Channel Estimation and Training Design for Two-Way Relay Networks in Time-Selective Fading Environments

Gongpu Wang, Feifei Gao, Wen Chen, and Chintha Tellambura, *Fellow, IEEE*

***Abstract*—In this paper, channel estimation and training sequence design are considered for *amplify-and-forward* (AF)-based two-way relay networks (TWRNs) in a time-selective fading environment. A new complex-exponential basis expansion model (CE-BEM) is proposed to represent the mobile-to-mobile time-varying channels. To estimate such channels, a novel pilot symbol-aided transmission scheme is developed such that a low complex linear approach can estimate the BEM coefficients of the convoluted channels. More essentially, two algorithms are designed to extract the BEM coefficients of the individual channels. The optimal training parameters, including the number of the pilot symbols, the placement of the pilot symbols, and the power allocation to the pilot symbols, are derived by minimizing the channel mean-square error (MSE). The selections of the system parameters are thoroughly discussed in order to guide practical system design. Finally, extensive numerical results are provided to corroborate the proposed studies.**

***Index Terms*—Channel estimation, optimal training design, time-varying channel, two-way relay network, basis expansion model.**

I. Introduction

THE two-way relay network (TWRN) concept [1], [2] has attracted much attention recently due to its enhanced spectral efficiency compared to the one-way relay network (OWRN) [3]. In a TWRN, two source nodes simultaneously send their signals to the relay node, which retransmits to both source nodes after a "network coding"-like process [4]. By removing the self-signal component, each source node obtains the other source node's information.

Manuscript received August 5, 2010; revised April 5, 2011; accepted April 20, 2011. The associate editor coordinating the review of this paper and approving it for publication was G. Li.

G. Wang and C. Tellambura are with the Department of Electrical and Computer Engineering, University of Alberta, Edmonton, AB, Canada, T6G 2V4 (e-mail: {gongpu, chintha}@ece.ualberta.ca).

F. Gao is with the Department of Automation, Tsinghua University, Beijing 100084, China, and is also with the School of Engineering and Science, Jacobs University, Germany (e-mail: feifeigao@ieee.org).

W. Chen is with the Department of Electronic Engineering, Shanghai Jiao Tong University, Shanghai 200240, the SKL for Mobile Communications, Southeast University, and the SKL for ISN, Xidian University, P.R. China (e-mail: wenchen@sjtu.edu.cn).

The work of G. Wang was supported in part by the China Scholarship Council.

The work of F. Gao was supported in part by the German Research Foundation (DFG) under Grant GA 1654/1-1.

This work of W. Chen was supported by NSF China #60972031, by SEU SKL project #W200907, by ISN project #ISN11-01, and by National 973 project #2009CB824900.

Digital Object Identifier 10.1109/TWC.2011.060711.101407

Like any other wireless communication system, a TWRN performs better with better channel estimates. Channel estimation for TWRN flat fading channels was first studied in [5]. Shortly after, the frequency selective case was treated in [6] for an orthogonal frequency division multiplexing (OFDM)-based TWRN. Both these works revealed that channel estimation for TWRNs is quite different from that for the conventional point-to-point wireless systems [7], or even the OWRNs [8]. There are more than one channel in a TWRN and usually the source nodes can only estimate the convoluted channel unless special methods are taken by the relay node [9]. Furthermore, to facilitate channel estimation and data detection, the relay node has to change some received symbols or add some symbols before forwarding to the source nodes [6],[10].

In these previous works [5], [6], [9], [10], only the slow-fading or the block-fading scenarios are considered. However, a TWRN is more susceptible to the changing environment because the relay and the two sources can all be mobile and the relative motion between any two nodes doubles the Doppler spread [11]. This fact places additional demands on the estimation of time-varying channels, which are usually tracked by using periodic training signals, also known as pilot symbol aided modulation (PSAM) [12]. To the best of our knowledge, time-varying channel estimation in TWRNs has not yet been reported. The need for such techniques motivates our current work.

Time-varying channels are typically represented in two ways: by using the Gauss-Markov model [14], which tracks channel variation through symbol-by-symbol updating, and by using the basis expansion model (BEM) [15], which decomposes the channel into the superposition of the time-varying basis functions weighted by time-invariant coefficients. Using the Gauss-Markov channel model for a flat-fading time-varying channel, the channel estimation methods [16] and [17] minimize the mean square error (MSE) and maximize the data throughput criteria, respectively. The optimal training design for doubly selective channels based on the BEM [18]–[20] adopted the same criteria. The equivalence between these two criteria was pointed out in [18].

In this paper, we address the problem of channel estimation and training design for time-varying TWRN channels. We adopt the complex-exponential BEM (CE-BEM) [21], which represents the time-varying channels by a finite set of parameters and Fourier bases. To handle the special features of the TWRN, we propose a new data frame structure, which enables

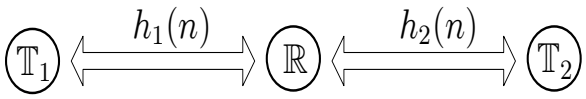


Fig. 1. A two-way relay network over time-selective flat-fading channels.

the periodical reception of pilot symbols at the source nodes. The conventional TWRN transmission structure [5], [6] is a special case of our proposed structure. To reduce the estimation complexity, the interval between the pilot sending and pilot receiving is fixed, and the resultant BEM coefficients of the cascaded channels can be estimated linearly. The optimal training parameters, including the number of pilot symbols, the placement of pilot symbols, and the power of pilots symbols, are derived by minimizing the channel MSE criterion. Two algorithms are then designed to recover the BEM coefficients of individual channel coefficients. An iterative method is also used to refine the estimates. We also provide a thorough discussion of the system parameter selection and reveal many interesting results. Finally, simulation results are provided to corroborate our studies.

The novel contributions of this paper are summarized as follows:

1) A new training structure suitable for TWRNs is proposed.
2) The CE-BEM is modified and applied to TWRNs.
3) New algorithms are designed to recover the individual channel knowledge from the estimated of the effective channels.
4) Optimal training sequences are designed.

The rest of the paper is organized as follows. Section II introduces the TWRN channel model, presents the PSAM transmission structure, and identifies the parameters to be estimated with the aid of the CE-BEM model. The channel estimation algorithms and the optimal training design are then provided in Section III. Useful discussions can be found in the same section. In Section IV, our simulation results are presented to corroborate our study, and, finally, our conclusions are provided in Section VI.

Notations: Vectors and matrices are boldface small and capital letters; the transpose, complex conjugate, Hermitian, inverse, and pseudo-inverse of the matrix $\mathbf{A}$ are denoted by $\mathbf{A}^T$, $\mathbf{A}^*$, $\mathbf{A}^H$, $\mathbf{A}^{-1}$ and $\mathbf{A}^\dagger$, respectively; $\Re\{\mathbf{A}\}$ and $\Im\{\mathbf{A}\}$ are the real and the imaginary part of $\mathbf{A}$; $\mathrm{diag}\{\mathbf{a}\}$ denotes a diagonal matrix with the diagonal elements constructed from $\mathbf{a}$; $\otimes$ represents the linear convolution between two vectors; $\mathrm{tr}(\mathbf{A})$ is the trace of $\mathbf{A}$ and $\mathrm{E}\{\cdot\}$ denotes the statistical expectation; $\lceil\cdot\rceil$ and $\lfloor\cdot\rfloor$ are the integer ceiling and integer floor, respectively; the entry indices of vectors and matrices start from 0.

II. Problem Formulation

Consider a TWRN with two source nodes $\mathbb{T}_1$, $\mathbb{T}_2$ and one relay node $\mathbb{R}$ (Fig.1). Each node has only one half-duplex antenna. The baseband channel from $\mathbb{T}_i$, $i = 1, 2$ to $\mathbb{R}$ is assumed to be time-selective flat-fading and is denoted by $h_i(n)$, where n is the discrete time index. Moreover, the channels are modeled as wide-sense stationary (WSS) zero mean complex Gaussian (ZMCG) random processes with variances $\sigma^2_{h_i}$. The channel from $\mathbb{R}$ to $\mathbb{T}_i$ is also denoted as $h_i(n)$. Perfect synchronization is assumed in this paper as did in [10], [15],[21].

A. Time-varying Relay Channels

The channel statistics in a relay network depend on the mobility of the three nodes, i.e., the fixed nodes or the moving nodes [11]. Denote f_{d1}, f_{d2} and f_{dr} as the maximum Doppler shifts due to the motion of $\mathbb{T}_1$, $\mathbb{T}_2$, and $\mathbb{R}$, respectively. The discrete autocorrelation functions of $h_i(n)$'s can be represented as [22]

$$\begin{aligned} R_{h_i}(m) =& \mathrm{E}\{h_i(n+m)h_i^*(n)\} \\ =& \sigma^2_{h_i} J_0(2\pi f_{di} m T_s) J_0(2\pi f_{dr} m T_s), \quad i = 1, 2 \end{aligned} \tag{1}$$

where $J_0(\cdot)$ is the zero-th order Bessel function of the first kind, and T_s is the symbol sampling time. The correlation function in (1) has been widely adopted to describe the mobile-to-mobile link, e.g., [11], [22]. If one node is fixed, i.e., if the corresponding Doppler shift is zero, then (1) reduces to the well-known Jakes model [13]. Meanwhile, (1) reveals that the power spectra of $h_1(n)$ and $h_2(n)$ span over the bandwidths $f_1 = f_{d1} + f_{dr}$ and $f_2 = f_{d2} + f_{dr}$, respectively, which indicates an increased Doppler effect for the mobile-to-mobile transmission.

The parsimonious finite-parameter BEM [15] can be applied to approximate the two time-varying channels, respectively, so that during any time interval of NT_s, $h_i(n)$'s can be modeled by

$$h_1(n) = \sum_{q=0}^{Q_1} \lambda_q w_1(q), \quad h_2(n) = \sum_{q=0}^{Q_2} \mu_q w_2(q), \tag{2}$$

where $0 \le n \le N-1$, λ_q's and μ_q's are the BEM coefficients that remain invariant within one interval of NT_s but will change in the next interval, while $w_i(q)$'s are the bases that capture the time variation and will remain the same for any interval. The number of the bases Q_i is a function of the channel bandwidth f_i and the interval length NT_s. Specific choices for $\{w_i(q)\}_{q=0}^{Q_i}$ include the polynomial [23], wavelet [24], discrete prolate spheroid sequence [25], and Fourier bases [18]. In this paper, we choose the CE-BEM [21], a specific form of Fourier bases. Then (2) can be explicitly written as

$$h_1(n) = \sum_{q=0}^{Q_1} \lambda_q e^{j2\pi(q-Q_1/2)n/N}, \tag{3a}$$

$$h_2(n) = \sum_{q=0}^{Q_2} \mu_q e^{j2\pi(q-Q_2/2)n/N}. \tag{3b}$$

The CE-BEM (3) can be viewed as the Fourier series of the time-varying channels, and the number of bases Q_i should be at least $2\lceil f_i N T_s \rceil$ in order to provide sufficient degrees of freedom [18], [21].

To simplify the notation as well as the discussion, we assume $f_1 = f_2 = f_d$ and $Q_1 = Q_2 = Q$. Nonetheless, the

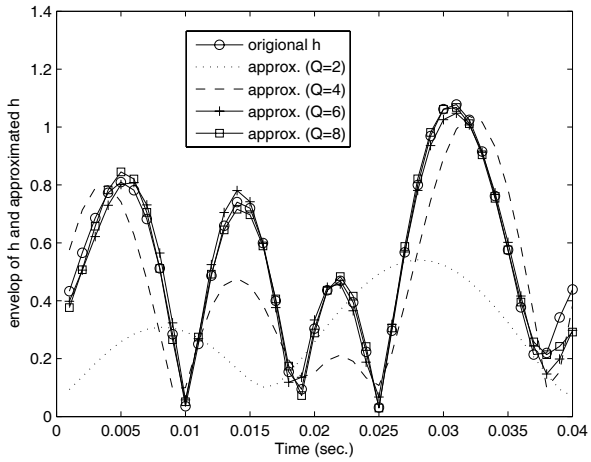


Fig. 2. Approximating the mobile-to-mobile channels with CE-BEM.

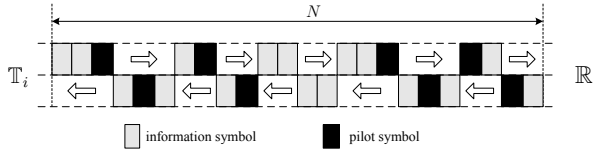


Fig. 3. Proposed transmission strategy for two-way relay network with time-varying channel.

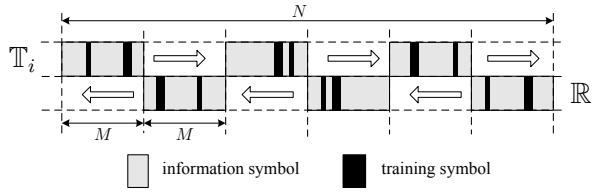


Fig. 4. Sub-block based transmission strategy.

extension to the general case is straightforward. We further denote $\omega_q = 2\pi(q - Q/2)/N$ and define

$$\boldsymbol{\lambda} = [\lambda_0, \lambda_1, \ldots, \lambda_Q]^T, \qquad \boldsymbol{\mu} = [\mu_0, \mu_1, \ldots, \mu_Q]^T$$

for subsequent use.

Remark 1: A detailed discussion of the accuracy of the CE-BEM representation of the time-varying TWRN channels is beyond the scope of this paper. A brief example is given in Fig. 2, where the system parameters are taken as $f_{d1} = f_{d2} = f_{dr} = 40$ Hz, $T_s = 1$ ms, and $N = 40$. Fig. 2 reveals that the larger the Q is, the better the approximation will be. As pointed out in [18], [21], Q must be at least $2\lceil f_d T_s N \rceil = 4$ in order to keep the shape of the envelope, i.e., with the sufficient degrees of freedom. However, for $Q = 2$ the ambiguous estimation appears due to the lack of sufficient sampling degrees of freedom. Nonetheless, one can always use a larger Q for a better approximation.

B. Transmission Strategy

To enable the use of PSAM in our TWRN, we propose a new transmission strategy over one interval NT_s, as depicted in Fig. 3. Let $\mathcal{D}_t$ and $\mathcal{T}_t$ be the time index sets for the transmitted information symbols and the pilot symbols from $\mathbb{T}_i$, $i = 1, 2$, respectively. Moreover, let $\mathcal{D}_r$ and $\mathcal{T}_r$ be the time index sets for the received information symbols and pilot symbols at $\mathbb{T}_i$, respectively. These four sets are disjoint with the property that $\mathcal{D}_t \bigcup \mathcal{T}_t \bigcup \mathcal{D}_r \bigcup \mathcal{T}_r = \{0, 1, \ldots, N-1\}$. Let us define the cardinality of the sets as $|\mathcal{D}_t| = |\mathcal{D}_r| = D$ and $|\mathcal{T}_t| = |\mathcal{T}_r| = T$. Then, $N = 2(D + T)$ is an even integer.

We assume that the relay node $\mathbb{R}$ forwards its received symbols on time slot $g(n)$ to both $\mathbb{T}_i$ on time slot n; i.e.,

$$\mathcal{D}_t \bigcup \mathcal{T}_t = \left\{ g(n) | n \in \mathcal{D}_r \bigcup \mathcal{T}_r \right\}. \tag{4}$$

By defining the one-to-one mapping function $g(n)$, we implicitly allow for the symbols' order to be changed when $\mathbb{R}$ forwards them back to $\mathbb{T}_i$. Hence, it is also possible to optimize $g(n)$ according to different criteria, i.e., data detection MSE, bit-error-rate (BER), throughput, and others.[1] A detailed discussion is beyond the scope of this paper and will be left to the future research. Note that $0 \le g(n) < n$ is required because $\mathbb{R}$ can only forward a symbol after receiving it. Interestingly, the conventional data transmission in a TWRN [5], [6] becomes a special case of our proposed scheme if $g(n) = n - N/2$ is selected.

Remark 2: A special yet important case involves evenly dividing NT_s intervals into several sub-blocks, as shown in Fig. 4. This case corresponds to setting $g(n) = n - M$, where M divides $N/2$, and will be separately discussed later. The decision on whether to adopt the general scheme (Fig. 3) or the sub-block-based scheme (Fig. 4) depends on the synchronization requirement in practical scenarios and other design issues.

Denote the symbols sent from $\mathbb{T}_i$ as $s_i(n)$, $n \in \mathcal{D}_t \bigcup \mathcal{T}_t$, of which the average power for the information symbols is P_i; i.e., $\mathrm{E}\{|s_i(n)|^2\} = P_i$, $\forall n \in \mathcal{D}_t$, while the total training power[2] is $P_{i,t}$; i.e., $\sum_{n \in \mathcal{T}_t} |s_i(n)|^2 = P_{i,t}$. With perfect synchronization, $\mathbb{R}$ receives

$$r(n) = h_1(n)s_1(n) + h_2(n)s_2(n) + w_r(n), \quad n \in \mathcal{D}_t \bigcup \mathcal{T}_t, \tag{5}$$

where $w_r(n)$ is the circularly symmetric complex Gaussian (CSCG) noise with the variance σ_r^2. If the average transmit power of $\mathbb{R}$ is P_r, then $r(n)$ will be scaled by

$$\alpha(n) = \begin{cases} \sqrt{\dfrac{P_r}{\sigma_{h_1}^2 P_1 + \sigma_{h_2}^2 P_2 + \sigma_r^2}} & n \in \mathcal{D}_r \\ \sqrt{\dfrac{P_r}{\sigma_{h_1}^2 P_{1,t}/T + \sigma_{h_2}^2 P_{2,t}/T + \sigma_r^2}} & n \in \mathcal{T}_r \end{cases} \tag{6}$$

before it is forwarded to $\mathbb{T}_i$'s to keep the power constraint.

Remark 3: More practical considerations should include the process delay at $\mathbb{R}$ as well as the path-delay between $\mathbb{T}_1$ and $\mathbb{T}_2$. These considerations require only slightly changing the channel from $h_i(n)$ to $h_i(n + \Delta n)$, and the remaining discussion holds the same.

[1]The dual problem of optimally re-ordering the subcarrier indices in a frequency-selective environment has been studied in [26].

[2]We should not consider the average power constraints for training because otherwise, the training length is trivially preferred to be as large as possible.

C. On Channel Estimation

Due to symmetry, we present only the estimation procedure at $\mathbb{T}_1$, and the received signal is

$$\begin{aligned} y(n) =& \alpha(n)h_1(n)r(g(n)) + w_1(n) \\ =& \alpha(n)\underbrace{h_1(n)h_1(g(n))}_{b_1(n)}s_1(g(n)) \\ &+ \alpha(n)\underbrace{h_1(n)h_2(g(n))}_{b_2(n)}s_2(g(n)) \\ &+ \underbrace{\alpha(n)h_1(n)w_r(g(n)) + w_1(n)}_{w(n)}, \quad n \in \mathcal{D}_r \bigcup \mathcal{T}_r, \end{aligned} \tag{7}$$

where $w_1(n)$ is the CSCG noise at $\mathbb{T}_1$ with variance σ_1^2; $w(n)$ denotes the overall noise; and $b_i(n)$, $i = 1, 2$ can be treated as the equivalent time-varying channel of $\mathbb{T}_i \to \mathbb{R} \to \mathbb{T}_1$. Obviously, if $b_i(n)$'s are known at $\mathbb{T}_1$, the self-signal component $s_1(g(n))$ can be subtracted from $y(n)$ in order to detect the desired information $s_2(g(n))$.

To gain more insight into the time-varying channels, we apply (3) and rewrite $b_i(n)$'s as

$$b_1(n) = \sum_{p=0}^{Q}\sum_{q=0}^{Q}\lambda_p\lambda_q e^{j(\omega_p n + \omega_q g(n))}, \tag{8a}$$

$$b_2(n) = \sum_{p=0}^{Q}\sum_{q=0}^{Q}\lambda_p\mu_q e^{j(\omega_p n + \omega_q g(n))}, \quad n \in \mathcal{D}_r \bigcup \mathcal{T}_r. \tag{8b}$$

The new expression (8) indicates that in order to obtain $b_i(n)$, $0 \le n \le N-1$, one needs to know either $2(Q+1)$ parameters λ_p, μ_p, $p = 0, \ldots, Q$ or $2(Q+1)^2$ parameters $\lambda_p\lambda_q$, $\lambda_p\mu_q$, $p, q = 0, \ldots, Q$. For a general mapping function $g(n)$, the former approach requires a non-linear search, which is computationally prohibitive, while the latter approach, though could be implemented from linear approach, possesses large redundancy in the number of estimated variables.

To facilitate the channel estimation, we propose to use

$$g(n) = n - M, \tag{9}$$

for $n \in \mathcal{T}_r$, while $g(n)$ for information transmission $n \in \mathcal{D}_r$ could still be designed from a certain optimization criterion. The condition (9) says that $\mathbb{R}$ retransmits each received pilot symbol with a delay of M-symbol interval, and this interval is common for all pilot symbols.

With (9), the received pilot symbols at $\mathbb{T}_1$ can be further expressed as

$$\begin{aligned} y(n) =& \alpha\sum_{m=0}^{2Q}\underbrace{\Big(\sum_{q=0}^{m}\lambda_{m-q}\lambda_q e^{-j\omega_q M}\Big)}_{x_1(m)}e^{j\theta_m n}s_1(n-M) \\ &+ \alpha\sum_{m=0}^{2Q}\underbrace{\Big(\sum_{q=0}^{m}\lambda_{m-q}\mu_q e^{-j\omega_q M}\Big)}_{x_2(m)}e^{j\theta_m n}s_2(n-M) + w(n), \end{aligned} \tag{10}$$

where $n \in \mathcal{T}_r$, $\theta_m = 2\pi(m-Q)/N$, $x_i(m)$ are defined as the corresponding items, and the index n in $\alpha(n)$ is omitted for brevity. When deriving (10), we use the property that $\omega_p + \omega_q = \omega_{p'} + \omega_{q'}$ whenever $p + q = p' + q'$.

Remark 4: If the sub-block transmission in Fig. 4 is applied, then (10) is also applicable for the received information symbols $n \in \mathcal{D}_r$.

Interestingly, we may treat $x_i(m)$'s as the equivalent BEM coefficients with $2Q+1$ carriers $e^{j\theta_m n}$ for the equivalent time-varying channel $b_i(n)$, $n \in \mathcal{T}_r$. The equivalent BEM sequence $x_1(m)$ is the convolution between the original BEM λ_p and $e^{-j\omega_q M}\lambda_q$, while $x_2(m)$ is the convolution between λ_p and $e^{-j\omega_p M}\mu_q$.

Define $\mathbf{x}_i = [x_i(0), x_i(1), \ldots, x_i(2Q)]^T$, $\mathbf{\Gamma} = \text{diag}\{e^{-j\omega_0 M}, e^{-j\omega_1 M}, e^{-j\omega_Q M}\}$, and define $\mathbf{\Lambda}$ as the $(2Q+1) \times (Q+1)$ Toeplitz matrix with the first column $[\boldsymbol{\lambda}^T, \mathbf{0}_{1\times Q}]^T$. We can explicitly express the convolutions as

$$\mathbf{x}_1 = \boldsymbol{\lambda} \otimes (\mathbf{\Gamma}\boldsymbol{\lambda}) = \mathbf{\Lambda}\mathbf{\Gamma}\boldsymbol{\lambda}, \;\; \mathbf{x}_2 = \boldsymbol{\lambda} \otimes (\mathbf{\Gamma}\boldsymbol{\mu}) = \mathbf{\Lambda}\mathbf{\Gamma}\boldsymbol{\mu}. \tag{11}$$

Based on (10), we may estimate the equivalent BEM coefficient $x_i(m)$ (with $4Q+2$ unknowns) and recover the original BEM λ_q, μ_q (with $2Q+2$ unknowns). Then, the equivalent time-varying channels $b_i(n)$, $n \in \mathcal{D}_r$ can be obtained from (8).

III. Channel Estimation Algorithms

Let us specify the indices in $\mathcal{T}_r$ as $n_0 < n_1 \ldots < n_{T-1}$, and define

$$\begin{aligned} \mathbf{y}_t =& [y(n_0), y(n_1), \ldots, y(n_{T-1})]^T, \\ \mathbf{w}_t =& [w(n_0), w(n_1), \ldots, w(n_{T-1})]^T, \\ \mathbf{t}_i =& [s_i(n_0 - M), s_i(n_1 - M), \ldots, s_i(n_{T-1} - M)]^T, \\ \mathbf{T}_i =& \text{diag}\{\mathbf{t}_i\}, \quad i = 1, 2, \end{aligned}$$

where $\mathbf{t}_i$ contains all the pilot symbols from $\mathbb{T}_i$. For notational simplicity, the m-th entry of $\mathbf{t}_i$ is also denoted by $t_i(m)$, $m = 0, \ldots, T-1$.

With the aid of (10), we can express $\mathbf{y}_t$ in matrix form as

$$\mathbf{y}_t = \alpha\mathbf{T}_1\mathbf{A}\mathbf{x}_1 + \alpha\mathbf{T}_2\mathbf{A}\mathbf{x}_2 + \mathbf{w}_t, \tag{12}$$

where $\mathbf{A}$ is the $T \times (2Q+1)$ matrix

$$\mathbf{A} = \begin{bmatrix} e^{j\theta_0 n_0} & e^{j\theta_1 n_0} & \ldots & e^{j\theta_{2Q} n_0} \\ e^{j\theta_0 n_1} & e^{j\theta_1 n_1} & \ldots & e^{j\theta_{2Q} n_1} \\ \vdots & \vdots & \ldots & \vdots \\ e^{j\theta_0 n_{T-1}} & e^{j\theta_1 n_{T-1}} & \ldots & e^{j\theta_{2Q} n_{T-1}} \end{bmatrix}. \tag{13}$$

A. Channel Estimation Algorithm

When $T \ge 4Q + 2$, there are sufficient observations to estimate all the unknown $x_i(m)$'s. In this case, one could choose a linear estimator, e.g., the least square (LS) or the linear minimum mean square error (LMMSE) estimator, to reduce the computational complexity. For example in [18]–[20], the authors assumed that the knowledge of the statistics of the BEM coefficients were available in order to derive the LMMSE estimator. Moreover, a closed-form training design requires the assumption that the BEM coefficients are uncorrelated among themselves [18]–[20]. Although the same assumption can be invoked here, we would rather choose the LS estimator in order to embrace more practical scenarios

where the statistics of the BEM coefficients are not available. This choice is further justified because the LS estimator performs similarly to the LMMSE estimator at a relatively high signal-to-noise ratio (SNR).

Let us define

$$\mathbf{T} = [\mathbf{T}_1\mathbf{A}, \mathbf{T}_2\mathbf{A}], \qquad \mathbf{x} = [\mathbf{x}_1^T, \mathbf{x}_2^T]^T.$$

The LS estimator of $\mathbf{x}$ is expressed as

$$\hat{\mathbf{x}} = \frac{1}{\alpha}\mathbf{T}^\dagger\mathbf{y} = \frac{1}{\alpha}(\mathbf{T}^H\mathbf{T})^{-1}\mathbf{T}^H\mathbf{y}, \tag{14}$$

with the error covariance matrix given by

$$\mathbf{W} = \mathbf{T}^\dagger\mathbf{M}_w(\mathbf{T}^\dagger)^H, \tag{15}$$

where

$$\mathbf{M}_w = \begin{bmatrix} \sigma_r^2|h_1(n_0)|^2 + \frac{\sigma_1^2}{\alpha^2} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \sigma_r^2|h_1(n_{T-1})|^2 + \frac{\sigma_1^2}{\alpha^2} \end{bmatrix}. \tag{16}$$

B. Optimal Training Design

The channel estimation MSE is defined as $\mathrm{tr}(\mathbf{W})$ and is related to the instant CSI. In this case, we propose to minimize the average MSE, which is defined as

$$\mathrm{AMSE} = \mathrm{E}_h\{\mathrm{tr}(\mathbf{W})\} = \left(\sigma_{h_1}^2\sigma_r^2 + \frac{\sigma_1^2}{\alpha^2}\right)\mathrm{tr}((\mathbf{T}^H\mathbf{T})^{-1}), \tag{17}$$

where the property $J_0(0) = 1$ is used. We further partition $(\mathbf{T}^H\mathbf{T})^{-1}$ as

$$(\mathbf{T}^H\mathbf{T})^{-1} = \begin{bmatrix} \mathbf{A}^H\mathbf{T}_1^H\mathbf{T}_1\mathbf{A} & \mathbf{A}^H\mathbf{T}_1^H\mathbf{T}_2\mathbf{A} \\ \mathbf{A}^H\mathbf{T}_2^H\mathbf{T}_1\mathbf{A} & \mathbf{A}^H\mathbf{T}_2^H\mathbf{T}_2\mathbf{A} \end{bmatrix}^{-1}. \tag{18}$$

The optimal training design amounts to selecting the number of the pilot symbols, their placement, and the power allocation for each pilot by minimizing the AMSE. The optimization problem is then formulated as

$$\text{(P1):} \quad \min_{\mathbf{t}_1,\mathbf{t}_2,\mathcal{T}_r} \quad \left(\sigma_{h_1}^2\sigma_r^2 + \frac{\sigma_1^2}{\alpha^2}\right)\mathrm{tr}((\mathbf{T}^H\mathbf{T})^{-1}) \tag{19}$$
$$\text{s.t.} \quad \sum_{m=0}^{T-1}|t_i(m)|^2 \le P_{i,t}, \quad i = 1, 2.$$

Since α is related only to T, we can first solve, for a given T, the following problem:

$$\text{(P2):} \quad \min_{\substack{\mathbf{t}_1,\mathbf{t}_2,\\ n_i:0\le i\le T}} \quad \mathrm{tr}((\mathbf{T}^H\mathbf{T})^{-1}) \tag{20}$$
$$\text{s.t.} \quad \sum_{m=0}^{T-1}|t_i(m)|^2 \le P_{i,t}, \quad i = 1, 2.$$

From [27], we know that

$$\mathrm{tr}((\mathbf{T}^H\mathbf{T})^{-1}) \ge \sum_{i=0}^{4Q+1}\frac{1}{[\mathbf{T}^H\mathbf{T}]_{i,i}} = \sum_{i=1}^{2}\frac{2Q+1}{\sum_{m=0}^{T-1}|t_i(m)|^2}, \tag{21}$$

where $[\mathbf{T}^H\mathbf{T}]_{i,i}$ is the ith diagonal elements of $\mathbf{T}^H\mathbf{T}$, and the equality holds when $\mathbf{T}^H\mathbf{T}$ is a diagonal matrix. However, this inequality does not directly show that the diagonal must hold for the optimal $\mathbf{T}^H\mathbf{T}$. Let us first formulate a new optimization problem:

$$\text{(P3):} \quad \min_{\mathbf{t}_1,\mathbf{t}_2} \quad \sum_{i=1}^{2}\frac{2Q+1}{\sum_{m=0}^{T-1}|t_i(m)|^2} \tag{22}$$
$$\text{s.t.} \quad \sum_{m=0}^{T-1}|t_i(m)|^2 \le P_{i,t}, \quad i = 1, 2.$$

Obviously, the optimal objective of (P3) serves as a lower bound for (P2). Since (P3) is a simple convex optimization, any training sequence satisfying $\sum_{m=0}^{T-1}|t_i(m)|^2 = P_{i,t}$ is optimal. Hence, if we can find $\mathbf{t}_i$'s that satisfy the equality constraints and make $\mathbf{T}^H\mathbf{T}$ diagonal, then these $\mathbf{t}_i$'s must also be the optimal solutions for problem (P2). In other words, the sufficient conditions for the optimal solutions to (P2) are

$$\mathbf{A}^H\mathbf{T}_i^H\mathbf{T}_i\mathbf{A} = P_{i,t}\mathbf{I}_{2Q+1}, \quad i = 1, 2, \tag{23a}$$
$$\mathbf{A}^H\mathbf{T}_1^H\mathbf{T}_2\mathbf{A} = \mathbf{0}_{2Q+1}. \tag{23b}$$

Observing the Vandermonde structure of $\mathbf{A}$ and the structure of θ_m, we know that if the pilot symbols are equi-powered and equi-spaced over $\{0, \dots, N-1\}$, then (23a) is satisfied; i.e.,

C1): $|t_i(m)|^2 = P_{i,t}/T$, $\forall m = 0, 1, \dots, T-1$, $i = 1, 2$,
C2): $n_m = mL + l_0$, $\forall l_0 \in [M, L-1]$,
and $L = N/T$ is an integer,

where we include the consideration that $n_0 \ge M$ in C2).[3] Combined with C1) and C2), the following condition can guarantee (23b):

$$\sum_{m=0}^{T-1} t_2^*(m)t_1(m)e^{-j(\theta_u+\theta_v)n_m} = 0, \ \forall u, v = 0, 1, \cdots, 2Q$$

which can be simplified as

$$\text{C3):} \quad \sum_{m=0}^{T-1} t_2^*(m)t_1(m)e^{j2\pi mk/T} = 0, \quad \forall k = -2Q, -2Q+1, \dots, 2Q.$$

One example of pilot sequences that satisfy conditions C1)–C3) is

$$\mathbf{t}_1 = \sqrt{\frac{P_{1,t}}{T}}[+1, +1, +1, \dots, +1, +1]^T, \tag{24a}$$
$$\mathbf{t}_2 = \sqrt{\frac{P_{2,t}}{T}}[1, e^{j2\pi v/T}, \dots, e^{j2\pi(T-1)v/T}]^T, \quad \forall v = 2Q+1, \dots, T-2Q-1. \tag{24b}$$

The minimum $\mathrm{tr}((\mathbf{T}^H\mathbf{T})^{-1})$ is then $(2Q+1)(1/P_{1,t} + 1/P_{2,t})$ and does not depend on T. Hence, the optimal value of T should be independently obtained from

$$T = \arg\min_T \left(\sigma_{h_1}^2\sigma_r^2 + \frac{\sigma_1^2}{\alpha^2}\right) = \arg\max_T \frac{P_r}{\sigma_{h_1}^2 P_{1,t}/T + \sigma_{h_2}^2 P_{2,t}/T + \sigma_r^2}. \tag{25}$$

[3] $n_0 = l_0$ denotes the index of the first symbol sent by $\mathbb{R}$. From the adopted $g(n)$, $n_0 = l_0 \ge M$ is required.

The objective function (25) is an increasing function of T, so the optimal T should be made as large as possible. Note that this result is different from the conventional training design in point-to-point systems, where the channel estimation MSE is related only with the total training power but not to the training length.

However, increasing T would reduce the efficiency of the data transmission and, consequently, the system throughput. Besides, the constant $\sigma_{h_1}^2\sigma_r^2$ will dominate the summation from $\left(\sigma_{h_1}^2\sigma_r^2+\frac{\sigma_1^2}{\alpha^2}\right)$ when T is greater than a certain threshold. Therefore, increasing T beyond a certain value cannot improve the channel estimation MSE, but the throughput will be linearly decreased. A more meaningful design of T can be obtained by maximizing the transmission throughput criterion [18]–[20]. Due to the page limit, we in this paper focus only on introducing the new channel estimation strategy in TWRN, and we simply consider achieving the minimum amount of training as our optimization goal.

The selection of the minimum possible T depends on many factors and will be discussed in the next subsection. When $T=4Q+2$ is allowed, the optimal pilot schemes become more specific:

$$\mathbf{t}_1=\sqrt{\frac{P_{1,t}}{4Q+2}}[+1,+1,+1,+1,\ldots,+1,+1]^T,$$

$$\mathbf{t}_2=\sqrt{\frac{P_{2,t}}{4Q+2}}[+1,-1,+1,-1,\ldots,+1,-1]^T,$$

and the corresponding minimum AMSE is

$$\text{AMSE}=\left(\frac{1}{P_{1,t}}+\frac{1}{P_{2,t}}\right)\left((2Q+1)\sigma_{h_1}^2\sigma_r^2+\vartheta\right), \quad (26)$$

where

$$\vartheta=\frac{\left(\sigma_{h_1}^2P_{1,t}+\sigma_{h_2}^2P_{2,t}+2\sigma_r^2(2Q+1)\right)\sigma_1^2}{2P_r}. \quad (27)$$

Remark 5: Importantly, it can be verified that the designed optimal pilot sequences for channel estimation at $\mathbb{T}_1$ are also optimal at $\mathbb{T}_2$. Hence, simultaneous optimal channel estimation can be achieved at both source nodes.

C. Parameter Selection

Observing C2), we know the following: (i) The pilot spacing L should at least be $M+1$; (ii) To transmit non-zero information symbols in one NT_s interval, we need $D=\frac{LT-2T}{2}\geq 1$, so the spacing L should be at least 3;[4] (iii) Since N must be even, either T or L should be an even integer.

The above discussion suggests the guidelines for choosing T, i.e., select the smallest integer that is greater than or equal to $4Q+2$, divides N, and satisfies $N/T\geq 3$.

Remark 6: Since $T\geq 4Q+2$ pilot symbols are needed to provide sufficient observations, and since $Q\geq 1$ for a time-varying channel, the TWRN requires that pilot symbols to be transmitted back-and-forth at least 6 times. Therefore, the conventional two-way frame transmission structure [1], [2], [6], i.e., sending and receiving the continuous data sequence only once, obviously does not work in time-selective environment. We thus re-emphasize the novelty of the proposed PSAM scheme in Fig. 3.

Remark 7: For the sub-block based frame structure in Fig. 4, the receiving equi-spaced pilot at $\mathbb{T}_i$ is possible only if each sub-block contains only one pilot symbol at the same position of each sub-block.

D. Doppler Shift and Transmission Efficiency

From $N=LT\geq L(4Q+2)$ and $Q\geq 2\lceil f_dNT_s\rceil$, a successful channel estimation requires

$$\frac{N-2L}{8LN}\geq f_dT_s. \quad (28)$$

To cope with more Doppler shift, the left-hand side (LHS) of (28) should be as large as possible. From $L\geq 3$, there is

$$\frac{N-2L}{8LN}\leq\frac{N-6}{24N}<\frac{1}{24}. \quad (29)$$

Then the proposed strategy can handle the time-selective scenario with

$$f_dT_s<\frac{1}{24}=0.0416. \quad (30)$$

This normalized Doppler shift, fortunately, lies in the acceptable range of most studies [16]–[20].

Moreover, the training requirement $4Q+2\leq T=N/L\leq N/3$ implies that the transmission efficiency has the range

$$\eta=\frac{N/2-T}{N/2}\leq 1-16f_dT_s-\frac{4}{N}<1-16f_dT_s, \quad (31)$$

and

$$\eta\geq\frac{N/2-N/3}{N/2}=1/3. \quad (32)$$

Therefore, the higher the f_dT_s is, the less the transmission efficiency will be. This result is intuitively satisfying.

E. Recovering the Original BEM Coefficients

After estimating $\mathbf{x}_i$'s, $i=1,2$, we need to obtain the original BEM coefficients λ_q and μ_q in order to build the time-varying channel $b_i(n)$, $n\in\mathcal{D}_r$. This is the key difference of TWRN from OWRN, as pointed out in [5], [6]. Retrieving λ_q and μ_q from $\mathbf{x}_i$ generally requires solving multivariate nonlinear equations, but doing so is computationally quite expensive. In the following, we propose two simple methods and describe them under a noise-free scenario.

1) Time-Domain Approach: Because of the structure of $x_1(m)$, a straightforward way is to estimate λ_q sequentially. Specifically, we first estimate λ_0 from

$$\lambda_0=I_s\left(x_1(0)e^{j\omega_0M}\right)^{1/2}, \quad (33)$$

where $I_s=\pm 1$ denotes the sign uncertainty. By choosing any of the positive or negative signs in (33), λ_1 can be computed from

$$\lambda_1=\frac{x_1(1)}{\lambda_0e^{-j\omega_0M}+\lambda_0e^{-j\omega_1M}}. \quad (34)$$

[4]This conclusion is also seen from the fact that if $L=2$, then the only choice for M is 1, in which case $\mathbb{T}_i$ alternatively transmits and receives pilot symbols while no information can be sent.

We then sequentially compute λ_q from $x_1(q)$ with the previous estimates of $\lambda_0, \ldots, \lambda_{q-1}$. The detailed steps are straightforward and are omitted here.

The above process uses only the first $Q+1$ entries in $\mathbf{x}_1$ and cannot provide satisfactory precision. Nevertheless, with this initial estimation, we can apply the gradient decent process [28] to improve the estimation accuracy. The objective is to minimize the distance between $\mathbf{x}$ and $\mathbf{\Lambda\Gamma\lambda}$; i.e., $\zeta = \|\mathbf{x} - \mathbf{\Lambda\Gamma\lambda}\|^2$. Then, $\boldsymbol{\lambda}$ can be updated according to

$$\boldsymbol{\lambda}^{(i+1)} = \boldsymbol{\lambda}^{(i)} - \epsilon \left.\frac{\partial \zeta}{\partial \boldsymbol{\lambda}^*}\right|_{\boldsymbol{\lambda}=\boldsymbol{\lambda}^{(i)}}, \tag{35}$$

where ϵ is the step size. See Appendix I for a brief illustration of the gradient decent method with complex variables. The partial differential in (35) can be explicitly expressed as

$$\frac{\partial \zeta}{\partial \boldsymbol{\lambda}^*} = -(\mathbf{\Lambda\Gamma} + \mathbf{\Omega})^H(\mathbf{x} - \mathbf{\Lambda\Gamma\lambda}), \tag{36}$$

where $\mathbf{\Omega}$ is a $(2Q+1) \times (Q+1)$ Toeplitz matrix with the first column $[(\mathbf{\Gamma\lambda})^T, \mathbf{0}_{1\times Q}]^T$.

Once $\boldsymbol{\lambda}$ is obtained, $\boldsymbol{\mu}$ can be found from

$$\boldsymbol{\mu} = \mathbf{\Gamma}^H \mathbf{\Lambda}^\dagger \mathbf{x}_2. \tag{37}$$

Remark 8: Note that there exists a simultaneous sign ambiguity (SSA) in the estimated results due to step (33); i.e., either $\{\boldsymbol{\lambda}, \boldsymbol{\mu}\}$ or $\{-\boldsymbol{\lambda}, -\boldsymbol{\mu}\}$ is found. Nonetheless, the SSA does not affect the data detection when we reconstruct $b_i(n)$'s. A similar observation is also made in [5], [6].

2) Frequency-Domain Approach: Let $\tilde{\boldsymbol{\lambda}}$ be the Z-point discrete Fourier transform (DFT) of $\boldsymbol{\lambda}$ with $Z \geq Q+1$, whose mth entry is defined as

$$\tilde{\lambda}_m = \sum_{q=0}^{Q} \lambda_q e^{-j2\pi qm/Z}, \quad m = 0, \ldots, Z-1. \tag{38}$$

On the other side, the mth element of the Z-point DFT of $\mathbf{\Gamma\lambda}$ is

$$\begin{aligned}\xi_m =& \sum_{q=0}^{Q} \lambda_q e^{-j2\pi \frac{(q-Q/2)M}{N}} e^{-j2\pi qm/Z} \\ =& e^{j\frac{\pi QM}{N}} \sum_{q=0}^{Q} \lambda_q e^{-j\frac{2\pi q(\frac{ZM}{N}+m)}{Z}}.\end{aligned} \tag{39}$$

If $R \triangleq \frac{ZM}{N}$ is an integer, then (39) becomes $e^{j\frac{\pi QM}{N}} \tilde{\lambda}_{\langle m+R\rangle_Z}$, where $\langle\cdot\rangle_Z$ denotes the modulo-Z operation. Then the mth element of the Z-point DFT of $\mathbf{x}_1(m)$ is

$$\tilde{x}_1(m) = \xi_m \tilde{\lambda}_m = e^{j\frac{\pi QM}{N}} \tilde{\lambda}_m \tilde{\lambda}_{\langle m+R\rangle_Z}. \tag{40}$$

Our target is to retrieve Z unknown $\tilde{\lambda}_m$'s, $m = 0, 1, \ldots, Z-1$ from Z equations

$$\tilde{\lambda}_m \tilde{\lambda}_{\langle m+R\rangle_Z} = \tilde{x}_1(m) e^{-j\frac{\pi QM}{N}} \triangleq c_m, \quad \forall m = 0, \ldots, Z-1, \tag{41}$$

where c_m is defined as the corresponding constant.

Theorem 1: If Z is odd and is co-prime with R, then $\tilde{\lambda}_m$'s can be found from (41) as

$$\tilde{\lambda}_m = \frac{\left(\prod_{i=0}^{Z-1} c_i\right)^{1/2}}{\prod_{i=0}^{\frac{Z-3}{2}} c_{\langle m+(2i+1)R\rangle_Z}}, \qquad \forall m = 0, \ldots, Z-1, \tag{42}$$

with only a SSA.

Proof: See Appendix II. ■

Note that the selection of R and Z is very important in implementing the frequency-domain approach. Let $N'/M' = N/M$ be the simplest form of the fraction; i.e., N' and M' are co-prime. The following ways to select Z are proposed:

- If N' is odd and is greater than $Q+1$, then we can choose $Z = N'$ and $R = M'$.
- Otherwise, pick any integer κ such that $\kappa N' \geq Q$ (κ can be 1 to account for the case when N' is greater than $Q+1$ but is even):
  - If $\kappa N'$ is even, then we choose $R = \kappa M'$ and $Z = \kappa N' + 1$, namely, $Z = RN/M + 1$. This choice will guarantee that Z is odd and is co-prime with R, while the consequence is that

$$\begin{aligned}\xi_m =& e^{j\frac{\pi QM}{N}} \sum_{q=0}^{Q} \lambda_q e^{-j\frac{2\pi q(R+m)}{Z}} e^{-j\frac{2\pi q\frac{M}{N}}{Z}} \\ \approx& \tilde{\lambda}_{\langle m+R\rangle_Z}.\end{aligned} \tag{43}$$

    Note that, a similar approximation has been used in many multi-carrier systems when the channel frequency response on the adjacent carriers is assumed to be the same [29]. Our approximation is more accurate since the distortion phase of each summand is only $\frac{2\pi q\frac{M}{N}}{Z} < \frac{2\pi q}{Z}$. Moreover, we can always choose a large enough Z such that the approximation becomes sufficiently accurate.
  - If $\kappa N'$ is odd, then we choose $R = \kappa M'$ and $Z = \kappa N' + 2$, namely, $Z = RN/M + 2$. This choice will guarantee that Z is odd and is co-prime with R. A similar approximation on ξ_m applies.

After obtaining $\tilde{\boldsymbol{\lambda}}$, we can find $\boldsymbol{\lambda}$ from the first $Q+1$ elements of the Z-point inverse Fourier transform (IDFT) of $\tilde{\boldsymbol{\lambda}}$. Since the frequency-domain approach fully utilizes all the information, the initial estimates of λ_q's are expected more accurate than those from the time-domain approach. However in the low SNR region, (42) is susceptible to error enhancement due to the products in both the denominator and numerator, as will be seen in our later simulations.

The same iteration (35) can be then applied to improve the accuracy of $\boldsymbol{\lambda}$. Finally, $\boldsymbol{\mu}$ can be found from (37).

IV. Simulation Results

In order to evaluate the inherent performance of our algorithms, the time-varying channels are generated directly from the BEM model (3). The same approach has been adopted in many other papers when testing the performance of channel estimation [25], [30]. However, the real channel generated from (1) will be applied for data detection [17], [18], [21].

A. Channel Estimation and Training Design

The parameters for channel estimation are taken as $Q = 4$, $N = 352$, $M = 8$, and $T = 22$. A total of 10000 Monte-Carlo trials are used for averaging. Optimal training is compared with two types of random training. In the first one, all pilots are equi-powered but randomly spaced. In the second one, the pilot power levels are random, but the pilots are uniformly spaced.

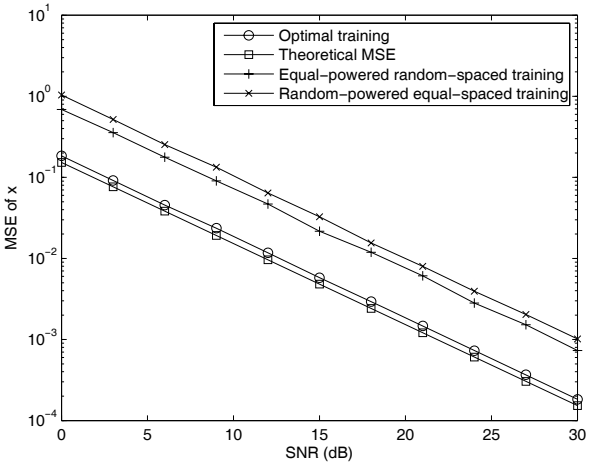


Fig. 5. Channel estimation MSEs versus SNR for $\mathbf{x}$.

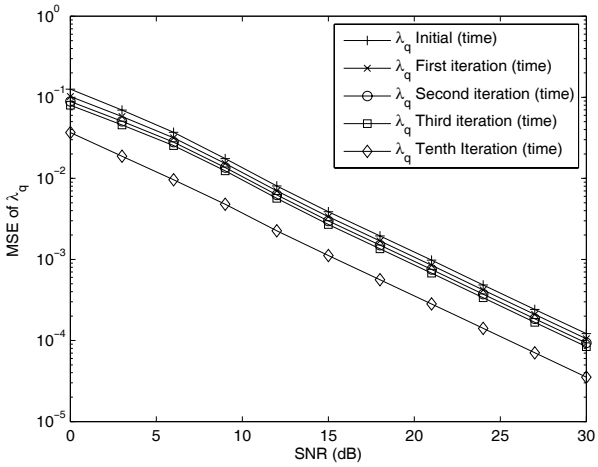


Fig. 6. Channel estimation MSEs versus SNR for $\boldsymbol{\lambda}$: time-domain approach.

1) Estimation of the equivalent BEM coefficients: The estimation MSEs of the equivalent channel $\mathbf{x} = [\mathbf{x}_1^T, \mathbf{x}_2^T]^T$ from the three different types of training are shown in Fig. 5. The theoretical MSE is also displayed for comparison. Clearly, the designed optimal training sequence given in (24) achieves the best performance, with the MSE being close to the theoretical one.

2) Recovery of the original BEM coefficients– time-domain approach: In this example, the coefficients $\boldsymbol{\lambda}$ are extracted from the estimated $\hat{\mathbf{x}}_1$. The MSEs versus the SNR for the initial estimate as well as those after several iterations are shown in Fig. 6, which reveals that the initial estimate is effective in the sense that the MSE curve linearly decreases with the increase of the SNR. Moreover, the iterations can significantly improve the estimation accuracy since the initial estimation utilizes only part of the observations in $\mathbf{x}_1$. After the tenth iteration, the improvement is negligible.

3) Recovery of the original BEM coefficients– frequency-domain approach: Next we choose the frequency-domain approach to recover the coefficients $\boldsymbol{\lambda}$. From (39), we take $R = 1, 2, 5$, respectively, and $Z = RN/M + 1$ is $45, 89, 221$, respectively. The estimation MSEs versus the SNR for the initial estimation as well as those from the 10-th iteration are shown in Fig. 7. The iteration in this case marginally improves the estimation accuracy because the initial frequency-domain estimate fully explores $\mathbf{x}_1$. Moreover, the choice of a different R does not affect the performance significantly. As mentioned previously, the performance of the frequency-domain approach degrades at a relatively low SNR, say, 8 dB in Fig. 7, due to the error enhancement.

4) Comparing the time- and the frequency-domain approaches: It is then of interest to compare the performances of the two different approaches in recovering $\boldsymbol{\lambda}$. To make this comparison clear, we present the results in a new figure, and apply fifty iterations for both methods. As Fig. 8 indicates, the initial frequency-domain estimation outperforms the time-domain results at the high SNR region, even if the latter apply iterations. Nonetheless, the performance gap is quite small. At a relatively low SNR, say SNR$= 8$ dB, the time-domain approach gives a better performance.

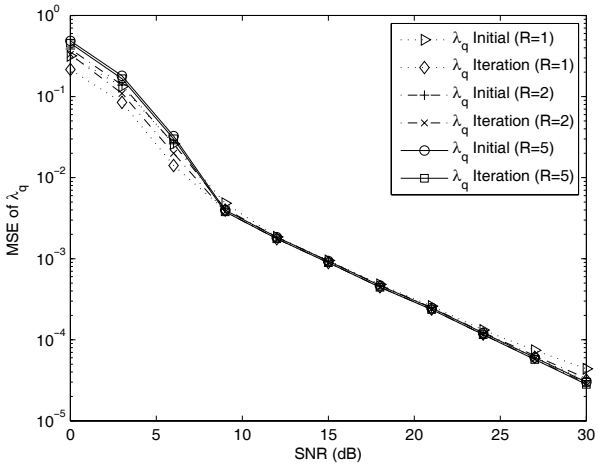


Fig. 7. Channel estimation MSEs versus SNR for $\boldsymbol{\lambda}$: frequency-domain approach.

5) Estimation of $\boldsymbol{\mu}$: After obtaining $\boldsymbol{\lambda}$, $\boldsymbol{\mu}$ can be estimated through (37). The corresponding MSEs versus the SNR from both time and frequency-domain approaches are shown in Fig. 9. It is seen that the estimated $\boldsymbol{\mu}$ contains a larger error than the estimated $\boldsymbol{\lambda}$ in Fig. 8. This difference is expected since $\boldsymbol{\mu}$ is obtained through the estimated $\boldsymbol{\lambda}$ so that the errors in $\boldsymbol{\lambda}$ propagate to the estimates of $\boldsymbol{\mu}$.

B. Data Detection

For data detection, the channel is generated by using the more realistic model (1). The bit error rate (BER) is the figure of merit. The system parameters shown in Fig. 2 are taken. We first apply the channel estimation method to find the BEM coefficients and to reconstruct the time-varying channels $b_i(n)$. Then, the self-signal component is canceled before the data detection. The error due to non-perfect removal of the self-signal will also affect the system performance. The time- and the frequency-domain approaches are used to estimate $\boldsymbol{\lambda}$ and $\boldsymbol{\mu}$. The BERs versus the SNR for different numbers of Q, and the BER under perfect channel knowledge are displayed

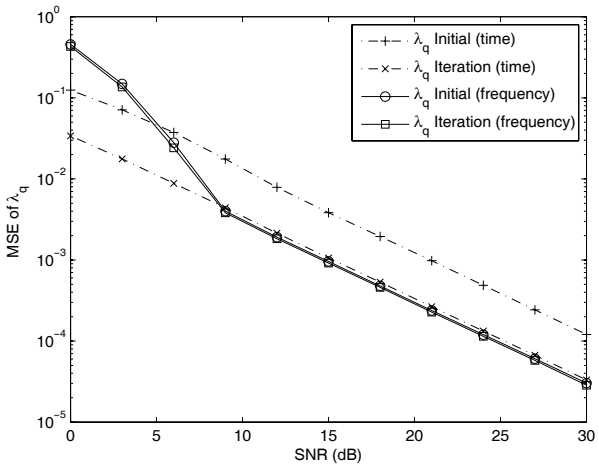


Fig. 8. Comparison between the time- and the frequency-domain approaches.

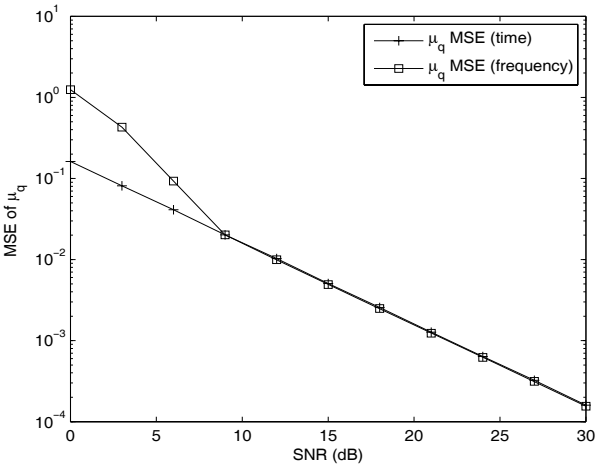


Fig. 9. Channel estimation MSEs versus SNR for $\boldsymbol{\mu}$.

in Fig. 10. Clearly, the proposed methods yield effective data detection. At high SNRs, the frequency-domain method yields better BER performance than the time-domain method since the former can provide better estimation results. 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 of Qs.

V. Conclusions

In this paper, we studied the problem of channel estimation for time-varying TWRN channels. A new PSAM scheme was designed, and the channel estimation was related to a finite number of variables by using CE-BEM. The LS estimator for the convolved BEM coefficients was derived along with the optimal training sequences. Time-domain and frequency-domain algorithms were then developed to recover the individual BEM coefficients from the convolved ones. The selection of the system parameters to guide the practical design was fully discussed. The simulation results clearly demonstrated the effectiveness of the proposed algorithms and corroborated the studies.

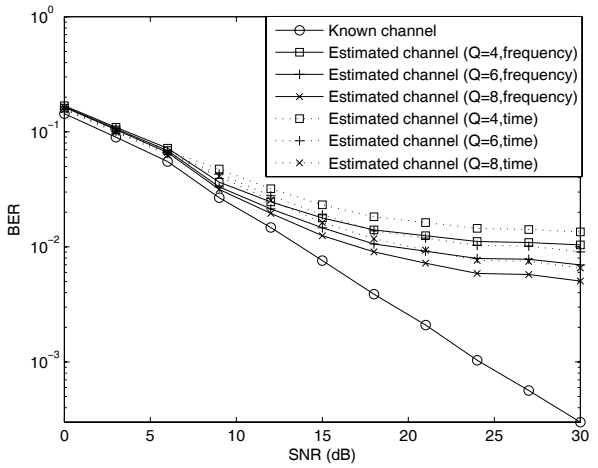


Fig. 10. BER versus SNR for realistic mobile-to-mobile channels with different values of Q.

Appendix I
Gradient Decent Method with Complex Variables

Let us define a new $2(Q+1)\times 1$ real vector $\boldsymbol{\rho} = [\Re\{\boldsymbol{\lambda}\}^T, \Im\{\boldsymbol{\lambda}\}^T]^T$. The gradient decent method to update $\boldsymbol{\rho}$ can be directly obtained as [28]

$$\boldsymbol{\rho}^{(i+1)} = \boldsymbol{\rho}^{(i)} - \epsilon' \left.\frac{\partial \zeta}{\partial \boldsymbol{\rho}}\right|_{\boldsymbol{\rho}=\boldsymbol{\rho}^{(i)}}, \tag{44}$$

or equivalently,

$$\begin{bmatrix}\Re\{\boldsymbol{\lambda}\}^{(i+1)}\\ \Im\{\boldsymbol{\lambda}\}^{(i+1)}\end{bmatrix} = \begin{bmatrix}\Re\{\boldsymbol{\lambda}\}^{(i)}\\ \Im\{\boldsymbol{\lambda}\}^{(i)}\end{bmatrix} - \epsilon' \left.\begin{bmatrix}\frac{\partial \zeta}{\partial \Re\{\boldsymbol{\lambda}\}}\\ \frac{\partial \zeta}{\partial \Im\{\boldsymbol{\lambda}\}}\end{bmatrix}\right|_{\boldsymbol{\lambda}=\boldsymbol{\lambda}^{(i)}}. \tag{45}$$

There is

$$\begin{aligned}\boldsymbol{\lambda}^{(i+1)} &= \Re\{\boldsymbol{\lambda}\}^{(i+1)} + j\Im\{\boldsymbol{\lambda}\}^{(i+1)}\\ &= \underbrace{\Re\{\boldsymbol{\lambda}\}^{(i)} + j\Im\{\boldsymbol{\lambda}\}^{(i)}}_{\boldsymbol{\lambda}^{(i)}} - \epsilon' \underbrace{\left(\frac{\partial \zeta}{\partial \Re\{\boldsymbol{\lambda}\}} + j\frac{\partial \zeta}{\partial \Im\{\boldsymbol{\lambda}\}}\right)}_{2\frac{\partial \zeta}{\partial \boldsymbol{\lambda}^*}} \Bigg|_{\boldsymbol{\lambda}=\boldsymbol{\lambda}^{(i)}},\end{aligned} \tag{46}$$

where the definition of the complex derivative [28] is used. Setting $\epsilon = 2\epsilon'$ yields (35).

Appendix II
Proof of Theorem 1

Let us first prove the following lemma:

Lemma 1: If Z and R are co-prime, then the index set $\mathcal{I} = \{\langle m+uR\rangle_Z\}_{u=0}^{Z-1}$ is the same as the universal set $\{0,\ldots,Z-1\}$, or equivalently,

$$\langle m+uR\rangle_Z \neq \langle m+vR\rangle_Z, \quad \text{for} \quad 0 \le u < v \le Z-1. \tag{47}$$

Proof: Let us first assume the contrary holds; i.e.,

$$\langle m+uR\rangle_Z = \langle m+vR\rangle_Z, \quad \exists u < v. \tag{48}$$

Then we know

$$(v-u)R = kZ \tag{49}$$

for some integer $k \neq 0$. Since Z and R are co-prime, their least common multiple must be ZR. However in (49), $(v-u) < Z$, so the equality (49) cannot hold. By the contradiction, we prove Lemma 1. ■

Define a new variable

$$c = \left(\prod_{i=0}^{Z-1} c_i \right)^{1/2} = I_s \prod_{i=0}^{Z-1} \tilde{\lambda}_i. \tag{50}$$

When Z is odd, the denominator in (42) can be expanded as

$$\begin{aligned} \prod_{i=0}^{\frac{Z-3}{2}} c_{\langle m+(2i+1)R \rangle_Z} &= \prod_{i=0}^{\frac{Z-3}{2}} \tilde{\lambda}_{\langle m+(2i+1)R \rangle_Z} \tilde{\lambda}_{\langle m+(2i+2)R \rangle_Z} \\ &= \prod_{i=1}^{Z-1} \tilde{\lambda}_{\langle m+iR \rangle_Z} = \prod_{\substack{i=0 \\ i \neq m}}^{Z-1} \tilde{\lambda}_i, \end{aligned} \tag{51}$$

where Lemma 1 is applied in the last equality.

Dividing (50) by (51) proves Theorem 1, where I_s serves as SSA for all $\tilde{\lambda}_m$.

REFERENCES

[1] S. Zhang, S.-C. Liew, and P. P. Lam, "Physical-layer network coding," in *Proc. MobiComm*, Sep. 2006, pp. 358-365.

[2] S. Katti, S. Gollakota, and D. Katabi, "Embracing wireless interference: analog network coding," Computer Science and Artificial Intelligence Laboratory Technical Report, MIT-CSAIL-TR-2007-012, Feb. 23, 2007.

[3] 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*, vol. 50, no. 12, pp. 3062-3080, Dec. 2004.

[4] S.-Y. R. Li, R. W. Yeung, and N. Cai, "Linear network coding," *IEEE Trans. Inf. Theory*, vol. 49, no. 2, pp.371-381, Feb. 2003.

[5] F. Gao, R. Zhang, and Y.-C. Liang, "Optimal channel estimation and training design for two-way relay networks," *IEEE Trans. Commun.*, vol. 57, no. 10, pp. 3024-3033, Oct. 2009.

[6] ——, "Channel estimation for OFDM modulated two-way relay networks," *IEEE Trans. Signal Process.*, vol. 57, no. 11, pp. 4443-4455, Nov. 2009.

[7] H. Minn and N. Al-Dhahir, "Optimal training signals for MIMO OFDM channel estimation," *IEEE Trans. Wireless Commun.*, vol. 5, no. 6, pp. 1158-1168, May 2006.

[8] F. Gao, T. Cui, and A. Nallanathan, "On channel estimation and optimal training design for amplify and forward relay network," *IEEE Trans. Wireless Commun.*, vol. 7, no. 5, pp. 1907-1916, May 2008.

[9] G. Wang, F. Gao, X. Zhang, and C. Tellambura, "Superimposed training based joint CFO and channel estimation for CP-OFDM modulated two-way relay networks," *EURASIP J. Wireless Commun. Netw.*, vol. 2010, pp. 1-9, July 2010.

[10] G. Wang, F. Gao, Y.-C. Wu, and C. Tellambura, "Joint carrier frequency offset and channel estimation for two-way relay networks," *IEEE Trans. Wireless Commun.*, vol. 10, no. 2, pp.456-465, Feb. 2011.

[11] C. S. Patel and G. L. Stuber, "Channel estimation for amplify and forward relay based cooperation diversity systems," *IEEE Trans. Wireless Commun.,* vol. 6, no. 6, pp. 3348-3356, June 2007.

[12] J. K. Cavers, "An analysis of pilot symbol assisted modulation for Rayleigh fading channels," *IEEE Trans. Veh. Technol.,* vol. 40, pp. 686-693, Nov. 1991.

[13] W. C. Jakes, *Microwave Mobile Communications*. Wiley, 1974.

[14] A. Gaston, W. Chriss, and E. Walker, "A multipath fading simulator for radio," *IEEE Trans. Veh. Technol.,* vol. VT-22, pp. 241-244, 1973.

[15] G. B. Giannakis and C. Tepedelenlioglu, "Basis expansion models and diversity techniques for blind identification and equalization of time varying channels," *Proc. IEEE,* vol. 86, pp. 1969-1986, Nov. 1998.

[16] M. Dong, L. Tong, and B. Sadler, "Optimal insertion of pilot symbols for transmissions over time-varying flat fading channels," *IEEE Trans. Signal Process.*, vol. 52, no. 5, pp. 1403-1418, May 2004.

[17] S. Savazzi and U. Spagnolini, "Optimizing training lengths and training intervals in time-varying fading channels," *IEEE Trans. Signal Process.*, vol. 57, no. 3, pp. 1098-1112, Mar. 2009.

[18] X. Ma, G. B. Giannakis, and S. Ohno, "Optimal training for block transmissions over doubly selective wireless fading channels," *IEEE Trans. Signal Process.,* vol. 41, no. 5, pp. 1351-1366, May 2003.

[19] A. P. Kannu and P. Schniter, "Design and analysis of MMSE pilot-aided cyclic-prefixed block transmissions for doubly selective channels," *IEEE Trans. Signal Process.,* vol. 56, no. 3, pp. 1148-1160, Mar. 2008.

[20] T. Whitworth, M. Ghogho, and D. McLernon, "Optimized training and basis expansion model parameters for doubly-selective channel estimation," *IEEE. Trans. Wireless Commun.,* vol. 8, no. 3, pp. 1490-1498, Mar. 2009.

[21] X. Ma and G. B. Giannakis, "Maximum-diversity transmissions over doubly-selective wireless channels," *IEEE Trans. Inf. Theory*, vol. 49, pp. 1832-1840, July 2003.

[22] A. S. Akki and F. Haber, "A statistical model for mobile-to-mobile land communication channel," *IEEE Trans. Veh. Technol.,* vol. VT-35, no. 1, pp. 2-7, Feb. 1986.

[23] D. K. Borah and B. T. Hart, "Frequency-selective fading channel estimation with a polynomial time-varying channel model," *IEEE Trans. Commun.,* vol. 47, no. 6, pp. 862-873, June 1999.

[24] M. Martone, "Wavelet-based separating kernels for sequences estimation with unknown rapidly time-varying channels," *IEEE Commun. Lett.* vol. 3, no. 3, pp. 78-80, Mar, 1999.

[25] T. Zemen and C. F. Mecklenbräuker, "Time-variant channel estimation using discrete prolate spheroidal sequences," *IEEE Trans. Signal Process.*, vol. 53, pp. 3597-3607, Sep. 2005.

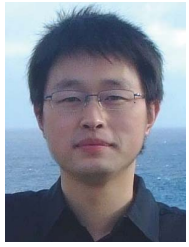
[26] C. K. Ho, R. Zhang, and Y.-C. Liang, "Two-way relaying over OFDM: optimized tone permutation and power allocation," in *Proc. IEEE ICC*, May 2008, pp. 3908-3912.

[27] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Prentice-Hall, 1993.

[28] S. Haykin, *Adaptive Filter Theory*. Prentice-Hall, 2002.

[29] M. Sandell, D. McNamara, and S. Parker, "Analysis of frequency-offset tracking in MIMO OFDM systems," *IEEE Trans. Commun.*, vol. 54, no. 8, pp. 1481-1491, Aug. 2006.

[30] H. Liu and G. B. Giannakis, "Deterministic approaches for blind equalization of time-varying channels with antenna arrays," *IEEE Trans. Signal Process.,* vol. 46, no. 11, pp. 3003-3013, Nov. 1998.



Gongpu Wang received the B.Eng. degree in communication engineering from Anhui University, Hefei, Anhui China, in 2001, the M.Sc. degree from Beijing University of Posts and Telecommunications, Beijing China, in 2004. From 2004 to 2007, he was a teacher in Beijing University of Posts and Telecommunications. He is currently working towards a Ph.D. degree at the Department of Electrical and Computer Engineering, University of Alberta, Edmonton, Canada. His research interests are wireless communication theory and signal processing for communications.

Feifei Gao (S'05-M'09) received the B.Eng. degree from Xi¡¯an Jiaotong University, Xi¡¯an, China, in 2002, the M.Sc. degree from McMaster University, Hamilton, ON, Canada, in 2004, and the Ph.D. degree from National University of Singapore, Singapore, in 2007. He was a Research Fellow with the Institute for Infocomm Research, A*STAR, Singapore, in 2008 and an Assistant Professor with the School of Engineering and Science, Jacobs University, Bremen, Germany, from 2009 to 2010. In 2011, he joined the Department of Automation, Tsinghua University, Beijing, China, where he is currently an Associate Professor. He also serves as an Adjunct Professor with the School of Engineering and Science, Jacobs University. He has coauthored more than 80 refereed IEEE journal and conference proceeding papers. His research interests are communication theory, broadband wireless communications, signal processing, multiple-input-multiple-output systems, and array signal processing.

Prof. Gao has served as a Technical Program Committee member for the IEEE International Conference on Communications, the IEEE Global Communications Conference, the IEEE Vehicular Technology Conference, and the IEEE Personal, Indoor, and Mobile Radio Communications Conference.

Wen Chen received BS and MS from Wuhan University, China in 1990 and 1993 respectively, and PhD from University of Electro-Communications, Tokyo, Japan in 1999. He was a researcher of Japan Society for the Promotion of Sciences (JSPS) from 1999 through 2001. In 2001, he joined University of Alberta, Canada, starting as a post-doctoral fellow in Information Research Laboratory and continuing as a research associate in the Department of Electrical and Computer Engineering. Since 2006, he has been a full professor in the Department of Electronic Engineering, Shanghai Jiaotong University, China, where he is also the director of Institute for Signal Processing and Systems. Dr. Chen was awarded the Ariyama Memorial Research Prize in 1997, the PIMS Post-Doctoral Fellowship in 2001. He received the honors of "New Century Excellent Scholar in China" in 2006 and "Pujiang Excellent Scholar in Shanghai" in 2007. He is elected to the vice general secretary of Shanghai Institute of Electronics in 2008. He is in the editorial board of the *International Journal of Wireless Communications and Networking*, and serves the *Journal of Communications*, *Journal of Computers*, *Journal of Networks* and *EURASIP Journal on Wireless Communications and Networking* as (lead) guest editors. He is the Technical Program Committee chair for IEEE-ICCSC2008, the General Conference Chair for IEEE-ICIS2009 and IEEE-WCNIS2010. He has published more than 100 papers in IEEE journals and conferences. His interests cover network coding, cooperative communications, cognitive radio, and MIMO-OFDM systems.

Chintha Tellambura (F'11) received the B.Sc. degree (with first-class honors) from the University of Moratuwa, Moratuwa, Sri Lanka, in 1986, the M.Sc. degree in electronics from the University of London, London, U.K., in 1988, and the Ph.D. degree in electrical engineering from the University of Victoria, Victoria, BC, Canada, in 1993.

He was a Postdoctoral Research Fellow with the University of Victoria (1993-1994) and the University of Bradford (1995-1996). He was with Monash University, Melbourne, Australia, from 1997 to 2002. Presently, he is a Professor with the Department of Electrical and Computer Engineering, University of Alberta. His research current interests include multiple antenna and multicarrier communication systems, cognitive radio and relay networks.

Prof. Tellambura is the Area Editor for Wireless Communications Systems and Theory in the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS and an Associate Editor for the IEEE TRANSACTIONS ON COMMUNICATIONS. He was Chair of the Communication Theory Symposium in Globecom'05 held in St. Louis, MO.