Iterative LMMSE Individual Channel Estimation over Relay Networks with Multiple Transmit and Receive Antennas

Jianpeng Ma, *Student Member, IEEE,* Shun Zhang, *Member, IEEE,* Hongyan Li, *Member, IEEE,* Nan Zhao, *Senior Member, IEEE,* and Arumugam Nallanathan, *Fellow, IEEE*

***Abstract*—In this paper, we investigate the individual channel estimation over the three-node one-way relay network (OWRN), where all nodes are equipped with multiple antennas. We first examine the scenario where the relay node is equipped with single antenna. An iterative linear minimum mean-square-error (LMMSE) method, which has fast convergence speed, is proposed to estimate the individual channels. The closed-form least square (LS) channel estimator is also derived through matrix unitary diagonalization to provide one good initialization point for the iterative LMMSE estimator. To evaluate the performance of the proposed algorithm, we present two performance lower bounds: Bayesian Cramér lower bound (BCRB) and linear estimation lower bound (LELB). Then, the training block and the relay amplification factor are optimized through minimizing the LELB. After that, our studies are extended to the general case where all the nodes are equipped with multiple nodes. Finally, numerical results are provided to corroborate our proposed studies.**

***Index Terms*—Bayesian Cramér lower bound, individual channel estimation, linear estimation lower bound, one-way relay network.**

I. Introduction

The cooperative relay network is efficient to combat the wireless channel fading and enhance the transmission capacity [2]–[4]. As one of the key technologies for the future wireless networks, the cooperative relay network has attracted more and more attention. Several efficient techniques, such as the relay beamforming [5], [6], the best relay selection [7], [8], the multi-antenna relay [9], [10], the distributed space time coding (DSTC) [11], [12], have been proposed to take the advantages of the wireless relay networks.

The accurate channel state information (CSI) at some or all the nodes in the relay networks is essential for achieving the promised performance of the above mentioned techniques. Several channel estimation algorithms at the destination nodes of the amplify-and-forward (AF) one-way relay networks (OWRN) have been proposed. In [15], the expectation conditional maximization (ECM) based iterative algorithm was applied to jointly estimate the channel and detect the data. In [16], the maximum a posteriori (MAP) based estimation schemes are developed to estimate the composite channel. In [17], Wang *et al.* investigated the effect of antenna correlations on the design of training sequences and channel estimation performance. In [18], one weighted least-square (WLS) channel estimator has been developed for the multiple-input multiple-output (MIMO) relay communication systems. The algorithms in [13]–[18] only estimate the cascaded channels from the source node to the relay node and then to the destination node. However, the individual channels (in-channels) along the link from the source node to the relay node and that from the relay node to the destination node are necessary for some optimal system designs in certain scenarios, such as the relay beamforming [19] and the subcarrier pairing [20].

To estimate the in-channels, Gao *et al.* proposed one superimposed training framework [22], where the relay superimposes its own training over the received one. Due to is advantages in efficiency and low-overhead, most of the recent in-channel estimation algorithms were proposed under the superimposed training framework [23]–[25]. In [23], Xie *et al.* designed two maximum likelihood (ML) estimators to recover the in-channels. In [24], Zhang *et al.* examined the performances of the two in-channel estimators under the superimposed training framework, i.e., the fully data-aided (FDA) estimator and the partially data-aided (PDA) estimator. The FDA estimator has knowledge about the training symbols from both the source and the relay, while the PDA estimator only has the statistical information about the data from the source and the full information of the training symbols superimposed by the relay. Moreover, Zhang *et al.* also studied the in-channel estimation under the time selective flat fading scenario [25]. Specifically, the unscented Kalman filter (UKF) was used to track the time-varying in-channels, while the unscented Rauch-Tung-Striebel smoother (URTSS) was adopted to smooth the UKF's results.

The works [22]–[25] only considered the classical relay scenario, where each node in the network is equipped with single antenna. Yet, some attentions have been paid to the in-channel estimation in the multiple-input and multiple-output (MIMO) based OWRN, where some or all nodes are equipped with multiple antennas. In [26], Rong *et al.* compared the two-stage and the superimposed training schemes for the in-channel estimation in the MIMO-based relay networks, and optimized the training structures through minimizing the mean-squared error (MSE) of the in-channel estimation. In [27], Jing *et al.* designed one low-complex approximated ML in-channel estimator for the three-node AF relay network

The authors are with the State Key Laboratory of Integrated Services Networks, Xidian University, Xi'an 710071, P. R. China. (Email: jpmaxdu@gmail.com; zhangshunsdu@gmail.com; hyli@mail.xidian.edu.cn).

N. Zhao is with the School of Inform. and Commun. Eng., Dalian University of Technology, Dalian, Liaoning, P. R. China (email: zhaonan@dlut.edu.cn).

A. Nallanathan is with the Department of Informatics, Kings College London, London, UK (Email: arumugam.nallanathan@kcl.ac.uk).

A preliminary version of this paper was published in part on proceedings of IEEE International Conference on Communications (ICC) [1].

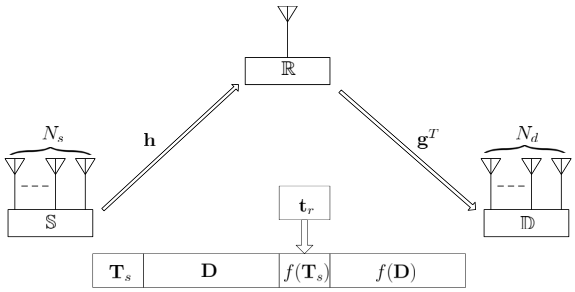

Fig. 1. The system model and transmission protocol for three-node OWRN with multiple antennas at both source and destination.

with multiple antennas at the source node and the destination node. In [28], the optimal training sequences in the sense of the minimal MSE is derived for the superimposed training based in-channel estimation under the MIMO-based relay networks. However, we still face some challenges. Since the number of parameters to be estimated in the multiple-antennas case is much larger than that in the single-antenna case, the ML, minimum MSE (MMSE) and maximum a posteriori probability (MAP) estimators for the in-channel estimation in multiple-antennas case require high dimensional searches or multiple integrals and cannot be implemented in an easy way. Theoretically, the linear MMSE (LMMSE) estimator is an alternative when the optimal ML, MMSE and MAP estimators are difficult to implement. Unfortunately, due to the product between the in-channels from the source to the relay and that from the relay to the destination, the observation data model at the destination is not a linear model with respect to each in-channel. Therefore, the LMMSE estimator cannot be directly utilized.

In this paper, we apply the superimposed training based in-channel estimation framework to the classical three-node OWRN, where both the source and the relay have multiple antennas. When the flat channel fading scenario is considered, we can achieve the following observation: for a specific realization of the in-channels along the link from the source to the relay (or from the relay to the destination), the observation data model at the destination is one Bayesian linear model with respect to the in-channels along the link from the relay to the destination (or from the source to the relay). Based on the above observation, we propose an iterative LMMSE in-channel estimator to address the challenges mentioned in the previous paragraph. Then, we formulate the closed-formed LS estimator through the matrix unitary diagonalization to provide one good initialization point for the iterative algorithm. In order to evaluate the proposed algorithm, we derive two performance lower bounds: the classical Bayesian Cramér lower bound (CRB) [29], [30], and the linear estimation lower bound (LELB). Interestingly, the LELB is proved to be more tighter than the Bayesian CRB. Moreover, we optimize the optimal training blocks and the relay's amplification factor through minimizing the LELB. Finally, our studies are extended to one general case, where all the three nodes (the source, the relay, and the destination) are equipped with multiple antennas.

The rest of the paper is organized as follows. The system model of the typical three-node OWRN with multiple-antennas at both the source and the destination is presented in Section II. In Section III, we develop the iterative LMMSE in-channel estimator under the superimposed training framework, complete its convergence analysis and derive the closed-formed LS estimator to provide the initialization point for the iterative estimator. Two performance lower bounds are given in Section IV. Then, we optimize the training structure and amplification factor in Section V. Section VI provide numerical simulation results to show the performance of the proposed algorithm. Moreover, in Section VII, our studies are extended to the scenario, where all the nodes in our system mode are equipped with multiple antennas. Finally, we draw our conclusions in Section VIII.

Notation: The vectors and the matrices are separately represented by the boldface small and capital letters. The matrix transpose, complex conjugate and Hermitian transpose are denoted by $(\cdot)^T, (\cdot)^*, (\cdot)^H$, respectively. $\mathrm{E}(\cdot)$ denotes the expectation operation. $[\mathbf{x}]_i$ is i-th entry of the vector $\mathbf{x}$, while $[\mathbf{X}]_{ij}$ is the (i,j)-th entry of the matrix $\mathbf{X}$. The $N\times N$ identity matrix is denoted by $\mathbf{I}_N$. $\mathrm{tr}(\mathbf{X})$ represents the trace of $\mathbf{X}$, and $\mathrm{vec}(\mathbf{X})$ denotes the column vector formed by stacking the columns of $\mathbf{X}$. $\otimes$ is the Kronecker product, and $\|\mathbf{x}\|^2$ denotes the two-norm of the vector $\mathbf{x}$. $|\mathbf{X}|$ is the determine of the matrix $\mathbf{X}$.

II. System Model

As shown in Fig. 1, we consider a typical three-node OWRN with one source $\mathbb{S}$, one relay $\mathbb{R}$, and one destination $\mathbb{D}$ under the flat fading channel scenario. There is only one antenna at $\mathbb{R}$, while the nodes $\mathbb{S}$ and $\mathbb{D}$ are equipped with N_s and N_d antennas, respectively. Without loss of generality, we do not consider the direct link between $\mathbb{S}$ and $\mathbb{D}$. The in-channel vectors along the links $\mathbb{S}\to\mathbb{R}$ and $\mathbb{R}\to\mathbb{D}$ can be separately written as $\mathbf{h}=[h_0,h_1,\ldots,h_{N_s-1}]^T\in\mathbb{C}^{N_s\times 1}$ and $\mathbf{g}^T=[g_0,g_1,\ldots,g_{N_d-1}]\in\mathbb{C}^{1\times N_d}$, where h_i represents the channel between the i-th antenna of $\mathbb{S}$ and $\mathbb{R}$, and g_j denotes the channel between $\mathbb{R}$ and the j-th antenna of $\mathbb{D}$. All the in-channels are assumed to be zero-mean circularly symmetric complex Gaussian (CSCG) random variables, and the covariance matrices for $\mathbf{h}$ and $\mathbf{g}$ can be separately written as $\mathbf{R}_{\mathbf{h}}=\sigma_h^2\mathbf{I}_{N_s}$ and $\mathbf{R}_{\mathbf{g}}=\sigma_g^2\mathbf{I}_{N_d}$. Furthermore, we assume that $\mathbf{h}$ is independent of $\mathbf{g}$, and both $\mathbf{h}$ and $\mathbf{g}$ are block fading, which means that they do not change within one phase but may change from phase to phase.

Under the superimposed training framework, one round of the training symbols transmission from $\mathbb{S}$ to $\mathbb{D}$ can be partitioned into two phases. During phase I, $\mathbb{S}$ sends one source-training block $\mathbf{T}_s\in\mathbb{C}^{\tau\times N_s}$ to $\mathbb{R}$, where τ is the number of the time slots for training block transmission. The received training at $\mathbb{R}$ during phase I can be expressed as

$$\mathbf{r}_r=\mathbf{T}_s\mathbf{h}+\mathbf{n}_r, \tag{1}$$

where $\mathbf{n}_r\in\mathbb{C}^{\tau\times 1}$ is the additive white Gaussian noise (AWGN) vector with zero mean and the covariance matrix $\sigma_n^2\mathbf{I}_\tau$.

During phase II, $\mathbb{R}$ superimposes its training sequence $\mathbf{t}_r \in \mathcal{C}^{\tau\times 1}$ over its received training $\mathbf{r}_r$, and the resultant training from $\mathbb{R}$ can be denoted as [22]–[25]

$$\mathbf{r}_t = \alpha\mathbf{r}_r + \mathbf{t}_r. \tag{2}$$

Then, $\mathbb{R}$ forwards $\mathbf{r}_t$ to $\mathbb{D}$, and the received training at $\mathbb{D}$ can be written as

$$\mathbf{Y} = \alpha\mathbf{T}_s\mathbf{h}\mathbf{g}^T + \mathbf{t}_r\mathbf{g}^T + \alpha\mathbf{n}_r\mathbf{g}^T + \mathbf{N}_d, \tag{3}$$

where $\mathbf{N}_d \in \mathcal{C}^{\tau\times N_d}$ denotes the AWGN matrix at $\mathbb{D}$, whose entries are i.i.d CSCG random variables with zero mean and variance σ_n^2.

We set the average transmitting power of $\mathbb{R}$ as P_r, i.e.,

$$\begin{aligned} P_r =& \mathbf{E}_{\mathbf{h}}\{\|\alpha(\mathbf{T}_s\mathbf{h}+\mathbf{n}_r)+\mathbf{t}_r\|^2\} \\ =& \alpha^2\sigma_h^2P_s + \tau\alpha^2\sigma_n^2 + P_t, \end{aligned} \tag{4}$$

where $P_s = \mathrm{tr}\{\mathbf{T}_s\mathbf{T}_s^H\}$ is the power of the source training block and $P_t = ||\mathbf{t}_r||^2$ is the power of the relay training sequence . Then the amplification factor α can be expressed as

$$\alpha = \sqrt{\frac{P_r - P_t}{\sigma_h{}^2P_s + \tau\sigma_n^2}}. \tag{5}$$

III. In-Channel Estimation

A. Iterative LMMSE In-Channel Estimator

Before preceding, let us define $\mathbf{y} = \mathrm{vec}(\mathbf{Y})$ and $\mathbf{n}_d = \mathrm{vec}(\mathbf{N}_d)$. Resorting to the Kronecker product property $\mathrm{vec}(\mathbf{ABC}) = (\mathbf{C}^T \otimes \mathbf{A})\mathrm{vec}(\mathbf{B})$ [31], we can obtain

$$\begin{aligned} \mathbf{y} =& \alpha(\mathbf{g}\otimes\mathbf{T}_s)\mathbf{h} + (\mathbf{g}\otimes\mathbf{I}_\tau)\mathbf{t}_r + \underbrace{\alpha(\mathbf{g}\otimes\mathbf{I}_\tau)\mathbf{n}_r + \mathbf{n}_d}_{\mathbf{n}} \\ =& \big(\alpha(\mathbf{I}_{N_d}\otimes\mathbf{T}_s\mathbf{h}) + (\mathbf{I}_{N_d}\otimes\mathbf{t}_r)\big)\mathbf{g} + \mathbf{n}, \end{aligned} \tag{6}$$

where the equivalent noise vector $\mathbf{n}$ defined as the corresponding term is related to the specific realization of $\mathbf{g}$.

When the in-channel statistics is known, the optimal in-channel estimation method is the MAP estimator, which can be explicitly expressed as

$$\{\hat{\mathbf{h}}, \hat{\mathbf{g}}\} = \arg\max_{\mathbf{h},\mathbf{g}} p(\mathbf{y}|\mathbf{h},\mathbf{g})p(\mathbf{h},\mathbf{g}), \tag{7}$$

where $p(\mathbf{y}|\mathbf{h},\mathbf{g})$ denotes the probability density function (PDF) of $\mathbf{y}$ conditioned on $\mathbf{h}$, $\mathbf{g}$, and $p(\mathbf{h},\mathbf{g})$ is the joint PDF of $\mathbf{h}$, $\mathbf{g}$. With the equation (6) and the statistical characteristics of $\mathbf{h}$ and $\mathbf{g}$, we can derive the following equations as

$$\ln p(\mathbf{y}|\mathbf{h},\mathbf{g}) = \mathrm{Const.} - \ln|\mathbf{C}_{\mathbf{n}|\mathbf{h},\mathbf{g}}| - (\mathbf{y}-\boldsymbol{\mu})^H\mathbf{C}_{\mathbf{n}|\mathbf{h},\mathbf{g}}^{-1}(\mathbf{y}-\boldsymbol{\mu}), \tag{8}$$

$$\ln p(\mathbf{h},\mathbf{g}) = \mathrm{Const.} - \sigma_h^{-2}\|\mathbf{h}\|^2 - \sigma_g^{-2}\|\mathbf{g}\|^2, \tag{9}$$

where

$$\boldsymbol{\mu} = \alpha(\mathbf{g}\otimes\mathbf{T}_s)\mathbf{h} + (\mathbf{g}\otimes\mathbf{I}_\tau)\mathbf{t}_r = \big(\alpha(\mathbf{I}_{N_d}\otimes\mathbf{T}_s\mathbf{h}) + (\mathbf{I}_{N_d}\otimes\mathbf{t}_r)\big)\mathbf{g},$$

and

$$\mathbf{C}_{\mathbf{n}|\mathbf{h},\mathbf{g}} = \mathrm{E}\{\mathbf{n}\mathbf{n}^H|\mathbf{h},\mathbf{g}\} = \sigma_n^2(\alpha^2\mathbf{g}\mathbf{g}^H + \mathbf{I}_{N_d})\otimes\mathbf{I}_\tau.$$

It is noticed that the Kronecker product property $(\mathbf{A}\otimes\mathbf{B})(\mathbf{C}\otimes\mathbf{D}) = \mathbf{AC}\otimes\mathbf{BD}$ is utilized in the above derivation. After straightforward calculation, the MAP in-channel estimator can be represented as

$$\begin{aligned} \{\hat{\mathbf{h}}, \hat{\mathbf{g}}\} =& \arg\min_{\mathbf{h},\mathbf{g}} (\mathbf{y}-\boldsymbol{\mu})^H\mathbf{C}_{\mathbf{n}|\mathbf{h},\mathbf{g}}^{-1}(\mathbf{y}-\boldsymbol{\mu}) \\ &+ \ln|\mathbf{C}_{\mathbf{n}|\mathbf{h},\mathbf{g}}| + \sigma_h^{-2}\|\mathbf{h}\|^2 + \sigma_g^{-2}\|\mathbf{g}\|^2. \end{aligned} \tag{10}$$

Unfortunately, due to its complicated structure, especially the presence of $\mathbf{C}_{\mathbf{n}|\mathbf{h},\mathbf{g}}$, the MAP estimator requires a high dimensional search and is difficult to implement. Theoretically, the LMMSE estimator, which minimizes the MSE for the unknown parameters' estimation under the constraint that the estimator must be linear, can be adopted as one suboptimal estimator under the Bayesian framework. However, due to the presence of nonlinear term $(\mathbf{g}\otimes\mathbf{T}_s)\mathbf{h}$ or $(\mathbf{I}_{N_d}\otimes\mathbf{T}_s\mathbf{h})\mathbf{g}$, the observation data model in (6) is not a Bayesian linear model with respect to $\mathbf{h}$ and $\mathbf{g}$. Thus, the Bayesian Gaussian-Markov theorem does not hold here, and the LMMSE estimator cannot be directly used to estimate the in-channels $\mathbf{h}$ and $\mathbf{g}$. Instead, we would like to conceive an iterative LMMSE in-channel estimator.

The proposed iterative LMMSE in-channel estimator is based on the following observation: for specific realization of $\mathbf{g}$ (or $\mathbf{h}$), the observation data model of $\mathbf{y}$ in (6) is a Bayesian linear model with respect to $\mathbf{h}$ (or $\mathbf{g}$). Hence, with the Bayesian Gaussian-Markov theorem [32], the LMMSE estimation of $\mathbf{h}$ conditioned on given $\mathbf{g}$ can be formulated as

$$\begin{aligned} \hat{\mathbf{h}}|\mathbf{g} =& \alpha\left(\mathbf{R}_{\mathbf{h}}^{-1} + \alpha^2(\mathbf{g}\otimes\mathbf{T}_s)^H\mathbf{C}_{\mathbf{n}|\mathbf{g}}^{-1}(\mathbf{g}\otimes\mathbf{T}_s)\right)^{-1} \\ &(\mathbf{g}\otimes\mathbf{T}_s)^H\mathbf{C}_{\mathbf{n}|\mathbf{g}}^{-1}(\mathbf{y} - (\mathbf{g}\otimes\mathbf{I}_\tau)\mathbf{t}_r), \end{aligned} \tag{11}$$

where the covariance matrix

$$\mathbf{C}_{\mathbf{n}|\mathbf{g}} = \mathrm{E}\{\mathbf{n}\mathbf{n}^H|\mathbf{g}\} = \sigma_n^2(\alpha^2\mathbf{g}\mathbf{g}^H + \mathbf{I}_{N_d})\otimes\mathbf{I}_\tau. \tag{12}$$

Utilizing the matrix property equations $(\mathbf{I}+\mathbf{AB})^{-1} = \mathbf{I} - \mathbf{A}(\mathbf{I}+\mathbf{BA})^{-1}\mathbf{B}$ and $(\mathbf{A}\otimes\mathbf{B})^{-1} = \mathbf{A}^{-1}\otimes\mathbf{B}^{-1}$, we can derive the inverse of $\mathbf{C}_{\mathbf{n}|\mathbf{g}}$ as

$$\begin{aligned} \mathbf{C}_{\mathbf{n}|\mathbf{g}}^{-1} =& \sigma_n^{-2}\Big(\mathbf{I}_{N_d} - \alpha^2\mathbf{g}(1+\alpha^2\mathbf{g}^H\mathbf{g}^*)^{-1}\mathbf{g}^H\Big)\otimes\mathbf{I}_\tau \\ =& \sigma_n^{-2}\left(\mathbf{I}_{N_d} - \frac{\alpha^2}{1+\alpha^2\|\mathbf{g}\|^2}\mathbf{g}\mathbf{g}^H\right)\otimes\mathbf{I}_\tau. \end{aligned} \tag{13}$$

Substituting (13) into (11), we can reexpress (11) as

$$\begin{aligned} \hat{\mathbf{h}}|\mathbf{g} =& \alpha\left(\mathbf{R}_{\mathbf{h}}^{-1} + \frac{\alpha^2\|\mathbf{g}\|^2}{\sigma_n^2(1+\alpha^2\|\mathbf{g}\|^2)}\mathbf{T}_s^H\mathbf{T}_s\right)^{-1} \\ &\left(\frac{\mathbf{g}^H}{\sigma_n^2(1+\alpha^2\|\mathbf{g}\|^2)}\otimes\mathbf{T}_s^H\right)\Big(\mathbf{y} - (\mathbf{g}\otimes\mathbf{I}_\tau)\mathbf{t}_r\Big) \\ =& \frac{\alpha}{\sigma_n^2(1+\alpha^2\|\mathbf{g}\|^2)}\left(\mathbf{R}_{\mathbf{h}}^{-1} + \frac{\alpha^2\|\mathbf{g}\|^2}{\sigma_n^2(1+\alpha^2\|\mathbf{g}\|^2)}\mathbf{T}_s^H\mathbf{T}_s\right)^{-1} \\ &\mathbf{T}_s^H\Big(\mathbf{Y} - \mathbf{t}_r\mathbf{g}^T\Big)\mathbf{g}^*, \end{aligned} \tag{14}$$

where the matrix equation $\mathrm{vec}(\mathbf{ABC}) = (\mathbf{C}^T\otimes\mathbf{A})\mathrm{vec}(\mathbf{B})$ is used in the derivation.

Let us further investigate the estimation of $\mathbf{g}$ for a specific $\mathbf{h}$. Similarly, the covariance matrix of $\mathbf{n}$ conditioned on $\mathbf{h}$ can be denoted as

$$\mathbf{C}_{\mathbf{n}|\mathbf{h}} = \mathrm{E}\{\mathbf{n}\mathbf{n}^H|\mathbf{h}\} = \sigma_n^2(\alpha^2\mathbf{R}_{\mathbf{g}} + \mathbf{I}_{N_d})\otimes\mathbf{I}_\tau. \tag{15}$$

Algorithm 1 The Iterative LMMSE In-Channel Estimator for In-Channels $\mathbf{h}$ and $\mathbf{g}$

Initialization: Initial LS in-channel estimation

1: Unitarily diagonalize the Hermitian matrix $\mathbf{Y}^H\mathbf{\Xi}\mathbf{Y}$ to obtain the eigenvector that corresponds to the maximal eigenvalue of $\mathbf{Y}^H\mathbf{\Xi}\mathbf{Y}$, i.e, $\mathbf{v}_0$, and derive $\widehat{\boldsymbol{g}}_{LS}$ according to (29).
2: Calculate $\widehat{\|\mathbf{g}\|}_{LS}$ according to (23), and obtain $\widehat{\mathbf{g}}_{LS}$ as $\widehat{\mathbf{g}}_{LS} = \widehat{\|\mathbf{g}\|}_{LS}\widehat{\boldsymbol{g}}_{LS}$.
3: Calculate $\widehat{\mathbf{h}}_{LS}$ according to (19) by substituting $\widehat{\mathbf{g}}_{LS}$ into it.
4: k= 0, $\hat{\mathbf{h}}_0 = \widehat{\mathbf{h}}_{LS}$, $\hat{\mathbf{g}}_0 = \widehat{\mathbf{g}}_{LS}$.
5: **repeat**
6: k=k+1.
7: Calculate the current $\hat{\mathbf{h}}_k$ with the linear estimator (14) by substituting the previous $\widehat{\mathbf{g}}_{k-1}$ into it.
8: Calculate the current $\hat{\mathbf{g}}_k$ with the linear estimator (17) by substituting the current $\hat{\mathbf{h}}_k$ into it.
9: **until** some termination criterion is satisfied.
10: **return** $\hat{\mathbf{h}}_k$,$\hat{\mathbf{g}}_k$.

Resorting to the fact $\mathbf{R}_{\mathbf{g}} = \sigma_g^2\mathbf{I}_{N_d}$, we can derive the inverse of $\mathbf{C}_{\mathbf{n}|\mathbf{h}}$ as

$$\begin{aligned}\mathbf{C}_{\mathbf{n}|\mathbf{h}}^{-1} =&\sigma_n^{-2}(\alpha^2\sigma_g^2\mathbf{I}_{N_d} + \mathbf{I}_{N_d})^{-1}\otimes\mathbf{I}_\tau\\ =&\frac{1}{\sigma_n^2(\alpha^2\sigma_g^2+1)}\mathbf{I}_{\tau N_d}.\end{aligned} \tag{16}$$

Then, for given $\mathbf{h}$, the standard LMMSE estimation of $\mathbf{g}$ is given by

$$\begin{aligned}\hat{\mathbf{g}}|\mathbf{h} =&\Big(\mathbf{R}_{\mathbf{g}}^{-1} + \big(\mathbf{I}_{N_d}\otimes(\alpha\mathbf{T}_s\mathbf{h}+\mathbf{t}_r)\big)^H\mathbf{C}_{\mathbf{n}|\mathbf{h}}^{-1}\big(\mathbf{I}_{N_d}\otimes\\ &(\alpha\mathbf{T}_s\mathbf{h}+\mathbf{t}_r)\big)\Big)^{-1}\Big(\mathbf{I}_{N_d}\otimes(\alpha\mathbf{T}_s\mathbf{h}+\mathbf{t}_r)\Big)^H\mathbf{C}_{\mathbf{n}|\mathbf{h}}^{-1}\mathbf{y}\\ =&\frac{1}{\sigma_n^2(\alpha^2\sigma_g^2+1)}\Big(\mathbf{R}_{\mathbf{g}}^{-1}+\frac{\|\alpha\mathbf{T}_s\mathbf{h}+\mathbf{t}_r\|^2}{\sigma_n^2(\alpha^2\sigma_g^2+1)}\mathbf{I}_{N_d}\Big)^{-1}\\ &\Big(\mathbf{I}_{N_d}\otimes(\alpha\mathbf{T}_s\mathbf{h}+\mathbf{t}_r)^H\Big)\mathbf{y}\\ =&\Big(\sigma_n^2(\alpha^2+\sigma_g^{-2})+\|\alpha\mathbf{T}_s\mathbf{h}+\mathbf{t}_r\|^2\Big)^{-1}\mathbf{Y}^T(\alpha\mathbf{T}_s\mathbf{h}+\mathbf{t}_r)^*\end{aligned} \tag{17}$$

B. Initial LS In-Channel Estimator

As is well known, a good initial point is essential for an iterative estimator since it affects the convergence point and speed of convergence. Obviously, the LS estimator

$$\begin{aligned}\{\hat{\mathbf{h}}_{LS},\hat{\mathbf{g}}_{LS}\} =& \arg\min_{\mathbf{h},\mathbf{g}}\|\mathbf{y}-\alpha(\mathbf{g}\otimes\mathbf{T}_s)\mathbf{h}-(\mathbf{g}\otimes\mathbf{I}_\tau)\mathbf{t}_r\|^2\\ =&\arg\min_{\mathbf{h},\mathbf{g}}\|\mathbf{Y}-\alpha\mathbf{T}_s\mathbf{h}\mathbf{g}^T-\mathbf{t}_r\mathbf{g}^T\|_F^2\end{aligned} \tag{18}$$

could serve as a good initial point for the iterative LMMSE in-channel estimator.

With given $\mathbf{g}$, the LS estimation of $\mathbf{h}$ can be expressed as

$$\begin{aligned}\hat{\mathbf{h}}_{LS} =&\alpha^{-1}\|\mathbf{g}\|^{-2}\Big(\mathbf{g}^H\otimes(\mathbf{T}_s^H\mathbf{T}_s)\mathbf{T}_s^H\Big)(\mathbf{y}-(\mathbf{g}\otimes\mathbf{I}_\tau)\mathbf{t}_r)\\ =&\alpha^{-1}\|\mathbf{g}\|^{-2}(\mathbf{T}_s^H\mathbf{T}_s)^{-1}\mathbf{T}_s^H(\mathbf{Y}-\mathbf{t}_r\mathbf{g}^T)\mathbf{g}^*.\end{aligned} \tag{19}$$

Incorporating (19) into (18) and recalling (6), we can rewrite (18) as

$$\widehat{\mathbf{g}}_{LS}=\arg\min_{\mathbf{g}}\left\|\mathbf{Y}-\mathbf{t}_r\mathbf{g}^T-\|\mathbf{g}\|^{-2}\mathbf{W}(\mathbf{Y}-\mathbf{t}_r\mathbf{g}^T)\mathbf{g}^*\mathbf{g}^T\right\|_F^2, \tag{20}$$

where $\mathbf{W}=\mathbf{T}_s(\mathbf{T}_s^H\mathbf{T}_s)^{-1}\mathbf{T}_s^H$. For simplicity, instead of minimizing the right hand side (RHS) of (20) through optimizing $\mathbf{g}$ directly, we can minimize it first with respect to the norm of $\mathbf{g}$, i.e., $\|\mathbf{g}\|$, and then with respect to the normalized version of $\mathbf{g}$, i.e., $\boldsymbol{g}=\frac{\mathbf{g}}{\|\mathbf{g}\|}$. Hence, we should reformulate (20) as

$$\begin{aligned}&\left\{\widehat{\boldsymbol{g}}_{LS},\widehat{\|\mathbf{g}\|}_{LS}\right\} \quad (21)\\ &=\arg\min_{\boldsymbol{g},\|\mathbf{g}\|}\left\|\mathbf{Y}-\|\mathbf{g}\|\mathbf{t}_r\boldsymbol{g}^T-\mathbf{W}(\mathbf{Y}-\|\mathbf{g}\|\mathbf{t}_r\boldsymbol{g}^T)\boldsymbol{g}^*\boldsymbol{g}^T\right\|_F^2\\ &=\arg\min_{\boldsymbol{g},\|\mathbf{g}\|}\left\|\mathbf{Y}-\|\mathbf{g}\|\mathbf{t}_r\boldsymbol{g}^T\right\|_F^2-\left\|\mathbf{W}(\mathbf{Y}-\|\mathbf{g}\|\mathbf{t}_r\boldsymbol{g}^T)\boldsymbol{g}^*\boldsymbol{g}^T\right\|_F^2\\ &=\arg\min_{\boldsymbol{g},\|\mathbf{g}\|}\|\mathbf{g}\|^2\|(\mathbf{I}_\tau-\mathbf{W})\mathbf{t}_r\|^2\\ &\quad-2\|\mathbf{g}\|\Re\left(\mathbf{t}_r^H(\mathbf{I}_\tau-\mathbf{W})\mathbf{Y}\boldsymbol{g}^*\right)-\|\mathbf{W}\mathbf{Y}\boldsymbol{g}^*\|^2.\end{aligned} \tag{22}$$

Fixing $\boldsymbol{g}$ and taking the first order derivative of (22)'s RHS term with respect to $\|\mathbf{g}\|$, we can derive the LS estimation of $\|\mathbf{g}\|$ as

$$\widehat{\|\mathbf{g}\|}_{LS}=\|(\mathbf{I}_\tau-\mathbf{W})\mathbf{t}_r\|^{-2}\Re\left(\mathbf{t}_r^H(\mathbf{I}_\tau-\mathbf{W})\mathbf{Y}\boldsymbol{g}^*\right). \tag{23}$$

Let us substitute (23) back into (22). Then, the LS estimation of $\mathbf{g}$ can be achieved as

$$\widehat{\boldsymbol{g}}_{LS}=\arg\max_{\boldsymbol{g}}L(\boldsymbol{g}), \tag{24}$$

where

$$L(\boldsymbol{g})=\Re^2(\mathbf{t}_r^H(\mathbf{I}_\tau-\mathbf{W})\mathbf{Y}\boldsymbol{g}^*)+\|(\mathbf{I}_\tau-\mathbf{W})\mathbf{t}_r\|^2\|\mathbf{W}\mathbf{Y}\boldsymbol{g}^*\|^2.$$

From the basic inequalities, we can prove that $L(\boldsymbol{g})$ satisfy

$$L(\boldsymbol{g})\le\boldsymbol{g}^T\mathbf{Y}^H\mathbf{\Xi}\mathbf{Y}\boldsymbol{g}^*, \tag{25}$$

where the Hermitian matrix $\mathbf{\Xi}=\|(\mathbf{I}_\tau-\mathbf{W})\mathbf{t}_r\|^2\mathbf{W}+(\mathbf{I}_\tau-\mathbf{W})\mathbf{t}_r\mathbf{t}_r^H(\mathbf{I}_\tau-\mathbf{W})$ is only related to the training matrices $\mathbf{T}_s$, $\mathbf{t}_r$, and the equality holds when the term $\mathbf{t}_r^H(\mathbf{I}_\tau-\mathbf{W})\mathbf{Y}\boldsymbol{g}^*$ is real. Note that the quadratic polynomial $\boldsymbol{g}^T\mathbf{Y}^H\mathbf{\Xi}\mathbf{Y}\boldsymbol{g}^*$ can serve as an upper bound of the objective function $L(\boldsymbol{g})$. Instead of directly solving (24), we can first look into the following optimization:

$$\boldsymbol{\nu}_{opt}=\arg\max_{\|\boldsymbol{\nu}\|^2=1}\boldsymbol{\nu}^H\mathbf{Y}^H\mathbf{\Xi}\mathbf{Y}\boldsymbol{\nu}, \tag{26}$$

where $\boldsymbol{\nu}$ denotes the $N_d\times 1$ column vector. It can be readily checked that $\mathbf{Y}^H\mathbf{\Xi}\mathbf{Y}$ is Hermitian and can be decomposed as [31]

$$\mathbf{Y}^H\mathbf{\Xi}\mathbf{Y}=\mathbf{V}\mathrm{diag}(\lambda_0,\lambda_1,\ldots,\lambda_{N_d-1})\mathbf{V}^H, \tag{27}$$

where $\mathbf{V}=[\mathbf{v}_0,\mathbf{v}_1,\ldots,\mathbf{v}_{N_d-1}]$ is an unitary matrix, i.e., $\mathbf{V}\mathbf{V}^H=\mathbf{V}^H\mathbf{V}=\mathbf{I}$ and the real variables $\lambda_0,\lambda_1,\ldots,\lambda_{N_d-1}$ are the eigenvalues of $\mathbf{Y}^H\mathbf{\Xi}\mathbf{Y}$. Without loss of generality, we assume $\lambda_0\ge\lambda_1\ge\ldots\ge\lambda_{N_d-1}$.

From the Rayleigh-Ritz theorem [31], it can be obtained that

$$\lambda_0 = \max_{\|\boldsymbol{\nu}\|^2=1} \boldsymbol{\nu}^H \mathbf{Y}^H \boldsymbol{\Xi} \mathbf{Y} \boldsymbol{\nu}, \tag{28}$$

where the maximal value is achieved, if $\boldsymbol{\nu}_{opt}$ equals the eigenvector of $\mathbf{Y}^H \boldsymbol{\Xi} \mathbf{Y}$ corresponding to the eigenvalue λ_0, i.e., $\boldsymbol{\nu}_{opt} = \mathbf{v}_0$.

With the above analysis, we can conclude that $L(\boldsymbol{g}) \leq \mathbf{v}_0^H \mathbf{Y}^H \boldsymbol{\Xi} \mathbf{Y} \mathbf{v}_0$ and the upper bound can be achieved only if

$$\widehat{\mathbf{g}}_{LS} = \mathbf{v}_0^* e^{j\angle \mathbf{W}\mathbf{Y}\mathbf{v}_0}. \tag{29}$$

Let us use $\hat{\mathbf{h}}_k$ and $\hat{\mathbf{g}}_k$ to denote the estimations of $\mathbf{h}$ and $\mathbf{g}$ during the k-th iteration, respectively. Then, the proposed iterative LMMSE in-channel estimation algorithm is outlined in Algorithm 1.

C. Convergence Analysis

To analyse the convergence property of proposed iterative LMMSE in-channel estimator, we derive the following lemma.

Lemma 1: Define the variable $a > 0$ and the $N \times N$ positive semi-definite Hermitian matrix $\mathbf{X}$. For the following matrix function

$$f(\mathbf{X}) = \left(a\mathbf{I} + \mathbf{A}^H \mathbf{X}^{-1} \mathbf{A}\right)^{-1}, \tag{30}$$

we have $\mathrm{tr}\left(f(\mathbf{X}_2)\right) - \mathrm{tr}\left(f(\mathbf{X}_1)\right) < 0$ if $\mathrm{tr}\left(\mathbf{X}_2\right) < \mathrm{tr}\left(\mathbf{X}_1\right)$.

Proof: See Appendix A. ■

It is clear that the current estimation of $\mathbf{h}$ ($\mathbf{g}$) is subjected to the error propagation from the previous imperfect estimation of $\mathbf{g}$ ($\mathbf{h}$) . During the k-th iteration, the average MSEs (AMSEs) for both $\hat{\mathbf{h}}_k$ and $\hat{\mathbf{g}}_k$ can be separately expressed as

$$\mathrm{tr}\left(\mathbf{C}_{\Delta\mathbf{h}_k}\right) = \mathrm{tr}\left(\mathrm{E}\left\{\Delta\mathbf{h}_k \Delta\mathbf{h}_k^H\right\}\right), \tag{31}$$

$$\mathrm{tr}\left(\mathbf{C}_{\Delta\mathbf{g}_k}\right) = \mathrm{tr}\left(\mathrm{E}\left\{\Delta\mathbf{g}_k \Delta\mathbf{g}_k^H\right\}\right). \tag{32}$$

where $\Delta\mathbf{h}_k = \mathbf{h} - \hat{\mathbf{h}}_k$ and $\Delta\mathbf{g}_k = \mathbf{g} - \hat{\mathbf{g}}_k$. Plugging $\mathbf{g} = \hat{\mathbf{g}}_k + \Delta\mathbf{g}_k$ into (6), we can obtain that

$$\begin{aligned}\mathbf{y} =& \alpha(\hat{\mathbf{g}}_k \otimes \mathbf{T}_s)\mathbf{h} + (\hat{\mathbf{g}}_k \otimes \mathbf{I}_\tau)\mathbf{t}_r \\ &+ \underbrace{\alpha(\Delta\mathbf{g}_k \otimes \mathbf{T}_s)\mathbf{h} + (\Delta\mathbf{g}_k \otimes \mathbf{I}_\tau)\mathbf{t}_r + \mathbf{n}}_{\boldsymbol{\omega}|\Delta\mathbf{g}_k},\end{aligned} \tag{33}$$

where $\boldsymbol{\omega}|\Delta\mathbf{g}_k$ represents the equivalent noise during the estimation of $\mathbf{h}$ conditioned on given $\Delta\mathbf{g}_k$. Correspondingly, its covariance matrix can be listed as

$$\begin{aligned}\mathbf{C}_{\boldsymbol{\omega}|\Delta\mathbf{g}_k} =& \alpha^2\sigma_h^2 \mathbf{C}_{\Delta\mathbf{g}_k} \otimes (\mathbf{T}_s {\mathbf{T}_s}^H) \\ &+ (\mathbf{I}_{N_d} \otimes \mathbf{t}_r)\mathbf{C}_{\Delta\mathbf{g}_k}(\mathbf{I}_{N_d} \otimes \mathbf{t}_r^H) + \mathbf{C}_{\mathbf{n}|\mathbf{g}}.\end{aligned} \tag{34}$$

Furthermore, we can derive the covariance matrix for the LMMSE estimation $\hat{\mathbf{h}}_{k+1}$ conditioned on given $\mathbf{C}_{\Delta\mathbf{g}_k}$ as [32]

$$\begin{aligned}&\mathbf{C}_{\Delta\mathbf{h}_{k+1}} | \mathbf{C}_{\Delta\mathbf{g}_k} \\ &= \left(\mathbf{R}_\mathbf{h}^{-1} + \alpha^2(\hat{\mathbf{g}}_k^H \otimes \mathbf{T}_s^H)(\mathbf{C}_{\boldsymbol{\omega}|\hat{\mathbf{g}}_k})^{-1}(\hat{\mathbf{g}}_k \otimes \mathbf{T}_s)\right)^{-1}.\end{aligned} \tag{35}$$

Since all the matrices terms in (34) and (35) are positive semi-definite, it can be concluded from Lemma 1 that the decrease of the AMSE for $\hat{\mathbf{g}}_k$, i.e. $\mathrm{tr}(\mathbf{C}_{\Delta\mathbf{g}_k})$, will reduce the AMSE of $\hat{\mathbf{h}}_{k+1}$, i.e. $\mathrm{tr}(\mathbf{C}_{\Delta\mathbf{h}_{k+1}})$. Similarly, substituting $\mathbf{h} = \hat{\mathbf{h}}_{k+1} + \Delta\mathbf{h}_{k+1}$ into (6),we get

$$\begin{aligned}\mathbf{y} =& \left(\alpha(\mathbf{I}_{N_d} \otimes \mathbf{T}_s\hat{\mathbf{h}}_{k+1}) + (\mathbf{I}_{N_d} \otimes \mathbf{t}_r)\right)\mathbf{g} \\ &+ \underbrace{\alpha(\mathbf{I}_{N_d} \otimes \mathbf{T}_s\Delta\mathbf{h}_{k+1})\mathbf{g} + \mathbf{n}}_{\boldsymbol{\omega}|\Delta\mathbf{h}_{k+1}},\end{aligned} \tag{36}$$

and

$$\mathbf{C}_{\boldsymbol{\omega}|\Delta\mathbf{h}_{k+1}} = \alpha^2\sigma_g^2(\mathbf{I}_{Nd} \otimes \mathbf{T}_s)\mathbf{C}_{\Delta\mathbf{h}_{k+1}}(\mathbf{I}_{Nd} \otimes \mathbf{T}_s^H) + \mathbf{C}_{\mathbf{n}|\mathbf{h}}. \tag{37}$$

Then, the covariance matrix for $\mathbf{g}$'s estimation error conditioned on given $\mathbf{C}_{\Delta\mathbf{h}_{k+1}}$ in $(k+1)$-th iteration step can be written as

$$\begin{aligned}\mathbf{C}_{\Delta\mathbf{g}_{k+1}} | \mathbf{C}_{\Delta\mathbf{h}_{k+1}} =& \Big(\mathbf{R}_\mathbf{g}^{-1} + \left(\mathbf{I}_{N_d} \otimes (\alpha\mathbf{T}_s\hat{\mathbf{h}}_{k+1} + \mathbf{t}_r)\right)^H \\ &\left(\mathbf{C}_{\boldsymbol{\omega}|\hat{\mathbf{h}}_{k+1}}\right)^{-1}\left(\mathbf{I}_{N_d} \otimes (\alpha\mathbf{T}_s\hat{\mathbf{h}}_{k+1} + \mathbf{t}_r)\right)\Big)^{-1}.\end{aligned} \tag{38}$$

Basing on (37) (38) and Lemma1, we can also find that decrease of the AMSE for $\hat{\mathbf{h}}_{k+1}$, i.e. $\mathrm{tr}(\mathbf{C}_{\Delta\mathbf{h}_{k+1}})$, will reduce the AMSE of $\hat{\mathbf{g}}_{k+1}$, i.e. $\mathrm{tr}(\mathbf{C}_{\Delta\mathbf{g}_{k+1}})$.

Obviously, the estimation error of $\mathbf{h}$ ($\mathbf{g}$) is lower bounded by estimation with exact given $\mathbf{g}$ ($\mathbf{h}$), which is given by

$$\begin{aligned}&\mathrm{tr}\left(\mathbf{C}_{\Delta\mathbf{h}|\mathbf{g}}\right) \\ &= \mathrm{tr}\left(\left(\mathbf{R}_\mathbf{h}^{-1} + \alpha^2(\mathbf{g} \otimes \mathbf{T}_s)^H \mathbf{C}_{\mathbf{n}|\mathbf{g}}^{-1}(\mathbf{g} \otimes \mathbf{T}_s)\right)^{-1}\right) \\ &= \mathrm{tr}\left(\left(\sigma_h^{-2}\mathbf{I}_{N_t} + \frac{\alpha^2\|\mathbf{g}\|^2}{\sigma_n^2(1+\alpha^2\|\mathbf{g}\|^2)}\mathbf{T}_s^H\mathbf{T}_s\right)^{-1}\right),\end{aligned} \tag{39}$$

$$\begin{aligned}&\mathrm{tr}\left(\mathbf{C}_{\Delta\mathbf{g}|\mathbf{h}}\right) \\ &= \mathrm{tr}\Bigg(\Big(\left(\mathbf{R}_\mathbf{g}^{-1} + \left(\mathbf{I}_{N_d} \otimes (\alpha\mathbf{T}_s\mathbf{h} + \mathbf{t}_r)\right)\right)^H \mathbf{C}_{\mathbf{n}|\mathbf{h}}^{-1} \\ &\qquad \left(\mathbf{I}_{N_d} \otimes (\alpha\mathbf{T}_s\mathbf{h} + \mathbf{t}_r)\right)\Big)^{-1}\Bigg) \\ &= \mathrm{tr}\left(\left(\mathbf{R}_\mathbf{g}^{-1} + \sigma_n^{-2}\|\alpha\mathbf{T}_s\mathbf{h} + \mathbf{t}_r\|^2(\alpha^2\mathbf{R}_\mathbf{g} + \mathbf{I}_{N_d})^{-1}\right)^{-1}\right) \\ &= \mathrm{tr}\left(\left(\sigma_g^{-2} + \frac{\|\alpha\mathbf{T}_s\mathbf{h} + \mathbf{t}_r\|^2}{\sigma_n^2(\alpha^2\sigma_g^2 + 1)}\right)^{-1}\mathbf{I}_{N_d}\right).\end{aligned} \tag{40}$$

With a good initial point provided by LS estimator, the estimation errors of $\mathbf{h}$ and $\mathbf{g}$ decrease alternately. Moreover, as the estimation error has a lower bound, it is easy to draw the conclusion that the proposed iterative LMMSE in-channel estimator converges to a minimum estimation error.

IV. Performance Lower Bounds

In this section, we first derive the LELB to evaluate the performance of the proposed iterative LMMSE in-channel estimator. Then, to solidify the derived LELB, we will also derive the Bayesian CRB, which is widely used to evaluate the performance of the estimators under the Bayesian framework. Interestingly, it will be proved that the LELB is tighter than the Bayesian CRB.

A. Linear Estimation Lower Bound (LELB)

To derive a closed-form LELB, we first introduce the following lemma.

Lemma 2: With the variable $a > 0$, $x > 0$ and the $N \times N$ matrix $\mathbf{A} \succ \mathbf{0}$, the matrix function

$$f(x) = \mathrm{tr}\left((a\mathbf{I} + x\mathbf{A})^{-1}\right) \tag{41}$$

is convex for $x > 0$.

Proof: See Appendix C. ■

From the convex optimization theory, we have the following inequality [33]

$$\mathrm{E}\{f(x)\} \geq f(\mathrm{E}\{x\}), \tag{42}$$

where $f(x)$ is one convex function, and x is a random variable.

As derived in section III.C, the lower bound of the estimation error $\mathrm{tr}\{\mathbf{C}_{\Delta\mathbf{h}_\mathbf{g}}\}$ ($\mathrm{tr}\{\mathbf{C}_{\Delta\mathbf{g}_\mathbf{h}}\}$) with a specific $\mathbf{g}$ ($\mathbf{h}$) is given by (39) and (40). By taking the expectation of $\mathrm{tr}\{\mathbf{C}_{\Delta\mathbf{h}_\mathbf{g}}\}$ and $\mathrm{tr}\{\mathbf{C}_{\Delta\mathbf{g}_\mathbf{h}}\}$ over $\mathbf{g}$ and $\mathbf{h}$, respectively, we can derived the closed-form LELBs for $\mathbf{h}$ and $\mathbf{g}$ with the help of Lemma 2 and inequality (42) as follows:

$$\begin{aligned}
&\mathrm{E}_\mathbf{g}\Big\{\mathrm{tr}\left(\mathbf{C}_{\Delta\mathbf{h}|\mathbf{g}}\right)\Big\} \\
&= \mathrm{E}_\mathbf{g}\Big\{\mathrm{tr}\Big[\Big(\sigma_h^{-2}\mathbf{I}_{N_s} + \frac{\alpha^2\|\mathbf{g}\|^2}{\sigma_n^2(1+\alpha^2\|\mathbf{g}\|^2)}\mathbf{T}_s^H\mathbf{T}_s\Big)^{-1}\Big]\Big\}, \\
&\geq \mathrm{tr}\left(\left(\sigma_h^{-2}\mathbf{I}_{N_s} + \omega_1\mathbf{T}_s^H\mathbf{T}_s\right)^{-1}\right) \\
&= \frac{N_s}{\omega_1 P_s/N_s + \sigma_h^{-2}} \\
&\triangleq \mathrm{LELB}_\mathbf{h}
\end{aligned} \tag{43}$$

$$\begin{aligned}
&\mathrm{E}_\mathbf{h}\Big\{\mathrm{tr}\left(\mathbf{C}_{\Delta\mathbf{g}|\mathbf{h}}\right)\Big\} \\
&= \mathrm{E}_\mathbf{h}\Big\{\mathrm{tr}\Big[\Big(\sigma_g^{-2} + \frac{\|\alpha\mathbf{T}_s\mathbf{h}+\mathbf{t}_r\|^2}{\sigma_n^2(\alpha^2\sigma_g^2+1)}\Big)^{-1}\mathbf{I}_{N_d}\Big]\Big\} \\
&\geq \mathrm{tr}\Big[\Big(\sigma_g^{-2} + \frac{\sigma_n^{-2}P_r - \tau}{\alpha^2\sigma_g^2+1}\Big)^{-1}\mathbf{I}_{N_d}\Big] \\
&= N_d\Big(\sigma_g^{-2} + \frac{\sigma_n^{-2}P_r - \tau}{\alpha^2\sigma_g^2+1}\Big)^{-1} \\
&\triangleq \mathrm{LELB}_\mathbf{g}.
\end{aligned} \tag{44}$$

B. Bayesian CRB (BCRB)

Since the iterative LMMSE in-channel estimator is proposed under Bayesian frameworks, the Bayesian CRB can be adopted to evaluate the performance of the proposed in-channel estimator. From [34], we know that the estimation error covariance of $\boldsymbol{\theta} = [\mathbf{g}^T, \mathbf{h}^T]^T$ is lower bounded as

$$\mathrm{E}_{\mathbf{h},\mathbf{g}}\{(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})^H\} \succeq \mathbf{F}^{-1} \tag{45}$$

where $\mathbf{A} \succeq \mathbf{B}$ means that the matrix $\mathbf{A} - \mathbf{B}$ is positive semi-definite, and $\mathbf{F}$ is the $(N_s + N_d) \times (N_s + N_d)$ Bayesian Fisher information matrix (BFIM).

Following the similar methods in [23]–[25], we can calculate $\mathbf{F}$ as

$$\begin{aligned}
\mathbf{F} =& \mathrm{E}_{\mathbf{h},\mathbf{g}}\left\{\begin{bmatrix} \frac{\partial^2 p(\mathbf{y}|\mathbf{h},\mathbf{g})}{\partial\mathbf{g}^*\partial\mathbf{g}^T} & \frac{\partial^2 p(\mathbf{y}|\mathbf{h},\mathbf{g})}{\partial\mathbf{g}^*\partial\mathbf{h}^T} \\ \frac{\partial^2 p(\mathbf{y}|\mathbf{h},\mathbf{g})}{\partial\mathbf{h}^*\partial\mathbf{g}^T} & \frac{\partial^2 p(\mathbf{y}|\mathbf{h},\mathbf{g})}{\partial\mathbf{h}^*\partial\mathbf{h}^T} \end{bmatrix}\right\} \\
&+ \mathrm{E}_{\mathbf{h},\mathbf{g}}\left\{\begin{bmatrix} \frac{\partial^2 p(\mathbf{h},\mathbf{g})}{\partial\mathbf{g}^*\partial\mathbf{g}^T} & \frac{\partial^2 p(\mathbf{h},\mathbf{g})}{\partial\mathbf{g}^*\partial\mathbf{h}^T} \\ \frac{\partial^2 p(\mathbf{h},\mathbf{g})}{\partial\mathbf{h}^*\partial\mathbf{g}^T} & \frac{\partial^2 p(\mathbf{h},\mathbf{g})}{\partial\mathbf{h}^*\partial\mathbf{h}^T} \end{bmatrix}\right\} \\
=& \mathrm{E}_{\mathbf{h},\mathbf{g}}\left\{\begin{bmatrix} \frac{\partial\boldsymbol{\mu}^H}{\partial\mathbf{g}^*}\mathbf{C}_{\mathbf{n}|\mathbf{h},\mathbf{g}}^{-1}\frac{\partial\boldsymbol{\mu}}{\partial\mathbf{g}^T} & \frac{\partial\boldsymbol{\mu}^H}{\partial\mathbf{g}^*}\mathbf{C}_{\mathbf{n}|\mathbf{h},\mathbf{g}}^{-1}\frac{\partial\boldsymbol{\mu}}{\partial\mathbf{h}^T} \\ \frac{\partial\boldsymbol{\mu}^H}{\partial\mathbf{h}^*}\mathbf{C}_{\mathbf{n}|\mathbf{h},\mathbf{g}}^{-1}\frac{\partial\boldsymbol{\mu}}{\partial\mathbf{g}^T} & \frac{\partial\boldsymbol{\mu}^H}{\partial\mathbf{h}^*}\mathbf{C}_{\mathbf{n}|\mathbf{h},\mathbf{g}}^{-1}\frac{\partial\boldsymbol{\mu}}{\partial\mathbf{h}^T} \end{bmatrix}\right\} \\
&+ \mathrm{E}_{\mathbf{h},\mathbf{g}}\Big\{\mathrm{blkdiag}\Big(\boldsymbol{\Sigma}, \mathbf{0}_{N_s\times N_s}\Big)\Big\} + \mathrm{blkdiag}\left(\mathbf{R}_\mathbf{g}^{-1}, \mathbf{R}_\mathbf{h}^{-1}\right) \\
=& \begin{bmatrix} \mathbf{F}_{11} & \mathbf{F}_{12} \\ \mathbf{F}_{21} & \mathbf{F}_{22} \end{bmatrix}
\end{aligned} \tag{46}$$

where $\mathbf{F}_{11} \in \mathbb{C}^{N_s\times N_s}$, $\mathbf{F}_{12} \in \mathbb{C}^{N_s\times N_d}$, $\mathbf{F}_{21} \in \mathbb{C}^{N_d\times N_s}$ and $\mathbf{F}_{22} \in \mathbb{C}^{N_d\times N_d}$ are the corresponding sub-matrices of $\mathbf{F}$, $\mathbf{C}_{\mathbf{n}|\mathbf{h},\mathbf{g}}$ equals $\mathbf{C}_{\mathbf{n}|\mathbf{g}}$ in (12), the partial derivatives are defined as

$$\begin{aligned}
\frac{\partial\boldsymbol{\mu}}{\partial\mathbf{g}^T} &= \Big(\frac{\partial\boldsymbol{\mu}^H}{\partial\mathbf{g}^*}\Big)^H = \Big[\frac{\partial\boldsymbol{\mu}}{\partial g_0}, \frac{\partial\boldsymbol{\mu}}{\partial g_1}, \ldots, \frac{\partial\boldsymbol{\mu}}{\partial g_{N_d-1}}\Big], \\
\frac{\partial\boldsymbol{\mu}}{\partial\mathbf{h}^T} &= \Big(\frac{\partial\boldsymbol{\mu}^H}{\partial\mathbf{h}^*}\Big)^H = \Big[\frac{\partial\boldsymbol{\mu}}{\partial h_0}, \frac{\partial\boldsymbol{\mu}}{\partial h_1}, \ldots, \frac{\partial\boldsymbol{\mu}}{\partial h_{N_s-1}}\Big],
\end{aligned}$$

and $\boldsymbol{\Sigma}$ is a $N_d \times N_d$ matrix with the (i,j)-th element as

$$\boldsymbol{\Sigma}_{ij} = \mathrm{tr}\left(\mathbf{C}_{\mathbf{n}|\mathbf{h},\mathbf{g}}^{-1}\frac{\partial\mathbf{C}_{\mathbf{n}|\mathbf{h},\mathbf{g}}}{\partial g_i^*}\mathbf{C}_{\mathbf{n}|\mathbf{h},\mathbf{g}}^{-1}\frac{\partial\mathbf{C}_{\mathbf{n}|\mathbf{h},\mathbf{g}}}{\partial g_j}\right). \tag{47}$$

Notice that the facts that $\mathbf{n}$ and $\mathbf{h}$ are separately independent on $\mathbf{h}$ and $\mathbf{g}$ are utilized in derivation. From (12), it can be derived

$$\frac{\partial\mathbf{C}_{\mathbf{n}|\mathbf{h},\mathbf{g}}}{\partial g_i} = \sigma_n^2\alpha^2\mathbf{e}_i\mathbf{g}^H \otimes \mathbf{I}_\tau, \tag{48}$$

where $\mathbf{e}_i$ denotes the $N_d \times 1$ basis vector with 1 for the i-th element and 0 otherwise. Plugging (13), (48) into (47) and taking some tedious mathematical manipulations, we can derive $\mathrm{E}_{\mathbf{h},\mathbf{g}}\{\boldsymbol{\Sigma}_{ij}\}$ as

$$\begin{aligned}
&\mathrm{E}_{\mathbf{h},\mathbf{g}}\{\boldsymbol{\Sigma}_{ij}\} \\
&= \mathrm{E}_\mathbf{g}\Big\{\frac{\tau\alpha^4}{1+\alpha^2\|\mathbf{g}\|^2}\mathrm{tr}\left(\mathbf{g}\mathbf{e}_i^H\mathbf{e}_j\mathbf{g}^H\right)\Big\} \\
&\qquad - \mathrm{E}_\mathbf{g}\Big\{\frac{\tau\alpha^6\|\mathbf{g}\|^2}{(1+\alpha^2\|\mathbf{g}\|^2)^2}\mathrm{tr}\left(\mathbf{g}\mathbf{e}_i^H\mathbf{g}\mathbf{g}^H\mathbf{e}_j\mathbf{g}^H\right)\Big\} \\
&= \mathrm{E}_\mathbf{g}\Big\{\frac{\tau\alpha^4\|\mathbf{g}\|^2}{1+\alpha^2\|\mathbf{g}\|^2}\big(1 - \frac{\alpha^2\|g_i\|^2}{1+\alpha^2\|\mathbf{g}^2\|}\big)\Big\}\sigma(i-j) \\
&= \mathrm{E}_\mathbf{g}\Big\{\frac{\tau\alpha^4\|\mathbf{g}\|^2}{1+\alpha^2\|\mathbf{g}\|^2}\big(1 - \frac{\alpha^2\|g_0\|^2}{1+\alpha^2\|\mathbf{g}^2\|}\big)\Big\}\sigma(i-j),
\end{aligned} \tag{49}$$

where$i, j = 0, 1, \cdots, N_d - 1$. The equations $\mathrm{tr}(\mathbf{AB}) = \mathrm{tr}(\mathbf{BA})$, $\mathrm{tr}(\mathbf{A}\otimes\mathbf{B}) = \mathrm{tr}(\mathbf{A})\,\mathrm{tr}(\mathbf{B})$ and the fact that g_0 and g_i has the same distrbution are used in the above derivation,

With the definition of $\boldsymbol{\mu}$ in (7), it can be achieved that

$$\begin{aligned}
\frac{\partial\boldsymbol{\mu}}{\partial\mathbf{g}} &= \alpha(\mathbf{I}_{N_d} \otimes \mathbf{T}_s\mathbf{h}) + (\mathbf{I}_{N_d} \otimes \mathbf{t}_r), \\
\frac{\partial\boldsymbol{\mu}}{\partial\mathbf{h}} &= \alpha(\mathbf{g} \otimes \mathbf{T}_s).
\end{aligned} \tag{50}$$

Combining (13) and (50), we have the following results:

$$\mathrm{E}_{\mathbf{h},\mathbf{g}}\Big\{\frac{\partial\boldsymbol{\mu}^H}{\partial\mathbf{h}^*}\mathbf{C}^{-1}_{\mathbf{n}|\mathbf{h},\mathbf{g}}\frac{\partial\boldsymbol{\mu}}{\partial\mathbf{h}^T}\Big\} = \sigma_n^{-2}\mathrm{E}_{\mathbf{g}}\Big\{\frac{\alpha^2\|\mathbf{g}\|^2}{1+\alpha^2\|\mathbf{g}\|^2}\Big\}\mathbf{T}_s^H\mathbf{T}_s \tag{51}$$

$$\begin{aligned}\mathrm{E}_{\mathbf{h},\mathbf{g}}\Big\{\frac{\partial\boldsymbol{\mu}^H}{\partial\mathbf{g}^*}\mathbf{C}^{-1}_{\mathbf{n}|\mathbf{h},\mathbf{g}}\frac{\partial\boldsymbol{\mu}}{\partial\mathbf{h}^T}\Big\} &= \Big(\mathrm{E}_{\mathbf{h},\mathbf{g}}\Big\{\frac{\partial\boldsymbol{\mu}^H}{\partial\mathbf{h}^*}\mathbf{C}^{-1}_{\mathbf{n}|\mathbf{h},\mathbf{g}}\frac{\partial\boldsymbol{\mu}}{\partial\mathbf{g}^T}\Big\}\Big)^H\\ &= \mathbf{0}_{N_d\times N_s}\end{aligned} \tag{52}$$

$$\begin{aligned}&\mathrm{E}_{\mathbf{h},\mathbf{g}}\Big\{\frac{\partial\boldsymbol{\mu}^H}{\partial\mathbf{g}^*}\mathbf{C}^{-1}_{\mathbf{n}|\mathbf{h},\mathbf{g}}\frac{\partial\boldsymbol{\mu}}{\partial\mathbf{g}^T}\Big\}\\ &= \sigma_n^{-2}\mathrm{E}_{\mathbf{h}}\big\{\|\alpha^2\mathbf{T}_s\mathbf{h}+\mathbf{t}_r\|^2\big\}\mathrm{E}_{\mathbf{g}}\Big\{1-\frac{\alpha^2\|g_0\|^2}{1+\alpha^2\|\mathbf{g}\|^2}\Big\}\mathbf{I}_{N_d}\\ &= (\sigma_n^{-2}P_r-\tau)\mathrm{E}_{\mathbf{g}}\Big\{1-\frac{\alpha^2\|g_0\|^2}{1+\alpha^2\|\mathbf{g}\|^2}\Big\}\mathbf{I}_{N_d}\end{aligned} \tag{53}$$

With and (46), (49) and (53), the sub-matrices $\mathbf{F}_{ij}(i,j=1,2)$ of the $\mathbf{F}$ can be derived as

$$\mathbf{F}_{12} = \mathbf{F}_{12}^H = \mathbf{0}_{N_s\times N_d}, \tag{54}$$

$$\begin{aligned}\mathbf{F}_{22} &= \sigma_n^{-2}\mathrm{E}_{\mathbf{g}}\Big\{\frac{\alpha^2\|\mathbf{g}\|^2}{1+\alpha^2\|\mathbf{g}\|^2}\Big\}\mathbf{T}_s^H\mathbf{T}_s+\mathbf{R}_{\mathbf{h}}^{-1}\\ &= \omega_1\mathbf{T}_s^H\mathbf{T}_s+\sigma_h^{-2}\mathbf{I}_{N_s},\end{aligned} \tag{55}$$

$$\begin{aligned}\mathbf{F}_{11} &= \mathrm{E}_{\mathbf{g}}\Big\{\frac{\tau\alpha^4\|\mathbf{g}\|^2}{1+\alpha^2\|\mathbf{g}\|^2}\Big(1-\frac{\alpha^2\|g_0\|^2}{1+\alpha^2\|\mathbf{g}\|^2}\Big)\Big\}\mathbf{I}_{N_d}\\ &\quad+(\sigma_n^{-2}P_r-\tau)\mathrm{E}_{\mathbf{g}}\Big\{1-\frac{\alpha^2\|g_0\|^2}{1+\alpha^2\|\mathbf{g}\|^2}\Big\}\mathbf{I}_{N_d}+\mathbf{R}_{\mathbf{g}}^{-1}\\ &= \omega_2\mathbf{I}_{N_d}+\omega_3\mathbf{I}_{N_d}+\sigma_g^{-2}\mathbf{I}_{N_d}.\end{aligned} \tag{56}$$

where ω_1, ω_2 and ω_3 are defined in Appendix B. Since the optimal source-training block $\mathbf{T}_s$ satisfies $\mathbf{T}_s^H\mathbf{T}_s=\frac{P_s}{N_s}\mathbf{I}_{N_s}$, which will be proved in next section, the Bayesian CRBs for $\mathbf{g}$ and $\mathbf{h}$ can be separately given by

$$\mathrm{BCRB}_{\mathbf{h}} = \mathrm{tr}(\mathbf{F}_{22}^{-1}) = \frac{N_s}{\omega_1P_s/N_s+\sigma_h^{-2}}, \tag{57}$$

$$\mathrm{BCRB}_{\mathbf{g}} = \mathrm{tr}(\mathbf{F}_{11}^{-1}) = \frac{N_d}{\omega_2+\omega_3+\sigma_g^{-2}}. \tag{58}$$

Theorem 1: In an AF-based OWRN with multiple antennas at the source $\mathbb{S}$ and the destination $\mathbb{D}$, the LELB for the in-channel along the link $\mathbb{S}\to\mathbb{R}$ equals the corresponding Bayesian CRB, while the LELB for the in-channel along the link $\mathbb{R}\to\mathbb{D}$ is tighter than the corresponding Bayesian CRB, i.e., $\mathrm{LELB}_{\mathbf{h}}=\mathrm{BCRB}_{\mathbf{h}}$ and $\mathrm{LELB}_{\mathbf{g}}>\mathrm{BCRB}_{\mathbf{g}}$.

Proof: The equation $\mathrm{LELB}_{\mathbf{h}}=\mathrm{BCRB}_{\mathbf{h}}$ can be readily checked from the expressions of $\mathrm{LELB}_{\mathbf{h}}$ and $\mathrm{BCRB}_{\mathbf{h}}$ in (43) and (57). Moreover, the inequality $\mathrm{LELB}_{\mathbf{g}}>\mathrm{BCRB}_{\mathbf{g}}$ can be obtained by the following derivations as

$$\begin{aligned}&\mathrm{BCRB}_{\mathbf{g}}\\ &<\mathrm{tr}\Big(\Big((\sigma_n^{-2}P_r-\tau)\mathrm{E}_{\mathbf{g}}\Big\{1-\frac{\alpha^2\|g_i\|^2}{1+\alpha^2\|\mathbf{g}\|^2}\Big\}\mathbf{I}_{N_d}+\sigma_g{}^{-2}\mathbf{I}_{N_d}\Big)^{-1}\Big)\\ &<\mathrm{tr}\Big(\Big((\sigma_n^{-2}P_r-\tau)\mathrm{E}_{\mathbf{g}}\Big\{\frac{1}{1+\alpha^2\|g_0\|^2}\Big\}\mathbf{I}_{N_d}+\sigma_g{}^{-2}\mathbf{I}_{N_d}\Big)^{-1}\Big)\\ &\le\mathrm{tr}\Big(\Big(\sigma_g^{-2}+\frac{\sigma_n^{-2}P_r-\tau}{\alpha^2\sigma_g^2+1}\Big)^{-1}\mathbf{I}_{N_d}\Big)\\ &= N_d\Big(\sigma_g^{-2}+\frac{\sigma_n^{-2}P_r-\tau}{\alpha^2\sigma_g^2+1}\Big)^{-1}\\ &= \mathrm{LELB}_{\mathbf{g}}.\end{aligned} \tag{59}$$

The proof of Theorem 1 is completed. ∎

V. Optimal Training and Amplification Factor Design

Since LELB is tighter than BCRB, we will adopt the LELB as the criterion to design the training block and amplification factor. Fortunately, the LELB has simple analytical solutions, which can be seen from (43) and (44) . Hence, it can provides sufficient insight on the training parameter optimization.

A. Training Design

It can be concluded from (43) and (44) that the source-training block $\mathbf{T}_s$ only affects the $\mathrm{LELB}_{\mathbf{h}}$, and that the structure of relay-training block $\mathbf{t}_r$ has no impact either on $\mathrm{LELB}_{\mathbf{h}}$ or on $\mathrm{LELB}_{\mathbf{g}}$. Therefore, the optimal training design can be done through minimizing the $\mathrm{LELB}_{\mathbf{h}}$ with respect to $\mathbf{T}_s$, which can be formulated as follows

$$\begin{aligned}\text{(P1)}\qquad &\min_{\mathbf{T}_s}\ \mathrm{LELB}_{\mathbf{h}}\\ &\text{s.t.}\ \ \mathrm{tr}\{\mathbf{T}_s^H\mathbf{T}_s\}=P_s\end{aligned} \tag{60}$$

For any positive-definite $N_s\times N_s$ matrix $\mathbf{A}$, we have [35]

$$\mathrm{tr}(\mathbf{A}^{-1})\ge\sum_{i=0}^{N_s-1}\frac{1}{[\mathbf{A}]_{ii}}, \tag{61}$$

where the equality holds when $\mathbf{A}$ is diagonal. With this inequality, it can be derived that

$$\begin{aligned}\mathrm{LELB}_{\mathbf{h}} &\overset{(a)}{\ge}\sum_{i=0}^{N_s-1}\frac{1}{\sigma_h^{-2}+\omega_1[\mathbf{T}_s^H\mathbf{T}_s]_{ii}}\\ &\overset{(b)}{\ge} N_s(\sigma_h^{-2}+\omega_1P_s/N_s)^{-1},\end{aligned} \tag{62}$$

where the equality in (a) holds when $\mathbf{T}_s^H\mathbf{T}_s$ is diagonal, and the equality in (b) holds when each diagonal entry of $\mathbf{T}_s^H\mathbf{T}_s$ equals P_s/N_s [36]. Thus, the optimal $\mathbf{T}_s$ should satisfy the following constraint.

$$\mathbf{T}_s\mathbf{T}_s^H=\frac{P_s}{N_s}\mathbf{I}_{N_s}. \tag{63}$$

Notice that the number τ of the time slots for training block transmission should be no less than N_s to satisfy the above constraint. Since the system throughput decreases with the increase of τ, the optimal choice of τ is $\tau=N_s+1$ to ensure that $\mathbf{t}_r$ are also orthogonal with $\mathbf{T}_s$. With the above analysis, an example of the optimal training blocks $\mathbf{T}_s$ and $\mathbf{t}_r$ can be constructed from the $(N_s+1)\times(N_s+1)$ discrete fourier transform (DFT) matrix as

$$[\mathbf{T}_s]_{i,j}=\sqrt{\frac{P_s}{\tau N_s}}e^{\frac{-2\pi ij}{\tau}},i=0,1,\cdots,\tau-1,j=0,1,\cdots,Ns-1. \tag{64}$$

$$[\mathbf{t}_r]_i=\sqrt{\frac{P_t}{\tau}}e^{\frac{-2\pi iN_s}{\tau}},\ \ i=0,1,\cdots,\tau-1. \tag{65}$$

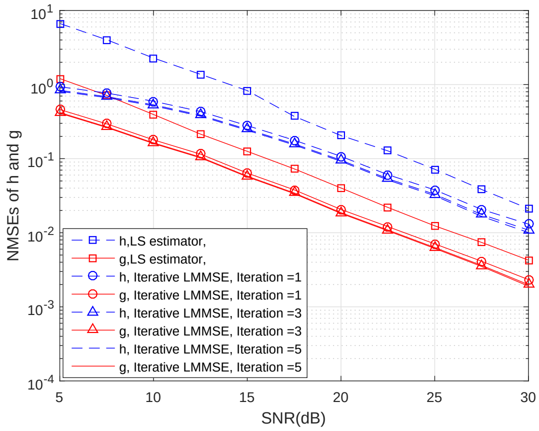

Fig. 2. The in-channels $\mathbf{h}$, $\mathbf{g}$ estimation NMSEs versus SNR with $N = 2$.

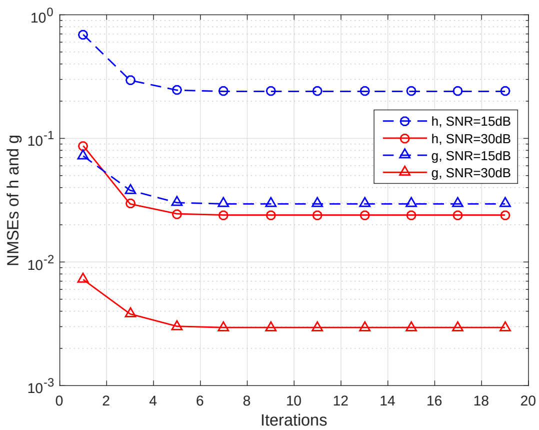

Fig. 3. The in-channels $\mathbf{h}$, $\mathbf{g}$ estimation NMSEs versus iteration with $SNR = 15, 30dB, N = 2$.

B. Amplification Factor α Design

It is clear that $\text{LELB}_{\mathbf{h}}$ and $\text{LELB}_{\mathbf{g}}$ are monotonely decreasing and increasing functions with respect to α, respectively. Therefore, a tradeoff between $\text{LELB}_{\mathbf{h}}$ and $\text{LELB}_{\mathbf{g}}$ should be made in designing α. A meaningful approach is optimizing α to minimize the weighted sum of $\text{LELB}_{\mathbf{h}}$ and $\text{LELB}_{\mathbf{g}}$, which can be formulated as follows

$$\text{(P2)} \qquad \min_{\alpha} \ f(\alpha) = \gamma_{\mathbf{h}}\text{LELB}_{\mathbf{h}} + \gamma_{\mathbf{g}}\text{LELB}_{\mathbf{g}}$$
$$\text{s.t. } 0 \le \alpha \le \alpha_{max},$$

where the positive variables $\gamma_{\mathbf{h}}$ and $\gamma_{\mathbf{g}}$ are the weights for $\text{LELB}_{\mathbf{h}}$ and $\text{LELB}_{\mathbf{g}}$, respectively, and $\alpha_{max} = \sqrt{P_r/({\sigma_h}^2 P_s + \tau\sigma_n^2)}$ ensures the power of $\mathbf{t}_r$ is nonnegative. In an OWRN, the in-channel $\mathbf{g}$ always plays a more important role than $\mathbf{h}$, because the destination can utilize $\mathbf{g}$ to predict the relay's operation, to achieve the relay's paring strategy, or to estimate the covariance matrix of the equivalent noise [37], [38]. Moreover, in some scenarios, the relay not only forwards the received data packet from source but also sends its own data to destination. Thus, the factor $\gamma_{\mathbf{g}}$ should be bigger than $\gamma_{\mathbf{h}}$ in general.

With the help of (43),(44) and (75), the closed-from expressions of $f(\alpha)$ and its first-order derivative $f'(\alpha)$ can be presented on the top of the next page, where $\xi = \frac{P_s e^{-\sigma^{-2}\alpha^{-2}}}{N_s N_d \sigma_n^2 \sigma_g^2 \alpha^2}$ and the derivative formula $\frac{\delta \ {}_1F_1(a;b;x)}{\delta x} = \frac{a}{b} \ {}_1F_1(a+1; b+1; x)$ of the confluent hypergeometric function is utilized in the derivation. Due to its complicated structure, it is difficult to derive the analytic expression of the optimal α through solving the equation $f'(\alpha) = 0$. However, its numerical solution can be conveniently obtained with the gradient descent algorithm [39].

VI. Simulations

In this section, the numerical results are provided to examine the proposed iterative LMMSE estimator. All the in-channels and noise are assumed to have unit variance, i.e., $\sigma_h^2 = \sigma_g^2 = \sigma_n^2 = 1$. Without loss of generality, we set $N_s = N_d = N$. We fix $P_r = 2P_s$ and the signal to noise ratio is defined as $\text{SNR} = P_s/\sigma_n = P_s$. The normalized MSEs, BCRBs, and LELBs (NMSEs, NBCRBs and NLELBs) of $\mathbf{h}$, $\mathbf{g}$ are used as the figures of the merit, which are defined as

$$\text{NMSE}_{\boldsymbol{\theta}} = \frac{\mathbf{E}\{\|\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}\|^2\}}{N}, \tag{68}$$

$$\text{NBCRB}_{\boldsymbol{\theta}} = \frac{\text{BCRB}_{\boldsymbol{\theta}}}{N}, \tag{69}$$

$$\text{NLELB}_{\boldsymbol{\theta}} = \frac{\text{LELB}_{\boldsymbol{\theta}}}{N}, \tag{70}$$

where $\boldsymbol{\theta} = \mathbf{h}, \mathbf{g}$. In total, 10^4 Monte-Carlo runs are adopted for numerical average.

Firstly, we examine the performance and convergence property of the proposed iterative LMMSE estimator. Here, we set $\alpha = 0.8$ and utilize the optimal orthogonal training block in (64) and (65).

Fig. 2 presents the NMSE curves of the initial LS estimation and the proposed iterative LMMSE in-channel estimator. It can be seen that the iterative LMMSE in-channel estimator can effectively improve the accuracy of the initial LS in-channel estimation in terms of NMSE of both $\mathbf{h}$ and $\mathbf{g}$. Compared with the cases of iterations 3 and 5, the iterative LMMSE in-channel estimator with iteration=1 can achieve larger performance gains. The iterative LMMSE algorithm only takes five iterations to arrive at a steady state, which shows that the algorithm has fast convergence speed. Moreover, the NMSE curves versus number of iteration is show in Fig. 3 to explicitly exhibit the convergence of proposed estimator.

Fig. 4 and Fig. 5 compare NLELBs and NBCRBs with $N = 2$ and $N = 6$, respectively. Five iterations are used for the iterative LMMSE algorithm. We obtain the following observations: the in-channels' ($\mathbf{h}$ and $\mathbf{g}$) NMSE curves of the iterative LMMSE estimator cannot approximate the NLELB curves, which can be explained by the in-channel estimation error propagation phenomena; the NLELB is more tighter than NBCRB, which matches the theoretical analysis in Theorem 1; When N increases from $N = 2$ to $N = 6$, NLELB of both $\mathbf{h}$ and $\mathbf{g}$ become quite close to the NMSE, but the $\text{NBCRB}_{\mathbf{g}}$ is still far away from $\text{NMSE}_{\mathbf{g}}$.

Next, we evaluate the designed optimal training structure. The NMSE with different τ, i.e., $N-1, N, N+1$ and $N+2$,

$$f(\alpha)=\gamma_{\mathbf{h}}N_s\Big(\sigma_h^{-2}+\frac{P_s}{N_s\sigma_n^2}-\frac{P_s}{N_sN_d\sigma_n^2\sigma_g^2\alpha^2}e^{-\sigma_g^{-2}\alpha^{-2}}{}_1F_1(N_d-1;N_d;\sigma_g^{-2}\alpha^{-2})\Big)^{-1}+\gamma_{\mathbf{g}}N_d\Big(\sigma_g^{-2}+\frac{\sigma_n^{-2}P_r-\tau}{\alpha^2\sigma_g^2+1}\Big)^{-1}. \tag{66}$$

$$f'(\alpha)=2\xi\gamma_{\mathbf{h}}N_s\Big(\sigma_h^{-2}+\frac{P_s}{N_s\sigma_n^2}-\frac{P_s}{N_sN_d\sigma_n^2\sigma_g^2\alpha^2}e^{-\sigma_g^{-2}\alpha^{-2}}\ {}_1F_1(N_d-1;N_d;\sigma_g^{-2}\alpha^{-2})\Big)^{-2}\Big((\alpha^{-3}-\alpha^{-5}\sigma_g^{-2})$$
$${}_1F_1(N_d-1;N_d;\sigma_g^{-2}\alpha^{-2})-\frac{N_d-1}{N_d\alpha^5\sigma_g^2}\ {}_1F_1(N_d;N_d+1;\sigma_g^{-2}\alpha^{-2})\Big)+2\alpha\sigma_g^2\gamma_{\mathbf{g}}N_d\Big(\sigma_g^{-2}+\frac{\sigma_n^{-2}P_r-\tau}{\alpha^2\sigma_g^2+1}\Big)^{-2}\frac{\sigma_n^{-2}P_r-\tau}{(\alpha^2\sigma_g^2+1)^2}. \tag{67}$$

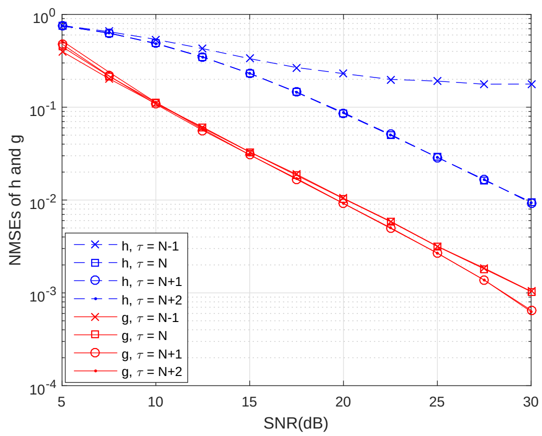

Fig. 6. Comparison the in-channels $\mathbf{h}$, $\mathbf{g}$ estimation NMSEs versus SNR under different τ.

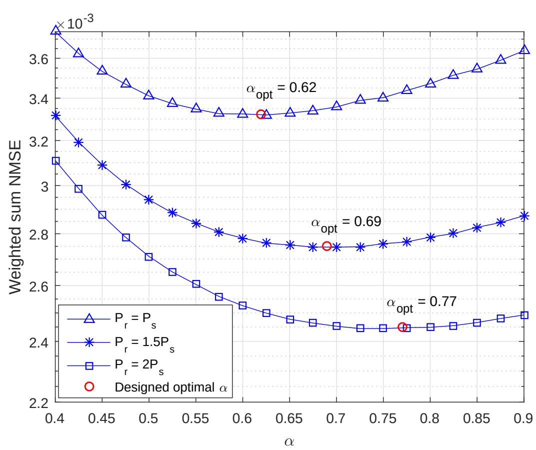

Fig. 7. The weighted sum NMSEs versus α with SNR = 30dB.

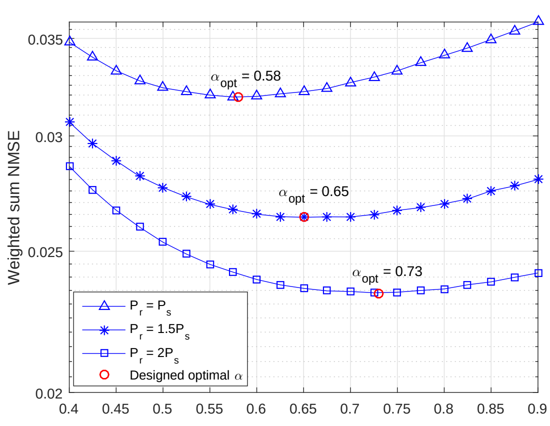

Fig. 8. The weighted sum NMSEs versus α with SNR = 20dB.

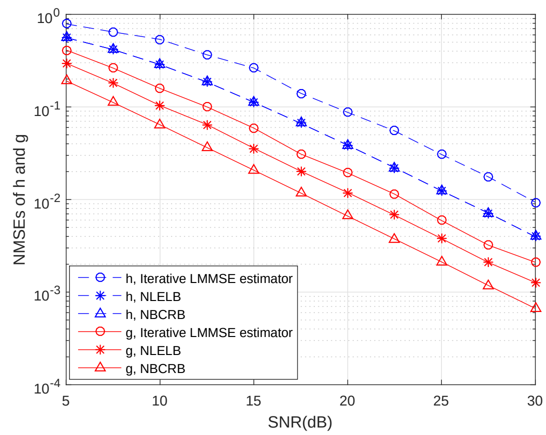

Fig. 4. Comparison between the LELB and BCRB versus SNR with $N = 2$.

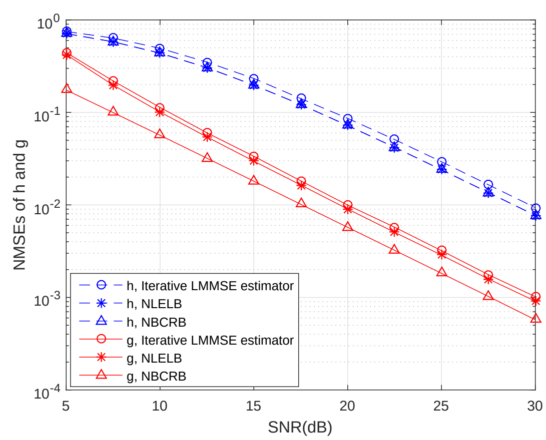

Fig. 5. Comparison between the LELB and BCRB versus SNR with $N = 6$.

are presented in Fig.6. Comparing the NMSE with different τ, we can find that the $\text{NMSE}_{\mathbf{h}}$ decreases significantly when τ increases from $N-1$ to N as the training sequences for different source antennas becomes orthogonal, and that $\text{NMSE}_{\mathbf{g}}$ decreases when τ increases from N to $N+1$ since the relay's training sequences becomes orthogonal with source's training block. However, both the $\text{NMSE}_{\mathbf{h}}$ and $\text{NMSE}_{\mathbf{g}}$ can't

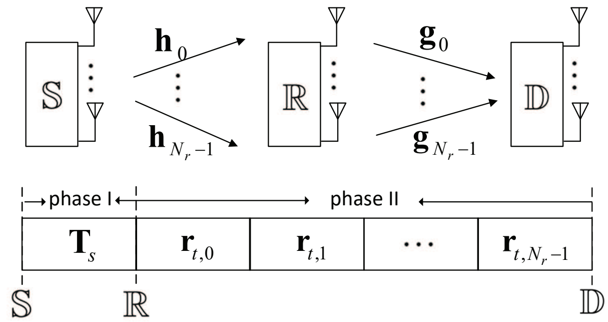

Fig. 9. The system model and training blocks transmission protocol for OWRM with multiple antennas at all nodes.

be lowered any further by continually increasing τ, which validates our analytical study. It can be also seen that $\text{NMSE}_{\mathbf{g}}$ increases with increasing τ at low SNR regime. This phenomenon is not strange and can be explained as follow: as the relay's total power is constrained, more noise power merges into received signal power with increase of τ, which leads to decrease in SNR at destination.

Lastly, we would like to check our designing of optimal α. Fig.7 and Fig.8 present the curves of weighted sum of $\text{NMSE}_{\mathbf{g}}$ and $\text{NMSE}_{\mathbf{h}}$ versus α with SNR $=$ 30dB and SNR $=$ 20dB, respectively. Three different relay power, i.e., $P_r = 2P_s$, $P_r = 1.5P_s$, and $P_r = P_s$, are adopted for both cases. We set the weight $\gamma_{\mathbf{h}} = 0.2$ and $\gamma_{\mathbf{g}} = 0.8$ as analyzed in main text that the weight of $\mathbf{g}$ should be greater than $\mathbf{h}$ in general. As observed from both figures, the optimal α obtained by our designed approach (marked as red circle) are well-matched with the lowest points of weighted sum NMSEs of simulation, which demonstrated our analytical study. It is clearly that α increases with the increase of relay power P_r, which matches the fact that more power should be allocated for forwarding the received source training blocks if relay have more power. Comparing the optimal α with different SNR, we can also conclude that optimal α decreases with decreas of SNR. This result comes from the fact that more power should be allocated for relay's training sequence to guarantee the estimation of $\mathbf{g}$ in low SNR as $\mathbf{g}$ has a greater weight.

VII. The General Case where Relay Has Multiple Antennas

In previous sections, we only consider the specific case where the relay has single antenna . In this section, following the similar approach in [27], we extend our studies to the general scenario, where the relay has $N_r > 1$ antennas.

As shown in Fig. 9, let us denote the in-channel vector along $\mathbb{S}$ to $\mathbb{R}$'s i-th ($i = 0, 1, \cdots, N_r - 1$) antenna and that along $\mathbb{R}$'s i-th antenna to $\mathbb{D}$ as $\mathbf{h}_i$ and $\mathbf{g}_i$, respectively. One round of the training symbols transmission from $\mathbb{S}$ to $\mathbb{D}$ is still partitioned into two phases.

During phase I, $\mathbb{S}$ sends one source-training block $\mathbf{T}_s$ to $\mathbb{R}$. Then, the received training at $\mathbb{R}$'s i-th antenna during phase I can be expressed as

$$\mathbf{r}_{r,i} = \mathbf{T}_s \mathbf{h}_i + \mathbf{n}_{r,i}, \tag{71}$$

Different from the single antenna case, we divide the phase II into N_r stages. In the i-th stage, $\mathbb{R}$ superimposes the relay-training sequence $\mathbf{t}_{r,i}$ in $\mathbf{r}_{r,i}$, i.e.

$$\mathbf{r}_{t,i} = \alpha_i \mathbf{r}_{r,i} + \mathbf{t}_{r,i}, \tag{72}$$

and forwards $\mathbf{r}_{r,i}$ to $\mathbb{D}$ through its i-th antenna. Then, the received training at $\mathbb{D}$ in stage i can be written as

$$\mathbf{Y}_i = \alpha_i \mathbf{T}_s \mathbf{h}_i {\mathbf{g}_i}^T + \mathbf{t}_{r,i} {\mathbf{g}_i}^T + \alpha_i \mathbf{n}_{r,i} {\mathbf{g}_i}^T + \mathbf{N}_{d,i}. \tag{73}$$

Since each relay's antenna forwards its own received training in each stage independently, with respect to each relay antenna, the problem becomes the same as that for the case where relay has single antenna. Thus, $\mathbf{h}_i$ and $\mathbf{g}_i$ can be estimated by the proposed iterative LMMSE in-channel estimator from observation (73) in stage i. The performance analysis, optimal training design, and optimal amplification factor design for the single antenna case also apply to the case where relay has multiple antennas. It is worth to point out that the the dimension of the source-training block $\mathbf{T}_s$ does not increase with relay's antenna number N_r.

VIII. Conclusion

In this paper, we investigated the in-channel estimation in the classical three-node AF OWRN with multiple antennas at all the nodes. The typical scenario, where the relay node is equipped with single antenna, was first examined. Under the superimposed training framework, we developed an iterative LMMSE in-channel estimator, whose initial point was provided by the LS in-channel estimator formed by the matrix unitary diagonalization. We further derived LELB and BCRB to evaluate the proposed algorithms. The optimal training and amplification factor was designed by minimizing the LELB. Furthermore, simulation results were provided to corroborate our studies. Finally, we extended our studies to the general case, where all the nodes are equipped with multiple antennas.

Appendix A
Proof of Lemma 1

The term $\text{tr}\left(f(\mathbf{X})\right)$ can be expressed as

$$\begin{aligned}\text{tr}\left(f(\mathbf{X})\right) &= \text{tr}\left((a\mathbf{I} + \mathbf{A}^H\mathbf{X}^{-1}\mathbf{A})^{-1}\right)\\ &= \frac{1}{aN + \text{tr}\left(\mathbf{A}^H\mathbf{X}^{-1}\mathbf{A}\right)}\\ &= \frac{1}{aN + \frac{1}{\text{tr}\left(\mathbf{X}(\mathbf{A}\mathbf{A}^H)^{-1}\right)}}.\end{aligned} \tag{74}$$

From the fact that $(\mathbf{A}\mathbf{A}^H)^{-1} \succ \mathbf{0}$, it can be derived that $\text{tr}\left(\mathbf{X}_2(\mathbf{A}\mathbf{A}^H)^{-1}\right) < \text{tr}\left(\mathbf{X}_1(\mathbf{A}\mathbf{A}^H)^{-1}\right)$ if $\text{tr}(\mathbf{X}_2) < \text{tr}(\mathbf{X}_1)$, where $\mathbf{X}_1 \succeq 0$ and $\mathbf{X}_2 \succeq 0$. Therefore, we can obtain $\text{tr}\left(f(\mathbf{X}_2)\right) < \text{tr}\left(f(\mathbf{X}_1)\right)$.

Appendix B
Definitions and Calculations of ω_1 ω_2 and ω_3

We first calculate the factor ω_1 as

$$\begin{aligned}\omega_1 &= \sigma_n^{-2}\mathbf{E}_{\mathbf{g}}\Big\{\frac{\alpha^2\|\mathbf{g}\|^2}{1+\alpha^2\|\mathbf{g}\|^2}\Big\}\\ &= \sigma_n^{-2} - \frac{2}{\sigma_n^2\sigma_g^2\alpha^2}\mathbf{E}_{\mathbf{g}}\Big\{\frac{1}{2\sigma_g^{-2}\alpha^{-2}+2\sigma_g^{-2}\|\mathbf{g}\|^2}\Big\}. \end{aligned} \tag{75}$$

With the statistics of $\mathbf{g}$, it can be checked that the term $2\sigma_g^{-2}\alpha^{-2}+2\sigma_g^{-2}\|\mathbf{g}\|^2$ is the non-central chi-squared distributed with the degrees of freedom $2N_d$ and the non-centrality parameter $2\sigma_g^{-2}\alpha^{-2}$. Thus, we can derive that

$$\begin{aligned}&\mathbf{E}_{\mathbf{g}}\Big\{\frac{1}{2\sigma_g^{-2}\alpha^{-2}+2\sigma_g^{-2}\|\mathbf{g}\|^2}\Big\}\\ &= \frac{1}{2N_d}e^{-\sigma_g^{-2}\alpha^{-2}}\ {}_1F_1(N_d-1;N_d;\sigma_g^{-2}\alpha^{-2}),\end{aligned} \tag{76}$$

where

$${}_1F_1(a;b;z) = \frac{\Gamma(b)}{\Gamma(a)\Gamma(b-a)}\int_0^1 t^{a-1}(1-t)^{b-a-1}e^{zt}dt \tag{77}$$

is the confluent hypergeometric function. Plugging (76) into (75), we get the closed-form expression of ω_1.

Following the similar methods, the variables ω_2 and ω_3 in (56) can be separately written as

$$\omega_2 = \mathbf{E}_{\mathbf{g}}\Big\{\frac{\tau\alpha^4\|\mathbf{g}\|^2}{1+\alpha^2\|\mathbf{g}\|^2}\Big(1-\frac{\alpha^2\|g_0\|^2}{1+\alpha^2\|\mathbf{g}^2\|}\Big)\Big\}, \tag{78}$$

$$\omega_3 = (\sigma_n^{-2}P_r-\tau)\mathbf{E}_{\mathbf{g}}\Big\{1-\frac{\alpha^2\|g_0\|^2}{1+\alpha^2\|\mathbf{g}\|^2}\Big\}. \tag{79}$$

Due to the complicated structures of the expectations in (78) and (79), it is difficult for us to achieve simple analytical solutions for both ω_2 and ω_3. Nonetheless, we can use the Gibbs sampling technique and generate samples $\mathbf{g}^1, \mathbf{g}^2, \cdots, \mathbf{g}^{N_c}$ from its PDF $\mathcal{CN}(0, \mathbf{R}_g)$. Then, the numerical scheme to calculate ω_2 and ω_3 can be listed as

$$\omega_2 \approx \frac{1}{N_c}\sum_{k=1}^{N_c}\Big\{\frac{\tau\alpha^4\|\mathbf{g}^k\|^2}{1+\alpha^2\|\mathbf{g}^k\|^2}\Big(1-\frac{\alpha^2\|g_0^k\|^2}{1+\alpha^2\|\mathbf{g}^k\|^2}\Big)\Big\} \tag{80}$$

$$\omega_3 \approx \frac{1}{N_c}\sum_{k=1}^{N_c}(\sigma_n^{-2}P_r-\tau)\Big\{1-\frac{\alpha^2\|g_0^k\|^2}{1+\alpha^2\|\mathbf{g}^k\|^2}\Big\} \tag{81}$$

Appendix C
Proof of Lemma 2

With the matrix property $\mathrm{tr}(\mathbf{X}^{-1}) = 1/\mathrm{tr}(\mathbf{X})$, $f(x)$ can be rewritten as

$$f(x) = \frac{1}{aN + x\,\mathrm{tr}(\mathbf{A})}. \tag{82}$$

Moreover, the second derivative of $f(x)$ with respect to x can be denoted as

$$f''(x) = \frac{2\,\mathrm{tr}^2(\mathbf{A})}{\big(aN+x\,\mathrm{tr}(\mathbf{A})\big)^3} > 0,\ (x>0) \tag{83}$$

where $\mathrm{tr}(\mathbf{A}) > 0$ for $\mathbf{A} \succ \mathbf{0}$ is utilized. From the fact that $f''(x) > 0$, we can conclude that $f(x)$ is convex for $x > 0$.

References

[1] S. Zhang, M. Sheng, and F. Gao, "Iterative LMMSE individual channel estimation with superimposed training over one-way relay networks," in *Proc. IEEE Int. Conf. Commun (ICC)*, pp. 4717–4721, 2014.

[2] J. N. Laneman, D. N. Tse, and G. W. Wornell, "Cooperative diversity in wireless networks: Efficient protocols and outage behavior," *IEEE Trans. Inf. Theory*, vol. 50, no. 12, pp. 3062–3080, Dec. 2004.

[3] T. Jiang and G. Zhu, "Complement block coding for reduction in peak-to-average power ratio of OFDM signals," *IEEE Commun. Mag.*, vol. 43, no. 9, pp. S17–S22, Sept. 2005.

[4] W. Cui, D. Qu, T. Jiang, and B. Farhang-Boroujeny, "Coded auxiliary pilots for channel estimation in fbmc-oqam systems," *IEEE Trans. Veh. Technol.*, vol. 65, no. 5, pp. 2936–2946, May 2016.

[5] G. Okeke, W. A. Krzymień, and Y. Jing, "Beamforming in non-regenerative MIMO broadcast relay networks," *IEEE Trans. Signal Process.*, vol. 60, no. 12, pp. 6641–6654, Dec. 2012.

[6] Q. Li, Y. Yang, W.-K. Ma, M. Lin, J. Ge, and J. Lin, "Robust cooperative beamforming and artificial noise design for physical-layer secrecy in AF multi-antenna multi-relay networks," *IEEE Trans. Signal Process.*, vol. 63, no. 1, pp. 206–220, Jan. 2015.

[7] Y. Liu, L. Wang, T. T. Duy, M. Elkashlan, and T. Q. Duong, "Relay selection for security enhancement in cognitive relay networks," *IEEE Wireless Commun. Lett.*, vol. 4, no. 1, pp. 46–49, Feb. 2015.

[8] V. N. Q. Bao, T. Q. Duong, D. Benevides da Costa, G. C. Alexandropoulos, and A. Nallanathan, "Cognitive amplify-and-forward relaying with best relay selection in non-identical Rayleigh fading," *IEEE Commun. Lett.*, vol. 17, no. 3, pp. 475–478, Mar. 2013.

[9] X. Chen, C. Zhong, C. Yuen, and H.-H. Chen, "Multi-antenna relay aided wireless physical layer security," *IEEE Commun. Mag.*, vol. 53, no. 12, pp. 40–46, Dec. 2015.

[10] J. Huang, Q. Li, Q. Zhang, G. Zhang, and J. Qin, "Relay beamforming for amplify-and-forward multi-antenna relay networks with energy harvesting constraint," *IEEE Trans. Signal Process.*, vol. 21, no. 4, pp. 454–458, Apr. 2014.

[11] P. A. Anghel and M. Kaveh, "On the performance of distributed space-time coding systems with one and two non-regenerative relays," *IEEE Trans. Wireless Commun.*, vol. 5, no. 3, pp. 682–692, Mar. 2006.

[12] M. Raeisi, S. M. Razavizadeh, and I. Lee, "Distributed space-time coding for two-way relay networks," in *Proc. IEEE 79th Veh. Technol Conf. (VTC Spring)*, pp. 1–5, Dec. 2014.

[13] M. Sheng, Y. Li, X. Wang, J. Li, and Y. Shi, "Energy efficiency and delay tradeoff in device-to-device communications underlaying cellular networks," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 1, pp. 92–106, Jan. 2016.

[14] Y. Li, M. Sheng, C. W. Tan, Y. Zhang, Y. Sun, X. Wang, Y. Shi, and J. Li, "Energy-efficient subcarrier assignment and power allocation in ofdma systems with max-min fairness guarantees," *IEEE Trans. Commun.*, vol. 63, no. 9, pp. 3183–3195, Sept 2015.

[15] T.-H. Pham, Y.-C. Liang, H. K. Garg, and A. Nallanathan, "Joint channel estimation and data detection for MIMO-OFDM two-way relay networks," in *Proc. IEEE Global Telecommunications Conf. (GLOBECOM)*, pp. 1–5, Dec. 2010.

[16] X. Xie, M. Peng, B. Zhao, W. Wang, and Y. Hua, "Maximum a posteriori based channel estimation strategy for two-way relaying channels," *IEEE Trans. Wireless Commun.*, vol. 13, no. 1, pp. 450–463, Jan. 2014.

[17] R. Wang, M. Tao, H. Mehrpouyan, and Y. Hua, "Channel estimation and optimal training design for correlated MIMO two-way relay systems in colored environment," *IEEE Trans. Wireless Commun.*, vol. 14, no. 5, pp. 2684–2699, May 2015.

[18] P. Lioliou, M. Viberg, and M. Coldrey, "Efficient channel estimation techniques for amplify and forward relaying systems," *IEEE Trans. Commun.*, vol. 60, no. 11, pp. 3150–3155, Nov. 2012.

[19] Q. Li, Q. Zhang, and J. Qin, "Secure relay beamforming for simultaneous wireless information and power transfer in nonregenerative relay networks," *IEEE Trans. Veh. Technol.*, vol. 63, no. 5, pp. 2462–2467, Jun. 2014.

[20] X. Li, Q. Zhang, G. Zhang, and J. Qin, "Joint power allocation and subcarrier pairing for cooperative OFDM AF multi-relay networks," *IEEE Commun. Lett.*, vol. 17, no. 5, pp. 872–875, May 2013.

[21] 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, Sept. 2011.

[22] F. Gao, B. Jiang, X. Gao, and X.-D. Zhang, “Superimposed training based channel estimation for OFDM modulated amplify-and-forward relay networks,” *IEEE Trans. Commun.*, vol. 59, no. 7, pp. 2029–2039, Jul. 2011.

[23] X. Xie, M. Peng, F. Gao, and W. Wang, “Superimposed training based channel estimation for uplink multiple access relay networks,” *IEEE Trans. Wireless Commun.*, vol. 14, no. 8, pp. 4439–4453, Aug. 2015.

[24] S. Zhang, F. Gao, J. Li, and M. Sheng, “Online and offline bayesian Cramér bounds for time-varying channel estimation in one-way relay networks,” *IEEE Trans. Signal Process.*, vol. 63, no. 8, pp. 1977–1992, Apr. 2015.

[25] S. Zhang, F. Gao, J. Li, and H. Li, “Time varying channel estimation for dstc-based relay networks: Tracking, smoothing and BCRBs,” *IEEE Trans. Wireless Commun.*, vol. 14, no. 9, pp. 5022–5037, Sept. 2015.

[26] C. W. R. 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.

[27] Y. Jing and X. Yu, “ML-based channel estimations for non-regenerative relay networks with multiple transmit and receive antennas,” *IEEE J. Sel. Areas Commun.*, vol. 30, no. 8, pp. 1428–1439, Sept. 2012.

[28] C. W. R. Chiong, Y. Rong, and Y. Xiang, “Channel estimation for time-varying MIMO relay systems,” *IEEE Trans. Wireless Commun.*, vol. 14, no. 12, pp. 6752–6762, Dec. 2015.

[29] D. Kong, D. Qu, and T. Jiang, “Time domain channel estimation for OQAM-OFDM systems: Algorithms and performance bounds,” *IEEE Trans. Signal Process.*, vol. 62, no. 2, pp. 322–330, Jan. 2014.

[30] S. Zhang, F. Gao, C. Pei, and X. He, “Segment training based individual channel estimation in one-way relay network with power allocation,” *IEEE Trans. Wireless Commun.*, vol. 12, no. 3, pp. 1300–1309, Mar. 2013.

[31] K. B. Petersen, M. S. Pedersen *et al.*, *The matrix cookbook*. Technical University of Denmark, 2008.

[32] S. K. Sengijpta, *Fundamentals of statistical signal processing: Estimation theory*. Prentice Hall, 1987.

[33] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge Univ. Press, 2004.

[34] 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, Sept. 2012.

[35] M. Biguesh and A. B. Gershman, “Training-based MIMO channel estimation: a study of estimator tradeoffs and optimal training signals,” *IEEE Trans. Signal Process.*, vol. 54, no. 3, pp. 884–893, Mar. 2006.

[36] F. Gao, T. Cui, and A. Nallanathan, “Optimal training design for channel estimation in decode-and-forward relay networks with individual and total power constraints,” *IEEE Trans. Signal Process.*, vol. 56, no. 12, pp. 5937–5949, Dec. 2008.

[37] M. Kashef and A. Ephremides, “Optimal partial relaying for energy-harvesting wireless networks,” *IEEE/ACM Trans. Networking*, vol. 24, no. 1, pp. 113–122, Feb. 2016.

[38] G. Li, F. Gong, N. Zhang, and Y. Wang, “A fuse-and-forward protocol for two-way relaying networks with relay having its own broadcasting information,” *IEEE Commun. Lett.*, vol. 19, no. 8, pp. 1450–1453, Aug. 2015.

[39] M. Avriel, *Nonlinear programming: analysis and methods*, Courier Corporation, 2003.

[40] T. Fujioka, “Asymptotic approximations of the inverse moment of the noncentral chi-squared variable,” *Journal of the Japan Statistical Society*, vol. 31, no. 1, pp. 99–109, Oct. 2001.