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

Feifei Gao, Tao Cui, and Arumugam Nallanathan, *Senior Member, IEEE*

Abstract— In this paper, we provide a complete study on the training based channel estimation issues for relay networks that employ the *amplify-and-forward* (AF) transmission scheme. We first point out that separately estimating the channel from source to relay and relay to destination suffers from many drawbacks. Then we provide a new estimation scheme that directly estimates the overall channels from the source to the destination. The proposed channel estimation well serves the AF based space time coding (STC) that was recently developed. There exists many differences between the proposed channel estimation and that in the traditional single input single out (SISO) and multiple input single output (MISO) systems. For example, a relay must linearly precode its received training sequence by a sophisticatedly designed matrix in order to minimize the channel estimation error. Besides, each relay node is individually constrained by a different power requirement because of the non-cooperation among all relay nodes. We study both the linear least-square (LS) estimator and the minimum mean-square-error (MMSE) estimator. The corresponding optimal training sequences, as well as the optimal precoding matrices are derived from an efficient convex optimization process.

Index Terms— Channel estimation, amplify and forward relay networks, cooperative communications, optimal training, optimal precoding.

I. INTRODUCTION

IT IS WELL known that in wireless communication systems, employing multiple antennas can boost the transmission capacity which, in turn, increases the overall data throughput [1], [2]. On the other hand, diversity techniques that apply the space time coding (STC) on multiple transmit antennas could combat the detrimental effects of the multiplicative time selective fading and thus enhances the transmission reliability [3], [4]. Unfortunately, packing many antennas onto a small mobile terminal normally faces the practical difficulty of the size limit. In order to overcome this limitation, one would refer to the relay network, where the space diversity could be exploited from relay nodes existing in the network [5]- [13]. These relay nodes can either be provided by the telecommunication agency or could be obtained from cooperating terminals of other users. The latter scenario is

F. Gao is with the Institute for Infocomm Research, 21 Heng Mui Keng Terrace, Singapore, 119613 (e-mail: feifeigao@ieee.org).

T. Cui is with the Department of Electrical Engineering, California Institute of Technology, Pasadena, CA 91125, USA (e-mail: taocui@caltech.edu).

A. Nallanathan is with the Division of Engineering, King's College London, United Kingdom (e-mail: arumugam.nallanathan@kcl.ac.uk).

sometimes referred to as "cooperative communication" since each user, although acted as a relay for a certain period, still has its own information to transmit.

The relay based transmission is usually divided into two phases. During Phase I, the source broadcasts 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 [7]- [13] for either AF or DF approach.

Before enjoying all those benefits brought by the relay network, 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 relay network could be built upon following the same approach between the individual transmitters and receivers [8], [9]. However, this belief is only valid for the DF scheme where Phase I and Phase II are conducted independently. For AF based transmission, separating the channel estimation from two phases will meet several drawbacks. For example, the relay must inform the destination the CSI of Phase I, which not only reduces the bandwidth efficiency but also consumes additional transmitting power. Besides, transmitting the estimated channel will suffer from further distortions.

In this work, we propose a novel training based channel estimation scheme, where the overall channel from source to destination is estimated at the destination only. Specifically, a unique training sequence will be broadcasted to all relays during Phase I, and the relays forward the received signals, after performing a linear transformation, to the destination during Phase II. Both the least square (LS) and the minimum mean square error (MMSE) channel estimation algorithms are derived. Since the relay networks conduct the distributed transmissions, each relay will have its own power constraint. This is a major difference from the traditional single-user multi-antenna transmission, where there is a total power consumption constraint for all transmit antennas. Consequently, the optimal training design may not have closed form solution all the time. The optimal training design includes designing both the training sequence sent from the source and the linear precoding matrix at each relay. It is shown that the
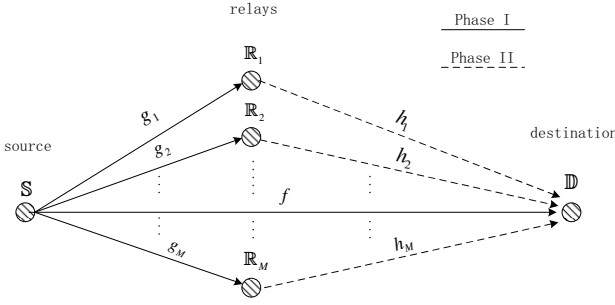
Fig. 1.   Wireless relay networks with one source, one destination and M relays.

optimal design is in fact a convex optimization problem, which could be efficiently solved by modern interior point method based convex optimization software [14], and the resulting complexity is bounded by the polynomial $O(M^{6.5}log(1/\epsilon))$, where M is the number of the relays and $\epsilon > 0$ is the solution accuracy [15]. Nevertheless, the closed form optimization has been obtained for two special cases; namely, when channel to or from different relays are mutually uncorrelated, and when the signal-to-noise ratio (SNR) is higher than a certain threshold. To simplify the relay design, we also study some popular STCs in the AF based relay network and propose uniform linear precoding matrices that could be applied over both the training and the data transmitting periods.

The rest of the paper is organized as follows. Section II presents the system model of AF based relay networks and illustrates our main purpose of developing a new channel estimation technique. Section III provides the proposed LS and MMSE channel estimation algorithm, as well as the optimal training and the design of precoding matrices. In section IV, we study some popular STCs for the AF based relay networks. Section V displays simulation results to validate the proposed studies. Finally conclusions are drawn in Section VI and related proofs are provided in the Appendix.

Notations: Vectors and matrices are boldface small and capital letters; the transpose, complex conjugate, Hermitian, inverse, and pseudo-inverse of the matrix $\mathbf{A}$ are denoted by $\mathbf{A}^T$, $\mathbf{A}^*$, $\mathbf{A}^H$, $\mathbf{A}^{-1}$ and $\mathbf{A}^\dagger$, respectively; $\mathrm{tr}(\mathbf{A})$ and $\|\mathbf{A}\|_F$ are the trace and the Frobenius norm of $\mathbf{A}$; $[\mathbf{A}]_{ij}$ is the (i, j)th entry of $\mathbf{A}$ and $\mathrm{diag}\{\mathbf{a}\}$ denotes a diagonal matrix with the diagonal element constructed from $\mathbf{a}$; $\mathbf{I}$ is the identity matrix; $\mathrm{E}\{\cdot\}$ denotes the statistical expectation.

II. SPACE TIME CODING IN AF RELAY NETWORKS

Consider a wireless network with M randomly placed relay nodes $\mathbb{R}_i$, $i = 1, ..., M$, one source node $\mathbb{S}$, and one destination node $\mathbb{D}$, as shown in Fig. 1. 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 $f \in \mathcal{CN}(0, \sigma_f^2)$, $g_i \in \mathcal{CN}(0, \sigma_{gi}^2)$, and $h_i \in \mathcal{CN}(0, \sigma_{hi}^2)$. Throughout this paper, we assume perfect

synchronization among all terminals as in [13]. Assume that $\mathbb{S}$ wishes to send signal block $\mathbf{s} = [s_1, ..., s_T]^T$ to $\mathbb{D}$ via the AF mode. The transmission is accomplished by the following two phases, each containing T consecutive time slots. 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 then be expressed as

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

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

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 $\mathrm{E}\{\mathbf{s}^H\mathbf{s}\} = TP_s$, where P_s is the average transmitting power of the 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. A linear dispersion (LD) based STC has been proposed in [13], 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 [13] to

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

where $(\cdot)^{(*)}$ represents the item itself if the ith relay operates on $\mathbf{r}_i$ whereas represents the conjugate of the item if the ith relay operates on $\mathbf{r}_i^*$. Note that this type of STC, where one relay operates on either $\mathbf{r}_i$ or $\mathbf{r}_i^*$, exclusively, has been proposed and analyzed in [13] and [16] for AF relay networks.

Basically, there are two different choices of α_i, which are listed as follows:

$$\alpha_i = \sqrt{\frac{P_{ri}}{|g_i|^2P_s + N_0}}, \tag{4a}$$

$$\alpha_i = \sqrt{\frac{P_{ri}}{\sigma_{gi}^2P_s + N_0}}. \tag{4b}$$

The second choice of α_i is recommended since it is not a random value while keeps the power constraint from the long term point of view. Besides, g_i in (4a) could only be replaced by its estimate $\hat{g}_i$ which may not keep the average relay power exactly as P_{ri}. In this work, we will adopt α_i in (4b). Nevertheless, the provided studies could be easily extended for α_i in (4a). The destination $\mathbb{D}$ in Phase II then receives

$$\begin{aligned}
\mathbf{d}_2 &= \sum_{i=1}^{M} h_i\mathbf{t}_i + \mathbf{n}_{d2} \\
&= \sum_{i=1}^{M} w_i\alpha_i\mathbf{P}_i\mathbf{s}_i^{(*)} + \sum_{i=1}^{M} h_i\alpha_i\mathbf{P}_i\mathbf{n}_{ri}^{(*)} + \mathbf{n}_{d2} \\
&= \mathbf{B}\mathbf{\Lambda}\mathbf{w} + \mathbf{n}_d,
\end{aligned} \tag{5}$$

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 $i \in \{1, \ldots, M\}^1$. Other variables are defined as follows:

$$\mathbf{w} = [w_1, ..., w_M]^T, \ w_i = h_i g_i^{(*)}, \ i = 1, \ldots, M,$$
$$\mathbf{\Lambda} = \text{diag}\{\alpha_1, ..., \alpha_M\},$$
$$\mathbf{B} = [\mathbf{P}_1 \mathbf{s}_1^{(*)}, \mathbf{P}_2 \mathbf{s}_2^{(*)}, ..., \mathbf{P}_M \mathbf{s}_M^{(*)}], \quad (6)$$
$$\mathbf{n}_d = \sum_{i=1}^{M} h_i \alpha_i \mathbf{P}_i \mathbf{n}_{ri}^{(*)} + \mathbf{n}_{d2}.$$

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

$$\text{Cov}(\mathbf{n}_d | h_i, i = 1, ..., M) = (\sum_{i=1}^{M} |h_i|^2 \alpha_i^2 + 1) N_0 \mathbf{I}, \quad (7)$$

where the property $\mathbf{P}_i \mathbf{P}_i^H = \mathbf{I}$ is utilized. Therefore, the overall noise, under a specific realization of h_i, is still white Gaussian but with a scaled covariance of $N_0 \mathbf{I}$. It has been shown in [13] that for well designed $\mathbf{P}_i$, the maximum possible diversity order can be approximated by $\min\{T, M\}$ at high SNR. Therefore, similar to the multi-antenna case, there is no point in having more relays than the coherence interval [3], [17]. In the following discussions, we will assume $T \geq M$.

III. TRAINING BASED CHANNEL ESTIMATION

For coherent detection in the AF mode [8]- [13], 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. This is reasonable since considering the statistics of $h_i \mathbf{n}_{ri}^{(*)}$ is not convenient. Besides, g_i, h_i, and w_i vary little in a frame based communication system. 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[2] is usually constant. Therefore, additional energy will be consumed when transmitting over 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 [18]. 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 during Phase II. Denote the $N \times N$ unitary precoding matrix at the ith relay as $\mathbf{A}_i$ and define

$$\mathbf{C} = [\mathbf{A}_1 \mathbf{z}_1^{(*)}, \mathbf{A}_2 \mathbf{z}_2^{(*)}, ..., \mathbf{A}_M \mathbf{z}_M^{(*)}]. \quad (8)$$

The transmitting model with other equations from (1) to (7) 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 replaced by $\mathbf{z}^H \mathbf{z} \leq N P_s = E_s$.

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

$$\hat{\mathbf{w}}_{LS} = \mathbf{\Lambda}^{-1} \mathbf{C}^\dagger \mathbf{d}_2 = \mathbf{w} + \Delta \mathbf{w} \quad (9)$$

with error

$$\Delta \mathbf{w} = \mathbf{\Lambda}^{-1} \mathbf{C}^\dagger \mathbf{n}_d. \quad (10)$$

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

$$\text{Cov}(\Delta \mathbf{w} | \mathbf{g}^{(*)}, \mathbf{h}) = N_0 \left(\sum_i |h_i|^2 |\alpha_i|^2 + 1 \right)$$
$$\times \mathbf{\Lambda}^{-1} (\mathbf{C}^H \mathbf{C})^{-1} \mathbf{\Lambda}^{-1}, \quad (11)$$

where $\mathbf{g}^{(*)} = [g_1^{(*)}, g_2^{(*)}, \ldots, g_M^{(*)}]^T$ and $\mathbf{h} = [h_1, h_2, \ldots, h_M]^T$ are defined for convenience. Meanwhile, we would like to define $\mathbf{g} = [g_1, g_2, \ldots, g_M]^T$ for future use. Since $\mathbf{\Lambda}$ is a constant matrix, the optimization is conducted by varying the value of $\mathbf{C}$. Noted that the diagonal elements of $\mathbf{C}$ must all be no greater than E_s. Therefore, the optimal $\mathbf{C}$ can be found by solving the following constrained optimization problem:

$$\min_{\mathbf{A}_i, \mathbf{z}} \text{tr} \left(\mathbf{\Lambda}^{-1} (\mathbf{C}^H \mathbf{C})^{-1} \mathbf{\Lambda}^{-1} \right)$$
$$\text{subject to} \quad [\mathbf{C}^H \mathbf{C}]_{ii} \leq E_s, \quad i = 1, ...M. \quad (12)$$

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 [19]. In the relay networks, since different relays could not share a common power pool, each relay will have its own power constraint P_{ri}, which is reflected by M individual constraints in (12).

Theorem 1: The optimal $\mathbf{C}^H \mathbf{C}$ in (12) must be $E_s \mathbf{I}$.

Proof: We first prove that $\mathbf{C}^H \mathbf{C}$ must be a diagonal matrix. The following inequality for an arbitrary $N \times N$ positive definite matrix $\mathbf{F}$ could be utilized [20]:

$$\text{tr} \left(\mathbf{F}^{-1} \right) \geq \sum_{i=1}^{N} \left([\mathbf{F}]_{ii} \right)^{-1}, \quad (13)$$

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 (12). Let $\mathbf{F} = \mathbf{\Lambda} \mathbf{D}_0 \mathbf{\Lambda}$ whose diagonal part is given by

$$\bar{\mathbf{F}} = \text{diag}\{\mathbf{F}\} = \mathbf{\Lambda} \bar{\mathbf{D}}_0 \mathbf{\Lambda}. \quad (14)$$

Using the inequality (13), 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 constraint in (12), we know that the optimal choice is $\mathbf{C}^H\mathbf{C} = E_s\mathbf{I}$. ■

Theorem 1 says that each relay should transmit orthogonal training whereas should meet their own power constraint. 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

$$(\mathbf{z}_i^{(*)})^H\mathbf{A}_i\mathbf{A}_j^H\mathbf{z}_j^{(*)} = 0 \quad \text{for } i \neq j. \tag{15}$$

There exists a number of such $\mathbf{z}$ and unitary matrices $\mathbf{A}_i$. For example, we could take $\mathbf{z}$ as $\sqrt{P_s}\mathbf{1}$ and $\mathbf{A}_i$ as $\text{diag}(\mathbf{a}_i)$, where $\mathbf{1}$ is an all one vector and $\mathbf{a}_i$ is the ith column of an $N \times N$ normalized discrete Fourier transform (DFT) matrix. We will provide a more general design in the later discussions. It needs to be mentioned that choosing $\mathbf{z} = \sqrt{P_s}\mathbf{1}$ is preferable since it releases the peak to average power ratio (PAPR) problem for the transmitter.

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

$$\text{var}(w_i|\mathbf{h},\mathbf{g}) = \frac{N_0(\sum_j |h_j|^2|\alpha_j|^2 + 1)}{|\alpha_i|^2 E_s}, \tag{16}$$

and

$$\text{var}(w_i) = \frac{N_0\sigma_{hi}^2}{E_s} + \frac{N_0(\sum_{j\neq i}\sigma_{hj}^2|\alpha_j|^2/|\alpha_i|^2)}{E_s} + \frac{N_0\sigma_{gi}^2}{E_{ri}} + \frac{N_0^2}{E_s E_{ri}}, \tag{17}$$

where $E_{ri} = NP_{ri}$ is defined for notation simplicity. It can be seen that the second term is the cross distortion since other relays also forward a noisy version of the training sequence to the destination.

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

$$\mathbf{R}_w = \text{E}\{\mathbf{w}\mathbf{w}^H\} = \mathbf{R}_{g^{(*)}} \odot \mathbf{R}_h, \tag{18}$$

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

$$\hat{\mathbf{w}}_{MMSE} = \text{E}\{\mathbf{w}\mathbf{d}_2^H\}(\text{E}\{\mathbf{d}_2\mathbf{d}_2^H\})^{-1}\mathbf{d}_2. \tag{19}$$

With straightforward calculations,

$$\text{E}\{\mathbf{w}\mathbf{d}_2^H\} = \mathbf{R}_w\mathbf{\Lambda}\mathbf{C}^H, \tag{20a}$$

$$\text{E}\{\mathbf{d}_2\mathbf{d}_2^H\} = \mathbf{C}\mathbf{\Lambda}\mathbf{R}_w\mathbf{\Lambda}\mathbf{C}^H + N_0\sum_i(\sigma_{hi}^2|\alpha_i|^2 + 1)\mathbf{I}. \tag{20b}$$

Then

$$\hat{\mathbf{w}}_{MMSE} = \mathbf{R}_w\mathbf{\Lambda}\mathbf{C}^H\Big(\mathbf{C}\mathbf{\Lambda}\mathbf{R}_w\mathbf{\Lambda}\mathbf{C}^H + N_0\sum_i(\sigma_{hi}^2|\alpha_i|^2 + 1)\mathbf{I}\Big)^{-1}\mathbf{d}_2. \tag{21}$$

The error covariance of the MMSE estimator is

$$\text{Cov}(\Delta\mathbf{w}) = \left(\mathbf{R}_w^{-1} + \frac{1}{N_0\sum_i(\sigma_{hi}^2|\alpha_i|^2 + 1)}\mathbf{\Lambda}\mathbf{C}^H\mathbf{C}\mathbf{\Lambda}\right)^{-1}. \tag{22}$$

The optimal training should be obtained from

$$\min_{\mathbf{A}_i,\mathbf{z}} \text{tr}(\text{Cov}(\Delta\mathbf{w})), \text{ subject to } [\mathbf{C}^H\mathbf{C}]_{ii} \leq E_s, \ i = 1,...,M. \tag{23}$$

The problem (23) 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

$$\min_{\mathbf{D}} \text{tr}\left(\left(\mathbf{R}_w^{-1} + \frac{1}{N_0\sum_i(\sigma_{hi}^2|\alpha_i|^2 + 1)}\mathbf{\Lambda}\mathbf{D}\mathbf{\Lambda}\right)^{-1}\right) \tag{24}$$

$$\text{s.t.} \quad [\mathbf{D}]_{i,i} \leq E_s,$$

$$\mathbf{D} \succcurlyeq \mathbf{0},$$

where $\mathbf{M} \succcurlyeq \mathbf{N}$ means that $(\mathbf{M} - \mathbf{N})$ is a positive semi-definite matrix. This is precisely a convex optimization problem. Using an auxiliary matrix $\mathbf{T}$, we can rewrite (24) in the following equivalent form:

$$\min_{\mathbf{T},\mathbf{D}} \quad \text{tr}(\mathbf{T}) \tag{25}$$

$$\text{s.t.} \quad \mathbf{T} \succcurlyeq \left(\mathbf{R}_w^{-1} + \frac{1}{N_0\sum_i(\sigma_{hi}^2|\alpha_i|^2 + 1)}\mathbf{\Lambda}\mathbf{D}\mathbf{\Lambda}\right)^{-1},$$

$$[\mathbf{D}]_{i,i} \leq E_s,$$

$$\mathbf{D} \succcurlyeq \mathbf{0}.$$

The equivalence of (25) and (24) can be established by observing that the optimal $\mathbf{T}$ must be equal to $(\mathbf{R}_w^{-1} + \frac{1}{N_0\sum_i(\sigma_{hi}^2|\alpha_i|^2+1)}\mathbf{\Lambda}\mathbf{D}\mathbf{\Lambda})^{-1}$. The reason is seen from the simple linear algebra that $\text{tr}(\mathbf{M}) \geq \text{tr}(\mathbf{N})$ for all $\mathbf{M} \succcurlyeq \mathbf{N}$. Moreover, note that the constraint

$$\mathbf{T} \succcurlyeq \left(\mathbf{R}_w^{-1} + \frac{1}{N_0\sum_i(\sigma_{hi}^2|\alpha_i|^2 + 1)}\mathbf{\Lambda}\mathbf{D}\mathbf{\Lambda}\right)^{-1} \tag{26}$$

can be rewritten, via Schur's complement [21], into the following linear matrix inequality (LMI):

$$\begin{bmatrix} \mathbf{T} & \mathbf{I} \\ \mathbf{I} & \mathbf{R}_w^{-1} + \frac{1}{N_0\sum_i(\sigma_{hi}^2|\alpha_i|^2+1)}\mathbf{\Lambda}\mathbf{D}\mathbf{\Lambda} \end{bmatrix} \succcurlyeq \mathbf{0}. \tag{27}$$

Therefore, (25) lies in the so called semi-definite programming (SDP) formulation. Since both the cost function and the constraints are convex, the SDP formulation could be solved efficiently by interior point methods [15]. The convexity of (25) ensures that its global minimum can be found in polynomial time without the usual headaches of step size selection, algorithm initialization, or the risk of local minima. The arithmetic complexity of the interior point methods for solving the SDP (25) is $O(M^{6.5}log(1/\epsilon))$, where $\epsilon > 0$ is the solution accuracy [15].

Although, the efficient polynomial time searching can be applied to (25), there exist closed form solutions for several important scenarios. First let us ignore the condition $\mathbf{D} \succcurlyeq \mathbf{0}$

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

$$L(\mathbf{C}, \mu_i) = \text{tr}\left(\left(\mathbf{R}_w^{-1} + \frac{1}{\beta}\mathbf{\Lambda}\mathbf{C}^H\mathbf{C}\mathbf{\Lambda}\right)^{-1}\right) + \sum_i \mu_i(\text{tr}(\mathbf{e}_i^H\mathbf{C}^H\mathbf{C}\mathbf{e}_i) - E_s), \tag{28}$$

where $\mathbf{e}_i$ is the ith column of an $N \times N$ identity matrix. Since (28) is in quadratic form, it is sufficient to consider only $\partial L(\mathbf{C}, \mu_i)/\partial\mathbf{C}$ [22].

Lemma 1: If a square matrix $\mathbf{F}$ is a function of another square matrix $\mathbf{G} = \mathbf{\Delta} + \mathbf{M}^H\mathbf{X}^H\mathbf{X}\mathbf{M}$, then the following chain rule is valid:

$$\frac{\partial\text{tr}\{\mathbf{F}\}}{\partial\mathbf{X}} = \mathbf{X}^*\mathbf{M}^*\frac{\partial\text{tr}\{\mathbf{F}\}}{\partial\mathbf{G}}\mathbf{M}^T, \tag{29}$$

where $\mathbf{M}$ is a constant matrix and $\mathbf{\Delta}$ is a constant positive semi-definite matrix.

Proof: See Appendix.

Furthermore, the following matrix derivatives will be used

$$\frac{\partial\text{tr}\{\mathbf{X}\mathbf{X}^H\}}{\partial\mathbf{X}} = \mathbf{X}^* \tag{30a}$$

$$\frac{\partial\text{tr}\{\mathbf{X}^{-1}\}}{\partial\mathbf{X}} = -\mathbf{X}^{-T}\mathbf{X}^{-T} = -\mathbf{X}^{-2T}. \tag{30b}$$

Inserting $\mathbf{X} = \mathbf{C}$, $\mathbf{F} = \text{tr}((\mathbf{R}_w^{-1} + \frac{1}{\beta}\mathbf{\Lambda}\mathbf{C}^H\mathbf{C}\mathbf{\Lambda})^{-1})$, $\mathbf{M} = \mathbf{\Lambda}/\sqrt{\beta}$ into (29) and applying (30b), we obtain

$$\frac{\partial\text{tr}\left(\left(\mathbf{R}_w^{-1} + \frac{1}{\beta}\mathbf{\Lambda}\mathbf{C}^H\mathbf{C}\mathbf{\Lambda}\right)^{-1}\right)}{\partial\mathbf{C}} \tag{31}$$

$$= \frac{1}{\beta}\mathbf{C}^*\mathbf{\Lambda}\frac{\partial\text{tr}\left(\left(\mathbf{R}_w^{-1} + \frac{1}{\beta}\mathbf{\Lambda}\mathbf{C}^H\mathbf{C}\mathbf{\Lambda}\right)^{-1}\right)}{\partial\left(\mathbf{R}_w^{-1} + \frac{1}{\beta}\mathbf{\Lambda}\mathbf{C}^H\mathbf{C}\mathbf{\Lambda}\right)}\mathbf{\Lambda}$$

$$= -\frac{1}{\beta}\mathbf{C}^*\mathbf{\Lambda}\left(\mathbf{R}_w^{-1} + \frac{1}{\beta}\mathbf{\Lambda}\mathbf{C}^H\mathbf{C}\mathbf{\Lambda}\right)^{-2T}\mathbf{\Lambda}. \tag{32}$$

Using (30a) to compute the second term in (28), we have

$$\frac{\text{tr}\left(\mathbf{e}_i^T\mathbf{C}^H\mathbf{C}\mathbf{e}_i\right) - E_s}{\partial\mathbf{C}} = \mathbf{C}^*\mathbf{e}_i\mathbf{e}_i^T. \tag{33}$$

Then, (28) can be rewritten as

$$\frac{\partial L(\mathbf{C}, \mu_i)}{\partial\mathbf{C}} = \mathbf{C}^*\left(\sum_i \mu_i\mathbf{e}_i\mathbf{e}_i^T - \frac{1}{\beta}\mathbf{\Lambda}\left(\mathbf{R}_w^{-1} + \frac{1}{\beta}\mathbf{\Lambda}\mathbf{C}^H\mathbf{C}\mathbf{\Lambda}\right)^{-2T}\mathbf{\Lambda}\right). \tag{34}$$

Therefore, any $\mathbf{C}$ can be the optimal solution if (34) is zero. Since $\mathbf{C}$ is a tall and full rank matrix, (34) is equivalent to

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

$$= \mathbf{\Omega}, \tag{35}$$

where

$$\mathbf{\Omega} = \text{diag}\left\{\frac{\beta}{\alpha_1^2\mu_1}, \frac{\beta}{\alpha_2^2\mu_2}, ..., \frac{\beta}{\alpha_M^2\mu_M}\right\}. \tag{36}$$

To get the explicit form of $\mathbf{R}_w^{-1} + \frac{1}{\beta}\mathbf{\Lambda}\mathbf{C}^H\mathbf{C}\mathbf{\Lambda}$, we can eigen-decompose it as

$$\mathbf{R}_w^{-1} + \frac{1}{\beta}\mathbf{\Lambda}\mathbf{C}^H\mathbf{C}\mathbf{\Lambda} = \mathbf{U}_b\mathbf{\Sigma}_b\mathbf{U}_b^H. \tag{37}$$

Then

$$\mathbf{\Omega} = \mathbf{U}_b\mathbf{\Sigma}_b^2\mathbf{U}_b^H, \tag{38}$$

which also represents the eigen-decomposition of $\mathbf{\Omega}$. Since $\mathbf{\Omega}$ is itself a diagonal matrix with non-negative diagonal entries, $\mathbf{\Sigma}_b^2$ must be $\mathbf{\Omega}$. Therefore,

$$\mathbf{R}_w^{-1} + \frac{1}{\beta}\mathbf{\Lambda}\mathbf{C}^H\mathbf{C}\mathbf{\Lambda} = \mathbf{\Omega}^{1/2}. \tag{39}$$

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

$$\begin{aligned}\mathbf{\Omega}^{1/2} &= \text{diag}\{\mathbf{\Omega}^{1/2}\} \\ &= \tilde{\mathbf{R}}_{w,\text{inv}} + \frac{1}{\beta}\mathbf{\Lambda}\text{diag}\{\mathbf{C}^H\mathbf{C}\}\mathbf{\Lambda} \\ &= \tilde{\mathbf{R}}_{w,\text{inv}} + \frac{E_s}{\beta}\mathbf{\Lambda}^2.\end{aligned} \tag{40}$$

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

$$\begin{aligned}\mathbf{C}^H\mathbf{C} &= \beta\mathbf{\Lambda}^{-1}(\mathbf{\Omega}^{1/2} - \mathbf{R}_w^{-1})\mathbf{\Lambda}^{-1} \\ &= E_s\mathbf{I} - \beta\mathbf{\Lambda}^{-1}\tilde{\mathbf{R}}_{w,off}\mathbf{\Lambda}^{-1},\end{aligned} \tag{41}$$

where $\tilde{\mathbf{R}}_{w,off}$ 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 $\mathbf{D} = \mathbf{C}^H\mathbf{C} \succeq \mathbf{0}$ in order to arrive at a true optimal solution. Therefore, (41) is the optimal solution only when $E_s\mathbf{I} - \beta\mathbf{\Lambda}^{-1}\tilde{\mathbf{R}}_{w,off}\mathbf{\Lambda}^{-1} \succeq \mathbf{0}$. This positive semi-definite requirement is satisfied under two important scenarios:

a) When channel w_i are uncorrelated with each other; namely, $\mathbf{R}_w$ is only a diagonal matrix. From (18), we know when either $\mathbf{R}_{g(*)}$ or $\mathbf{R}_h$ is diagonal, $\mathbf{R}_w$ becomes a diagonal matrix. In other words, if the channels from source to all relays are uncorrelated or if the channels from all relays to destination are uncorrelated, then $\mathbf{R}_w$ is a diagonal matrix. Note that the uncorrelated channel condition for either g_i or h_i is normally satisfied since relays are randomly distributed over a certain region such that the channels from either branches are uncorrelated. 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

$$\text{Cov}(\Delta\mathbf{w}) = \left(\mathbf{R}_w^{-1} + \frac{E_s}{\beta}\mathbf{\Lambda}^2\right)^{-1}, \tag{42}$$

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

$$\text{var}(w_i) = \frac{N_0(\sum_j\sigma_{hj}^2|\alpha_j|^2 + 1)}{|\alpha_i|^2E_s + N_0\sum_j\sigma_{hj}^2|\alpha_j|^2/\sigma_{wi}^2}. \tag{43}$$

Compared to the MSE of the LS estimation (17), the MSE in (43) 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

$$\text{Cov}(\Delta\mathbf{w}) = \left(\tilde{\mathbf{R}}_{w,\text{inv}} + \frac{E_s}{\beta}\mathbf{\Lambda}^2\right)^{-1}, \qquad (44)$$

and the MSE on w_i can be separately obtained as

$$\text{var}(w_i) = \frac{N_0(\sum_j \sigma_{hj}^2|\alpha_j|^2 + 1)}{|\alpha_i|^2 E_s + N_0\sum_j \sigma_{hj}^2|\alpha_j|^2/\tilde{\sigma}_{wi}^2}. \qquad (45)$$

where $\tilde{\sigma}_{wi}^2$ is the ith diagonal element of $\mathbf{R}_w^{-1}$. Similarly, the MSE in (45) is also smaller than (17).

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

Assume that the optimal $\mathbf{D}$ has been derived from either (25) or (41). 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:

$$(\mathbf{z}_i^{(*)})^H \mathbf{A}_i^H \mathbf{A}_j \mathbf{z}_j^{(*)} = [\mathbf{D}]_{ij}, \quad i,j = 1,...,M, \quad (46a)$$

$$\mathbf{A}_i^H \mathbf{A}_i = \mathbf{I}, \quad i = 1,...,M. \qquad (46b)$$

The standard procedure to achieve an approximate solution of a set of nonlinear equations using either the method of steepest descent [23] or the technique suggested by Newton and Raphson [24], requires the evaluation of partial derivative to each unknown variables. This involves considerably time consuming computation. Due to the speciality observed in (46), the following way could be proceeded to achieve the closed form solutions of $\mathbf{z}$ and $\mathbf{A}_i$. Since $\mathbf{D}$ is a positive semi-definite matrix, we can decompose $\mathbf{D}$ as

$$\mathbf{D} = \mathbf{F}^H\mathbf{F}, \qquad (47)$$

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

$$\mathbf{A}_i\mathbf{z}_i^{(*)} = \mathbf{f}_i, \quad i = 1,...,M, \qquad (48)$$

or equivalently

$$\mathbf{A}_i^{(*)}\mathbf{z} = \mathbf{f}_i^{(*)}, \quad i = 1,...,M. \qquad (49)$$

The problem now becomes to find M unitary matrices $\mathbf{A}_i^{(*)}$ and $\mathbf{z}$, such that all M non-linear equations (49) are satisfied.

Since the diagonal elements of the optimal $\mathbf{D}$ is E_s, the norm of $\mathbf{f}_i^{(*)}$ is then the same as the normal of $\mathbf{z}$. Hence, the unitary matrix $\mathbf{A}_i^{(*)}$ is the N dimensional 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 (49) 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 $\sqrt{E_s}$. Any other vector $\bar{\mathbf{z}}$ would only be a rotated version from $\mathbf{z}$, which could be represented as $\bar{\mathbf{z}} = \mathbf{U}\mathbf{z}$, where $\mathbf{U}$ is some unitary matrix. To this end, we could simply take $\mathbf{A}_i^{(*)}\mathbf{U}$ as the new rotation matrix while keep $\mathbf{z}$ unchanged.

Once $\mathbf{z}$ and $\mathbf{f}_i^{(*)}$ are fixed, we need to find matrix $\mathbf{A}_i^{(*)}$ from (49). It is noted that a direct derivation for $\mathbf{A}_i^{(*)}$ would

be difficult. We will first rotate $\mathbf{z}$ by a unitary matrix $\mathbf{U}$ to a simple reference vector, say $\sqrt{E_s}\mathbf{e}_1$, such that

$$\mathbf{U}\mathbf{z} = \sqrt{E_s}\mathbf{e}_1. \qquad (50)$$

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 $\mathbf{z}$. These $N-1$ rows could be easily obtained from the left singular vectors of $\mathbf{z}$.

Similarly, we will rotate $\mathbf{f}_i^{(*)}$ to $\sqrt{E_s}\mathbf{e}_1$, and there is

$$\mathbf{V}_i\mathbf{f}_i^{(*)} = \sqrt{E_s}\mathbf{e}_1, \qquad (51)$$

where $\mathbf{V}_i$ is the corresponding unitary matrix. After obtaining both $\mathbf{U}$ and $\mathbf{V}_i$, $\mathbf{A}_i^{(*)}$ can be derived from

$$\mathbf{A}_i^{(*)} = \mathbf{V}_i^H\mathbf{U} \quad \text{or} \quad \mathbf{A}_i = (\mathbf{V}_i^H\mathbf{U})^{(*)}. \qquad (52)$$

As mentioned before, the choice of $\mathbf{z}$ as $\sqrt{P_s}\mathbf{1}$ is preferred in order to alleviate the PAPR problem.

IV. CONSISTENCY WITH SPACE TIME CODING

Generally, the $\mathbf{A}_i$ designed for the channel estimation in the previous section is not the same as the LD codes $\mathbf{P}_i$ even if $N = T$. Therefore, each relay needs to store both $\mathbf{P}_i$, $\mathbf{A}_i$ and should perform different precoding at different stage. Although this is not a big deal with perfect synchronization, using unique precoding during the whole transmission is still preferred due to its reduction on complexity.

To make consistency between precoding at training and at data transmission, we will consider a special yet very popular transmission scenario, where $\mathbf{R}_w$ is diagonal and orthogonal space time coding (OSTC), or quasi-orthogonal space time coding (QOSTC) is applied. The OSTC and QOSTC are chosen because they could maximize the diversity gain while achieves a higher coding gain [16], but the proposed discussion is also applicable if other types of STC are adopted.

A. Real and Complex OSTC

From [25], [4], we know that the dimension of real OSTC could only be $M = 2, 4, 8$ and the dimension of complex OSTC could only be $M = 2$. Without loss of generality, we suppose the training block has the length $N = KM$ for some positive integer K. Let us partition $\mathbf{z}$ in to $\mathbf{z} = [\mathbf{z}_{(1)}^T, \mathbf{z}_{(2)}^T, ..., \mathbf{z}_{(K)}^T]^T$ with each $\mathbf{z}_{(k)}$ an $M \times 1$ vector. We use $(\cdot)$ in the subscript in order to discriminate $\mathbf{z}_{(k)}$ from $\mathbf{z}_k$. If $\mathbf{P}_i$ is also applied over the training period, the linear precoding matrix $\mathbf{A}_i$ is in fact

$$\mathbf{A}_i = \text{diag}\left\{\underbrace{\mathbf{P}_i, \mathbf{P}_i, ..., \mathbf{P}_i}_{K \text{ blocks}}\right\}. \qquad (53)$$

Since $\mathbf{P}_i$ is obtained from OSTC design, $\mathbf{C}^H\mathbf{C} = NP_s\mathbf{I} = E_s\mathbf{I}$ for any vector $\mathbf{z}$ that meets $\mathbf{z}^H\mathbf{z} = E_s$. With a diagonal matrix $\mathbf{R}_w$, any $\mathbf{z}$ is a kind of optimal training sequence.

B. Complex QOSTC

The 4×4 and 8×8 QOSTC have been proposed in [26] and [27], respectively, where certain constellation rotation should be conducted to ensure the transmission with full-rate and full diversity. The matrix $\mathbf{B}$ for 4×4 QOSTC is given by

$$\mathbf{B} = \begin{bmatrix} s_1 & -s_2^* & -s_3^* & s_4 \\ s_2 & s_1^* & -s_4^* & -s_3 \\ s_3 & -s_4^* & s_1^* & -s_2 \\ s_4 & s_3^* & s_2^* & s_1 \end{bmatrix}. \tag{54}$$

Then

$$\mathbf{P}_1 = \mathbf{I}, \qquad \mathbf{P}_2 = \begin{bmatrix} 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 \end{bmatrix},$$

$$\mathbf{P}_3 = \begin{bmatrix} 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}, \quad \mathbf{P}_4 = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \\ 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix},$$

where relays $\mathbb{R}_2$ and $\mathbb{R}_3$ operate on $\mathbf{r}_2^*$ and $\mathbf{r}_3^*$ respectively. From (46), the optimal $\mathbf{z}$ should be found from

$$\sum_{k=1}^{K} \mathbf{z}_{(k)}^H \mathbf{P}_4 \mathbf{z}_{(k)} = 0, \tag{55a}$$

$$\sum_{k=1}^{K} \mathbf{z}_{(k)}^T \mathbf{P}_2^H \mathbf{P}_3 \mathbf{z}_{(k)}^* = 0. \tag{55b}$$

The solutions to (55) are not unique. Considering the PAPR problem, we can simply take $\mathbf{z} = \sqrt{P_s}\mathbf{1}$. Similar argument could be conducted for 8×8 QOSTC [27]. The detailed discussion is omitted here. Nevertheless, $\mathbf{z} = \sqrt{P_s}\mathbf{1}$ is still one of the optimal solutions.

V. SIMULATION RESULTS

In this section, we numerically study the performance of our proposed channel estimation algorithms as well as the optimal training designs under various scenarios. The channels g_i, h_i and the noise at the relays and destination are assumed as circularly symmetric complex Gaussian random variables with unit variances. The channel covariance matrices $\mathbf{R}_h$ and $\mathbf{R}_g$ have the following structures [19]:

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

where ε_1 and ε_2 are two real scalars. Since g_i is circularly symmetric, it is known that $\mathrm{E}\{g_i g_j\} = 0$. 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 $\mathbf{r}_i^*$. For convenience, we set $\varepsilon_1 = \varepsilon_2 = \varepsilon$ in all the examples. The signal to noise ratio is defined as 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 [14] running under MATLAB environment.
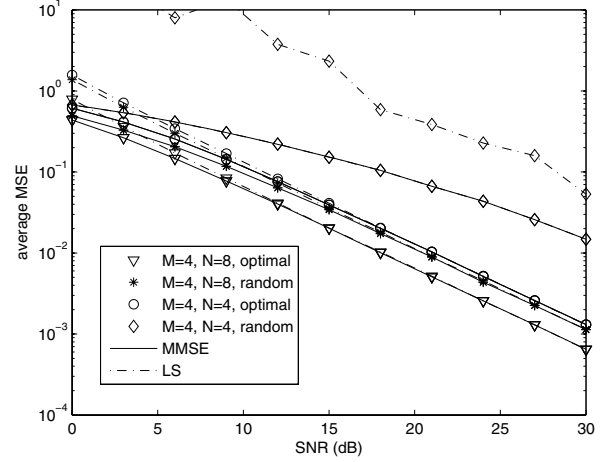

Fig. 2. Channel estimation MSEs versus SNR for both optimal training and random training.

A. Channel Estimation

The performance of channel estimation mean square error (MSE) is tested here. For all examples in this subsection, 1000 independent Monte-Carlo runs are performed for averaging.

1) Optimal Precoding vs Random Precoding. In Fig. 2, we display the MSEs versus SNR of both the LS estimator (9) and the MMSE estimator (21) for $M = 4$ relay nodes. The relay powers are chosen as $\{0.8P_s, P_s, 0.8P_s, P_s\}$, respectively. It is assumed that the QOSTC will be applied for the later data transmission so that $\mathbb{R}_2$ and $\mathbb{R}_3$ operate on the conjugate of their received signals. The parameter ε is 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. 2, 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 SISO or MISO channel estimation [19]. Moreover, the performance with optimal precoding is about 2 dB 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 2.5 dB. However, the channel estimation with random precoding degrades significantly. This clearly shows the importance of the optimal training design in our study.

2) Different Parameters M, N. We then provide three simulations figures considering three different simulation scenarios, e.g. same N different M in Fig. 3, same M different N in Fig. 4, different M and N with $M = N$ in Fig. 5. The ε is still set as 0.1. For each M, half of the relays have the power $0.8P_s$ and the other half have the power P_s. From Fig. 3, it is seen that the relay number affects the channel estimation even when the same power of source and same average power of relays are applied. This is a direct reflection from (17) and (43) where it says that the channel estimation MSE increases with M for each w_i. Note that, in the traditional MISO system, the channel estimation MSE is only related with the transmit power but is not related with the number of antennas. The reason for the
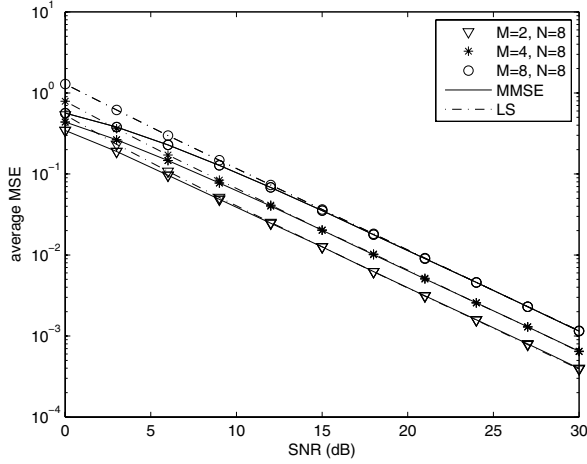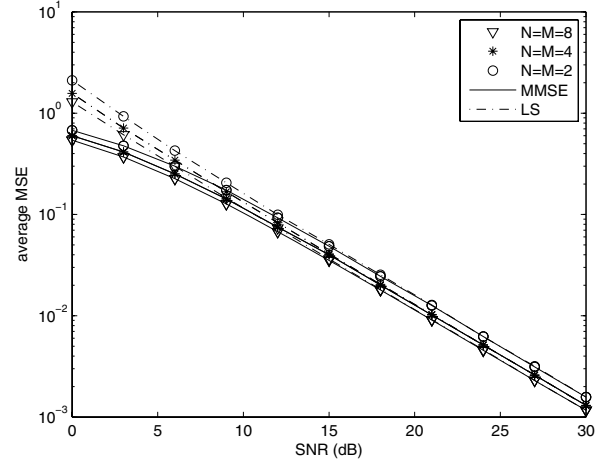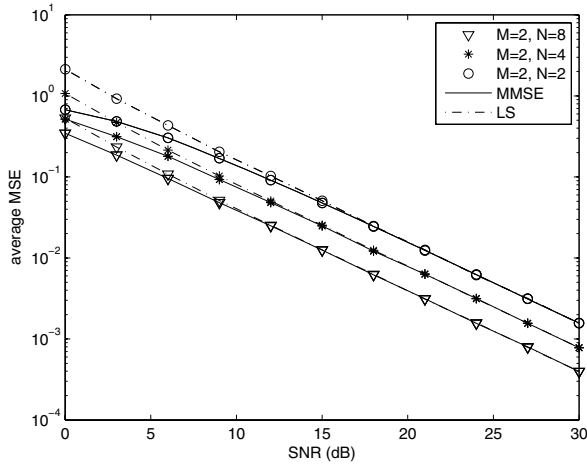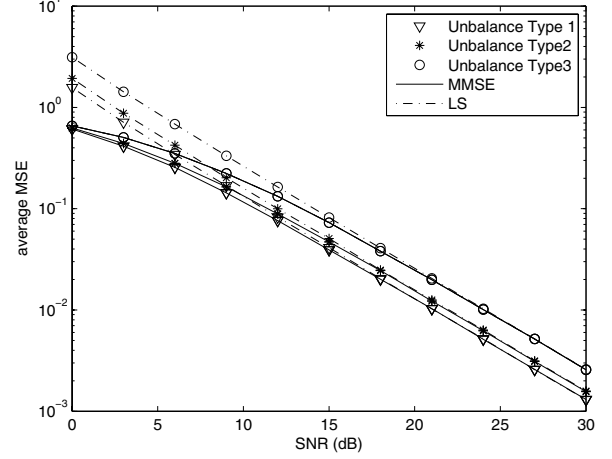
Fig. 3. Channel estimation MSEs versus SNR: different M, same N.

Fig. 5. Channel estimation MSEs versus SNR: different pairs of $N = M$.

Fig. 4. Channel estimation MSEs versus SNR: different N, same M.

Fig. 6. Channel estimation MSEs versus SNR: unbalanced relay power distribution.

degrading performance in AF relay networks is because the relays forward additional noise to the destination too. From Fig. 4, it is seen that the channel estimation MSE decreases when N increases. This is expected since the transmitting energy E_s, E_{ri} are linear functions of N. From Fig. 3 and Fig. 4, the performance of the channel estimation degrades when M increases while improves when N increases. Therefore, one may be curious to what happens when both N and M increase. The performances for different M, N but with $M = N$ are shown in Fig. 5, where it can be seen that lager $M = N$ gives better performance. This validates the simultaneous training for all relay channels instead of trivially training each relay channel individually.

3) Unbalanced Relay Power: In practical transmissions, it often happens that different relays have different powers according to their protocols, category, longevity, etc. We then characterize the effect of these unbalanced power distributions on the channel estimation. To be fair, we study the four relays case with three types of power distribution, e.g. $\{0.8P_s, 0.8P_s, P_s, P_s\}$, $\{0.4P_s, 0.8P_s, P_s, 1.4P_s\}$, $\{0.2P_s, 0.4P_s, P_s, 2P_s\}$ with each type possessing the same

amount of total power. The average channel estimation MSEs versus SNR are show in Fig. 6. It is seen from this figure that the unbalanced power distribution will deteriorate the accuracy of the channel estimation. Meanwhile, the gaps between the MMSE estimator and the LS estimator at lower SNR range are enlarged.

B. Performance of Space Time Coding

We apply the proposed channel estimation in Section IV here. The parameter ε is, thus, set as 0. The performance of the data transmission will be evaluated by the bit-error rate (BER). Each point in the figures is determined after 1000 bit errors are made.

We study two different cases where the number of relays are chosen as $M = 2$ and $M = 4$ respectively. The simulation results are shown in Fig. 7 and Fig. 8, respectively. For 2-relay networks, the Alamouti code is applied whereas for 4-relay networks, the QOSTC is applied. During the training period, the precoding matrix at each relay is constructed according to (53) such that the precoding scheme for both the training period and the data transmission are exactly the same. For each
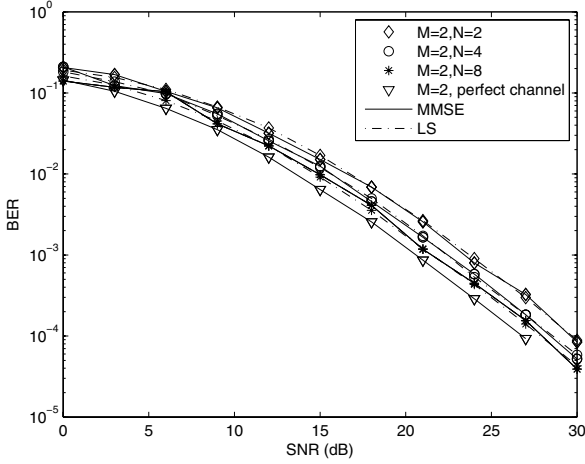
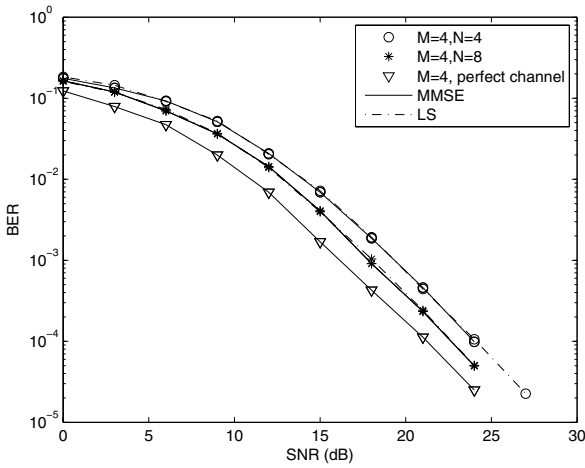Fig. 7.   BER performance for two relays with Alamouti STC applied.

Fig. 8.   BER performance for four relays with QOSTC applied.

scenario, we also choose different training length to show how the BER is affected by the channel estimation errors. To make a further comparison, the BER curve under perfect channel knowledge is also plotted for both scenarios. It can be seen that the length of the training affects the BER in a range of 0 to 3 dB, namely, the performance with minimum training length $N = M$ is about 3 dB worse than the performance from perfect channel knowledge. On the other hand, increasing the length N could not improve the BER over 3 dB. In addition, it is seen that 4-relay network possesses a higher diversity order than the 2-relay network.

VI. CONCLUSIONS

In this paper, we study 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 matrices for all relay nodes. The whole design process is efficiently conducted by dividing it into a convex optimization problem plus a matrix calculation problem. We also consider several practical AF

STCs and develop the optimal training that could simplify the relay design. Numerical examples have been provided to validate the proposed studies.

APPENDIX
PROOF FOR LEMMA 1

First of all, consider a particular case when $\mathbf{G} = \mathbf{NXM}$, where $\mathbf{N}$ is some constant matrix. Expand the (i, j)th element of $\mathbf{G}$ as $[\mathbf{G}]_{ij} = \sum_p \sum_q [\mathbf{N}]_{ip} [\mathbf{X}]_{pq} [\mathbf{M}]_{qj}$. Therefore

$$\frac{\partial [\mathbf{G}]_{ij}}{\partial [\mathbf{X}]_{pq}} = [\mathbf{N}]_{ip} [\mathbf{M}]_{qj}. \tag{56}$$

Applying the extended derivative chain rule [28] results in

$$\left[\frac{\partial \mathrm{tr}(\mathbf{F})}{\partial \mathbf{X}} \right]_{pq} = \frac{\partial \mathrm{tr}(\mathbf{F})}{\partial [\mathbf{X}]_{pq}} = \sum_i \sum_j \frac{\partial \mathrm{tr}(\mathbf{F})}{\partial [\mathbf{G}]_{ij}} \frac{\partial [\mathbf{G}]_{ij}}{\partial [\mathbf{X}]_{pq}}$$

$$= \sum_i \sum_j \frac{\partial \mathrm{tr}(\mathbf{F})}{\partial [\mathbf{G}]_{ij}} [\mathbf{N}]_{ip} [\mathbf{M}]_{qj} \tag{57}$$

$$= \left[\mathbf{M} \left[\frac{\partial \mathrm{tr}(\mathbf{F})}{\partial \mathbf{G}} \right]^T \mathbf{N} \right]_{qp}. \tag{58}$$

Equivalently, we derive that

$$\frac{\partial \mathrm{tr}(\mathbf{F})}{\partial \mathbf{X}} = \mathbf{N}^T \frac{\partial \mathrm{tr}(\mathbf{F})}{\partial \mathbf{G}} \mathbf{M}^T. \tag{59}$$

Substitute $\mathbf{N} = \mathbf{M}^H \mathbf{X}^H$ into (59) and bearing in the mind that $\partial \mathrm{tr}(\mathbf{M}^H \mathbf{X}^H)/\partial \mathbf{X} = \mathbf{0}$ yield the desired rule in (29).

REFERENCES

[1] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," *Eur. Trans. Telecommun.*, 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. J.*, 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. A. El Gamal, "Capacity theorems for the relay channel," *IEEE Trans. Inform. Theory*, vol. IT-25, pp. 572-584, Sept. 1979.

[6] R. U. Nabar, H. Bolcskei, and F. W. Kneubuhler, "Fading relay channels: performance limits and space time signal design," *IEEE J. Select. Areas Commun.*, vol. 22, pp. 1099-1109, Aug. 2004.

[7] J. Boyer, D. D. Falconer, and H. Yanikomeroglu, "'Multihop diversity in wireless relaying channels," *IEEE Trans. Commun.*, vol. 52, pp. 1820-1830, Oct. 2004.

[8] 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.

[9] J. N. Laneman, D. N. C. Tse, and G. W. Wornell, "Cooperative diversity in wireless networks: efficient protocols and outage behavior," *IEEE Trans. Inform. Theory*, vol. 50, pp. 3062-3080, Dec. 2004.

[10] A. Sendonaris, E. Erkip, and B. Aazhang, "User cooperation diversity—part I: system description," *IEEE Trans. Commun.*, vol. 51, pp. 1927-1938, Nov. 2003.

[11] ——, "User cooperation diversity–part II: system description," *IEEE Trans. Commun.*, vol. 51, pp. 1939-1948, Nov. 2003.

[12] S. Yiu, R. Schober, and L. Lampe, "Distributed space time block coding," *IEEE Trans. Commun.*, vol. 54, pp. 1195-1206, July 2006.

[13] Y. Jing and B. Hassibi, "Distributed space time coding in wireless relay networks," *IEEE Trans. Wireless Commun.*, vol. 5, pp. 3524-3536, Dec. 2006.

[14] J. F. Sturm, "Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones," *Optim. Meth. Softw.*, vol. 11-12, pp. 625-653, Aug. 1999.

[15] L. Vandenberghe and S. Boyd, "Semidefinite programming," *SIAM Rev.*, vol. 39, pp. 49-95, Mar. 1996.

[16] Y. Jing, and H. Jafarkhani, "Using orthogonal and quasi-orthogonal designs in wireless relay networks," in *Proc. GLOBECOM'06*, Nov. 2006.

[17] B. M. Hochwald, and T. L. Marzetta, "Unitary space time modulation for multiple antenna communication in Rayleigh flat fading," *IEEE Trans. Inform. Theory*, vol. 46, pp. 543-564, Mar. 2000.

[18] J.-J. Xiao and Z.-Q. Luo, "Universal decentralized estimation in an inhomogeneous sensing environment," *IEEE Trans. Inform. Theorey*, vol. 51, pp. 3564-3575, Oct. 2005.

[19] M. Biguesh, 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.

[20] S. M. Kay, *Fundumentals of Statistical Signal Processing: Estimation Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1993.

[21] 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.

[22] S. Haykin, *Adaptive Filter Theory*, 4th edition. Englewood Cliffs, NJ: Prentice-Hall, 1997.

[23] T. L. Saaty and J. Bram, *Nonlinear Mathematics*. New York: McGraw-Hill, 1964.

[24] K. J. Nielsen, *Methods in Numerical Analysis*. New York: Macmillan, 1957.

[25] V. Tarokh, H. Jafarkhani, and A. R. Calderbank, "Space time block codes from orthogonal designs," *IEEE Trans. Inform. Theory*, vol. 45, pp. 1456-1467, July 1999.

[26] H. Jafarkhani, "A quasi-rothogonal space time block code," *IEEE Trans. Commun.*, vo. 49, pp. 1-4, Jan. 2001.

[27] C. Yuen, Y. L. Guan, and T. T. Tjhung, "Full rate full diversity STBC with constellation rotation," in *Proc. VTC'03, Spring*, Jeju, Korea, vol. 1, pp. 296-300, 2003.

[28] G. A. Korn and T. M. Korn, *Mathematical Handbook for Scientists and Engineers*. Mineola, NY: Dover Publications, 2000.

Feifei Gao (S'05) received the B.Eng. degree from Xi'an Jiaotong University, Xi'an, Shaanxi China, in 2002, the M.Sc. degree from the McMaster University, Hamilton, ON, Canada in 2004, and the Ph.D degree from National University of Singapore in 2007. His research interests are in communication theory, broadband wireless communications, signal processing for communications, MIMO systems, and array signal processing.

Mr. Gao was a recipient of the president scholarship from the National University of Singapore. He has co-authored more than 30 refereed IEEE journal and conference papers and has served as a TPC member for IEEE ICC (2008), IEEE VTC (2008), and IEEE GLOBECOM (2008).

Tao Cui (S'04) received the M.Sc. degree in the Department of Electrical and Computer Engineering, University of Alberta, Edmonton, AB, Canada, in 2005, and the M.S. degree from the Department of Electrical Engineering, California Institute of Technology, Pasadena, USA, in 2006. He is currently working toward the Ph.D. degree at the Department of Electrical Engineering, California Institute of Technology, Pasadena. His research interests are in the interactions between networking theory, communication theory, and information theory.

Mr. Cui received the Best Paper Award at the IEEE International Conference on Mobile Ad-hoc and Sensor Systems (MASS) in 2007 and the Second Place in the ACM Student Research Competition at the 2007 Richard Tapia Celebration of Diversity in Computing Conference. Mr. Cui was a recipient of postgraduate scholarships from the Alberta Ingenuity Fund and the Alberta Informatics Circle of Research Excellence.

Arumugam Nallanathan (S'97−M'00−SM'05) received the B.Sc. with honors from the University of Peradeniya, Sri-Lanka, in 1991, the CPGS from the Cambridge University, United Kingdom, in 1994 and the Ph.D. from the University of Hong Kong, Hong Kong, in 2000, all in Electrical Engineering. He was an Assistant Professor in the Department of Electrical and Computer Engineering, National University of Singapore, Singapore from August 2000 to December 2007. Currently, he is a Senior Lecturer in the Department of Electronic Engineering at King's College London, United Kingdom. His research interests include OFDM systems, ultra-wide bandwidth (UWB) communication and localization, MIMO systems, and cooperative diversity techniques. In these areas, he has published over 120 journal and conference papers. He is a co-recipient of the Best Paper Award presented at 2007 IEEE International Conference on Ultra-Wideband (ICUWB'2007).

He currently serves on the Editorial Board of *IEEE Transactions on Wireless Communications*, *IEEE Transactions on Vehicular Technology*, John Wiley's *Wireless Communications and Mobile Computing* and *EURASIP Journal of Wireless Communications and Networking* as an Associate Editor. He served as a Guest Editor for *EURASIP Journal of Wireless Communications and Networking* Special issue on UWB Communication Systems- Technology and Applications. He also served as a technical program committee member for more than 25 IEEE international conferences. He currently serves as the General Track Chair for IEEE VTC'2008-Spring, Co-Chair for the IEEE GLOBECOM'2008 Signal Processing for Communications Symposium, and IEEE ICC'2009 Wireless Communications Symposium.