# Optimal Training Sequences for Channel Estimation in Bi-Directional Relay Networks with Multiple Antennas

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

***Abstract*—In this letter, we consider a bi-directional relay network in which two users, $\mathrm{U}_1$ and $\mathrm{U}_2$, exchange their information via a relay station, RS. Multiple antennas are deployed at both users and at RS. Single carrier cyclic prefix (SCCP) is used to combat intersymbol interference (ISI) in frequency-selective fading channels. The transmission process is divided into two time slots. At the first time slot, both users send their information to RS concurrently. RS then amplifies and broadcasts its received signals at the second time slot. We propose an algorithm to estimate the channel information at end users based on the least-square (LS) principle. To further minimize the mean square error (MSE) of the estimate, a method to design the optimal training sequences is also proposed. Simulation results show that the performance achieved by our optimal design is close to that with perfect channel information.**

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

## I. Introduction

BI-DIRECTIONAL relay networks have attracted increasing research attention due to their capability to improve spectral efficiency upon one-way relaying networks [1]. The scenario for both kinds of networks is that two users, $\mathrm{U}_1$ and $\mathrm{U}_2$, exchange information via a relay station, RS. In conventional one-way relay networks, 4 time slots are needed to accomplish one information exchange process [2]. In contrast, bi-directional relay networks require only 2 time slots to carry out an information exchange process. At the first time slot, *both* users send their information to RS on the same frequency band. RS receives a superposition of the signals transmitted from the two users. RS then, at the second time slot, amplifies its received signals and transmits it back to the end users. The end users, after receiving the signals from RS, subtract their own signals and perform the detection process to recover the data transmitted from the other party.

Multi-antenna transmission over multi-input multi-output (MIMO) channels has been proved to be effective in combating multipath fading, as well as in increasing the channel capacity. Hence, in [3], authors consider a bi-directional relay network which works in flat-fading channels and deploys multiple antennas at both users and at RS as well. A linear spatial filter is designed at RS to separate the signals from two users when RS transmits at the second time slot. This method demands that i) the channel information from two users to RS has to be available at RS and ii) the number of antennas at RS must be at least equal to the summation of the numbers of antennas of two users. The above two requirements may place a heavy burden on RS.

When systems operate in frequency-selective fading channels, intersymbol interference (ISI) may arise. The single carrier cyclic prefix (SCCP) [4], a block-based transmission scheme, is a promising one to suppress ISI. Moreover, SCCP enjoys a low peak-to-average power ratio (PAPR) as compared with multi-carrier schemes. Hence, we consider SCCP in our systems.

In this letter, we consider the channel estimation problem in bi-directional relay networks with multiple antennas at RS and both users. The channel estimation is performed at each of the two end users but not at RS. By doing so, we do not require a large number of antennas at RS; moreover, RS only forwards the signals it receives without any further processing which simplifies the computations at RS. The least-square (LS) principle is applied to estimate the channel information. Based on the mean-square error (MSE) criterion, an optimal method is proposed to design the training signals. Instead of separately estimating the channels (i.e., the channels from $\mathrm{U}_1$ to RS and those from $\mathrm{U}_2$ to RS), we estimate the *composite* channels resulting from the nature of signaling [5]. More specifically, we have two kinds of composite channels. The first one consists of overall channels from one user to RS and from RS back to itself. The second one is the overall channel from the remaining user to RS and from RS to the interested user.

The rest of the letter is structured as follows. Section II describes the system model. A channel estimation method is presented based on the LS principle in Section III. In this section, the MSE of channel estimate is derived. The requirements on training signals to minimize MSE are stated. Section IV provides the optimal training signal design method. Simulation results are given in Section V and necessary derivations are given in Appendix.

Notations: The capital bold letters denote matrices and the small bold letters denote row/column vectors. Transpose, Hermitian transpose of a vector/matrix are denoted by $(\cdot)^{\mathcal{T}}$ and $(\cdot)^{\mathcal{H}}$, respectively. The identity matrix of size $N$ is denoted

Paper approved by G.-H. Im, the Editor for Equalization and Multicarrier Techniques of the IEEE Communications Society. Manuscript received August 28, 2008; revised May 7, 2009 and August 7, 2009.

T.-H. Pham and H. K. Garg are with the Department of Electrical & Computer Engineering, National University of Singapore, Singapore 119260 (e-mail: {elepth, eleghk}@nus.edu.sg).

Y.-C. Liang is with the Institute for Inforcomm Research ($\mathrm{I}^2\mathrm{R}$), 1 Fusionopolis Way, #21-01 Connexis, Singapore 138632 (e-mail: ycliang@i2r.a-star.edu.sg).

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

This research was supported by NUS research grants, grant numbers are R-263-000-421-112 and R-263-000-436-112.

Digital Object Identifier 10.1109/TCOMM.2010.02.080431

by $\boldsymbol{I}_N$. $\mathbf{0}_{m,n}$ stands for a zero matrix of size $m \times n$ and $n$-order zero matrix is denoted by $\mathbf{0}_n$. $\mathbf{1}_{m\times n}$ is a $m \times n$ matrix of all ones. For a matrix $\boldsymbol{B}$, $[\boldsymbol{B}]_{m,n}$ is the $(m,n)$th element of $\boldsymbol{B}$. $\boldsymbol{\mathcal{D}}(\boldsymbol{b})$ is a diagonal matrix whose diagonal entries are from vector $\boldsymbol{b}$. $\odot$ is the Hadamard product. $\|\cdot\|$ is the Frobenius norm. $\boldsymbol{W}$ is the discrete Fourier transform (DFT) matrix of size $N$ with $[\boldsymbol{W}]_{m,n} = \frac{1}{\sqrt{N}}\exp\{-j\frac{2\pi mn}{N}\}$, $m,n = 0,\ldots,N-1$. Convolution of two vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ is denoted as $\boldsymbol{a} * \boldsymbol{b}$. $\mathbb{E}\{\cdot\}$ denotes expectation operation.

## II. System model

Let us consider a bi-directional relay network in which the number of antennas at each user is $N_t$ and the number of antennas at RS is $N_r$, $N_r \geq N_t$. At this section and the first half of the next section, for the ease of presentation, we consider the case of $N_t = 2$. The results for other scenarios will be generalized from this consideration later on. The channel between the $k$th antenna of $\mathrm{U}_1$ and the $m$th antenna of RS is denoted by $\boldsymbol{h}_{m,k}^{(1)} = \big[h_{m,k}^{(1)}(0)\ \cdots\ h_{m,k}^{(1)}(L-1)\big]^{\mathcal{T}} \in \mathbb{C}^{L\times 1}$ where $m = 1,\ldots,N_r$, $k = 1,2$ and $L$ is the channel length. The corresponding channel from $\mathrm{U}_2$ to RS is denoted by $\boldsymbol{h}_{m,k}^{(2)}$. Each element $h_{m,k}^{(u)}(l) \sim \mathcal{CN}\big(0,\sigma^{(u)}(l)\big)$ where $u = 1,2$ and $l = 0,\ldots,L-1$. The training signal vectors transmitted from the $k$th antenna of $\mathrm{U}_1$ and $\mathrm{U}_2$ are denoted by $\boldsymbol{s}_k^{(1)}$ and $\boldsymbol{s}_k^{(2)}$, respectively. The vector $\boldsymbol{s}_k^{(u)}$ satisfies $\mathbb{E}\{\|\boldsymbol{s}_k^{(u)}\|^2\} = NE_s$.

At Phase 1, those vectors, $\boldsymbol{s}_k^{(u)}$'s, are CP-added with CP length $L_{\mathrm{CP}}$ before transmission, $L_{\mathrm{CP}} \geq (L-1)$. The received signal vector at the $m$th antenna of RS (after CP removal) is denoted by $\boldsymbol{y}_m$. If we construct a vector $\boldsymbol{y} \triangleq [\boldsymbol{y}_1^{\mathcal{T}}\ \cdots\ \boldsymbol{y}_m^{\mathcal{T}}\ \cdots\ \boldsymbol{y}_{N_r}^{\mathcal{T}}]^{\mathcal{T}} \in \mathbb{C}^{NN_r\times 1}$, this vector can be written as follows

$$\boldsymbol{y} = \boldsymbol{H}^{(1)}\boldsymbol{s}^{(1)} + \boldsymbol{H}^{(2)}\boldsymbol{s}^{(2)} + \boldsymbol{n}, \tag{1}$$

where $\boldsymbol{s}^{(1)} = \big[(\boldsymbol{s}_1^{(1)})^{\mathcal{T}}\ (\boldsymbol{s}_2^{(1)})^{\mathcal{T}}\big]^{\mathcal{T}}$, $\boldsymbol{s}^{(2)} = \big[(\boldsymbol{s}_1^{(2)})^{\mathcal{T}}\ (\boldsymbol{s}_2^{(2)})^{\mathcal{T}}\big]^{\mathcal{T}}$, $\boldsymbol{n} = \big[\boldsymbol{n}_1^{\mathcal{T}}\ \cdots\ \boldsymbol{n}_m^{\mathcal{T}}\ \cdots\ \boldsymbol{n}_2^{\mathcal{T}}\big]^{\mathcal{T}}$ and

$$\boldsymbol{H}^{(1)} = \begin{bmatrix} \boldsymbol{H}_{1,1}^{(1)} & \boldsymbol{H}_{1,2}^{(1)} \\ \vdots & \vdots \\ \boldsymbol{H}_{N_r,1}^{(1)} & \boldsymbol{H}_{N_r,2}^{(1)} \end{bmatrix} \in \mathbb{C}^{NN_r\times 2N}, \tag{2}$$

$$\boldsymbol{H}^{(2)} = \begin{bmatrix} \boldsymbol{H}_{1,1}^{(2)} & \boldsymbol{H}_{1,2}^{(2)} \\ \vdots & \vdots \\ \boldsymbol{H}_{N_r,1}^{(2)} & \boldsymbol{H}_{N_r,2}^{(2)} \end{bmatrix} \in \mathbb{C}^{NN_r\times 2N}. \tag{3}$$

The matrix $\boldsymbol{H}_{m,k}^{(u)}$ is a circulant matrix with $\big[h_{m,k}^{(u)}(0)\ h_{m,k}^{(u)}(1)\ \cdots\ h_{m,k}^{(u)}(L-1)\ \mathbf{0}_{1\times(N-L)}\big]^{\mathcal{T}}$ as its first column. The noise vector $\boldsymbol{n}_m$ is a complex Gaussian random vector with zero mean and covariance matrix $\mathbb{E}\{\boldsymbol{n}_m\boldsymbol{n}_m^{\mathcal{H}}\} = N_0\boldsymbol{I}_N$. The vector $\boldsymbol{y}$ is then amplified by a real coefficients $\alpha$ such that $\alpha^2\mathbb{E}\{\boldsymbol{y}^{\mathcal{H}}\boldsymbol{y}\} = N_r \times NE_r$. The coefficient $\alpha$ is given in (4) on the top of next page.
The vector $\alpha\boldsymbol{y}_m$ is CP-added with CP length $L_{\mathrm{CP}}$ before being broadcasted back to $\mathrm{U}_1$ and $\mathrm{U}_2$. Without loss of generality, only the channel estimation problem at $\mathrm{U}_1$ is considered. A similar procedure can be applied at $\mathrm{U}_2$.
Let $\boldsymbol{r}_k^{(1)}$ and $\boldsymbol{n}_k^{(1)}$ are the received signal and noise vector at the $k$th antenna of $\mathrm{U}_1$ at Phase 2, $k = 1,2$ and $\boldsymbol{n}_k^{(1)} \sim \mathcal{CN}(\mathbf{0}_{N\times 1}, N_0\boldsymbol{I}_N)$. If we construct a vector $\boldsymbol{r}^{(1)} \triangleq \big[(\boldsymbol{r}_1^{(1)})^{\mathcal{T}}\ (\boldsymbol{r}_2^{(1)})^{\mathcal{T}}\big]^{\mathcal{T}}$, it is given by

$$\boldsymbol{r}^{(1)} = \alpha\boldsymbol{\mathcal{H}}^{(1)}\boldsymbol{H}^{(1)}\boldsymbol{s}^{(1)} + \alpha\boldsymbol{\mathcal{H}}^{(1)}\boldsymbol{H}^{(2)}\boldsymbol{s}^{(2)} + \big(\alpha\boldsymbol{\mathcal{H}}^{(1)}\boldsymbol{n} + \boldsymbol{n}^{(1)}\big), \tag{5}$$

where $\boldsymbol{n}^{(1)} = \big[(\boldsymbol{n}_1^{(1)})^{\mathcal{T}}\ (\boldsymbol{n}_2^{(1)})^{\mathcal{T}}\big]^{\mathcal{T}}$ and

$$\boldsymbol{\mathcal{H}}^{(1)} = \begin{bmatrix} \boldsymbol{H}_{1,1}^{(1)} & \cdots & \boldsymbol{H}_{m,1}^{(1)} & \cdots & \boldsymbol{H}_{N_r,1}^{(1)} \\ \boldsymbol{H}_{1,2}^{(1)} & \cdots & \boldsymbol{H}_{m,2}^{(1)} & \cdots & \boldsymbol{H}_{N_r,2}^{(1)} \end{bmatrix} \in \mathbb{C}^{2N\times NN_r}. \tag{6}$$

The circulant matrices, $\boldsymbol{H}_{m,k}^{(u)}$'s, can be decomposed as $\boldsymbol{H}_{m,k}^{(u)} = \boldsymbol{W}^{\mathcal{H}}\boldsymbol{\Lambda}_{m,k}^{(u)}\boldsymbol{W}$ where $\boldsymbol{\Lambda}_{m,k}^{(u)} = \mathrm{diag}\big\{H_{m,k}^{(u)}(0),\cdots,H_{m,k}^{(u)}(v),\cdots,H_{m,k}^{(u)}(N-1)\big\} \in \mathbb{C}^{N\times N}$ and $H_{m,k}^{(u)}(v) = \sum_{l=0}^{L-1} h_{m,k}^{(u)}(l)e^{-j\frac{2\pi vl}{N}}$. It is easy to see that the product of $\boldsymbol{H}_{m_1,k_1}^{(u_1)}$, $\boldsymbol{H}_{m_2,k2}^{(u_2)}$ and $\alpha$ can be written as

$$\alpha\boldsymbol{H}_{m_1,k_1}^{(u_1)}\boldsymbol{H}_{m_2,k_2}^{(u_2)} = \boldsymbol{W}^{\mathcal{H}}\alpha\boldsymbol{\Lambda}_{m_1,k_1}^{(u_1)}\boldsymbol{\Lambda}_{m_2,k_2}^{(u_2)}\boldsymbol{W}, \tag{7}$$

where $k_1,k_2,u_1,u_2 = 1,2$ and $m_1,m_2 = 1,\ldots,N_r$.
The right hand side (RHS) of (7), with the condition $(2L-1) \leq N$, can be regarded as the decomposition of a circular matrix which has the first column as $\big[\big(\boldsymbol{h}_{(m_1,k_1),(m_2,k_2)}^{(u_1,u_2)}\big)^{\mathcal{T}}\ \ \mathbf{0}_{1\times(N-2L+1)}\big]^{\mathcal{T}}$, $\boldsymbol{h}_{(m_1,k_1),(m_2,k_2)}^{(u_1,u_2)} \triangleq \alpha\big(\boldsymbol{h}_{m_1,k1}^{(u_1)} * \boldsymbol{h}_{m_2,k_2}^{(u_2)}\big)$. Based on (2), (5) and (6), we have three composite channels resulting from $\alpha\boldsymbol{\mathcal{H}}^{(1)}\boldsymbol{H}^{(1)}$: $\boldsymbol{h}_1 = \sum_{m=1}^{N_r}\boldsymbol{h}_{(m,1),(m,1)}^{(1,1)}$, $\boldsymbol{h}_2 = \sum_{m=1}^{N_r}\boldsymbol{h}_{(m,2),(m,1)}^{(1,1)}$ and $\boldsymbol{h}_3 = \sum_{m=1}^{N_r}\boldsymbol{h}_{(m,2),(m,2)}^{(1,1)}$. Similarly, $\alpha\boldsymbol{\mathcal{H}}^{(1)}\boldsymbol{H}^{(2)}$ provides us the following four composite channels: $\boldsymbol{h}_4 = \sum_{m=1}^{N_r}\boldsymbol{h}_{(m,1),(m,1)}^{(1,2)}$, $\boldsymbol{h}_5 = \sum_{m=1}^{N_r}\boldsymbol{h}_{(m,1),(m,2)}^{(1,2)}$, $\boldsymbol{h}_6 = \sum_{m=1}^{N_r}\boldsymbol{h}_{(m,2),(m,1)}^{(1,2)}$ and $\boldsymbol{h}_7 = \sum_{m=1}^{N_r}\boldsymbol{h}_{(m,2),(m,2)}^{(1,2)}$.
If $\boldsymbol{r}_1^{(1)}$ and $\boldsymbol{r}_2^{(1)}$ are pre-multiplied with $\boldsymbol{W}$, we obtain

$$\boldsymbol{z}^{(1)} = \big(\boldsymbol{I}_2 \otimes \boldsymbol{W}\big)\boldsymbol{r}^{(1)} = \boldsymbol{S}^{(1)}\boldsymbol{g} + \boldsymbol{S}^{(2)}\boldsymbol{q} + \boldsymbol{w}^{(1)}, \tag{8}$$

where $\boldsymbol{g} = \big[\boldsymbol{h}_1^{\mathcal{T}}\ \boldsymbol{h}_2^{\mathcal{T}}\ \boldsymbol{h}_3^{\mathcal{T}}\big]^{\mathcal{T}}$, $\boldsymbol{q} = \big[\boldsymbol{h}_4^{\mathcal{T}}\ \boldsymbol{h}_5^{\mathcal{T}}\ \boldsymbol{h}_6^{\mathcal{T}}\ \boldsymbol{h}_7^{\mathcal{T}}\big]^{\mathcal{T}}$, $\boldsymbol{w}^{(1)} = (\boldsymbol{I}_2\otimes\boldsymbol{W})\big(\alpha\boldsymbol{\mathcal{H}}^{(1)}\boldsymbol{n}+\boldsymbol{n}^{(1)}\big)$, $\boldsymbol{\mathcal{D}}_k^{(u)} = \mathrm{diag}\big\{\boldsymbol{W}\boldsymbol{s}_k^{(u)}\big\}$ for $u,k = 1,2$, and

$$\boldsymbol{S}^{(1)} = \begin{bmatrix} \boldsymbol{\mathcal{D}}_1^{(1)}\boldsymbol{F} & \boldsymbol{\mathcal{D}}_2^{(1)}\boldsymbol{F} & \mathbf{0}_{N\times(2L-1)} \\ \mathbf{0}_{N\times(2L-1)} & \boldsymbol{\mathcal{D}}_1^{(1)}\boldsymbol{F} & \boldsymbol{\mathcal{D}}_2^{(1)}\boldsymbol{F} \end{bmatrix}, \tag{9}$$

$$\boldsymbol{S}^{(2)} = \begin{bmatrix} \boldsymbol{\mathcal{D}}_1^{(2)}\boldsymbol{F} & \mathbf{0}_{N\times(2L-1)} & \boldsymbol{\mathcal{D}}_2^{(2)}\boldsymbol{F} & \mathbf{0}_{N\times(2L-1)} \\ \mathbf{0}_{N\times(2L-1)} & \boldsymbol{\mathcal{D}}_1^{(2)}\boldsymbol{F} & \mathbf{0}_{N\times(2L-1)} & \boldsymbol{\mathcal{D}}_2^{(2)}\boldsymbol{F} \end{bmatrix}. \tag{10}$$

where $\boldsymbol{F}$ is the first $(2L-1)$ columns of matrix $\sqrt{N}\boldsymbol{W}$. The noise vector $\boldsymbol{w}^{(1)}$ is a complex Gaussian random vector with zero mean and covariance matrix $\boldsymbol{\Sigma} = N_0\big(\alpha^2\boldsymbol{\Phi}+\boldsymbol{I}_{2N}\big)$ where

$$\boldsymbol{\Phi} = \begin{bmatrix} \sum_{m=1}^{N_r}|\boldsymbol{\Lambda}_{m,1}^{(1)}|^2 & \sum_{m=1}^{N_r}\boldsymbol{\Lambda}_{m,1}^{(1)}\big(\boldsymbol{\Lambda}_{m,2}^{(1)}\big)^{\mathcal{H}} \\ \sum_{m=1}^{N_r}\boldsymbol{\Lambda}_{m,2}^{(1)}\big(\boldsymbol{\Lambda}_{m,1}^{(1)}\big)^{\mathcal{H}} & \sum_{m=1}^{N_r}|\boldsymbol{\Lambda}_{m,2}^{(1)}|^2 \end{bmatrix}. \tag{11}$$

Equation (8) can be written in another form as

$$\boldsymbol{z}^{(1)} = \boldsymbol{S}\boldsymbol{k} + \boldsymbol{w}^{(1)}, \tag{12}$$

where $\boldsymbol{S} = \big[\boldsymbol{S}^{(1)}\ \ \boldsymbol{S}^{(2)}\big] \in \mathbb{C}^{2N\times 7(2L-1)}$ and $\boldsymbol{k} = \big[\boldsymbol{g}^{\mathcal{T}}\ \boldsymbol{q}^{\mathcal{T}}\big]^{\mathcal{T}} \in \mathbb{C}^{7(2L-1)}$.

$$\alpha = \sqrt{\frac{E_r}{\left(\sum_{k=1}^{2}\sum_{m=1}^{N_r}\sum_{l=0}^{L-1}\sigma_{m,k}^{(1)}(l) + \sum_{k=1}^{2}\sum_{m=1}^{N_r}\sum_{l=0}^{L-1}\sigma_{m,k}^{(2)}(l)\right)E_S + N_0}}. \quad (4)$$

## III. Least-Square (LS) Channel Estimation

The LS principle is applied to (12) to get the estimated $\hat{\boldsymbol{k}}$ of $\boldsymbol{k}$ as follows

$$\hat{\boldsymbol{k}} \triangleq \boldsymbol{S}^{\dagger}\boldsymbol{z}^{(1)} = \boldsymbol{k} + \boldsymbol{S}^{\dagger}\boldsymbol{w}^{(1)}, \quad (13)$$

where $\boldsymbol{S}^{\dagger} \triangleq \left(\boldsymbol{S}^{\mathcal{H}}\boldsymbol{S}\right)^{-1}\boldsymbol{S}^{\mathcal{H}}$ is the Moore-Penrose pseudo-inverse of $\boldsymbol{S}$. According to the LS principle, the number of rows of $\boldsymbol{S}$ should be greater than its number of columns, i.e., $L < (2N+7)/14$. The MSE of the estimation is defined as $\text{MSE} \triangleq \mathbb{E}\{(\boldsymbol{k}-\hat{\boldsymbol{k}})^{\mathcal{H}}(\boldsymbol{k}-\hat{\boldsymbol{k}})\}$ and it is determined for two cases. The first case is that the channel statistics are available at $\text{U}_1$. The second one is that such information is unavailable.

Case 1: In this case, MSE is written as

$$\text{MSE} = N_0\alpha^2\text{tr}\{\boldsymbol{S}^{\dagger}\mathbb{E}\{\boldsymbol{\Phi}\}(\boldsymbol{S}^{\dagger})^{\mathcal{H}}\} + N_0\text{tr}\{(\boldsymbol{S}^{\mathcal{H}}\boldsymbol{S})^{-1}\}. \quad (14)$$

We have

$$\mathbb{E}\{|H_{m,k}^{(1)}(v)|^2\} = \sum_{l_1=0}^{L-1}\sum_{l_2=0}^{L-1}\mathbb{E}\{h_{m,k}^{(1)}(l_1)(h_{m,k}^{(1)}(l_2))^*\}e^{-j\frac{2\pi v(l_1-l_2)}{N}} = \sum_{l=0}^{L-1}\sigma^{(1)}(l), \quad (15)$$

for $m = 1,\ldots,N_r$, $k = 1,2$ and $v = 0,\ldots,N-1$. Besides,

$$\mathbb{E}\{H_{m_1,k_1}^{(1)}(v)(H_{m_2,k_2}^{(1)}(v))^*\} = \sum_{l_1=0}^{L-1}\sum_{l_2=0}^{L-1}\mathbb{E}\{h_{m_1,k_1}^{(1)}(l_1)(h_{m_2,k_2}^{(1)}(l_2))^*e^{-j\frac{2\pi v(l_1-l_2)}{N}}\} = 0, \quad (16)$$

if $m_1 \neq m_2$ or $k_1 \neq k_2$, $m_1, m_2 = 1,\ldots,N_r$ and $k_1, k_2 = 1,2$. Therefore, we have $\mathbb{E}\{\sum_{m=1}^{N_r}|\boldsymbol{\Lambda}_{m,1}^{(1)}|^2\} = \mathbb{E}\{\sum_{m=1}^{N_r}|\boldsymbol{\Lambda}_{m,2}^{(1)}|^2\} = N_r\left(\sum_{l=0}^{L-1}\sigma^{(1)}(l)\right)\boldsymbol{I}_N$ and $\mathbb{E}\{\sum_{m=1}^{N_r}\boldsymbol{\Lambda}_{m,1}^{(1)}(\boldsymbol{\Lambda}_{m,2}^{(1)})^{\mathcal{H}}\} = \mathbb{E}\{\sum_{m=1}^{N_r}\boldsymbol{\Lambda}_{m,2}^{(1)}(\boldsymbol{\Lambda}_{m,1}^{(1)})^{\mathcal{H}}\} = \boldsymbol{0}_N$. Hence, we obtain $\mathbb{E}\{\boldsymbol{\Phi}\} = N_r\left(\sum_{l=0}^{L-1}\sigma^{(1)}(l)\right)\boldsymbol{I}_{2N}$. In other words, (14) can be written as

$$\text{MSE} = N_0\left(\alpha^2 N_r\left(\sum_{l=0}^{L-1}\sigma^{(1)}(l)\right) + 1\right)\text{tr}\{(\boldsymbol{S}^{\mathcal{H}}\boldsymbol{S})^{-1}\}. \quad (17)$$

We observe from (17) that minimizing the MSE is equivalent to minimizing $J \triangleq \text{tr}\{(\boldsymbol{S}^{\mathcal{H}}\boldsymbol{S})^{-1}\}$.

Case 2: In this case, MSE is written as

$$\text{MSE} = N_0\alpha^2\text{tr}\{\boldsymbol{S}^{\dagger}\boldsymbol{\Phi}(\boldsymbol{S}^{\dagger})^{\mathcal{H}}\} + N_0\text{tr}\{(\boldsymbol{S}^{\mathcal{H}}\boldsymbol{S})^{-1}\}. \quad (18)$$

To minimize the second term on RHS of (18), $I_2 \triangleq N_0\text{tr}\{(\boldsymbol{S}^{\mathcal{H}}\boldsymbol{S})^{-1}\}$, we must minimize the quantity $J$. The first term on RHS of (18), $I_1 \triangleq N_0\alpha^2\text{tr}\{\boldsymbol{S}^{\dagger}\boldsymbol{\Phi}(\boldsymbol{S}^{\dagger})^{\mathcal{H}}\}$, can be written in another form as $I_1 = N_0\alpha^2\text{tr}\{(\boldsymbol{S}^{\dagger})^{\mathcal{H}}\boldsymbol{S}^{\dagger}\boldsymbol{\Phi}\}$. In Appendix, we prove that $(\boldsymbol{S}^{\dagger})^{\mathcal{H}}\boldsymbol{S}^{\dagger}$ and $\boldsymbol{\Phi}$ are positive definite and positive semidefinite, respectively. Hence, from [6, Theorem 6.5.3], we have

$$I_1 \le N_0\alpha^2\text{tr}\{\boldsymbol{\Phi}\}\text{tr}\{(\boldsymbol{S}^{\dagger})^{\mathcal{H}}\boldsymbol{S}^{\dagger}\} = N_0\alpha^2\text{tr}\{\boldsymbol{\Phi}\}\text{tr}\{(\boldsymbol{S}^{\mathcal{H}}\boldsymbol{S})^{-1}\}. \quad (19)$$

It is observed that $I_1$ depends on channel realizations. Based on (19), minimizing $I_1$ can be done by minimizing its upper limit. In other words, we also minimize quantity $J$ as defined in Case 1.

In summary, the optimal training signal vectors transmitted from $\text{U}_1$ and $\text{U}_2$ are the ones that minimize $J = \text{tr}\{(\boldsymbol{S}^{\mathcal{H}}\boldsymbol{S})^{-1}\}$. These vectors must satisfy the following theorem.

*Theorem 1:* The training signal vectors $\boldsymbol{s}_k^{(u)}$ for $u, k = 1,2$, which minimize MSE, must be designed to satisfy the following condition for $u_1, u_2, k_1, k_2 = 1,2$

$$\boldsymbol{C}_{k_1,k_2}^{(u_1,u_2)} \triangleq \boldsymbol{F}^{\mathcal{H}}(\boldsymbol{\mathcal{D}}_{k_1}^{(u_1)})^{\mathcal{H}}\boldsymbol{\mathcal{D}}_{k_2}^{(u_2)}\boldsymbol{F} = \begin{cases}\boldsymbol{0}_{2L-1} & \text{if } u_1 \neq u_2 \text{ or } k_1 \neq k_2 \\ NE_s\boldsymbol{I}_{2L-1} & \text{if } u_1 = u_2 \text{ and } k_1 = k_2\end{cases}. \quad (20)$$

*Proof:* We define $\boldsymbol{A}^{-1}$ at the top of next page. If we apply an inequality[1] in [7], we have $J \ge \sum_{l=0}^{7(2L-1)-1}[\boldsymbol{A}]_{l,l}^{-1}$ where the equality holds if and only if $\boldsymbol{A}$ is a diagonal matrix. From the expression of $\boldsymbol{A}$, we observe that $\boldsymbol{C}_{k_1,k_2}^{(u_1,u_2)}$ must be $\boldsymbol{0}_{2L-1}$ if $u_1 \neq u_2$ or $k_1 \neq k_2$. For the cases that $u_1 = u_2$ and $k_1 = k_2$, $\boldsymbol{C}_{k_1,k_2}^{(u_1,u_2)}$'s, first of all, must be diagonal matrices. Hence, $J$ is written as

$$J \ge \sum_{l=0}^{(2L-1)-1}\Big([\boldsymbol{C}_{1,1}^{(1,1)}]_{l,l}^{-1} + [\boldsymbol{C}_{2,2}^{(1,1)}]_{l,l}^{-1} + 2[\boldsymbol{C}_{1,1}^{(2,2)}]_{l,l}^{-1} + 2[\boldsymbol{C}_{2,2}^{(2,2)}]_{l,l}^{-1} + [\boldsymbol{C}_{2,2}^{(1,1)} + \boldsymbol{C}_{1,1}^{(1,1)}]_{l,l}^{-1}\Big). \quad (22)$$

Now we apply a well-known inequality that the arithmetic mean of $n$ numbers is greater or equal than the geometric mean with equality holds if all numbers are equal. Therefore, to further minimize $J$, it is easy to deduce from (22) that $\boldsymbol{C}_{k_1,k_2}^{(u_1,u_2)}$ with $u_1 = u_2$ and $k_1 = k_2$ must be weighted identity matrices. With the power of $\boldsymbol{s}_k^{(u)}$, $u, k = 1,2$, is $NE_s$, it is straightforward to derive the result $\boldsymbol{C}_{k_1,k_2}^{(u_1,u_2)} = NE_s\boldsymbol{I}_{2L-1}$ with $u_1 = u_2$ and $k_1 = k_2$. This provides us $J_{\min} = (2L-1)\frac{13}{2NE_s}$. ■

Several remarks on generalization of the above results for the case with $N_t > 2$ are as follows:

*Remark 1:* The constraint on $L$ as stated by the least-square estimation is $L < \frac{2N+3N_t+1}{2(3N_t+1)}$.

*Remark 2:* Theorem 1 on the requirements of training signal vectors is also applicable for $N_t > 2$. Proof for the general case is also the same. For $N_t > 2$, we have $J_{\min} = (2L-1)\frac{5N_t^2+3N_t}{4NE_s}$. In case the channel statistics are available at $\text{U}_1$, we have $\text{MSE}_{\min} = N_0\left(\alpha^2 N_r\left(\sum_{l=0}^{L-1}\sigma^{(1)}(l)\right) + 1\right) \times (2L-1)\frac{5N_t^2+3N_t}{2NE_s}$. In case channel statistics are unavailable, we have the minimum of second term $I_2$ as $I_{2,\min} = N_0 \times (2L-1)\frac{5N_t^2+3N_t}{4NE_s}$.

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

$$\boldsymbol{A}^{-1} \triangleq \left(\boldsymbol{S}^{\mathcal{H}}\boldsymbol{S}\right)^{-1} = \begin{bmatrix} \boldsymbol{C}_{1,1}^{(1,1)} & \boldsymbol{C}_{1,2}^{(1,1)} & \mathbf{0}_{2L-1} & \mathbf{0}_{2L-1} & \mathbf{0}_{2L-1} & \mathbf{0}_{2L-1} & \mathbf{0}_{2L-1} \\ \boldsymbol{C}_{2,1}^{(1,1)} & \boldsymbol{C}_{2,2}^{(1,1)} + \boldsymbol{C}_{1,1}^{(1,1)} & \boldsymbol{C}_{1,2}^{(1,1)} & \boldsymbol{C}_{2,1}^{(1,2)} & \boldsymbol{C}_{1,1}^{(1,2)} & \boldsymbol{C}_{2,2}^{(1,2)} & \boldsymbol{C}_{1,2}^{(1,2)} \\ \mathbf{0}_{2L-1} & \boldsymbol{C}_{2,1}^{(1,1)} & \boldsymbol{C}_{2,2}^{(1,1)} & \mathbf{0}_{2L-1} & \boldsymbol{C}_{2,1}^{(1,2)} & \mathbf{0}_{2L-1} & \boldsymbol{C}_{2,2}^{(1,2)} \\ \boldsymbol{C}_{1,1}^{(2,1)} & \boldsymbol{C}_{1,2}^{(2,1)} & \mathbf{0}_{2L-1} & \boldsymbol{C}_{1,1}^{(2,2)} & \mathbf{0}_{2L-1} & \boldsymbol{C}_{1,2}^{(2,2)} & \mathbf{0}_{2L-1} \\ \mathbf{0}_{2L-1} & \boldsymbol{C}_{1,1}^{(2,1)} & \boldsymbol{C}_{1,2}^{(2,1)} & \mathbf{0}_{2L-1} & \boldsymbol{C}_{1,1}^{(2,2)} & \mathbf{0}_{2L-1} & \boldsymbol{C}_{1,2}^{(2,2)} \\ \boldsymbol{C}_{2,1}^{(2,1)} & \boldsymbol{C}_{2,2}^{(2,1)} & \mathbf{0}_{2L-1} & \boldsymbol{C}_{2,1}^{(2,2)} & \mathbf{0}_{2L-1} & \boldsymbol{C}_{2,2}^{(2,2)} & \mathbf{0}_{2L-1} \\ \mathbf{0}_{2L-1} & \boldsymbol{C}_{2,1}^{(2,1)} & \boldsymbol{C}_{2,2}^{(2,1)} & \mathbf{0}_{2L-1} & \boldsymbol{C}_{2,1}^{(2,2)} & \mathbf{0}_{2L-1} & \boldsymbol{C}_{2,2}^{(2,2)} \end{bmatrix}^{-1}. \tag{20}$$

*Remark 3:* Another quantity that can be used to evaluate the effectiveness of an unbiased estimator is the Cramér-Rao Lower Bound (CRLB). The derivations of CRLB is omitted due to space constraint and the result is $\mathrm{CRLB}(\boldsymbol{k}) = \mathrm{tr}\{(\boldsymbol{S}^{\mathcal{H}}\boldsymbol{\Sigma}^{-1}\boldsymbol{S})^{-1}\}$. The CRLB varies from channel realizations to channel realizations; hence, in Section V, we use the *average* CRLB which is the average of many individual CRLBs.

## IV. Optimal Training Sequence Design

In this section we present our proposed design using special sequences known as Zadoff-Chu sequences [8]. Zadoff-Chu sequences have been applied in training design problems in conventional MIMO systems [9, 10] and can be considered as generalized chirp-like polyphase sequences [11]. The general expression of $M$-length Zadoff-Chu sequences is given by

$$x(n) = \begin{cases} e^{-j\frac{\pi G n(n+2g)}{M}} & n = 0, 1, \ldots, M-1;\ M \text{ is even} \\ e^{-j\frac{\pi G n(n+1+2g)}{M}} & n = 0, 1, \ldots, M-1;\ M \text{ is odd} \end{cases}, \tag{23}$$

where $g$ and $G$ are integer in which $G$ is an integer relatively prime to $M$. We have the following theorem to design $\boldsymbol{s}_k^{(u)}$'s.

*Theorem 2:* Suppose we have a $N/U$-length Zadoff-Chu sequence $\boldsymbol{x}$ ($N/U$ and $U$ are both integer numbers), $\|\boldsymbol{x}\|^2 = (N/U)E_s$. Let $\boldsymbol{d}_i = \boldsymbol{d}_0 \odot \boldsymbol{e}_i$ where $i = 1, \ldots, U-1$, $\boldsymbol{e}_i = \left[1\ e^{j\frac{2\pi i}{N}} \cdots\ e^{j\frac{2\pi(N-1)i}{N}}\right]^{\mathcal{T}}$ and $\boldsymbol{d}_0 = \mathbf{1}_{U\times 1} \otimes \boldsymbol{x}$. The training sequences from $\mathrm{U}_1$ and $\mathrm{U}_2$ can be chosen from $2N_t$ different sequences belonging to $\{\boldsymbol{d}_i\}_{i=0}^{U-1}$ ($U \geq 2N_t$). By doing so, we have a new constraint on $L$ as $L < (N+2U)/U$.

*Proof:* The $k$th element of DFT of $\boldsymbol{d}_i$, $d_{i,f}(k)$, equals $\sqrt{U}x_f\big((k-i)/U\big)$ if $k$ belongs to a index set $\mathcal{S}_i = \{i, i+U, \ldots, i+(N/U-1)U\}$ and 0 elsewhere; $x_f(k)$, $k = 0, \cdots, N/U-1$, is the $k$th element of DFT of $\boldsymbol{x}$.

Two arbitrary training signal vectors $\boldsymbol{s}_{k_1}^{(u_1)}$ and $\boldsymbol{s}_{k_2}^{(u_2)}$, $u_1 \neq u_2$ or $k_1 \neq k_2$, are two distinct vectors belonging to the set $\{\boldsymbol{d}_i\}_{i=0}^{U-1}$. The intersection between these two vectors' index sets is the null set. Hence, $\boldsymbol{C}_{k_1,k_2}^{(u_1,u_2)} = \boldsymbol{F}^{\mathcal{H}}(\boldsymbol{\mathcal{D}}_{k_1}^{(u_1)})^{\mathcal{H}}\boldsymbol{\mathcal{D}}_{k_2}^{(u_2)}\boldsymbol{F} = \mathbf{0}_{2L-1}$. Suppose that the training signal vector at antenna $k$ of user $u$ is $\boldsymbol{d}_i$. Hence, $\left[(\boldsymbol{\mathcal{D}}_k^{(u)})^{\mathcal{H}}\boldsymbol{\mathcal{D}}_k^{(u)}\right]_{l,l} = UE_s$ if $l \in \mathcal{S}_i$. The $(p,q)$ element of $\left[\boldsymbol{C}_{k,k}^{(u,u)}\right]_{p,q}$, $p, q = 0, 1, \ldots, 2L-2$, is calculated as

$$\begin{aligned} \left[\boldsymbol{C}_{k,k}^{(u,u)}\right]_{p,q} &= UE_s \sum_{l\in \mathrm{S}_i} e^{j\frac{2\pi pl}{N}} e^{-j\frac{2\pi ql}{N}} \\ &= \begin{cases} NE_s & \text{if } p = q \\ UE_s e^{j\frac{2\pi i(p-q)}{N}} \times \frac{e^{j2\pi(p-q)}-1}{e^{j\frac{2\pi U(p-q)}{N}}-1} & \text{if } p \neq q \end{cases}. \end{aligned} \tag{24}$$

If $p \neq q$, we have $0 < |p-q| \leq 2L-2$. According to Theorem 1, $\left[\boldsymbol{C}_{k,k}^{(u,u)}\right]_{p,q} = 0$ if $p \neq q$. Based on the second line of (24), it means $(e^{j\frac{2\pi U(p-q)}{N}} - 1) \neq 0$. To satisfy this condition, we have another restriction on $L$ as $L < \frac{N+2U}{2U}$. This restriction shows that if we want to maximize the channel length $L$, we must minimize $U$ under the condition that $N/U$ is an integer. ■

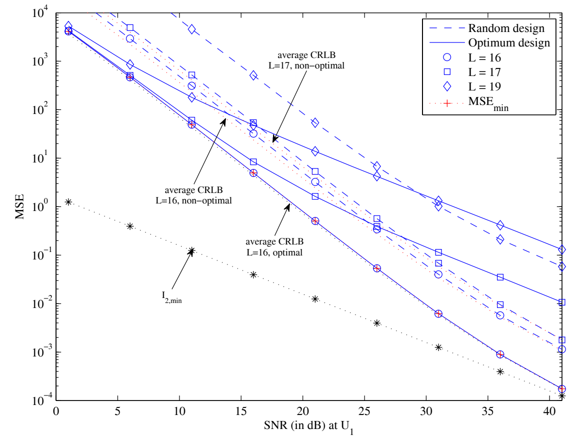


Fig. 1. MSE performance of channel estimation for the case when $N_t = N_r = 2$, $\xi = 0$dB.

## V. Simulation Results

In this section, we present the simulation results for our system with $N_t = N_r = 2$. The size of training signal vectors $N = 128$. The channel response $\boldsymbol{h}_{m,k}^{(u)}$ has length $L$ and the uniform power delay profile is assumed. We define the relative power gain factor between two links as $\xi \triangleq 10\log\left(\frac{\sum_{l=0}^{L-1}\sigma^{(1)}(l)}{\sum_{l=0}^{L-1}\sigma^{(2)}(l)}\right)$ in dB.

The channels are assumed to be constant over 20 time slots in which 2 time slots are used for training purpose and the left slots are used for data transmission. For data transmission, QPSK signaling is used. We choose $E_s = 2$ and define signal-to-noise ratio (SNR) at $\mathrm{U}_1$ (and $\mathrm{U}_2$) as $E_s/N_0$ and SNR at RS as $E_r/N_0$ which is 40dB in all our simulations.

Firstly, we consider a "symmetric" scenario, i.e., $\xi = 0$dB. In this case, both users have the same performances, thus the results of $\mathrm{U}_1$ are given. As in Theorem 2, we must choose the minimum value of $U$ but that value must satisfy the condition that $N/U$ is an integer. For $N_t = 2$, we choose $U = 4$. In case $N = 128$, the optimal design can support channel length

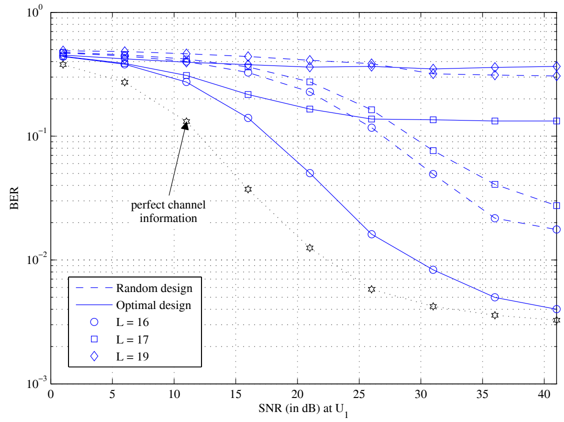


Fig. 2. BER performance for the case when $N_t = N_r = 2$, $\xi$ = 0dB.

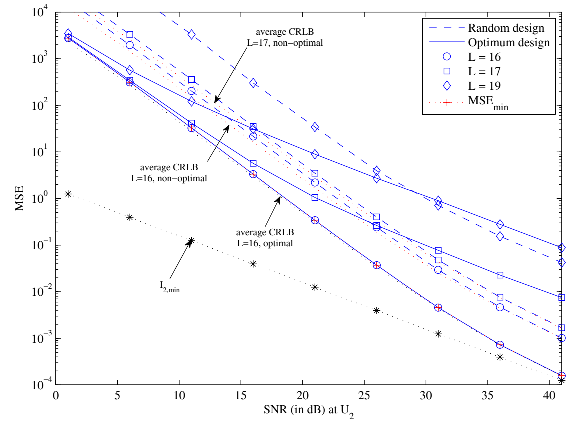


Fig. 4. MSE performance of channel estimation at $U_2$ for the case when $N_t = N_r = 2$, $\xi$ = 3dB.

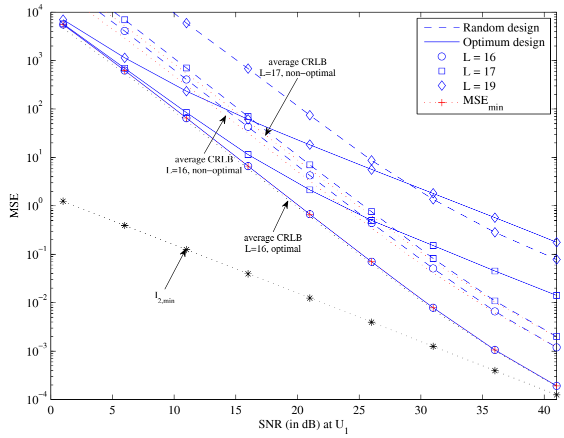


Fig. 3. MSE performance of channel estimation at $U_1$ for the case when $N_t = N_r = 2$, $\xi$ = 3dB.

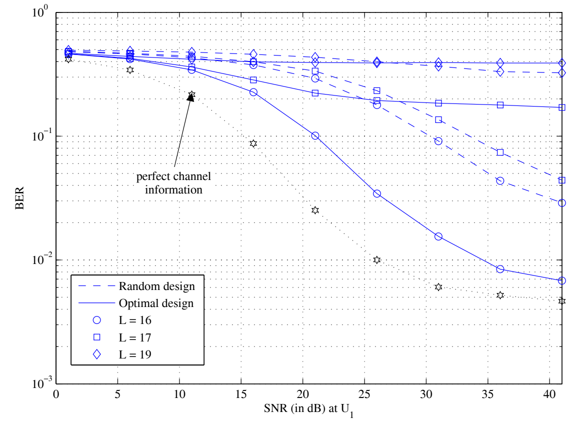


Fig. 5. BER performance of channel estimation at $U_1$ for the case when $N_t = N_r = 2$, $\xi$ = 3dB.

$L$ up to $L_1 = 16$; meanwhile that of the LS method without training design is $L_2 = 18$.

Fig. 1 provides the MSE performance for two kinds of training signals: i) our proposed optimal design and ii) random design where elements of training vectors are randomly chosen from QPSK constellation. From Fig. 1, it is observed that when $L = 16$, which lies in the range that our proposed optimal design can handle, the MSE performance given by our design outperforms the performance obtained by random training vectors. When $L = 17$, which is greater than the maximum channel length supported by our optimal design, the MSE performance of the optimal design is worse (compared to that with $L = 16$). Meanwhile, random design can still provide better MSE performance when $L = 17$ in the high-SNR region. Random design only blows up when $L = 19$ which is bigger than the maximum value $L_2$. In this figure, the average CRLB for different scenarios are also provided. It is observed that the MSE of the channel estimation algorithm is very close to CRLB. With the optimal design, the MSE and CRLB curves are almost the same. BER performances are illustrated in Fig. 2. The detection of signals transmitted from $U_2$ at $U_1$ is as follows: firstly, $U_1$ deducts its contribution in the signals received from RS; then, $U_1$ uses the Zero-Forcing (ZF) frequency domain equalizer to detect the signals of $U_2$. Similarly to MSE performances, if the channel length $L$ is in the supported range of our proposed design, the obtained BER performance is superior to the random design.

Secondly, we investigate an "asymmetric" scenario with $\xi$ = 3dB. In the following, we provide the performance of both $U_1$ and $U_2$. The MSE performances for two kinds of training signal vectors at $U_1$ and $U_2$ are given in Fig. 3 and Fig. 4, respectively. The conclusions drawn in the "symmetric" scenario are still applicable here. The BER performance at $U_1$ and $U_2$ are provided in Fig. 5 and Fig. 6, respectively. At both users, the performance given by the optimal design is much better than that of the random design in the supported range of channel lengths. The performance given by the optimal design is close to that given by perfect channel information.

## VI. Conclusions

In this letter, channel estimation problem in a bi-directional relay network with multiple antennas is investigated. The LS principle is used to estimate the channel information. Based on Zadoff-Chu sequence, an optimal training signal design has been proposed to minimize the MSE of channel

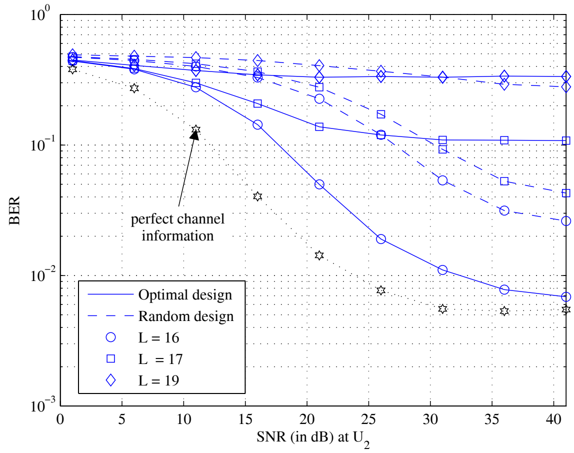


Fig. 6. BER performance of channel estimation at $\mathrm{U}_2$ for the case when $N_t = N_r = 2$, $\xi = 3$dB.

estimate. Simulation results have shown that our proposed design provides better performances in terms of MSE and BER than the random design.

## Appendix

In this appendix, we prove the positive semidefiniteness of $\boldsymbol{\Phi}$ and positive definiteness of $(\boldsymbol{S}^{\dagger})^{\mathcal{H}}\boldsymbol{S}^{\dagger}$. Matrix $\boldsymbol{\Phi}$ is given in (11). Clearly, two diagonal block matrices of $\boldsymbol{\Phi}$, $\sum_{m=1}^{N_r}|\boldsymbol{\Lambda}_{m,1}^{(1)}|^2$ and $\sum_{m=1}^{N_r}|\boldsymbol{\Lambda}_{m,2}^{(1)}|^2$, are two positive definite matrix. From [12, p. 89], $\boldsymbol{\Phi}$ is a semidefinite matrix.

Based on the definition of $\boldsymbol{S}^{\dagger}$, we have $(\boldsymbol{S}^{\dagger})^{\mathcal{H}}\boldsymbol{S}^{\dagger} = \boldsymbol{S}(\boldsymbol{S}^{\mathcal{H}}\boldsymbol{S})^{-1}(\boldsymbol{S}^{\mathcal{H}}\boldsymbol{S})^{-1}\boldsymbol{S}^{\mathcal{H}}$. We define $\boldsymbol{V} \triangleq (\boldsymbol{S}^{\mathcal{H}}\boldsymbol{S})^{-1}$. It is observed that $\boldsymbol{V}^{\mathcal{H}} = \boldsymbol{V}$. Let $\boldsymbol{y}$ be an arbitrary non-zero vector and we define $\boldsymbol{x} \triangleq \boldsymbol{S}^{\mathcal{H}}\boldsymbol{y}$. The product $\boldsymbol{y}^{\mathcal{H}}(\boldsymbol{S}^{\dagger})^{\mathcal{H}}\boldsymbol{S}^{\dagger}\boldsymbol{y}$ can be written as

$$\begin{aligned}\boldsymbol{y}^{\mathcal{H}}(\boldsymbol{S}^{\dagger})^{\mathcal{H}}\boldsymbol{S}^{\dagger}\boldsymbol{y} &= \boldsymbol{x}^{\mathcal{H}}(\boldsymbol{S}^{\mathcal{H}}\boldsymbol{S})^{-1}(\boldsymbol{S}^{\mathcal{H}}\boldsymbol{S})^{-1}\boldsymbol{x}\\ &= \boldsymbol{x}^{\mathcal{H}}\boldsymbol{G}^{\mathcal{H}}\boldsymbol{G}\boldsymbol{x} = \|\boldsymbol{G}\boldsymbol{x}\|^2 > 0, \quad \text{(A-1)}\end{aligned}$$

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

## References

[1] P. Larsson, N. Johansson, and K. Sunell, “Coded bi-directional relaying," in *Proc. IEEE VTC-Spring'2006*, May 2006, pp. 851-855.
[2] J. N. Laneman, D. N. C. Tse, and G. W. Wornell, “Cooperative diversity in wireless networks: efficient protocols and outage behavior," *IEEE Trans. Inf. Theory*, no. 12, pp. 3062-3080, Dec. 2004.
[3] 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 Wireless Summit*, Budapest, Hungary, July 2007.
[4] H. Sari, G. Karam, and I. Jeanclaude, “Frequency-domain equalization of mobile radio and terrestrial broadcast channels," in *Proc. IEEE GLOBECOM'1994*, Nov-Dec 1994.
[5] F. Gao, T. Cui, and A. Nallanathan, “On channel estimation and optimal training design for amplify and forward relay networks," *IEEE Trans. Wireless Commun.*, vol. 7, no. 5, pp. 1907-1916, May 2008.
[6] F. Zhang, *Matrix Theory: Basic Results and Techniques*. Springer-Verlag, 1999.
[7] S. M. Kay, *Fundamentals of Statistical Signal Processing, Volume I: Estimation Theory*. Prentice Hall, 1993.
[8] R. L. Frank and S. A. Zadoff, “Phase shift pulse codes with good periodic correlation properties," *IEEE Trans. Inf. Theory*, no. 6, pp. 381-382, Oct. 1992.
[9] J. Siew, J. Coon, R. J. Piechocki, A. Dowler, A. Nix, M. Beach, S. Armour, and J. McGeehan, “A channel estimation algorithm for MIMO-SCFDE," *IEEE Commun. Lett.*, vol. 8, no. 9, pp. 555-557, 2004.
[10] J. Coon, M. Beach, and J. McGeehan, “Optimal training sequences for channel estimation in cyclic-prefix based single-carrier systems with transmit diversity," *IEEE Signal Process. Lett.*, vol. 11, no. 9, pp. 729-732, 2004.
[11] B. M. Popovic, “Generalized chirp-like polyphase sequences with optimum correlation properties," *IEEE Trans. Inf. Theory*, vol. 38, no. 4, pp. 1406-1409, July 1992.
[12] F. Zhang, *The Schur Complement and its Applications*. Springer, 2005.