$, $\mathbf{N}_i \triangleq \left[\sqrt{\eta}\tilde{\mathbf{P}}_i^T \otimes \mathbf{I}_{N_i}, \ \sqrt{\eta}\tilde{\mathbf{P}}_{\bar{i}}^T \otimes \mathbf{I}_{N_i}\right]$, and $\boldsymbol{\theta}_i \triangleq \left[\mathbf{g}_{ii}^T, \ \mathbf{g}_{i\bar{i}}^T\right]^T$ is the vector of unknown variables at node i.

Using a linear MMSE receiver to estimate $\boldsymbol{\theta}_i$, we have

$$\hat{\boldsymbol{\theta}}_i = \mathbf{W}_{i,2}^H\mathbf{y}_{i,2}, \qquad i = 1, 2 \quad (54)$$

where $\hat{\boldsymbol{\theta}}_i$ stands for an estimation of $\boldsymbol{\theta}_i$, $\mathbf{W}_{i,2}$ is the weight matrix of the MMSE estimator and given by

$$\mathbf{W}_{i,2} = \left(\mathbf{N}_i\mathbf{R}_{\theta_i}\mathbf{N}_i^H + \mathbf{R}_{\bar{v}_{i,2}}\right)^{-1}\mathbf{N}_i\mathbf{R}_{\theta_i}, \qquad i = 1, 2. \quad (55)$$

From (53) and (54), we find that since a linear estimator is used, there is $T_2 \geq N_1 + N_2$. In (55), $\mathbf{R}_{\theta_i} \triangleq \mathrm{E}[\boldsymbol{\theta}_i\boldsymbol{\theta}_i^H]$ is the covariance matrix of $\boldsymbol{\theta}_i$, which can be calculated similar to $\mathbf{R}_{\gamma_i}$ in (16) and written as

$$\mathbf{R}_{\theta_i} = \mathrm{Bdiag}\left[\boldsymbol{\Lambda}_i \otimes b_i\mathbf{R}_{ir}, \ \boldsymbol{\Lambda}_{\bar{i}} \otimes b_{\bar{i}}\mathbf{R}_{ir}\right], \qquad i = 1, 2.$$

In (55), $\mathbf{R}_{\bar{v}_{i,2}} \triangleq \mathrm{E}[\bar{\mathbf{v}}_{i,2}\bar{\mathbf{v}}_{i,2}^H]$ is the noise covariance matrix which can be calculated similar to $\mathbf{R}_{\bar{v}_i}$ (13) as

$$\mathbf{R}_{\bar{v}_{i,2}} = \mathbf{I}_{T_2} \otimes \left(\eta\mathrm{tr}(\mathbf{C}_r^T)\mathbf{R}_{ir} + \mathbf{I}_{N_i}\right), \qquad i = 1, 2.$$

Using (55), the MSE of estimating $\boldsymbol{\theta}_i$ can be obtained as

$$\begin{aligned}
\mathrm{MSE}_{i,2} &= \mathrm{E}\left[\mathrm{tr}\left((\hat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_i)(\hat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_i)^H\right)\right] \\
&= \mathrm{tr}\left(\left[\mathbf{R}_{\theta_i}^{-1} + \mathbf{N}_i^H\mathbf{R}_{\bar{v}_{i,2}}^{-1}\mathbf{N}_i\right]^{-1}\right), \quad i = 1, 2. \quad (56)
\end{aligned}$$

The transmission power consumed at nodes 1 and 2 is

$$\mathrm{tr}(\mathbf{P}_i\mathbf{P}_i^H) = \mathrm{tr}(\tilde{\mathbf{P}}_i\tilde{\mathbf{P}}_i^H), \qquad i = 1, 2. \quad (57)$$

And the power consumed at the relay node is given by

$$\begin{aligned}
&\eta\mathrm{E}\left[\mathrm{tr}\left(\sum_{i=1}^{2}\mathbf{H}_{ri}\mathbf{P}_i\mathbf{P}_i^H\mathbf{H}_{ri}^H + \mathbf{I}_{N_r}\right)\right] \\
&= \eta N_r + \eta\sum_{i=1}^{2}\mathrm{tr}(\boldsymbol{\Lambda}_i\tilde{\mathbf{P}}_i\tilde{\mathbf{P}}_i^H)\mathrm{tr}(\mathbf{R}_{ri}). \quad (58)
\end{aligned}$$

From (56)-(58), the optimal training matrices $\mathbf{P}_i$, $i = 1, 2$, and the optimal η can be obtained through solving the following optimization problem

$$\min_{\eta,\tilde{\mathbf{P}}_1,\tilde{\mathbf{P}}_2} \sum_{i=1}^{2}\mathrm{tr}\left(\left[\mathbf{R}_{\theta_i}^{-1} + \mathbf{N}_i^H\mathbf{R}_{\bar{v}_{i,2}}^{-1}\mathbf{N}_i\right]^{-1}\right) \quad (59)$$

$$\text{s.t. } \mathrm{tr}(\tilde{\mathbf{P}}_i\tilde{\mathbf{P}}_i^H) \leq q_i, \qquad i = 1, 2 \quad (60)$$

$$\eta\left[N_r + \sum_{i=1}^{2}\mathrm{tr}(\boldsymbol{\Lambda}_i\tilde{\mathbf{P}}_i\tilde{\mathbf{P}}_i^H)\mathrm{tr}(\mathbf{R}_{ri})\right] \leq q_{r,2} \quad (61)$$

where q_i is the transmission power available at node i, $i = 1, 2$, and $q_{r,2}$ is the transmission power available at the relay node at the second stage. Note that for a fair comparison with the superimposed channel training algorithm, the power at three nodes should satisfy

$$q_{r,1}T_1 + q_{r,2}T_2 = p_rT, \qquad q_iT_2 = p_iT, \qquad i = 1, 2. \quad (62)$$

The following theorem establishes the optimal structure of $\mathbf{P}_1$ and $\mathbf{P}_2$ as the solution to the problem (59)-(61).

THEOREM 3: The optimal training sequences $\mathbf{P}_1$ and $\mathbf{P}_2$ satisfy $\mathbf{P}_1\mathbf{P}_2^H = \mathbf{0}$ and $\mathbf{P}_i\mathbf{P}_i^H = \mathbf{U}_i\boldsymbol{\Xi}_i\mathbf{U}_i^H$, $i = 1, 2$, where $\boldsymbol{\Xi}_i$ is an $N_i \times N_i$ diagonal matrix.

PROOF: Similar to the proof of Theorem 1. $\square$

The optimal structure of $\mathbf{P}_i$ can be obtained from Theorem 3 as $\mathbf{P}_i = \mathbf{U}_i\boldsymbol{\Xi}_i^{\frac{1}{2}}\boldsymbol{\Omega}_{i,2}$, where $\boldsymbol{\Omega}_{i,2}$ is an $N_i \times T_2$ semi-unitary matrix satisfying $\boldsymbol{\Omega}_{i,2}\boldsymbol{\Omega}_{i,2}^H = \mathbf{I}_{N_i}$, $i = 1, 2$, and $\boldsymbol{\Omega}_{1,2}\boldsymbol{\Omega}_{2,2}^H = \mathbf{0}$. Such $\boldsymbol{\Omega}_{1,2}$ and $\boldsymbol{\Omega}_{2,2}$ can be easily constructed from the normalized DFT matrix when $T_2 \geq N_1 + N_2$. Using Theorem 3, the problem (59)-(61) is equivalently converted to the following problem

$$\min_{\eta,\boldsymbol{\Xi}_1,\boldsymbol{\Xi}_2} \sum_{i=1}^{2}\sum_{j=1}^{2}\mathrm{tr}\left(\left[\mathbf{D}_{ij} + \eta\boldsymbol{\Xi}_j \otimes \mathbf{E}_{ri}\right]^{-1}\right) \quad (63)$$

$$\text{s.t. } \mathrm{tr}(\boldsymbol{\Xi}_i) \leq q_i, \qquad i = 1, 2 \quad (64)$$

$$\eta N_r + \eta\sum_{i=1}^{2}\mathrm{tr}(\boldsymbol{\Lambda}_i\boldsymbol{\Xi}_i)\mathrm{tr}(\mathbf{R}_{ri}) \leq q_{r,2} \quad (65)$$

$$\eta > 0, \quad \boldsymbol{\Xi}_i \geq 0, \quad i = 1, 2 \quad (66)$$

where $\mathbf{E}_{ri} \triangleq \left(\eta\mathrm{tr}(\mathbf{C}_r^T)\boldsymbol{\Delta}_i + \mathbf{I}_{N_i}\right)^{-1}$, $i = 1, 2$, are diagonal matrices. The problem (63)-(66) can be equivalently rewritten as the following problem with scalar variables

$$\min_{\eta,\boldsymbol{\xi}_1,\boldsymbol{\xi}_2} \sum_{i=1}^{2}\sum_{j=1}^{2}\sum_{m=1}^{N_j}\sum_{n=1}^{N_i}\left[\frac{1}{b_j\lambda_{j,m}\delta_{i,n}} + \eta f_{i,n}\xi_{j,m}\right]^{-1} \quad (67)$$

$$\text{s.t. } \sum_{m=1}^{N_i}\xi_{i,m} \leq q_i, \quad i = 1, 2 \quad (68)$$

$$\eta N_r + \eta\sum_{i=1}^{2}\left(\sum_{m=1}^{N_i}\lambda_{i,m}\xi_{i,m}\mathrm{tr}(\mathbf{R}_{ri})\right) \leq q_{r,2} \quad (69)$$

$$\eta > 0, \ \xi_{i,m} \geq 0, \quad m = 1, \cdots, N_i, \quad i = 1, 2 \quad (70)$$

where $f_{i,n} \triangleq 1/(\eta\mathrm{tr}(\mathbf{C}_r^T)\delta_{i,n} + 1)$, $\boldsymbol{\xi}_i \triangleq \left[\xi_{i,1}, \cdots, \xi_{i,N_i}\right]^T$, $i = 1, 2$, and $\xi_{i,m}$ is the mth diagonal element of $\boldsymbol{\Xi}_i$.

With fixed η, the objective function (67) can be rewritten as

$$\min_{\boldsymbol{\xi}_1,\boldsymbol{\xi}_2} \sum_{i=1}^{2}\sum_{j=1}^{2}\sum_{m=1}^{N_j}\sum_{n=1}^{N_i}\frac{1}{a_{i,j,m,n} + h_{i,n}\xi_{j,m}}$$

where $a_{i,j,m,n}$ and $h_{i,n} \triangleq \eta f_{i,n}$ are known variables. It can be seen from the above equation that the summation terms are monotonically decreasing and convex with respect to $\xi_{1,m}$ and $\xi_{2,m}$. Moreover, with fixed η, the constraints in (68)-(70) are linear inequality constraints. Therefore, the problem (67)-(70) is a convex optimization problem with respect to $\boldsymbol{\xi}_1$ and $\boldsymbol{\xi}_2$ when η is fixed. For a given η, the optimal $\boldsymbol{\xi}_1$ and $\boldsymbol{\xi}_2$ can be efficiently obtained through the KKT optimality conditions associated with the problem (67)-(70). The gradient conditions are given by

$$\sum_{j=1}^{2}\sum_{n=1}^{N_j}\frac{\eta f_{j,n}}{\left[(b_i\lambda_{i,m}\delta_{j,n})^{-1} + \eta f_{j,n}\xi_{i,m}\right]^2} = \nu_i + \nu_3 c_{i,m},$$
$$m = 1, \cdots, N_i, \quad i = 1, 2 \quad (71)$$

Fig. 3. Two-stage channel estimation: NMSE versus η for different $q_1 = q_2$ with $N = 4$ and $q_{r,2} = 20$dB.
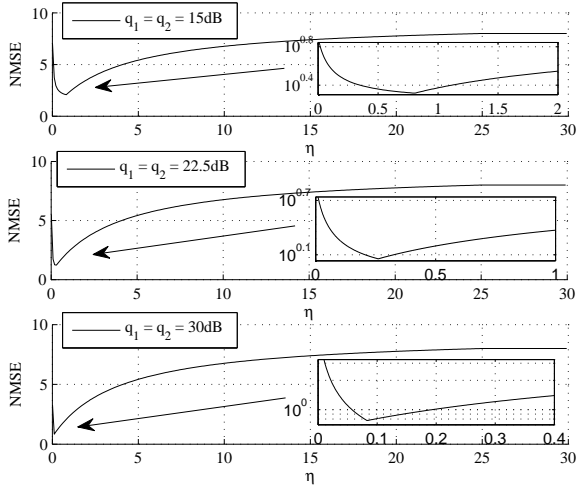

Fig. 4. Two-stage channel estimation: NMSE versus β for different $q_1 = q_2$ with $N = 4$ and $p_r = 20$dB.

where $c_{i,m} \triangleq \eta \text{tr}(\mathbf{R}_{ri})\lambda_{i,m}$, $i = 1, 2$, and $\nu_i \geq 0$, $i = 1, 2, 3$, are Lagrange multipliers such that the complementary slackness conditions given by

$$\nu_i \left(q_i - \sum_{m=1}^{N_i} \xi_{i,m} \right) = 0, \quad i = 1, 2 \tag{72}$$

$$\nu_3 \left(q_{r,2} - \eta N_r - \eta \sum_{i=1}^{2} \sum_{m=1}^{N_i} \lambda_{i,m} \xi_{i,m} \text{tr}(\mathbf{R}_{ri}) \right) = 0 \tag{73}$$

are satisfied.

With fixed η and ν_i, $i = 1, 2, 3$, for each m, the non-negative $\xi_{1,m}$ and $\xi_{2,m}$ can be found by using the bi-section search, since the LHS of (71) is a monotonically decreasing function of $\xi_{1,m}$ and $\xi_{2,m}$. To find the optimal ν_i, $i = 1, 2, 3$, an outer bi-section search is used as the LHS of (68) and (69) are increasing functions of $\xi_{1,m}$ and $\xi_{2,m}$, while in (71), $\xi_{i,m}$ is monotonically decreasing with respect to ν_i and ν_3.

The problem (67)-(70) as a whole is is non-convex with respect to $\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \eta$. However, based on a similar analysis used in the problem (28)-(31), it can be shown that (67) subjecting to (68)-(70) is a unimodal (quasi-convex) function with respect to η. To validate our analysis, a plot of the MSE value over a range of feasible values of η is generated in Fig. 3 for the case where all nodes have the same number of antennas, i.e., $N_i = N = 4$, $i = 1, 2, r$. The channel matrices have i.i.d. entries, i.e., $\mathbf{T}_{ri} = \mathbf{R}_{ri} = \mathbf{R}_{ir} = \mathbf{C}_r = \mathbf{I}_N$, $i = 1, 2$. Fig. 3 shows the NMSE value versus η for different $q_1 = q_2$ with $q_{r,2}$ set to be 20dB. It can be observed from Fig. 3 that (67) is a unimodal function of η. For a unimodal function, the minimum value can be efficiently found by the GSS algorithm [21]. Hence, the optimal η for the problem (67)-(70) can be obtained by applying the GSS technique similar to the procedure listed in Table I.

Now let us investigate the optimal power allocation $q_{r,1}$ and $q_{r,2}$ at the relay node during two stages of channel training. Based on (62), we let $q_{r,1} T_1 = \beta p_r T$ and $q_{r,2} T_2 = (1-\beta) p_r T$, where $0 < \beta < 1$. The aim is to find the optimal β to minimize
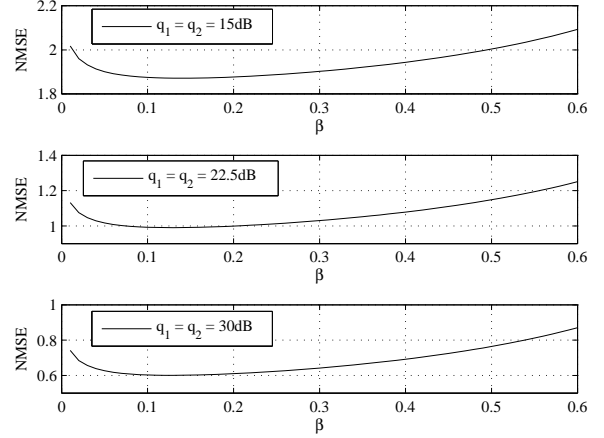
the overall MSE of channel estimation over two-stages which is given by the summation of (44) and (59), and can be written as

$$\sum_{i=1}^{2} \text{tr} \Big(\big[(\mathbf{C}_r \otimes \mathbf{R}_{ir})^{-1} + (\mathbf{P}_r^* \mathbf{P}_r^T) \otimes \mathbf{I}_{N_i} \big]^{-1} + \big[\mathbf{R}_{\theta_i}^{-1} + \mathbf{N}_i^H \mathbf{R}_{\bar{v}_{i,2}}^{-1} \mathbf{N}_i \big]^{-1} \Big). \tag{74}$$

Fig. 4 shows the value of (74) over a range of feasible values of β for different $q_1 = q_2$ with $T_1 = N_r$, $T_2 = N_1 + N_2$, $T = N_1 + N_2 + N_r$, and $p_r = 20$dB. We assume that $N_i = N = 4$, $i = 1, 2, r$, and $\mathbf{T}_{ri} = \mathbf{R}_{ri} = \mathbf{R}_{ir} = \mathbf{C}_r = \mathbf{I}_N$, $i = 1, 2$. Here for each β, the problem (44)-(45) and the problem (59)-(61) are solved to obtain the optimal $\mathbf{P}_r$, $\mathbf{P}_1$, $\mathbf{P}_2$, and η. It can be seen from Fig. 4 that (74) is a unimodal function of β. Hence, the GSS technique described in Table I can be applied to obtain the optimal β.

The complexity of the two-stage channel estimation algorithm can be estimated as $\mathcal{O}(d_\beta d_\mu d_\xi N_r (N_1 + N_2) + d_\beta d_\eta d_\nu d_\xi (N_1 + N_2)^2)$, where the first term is the complexity of stage one, and the second term represents the complexity involved in stage two. Here d_β, d_η, and d_μ stand for the numbers of iterations required to obtain the optimal β, η, and μ, respectively, d_ν is the number of iterations in the outer bi-section loop to obtain the optimal ν_1, ν_2, and μ_3, and d_ξ represents the number of bi-section operations required to obtain the optimal $\boldsymbol{\xi}_1$, $\boldsymbol{\xi}_2$, and $\boldsymbol{\xi}_r$.

V. NUMERICAL EXAMPLES

In this section, we study the performance of the proposed superimposed channel training algorithm and two-stage channel estimation algorithm 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., $N_i = N$, $i = 1, 2, r$. We also assume that all nodes have the same transmission power $p_i = p$, $i = 1, 2, r$, and use the shortest training sequence possible with $T_1 = N$, $T_2 = 2N$, $T = 3N$. Thus, based on (62), there are $q_1 = q_2 = 1.5p$ and $(q_{r,1} + 2q_{r,2})/3 = p$ for the two-stage

channel estimation algorithm. The channel covariance matrices have the commonly used exponential Toeplitz structure [16] such that $[\mathbf{T}_{ri}]_{m,n} = \rho^{|m-n|}$, $i = 1, 2$, $[\mathbf{R}_{ri}]_{m,n} = \rho^{|m-n|}$, $i = 1, 2$, $[\mathbf{R}_{ir}]_{m,n} = \rho^{|m-n|}$, $i = 1, 2$, and $[\mathbf{C}_r]_{m,n} = \rho^{|m-n|}$, where ρ is the correlation coefficient with magnitude $|\rho| < 1$. For all scenarios, the normalized MSE (NMSE) of channel estimation at nodes 1 and 2 are computed. The optimal training sequences for the superimposed channel training method and the two-stage channel estimation algorithm are generated by using Theorem 1 and Theorems 2 and 3, respectively. In particular, the semi-unitary matrices in the superimposed channel training method are set based on the normalized DFT matrix as $[\mathbf{\Omega}_1]_{m,n} = \frac{1}{\sqrt{3N}}e^{-j\frac{2\pi mn}{3N}}$, $[\mathbf{\Omega}_2]_{m,n} = \frac{1}{\sqrt{3N}}e^{-j\frac{2\pi(m+N)n}{3N}}$, $[\mathbf{\Omega}_r]_{m,n} = \frac{1}{\sqrt{3N}}e^{-j\frac{2\pi(m+2N)n}{3N}}$, $m = 1, \cdots, N$, $n = 1, \cdots, 3N$. Matrices $\mathbf{\Omega}_{r,1}$ and $\mathbf{\Omega}_{i,2}$, $i = 1, 2$, in the two-stage channel estimation algorithm are chosen as $[\mathbf{\Omega}_{r,1}]_{m,n} = \frac{1}{\sqrt{N}}e^{-j\frac{2\pi mn}{N}}$, $m, n = 1, \cdots, N$, and $[\mathbf{\Omega}_{1,2}]_{m,n} = \frac{1}{\sqrt{2N}}e^{-j\frac{2\pi mn}{2N}}$, $[\mathbf{\Omega}_{2,2}]_{m,n} = \frac{1}{\sqrt{2N}}e^{-j\frac{2\pi(m+N)n}{2N}}$, $m = 1, \cdots, N$, $n = 1, \cdots, 2N$.

In the first example, we study the performance of the superimposed channel training algorithm with respect to α. Fig. 5 shows the NMSE of this algorithm versus p with different α when $N = 4$ and $\rho = 0.8$. The curve associated with the optimal α is obtained by applying the GSS algorithm on the proposed superimposed channel training technique to find the optimal α for different p. It can be seen from Fig. 5 that the curve associated with the optimal α has the lowest MSE level. This justifies that the GSS algorithm can be applied to obtain the optimal α at different p efficiently. Interestingly, we observe from Fig. 5 that the optimal α vary with respect to p, indicating that using constant α is strictly suboptimal. In fact, the optimal α at low p level is smaller compared with the optimal α for large p. The reason is that the estimation of the first-hop channels $\mathbf{H}_{ri}$ is based on that of the second-hop channels $\mathbf{H}_{ir}$. When p is small, at the relay node, more power should be allocated for the estimation of $\mathbf{H}_{ir}$, which is also beneficial to the estimation of $\mathbf{H}_{ri}$. When a large amount of power p is available, the MSE of estimating $\mathbf{H}_{ir}$ is smaller compared with that of $\mathbf{H}_{ri}$. Therefore, more power should be allocated at the relay node to assist the estimation of $\mathbf{H}_{ri}$.

In the second example, we investigate the performance of the two-stage channel estimation algorithm with respect to β. A plot of the NMSE of this algorithm for different β is shown in Fig. 6, where the curve with the optimal β is obtained from the GSS algorithm. Similar to Fig. 5, it can be seen from Fig. 6 that the curve associated with the optimal β has the lowest MSE level.

In the third example, we compare the performance of the superimposed and two-stage channel estimation algorithms when the optimal α and β are used. We also show the performance of the conventional two-stage channel estimator, where random orthogonal pilot sequences are used to estimate the channel matrices and the transmission power at the relay node is equally distributed between two stages. Fig. 7 demonstrates the MSE performance of all algorithms with $\rho = 0.2$ for different N, while Fig. 8 shows the MSE results at $\rho = 0.8$. It can be seen from Figs. 7 and 8 that the proposed algorithms yield
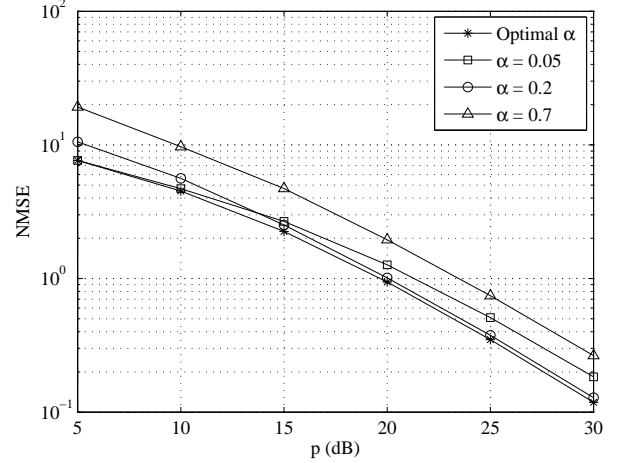

Fig. 5. Example 1. Superimposed channel training: NMSE versus p for different α with $N = 4$ and $\rho = 0.8$.
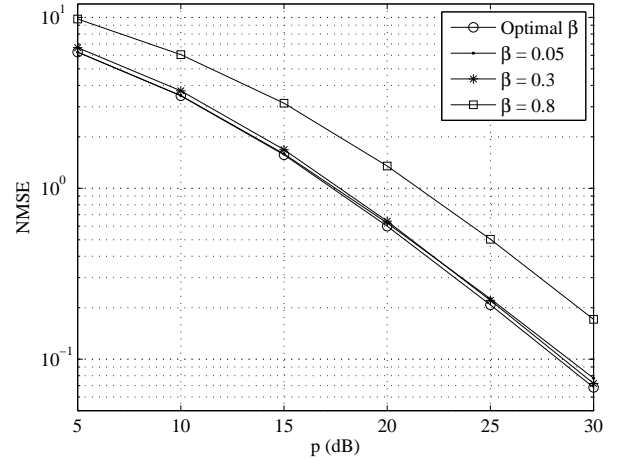

Fig. 6. Example 2. Two-stage channel estimation: NMSE versus p for different β with $N = 4$ and $\rho = 0.8$.

much smaller estimation error compared with the conventional two-stage channel estimator, especially at high p level. It can also be observed from Figs. 7 and 8 that for both scenarios, the two-stage channel estimation algorithm yields smaller MSEs than the superimposed channel training scheme. This is mainly due to the fact that in the superimposed channel training algorithm, the estimation of $\mathbf{H}_{ir}$ is affected by the noise at the relay node, which is not the case in the two-stage channel estimation scheme. However, the two-stage channel estimation algorithm has a higher computational complexity than that of the superimposed channel training scheme, since both β and η need to be optimized in the former algorithm. Such performance-complexity tradeoff can be exploited in practical two-way MIMO relay communication systems.

VI. CONCLUSIONS

In this paper, we have proposed and investigated the performance of two channel estimation algorithms, namely, the
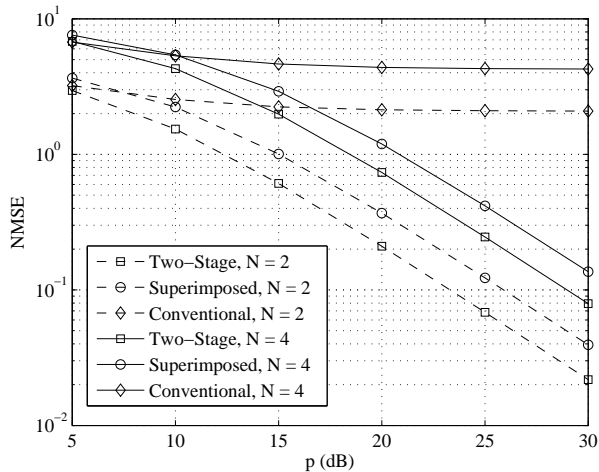
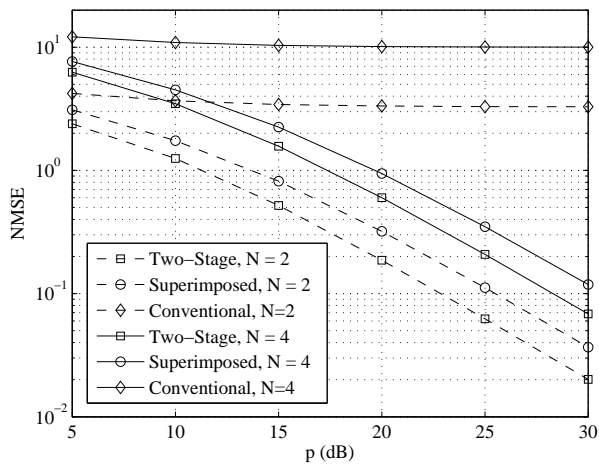Fig. 7. Example 3. NMSE versus p for $\rho = 0.2$ and different N.

Fig. 8. Example 3. NMSE versus p for $\rho = 0.8$ and different N.

superimposed channel training and two-stage channel estimation schemes, for two-way MIMO relay communication systems. The proposed algorithms can efficiently estimate the individual CSI for two-way MIMO relay systems, with the two-stage channel estimation algorithm performs better than the superimposed channel training scheme at a higher computational complexity.

APPENDIX A
PROOF OF THEOREM 1

Let us introduce the EVD of $\mathbf{R}_{ir} = \mathbf{Q}_i \boldsymbol{\Delta}_i \mathbf{Q}_i^H$. We can equivalently rewrite (13) and (16) as

$$\mathbf{R}_{\bar{v}_i} = \mathbf{I}_T \otimes \big(\mathbf{Q}_i(\alpha \mathrm{tr}(\mathbf{C}_r^T)\boldsymbol{\Delta}_i + \mathbf{I}_{N_i})\mathbf{Q}_i^H\big), \quad i = 1, 2 \quad (75)$$

$$\mathbf{R}_{\gamma_i} = \mathbf{U}_{\gamma_i} \mathrm{Bdiag}\big[\boldsymbol{\Lambda}_i \otimes b_i \boldsymbol{\Delta}_i, \ \boldsymbol{\Lambda}_{\bar{i}} \otimes b_{\bar{i}} \boldsymbol{\Delta}_i, \ \boldsymbol{\Lambda}_r \otimes \boldsymbol{\Delta}_i\big] \mathbf{U}_{\gamma_i}^H,$$
$$i = 1, 2 \quad (76)$$

where $\mathbf{U}_{\gamma_i} \triangleq \mathrm{Bdiag}\big[\mathbf{I}_{N_i} \otimes \mathbf{Q}_i, \ \mathbf{I}_{N_{\bar{i}}} \otimes \mathbf{Q}_i, \ \mathbf{I}_{N_r} \otimes \mathbf{Q}_i\big], \ i = 1, 2$. Substituting (75) and (76) back into (18), MSE_i can be

rewritten as

$$\mathrm{MSE}_i = \mathrm{tr}\Bigg(\Bigg[\begin{pmatrix} \mathbf{D}_{ii} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{D}_{i\bar{i}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{D}_{si} \end{pmatrix} + \begin{pmatrix} \sqrt{\alpha}\tilde{\mathbf{S}}_i^* \otimes \mathbf{I}_{N_i} \\ \sqrt{\alpha}\tilde{\mathbf{S}}_{\bar{i}}^* \otimes \mathbf{I}_{N_i} \\ \tilde{\mathbf{S}}_r^* \otimes \mathbf{I}_{N_i} \end{pmatrix} \big(\mathbf{I}_T \otimes \mathbf{D}_{ri}\big)$$
$$\times \big(\sqrt{\alpha}\tilde{\mathbf{S}}_i^T \otimes \mathbf{I}_{N_i}, \ \sqrt{\alpha}\tilde{\mathbf{S}}_{\bar{i}}^T \otimes \mathbf{I}_{N_i}, \ \tilde{\mathbf{S}}_r^T \otimes \mathbf{I}_{N_i}\big) \Bigg]^{-1} \Bigg) (77)$$

where

$$\mathbf{D}_{ij} \triangleq \boldsymbol{\Lambda}_j^{-1} \otimes (b_j \boldsymbol{\Delta}_i)^{-1}, \quad j = i, \bar{i}, \qquad \mathbf{D}_{si} \triangleq \boldsymbol{\Lambda}_r^{-1} \otimes \boldsymbol{\Delta}_i^{-1}$$
$$\mathbf{D}_{ri} \triangleq \big(\alpha \mathrm{tr}(\mathbf{C}_r^T)\boldsymbol{\Delta}_i + \mathbf{I}_{N_i}\big)^{-1} \quad (78)$$

are all diagonal matrices. It can be seen from (77) that the objective function (21) is minimized only if

$$(\tilde{\mathbf{S}}_i^* \otimes \mathbf{I}_{N_i})(\mathbf{I}_T \otimes \mathbf{D}_{ri})(\tilde{\mathbf{S}}_{\bar{i}}^T \otimes \mathbf{I}_{N_i}) = \big(\tilde{\mathbf{S}}_i^* \tilde{\mathbf{S}}_{\bar{i}}^T\big) \otimes \mathbf{D}_{ri} = \mathbf{0} \ (79)$$
$$(\tilde{\mathbf{S}}_j^* \otimes \mathbf{I}_{N_i})(\mathbf{I}_T \otimes \mathbf{D}_{ri})(\tilde{\mathbf{S}}_r^T \otimes \mathbf{I}_{N_i}) = \big(\tilde{\mathbf{S}}_j^* \tilde{\mathbf{S}}_r^T\big) \otimes \mathbf{D}_{ri} = \mathbf{0} \ (80)$$

for $i = 1, 2$, and $j = i, \bar{i}$. Equations (79) and (80) hold if and only if $\tilde{\mathbf{S}}_1 \tilde{\mathbf{S}}_2^T = \mathbf{0}$ and $\tilde{\mathbf{S}}_i^* \tilde{\mathbf{S}}_r^T = \mathbf{0}$, $i = 1, 2$, or equivalently $\mathbf{S}_1 \mathbf{S}_2^H = \mathbf{0}$ and $\mathbf{S}_i \mathbf{S}_r^H = \mathbf{0}$, $i = 1, 2$. Then the objective function (21) can be written as

$$\sum_{i=1}^{2} \mathrm{tr}\Bigg(\sum_{j=1}^{2} \big[\mathbf{D}_{ij} + \alpha \tilde{\mathbf{S}}_j^* \tilde{\mathbf{S}}_j^T \otimes \mathbf{D}_{ri}\big]^{-1} + \big[\mathbf{D}_{si} + \tilde{\mathbf{S}}_r^* \tilde{\mathbf{S}}_r^T \otimes \mathbf{D}_{ri}\big]^{-1} \Bigg).$$
$$(81)$$

Since $\mathbf{D}_{ij}$, $\mathbf{D}_{si}$, and $\mathbf{D}_{ri}$ are all diagonal, to minimize (81), $\tilde{\mathbf{S}}_i^* \tilde{\mathbf{S}}_i^T$, $i = 1, 2, r$, must be diagonal. Note that the diagonality of $\tilde{\mathbf{S}}_i \tilde{\mathbf{S}}_i^H$ does not change $\mathrm{tr}(\tilde{\mathbf{S}}_i \tilde{\mathbf{S}}_i^H)$, $i = 1, 2, r$, in the constraints (22) and (23). We would like to note that $\mathrm{tr}(\boldsymbol{\Lambda}_i \tilde{\mathbf{S}}_i \tilde{\mathbf{S}}_i^H)$ in the constraints (23) is minimized if $\tilde{\mathbf{S}}_i \tilde{\mathbf{S}}_i^H$ is diagonal and its diagonal entries are in the inverse order of that of $\boldsymbol{\Lambda}_i$ [22]. Denoting $\tilde{\mathbf{S}}_i \tilde{\mathbf{S}}_i^H = \boldsymbol{\Sigma}_i$, $i = 1, 2, r$, then we have $\mathbf{S}_i \mathbf{S}_i^H = \mathbf{U}_i \boldsymbol{\Sigma}_i \mathbf{U}_i^H$, $i = 1, 2, r$. $\qquad \square$

REFERENCES

[1] Y. Fan and J. Thompson, "MIMO configurations for relay channels: Theory and practice," *IEEE Trans. Wireless Commun.*, vol. 6, pp. 1774-1786, May 2007.

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

[3] I. Hammerström and A. Wittneben, "Power allocation schemes for amplify-and-forward MIMO-OFDM relay links," *IEEE Trans. Wireless Commun.*, vol. 6, pp. 2798-2802, Aug. 2007.

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

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

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

[7] T. Kong and Y. Hua, "Optimal design of source and relay pilots for MIMO relay channel estimation," *IEEE Trans. Signal Process.*, vol. 59, pp. 4438-4446, Sep. 2011.

[8] 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, pp. 2224-2233, June 2012.

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

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

[11] 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.

[12] C. K. Ho, R. Zhang, and Y.-C. Liang, "Two-way relaying over OFDM: Optimized tone permutation and power allocation," in *Proc. IEEE ICC*, pp. 3908-3912, May 2008.

[13] Z. Fang, J. Shi, and H. Shan, "Comparison of channel estimation schemes for MIMO two-way relaying systems," in *Proc. Cross Strait Quad-Regional Radio Science and Wireless Technology Conference (CSQRWC)*, vol. 1, pp. 719-722, Jul. 2011.

[14] M. Biguesh and A. B. Gershman, "Training-based MIMO channel estimation: A study of estimator tradeoffs and optimal training signals," *IEEE Trans. Signal Process.*, vol. 54, pp. 884-893, Mar. 2006.

[15] M. Biguesh, S. Gazor, and M. H. Shariat, "Optimal training sequence for MIMO wireless systems in colored environment," *IEEE Trans. Signal Process.*, vol. 57, pp. 3144-3153, Aug. 2009.

[16] D. S. Shiu, G. Foschini, M. Gans, and J. Kahn, "Fading correlation and its effect on the capacity of multielement antenna systems," *IEEE Trans. Commun.*, vol. 48, pp. 503-513, Mar. 2000.

[17] J. W. Brewer, "Kronecker products and matrix calculus in system theory," *IEEE Trans. Circuits Syst.*, vol. 25, pp. 772-781, Sep. 1978.

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

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

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

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

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

Choo W. R. Chiong received the B.E. (Hons.) degree in electrical power engineering from Curtin University, Australia, in 2010.

In 2009, he was with Curtin University Sarawak Campus, Malaysia as a research assistant, mainly worked on the development of Smart Drive Build in AC Motors for Optimal Operations. He is currently working toward the Ph.D. degree at the Department of Electrical and Computer Engineering, Curtin University, Australia. 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.

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.

Yong Xiang (SM'12) received the B.E. and M.E. degrees from the University of Electronic Science and Technology of China, Chengdu, China, in 1983 and 1989, respectively. In 2003, he received the Ph.D. degree from The University of Melbourne, Melbourne, Australia.

He was with the Southwest Institute of Electronic Equipment of China, Chengdu, from 1983 to 1986. In 1989, he joined the University of Electronic Science and Technology of China, where he was a Lecturer from 1989 to 1992 and an Associate Professor from 1992 to 1997. He was a Senior Communications Engineer with Bandspeed Inc., Melbourne, Australia, from 2000 to 2002. He is currently an Associate Professor with the School of Information Technology at Deakin University, Melbourne, Australia. His research interests include blind signal/system estimation, information and network security, communication signal processing, and multimedia processing.