

# Channel Estimation and Optimal Training Design for Correlated MIMO Two-Way Relay Systems in Colored Environment

Rui Wang, *Member, IEEE*, Meixia Tao, *Senior Member, IEEE*,  
Hani Mehrpouyan, *Member, IEEE*, and Yingbo Hua, *Fellow, IEEE*

**Abstract**—In this paper, while considering the impact of the antenna correlation and the interference from neighboring users, we analyze channel estimation and training sequence design for multi-input multi-output (MIMO) two-way relay systems. To this end, we propose to decompose the bidirectional transmission links into two phases, i.e., the multiple access (MAC) phase and the broadcast (BC) phase. By considering the Kronecker-structured channel model, we derive the optimal linear minimum mean-square-error (LMMSE) channel estimators. The corresponding training designs for the MAC phase and the BC phase are then formulated and solved to improve channel estimation accuracy. For the general scenario of the training sequence design for both phases, two iterative training design algorithms are proposed that are verified to produce training sequences achieving near optimal channel estimation performance. Furthermore, for specific practical scenarios, where the covariance matrices of the channel or disturbances are of particular structures, the optimal training sequence design guidelines are obtained. The minimum required training lengths for channel estimation in both the MAC phase and the BC phase are also analyzed. Comprehensive simulations are carried out to demonstrate the effectiveness of the proposed training designs.

**Index Terms**—MIMO, two-way relaying, channel estimation, linear minimum mean-square-error, convex optimization.

## I. INTRODUCTION

RELAY assisted cooperative communication has been regarded as one of the most promising techniques in combating long distance channel fading in complex wireless communication systems. One popular example is one-way relaying, which has been well studied in the past decade [2], [3]. Although one-way relaying shows great potential in reducing

power consumption, enhancing reliability, and extending coverage, it suffers from low spectral efficiency due to the half-duplex nature of the network. To overcome this disadvantage, by using the idea of network coding, two-way relaying (TWR) has been proposed and has received great attention recently. In fact, TWR can maintain the advantages of traditional relaying while doubling spectral efficiency.

The improvement in spectral efficiency in TWR is achieved by applying self-interference cancellation at each source node and extracting the desired information from the received network-coded messages. In this case, the accuracy of the self-interference cancellation process significantly affects the performance of TWR systems. Moreover, when using the popular amplify-and-forward (AF) relaying strategy, the accuracy of self-interference cancellation process heavily depends on the precision of the channel estimation. Thus, obtaining highly accurate channel state information (CSI) becomes more important in TWR systems compared to traditional one-way relaying systems. In fact, devising new channel estimation schemes for TWR systems has received great attention recently. For example, in [4], the authors propose to estimate the cascaded channel of TWR systems under the AF relaying strategy. By using multiple phase shift keying (M-PSK) training symbols, blind and partially-blind channel estimators are investigated in [5], [6]. Different from [4]–[6], where flat fading channels are assumed, the authors in [7] investigate time varying channel estimation via a new complex-exponential basis expansion model. Moreover, in [8], [9], the channel estimation for TWR is extended to the scenario of orthogonal frequency division multiplexing (OFDM) systems.

It is worth noting that the works summarized above are concerned with single-antenna TWR systems. As expected, the multi-antenna or multi-input multi-output (MIMO) technique can be introduced into TWR systems to further improve transmission reliability and bandwidth efficiency. One efficient way to realize such performance improvement is to exploit the estimated CSI for the application of source and relay precoding [10], [11]. Therefore, in MIMO TWR systems, in addition to affecting the performance of self-interference cancellation, inaccurate channel estimation also imposes a negative effect on the precoder design.

Fig. 1 depicts a MIMO TWR setup. Let us denote the process of data transmission from the source nodes to the relay and the relay to the source nodes as multiple access (MAC) phase and

Manuscript received March 21, 2014; revised September 15, 2014 and December 20, 2014; accepted December 24, 2014. Date of publication January 12, 2015; date of current version May 7, 2015. The work of Rui Wang was supported by the National Natural Science Foundation of China under Grant 61401313. The work of Meixia Tao was supported by the NSF of China under grant 61322102. Part of this work was presented at IEEE Globecom, Austin, TX, USA, December 2014. The associate editor coordinating the review of this paper and approving it for publication was S. Bhashyam.

R. Wang is with the College of Electronics and Information Engineering, Tongji University, Shanghai 200092, China (e-mail: ruiwang@tongji.edu.cn).

M. Tao is with the Department of Electronic Engineering at Shanghai Jiao Tong University, Shanghai 200240, China (e-mail: mxtao@sjtu.edu.cn).

H. Mehrpouyan is with the Department of Computer and Electrical Engineering and Computer Science at California State University, Bakersfield, CA 93311, USA (e-mail: hani.mehr@ieee.org).

Y. Hua is with the University of California, Riverside, CA 92521 USA (e-mail: yhua@ee.ucr.edu).

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.2015.2390645

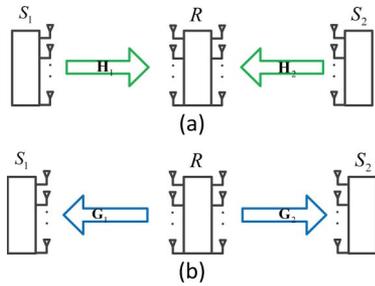


Fig. 1. An illustration of MIMO two-way relay system. (a) 1-st time slot (MAC phase); (b) 2-nd time slot (BC phase).

broadcast (BC) phase, respectively. In [12], the authors propose a MIMO channel estimator that uses the self-interference as a training sequence to estimate the channel matrices corresponding to the BC phase. In [13], the performances of different channel estimators, including individual and cascaded channel estimators, are compared based on the least squares (LS) criterion. In [14], an LS estimator is used to obtain the cascaded channel matrices corresponding to the BC and MAC phases using a single carrier cyclic prefix. Note that in the contributions of [12]–[14], the channel statistics of random channel matrices, whether cascaded channels or individual channels, are assumed to be unknown. Based on the estimation theory, if channel statistics are known, the channel estimation can be conducted under the Bayesian framework and the estimation accuracy can be further enhanced. Hence, by taking these statistics into account, we seek to improve upon the channel estimators in [12]–[14].

Very recently, the authors in [15], [16] independently investigate the minimum mean-square-error (MMSE) channel estimation for TWR systems based on a correlated Gaussian MIMO channel model. In particular, in [15], the cascaded channel matrices for AF TWR systems are estimated and the training sequences at the two source nodes are optimized to minimize the total channel estimation MSE. Different from [15], the authors in [16] aim to estimate the individual channel matrices for each link. To reach this goal, two different estimation schemes, i.e., a superimposed channel training scheme and a two-stage channel estimation scheme, are proposed. In addition, the training sequences at the two source nodes, as well as, at the relay node are jointly optimized to improve channel estimation accuracy.

In this paper, similar to [1], [12]–[16], while assuming that the channel statistics are known, we analyze and devise channel estimators for correlated MIMO TWR systems. Specifically, we consider the Kronecker-structured channel model, such that the individual channel matrices can be estimated based on the Bayesian framework. However, *unlike* [15], [16], we also take into account the interference from the nearby users. Thus, in this model, the disturbance at each of the source nodes or at the relay node consists of both noise and interference. Note that the considered colored estimation environment may be more practical for applications in today's more densely deployed wireless networks. Although channel estimation in point-to-point MIMO systems in colored environments has been studied in [17], [18], to the best of our knowledge, this topic has not been addressed in the TWR scenario.

To enhance TWR performance, we seek to estimate the individual channel matrices corresponding to the source-to-relay link and the relay-to-source link, see Fig. 1. To this end, a new two-phase estimation scheme is proposed, where the bidirectional transmission of a TWR system is decomposed into the MAC phase and the BC phase. For the MAC and BC phases, the channel estimations are performed at the relay node and at the two source nodes, respectively. The proposed estimation scheme is different from the ones in [15], [16], where the channel estimation is assumed to only be conducted at the source nodes. As such, our proposed estimation scheme can more efficiently support precoding at the relay since it requires significantly less feedback overhead [10], [19], [20]. Based on the proposed estimation scheme, we derive the optimal linear MMSE (LMMSE) estimator for each phase. Next, the corresponding training design problems are formulated with the aim of minimizing the total MSE of channel estimation process for each phase. The training design problem considered here is different from that of [15], [16], since we take into account the effect of colored disturbances caused by interference at the relay node and at the source nodes. Moreover, the training design scenarios for point-to-point systems in [17], [18] are different from the scenario under consideration in this paper, since our proposed training sequence design is optimized to simultaneously enhance channel estimation accuracy over both links in the BC phase and in the MAC phase. Although, for the general scenario, it is difficult to derive the optimal training sequence structures as in [15]–[18], we propose two iterative design algorithms to solve the training design problems. These algorithms are verified to converge quickly to the near optimal solution and to not be sensitive to the initialization process. For some special cases, where the covariance matrices of the channels or disturbances have specific forms, two specific approaches are applied to obtain the optimal training sequences: 1) the original problem is converted into a standard convex optimization problem; 2) the optimal structures of the training sequences are first derived and then used to reduce the original non-convex problem into a simple power allocation problem. Finally, to reduce training overhead, the minimum required training lengths for channel estimation in both the MAC and BC phases are analyzed. Extensive simulations are carried out to support the findings of the paper.

The rest of the paper is organized as follows. In Section II, we present the system model. The LMMSE estimators for both the MAC and BC phases are derived in Section III. The training designs for the MAC and BC phases are analyzed in Section IV and V, respectively. Simulation results are provided in Section VI. We conclude the paper in Section VII.

*Notations:*  $\mathcal{E}(\cdot)$  denotes the expectation operator.  $\otimes$  denotes the Kronecker operator.  $\text{vec}(\cdot)$  signifies the matrix vectorization operator. Superscripts  $\mathbf{A}^T$ ,  $\mathbf{A}^*$ , and  $\mathbf{A}^H$  denote the transpose, conjugate, and conjugate transpose of matrix  $\mathbf{A}$ , respectively.  $\text{Tr}(\mathbf{A})$ ,  $\mathbf{A}^{-1}$ ,  $\det(\mathbf{A})$ , and  $\text{Rank}(\mathbf{A})$  stand for the trace, inverse, determinant, and rank of  $\mathbf{A}$ , respectively.  $\lambda(\mathbf{A})$  denotes a vector containing eigenvalues of  $\mathbf{A}$ .  $\text{Blkdiag}(\mathbf{A}_1, \mathbf{A}_2, \dots, \mathbf{A}_N)$  denotes a block diagonal matrix constructed by matrices  $\mathbf{A}_i$ , for  $\forall i$ .  $\text{Diag}(\mathbf{a})$  denotes a diagonal matrix with  $\mathbf{a}$  being its diagonal entries.  $\mathbf{A}(n:m, :)$  and  $\mathbf{A}(:, n:m)$  denote the sub-matrices

constructed by  $n$  to  $m$  rows and  $n$  to  $m$  columns of  $\mathbf{A}$ , respectively.  $\|\mathbf{A}\|_F^2$  denotes the Frobenius norm of  $\mathbf{A}$ .  $\Re(z)$  denotes the real part of a complex variable  $z$ . The distribution of a circular symmetric complex Gaussian vector with mean vector  $\mathbf{x}$  and covariance matrix  $\Sigma$  is denoted by  $C\mathcal{N}(\mathbf{x}, \Sigma)$ .  $\mathbb{C}^{x \times y}$  denotes the space of complex  $x \times y$  matrices.  $\mathbb{S}^N$  and  $\mathbb{S}_+^N$  denote the set of symmetric  $N \times N$  matrices and the set of positive semidefinite  $N \times N$  matrices, respectively.  $\mathbf{x} \preceq \mathbf{y}$  denotes that the vector  $\mathbf{y}$  majorizes the vector  $\mathbf{x}$ .

## II. SYSTEM MODEL

Consider a TWR system, where sources  $S_1$  and  $S_2$  exchange messages with one another through a relay  $R$ .  $S_1$ ,  $R$ , and  $S_2$  are assumed to be equipped with  $N_1$ ,  $M$ , and  $N_2$  antennas, respectively. Channel matrices from  $S_1$  and  $S_2$  to  $R$  are denoted by  $\mathbf{H}_1$  and  $\mathbf{H}_2$ , respectively, and channel matrices from  $R$  to  $S_1$  and  $S_2$  are denoted by  $\mathbf{G}_1$  and  $\mathbf{G}_2$ , respectively.

Signal transmission within the TWR system is achieved in two time slots. In the first phase, referred to as the MAC phase, the source node  $S_i$ , for  $i = 1, 2$ , transmit their signals to the relay node  $R$ , while in the second phase, referred to as the BC phase, the relay node  $R$  forwards its combined received signal to the two source nodes  $S_1$  and  $S_2$ . The proposed channel estimator aims to obtain the individual channels of the two hops, i.e.,  $\{\mathbf{H}_1, \mathbf{H}_2, \mathbf{G}_1, \mathbf{G}_2\}$ , which is different from the cases studied in [4], [15] to estimate the cascaded channel. This approach facilitates the precoding design and/or the power allocation at the relay node, which can further improve the overall system performance [10], [11], [19], and [20].

Following the transmission in Fig. 1, we assume that  $\mathbf{H}_1$  and  $\mathbf{H}_2$  are estimated in the MAC phase via the training signals sent from the two sources, and  $\mathbf{G}_1$  and  $\mathbf{G}_2$  are estimated in the BC phase via the training signal transmitted from the relay. Moreover, the channels follow quasi-static fadings, i.e., the channel matrices  $\{\mathbf{H}_1, \mathbf{H}_2, \mathbf{G}_1, \mathbf{G}_2\}$  are random but stay constant during the whole duration of transmission.

The received training signals at the relay node in the MAC phase can be expressed as

$$\mathbf{y}_R(t) = \mathbf{H}_1 \mathbf{s}_1(t) + \mathbf{H}_2 \mathbf{s}_2(t) + \mathbf{n}_R(t), \quad (1)$$

where  $\mathbf{s}_i(t)$  denotes the training signal at  $S_i$  and  $\mathbf{n}_R(t)$  represents the correlated disturbance at  $R$ .  $\mathbf{n}_R(t)$  includes the total background noise as well as the interference from adjacent communication links, and is modeled as a stochastic process with respect to the time variable  $t$  [17], [18], [21]. To estimate the channel matrices at the relay, the sources typically need to send a sequence of known training signals. Assuming training sequences have a length of  $L_S$ , the received signal in (1) can be written in a matrix form as

$$\mathbf{Y}_R = \mathbf{H}_1 \mathbf{S}_1 + \mathbf{H}_2 \mathbf{S}_2 + \mathbf{N}_R, \quad (2)$$

where  $\mathbf{Y}_R \triangleq [\mathbf{y}_R(1), \mathbf{y}_R(2), \dots, \mathbf{y}_R(L_S)]$ ,  $\mathbf{S}_i \triangleq [\mathbf{s}_i(1), \mathbf{s}_i(2), \dots, \mathbf{s}_i(L_S)]$  and  $\mathbf{N}_R \triangleq [\mathbf{n}_R(1), \mathbf{n}_R(2), \dots, \mathbf{n}_R(L_S)]$ . Suppose that the source node  $S_i$  has a maximum power of  $\tau_i$  during the channel

estimation phase, the training sequence  $\mathbf{S}_i$  should fulfill the following power constraint

$$\text{Tr}(\mathbf{S}_i \mathbf{S}_i^H) \leq \tau_i, \quad i = 1, 2. \quad (3)$$

Now we give the distributions of the channel matrices  $\mathbf{H}_i$ , for  $i = 1, 2$ , and the disturbance  $\mathbf{N}_R$  in (2). We assume that the channel matrices  $\mathbf{H}_i$ , for  $i = 1, 2$ , follow the *Kronecker-structured* model [15]–[18], [21] and are expressed as

$$\mathbf{H}_i = \mathbf{C}_{r,H} \mathbf{W}_{H_i} \mathbf{C}_{t,H_i}^T, \quad i = 1, 2 \quad (4)$$

where indexes ‘ $t$ ’ and ‘ $r$ ’ denote the ‘transmitter’ and ‘receiver’, respectively;  $\mathbf{W}_{H_i}$ , for  $i = 1, 2$ , are assumed to be unknown matrices, where their entries are modeled by  $C\mathcal{N}(0, 1)$ ;  $\mathbf{C}_{r,H}$  denotes the receive correlation matrix of both  $\mathbf{H}_1$  and  $\mathbf{H}_2$ ;<sup>1</sup> and  $\mathbf{C}_{t,H_i}$  denotes the transmit correlation matrix of  $\mathbf{H}_i$ . In (4), the correlation amongst the channel parameters can be caused by insufficient antenna spacing as verified by the measurements in [22], [23]. Based on the channel model in (4), we have  $\text{vec}(\mathbf{H}_i) \sim C\mathcal{N}(\mathbf{0}, \mathbf{Z}_{H_i})$  with

$$\mathbf{Z}_{H_i} = \mathbf{Z}_{t,H_i} \otimes \mathbf{Z}_{r,H}, \quad i = 1, 2, \quad (5)$$

where  $\mathbf{Z}_{t,H_i} = \mathbf{C}_{t,H_i} \mathbf{C}_{t,H_i}^H$  and  $\mathbf{Z}_{r,H} = \mathbf{C}_{r,H} \mathbf{C}_{r,H}^H$ . Here,  $\mathbf{Z}_{t,H_i}$ , for  $i = 1, 2$ , denote the covariance matrices at the transmitter side, and  $\mathbf{Z}_{r,H}$  denotes the covariance matrix at the receiver side. We assume that the correlation matrices  $\mathbf{C}_{r,H}$  and  $\mathbf{C}_{t,H_i}$ , for  $i = 1, 2$ , are known in prior. This also implies that the channel covariance matrices  $\mathbf{Z}_{r,H}$  and  $\mathbf{Z}_{H_i}$ , for  $i = 1, 2$ , are known and are in turn used for the training design in the MAC phase.

Regarding the disturbance  $\mathbf{N}_R$ , we denote  $\mathbf{n}_R = \text{vec}(\mathbf{N}_R)$  and assume that  $\mathbf{n}_R$  is a random vector with mean  $\mathbf{0}$  and  $\mathcal{E}(\mathbf{n}_R \mathbf{n}_R^H) = \mathbf{K}_R$ . Moreover,  $\mathbf{K}_R$  is structured and modeled by [17], [18]

$$\mathbf{K}_R = \mathbf{K}_{q,R} \otimes \mathbf{K}_{r,R}, \quad (6)$$

where  $\mathbf{K}_{q,R} \in \mathbb{C}^{L_S \times L_S}$  denotes the temporal covariance matrix;  $\mathbf{K}_{r,R} \in \mathbb{C}^{M \times M}$  denotes the receive spatial covariance matrix. Moreover, the covariances of disturbance  $\mathbf{K}_{q,R}$  and  $\mathbf{K}_{r,R}$  are also known in prior. This further implies that  $\mathbf{K}_R$  is known.  $\mathbf{K}_{q,R}$  and  $\mathbf{K}_{r,R}$  shall be utilized for the training design in what follows.

In the BC phase, the received training signals at the source nodes are given by

$$\mathbf{y}_i(t) = \mathbf{G}_i \mathbf{s}_R(t) + \mathbf{n}_i(t), \quad i = 1, 2 \quad (7)$$

where  $\mathbf{s}_R(t)$  denotes the training signal at the relay node and  $\mathbf{n}_i(t)$  represents the correlated disturbance at the source node  $S_i$ . By rewriting (7) into the matrix form, we have

$$\mathbf{Y}_i = \mathbf{G}_i \mathbf{S}_R + \mathbf{N}_i, \quad i = 1, 2, \quad (8)$$

where  $\mathbf{Y}_i \triangleq [\mathbf{y}_i(1), \mathbf{y}_i(2), \dots, \mathbf{y}_i(L_R)]$ ,  $\mathbf{S}_R \triangleq [\mathbf{s}_R(1), \mathbf{s}_R(2), \dots, \mathbf{s}_R(L_R)]$  and  $\mathbf{N}_i \triangleq [\mathbf{n}_i(1), \mathbf{n}_i(2), \dots, \mathbf{n}_i(L_R)]$  with  $L_R$  being the

<sup>1</sup>Here,  $\mathbf{H}_1$  and  $\mathbf{H}_2$  share a common receive correlation matrix as they terminate at the relay node.

training sequence length at the relay. The following condition must be met to satisfy the power constraint at the relay node

$$\text{Tr}(\mathbf{S}_R \mathbf{S}_R^H) \leq \tau_R. \quad (9)$$

In (9),  $\tau_R$  denotes the maximum power at the relay node during the training phase.

Similar to the MAC phase, the channel matrices in the BC phase also follow the *Kronecker-structured* model, i.e.,

$$\mathbf{G}_i = \mathbf{C}_{r,G_i} \mathbf{W}_{G_i} \mathbf{C}_{t,G}^T, \quad i = 1, 2, \quad (10)$$

where  $\mathbf{W}_{G_i}$ , for  $i = 1, 2$ , are unknown matrices in which the entries are modeled by  $C\mathcal{N}(0, 1)$ ;  $\mathbf{C}_{t,G}$  denotes the transmit correlation matrix of both  $\mathbf{G}_1$  and  $\mathbf{G}_2$ ; <sup>2</sup> and  $\mathbf{C}_{r,G_i}$  denotes the receive correlation matrix of  $\mathbf{G}_i$ . We also assume that the correlation matrices  $\mathbf{C}_{t,G}$  and  $\mathbf{C}_{r,G_i}$ , for  $i = 1, 2$ , are known in prior and will be used for the training sequence design in the BC phase. Based on (10), we then have  $\text{vec}(\mathbf{G}_i) \sim C\mathcal{N}(\mathbf{0}, \mathbf{Z}_{G_i})$  with

$$\mathbf{Z}_{G_i} = \mathbf{Z}_{t,G} \otimes \mathbf{Z}_{r,G_i}, \quad i = 1, 2, \quad (11)$$

where  $\mathbf{Z}_{t,G} = \mathbf{C}_{t,G} \mathbf{C}_{t,G}^H$  and  $\mathbf{Z}_{r,G_i} = \mathbf{C}_{r,G_i} \mathbf{C}_{r,G_i}^H$ .

As in (6), we denote  $\mathbf{n}_i = \text{vec}(\mathbf{N}_i)$ , for  $i = 1, 2$ , and assume that  $\mathbf{n}_i$  is a random vector with mean  $\mathbf{0}$  and  $\mathcal{E}(\mathbf{n}_i \mathbf{n}_i^H) = \mathbf{K}_i$  with  $\mathbf{K}_i$  having a structure of

$$\mathbf{K}_i = \mathbf{K}_{q,i} \otimes \mathbf{K}_{r,i} \in \mathbb{S}^{N_i L_R \times N_i L_R}, \quad i = 1, 2, \quad (12)$$

where  $\mathbf{K}_{q,i} \in \mathbb{C}^{L_R \times L_R}$ , for  $i = 1, 2$ , denote the temporal covariance matrices and  $\mathbf{K}_{r,i} \in \mathbb{C}^{N_i \times N_i}$ , for  $i = 1, 2$ , denote the received spatial covariance matrices. Also, we assume that the statistics of  $\mathbf{K}_{q,i}$  and  $\mathbf{K}_{r,i}$ , for  $i = 1, 2$ , are known in prior.

Other assumptions made throughout this paper are listed as follows. We assume that  $\mathbf{K}_{r,1}$ ,  $\mathbf{K}_{r,2}$ , and  $\mathbf{K}_{r,R}$  share the same eigenvectors with  $\mathbf{Z}_{r,G_1}$ ,  $\mathbf{Z}_{r,G_2}$  and  $\mathbf{Z}_{r,H}$ , respectively. This assumption models the following scenarios as summarized in [18], [21]: 1) Additive noise-limited scenario,  $\mathbf{K}_R = \mu_R \mathbf{I}_{ML_S} = \mathbf{I}_{L_S} \otimes \mu_R \mathbf{I}_M$  and  $\mathbf{K}_i = \mu_i \mathbf{I}_{N_i L_R} = \mathbf{I}_{L_R} \otimes \mu_i \mathbf{I}_{N_i}$ , for  $i = 1, 2$ , which implies  $\mathbf{K}_{r,R} = \mu_R \mathbf{I}_M$  and  $\mathbf{K}_{r,i} = \mu_i \mathbf{I}_{N_i}$  with  $\mu_R$  and  $\mu_i$  being the variances of the additive noises at the relay and the sources  $S_i$ , respectively; 2) Interference-limited scenario,  $\mathbf{K}_R = \mathbf{Q}_R \otimes \mathbf{Z}_{r,H}$ ,  $\mathbf{K}_i = \mathbf{Q}_{G_i} \otimes \mathbf{Z}_{r,G_i}$  with  $\mathbf{Q}_R$  and  $\mathbf{Q}_{G_i}$ ,  $i = 1, 2$ , denoting the temporal covariances from nearby interfering users, which implies  $\mathbf{K}_{r,R} = \mathbf{Z}_{r,H}$  and  $\mathbf{K}_{r,i} = \mathbf{Z}_{r,G_i}$ , for  $i = 1, 2$ ; 3) Additive noise and temporally uncorrelated interference scenario,  $\mathbf{K}_R = \mu_R \mathbf{I}_{ML_S} + \nu_R \mathbf{I}_{L_S} \otimes \mathbf{Z}_{r,H} = \mathbf{I}_{L_S} \otimes (\mu_R \mathbf{I}_M + \nu_R \mathbf{Z}_{r,H})$  and  $\mathbf{K}_i = \mu_i \mathbf{I}_{N_i L_R} + \nu_i \mathbf{I}_{L_R} \otimes \mathbf{Z}_{r,G_i} = \mathbf{I}_{L_R} \otimes (\mu_i \mathbf{I}_{N_i} + \nu_i \mathbf{Z}_{r,G_i})$ , which implies  $\mathbf{K}_{r,R} = \mu_R \mathbf{I}_M + \nu_R \mathbf{Z}_{r,H}$  and  $\mathbf{K}_{r,i} = \mu_i \mathbf{I}_{N_i} + \nu_i \mathbf{Z}_{r,G_i}$ , for  $i = 1, 2$ , with  $\nu_R$  and  $\nu_i$  being the powers of the interfering users at the relay and at the source  $S_i$ , respectively,  $\mu_R$  and  $\mu_i$  being the variances of the additive noises at the relay and at the source  $S_i$ , respectively; 4) Additive noise and spatially uncorrelated interference scenario,  $\mathbf{K}_R = \mu_R \mathbf{I}_{ML_S} + \mathbf{Q}_R \otimes \mathbf{I}_M = (\mu_R \mathbf{I}_{L_S} + \mathbf{Q}_R) \otimes \mathbf{I}_M$ ,  $\mathbf{K}_i = \mu_i \mathbf{I}_{N_i L_R} + \mathbf{Q}_{G_i} \otimes \mathbf{I}_{N_i} = (\mu_i \mathbf{I}_{L_R} + \mathbf{Q}_{G_i}) \otimes \mathbf{I}_{N_i}$ , which implies  $\mathbf{K}_{r,R} = \mathbf{I}_M$  and  $\mathbf{K}_{r,i} = \mathbf{I}_{N_i}$ , for  $i = 1, 2$ .

<sup>2</sup>Here,  $\mathbf{G}_1$  and  $\mathbf{G}_2$  share a common transmit correlation matrix as they begin at the relay node.

For ease of the following training designs, we give the following definitions. The eigenvalue decomposition (EVD) of  $\mathbf{Z}_{t,H_i}$ ,  $\mathbf{Z}_{r,H}$ ,  $\mathbf{Z}_{t,G}$ , and  $\mathbf{Z}_{r,G_i}$  are given by

$$\mathbf{Z}_{a,b} = \mathbf{U}_{a,b} \boldsymbol{\Sigma}_{a,b} \mathbf{U}_{a,b}^H, \quad (13)$$

where  $a \in \{r, t\}$ ,  $b \in \{H, H_1, H_2, G, G_1, G_2\}$ ,  $\mathbf{U}_{a,b}$  denotes the unitary eigenvector matrix and  $\boldsymbol{\Sigma}_{a,b}$  is a diagonal matrix with  $[\boldsymbol{\Sigma}_{a,b}]_{n,n} = \sigma_{a,b,n}$  being the  $n$ -th eigenvalue of  $\mathbf{Z}_{a,b}$ . Accordingly, the singular value decomposition (SVD) of  $\mathbf{C}_{a,b}$  is denoted by  $\mathbf{C}_{a,b} = \mathbf{U}_{a,b} \boldsymbol{\Sigma}_{a,b}^{1/2} \tilde{\mathbf{U}}_{a,b}^H$  with  $\tilde{\mathbf{U}}_{a,b}$  representing a unitary matrix. The EVD of  $\mathbf{K}_{q,i}$  and  $\mathbf{K}_{r,i}$  is denoted by

$$\mathbf{K}_{a,b} = \mathbf{V}_{a,b} \boldsymbol{\Delta}_{a,b} \mathbf{V}_{a,b}^H, \quad a \in \{r, t\}, b \in \{1, 2, R\}, \quad (14)$$

where  $\mathbf{V}_{a,b}$  denotes the unitary eigenvector matrix,  $\boldsymbol{\Delta}_{a,b}$  is a diagonal matrix with  $[\boldsymbol{\Delta}_{a,b}]_{n,n} = \delta_{a,b,n}$  being the  $n$ -th eigenvalue of  $\mathbf{K}_{a,b}$ . As mentioned before, it is assumed that  $\mathbf{V}_{r,1} = \mathbf{U}_{r,G_1}$ ,  $\mathbf{V}_{r,2} = \mathbf{U}_{r,G_2}$  and  $\mathbf{V}_{r,R} = \mathbf{U}_{r,H}$ .

### III. CHANNEL ESTIMATION FOR TWO-WAY RELAY SYSTEMS

Following the proposed estimation scheme in Section II, we next obtain the channel estimates based on (2) and (8). For the estimation during the MAC phase, we rewrite (2) as

$$\begin{aligned} \mathbf{Y}_R &= \mathbf{C}_{r,H} \mathbf{W}_{H_1} \mathbf{C}_{t,H_1}^T \mathbf{S}_1 + \mathbf{C}_{r,H} \mathbf{W}_{H_2} \mathbf{C}_{t,H_2}^T \mathbf{S}_2 + \mathbf{N}_R \\ &= \mathbf{C}_{r,H} \mathbf{W}_H \mathbf{C}_{t,H}^T \mathbf{S} + \mathbf{N}_R, \end{aligned} \quad (15)$$

where  $\mathbf{W}_H \triangleq [\mathbf{W}_{H_1}, \mathbf{W}_{H_2}]$ ,  $\mathbf{C}_{t,H} \triangleq \text{Blkdiag}(\mathbf{C}_{t,H_1}, \mathbf{C}_{t,H_2})$ , and  $\mathbf{S} \triangleq [\mathbf{S}_1^T, \mathbf{S}_2^T]^T$ . Vectorizing  $\mathbf{Y}_R$  in (15) and applying the identity

$$\text{vec}(\mathbf{ABC}) = (\mathbf{C}^T \otimes \mathbf{A}) \text{vec}(\mathbf{B}), \quad (16)$$

we can rewrite (15) into

$$\mathbf{y}_R = (\mathbf{S}^T \mathbf{C}_{t,H} \otimes \mathbf{C}_{r,H}) \mathbf{w}_H + \mathbf{n}_R, \quad (17)$$

where  $\mathbf{y}_R \triangleq \text{vec}(\mathbf{Y}_R)$  and  $\mathbf{w}_H \triangleq \text{vec}(\mathbf{W}_H)$ . The estimation of  $\mathbf{w}_H$  based on the LMMSE criterion can be obtained as  $\hat{\mathbf{w}}_H = \mathbf{T}_R \mathbf{y}_R$ . The estimation matrix  $\mathbf{T}_R$  has the following form [24]

$$\mathbf{T}_R = \mathfrak{R}_{\mathbf{w}_H \mathbf{y}_R} \mathfrak{R}_{\mathbf{y}_R \mathbf{y}_R}^{-1}. \quad (18)$$

where  $\mathfrak{R}_{\mathbf{w}_H \mathbf{y}_R} \triangleq \mathcal{E}(\mathbf{w}_H \mathbf{y}_R^H) = \mathbf{C}_{t,H}^H \mathbf{S}^* \otimes \mathbf{C}_{r,H}^H$ ,  $\mathfrak{R}_{\mathbf{y}_R \mathbf{y}_R} \triangleq \mathcal{E}(\mathbf{y}_R \mathbf{y}_R^H) = (\mathbf{S}^T \mathbf{C}_{t,H} \otimes \mathbf{C}_{r,H}) (\mathbf{S}^T \mathbf{C}_{t,H} \otimes \mathbf{C}_{r,H})^H + \mathbf{K}_R = \mathbf{S}^T \mathbf{C}_{t,H} \mathbf{C}_{t,H}^H \mathbf{S}^* \otimes \mathbf{C}_{r,H} \mathbf{C}_{r,H}^H + \mathbf{K}_R$ . With  $\mathbf{H} \triangleq [\mathbf{H}_1, \mathbf{H}_2]$ , we define  $\mathbf{h} \triangleq \text{vec}(\mathbf{H}) = (\mathbf{C}_{t,H} \otimes \mathbf{C}_{r,H}) \mathbf{w}_H$  where  $\mathbf{C}_{r,H}$  and  $\mathbf{C}_{t,H}$  are defined in (4) and (15), respectively.

Then, the resulting estimation error, or mean-square-error (MSE),  $e_R$  can be derived as

$$\begin{aligned} e_R &= \mathcal{E}(\|\mathbf{h} - \hat{\mathbf{h}}\|_2^2) \\ &= \mathcal{E}(\text{Tr}[\mathbf{C}_{0,H} (\mathbf{w}_H - \hat{\mathbf{w}}_H) (\mathbf{w}_H - \hat{\mathbf{w}}_H)^H]) \\ &= \mathcal{E}(\text{Tr}[\mathbf{C}_{0,H} (\mathbf{w}_H - \mathbf{T}_R \mathbf{y}_R) (\mathbf{w}_H - \mathbf{T}_R \mathbf{y}_R)^H]), \end{aligned} \quad (19)$$

where  $\mathbf{C}_{0,H} \triangleq \mathbf{C}_{t,H}^H \mathbf{C}_{t,H} \otimes \mathbf{C}_{r,H}^H \mathbf{C}_{r,H}$ . Substituting  $\mathbf{T}_R$  into (19) and using the rule  $(\mathbf{I} + \mathbf{A}\mathbf{B})^{-1} = \mathbf{I} - \mathbf{A}(\mathbf{I} + \mathbf{B}\mathbf{A})^{-1}\mathbf{B}$ , we obtain the following more compact form for  $e_R$

$$e_R = \text{Tr} \left[ \mathbf{C}_{0,H} \left( \mathbf{I} + (\mathbf{S}^T \mathbf{C}_{t,H} \otimes \mathbf{C}_{r,H})^H \mathbf{K}_R^{-1} \right. \right. \\ \left. \left. \times (\mathbf{S}^T \mathbf{C}_{t,H} \otimes \mathbf{C}_{r,H}) \right)^{-1} \right]. \quad (20)$$

During the BC phase, the channel estimation should be based on the received signal in (8). By vectorizing  $\mathbf{Y}_i$  in (8), we have

$$\mathbf{y}_i = (\mathbf{S}_R^T \mathbf{C}_{t,G} \otimes \mathbf{C}_{r,G_i}) \mathbf{w}_{G_i} + \mathbf{n}_i, \quad i = 1, 2 \quad (21)$$

where  $\mathbf{y}_i \triangleq \text{vec}(\mathbf{Y}_i)$  and  $\mathbf{w}_{G_i} \triangleq \text{vec}(\mathbf{W}_{G_i})$ . Similar to steps above, the estimation MSE of  $\mathbf{g}_i$  with  $\mathbf{g}_i \triangleq \text{vec}(\mathbf{G}_i)$  can be obtained as

$$e_i = \mathcal{E} (\|\mathbf{g}_i - \hat{\mathbf{g}}_i\|_2^2) \\ = \mathcal{E} (\text{Tr} [\mathbf{C}_{0,G_i} (\mathbf{w}_{G_i} - \hat{\mathbf{w}}_{G_i}) (\mathbf{w}_{G_i} - \hat{\mathbf{w}}_{G_i})^H]) \\ = \mathcal{E} (\text{Tr} [\mathbf{C}_{0,G_i} (\mathbf{w}_H - \mathbf{T}_i \mathbf{y}_i) (\mathbf{w}_H - \mathbf{T}_i \mathbf{y}_i)^H]), \quad (22)$$

where  $\mathbf{C}_{0,G_i} \triangleq \mathbf{C}_{t,G}^H \mathbf{C}_{t,G} \otimes \mathbf{C}_{r,G_i}^H \mathbf{C}_{r,G_i}$  and

$$\mathbf{T}_i = \mathfrak{R}_{\mathbf{w}_{G_i} \mathbf{y}_i}^{-1} \mathfrak{R}_{\mathbf{y}_i \mathbf{y}_i}^{-1}, \quad i = 1, 2. \quad (23)$$

In (23),  $\mathfrak{R}_{\mathbf{w}_{G_i} \mathbf{y}_i} = \mathcal{E}(\mathbf{w}_{G_i} \mathbf{y}_i^H) = \mathbf{C}_{t,G}^H \mathbf{S}_R^* \otimes \mathbf{C}_{r,G_i}^H$ ,  $\mathfrak{R}_{\mathbf{y}_i \mathbf{y}_i} = \mathcal{E}(\mathbf{y}_i \mathbf{y}_i^H) = (\mathbf{S}_R^T \mathbf{C}_{t,G} \otimes \mathbf{C}_{r,G_i}) (\mathbf{S}_R^T \mathbf{C}_{t,G} \otimes \mathbf{C}_{r,G_i})^H + \mathbf{K}_i = \mathbf{S}_R^T \mathbf{C}_{t,G} \mathbf{C}_{t,G}^H \mathbf{S}_R^* \otimes \mathbf{C}_{r,G_i} \mathbf{C}_{r,G_i}^H + \mathbf{K}_i$ . By substituting  $\mathbf{T}_i$  into (22), we obtain

$$e_i = \text{Tr} \left[ \mathbf{C}_{0,G_i} \left( \mathbf{I} + (\mathbf{S}_R^T \mathbf{C}_{t,G} \otimes \mathbf{C}_{r,G_i})^H \mathbf{K}_i^{-1} \right. \right. \\ \left. \left. \times (\mathbf{S}_R^T \mathbf{C}_{t,G} \otimes \mathbf{C}_{r,G_i}) \right)^{-1} \right]. \quad (24)$$

#### IV. TRAINING SEQUENCE DESIGN FOR MAC PHASE

In this section, the design of the training sequences for the MAC phase is analyzed. Namely, we shall optimize the training sequences  $\mathbf{S}_1$  and  $\mathbf{S}_2$  subject to two source power constraints to minimize the total estimation MSE, i.e.,  $e_R$  in (19). The corresponding training sequence optimization problem can be formulated as

$$\min_{\mathbf{S}_1, \mathbf{S}_2} e_R \text{ in (19)} \\ \text{s.t. } \text{Tr}(\mathbf{S}_i \mathbf{S}_i^H) \leq \tau_i, \quad i = 1, 2. \quad (25)$$

Before solving (25), we first introduce the following lemma that deals with the minimum length of  $\mathbf{S}$ ,  $L_S$ .

*Lemma 1:* For an arbitrary  $\mathbf{K}_{q,R}$ , for the total MSE to tend to zero as the power at each of source tends to infinity, the minimum length of the source training sequence must be  $L_S = N_1 + N_2$ . Otherwise, when  $L_S < N_1 + N_2$ , the total MSE is lower bounded by  $\sum_{n=1}^M \sigma_{r,H,n} \sum_{m=L_S+1}^{N_1+N_2} \sigma_{t,H,m}$ , where  $\sigma_{t,H,n}$  is the  $n$ -th element of  $\lambda(\mathbf{Z}_{r,H})$  with  $\mathbf{Z}_{t,H} = \text{Blkdiag}(\mathbf{Z}_{t,H_1}, \mathbf{Z}_{t,H_2})$  as the power at each of source tends to infinity.

For  $\mathbf{K}_{q,R} = q\mathbf{I}$  and with finite power at the source nodes, the minimum length of the training sequences transmitted from the source nodes must be  $r \leq N_1 + N_2$ . It is important to note that  $r$  denotes the rank of the optimal solution of  $\mathbf{S} \in \mathbb{C}^{(N_1+N_2) \times L_S}$ , where  $L_S$  is an arbitrary integer satisfying  $L_S \geq N_1 + N_2$  in (25).<sup>3</sup>

*Proof:* See Appendix A.

Next, we seek to solve the non-convex optimization problem in (25) with respect to  $\mathbf{S}_1$  and  $\mathbf{S}_2$ . Although the objective function in (25) has a similar form to that of point-to-point systems, there are two power constraints in (25) that make the problem of solving this non-convex optimization problem more difficult than that of point-to-point systems in [17] and [18]. To proceed, we first note that  $e_R$  in (20) can be obtained by substituting (18) into (19). Thus, to make the problem tractable, we propose an iterative algorithm, which decouples the primal problem into two sub-problems such that we can solve each of them in an alternating manner. Let us rewrite (20) into the following form

$$e_R = \mathcal{E} (\text{Tr} [\mathbf{C}_{0,H} (\mathbf{w}_H - \mathbf{T}_R \mathbf{y}_R) (\mathbf{w}_H - \mathbf{T}_R \mathbf{y}_R)^H]) \\ = \text{Tr} \left[ \mathbf{C}_{0,H} - (\mathbf{S}^T \mathbf{C}_{t,H} \otimes \mathbf{C}_{r,H})^H \mathbf{T}_R^H \mathbf{C}_{0,H}^H - \mathbf{C}_{0,H} \mathbf{T}_R \right. \\ \left. \times (\mathbf{S}^T \mathbf{C}_{t,H} \otimes \mathbf{C}_{r,H}) + \mathbf{C}_{0,H} \mathbf{T}_R (\mathbf{S}^T \mathbf{C}_{t,H} \otimes \mathbf{C}_{r,H}) \right. \\ \left. \times (\mathbf{S}^T \mathbf{C}_{t,H} \otimes \mathbf{C}_{r,H})^H \mathbf{T}_R^H + \mathbf{C}_{0,H} \mathbf{T}_R \mathbf{K}_R \mathbf{T}_R^H \right]. \quad (26)$$

Then, the optimization problem in (25) is equivalent to

$$\min_{\mathbf{T}_R, \mathbf{S}_1, \mathbf{S}_2} e_R \text{ in (26)} \\ \text{s.t. } \text{Tr}(\mathbf{S}_i \mathbf{S}_i^H) \leq \tau_i, \quad i = 1, 2. \quad (27)$$

In the first subproblem, we intend to optimize the LMMSE estimator matrix  $\mathbf{T}_R$  for a given  $\mathbf{S}_1$  and  $\mathbf{S}_2$ . Since  $\mathbf{T}_R$  is not related to the the power constraint, the problem simplifies to an unconstrained optimization problem given by

$$\min_{\mathbf{T}_R} e_R \text{ in (26)}. \quad (28)$$

Given that (28) is convex with respect to  $\mathbf{T}_R$ , the optimal  $\mathbf{T}_R$  can be obtained as given in (18).

In the second subproblem, the training sequences  $\mathbf{S}_1$  and  $\mathbf{S}_2$  need to be optimized for a given  $\mathbf{T}_R$  by solving the following optimization problem

$$\min_{\mathbf{S}_1, \mathbf{S}_2} e_R \text{ in (26)} \\ \text{s.t. } \text{Tr}(\mathbf{S}_i \mathbf{S}_i^H) \leq \tau_i, \quad i = 1, 2. \quad (29)$$

We next show that the optimization problem in (29) can be transformed into a convex quadratically constrained quadratic

<sup>3</sup>We note that for an arbitrary  $\mathbf{K}_{q,R}$  without a form of  $q\mathbf{I}$ , it is difficult to derive the minimum length of the training sequence with finite source powers as it highly depends on the structure of  $\mathbf{K}_{q,R}$ .

programmable (QCQP) problem [25]. To achieve this goal, we first reformulate the last term in (26) as

$$\begin{aligned} & \text{Tr} \left[ \mathbf{C}_{0,H} \mathbf{T}_R (\mathbf{S}^T \mathbf{C}_{i,H} \otimes \mathbf{C}_{r,H}) (\mathbf{S}^T \mathbf{C}_{i,H} \otimes \mathbf{C}_{r,H})^H \mathbf{T}_R^H \right] \\ & \stackrel{(a)}{=} \text{Tr} \left[ \mathbf{T}_R^H \mathbf{C}_{0,H} \mathbf{T}_R (\mathbf{S}^T \otimes \mathbf{I}_M) \right. \\ & \quad \left. \times (\mathbf{C}_{i,H} \mathbf{C}_{i,H}^H \otimes \mathbf{C}_{r,H} \mathbf{C}_{r,H}^H) (\mathbf{S}^* \otimes \mathbf{I}_M) \right] \\ & \stackrel{(b)}{=} \text{vec}(\mathbf{S} \otimes \mathbf{I}_M)^H (\mathbf{T}_R^H \mathbf{C}_{0,H} \mathbf{T}_R \otimes \mathbf{C}_{ir}^T) \text{vec}(\mathbf{S} \otimes \mathbf{I}_M) \\ & \stackrel{(c)}{=} \mathbf{s}^H \mathbf{E}^H (\mathbf{T}_R^H \mathbf{C}_{0,H} \mathbf{T}_R \otimes \mathbf{C}_{ir}^T) \mathbf{E} \mathbf{s}, \end{aligned} \quad (30)$$

where  $\mathbf{C}_{ir} \triangleq \mathbf{C}_{i,H} \mathbf{C}_{i,H}^H \otimes \mathbf{C}_{r,H} \mathbf{C}_{r,H}^H$ ,  $\mathbf{s} \triangleq \text{vec}(\mathbf{S})$ ,  $\mathbf{E} \triangleq \text{Blkdiag}(\tilde{\mathbf{E}}_{(1)}, \tilde{\mathbf{E}}_{(2)}, \dots, \tilde{\mathbf{E}}_{(L_S)})$ ,  $\tilde{\mathbf{E}}_{(i)} = \tilde{\mathbf{E}}$ ,  $\tilde{\mathbf{E}} \triangleq [\tilde{\mathbf{E}}_{(1)}; \tilde{\mathbf{E}}_{(2)}; \dots; \tilde{\mathbf{E}}_{(M)}]$ ,  $\tilde{\mathbf{E}}_{(i)} \triangleq \text{Blkdiag}(\underbrace{\mathbf{e}_i, \mathbf{e}_i, \dots, \mathbf{e}_i}_{N_1+N_2 \text{ elements}}, \underbrace{1}_{i\text{-th element}}, \dots, 0)^T$ .

In (30), Eq. (a) is obtained by using the circular property  $\text{Tr}\{\mathbf{A}\mathbf{B}\} = \text{Tr}\{\mathbf{B}\mathbf{A}\}$  and the matrix identity

$$(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = \mathbf{A}\mathbf{C} \otimes \mathbf{B}\mathbf{D}, \quad (31)$$

Eq. (b) is obtained by using the identity  $\text{Tr}(\mathbf{A}\mathbf{B}\mathbf{C}\mathbf{D}) = \text{vec}(\mathbf{D})^T (\mathbf{A} \otimes \mathbf{C}^T) \text{vec}(\mathbf{B}^T)$  and  $(\mathbf{A} \otimes \mathbf{B})^H = \mathbf{A}^H \otimes \mathbf{B}^H$ , and Eq. (c) is obtained by using  $\text{vec}(\mathbf{S} \otimes \mathbf{I}_M) = \mathbf{E}\mathbf{s}$ . Similarly, the term  $\text{Tr}[\mathbf{C}_{0,H} \mathbf{T}_R (\mathbf{S}^T \mathbf{C}_{i,H} \otimes \mathbf{C}_{r,H})]$  can be expressed as

$$\begin{aligned} & \text{Tr} [\mathbf{C}_{0,H} \mathbf{T}_R (\mathbf{S}^T \mathbf{C}_{i,H} \otimes \mathbf{C}_{r,H})] \\ & = \text{Tr} [(\mathbf{S} \otimes \mathbf{I}_M)^T (\mathbf{C}_{i,H} \otimes \mathbf{C}_{r,H}) \mathbf{C}_{0,H} \mathbf{T}_R] \\ & = \text{vec}(\mathbf{S} \otimes \mathbf{I}_M)^T \text{vec}(\mathbf{C}_T) \\ & = \text{vec}(\mathbf{C}_T)^T \mathbf{E} \mathbf{s}, \end{aligned} \quad (32)$$

where  $\mathbf{C}_T \triangleq (\mathbf{C}_{i,H} \otimes \mathbf{C}_{r,H}) \mathbf{C}_{0,H} \mathbf{T}_R$ . To obtain (32), we use

$$\text{Tr}(\mathbf{A}^T \mathbf{B}) = \text{vec}(\mathbf{A})^T \text{vec}(\mathbf{B}). \quad (33)$$

The source power constraint in (29) can be rewritten as

$$\text{Tr}(\mathbf{S}_i \mathbf{S}_i^H) = \text{Tr}(\mathbf{E}_i \mathbf{S} \mathbf{S}^H), \quad (34)$$

where  $\mathbf{E}_1 \triangleq \text{Blkdiag}(\mathbf{I}_{N_1}, \mathbf{0}_{N_2 \times N_2})$  and  $\mathbf{E}_2 \triangleq \text{Blkdiag}(\mathbf{0}_{N_1 \times N_1}, \mathbf{I}_{N_2})$ . Based on the property that  $\text{Tr}(\mathbf{A}\mathbf{B}\mathbf{C}\mathbf{D}) = \text{vec}(\mathbf{D}^T)^T (\mathbf{C}^T \otimes \mathbf{A}) \text{vec}(\mathbf{B})$ , (34) can be further modified as

$$\text{Tr}(\mathbf{S}_i \mathbf{S}_i^H) = \mathbf{s}^H (\mathbf{I} \otimes \mathbf{E}_i) \mathbf{s}. \quad (35)$$

According to (30), (32), and (35), the optimization problem in (29) can be transformed into

$$\begin{aligned} \min_{\mathbf{s}} & \mathbf{s}^H \mathbf{E}^H (\mathbf{T}_R^H \mathbf{C}_{0,H} \mathbf{T}_R \otimes \mathbf{C}_{ir}^T) \mathbf{E} \mathbf{s} - 2\Re(\text{vec}(\mathbf{C}_T)^T \mathbf{E} \mathbf{s}) \\ \text{s.t.} & \mathbf{s}^H (\mathbf{I} \otimes \mathbf{E}_i) \mathbf{s} \leq \tau_i, \quad i = 1, 2. \end{aligned} \quad (36)$$

Since both  $\mathbf{E}^H (\mathbf{T}_R^H \mathbf{T}_R \otimes \mathbf{C}_{ir}^T) \mathbf{E}$  and  $\mathbf{I} \otimes \mathbf{E}_i$  are positive semidefinite matrices, we conclude that the optimization problem in (36) is a convex QCQP problem, which can be easily solved by ap-

plying the available software package. In summary, we outline the proposed iterative training design algorithm as follows:

---

**Algorithm 1**

---

- **Initialize**  $\mathbf{S}_1, \mathbf{S}_2$
  - **Repeat**
    - Update the LMMSE estimator matrix  $\mathbf{T}_R$  using (18) for fixed  $\mathbf{S}_1$  and  $\mathbf{S}_2$ ;
    - For a fixed  $\mathbf{T}_R$ , solve the convex QCQP problem in (36) to get the optimal  $\mathbf{S}_1$  and  $\mathbf{S}_2$ ;
  - **Until** The difference between the MSE from one iteration to another is smaller than a certain predetermined threshold.
- 

*Theorem 1:* The proposed iterative precoding design in Algorithm 1 is convergent and the limit point of the iteration is a stationary point of (27).

*Proof:* We first prove that Algorithm 1 is convergent. To this end, we show that the sequence of the updates of the value of the objective function in (27) is convergent. Let us denote the value of the objective function in (27) by MSE and define  $\mathbf{S} \triangleq [\mathbf{S}_1^T, \mathbf{S}_2^T]^T$  as in (15). We also denote  $\mathbf{T}_R^{(n)}$  and  $\mathbf{S}^{(n)}$  as the  $n$ -th updates of  $\mathbf{T}_R$  and  $\mathbf{S}$ , respectively.  $\text{MSE}^{(2n-1)}$  is used to represent the value of the  $(2n-1)$ -th update of MSE when  $\mathbf{T}_R = \mathbf{T}_R^{(n)}$  and  $\mathbf{S} = \mathbf{S}^{(n-1)}$ . Similarly,  $\text{MSE}^{(2n)}$  indicates the value of the  $2n$ -th update of MSE when  $\mathbf{T}_R = \mathbf{T}_R^{(n)}$  and  $\mathbf{S} = \mathbf{S}^{(n)}$ . Consider that the solutions of  $\mathbf{T}_R$  and  $\mathbf{S}$  from subproblems (28) and (29), respectively, are optimal. For each update of  $\mathbf{T}_R$  or  $\mathbf{S}$ , the value of MSE always decreases, i.e.,  $\text{MSE}^{(n)} \leq \text{MSE}^{(n-1)}$ . This further indicates that the sequence  $\{\text{MSE}^{(n)}\}_{n=1}^\infty$  decreases monotonically. Moreover, as the value of MSE is lower bounded (at least by zero), it can be concluded that the sequence  $\{\text{MSE}^{(n)}\}_{n=1}^\infty$  is convergent based on the monotone convergence theorem in [26], i.e., for an arbitrary  $\epsilon$ , we can always find an  $N$  such that  $|\text{MSE}^{(n)} - \text{MSE}^{(m)}| \leq \epsilon$  when  $n > m \geq N$ .

Now we prove the convergence of the sequences  $\{\mathbf{T}_R^{(n)}\}_{n=1}^\infty$  and  $\{\mathbf{S}^{(n)}\}_{n=1}^\infty$ . We first prove the convergence of  $\{\mathbf{S}^{(n)}\}_{n=1}^\infty$  by showing  $\lim_{n \rightarrow \infty} \mathbf{S}^{(n)} - \mathbf{S}^{(n-1)} = \mathbf{X}^{(n)} = \mathbf{0}$  using a contradiction. Assume  $\mathbf{X}^{(n)} \neq \mathbf{0}$  with  $n \rightarrow \infty$ , we then have  $|\text{MSE}^{(2n-1)} - \text{MSE}^{(2n)}| = |\text{Tr}[(\mathbf{X}^{(n)T} \mathbf{C}_{i,H} \otimes \mathbf{C}_{r,H})^H \mathbf{T}_R^{(n)H} \mathbf{C}_{0,H}^H + \mathbf{C}_{0,H} \mathbf{T}_R^{(n)} (\mathbf{X}^{(n)T} \mathbf{C}_{i,H} \otimes \mathbf{C}_{r,H}) + \mathbf{C}_{0,H} \mathbf{T}_R^{(n)} (\mathbf{S}^{(n)T} \mathbf{Z}_{i,H} \mathbf{S}^{(n)*} \otimes \mathbf{Z}_{r,H}) \mathbf{T}_R^{(n)H} - \mathbf{C}_{0,H} \mathbf{T}_R^{(n)} (\mathbf{S}^{(n-1)T} \mathbf{Z}_{i,H} \mathbf{S}^{(n-1)*} \otimes \mathbf{Z}_{r,H}) \mathbf{T}_R^{(n)H}]|$ , which can not approach zero even as  $n \rightarrow \infty$ . This contradicts the result that the sequence  $\{\text{MSE}^{(n)}\}_{n=1}^\infty$  is convergent. This implies that the sequence  $\{\mathbf{S}^{(n)}\}_{n=1}^\infty$  is convergent. Repeating the same argument, we can also prove the convergence of the sequence  $\{\mathbf{T}_R^{(n)}\}_{n=1}^\infty$ . As a convergent sequence always has a unique limit point, we, thus, obtain that the sequences  $\{\mathbf{T}_R^{(n)}\}_{n=1}^\infty$  and  $\{\mathbf{S}^{(n)}\}_{n=1}^\infty$  always have limit points, which are denoted by  $\tilde{\mathbf{T}}_R$  and  $\tilde{\mathbf{S}} = [\tilde{\mathbf{S}}_1^T, \tilde{\mathbf{S}}_2^T]^T$ , respectively.

We now prove that the limit point is a stationary point of (27). To proceed, we denote  $\bar{\mathbf{Y}} = \{\bar{\mathbf{T}}_R, \bar{\mathbf{S}}_1, \bar{\mathbf{S}}_2\}$ . Since at the limit point  $\bar{\mathbf{Y}}$ ,  $\bar{\mathbf{T}}_R$  is the local minimizer of subproblem (28), which implies that  $\bar{\mathbf{T}}_R$  satisfies the following Karush-Kuhn-Tucker (KKT) condition [25]

$$\frac{\partial e_R(\bar{\mathbf{S}}_1, \bar{\mathbf{S}}_2)}{\partial \mathbf{T}_R} \Big|_{\mathbf{T}_R = \bar{\mathbf{T}}_R} = \mathbf{0}, \quad (37)$$

where  $e_R(\bar{\mathbf{S}}_1, \bar{\mathbf{S}}_2)$  denotes the function  $e_R$  with  $\mathbf{S}_1$  and  $\mathbf{S}_2$  being evaluated at  $\bar{\mathbf{S}}_1$  and  $\bar{\mathbf{S}}_2$ , respectively. Similarly,  $\bar{\mathbf{S}}_i$ , for  $i = 1, 2$ , is the local minimizer of subproblem (29), which satisfies the following KKT conditions

$$\begin{aligned} \frac{\partial e_R(\bar{\mathbf{T}}_R)}{\partial \mathbf{S}_i} \Big|_{\mathbf{S}_i = \bar{\mathbf{S}}_i} &= \mathbf{0}, \\ \lambda_i (\text{Tr}(\bar{\mathbf{S}}_i \bar{\mathbf{S}}_i^H) - \tau_i) &= 0, \quad i = 1, 2 \end{aligned} \quad (38)$$

where  $\lambda_i$  is the Lagrangian multiplier associated with the source power constraints. By summing up the KKT conditions given in (37) and (38), it can be concluded that the limit point  $\bar{\mathbf{Y}}$  satisfies the KKT conditions of the primal problem in (27), which further means that  $\bar{\mathbf{Y}}$  is a stationary point of (27).

To this point, it is shown that the joint source training design can be solved via *Algorithm 1*. In the following, we illustrate that for some special cases, the optimal solution of (25) can be obtained in closed-form.

#### A. When $\mathbf{K}_{r,R} = \mathbf{Z}_{r,H}$

We first consider the case with  $\mathbf{K}_{r,R} = \mathbf{Z}_{r,H}$ . This case is applicable in the scenario where the disturbance is dominated by the interference from neighboring users as shown in [21]. Accordingly, the LMMSE estimator given in (18) can be rewritten as

$$\begin{aligned} \mathbf{T}_R &= [\mathbf{C}_{t,H}^H \mathbf{S}^* \otimes \mathbf{C}_{r,H}^H] \\ &\quad \times [\mathbf{S}^T \mathbf{C}_{t,H} \mathbf{C}_{t,H}^H \mathbf{S}^* \otimes \mathbf{C}_{r,H} \mathbf{C}_{r,H}^H + \mathbf{K}_R]^{-1} \\ &\stackrel{(a)}{=} [\mathbf{C}_{t,H}^H \mathbf{S}^* \otimes \mathbf{C}_{r,H}^H] \\ &\quad \times [(\mathbf{S}^T \mathbf{C}_{t,H} \mathbf{C}_{t,H}^H \mathbf{S}^* + \mathbf{K}_{q,R})^{-1} \otimes \mathbf{Z}_{r,H}^{-1}] \\ &= \mathbf{T}_{R,1} \otimes \mathbf{C}_{r,H}^{-1}, \end{aligned} \quad (39)$$

where  $\mathbf{T}_{R,1} = \mathbf{C}_{t,H}^H \mathbf{S}^* (\mathbf{S}^T \mathbf{C}_{t,H} \mathbf{C}_{t,H}^H \mathbf{S}^* + \mathbf{K}_{q,R})^{-1}$  and in obtaining Eq. (a), we have used the fact  $\mathbf{K}_R = \mathbf{K}_{q,R} \otimes \mathbf{K}_{r,R} = \mathbf{K}_{q,R} \otimes \mathbf{Z}_{r,H}$ . The new form of  $\mathbf{T}_R$  in (39) further leads to  $\mathbf{T}_{R,1}^H \mathbf{C}_{0,H} \mathbf{T}_R = (\mathbf{T}_{R,1} \otimes \mathbf{C}_{r,H}^{-1})^H (\mathbf{C}_{t,H}^H \mathbf{C}_{t,H} \otimes \mathbf{C}_{r,H}^H \mathbf{C}_{r,H}) (\mathbf{T}_{R,1} \otimes \mathbf{C}_{r,H}^{-1}) = \mathbf{T}_{R,1}^H \mathbf{C}_{t,H}^H \mathbf{C}_{t,H} \mathbf{T}_{R,1} \otimes \mathbf{I}_M$ , and

$$\begin{aligned} &\text{Tr} [\mathbf{C}_{0,H} \mathbf{T}_R (\mathbf{S}^T \mathbf{C}_{t,H} \otimes \mathbf{C}_{r,H}) (\mathbf{S}^T \mathbf{C}_{t,H} \otimes \mathbf{C}_{r,H})^H \mathbf{T}_R^H] \\ &= \text{Tr} [(\mathbf{S}^T \mathbf{C}_{t,H} \mathbf{C}_{t,H}^H \mathbf{S}^* \otimes \mathbf{C}_{r,H} \mathbf{C}_{r,H}^H) \\ &\quad \times (\mathbf{T}_{R,1}^H \mathbf{C}_{t,H}^H \mathbf{C}_{t,H} \mathbf{T}_{R,1} \otimes \mathbf{I}_M)] \\ &\stackrel{(a)}{=} \text{Tr}(\mathbf{Z}_{r,H}) \text{Tr} [\mathbf{S}^T \mathbf{C}_{t,H} \mathbf{C}_{t,H}^H \mathbf{S}^* \mathbf{T}_{R,1}^H \mathbf{C}_{t,H}^H \mathbf{C}_{t,H} \mathbf{T}_{R,1}] \\ &\stackrel{(b)}{=} \text{Tr}(\mathbf{Z}_{r,H}) \sum_{i=1}^2 \text{Tr} (\mathbf{S}_i^T \mathbf{Z}_{t,H_i} \mathbf{S}_i^* \mathbf{T}_{R,1}^H \mathbf{C}_{t,H}^H \mathbf{C}_{t,H} \mathbf{T}_{R,1}). \end{aligned} \quad (40)$$

In (40), Eq. (a) is obtained with  $\text{Tr}(\mathbf{A} \otimes \mathbf{B}) = \text{Tr}(\mathbf{A})\text{Tr}(\mathbf{B})$ , and Eq. (b) is derived based on the fact that  $\mathbf{C}_{t,H}$  is a block diagonal matrix as shown in (15). In addition, the term  $\text{Tr}[\mathbf{C}_{0,H} \mathbf{T}_R (\mathbf{S}^T \mathbf{C}_{t,H} \otimes \mathbf{C}_{r,H})]$  can be reexpressed as

$$\begin{aligned} &\text{Tr} [\mathbf{C}_{0,H} \mathbf{T}_R (\mathbf{S}^T \mathbf{C}_{t,H} \otimes \mathbf{C}_{r,H})] \\ &= \text{Tr} [(\mathbf{S}^T \mathbf{C}_{t,H} \mathbf{C}_{t,H}^H \mathbf{C}_{t,H} \mathbf{T}_{R,1}) \otimes \mathbf{C}_{r,H} \mathbf{C}_{r,H}^H] \\ &= \text{Tr}(\mathbf{Z}_{r,H}) \sum_{i=1}^2 \text{Tr} (\mathbf{S}_i^T \mathbf{Z}_{t,H_i} \mathbf{C}_{t,H_i} \mathbf{T}_{R,1,i}), \end{aligned} \quad (41)$$

where  $\mathbf{T}_{R,1,1} \triangleq \mathbf{T}_{R,1}(1 : N_1, :)$  and  $\mathbf{T}_{R,1,2} \triangleq \mathbf{T}_{R,1}(N_1 + 1 : N_1 + N_2, :)$ . Based on (40) and (41), (29) is equivalent to the following optimization problem

$$\begin{aligned} \min_{\mathbf{S}_1, \mathbf{S}_2} &\sum_{i=1}^2 \{ \text{Tr} (\mathbf{S}_i^T \mathbf{Z}_{t,H_i} \mathbf{S}_i^* \mathbf{T}_{R,1}^H \mathbf{C}_{t,H}^H \mathbf{C}_{t,H} \mathbf{T}_{R,1}) \\ &\quad - \text{Tr} (\mathbf{S}_i^T \mathbf{Z}_{t,H_i} \mathbf{C}_{t,H_i} \mathbf{T}_{R,1,i}) \\ &\quad - \text{Tr} (\mathbf{T}_{R,1,i}^H \mathbf{C}_{t,H_i}^H \mathbf{Z}_{t,H_i} \mathbf{S}_i^*) \} \\ \text{s.t.} &\text{Tr}(\mathbf{S}_i \mathbf{S}_i^H) \leq \tau_i, \quad i = 1, 2. \end{aligned} \quad (42)$$

Although the optimization in (42) is convex and can be converted to a convex QCQP problem similar to (36). A closer investigation shows that the optimization problem in (42) can be more easily solved compared to (29). This follows from the fact that the training matrices  $\mathbf{S}_1$  and  $\mathbf{S}_2$  are fully decoupled in the objective function. As shown below, this specific condition of the optimization in (42) allows us to find a closed-form solution for this problem via the KKT conditions. Such an approach significantly reduces the complexity of solving this optimization problem compared to using the QCQP approach.

The Lagrangian function of (42) is first derived as

$$\begin{aligned} \mathcal{L} &= \sum_{i=1}^2 \{ \text{Tr} (\mathbf{S}_i^T \mathbf{Z}_{t,H_i} \mathbf{S}_i^* \mathbf{T}_{R,1}^H \mathbf{C}_{t,H}^H \mathbf{C}_{t,H} \mathbf{T}_{R,1}) \\ &\quad - \text{Tr} (\mathbf{S}_i^T \mathbf{Z}_{t,H_i} \mathbf{C}_{t,H_i} \mathbf{T}_{R,1,i}) - \text{Tr} (\mathbf{T}_{R,1,i}^H \mathbf{C}_{t,H_i}^H \mathbf{Z}_{t,H_i} \mathbf{S}_i^*) \} \\ &\quad + \sum_{i=1}^2 \lambda_i [\text{Tr}(\mathbf{S}_i \mathbf{S}_i^H) - \tau_i], \end{aligned}$$

where  $\lambda_i$  is the Lagrangian multiplier associated with the power constraint at the source  $S_i$ . The KKT conditions for (42) can be derived as follows [25]

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial \mathbf{S}_i^*} &= (\mathbf{T}_{R,1}^H \mathbf{C}_{t,H}^H \mathbf{C}_{t,H} \mathbf{T}_{R,1} \mathbf{S}_i^T \mathbf{Z}_{t,H_i})^T \\ &\quad - (\mathbf{T}_{R,1,i}^H \mathbf{C}_{t,H_i}^H \mathbf{Z}_{t,H_i})^T + \lambda_i \mathbf{S}_i = \mathbf{0}, \end{aligned} \quad (43a)$$

$$\lambda_i [\text{Tr}(\mathbf{S}_i \mathbf{S}_i^H) - \tau_i] = 0, \quad (43b)$$

$$\text{Tr}(\mathbf{S}_i \mathbf{S}_i^H) \leq \tau_i, \quad i = 1, 2. \quad (43c)$$

Based on the KKT conditions shown in (43a) and by using (16), the optimal  $\mathbf{s}_i \triangleq \text{vec}(\mathbf{S}_i)$  can be obtained as

$$\mathbf{s}_i = [\mathbf{X}_{s,1} \otimes \mathbf{X}_{s,2,i} + \lambda_i \mathbf{I}]^{-1} \mathbf{x}_{s,3,i}, \quad (44)$$

where  $\mathbf{X}_{s,1} \triangleq \mathbf{T}_{R,1}^H \mathbf{C}_{t,H}^H \mathbf{C}_{t,H} \mathbf{T}_{R,1}$ ,  $\mathbf{X}_{s,2,i} \triangleq \mathbf{Z}_{t,H_i}^T$ , and  $\mathbf{x}_{s,3,i} \triangleq \text{vec}((\mathbf{T}_{R,1,i}^H \mathbf{C}_{t,H_i}^H \mathbf{Z}_{t,H_i}^H)^T)$ . The optimal  $\lambda_i$  in (44) can be zero or should be chosen to activate the power constraint in (43c). For the case where  $\lambda_i \neq 0$ , the following lemma is introduced.

*Lemma 2:* The function  $g(\lambda_i) = \text{Tr}\{\mathbf{S}_i \mathbf{S}_i^H\} = \text{Tr}\{\mathbf{s}_i \mathbf{s}_i^H\}$ , with  $\mathbf{s}_i$  defined above (44), monotonically decreases with respect to  $\lambda_i$  and the optimal  $\lambda_i$  is upper bounded by  $\sqrt{\frac{\sigma_{s,3,i}}{\tau_i}} - \sigma_{s,min,i}$ . Here,  $\sigma_{s,min,i}$  denotes the smallest eigenvalue of  $\mathbf{X}_{s,1} \otimes \mathbf{X}_{s,2,i}$  and  $\sigma_{s,3,i} = \|\mathbf{x}_{s,3,i}\|_2^2$ .

*Proof:* See Appendix B. ■

By applying Lemma 2, the optimal  $\lambda_i$  that meets the condition  $\text{Tr}\{\mathbf{S}_i \mathbf{S}_i^H\} = \tau_i$  can be readily obtained via the bisection search algorithm.

### B. When $\mathbf{K}_{q,R} = q\mathbf{I}_M$

This scenario corresponds to the practical case, where the disturbance consists of both the additive white Gaussian noise and the temporally uncorrelated interference. Similar to the derivations in Appendix A, the total MSE can be derived as

$$\begin{aligned} e_R &= \text{Tr} \left[ \mathbf{C}_{0,H} \left( \mathbf{I}_{M(N_1+N_2)} + (\mathbf{S}^T \mathbf{C}_{t,H} \otimes \mathbf{C}_{r,H})^H \mathbf{K}_R^{-1} \right. \right. \\ &\quad \left. \left. \times (\mathbf{S}^T \mathbf{C}_{t,H} \otimes \mathbf{C}_{r,H}) \right)^{-1} \right] \\ &= \text{Tr} \left[ \mathbf{C}_{0,H} \left( \mathbf{I}_{M(N_1+N_2)} + \frac{1}{q} \mathbf{C}_{t,H}^H \mathbf{S}^* \mathbf{S}^T \mathbf{C}_{t,H} \right. \right. \\ &\quad \left. \left. \otimes \mathbf{C}_{r,H}^H \mathbf{K}_{r,R}^{-1} \mathbf{C}_{r,H} \right)^{-1} \right] \\ &= \sum_{n=1}^M \sigma_{r,H,n} \text{Tr} \left[ \left( \mathbf{Z}_{t,H}^{-1} + \alpha_n \mathbf{S}^* \mathbf{S}^T \right)^{-1} \right], \end{aligned}$$

where  $\alpha_n = \frac{\sigma_{r,H,n}}{q\sigma_{r,R,n}}$ . Subsequently, the optimization problem in (25) can be rewritten as

$$\begin{aligned} \min_{\mathbf{S}_1, \mathbf{S}_2} \quad & \sum_{n=1}^M \sigma_{r,H,n} \text{Tr} \left[ \left( \mathbf{Z}_{t,H}^{-1} + \alpha_n \mathbf{S}^* \mathbf{S}^T \right)^{-1} \right] \\ \text{s.t.} \quad & \text{Tr}(\mathbf{E}_i \mathbf{S}^* \mathbf{S}^T) \leq \tau_i, \quad i = 1, 2 \end{aligned} \quad (45)$$

where  $\mathbf{E}_i$  is defined in (34). Although the optimization problem in (45) is non-convex with respect to  $\mathbf{S}_i$ , it is noted that one may optimize (45) with respect to the positive semidefinite matrix  $\mathbf{S}^* \mathbf{S}^T$  instead of the training sequence  $\mathbf{S}_i$ . This approach is preferable since in (45), the objective function and the constraint both depend on  $\mathbf{S}^* \mathbf{S}^T$  and not  $\mathbf{S}_i$ . Hence, by defining  $\mathbf{Q}_S \triangleq \mathbf{S}^* \mathbf{S}^T$ , the following equivalent problem can be obtained

$$\begin{aligned} \min_{\mathbf{Q}_S \succeq \mathbf{0}} \quad & \sum_{n=1}^M \sigma_{r,H,n} \text{Tr} \left[ \left( \mathbf{Z}_{t,H}^{-1} + \alpha_n \mathbf{Q}_S \right)^{-1} \right] \\ \text{s.t.} \quad & \text{Tr}(\mathbf{E}_i \mathbf{Q}_S) \leq \tau_i, \quad i = 1, 2. \end{aligned} \quad (46)$$

Note that although the mapping from  $\mathbf{S}$  to  $\mathbf{Q}_S$  is a many-to-one mapping, the optimal value of the objective function in (46) is equal to the one in (45). Based on this fact, if we find the

optimal solution of (46), which is denoted as  $\tilde{\mathbf{Q}}_S$ , we can always obtain an  $\tilde{\mathbf{S}}_S$  with  $\tilde{\mathbf{Q}}_S = \tilde{\mathbf{S}}^* \tilde{\mathbf{S}}^T$ , which is also the optimal solution of (45).

*Theorem 2:* The optimization problem in (46) is convex with respect to the positive semidefinite matrix  $\mathbf{Q}_S$ .

*Proof:* See Appendix C. ■

Next, we further show that the optimization problem in (46) can be solved by transforming it into a semidefinite programming (SDP) problem. By introducing the variables  $\mathbf{X}_n$  for  $n = 1, 2, \dots, M$ , the problem in (46) can be rewritten in an equivalent form as

$$\begin{aligned} \min_{\mathbf{Q}_S \succeq \mathbf{0}, \mathbf{X}_n} \quad & \sum_{n=1}^M \sigma_{r,H,n} \text{Tr}(\mathbf{X}_n) \\ \text{s.t.} \quad & \text{Tr}(\mathbf{E}_i \mathbf{Q}_S) \leq \tau_i, \quad i = 1, 2 \\ & \left( \mathbf{Z}_{t,H}^{-1} + \alpha_n \mathbf{Q}_S \right)^{-1} \preceq \mathbf{X}_n, \quad \forall n \end{aligned} \quad (47)$$

By using the Schur complement, (47) can be further transformed into the following SDP problem

$$\begin{aligned} \min_{\mathbf{Q}_S \succeq \mathbf{0}, \mathbf{X}_n} \quad & \sum_{n=1}^M \sigma_{r,R,n} \text{Tr}(\mathbf{X}_n) \\ \text{s.t.} \quad & \text{Tr}(\mathbf{E}_i \mathbf{Q}_S) \leq \tau_i, \quad i = 1, 2 \\ & \begin{bmatrix} \mathbf{Z}_{t,H}^{-1} + \alpha_n \mathbf{Q}_S & \mathbf{I}_{N_1+N_2} \\ \mathbf{I}_{N_1+N_2} & \mathbf{X}_n \end{bmatrix} \succeq \mathbf{0}, \quad \forall n \end{aligned} \quad (48)$$

By solving the SDP problem in (48), the optimal solution to the optimization problem in (46) can be obtained. However, this numerical method of solving this optimization problem has a relatively high computational complexity. As such, to obtain the optimal structure of  $\mathbf{S}_i$  and gain a better understanding of the optimization in (45), the following theorem is introduced.

*Theorem 3:* With  $L_S \geq N_1 + N_2$ , the optimal training sequence  $\mathbf{S}_i$  in (45) must satisfy the condition  $\mathbf{S}_1^* \mathbf{S}_2^T = \mathbf{0}$ . In addition, the optimal  $\mathbf{S}_i$  has a form of  $\mathbf{S}_i = \mathbf{U}_{t,H_i}^* \Sigma_{s_i} \mathbf{V}_{s_i}^H$ . The condition  $\mathbf{S}_1^* \mathbf{S}_2^T = \mathbf{0}$  can be achieved by choosing  $\mathbf{V}_{s_1}$  and  $\mathbf{V}_{s_2}$  such that  $\mathbf{V}_{s_1}^H \mathbf{V}_{s_2} = \mathbf{0}$ . Moreover,  $\Sigma_{s_i}$  is a diagonal matrix where  $[\Sigma_{s_i}]_{m,m} = \sigma_{s_i,m}$ , and  $\Sigma_{s_i}$ , for  $i = 1, 2$ , are obtained by solving the following water-filling problem

$$\sum_{m=1}^M \frac{\alpha_n \sigma_{r,H,n} \sigma_{t,H_i,m}^2}{(1 + \alpha_n \sigma_{t,H_i,m} \sigma_{s_i,m}^2)^2} = \lambda_i. \quad (49)$$

In (49), the optimal  $\lambda_i$  is nonnegative and should be selected to satisfy  $\sum_{m=1}^{N_i} \sigma_{s_i,m}^2 = \tau_i$ .

*Proof:* See Appendix D. ■

*Remark 1:* The optimal  $\lambda_i$  in (49) can be found via the bisection search algorithm and the optimal  $\lambda_i$  is bounded by  $(0, \min_m \sum_{n=1}^M \alpha_n \sigma_{r,H,n} \sigma_{t,H_i,m}^2)$ . The upper limit of this bound is obtained via the following relationship  $\lambda_i = \sum_{n=1}^M \frac{\alpha_n \sigma_{r,H,n} \sigma_{t,H_i,m}^2}{(1 + \alpha_n \sigma_{t,H_i,m} \sigma_{s_i,m}^2)^2} \leq \sum_{n=1}^M \alpha_n \sigma_{r,H,n} \sigma_{t,H_i,m}^2 \leq \min_m \sum_{n=1}^M \alpha_n \sigma_{r,H,n} \sigma_{t,H_i,m}^2$ .

## V. TRAINING SEQUENCE DESIGN FOR BC PHASE

In this section, we intend to optimize the training sequence  $\mathbf{S}_R$  by minimizing the total estimation MSE at the two source ends subject to the relay power constraint. According to the MSE derived in (24), the corresponding problem can be formulated as

$$\begin{aligned} \min_{\mathbf{S}_R} \quad & \sum_{i=1}^2 \text{Tr} \left[ \mathbf{C}_{0,G_i} \left( \mathbf{I} + (\mathbf{S}_R^T \mathbf{C}_{l,G} \otimes \mathbf{C}_{r,G_i})^H \right. \right. \\ & \left. \left. \times \mathbf{K}_i^{-1} (\mathbf{S}_R^T \mathbf{C}_{l,G} \otimes \mathbf{C}_{r,G_i}) \right)^{-1} \right] \\ \text{s.t.} \quad & \text{Tr}(\mathbf{S}_R \mathbf{S}_R^H) \leq \tau_R. \end{aligned} \quad (50)$$

It is also worth noting that the training designs for point-to-point systems in [17], [18] are not applicable to the scenario under consideration here, since the training sequence at the relay,  $\mathbf{S}_R$ , needs to be optimized to enhance channel estimation over both links that connect the relay to the sources nodes. Prior to solving (50), let us first present the minimum training sequence length required for channel estimation in the BC phase.

*Lemma 3:* For arbitrary  $\mathbf{K}_{q,i}$ ,  $i = 1, 2$ , for the MSE at the relay to tend to zero as the relay power tends to infinity, the minimum length of the relay training sequence must be  $L_R = M$ . Otherwise, when  $L_R < M$ , the total MSE is always lower bounded by  $\sum_{m=L_R+1}^M \sigma_{l,G,m} (\sum_{n=1}^{N_1} \sigma_{r,G_1,n} + \sum_{n=1}^{N_2} \sigma_{r,G_2,n})$ , with  $\sigma_{l,G,m}$  being the  $m$ -th largest eigenvalue of  $\mathbf{Z}_{l,G}$  as the relay power tends to infinity.

For  $\mathbf{K}_{q,i} = q_i \mathbf{I}$ ,  $i = 1, 2$  and with finite power at the relay, the minimum length of the training sequence transmitted from the relay must be  $r \leq M$ . It is important to note that  $r$  denotes the optimal solution of  $\mathbf{S}_R \in \mathbb{C}^{M \times L_R}$ , where  $L_R$  is an arbitrary integer satisfying  $L_R \geq M$  in (50).

*Proof:* Since the proof is similar to the proof of *Lemma 1*, it is omitted for brevity. ■

In general, the optimization in (50) is non-convex. Hence, as in *Algorithm 1*, we first propose an iterative approach to optimize the design of the training sequence at the relay. To this end, the MSE, i.e.,  $e_i$ , given in (24) is rewritten as

$$\begin{aligned} e_i &= \mathcal{E} \left( \text{Tr} \left[ \mathbf{C}_{0,G_i} (\mathbf{w}_{G_i} - \mathbf{T}_i \mathbf{y}_i) (\mathbf{w}_{G_i} - \mathbf{T}_i \mathbf{y}_i)^H \right] \right) \\ &= \text{Tr} \left[ \mathbf{C}_{0,G_i} - (\mathbf{S}_i^T \mathbf{C}_{l,G} \otimes \mathbf{C}_{r,G_i})^H \mathbf{T}_i^H \mathbf{C}_{0,G_i}^H \right. \\ &\quad - \mathbf{C}_{0,G_i} \mathbf{T}_i (\mathbf{S}_i^T \mathbf{C}_{l,G} \otimes \mathbf{C}_{r,G_i}) + \mathbf{C}_{0,G_i} \mathbf{T}_i \\ &\quad \times (\mathbf{S}_i^T \mathbf{C}_{l,G} \otimes \mathbf{C}_{r,G_i}) (\mathbf{S}_i^T \mathbf{C}_{l,G} \otimes \mathbf{C}_{r,G_i})^H \mathbf{T}_i^H \\ &\quad \left. + \mathbf{C}_{0,G_i} \mathbf{T}_i \mathbf{K}_i \mathbf{T}_i^H \right]. \end{aligned} \quad (51)$$

Since (22) and (24) are in equivalent form, the optimization problem in (50) can be rewritten as

$$\begin{aligned} \min_{\mathbf{T}_1, \mathbf{T}_2, \mathbf{S}_R} \quad & e_1 + e_2 \\ \text{s.t.} \quad & \text{Tr}(\mathbf{S}_R \mathbf{S}_R^H) \leq \tau_R. \end{aligned} \quad (52)$$

In the first subproblem, for a given  $\mathbf{S}_R$ , the optimal LMMSE estimators  $\mathbf{T}_1$  and  $\mathbf{T}_2$  at the two source ends are obtained as given in (23). Thus, we focus on solving the second subproblem, where the relay training sequence is optimized for a fixed

LMMSE estimator. Similar to (30) and (32), the MSE in (51) is reexpressed as

$$\begin{aligned} e_i &= \mathbf{s}_R^H \mathbf{E}_i^H (\mathbf{T}_i^H \mathbf{C}_{0,G_i} \mathbf{T}_i \otimes \mathbf{C}_{l,r,G_i}^T) \mathbf{E}_i \mathbf{s}_R \\ &\quad - \text{vec}(\mathbf{C}_{T,G_i})^T \mathbf{E}_i \mathbf{s}_R - \text{vec}(\mathbf{C}_{T,G_i}^*)^T \mathbf{E}_i \mathbf{s}_R^* \\ &\quad + \text{Tr}(\mathbf{C}_{0,G_i} \mathbf{T}_i \mathbf{K}_i \mathbf{T}_i^H) + \text{Tr}(\mathbf{C}_{0,G_i}), \end{aligned}$$

where  $\mathbf{s}_R \triangleq \text{vec}(\mathbf{S}_R)$ ,  $\mathbf{C}_{l,r,G_i} \triangleq \mathbf{C}_{l,G} \mathbf{C}_{l,G}^H \otimes \mathbf{C}_{r,G_i} \mathbf{C}_{r,G_i}^H$ , and  $\mathbf{C}_{T,G_i} \triangleq (\mathbf{C}_{l,G} \otimes \mathbf{C}_{r,G_i}) \mathbf{C}_{0,G_i} \mathbf{T}_i$ . In this case,  $\mathbf{E}_i$  is an  $N_i L_R \times M L_R$  matrix that is constructed as in (30). Accordingly, (52) can be rewritten as

$$\begin{aligned} \min_{\mathbf{S}_R} \quad & \mathbf{s}_R^H \mathbf{A}_R \mathbf{s}_R - \mathbf{a}_R^T \mathbf{s}_R - \mathbf{a}_R^H \mathbf{s}_R^* \\ \text{s.t.} \quad & \mathbf{s}_R^H \mathbf{s}_R \leq \tau_R, \end{aligned} \quad (53)$$

where  $\mathbf{A}_R \triangleq \sum_{i=1}^2 \mathbf{E}_i^H (\mathbf{T}_i^H \mathbf{C}_{0,G_i} \mathbf{T}_i \otimes \mathbf{C}_{l,r,G_i}^T) \mathbf{E}_i$  and  $\mathbf{a}_R \triangleq \mathbf{E}_1^T \text{vec}(\mathbf{C}_{T,G_1}) + \mathbf{E}_2^T \text{vec}(\mathbf{C}_{T,G_2})$ . Note that different from the source training design, (53) has only one power constraint. Thus, its solution can be obtained via the KKT conditions given by

$$\begin{aligned} \mathbf{A}_R \mathbf{s}_R - \mathbf{a}_R^* + \lambda \mathbf{s}_R &= \mathbf{0}, \\ \lambda (\mathbf{s}_R^H \mathbf{s}_R - \tau_R) &= 0, \\ \mathbf{s}_R^H \mathbf{s}_R - \tau_R &\leq 0, \end{aligned} \quad (54)$$

where  $\lambda$  is the Lagrangian multiplier associated with the relay power constraint. The solution of (53) is obtained as

$$\mathbf{s}_R = (\mathbf{A}_R + \lambda \mathbf{I})^{-1} \mathbf{a}_R^*. \quad (55)$$

In (55), if the solution  $\mathbf{s}_R$  with  $\lambda = 0$  violates the KKT conditions given in (54),  $\lambda$  should be chosen to meet  $\mathbf{s}_R^H \mathbf{s}_R = \tau_R$ . Consequently, we introduce the following lemma.

*Lemma 4:* The function  $g(\lambda) = \mathbf{s}_R^H \mathbf{s}_R$ , with  $\mathbf{s}_R$  defined in (55), monotonically decreases with respect to  $\lambda$ . Moreover, the optimal  $\lambda$  is upper-bounded by  $\sqrt{\frac{\sigma_{a,R}}{\tau_R}} - \sigma_{R,\min}$  with  $\sigma_{a,R} = \mathbf{a}_R^T \mathbf{a}_R^*$  and  $\sigma_{R,\min}$  denoting the smallest eigenvalue of  $\mathbf{A}_R$ .

*Proof:* Since the proof is similar to that of *Lemma 2*, it is omitted for brevity.

Using the above steps, the overall relay training design algorithm can be summarized as follows:

---

### Algorithm 2

---

- **Initialize**  $\mathbf{S}_R$
  - **Repeat**
    - Update the LMMSE estimator matrix  $\mathbf{T}_i$ , for  $i = 1, 2$ , using (23) for a fixed  $\mathbf{S}_R$ ;
    - Update the training signal  $\mathbf{S}_R$  using (55) for a fixed  $\mathbf{T}_i$ , for  $i = 1, 2$ ;
  - **Until** The difference between the MSE from one iteration to another is smaller than a certain predetermined threshold.
-

Note that the convergence property proven for Algorithm 1 also applies to Algorithm 2. Although for the general case the solution of  $\mathbf{S}_R$  can only be obtained via an iterative approach, it is shown that for some special cases, the optimal training sequence  $\mathbf{S}_R$  can be found in closed-form.

#### A. When $\mathbf{K}_{q,i} = q_i \mathbf{I}$

In this subsection, we consider that the temporal covariance matrix,  $\mathbf{K}_{q,i}$ , is a scalar multiple of the identity matrix. This scenario corresponds to the practical case, where the disturbance consists of both the additive white Gaussian noise and the temporally uncorrelated interference. Using similar steps as in Appendix A, we rewrite the MSE in (24) as

$$\begin{aligned} e_i &= \sum_{n=1}^{N_i} \sigma_{r,G_i,n} \text{Tr} \left[ \left( \mathbf{Z}_{t,G}^{-1} + \beta_{i,n} \mathbf{S}_R^* \mathbf{K}_{q,i}^{-1} \mathbf{S}_R^T \right)^{-1} \right], \\ &= \sum_{n=1}^{N_i} \sigma_{r,G_i,n} \text{Tr} \left[ \left( \mathbf{Z}_{t,G}^{-1} + \tilde{\beta}_{i,n} \mathbf{S}_R^* \mathbf{S}_R^T \right)^{-1} \right] \end{aligned}$$

where  $\beta_{i,n} = \frac{\sigma_{r,G_i,n}}{\delta_{r,i,n}}$  with  $\delta_{r,i,n}$  being the  $n$ -th eigenvalue of  $\mathbf{K}_{r,i}$  as defined in (14), and  $\tilde{\beta}_{i,n} = \beta_{i,n}/q_i$ , for  $i = 1, 2$ . Then, (50) can be further simplified as

$$\begin{aligned} \min_{\mathbf{S}_R} \quad & \sum_{n=1}^{N_1} \sigma_{r,G_1,n} \text{Tr} \left[ \left( \mathbf{Z}_{t,G}^{-1} + \tilde{\beta}_{1,n} \mathbf{S}_R^* \mathbf{S}_R^T \right)^{-1} \right] \\ & + \sum_{n=1}^{N_2} \sigma_{r,G_2,n} \text{Tr} \left[ \left( \mathbf{Z}_{t,G}^{-1} + \tilde{\beta}_{2,n} \mathbf{S}_R^* \mathbf{S}_R^T \right)^{-1} \right] \\ \text{s.t.} \quad & \text{Tr}(\mathbf{S}_R \mathbf{S}_R^H) \leq \tau_R, \end{aligned} \quad (56)$$

The optimal solution of (56) is given in the following lemma.

*Lemma 5:* The optimal  $\mathbf{S}_R$  in (56) is of the form

$$\mathbf{S}_R = \mathbf{U}_{t,G}^* \Sigma_{s,R} \mathbf{V}_{s,R}^T, \quad (57)$$

where  $\Sigma_{s,R}$  is a positive real diagonal matrix with  $[\Sigma_{s,R}]_{n,n} = \sigma_{s,R,n}$ ,  $\mathbf{U}_{t,G}$  is the eigenvector matrix of  $\mathbf{Z}_{t,G}$ , and  $\mathbf{V}_{s,R}$  is an arbitrary unitary matrix. The eigenvalues of  $\mathbf{Z}_{t,G}$  corresponding to the eigenvector matrix  $\mathbf{U}_{t,G}$  are arranged in the same order as the diagonal elements of  $\Sigma_{s,R}$ . The optimal  $\Sigma_{s,R}$  can be obtained from

$$\begin{aligned} \lambda &= \sum_{n=1}^{N_1} \frac{\sigma_{r,G_1,n} \tilde{\beta}_{1,n} \sigma_{t,G,m}^2}{(1 + \tilde{\beta}_{1,n} \sigma_{t,G,m} \sigma_{s,R,m})^2} \\ &+ \sum_{n=1}^{N_2} \frac{\sigma_{r,G_2,n} \tilde{\beta}_{2,n} \sigma_{t,G,m}^2}{(1 + \tilde{\beta}_{2,n} \sigma_{t,G,m} \sigma_{s,R,m})^2}, \end{aligned} \quad (58)$$

where  $\lambda \in (0, \max_m \{ \sum_{n=1}^{N_1} (\sigma_{r,G_1,n} \tilde{\beta}_{1,n} \sigma_{t,G,m}^2 + \sigma_{r,G_2,n} \tilde{\beta}_{2,n} \sigma_{t,G,m}^2) \})$  can be found via the bisection search to meet the relay's power constraint.

*Proof:* The expression of MSE, i.e.,  $e_i$ , for  $i = 1, 2$ , in (56) contains a term  $\text{Tr}[(\mathbf{Z}_{t,G}^{-1} + \beta_{i,n} \mathbf{S}_R^* \mathbf{S}_R^T)^{-1}]$ . If the eigenvalues of  $\mathbf{Z}_{t,G}$  and  $\mathbf{S}_R^* \mathbf{S}_R^T$  are arranged in the same order, we have [27]

$$\lambda(\mathbf{Z}_{t,G}^{-1}) + \lambda(\tilde{\beta}_{i,n} \mathbf{S}_R^* \mathbf{S}_R^T) \preceq \lambda(\mathbf{Z}_{t,G}^{-1} + \tilde{\beta}_{i,n} \mathbf{S}_R^* \mathbf{S}_R^T). \quad (59)$$

Denote by  $\mathbf{S}_R^* \mathbf{S}_R^T = \mathbf{U}_{s,R} \Sigma_{s,R}^2 \mathbf{U}_{s,R}^H$  the EVD of  $\mathbf{S}_R^* \mathbf{S}_R^T$ . Since function  $\text{Tr}(\mathbf{A}^{-1})$  is a schur convex function with respect to the eigenvalues of  $\mathbf{A}$  [18], based on (59), we see that the minimum value can be reached if  $\mathbf{U}_{s,R} = \mathbf{U}_{t,G}$  and the eigenvalues of  $\mathbf{Z}_{t,G}$ , i.e.,  $\sigma_{t,G,1}, \sigma_{t,G,2}, \dots, \sigma_{t,G,M}$ , are arranged in the same order with diagonal elements of  $\Sigma_{s,R}$ , i.e.,  $\sigma_{s,R,1}, \sigma_{s,R,2}, \dots, \sigma_{s,R,M}$ . With the structure of training sequence given in (57), the original optimization problem in (56) is reduced to the following power allocation problem

$$\begin{aligned} \min_{\sigma_{s,R,n}, \forall n} \quad & \sum_{n=1}^{N_1} \sum_{m=1}^M \frac{\sigma_{r,G_1,n} \sigma_{t,G,m}}{1 + \tilde{\beta}_{1,n} \sigma_{t,G,m} \sigma_{s,R,m}} \\ & + \sum_{n=1}^{N_2} \sum_{m=1}^M \frac{\sigma_{r,G_2,n} \sigma_{t,G,m}}{1 + \tilde{\beta}_{2,n} \sigma_{t,G,m} \sigma_{s,R,m}} \\ \text{s.t.} \quad & \sum_{m=1}^M \sigma_{s,R,m} \leq \tau_R \end{aligned} \quad (60)$$

The Lagrangian function of (60) can be written as  $\mathcal{L} = \sum_{n=1}^{N_1} \sum_{m=1}^M \frac{\sigma_{r,G_1,n} \sigma_{t,G,m}}{1 + \tilde{\beta}_{1,n} \sigma_{t,G,m} \sigma_{s,R,m}} + \sum_{n=1}^{N_2} \sum_{m=1}^M \frac{\sigma_{r,G_2,n} \sigma_{t,G,m}}{1 + \tilde{\beta}_{2,n} \sigma_{t,G,m} \sigma_{s,R,m}} + \lambda(\sum_{m=1}^M \sigma_{s,R,m} - \tau_R)$ , where  $\lambda$  is Lagrangian multiplier associated with the power constraint given in (60). Based on the KKT conditions, we obtain (58). Then, by setting  $\sigma_{s,R,m} = 0$  for  $\forall m$ , we obtain the range of  $\lambda$  as shown in *Lemma 2*.

#### B. When $\mathbf{Z}_{t,G} = a \mathbf{I}$

This case corresponds to a scenario, where the relay antennas are far enough from one to another such that they are spatially uncorrelated. In this case, the corresponding training design problem can be formulated as

$$\begin{aligned} \min_{\mathbf{S}_R} \quad & \sum_{n=1}^{N_1} \tilde{\sigma}_{r,G_1,n} \text{Tr} \left[ \left( \mathbf{I} + \tilde{\beta}_{1,n} \mathbf{S}_R^* \mathbf{K}_{q,1}^{-1} \mathbf{S}_R^T \right)^{-1} \right] \\ & + \sum_{n=1}^{N_2} \tilde{\sigma}_{r,G_2,n} \text{Tr} \left[ \left( \mathbf{I} + \tilde{\beta}_{2,n} \mathbf{S}_R^* \mathbf{K}_{q,2}^{-1} \mathbf{S}_R^T \right)^{-1} \right] \\ \text{s.t.} \quad & \text{Tr}(\mathbf{S}_R^* \mathbf{S}_R^T) \leq \tau_R, \end{aligned} \quad (61)$$

where  $\tilde{\sigma}_{r,G_i,n} = a \sigma_{r,G_i,n}$ , and  $\tilde{\beta}_{i,n} = a \beta_{i,n}$  for  $i = 1, 2$ . In general, the optimization problem in (61) is non-convex with respect to  $\mathbf{S}_R$ . However, when the relay power is large enough, we approximately use the minimum training sequence length in the case of an infinite relay power derived in *Lemma 3*, i.e.,  $L_R = M$ . Then, (61) is equivalent to the following problem

$$\begin{aligned} \min_{\mathbf{S}_R} \quad & \sum_{n=1}^{N_1} \tilde{\sigma}_{r,G_1,n} \text{Tr} \left[ \left( \mathbf{I} + \tilde{\beta}_{1,n} \mathbf{S}_R^T \mathbf{S}_R^* \mathbf{K}_{q,1}^{-1} \right)^{-1} \right] \\ & + \sum_{n=1}^{N_2} \tilde{\sigma}_{r,G_2,n} \text{Tr} \left[ \left( \mathbf{I} + \tilde{\beta}_{2,n} \mathbf{S}_R^T \mathbf{S}_R^* \mathbf{K}_{q,2}^{-1} \right)^{-1} \right] \\ \text{s.t.} \quad & \text{Tr}(\mathbf{S}_R^T \mathbf{S}_R^*) \leq \tau_R. \end{aligned} \quad (62)$$

In obtaining (62) from (61), we have used the identity  $\text{Tr}([\mathbf{I} + \mathbf{A}\mathbf{B}]^{-1}) = \text{Tr}([\mathbf{I} + \mathbf{B}\mathbf{A}]^{-1}) + m - n$ , where  $\mathbf{A}$  and  $\mathbf{B}$  are  $m \times n$  and  $n \times m$  matrices, respectively, [28]. Similar to Section IV-B, it is

observed that in (62), we can directly optimize the matrix  $\mathbf{S}_R^T \mathbf{S}_R^*$  instead of  $\mathbf{S}_R$  by solving the following optimization

$$\begin{aligned} \min_{\mathbf{Q}_R \succeq \mathbf{0}} \quad & \sum_{n=1}^{N_1} \tilde{\sigma}_{r,G_1,n} \text{Tr} \left[ \left( \mathbf{I} + \tilde{\beta}_{1,n} \mathbf{K}_{q,1}^{-1/2} \mathbf{Q}_R \mathbf{K}_{q,1}^{-1/2} \right)^{-1} \right] \\ & + \sum_{n=1}^{N_2} \tilde{\sigma}_{r,G_2,n} \text{Tr} \left[ \left( \mathbf{I} + \tilde{\beta}_{2,n} \mathbf{K}_{q,2}^{-1/2} \mathbf{Q}_R \mathbf{K}_{q,2}^{-1/2} \right)^{-1} \right] \\ \text{s.t.} \quad & \text{Tr}(\mathbf{Q}_R) \leq \tau_R, \end{aligned} \quad (63)$$

where  $\mathbf{Q}_R \triangleq \mathbf{S}_R^T \mathbf{S}_R^*$ . By using similar steps as that in Theorem 2, it can be shown that (63) is convex and it can be readily solved via the following SDP problem

$$\begin{aligned} \min_{\mathbf{Q}_R \succeq \mathbf{0}, \mathbf{X}_n, \mathbf{Y}_n} \quad & \sum_{n=1}^M \tilde{\sigma}_{r,G_1,n} \text{Tr}(\mathbf{X}_n) + \tilde{\sigma}_{r,G_2,n} \text{Tr}(\mathbf{Y}_n) \\ \text{s.t.} \quad & \text{Tr}(\mathbf{Q}_R) \leq \tau_R \\ & \begin{bmatrix} \mathbf{I} + \tilde{\beta}_{1,n} \mathbf{K}_{q,1}^{-1/2} \mathbf{Q}_R \mathbf{K}_{q,1}^{-1/2} & \mathbf{I} \\ & \mathbf{X}_n \end{bmatrix} \succeq \mathbf{0}, \forall n \\ & \begin{bmatrix} \mathbf{I} + \tilde{\beta}_{2,n} \mathbf{K}_{q,2}^{-1/2} \mathbf{Q}_R \mathbf{K}_{q,2}^{-1/2} & \mathbf{I} \\ & \mathbf{Y}_n \end{bmatrix} \succeq \mathbf{0}, \forall n. \end{aligned}$$

## VI. SIMULATION RESULTS

In this section, we present simulation results to verify the performance of the proposed training design algorithms. The total normalized MSE (NMSE), defined as either  $\frac{1}{M(N_1+N_2)} \sum_{i=1}^2 \mathcal{E}\{\|\mathbf{H}_i - \hat{\mathbf{H}}_i\|_F^2\}$  or  $\frac{1}{M(N_1+N_2)} \sum_{i=1}^2 \mathcal{E}\{\|\mathbf{G}_i - \hat{\mathbf{G}}_i\|_F^2\}$ , is utilized to illustrate the performance of the proposed algorithms. In all simulations, the channel covariance matrices are assumed to have the following structures

$$\begin{aligned} [\mathbf{Z}_{t,b}]_{n,m} &= z_{t,b} J_0(d_{t,b}|n-m|), b \in \{H_1, H_2, G\}, \\ [\mathbf{Z}_{r,b}]_{n,m} &= z_{r,b} J_0(d_{r,b}|n-m|), b \in \{H, G_1, G_2\}, \end{aligned}$$

where  $J_0(\cdot)$  is the zeroth-order Bessel function of the first kind,  $d_{t,b}$  and  $d_{r,b}$  are proportional to the carrier frequency and the antenna separation vectors at the transmitter and the receiver, respectively [17]. Moreover, the scalars  $z_{t,b}$  and  $z_{r,b}$  are normalization factors such that  $\text{Tr}(\mathbf{Z}_{t,H_i}) = N_i$ ,  $\text{Tr}(\mathbf{Z}_{r,H}) = M$ ,  $\text{Tr}(\mathbf{Z}_{r,G}) = M$  and  $\text{Tr}(\mathbf{Z}_{r,G_i}) = N_i$ . The temporal covariance matrix of the disturbance is assumed to be modeled via a first order autoregressive (AR) filter, i.e., AR(1), that is denoted by  $[\mathbf{K}_{q,b}]_{n,m} = I_{q,b} k_{q,b} \eta_{q,b}^{|n-m|}$  for  $b \in \{1, 2, R\}$  [17]. Here, the scalar  $k_{q,b}$  is a normalization factor similar to the ones used in  $\mathbf{Z}_{t,b}$  and  $\mathbf{Z}_{r,b}$ . Moreover,  $I_{q,b}$  indicates the strength of the interference from the nearby users. Following the approach in [18], it is assumed that the received spatial covariance matrix of the disturbance,  $\mathbf{K}_{r,b}$ , shares the same eigenvalue vectors with  $\mathbf{Z}_{r,b}$  but with different eigenvalues. For simplicity, the length of the source and relay training sequences are assumed to be  $L_S = N_1 + N_2$  and  $L_R = M$ , respectively. The sum power at the two sources are assumed to be  $\tau_1 + \tau_2 = 2P$ . If not specified otherwise, we assume that  $N_1 = N_2 = M = 3$ . Furthermore, the system parameters for the MAC phase are set to:  $d_{t,H_1} = 1.5$ ,

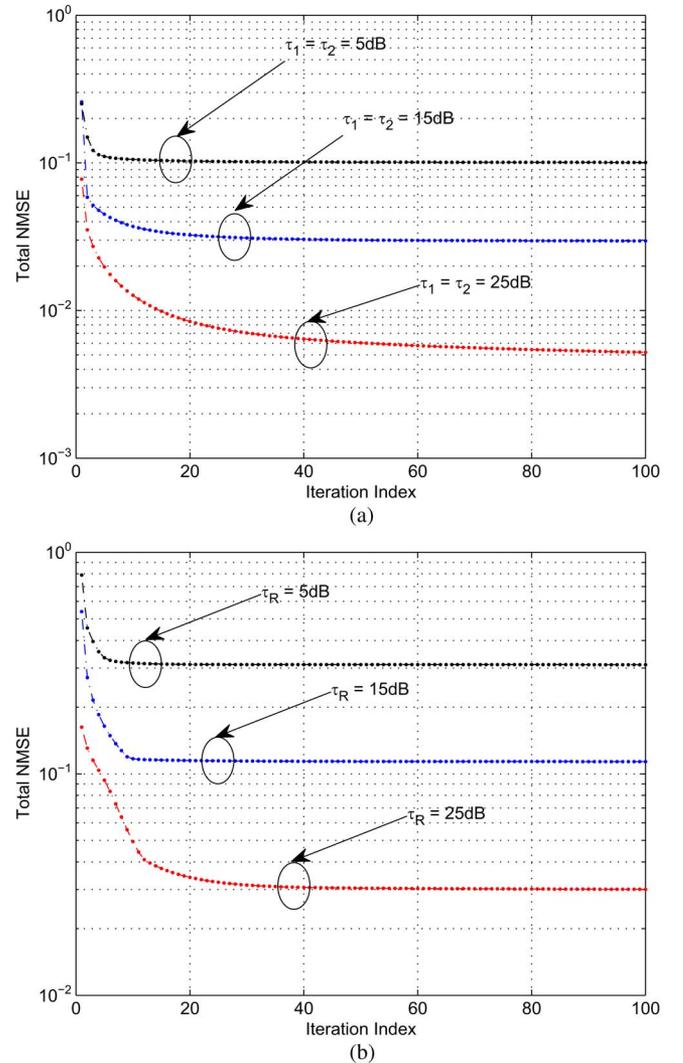


Fig. 2. Convergence behaviors of the proposed iterative designs. (a) MAC phase; (b) BC phase.

$d_{t,H_2} = 1.8$ ,  $d_{r,G} = 1.3$ ,  $\eta_{q,R} = 0.9$  and  $I_{q,R} = 1$ , while for the BC phase, we choose  $d_{t,G} = 1.9$ ,  $d_{r,G_1} = 1.95$ ,  $d_{r,G_2} = 0.3$ ,  $\eta_{q,1} = 0.9$ ,  $\eta_{q,2} = -0.9$  and  $I_{q,1} = I_{q,2} = 1$ .

In Fig. 2, the convergence behaviors of *Algorithms 1* and *2* for different SNRs are shown in subfigures (a) and (b), respectively. It is illustrated that in general, the proposed iterative algorithms converge very quickly and at most 60 iterations are required for them to converge. These results also indicate that as the SNR increases more iterations are needed for the proposed algorithms to converge. In Figs. (3a) and (3b), the convergence of the proposed algorithms are verified for different sets of initializations. In this setup, “Random-1” indicates that a random initial point is selected, “Random- $N$ ” implies that  $N$  random initial points are tested but the one with the best performance is selected, and “Identity” indicates that  $\mathbf{S} = \text{Blkdiag}(a\mathbf{I}_{N_1}, b\mathbf{I}_{N_2})$  and  $\mathbf{S}_R = c\mathbf{I}_M$ , where  $a$ ,  $b$ , and  $c$  are used to satisfy the source and relay power constraints. The results in Figs. (3a) and (3b) indicate that for various SNR values, the proposed iterative training design algorithms are not sensitive to the selected initial point. Furthermore, it is observed that the initialization process denoted by “Identity” performs well

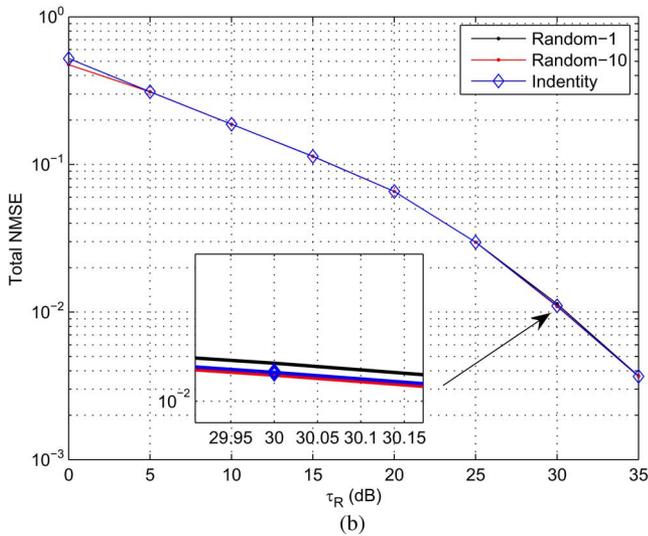
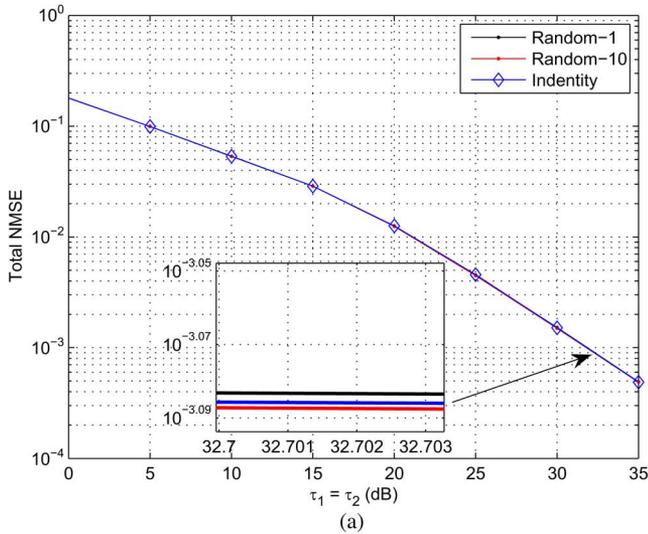


Fig. 3. Optimality for the proposed iterative designs with different initiations. (a) MAC phase. (b) BC phase.

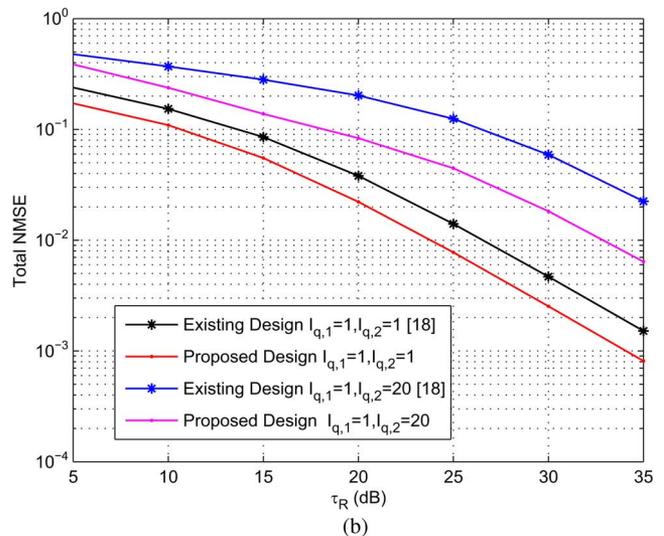
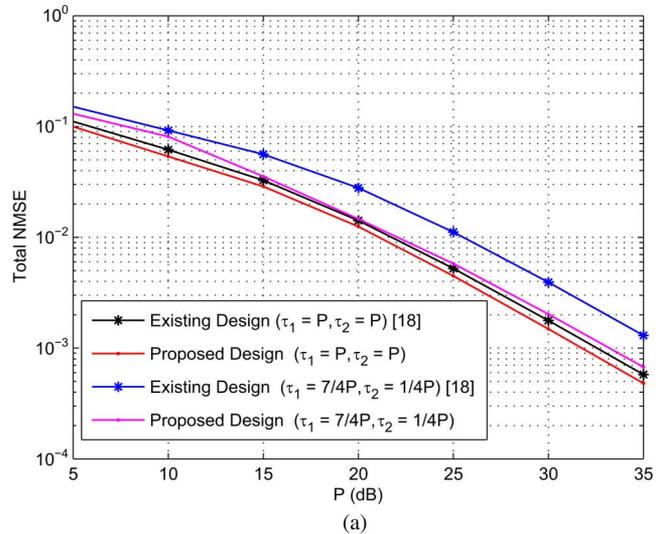


Fig. 4. Performance comparison with the existing design. (a) MAC phase. (b) BC phase.

as an initial point and can approach the initialization scenario denoted by “Random-10”. Hence, in the following, if not state otherwise, the “Identity” initial point is used.

In Fig. 4, we compare the total NMSE of the proposed iterative training design algorithm with that of [18], which is intended for point-to-point systems. To make this comparison possible, for the training sequence design in the MAC phase, it is assumed that two source nodes transmit their training sequences in two orthogonal time intervals, i.e.,  $\mathbf{s}_1(t)$  with  $t \in [1, 2, \dots, N_1]$  and  $\mathbf{s}_2(t)$  with  $t \in [N_1 + 1, N_1 + 2, \dots, N_1 + N_2]$ . In the BC phase, the training sequence,  $\mathbf{S}_R$  is designed according to the channel from the relay to the source  $S_1$ . The plots in Figs. (4a) and (4b) illustrate that compared to the approach in [18], the proposed training design can significantly improve the accuracy of channel estimation in TWR systems. This gain is even more pronounced when the two source nodes operate at different transmit power levels during the MAC phase and when the strengths of the disturbances at the two source ends are asymmetric, i.e.,  $I_{q,1} \neq I_{q,2}$  during the BC phase. This can be mainly attributed to the fact that the proposed training design

algorithm, i.e., *Algorithm 1*, takes into account the temporal correlation of the disturbance at the relay node in the MAC phase, while ensuring that the training sequences transmitted from the relay node simultaneously match the channels corresponding to relay-to-source links during the BC phase.

In Fig. 5, the performance of the proposed training sequence design algorithms and channel estimators in the MAC phase for  $\mathbf{K}_{q,R} = q\mathbf{I}_M$  is demonstrated. Three training sequence design approaches are taken into consideration: 1) The iterative design based on *Algorithm 1*; 2) The SDP design based on (48); and 3) The SVD design based on *Theorem 1*. As shown in *Theorem 3*, in this case, the optimization problem for finding the optimal training sequences is convex. Hence, it is well-known that both the SDP and the SVD design schemes can achieve optimal channel estimation performance. This outcome is also verified by the results in Fig. 5. However, it is interesting to note that the proposed iterative algorithm denoted by *Algorithm 1* can also achieve optimal performance, which further verifies its effectiveness for designing the training sequences in the MAC phase.

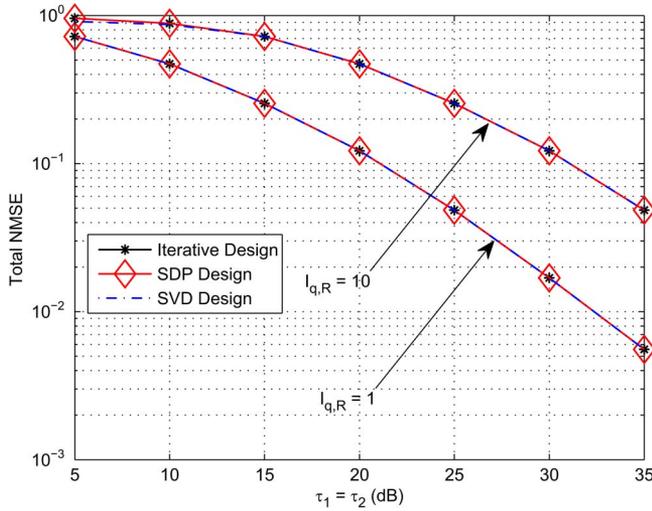


Fig. 5. Total NMSE of channel estimation during the MAC phase with  $\mathbf{K}_{q,R} = q\mathbf{I}_M$ .

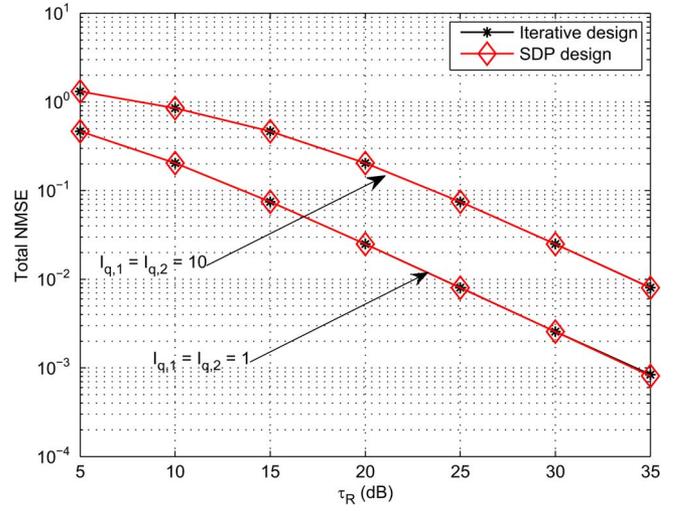


Fig. 7. Total NMSE of channel estimation during the BC phase with  $\mathbf{C}_{t,G} = \sqrt{a}\mathbf{I}_M$ .

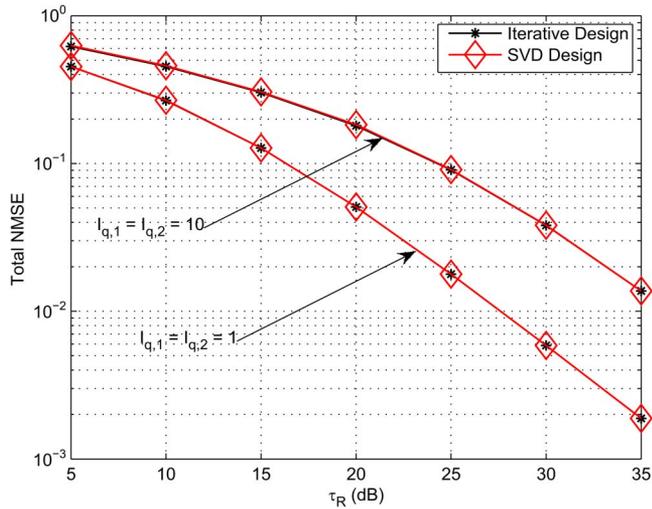


Fig. 6. Total NMSE of channel estimation during the BC phase with  $\mathbf{K}_{q,1} = q_1\mathbf{I}_{N_1}$  and  $\mathbf{K}_{q,2} = q_2\mathbf{I}_{N_2}$ .

Figs. 6 and 7 present the channel estimation performance for the special cases presented in Sections V-A and V-B during the BC phase. More specifically, in Fig. 6, the plot with the legend ‘‘SVD design’’ refers to the results in Lemma 5. The results in Fig. 6 show that the proposed iterative design algorithm closely matches the optimal performance obtained in the proposed ‘‘SVD design’’ method. Moreover, the optimality of Algorithm 2 is further verified via the results in Fig. 7. In this setup we consider the special case presented in Section V-B. Note that since the proposed ‘‘SDP design’’ in Section V-B is optimal, we conclude that the iterative design in Fig. 7 can approach the optimal solution in this special case.

VII. CONCLUSION

In this paper, the problem of channel estimation in MIMO TWR systems was analyzed. Unlike prior work in this field, the impact of the interference from neighboring devices and

the effect of antenna correlations on the design of training sequences and channel estimation performance were taken into consideration. To obtain the channel parameters corresponding to each individual link, we have proposed to carry out the channel estimation process in two phases: the BC phase and the MAC phase. Next, the optimal LMMSE channel estimators for both phases were derived and the corresponding training sequence design problems for both phases were formulated. Subsequently, to ensure accurate channel estimation in TWR systems, the minimum required length of the training sequences were also analyzed. Since the resulting optimization problems were non-convex in their general form, specific transformations were used to obtain near optimal iterative algorithms for the design of the training sequences. Further analysis showed that the optimal structures of the training sequences can be obtained in closed-form when the channel or the noise temporal covariance matrices have special structures. Simulation results show that the proposed training sequence design algorithms can significantly enhance channel estimation performance in TWR systems.

APPENDIX A  
PROOF OF LEMMA 1

To prove Lemma 1, we first rewrite (20) into the following form

$$e_R = \text{Tr} \left[ \left( \mathbf{Z}_{r,H}^{-1} \otimes \mathbf{Z}_{t,H}^{-1} + \mathbf{K}_{r,R}^{-1} \otimes \mathbf{S}^* \mathbf{K}_{q,R}^{-1} \mathbf{S}^T \right)^{-1} \right] \\ = \sum_{n=1}^M \sigma_{r,H,n} \text{Tr} \left[ \left( \mathbf{Z}_{t,H}^{-1} + \beta_{R,n} \mathbf{S}^* \mathbf{K}_{q,R}^{-1} \mathbf{S}^T \right)^{-1} \right], \quad (\text{A.1})$$

where  $\beta_{R,n} \triangleq \frac{\sigma_{r,H,n}}{\delta_{r,R,n}}$ . To obtain (A.1), we have used the rules  $(\mathbf{A} \otimes \mathbf{B})^{-1} = \mathbf{A}^{-1} \otimes \mathbf{B}^{-1}$  and (31). Since  $\mathbf{S}^* \mathbf{K}_{q,R}^{-1} \mathbf{S}^T \in \mathbb{C}^{(N_1+N_2) \times (N_1+N_2)}$ , to achieve an arbitrary small total MSE with infinite source powers, we must design a training sequence

$\mathbf{S}$  achieving a full rank  $\mathbf{S}^* \mathbf{K}_{q,R}^{-1} \mathbf{S}^T$ , which implies that the minimum length of  $\mathbf{S}_i$  must satisfy  $L_s \geq N_1 + N_2$ . Otherwise, based on the fact  $\text{Rank}(\mathbf{AB}) \leq \min\{\text{Rank}(\mathbf{A}), \text{Rank}(\mathbf{B})\}$ , we must have  $\text{Rank}(\mathbf{S}^* \mathbf{K}_{q,R}^{-1} \mathbf{S}^T) \leq L_s < N_1 + N_2$ . Let us consider the case where  $\mathbf{S}^* \mathbf{K}_{q,R}^{-1} \mathbf{S}^T$  has a maximum rank of  $L_s$  (which will lead to the MSE lower bound as shown below). In this case, the MSE in (A.1) can be lower bounded by

$$e_R \geq \sum_{n=1}^M \sigma_{r,H,n} \times \left( \sum_{m=1}^{L_s} \frac{1}{\sigma_{t,H,m}^{-1} + \beta_{R,n} \lambda_{SK,m}} + \sum_{m=L_s+1}^{N_1+N_2} \sigma_{t,H,m} \right), \quad (\text{A.2})$$

where  $\lambda_{SK,m}$  is the  $m$ -th element of  $\lambda(\mathbf{S}^* \mathbf{K}_{q,R}^{-1} \mathbf{S}^T)$ . Moreover, the eigenvalues in  $\{\sigma_{t,H,m}\}$  and  $\{\lambda_{SK,m}\}$  are assumed to be arranged in a decreasing order, respectively. To obtain (A.2), we have used the fact that the function  $\text{Tr}(\mathbf{A}^{-1})$  is a schur convex function with respect to the eigenvalues of  $\mathbf{A}$  and the following result from [27]  $\lambda(r,H) + \lambda(\mathbf{S}^* \mathbf{K}_{q,R}^{-1} \mathbf{S}^T) \preceq \lambda(\mathbf{Z}_{r,H}^{-1} + \mathbf{S}^* \mathbf{K}_{q,R}^{-1} \mathbf{S}^T)$ . When the source power tends to infinity, the term  $\sum_{m=1}^{L_s} \frac{1}{\sigma_{t,H,m}^{-1} + \beta_{R,n} \lambda_{SK,m}}$  in (A.2) approaches zero. Thus,  $e_R$  in (A.2) is lower bounded by  $\sum_{n=1}^M \sigma_{r,H,n} \sum_{m=L_s+1}^{N_1+N_2} \sigma_{t,H,m}$ .

If  $\mathbf{K}_{q,R} = q\mathbf{I}$ , the total MSE in (A.1) can be written as

$$e_R = \sum_{n=1}^M \sigma_{r,H,n} \text{Tr} \left[ \left( \mathbf{Z}_{t,H}^{-1} + \tilde{\beta}_{R,n} \mathbf{S}^* \mathbf{S}^T \right)^{-1} \right], \quad (\text{A.3})$$

where  $\tilde{\beta}_{R,n} \triangleq \beta_{R,n}/q$ . With a finite power at the source, it is assumed that the optimal solution of  $\mathbf{S}_1$  and  $\mathbf{S}_2$  in (25) results in the optimal  $\mathbf{S}$ , denoted by  $\tilde{\mathbf{S}}$ , to have a rank of  $r \leq N_1 + N_2$ . By using the SVD decomposition, the optimal  $\tilde{\mathbf{S}}$  can be decomposed to  $\tilde{\mathbf{S}} = \mathbf{U}_S \Sigma_S \mathbf{V}_S^H$ , where  $\mathbf{U}_S$  and  $\mathbf{V}_S$  are matrices of size  $(N_1 + N_2) \times r$  and  $L_s \times r$ , respectively, with  $\mathbf{U}_S^H \mathbf{U}_S = \mathbf{I}_r$ ,  $\mathbf{V}_S^H \mathbf{V}_S = \mathbf{I}_r$ ;  $\Sigma_S$  is an  $r \times r$  diagonal singular-value matrix. The optimal  $\mathbf{S}_1$  and  $\mathbf{S}_2$  can be denoted as  $\tilde{\mathbf{S}}_1 = \mathbf{U}_{S,1} \Sigma_S \mathbf{V}_S^H$  and  $\tilde{\mathbf{S}}_2 = \mathbf{U}_{S,2} \Sigma_S \mathbf{V}_S^H$ , where  $\mathbf{U}_{S,1} \triangleq \mathbf{U}_S(1 : N_1, :)$  and  $\mathbf{U}_{S,2} \triangleq \mathbf{U}_S(N_1 + 1 : N_1 + N_2, :)$ . Subsequently, a new  $\tilde{\mathbf{S}}$  given by  $\tilde{\mathbf{S}} = \mathbf{U}_S \Sigma_S$ , can be obtained that achieves an equal total MSE as that of  $\tilde{\mathbf{S}}$ , however, with a shorter training sequence length of  $L_s = r$ . Furthermore, the new  $\tilde{\mathbf{S}}_1 = \mathbf{U}_{S,1} \Sigma_S$  and  $\tilde{\mathbf{S}}_2 = \mathbf{U}_{S,2} \Sigma_S$  require the same power at the sources nodes compared to  $\tilde{\mathbf{S}}_1$  and  $\tilde{\mathbf{S}}_2$ . This completes the proof of Lemma 1.

APPENDIX B  
PROOF OF LEMMA 2

By taking the gradient of  $g(\lambda_i)$ , we can easily verify that  $g(\lambda_i)$  decreases with  $\lambda_i$ . Next, we mainly focus on deriving the upper bound of  $\lambda_i$ . The source power constraint can be rewritten as

$$\begin{aligned} \text{Tr}(\mathbf{s}_i \mathbf{s}_i)^H &= \text{Tr} \left[ (\mathbf{X}_{s,1} \otimes \mathbf{X}_{s,2,i} + \lambda_i \mathbf{I})^{-2} \mathbf{x}_{s,3,i} \mathbf{x}_{s,3,i}^H \right] \\ &\leq \frac{\sigma_{s,3,i}}{(\sigma_{s,\min,i} + \lambda_i)^2}, \end{aligned} \quad (\text{B.1})$$

where  $\sigma_{s,\min,i}$  and  $\sigma_{s,3,i}$  are defined in Lemma 2. In (B.1), the inequality is obtained based on the identity  $\text{Tr}(\mathbf{AB}) \leq \sum_i \sigma_{A,i} \sigma_{B,i}$  [29]. Here,  $\sigma_{A,i}$  and  $\sigma_{B,i}$  are the eigenvalues of the  $n \times n$  matrices  $\mathbf{A}$  and  $\mathbf{B}$ , respectively,  $\{\sigma_{A,1}, \sigma_{A,2}, \dots, \sigma_{A,n}\}$  and  $\{\sigma_{B,1}, \sigma_{B,2}, \dots, \sigma_{B,n}\}$  are arranged in the same order, and the equality is achieved when  $\mathbf{A}$  and  $\mathbf{B}$  are diagonal matrices. Hence we have  $\frac{\sigma_{s,3,i}}{(\sigma_{s,\min,i} + \lambda_i)^2} \geq \tau_i$ , which further implies  $\lambda_i \leq \sqrt{\frac{\sigma_{s,3,i}}{\tau_i}} - \sigma_{s,\min,i}$ .

APPENDIX C  
PROOF OF THEOREM 2

In (46), the feasible set established by the power constraints is convex since the function  $\text{Tr}(\mathbf{E}_i \mathbf{Q}_S)$  is linear [25]. To prove the convexity of (46), it is sufficient to show that the objective function is convex. Without loss of generality, we denote that  $f(\mathbf{Q}_S) = \text{Tr}[(\mathbf{Z}_{t,H}^{-1} + \alpha_n \mathbf{Q}_S)^{-1}]$ . Using the fact that a function is convex if and only if it is convex when restricted to any line that intersects its domain [25], we can prove the convexity of  $f(\mathbf{Q}_S)$  by restricting it to an arbitrary line that intersects the domain  $\mathbb{S}_+^N$ , i.e., prove the convexity of  $g(t) = \text{Tr}[(\mathbf{Z}_{t,H}^{-1} + \alpha_n((1-t)\mathbf{Q}_{S,1} + t\mathbf{Q}'_{S,2}))^{-1}]$  with respect to  $t$  where  $t \geq 0$ ,  $\mathbf{Q}_{S,1}$  and  $\mathbf{Q}'_{S,2}$  are two arbitrary given elements in  $\mathbb{S}_+^N$ . Considering that  $g(t) = \text{Tr}[(\mathbf{Z}_{t,H}^{-1} + \alpha_n(\mathbf{Q}_{S,1} + t(\mathbf{Q}'_{S,2} - \mathbf{Q}_{S,1})))^{-1}]$  with  $\mathbf{Q}'_{S,2} - \mathbf{Q}_{S,1} \in \mathbb{S}^N$ , it is equivalent to prove  $g(t) = \text{Tr}[(\mathbf{Z}_{t,H}^{-1} + \alpha_n(\mathbf{Q}_{S,1} + t\mathbf{Q}_{S,2}))^{-1}]$  with  $\mathbf{Q}_{S,2} \in \mathbb{S}^N$  is convex with respect to  $t$ . To this end, we first have  $\frac{dg(t)}{dt} = -\text{Tr}(\alpha_n(\mathbf{Z}_{t,H}^{-1} + \alpha_n(\mathbf{Q}_{S,1} + t\mathbf{Q}_{S,2}))^{-2} \mathbf{Q}_{S,2})$ . Based on that, we can further reach

$$\begin{aligned} \frac{d^2g(t)}{dt^2} &= 2\alpha_n^2 \text{Tr} \left( \left( \mathbf{Z}_{t,H}^{-1} + \alpha_n(\mathbf{Q}_{S,1} + t\mathbf{Q}_{S,2}) \right)^{-2} \mathbf{Q}_{S,2} \right. \\ &\quad \left. \times \left( \mathbf{Z}_{t,H}^{-1} + \alpha_n(\mathbf{Q}_{S,1} + t\mathbf{Q}_{S,2}) \right)^{-1} \mathbf{Q}_{S,2} \right) \\ &\geq 0. \end{aligned} \quad (\text{C.1})$$

To obtain (C.1), we use the fact that  $(\mathbf{Z}_{t,H}^{-1} + \alpha_n(\mathbf{Q}_{S,1} + t\mathbf{Q}_{S,2}))^{-2}$  and  $\mathbf{Q}_{S,2}(\mathbf{Z}_{t,H}^{-1} + \alpha_n(\mathbf{Q}_{S,1} + t\mathbf{Q}_{S,2}))^{-1} \mathbf{Q}_{S,2}$  are positive semidefinite matrices. Hence, we conclude that the function  $f(\mathbf{Q}_S)$  is convex with respect to the positive semidefinite matrix  $\mathbf{Q}_S$ , which further implies that the objective function in (46) is convex since the sum of multiple convex functions is a still a convex function.

APPENDIX D  
PROOF OF THEOREM 3

For notation convenience, we define  $\mathbf{D}_0 \triangleq \mathbf{Z}_{t,H_1}^{-1} + \alpha_n \mathbf{S}^* \mathbf{S}^T$  and let  $\mathbf{D}_0 = \mathbf{Z}_{t,H}^{-1} + \alpha_n \mathbf{S}^* \mathbf{S}^T = \begin{bmatrix} \mathbf{Z}_{t,H_1}^{-1} + \alpha_n \mathbf{S}_1^* \mathbf{S}_1^T & \alpha_n \mathbf{S}_1^* \mathbf{S}_2^T \\ \alpha_n \mathbf{S}_2^* \mathbf{S}_1^T & \mathbf{Z}_{t,H_2}^{-1} + \alpha_n \mathbf{S}_2^* \mathbf{S}_2^T \end{bmatrix} \triangleq \begin{bmatrix} \mathbf{D}_1 & \mathbf{D}_2^H \\ \mathbf{D}_2 & \mathbf{D}_3 \end{bmatrix}$ . According to the matrix inverse identity, we have  $\mathbf{D}_0^{-1} = \begin{bmatrix} \mathbf{A}_0 & \mathbf{B}_0 \\ \mathbf{C}_0 & \mathbf{D}_0 \end{bmatrix}$ , where  $\mathbf{A}_0 = (\mathbf{D}_1 - \mathbf{D}_2^H \mathbf{D}_3^{-1} \mathbf{D}_2)^{-1}$ ,  $\mathbf{B}_0 = -\mathbf{D}_1^{-1} \mathbf{D}_2^H (\mathbf{D}_3 - \mathbf{D}_2 \mathbf{D}_1^{-1} \mathbf{D}_2^H)^{-1}$ ,  $\mathbf{C}_0 = -\mathbf{D}_3^{-1} \mathbf{D}_2 (\mathbf{D}_1 - \mathbf{D}_2^H \mathbf{D}_3^{-1} \mathbf{D}_2)^{-1}$ , and  $\mathbf{D}_0 = (\mathbf{D}_3 - \mathbf{D}_2 \mathbf{D}_1^{-1} \mathbf{D}_2^H)^{-1}$ . Subsequently,

we obtain that  $\text{Tr}(\mathbf{D}_0^{-1}) = \text{Tr}[(\mathbf{D}_1 - \mathbf{D}_2^H \mathbf{D}_3^{-1} \mathbf{D}_2)^{-1}] + \text{Tr}[(\mathbf{D}_3 - \mathbf{D}_2 \mathbf{D}_1^{-1} \mathbf{D}_2^H)^{-1}]$ . Since  $\mathbf{D}_2^H \mathbf{D}_3^{-1} \mathbf{D}_2 \succeq \mathbf{0}$  and  $\mathbf{D}_2 \mathbf{D}_1^{-1} \mathbf{D}_2^H \succeq \mathbf{0}$ , the inequalities  $\mathbf{D}_1 - \mathbf{D}_2^H \mathbf{D}_3^{-1} \mathbf{D}_2 \preceq \mathbf{D}_1$  and  $\mathbf{D}_3 - \mathbf{D}_2 \mathbf{D}_1^{-1} \mathbf{D}_2^H \preceq \mathbf{D}_3$  hold, which further implies that

$$\begin{aligned} (\mathbf{D}_1 - \mathbf{D}_2^H \mathbf{D}_3^{-1} \mathbf{D}_2)^{-1} &\succeq \mathbf{D}_1^{-1}, \\ (\mathbf{D}_3 - \mathbf{D}_2 \mathbf{D}_1^{-1} \mathbf{D}_2^H)^{-1} &\succeq \mathbf{D}_3^{-1}, \end{aligned} \quad (\text{D.1})$$

where in obtaining the above, we have used the fact that the matrices  $\mathbf{D}_1$ ,  $\mathbf{D}_3$ ,  $\mathbf{D}_1 - \mathbf{D}_2^H \mathbf{D}_3^{-1} \mathbf{D}_2$  and  $\mathbf{D}_3 - \mathbf{D}_2 \mathbf{D}_1^{-1} \mathbf{D}_2^H$  are positive semidefinite. From (D.1), we have

$$\begin{aligned} \text{Tr}[(\mathbf{D}_1 - \mathbf{D}_2^H \mathbf{D}_3^{-1} \mathbf{D}_2)^{-1}] &\succeq \text{Tr}(\mathbf{D}_1^{-1}), \\ \text{Tr}[(\mathbf{D}_3 - \mathbf{D}_2 \mathbf{D}_1^{-1} \mathbf{D}_2^H)^{-1}] &\succeq \text{Tr}(\mathbf{D}_3^{-1}). \end{aligned} \quad (\text{D.2})$$

Hence, if  $\mathbf{D}_2 = \mathbf{0}$ , i.e.,  $\mathbf{S}_2^* \mathbf{S}_1^T = \mathbf{0}$ , the value of the objective function in (45) can be always reduced.

Next, we show that for any  $\mathbf{S}_1$  and  $\mathbf{S}_2$ , letting  $\mathbf{S}_2^* \mathbf{S}_1^T = \mathbf{0}$  does not increase the need for power at the source nodes. Since in (45), the value of the objective function and the power constraints are only affected by  $\mathbf{S}_1^* \mathbf{S}_1^T$  and  $\mathbf{S}_2^* \mathbf{S}_2^T$ , the optimal  $\mathbf{S}_1$  and  $\mathbf{S}_2$ , denoted by  $\bar{\mathbf{S}}_1$  and  $\bar{\mathbf{S}}_2$ , can be determined as

$$\bar{\mathbf{S}}_i = \mathbf{U}_{t,s_i} \Sigma_{s_i} \mathbf{V}_{s_i}^H, \quad i = 1, 2 \quad (\text{D.3})$$

where  $\mathbf{V}_{s_i}$  can be any matrix satisfying  $\mathbf{V}_{s_i}^H \mathbf{V}_{s_i} = \mathbf{I}$ . It is worth noting that as  $L_S \geq N_1 + N_2$ , one can always find a specific  $\mathbf{V}_{s_1}$  and  $\mathbf{V}_{s_2}$  such that

$$\mathbf{V}_{s_1}^H \mathbf{V}_{s_2} = \mathbf{0}, \quad (\text{D.4})$$

which further results in  $\mathbf{S}_2^* \mathbf{S}_1^T = \mathbf{0}$ . In this case, we do not change the value of  $\mathbf{S}_i^* \mathbf{S}_i^T$  and the power constraint, while decrease the value of the objective function according to (D.2). This indicates that the condition  $\mathbf{S}_1^* \mathbf{S}_2^T = \mathbf{0}$  is sufficient and necessary for obtaining the optimal solution of (45).

It is noticed that the orthogonal training sequences have also been proven to be optimal for the cascaded channel estimation in two-way relaying system in [15] and [16]. Here, we show that similar results hold for individual channel estimation. With the optimal condition  $\mathbf{S}_1^* \mathbf{S}_2^T = \mathbf{0}$ , the optimization problem in (45) can be decomposed into two subproblems given by

$$\begin{aligned} \min_{\mathbf{S}_i} \quad & \sum_{n=1}^M \sigma_{r,H,n} \text{Tr} \left[ \left( \mathbf{Z}_{t,H_i}^{-1} + \alpha_n \mathbf{S}_i^* \mathbf{S}_i^T \right)^{-1} \right] \\ \text{s.t.} \quad & \text{Tr}(\mathbf{S}_i^* \mathbf{S}_i^T) \leq \tau_i, \quad i = 1, 2 \end{aligned} \quad (\text{D.5})$$

Based on the results derived for point-to-point MIMO systems [18], we obtain that the unitary matrix  $\mathbf{U}_{t,s_i}$  given in (D.3) should be of the form  $\mathbf{U}_{t,s_i} = \mathbf{U}_{t,H_i}^*$ , and  $\mathbf{V}_{s_1}$  should satisfy (D.4). Then, solving (D.5) just reduces to solving the a power allocation problem with the optimal  $\sigma_{s_i,m}$  being obtained via the water-filling problem in (49).

## REFERENCES

- [1] R. Wang, H. Mehrpouyan, M. Tao, and Y. Hua, "Optimal training design and individual channel estimation for MIMO two-way relay systems in colored environment," in *Proc. IEEE Globecom*. Austin, Texas, USA, 2014.
- [2] Y. Hua, D. W. Bliss, S. Gazor, Y. Rong, and Y. Sung, "Theories and methods for advanced wireless relays: Issue I," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 8, pp. 1297–1303, Sep. 2012.
- [3] 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.
- [4] F. Gao, R. Zhang, and Y.-C. Liang, "Optimal channel estimation and training design for two-way relay networks," *IEEE Trans. Commun.*, vol. 57, no. 10, pp. 3024–3033, Oct. 2009.
- [5] S. Abdallah and I. Psaromiligkos, "Partially-blind estimation of reciprocal channels for AF two-way relay networks employing M-PSK modulation," *IEEE Trans. Wireless Commun.*, vol. 11, no. 5, pp. 1649–1654, May 2012.
- [6] S. Abdallah and I. Psaromiligkos, "Blind channel estimation for amplify-and-forward two-way relay networks employing M-PSK modulation," *IEEE Trans. Signal Process.*, vol. 60, no. 7, pp. 3604–3615, Mar. 2012.
- [7] 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, Aug. 2011.
- [8] F. Gao, R. Zhang, and Y.-C. Liang, "Channel estimation for OFDM modulated two-way relay networks," *IEEE Trans. Signal Process.*, vol. 57, no. 11, pp. 4443–4455, Nov. 2009.
- [9] S. Zhang, F. Gao, and C.-X. Pei, "Optimal training design for individual channel estimation in two-way relay networks," *IEEE Trans. Signal Process.*, vol. 60, no. 9, pp. 4987–4991, Sep. 2012.
- [10] R. Wang and M. Tao, "Joint source and relay precoding designs for MIMO two-way relaying based on MSE criterion," *IEEE Trans. Signal Process.*, vol. 60, no. 3, pp. 1352–1365, 2012.
- [11] S. Xu and Y. Hua, "Optimal design of spatial source-and-relay matrices for a non-regenerative two-way MIMO relay system," *IEEE Trans. Wireless Commun.*, vol. 10, no. 5, pp. 1645–1655, May 2011.
- [12] J. Zhao, M. Kuhn, A. Wittneben, and G. Bauch, "Self-interference aided channel estimation in two-way relaying systems," in *Proc. IEEE GLOBECOM*, New Orleans, LO, USA, Dec. 2008, pp. 1–6.
- [13] Z. Fang, J. Shi, and H. Shan, "Comparison of channel estimation schemes for MIMO two-way relaying systems," in *Proc. CSQRWC*, Harbin, China, Jul. 2011, pp. 719–722.
- [14] T.-H. Pham, Y.-C. Liang, A. Nallanathan, and H. K. Garg, "Optimal training sequences for channel estimation in bi-directional relay networks with multiple antennas," *IEEE Trans. Commun.*, vol. 58, no. 2, pp. 474–479, Feb. 2010.
- [15] D.-H. Kim, M. Ju, and H.-M. Kim, "Optimal training signal design for estimation of correlated MIMO channels in two-way amplify-and-forward relay systems," *IEEE Commun. Lett.*, vol. 17, no. 3, pp. 491–494, Mar. 2013.
- [16] C. Chiong, Y. Rong, and Y. Xiang, "Channel training algorithms for two-way MIMO relay systems," *IEEE Trans. Signal Process.*, vol. 61, no. 16, pp. 3988–3998, Aug. 2013.
- [17] M. Biguesh, S. Gazor, and M. Shariat, "Optimal training sequence for MIMO wireless systems in colored environments," *IEEE Trans. Signal Process.*, vol. 57, no. 8, pp. 3144–3153, Aug. 2009.
- [18] E. Bjornson and B. Ottersten, "A framework for training-based estimation in arbitrarily correlated rician MIMO channels with rician disturbance," *IEEE Trans. Signal Process.*, vol. 58, no. 3, pp. 1807–1820, Mar. 2010.
- [19] R. Wang, M. Tao, and Z. Xiang, "Nonlinear precoding design for MIMO amplify-and-forward two-way relay systems," *IEEE Trans. Veh. Technol.*, vol. 61, no. 9, pp. 3984–3995, Nov. 2012.
- [20] R. Wang, M. Tao, and Y. Huang, "Linear precoding designs for amplify-and-forward multiuser two-way relay systems," *IEEE Trans. Wireless Commun.*, vol. 11, no. 12, pp. 4457–4469, Dec. 2012.
- [21] Y. Liu, T. Wong, and W. W. Hager, "Training signal design for estimation of correlated MIMO channels with colored interference," *IEEE Trans. Signal Process.*, vol. 55, no. 4, pp. 1486–1497, Apr. 2007.
- [22] D. Chizhik *et al.*, "Multiple-input-multiple-output measurements and modeling in manhattan," *IEEE J. Sel. Areas Commun.*, vol. 21, no. 3, pp. 321–331, Apr. 2003.
- [23] K. Yu *et al.*, "A wideband statistical model for nlos indoor MIMO channels," in *Proc. IEEE VTC Spring*, Birmingham, U.K., May 2002, pp. 370–374.
- [24] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1993.

- [25] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.
- [26] J. Yeh, *Real Analysis: Theory of Measure and Integration*, 2nd ed. Singapore: World Scientific, 2006.
- [27] E. Jorswieck and H. Boche, *Majorization and Matrix-Monotone Functions in Wireless Communications*. Delft, The Netherlands: Now Publishers Inc., 2007.
- [28] K. B. Petersen and M. S. Pedersen, "The matrix cookbook." Technical University of Denmark, Nov. 2012, Version 20121115. [Online]. Available: <http://www2.imm.dtu.dk/pubdb/p.php?3274>.
- [29] J.-B. Lasserre, "A trace inequality for matrix product," *IEEE Trans. Autom. Control*, vol. 40, no. 8, pp. 1500–1501, Aug. 1995.



**Rui Wang** (M'14) received the B.S. degree from Anhui Normal University, Wuhu, China, in 2006, and the M.S. degree from Shanghai University, Shanghai, China, in 2009 and the Ph.D. degree from Shanghai Jiao Tong University, China, in 2013, all in electronic engineering. From August 2012 to February 2013, he was a visiting Ph.D. student with the Department of Electrical Engineering, University of California, Riverside, CA, USA. From October 2013 to October 2014, he was with the Institute of Network Coding, The Chinese University of

Hong Kong as a Postdoctoral Research Associate. Since October 2014, he has been an Assistant Professor with the College of Electronics and Information Engineering, Tongji University. His research interests include wireless cooperative communications, MIMO technique, network coding, and OFDM etc.



**Meixia Tao** (S'00–M'04–SM'10) received the B.S. degree in electronic engineering from Fudan University, Shanghai, China, in 1999, and the Ph.D. degree in electrical and electronic engineering from Hong Kong University of Science and Technology in 2003. She is currently a Professor with the Department of Electronic Engineering, Shanghai Jiao Tong University, China. Prior to that, she was a Member of Professional Staff at Hong Kong Applied Science and Technology Research Institute during 2003–2004, and a Teaching Fellow then an Assistant Professor at

the Department of Electrical and Computer Engineering, National University of Singapore from 2004 to 2007. Her current research interests include cooperative communications, wireless resource allocation, MIMO techniques, and physical layer security.

Dr. Tao is an Editor for the *IEEE TRANSACTIONS ON COMMUNICATIONS* and the *IEEE WIRELESS COMMUNICATIONS LETTERS*. She was on the Editorial Board of the *IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS* from 2007 to 2011 and the *IEEE COMMUNICATIONS LETTERS* from 2009 to 2012. She also served as Guest Editor for *IEEE COMMUNICATIONS MAGAZINE* with feature topic on LTE-Advanced and 4G Wireless Communications in 2012, and Guest Editor for *EURISAP J WCN* with special issue on Physical Layer Network Coding for Wireless Cooperative Networks in 2010.

Dr. Tao is the recipient of the IEEE Heinrich Hertz Award for Best Communications Letters in 2013, the IEEE ComSoc Asia-Pacific Outstanding Young Researcher Award in 2009, and the International Conference on Wireless Communications and Signal Processing (WCSP) Best Paper Award in 2012.



**Hani Mehrpouyan** (S'05–M'10) received the B.Sc. honours degree from Simon Fraser University, Burnaby, Canada, in 2004 and the Ph.D. degree from Queens University, Kingston, Canada, in electrical engineering in 2010. From September 2010 to March 2012 he was a Post-Doc at Chalmers University of Technology, where he led the MIMO aspects of the microwave backhauling for next generation wireless networks project. He next joined the University of Luxembourg as a Research Associate from April 2012 to August 2012, where he was responsible for new interference cancellation and synchronization schemes for satellite communication links. Since 2012, he has been an Assistant Professor at California State University, Bakersfield, CA, USA. His current research interests lie in the area of applied signal processing and physical layer of wireless communication systems, including millimeter-wave systems, reconfigurable antennas, energy harvesting systems, synchronization, and channel estimation. For more information refer to [www.mehrpouyan.info](http://www.mehrpouyan.info).

Dr. Mehrpouyan has more than 40 publications in prestigious IEEE journals and conferences. He is an Associate Editor of *IEEE COMMUNICATION LETTERS* and he has also served as a TPC member for *IEEE Globecom*, *ICC*, *VTC*, etc. He has also been involved with industry leaders such as Ericsson AB, Blackberry etc. He has received many scholarships and awards, e.g., IEEE Globecom Early Bird Student Award, NSERC-IRDF, NSERC PGS-D, NSERC CGS-M Alexander Graham Bell, B.C. Wireless Innovation, and more.



**Yingbo Hua** (S'86–M'88–SM'92–F'02) received the B.S. degree from Southeast University, Nanjing, China, in 1982 and the Ph.D. degree from Syracuse University, Syracuse, NY, USA, in 1988. He held a faculty position with the University of Melbourne, Australia, where he was promoted to the rank of Reader and Associate Professor from 1996. He was a Visiting Professor with Hong Kong University of Science and Technology (1999–2000), and a Consultant with Microsoft Research, WA, USA (summer 2000). Since 2001, he has been a Professor with the

University of California at Riverside.

Dr. Hua has served as Editor, Guest Editor, Member of Editorial Board and/or Member of Steering Committee for *IEEE TRANSACTIONS ON SIGNAL PROCESSING*, *IEEE SIGNAL PROCESSING LETTERS*, *EURASIP Signal Processing*, *IEEE SIGNAL PROCESSING MAGAZINE*, *IEEE JOURNAL OF SELECTED AREAS IN COMMUNICATIONS*, and *IEEE WIRELESS COMMUNICATION LETTERS*. He has been a Member of IEEE Signal Processing Society's Technical Committees for Underwater Acoustic Signal Processing, Sensor Array and Multichannel Signal Processing, and Signal Processing for Communication and Networking. He has served on Technical and/or Organizing Committees for over 50 international conferences and workshops. He has authored/coauthored over 300 articles and coedited three volumes of books, with more than 8000 citations, in the fields of Sensing, Signal Processing and Communications. He is a Fellow of AAAS.