On the Design of Optimal Training Sequence for Bi-Directional Relay Networks

The-Hanh Pham, Ying-Chang Liang, *Senior Member, IEEE*, A. Nallanathan, *Senior Member, IEEE*, and Garg Hari Krishna, *Senior Member, IEEE*

***Abstract*—In this letter, the problem of training sequence design in a bi-directional relay network is investigated. The two nodes, Tx1 and Tx2, communicate with each other via a relay (RL). The RL receives the signals that are transmitted simultaneously from both nodes and then broadcasts the superposition back to the two nodes. Single carrier cyclic prefix (SCCP) is deployed in the transmission from the two nodes and the RL as well for combating intersymbol interference (ISI). We are interested in the channel estimation problem in this scenario. We propose a design of training sequences from two nodes to minimize the mean-square error (MSE) of the channel estimation according to zero forcing (ZF) criterion. We do not separate the links between each node to RL but incorporate them into the composite channels. Simulation results show that our design performance suffers only 1-dB loss in terms of bit-error rate (BER) as compared to the case with perfect channel information.**

***Index Terms*—Bi-directional relay networks, channel estimation, training signal design, Zadoff–Chu sequences.**

I. Introduction

Many researchers have paid attention to communications based on relays recently. In this letter, a two-hop relaying network is investigated where two nodes, Tx1 and Tx2, exchange information based on the assistance of a single relay (RL). Conventionally, if time division multiplexing (TDD) is deployed, four time slots are required to accomplish the information exchange between Tx1 and Tx2. However, in [1]–[3], bi-directional relaying has been investigated. For this kind of relaying, the two nodes, Tx1 and Tx2, transmit their own data simultaneously at the first time slot and RL receives the superposition of both information. At the second time slot, RL broadcasts the scaled version of this superposition. The information of one node can be detected at the other node when the latter subtracts its own information from the signal received from RL. It is obvious that bi-directional relaying requires half of the number of the time slots compared with the conventional one. In [4], the authors design the linear spatial filters at RL of a bi-directional relaying system working under flat-fading channels to separate the transmitted signals from the two nodes. This method requires the number of antennas at RL to be at least more than the summation of the number of antennas at two nodes. Besides, RL has to estimate the channels from the two nodes to itself. This approach demands a lot of resources at RL.

In a frequency-selective fading environment where intersymbol inteference (ISI) appears, single carrier cyclic prefix (SCCP) [5] proves to be efficient in suppressing ISI and enjoys a low peak-to-average power ratio (PAPR). SCCP has been selected for IEEE 802.16 [6]. Hence, SCCP is deployed at two nodes as well as at RL in the bi-directional relaying scenario. In the detection process, each node not only needs to know the channel information from RL to itself (for subtraction purpose) but also the channel information from the other node to RL. Hence, the channel estimation problem becomes complicated at the nodes.

In this letter, we design the training sequences for the nodes Tx1 and Tx2 to estimate the necessary channel information for detection. We do not separately estimate the two links from Tx1 to RL and Tx2 to RL. Here, due to the nature of the signaling, we estimate two composite channels. The first one is the overall channel from one node to RL and back from RL to itself. The other is the composite channel consisting of the channels from the remaining node to RL and RL to the interested node. The zero forcing (ZF) criterion is applied to estimate those composite channels. Based on the mean-square error (MSE) criterion, a design method for training sequences of the two nodes is proposed. Simulation results are also provided to illustrate our proposed method.

The structure of the rest of the letter is as follows. The system model is given in Section II. In Section III, the least-square estimation is applied in our system. In this section, we state the requirements on the training sequences from the two nodes. The training design is given in Section IV. Finally, the simulation results are provided in Section V and conclusions are drawn in Section VI.

Notation: $\mathbf{0}_{m,n}$ is $m \times n$ zero matrix and $n \times n$ zero matrix is denoted by $\mathbf{0}_n$. The identity matrix of size n is denoted by $\mathbf{I}_n$. $[\mathbf{A}]_{m,n}$ is the (m,n)th element of $\mathbf{A}$. $|a|$ is the absolute value of a number a; meanwhile, $|\mathbf{A}|$ is a matrix whose elements are absolute values of corresponding elements of $\mathbf{A}$. $\boldsymbol{\mathcal{D}}(\mathbf{a})$ is a diagonal matrix whose diagonal entries are from vector $\mathbf{a}$. Transpose, Hermitian transpose of a vector/matrix are denoted by $(\cdot)^{\mathcal{T}}$ and $(\cdot)^{\mathcal{H}}$. $\mathrm{tr}\{\boldsymbol{A}\}$ stands for the trace operation. $\odot$ is the Hadamard product. $\|\cdot\|$ is the Frobenius norm. The convolution between vectors $\mathbf{a}$ and $\mathbf{b}$ is denoted as $\mathbf{a} * \mathbf{b}$.

II. System Model

Consider a relay network where the two nodes, Tx1 and Tx2, exchange information relying on the help of an RL. The RL and two nodes are assumed to have one antenna each. Let $\mathbf{h}_1 = [h_1(0) \cdots h_1(L-1)]^{\mathcal{T}}$ and $\mathbf{h}_2 = [h_2(0) \cdots h_2(L-1)]^{\mathcal{T}}$ be the impulse responses of the frequency-selective fading channels

Manuscript received April 24, 2008; revised September 19, 2008. Current version published February 11, 2009. This work was supported by the National University of Singapore under Grant R-263-000-421-112. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Zhengdao Wang.

T.-H. Pham and G. H. Krishna are with the Department of Electrical and Computer Engineering, National University of Singapore, Singapore 119260 (e-mail: elepth@nus.edu.sg; eleghk@nus.edu.sg).

Y.-C. Liang is with the Institute for Infocomm Research ($\mathrm{I}^2\mathrm{R}$), Singapore 138632 (e-mail: ycliang@i2r.a-star.edu.sg).

A. Nallanathan is with the Division of Engineering, King's College London, Strand, London WC2R 2LS, U.K. (e-mail: nallanathan@ieee.org).

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

Digital Object Identifier 10.1109/LSP.2008.2008949

between Tx1 and Tx2 to RL, respectively. Each element of $\mathbf{h}_1$ or $\mathbf{h}_2$ is modeled as a zero-mean complex Gaussian random variable with variance of $\sigma_{i,l}^2$ for $i = 1, 2$ and $l = 0, \ldots, L-1$. These elements are also assumed to be independent. We use $\mathbf{s}_1 = [s_1(0) \cdots s_1(N-1)]^{\mathcal{T}}$ to denote the training signal vector transmitted from Tx1. The counterpart from Tx2 is defined as $\mathbf{s}_2 = [s_2(0) \cdots s_2(N-1)]^{\mathcal{T}}$. The power constraints of the transmission is $E\{\mathbf{s}_1^{\mathcal{H}}\mathbf{s}_1\} = E\{\mathbf{s}_2^{\mathcal{H}}\mathbf{s}_2\} = NP$, where P is the average transmitting power of Tx1 and Tx2. The transmission is divided into two phases. At the first phase, a cyclic prefix (CP) is added to the top of each signal vector. The length of the CP, L_{CP}, is greater than or equal to the channel memory to avoid the inter-block interference (IBI), $L_{\mathrm{CP}} \geq L-1$. These new signal vectors are transmitted simultaneously from Tx1 and Tx2 to RL. At RL, the received signals associated with the CP portion are discarded first. Then the remaining received signals are scaled by a real factor α to keep the average power of RL to be P_{r}. The resultant signals are CP-added and transmitted to Tx1 and Tx2 during Phase 2.

The received signal vector at RL after Phase 1 is

$$\mathbf{y} = \mathbf{H}_1\mathbf{s}_1 + \mathbf{H}_2\mathbf{s}_2 + \mathbf{n}_r \tag{1}$$

where $\mathbf{H}_1$ and $\mathbf{H}_2$ are two $N \times N$ circulant matrices with $[h_1(0) \cdots h_1(L-1)\ \mathbf{0}_{1\times(N-L)}]^{\mathcal{T}}$ and $[h_2(0) \cdots h_2(L-1)\ \mathbf{0}_{1\times(N-L)}]^{\mathcal{T}}$ as their first columns, respectively; $\mathbf{n}_r$ is a realization of a complex Gaussian random vector with zero mean and covariance matrix $E\{\mathbf{n}_r\mathbf{n}_r^{\mathcal{H}}\} = N_0\mathbf{I}_N$. The received signal vector $\mathbf{y}$ in (1) is then amplified by a real coefficient α which is given by

$$\alpha = \sqrt{\frac{P_{\mathrm{r}}}{\left(\sum_{l=0}^{L-1}\sigma_{1,l}^2 + \sum_{l=0}^{L-1}\sigma_{2,l}^2\right)P + N_0}}. \tag{2}$$

The last L_{CP} components of the vector $\alpha\mathbf{y}$ are appended to the top of itself and the resultant vector is transmitted from RL to both ends, Tx1 and Tx2. Without loss of generality, we only consider the channel estimation problem at Tx1. A similar operation can be applied at Tx2.

The received signal vector at Tx1 after CP removal can be written as follows:

$$\mathbf{r} = \alpha\mathbf{H}_1\mathbf{y} + \mathbf{n}_1 = \alpha\mathbf{H}_1\mathbf{H}_1\mathbf{s}_1 + \alpha\mathbf{H}_1\mathbf{H}_2\mathbf{s}_2 + \mathbf{n}' \tag{3}$$

where $\mathbf{n}' = \alpha\mathbf{H}_1\mathbf{n}_r + \mathbf{n}_1$, $\mathbf{n}_1$ is a realization of a complex Gaussian random vector with zero mean and covariance matrix $E\{\mathbf{n}_1\mathbf{n}_1^{\mathcal{H}}\} = N_0\mathbf{I}_N$. According to matrix theory, the circulant matrices $\mathbf{H}_1$ and $\mathbf{H}_2$ can be decomposed as $\mathbf{H}_i = \mathbf{W}^{\mathcal{H}}\mathbf{\Lambda}_i\mathbf{W}$, $i = 1, 2$, where $\mathbf{\Lambda}_i = \mathrm{diag}\{H_{i,0}, \cdots, H_{i,k}, \cdots, H_{i,N-1}\} \in \mathbb{C}^{N\times N}$, $H_{i,k} = \sum_{l=0}^{L-1} h_i(l)e^{-j(2\pi kl/N)}$, $k = 0, \ldots, N-1$, and $\mathbf{W}$ is the (unitary) discrete Fourier transform matrix with $[\mathbf{W}]_{m,n} = (1/\sqrt{N})e^{-j(2\pi mn/N)}$, $m, n = 0, \ldots, N-1$. Therefore, (3) can be written as

$$\mathbf{r} = \mathbf{W}^{\mathcal{H}}\alpha\mathbf{\Lambda}_1\mathbf{\Lambda}_1\mathbf{W}\mathbf{s}_1 + \mathbf{W}^{\mathcal{H}}\alpha\mathbf{\Lambda}_1\mathbf{\Lambda}_2\mathbf{W}\mathbf{s}_2 + \mathbf{n}'. \tag{4}$$

Based on the DFT theory, with a condition of $(2L-1) \leq N$, we observe that $\mathbf{W}^{\mathcal{H}}\alpha\mathbf{\Lambda}_1\mathbf{\Lambda}_1\mathbf{W}$ is the decomposition of a circulant matrix which is constructed from a composite impulse response $\mathbf{h} \triangleq \alpha(\mathbf{h}_1 * \mathbf{h}_1)$. Similarly, $\mathbf{W}^{\mathcal{H}}\alpha\mathbf{\Lambda}_1\mathbf{\Lambda}_2\mathbf{W}$ stems from a circulant matrix which is constructed from a composite impulse response $\mathbf{g} \triangleq \alpha(\mathbf{h}_1 * \mathbf{h}_2)$. The length of $\mathbf{h}$ or $\mathbf{g}$ is $(2L-1)$, and we incorporate the scaling factor α into our composite channels, $\mathbf{h}$ and $\mathbf{g}$. If (4) is premultiplied by $\mathbf{W}$, we obtain

$$\mathbf{z} = \boldsymbol{\mathcal{D}}(\mathbf{W}\mathbf{s}_1)\mathbf{F}\mathbf{h} + \boldsymbol{\mathcal{D}}(\mathbf{W}\mathbf{s}_2)\mathbf{F}\mathbf{g} + \mathbf{n} \tag{5}$$

where $\mathbf{n} = \alpha\mathbf{\Lambda}_1\mathbf{W}\mathbf{n}_r + \mathbf{W}\mathbf{n}_1$ and $\mathbf{F}$ is the first $(2L-1)$ columns of $\sqrt{N}\mathbf{W}$. The noise term $\mathbf{n}$, therefore, is a realization of a complex Gaussian random vector with zero mean and covariance matrix of $E\{\mathbf{n}\mathbf{n}^{\mathcal{H}}\} = N_0(\alpha^2|\mathbf{\Lambda}_1|^2 + \mathbf{I}_N)$. Equation (5) can be written in another form as

$$\mathbf{z} = [\boldsymbol{\mathcal{D}}(\mathbf{W}\mathbf{s}_1)\mathbf{F} \quad \boldsymbol{\mathcal{D}}(\mathbf{W}\mathbf{s}_2)\mathbf{F}]\begin{bmatrix}\mathbf{h}\\ \mathbf{g}\end{bmatrix} + \mathbf{n} = \mathbf{S}\mathbf{k} + \mathbf{n} \tag{6}$$

where $\mathbf{S} = [\boldsymbol{\mathcal{D}}(\mathbf{W}\mathbf{s}_1)\mathbf{F}\ \boldsymbol{\mathcal{D}}(\mathbf{W}\mathbf{s}_2)\mathbf{F}] \in \mathbb{C}^{N\times 2(2L-1)}$ and $\mathbf{k} = [\mathbf{h}^{\mathcal{T}}\ \mathbf{g}^{\mathcal{T}}]^{\mathcal{T}} \in \mathbb{C}^{2(2L-1)\times 1}$.

III. Least-Square Channel Estimation

From (6), the least-square estimation of the composite channel $\mathbf{k}$ is given by

$$\hat{\mathbf{k}} \triangleq \mathbf{S}^{\dagger}\mathbf{z} = \mathbf{k} + \mathbf{S}^{\dagger}\mathbf{n} \tag{7}$$

where $\mathbf{S}^{\dagger} = \left(\mathbf{S}^{\mathcal{H}}\mathbf{S}\right)^{-1}\mathbf{S}^{\mathcal{H}}$ is the Moore–Penrose pseudoinverse of $\mathbf{S}$. The MSE of the estimation is defined as

$$\mathrm{MSE} \triangleq \frac{1}{2(2L-1)}E\left\{(\hat{\mathbf{k}} - \mathbf{k})^{\mathcal{H}}(\hat{\mathbf{k}} - \mathbf{k})\right\}. \tag{8}$$

We now consider two cases: the first case is that Tx1 can access the statistics of channel $\mathbf{h}_1$, and the other is that Tx1 cannot do so.

- Case 1) In this case, (8) can be written as

$$\mathrm{MSE} = \frac{N_0\alpha^2}{2(2L-1)}\mathrm{tr}\left\{\mathbf{S}^{\dagger}E\left\{|\mathbf{\Lambda}_1|^2\right\}(\mathbf{S}^{\dagger})^{\mathcal{H}}\right\} + \frac{N_0}{2(2L-1)}\mathrm{tr}\left\{(\mathbf{S}^{\mathcal{H}}\mathbf{S})^{-1}\right\}. \tag{9}$$

Since $H_{1,k} = \sum_{l=0}^{L-1} h_1(l)e^{-j(2\pi kl/N)}$, we have

$$E\left\{|H_{1,k}|^2\right\} = \sum_{l_1=0}^{L-1}\sum_{l_2=0}^{L-1} E\left\{h_1(l_1)h_1^*(l_2)\right\}e^{-j\frac{2\pi k(l_1-l_2)}{N}} = \sum_{l=0}^{L-1}\sigma_{1,l}^2. \tag{10}$$

Therefore, $E\{|\mathbf{\Lambda}_1|^2\} = (\sum_{L=0}^{L-1}\sigma_{1,l}^2)\mathbf{I}_N$ and (9) can be written as

$$\mathrm{MSE} = \frac{\left(\alpha^2\left(\sum_{l=0}^{L-1}\sigma_{1,l}^2\right) + 1\right)N_0}{2(2L-1)}\mathrm{tr}\left\{(\mathbf{S}^{\mathcal{H}}\mathbf{S})^{-1}\right\}. \tag{11}$$

It is clear from (11) that minimizing the MSE is equivalent to minimizing $Q \triangleq \mathrm{tr}\{(\mathbf{S}^{\mathcal{H}}\mathbf{S})^{-1}\}$.

- Case 2) For this case, (8) becomes

$$\mathrm{MSE} = \frac{N_0}{2(2L-1)}\left[\alpha^2\mathrm{tr}\left\{\mathbf{S}^{\dagger}|\mathbf{\Lambda}_1|^2(\mathbf{S}^{\dagger})^{\mathcal{H}}\right\} + \mathrm{tr}\left\{(\mathbf{S}^{\mathcal{H}}\mathbf{S})^{-1}\right\}\right]. \tag{12}$$

To minimize the MSE in this case we have to minimize two terms on the right-hand side (RHS) of (12). It is easy to see that in order to minimize the second term, $J \triangleq (N_0/2(2L-1))\mathrm{tr}\{(\mathbf{S}^{\mathcal{H}}\mathbf{S})^{-1}\}$, we must minimize the quantity Q defined in Case 1. Let $V \triangleq \mathrm{tr}\{\mathbf{S}^{\dagger}|\mathbf{\Lambda}_1|^2(\mathbf{S}^{\dagger})^{\mathcal{H}}\} = \mathrm{tr}\{(\mathbf{S}^{\dagger})^{\mathcal{H}}\mathbf{S}^{\dagger}|\mathbf{\Lambda}_1|^2\}$. In the Appendix, we prove that $(\mathbf{S}^{\dagger})^{\mathcal{H}}\mathbf{S}^{\dagger}$ is a Hermitian positive definite matrix. $|\mathbf{\Lambda}_1|^2$, clearly, is a Hermitian positive definite matrix, too. Hence, from [7], we have

$$\begin{aligned} V &\le \mathrm{tr}\left\{|\mathbf{\Lambda}_1|^2\right\}\mathrm{tr}\left\{(\mathbf{S}^{\dagger})^{\mathcal{H}}\mathbf{S}^{\dagger}\right\} = \mathrm{tr}\left\{|\mathbf{\Lambda}_1|^2\right\}\mathrm{tr}\left\{\mathbf{S}^{\dagger}(\mathbf{S}^{\dagger})^{\mathcal{H}}\right\} \\ &= \mathrm{tr}\left\{|\mathbf{\Lambda}_1|^2\right\}\mathrm{tr}\left\{(\mathbf{S}^{\mathcal{H}}\mathbf{S})^{-1}\right\}. \end{aligned} \tag{13}$$

From (13), to minimize the first term on the RHS of (12), we turn to minimize the upper limit of V. In other words, we have to minimize the quantity Q.

From the above, we design the training signal vectors $\mathbf{s}_1$ and $\mathbf{s}_2$ such that Q is minimized. The design must satisfy the following theorem.

Theorem 1: To achieve minimum MSE, the training signal vectors $\mathbf{s}_1$ and $\mathbf{s}_2$ must be designed to meet the following conditions:

- $\mathbf{C} \triangleq \mathbf{F}^{\mathcal{H}}\boldsymbol{\mathcal{D}}^{\mathcal{H}}(\mathbf{W}\mathbf{s}_1)\boldsymbol{\mathcal{D}}(\mathbf{W}\mathbf{s}_2)\mathbf{F} = \mathbf{0}_{2L-1}$;
- $\mathbf{B} \triangleq \mathbf{F}^{\mathcal{H}}\boldsymbol{\mathcal{D}}^{\mathcal{H}}(\mathbf{W}\mathbf{s}_1)\boldsymbol{\mathcal{D}}(\mathbf{W}\mathbf{s}_1)\mathbf{F}$ and $\mathbf{D} \triangleq \mathbf{F}^{\mathcal{H}}\boldsymbol{\mathcal{D}}^{\mathcal{H}}(\mathbf{W}\mathbf{s}_2)\boldsymbol{\mathcal{D}}(\mathbf{W}\mathbf{s}_2)\mathbf{F}$ are two diagonal matrices with equal diagonal elements.

Proof: Let

$$\begin{aligned} \mathbf{A}^{-1} &\triangleq (\mathbf{S}^{\mathcal{H}}\mathbf{S})^{-1} = \begin{bmatrix} \mathbf{B} & \mathbf{C} \\ \mathbf{C}^{\mathcal{H}} & \mathbf{D} \end{bmatrix}^{-1} \\ &= \begin{bmatrix} \mathbf{F}^{\mathcal{H}}\boldsymbol{\mathcal{D}}^{\mathcal{H}}(\mathbf{W}\mathbf{s}_1)\boldsymbol{\mathcal{D}}(\mathbf{W}\mathbf{s}_1)\mathbf{F} & \mathbf{F}^{\mathcal{H}}\boldsymbol{\mathcal{D}}^{\mathcal{H}}(\mathbf{W}\mathbf{s}_1)\boldsymbol{\mathcal{D}}(\mathbf{W}\mathbf{s}_2)\mathbf{F} \\ \mathbf{F}^{\mathcal{H}}\boldsymbol{\mathcal{D}}^{\mathcal{H}}(\mathbf{W}\mathbf{s}_2)\boldsymbol{\mathcal{D}}(\mathbf{W}\mathbf{s}_1)\mathbf{F} & \mathbf{F}^{\mathcal{H}}\boldsymbol{\mathcal{D}}^{\mathcal{H}}(\mathbf{W}\mathbf{s}_2)\boldsymbol{\mathcal{D}}(\mathbf{W}\mathbf{s}_2)\mathbf{F} \end{bmatrix}^{-1}. \end{aligned} \tag{14}$$

Applying an inequality[1] in [8], we have

$$Q \ge \sum_{k=0}^{2(2L-1)-1} \left([\mathbf{A}]_{k,k}\right)^{-1} \tag{15}$$

where equality holds if and only if $\mathbf{A}$ is diagonal. This means that matrix $\mathbf{C}$ must be a zero matrix. If we apply the Cauchy–Schwartz inequality on the RHS of (15), we further obtain

$$Q \ge 2(2L-1)\sqrt[2(2L-1)]{\prod_{k=0}^{2(2L-1)-1}\left([\mathbf{A}]_{k,k}\right)^{-1}} \tag{16}$$

where equality holds if and only if $[\mathbf{A}]_{k,k}$'s are equal. ■

From (15) and (16), the minimum Q is

$$Q_{\min} = \frac{2(2L-1)}{[\mathbf{A}]_{k,k}} \tag{17}$$

for an arbitrary k belonging to the interval $[0,\ldots,2(2L-1)-1]$.

IV. Training Sequence Design

To reduce PAPR, the time-domain signals should have constant magnitude. In [9], a special sequence, called Zadoff–Chu sequence, is proposed. All elements of this sequence have the same magnitude in both time and frequency domain. The general form of a Zadoff–Chu sequence of length M is given as follows:

$$m(k) = \begin{cases} e^{-j\frac{\pi G n(n+2g)}{M}}, & n = 0,\cdots,M-1; M \text{ is even} \\ e^{-j\frac{\pi G n(n+1+2g)}{M}}, & n = 0,\cdots,M-1; M \text{ is odd} \end{cases} \tag{18}$$

where g is an integer and G is an integer relatively prime to M. Based on Zadoff–Chu sequence, we have the following theorem about choosing $\mathbf{s}_1$ and $\mathbf{s}_2$ to satisfy Theorem 1.

Theorem 2: The optimal training signal vectors $\mathbf{s}_1$ and $\mathbf{s}_2$ can be selected from two distinct vectors $\mathbf{n}_v$ and $\mathbf{n}_l$ belonging to the set $\{\mathbf{n}_i\}_{i=0}^{U-1}$ in which $\mathbf{n}_i = \mathbf{n}_0 \odot \mathbf{e}_i$, where $\mathbf{n}_0 = [\underbrace{\mathbf{m}^{\mathcal{T}}\ \mathbf{m}^{\mathcal{T}}\ \cdots\ \mathbf{m}^{\mathcal{T}}}_{U \text{ times}}]^{\mathcal{T}}$, $\mathbf{e}_i = [1\ e^{j(2\pi i/N)}\ \ldots\ e^{j(2\pi(N-1)i/N)}]^{\mathcal{T}}$, $i = 1,\ldots,U-1$, and $\mathbf{m}$ is a Zadoff–Chu sequence of length N/U (N/U and U are integer numbers) with $\|\mathbf{m}\|^2 = (N/U)P$. Here, the L satisfies the condition $L < ((N+2U)/2U)$.

Proof: Denote $\mathbf{m}_f = [m_f(0)\, m_f(1)\, \cdots\, m_f(N/U-1)]^{\mathcal{T}}$ as the corresponding frequency transform of $\mathbf{m}$. We have, from properties of a Zadoff–Chu sequence, $|m_f(0)|^2 = |m_f(1)|^2 = \cdots = |m_f(N/U-1)|^2 = P$. By repeating $\mathbf{m}$ U times to construct $\mathbf{n}_0$ and then element-wise multiplying $\mathbf{n}_0$ with $\mathbf{e}_i$ to obtain $\mathbf{n}_i$, the frequency transform of $\mathbf{n}_i$, $\mathbf{n}_{i,f} = [n_{i,f}(0)\, n_{i,f}(1)\, \cdots\, n_{i,f}(N-1)]^{\mathcal{T}}$, can be determined by

$$n_{i,f}(k) = \begin{cases} \sqrt{U} m_f\left((k-i)/U\right), & \text{if } k \in \mathcal{I}_i \\ 0, & \text{elsewhere} \end{cases} \tag{19}$$

where $\mathcal{I}_i = \{i, i+U,\cdots,i+(N/U-1)U\}$, $i = 0,\ldots,U-1$.

The training signal vectors $\mathbf{s}_1$ and $\mathbf{s}_2$ are chosen from two distinct vectors $\mathbf{n}_v$ and $\mathbf{n}_l$, $v \neq l$, so we have the following.

- Because of $v \neq l$, hence $\mathcal{I}_v \cap \mathcal{I}_l = \emptyset$. In other words, $\boldsymbol{\mathcal{D}}^{\mathcal{H}}(\mathbf{W}\mathbf{s}_1)\boldsymbol{\mathcal{D}}(\mathbf{W}\mathbf{s}_2) = \mathbf{0}_N$; therefore, $\mathbf{C} = \mathbf{0}_{2L-1}$.
- We have $\boldsymbol{\mathcal{D}}^{\mathcal{H}}(\mathbf{W}\mathbf{s}_1)\boldsymbol{\mathcal{D}}(\mathbf{W}\mathbf{s}_1)$ is a diagonal matrix whose the kth diagonal component is UP if $k \in \mathcal{I}_v$ and is 0 if $k \notin \mathcal{I}_v$. Hence, $[\mathbf{B}]_{p,q}$ for $p,q = 0,\ldots,2L-2$ is determined as follows:

$$\begin{aligned} [\mathbf{B}]_{p,q} &= UP\sum_{k\in\mathcal{I}_v} e^{j\frac{2\pi kp}{N}} e^{-j\frac{2\pi kq}{N}} = UP\sum_{k\in\mathcal{I}_v} e^{j\frac{2\pi k(p-q)}{N}} \\ &= \begin{cases} PN, & \text{if } p = q \\ UPe^{j\frac{2\pi v(p-q)}{N}} \cdot \frac{e^{j2\pi(p-q)}-1}{e^{j\frac{2\pi U(p-q)}{N}}-1}, & \text{if } p \neq q. \end{cases} \end{aligned} \tag{20}$$

It is easy to see that, for $p \neq q$, $(p-q)$ belongs to the interval of $\{-2L+2,\cdots,-1,1,\cdots,2L-2\}$; in other words, $0 < |p-q| \le 2L-2$. To make the quantity $(e^{j(2\pi U(p-q)/N)}-1)$ not to be 0 (hence, $[\mathbf{B}]_{p,q} = 0$ for $p \neq q$), we have to restrict L as follows:

$$2L-2 < \frac{N}{U} \Longleftrightarrow L < \frac{N+2U}{2U}. \tag{21}$$

With the condition of (21), we have $[\mathbf{B}]_{p,q} = PN$ if $p = q$ and 0 if $p \neq q$. Similarly, we also have $[\mathbf{D}]_{p,q} = PN$ if $p = q$ and 0 if $p \neq q$.

■

It is easy to see that $\mathbf{A}$ is a diagonal matrix with equal diagonal elements of PN. Hence, in Case 1, we have the minimum MSE as $\mathrm{MSE}_{\min} = [((\alpha^2(\sum_{l=0}^{L-1}\sigma_{1,l}^2) +$

[1]For a $n \times n$ positive definite matrix $\mathbf{A}$, we have $\mathrm{tr}\{\mathbf{A}^{-1}\} \ge \sum_{i=1}^{n}([\mathbf{A}]_{i,i})^{-1}$ and the equality holds if and only if $\mathbf{A}$ is diagonal.

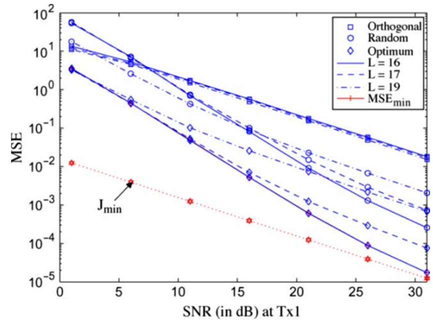

Fig. 1. MSE performance for different channel lengths, $U = 2$.

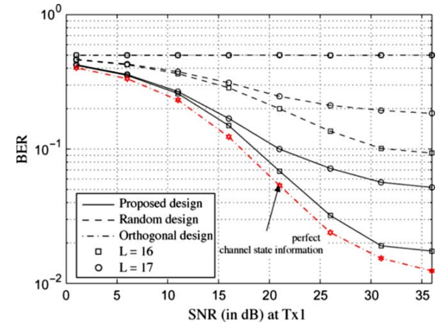

Fig. 2. BER performance for different channel lengths, $U = 2$.

$1)N_0)/(2(2L-1))]Q_{\min} = ((\alpha^2(\sum_{l=0}^{L-1}\sigma_{1,l}^2)+1)N_0)/(PN)$. Similarly, in Case 2, we have the minimum J as $J_{\min} = (N_0/(2(2L-1)))Q_{\min} = N_0/(PN)$.

From (21), it is observed that for a given N, in order to maximize the channel length that the system can handle, we have to minimize U as much as possible. For our system where we only have to design $\mathbf{s}_1$ and $\mathbf{s}_2$, we can choose $U = 2$; therefore, $L < (N/4 + 1)$.

V. Simulation Results

In this section, we provide simulation results for the bi-directional relaying system having one antenna at the two nodes, Tx1 and Tx2, as well as at RL. The size of training vectors, $\mathbf{s}_1$ and $\mathbf{s}_2$, is $N = 64$. The channel between Tx1 (Tx2) to RL is modeled as a frequency-selective fading channel characterized by a channel response of length L. We further assume that the power delay profile of each channel is uniform, i.e., each tap of $\mathbf{h}_1$ or $\mathbf{h}_2$ is modeled as a zero-mean Gaussian random variable with variance $E\{|h_i(l)|^2\} = 1/L$ for $i = 1, 2$ and $l = 0, \ldots, L-1$. The channels are assumed to be static over 20 time slots. The first two time slots are used to estimate the channel information and the remaining are used for data transmission. We consider both MSE performance as well as bit-error rate (BER) performance. For data transmission, QPSK signaling is deployed. We choose $P = 1$ and define the signal-to-noise ratio (SNR) at Tx1 as P/N_0 and the SNR at RL as P_{r}/N_0 which equals 30 dB in all simulations.

Fig. 1 shows the MSE between the estimated $\hat{\mathbf{k}}$ and $\mathbf{k}$ for different values of channel length L when $U = 2$. From (21), if $N = 64$ and $U = 2$, the maximum channel length that can be supported is $L_{\max} = 16$. From Fig. 1, it is observed that with $L = 17$, the MSE performance becomes worse compared with that of $L = 16$. In Fig. 1, we also present the MSE performances of the random training sequences whose elements are randomly chosen from QPSK signaling and the orthogonal design whose training sequences are distinct columns of Hadamard matrix of size N. The three kinds of training signal vectors have the same power. The MSE performances obtained using random sequences or orthogonal sequences are worse compared with those of our proposed design.

BER performance for the above system set-up is illustrated in Fig. 2 for $U = 2$. The detection process of the transmitted signals belonging to Tx2 at Tx1 is performed as follows: First of all, Tx1 cancels its contribution in the signal received from RL; then, a ZF frequency domain equalizer is applied to reconstruct the signal of Tx2. Fig. 2 shows that the BER obtained using our design can be around 1 dB away from that of perfect channel information at Tx1 over the nonsaturation region. Compared with the random and orthogonal design, our proposed design performs superiorly.

VI. Conclusions

In this letter, the channel estimation issue in bi-directional relay networks is investigated. More specifically, relying on the Zadoff–Chu sequence, we design the training sequences from nodes Tx1 and Tx2 to minimize the MSE of channel estimation. Simulation results show that BER performance using the channel information estimated from our design suffers only 1-dB loss as compared to that with perfect channel information.

Appendix

We have $(\mathbf{S}^{\dagger})^{\mathcal{H}}\mathbf{S}^{\dagger} = \mathbf{S}(\mathbf{S}^{\mathcal{H}}\mathbf{S})^{-1}(\mathbf{S}^{\mathcal{H}}\mathbf{S})^{-1}\mathbf{S}^{\mathcal{H}}$. Let $\mathbf{G} \triangleq (\mathbf{S}^{\mathcal{H}}\mathbf{S})^{-1}$. Note that $\mathbf{G}^{\mathcal{H}} = \mathbf{G}$. Hence, for any nonzero vector $\mathbf{x}$ and $\mathbf{v} \triangleq \mathbf{S}^{\mathcal{H}}\mathbf{x}$, we have

$$\begin{aligned}\mathbf{x}^{\mathcal{H}}(\mathbf{S}^{\dagger})^{\mathcal{H}}\mathbf{S}^{\dagger}\mathbf{x}^{\mathcal{H}} &= \mathbf{v}^{\mathcal{H}}(\mathbf{S}^{\mathcal{H}}\mathbf{S})^{-1}(\mathbf{S}^{\mathcal{H}}\mathbf{S})^{-1}\mathbf{v} = \mathbf{v}^{\mathcal{H}}\mathbf{G}^{\mathcal{H}}\mathbf{G}\mathbf{v} \\ &= \|\mathbf{G}\mathbf{v}\|^2 > 0 \end{aligned} \tag{A1}$$

i.e., $(\mathbf{S}^{\dagger})^{\mathcal{H}}\mathbf{S}^{\dagger}$ is a Hermitian positive definite matrix.

References

[1] B. Rankov and A. Wittneben, “Spectral efficient signaling for halfduplex relay channels,” in *Proc. 39th IEEE Asilomar Conf. Signals, Systems and Computers*, Nov. 2005.

[2] P. Popovski and H. Yomo, “Bi-directional amplification of throughput in a wireless multi-hop network,” in *Proc. IEEE VTC-Spring 2006*, pp. 851–855.

[3] Y.-C. Liang and R. Zhang, “Optimal analogue relaying with multi-antennas for physical layer network coding,” in *Proc. IEEE ICC 2008*, May 2008, pp. 3893–3897.

[4] T. Unger and A. Klein, “Linear transceive filters for relay stations with multiple antennas in the two-way relay channel,” in *Proc. 16th IST Mobile and Wireless Summit*, Budapest, Hungary, Jul. 2007.

[5] D. Falconer, S. L. Ariyavisitakul, A. Benyamin-Seeyar, and B. Edison, “Frequency domain equalization for single-carrier broadband wireless systems,” *IEEE Commun. Mag.*, vol. 40, no. 4, pp. 58–66, Apr. 2002.

[6] *IEEE Standard for Local and Metropolitan Area Networks Part 16: Air Interface for Fixed Broadband Wireless Access Systems*, IEEE Std 802.16-2004, 2004.

[7] Z. P. Yang and X. X. Feng, “A note on the trace inequality for products of Hermitian matrix power,” *J. Ineq. Pure Appl. Math*, vol. 3, no. 5, 2002, article 78.

[8] S. M. Kay, *Fundamentals of Statistical Signal Processing, Volume I: Estimation Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1993.

[9] R. L. Frank and S. A. Zadoff, “Phase shift pulse codes with good periodic correlation properties,” *IEEE Trans. Inf. Theory*, vol. 38, no. 6, pp. 381–382, Oct. 1992.