Channel Estimation of Dual-Hop MIMO Relay System via Parallel Factor Analysis

Yue Rong, *Senior Member, IEEE*, Muhammad R. A. Khandaker, *Student Member, IEEE*, and Yong Xiang

***Abstract*—The optimal source precoding matrix and relay amplifying matrix have been developed in recent works on multiple-input multiple-output (MIMO) relay communication systems assuming that the instantaneous channel state information (CSI) is available. However, in practical relay communication systems, the instantaneous CSI is unknown, and therefore, has to be estimated at the destination node. In this paper, we develop a novel channel estimation algorithm for two-hop MIMO relay systems using the parallel factor (PARAFAC) analysis. The proposed algorithm provides the destination node with full knowledge of all channel matrices involved in the communication. Compared with existing approaches, the proposed algorithm requires less number of training data blocks, yields smaller channel estimation error, and is applicable for both one-way and two-way MIMO relay systems with single or multiple relay nodes. Numerical examples demonstrate the effectiveness of the PARAFAC-based channel estimation algorithm.**

***Index Terms*—Channel estimation, MIMO relay, PARAFAC.**

I. Introduction

RECENTLY, there have been many research efforts on multiple-input multiple-output (MIMO) relay systems [1]-[6]. For a three-node two-hop MIMO relay system where the direct source-destination link is omitted, the optimal relay amplifying matrix is obtained in [2]-[3] to maximize the mutual information between source and destination. In [4], optimal relay matrices are developed to minimize the mean-squared error (MSE) of the signal waveform estimation at the destination node for a two-hop MIMO relay system with multiple parallel relay nodes. A unified framework is established for optimizing the source precoding matrix and the relay amplifying matrix of two-hop linear non-regenerative MIMO relay systems with a broad class of objective functions [5]. Recently, it has been shown in [6] that by using a nonlinear decision feedback equalizer (DFE) based on the minimal MSE (MMSE) criterion at the destination node, the system bit-error-rate (BER) can be significantly reduced.

For the aforementioned MIMO relay systems, the instantaneous channel state information (CSI) knowledge of both the source-relay link and the relay-destination link is required at the destination node to estimate the source signals. Moreover, in order to optimize the source and/or relay matrices in [1]-[6], the instantaneous CSI knowledge of both links is needed to carry out the optimization procedure. When the direct source-destination link is considered, the CSI knowledge of the direct link is also required at the destination node to estimate the source signals [7]. However, in practical relay communication systems, the instantaneous CSI is unknown, and therefore, has to be estimated. Recently, a tensor-based channel estimation algorithm is developed in [8] for a two-way MIMO relay system. Since the algorithm in [8] exploits the channel reciprocity in a two-way relay system, its application in one-way MIMO relay systems is not straightforward. In [9], a relay channel estimation algorithm using the least-squares (LS) fitting is proposed. The performance of the algorithm in [9] is further analyzed and improved by using the weighted least-squares (WLS) fitting in [10]. However, the number of training data blocks required in [9] and [10] is at least equal to the number of relay nodes (antennas), resulting in a low system spectral efficiency. For amplify-and-forward relay networks with single-antenna source, relay, and destination nodes, the optimal training sequence is developed in [11]. A superimposed training based channel estimation algorithm has been developed recently for OFDM modulated relay systems in [12]. The optimal training sequence is derived in [13] for a MIMO relay system with one multi-antenna relay node. However, for systems with distributed relay nodes which do not cooperate with each other, the result in [13] can not be used.

There are two major challenges in channel estimation for MIMO relay systems. Firstly, for most applications, the CSI on the compound source-relay-destination channel alone is not sufficient. In fact, the CSI of each hop is required at the destination node to perform signal retrieving and system optimization. Secondly, relay nodes (in particular, non-regenerative distributed relays) often have limited computation capacity. Thus, channel estimation is usually carried out at the destination node, not at the relay nodes [8]-[12]. In this paper, we address these two challenges by proposing a novel MIMO relay channel estimation algorithm based on the parallel factor (PARAFAC) analysis [14]-[16]. The proposed algorithm provides the destination node with full knowledge of all channel matrices involved in the communication. The contributions of this paper can be summarized as follows. Firstly, compared with algorithms in [9] and [10] where the number of training data blocks should be at least equal to the number of relay nodes (antennas), the number of training data blocks required in the proposed algorithm can be less than the number of relay nodes (antennas). In particular, we show that when the number of relay nodes (antennas) is smaller than

Manuscript received June 30, 2011; revised October 18, 2011 and January 30, 2012; accepted March 4, 2012. The associate editor coordinating the review of this paper and approving it for publication was D. Hong.

Y. Rong and M. R. A. Khandaker are with the Department of Electrical and Computer Engineering, Curtin University, Bentley, WA 6102, Australia (e-mail: y.rong@curtin.edu.au; m.khandaker@postgrad.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).

This research was supported under the Australian Research Council's Discovery Projects funding scheme (project numbers DP110100736, DP110102076).

Digital Object Identifier 10.1109/TWC.2012.032712.111251

the number of antennas at the source node and the destination node, as few as two training data blocks are sufficient to estimate all channels. Thus, the proposed algorithm has a higher spectral efficiency than those in [9] and [10]. Secondly, in this paper, the initial estimation of channel matrices is improved by a linear MMSE (LMMSE) algorithm, which yields a smaller estimation error than the WLS fitting applied in [10]. Thirdly, in contrast to [8], the proposed algorithm is applicable for both one-way and two-way relay systems with single or multiple relay nodes.

In the proposed algorithm, the MIMO channel matrix of the direct source-destination link in one-way relay systems is estimated by the LS approach. For the source-relay-destination link in both one-way and two-way relay systems, we show that under a mild condition of the channel training data block length, the MIMO channel matrices of both hops can be estimated up to permutation and scaling ambiguities, which are inherent to the PARAFAC model. To remove the permutation ambiguity, we exploit the knowledge of the relay factors available at the destination node during the channel training period. Then by using a bilinear alternating least-squares (BALS) algorithm, the channel matrix of each hop can be estimated up to some scaling ambiguity, which can be resolved through normalization as in [9], [10], [15].

Since during the training period, the noise at the relay nodes is amplified and forwarded to the destination node, the effective noise vector at the destination node is non-white. Taking this fact into account, we propose an LMMSE approach to further improve the channel estimation, by exploiting the initial estimate of the relay-destination channel. We show that the proposed BALS and LMMSE algorithms can also be applied for channel estimation in two-way MIMO relay systems. Numerical examples demonstrate the effectiveness of the proposed PARAFAC-based channel estimation algorithm compared with existing techniques. We would like to mention that in this paper, for notational convenience, we consider a narrowband single-carrier system. However, our algorithm can be straightforwardly applied to estimate the MIMO channel matrices in each subcarrier of a broadband multi-carrier relay communication system[1].

The rest of this paper is organized as follows. In Section II, we introduce the model of a two-hop amplify-and-forward MIMO relay communication system. The proposed channel estimation algorithm is developed in Sections III. In Section IV, we show some numerical examples. Conclusions are drawn in Section V.

II. System Model

We consider a two-hop MIMO communication system where the source node transmits information to the destination node with the aid of R relay nodes as shown in Fig. 1. The source node and the destination node are equipped with $N_s \geq 2$ and $N_d \geq 2$ antennas, respectively, while the ith relay node has M_i antennas, $i = 1, \cdots, R$. Since several practical constraints such as power consumption, implementation costs

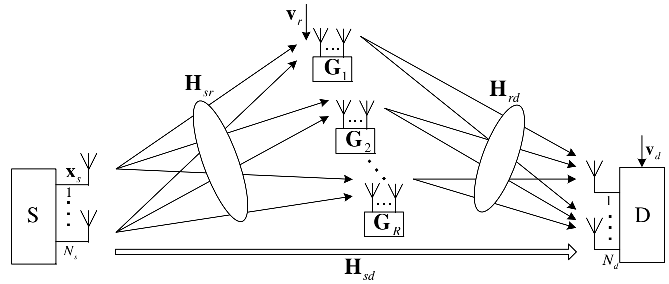


Fig. 1. Two-Hop MIMO relay system with R relay nodes.

and spatial efficiency make half-duplex relays more appealing for wireless applications than full-duplex relays, in this paper, we consider half-duplex relays as in [2]-[13] (i.e., each relay node does not receive and transmit signals simultaneously). Thus, the communication process between the source and destination nodes is completed in two time slots. In the first time slot, the $N_s \times 1$ modulated signal vector $\mathbf{u}_s(t)$ is transmitted to all relay nodes and the destination node, and the received signal vectors are respectively given by

$$\begin{aligned}\mathbf{y}_{r,i}(t) &= \mathbf{H}_{sr,i}\mathbf{u}_s(t) + \mathbf{v}_{r,i}(t), \quad i = 1, \cdots, R \\ \mathbf{y}_d(t) &= \mathbf{H}_{sd}\mathbf{u}_s(t) + \mathbf{v}_d(t)\end{aligned} \tag{1}$$

where $\mathbf{y}_{r,i}(t)$ is an $M_i \times 1$ received signal vector at the ith relay node, $\mathbf{y}_d(t)$ is an $N_d \times 1$ received signal vector at the destination node, $\mathbf{H}_{sr,i}$ is the $M_i \times N_s$ MIMO fading channel matrix between the source node and the ith relay node, $\mathbf{H}_{sd}$ is the $N_d \times N_s$ MIMO source-destination channel matrix, $\mathbf{v}_{r,i}(t)$ is an $M_i \times 1$ noise vector at the ith relay node, and $\mathbf{v}_d(t)$ is the $N_d \times 1$ noise vector at the destination node. We assume that all noises are independent identically distributed (i.i.d.) complex Gaussian noise with zero mean and unit variance.

In the second time slot, the source node is silent, and each relay node amplifies the received signal vector with matrix $\mathbf{G}_i$ and forwards the amplified signals to the destination node. We assume that relay nodes are synchronized during transmission[2] as in [4], [9], and [10]. The received signal vector at the destination node is

$$\begin{aligned}\mathbf{y}_d(t+1) &= \sum_{i=1}^{R} \mathbf{H}_{rd,i}\mathbf{G}_i\mathbf{H}_{sr,i}\mathbf{u}_s(t) \\ &\quad + \sum_{i=1}^{R} \mathbf{H}_{rd,i}\mathbf{G}_i\mathbf{v}_{r,i}(t) + \mathbf{v}_d(t+1) \\ &= \mathbf{H}_{rd}\mathbf{G}\mathbf{H}_{sr}\mathbf{u}_s(t) + \mathbf{H}_{rd}\mathbf{G}\mathbf{v}_r(t) + \mathbf{v}_d(t+1)\end{aligned} \tag{2}$$

where $\mathbf{H}_{rd,i}$ is the $N_d \times M_i$ MIMO fading channel matrix between the destination node and the ith relay node, and $\mathbf{v}_d(t+1)$ is an $N_d \times 1$ noise vector at the destination node at time $t+1$. Here $\mathbf{H}_{sr} \triangleq \left[\mathbf{H}_{sr,1}^T, \cdots, \mathbf{H}_{sr,R}^T\right]^T$ is the $M \times N_s$ ($M = \sum_{i=1}^{R} M_i$) MIMO channel from the source node to all relay nodes, $\mathbf{H}_{rd} \triangleq [\mathbf{H}_{rd,1}, \cdots, \mathbf{H}_{rd,R}]$ is the $N_d \times M$ channel matrix between all relay nodes and the destination node, $\mathbf{v}_r(t) \triangleq \left[\mathbf{v}_{r,1}^T(t), \cdots, \mathbf{v}_{r,R}^T(t)\right]^T$ is an $M \times 1$ vector

[1]In a multicarrier communication system, the spectral correlation among subcarriers can be exploited to reduce the computational complexity and improve the quality of channel estimation [17]. Exploiting such correlation in multicarrier MIMO relay channel estimation is an interesting future topic.

[2]If a blind synchronization technique is applied, relay synchronization and channel estimation can be jointly designed to improve the system performance [18]. While in pilot symbols-based synchronization methods, these pilot symbols can be exploited to assist channel estimation.

stacking the noise at all relay nodes on top of each other, and $\mathbf{G} \triangleq \mathrm{bd}[\mathbf{G}_1, \cdots, \mathbf{G}_R]$ is an $M \times M$ block diagonal matrix containing all relay matrices. Here $(\cdot)^T$ denotes matrix (vector) transpose and $\mathrm{bd}[\cdot]$ stands for a block diagonal matrix. We assume that $\mathbf{H}_{sr}$, $\mathbf{H}_{rd}$, and $\mathbf{H}_{sd}$ have complex Gaussian entries with zero-mean and variances of $1/N_s$, $1/M$, $1/(8N_s)$, respectively[3]. Depending on the environment, the elements in each channel matrix can be independent or correlated [19]. We assume that the channel correlation knowledge is not available at the destination node and thus can not be exploited. All channels are quasi-static block fading which means they are constant over some time interval before changing to another realization. Combining (1) and (2), the received signals at the destination node over two time slots are given by

$$\mathbf{y}(t) = \begin{bmatrix} \mathbf{H}_{rd}\mathbf{G}\mathbf{H}_{sr} \\ \mathbf{H}_{sd} \end{bmatrix} \mathbf{u}_s(t) + \begin{bmatrix} \mathbf{H}_{rd}\mathbf{G}\mathbf{v}_r(t) + \mathbf{v}_d(t+1) \\ \mathbf{v}_d(t) \end{bmatrix}. \tag{3}$$

Due to its lower computational complexity, a linear receiver is used at the destination node to retrieve the transmitted signal vector $\mathbf{u}_s(t)$ [2]-[5]. The estimated signal waveform vector is given by $\hat{\mathbf{u}}_s(t) = \mathbf{W}^H\mathbf{y}(t)$, where $\mathbf{W}$ is the $2N_d \times N_s$ weight matrix. From (3), the MSE of the signal waveform estimation can be written as

$$e = \mathrm{tr}\big(\mathrm{E}\big[(\hat{\mathbf{u}}_s(t) - \mathbf{u}_s(t))(\hat{\mathbf{u}}_s(t) - \mathbf{u}_s(t))^H\big]\big) \tag{4}$$

where $\mathrm{E}[\cdot]$ stands for statistical expectation, $\mathrm{tr}(\cdot)$ and $(\cdot)^H$ denote matrix trace, and matrix Hermitian transpose, respectively. Assuming that $\mathrm{E}\big[\mathbf{u}_s(t)\mathbf{u}_s(t)^H\big] = \mathbf{I}_{N_s}$, the receiver weight matrix which minimizes (4) is the Wiener filter given by [7]

$$\mathbf{W} = \left(\bar{\mathbf{H}}\bar{\mathbf{H}}^H + \bar{\mathbf{C}}\right)^{-1} \bar{\mathbf{H}} \tag{5}$$

where

$$\bar{\mathbf{H}} \triangleq \begin{bmatrix} \mathbf{H}_{rd}\mathbf{G}\mathbf{H}_{sr} \\ \mathbf{H}_{sd} \end{bmatrix}, \bar{\mathbf{C}} \triangleq \begin{bmatrix} \mathbf{H}_{rd}\mathbf{G}\mathbf{G}^H\mathbf{H}_{rd}^H + \mathbf{I}_{N_d} & \mathbf{0}_{N_d \times N_d} \\ \mathbf{0}_{N_d \times N_d} & \mathbf{I}_{N_d} \end{bmatrix}. \tag{6}$$

Here $(\cdot)^{-1}$ stands for the matrix inversion, $\mathbf{0}_{m \times n}$ denotes an $m \times n$ matrix with all zero entries, and $\mathbf{I}_n$ denotes an $n \times n$ identity matrix. We assume that the destination node knows the relay amplifying matrix $\mathbf{G}$.

It can be clearly seen from (5) and (6) that in order to compute $\mathbf{W}$, the CSI knowledge of the compound channel $\bar{\mathbf{H}}$ alone is not sufficient. In fact, the CSI of $\mathbf{H}_{rd}$ is also needed at the destination node to obtain $\mathbf{W}$ in (5). Moreover, it has been shown in [7] that the CSI of $\mathbf{H}_{sr}$, $\mathbf{H}_{rd}$, and $\mathbf{H}_{sd}$ is required to optimize the source precoding matrix and the relay amplifying matrix.

It is shown in [13] that the CSI required above can be obtained through a two-stage training (TST) approach. At the first stage, $\mathbf{H}_{rd}$ is estimated by transmitting an $M \times L_1$ training sequence $\mathbf{S}_1$ from all R relay nodes to the destination node, where L_1 ($L_1 \geq M$) is the length of the training sequence. The received signal matrix at the destination node is given by $\mathbf{Y}_d = \mathbf{H}_{rd}\mathbf{S}_1 + \mathbf{V}_d(1)$, where $\mathbf{V}_d(1)$ is the noise matrix at the destination node. According to [20], the optimal $\mathbf{S}_1$ minimizing the MSE of channel estimation is orthogonal, i.e., $\mathbf{S}_1\mathbf{S}_1^H = \mathbf{I}_M$. Such $\mathbf{S}_1$ can be constructed, for example, from the normalized discrete Fourier transform (DFT) matrix [20]. The estimation of $\mathbf{H}_{rd}$ is given by

$$\hat{\mathbf{H}}_{rd} = \mathbf{Y}_d\mathbf{S}_1^H. \tag{7}$$

At the second stage, the source node transmits an $N_s \times L_2$ ($L_2 \geq N_s$) orthogonal training sequence $\mathbf{S}_2$ ($\mathbf{S}_2\mathbf{S}_2^H = \mathbf{I}_{N_s}$) to all relay nodes which then forward it to the destination node. From (3), the received signal matrix at the destination node is

$$\mathbf{Y} = \begin{bmatrix} \mathbf{H}_{rd}\mathbf{G}\mathbf{H}_{sr} \\ \mathbf{H}_{sd} \end{bmatrix} \mathbf{S}_2 + \begin{bmatrix} \mathbf{H}_{rd}\mathbf{G}\mathbf{V}_r + \mathbf{V}_d(2) \\ \mathbf{V}_d(3) \end{bmatrix} \tag{8}$$

where $\mathbf{V}_r$ is the noise matrix at the relay nodes, $\mathbf{V}_d(2)$ and $\mathbf{V}_d(3)$ are the noise matrices at the destination node. The estimation of the compound channel $\bar{\mathbf{H}}$ is obtained from (8) as $\hat{\bar{\mathbf{H}}} = \mathbf{Y}\mathbf{S}_2^H$. Then an estimation of $\mathbf{H}_{sr}$ can be obtained from $\hat{\bar{\mathbf{H}}}$ as

$$\hat{\mathbf{H}}_{sr} = (\hat{\mathbf{H}}_{rd}\mathbf{G})^{\dagger}\hat{\bar{\mathbf{H}}}^{(1)} \tag{9}$$

where $(\cdot)^{\dagger}$ stands for matrix pseudo-inverse, and $\hat{\bar{\mathbf{H}}}^{(1)}$ contains the first N_d rows of $\hat{\bar{\mathbf{H}}}$. It can be seen from (9) that the error in estimating $\mathbf{H}_{sr}$ can be very large since $\hat{\mathbf{H}}_{sr}$ depends on $\hat{\mathbf{H}}_{rd}$, which is also an estimated matrix. To overcome this difficulty, in the following, we develop a PARAFAC analysis based algorithm to directly estimate all channel matrices $(\mathbf{H}_{sr}, \mathbf{H}_{rd}, \mathbf{H}_{sd})$ at the destination node.

III. Proposed Channel Estimation Algorithm

In order to estimate the channel matrices, training sequences are transmitted from the source node. The overall channel training period is divided into K time blocks (the minimal K required will be determined later). In each time block, the same $N_s \times L$ ($L \geq N_s$) orthogonal channel training sequence $\mathbf{S}$ with $\mathbf{S}\mathbf{S}^H = \mathbf{I}_{N_s}$ is transmitted by the source node. In the kth time block, the ith relay node amplifies the received signal vector with a *diagonal* matrix $\mathbf{E}_{k,i}$ and forwards the amplified signal to the destination node[4]. Thus, the overall amplifying matrix from all relay nodes is $\mathbf{E}_k = \mathrm{bd}[\mathbf{E}_{k,1}, \cdots, \mathbf{E}_{k,R}]$, which is in fact a diagonal matrix. From (3), the received signal matrices at the destination node over K time blocks are given by

$$\begin{aligned} \mathbf{Y}_k &\triangleq \begin{bmatrix} \mathbf{Y}_k^{(1)} \\ \mathbf{Y}_k^{(2)} \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{H}_{rd}\mathcal{D}_k\{\mathbf{F}\}\mathbf{H}_{sr} \\ \mathbf{H}_{sd} \end{bmatrix} \mathbf{S} + \begin{bmatrix} \mathbf{H}_{rd}\mathcal{D}_k\{\mathbf{F}\}\mathbf{V}_{r,k} + \mathbf{V}_{d,k}^{(1)} \\ \mathbf{V}_{d,k}^{(2)} \end{bmatrix} \\ &\quad k = 1, \cdots, K \end{aligned} \tag{10}$$

where $\mathcal{D}_k\{\mathbf{F}\} \triangleq \mathbf{E}_k$, $\mathbf{F}$ is a $K \times M$ matrix whose kth row contains the amplifying factors of all M relay antennas at the kth time block, $\mathcal{D}_k\{\cdot\}$ is the operator that makes a diagonal matrix by selecting the kth row and putting it on the main diagonal while putting zeros elsewhere, $\mathbf{V}_{r,k}$ is the $M \times L$ noise matrix at the relay nodes during the kth time block, $\mathbf{V}_{d,k}^{(1)}$

[3]The variances are set to normalize the effect of number of transmit antennas to the receive signal-to-noise ratio. The relay nodes are assumed to be of equal distance to the source and the destination nodes with a path loss factor of 3.

[4]Diagonal relay amplifying matrix is only used for the purpose of channel estimation. During the normal communication period, however, the relay amplifying matrix does not need to be diagonal.

and $\mathbf{V}_{d,k}^{(2)}$ are $N_d \times L$ noise matrices at the destination node during the kth time block, and $\mathbf{Y}_k^{(1)}$ and $\mathbf{Y}_k^{(2)}$ are matrices containing the first and the last N_d rows of $\mathbf{Y}_k$, respectively.

At the destination node, by multiplying both sides of (10) with $\mathbf{S}^H$, we obtain

$$\mathbf{Y}_k\mathbf{S}^H = \begin{bmatrix} \mathbf{H}_{rd}\mathcal{D}_k\{\mathbf{F}\}\mathbf{H}_{sr} \\ \mathbf{H}_{sd} \end{bmatrix} + \begin{bmatrix} \mathbf{H}_{rd}\mathcal{D}_k\{\mathbf{F}\}\mathbf{V}_{r,k}\mathbf{S}^H + \mathbf{V}_{d,k}^{(1)}\mathbf{S}^H \\ \mathbf{V}_{d,k}^{(2)}\mathbf{S}^H \end{bmatrix}, \quad k = 1, \cdots, K. \quad (11)$$

From (11), an LS estimate of $\mathbf{H}_{sd}$ is given by

$$\hat{\mathbf{H}}_{sd} = \frac{1}{K}\mathbf{Y}^{(2)}(\mathbf{1}_K \otimes \mathbf{S})^H = \mathbf{H}_{sd} + \frac{1}{K}\mathbf{V}_d^{(2)}(\mathbf{1}_K \otimes \mathbf{S})^H$$

where $\mathbf{Y}^{(2)} \triangleq \left[\mathbf{Y}_1^{(2)}, \mathbf{Y}_2^{(2)}, \cdots, \mathbf{Y}_K^{(2)}\right]$, $\mathbf{V}_d^{(2)} \triangleq \left[\mathbf{V}_{d,1}^{(2)}, \mathbf{V}_{d,2}^{(2)}, \cdots, \mathbf{V}_{d,K}^{(2)}\right]$, $\mathbf{1}_K$ denotes a $1 \times K$ vector with all 1 elements, and $\otimes$ stands for the Kronecker matrix product [21]. In the following, we show how to estimate $\mathbf{H}_{rd}$ and $\mathbf{H}_{sr}$ at the destination node.

A. PARAFAC model and identifiability of channel matrices

Let us introduce

$$\tilde{\mathbf{X}}_k \triangleq \mathbf{Y}_k^{(1)}\mathbf{S}^H = \mathbf{X}_k + \mathbf{V}_k, \quad k = 1, \cdots, K \quad (12)$$

$$\mathbf{X}_k \triangleq \mathbf{H}_{rd}\mathcal{D}_k\{\mathbf{F}\}\mathbf{H}_{sr}, \quad k = 1, \cdots, K \quad (13)$$

$$\mathbf{V}_k \triangleq \mathbf{H}_{rd}\mathcal{D}_k\{\mathbf{F}\}\mathbf{V}_{r,k}\mathbf{S}^H + \mathbf{V}_{d,k}^{(1)}\mathbf{S}^H, \quad k = 1, \cdots, K \quad (14)$$

where $\mathbf{X}_k$ is the matrix-of-interest containing both $\mathbf{H}_{rd}$ and $\mathbf{H}_{sr}$, $\mathbf{V}_k$ is the effective noise matrix, and $\tilde{\mathbf{X}}_k$ is a noisy observation of $\mathbf{X}_k$. We would like to mention that $\mathbf{F}$ is chosen beforehand and is known at the destination node. The optimal $\mathbf{F}$ is very difficult to obtain for the PARAFAC-based channel estimation algorithm. Nevertheless, an intuitive way of designing $\mathbf{F}$ will be discussed later. By assembling the set of K matrices in (13) together along the direction of the index k (the third dimension), we obtain an $N_d \times N_s \times K$ three-way array $\underline{\mathbf{X}}$, whose (i, j, k)-th element is given by

$$x(i, j, k) = \sum_{m=1}^{M} h_{rd}(i, m) f(k, m) h_{sr}(m, j) \quad (15)$$

for all $i = 1, \cdots, N_d$, $j = 1, \cdots, N_s$, and $k = 1, \cdots, K$. Here $h_{rd}(i, m)$, $f(k, m)$, and $h_{sr}(m, j)$ stand for the (i, m)-th, (k, m)-th, and (m, j)-th elements of $\mathbf{H}_{rd}$, $\mathbf{F}$, and $\mathbf{H}_{sr}$, respectively. Equation (15) expresses $x(i, j, k)$ as a sum of M rank-1 triple products, which is known as the trilinear decomposition, or PARAFAC[5] analysis of $x(i, j, k)$ [14]-[16]. Correspondingly, assembling K matrices of $\tilde{\mathbf{X}}_k$ in (12) along the index k leads to a noise-contaminated $\underline{\mathbf{X}}$ given by $\underline{\tilde{\mathbf{X}}} = \underline{\mathbf{X}} + \underline{\mathbf{V}}$, where $\underline{\mathbf{V}}$ is obtained by assembling K noise matrices in (14).

Let us denote the Kruskal rank (or k-rank) [22] of a matrix $\mathbf{A}$ as $k_{\mathbf{A}}$, which is the maximum integer k, such that *any* k columns drawn from $\mathbf{A}$ are linearly independent. Note that Kruskal rank is always less than or equal to the conventional matrix rank. It can be easily checked that if $\mathbf{A}$ is full column rank, then it is also full Kruskal rank. It can be shown by using the identifiability theorem of the PARAFAC model in [15] and [22] that if

$$k_{\mathbf{H}_{rd}} + k_{\mathbf{F}} + k_{\mathbf{H}_{sr}} \geq 2M + 2 \quad (16)$$

then the triple $(\mathbf{H}_{rd}, \mathbf{F}, \mathbf{H}_{sr})$ is unique up to permutation and scaling ambiguities, i.e., if there exists any other triple $(\bar{\mathbf{H}}_{rd}, \bar{\mathbf{F}}, \bar{\mathbf{H}}_{sr})$ that gives rise to (13), then it is related to $(\mathbf{H}_{rd}, \mathbf{F}, \mathbf{H}_{sr})$ via

$$\bar{\mathbf{H}}_{rd} = \mathbf{H}_{rd}\mathbf{\Pi}\mathbf{\Delta}_1, \quad \bar{\mathbf{F}} = \mathbf{F}\mathbf{\Pi}\mathbf{\Delta}_2, \quad \bar{\mathbf{H}}_{sr}^T = \mathbf{H}_{sr}^T\mathbf{\Pi}\mathbf{\Delta}_3 \quad (17)$$

where $\mathbf{\Pi}$ is an $M \times M$ permutation matrix, and $\mathbf{\Delta}_i$, $i = 1, 2, 3$, are $M \times M$ diagonal (complex) scaling matrices satisfying

$$\mathbf{\Delta}_1\mathbf{\Delta}_2\mathbf{\Delta}_3 = \mathbf{I}_M. \quad (18)$$

Inequality (16) establishes the sufficient condition for the identifiability of $(\mathbf{H}_{rd}, \mathbf{F}, \mathbf{H}_{sr})$. Since $\mathbf{F}$ is chosen beforehand (e.g., based on the DFT matrix as shown later), one can guarantee that $\mathbf{F}$ has full k-rank. Moreover, both $\mathbf{H}_{sr}$ and $\mathbf{H}_{rd}$ are random matrices, and hence have full k-rank. Therefore, in such case, condition (16) becomes

$$\min(N_d, M) + \min(K, M) + \min(N_s, M) \geq 2M + 2. \quad (19)$$

From (19), the identifiability condition can be summarized in the following theorem.

THEOREM 1: The PARAFAC model (15) is identifiable only if $N_s \geq 2$, $N_d \geq 2$, and $2 \leq M \leq N_s + N_d - 2$. Moreover, for different N_s, N_d, and M, the lower bound of K satisfying (19) is given by

$$K \geq \begin{cases} 2M + 2 - N_s - N_d & M \geq N_s, N_d \\ M + 2 - N_d & N_d \leq M \leq N_s \\ M + 2 - N_s & N_s \leq M \leq N_d \\ 2 & M \leq N_s, N_d \end{cases}. \quad (20)$$

For all four cases in (20), the lower bound of K is no greater than M.

PROOF: The proof can be done by expanding the three $\min(\cdot)$ operators in (19).

- If $M \geq N_s, N_d \longrightarrow \min(K, M) \geq 2M + 2 - N_d - N_s \longrightarrow K \geq 2M + 2 - N_d - N_s$, and $M \geq 2M + 2 - N_d - N_s \longrightarrow M \leq N_s + N_d - 2$. This together with $M \geq N_s, N_d$, we have $N_s, N_d \geq 2$ and $2 \leq M \leq N_s + N_d - 2$. Since $M \leq N_s + N_d - 2$, it holds that $2M + 2 - N_s - N_d \leq M$;
- If $N_d \leq M \leq N_s \longrightarrow \min(K, M) \geq M + 2 - N_d \longrightarrow K \geq M + 2 - N_d$, and $M \geq M + 2 - N_d \longrightarrow N_d \geq 2$. This together with $N_d \leq M \leq N_s$, we have $N_s, N_d \geq 2$ and $2 \leq M \leq N_s + N_d - 2$. Since $N_d \geq 2$, there is $M + 2 - N_d \leq M$;
- If $N_s \leq M \leq N_d \longrightarrow \min(K, M) \geq M + 2 - N_s \longrightarrow K \geq M + 2 - N_s$, and $M \geq M + 2 - N_s \longrightarrow N_s \geq 2$.

[5]PARAFAC is a multi-way method originating from psychometrics [14] and has recently found applications in array signal processing [15] and communications [16]. Generalizing the concept of low-rank decomposition to higher way arrays or tensors, PARAFAC is instrumental in the analysis of data arrays indexed by three or more independent variables, just like singular value decomposition (SVD) is instrumental in ordinary matrix (two-way array) analysis. Unlike SVD, PARAFAC does not impose orthogonality constraints. The reason is that in contrast to low-rank matrix decomposition, low-rank decomposition of higher order tensorial data is essentially unique under certain conditions.

This together with $N_s \le M \le N_d$, we have $N_s, N_d \ge 2$ and $2 \le M \le N_s + N_d - 2$. Since $N_s \ge 2$, it holds that $M + 2 - N_s \le M$;

- If $M \le N_s, N_d \longrightarrow \min(K, M) \ge 2 \longrightarrow K \ge 2$, and $M \ge 2$. This together with $M \le N_s, N_d$, we have $N_s, N_d \ge 2$ and $2 \le M \le N_s + N_d - 2$.

Summarizing the four cases above, we obtain the necessary conditions for identifiability in the PARAFAC model (15) as $N_s \ge 2$, $N_d \ge 2$, and $2 \le M \le N_s + N_d - 2$. The lower bound of K in each case, which is less than or equal to M, is also given above. □

Interestingly, it is shown in Theorem 1 that under the mild condition of $N_s, N_d \ge 2$ and $2 \le M \le N_s + N_d - 2$, the minimal K required in the proposed PARAFAC-based channel estimation algorithm can be less than M. While in [9] and [10], at least $K = M$ training data blocks are required to perform the channel estimation. Therefore, the proposed algorithm has a higher spectral efficiency than those in [9] and [10]. Moreover, Theorem 1 shows that if $N_d \ge M$ and $N_s \ge M$, then two training data blocks ($K = 2$) are sufficient to estimate both $\mathbf{H}_{rd}$ and $\mathbf{H}_{sr}$ at the destination node. We also observe that if (20) is satisfied, then it holds that $KN_d > M$ and $KN_s > M$. We would like to mention that since $N_s \ge 2$ and $N_d \ge 2$ are required, which implies that $\mathbf{H}_{sr}$ and $\mathbf{H}_{rd}$ need to be matrices, the PARAFAC-based MIMO relay channel estimation algorithm can not be straightforwardly applied to relay systems with $N_s = 1$ and/or $N_d = 1$.

B. Bilinear alternating least-squares (BALS) fitting

In this subsection, we develop a BALS algorithm to estimate $\mathbf{H}_{sr}$ and $\mathbf{H}_{rd}$ by carrying out the PARAFAC model fitting with known $\mathbf{F}$. First we show some rearrangements of three-way arrays $\underline{\mathbf{X}}$, $\underline{\mathbf{V}}$, and $\underline{\tilde{\mathbf{X}}}$ which will be used later.

By stacking K matrices of $\mathbf{X}_k$ in (13) on top of each other, we obtain

$$\mathbf{X} \triangleq \begin{bmatrix} \mathbf{X}_1 \\ \vdots \\ \mathbf{X}_K \end{bmatrix} = \begin{bmatrix} \mathbf{H}_{rd}\mathcal{D}_1\{\mathbf{F}\} \\ \vdots \\ \mathbf{H}_{rd}\mathcal{D}_K\{\mathbf{F}\} \end{bmatrix} \mathbf{H}_{sr} = (\mathbf{F} \odot \mathbf{H}_{rd})\mathbf{H}_{sr} \tag{21}$$

where $\odot$ stands for the Khatri-Rao (column-wise Kronecker) matrix product [21]. Correspondingly, stacking matrices $\tilde{\mathbf{X}}_k$ in (12) on top of each other gives rise to

$$\tilde{\mathbf{X}} = \begin{bmatrix} \mathbf{X}_1 \\ \vdots \\ \mathbf{X}_K \end{bmatrix} + \begin{bmatrix} \mathbf{V}_1 \\ \vdots \\ \mathbf{V}_K \end{bmatrix} = \mathbf{X} + \mathbf{V}. \tag{22}$$

By slicing $\underline{\mathbf{X}}$ perpendicular to the dimension of j, we obtain a set of N_s matrices $\mathbf{Z}_j = \mathbf{F}\mathcal{D}_j\{\mathbf{H}_{sr}^T\}\mathbf{H}_{rd}^T$, $j = 1, \cdots, N_s$. By stacking N_s matrices of $\mathbf{Z}_j$ on top of each other, we have

$$\mathbf{Z} \triangleq \begin{bmatrix} \mathbf{Z}_1 \\ \vdots \\ \mathbf{Z}_{N_s} \end{bmatrix} = \begin{bmatrix} \mathbf{F}\mathcal{D}_1\{\mathbf{H}_{sr}^T\} \\ \vdots \\ \mathbf{F}\mathcal{D}_{N_s}\{\mathbf{H}_{sr}^T\} \end{bmatrix} \mathbf{H}_{rd}^T = (\mathbf{H}_{sr}^T \odot \mathbf{F})\mathbf{H}_{rd}^T. \tag{23}$$

Similarly, by slicing $\underline{\tilde{\mathbf{X}}}$ perpendicular to the dimension of j and stacking the resulting matrices on top of each other, we have

$$\tilde{\mathbf{Z}} = \begin{bmatrix} \mathbf{Z}_1 \\ \vdots \\ \mathbf{Z}_{N_s} \end{bmatrix} + \begin{bmatrix} \mathbf{N}_1 \\ \vdots \\ \mathbf{N}_{N_s} \end{bmatrix} \tag{24}$$

where $\mathbf{N}_j$, $j = 1, \cdots, N_s$, are the slabs of $\underline{\mathbf{V}}$ along the dimension of j.

The BALS fitting starts at a random $\hat{\mathbf{H}}_{rd}$. In each iteration, we first update $\mathbf{H}_{sr}$ using the LS fitting with fixed $\mathbf{F}$ and $\hat{\mathbf{H}}_{rd}$. Using (21) and (22), we obtain an updated $\mathbf{H}_{sr}$ as

$$\hat{\mathbf{H}}_{sr} = \arg\min_{\mathbf{H}_{sr}} \left\|\tilde{\mathbf{X}} - (\mathbf{F} \odot \hat{\mathbf{H}}_{rd})\mathbf{H}_{sr}\right\| = (\mathbf{F} \odot \hat{\mathbf{H}}_{rd})^{\dagger}\, \tilde{\mathbf{X}} \tag{25}$$

where $\|\cdot\|$ denotes the matrix Frobenius norm. Then we update $\mathbf{H}_{rd}$ through the LS fitting with known $\mathbf{F}$ and $\hat{\mathbf{H}}_{sr}$, and obtain $\hat{\mathbf{H}}_{rd}$ using (23) and (24) as

$$\hat{\mathbf{H}}_{rd} = \arg\min_{\mathbf{H}_{rd}} \left\|\tilde{\mathbf{Z}} - (\hat{\mathbf{H}}_{sr}^T \odot \mathbf{F})\mathbf{H}_{rd}^T\right\| = \left[(\hat{\mathbf{H}}_{sr}^T \odot \mathbf{F})^{\dagger}\, \tilde{\mathbf{Z}}\right]^T. \tag{26}$$

Since the conditional update of matrices in (25) and (26) may either improve or maintain but can not worsen the current LS fit, a monotonic convergence of the BALS procedure to (at least) a locally optimal solution follows directly from this observation [15]. The procedure of the BALS fitting is listed in Table I, where ε is a positive constant close to 0, and the matrix with superscript (n) denotes the estimated matrix at the nth iteration. Theoretically, for some particular data sets, the convergence of the BALS algorithm can be extremely slow. However, since both $\mathbf{H}_{sr}$ and $\mathbf{H}_{rd}$ are random matrices, the probability that both matrices fall in such data sets is very small. For large values of N_s, M, and N_d, such probability is almost zero. It will be shown in Section IV that the BALS algorithm typically converges in only a few iterations.

TABLE I
PROCEDURE OF THE BALS FITTING

1) Initialize the algorithm with a given $\mathbf{F}$ and a random $\mathbf{H}_{rd}^{(0)}$; Set $\delta(0) = \infty$ and $n = 1$.
2) Update $\mathbf{H}_{sr}^{(n)}$ as (25) using $\mathbf{H}_{rd}^{(n-1)}$; Update $\mathbf{H}_{rd}^{(n)}$ as (26) using $\mathbf{H}_{sr}^{(n)}$; Calculate $\delta(n) = \left\|\tilde{\mathbf{X}} - (\mathbf{F} \odot \mathbf{H}_{rd}^{(n)})\mathbf{H}_{sr}^{(n)}\right\|$.
3) If $\left[\delta(n-1) - \delta(n)\right]/\delta(n) \le \varepsilon$, then end.
Otherwise, let $n := n + 1$ and go to step 2).

Since $\mathbf{F}$ is known, the BALS algorithm delivers an estimation of $\mathbf{H}_{sr}$ and $\mathbf{H}_{rd}$ with only a scaling ambiguity $\mathbf{\Delta}_1$ at the convergence point, i.e., $\mathbf{\Pi} = \mathbf{I}_M$, $\mathbf{\Delta}_2 = \mathbf{I}_M$ in (17), and $\mathbf{\Delta}_3 = \mathbf{\Delta}_1^{-1}$ according to (18). This scaling ambiguity also exists in [9] and [10], and can be resolved through normalization as in [9], [10], [15].

The major computation task in the proposed BALS algorithm lies in the LS fittings in (25) and (26). Thus the per-iteration complexity of the BALS algorithm can be estimated as $\mathcal{O}(MKN_dN_s + M^3)$. The overall complexity of the BALS algorithm depends on the number of iterations and will be further discussed in Section IV. Note that the computational complexity of the LS-based algorithm in [9] can be estimated as $\mathcal{O}(MKN_dN_s + M^3 + MN_d^2N_s)$, where the three terms are from matrix multiplications, matrix inversion, and M matrix SVDs, respectively.

Now we present an intuitive choice of $\mathbf{F}$. By slicing $\underline{\tilde{\mathbf{X}}}$ perpendicular to the dimension of i and stacking the resulting matrices on top of each other, we have

$$\mathbf{P} = (\mathbf{H}_{rd} \odot \mathbf{H}_{sr}^T)\mathbf{F}^T + \mathbf{M}$$

where $\mathbf{M}$ is the corresponding noise matrix obtained by slicing $\underline{\mathbf{V}}$ perpendicular to the dimension of i and stacking the resulting matrices on top of each other. Let us denote $\mathbf{H}_{srd} \triangleq \mathbf{H}_{rd}\odot\mathbf{H}_{sr}^T$. Since $\mathrm{E}\big[\mathbf{H}_{srd}\mathbf{H}_{srd}^H\big] = \mathbf{I}_{N_sN_d}$, if $K \geq M$ and $\mathbf{M}$ have i.i.d. entries, the optimal $\mathbf{F}$ minimizing the MSE of a linear estimation of $\mathbf{H}_{srd}$ is unitary ($\mathbf{F}^H\mathbf{F} = \mathbf{I}_M$). However, it can be shown that the elements in $\mathbf{M}$ are correlated and the covariance matrix of $\mathbf{M}$ is a complicated function of $\mathbf{F}$. Thus, strictly speaking, a unitary $\mathbf{F}$ is not optimal in general. Nevertheless, such $\mathbf{F}$ is still a good choice especially when the signal-to-noise ratio is medium to high at channel training stage. In numerical simulations, we also find that the DFT matrix (which satisfies $\mathbf{F}^H\mathbf{F} = \mathbf{I}_M$) is a good choice for $\mathbf{F}$.

C. Linear minimal mean-squared error (LMMSE) estimation

It can be seen from (14) that the covariance matrix of the effective noise $\mathbf{V}_k$ at the destination node is given by $\mathbf{C}_k \triangleq \mathrm{E}[\mathbf{V}_k\mathbf{V}_k^H] = N_s\big(\mathbf{H}_{rd}\mathcal{D}_k\{\mathbf{F}\}(\mathcal{D}_k\{\mathbf{F}\})^H\mathbf{H}_{rd}^H + \mathbf{I}_{N_d}\big)$, $k = 1, \cdots, K$. Obviously, $\mathbf{V}_k$ is non-white due to the channel $\mathbf{H}_{rd}$. Therefore, after an initial estimation of $\mathbf{H}_{rd}$ by the BALS algorithm in Section III-B, an improved estimation of $\mathbf{H}_{sr}$ can be obtained by the LMMSE approach as $\breve{\mathbf{H}}_{sr} = \mathbf{T}_{sr}^H\tilde{\mathbf{X}}$, where $\mathbf{T}_{sr}$ is the $KN_d \times M$ weight matrix. The MSE of channel estimation can be written as

$$\mathrm{E}\Big[\mathrm{tr}\big((\breve{\mathbf{H}}_{sr} - \mathbf{H}_{sr})(\breve{\mathbf{H}}_{sr} - \mathbf{H}_{sr})^H\big)\Big] = \mathrm{tr}\Big(\big(\mathbf{T}_{sr}^H(\mathbf{F}\odot\hat{\mathbf{H}}_{rd}) - \mathbf{I}_M\big)\big(\mathbf{T}_{sr}^H(\mathbf{F}\odot\hat{\mathbf{H}}_{rd}) - \mathbf{I}_M\big)^H + \mathbf{T}_{sr}^H\hat{\mathbf{C}}\mathbf{T}_{sr}\Big) \quad (27)$$

where $\hat{\mathbf{C}} = \mathrm{bd}\big[\hat{\mathbf{C}}_1, \hat{\mathbf{C}}_2, \cdots, \hat{\mathbf{C}}_K\big]$, and $\hat{\mathbf{C}}_k = N_s\big(\hat{\mathbf{H}}_{rd}\mathcal{D}_k\{\mathbf{F}\}(\mathcal{D}_k\{\mathbf{F}\})^H\hat{\mathbf{H}}_{rd}^H + \mathbf{I}_{N_d}\big)$, $k = 1, \cdots, K$, is an estimate of $\mathbf{C}_k$ using $\hat{\mathbf{H}}_{rd}$. The weight matrix minimizing (27) is given by

$$\mathbf{T}_{sr} = \big((\mathbf{F}\odot\hat{\mathbf{H}}_{rd})(\mathbf{F}\odot\hat{\mathbf{H}}_{rd})^H + \hat{\mathbf{C}}\big)^{-1}(\mathbf{F}\odot\hat{\mathbf{H}}_{rd}). \quad (28)$$

It will be seen in Section IV that there is an obvious improvement in the estimation of $\mathbf{H}_{sr}$ by using (28) after the convergence of the BALS algorithm.

Similarly, we expect that the initial estimation of $\mathbf{H}_{rd}$ can be improved by the LMMSE approach. It can be shown from (24) that the covariance matrix of the noise $\mathbf{N}_j$, denoted as $\mathbf{\Theta}_j \triangleq \mathrm{E}[\mathbf{N}_j\mathbf{N}_j^H]$, $j = 1, \cdots, N_s$, is a diagonal matrix whose (k, k)-th diagonal element is given by $\sum_{m=1}^{M}\|f(k,m)\mathbf{h}_{rd,m}\|^2 + N_d$, where $\mathbf{h}_{rd,m}$ is the mth column of $\mathbf{H}_{rd}$. Thus, an improved LMMSE estimate of $\mathbf{H}_{rd}$ can be obtained as $\breve{\mathbf{H}}_{rd} = [\mathbf{T}_{rd}^H\tilde{\mathbf{Z}}]^T$, where $\mathbf{T}_{rd}$ is the $KN_s \times M$ weight matrix with

$$\mathbf{T}_{rd} = \big((\breve{\mathbf{H}}_{sr}^T\odot\mathbf{F})(\breve{\mathbf{H}}_{sr}^T\odot\mathbf{F})^H + \hat{\mathbf{\Theta}}\big)^{-1}(\breve{\mathbf{H}}_{sr}^T\odot\mathbf{F}). \quad (29)$$

Here $\hat{\mathbf{\Theta}} = \mathrm{bd}\big[\hat{\mathbf{\Theta}}_1, \hat{\mathbf{\Theta}}_2, \cdots, \hat{\mathbf{\Theta}}_{N_s}\big]$, and $[\hat{\mathbf{\Theta}}_j]_{k,k} = \sum_{m=1}^{M}\|f(k,m)\hat{\mathbf{h}}_{rd,m}\|^2 + N_d$, $k = 1, \cdots, K$, $j = 1, \cdots, N_s$, is an estimate of $[\mathbf{\Theta}_j]_{k,k}$ using $\hat{\mathbf{H}}_{rd}$.

We would like to mention that in [10], the WLS approach is used to improve the channel estimation after the LS algorithm. It will be shown in Section IV that the LMMSE algorithm yields a smaller MSE of channel estimation (particularly for estimating $\mathbf{H}_{rd}$) than that of the WLS method in [10].

D. Extension to channel estimation in two-way MIMO relay systems

In the following, we show that the proposed algorithm can also be used for channel estimation in two-way MIMO relay systems.

In a two-way relay system, two users exchange their information through one or multiple relay nodes [23]. The received signal matrices at two users during the kth time block of the channel training period are given respectively by

$$\mathbf{Y}_{1,k} = \mathbf{H}_{1,r}\mathcal{D}_k\{\mathbf{F}\}\mathbf{H}_{r,2}\mathbf{S}_2 + \mathbf{H}_{1,r}\mathcal{D}_k\{\mathbf{F}\}\mathbf{H}_{r,1}\mathbf{S}_1 + \mathbf{H}_{1,r}\mathcal{D}_k\{\mathbf{F}\}\mathbf{V}_{r,k} + \mathbf{V}_{1,k}, \quad k = 1, \cdots, K \quad (30)$$

$$\mathbf{Y}_{2,k} = \mathbf{H}_{2,r}\mathcal{D}_k\{\mathbf{F}\}\mathbf{H}_{r,1}\mathbf{S}_1 + \mathbf{H}_{2,r}\mathcal{D}_k\{\mathbf{F}\}\mathbf{H}_{r,2}\mathbf{S}_2 + \mathbf{H}_{2,r}\mathcal{D}_k\{\mathbf{F}\}\mathbf{V}_{r,k} + \mathbf{V}_{2,k}, \quad k = 1, \cdots, K \quad (31)$$

where $\mathbf{H}_{r,i}$, $i = 1, 2$, is the MIMO channel from user i to all relay nodes, $\mathbf{H}_{i,r}$, $i = 1, 2$, is the MIMO channel from all relay nodes to user i, and $\mathbf{V}_{i,k}$, $i = 1, 2$, is the noise matrix at user i during the kth time block.

The $N_i \times L$ training sequence $\mathbf{S}_i$ chosen by user i, $i = 1, 2$, in (30) and (31) is designed such that

$$\mathbf{S}_i\mathbf{S}_i^H = \mathbf{I}_{N_i}, \quad i = 1, 2, \qquad \mathbf{S}_1\mathbf{S}_2^H = \mathbf{0}_{N_1\times N_2} \quad (32)$$

where N_i is the number of antennas at user i. Note that $\mathbf{S}_1$ and $\mathbf{S}_2$ satisfying (32) can be easily constructed from the normalized DFT matrix with $L \geq N_1 + N_2$. Multiplying both sides of (30) with $\mathbf{S}_2^H$ and both sides of (31) with $\mathbf{S}_1^H$, we have

$$\mathbf{Y}_{1,k}\mathbf{S}_2^H = \mathbf{H}_{1,r}\mathcal{D}_k\{\mathbf{F}\}\mathbf{H}_{r,2} + \mathbf{H}_{1,r}\mathcal{D}_k\{\mathbf{F}\}\mathbf{V}_{r,k}\mathbf{S}_2^H + \mathbf{V}_{1,k}\mathbf{S}_2^H, \quad k = 1, \cdots, K \quad (33)$$

$$\mathbf{Y}_{2,k}\mathbf{S}_1^H = \mathbf{H}_{2,r}\mathcal{D}_k\{\mathbf{F}\}\mathbf{H}_{r,1} + \mathbf{H}_{2,r}\mathcal{D}_k\{\mathbf{F}\}\mathbf{V}_{r,k}\mathbf{S}_1^H + \mathbf{V}_{2,k}\mathbf{S}_1^H, \quad k = 1, \cdots, K. \quad (34)$$

Now the proposed PARAFAC-based algorithm developed in Section III-A to Section III-C can be applied at user 1 to estimate $\mathbf{H}_{1,r}$ and $\mathbf{H}_{r,2}$ from (33) and at user 2 to estimate $\mathbf{H}_{2,r}$ and $\mathbf{H}_{r,1}$ from (34).

IV. Numerical Examples

In this section, we study the performance of the proposed channel estimation algorithm through numerical simulations. In particular, we compare the proposed algorithm with the conventional TST scheme in Section II, the LS-based algorithm in [9], and the WLS fitting algorithm in [10]. Note that the purpose of the WLS fitting in [10] is to improve the performance of the LS algorithm in [9]. To ensure a fair comparison, a factor of $\sqrt{K}$ is used to scale the training sequences $\mathbf{S}_1$ and $\mathbf{S}_2$ in the TST scheme such that the total energy spent on channel training is identical for all approaches. In the simulations, $\mathbf{F}$ is generated based on the DFT matrix, and the BALS algorithm is performed following the procedure in Table I with $\varepsilon = 1 \times 10^{-5}$. Similar to [9] and [10], the scaling ambiguity $\mathbf{\Delta}_1 = \mathbf{\Delta}_3^{-1}$ in the proposed algorithm is

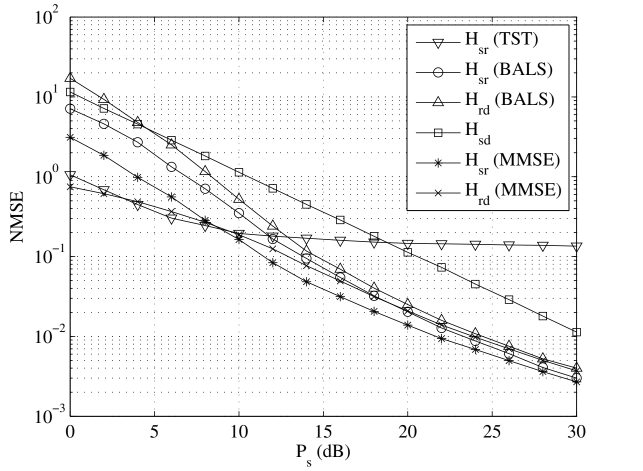


Fig. 2. Example 1: Normalized MSE versus P_s for i.i.d. MIMO channels. $K = 3$.

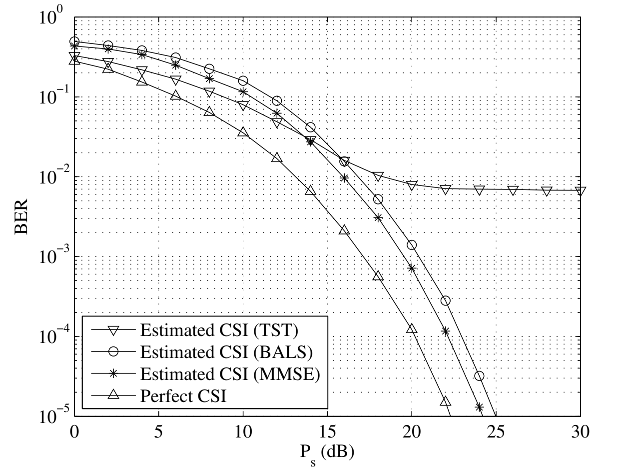


Fig. 3. Example 1: BER versus P_s for i.i.d. MIMO channels. $K = 3$.

removed by assuming that the first column of $\mathbf{H}_{sr}$ contains all one elements[6]. For each channel realization, the normalized MSE (NMSE) of channel estimation for different algorithms is calculated as $\|\hat{\mathbf{H}}_{sr} - \mathbf{H}_{sr}\|^2/\|\mathbf{H}_{sr}\|^2$ for the channel $\mathbf{H}_{sr}$, where $\hat{\mathbf{H}}_{sr}$ is the estimated value. The channel estimation errors of $\mathbf{H}_{rd}$ and $\mathbf{H}_{sd}$ are calculated in a similar way to that of $\mathbf{H}_{sr}$. All simulation results are averaged over 2000 independent channel realizations.

We consider a two-hop MIMO relay communication system with $M = 4$ single-antenna relay nodes, and the source and destination nodes are equipped with $N_s = N_d = 4$ antennas. Throughout the simulations, we use the minimal L, i.e., $L = N_s = 4$. The transmission power at the relay node is set to be 20dB above the noise level.

In the first example, we study the performance of the proposed algorithm and the TST approach with $K = 3$ where all channel matrices have i.i.d. complex Gaussian entries with zero-mean and variances of $1/N_s$, $1/M$, $1/(8N_s)$ for $\mathbf{H}_{sr}$, $\mathbf{H}_{rd}$, and $\mathbf{H}_{sd}$, respectively. Note that since $K < M$, the algorithms in [9] and [10] can not be applied in this case. The NMSE of both algorithms versus the source node transmission power P_s is shown in Fig. 2. Since the NMSE for the estimation of $\mathbf{H}_{rd}$ by the TST scheme does not change with P_s (see (7)), it is not displayed in Fig. 2 (neither in Figs. 4 and 6 later on). It can be seen that for the proposed algorithm, the NMSE of channel estimation decreases as P_s increases. As expected, the estimation of $\mathbf{H}_{sr}$ and $\mathbf{H}_{rd}$ is improved by carrying out the additional MMSE estimation. At the low P_s level, the TST scheme is better than the proposed algorithm. While at medium to high P_s levels, the proposed algorithm significantly outperforms the TST scheme even without using the additional MMSE estimation. In fact, the TST scheme has an error floor in estimating $\mathbf{H}_{sr}$. The reason is that as can be seen from (9), the estimation of $\mathbf{H}_{sr}$ in the TST scheme is extracted from the estimation of the compound channel $\bar{\mathbf{H}}$ and the estimation of $\mathbf{H}_{rd}$. Thus, the accuracy of $\hat{\bar{\mathbf{H}}}$ and $\hat{\mathbf{H}}_{rd}$ has a great impact on the estimation of $\mathbf{H}_{sr}$. While in the proposed algorithm, $\mathbf{H}_{sr}$ is estimated together with $\mathbf{H}_{rd}$. We also observe from Fig. 2 that for the proposed algorithm, the NMSE of estimating $\mathbf{H}_{sd}$ is larger than that of $\mathbf{H}_{sr}$ and $\mathbf{H}_{rd}$. This is due to the lower signal-to-noise ratio at the direct link as the source-destination distance is twice of the source-relay (or relay-destination) distance.

The impact of channel estimation on the system BER performance in this example is shown in Fig. 3. QPSK constellations are used to modulate the source symbols, and 3000 randomly generated bits are transmitted for each channel realization. It can be seen that at medium to high P_s levels, the proposed algorithm significantly outperforms the TST scheme even without using the additional MMSE estimation, and the TST scheme shows a high error floor. We also observe in Fig. 3 that around 1dB gain in P_s is obtained by using the additional MMSE estimation after the convergence of the BALS algorithm. At a BER of 1×10^{-4}, there is only around 2dB loss in P_s by using the estimated CSI obtained from the MMSE algorithm compared with the system using the perfect CSI.

In the second example, we consider correlated MIMO channels. Based on [19], we assume that $\mathbf{H}_l = \mathbf{R}_l^{\frac{1}{2}} \mathbf{H}_l^{\mathrm{w}} \mathbf{C}_l^{\frac{T}{2}}$, where $l \in [sr, rd, sd]$ denotes the link index. Here $\mathbf{H}_{sr}^{\mathrm{w}}$, $\mathbf{H}_{rd}^{\mathrm{w}}$, and $\mathbf{H}_{sd}^{\mathrm{w}}$ are complex Gaussian random matrices having i.i.d. entries with zero mean and variances of $1/N_s$, $1/M$, $1/(8N_s)$, respectively, $\mathbf{R}_l$ and $\mathbf{C}_l$ characterize the channel correlation at the receive side and the transmit side of link l, respectively. We adopt the commonly used exponential Toeplitz structure in [19] such that $[\mathbf{R}_l]_{m,n} = \mathcal{J}_0(2\pi|m-n|/r_l)$ and $[\mathbf{C}_l]_{m,n} = \mathcal{J}_0(2\pi|m-n|/c_l)$, where $\mathcal{J}_0(\cdot)$ is the zeroth order Bessel function of the first kind, r_l and c_l stand for the correlation coefficients which depend on physical factors such as the angle of arrival spread, spacing between antenna elements, and the wavelength at the center frequency [19]. For the sake of simplicity, we choose $c_l = r_l = 2$ for all $l \in [sr, rd, sd]$. The NMSE and BER performance of different algorithms in this example are displayed in Fig. 4 and Fig. 5, respectively. It can be observed that similar to Fig. 2 and Fig. 3, the proposed

[6]The scaling ambiguity is represented as $\hat{\mathbf{H}}_{sr} = \mathbf{\Delta}_3 \mathbf{H}_{sr}$ in (17) (with $\mathbf{\Pi} = \mathbf{I}_M$). Since the first column of $\mathbf{H}_{sr}$ contains all one elements, it can be seen that $[\mathbf{\Delta}_3]_{i,i} = \left[\hat{\mathbf{H}}_{sr}\right]_{i,1}$. Here $[\mathbf{A}]_{i,j}$ stands for the (i,j)-th element of matrix $\mathbf{A}$.

TABLE II
EXAMPLE 3: NMSE OF THE LS [9], THE WLS [10], AND THE PROPOSED BALS ALGORITHM

P_s (dB)	0	4	8	12	16	20	24	28
BALS ($\mathbf{H}_{sr}$)	4.4316	1.2800	0.3212	0.0891	0.0296	0.0118	0.0050	0.0025
LS [9] ($\mathbf{H}_{sr}$)	4.4316	1.2800	0.3212	0.0891	0.0296	0.0118	0.0050	0.0025
BALS ($\mathbf{H}_{rd}$)	0.9208	0.3678	0.1359	0.0527	0.0206	0.0086	0.0038	0.0020
LS [9] ($\mathbf{H}_{rd}$)	0.9208	0.3679	0.1359	0.0527	0.0206	0.0086	0.0038	0.0020
WLS [10] ($\mathbf{H}_{rd}$)	0.9207	0.3678	0.1358	0.0526	0.0204	0.0084	0.0037	0.0020

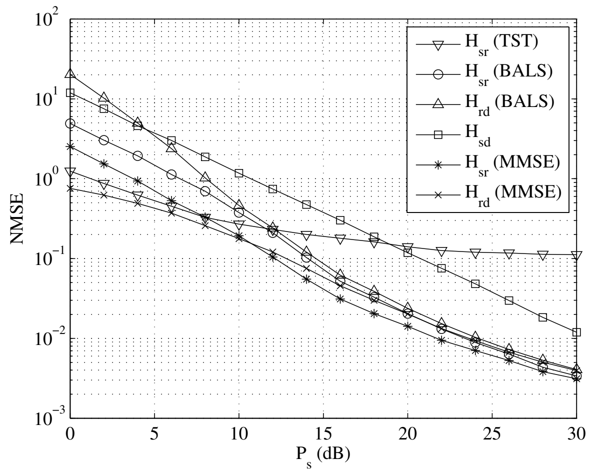


Fig. 4. Example 2: Normalized MSE versus P_s for correlated MIMO channels. $K = 3$.

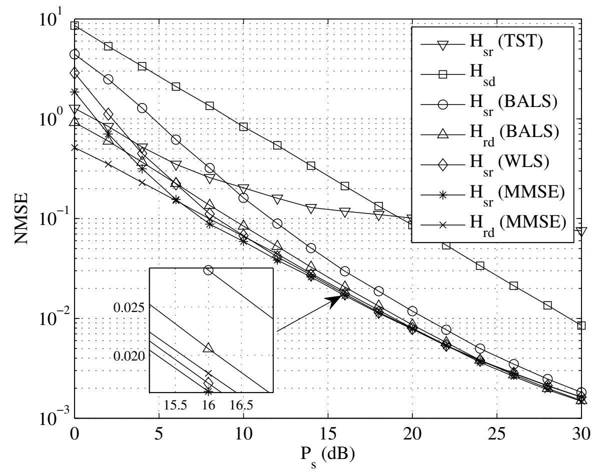


Fig. 6. Example 3: Normalized MSE versus P_s for i.i.d. MIMO channels. $K = 4$.

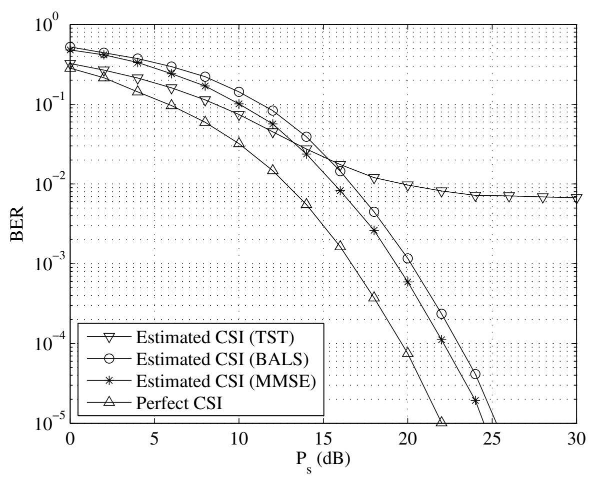


Fig. 5. Example 2: BER versus P_s for correlated MIMO channels. $K = 3$.

algorithm performs better than the TST algorithm.

In the third example, we simulate all algorithms with $K = 4$ and i.i.d. channel matrices. Since $K = M$, now we can compare the performance of the proposed algorithm with the algorithms developed in [9] and [10]. The NMSE of the LS algorithm in [9], the additional WLS fitting in [10], and the proposed algorithm is shown in Table II. It can be seen that the proposed BALS fitting yields the same NMSE as the LS approach. The NMSE of the proposed algorithm, the TST scheme, and the WLS fitting versus P_s is shown in Fig. 6, where we observe that as K is increased from 3 to 4, the NMSE of all algorithms is reduced compared with that in Fig. 2. Moreover, we see from Table II and Fig. 6 that the improvement in NMSE of the WLS fitting over the LS algorithm is obvious for the estimation of $\mathbf{H}_{sr}$, while the improvement for that of $\mathbf{H}_{rd}$ is negligible. The reason is that $\hat{\mathbf{C}}_k$ in (28) is non-diagonal, while $\hat{\mathbf{\Theta}}_j$ in (29) is diagonal. In contrast to the WLS fitting (Table II), it can be seen from Fig. 6 that the MMSE approach greatly reduces the NMSE of estimating $\mathbf{H}_{rd}$. From Fig. 6 we also observe that the MMSE approach yields a smaller NMSE in estimating $\mathbf{H}_{sr}$ compared with the WLS fitting.

For this example, with a random initialization of $\mathbf{H}_{rd}$, the average and maximum number of iterations over 2000 independent channel realizations required by the proposed BALS algorithm till convergence at different P_s level are listed in Table III. Based on the analysis of the overall complexity of the LS-based algorithm and the per-iteration complexity of the BALS algorithm in Section III-B, it can be seen from the second row of Table III that in average at medium and high P_s levels, the overall complexity of the proposed BALS algorithm is similar to that of the LS algorithm. When P_s is low, the complexity of the BALS algorithm is slightly higher than that of the LS algorithm. It can also be seen from the third row of Table III that at medium to high P_s levels (which is the P_s range in practical systems), the maximum number of iterations is only twice of or almost identical to the average ones.

For the third example, the comparison of BERs among the system using different channel estimation algorithms is demonstrated in Fig. 7. We observe from Fig. 7 that as K is increased from 3 to 4, the BER of all algorithms is reduced

TABLE III
ITERATIONS REQUIRED TILL CONVERGENCE BY THE PROPOSED BALS ALGORITHM

P_s (dB)	0	4	8	12	16	20	24	28
Iterations (average)	7	6	6	5	4	4	3	3
Iterations (maximum)	23	18	13	10	8	6	5	4

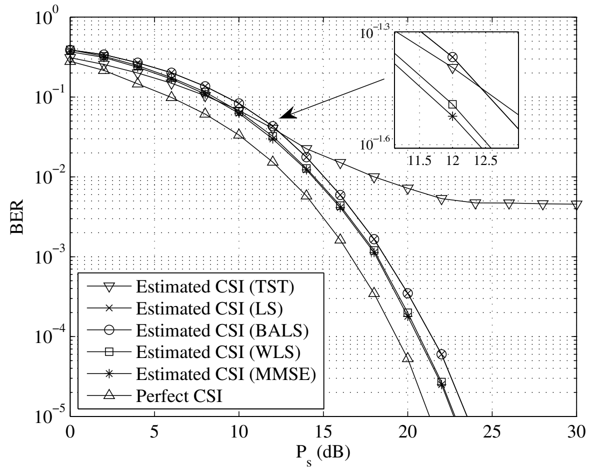


Fig. 7. Example 3: BER versus P_s for i.i.d. MIMO channels. $K = 4$.

compared with that in Fig. 3. Similar to Fig. 3, we see from Fig. 7 that the proposed algorithm significantly outperforms the TST scheme at medium to high P_s levels, where the latter scheme shows a high error floor. The LS approach in [9] has the same BER performance as the proposed BALS approach, while the proposed MMSE algorithm performs slightly better than that of the WLS algorithm in [10]. At a BER of 1×10^{-4}, the loss in P_s using the estimated CSI from the MMSE algorithm is less than 2dB compared with the system using the perfect CSI. This is quite reasonable for practical systems.

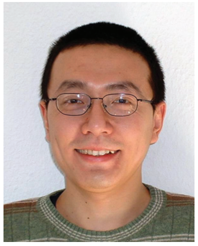
Based on the three numerical examples above, one can draw the following conclusions: (1) The proposed algorithm performs well in case of $K < M$ where the algorithms in [9] and [10] stop working; (2) When $K \geq M$, the proposed MMSE approach outperforms the algorithm in [10]; (3) The computational complexity of the proposed algorithm is similar to that of [9]; (4) The proposed algorithm performs well for both i.i.d. and correlated fading MIMO channels.

V. CONCLUSIONS

We have developed a novel PARAFAC-based channel estimation method for two-hop MIMO relay communication systems. The proposed algorithm provides the destination node with full knowledge of all channel matrices involved in the communication. Compared with existing approaches, the proposed algorithm requires less number of training data blocks, yields smaller channel estimation error, and is applicable for both one-way and two-way MIMO relay systems. Simulation results demonstrate the effectiveness of the proposed channel estimation algorithm.

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] A. S. Behbahani, R. Merched, and A. M. Eltawil, "Optimizations of a MIMO relay network," *IEEE Trans. Signal Process.*, vol. 56, pp. 5062–5073, Oct. 2008.
[5] Y. Rong, X. Tang, and Y. Hua, "A unified framework for optimizing linear non-regenerative multicarrier MIMO relay communication systems," *IEEE Trans. Signal Process.*, vol. 57, pp. 4837–4851, Dec. 2009.
[6] Y. Rong, "Optimal linear non-regenerative multi-hop MIMO relays with MMSE-DFE receiver at the destination," *IEEE Trans. Wireless Commun.*, vol. 9, pp. 2268–2279, July 2010.
[7] Y. Rong, "Optimal joint source and relay beamforming for MIMO relays with direct link," *IEEE Commun. Lett.*, vol. 14, pp. 390–392, May 2010.
[8] F. Roemer and M. Haardt, "Tensor-based channel estimation and iterative refinements for two-way relaying with multiple antennas and spatial reuse," *IEEE Trans. Signal Process.*, vol. 58, pp. 5720–5735, Nov. 2010.
[9] P. Lioliou and M. Viberg, "Least-squares based channel estimation for MIMO relays," in *Proc. 2008 IEEE WSA*, pp. 90–95.
[10] P. Lioliou, M. Viberg, and M. Coldrey, "Performance analysis of relay channel estimation," in *Proc. 2009 IEEE Asilomar Conference on Signals, Systems, and Computers*, pp. 1533–1537.
[11] F. Gao, T. Cui, and A. Nallanathan, "On channel estimation and optimal training design for amplify and forward relay networks," *IEEE Trans. Wireless Commun.*, vol. 7, pp. 1907–1916, May 2008.
[12] F. Gao, B. Jiang, X. Gao, and X.-D. Zhang, "Superimposed training based channel estimation for OFDM modulated amplify-and-forward relay networks," *IEEE Trans. Commun.*, vol. 59, pp. 2029–2039, July 2011.
[13] T. Kong and Y. Hua, "Optimal channel estimation and training design for MIMO relays," in *Proc. 2010 IEEE Asilomar Conference on Signals, Systems, and Computers*, pp. 663–667.
[14] R. A. Harshman, "Foundations of PARAFAC procedure: Models and conditions for an 'explanatory' multi-mode factor analysis," *UCLA Working Papers in Phonetics*, vol. 16, pp. 1–84, 1970.
[15] N. D. Sidiropoulos, R. Bro, and G. B. Giannakis, "Parallel factor analysis in sensor array processing," *IEEE Trans. Signal Process.*, vol. 48, pp. 2377–2388, Aug. 2000.
[16] N. D. Sidiropoulos, G. B. Giannakis, and R. Bro, "Blind PARAFAC receivers for DS-CDMA systems," *IEEE Trans. Signal Process.*, vol. 48, pp. 810–823, Mar. 2000.
[17] H. Wang, Y. Lin, and B. Chen, "Data-efficient blind OFDM channel estimation using receiver diversity," *IEEE Trans. Signal Process.*, vol. 51, pp. 2613–2623, Oct. 2003.
[18] A. A. Nasir, S. Durrani, and R. A. Kennedy, "Blind timing and carrier synchronization in decode and forward cooperative systems," in *Proc. 2011 IEEE Int. Conf. Commun.*
[19] 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.
[20] 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.
[21] J. W. Brewer, "Kronecker products and matrix calculus in system theory," *IEEE Trans. Circuits Syst.*, vol. 25, pp. 772–781, Sep. 1978.
[22] J. B. Kruskal, "Three-way arrays: rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics," *Linear Algebra Applicat.*, vol. 16, pp. 95–138, 1977.
[23] R. Zhang, Y.-C. Liang, C. C. Chai, and S. Cui, "Optimal beamforming for two-way multi-antenna relay channel with analogue network coding," *IEEE J. Sel. Areas Commun.*, vol. 27, pp. 699–712, June 2009.



Yue Rong (S'03-M'06-SM'11) received the B.E. degree from Shanghai Jiao Tong University, China, the M.Sc. degree from the University of Duisburg-Essen, Duisburg, Germany, and the Ph.D. degree (summa cum laude) from Darmstadt University of Technology, Darmstadt, Germany, all in electrical engineering, in 1999, 2002, and 2005, respectively.

From April 2001 to October 2001, he was a research assistant at the Fraunhofer Institute of Microelectronic Circuits and Systems, Duisburg, Germany. From October 2001 to March 2002, he was with Nokia Ltd., Bochum, Germany. From November 2002 to March 2005, he was a Research Associate at the Department of Communication Systems in the University of Duisburg-Essen. From April 2005 to January 2006, he was with the Institute of Telecommunications at Darmstadt University of Technology, as a Research Associate. From February 2006 to November 2007, he was a Postdoctoral Researcher with the Department of Electrical Engineering, University of California, Riverside. Since December 2007, he has been with the Department of Electrical and Computer Engineering, Curtin University of Technology, Perth, Australia, where he is now a Senior Lecturer. His research interests include signal processing for communications, wireless communications, wireless networks, applications of linear algebra and optimization methods, and statistical and array signal processing.

Dr. Rong received the Best Paper Award at the Third International Conference on Wireless Communications and Signal Processing, Nanjing, China, 2011, the Best Paper Award at the 16th Asia-Pacific Conference on Communications, Auckland, New Zealand, 2010, the 2010 Young Researcher of the Year Award of the Faculty of Science and Engineering at Curtin University, the 2004 Chinese Government Award for Outstanding Self-Financed Students Abroad (China), and the 2001-2002 DAAD/ABB Graduate Sponsoring Asia Fellowship (Germany). He has co-authored more than 70 referred IEEE journal and conference papers. He is an Editor of 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 to be published in 2012, a Guest Editor of the EURASIP JASP Special Issue on Signal Processing Methods for Diversity and Its Applications to be published in 2012, and has served as a TPC member for IEEE ICC, IWCMC, and ChinaCom.

Muhammad R. A. Khandaker (S'10) received the Bachelor of Science (Honours) degree in Computer Science and Engineering from Jahangirnagar University, Dhaka, Bangladesh in 2006 and Master of Science in Telecommunications Engineering from East West University, Dhaka, Bangladesh in 2007. He is currently working toward the Ph.D. degree at the Department of Electrical and Computer Engineering, Curtin University, Australia.

Mr. Khandaker started his career as a Jr. Hardware Design Engineer in Visual Magic Corporation Limited (VMCL-BD) in November 2005. He joined as a Lecturer in the Department of Computer Science and Engineering, IBAIS University, Dhaka, Bangladesh in January 2006. In November 2007, he joined as a Lecturer in the Department of Information and Communication Technology of Mawlana Bhasani Science and Technology University, Bangladesh. Finally, he joined as a Lecturer in the Institute of Information Technology, Jahangirnagar University, Dhaka, Bangladesh in January 2008.

He was awarded Curtin International Postgraduate Research Scholarship (CIPRS) for his Ph.D. study in 2009. He also received the Best Paper Award at the 16th Asia-Pacific Conference on Communications, Auckland, New Zealand, 2010.

Yong Xiang 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, multimedia processing, pattern recognition, and biomedical signal processing.