

Channel Estimation for Time-Varying MIMO Relay Systems

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

Abstract—In this paper, we investigate the channel estimation problem for multiple-input multiple-output (MIMO) relay communication systems with time-varying channels. The time-varying characteristic of the channels is described by the complex-exponential basis expansion model (CE-BEM). We propose a superimposed channel training algorithm to estimate the individual first-hop and second-hop time-varying channel matrices for MIMO relay systems. In particular, the estimation of the second-hop time-varying channel matrix is performed by exploiting the superimposed training sequence at the relay node, while the first-hop time-varying channel matrix is estimated through the source node training sequence and the estimated second-hop channel. To improve the performance of channel estimation, we derive the optimal structure of the source and relay training sequences that minimize the mean-squared error (MSE) of channel estimation. We also optimize the relay amplification factor that governs the power allocation between the source and relay training sequences. Numerical simulations demonstrate that the proposed superimposed channel training algorithm for MIMO relay systems with time-varying channels outperforms the conventional two-stage channel estimation scheme.

Index Terms—Channel estimation, MIMO relay, superimposed training, MMSE, time-varying channel, CE-BEM.

I. INTRODUCTION

WIRELESS communication networks have been a revolutionary part in the field of communication, and there is a strong demand for high speed and reliable wireless communication. Multiple-input multiple-output (MIMO) relay communication systems have been identified to be one of the promising solutions to high rate wireless communication [1], [2]. Many research works have been carried out to optimize the MIMO relay communication systems [3]–[12]. The optimal relay precoding matrix that maximizes the mutual information between the source and destination nodes for a three-node two-hop MIMO relay communication system has been derived in

[3] and [4]. In [5], the capacity of MIMO relay networks for Gaussian channel and Rayleigh fading channel has been investigated. Recently, a survey on transceiver design for amplify-and-forward (AF) MIMO relay systems has been presented in [6].

In [7], a simple protocol that adopts the transmit antenna selection with maximal-ratio combining has been proposed for multiuser MIMO relay networks. It has been shown in [7] that to maximize the performance of multiuser MIMO relay networks, precise channel state information (CSI) in real time is required. In [8], transmit antenna selection for MIMO multi-relay networks with different fading environments has been investigated. The effect of feedback delay and interference on AF relay networks using transmit beamforming has been discussed in [9]. Other works on the optimization of MIMO relay systems can be found in [10]–[12].

For MIMO relay systems discussed in [2]–[12], the knowledge of the instantaneous CSI is required for the retrieval of the source information at the destination node and the optimization of the MIMO relay systems by deriving the optimal source and relay precoding matrices. However, in practical MIMO relay communication systems, the instantaneous CSI is not available at any node, and hence, it needs to be estimated.

A least-squares (LS) channel estimator for MIMO relay communication systems has been developed in [13], where the individual source-relay and relay-destination channel matrices are estimated from the observed compound source-relay-destination channel matrix. This work has been improved in [14] by applying the weighted least-squares (WLS) fitting method. In [15], the individual channel matrices are estimated in two stages based on the linear minimum mean-squared error (MMSE) criterion, where the relay-destination channel matrix is estimated in the first stage, while the source-relay channel matrix is derived in the second stage. A parallel factor (PARAFAC) analysis based channel estimator has been developed in [16] for MIMO relay systems.

The channel estimation algorithms in [13]–[16] were developed assuming that the channels in MIMO relay systems are quasi-static block fading. However, with the increasing number of mobile wireless devices, the channels in MIMO relay systems are more likely to be affected by the changing environment caused by the movement of users. It is stated in [17] that in wireless relay systems, the Doppler spread is doubled when there is a relative motion between any two nodes. Thus, it is essential to address the channel estimation issues for MIMO relay systems with time-varying channels. An MMSE-based estimator for time-varying channels has been proposed in [18] based on the Gauss-Markov model, where the channel variation is captured through symbol-by-symbol updating [19].

Manuscript received December 14, 2014; revised April 20, 2015; accepted July 15, 2015. Date of publication July 22, 2015; date of current version December 8, 2015. This research was supported by the Australian Research Council's Discovery Projects funding scheme (project numbers DP110102076 and DP140102131). The associate editor coordinating the review of this paper and approving it for publication was A. Wyglinski.

C. W. R. Chiong and Y. Rong 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).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TWC.2015.2459700

In [20], the channel estimation problem has been investigated for doubly selective wireless fading channels, where the time-varying characteristic of channels is represented by the basis expansion model (BEM). In BEM, the channel is represented by the superposition of the time-varying basis functions weighted by time-invariant coefficients [21]. Note that the channel estimation algorithms in [18] and [20] were developed for single-hop single-antenna communication networks.

Channel estimation in AF relay networks under time-varying environment has been addressed in [22], with the derivation of optimal training sequences that minimize the MSE of channel estimation. However, the algorithm in [22] focuses on the estimation of the composite channel parameters. Further works on the estimation of the individual channel for AF relay networks have been done in [23]–[25]. Channel estimation and training sequence design for time-varying relay networks have been discussed in [23], where the complex-exponential BEM (CE-BEM) [20] was adopted to represent the time-varying channels in terms of Fourier bases. In [24], the Bayesian Cramér–Rao lower bounds for the estimation of individual channel matrices have been derived for time-varying AF relay networks. A time-multiplexed-superimposed training scheme for zero-prefixed block transmission has been proposed in [25] to estimate the individual time-varying channels in AF relay networks. However, the algorithms in [23]–[25] were developed for single-antenna relay networks, and the extension to MIMO relay networks is not straightforward.

Compared with single-antenna relay networks, channel estimation for time-varying MIMO relay communication systems is more challenging, as the number of variables to be estimated is much larger in MIMO relay systems than that of single-antenna relay networks. To our best knowledge, the problem of individual channel matrices estimation has not been addressed for time-varying MIMO relay systems. To fill the niche, in this paper, we propose a superimposed channel training algorithm to estimate the individual first-hop and second-hop time-varying channel matrices for MIMO relay communication systems. We use the CE-BEM adopted in [23] to capture the time-varying characteristic of the channel, and apply the pilot symbol aided modulation (PSAM) technique [26] for channel estimation. At the source and relay nodes, training symbols are inserted among information-carrying symbols in each data frame for channel estimation. In particular, the estimation of the second-hop time-varying channel matrix is performed by exploiting the superimposed training sequence at the relay node, while the first-hop time-varying channel matrix is estimated through the source node training sequence and the estimated second-hop channel. We would like to mention that all channel estimation works are implemented at the destination node to minimize the amount of signal processing at the relay node. We derive the optimal structure of the source and relay training sequences that minimize the MSE of the channel estimation. We also optimize the amplification factor at the relay node that determines the power allocation between the source and relay training sequences.

We would like to note that the algorithm developed in [23] is only applicable to single-antenna time-varying relay systems. Our paper investigates the individual channel matrices estimation for MIMO relay systems with time-varying channels, and

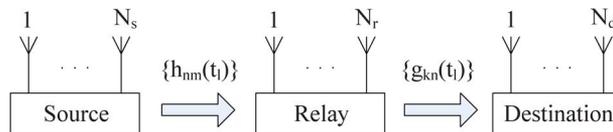


Fig. 1. Block diagram of a three-node two-hop MIMO relay communication system.

proposes a channel estimation approach different from [23]. Using the PSAM method proposed in [26] as the fundamental, we superimpose a training sequence at the relay node to complete the channel estimation process in one transmission cycle. Note that the superimposed scheme is not discussed in [26], and the system model in [26] is much simpler than the one in this paper. Compared with [23], it is more challenging to derive the optimal training sequences and the optimal amplification factor due to the large number of parameters.

The rest of the paper is organized as follows. The system model of a three-node two-hop MIMO relay system with time-varying channels is introduced in Section II. The superimposed channel training algorithm is proposed in Section III, where the optimal structure of the source and relay training sequences and the optimal relay amplification factor are discussed. Section IV shows numerical simulations to demonstrate the performance of the proposed algorithm. Finally, conclusions are drawn in Section V.

II. SYSTEM MODEL

We consider a three-node two-hop time-varying MIMO relay communication system where the source node transmits information to the destination node through a relay node as shown in Fig. 1. The source, relay, and destination nodes are equipped with N_s , N_r , and N_d antennas, respectively. In this paper, the direct link between the source node and destination node is assumed to be sufficiently weak and thus can be neglected.¹ For notational convenience, we consider a narrow-band single-carrier system. However, our results can be straightforwardly generalized to each subcarrier of a broadband multi-carrier MIMO relay system.

The relay node can either amplify-and-forward (AF), decode-and-forward (DF), or compress-and-forward (CF) relayed signals. In the AF protocol, the received signals are simply amplified (including a possible linear transformation) through the relay precoding matrix before being forwarded to the destination node. On the other hand, in the DF and CF protocols, received signals must be either decoded or compressed, and then re-encoded before being forwarded to the destination node. Therefore, in general, the AF protocol has lower complexity, shorter processing delay, and is easier to implement than the DF and CF protocols. As the optimal relay strategy is not the focus of this paper, the AF protocol is chosen for its simplicity in implementation.

¹When the direct link is strong and cannot be neglected, it can be estimated through single-hop MIMO channel estimation algorithms such as [27] and the references therein. The main focus of this paper is to estimate the individual time-varying channels for MIMO relay communication systems.

We consider that the relay node is operating in the half-duplex mode, i.e., the relay node does not transmit and receive signals at the same time. The choice of half-duplex relay scheme in this paper is motivated by its simpler practical implementation, as the transmit and receive signals at the relay node do not interfere with each other. The change of channels between two time phases in the half-duplex mode does not affect the channel estimation process, and subsequently the retrieval of the source signal. The half-duplex relay scheme has been widely used in MIMO relay communication systems [6].

The pilot symbol aided modulation (PSAM) technique [26] is applied to estimate the time-varying channels, where pilot symbols are inserted among information-carrying symbols in each data frame for channel estimation. Let us denote t_l , $l = 1, \dots, L$, as the time indexes (locations) of the pilot symbols within one data frame of duration T ($T > L$), where L is the length of the pilot sequences. The channel estimation process is completed in two time blocks. In the first time block, pilot symbols $s_m(t_l)$ (along with the information-carrying symbols in each data frame) are transmitted from the m th antenna of the source node, $m = 1, \dots, N_s$, $l = 1, \dots, L$. The received pilot symbols at the n th antenna of the relay node is given by

$$y_{r,n}(t_l) = \sum_{m=1}^{N_s} h_{nm}(t_l) s_m(t_l) + v_{r,n}(t_l),$$

$$n = 1, \dots, N_r, \quad l = 1, \dots, L \quad (1)$$

where $h_{nm}(t_l)$ is the first-hop time-varying channel from the m th antenna at the source node to the n th antenna at the relay node at time t_l , $v_{r,n}(t_l)$ is the noise at the n th antenna of the relay node. In the second time block, the relay node amplifies $y_{r,n}(t_l)$, $n = 1, \dots, N_r$, $l = 1, \dots, L$, and superimposes its own training sequence $r_n(t_l)$ before retransmitting the signal to the destination node. The transmitted signal at the n th antenna of the relay node can be written as

$$x_{r,n}(t_l) = \sqrt{\alpha} y_{r,n}(t_l) + r_n(t_l), \quad n = 1, \dots, N_r, \quad l = 1, \dots, L \quad (2)$$

where $\alpha > 0$ is the relay amplifying factor.

The pilot symbol received at the k th antenna of the destination node is given by

$$y_k(t_l) = \sum_{n=1}^{N_r} g_{kn}(t_l) x_{r,n}(t_l) + v_{d,k}(t_l),$$

$$k = 1, \dots, N_d, \quad l = 1, \dots, L \quad (3)$$

where $g_{kn}(t_l)$ is the second-hop time-varying channel from the n th antenna at the relay node to the k th antenna at the destination node at time t_l , $v_{d,k}(t_l)$ is the noise at the k th antenna of the destination node. All noises are assumed to be

independent and identically distributed (i.i.d.) additive white Gaussian noise (AWGN) with zero mean and unit variance. By substituting (1) and (2) into (3), we have

$$y_k(t_l) = \sqrt{\alpha} \sum_{m=1}^{N_s} \sum_{n=1}^{N_r} g_{kn}(t_l) h_{nm}(t_l) s_m(t_l) + \sum_{n=1}^{N_r} g_{kn}(t_l) r_n(t_l)$$

$$+ \sqrt{\alpha} \sum_{n=1}^{N_r} g_{kn}(t_l) v_{r,n}(t_l) + v_{d,k}(t_l)$$

$$k = 1, \dots, N_d, \quad l = 1, \dots, L. \quad (4)$$

The complex-exponential basis expansion model (CE-BEM) is used to represent the time-varying characteristic of channels as

$$h_{nm}(t) = \sum_{q=0}^Q \mu_{nm}(q) e^{j2\pi t(q - \frac{Q}{2})/T},$$

$$m = 1, \dots, N_s, \quad n = 1, \dots, N_r, \quad t = 1, \dots, T \quad (5)$$

$$g_{kn}(t) = \sum_{q=0}^Q \lambda_{kn}(q) e^{j2\pi t(q - \frac{Q}{2})/T},$$

$$n = 1, \dots, N_r, \quad k = 1, \dots, N_d, \quad t = 1, \dots, T \quad (6)$$

where $j = \sqrt{-1}$, $\mu_{nm}(q)$ and $\lambda_{kn}(q)$ are the BEM coefficients that do not change within the duration of one data frame T , and Q is the number of bases. The complex exponential terms in (5) and (6) are the Fourier bases that describe the time-varying characteristic of channels. Note that the value of Q depends on T and the channel bandwidth f , and should be at least $2\lceil fT \rceil$ [20], where $\lceil a \rceil$ denotes the smallest integer greater than a . In this paper, we assume that all channel links have the same number of bases Q , and the generalization to the case where each channel link has a different number of bases is straightforward. We also assume that the channels $h_{nm}(t)$ and $g_{kn}(t)$ are wide-sense stationary (WSS) zero mean complex Gaussian (ZMCG) random process with constant variances, $\mu_{nm}(q)$ and $\lambda_{kn}(q)$ are random variables with zero mean and independent with each other.

The idea of the proposed superimposed channel training algorithm is to estimate the second-hop channels $\{g_{kn}(t)\} \triangleq \{g_{kn}(t), k = 1, \dots, N_d, n = 1, \dots, N_r, t = 1, \dots, T\}$ through estimating $\{\lambda_{kn}(q)\} \triangleq \{\lambda_{kn}(q), k = 1, \dots, N_d, n = 1, \dots, N_r, q = 1, \dots, Q\}$ using the superimposed relay training sequence $\{r_n(t_l)\} \triangleq \{r_n(t_l), n = 1, \dots, N_r, l = 1, \dots, L\}$. Then, the first-hop channels $\{h_{nm}(t)\} \triangleq \{h_{nm}(t), n = 1, \dots, N_r, m = 1, \dots, N_s, t = 1, \dots, T\}$ are estimated through estimating $\{\mu_{nm}(q)\} \triangleq \{\mu_{nm}(q), n = 1, \dots, N_r, m = 1, \dots, N_s, q = 1, \dots, Q\}$ using the source training sequence $\{s_m(t_l)\} \triangleq \{s_m(t_l), m = 1, \dots, N_s, l = 1, \dots, L\}$ and the estimated $\{\lambda_{kn}(q)\}$. An advantage of the proposed superimposed channel training algorithm is that the channel estimation process is completed in one transmission cycle.

III. MMSE-BASED OPTIMAL TRAINING MATRICES

In this section, the optimal training sequences $\{s_m(t_l)\}$ and $\{r_n(t_l)\}$, and the relay amplifying factor α that minimize the MSE of channel estimation are derived. From (5) and (6), we have

$$\begin{aligned} g_{kn}(t_l)h_{nm}(t_l) &= \sum_{p=0}^Q \sum_{q=0}^Q \lambda_{kn}(p)\mu_{nm}(q)e^{j2\pi t_l(p+q-Q)/T} \\ &= \sum_{i=0}^{2Q} x_{knm}(i)e^{j\theta_i t_l} \end{aligned} \quad (7)$$

where $m = 1, \dots, N_s$, $n = 1, \dots, N_r$, $k = 1, \dots, N_d$, $l = 1, \dots, L$, $\theta_i \triangleq 2\pi(i - Q)/T$ and

$$x_{knm}(i) \triangleq \sum_{p=0}^i \lambda_{kn}(p)\mu_{nm}(i - p), \quad i = 0, \dots, 2Q. \quad (8)$$

By substituting (6) and (7) back into (4), we obtain

$$\begin{aligned} y_k(t_l) &= \sqrt{\alpha} \sum_{m=1}^{N_s} \sum_{n=1}^{N_r} \sum_{i=0}^{2Q} x_{knm}(i)e^{j\theta_i t_l} s_m(t_l) \\ &\quad + \sum_{n=1}^{N_r} \sum_{q=0}^Q \lambda_{kn}(q)e^{j\phi_q t_l} r_n(t_l) + \bar{v}_k(t_l), \\ k &= 1, \dots, N_d, \quad l = 1, \dots, L \end{aligned} \quad (9)$$

where $\phi_q \triangleq 2\pi(q - Q/2)/T$ and

$$\begin{aligned} \bar{v}_k(t_l) &\triangleq \sqrt{\alpha} \sum_{n=1}^{N_r} g_{kn}(t_l)v_{r,n}(t_l) + v_{d,k}(t_l), \\ k &= 1, \dots, N_d, \quad l = 1, \dots, L \end{aligned} \quad (10)$$

is the equivalent noise at the k th antenna of the destination node.

Let us denote $\mathbf{y}_k \triangleq [y_k(t_1), y_k(t_2), \dots, y_k(t_L)]^T$, $\mathbf{S}_m \triangleq \text{diag}[s_m(t_1), s_m(t_2), \dots, s_m(t_L)]$, $\mathbf{R}_n \triangleq \text{diag}[r_n(t_1), r_n(t_2), \dots, r_n(t_L)]$, $\bar{\mathbf{v}}_k \triangleq [\bar{v}_k(t_1), \bar{v}_k(t_2), \dots, \bar{v}_k(t_L)]^T$, and

$$\boldsymbol{\lambda}_{kn} \triangleq [\lambda_{kn}(0), \lambda_{kn}(1), \dots, \lambda_{kn}(Q)]^T, \quad (11)$$

$$\mathbf{x}_{km} \triangleq \left[\sum_{n=1}^{N_r} x_{knm}(0), \sum_{n=1}^{N_r} x_{knm}(1), \dots, \sum_{n=1}^{N_r} x_{knm}(2Q) \right]^T, \quad (12)$$

$$\boldsymbol{\Theta} \triangleq \begin{bmatrix} e^{j\theta_0 t_1} & \dots & e^{j\theta_{2Q} t_1} \\ \vdots & \ddots & \vdots \\ e^{j\theta_0 t_L} & \dots & e^{j\theta_{2Q} t_L} \end{bmatrix}, \quad \boldsymbol{\Phi} \triangleq \begin{bmatrix} e^{j\phi_0 t_1} & \dots & e^{j\phi_{2Q} t_1} \\ \vdots & \ddots & \vdots \\ e^{j\phi_0 t_L} & \dots & e^{j\phi_{2Q} t_L} \end{bmatrix} \quad (13)$$

where $(\cdot)^T$ denotes the vector (matrix) transpose and $\text{diag}[\cdot]$ stands for a diagonal matrix. Then we can equivalently rewrite

(9) in matrix vector form as

$$\mathbf{y}_k = \sqrt{\alpha} \sum_{m=1}^{N_s} \mathbf{S}_m \boldsymbol{\Theta} \mathbf{x}_{km} + \sum_{n=1}^{N_r} \mathbf{R}_n \boldsymbol{\Phi} \boldsymbol{\lambda}_{kn} + \bar{\mathbf{v}}_k \quad (14)$$

$$= \mathbf{A} \mathbf{b}_k + \bar{\mathbf{v}}_k, \quad k = 1, \dots, N_d \quad (15)$$

where

$$\begin{aligned} \mathbf{A} &\triangleq [\sqrt{\alpha} \mathbf{S}_1 \boldsymbol{\Theta}, \dots, \sqrt{\alpha} \mathbf{S}_{N_s} \boldsymbol{\Theta}, \mathbf{R}_1 \boldsymbol{\Phi}, \dots, \mathbf{R}_{N_r} \boldsymbol{\Phi}] \\ &\in \mathcal{C}^{L \times [N_s(2Q+1) + N_r(Q+1)]} \end{aligned} \quad (16)$$

$$\mathbf{b}_k \triangleq [\mathbf{x}_{k1}^T, \dots, \mathbf{x}_{kN_s}^T, \boldsymbol{\lambda}_{k1}^T, \dots, \boldsymbol{\lambda}_{kN_r}^T]^T. \quad (17)$$

Here \mathbf{b}_k in (17) is the vector of unknowns containing the BEM coefficients that need to be estimated at the destination node.

We apply a linear estimator at the destination node to estimate \mathbf{b}_k due to its simplicity. The estimated \mathbf{b}_k is given by

$$\hat{\mathbf{b}}_k = \mathbf{W}_k^H \mathbf{y}_k, \quad k = 1, \dots, N_d \quad (18)$$

where \mathbf{W}_k is the weight matrix of the linear receiver and $(\cdot)^H$ denotes the matrix (vector) Hermitian transpose. From (16), we have $L \geq N_s(2Q+1) + N_r(Q+1)$ since a linear estimator is used. From (15) and (18), the MSE of estimating \mathbf{b}_k , $k = 1, \dots, N_d$, is given by

$$\begin{aligned} \text{MSE} &= \sum_{k=1}^{N_d} \text{tr} \left(\mathbb{E} \left[(\hat{\mathbf{b}}_k - \mathbf{b}_k)(\hat{\mathbf{b}}_k - \mathbf{b}_k)^H \right] \right) \\ &= \sum_{k=1}^{N_d} \text{tr} \left(\mathbb{E} \left[(\mathbf{W}_k^H \mathbf{A} - \mathbf{I}_D) \mathbf{C}_{b,k} (\mathbf{W}_k^H \mathbf{A} - \mathbf{I}_D)^H \right. \right. \\ &\quad \left. \left. + \mathbf{W}_k^H \mathbf{C}_{\bar{v},k} \mathbf{W}_k \right] \right) \end{aligned} \quad (19)$$

where $\text{tr}(\cdot)$ denotes the matrix trace, $\mathbb{E}[\cdot]$ stands for the statistical expectation, \mathbf{I}_n represents the $n \times n$ identity matrix, $D \triangleq N_s(2Q+1) + N_r(Q+1)$, $\mathbf{C}_{b,k} \triangleq \mathbb{E}[\mathbf{b}_k \mathbf{b}_k^H]$ is the covariance matrix of \mathbf{b}_k , and $\mathbf{C}_{\bar{v},k} \triangleq \mathbb{E}[\bar{\mathbf{v}}_k \bar{\mathbf{v}}_k^H]$ is the noise covariance matrix.

To find $\mathbf{C}_{b,k}$, we first compute the covariance matrix of \mathbf{x}_{km} using (8) and (12) as

$$\begin{aligned} \mathbf{C}_{b,k}^{x,m} &= \mathbb{E}[\mathbf{x}_{km} \mathbf{x}_{km}^H] \\ &= \mathbb{E} \left[\left(\sum_{n=1}^{N_r} \boldsymbol{\lambda}_{kn} * \boldsymbol{\mu}_{nm} \right) \left(\sum_{n=1}^{N_r} \boldsymbol{\lambda}_{kn} * \boldsymbol{\mu}_{nm} \right)^H \right] \\ &= \sum_{n=1}^{N_r} \text{diag}[\boldsymbol{\sigma}_{k,n}^\lambda * \boldsymbol{\sigma}_{n,m}^\mu], \quad m = 1, \dots, N_s \end{aligned} \quad (20)$$

where $\mathbf{a} * \mathbf{b}$ denotes the linear convolution between vectors \mathbf{a} and \mathbf{b} , $\boldsymbol{\sigma}_{k,n}^\lambda = [\sigma_{k,n}^{\lambda,0}, \dots, \sigma_{k,n}^{\lambda,Q}]^T$, $\boldsymbol{\sigma}_{n,m}^\mu = [\sigma_{n,m}^{\mu,0}, \dots, \sigma_{n,m}^{\mu,Q}]^T$, and $\sigma_{k,n}^{\lambda,q} = \mathbb{E}[\lambda_{kn}(q)\lambda_{kn}^*(q)]$ and $\sigma_{n,m}^{\mu,q} = \mathbb{E}[\mu_{nm}(q)\mu_{nm}^*(q)]$ are

the variances of $\lambda_{kn}(q)$ and $\mu_{nm}(q)$, respectively. Here, $(\cdot)^*$ denotes complex conjugate. Similarly, from (11), the covariance matrix of λ_{kn} can be written as

$$\begin{aligned} \mathbf{C}_{b,k}^{\lambda,n} &= \mathbb{E}[\lambda_{kn}\lambda_{kn}^H] \\ &= \text{diag}[\sigma_{k,n}^{\lambda,0}, \dots, \sigma_{k,n}^{\lambda,Q}], \quad n = 1, \dots, N_r. \end{aligned} \quad (21)$$

Based on (20) and (21), the covariance matrix $\mathbf{C}_{b,k}$ can be rewritten as

$$\mathbf{C}_{b,k} = \text{bd}[\mathbf{C}_{b,k}^x, \mathbf{C}_{b,k}^\lambda], \quad k = 1, \dots, N_d$$

where $\text{bd}[\cdot]$ represents a block diagonal matrix and

$$\begin{aligned} \mathbf{C}_{b,k}^x &= \text{bd}[\mathbf{C}_{b,k}^{x,1}, \dots, \mathbf{C}_{b,k}^{x,N_s}], \\ \mathbf{C}_{b,k}^\lambda &= \text{bd}[\mathbf{C}_{b,k}^{\lambda,1}, \dots, \mathbf{C}_{b,k}^{\lambda,N_r}]. \end{aligned} \quad (22)$$

Since all noises are assumed to be i.i.d. with zero mean and unit variance, from (10) we have

$$\mathbf{C}_{\bar{v},k} = \left(\alpha \sum_{n=1}^{N_r} \sigma_{k,n}^g + 1 \right) \mathbf{I}_L, \quad k = 1, \dots, N_d$$

where $\sigma_{k,n}^g = \mathbb{E}[g_{k,n}(t_l)g_{k,n}^*(t_l)]$ is the variance of $g_{k,n}(t_l)$, $l = 1, \dots, L$.

The optimal weight matrices \mathbf{W}_k , $k = 1, \dots, N_d$ that minimize the MSE in (19) are given by

$$\mathbf{W}_k = (\mathbf{A}\mathbf{C}_{b,k}\mathbf{A}^H + \mathbf{C}_{\bar{v},k})^{-1} \mathbf{A}\mathbf{C}_{b,k}, \quad k = 1, \dots, N_d \quad (23)$$

where $(\cdot)^{-1}$ stands for matrix inversion. Substituting (23) back into (19), the MSE of channel estimation is given by

$$\text{MSE} = \sum_{k=1}^{N_d} \text{tr} \left(\left[\mathbf{C}_{b,k}^{-1} + \mathbf{A}^H \mathbf{C}_{\bar{v},k}^{-1} \mathbf{A} \right]^{-1} \right). \quad (24)$$

The transmission power constraint at the source node can be written as

$$\sum_{m=1}^{N_s} \mathbf{s}_m^H \mathbf{s}_m \leq P_s \quad (25)$$

where $\mathbf{s}_m \triangleq [s_m(t_1), s_m(t_2), \dots, s_m(t_L)]^T$ and P_s is the transmission power available at the source node. The transmission power constraint at the relay node can be calculated from (2) as

$$\begin{aligned} &\mathbb{E} \left[\sum_{n=1}^{N_r} \sum_{l=1}^L |x_{r,n}(t_l)|^2 \right] \\ &= \alpha \sum_{m=1}^{N_s} \sum_{n=1}^{N_r} \sigma_{n,m}^h \mathbf{s}_m^H \mathbf{s}_m + \sum_{n=1}^{N_r} \mathbf{r}_n^H \mathbf{r}_n + \alpha N_r L \leq P_r \end{aligned} \quad (26)$$

where $\sigma_{n,m}^h = \mathbb{E}[h_{n,m}(t_l)h_{n,m}^*(t_l)]$ is the variance of $h_{n,m}(t_l)$, $l = 1, \dots, L$, P_r is the transmission power available at the relay node, and $\mathbf{r}_n \triangleq [r_n(t_1), \dots, r_n(t_L)]^T$.

Based on (24)–(25), the optimal pilot sequences and the optimal α design problem can be written as

$$\min_{\{\mathbf{s}_m\}, \{\mathbf{r}_n\}, \alpha > 0} \sum_{k=1}^{N_d} \text{tr} \left(\left[\mathbf{C}_{b,k}^{-1} + \mathbf{A}^H \mathbf{C}_{\bar{v},k}^{-1} \mathbf{A} \right]^{-1} \right) \quad (27)$$

$$\text{s.t.} \quad \sum_{m=1}^{N_s} \mathbf{s}_m^H \mathbf{s}_m \leq P_s \quad (28)$$

$$\alpha \sum_{m=1}^{N_s} \sum_{n=1}^{N_r} \sigma_{n,m}^h \mathbf{s}_m^H \mathbf{s}_m + \sum_{n=1}^{N_r} \mathbf{r}_n^H \mathbf{r}_n + \alpha N_r L \leq P_r \quad (29)$$

where $\{\mathbf{s}_m\} \triangleq \{\mathbf{s}_m, m = 1, \dots, N_s\}$ and $\{\mathbf{r}_n\} \triangleq \{\mathbf{r}_n, n = 1, \dots, N_r\}$. The optimal structures of $\{\mathbf{s}_m\}$ and $\{\mathbf{r}_n\}$ as the solution to the optimization problem (27)–(29) are derived in the following theorem.

Theorem 1: The optimal training matrices $\{\mathbf{s}_m\}$ and $\{\mathbf{r}_n\}$ satisfy the following equations for all $m, p = 1, \dots, N_s$ and $n, q = 1, \dots, N_r$

$$(\mathbf{S}_m \mathbf{\Theta})^H \mathbf{S}_m \mathbf{\Theta} = \beta_m \mathbf{I}_{2Q+1} \quad (30)$$

$$(\mathbf{R}_n \mathbf{\Phi})^H \mathbf{R}_n \mathbf{\Phi} = \gamma_n \mathbf{I}_{Q+1} \quad (31)$$

$$(\mathbf{S}_m \mathbf{\Theta})^H \mathbf{S}_p \mathbf{\Theta} = \mathbf{0}, \quad m \neq p \quad (32)$$

$$(\mathbf{R}_n \mathbf{\Phi})^H \mathbf{R}_q \mathbf{\Phi} = \mathbf{0}, \quad n \neq q \quad (33)$$

$$(\mathbf{S}_m \mathbf{\Theta})^H \mathbf{R}_n \mathbf{\Phi} = \mathbf{0} \quad (34)$$

where $\beta_m = \mathbf{s}_m^H \mathbf{s}_m$, $m = 1, \dots, N_s$, and $\gamma_n = \mathbf{r}_n^H \mathbf{r}_n$, $n = 1, \dots, N_r$.

Proof: See Appendix A. \square

One example of achieving (30)–(34) is given as follows. By exploiting the Vandermonde structure of $\mathbf{\Theta}$ and $\mathbf{\Phi}$, and the structure of θ_q and ϕ_q , the conditions (30) and (31) are satisfied by

$$\begin{aligned} t_i &= iT/L, \quad |s_{m,i}| = \sqrt{\beta_m/L}, \quad |r_{n,i}| = \sqrt{\gamma_n/L} \\ i &= 1, \dots, L, \quad m = 1, \dots, N_s, \quad n = 1, \dots, N_r \end{aligned} \quad (35)$$

where $s_{m,i} \triangleq s_m(iT/L)$ and $r_{n,i} \triangleq r_n(iT/L)$. Interestingly, (35) indicates that it is optimal in terms of the MSE of the channel estimation to place pilot symbols in each data frame with equal intervals. Moreover, with a given power allocation β_m at the m th antennas of the source node and γ_n at the n th antenna of the relay node, it is optimal to equally allocate the power to the pilot symbols. In the following, we solve (32)–(34) based on $t_i = iT/L$, $i = 1, \dots, L$.

To satisfy the condition (32), there should be

$$\sum_{l=1}^L s_{m,l}^* e^{-j2\pi l(p-Q)/L} s_{n,l} e^{j2\pi l(q-Q)/L} = 0,$$

$$p, q = 0, \dots, 2Q, \quad m, n = 1, \dots, N_s, m \neq n$$

which is equivalent to

$$\sum_{l=1}^L s_{m,l}^* s_{n,l} e^{j2\pi kl/L} = 0,$$

$$k = -2Q, \dots, 2Q, \quad m, n = 1, \dots, N_s, m \neq n. \quad (36)$$

Similarly, to satisfy (33) and (34), we need to have

$$\sum_{l=1}^L r_{m,l}^* r_{n,l} e^{j2\pi pl/L} = 0,$$

$$p = -Q, \dots, Q, \quad m, n = 1, \dots, N_r, m \neq n \quad (37)$$

$$\sum_{l=1}^L s_{m,l}^* r_{n,l} e^{j2\pi ql/L} = 0,$$

$$q = -3Q/2, \dots, 3Q/2, \quad m = 1, \dots, N_s, n = 1, \dots, N_r. \quad (38)$$

One way to satisfy the conditions (35)–(38) is

$$r_{n,l} = \sqrt{\gamma_n/L} e^{j2\pi l(n-1)(Q+1)/L},$$

$$n = 1, \dots, N_r, \quad l = 1, \dots, L \quad (39)$$

$$s_{m,l} = \sqrt{\beta_m/L} e^{j2\pi l(N_r(Q+1)+Q/2+(m-1)(2Q+1))/L},$$

$$m = 1, \dots, N_s, \quad l = 1, \dots, L. \quad (40)$$

The training symbols shown in (39) and (40) have the advantage that they are easy to implement as the elements of \mathbf{s}_m (also \mathbf{r}_n) have a constant magnitude.

Using Theorem 1 and (53), the MSE function in (24) can be written as

$$\begin{aligned} \text{MSE} &= \sum_{k=1}^{N_d} \text{tr} \left(\sum_{m=1}^{N_s} \left[(\mathbf{C}_{b,k}^{x,m})^{-1} + \alpha \beta_m \eta_k \mathbf{I}_{2Q+1} \right]^{-1} \right. \\ &\quad \left. + \sum_{n=1}^{N_r} \left[(\mathbf{C}_{b,k}^{\lambda,n})^{-1} + \gamma_n \eta_k \mathbf{I}_{Q+1} \right]^{-1} \right) \\ &= \sum_{m=1}^{N_s} \sum_{k=1}^{N_d} \sum_{q=1}^{2Q+1} \frac{1}{c_{k,m,q} + \alpha \beta_m \eta_k} \\ &\quad + \sum_{n=1}^{N_r} \sum_{k=1}^{N_d} \sum_{p=1}^{Q+1} \frac{1}{d_{k,n,p} + \gamma_n \eta_k} \end{aligned} \quad (41)$$

where η_k is defined in (52), $c_{k,m,q} \triangleq [(\mathbf{C}_{b,k}^{x,m})^{-1}]_{q,q}$, $d_{k,n,p} \triangleq [(\mathbf{C}_{b,k}^{\lambda,n})^{-1}]_{p,p}$ and $[\cdot]_{j,j}$ denotes the j th diagonal element of a matrix. From (41), the problem (27)–(29) can be equivalently

rewritten as the following optimization problem with scalar variables

$$\begin{aligned} \min_{\{\beta_m\}, \{\gamma_n\}, \alpha} & \sum_{m=1}^{N_s} \sum_{k=1}^{N_d} \sum_{q=1}^{2Q+1} \frac{1}{c_{k,m,q} + \alpha \beta_m \eta_k} \\ & + \sum_{n=1}^{N_r} \sum_{k=1}^{N_d} \sum_{p=1}^{Q+1} \frac{1}{d_{k,n,p} + \gamma_n \eta_k} \end{aligned} \quad (42)$$

$$\text{s.t.} \quad \sum_{m=1}^{N_s} \beta_m \leq P_s \quad (43)$$

$$\alpha \sum_{m=1}^{N_s} \kappa_m \beta_m + \sum_{n=1}^{N_r} \gamma_n + \alpha N_r L \leq P_r \quad (44)$$

$$\alpha > 0, \quad \beta_m \geq 0, \quad m = 1, \dots, N_s,$$

$$\gamma_n \geq 0, \quad n = 1, \dots, N_r \quad (45)$$

where $\{\beta_m\} \triangleq \{\beta_m, m=1, \dots, N_s\}$, $\{\gamma_n\} \triangleq \{\gamma_n, n=1, \dots, N_r\}$, and $\kappa_m \triangleq \sum_{n=1}^{N_r} \sigma_{n,m}^h$.

As $c_{k,m,q}$, $d_{k,n,p}$, and η_k are known variables, when α is fixed, it can be seen that the first and second set of fractions in the objective function (42) are monotonically decreasing and convex with respect to β_m and γ_n , respectively. Moreover, the constraints in (43)–(45) are linear inequality constraints when α is fixed. We can conclude that the problem (42)–(45) is a convex optimization problem with respect to β_m and γ_n when α is fixed. The optimal β_m and γ_n can be efficiently derived using the Karush-Kuhn-Tucker (KKT) optimality conditions [28] of the problem (42)–(45), where the gradient conditions are given by

$$\sum_{k=1}^{N_d} \sum_{q=1}^{2Q+1} \frac{\alpha \eta_k}{(c_{k,m,q} + \alpha \beta_m \eta_k)^2} = \mu_1 + \mu_2 \alpha \kappa_m, \quad m = 1, \dots, N_s \quad (46)$$

$$\sum_{k=1}^{N_d} \sum_{p=1}^{Q+1} \frac{\eta_k}{(d_{k,n,p} + \gamma_n \eta_k)^2} = \mu_2, \quad n = 1, \dots, N_r \quad (47)$$

where $\mu_1 \geq 0$ and $\mu_2 \geq 0$ are Lagrange multipliers such that the complementary slackness conditions given by

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

$$\mu_2 \left(P_r - \alpha \sum_{m=1}^{N_s} \kappa_m \beta_m - \sum_{n=1}^{N_r} \gamma_n - \alpha N_r L \right) = 0$$

are satisfied.

For given α , μ_1 , and μ_2 , the non-negative $\{\beta_m\}$ and $\{\gamma_n\}$ can be found by applying the bi-section search, since the left-hand side (LHS) of (46) and (47) are monotonically decreasing functions of β_m and γ_n , respectively. To find the optimal μ_1 and μ_2 , we apply an outer bi-section search as the LHS of (43) is an

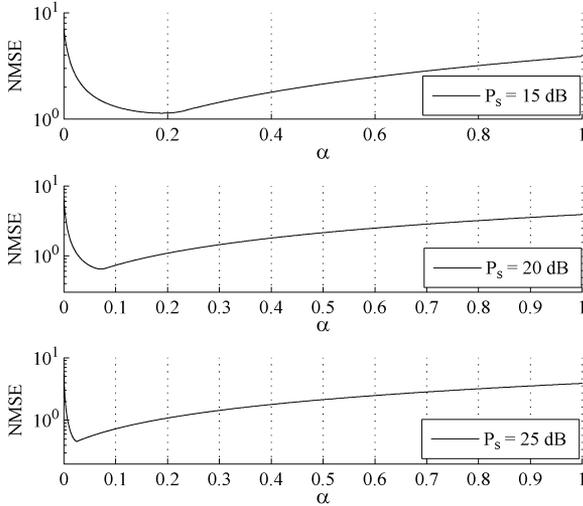


Fig. 2. NMSE versus α for different P_s and $P_r=20$ dB with $N=2$ and $Q=4$.

increasing function of β_m , and the LHS of (44) is an increasing function of β_m and γ_n , while in (46), β_m is a monotonically decreasing function of μ_1 and μ_2 , and γ_n is a monotonically decreasing function of μ_2 in (47).

When α is not fixed and needs to be optimized, the problem (42)–(45) as a whole is not a convex optimization problem. However, it is proven in the following theorem that (42) is a unimodal function of α .

Theorem 2: The objective function (42) with the constraints (43)–(45) is a unimodal (quasi-convex) function with respect to α .

Proof: See Appendix B. \square

A plot of the MSE value over a range of feasible value of α is shown in Fig. 2 to validate Theorem 2. We consider the case where all the nodes have the same number of antennas, i.e., $N_s = N_r = N_d = N = 2$, and $Q = 4$. We obtain the normalized MSE (NMSE) by dividing the objective function (42) with $L = N_s(2Q + 1) + N_r(Q + 1)$. Fig. 2 shows the NMSE versus α for different P_s and $P_r = 20$ dB. It is obvious from Fig. 2 that the objective function (42) is a unimodal function with respect to α . Moreover, the optimal value of α varies with P_s , where the optimal α has a smaller value when P_s increases. The minimum value of a unimodal function can be efficiently found by using the golden section search (GSS) method [29]. Thus, we can apply the GSS method described in Table I to obtain the optimal α for the problem (42)–(45), where $\phi > 0$ is the reduction factor and ε is a positive constant close to zero. It is shown in [29] that the optimal $\phi = 1.618$, which is also known as the golden ratio.

Once the estimated \mathbf{x}_{km} and λ_{kn} are obtained from (14), we can estimate μ_{nm} as follows. From (8) we have

$$\begin{aligned} \mathbf{x}_{km} &= \sum_{n=1}^{N_r} \lambda_{kn} * \mu_{nm} \\ &= \sum_{n=1}^{N_r} \mathbf{C}(\lambda_{kn}) \mu_{nm}, \quad k=1, \dots, N_d, \quad m=1, \dots, N_s \end{aligned} \quad (48)$$

TABLE I
PROCEDURE OF APPLYING THE GOLDEN SECTION SEARCH (GSS)
TO FIND THE OPTIMAL α IN THE PROBLEM (42)–(45)

- 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 (42)–(45) for $\alpha = c_1$;
Compute the MSE value defined in (42), $f_{\text{MSE}}(c_1)$ for $\alpha = c_1$.
- 4) Repeat Step 3 for $\alpha = c_2$.
- 5) If $f_{\text{MSE}}(c_1) < f_{\text{MSE}}(c_2)$, then assign $b = c_2$.
Otherwise, assign $a = c_1$.
- 6) If $|b - a| \leq \varepsilon$, then end.
Otherwise, go to step 2.

where $\mathbf{C}(\mathbf{a})$ is a $(2Q + 1) \times (Q + 1)$ column-wise circulant matrix taking $[\mathbf{a}^T, \mathbf{0}_{1 \times Q}]^T$ as its first column. We can rewrite (48) in matrix vector form as

$$\mathbf{Y} \mu_m = \mathbf{x}_m, \quad m = 1, \dots, N_s \quad (49)$$

where

$$\mathbf{Y} \triangleq \begin{bmatrix} \mathbf{C}(\lambda_{11}) & \cdots & \mathbf{C}(\lambda_{1N_r}) \\ \vdots & \ddots & \vdots \\ \mathbf{C}(\lambda_{N_d 1}) & \cdots & \mathbf{C}(\lambda_{N_d N_r}) \end{bmatrix}, \quad \mu_m \triangleq \begin{bmatrix} \mu_{1m} \\ \vdots \\ \mu_{N_r m} \end{bmatrix},$$

$$\mathbf{x}_m \triangleq \begin{bmatrix} \mathbf{x}_{1m} \\ \vdots \\ \mathbf{x}_{N_d m} \end{bmatrix}.$$

Then, from (49), the estimated μ_{nm} is given by

$$\hat{\mu}_m = \hat{\mathbf{Y}}^\dagger \hat{\mathbf{x}}_m, \quad m = 1, \dots, N_s \quad (50)$$

where $(\cdot)^\dagger$ denotes the matrix pseudo-inverse. Since \mathbf{Y} is an $N_d(2Q + 1) \times N_r(Q + 1)$ matrix, the estimation (50) holds if $N_d(2Q + 1) \geq N_r(Q + 1)$.

Once the estimated $\{\mu_{nm}(q)\}$ and $\{\lambda_{kn}(q)\}$ are obtained, an estimation of the time-varying channels $\{h_{nm}(t)\}$ and $\{g_{kn}(t)\}$ can be obtained using (5) and (6), respectively.

IV. NUMERICAL EXAMPLES

In this section, we study the performance of the proposed superimposed channel training algorithm for time-varying MIMO relay systems through numerical simulations. We consider a three-node two-hop MIMO relay system where the source, relay, and destination node are equipped with N_s, N_r , and N_d antennas, respectively. For simplicity, it is assumed that all BEM coefficients have unit variance. We use the shortest training sequence possible, i.e., $L = N_s(2Q + 1) + N_r(Q + 1)$. For all scenarios, we compute the NMSE of channel estimation at the destination node.

Unless explicitly mentioned, we assume that all nodes are equipped with the same number of antennas, i.e., $N_s = N_r = N_d = N$, and the source and relay nodes have the same transmission power, i.e., $P_s = P_r = P$. Then the shortest training sequence possible is given by $L = N(3Q + 2)$. In the first example, we study the impact of α on the performance of the proposed superimposed channel training algorithm. Fig. 3 shows the NMSE of the proposed algorithm versus P for different α

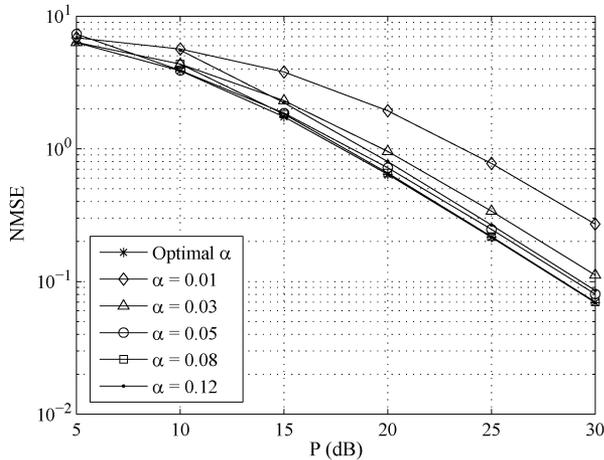


Fig. 3. Example 1: NMSE versus P for different α with $N = 2$ and $Q = 4$.

when $N = 2$ and $Q = 4$. We apply the GSS technique in the superimposed channel training algorithm to obtain the optimal α for different P , and thus the optimal α curve in Fig. 3. It can be observed from Fig. 3 that the optimal α curve always has the lowest MSE value for all P , and hence proves the efficiency of the GSS method in obtaining the optimal α for different P . Moreover, it can be seen from Fig. 3 that the optimal α at low P level is smaller compared with the optimal α at high P level. This indicates that the use of a constant α is strictly suboptimal. Note that the starting point for the curves associated with $\alpha = 0.08$ and $\alpha = 0.12$ is at $P = 10$ dB, as these values of α have exceeded the upper bound limit of the α for $P = 5$ dB. The optimal α is used for the following numerical simulations of the superimposed channel training algorithm.

In the following simulation examples, we compare the performance of the proposed algorithm with the conventional two-stage MMSE-based channel estimation algorithm. At the first stage of the two-stage channel estimation algorithm, a training sequence is transmitted from the relay node to estimate the relay-destination channel matrix, while at the second stage, the source-relay channel matrix is estimated by using the training sequence sent from the source node [15]. To the best of our knowledge, no other channel estimation algorithm has been proposed to estimate the individual time-varying channel matrices of the source-relay and relay-destination links in MIMO relay systems. The conventional two-stage channel estimation algorithm is chosen as a benchmark due to its simple implementation. We would like to note that to make a fair comparison, we have extended the two-stage channel estimation algorithm in [15] from block fading MIMO relay channel to time-varying MIMO relay channel.

In the second example, we investigate the performance of the two algorithms at various number of antennas. Fig. 4 shows the NMSE performance of both algorithms versus P for different N when $Q = 4$. As expected, the NMSE of the channel estimation increases when the number of antenna increases due to a larger number of unknowns to be estimated. It can also be observed from Fig. 4 that the proposed algorithm consistently has better NMSE performance compared with the conventional two-stage channel estimation algorithm, especially at high P level.

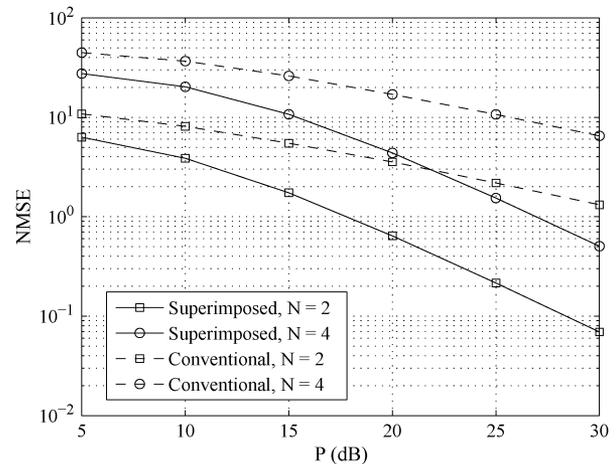


Fig. 4. Example 2: NMSE versus P for different N with $Q = 4$.

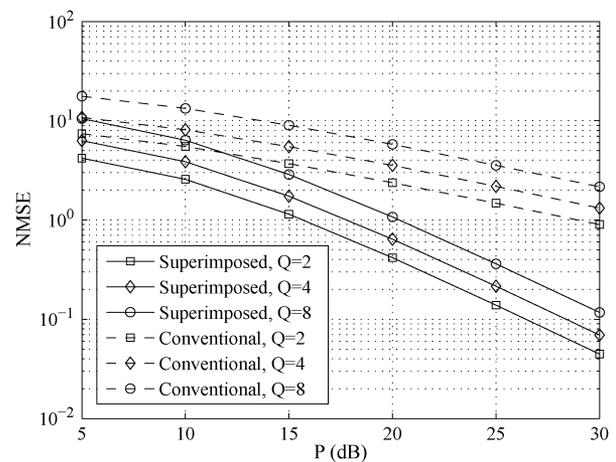


Fig. 5. Example 3: NMSE versus P for different Q and $N = 2$.

In the third example, we study the effect of the number of bases Q on the performance of the proposed superimposed channel training algorithm and the conventional two-stage channel estimation algorithm. Fig. 5 shows the NMSE of both algorithms versus P for different Q when $N = 2$. It can be seen that when Q increases, the NMSE of both algorithms also increases, with our proposed algorithm performing better than the conventional two-stage channel estimation algorithm. This is because with larger Q , there are more BEM coefficients to be estimated. The selection of Q is beyond the scope of this paper. However, it has been stated in [23] that the time-varying channels are more accurately represented by the CE-BEM when a larger Q is used.

In the fourth example, we consider that the power available for transmission at the source node and relay node is different. Fig. 6 demonstrates the NMSE performance of the proposed algorithm and the conventional two-stage method versus P_s for different fixed P_r when $N = 4$ and $Q = 4$. As expected, it can be seen from Fig. 6 that the NMSE performance of both algorithms improves when the power at the source and/or relay node increases, and our proposed algorithm outperforms the conventional two-stage method for all scenarios.

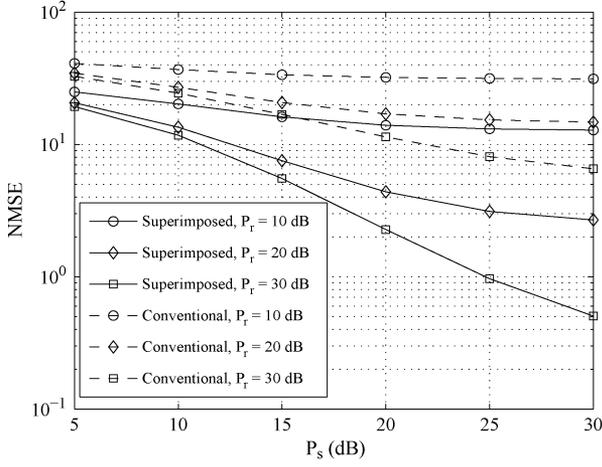


Fig. 6. Example 4: NMSE versus P_s for different P_r with $N = 4$ and $Q = 4$.

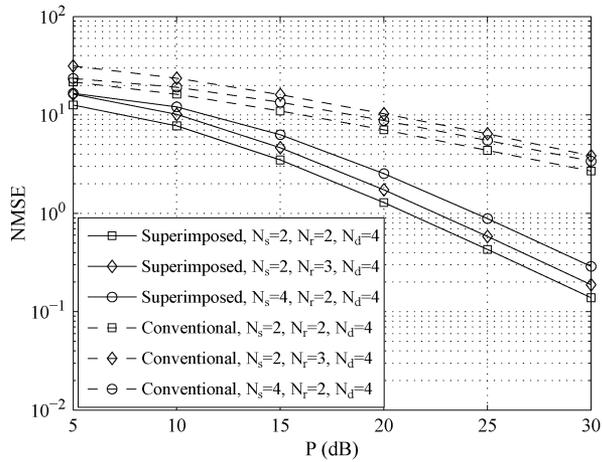


Fig. 7. Example 5: NMSE versus P for different number of antennas at each node and $Q = 4$.

In the fifth example, we study the performance of the proposed superimposed channel training algorithm and the conventional two-stage channel estimation algorithm when the nodes in the relay system have different number of antennas. Fig. 7 shows the NMSE performance of both algorithms versus P for different N_s , N_r , and N_d when $Q = 4$. It can be observed from Fig. 7 that the proposed algorithm consistently has a better NMSE performance than the conventional two-stage channel estimation algorithm for various combinations of N_s , N_r , and N_d .

Finally, we compare the complexity of the proposed superimposed channel training algorithm with the conventional two-stage channel estimation algorithm. The complexity of the proposed superimposed channel training algorithm can be estimated as $\mathcal{O}(c_\alpha c_\mu c_\beta \gamma N_d (N_s (2Q + 1) + N_r (Q + 1)))$, where c_α is the number of GSS iterations required to obtain the optimal α , c_μ represents the number of loops in the outer bi-section search to obtain the optimal μ_1 and μ_2 , and $c_\beta \gamma$ stands for the number of bi-section operations required to obtain the optimal β_m and γ_n . The complexity of the conventional two-stage channel estimation algorithm can be estimated as

$\mathcal{O}(N_d (N_s (2Q + 1) + N_r (Q + 1)))$. It can be clearly seen that the proposed algorithm has a similar computational complexity order as the conventional algorithm. However, the performance of the proposed algorithm is much better than the conventional two-stage channel estimation algorithm.

V. CONCLUSION

We have proposed a superimposed channel training algorithm to estimate the time-varying channels in MIMO relay communication systems. The proposed algorithm takes into account the time-varying characteristic of the channels and can efficiently estimate the individual CSI for MIMO relay communication systems. The optimal structure of the source and relay training sequences and the optimal power allocation between the source and relay training sequences are derived.

APPENDIX A PROOF OF THEOREM 1

Let us define $\Psi_S \triangleq [\mathbf{S}_1 \Theta, \mathbf{S}_2 \Theta, \dots, \mathbf{S}_{N_s} \Theta]$ and $\Psi_R \triangleq [\mathbf{R}_1 \Phi, \mathbf{R}_2 \Phi, \dots, \mathbf{R}_{N_r} \Phi]$. From (27), the MSE can be rewritten as

$$\text{MSE} = \sum_{k=1}^{N_d} \text{tr} \left(\left[\begin{pmatrix} \mathbf{C}_{b,k}^x & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_{b,k}^\lambda \end{pmatrix}^{-1} + \eta_k \begin{pmatrix} \sqrt{\alpha} \Psi_S^H \\ \Psi_R^H \end{pmatrix} (\sqrt{\alpha} \Psi_S, \Psi_R) \right]^{-1} \right) \quad (51)$$

where

$$\eta_k \triangleq \left(\alpha \sum_{n=1}^{N_r} \sigma_{k,n}^g + 1 \right)^{-1}, \quad k = 1, \dots, N_d. \quad (52)$$

From (20)–(22), $\mathbf{C}_{b,k}^x$ and $\mathbf{C}_{b,k}^\lambda$ are all diagonal matrices. Hence, the MSE in (51) is minimized only if all off-diagonal matrices of the second term are zero, i.e.,

$$\Psi_S^H \Psi_S = \mathbf{D}_s, \quad \Psi_R^H \Psi_R = \mathbf{D}_r, \quad \Psi_S^H \Psi_R = \mathbf{0}$$

where $\mathbf{D}_s \triangleq \text{diag}[\mathbf{D}_{s,1}, \mathbf{D}_{s,2}, \dots, \mathbf{D}_{s,N_s}]$ and $\mathbf{D}_r \triangleq \text{diag}[\mathbf{D}_{r,1}, \mathbf{D}_{r,2}, \dots, \mathbf{D}_{r,N_r}]$ are diagonal matrices, and $\mathbf{0}$ denotes a zero matrix. From the definition of Ψ_S and Ψ_R , we have that for $m, p = 1, \dots, N_s$ and $n, q = 1, \dots, N_r$,

$$\begin{aligned} (\mathbf{S}_m \Theta)^H \mathbf{S}_m \Theta &= \mathbf{D}_{s,m}, & (\mathbf{S}_m \Theta)^H \mathbf{S}_p \Theta &= \mathbf{0}, & m \neq p \\ (\mathbf{R}_n \Phi)^H \mathbf{R}_n \Phi &= \mathbf{D}_{r,n}, & (\mathbf{R}_n \Phi)^H \mathbf{R}_q \Phi &= \mathbf{0}, & n \neq q \\ (\mathbf{S}_m \Theta)^H \mathbf{R}_n \Phi &= \mathbf{0}. \end{aligned}$$

Then we can rewrite the MSE in (51) as

$$\begin{aligned} \text{MSE} &= \sum_{k=1}^{N_d} \text{tr} \left(\sum_{m=1}^{N_s} \left[(\mathbf{C}_{b,k}^{x,m})^{-1} + \alpha \eta_k \mathbf{D}_{s,m} \right]^{-1} \right. \\ &\quad \left. + \sum_{n=1}^{N_r} \left[(\mathbf{C}_{b,k}^{\lambda,n})^{-1} + \eta_k \mathbf{D}_{r,n} \right]^{-1} \right). \end{aligned} \quad (53)$$

Moreover, due to the structure of Θ and Φ in (13), we have $\mathbf{D}_{s,m} = \beta_m \mathbf{I}_{2Q+1}$ and $\mathbf{D}_{r,n} = \gamma_n \mathbf{I}_{Q+1}$. This indicates that

$$\mathbf{s}_m^H \mathbf{s}_m = \beta_m, \quad m = 1, \dots, N_s, \quad \mathbf{r}_n^H \mathbf{r}_n = \gamma_n, \quad n = 1, \dots, N_r. \quad (54)$$

We would like to note that the transmission power constraints in (28) and (29) are not affected by (54).

APPENDIX B
PROOF OF THEOREM 2

Let us denote $\xi_m \triangleq \alpha \beta_m$, $m = 1, \dots, N_s$, the problem (42)–(45) can be rewritten as

$$\min_{\{\xi_m\}, \{\gamma_n\}, \alpha} \sum_{m=1}^{N_s} \sum_{k=1}^{N_d} \sum_{q=1}^{2Q+1} \frac{1}{c_{k,m,q} + \xi_m \eta_k} + \sum_{n=1}^{N_r} \sum_{k=1}^{N_d} \sum_{p=1}^{Q+1} \frac{1}{d_{k,n,p} + \gamma_n \eta_k} \quad (55)$$

$$\text{s.t.} \quad \sum_{m=1}^{N_s} \xi_m \leq \alpha P_s \quad (56)$$

$$\sum_{m=1}^{N_s} \kappa_m \xi_m + \sum_{n=1}^{N_r} \gamma_n \leq P_r - \alpha N_r L \quad (57)$$

$$\alpha > 0, \quad \xi_m \geq 0, \quad m = 1, \dots, N_s, \quad \gamma_n \geq 0, \quad n = 1, \dots, N_r \quad (58)$$

where $\{\xi_m\} \triangleq \{\xi_m, m = 1, \dots, N_s\}$. We first ignore the effect of α on all η_k by assuming them as known variables. Then the problem (55)–(58) is a convex optimization problem as the objective function (55) is convex with respect to $\{\xi_m\}$ and $\{\gamma_n\}$, and (56)–(58) are linear inequality constraints. From (56), (57), we can see that when α is sufficiently small, the objective function (55) is strongly governed by (56), as (57) is inactive in this case. The increment of α from a small value results in a decrement of (55), as the feasible region specified by (56) expands. On the other hand, when α is large, i.e., close to $P_r/(N_r L)$, the objective function (55) strongly depends on the constraint (57), as the constraint (56) is inactive in this case. When α decreases from a large value, (55) also decreases as the feasible region specified by (57) expands.

Now we investigate the effect of α on η_k . From (52), it can be seen that η_k is a monotonically decreasing function of α , and the objective function (55) increases when η_k decreases. Then, from the discussion above, we can see that when α increases from a significantly small positive number, the objective function (55) decreases since the potential decrement of (55) caused by the expanded feasible region of (56) dominates the potential increment of (55) due to the decreasing of η_k . The objective function (55) continues to decrease till a ‘turning point’ where the constraint (56) loses its dominance. After such turning point, the value of (55) is monotonically increasing with an increasing α . Thus, we can conclude that the objective function (42) with the constraints (43)–(45) is a unimodal function with respect to α .

ACKNOWLEDGMENT

The authors would like to thank the editor and anonymous reviewers for their valuable comments and suggestions that improved the quality of the paper.

REFERENCES

- [1] M. Yuksel and E. Erkip, “Multi-antenna cooperative wireless systems: A diversity-multiplexing tradeoff perspective,” *IEEE Trans. Inf. Theory*, vol. 53, no. 10, pp. 3371–3393, Oct. 2007.
- [2] Y. Fan and J. Thompson, “MIMO configurations for relay channels: Theory and practice,” *IEEE Trans. Wireless Commun.*, vol. 6, no. 5, pp. 1774–1786, May 2007.
- [3] X. Tang and Y. Hua, “Optimal design of non-regenerative MIMO wireless relays,” *IEEE Trans. Wireless Commun.*, vol. 6, no. 4, pp. 1398–1407, Apr. 2007.
- [4] I. Hammerström and A. Wittneben, “Power allocation schemes for amplify-and-forward MIMO-OFDM relay links,” *IEEE Trans. Wireless Commun.*, vol. 6, no. 8, pp. 2798–2802, Aug. 2007.
- [5] B. Wang, J. Zhang, and A. Høst-Madsen, “On the capacity of MIMO relay channels,” *IEEE Trans. Inf. Theory*, vol. 51, no. 1, pp. 29–43, Jan. 2005.
- [6] L. Sanguinetti, A. A. D’Amico, and Y. Rong, “A tutorial on transceiver design for amplify-and-forward MIMO relay systems,” *IEEE J. Sel. Areas Commun.*, vol. 30, no. 8, pp. 1331–1346, Sep. 2012.
- [7] N. Yang, M. ElKashlan, P. L. Yeoh, and J. Yuan, “Multiuser MIMO relay networks in Nakagami-m fading channels,” *IEEE Trans. Commun.*, vol. 60, no. 11, pp. 3298–3310, Nov. 2012.
- [8] P. L. Yeoh, M. ElKashlan, N. Yang, D. B. da Costa, and T. Q. Duong, “Unified analysis of transmit antenna selection in MIMO multirelay networks,” *IEEE Trans. Veh. Technol.*, vol. 62, no. 2, pp. 933–939, Feb. 2013.
- [9] H. Phan, T. Q. Duong, M. ElKashlan, and H.-J. Zepernick, “Beamforming amplify-and-forward relay networks with feedback delay and interference,” *IEEE Signal Process. Lett.*, vol. 19, no. 1, pp. 16–19, Jan. 2012.
- [10] Y. Rong, X. Tang, and Y. Hua, “A unified framework for optimizing linear non-regenerative multicarrier MIMO relay communication systems,” *IEEE Trans. Signal Process.*, vol. 57, no. 12, pp. 4837–4851, Dec. 2009.
- [11] M. R. A. Kandaker and Y. Rong, “Precoding design for MIMO relay multicasting,” *IEEE Trans. Wireless Commun.*, vol. 12, no. 7, pp. 3544–3555, Jul. 2013.
- [12] A. S. Behbahani, R. Merched, and A. M. Eltawil, “Optimization of a MIMO relay network,” *IEEE Trans. Signal Process.*, vol. 56, no. 10, pp. 5062–5073, Oct. 2008.
- [13] P. Lioliou and M. Viberg, “Least-squares based channel estimation for MIMO relays,” in *Proc. Int. ITG Workshop Smart Antennas*, Feb. 2008, pp. 90–95.
- [14] P. Lioliou, M. Viberg, and M. Coldrey, “Efficient channel estimation techniques for amplify and forward relaying systems,” *IEEE Trans. Commun.*, vol. 60, no. 11, pp. 3150–3155, Nov. 2012.
- [15] T. Kong and Y. Hua, “Optimal design of source and relay pilots for MIMO relay channel estimation,” *IEEE Trans. Signal Process.*, vol. 59, no. 9, pp. 4438–4446, Sep. 2011.
- [16] Y. Rong, M. R. A. Kandaker, and Y. Xiang, “Channel estimation of dual-hop MIMO relay systems via parallel factor analysis,” *IEEE Trans. Wireless Commun.*, vol. 11, no. 6, pp. 2224–2233, Jun. 2012.
- [17] C. S. Patel and G. L. Stuber, “Channel estimation for amplify and forward relay based cooperation diversity systems,” *IEEE Trans. Wireless Commun.*, vol. 6, no. 6, pp. 2348–2356, Aug. 2007.
- [18] M. Dong, L. Tong, and B. M. Sadler, “Optimal insertion of pilot symbols for transmissions over time-varying flat fading channels,” *IEEE Trans. Signal Process.*, vol. 52, no. 5, pp. 1403–1418, Apr. 2004.
- [19] G. Arredondo, W. H. Chriss, and E. H. Walker, “A multipath fading simulator for mobile radio,” *IEEE Trans. Commun.*, vol. 21, no. 11, pp. 1325–1328, Jan. 2003.
- [20] X. Ma, G. B. Giannakis, and S. Ohno, “Optimal training for block transmissions over doubly selective wireless fading channels,” *IEEE Trans. Signal Process.*, vol. 51, no. 5, pp. 1351–1366, Apr. 2003.
- [21] G. B. Giannakis and C. Tepedelenlioglu, “Basis expansion models and diversity techniques for blind identification and equalization of time-varying channels,” in *Proc. IEEE*, vol. 86, no. 10, pp. 1969–1986, Aug. 2002.
- [22] G. Wang, F. Gao, J. Lin, and C. Tellambura, “Doubly selective channel estimation for amplify-and-forward relay networks,” in *Proc. IEEE Wireless Commun. Netw. Conf.*, Paris, France, Apr. 1–4, 2012, pp. 326–330.

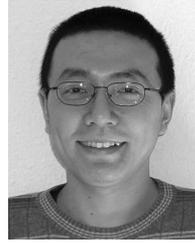
- [23] G. Wang, F. Gao, W. Chen, and C. Tellambura, "Channel estimation and training design for two-way relay networks in time-selective fading environments," *IEEE Trans. Wireless Commun.*, vol. 10, no. 8, pp. 2681–2691, Jun. 2011.
- [24] S. Zhang, F. Gao, J. Li, and M. Sheng, "Online and offline Bayesian Cramér-Rao bounds for time-varying channel estimation in one-way relay networks," *IEEE Trans. Signal Process.*, vol. 63, no. 8, pp. 1977–1992, Apr. 2015.
- [25] S. Zhang, F. Gao, H. Wang, and C. Pei, "Dynamic individual channel estimation for one-way relay networks with time-multiplexed-superimposed training," *IEEE Trans. Veh. Technol.*, vol. 63, no. 8, pp. 3841–3852, Oct. 2014.
- [26] J. K. Cavers, "An analysis of pilot symbol assisted modulation for Rayleigh fading channels," *IEEE Trans. Veh. Technol.*, vol. 40, no. 4, pp. 686–693, Nov. 1991.
- [27] J. Gao and H. Liu, "Decision-directed estimation of MIMO time-varying Rayleigh fading channels," *IEEE Trans. Wireless Commun.*, vol. 4, no. 4, pp. 1412–1417, Jul. 2005.
- [28] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.
- [29] A. Antoniou and W.-S. Lu, *Practical Optimization: Algorithms and Engineering Applications*. New York, NY, USA: Springer Science+Business Media, LCC, 2007.



Choo W. R. Chiong received the B.E. (Hons.) and Ph.D. degrees in electrical engineering from Curtin University, Bentley, WA, Australia, in 2010 and 2015, respectively.

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

Dr. 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.

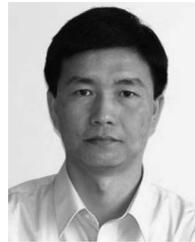


Yue Rong (S'03–M'06–SM'11) received the Ph.D. degree (*summa cum laude*) in electrical engineering from the Darmstadt University of Technology, Darmstadt, Germany, in 2005.

He was a Post-Doctoral Researcher with the Department of Electrical Engineering, University of California, Riverside, CA, USA, from February 2006 to November 2007. Since December 2007, he has been with the Department of Electrical and Computer Engineering, Curtin University, Bentley, WA, 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. He has published over 120 journal and conference paper in these areas.

Dr. Rong was a recipient of the Best Paper Award at the 2011 International Conference on Wireless Communications and Signal Processing, the Best Paper Award at the 2010 Asia-Pacific Conference on Communications, and the Young Researcher of the Year Award of the Faculty of Science and Engineering at Curtin University in 2010. He is an Associate Editor of the *IEEE TRANSACTIONS ON SIGNAL PROCESSING*. He was an Editor of the *IEEE WIRELESS COMMUNICATIONS LETTERS* from 2012 to 2014, a Guest Editor of the *IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS* special issue on theories and methods for advanced wireless relays, and was a TPC Member for the *IEEE ICC*, *WCSP*, *IWCMC*, and *ChinaCom*.



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

He currently serves as Associate Editor of *IEEE Access*.