

RESEARCH

Open Access

Doubly selective channel estimation for amplify-and-forward relay networks

Gongpu Wang^{1,2*}, Feifei Gao³, Rongtao Xu² and Chintla Tellambura⁴

Abstract

In this article, doubly selective channel estimation is considered for *amplify-and-forward*-based relay networks. The complex exponential basis expansion model is chosen to describe the time-varying channel, from which the infinite channel parameters are mapped onto finite ones. Since direct estimation of these coefficients encounters high computational complexity and large spectral cost, we develop an efficient estimator that only targets at *useful* channel parameters that could guarantee the later data detection. The training sequence design that can minimize the channel estimation mean-square error is also proposed. Finally, numerical results are provided to corroborate the study.

Keywords: Doubly selective channel, Relay network, Channel estimation, Training sequence design, Basis expansion model

1 Introduction

Wireless relay networks have been a highly active research field ever since the pioneer work [1-3]. A typical relay network consists of a source node, one or several relay nodes, and a destination node. The transmission from the source node to the destination node involves two phases. In the first phase, the source node broadcasts signals to the relay nodes and possibly to the destination node. In the second phase, the relay nodes re-transmit its received signals in the first phase to the destination node under a certain relaying strategy such as amplify-and-forward (AF) and decode-and-forward (DF) [2]. It has been shown that such a relay network can enhance the system throughput [1], improve the transmission coverage [2], and increase the multiplexing gain [3]. Several standards that have been or are being specified for the next-generation mobile broadband communication systems [4] already included the relay-aided transmission, e.g., long-term evolution-advanced (LTE-A), IEEE 802.16j, and IEEE 802.16m.

Like any other wireless communication system, a relay network performs better with better channel estimates,

and the quality of channel acquisition has a significant effect on the overall system performance. In addition, knowledge of channel state information is often a prerequisite for some physical layer approaches such as optimal strategy selection and the precoding design.

Assuming block fading scenarios, several channel estimation schemes were proposed for relay network with one or multiple-relay nodes. For example, the authors of [5,6] studied the channel estimation for relay networks and pointed out that there exist many differences in channel estimation between the AF-based relay networks and the traditional point-to-point networks. Shortly later, channel estimations under frequency-selective environment were developed in [7,8].

However, in many practical cases the source node, the relay node, and the destination node can be mobile. The relative motion between any two nodes will cause Doppler shift and thus make the channel time-varying [9]. Time-varying channel estimations for relay networks are studied in [10,11]. It is shown in [10] that for time-varying channels in relay networks, transmission of several subblocks can obtain better estimation performance than that of a single subblock.

Furthermore, when the transmission data rates are high and nodes are mobile, the relay network is expected to operate over doubly selective channels. Canonne-Velasquez et al. [12] suggest a channel estimation algorithm for orthogonal frequency division

*Correspondence: prof.wang@ieee.org

¹School of Computer and Information Technology, Beijing Jiaotong University, Beijing 100044, China

²State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, Beijing 100044, China

Full list of author information is available at the end of the article

multiplexing-based AF relay systems over doubly selective environments. However, the proposed channel estimation algorithm neglects the interference from the data symbols. Moreover, the optimal training sequence design for doubly selective AF relay channels remains an open problem. To the best of the authors' knowledge, estimation techniques considering inter symbol interference between data and training symbols, as well as training sequence design, have not yet been developed. This motivates our current work.

The doubly selective channel can typically be represented in two ways: the autoregressive (AR) process [13] or the basis expansion model (BEM) [14]. AR models describe channel variation through a symbol-by-symbol update manner. Though second- and third-order AR models can provide excellent fits to the Jake model, the first-order AR process is usually adopted [15] due to its tractability. On the other hand, BEM describes the doubly selective channel as the superpositions of time-varying basis functions weighted by time-invariant coefficients. The candidate basis functions include complex exponential (Fourier) functions [14,16], polynomials [17], wavelet [18], discrete prolate spheroidal sequences [19,20], etc. BEMs can well describe the time variations of channel, and thus achieve better approximation performance than symbol-wise AR models [21].

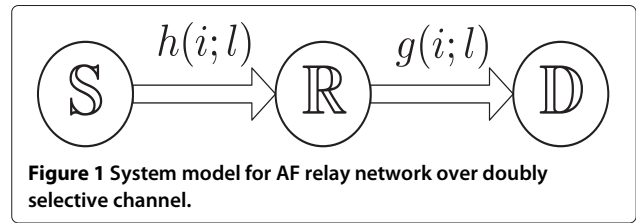
In this article, we focus on complex exponential BEM (CE-BEM) [16] due to its popularity and clear physical meaning. The data frame structure is designed to adapt to the transmission in doubly selective channels and to facilitate both channel estimation and data detection. We first develop an estimator that targets the combined channel parameters and then propose a detection algorithm. The training sequence that can minimize the channel estimation mean-square error (MSE) is also found.

The rest of this article is organized as follows. Section 2 presents the relay system model over doubly selective channels. Section 3 discusses the channel statistics and CE-BEM approximation accuracy. Section 4 develops the channel estimator and data detector, and suggests the optimal training sequence design. Simulation results are provided in Section 5. Finally, conclusions are drawn in Section 6.

Notations: Vectors and matrices are given in boldface letters; the transpose, Hermitian, and inverse of \mathbf{A} are \mathbf{A}^T , \mathbf{A}^H , and \mathbf{A}^{-1} , respectively; $\text{diag}\{\mathbf{a}\}$ is the diagonal matrix formed by \mathbf{a} , $\text{tr}(\mathbf{A})$ is the trace of the square matrix \mathbf{A} , $\mathbf{0}_L$ is the $L \times L$ matrix with all entries 0, and \mathbf{I}_L is the $L \times L$ identity matrix.

2 System model

Consider an AF relay network with one source node \mathbb{S} , one relay node \mathbb{R} , and one destination node \mathbb{D} (Figure 1). Let $h(i; l)$ denote the doubly selective channel between



the source node \mathbb{S} and the relay node \mathbb{R} , $g(i; l)$ denote the doubly selective channel between the relay node \mathbb{R} and the destination node \mathbb{D} .^a Without loss of generality, we assume that the channel length of both $h(i; l)$ and $g(i; l)$ as $L + 1$, and each tap is modeled as a zero mean complex Gaussian random process with power $\sigma_{h,l}^2$ (or $\sigma_{g,l}^2$).

We propose a new transmission scheme as shown in Figure 2. Each transmission block that contains N symbols is divided into P subblocks. Assume the k th subblock contains N_k symbols of which N_{s_k} symbols are data and are represented by \mathbf{s}_k , while N_{b_k} symbols are pilots and are represented by \mathbf{b}_k . The total number of data symbols is $N_s = \sum_{k=1}^P N_{s_k}$ and the total number of pilots is $N_p = \sum_{k=1}^P N_{b_k}$. With such a structure, we can represent the whole block as a vector

$$\mathbf{x} = [\mathbf{s}_1^T, \mathbf{b}_1^T, \dots, \mathbf{s}_P^T, \mathbf{b}_P^T]^T. \quad (1)$$

During the first phase, the relay node \mathbb{R} receives

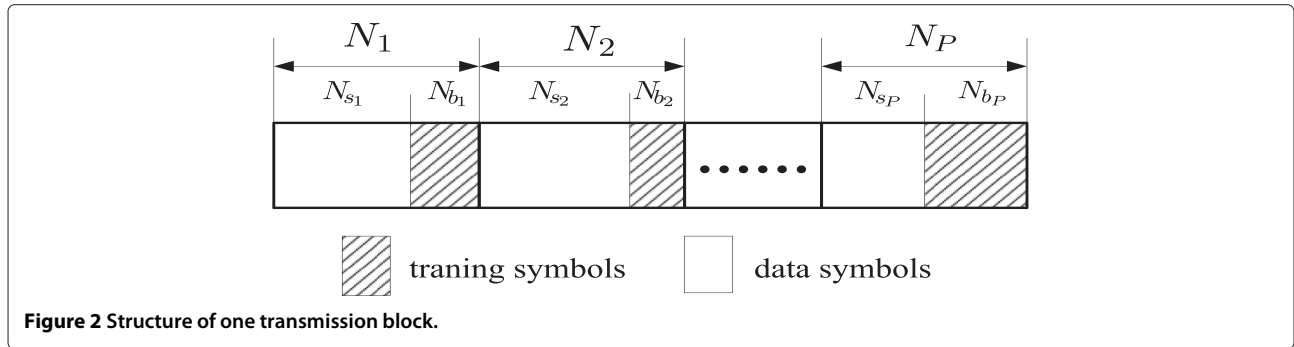
$$r(i) = \sum_{l=0}^L h(i; l)x(i-l) + w_1(i), \quad (2)$$

where $w_1(i)$ is the additive complex white Gaussian noise (ACWGN) with mean zero and variance $\sigma_{w_1}^2$, i.e., $w_1 \sim \mathcal{CN}(0, \sigma_{w_1}^2)$. During the second phase, the relay node \mathbb{R} amplifies $r(i)$ with a constant factor α and then retransmit it to the destination node \mathbb{D} . The signal obtained by \mathbb{D} is

$$\begin{aligned} y(i) &= \sum_{l=0}^L g(i; l)\alpha r(i-l) + w_2(i) \\ &= \alpha \sum_{l=0}^L g(i; l) \left(\sum_{l=0}^L h(i; l)x(i-l) \right) \\ &\quad + \underbrace{\alpha \sum_{l=0}^L g(i; l)w_1(i-l)}_{w(i)} + w_2(i), \end{aligned} \quad (3)$$

where $w_1(i)$ is the ACWGN with mean zero and variance $\sigma_{w_1}^2$, i.e., $w_1(i) \sim \mathcal{C}(0, \sigma_{w_1}^2)$ and $w(i)$ is defined as the combined noise.

Remark 1. Suppose the average power of the source node is \mathcal{P}_1 , i.e., $E\{|x_i(n)|^2\} = \mathcal{P}_1$ and the average power of the



relay node is \mathcal{P}_r . Then the amplifier factor α can be chosen as

$$\alpha = \sqrt{\frac{\mathcal{P}_r}{\mathcal{P}_1 \sum_{l=0}^L \sigma_{h,l}^2 + \sigma_{w_1}^2}}. \quad (4)$$

3 Doubly selective channel in relay networks

3.1 Relay channel statistics and CE-BEM

It was shown in [5,22] that for relay networks, the channel statistics depend on the mobility of the three nodes. Denote f_{ds}, f_{dd} , and f_{dr} as the maximum Doppler shifts due to the motion of \mathbb{S}, \mathbb{D} , and \mathbb{R} , respectively. The discrete autocorrelation functions for the l th tap of $h(i; l)$ can be represented as [22,23]

$$\begin{aligned} R_{h,l}(m) &= \sigma_{h,l}^2 E(h(n+m; l)h^*(n; l)) \\ &= \sigma_{h,l}^2 J_0(2\pi f_{ds} m T_s) J_0(2\pi f_{dr} m T_s), \end{aligned} \quad (5)$$

$$\begin{aligned} R_{g,l}(m) &= \sigma_{g,l}^2 E(g(n+m; l)g^*(n; l)) \\ &= \sigma_{g,l}^2 J_0(2\pi f_{dr} m T_s) J_0(2\pi f_{dd} m T_s), \end{aligned} \quad (6)$$

where $J_0(\cdot)$ is the zeroth-order Bessel function of the first kind, and T_s is the symbol sampling duration. If one node is fixed, i.e., the corresponding Doppler shift becomes zero, then (5) and (6) reduce to the well-known Jakes model [9].

In fact, (5) and (6) reveal that the power spectra of $h(i; l)$ and $g(i; l)$ span over the bandwidth $f_{d_1} = f_{ds} + f_{dr}$ and $f_{d_2} = f_{dr} + f_{dd}$, respectively. According to the analysis of CE-BEM in [14,16], we can express the doubly selective channel as

$$h(i; l) = \sum_{q=0}^{Q_1} h_q(l) e^{j2\pi(q-Q_1/2)i/N}, \quad (7)$$

$$g(i; l) = \sum_{q=0}^{Q_2} g_q(l) e^{j2\pi(q-Q_2/2)i/N}, \quad (8)$$

where $0 \leq i \leq N-1, 0 \leq l \leq L, Q_m (m = 1, 2) \triangleq 2\lceil f_{d_m} N T_s \rceil$ is the number of basis. The CE-BEM coefficients $h_q(l)$ and $g_q(l)$ are assumed as zero-mean, complex

Gaussian random variables with variance $\sigma_{h,q,l}^2$ and $\sigma_{g,q,l}^2$, respectively [14,16,24].

To simplify the notation as well as the following discussion, we assume $f_{d_1} = f_{d_2} = f_d$ and $Q_1 = Q_2 = Q$. We further denote $w_q = 2\pi(q - Q/2)/N$ and define

$$\mathbf{h}_q = [h_q(0), h_q(1), \dots, h_q(L)]^T, \quad (9)$$

$$\mathbf{g}_q = [g_q(0), g_q(1), \dots, g_q(L)]^T, \quad q \in [0, Q]. \quad (10)$$

3.2 CE-BEM approximation accuracy

Currently, the approximation accuracy about CE-BEM to time-varying channel is only shown through simulations with the merit of MSE [16]. Here, we take one step further by deriving the theoretical MSE of the CE-BEM approximation.

Without loss of generality, let us consider the 1st tap of the channel $h(i; 1)$. Define $\hat{\mathbf{h}}_{l=1} = [h(0; 1), \dots, h(N-1; 1)]^T$ and $\bar{\mathbf{h}}_l = [h_0(l), \dots, h_Q(l)]^T$. From (7) we know that the approximation error is

$$\mathbf{e}_1 = \mathbf{A} \bar{\mathbf{h}}_l - \hat{\mathbf{h}}_{l=1}, \quad (11)$$

where

$$\mathbf{A} = \begin{bmatrix} 1 & 1 & \dots & 1 \\ e^{jw_0} & e^{jw_1} & \dots & e^{jw_Q} \\ \vdots & \vdots & \vdots & \vdots \\ e^{j(N-1)w_0} & e^{j(N-1)w_1} & \dots & e^{j(N-1)w_Q} \end{bmatrix}. \quad (12)$$

Assume the singular value decomposition of the matrix \mathbf{A} is $\mathbf{A} = \mathbf{U}_0 \Sigma_0 \mathbf{V}_0^H$ where \mathbf{U}_0 is an $N \times N$ unitary matrix, \mathbf{V}_0 is a $(Q+1) \times (Q+1)$ unitary matrix, and Σ_0 is an $N \times (Q+1)$ matrix with the $(Q+1)$ diagonal entries as a constant c_0 and other entries as zero.^b Let \mathbf{u}_k denotes the k th column vector of the matrix \mathbf{U}_0 .

Lemma 1. *The MSE of the approximation error is*

$$\mathcal{V} = \frac{1}{N} E(\mathbf{e}_1^H \mathbf{e}_1) = \frac{N-Q-1}{N} \sum_{k=Q+2}^N \mathbf{u}_k^H \mathbf{R}_l \mathbf{u}_k, \quad (13)$$

where \mathbf{R}_l is the correlation function of $\hat{\mathbf{h}}_{l=1}$.

Proof. See Appendix 1. □

Lemma 1 gives the theoretical MSE of CE-BEM approximation to time-varying channels, which enables us to obtain the approximation accuracy without simulations.

A brief example about theoretical MSE of CE-BEM approximation to time-varying channels is shown in Figure 3, where the system parameters are taken as $f_{ds} = f_{dr} = 50$ Hz, $f_{d1} = 100$ Hz, $T_s = 100$ μ s, and $N = 200$. For comparison, simulation approximation MSE is also given by averaging 100 trials. It shows that the theoretical MSE (13) agrees with the simulation MSE. It also reveals that the better approximation comes with larger Q , which indicates that CE-BEM is a good choice for the doubly selective channel.

3.3 Problem formulation

Next we apply CE-BEM (7) and (8) into (4) for channel estimation and data detection. Our tasks are (i) estimate the parameters \mathbf{h}_q and \mathbf{g}_q so that the channel $h(i; l)$ and $g(i; l)$ can be recovered for each time index $i \in [0, N - 1]$, or estimate the equivalent channel parameters that still enable the successful data detection as did in [6,25]; (ii) find the optimal training sequence that can minimize the channel estimation error; (iii) recover the data $\mathbf{s}_k, k \in [1, P]$ from the estimated channel.

4 Channel estimation

Let us construct $N \times 1$ vectors \mathbf{r}, \mathbf{y} , and construct $N \times N$ matrices \mathbf{H}, \mathbf{G} from $g(i; l)$ in the following way:

$$\mathbf{r} = [r(0), r(1), \dots, r(N - 1)]^T, \quad (14)$$

$$\mathbf{y} = [y(0), y(1), \dots, y(N - 1)]^T, \quad (15)$$

$$\mathbf{H}_{i,j} = h(i; i - j), \quad \mathbf{G}_{i,j} = g(i; i - j), \quad (16)$$

for $i, j = 1, 2, \dots, N$. We can write (2) and (4) as

$$\mathbf{r} = \mathbf{H}\mathbf{x} + \mathbf{w}_1, \quad (17)$$

$$\mathbf{y} = \alpha\mathbf{G}\mathbf{r} + \mathbf{w}_2 = \alpha\mathbf{G}\mathbf{H}\mathbf{x} + \mathbf{w}, \quad (18)$$

where $\mathbf{w}_i = [w_i(0), w_i(1), \dots, w_i(N - 1)]^T, i = 1, 2$ and $\mathbf{w} = [w(0), w(1), \dots, w(N - 1)]^T$.

4.1 Channel partition

Following the channel partition method in [24], we can split the channel matrix \mathbf{H} into three matrices, namely, $\mathbf{H}_s, \mathbf{H}_b,$ and $\mathbf{H}_{\bar{b}}$, which are shown in Figure 4. Similarly, the channel $\mathbf{H}_k,$ the k th ($1 \leq k \leq P$) part of \mathbf{H} corresponding to the k th sub-block input of $[\mathbf{s}_k, \mathbf{b}_k],$ can also be partitioned into three matrices $\mathbf{H}_k^s, \mathbf{H}_k^b,$ and $\mathbf{H}_k^{\bar{b}}$ (Figure 5). After the separation of these channels, we derive two input-output relationships at the relay node

$$\mathbf{r}_s = \mathbf{H}_s\mathbf{s} + \mathbf{H}_{\bar{b}}\bar{\mathbf{b}} + \mathbf{w}_1^s, \quad (19)$$

$$\mathbf{r}_b = \mathbf{H}_b\mathbf{b} + \mathbf{w}_1^b, \quad (20)$$

where $\mathbf{r}_s = [(\mathbf{r}_1^s)^T, \dots, (\mathbf{r}_p^s)^T]^T, \mathbf{r}_b = [(\mathbf{r}_1^b)^T, \dots, (\mathbf{r}_p^b)^T]^T, \bar{\mathbf{b}}$ contains the first L and the last L entries of \mathbf{b}_k for $1 \leq k \leq P,$ while \mathbf{w}_1^s and \mathbf{w}_1^b denote the corresponding noise vectors.

Repeat the partition process for the channel \mathbf{G} and $\mathbf{G}_k.$ That is, split \mathbf{G} into $\mathbf{G}_s, \mathbf{G}_{\bar{b}},$ and $\mathbf{G}_b,$ while split $\mathbf{G}_k,$ the k th component of $\mathbf{G},$ into $\mathbf{G}_k^s, \mathbf{G}_k^{\bar{b}},$ and $\mathbf{G}_k^b.$ We obtain two input-output relationships at the destination node

$$\mathbf{y}_s = \alpha\mathbf{G}_s\mathbf{r}_s + \alpha\mathbf{G}_{\bar{b}}\bar{\mathbf{r}}_b + \mathbf{w}_2^s, \quad (21)$$

$$\mathbf{y}_b = \alpha\mathbf{G}_b\mathbf{r}_b + \mathbf{w}_2^b, \quad (22)$$

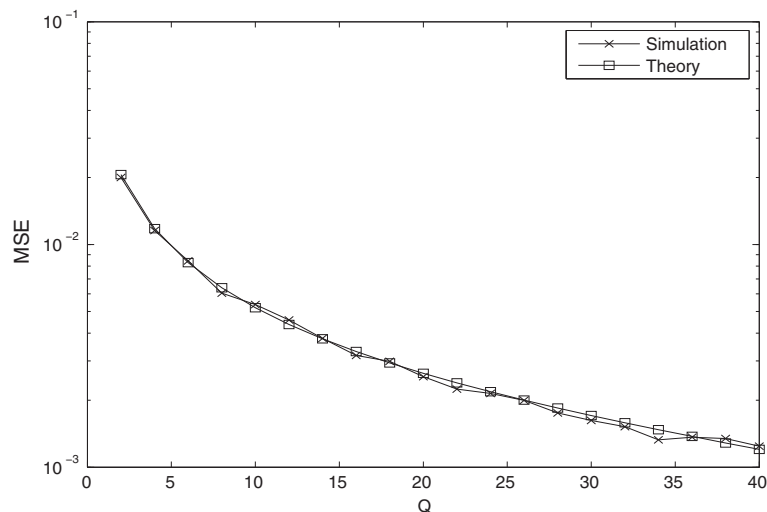
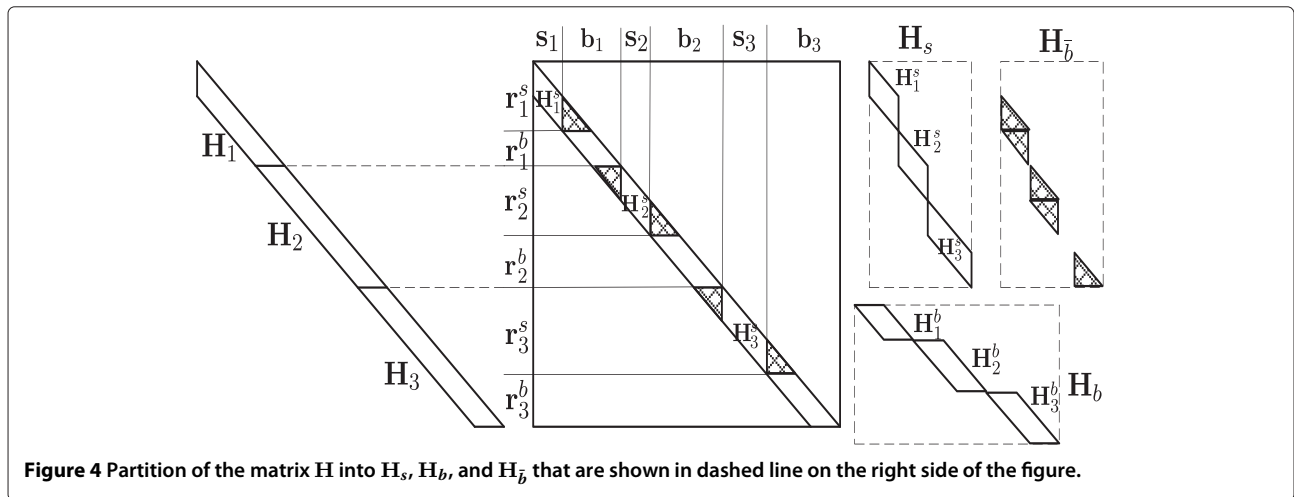


Figure 3 Theoretical and simulation MSE of CE-BEM approximation to the time-varying channel $h(i; 1)$.



where $\mathbf{y}_s = [(\mathbf{y}_1^s)^T, \dots, (\mathbf{y}_p^s)^T]^T$, $\mathbf{y}_b = [(\mathbf{y}_1^b)^T, \dots, (\mathbf{y}_p^b)^T]^T$, $\mathbf{r}_{\bar{b}}$ contains the first L and the last L entries of \mathbf{r}_k^b for all $1 \leq k \leq P$, \mathbf{w}_2^s , and \mathbf{w}_2^b denote the corresponding noise vectors.

Combining (20) and (22) yields

$$\mathbf{y}_b = \alpha \mathbf{G}_b \mathbf{H}_b \mathbf{b} + \underbrace{\alpha \mathbf{G}_b \mathbf{w}_1^b + \mathbf{w}_2^b}_{\mathbf{w}_b} \quad (23)$$

where \mathbf{w}_b is defined as the corresponding item.

It can readily be checked that (23) is equivalent to

$$\mathbf{y}_b = \begin{bmatrix} \mathbf{y}_1^b \\ \vdots \\ \mathbf{y}_p^b \end{bmatrix} = \begin{bmatrix} \alpha \mathbf{G}_1^b \mathbf{H}_1^b \mathbf{b}_1 \\ \vdots \\ \alpha \mathbf{G}_p^b \mathbf{H}_p^b \mathbf{b}_p \end{bmatrix} + \mathbf{w}_b. \quad (24)$$

Note that in (24) \mathbf{H}_k^b is an $(N_{b_k} - L) \times N_{b_k}$ matrix and \mathbf{G}_k^b is an $(N_{b_k} - 2L) \times (N_{b_k} - L)$ matrix. To perform channel estimation, the individual channel matrix \mathbf{G}_k^b should be a valid matrix. Thus, we require $N_{b_k} - 2L > 0$, i.e.,

the training length for the k th subblock should be $N_{b_k} \geq 2L + 1$.

4.2 Estimation algorithm

Let us define $\Lambda_M^{(w_q)} = \text{diag}\{1, e^{jw_q}, \dots, e^{jw_q(M-1)}\}$. For any $(L+1) \times 1$ vector $\mathbf{a} = [a_0, a_1, \dots, a_L]^T$, define an $M \times (M+L)$ Toeplitz matrix as

$$\mathbf{T}_{M+L}^{(\mathbf{a})} = \begin{bmatrix} a_L & \cdots & a_0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & a_L & \cdots & a_0 \end{bmatrix}. \quad (25)$$

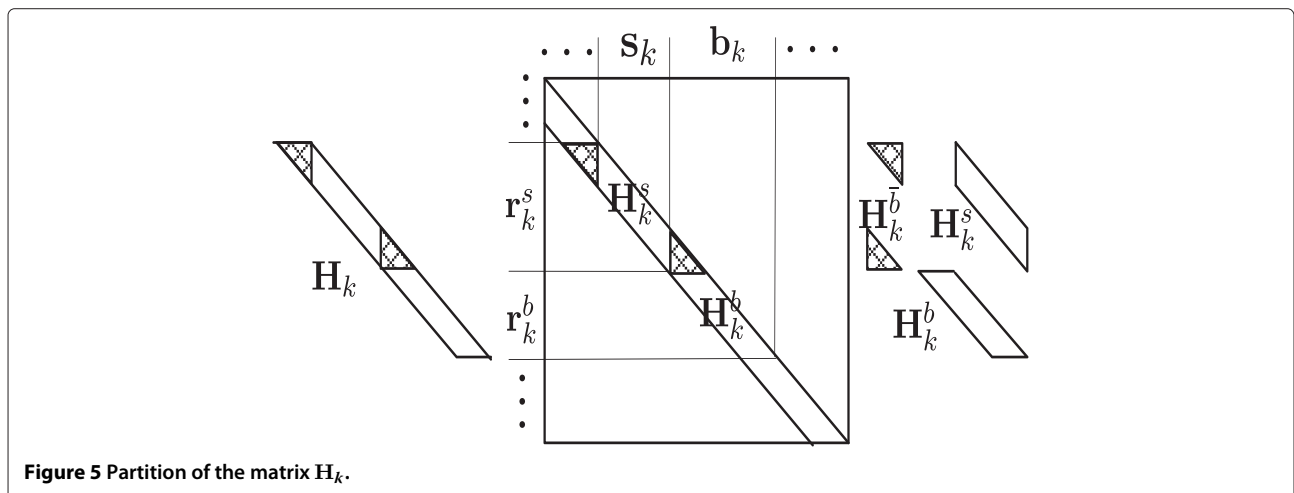
$M+L$ columns

We provide the following two lemmas.

Lemma 2.

$$\mathbf{T}_{M+L}^{(\mathbf{a})} \Lambda_{M+L}^{(w_q)} = \Lambda_M^{(w_q)} \mathbf{T}_{M+L}^{(\boldsymbol{\mu}_a)}, \quad (26)$$

where $\boldsymbol{\mu}_a = [a_0 e^{jw_q L}, a_1 e^{jw_q(L-1)}, \dots, a_L]$.



Proof. Proved from straight calculations due to the special structures of $\Lambda_M^{(w_q)}$ and $\mathbf{T}_{M+L}^{(\mu_a)}$. \square

Lemma 3. For two vectors $\mathbf{a}_i = [a_{i,0}, a_{i,1}, \dots, a_{i,L}]^T, i = 1, 2$, there is

$$\mathbf{T}_{M+L}^{(\mathbf{a}_1)} \mathbf{T}_{M+2L}^{(\mathbf{a}_2)} = \mathbf{T}_{M+2L}^{(\mathbf{a}_1 * \mathbf{a}_2)}, \quad (27)$$

where $*$ denotes linear convolution.

Proof. Note that both $\mathbf{T}_{M+L}^{(\mathbf{a}_1)}$ and $\mathbf{T}_{M+2L}^{(\mathbf{a}_2)}$ are circulant matrix. Proved from straight calculations. \square

According to these definitions and (7), it can readily be checked that

$$\mathbf{H} = \sum_{q=0}^Q \Lambda_N^{(w_q)} \Phi_q, \quad (28)$$

where Φ_q is a lower triangular Toeplitz matrix with the first column $[h_q(0), \dots, h_q(L), 0, \dots, 0]^T$. Furthermore, noticing that $h(i+n; l) = \sum_{q=0}^Q h_q(l) e^{jw_q n} e^{jw_q i}$, we can find

$$\mathbf{H}_k^b = \sum_{q=0}^Q e^{jw_q(N_{s_k} + 2L + \sum_{i=1}^{k-1} N_i)} \Lambda_{N_{b_k}-2L}^{(w_q)} \mathbf{T}_{N_{b_k}}^{(\mathbf{h}_q)}, \quad (29)$$

where $\mathbf{T}_{N_{b_k}}^{(\mathbf{h}_q)}$ is $(N_{b_k} - L) \times N_{b_k}$ Toeplitz matrix as defined in (25).

Similarly, based on (8) and $g(i+n; l) = \sum_{q=0}^Q g_q(l) e^{jw_q n} e^{jw_q i}$ we can obtain

$$\mathbf{G} = \sum_{q=0}^Q \Lambda_N^{(w_q)} \Omega_q, \quad (30)$$

$$\mathbf{G}_k^b = \sum_{q=0}^Q e^{jw_q(N_{s_k} + 2L + \sum_{i=1}^{k-1} N_i)} \Lambda_{N_{b_k}-2L}^{(w_q)} \mathbf{T}_{N_{b_k}-L}^{(\mathbf{g}_q)}, \quad (31)$$

where Ω_q is a lower triangular Toeplitz matrix with the first column $[g_q(0), \dots, g_q(L), 0, \dots, 0]^T$, and $\mathbf{T}_{N_{b_k}-L}^{(\mathbf{g}_q)}$ is an $(N_{b_k} - 2L) \times (N_{b_k} - L)$ Toeplitz matrix as defined in (25).

Combining (29) and (31) gives

$$\begin{aligned} \mathbf{G}_k^b \mathbf{H}_k^b &= \sum_{m=0}^Q e^{jw_m(N_{s_k} + 2L + \sum_{i=1}^{k-1} N_i)} \Lambda_{N_{b_k}-2L}^{(w_m)} \mathbf{T}_{N_{b_k}-L}^{(\mathbf{g}_m)} \\ &\quad \times \sum_{n=0}^Q e^{jw_n(N_{s_k} + 2L + \sum_{i=1}^{k-1} N_i)} \Lambda_{N_{b_k}-L}^{(w_n)} \mathbf{T}_{N_{b_k}}^{(\mathbf{h}_n)} \\ &= \sum_{m=0}^Q \sum_{n=0}^Q \underbrace{\theta_{m,n,k} \Lambda_{N_{b_k}-2L}^{(w_m)} \mathbf{T}_{N_{b_k}-L}^{(\mathbf{g}_m)} \Lambda_{N_{b_k}-L}^{(w_n)} \mathbf{T}_{N_{b_k}}^{(\mathbf{h}_n)}}_{\Xi_{m,n,k}}, \end{aligned} \quad (32)$$

where

$$\theta_{m,n,k} = e^{jw_m(N_{s_k} + 2L + \sum_{i=1}^{k-1} N_i) + jw_n(N_{s_k} + 2L + \sum_{i=1}^{k-1} N_i)}, \quad (33)$$

and $\Xi_{m,n,k}$ is defined as the corresponding item. Using Lemmas 2 and 3, $\Xi_{m,n,k}$ can be simplified as

$$\begin{aligned} \Xi_{m,n,k} &= \Lambda_{N_{b_k}-2L}^{(w_m)} \Lambda_{N_{b_k}-2L}^{(w_n)} \mathbf{T}_{N_{b_k}-L}^{(\mu_{g_m})} \mathbf{T}_{N_{b_k}}^{(\mathbf{h}_n)} \\ &= \Lambda_{N_{b_k}-2L}^{(w_m+w_n)} \mathbf{T}_{N_{b_k}}^{(\lambda_{m,n})}, \end{aligned} \quad (34)$$

where

$$\mu_{g_m} = [g_m(0) e^{jw_n L}, g_m(1) e^{jw_n(L-1)}, \dots, g_m(L)]^T, \quad (35)$$

$$\lambda_{m,n} = \mu_{g_m} * \mathbf{h}_n. \quad (36)$$

Since $\mathbf{T}_{N_{b_k}}^{(\lambda_{m,n})}$ is a Toeplitz matrix, we obtain

$$\begin{aligned} \mathbf{G}_k^b \mathbf{H}_k^b \mathbf{b}_k &= \sum_{m=0}^Q \sum_{n=0}^Q \theta_{m,n,k} \Lambda_{N_{b_k}-2L}^{(w_m+w_n)} \mathbf{T}_{N_{b_k}}^{(\lambda_{m,n})} \mathbf{b}_k \\ &= \sum_{m=0}^Q \sum_{n=0}^Q \theta_{m,n,k} \Lambda_{N_{b_k}-2L}^{(w_m+w_n)} \mathbf{B}_{N_{b_k}}^{(\mathbf{b}_k)} \lambda_{m,n}, \end{aligned} \quad (37)$$

where $\mathbf{B}_{N_{b_k}}^{(\mathbf{b}_k)}$ is defined as

$$\mathbf{B}_{N_{b_k}}^{(\mathbf{b}_k)} = \begin{bmatrix} b_k(2L), & \dots, & b_k(0) \\ b_k(2L+1), & \dots, & b_k(1) \\ \vdots & \vdots & \vdots \\ b_k(N_{b_k}-1), & \dots, & b_k(N_{b_k}-2L-1) \end{bmatrix}. \quad (38)$$

Unfortunately, it remains challenging to estimate $\lambda_{m,n}$ from (37). The number of unknown elements for all $\lambda_{m,n}$ ($m, n \in [0, Q]$) is $(Q+1)^2(2L+1)$. A direct way to estimate all $\lambda_{m,n}$ requires N_b to be no less than $2PL + (Q+1)^2(2L+1)$, which is too large and the transmission efficiency will be reduced. To solve this problem, we choose to estimate other type of channel information that requires smaller training length but at the same time guarantees the data detection.

Let us introduce two variables $\zeta_{q,k}$ and ϖ_q as

$$\begin{aligned} \varpi_q &= w_m + w_n = 2\pi(m+n-Q)/N \\ &= 2\pi(q-Q)/N, \quad m, n \in [0, Q], q \in [0, 2Q], \end{aligned} \quad (39)$$

$$\zeta_{q,k} = e^{j\varpi_q(N_{s_k} + 2L + \sum_{i=1}^{k-1} N_i)}, \quad k \in [1, P]. \quad (40)$$

It can readily be checked that $\theta_{m,n,k} = \zeta_{m+n,k} e^{-jw_n L}$. Then we can combine those items that satisfy $m+n=q$ in (37) and obtain

$$\mathbf{G}_k^b \mathbf{H}_k^b \mathbf{b}_k = \sum_{q=0}^{2Q} \zeta_{q,k} \Lambda_{N_{b_k}-2L}^{(\varpi_q)} \mathbf{B}_{N_{b_k}}^{(\mathbf{b}_k)} \eta_q, \quad (41)$$

where

$$\eta_q = \sum_{m+n=q} e^{-jw_n L} \lambda_{m,n}. \quad (42)$$

Define

$$\eta = [\eta_0^T, \eta_1^T, \dots, \eta_{2Q}^T]^T, \quad (43)$$

as the parameters to be estimated. Substituting (41) into (24) provides a simple model

$$\mathbf{y}_b = \alpha \Psi_b \eta + \mathbf{w}_b, \quad (44)$$

where Ψ_b is defined as

$$\Psi_b = \begin{bmatrix} \zeta_{0,1} \Lambda_{N_{b_1}-2L}^{(\overline{w}_0)} \mathbf{B}_{N_{b_1}}^{(\mathbf{b}_1)}, & \dots, & \zeta_{2Q,1} \Lambda_{N_{b_1}-2L}^{(\overline{w}_{2Q})} \mathbf{B}_{N_{b_1}}^{(\mathbf{b}_1)} \\ \vdots & & \vdots \\ \zeta_{0,P} \Lambda_{N_{b_P}-2L}^{(\overline{w}_0)} \mathbf{B}_{N_{b_P}}^{(\mathbf{b}_P)}, & \dots, & \zeta_{2Q,P} \Lambda_{N_{b_P}-2L}^{(\overline{w}_{2Q})} \mathbf{B}_{N_{b_P}}^{(\mathbf{b}_P)} \end{bmatrix}. \quad (45)$$

Thus, instead of estimating the coefficients \mathbf{h}_q and \mathbf{g}_q , we could estimate another parameter η from

$$\hat{\eta} = \frac{1}{\alpha} (\Psi_b^H \Psi_b)^{-1} \Psi_b^H \mathbf{y}_b. \quad (46)$$

Moreover, $\hat{\eta}_q$ can directly be obtained from $\hat{\eta}$ for each $q \in [0, 2Q]$.

Remark 2. Since η is a vector with $(2Q + 1)(2L + 1)$ entries, N_b should be at least $(2Q + 1)(2L + 1) + 2PL$.

4.3 Data detection

Substituting (19) into (21) yields

$$\mathbf{y}_s = \alpha \mathbf{G}_s \mathbf{H}_s \mathbf{s} + \alpha \mathbf{G}_s \mathbf{H}_b \bar{\mathbf{b}} + \alpha \mathbf{G}_s \mathbf{w}_1^s + \alpha \mathbf{G}_b \mathbf{r}_b + \mathbf{w}_2^s. \quad (47)$$

Note that \mathbf{r}_b in (47) can be further decomposed as

$$\mathbf{r}_b = \mathbf{H}_b \check{\mathbf{b}} + \mathbf{w}_1^b, \quad (48)$$

where \mathbf{H}_b contains the first L and the last L rows of every \mathbf{H}_k^b , $\check{\mathbf{b}}$ contains the first $2L$ and the last $2L$ entries of every \mathbf{b}_k , $1 \leq k \leq P$ and \mathbf{w}_1^b denotes the corresponding noise. Then (47) can be written as

$$\mathbf{y}_s = \alpha \mathbf{G}_s \mathbf{H}_s \mathbf{s} + \alpha \mathbf{G}_s \mathbf{H}_b \bar{\mathbf{b}} + \alpha \mathbf{G}_b \mathbf{H}_b \check{\mathbf{b}} + \mathbf{w}_s, \quad (49)$$

where the combined noise vector $\mathbf{w}_s = \alpha \mathbf{G}_s \mathbf{w}_1^s + \alpha \mathbf{G}_b \mathbf{w}_1^b + \mathbf{w}_2^s$.

Lemma 4. Among all training choices that lead to identical covariance matrix of the channel estimation error, if the training length N_{b_k} is greater than $4L + 1$ and if the training has the first $2L$ and the last $2L$ entries equal to zero, then the interference to the data detection is minimized.

Proof. See Appendix 2. \square

Following Lemma 4, we can simplify (49) as

$$\mathbf{y}_s = \alpha \mathbf{G}_s \mathbf{H}_s \mathbf{s} + \mathbf{w}_s, \quad (50)$$

which is equivalent to

$$\mathbf{y}_s = \begin{bmatrix} \mathbf{y}_1^s \\ \vdots \\ \mathbf{y}_P^s \end{bmatrix} = \begin{bmatrix} \alpha \mathbf{G}_1^s \mathbf{H}_1^s \mathbf{s}_1 \\ \vdots \\ \alpha \mathbf{G}_P^s \mathbf{H}_1^s \mathbf{s}_P \end{bmatrix} + \mathbf{w}_s. \quad (51)$$

Define $\mathbf{U}_M^{(\mathbf{h}_q)}$ as a Toeplitz matrix generated by the vector \mathbf{h}_q in the following way:

$$\mathbf{U}_M^{(\mathbf{h}_q)} = \underbrace{\begin{bmatrix} h_q(0), & \dots, & 0 \\ \vdots & \ddots & \vdots \\ h_q(L), & \ddots, & h_q(0) \\ \vdots & \ddots & \vdots \\ 0 & \dots & h_q(L) \end{bmatrix}}_{M \text{ columns}}. \quad (52)$$

We have the following lemmas.

Lemma 5.

$$\mathbf{U}_M^{(\mathbf{h}_q)} \Lambda_M^{(w_q)} = e^{-jw_q L} \Lambda_{M+L}^{(w_q)} \mathbf{U}_M^{(\mu_{h_q})}, \quad (53)$$

where $\mu_{h_q} = [h_q(0)e^{jw_q L}, h_q(1)e^{jw_q(L-1)}, \dots, h_q(L)]^T$.

Proof. Proved from straight calculation. \square

Lemma 6.

$$\mathbf{U}_{M+L}^{(\mathbf{g}_q)} \mathbf{U}_M^{(\mathbf{h}_q)} = \mathbf{U}_M^{(\mathbf{g}_q * \mathbf{h}_q)} \quad (54)$$

Proof. Proved from straight calculation. \square

According to (7) and (8), we obtain

$$\mathbf{G}_k^s = \sum_{q=0}^Q e^{jw_q \sum_{i=1}^{k-1} N_i} \Lambda_{N_{s_k}+2L}^{(w_q)} \mathbf{U}_{N_{s_k}+L}^{(\mathbf{g}_q)}, \quad (55)$$

$$\mathbf{H}_k^s = \sum_{q=0}^Q e^{jw_q \sum_{i=1}^{k-1} N_i} \Lambda_{N_{s_k}+L}^{(w_q)} \mathbf{U}_{N_{s_k}}^{(\mathbf{h}_q)}. \quad (56)$$

Next it can be verified that

$$\mathbf{G}_k^s \mathbf{H}_k^s = \sum_{m=0}^Q \sum_{n=0}^Q \phi_{m,n,k} \Lambda_{N_{s_k}+2L}^{(w_m)} \mathbf{U}_{N_{s_k}+L}^{(\mathbf{g}_m)} \Lambda_{N_{s_k}+L}^{(w_n)} \mathbf{U}_{N_{s_k}}^{(\mathbf{h}_n)} \quad (57)$$

where $\phi_{m,n,k} = e^{j(w_m+w_n) \sum_{i=1}^{k-1} N_i}$. Using Lemmas 5 and 6, it can be derived that

$$\begin{aligned} \mathbf{U}_{N_{s_k}+L}^{(\mathbf{g}_m)} \Lambda_{N_{s_k}+L}^{(w_n)} \mathbf{U}_{N_{s_k}}^{(\mathbf{h}_n)} &= e^{-jw_n L} \Lambda_{N_{s_k}+2L}^{(w_n)} \mathbf{U}_{N_{s_k}+L}^{(\mu_{g_m})} \mathbf{U}_{N_{s_k}}^{(\mathbf{h}_n)} \\ &= e^{-jw_n L} \Lambda_{N_{s_k}+2L}^{(w_n)} \mathbf{U}_{N_{s_k}}^{(\lambda_{m,n})}, \end{aligned} \quad (58)$$

where μ_{g_m} and $\lambda_{m,n}$ are defined in (35) and (36), respectively. Substituting (58) into (57), we can obtain

$$\begin{aligned} \mathbf{G}_k^s \mathbf{H}_k^s &= \sum_{m=0}^Q \sum_{n=0}^Q \phi_{m,n,k} e^{-jw_n L} \mathbf{\Lambda}_{N_{s_k}+2L}^{(w_m+w_n)} \mathbf{U}_{N_{s_k}}^{(\lambda_{m,n})} \\ &= \sum_{q=0}^{2Q} e^{j\varpi_q} \sum_{i=1}^{k-1} N_i \mathbf{\Lambda}_{N_{s_k}+2L}^{(\varpi_q)} \mathbf{U}_{N_{s_k}}^{(\eta_q)}. \end{aligned} \quad (59)$$

Clearly, given the estimates of η_q , $\mathbf{G}_k^s \mathbf{H}_k^s$ can be reconstructed from (59). Hence, the data \mathbf{s}_k can be detected with the reconstructed channel information $\mathbf{G}_k^s \mathbf{H}_k^s$.

4.4 Training sequence design

The estimation error of η can be expressed as

$$\mathbf{e} = \hat{\eta} - \eta = (\Psi_b^H \Psi_b)^{-1} \Psi_b^H \mathbf{w}_b. \quad (60)$$

The correlation matrix of \mathbf{w}_b is then computed from (31) as

$$\mathbf{R}_{\mathbf{w}_b} = E(\mathbf{w}_b \mathbf{w}_b^H) = \left(\sigma_{w_2}^2 \sum_{q=0}^Q \sum_{l=0}^L |g_q(l)|^2 + \sigma_{w_1}^2 \right) \mathbf{I}_{N_b-2PL}. \quad (61)$$

Thus, the MSE of \mathbf{e} is

$$\sigma_e^2 = \text{tr}(E(\mathbf{e}\mathbf{e}^H)) = C_e \text{tr}(\Psi_b^H \Psi_b)^{-1} \quad (62)$$

where $C_e = \left(\sigma_{w_2}^2 \sum_{q=0}^Q \sum_{l=0}^L |g_q(l)|^2 + \sigma_{w_1}^2 \right) / \alpha^2$.

According to ([26], Appendix A), we know that σ_e^2 in (62) is lower bounded as follows:

$$C_e \text{tr}(\Psi_b^H \Psi_b)^{-1} \geq \sum_m \frac{C_e}{[\Psi_b^H \Psi_b]_{m,m}}, \quad (63)$$

where the equality holds if and only if $(\Psi_b^H \Psi_b)$ is a diagonal matrix. We then need to design the training sequence that can diagonalize $(\Psi_b^H \Psi_b)$.

Based on the definition of Ψ_b (45), the optimal training sequence that can minimize the σ_e^2 requires the following conditions to be satisfied:

$$\sum_{k=1}^P (\mathbf{B}_{N_{b_k}}^{(\mathbf{b}_k)})^H \mathbf{B}_{N_{b_k}}^{(\mathbf{b}_k)} = \mathcal{P}_b \mathbf{I}_{2L+1}, \quad (64)$$

$$\sum_{k=1}^P (\mathbf{B}_{N_{b_k}}^{(\mathbf{b}_k)})^H \mathbf{\Lambda}_{N_{b_k}-2L}^{(-\varpi_{q_1})} \zeta_{q_1,k}^H \zeta_{q_2,k} \mathbf{\Lambda}_{N_{b_k}-2L}^{(\varpi_{q_2})} \mathbf{B}_{N_{b_k}}^{(\mathbf{b}_k)} = \mathbf{0}_{2L+1}, \quad (65)$$

$$\forall q_1 \neq q_2, q_1, q_2 \in [0, 2Q]$$

where \mathcal{P}_b is the power allocated to the training sequence.

Let us first focus on (64). Observing the structure of $\mathbf{B}_{N_{b_k}}^{(\mathbf{b}_k)}$, we know that (64) can be fulfilled if the following conditions are satisfied:

$$(C1): N_{b_k} = 4L + 1, \quad \forall k \in [1, P], \quad (66)$$

$$(C2): \mathbf{b}_k = \sqrt{\mathcal{P}_b/P} [0, \dots, 0, 1, 0, \dots, 0]^T. \quad (67)$$

With conditions (C1) and (C2), we can further simplify (65) as

$$\begin{aligned} &\frac{\mathcal{P}_b}{P} \sum_{k=1}^P \mathbf{\Lambda}_{2L+1}^{(-\varpi_{q_1})} \zeta_{q_1,k}^H \zeta_{q_2,k} \mathbf{\Lambda}_{2L+1}^{(\varpi_{q_2})} \\ &= \frac{\mathcal{P}_b}{P} \sum_{k=1}^P e^{j\frac{2\pi}{N}(q_2-q_1)(N_{s_k}+2L+\sum_{i=1}^{k-1} N_i)} \mathbf{\Lambda}_{2L+1}^{(\varpi_{q_2}-\varpi_{q_1})} \\ &= \mathbf{0}_{2L+1}, \quad \forall q_1 \neq q_2, q_1, q_2 \in [0, 2Q]. \end{aligned} \quad (68)$$

It can readily be checked that the sufficient conditions to achieve (68) are

$$(C3): N = P(N_{s_k} + 4L + 1), \quad N_{s_k} = N_s/P, \quad \forall k \in [1, P]. \quad (69)$$

Conditions (C1), (C2), and (C3) imply that the equal-spaced and equal-powered training sequence can minimize the estimation MSE. This coincides with the optimal training sequence design in the traditional point-to-point channel [24].

Remark 3. Note that Equation (68) cannot hold when $P = 1$. It indicates that the traditional transmission frame [1,2], i.e., sending and receiving the continuous data sequence only once, is not optimal in minimizing the estimation MSE.

4.5 Block parameters

Remark 2 shows $N_b \geq 2PL + (2Q + 1)(2L + 1)$ and the optimal training design requires $N_b = P(4L + 1)$ to minimize the mean-square channel estimation error. Thus, we know that $P \geq (2Q + 1)$ and $N \geq (N_{s_k} + 4L + 1)(2Q + 1)$.

Suppose a 3-tap channel and $N_{s_k} = 4L + 1 = 9$, we can obtain

$$2Q + 1 = 4[f_d T_s N] + 1 \leq N/18. \quad (70)$$

It can be found

$$f_d T_s \leq \frac{1}{72} + \frac{1}{4N} \approx 0.0139 + \frac{1}{4N}. \quad (71)$$

This is the maximum normalized Doppler shift that our estimation scheme can handle. With the following parameters

- carrier frequency $f_c = 900$ MHz and thus the wavelength $\lambda = 1/3$ m,

- data rate 20 Kbps and thus the symbol period $T_s = 50 \mu s$,

we can compute from $f_d = V/\lambda$ that the mobile speed V should not exceed 66 m/s, which can be satisfied in most application.

5 Simulation results

In order to evaluate the inherent performance of our algorithms, the doubly selective channels are generated directly from the CE-BEM (7) and (8). The same approach has been adopted in many other articles when testing the performance of channel estimation [24,27]. However, the real channel will be used in date detection.

We assume that carrier frequency $f_c = 900$ MHz, one symbol period $T_s = 50 \mu s$ and the maximum mobility speed is 90 km/h. Thus, the maximum Doppler shift is $f_d = 75$ Hz and $f_d T_s = 3.75 \times 10^{-3}$. Suppose one block contains 360 symbols, i.e., $N = 360$. Then $Q = 2 \lceil N f_d T_s \rceil = 4$. We also assume that both doubly selective channels $h(i; l)$ and $g(i; l)$ has three taps, i.e., $L = 2$. Thus, we know that $P \geq (2Q + 1) = 9$ and $N_b \geq P(4L + 1) = 81$. The variance of each tap for channel $h(i; l)$ is $\sigma_{h,l}^2 = \sum_{q=0}^Q \sigma_{h,q,l}^2 = e^{-l/10}$ and that for channel $g(i; l)$ is $\sigma_{g,l}^2 = \sum_{q=0}^Q \sigma_{g,q,l}^2 = e^{-l/10}$. The variance of the noise is taken as $\sigma_{w_1}^2 = \sigma_{w_2}^2 = 1$. The SNR is defined as the ratio of symbol power to the noise power, i.e., E_s/N_0 . BPSK constellation is utilized for both training and data symbols. One thousand Monte-Carlo trials are used for the averaging.

First we set the total number of trainings $N_b = 120$ and adopt three types of training: equi-powered and equi-spaced (our optimal design); equi-powered but with random length; equi-spaced but with random power. For

performance comparison, the total power for each types of trainings is the same. For each type of training, we find the MSE of our specially defined channel η . The estimation MSEs versus SNR for each type are plotted in Figure 6. The lower bound of σ_e^2 (63) is also plotted for comparison. It can be seen that the equi-spaced equi-powered training achieves the minimum estimation MSE among all the three trainings and its MSE almost approaches the lower bound in (63).

Next we use the estimated channel $\hat{\eta}$ to perform data detection. Define bit error rate (BER) as the ratio of number of successfully decoded data symbols over N_s the number of transmitted data symbols. The BER versus SNR is plotted in Figure 7. The BER curve in the case of perfectly known channel η is also plotted for comparison. It can be seen that our detection method works well, and at high SNR our BER curve approaches that of the ideal case when the channel is perfectly known at the receiver.

We also examine the performance of the suggested estimation and detection methods under real channel situations. That is, the channel are generated according to (5) and (6), and next our suggested estimation and detection methods are utilized. Three different number of bases Q are chosen as 4, 6, and 8, respectively, and hence the corresponding number of data symbols N_s is 279, 243, and 207. The BER versus SNR is plotted in Figure 8. For comparison, the BER curve under perfect channel knowledge at the receiver is also displayed. Clearly, the proposed methods yield effective data detection. An error floor is observed in the high SNR region due to the mismatch between the BEM model and the real channels. Obviously, the place where the floor begins could be improved by increasing the number Q .

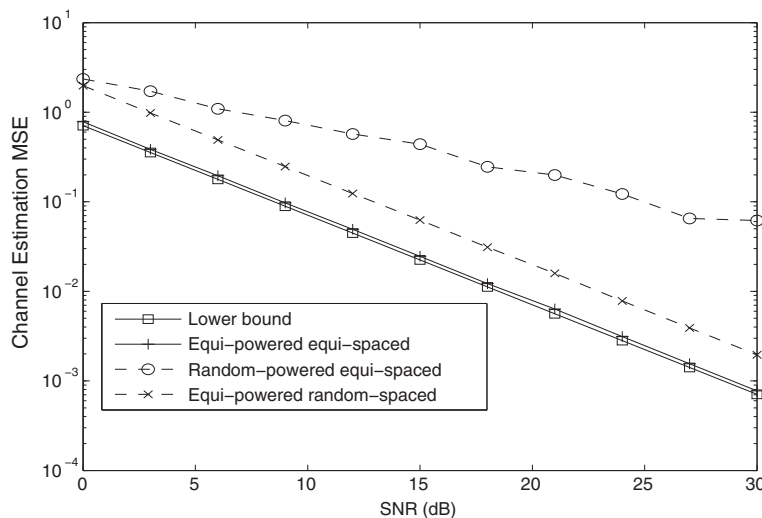


Figure 6 Channel MSE versus the SNR.

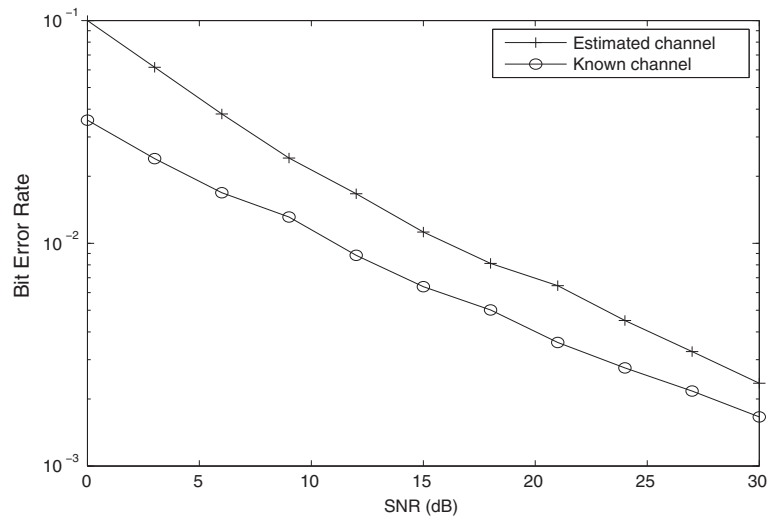


Figure 7 BER versus the SNR: CE-BEM channel.

In the last example, we choose three different number of subblocks P as $2Q+1$, $2Q+2$, and $2Q+4$, respectively, and the space left for data transmission is $N_{s_k} = N - P(4L + 1) = 279, 270$, and 252 , respectively. Define the transmission efficiency as the ratio of the number of successfully decoded data symbols over total number of symbols, i.e., $N_s \times \text{BER}/N$. We run the simulation process as SNR ranges from -10 to 30 dB. The transmission efficiency at different SNR for each P is plotted in Figure 9. It is shown that when the number of subblocks P equals $2Q + 1$, the best transmission efficiency is achieved at all SNR. It can be explained that when P increase by one unit, the data

loss will be $4L + 1$, which cannot be compensated even if channel estimation performance can be improved by larger P .

6 Conclusion

In this article, doubly selective channel estimation was considered for AF-based relay networks. Based on the CE-BEM, we designed an efficient method to estimate the channel coefficients and detect data symbols. The optimal training sequence that can minimize the estimation MSE was also derived. Finally, extensive numerical results are provided to corroborate the proposed studies.

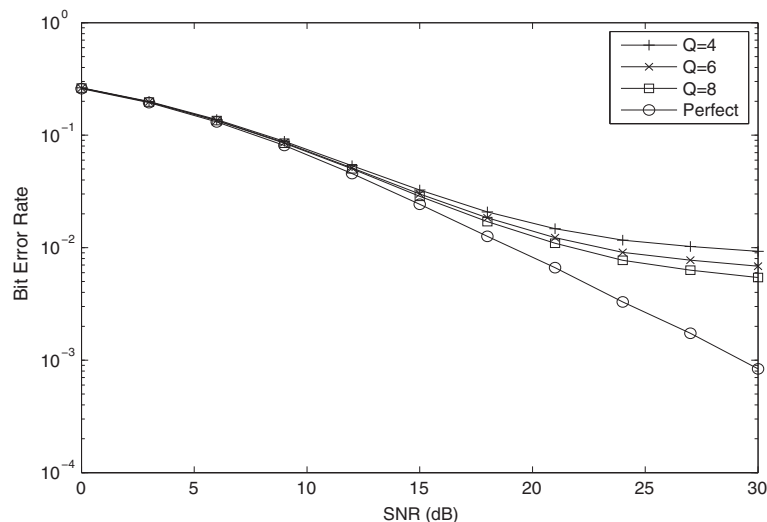


Figure 8 BER versus the SNR: real channel.

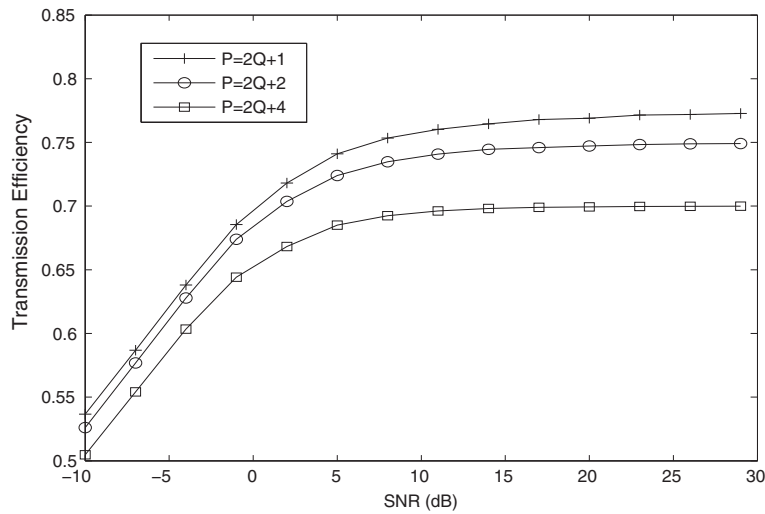


Figure 9 Transmission efficiency versus SNR.

Appendix 1

Proof of Lemma 1

Assuming that $\hat{\mathbf{h}}_{l=1}$ is perfectly known, the best CE-BEM coefficients $\bar{\mathbf{h}}_l$ are obtained from the LS criterion as

$$\bar{\mathbf{h}}_l = (\mathbf{A}^H \mathbf{A})^{-1} \mathbf{A}^H \hat{\mathbf{h}}_{l=1}. \quad (72)$$

Thus, the approximation error can be simplified as

$$\mathbf{e}_1 = (\mathbf{A}(\mathbf{A}^H \mathbf{A})^{-1} \mathbf{A}^H - \mathbf{I}_N) \hat{\mathbf{h}}_{l=1} = \mathbf{U}_0 \mathbf{M} \mathbf{U}_0^H \hat{\mathbf{h}}_{l=1}, \quad (73)$$

where \mathbf{M} is defined as

$$\mathbf{M} = \begin{bmatrix} \mathbf{0}_{Q+1} & \mathbf{0}_{(Q+1) \times (N-Q-1)} \\ \mathbf{0}_{(N-Q-1) \times (Q+1)} & -\mathbf{I}_{N-Q-1} \end{bmatrix}. \quad (74)$$

The MSE of the approximation error \mathbf{e}_1 is

$$\begin{aligned} \mathcal{V} &= \frac{1}{N} E \left(\hat{\mathbf{h}}_{l=1}^H \mathbf{U}_0 \mathbf{M}^2 \mathbf{U}_0^H \hat{\mathbf{h}}_{l=1} \right) \\ &= \frac{1}{N} \sum_{k=Q+2}^N E \left(\hat{\mathbf{h}}_{l=1}^H \mathbf{u}_k \mathbf{u}_k^H \hat{\mathbf{h}}_{l=1} \right). \end{aligned} \quad (75)$$

It can readily be checked that

$$\begin{aligned} E \left(\hat{\mathbf{h}}_{l=1}^H \mathbf{u}_k \mathbf{u}_k^H \hat{\mathbf{h}}_{l=1} \right) &= E \left(\text{tr} \left(\hat{\mathbf{h}}_{l=1}^H \mathbf{u}_k \mathbf{u}_k^H \hat{\mathbf{h}}_{l=1} \right) \right) \\ &= E \left(\mathbf{u}_k^H \hat{\mathbf{h}}_{l=1} \hat{\mathbf{h}}_{l=1}^H \mathbf{u}_k \right) = \mathbf{u}_k^H \mathbf{R}_h \mathbf{u}_k. \end{aligned} \quad (76)$$

Substituting (76) into (75) yields (13).

Appendix 2

Proof of Lemma 4

The interference during the data detection can be expressed as

$$\mathbf{v} = \alpha \Delta(\mathbf{G}_s \mathbf{H}_s) \mathbf{s} + \mathbf{w}_s + \alpha \Delta(\mathbf{G}_s \mathbf{H}_{\bar{b}}) \bar{\mathbf{b}} + \alpha \Delta(\mathbf{G}_{\bar{b}} \mathbf{H}_{\bar{b}}) \check{\mathbf{b}}, \quad (77)$$

where $\Delta(\cdot)$ denotes the estimation error of the inside item. The correlation function of the interference \mathbf{v} is given by

$$\begin{aligned} \mathbf{R}_v &= \alpha^2 \mathcal{P}_s E(\Delta(\mathbf{G}_s \mathbf{H}_s) \Delta(\mathbf{G}_s \mathbf{H}_s)^H) + E(\mathbf{w}_s \mathbf{w}_s^H) \\ &\quad + \alpha^2 E(\Delta(\mathbf{G}_s \mathbf{H}_{\bar{b}}) \bar{\mathbf{b}} \bar{\mathbf{b}}^H \Delta(\mathbf{G}_s \mathbf{H}_{\bar{b}})^H) \\ &\quad + \alpha^2 E(\Delta(\mathbf{G}_{\bar{b}} \mathbf{H}_{\bar{b}}) \check{\mathbf{b}} \check{\mathbf{b}}^H \Delta(\mathbf{G}_{\bar{b}} \mathbf{H}_{\bar{b}})^H), \end{aligned} \quad (78)$$

where \mathcal{P}_s is the power allocated to the data sequence. We need to find the training scheme that can minimize the trace of \mathbf{R}_v .

Suppose there are two training schemes with identical $E((\boldsymbol{\eta} - \hat{\boldsymbol{\eta}})(\boldsymbol{\eta} - \hat{\boldsymbol{\eta}})^H)$. Thus, the first and the second items in (78) are the same for both training schemes. Clearly, the trace of the two items for both training schemes is the same. If the training scheme has the first $2L$ and the last $2L$ entries equal to zero, the third and fourth item in (78) will become zero; if the training scheme does not has such condition, then it cannot null these two semi-definite items.

Endnotes

^aThere may exists a switching time at the relay node \mathbb{R} , which results in a delay Δ in the second retransmission phase. However, our model can be well adapted by setting $g(i + \Delta; l) = g(i_0; l)$ where i_0 is the new starting point for the following CE-BEM approximation.

^bIt is because that $e^{j2\pi Q/(2N)} \mathbf{A}$ can be considered as part of a discrete Fourier transform matrix.

Competing interests

The authors declare that they have no competing interests.

Acknowledgements

We would like to thank the anonymous reviewers for their helpful and constructive comments, which significantly improve this article. This study was supported by the National Natural Science Foundation of China (Grant Nos. 61201202 and 60971124), by the Key grant Project of Chinese Ministry of

education (No. 313006), by Tsinghua-Tencent Joint Laboratory for Internet Innovation Technology, and by the State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University (No. RCS2009ZT006).

Author details

¹School of Computer and Information Technology, Beijing Jiaotong University, Beijing 100044, China. ²State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, Beijing 100044, China. ³Department of Automation, Tsinghua National Laboratory for Information Science and Technology, Tsinghua University, Beijing 100084, China. ⁴Department of Electrical and Computer Engineering, University of Alberta, Edmonton, AB T6G 2V4, Canada.

Received: 5 December 2011 Accepted: 1 October 2012

Published: 28 November 2012

References

1. J Laneman, G Wornell, Distributed space time block coded protocols for exploiting cooperative diversity in wireless networks. *IEEE Trans. Inf. Theory*. **49**(10), 2415–2425 (2003)
2. J Laneman, D Tse, G Wornell, Cooperative diversity in wireless networks: efficient protocols and outage behavior. *IEEE Trans. Inf. Theory*. **50**(12), 3062–3080 (2004)
3. J Boyer, DD Falconer, H Yanikomeroglu, Multihop diversity in wireless relaying channels. *IEEE Trans. Commun.* **52**(9), 1820–1830 (2004)
4. Y Yang, H Hu, J Xu, G Mao, Relay technologies for WiMax and LTE-advanced mobile systems. *IEEE Commun. Mag.* **47**(10), 100–105 (2009)
5. CS Patel, GL Stuber, Channel estimation for amplify and forward relay based cooperation diversity systems. *IEEE Trans. Wirel. Commun.* **6**(6), 3348–3356 (2007)
6. F Gao, T Cui, A Nallanathan, On channel estimation and optimal training design for amplify and forward relay network. *IEEE Trans. Wirel. Commun.* **7**(5), 1907–1916 (2008)
7. AS Lalos, AA Rontogiannis, K Berberidis, Frequency domain channel estimation for cooperative communication networks. *IEEE Trans. Signal Process.* **58**(6), 3400–3405 (2010)
8. MG Song, D Kim, GH Im, Recursive channel estimation method for OFDM-based cooperative systems. *IEEE Commun. Lett.* **14**(11), 1029–1031 (2010)
9. WC Jakes, *Microwave Mobile Communications*. (Wiley, New York, 1974)
10. G Wang, F Gao, W Chen, C Tellambura, Channel estimation and training design for two-way relay networks in time-selective fading environment. *IEEE Trans. Wirel. Commun.* **10**(8), 2681–2691 (2011)
11. L He, Y-C Wu, S Ma, T-S Ng, HV Poor, Superimposed training based channel estimation and data detection for OFDM amplify-and-forward cooperative systems under high mobility. *IEEE Trans. Signal Process.* **60**(1), 274–284 (2012)
12. L Canonne-Velasquez, H Nguyen-Le, T Le-Ngoc, in *Proc. IEEE ICC*. Doubly selective cascaded channel estimation in amplify-and-forward relay systems, (Kyoto, Japan, 5), pp. 1–5
13. L Tong, BM Sadler, M Dong, Pilot-assisted wireless transmissions. *IEEE Signal Process. Mag.* **21**(6), 12–25 (2004)
14. GB Giannakis, C Tepedelenlioğlu, Basis expansion models and diversity techniques for blind identification and equalization of time-varying channels. *Proc. IEEE*. **86**, 1969–1986 (1998)
15. M Dong, L Tong, B Sadler, Optimal insertion of pilot symbols for transmissions over time-varying flat fading channels. *IEEE Trans. Signal Process.* **52**(5), 1403–1418 (2004)
16. X Ma, GB Giannakis, Maximum-diversity transmissions over doubly-selective wireless channels. *IEEE Trans. Inf. Theory*. **49**, 1832–1840 (2003)
17. DK Borah, BD Hart, Frequency-selective fading channel estimation with a polynomial time-varying channel model. *IEEE Trans. Commun.* **47**, 862–873 (1999)
18. M Martone, Wavelet-based separating kernels for sequence estimation with unknown rapidly time-varying channels. *IEEE Commun. Lett.* **3**, 78–80 (1999)
19. D Slepian, Prolate spheroidal wave functions, Fourier analysis, and uncertainty-V: the discrete case. *Bell Syst. Technol. J.* **57**(5), 1371–1430 (1978)
20. S He, J Tugnait, On doubly selective channel estimation using superimposed training and discrete prolate spheroidal sequences. *IEEE Trans. Signal Process.* **56**, 3214–3228 (2008)
21. JK Tugnait, S He, H Kim, Doubly selective channel estimation using exponential basis models and subblock tracking. *IEEE Trans. Signal Process.* **58**(3), 1275–1289 (2010)
22. CS Patel, GL Stuber, TG Pratt, Statistical properties of amplify and forward relay fading channels. *IEEE Trans. Veh. Technol.* **55**(1), 1–9 (2006)
23. AS Akki, F Haber, A statistical model for mobile-to-mobile land communication channel. *IEEE Trans. Veh. Technol.* **VT-35**(1), 2–7 (1986)
24. X Ma, GB Giannakis, S Ohno, Optimal training for block transmissions over doubly selective wireless fading channels. *IEEE Trans. Signal Process.* **41**(5), 1351–1366 (2003)
25. G Wang, F Gao, Y-C Wu, C Tellambura, Joint carrier frequency offset and channel estimation for two-way relay networks. *IEEE Trans. Wirel. Commun.* **10**(2), 456–465 (2011)
26. S Ohno, GB Giannakis, Capacity maximizing MMSE-optimal pilots for wireless OFDM over frequency-selective block Rayleigh-fading channels. *IEEE Trans. Inf. Theory*. **50**(9), 2138–2145 (2004)
27. T Zemen, CF Mecklenbräuker, Time-variant channel estimation using discrete prolate spheroidal sequences. *IEEE Trans. Signal Process.* **53**, 3597–3607 (2005)

doi:10.1186/1687-6180-2012-251

Cite this article as: Wang et al.: Doubly selective channel estimation for amplify-and-forward relay networks. *EURASIP Journal on Advances in Signal Processing* 2012 **2012**:251.

Submit your manuscript to a SpringerOpen® journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Immediate publication on acceptance
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► springeropen.com