

Optimal Training Design for Channel Estimation in Amplify and Forward Relay Networks

Tao Cui[†], Feifei Gao[‡], and Arumugam Nallanathan[‡]

[†]Department of Electrical Engineering, California Institute of Technology, Pasadena, CA 91125

Email: taocui@caltech.edu

[‡]Department of Electrical and Computer Engineering, National University of Singapore, Singapore 119260

Email: {feifeigao, elena}@nus.edu.sg

Abstract—In this paper, we study training based channel estimation for relay networks using the amplify-and-forward (AF) transmission scheme. We first point out that separately estimating the channel from source to relay and the channel from relay to destination incurs several problems. We then propose a new estimation scheme that directly estimates the overall channel from source to destination. The proposed channel estimation well matches the AF based space time coding that was developed recently. Both linear least-square estimator and minimum mean-square-error estimator are studied. The corresponding optimal training sequences and the optimal precoding matrices are also derived.

I. INTRODUCTION

Placing multiple antennas at the transmitters can increase the system capacity [1] enhance the system reliability [3], [4]. Unfortunately, packing many antennas onto a small mobile terminal faces the practical difficulty of the size limit. In order to overcome this limitation, one would refer to the relay network [5]- [8].

The relay based transmission is usually divided into two phases. During Phase I, the source will broadcast its own information bits to all relays. During Phase II, the relays would either choose to purely amplify and retransmit the information to the destination, or to decode the information first and then transmit these information bits to the destination. The former process is referred as *amplify-and-forward* (AF) and the latter is referred as *decode-and-forward* (DF). Various cooperative diversity schemes and space time coding (STC) techniques have been developed in [6]- [8] for either AF or DF approach.

Before enjoying all those benefits brought by the relay networks, an accurate channel state information (CSI) is required at the destination (for AF) or at both relay and destination (for DF). However, almost all the existing works assume that the perfect channel knowledge is available, and no attempt on channel estimation for the relay networks has yet been reported. One major reason of the lack of the literature is the common belief that the channel estimation for network could be built upon following the same approach between the individual transmitters and receivers. However, this belief is only valid for the DF scheme, whereas for AF based transmission, separating the channel estimation from two phases will incur several problems.

This work was supported by the National University of Singapore and Defence Science and Technology Agency (DSTA), Singapore under Grant MINDEF-NUS-JPP-07-05.

In this work, we propose a training based channel estimation scheme, where the overall channel from source to destination is estimated at the destination only. Both the least square (LS) and the minimum mean square error (MMSE) channel estimators are derived. We also derive optimal training designs including both the training sequence sent from the source and the linear precoding matrix at each relay.

II. SPACE TIME CODING IN AF RELAY NETWORK

Consider a wireless network with M randomly placed relay nodes \mathbb{R}_i , $i = 1, \dots, M$, one source node \mathbb{S} , and one destination node \mathbb{D} . Every node has only a single antenna that cannot transmit and receive simultaneously. The channel between each node pair is assumed quasi-stationary Rayleigh flat fading which is constant within one frame but may vary from frame to frame. Denote the channel from \mathbb{S} to \mathbb{D} as f , from \mathbb{S} to \mathbb{R}_i as g_i , from \mathbb{R}_i to \mathbb{D} as h_i , respectively; namely $f \in \mathcal{CN}(0, \sigma_f^2)$, $g_i \in \mathcal{CN}(0, \sigma_{g_i}^2)$, and $h_i \in \mathcal{CN}(0, \sigma_{h_i}^2)$. Assume that \mathbb{S} wants to send the signal block $\mathbf{s} = [s_1, \dots, s_T]^T$ to \mathbb{D} via AF mode. The transmission will be accomplished by the following two phases. For Phase I, the transmitter broadcasts the signal \mathbf{s} to relays and the destination. The received signals at \mathbb{R}_i and \mathbb{D} could be expressed as

$$\mathbf{r}_i = g_i \mathbf{s} + \mathbf{n}_{ri}, \quad (1)$$

$$\mathbf{d}_1 = f \mathbf{s} + \mathbf{n}_{d1}, \quad (2)$$

where \mathbf{n}_{ri} and \mathbf{n}_{d1} are the independent white complex Gaussian noise at the relays and the destination, respectively. For convenience, all noise variances are assumed as N_0 , namely, $\mathbf{n}_{ri}, \mathbf{n}_{d1} \in \mathcal{CN}(0, N_0 \mathbf{I})$. The power constraint of the transmission is $E\{\mathbf{s}^H \mathbf{s}\} = TP_s$, where P_s is the average transmitting power of source. To exploit the diversity of the relay system, a linear transformation of \mathbf{r}_i , denoted as \mathbf{t}_i , should be transmitted from each relay, instead of sending the \mathbf{r}_i to the destination directly. A linear dispersion (LD) based STC has been proposed in [8], where \mathbf{r}_i is first precoded by a unitary matrix \mathbf{P}_i and is then scaled by a real factor α_i to keep the average power of \mathbb{R}_i as P_{ri} . To enable complex STC, we here modify the precoding scheme in [8] into

$$\mathbf{t}_i = \alpha_i \mathbf{P}_i \mathbf{r}_i^{(*)}, \quad (3)$$

where $(\cdot)^{(*)}$ represents the item itself if the i th relay operates on \mathbf{r}_i whereas represents the conjugate of the item if the i th relay operates on \mathbf{r}_i^* . Note that this type of STC, where one

relay operates only in either \mathbf{r}_i or \mathbf{r}_i^* , exclusively, has been proposed and analyzed in [8] for AF relay networks. Since the specific realization of the channel is unknown, α_i will be chosen as

$$\alpha_i = \sqrt{\frac{P_{ri}}{\sigma_{gi}^2 P_s + N_0}}, \quad (4)$$

which keeps the power constraint from the long term point of view. The destination \mathbb{D} in Phase II then receives

$$\mathbf{d}_2 = \sum_{i=1}^M h_i \mathbf{t}_i + \mathbf{n}_{d2} = \mathbf{B}\mathbf{\Lambda}\mathbf{w} + \mathbf{n}_d, \quad (5)$$

where $\mathbf{n}_{d2} \in \mathcal{CN}(0, N_0 \mathbf{I})$ represents the complex white Gaussian noise vector at \mathbb{D} in the second phase, and $\mathbf{s}_i \triangleq \mathbf{s}$ for all i . Other variables are defined as follows:

$$\begin{aligned} w_i &= h_i g_i^{(*)}, & \mathbf{w} &= [w_1, \dots, w_M]^T, \\ \mathbf{\Lambda} &= \text{diag}\{\alpha_1, \dots, \alpha_M\}, & \mathbf{n}_d &= \sum_i h_i \alpha_i \mathbf{P}_i \mathbf{n}_{r_i}^{(*)} + \mathbf{n}_{d2}, \\ \mathbf{B} &= [\mathbf{P}_1 \mathbf{s}_1^{(*)}, \mathbf{P}_2 \mathbf{s}_2^{(*)}, \dots, \mathbf{P}_M \mathbf{s}_M^{(*)}]. \end{aligned}$$

Furthermore, it could easily be checked that the covariance of \mathbf{n}_d conditioned on a specific realization of h_i is

$$\text{Cov}(\mathbf{n}_d | h_i, i = 1, \dots, M) = \left(\sum_{i=1}^M |h_i|^2 \alpha_i^2 + 1 \right) N_0 \mathbf{I}, \quad (6)$$

where the property $\mathbf{P}_i \mathbf{P}_i^H = \mathbf{I}$ is utilized.

III. TRAINING BASED CHANNEL ESTIMATION

For coherent detection in the AF mode [6]- [8], the destination \mathbb{D} performs the maximum likelihood (ML) detection based only on a specific channel realization w_i while treating \mathbf{n}_d as the overall white Gaussian noise. Therefore, the task of the channel estimation focuses only on estimating w_i at \mathbb{D} . Based on this fact, two different channel estimation schemes could be considered.

One is to separately estimate g_i , h_i and then construct w_i from $g_i^{(*)} h_i$. This approach is not as trivial as it seems to be. For example, each relay should spend at least M additional time slots to send the estimated g_i to the destination after h_i was estimated. In practical transmissions, the power in one frame is usually constant. Therefore, additional energy will be consumed when transmitting additional time slots. Moreover, transmitting the estimated g_i will suffer from further distortion because of both the noise at the destination and the error in the estimated channel h_i . Sometimes, transmitting a real value is not convenient for relay nodes. Then g_i has to be quantized before the transmission [9]. The quantization error must also be counted.

The other way is to directly estimate the overall channel w_i at \mathbb{D} . We assume that the length of the training sequence sent from \mathbb{S} is N , which may be different from the data block size T . The training sequence, denoted as \mathbf{z} , will be broadcasted from \mathbb{S} at Phase I, and a linear transformation will be performed at each relay nodes before they forward the training to the destination at Phase II. Denote the $N \times N$ unitary precoding matrix at the i th relay as \mathbf{A}_i and define

$$\mathbf{C} = [\mathbf{A}_1 \mathbf{z}_1^{(*)}, \mathbf{A}_2 \mathbf{z}_2^{(*)}, \dots, \mathbf{A}_M \mathbf{z}_M^{(*)}]. \quad (7)$$

The transmitting model with other equations from (1) to (6) could be applied straightforwardly. With slight abuse of notations, we will keep all other notations unchanged from the previous section. During the training period, the power constraint is $\mathbf{z}^H \mathbf{z} \leq N P_s = E_s$.

A. Channel Estimation of \mathbf{w}

1) *LS Estimation.* The LS estimation assumes a specific realization of g_i, h_i , namely, the deterministic channel scenario. From (5), the optimal estimate of \mathbf{w} should be obtained from

$$\hat{\mathbf{w}}_{LS} = \mathbf{\Lambda}^{-1} \mathbf{C}^\dagger \mathbf{d}_2 = \mathbf{w} + \Delta \mathbf{w} \quad (8)$$

with error

$$\Delta \mathbf{w} = \mathbf{\Lambda}^{-1} \mathbf{C}^\dagger \mathbf{n}_d. \quad (9)$$

The covariance of $\Delta \mathbf{w}$ is then

$$\text{Cov}(\Delta \mathbf{w} | \mathbf{g}^{(*)}, \mathbf{h}) = N_0 \left(\sum_i |h_i|^2 |\alpha_i|^2 + 1 \right) \mathbf{\Lambda}^{-1} (\mathbf{C}^H \mathbf{C})^{-1} \mathbf{\Lambda}^{-1}, \quad (10)$$

where $\mathbf{g}^{(*)} = [g_1^{(*)}, \dots, g_M^{(*)}]^T$ and $\mathbf{h} = [h_1, \dots, h_M]^T$ are defined for convenience. Meanwhile, we would like to define $\mathbf{g} = [g_1, g_2, \dots, g_M]^T$ for future use. Since $\mathbf{\Lambda}$ is a constant matrix, the optimization is conducted by varying the value of \mathbf{C} . Note that the diagonal elements of \mathbf{C} must all be less than or equal to E_s . Therefore, the optimal \mathbf{C} can be found by solving the following constrained optimization problem:

$$\min_{\mathbf{A}_i, \mathbf{z}} \text{tr}(\mathbf{\Lambda}^{-1} (\mathbf{C}^H \mathbf{C})^{-1} \mathbf{\Lambda}^{-1}), \quad \text{s.t.} \quad [\mathbf{C}^H \mathbf{C}]_{ii} \leq E_s. \quad (11)$$

Note that, the above optimization problem is different from that of traditional MISO system, where there is a total power constraint over all transmit antennas and the solution follows a water-filling like structure [10]. In the relay networks, since different relays could not cooperate with each other, each relay has its own power constraint P_{ri} , which is reflected by M individual constraints in (11).

Theorem 1: The optimal $\mathbf{C}^H \mathbf{C}$ in (11) must be a diagonal matrix.

Proof: The following inequality for an arbitrary $N \times N$ positive definite matrix \mathbf{F} could be utilized [11]:

$$\text{tr}(\mathbf{F}^{-1}) \geq \sum_{i=1}^N ([\mathbf{F}]_{ii})^{-1}, \quad (12)$$

and the equality holds if and only if \mathbf{F} is diagonal.

Suppose \mathbf{C}_0 is the optimal solution and define $\mathbf{D}_0 = \mathbf{C}_0^H \mathbf{C}_0$. Meanwhile, we suppose \mathbf{D}_0 is not diagonal. Let $\bar{\mathbf{D}}_0 = \text{diag}\{\mathbf{D}_0\}$. Then $[\bar{\mathbf{D}}_0]_{ii}$ must be smaller than or equal to E_s , and therefore $\bar{\mathbf{D}}_0$ is also in the feasible set of (11). Let $\bar{\mathbf{F}} = \mathbf{\Lambda} \bar{\mathbf{D}}_0 \mathbf{\Lambda}$ whose diagonal part is given by

$$\bar{\mathbf{F}} = \text{diag}\{\mathbf{F}\} = \mathbf{\Lambda} \bar{\mathbf{D}}_0 \mathbf{\Lambda}. \quad (13)$$

Using the inequality (12), we obtain $\text{tr}(\mathbf{\Lambda}^{-1} \bar{\mathbf{D}}_0^{-1} \mathbf{\Lambda}^{-1}) = \text{tr}(\bar{\mathbf{F}}^{-1}) < \text{tr}(\mathbf{F}^{-1}) = \text{tr}(\mathbf{\Lambda}^{-1} \mathbf{D}_0^{-1} \mathbf{\Lambda}^{-1})$. This contradicts the assertion that the non-diagonal matrix \mathbf{D}_0 was optimal. Hence the optimal $\mathbf{C}^H \mathbf{C}$ must be diagonal. ■

From the above discussions, we know that the only choice is that $\mathbf{C}^H \mathbf{C} = E_s \mathbf{I}$. This result is to say that each relay should

transmit orthogonal training whereas should meet their own power constraints. Then, we only need to design \mathbf{z} and \mathbf{A}_i such that all off-diagonal elements of $\mathbf{C}^H \mathbf{C}$ are zero, namely

$$(\mathbf{z}_i^{(*)})^H \mathbf{A}_i \mathbf{A}_j^H \mathbf{z}_j^{(*)} = 0, \quad \text{for } i \neq j. \quad (14)$$

There exist a number of such \mathbf{z} and unitary matrix \mathbf{A}_i . The detailed discussion will be provided later.

Then the mean square error (MSE) on w_i , separately, can be obtained as

$$\text{var}(w_i | \mathbf{h}, \mathbf{g}) = \frac{N_0 (\sum_j |h_j|^2 |\alpha_j|^2 + 1)}{|\alpha_i|^2 E_s}, \quad (15)$$

and

$$\text{var}(w_i) = \frac{N_0 \sigma_{hi}^2}{E_s} + \frac{N_0 (\sum_{j \neq i} \sigma_{hj}^2 |\alpha_j|^2 / |\alpha_i|^2)}{E_s} + \frac{N_0 \sigma_{gi}^2}{E_{ri}} + \frac{N_0^2}{E_s E_{ri}}, \quad (16)$$

where $E_{ri} = NP_{ri}$ is defined for notation simplicity.

2) *MMSE Estimation.* The MMSE estimation assumes that the statistics of g_i, h_i are known at \mathbb{D} , namely, the statistical channel scenario. Denote the covariance of \mathbf{h} and $\mathbf{g}^{(*)}$ as \mathbf{R}_h and $\mathbf{R}_{g^{(*)}}$ respectively. Then, the covariance matrix of \mathbf{w} , assuming channels of Phase I are independent from channels of Phase II, is

$$\mathbf{R}_w = \mathbb{E}\{\mathbf{w}\mathbf{w}^H\} = \mathbf{R}_{g^{(*)}} \odot \mathbf{R}_h, \quad (17)$$

where \odot denotes the Hadamard product. The linear MMSE estimator of \mathbf{w} is expressed as

$$\begin{aligned} & \hat{\mathbf{w}}_{MMSE} \\ &= \mathbf{R}_w \mathbf{\Lambda} \mathbf{C}^H (\mathbf{C} \mathbf{\Lambda} \mathbf{R}_w \mathbf{\Lambda} \mathbf{C}^H + N_0 \sum_i (\sigma_{hi}^2 |\alpha_i|^2 + 1) \mathbf{I})^{-1} \mathbf{d}_2. \end{aligned} \quad (18)$$

The error covariance of the MMSE estimator is

$$\text{Cov}(\Delta \mathbf{w}) = \left(\mathbf{R}_w^{-1} + \frac{1}{N_0 \sum_i (\sigma_{hi}^2 |\alpha_i|^2 + 1)} \mathbf{\Lambda} \mathbf{C}^H \mathbf{C} \mathbf{\Lambda} \right)^{-1}. \quad (19)$$

Therefore, the optimal training should be obtained from

$$\arg \min_{\mathbf{A}_i, \mathbf{z}} \text{tr}(\text{Cov}(\Delta \mathbf{w})), \quad \text{s.t. } [\mathbf{C}^H \mathbf{C}]_{ii} \leq E_s. \quad (20)$$

The problem (20) can be reformulated as a convex optimization problem. Let us define $\mathbf{D} = \mathbf{C}^H \mathbf{C}$ and proceed the optimization under the unknown variable \mathbf{D} . As will be shown in Section III-B, there exist a number of feasible sets $\{\mathbf{A}_i, \mathbf{z}\}$ for any positive semi-definite \mathbf{D} , which validates our direct optimization with respect to \mathbf{D} . The optimization problem then becomes

$$\begin{aligned} \min_{\mathbf{T}, \mathbf{D}} \quad & \text{tr}(\mathbf{T}) \\ \text{s.t.} \quad & \mathbf{T} \succ \left(\mathbf{R}_w^{-1} + \frac{1}{N_0 \sum_i (\sigma_{hi}^2 |\alpha_i|^2 + 1)} \mathbf{\Lambda} \mathbf{D} \mathbf{\Lambda} \right)^{-1}, \\ & [\mathbf{D}]_{i,i} \leq E_s, \\ & \mathbf{D} \succ \mathbf{0}. \end{aligned} \quad (21)$$

This is precisely a convex optimization problem. Moreover, note that the constraint

$$\mathbf{T} \succ \left(\mathbf{R}_w^{-1} + \frac{1}{N_0 \sum_i (\sigma_{hi}^2 |\alpha_i|^2 + 1)} \mathbf{\Lambda} \mathbf{D} \mathbf{\Lambda} \right)^{-1} \quad (22)$$

can be rewritten, via Schur's complement [13], into the following linear matrix inequality (LMI):

$$\begin{bmatrix} \mathbf{T} & \mathbf{I} \\ \mathbf{I} & \mathbf{R}_w^{-1} + \frac{1}{N_0 \sum_i (\sigma_{hi}^2 |\alpha_i|^2 + 1)} \mathbf{\Lambda} \mathbf{D} \mathbf{\Lambda} \end{bmatrix} \succ \mathbf{0}. \quad (23)$$

Therefore, (21) lies in the so called semi-definite programming (SDP) formulation which can be solved efficiently from [12]. Meanwhile, there exist closed form solutions for several important scenarios. First let us ignore the condition $\mathbf{D} \succ \mathbf{0}$ and represent \mathbf{D} back to $\mathbf{C}^H \mathbf{C}$. For simplicity, let us further represent $N_0 \sum_i (\sigma_{hi}^2 |\alpha_i|^2 + 1)$ by β . Then the Lagrange function with the diagonal constraint can be written as

$$\begin{aligned} L(\mathbf{C}, \mu_i) = & \text{tr} \left(\left(\mathbf{R}_w^{-1} + \frac{1}{\beta} \mathbf{\Lambda} \mathbf{C}^H \mathbf{C} \mathbf{\Lambda} \right)^{-1} \right. \\ & \left. + \sum_i \mu_i (\text{tr}(\mathbf{e}_i^H \mathbf{C}^H \mathbf{C} \mathbf{e}_i) - E_s) \right), \end{aligned} \quad (24)$$

where \mathbf{e}_i is the i th column of an $N \times N$ identity matrix. Since (24) is in quadratic form, it is sufficient to consider only $\partial L(\mathbf{C}, \mu_i) / \partial \mathbf{C}$ [14], which is calculated as

$$\begin{aligned} & \frac{\partial L(\mathbf{C}, \mu_i)}{\partial \mathbf{C}} \\ &= \mathbf{C}^* \left(\sum_i \mu_i \mathbf{e}_i \mathbf{e}_i^T - \frac{1}{\beta} \mathbf{\Lambda} (\mathbf{R}_w^{-1} + \frac{1}{\beta} \mathbf{\Lambda} \mathbf{C}^H \mathbf{C} \mathbf{\Lambda})^{-2T} \mathbf{\Lambda} \right). \end{aligned} \quad (25)$$

Therefore, any \mathbf{C} can be the optimal solution if (25) is zero. Since \mathbf{C} is a tall and full rank matrix, (25) is equivalent to

$$\left(\mathbf{R}_w^{-1} + \frac{1}{\beta} \mathbf{\Lambda} \mathbf{C}^H \mathbf{C} \mathbf{\Lambda} \right)^2 = \beta \mathbf{\Lambda}^{-1} \left(\sum_i \mu_i \mathbf{e}_i \mathbf{e}_i^T \right)^{-1} \mathbf{\Lambda}^{-1} = \mathbf{\Omega}, \quad (26)$$

where

$$\mathbf{\Omega} = \text{diag} \left\{ \frac{\beta}{\alpha_1^2 \mu_1}, \frac{\beta}{\alpha_2^2 \mu_2}, \dots, \frac{\beta}{\alpha_M^2 \mu_M} \right\}. \quad (27)$$

It is proved in [15] that

$$\mathbf{R}_w^{-1} + \frac{1}{\beta} \mathbf{\Lambda} \mathbf{C}^H \mathbf{C} \mathbf{\Lambda} = \mathbf{\Omega}^{1/2}. \quad (28)$$

Let $\tilde{\mathbf{R}}_{w, \text{inv}}$ denote the diagonal matrix with diagonal elements obtained from \mathbf{R}_w^{-1} . Then

$$\mathbf{\Omega}^{1/2} = \text{diag}\{\mathbf{\Omega}^{1/2}\} = \tilde{\mathbf{R}}_{w, \text{inv}} + \frac{E_s}{\beta} \mathbf{\Lambda}^2. \quad (29)$$

Then, the optimal $\mathbf{C}^H \mathbf{C}$ should be obtained from (28) as,

$$\mathbf{C}^H \mathbf{C} = E_s \mathbf{I} - \beta \mathbf{\Lambda}^{-1} \tilde{\mathbf{R}}_{w, \text{off}} \mathbf{\Lambda}^{-1}, \quad (30)$$

where $\tilde{\mathbf{R}}_{w, \text{off}}$ represents the remaining matrix after setting all diagonal elements of \mathbf{R}_w^{-1} to be zero. However, we need to further consider the positive semi-definite constraint $\mathbf{D} = \mathbf{C}^H \mathbf{C} \succ \mathbf{0}$ in order to arrive at a true optimal solution. Therefore, (30) is the optimal solution only when $E_s \mathbf{I} - \beta \mathbf{\Lambda}^{-1} \tilde{\mathbf{R}}_{w, \text{off}} \mathbf{\Lambda}^{-1} \succ \mathbf{0}$. This positive semi-definite requirement is satisfied under two important scenarios:

a) When channel w_i are uncorrelated with each other. In this case, the optimal solution of $\mathbf{C}^H \mathbf{C}$ is $E_s \mathbf{I}$, which coincides

with the optimal training in the LS channel estimation. Therefore, the orthogonal training sequence is, again, employed. Then the error covariance matrix is

$$\text{Cov}(\Delta \mathbf{w}) = \left(\mathbf{R}_w^{-1} + \frac{E_s}{\beta} \mathbf{\Lambda}^2 \right)^{-1}, \quad (31)$$

and the estimation error for each w_i could be separately obtained as

$$\text{var}(w_i) = \frac{N_0 \left(\sum_j \sigma_{h_{ij}}^2 |\alpha_j|^2 + 1 \right)}{|\alpha_i|^2 E_s + N_0 \sum_j \sigma_{h_{ij}}^2 |\alpha_j|^2 / \sigma_{w_i}^2}. \quad (32)$$

Compared to the MSE of the LS estimation (16), the MSE in (32) is smaller, as expected.

b) When the transmit SNR is high enough such that $E_s \mathbf{I} - \beta \tilde{\mathbf{R}}_{w,off} \succcurlyeq \mathbf{0}$ could be satisfied. The error covariance matrix is

$$\text{Cov}(\Delta \mathbf{w}) = \left(\tilde{\mathbf{R}}_{w,inv} + \frac{E_s}{\beta} \mathbf{\Lambda}^2 \right)^{-1}, \quad (33)$$

and the MSE on w_i can be separately obtained as

$$\text{var}(w_i) = \frac{N_0 \left(\sum_j \sigma_{h_{ij}}^2 |\alpha_j|^2 + 1 \right)}{|\alpha_i|^2 E_s + N_0 \sum_j \sigma_{h_{ij}}^2 |\alpha_j|^2 / \tilde{\sigma}_{w_i}^2}. \quad (34)$$

where $\tilde{\sigma}_{w_i}^2$ is the i th diagonal element of $\tilde{\mathbf{R}}_w^{-1}$. The MSE in (34) is also smaller than (16).

B. Design of \mathbf{z} and \mathbf{A}_i

Assume that the optimal \mathbf{D} has been derived from either (21) or (30). Then, the remaining task is to find the training sequence \mathbf{z} as well as the unitary linear precoding matrices \mathbf{A}_i . The following equations should be considered simultaneously:

$$(\mathbf{z}_i^{(*)})^H \mathbf{A}_i^H \mathbf{A}_j \mathbf{z}_j^{(*)} = [\mathbf{D}]_{ij}, \quad i, j = 1, \dots, M, \quad (35a)$$

$$\mathbf{A}_i^H \mathbf{A}_i = \mathbf{I}, \quad i = 1, \dots, M. \quad (35b)$$

Since \mathbf{D} is a positive semi-definite matrix, we can decompose \mathbf{D} as

$$\mathbf{D} = \mathbf{F}^H \mathbf{F}, \quad (36)$$

where \mathbf{F} is an $N \times M$ matrix. Denote the i th column of \mathbf{F} by \mathbf{f}_i . Then,

$$\mathbf{A}_i \mathbf{z}_i^{(*)} = \mathbf{f}_i, \quad i = 1, \dots, M, \quad (37)$$

or equivalently

$$\mathbf{A}_i^{(*)} \mathbf{z} = \mathbf{f}_i^{(*)} \quad i = 1, \dots, M. \quad (38)$$

Since the diagonal elements of the optimal \mathbf{D} is E_s , the norm of $\mathbf{f}_i^{(*)}$ is the same as the normal of \mathbf{z} . Then, the unitary matrix $\mathbf{A}_i^{(*)}$ is the $N \times N$ rotation matrix which rotates the vector \mathbf{z} to the vector $\mathbf{f}_i^{(*)}$. For any fixed \mathbf{z} , there could exist many different $\mathbf{A}_i^{(*)}$ that satisfy (38) because the rotation can be performed from any direction in the N -dimensional space. Based on this fact, we can assign an arbitrary value to \mathbf{z} while keep its norm $\sqrt{E_s}$. Once \mathbf{z} and $\mathbf{f}_i^{(*)}$ are fixed, we need to find matrix $\mathbf{A}_i^{(*)}$ from (38). It is noted that a direct derivation for $\mathbf{A}_i^{(*)}$ would be difficult. Therefore, we will first rotate \mathbf{z} by

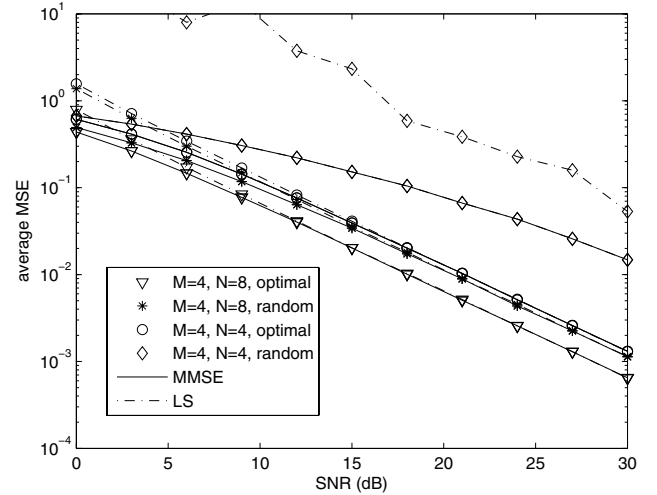


Fig. 1. Channel estimation MSEs versus SNR for both optimal training and random training.

a unitary matrix \mathbf{U} to a simple reference vector, say $\sqrt{E_s} \mathbf{e}_1$, such that

$$\mathbf{U} \mathbf{z} = \sqrt{E_s} \mathbf{e}_1. \quad (39)$$

Obviously, the first row of \mathbf{U} is the normalized vector $\mathbf{z}^H / \|\mathbf{z}\|$, and the remaining $N - 1$ rows of \mathbf{U} is just basis vectors that span the orthogonal complement space of \mathbf{z} . Similarly, we can rotate $\mathbf{f}_i^{(*)}$ to $\sqrt{E_s} \mathbf{e}_1$, and there is

$$\mathbf{V}_i \mathbf{f}_i^{(*)} = \sqrt{E_s} \mathbf{e}_1, \quad (40)$$

where \mathbf{V}_i is the corresponding unitary matrix. Then,

$$\mathbf{A}_i^{(*)} = \mathbf{V}_i^H \mathbf{U} \quad \text{or} \quad \mathbf{A}_i = (\mathbf{V}_i^H \mathbf{U})^{(*)}. \quad (41)$$

IV. SIMULATION RESULTS

In this section, we numerically study the performance of our proposed channel estimation algorithms. The channels g_i , h_i and the noise at the relays or destinations are assumed as circular complex Gaussian random variables with variances 1. The channel covariance matrices \mathbf{R}_h and \mathbf{R}_g have the following structures [10]:

$$[\mathbf{R}_g]_{a,b} = \varepsilon_1^{|a-b|}, \quad [\mathbf{R}_h]_{a,b} = \varepsilon_2^{|a-b|},$$

where ε_1 and ε_2 are two real scalars. For all examples, relays with odd indices operate on \mathbf{r} while relays with even indices operate on \mathbf{r}^* . Since g_i is circularly symmetric, it is known that $E\{g_i g_j\} = 0$. Therefore, $\mathbf{R}_{g^{(*)}}$ can be found from \mathbf{R}_g by setting the appropriate entries to be zero and interchanging some symmetric entries, depending on which relays operate on \mathbf{r}_i^* . The signal to noise ratio is defined as $\text{SNR} = (P_s \times 1) / N_0 = P_s$. The convex optimization involved in MMSE estimation is conducted by the SDP tool SeDuMi v1.1 running under MATLAB environment. One thousands Monte-Carlo runs are taken for average.

1) *Optimal Precoding vs Random Precoding.* In Fig. 1, we display the channel estimation mean square error (MSE) versus SNR of both the LS estimator (8) and the MMSE estimator (18) for $M = 4$ relay nodes. The relay powers are

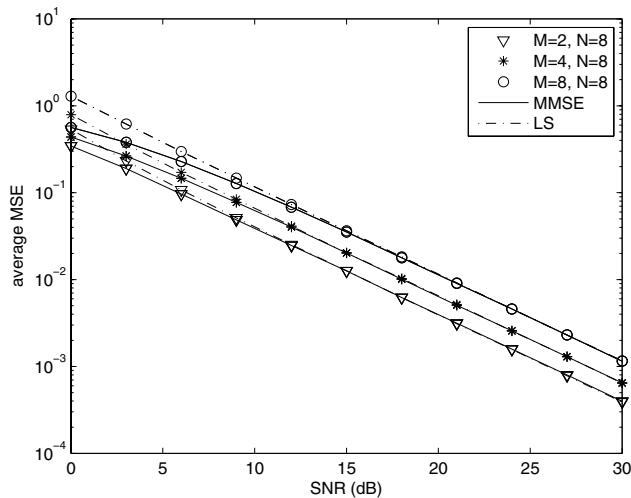


Fig. 2. Channel estimation MSEs versus SNR: different M , same N .

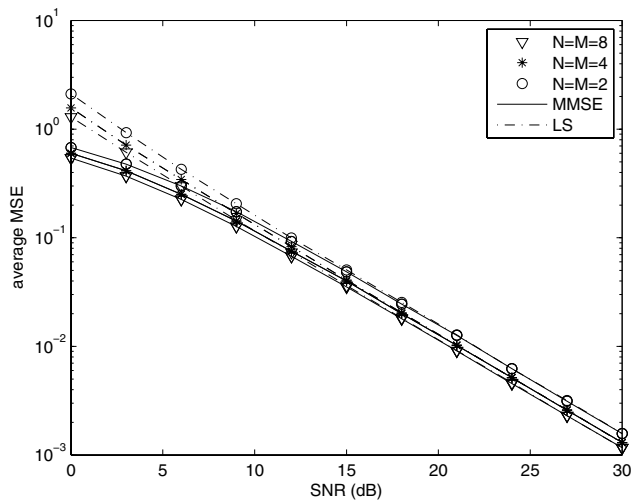


Fig. 3. Channel estimation MSEs versus SNR: different pairs of $N = M$.

chosen as $\{0.8P_s, P_s, 0.8P_s, P_s\}$, respectively. The parameter ε_1 and ε_2 are set as 0.1 which represents a very low correlation between different relays. We show the performance of both the optimal precoding and the random precoding where \mathbf{A}_i are drawn randomly in the latter case. From Fig. 1, we see that when $N = 8$ and with optimal precoding, the MMSE estimator outperforms the LS estimator at lower SNR region, whereas the two estimators have nearly the same performance at higher SNR range. This is a consistent phenomenon as in the traditional multi-input single-output (MISO) channel estimation [10]. Besides, the performance with optimal precoding is about 2dB better than that with random precoding. For the extreme case where N is as small as 4, the performance of both LS estimator and MMSE estimator with optimal precoding degrade somewhat around 2.5dB. However, the channel estimation with random precoding degrades significantly for both estimators.

2) *Different Parameters M, N .* We then consider two different simulation scenarios, e.g. same N , different M in Fig. 2, and different M, N with $M = N$ in Fig. 3. For each

M , half of the relays have the power $0.8P_s$ and the other half have the power P_s . From Fig. 2, it can be seen that the relay number affects the channel estimation even when the same powers of source and same average power of relays are applied. This is a direct reflection from (16) and (32) where it says that the channel estimation MSE increases with M for each w_i . Note that, in traditional MISO systems, the channel estimation MSE is only related with the transmit power but is not related with the number of antennas [10]. The reason for the degrading performance in AF relay networks is because the relays forward additional noise to the destination too. The performances for different M, N but with $M = N$ are shown in Fig. 3, where it can be seen that larger $M = N$ gives better performance. This validates the simultaneous training for all relay channels instead of trivially training each relay channel individually.

V. CONCLUSIONS

In this paper, we studied the training based channel estimation for AF based relay networks. The popular LS and MMSE approaches have both been considered. We show that the optimal training can be achieved from an arbitrary sequence and a set of well designed precoding matrix for each relay node. The whole design process is efficiently conducted by dividing it into a convex optimization problem plus a matrix calculation problem. Numerical examples have been provided to validate the proposed studies.

REFERENCES

- [1] I.E. Telatar, "Capacity of multi-antenna Gaussian channels," *Eur. Trans. Telecom.*, vol. 10, pp. 585-595, Nov. 1999.
- [2] G.J. Foschini, "Layered space time architecture for wireless communication in a fading environment when using multi-element antennas," *Bell Labs. Tech. Jour.*, vol. 1, pp. 41-59, 1996.
- [3] V. Tarokh, N. Seshadri, and A.R. Calderbank, "Space time codes for high data rate wireless communication: performance criterion and code construction," *IEEE Trans. Inform. Theory*, vol. 44, pp. 744-765, 1998.
- [4] S. Alamouti, "A simple transmit diversity technique for wireless communications," *IEEE J. Select. Areas Commun.*, vol. 16, pp. 1451-1458, Oct. 1998.
- [5] T.M. Cover, and A. El Gamal, "Capacity theorems for the relay channel," *IEEE Trans. Inform. Theory*, vol. 25, pp. 572-584, Sept. 1979.
- [6] J.N. Laneman and G.W. Wornell, "Distributed space time block coded protocols for exploiting cooperative diversity in wireless networks," *IEEE Trans. Inform. Theory*, vol. 49, pp. 2415-2425, Oct. 2003.
- [7] S. Yiu, R. Schober, and L. Lampe, "Distributed space time block coding," *IEEE Trans. Commun.*, vol. 54, pp. 1195-1206, July, 2006.
- [8] Y. Jing, and B. Hassibi, "Distributed space time coding in wireless relay networks," *IEEE Trans. Wireless Commun.*, vol. 5, pp. 3524-3536, Dec. 2006.
- [9] J.-J. Xiao and Z.-Q. Luo, "Universal decentralized estimation in an inhomogeneous sensing environment," *IEEE Trans. Inform. Theory*, vol. 51, pp. 3564-3575, Oct. 2005.
- [10] M. Buehner, and A.B. Gershman, "Training based MIMO channel estimation: a study of estimator tradeoffs and optimal training signals," *IEEE Trans. Signal Processing*, vol. 54, pp. 884-893, Mar. 2006.
- [11] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*, Prentice-Hall, 1993.
- [12] L. Vandenberghe and S. Boyd, "Semidefinite programming," *SIAM Rev.*, vol. 39, pp. 49-95, Mar. 1996.
- [13] M. S. Lobo, L. Vandenberghe, S. Boyd, and H. Lebret, "Applications of second-order cone programming," *Linear Algebra Appl.*, vol. 284, pp. 193-228, 1998.
- [14] S. Haykin, *Adaptive Filter Theory*, 4th ed., Prentice-Hall, 1997.
- [15] F. Gao, T. Cui, and A. Nallanathan, "On channel estimation and optimal training design for amplify and forward relay networks," submitted to *IEEE Trans. Wireless Commun.*