Research Article

Channel Frequency Response Estimation for MIMO Systems with Frequency-Domain Equalization

Yang Yang,[1] Zhiping Shi,[2] Yong Huat Chew,[3] and Tjeng Thiang Tjhung[3]

[1] *Department of Electrical and Computer Engineering, Lehigh University, 19 Memorial Drive West, Bethlehem, PA 18015, USA*
[2] *National Key Laboratory of Communication, University of Electronic Science and Technology of China, Chengdu 610054, Sichuan, China*
[3] *Institute for Infocomm Research, 1 Fusionpolis Way, #21-01 Connexis, Singapore 138632*

Correspondence should be addressed to Yang Yang, yay204@lehigh.edu

Since its recent adoption for the uplink transmissions in the next-generation cellular systems 3GPP long-term evolution (LTE) and LTE advanced, single-carrier frequency-domain equalization (SC-FDE), an effective technique to mitigate the distortion induced by long-spanning intersymbol interference has seen a surge of interest in the research community. Implementation of SC-FDE in multiple-input multiple-output (MIMO) systems usually requires, in advance, the channel information in terms of the channel frequency response (CFR). In this paper, we present a training-based CFR estimation scheme, which is hardware efficient when integrated with SC-FDE and space-time coding (STC) in MIMO systems. A thorough mean square error (MSE) analysis of this CFR estimation scheme is provided, where we consider linear estimators based on both least squares (LS) and minimum MSE (MMSE) criteria by assuming different knowledge of the channel statistics. More specifically, for the LS-based approach, we assume no *a priori* knowledge of the channel statistics is given other than the noise statistics, while for the MMSE-based method, we assume both the channel covariance matrix and the noise statistics are known. Given a constraint which effectively limits the transmit power of training signals, we also investigate the optimal design of training signals under both criteria. For the special case when the number of transmit antennas is equal to 2, we further demonstrate that the CFR estimation could be implemented in an adaptive manner by means of certain block-wise recursive algorithms. Extensive simulation results are provided, which demonstrate the efficacy of this CFR estimation scheme.

1. Introduction

The severe frequency selectivity often characterizing wideband radio channels would inevitably induce intersymbol interference (ISI) which can span over many symbol intervals. High-speed broadband wireless systems targeting data rate of tens of megabits or beyond should be, as a result, designed to mitigate the effect of such intense ISI. Traditionally, time-domain equalization (TDE) is a popular approach to compensate for ISI in single-carrier communication systems. But for wideband channels, TDE becomes unattractive as its complexity grows exponentially with channel memory or it requires very long finite impulse response filters to achieve acceptable performance. An alternative approach is the single-carrier frequency-domain equalization (SC-FDE), which has the advantage of large reduction in the computational complexity due to the use of the computationally efficient fast Fourier transform (FFT) (see [1–3] for a tutorial treatment). Even compared with orthogonal frequency-division multiplexing (OFDM), a well-recognized multicarrier solution to combat channel delay spread which also uses FFT, single-carrier transmission with FDE can handle the same channels with similar performance and essentially the same overall complexity but smaller peak-to-average transmitted power ratio [1]. This is particularly advantageous to mobile terminals and mobile personal assistants, as it can greatly alleviate the requirements on the radio frequency hardware at the transmitter, such as the digital-to-analog converter and the power amplifier, to name a few. For that reason, a technology named single-carrier frequency division multiple access (SC-FDMA), which is essentially based on SC-FDE, has been
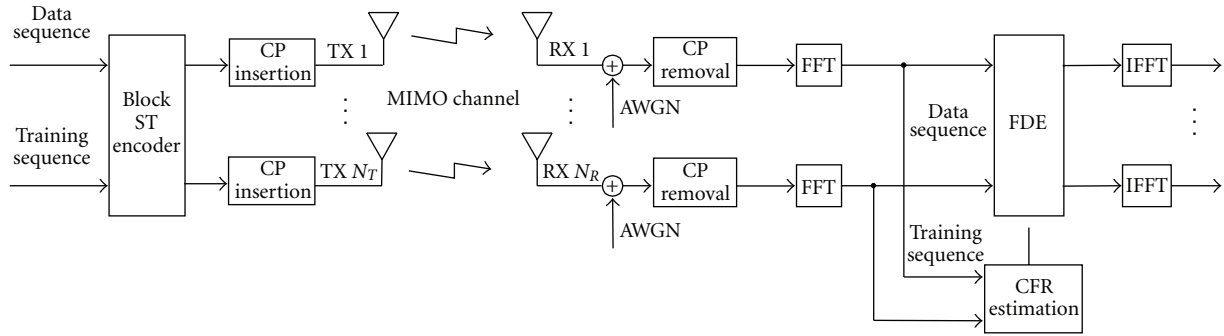
Figure 1: Block diagram of the CFR estimation for MIMO system with STC and SC-FDE.

adopted for the uplink transmissions in the next-generation cellular systems 3 GPP long-term evolution (LTE) and LTE advanced [4]. SC-FDE has thus grasped more attention in both academic and industrial circles.

SC-FDE has also been applied to multiple-input multiple-output (MIMO) communication systems. This, however, is often done jointly with space-time coding (STC), in order that the spatial diversity available in a MIMO system can be exploited to further mitigate the frequency selectivity, for example, [5–9]. For this case, properly designed ST block codes (STBCs) are generally required and there exist some works in that regard. For example, a time-reversal Alamouti-like STBC scheme with FDE was proposed firstly in [5]. This scheme is attractive as it can achieve full spatial diversity, and nearly full transmit rate if the cyclic prefix (CP) overhead is ignored. For SC-FDE in MIMO systems with more than 2 transmit antennas, a general block-level STC was proposed in [6] and a method based on quasi-orthogonal STBCs was proposed in [7].

Note that when performing FDE in MIMO systems, the channel frequency response between each transmit-receive antenna pair is usually required at the receiver to recover the transmitted signals [2, 3]. To obtain such channel frequency response (CFR) knowledge, one approach is to obtain the channel impulse response (CIR) firstly and then transfer it back to the frequency domain through FFT processing. As a result, the CFR estimation problem merely reduces to the problem of estimating the CIR in MIMO systems, which has been vigorously investigated over the years, for example, see [10] and references therein. As an alternative, one can apply the FFT firstly, and then estimate the CFR directly afterwards. In fact, we notice that this alternative approach, or the CFR estimation problem, has been studied, for example, in [11] for systems with single transmit and single receive antenna, and in [12] for SC-FDE in ultra-wideband communication systems. However, there does not seem to exist a lot of works which explore this alternative approach particularly for MIMO systems employing both STC and SC-FDE. This line of work merits interest on its own terms, for not only can it advance the existing knowledge on the subject of CFR estimation, but the CFR estimation scheme, when designed in a manner to be integrated with the techniques of STC and FDE in MIMO systems, can be amenable to system implementation, and has the potential

to induce less hardware complexity and cost. This basically motivates our work as detailed next.

In this paper, we present and investigate a CFR estimation scheme for MIMO systems with both STC and FDE. In this scheme, training sequences are encoded in space and time in a similar manner as data sequences. (We notice that the CIR estimation for MIMO channels using ST codes was considered in [13, 14].) In fact, the same set of coding hardware can be reused; thus, no additional hardware complexity is introduced at the transmitter and this is particularly suitable for mobile terminals. At the receiver, different from the traditional approach where CIR is obtained first then transferred to CFR, these training sequences are simply processed in a similar fashion as the data sequences, for example, CP removal and FFT processing. Following these procedures, estimation of the CFR can thus be done directly in the frequency domain. As the CFR estimation can make use of the existing FFT modules for FDE, fewer complexity or cost would be required at the receiver. This scheme is illustrated in Figure 1. Further, in this paper, we provide a thorough mean square error (MSE) analysis for the CFR estimation based on two criteria, least squares (LS) and minimum MSE (MMSE), by assuming different *a priori* knowledge of the channel statistics. More specifically, for the LS-based approach, we assume no *a priori* knowledge of the channel statistics is given other than the noise statistics, while for the MMSE-based method, we assume both the channel covariance matrix and the noise statistics are known. Under both criteria, we also study the optimal training sequence design by imposing a constraint on the transmit power of training sequences. Finally, we investigate the adaptive implementation of the proposed CFR estimation scheme for Alamouti-like transmissions. We provide several block-wise recursive algorithms to update the adaptive filter, and also study the convergence behaviors of these recursive algorithms.

The remainder of this paper is structured as follows. In Section 2, we describe the system model and the transmission scheme of the training sequences. In Section 3, we describe in detail the CFR estimation scheme for MIMO systems with more than 2 transmit antennas. We also investigate the optimal training sequence design under both LS and MMSE criteria. In Section 4, we focus on the special Alamouti case with 2 transmit antennas. We discuss an adaptive implementation of the CFR estimation scheme for

this special case, and provide a brief convergence analysis. In Section 5, we provide extensive simulation results and also compare with others' work to demonstrate the efficacy of this estimation approach. Section 6 concludes this paper.

Notation. Throughout this paper, we use bold upper case letters to denote matrices and bold lower case letters to signify column vectors. Superscript $\{\cdot\}^H$, $\{\cdot\}^*$, and $\{\cdot\}^T$ will be used to denote the complex conjugate transpose, conjugate, and transpose of a matrix or vector, respectively. We use diag$\{\mathbf{a}\}$ for a diagonal matrix with its diagonal vector given by $\mathbf{a}$, and $\otimes$ for Kronecker product. $\mathbf{I}_K$ denotes the identity matrix of size $K \times K$, and $\mathbf{0}_{M \times N}$ for a zero matrix of size $M \times N$. We use the subscript $\{\cdot\}_{\mathcal{F}}$ to denote the matrices or vectors in the frequency domain, and $(\cdot)^+$ for the nonnegative part of a real-valued scalar or matrix.

2. Signal and System Model

We consider an ST-coded MIMO system equipped with N_T transmit and N_R receive antennas. With symbol rate sampling, let $\mathbf{h}^{(p,q)} = [h^{(p,q)}(0), \ldots, h^{(p,q)}(\nu)]^T$ denote the equivalent baseband discrete-time CIR (including the transmit and receive filters as well as the multipath effect) between the pth transmit antenna and the qth receive antenna, where $1 \leq p \leq N_T$, $1 \leq q \leq N_R$, and ν is the channel order. We assume the channel is quasistatic, that is, its response remains time invariant within one ST-coded frame but can vary from frame to frame. We define N_S vectors of dimension $L \times 1$, $\{\mathbf{s}_i\}_{i=1}^{N_S}$ as the training sequences, where the symbols in $\mathbf{s}_i$ belong to the same alphabet $\mathcal{A}$, and L denotes the sequence length and is assumed to be at least equal to the number of multipaths, that is, $L \geq \nu + 1$. In this proposed CFR estimation scheme, the training sequence $\mathbf{s}_i$ is encoded in space and time, using the same ST block encoder for data sequences, as depicted in Figure 1. As a result of this, the same set of hardware can be reused without additional complexity and cost. As for the ST encoder, we adopt the code design described in [6]. It is an extension of the original orthogonal STBCs in [15, 16] for frequency-selective fading channels. This type of STBCs are capable of achieving full spatial diversity and are particularly amenable to FDE.

Without loss of generality, suppose the N_S training blocks are ST coded in a manner that they are transmitted over $N_c = 2N_S$ time slots, where a time slot is defined as the duration required to transmit a CP appended training block. Thus, the code rate is given by $\mathcal{R} = N_S/N_c = 1/2$. There exist some sporadic code designs which could achieve code rate higher than 1/2. For example, when $N_T = 3$ and 4, the code design with $\mathcal{R} = 3/4$ can be found in [16]. However, it has been proved in [17] that with complex signal constellation and under the orthogonality assumption, $\mathcal{R}$ cannot be greater than 3/4 for $N_T > 2$. For simplicity, in this part we only focus on the case of $\mathcal{R} = 1/2$ for $N_T > 2$. The special case of $\mathcal{R} = 1$ for $N_T = 2$ will be discussed in detail in Section 4.

Let $\{\mathbf{\Pi}_i\}_{i=1}^{N_S}$ be a set of $N_S \times N_T$ real-valued matrices of a full-rate generalized orthogonal STBC design for real

symbols. Entries of $\mathbf{\Pi}_i$ are either 0 or ± 1, and $\mathbf{\Pi}_i$ further satisfies the following conditions [18, Chapter 7]:

$$\mathbf{\Pi}_i^T \mathbf{\Pi}_i = \mathbf{I}_{N_T},$$

$$\mathbf{\Pi}_i^T \mathbf{\Pi}_j = -\mathbf{\Pi}_j^T \mathbf{\Pi}_i, \quad i \neq j. \tag{1}$$

Then, for the block-level generalized complex orthogonal STBC that is employed in our work, the code matrix, if denoted as $\mathcal{G} \in \mathbb{C}^{N_c L \times N_T}$, can be written as

$$\mathcal{G} = \sum_{i=1}^{N_S} \left(\mathbf{\Gamma}_{A_i} \otimes \mathbf{s}_i + \mathbf{\Gamma}_{B_i} \otimes \mathbf{P}_L^{(1)} \mathbf{s}_i^* \right), \tag{2}$$

where $\mathbf{\Gamma}_{A_i}$ and $\mathbf{\Gamma}_{B_i}$ are both $N_c \times N_T$ matrices, and are, respectively, defined as

$$\mathbf{\Gamma}_{A_i} = \begin{bmatrix} \mathbf{\Pi}_i \\ \mathbf{0}_{N_S \times N_T} \end{bmatrix}, \qquad \mathbf{\Gamma}_{B_i} = \begin{bmatrix} \mathbf{0}_{N_S \times N_T} \\ \mathbf{\Pi}_i \end{bmatrix}. \tag{3}$$

In (2), $\mathbf{P}_L^{(1)}$ is an $L \times L$ permutation matrix which performs a reverse cyclic shift when applied to an arbitrary $L \times 1$ vector, for example, suppose $\mathbf{s} = [s(0), s(1), \ldots s(L-1)]^T$, we then have

$$\mathbf{P}_L^{(1)} \mathbf{s}^* = [s^*(0), s^*(L-1), s^*(L-2), \ldots, s^*(1)]^T. \tag{4}$$

Given the properties of $\mathbf{\Pi}_i$ in (1), it can be easily verified that $\mathbf{\Gamma}_{A_i}$ and $\mathbf{\Gamma}_{B_i}$ have the following properties:

$$\mathbf{\Gamma}_{A_i}^T \mathbf{\Gamma}_{A_i} = \mathbf{I}_{N_T}, \qquad \mathbf{\Gamma}_{B_i}^T \mathbf{\Gamma}_{B_i} = \mathbf{I}_{N_T},$$

$$\mathbf{\Gamma}_{A_i}^T \mathbf{\Gamma}_{A_j} = -\mathbf{\Gamma}_{A_j}^T \mathbf{\Gamma}_{A_i}, \quad \mathbf{\Gamma}_{B_i}^T \mathbf{\Gamma}_{B_j} = -\mathbf{\Gamma}_{B_j}^T \mathbf{\Gamma}_{B_i}, \quad i \neq j, \tag{5}$$

$$\mathbf{\Gamma}_{A_i}^T \mathbf{\Gamma}_{B_j} = \mathbf{0}_{N_T \times N_T}, \quad \forall i, j.$$

Let $\mathcal{G}(:, i)$ denote the ith column of $\mathcal{G}$ that corresponds to the training blocks to be transmitted from the ith transmit antenna over N_c time slots. For notational convenience, we express the ith column of $\mathcal{G}$ as follows:

$$\mathcal{G}(:, i) = \sum_{m=1}^{N_S} \left[\mathbf{\Gamma}(:, i) \otimes \mathbf{s}_m + \mathbf{\Gamma}(:, i) \otimes \mathbf{P}_L^{(1)} \mathbf{s}_m^* \right]$$

$$= \left[\bar{\mathbf{s}}_i^T(1), \bar{\mathbf{s}}_i^T(2), \ldots, \bar{\mathbf{s}}_i^T(N_c) \right]^T, \tag{6}$$

where $i = 1, \ldots, N_T$. To give an example of $\mathcal{G}$, let us consider a code design with rate $\mathcal{R} = 1/2$ for $N_T = 3$, where $N_S = 4$ and $N_c = 8$. For this instance, $\mathcal{G}$ is illustrated as below

$$\mathcal{G} = \begin{pmatrix} \mathbf{s}_1 & \mathbf{s}_2 & \mathbf{s}_3 \\ -\mathbf{s}_2 & \mathbf{s}_1 & -\mathbf{s}_4 \\ -\mathbf{s}_3 & \mathbf{s}_4 & \mathbf{s}_1 \\ -\mathbf{s}_4 & -\mathbf{s}_3 & \mathbf{s}_2 \\ \mathbf{P}_L^{(1)} \mathbf{s}_1^* & \mathbf{P}_L^{(1)} \mathbf{s}_2^* & \mathbf{P}_L^{(1)} \mathbf{s}_3^* \\ -\mathbf{P}_L^{(1)} \mathbf{s}_2^* & \mathbf{P}_L^{(1)} \mathbf{s}_1^* & -\mathbf{P}_L^{(1)} \mathbf{s}_4^* \\ -\mathbf{P}_L^{(1)} \mathbf{s}_3^* & \mathbf{P}_L^{(1)} \mathbf{s}_4^* & \mathbf{P}_L^{(1)} \mathbf{s}_1^* \\ -\mathbf{P}_L^{(1)} \mathbf{s}_4^* & -\mathbf{P}_L^{(1)} \mathbf{s}_3^* & \mathbf{P}_L^{(1)} \mathbf{s}_2^* \end{pmatrix} = \begin{pmatrix} \bar{\mathbf{s}}_1(1) & \bar{\mathbf{s}}_2(1) & \bar{\mathbf{s}}_3(1) \\ \bar{\mathbf{s}}_1(2) & \bar{\mathbf{s}}_2(2) & \bar{\mathbf{s}}_3(2) \\ \bar{\mathbf{s}}_1(3) & \bar{\mathbf{s}}_2(3) & \bar{\mathbf{s}}_3(3) \\ \bar{\mathbf{s}}_1(4) & \bar{\mathbf{s}}_2(4) & \bar{\mathbf{s}}_3(4) \\ \bar{\mathbf{s}}_1(5) & \bar{\mathbf{s}}_2(5) & \bar{\mathbf{s}}_3(5) \\ \bar{\mathbf{s}}_1(6) & \bar{\mathbf{s}}_2(6) & \bar{\mathbf{s}}_3(6) \\ \bar{\mathbf{s}}_1(7) & \bar{\mathbf{s}}_2(7) & \bar{\mathbf{s}}_3(7) \\ \bar{\mathbf{s}}_1(8) & \bar{\mathbf{s}}_2(8) & \bar{\mathbf{s}}_3(8) \end{pmatrix}. \tag{7}$$

After ST coded, the transmission structure of the training sequences is shown in Table 1.

To avoid the interblock interference from preceding information or training sequences, a CP with a length of ν is inserted for each block before transmission. Then, at time slot k, the training sequence $\bar{\mathbf{s}}_p(k)$ is forwarded to the pth transmit antenna after CP insertion. The length of total training symbols from each transmit antenna, denoted as N_b, is equal to $N_b = N_c(L + \nu)$, and its minimum length is $N_b = N_c(2\nu + 1)$ when L is chosen to be equal to $\nu + 1$.

3. CFR Estimation for MIMO Transmissions ($N_T > 2$)

At the receiver, symbols corresponding to the CP are discarded. Thus, the received signal at the qth receive antenna at time slot k can be written as

$$\mathbf{x}_q(k)$$
$$= \sum_{p=1}^{N_T} \mathbf{H}^{(p,q)} \bar{\mathbf{s}}_p(k) + \mathbf{n}_q(k), \quad q = 1, \ldots, N_R, \ k = 1, \ldots, N_c, \tag{8}$$

where $\mathbf{H}^{(p,q)}$ is an $L \times L$ channel matrix with its (k, l)th entry given by $h^{(p,q)}((k-l) \bmod L)$, and $\mathbf{n}_q(k)$ denotes the additive white Gaussian noise (AWGN) vector. It is easy to verify that $\mathbf{H}^{(p,q)}$ is a circulant matrix. Thus, its eigen matrix is the FFT matrix, or in other words, its eigendecomposition can be written as

$$\mathbf{H}^{(p,q)} = \mathbf{F}_L^H \cdot \mathrm{diag}\left(\mathbf{h}_{\mathcal{F}}^{(p,q)}\right) \cdot \mathbf{F}_L. \tag{9}$$

$\mathbf{F}_L$ is the orthonormal FFT matrix whose (k, l)th entry is given by

$$\mathbf{F}_L(k, l) = \frac{1}{\sqrt{L}} \exp\left\{ \frac{-j 2\pi (k-1)(l-1)}{L} \right\}, \tag{10}$$

where $k = 1, \ldots, L$ and $l = 1, \ldots, L$. If denoting $\mathbf{D}_{\mathcal{F}}^{(p,q)} = \mathrm{diag}(\mathbf{h}_{\mathcal{F}}^{(p,q)})$, we have

$$\mathbf{D}_{\mathcal{F}}^{(p,q)}(i, i) = \mathbf{h}_{\mathcal{F}}^{(p,q)}(i) = \sum_{k=0}^{\nu} h^{(p,q)}(k) e^{-j 2\pi k(i-1)/L}, \tag{11}$$

where $i = 1, \ldots, L$. Applying the FFT operations on both sides of (8), we obtain

$$\mathbf{x}_{q\mathcal{F}}(k) = \sum_{p=1}^{N_T} \mathbf{D}_{\mathcal{F}}^{(p,q)} \bar{\mathbf{s}}_{p\mathcal{F}}(k) + \mathbf{n}_{q\mathcal{F}}(k), \tag{12}$$

where $\mathbf{x}_{q\mathcal{F}}(k) = \mathbf{F}_L \mathbf{x}_q(k)$, $\bar{\mathbf{s}}_{p\mathcal{F}}(k) = \mathbf{F}_L \bar{\mathbf{s}}_p(k)$, and $\mathbf{n}_{q\mathcal{F}}(k) = \mathbf{F}_L \mathbf{n}_q(k)$.

Since $\mathbf{D}_{\mathcal{F}}^{(p,q)}$ is diagonal, we can rewrite (12) into

$$\mathbf{x}_{q\mathcal{F}}(k) = \sum_{p=1}^{N_T} \bar{\mathbf{S}}_{p\mathcal{F}}(k) \mathbf{h}_{\mathcal{F}}^{(p,q)} + \mathbf{n}_{q\mathcal{F}}(k), \tag{13}$$

TABLE 1: Transmission structure of training sequences ($N_T > 2$).

	1	$\cdots$	N_c
TX 1	$\bar{\mathbf{s}}_1(1)$	$\cdots$	$\bar{\mathbf{s}}_1(N_c)$
$\vdots$	$\vdots$	$\ddots$	$\vdots$
TX N_T	$\bar{\mathbf{s}}_{N_T}(1)$	$\cdots$	$\bar{\mathbf{s}}_{N_T}(N_c)$

where $\bar{\mathbf{S}}_{p\mathcal{F}}(k) = \mathrm{diag}\{\bar{\mathbf{s}}_{p\mathcal{F}}(k)\}$. Stacking N_c blocks of received signals at the qth receive antenna, we have

$$\underbrace{\begin{bmatrix} \mathbf{x}_{q\mathcal{F}}(1) \\ \vdots \\ \mathbf{x}_{q\mathcal{F}}(N_c) \end{bmatrix}}_{\mathbf{x}_{\mathcal{F}}^q} = \underbrace{\begin{bmatrix} \bar{\mathbf{S}}_{1\mathcal{F}}(1) & \cdots & \bar{\mathbf{S}}_{N_T\mathcal{F}}(1) \\ \vdots & \ddots & \vdots \\ \bar{\mathbf{S}}_{1\mathcal{F}}(N_c) & \cdots & \bar{\mathbf{S}}_{N_T\mathcal{F}}(N_c) \end{bmatrix}}_{\mathbf{\mathcal{S}}_{\mathcal{F}}} \times \underbrace{\begin{bmatrix} \mathbf{h}_{\mathcal{F}}^{(1,q)} \\ \vdots \\ \mathbf{h}_{\mathcal{F}}^{(N_T,q)} \end{bmatrix}}_{\mathbf{h}_{\mathcal{F}}^q}$$
$$+ \underbrace{\begin{bmatrix} \mathbf{n}_{1\mathcal{F}}(1) \\ \vdots \\ \mathbf{n}_{q\mathcal{F}}(N_c) \end{bmatrix}}_{\mathbf{n}_{\mathcal{F}}^q} \tag{14}$$

or in a more simplified form

$$\mathbf{x}_{\mathcal{F}}^q = \mathbf{\mathcal{S}}_{\mathcal{F}} \mathbf{h}_{\mathcal{F}}^q + \mathbf{n}_{\mathcal{F}}^q. \tag{15}$$

Collecting the received signals across all those N_R receive antennas, we obtain the received data matrix $\mathbf{X}_{\mathcal{F}} = [\mathbf{x}_{\mathcal{F}}^1, \ldots, \mathbf{x}_{\mathcal{F}}^{N_R}]$, which is expressed as

$$\mathbf{X}_{\mathcal{F}} = \mathbf{\mathcal{S}}_{\mathcal{F}} \mathbf{H}_{\mathcal{F}} + \mathbf{N}_{\mathcal{F}}, \tag{16}$$

where $\mathbf{H}_{\mathcal{F}} = [\mathbf{h}_{\mathcal{F}}^1, \ldots, \mathbf{h}_{\mathcal{F}}^{N_R}]$ and $\mathbf{N}_{\mathcal{F}} = [\mathbf{n}_{\mathcal{F}}^1, \ldots, \mathbf{n}_{\mathcal{F}}^{N_R}]$. Thus, our task is to recover the CFR $\mathbf{H}_{\mathcal{F}}$ from (16).

Additionally, let us denote $\mathbf{h}^q = [\mathbf{h}^{(1,q)^T}, \ldots, \mathbf{h}^{(N_T,q)^T}]^T$ as the corresponding CIR associated with the qth antenna, and stack all the CIR across N_R receive antennas in matrix $\mathbf{H} = [\mathbf{h}^1, \ldots, \mathbf{h}^{N_R}]$. We further define the compound inverse FFT (IFFT) matrix $\mathbf{\mathcal{F}}_{N_T}^H = \mathbf{I}_{N_T} \otimes \mathbf{F}_L^H$, and the compound transmit matrix $\mathbf{\mathcal{T}}_{N_T} = \mathbf{I}_{N_T} \otimes [\mathbf{I}_{\nu+1} \mid \mathbf{0}_{(\nu+1) \times (L-\nu-1)}]$. Therefore, the corresponding CIR estimate can be computed by

$$\hat{\mathbf{H}} = \frac{1}{\sqrt{L}} \mathbf{\mathcal{T}}_{N_T} \mathbf{\mathcal{F}}_{N_T}^H \hat{\mathbf{H}}_{\mathcal{F}}, \tag{17}$$

where $\hat{\mathbf{H}}_{\mathcal{F}}$ is the CFR estimate for $\mathbf{H}_{\mathcal{F}}$. In the sequel, we discuss the linear CFR estimators based on both LS and MMSE criteria, along with the respective optimal designs of training sequences.

3.1. LS Estimator with Power Constraint.
For the convenience of ensuing analysis, we explicitly make the following assumption.

(A1) All noise components are assumed to be complex, independently and identically Gaussian distributed with zero mean and variance σ_n^2. Thus,

we have $\mathbf{n}_{\mathcal{F}}^q \sim \mathcal{CN}(\mathbf{0}_{N_cL \times 1}, \sigma_n^2 \mathbf{I}_{N_cL})$ and $\mathbf{N}_{\mathcal{F}} \sim \mathcal{CN}(\mathbf{0}_{N_cL \times N_R}, \sigma_n^2 N_R \mathbf{I}_{N_cL})$.

Except for the noise statistics, we assume no *a priori* knowledge of the channel parameters (e.g., the covariance matrix of the CFR) is given, and we only consider the conventional LS method. Therefore, the unique LS solution $\hat{\mathbf{H}}_{\mathcal{F}}$ that minimizes the cost function defined by $\|\mathbf{X}_{\mathcal{F}} - \mathbf{\mathcal{S}}_{\mathcal{F}} \mathbf{H}_{\mathcal{F}}\|^2$ can be written as

$$\hat{\mathbf{H}}_{\mathcal{F}} = \left(\mathbf{\mathcal{S}}_{\mathcal{F}}^H \mathbf{\mathcal{S}}_{\mathcal{F}}\right)^{-1} \mathbf{\mathcal{S}}_{\mathcal{F}}^H \mathbf{X}_{\mathcal{F}}. \tag{18}$$

It should be noted that if we want to obtain the CFR with a length greater than the default length L, interpolation is needed.

Based on assumption (A1), it is clear that this estimate is *unbiased* since $\mathrm{E}\{\hat{\mathbf{H}}_{\mathcal{F}}\} = \mathbf{H}_{\mathcal{F}}$. Let us define the CFR estimation error as $\mathbf{E}_{\mathcal{F}} = \hat{\mathbf{H}}_{\mathcal{F}} - \mathbf{H}_{\mathcal{F}}$. Using (16) and (18), we obtain

$$\mathbf{E}_{\mathcal{F}} = \left(\mathbf{\mathcal{S}}_{\mathcal{F}}^H \mathbf{\mathcal{S}}_{\mathcal{F}}\right)^{-1} \mathbf{\mathcal{S}}_{\mathcal{F}}^H \mathbf{N}_{\mathcal{F}}. \tag{19}$$

Its correlation matrix, $\mathbf{R}_{\mathbf{E}_{\mathcal{F}}} = \mathrm{E}\{\mathbf{E}_{\mathcal{F}} \mathbf{E}_{\mathcal{F}}^H\}$, can be calculated through

$$\mathbf{R}_{\mathbf{E}_{\mathcal{F}}} = \sigma_n^2 N_R \left(\mathbf{\mathcal{S}}_{\mathcal{F}}^H \mathbf{\mathcal{S}}_{\mathcal{F}}\right)^{-1}. \tag{20}$$

Thus, the MSE for this CFR estimation is given by

$$\mathrm{E}\left\{\|\mathbf{E}_{\mathcal{F}}\|^2\right\} = \mathrm{tr}\{\mathbf{R}_{\mathbf{E}_{\mathcal{F}}}\} = \sigma_n^2 N_R \cdot \mathrm{tr}\left\{\left(\mathbf{\mathcal{S}}_{\mathcal{F}}^H \mathbf{\mathcal{S}}_{\mathcal{F}}\right)^{-1}\right\}. \tag{21}$$

Now we consider the problem of designing the matrix $\mathbf{\mathcal{S}}_{\mathcal{F}}$ so that the estimation error is minimized. To have a reasonable solution, it is necessary to impose a constraint to limit the power of training sequences. Let such a constraint be $\|\mathbf{\mathcal{S}}_{\mathcal{F}}\|^2 \le \mathcal{P}_0$, where $\mathcal{P}_0$ is a given constant. Note that the power used in the cyclic prefix is not included in this formulation. Mathematically, this power constraint can also be written as $\mathrm{tr}\{\mathbf{\mathcal{S}}_{\mathcal{F}}^H \mathbf{\mathcal{S}}_{\mathcal{F}}\} \le \mathcal{P}_0$. For simplicity, we start with a general problem formulation, without examining the structure of the data matrix $\mathbf{\mathcal{S}}_{\mathcal{F}}$ but only assuming it has full rank. Therefore, our task is to find $\mathbf{\mathcal{S}}_{\mathcal{F}}$ that minimizes the MSE subject to the power constraint given above. This constrained optimization problem can be cast as

$$\min_{\mathbf{\mathcal{S}}_{\mathcal{F}}} \quad \mathrm{tr}\left\{\left(\mathbf{\mathcal{S}}_{\mathcal{F}}^H \mathbf{\mathcal{S}}_{\mathcal{F}}\right)^{-1}\right\}, \tag{22}$$
$$\mathrm{s.t.} \quad \mathrm{tr}\left\{\mathbf{\mathcal{S}}_{\mathcal{F}}^H \mathbf{\mathcal{S}}_{\mathcal{F}}\right\} \le \mathcal{P}_0.$$

To solve this problem, the following lemma will be useful.

Lemma 1. *For any $M \times M$ positive semidefinite Hermitian matrix $\mathbf{A}$ with its (i, j)th entry given by a_{ij}, the following inequality holds*

$$\mathrm{tr}(\mathbf{A}^{-1}) \ge \sum_{i=1}^M \frac{1}{a_{ii}}, \tag{23}$$

where the equality is achieved if and only if $\mathbf{A}$ is diagonal.

Applying this lemma and the method of Lagrange multipliers [19], we could readily solve this optimization problem. For brevity, we omit the details and simply provide the solution

$$\mathbf{\mathcal{S}}_{\mathcal{F}}^H \mathbf{\mathcal{S}}_{\mathcal{F}} = \frac{\mathcal{P}_0}{N_T L} \mathbf{I}_{N_T L}, \tag{24}$$

which means that the diagonal entries of $\mathbf{\mathcal{S}}_{\mathcal{F}}^H \mathbf{\mathcal{S}}_{\mathcal{F}}$ have the same value. Re-examining the matrix $\mathbf{\mathcal{S}}_{\mathcal{F}}$ as defined in (14) and its relation to $\mathbf{\mathcal{G}}$ in (2), we find that due to the orthogonal structure of the ST code, $\mathbf{\mathcal{S}}_{\mathcal{F}}^H \mathbf{\mathcal{S}}_{\mathcal{F}}$ is precisely diagonal. Moreover, recall $\{\mathbf{s}_i\}_{i=1}^{N_S}$ are training sequences, we define $\mathbf{s}_{i\mathcal{F}} = \mathbf{F}_L \mathbf{s}_i$ and $\mathbf{S}_{i\mathcal{F}} = \mathrm{diag}\{\mathbf{s}_{i\mathcal{F}}\}$ for $i = 1, \ldots, N_S$. Then, we arrive at the following result.

Theorem 1. *The following equality holds*

$$\mathbf{\mathcal{S}}_{\mathcal{F}}^H \mathbf{\mathcal{S}}_{\mathcal{F}} = \mathbf{I}_{N_T} \otimes \left\{2 \sum_{i=1}^{N_S} \mathbf{S}_{i\mathcal{F}}^H \mathbf{S}_{i\mathcal{F}}\right\}. \tag{25}$$

Proof. $\mathbf{\mathcal{S}}_{\mathcal{F}}^H \mathbf{\mathcal{S}}_{\mathcal{F}}$ is an $N_T L \times N_T L$ matrix and can be expressed in the block matrix form as

$$\mathbf{\mathcal{S}}_{\mathcal{F}}^H \mathbf{\mathcal{S}}_{\mathcal{F}} = \begin{pmatrix} \mathbf{\Xi}_{1,1} & \cdots & \mathbf{\Xi}_{1,N_T} \\ \vdots & \ddots & \vdots \\ \mathbf{\Xi}_{N_T,1} & \cdots & \mathbf{\Xi}_{N_T,N_T} \end{pmatrix}, \tag{26}$$

where $\mathbf{\Xi}_{i,j}$, $i = 1, \ldots, N_T$, $j = 1, \ldots, N_T$, is a square matrix of size $L \times L$. According to both (6) and (14), $\mathbf{\Xi}_{i,j}$ can be expressed as

$$\begin{aligned}
\mathbf{\Xi}_{i,j} &= \left[(\mathbf{I}_{N_c} \otimes \mathbf{F}_L) \cdot \mathbf{\mathcal{G}}(:, i)\right]^H \left[(\mathbf{I}_{N_c} \otimes \mathbf{F}_L) \cdot \mathbf{\mathcal{G}}(:, j)\right] \\
&= \left\{\sum_{m=1}^{N_s} \left[\mathbf{\Gamma}_{A_m}^T(:, i) \otimes \mathbf{S}_{m\mathcal{F}}^H + \mathbf{\Gamma}_{B_m}^T(:, i) \otimes \mathbf{S}_{m\mathcal{F}}\right]\right\} \\
&\quad \times \left\{\sum_{n=1}^{N_s} \left[\mathbf{\Gamma}_{A_n}(:, j) \otimes \mathbf{S}_{n\mathcal{F}} + \mathbf{\Gamma}_{B_n}(:, j) \otimes \mathbf{S}_{n\mathcal{F}}^H\right]\right\}.
\end{aligned} \tag{27}$$

To simplify (27), we need to use the *mixed-product property* of Kronecker product, that is, $(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = \mathbf{AC} \otimes \mathbf{BD}$, where $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$, and $\mathbf{D}$ are matrices of such size that one can form the matrix products $\mathbf{AC}$ and $\mathbf{BD}$. Further, given the properties of $\mathbf{\Gamma}_{A_m}$ and $\mathbf{\Gamma}_{B_n}$ in (5), we have the following:

$$\mathbf{\Gamma}_{A_m}^T(:, i) \mathbf{\Gamma}_{A_n}(:, j) = \begin{cases} 1, & m = n, \ i = j, \\ 0, & m = n, \ i \ne j, \\ -\mathbf{\Gamma}_{A_n}^T(:, i) \mathbf{\Gamma}_{A_m}(:, j), & m \ne n. \end{cases} \tag{28}$$

Similar properties also hold for $\mathbf{\Gamma}_{B_m}^T(:, i) \mathbf{\Gamma}_{B_n}(:, j)$. Moreover, we have

$$\mathbf{\Gamma}_{A_m}^T(:, i) \mathbf{\Gamma}_{B_n}(:, j) = \mathbf{\Gamma}_{B_m}^T(:, i) \mathbf{\Gamma}_{A_n}(:, j) = 0, \quad \forall i, j, m, n. \tag{29}$$

Based on the above properties, (27) can be simplified into

$$\Xi_{i,j} = \begin{cases} 2\sum_{i=1}^{N_S} \mathbf{S}_{i\mathcal{F}}^H \mathbf{S}_{i\mathcal{F}}, & i = j, \\ 0, & i \neq j. \end{cases} \tag{30}$$

Plugging (30) into (26), we then obtain (25). □

Based on (24) and (25), we summarize the following result.

Theorem 2. *The optimal training signals under the LS criterion should satisfy the following condition:*

$$\sum_{i=1}^{N_S} \mathbf{S}_{i\mathcal{F}}^H \mathbf{S}_{i\mathcal{F}} = \frac{\mathcal{P}_0}{2N_T L} \mathbf{I}_L. \tag{31}$$

This condition is the same as

$$\sum_{i=1}^{N_S} |s_{i\mathcal{F}}(j)|^2 = \frac{\mathcal{P}_0}{2N_T L}, \quad \forall j \in [1, L], \tag{32}$$

where $s_{i\mathcal{F}}(j)$ denotes the jth element of $\mathbf{s}_{i\mathcal{F}}$.

Of note is that although Theorem 2 states the conditions for training signals to be optimal in the sense of achieving the minimum value of MSE, it does not mean any sequences which satisfy (32) would be suitable for practical applications. This is because practical implementation of communication systems will inevitably impose some additional constraints on the sequences. To give an example, let us consider the CP-based communication systems. These systems are usually plagued by the well-known peak-to-average ratio (PAR) problem; thus, sequences with lower PAR values are, in general, more preferred in practice, for they can greatly alleviate the requirement on the power amplifier. Under this circumstance, training sequences which not only satisfy (32) but have a constant magnitude in both the time domain and the frequency domain would lend themselves to be a superior choice, for they are able to successfully preclude the PAR problem while achieving the minimum value of MSE. Chu sequences [20] and the class of training sequences proposed in [21] are examples of those sequences. Finally, the resulting minimum value of MSE can be calculated by

$$\mathrm{E}\left\{\|\mathbf{E}_{\mathcal{F}}\|^2\right\} = \sum_{j=1}^{L} \frac{N_T N_R \sigma_n^2}{2\sum_{i=1}^{N_S} |s_{i\mathcal{F}}(j)|^2} = \frac{\sigma_n^2 N_R (N_T L)^2}{\mathcal{P}_0}. \tag{33}$$

3.2. MMSE Estimator with Power Constraint. In this section, we consider the linear MMSE estimation of the CFR as well as the optimal training sequence design. For simplicity, we consider only the CFR associated with the qth receive antenna, that is, $\mathbf{h}_{\mathcal{F}}^q$, which was defined in (14). Besides assumption (A1), we make one additional assumption about the channel statistics as follows.

(A2) The CFR $\mathbf{h}_{\mathcal{F}}^q$ is a Gaussian random vector with zero mean and full-rank covariance matrix $\boldsymbol{\Sigma}_q$.

For convenience, we denote $\boldsymbol{\Sigma}_q$ by $\boldsymbol{\Sigma}$. Since $\mathbf{h}_{\mathcal{F}}^q = (\mathbf{I}_{N_T} \otimes \mathbf{F}_L)\mathbf{h}^q$, we have

$$\boldsymbol{\Sigma} = (\mathbf{I}_{N_T} \otimes \mathbf{F}_L) \cdot \mathrm{E}\left\{\mathbf{h}^q(\mathbf{h}^q)^H\right\} \cdot \left(\mathbf{I}_{N_T} \otimes \mathbf{F}_L^H\right), \tag{34}$$

where $\mathrm{E}\{\mathbf{h}^q(\mathbf{h}^q)^H\}$ is the covariance matrix of the corresponding CIR.

The MMSE estimate of the CFR can be computed through

$$\hat{\mathbf{h}}_{\mathcal{F}}^q = \left(\boldsymbol{\mathcal{S}}_{\mathcal{F}}^H \boldsymbol{\mathcal{S}}_{\mathcal{F}} + \sigma_n^2 \boldsymbol{\Sigma}^{-1}\right)^{-1} \boldsymbol{\mathcal{S}}_{\mathcal{F}}^H \mathbf{x}_{\mathcal{F}}^q. \tag{35}$$

We define the CFR estimation error as $\mathbf{e}_{\mathcal{F}}^q = \hat{\mathbf{h}}_{\mathcal{F}}^q - \mathbf{h}_{\mathcal{F}}^q$, then the resulting MSE can be expressed as

$$\mathrm{E}\left\{\left\|\mathbf{e}_{\mathcal{F}}^q\right\|^2\right\} = \mathrm{tr}\left\{\left(\sigma_n^{-2}\boldsymbol{\mathcal{S}}_{\mathcal{F}}^H \boldsymbol{\mathcal{S}}_{\mathcal{F}} + \boldsymbol{\Sigma}^{-1}\right)^{-1}\right\}. \tag{36}$$

Similar to the approach that we took in Section 3.1, we also impose a power constraint, and the design problem can be formulated into

$$\min_{\boldsymbol{\mathcal{S}}_{\mathcal{F}}} \quad \mathrm{tr}\left\{\left(\sigma_n^{-2}\boldsymbol{\mathcal{S}}_{\mathcal{F}}^H \boldsymbol{\mathcal{S}}_{\mathcal{F}} + \boldsymbol{\Sigma}^{-1}\right)^{-1}\right\}$$
$$\text{s.t.} \quad \mathrm{tr}\left\{\boldsymbol{\mathcal{S}}_{\mathcal{F}}^H \boldsymbol{\mathcal{S}}_{\mathcal{F}}\right\} \leq \mathcal{P}_0. \tag{37}$$

Note that $\boldsymbol{\Sigma}$ can be diagonalized through its eigenvalue decomposition, that is,

$$\boldsymbol{\Sigma} = \mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^H, \tag{38}$$

where $\mathbf{V}$ is a unitary matrix whose columns are eigenvectors of $\boldsymbol{\Sigma}$, and $\boldsymbol{\Lambda}$ is a nonnegative and diagonal matrix consisting of all the eigenvalues of $\boldsymbol{\Sigma}$. Then, (36) can be reformulated into

$$\mathrm{E}\left\{\left\|\mathbf{e}_{\mathcal{F}}^q\right\|^2\right\} = \mathrm{tr}\left\{\left(\sigma_n^{-2}\boldsymbol{\Psi}^H \boldsymbol{\Psi} + \boldsymbol{\Lambda}^{-1}\right)^{-1}\right\}, \tag{39}$$

where $\boldsymbol{\Psi} = \boldsymbol{\mathcal{S}}_{\mathcal{F}}\mathbf{V}$ is an $N_c L \times N_T L$ matrix. As $\mathbf{V}$ is unitary, it follows $\mathrm{tr}\{\boldsymbol{\mathcal{S}}_{\mathcal{F}}^H \boldsymbol{\mathcal{S}}_{\mathcal{F}}\} = \mathrm{tr}\{\boldsymbol{\Psi}^H \boldsymbol{\Psi}\}$. According to Lemma 1, the minimum value of $\mathrm{E}\{\|\mathbf{e}_{\mathcal{F}}^q\|^2\}$ is attained when $(\sigma_n^{-2}\boldsymbol{\Psi}^H \boldsymbol{\Psi} + \boldsymbol{\Lambda}^{-1})$ is diagonal. Let $\mathbf{Q} = \boldsymbol{\Psi}^H \boldsymbol{\Psi}$, then $\mathbf{Q}$ must be a diagonal matrix with elements $\mathbf{Q}_{ii} \geq 0$, for $i = 1, \ldots, N_T L$. Consequently, we can reformulate the optimization problem into

$$\min_{\mathbf{Q}} \quad \mathrm{tr}\left\{\left(\sigma_n^{-2}\mathbf{Q} + \boldsymbol{\Lambda}^{-1}\right)^{-1}\right\},$$
$$\text{s.t.} \quad \mathrm{tr}\{\mathbf{Q}\} \leq \mathcal{P}_0. \tag{40}$$

Using the method of Lagrange multipliers [19], we can obtain the following solution to the modified optimization problem

$$\mathbf{Q}_{ii} = \left(\tau - \frac{\sigma_n^2}{\boldsymbol{\Lambda}_{ii}}\right)^+, \quad \forall i \in [1, N_T L], \tag{41}$$

where $\boldsymbol{\Lambda}_{ii}$ denotes the (i, i)th element of $\boldsymbol{\Lambda}$, and the value of τ can be found by solving

$$\sum_{i=1}^{N_T L} \left(\tau - \frac{\sigma_n^2}{\boldsymbol{\Lambda}_{ii}}\right)^+ = \mathcal{P}_0. \tag{42}$$

Alternatively, $\mathbf{Q}$ can be rewritten as

$$\mathbf{Q} = \left(\tau \mathbf{I}_{N_T L} - \sigma_n^2 \mathbf{\Lambda}^{-1} \right)^+. \tag{43}$$

Thus, the resulting MSE can be computed through

$$\mathrm{E}\left\{ \left\| \mathbf{e}_{\mathcal{F}}^q \right\|^2 \right\} = \sum_{i=1}^{N_T L} \frac{\mathbf{\Lambda}_{ii}}{(\mathbf{\Lambda}_{ii} \sigma_n^{-2} \tau - 1)^+ + 1}. \tag{44}$$

It is worth noting that $\mathbf{\Psi}^H \mathbf{\Psi}$ is invariant to the post-multiplication of $\mathbf{\Psi}$ by a semi-orthogonal matrix. Thus, given the optimal solution for $\mathbf{Q}$ in (43), a general solution for $\mathbf{\Psi}$ can be composed as $\mathbf{\Psi} = \mathbf{Z}\mathbf{Q}^{1/2}$, where $\mathbf{Z}$ is an $N_c L \times N_T L$ matrix with its column forming an orthonormal basis. Since $\mathbf{\Psi} = \mathcal{S}_{\mathcal{F}} \mathbf{V}$, it is clear that the necessary condition for $\mathcal{S}_{\mathcal{F}}$ to be optimum is $\mathcal{S}_{\mathcal{F}} = \mathbf{Z}\mathbf{Q}^{1/2}\mathbf{V}^H$. Meanwhile, we have $\mathcal{S}_{\mathcal{F}}^H \mathcal{S}_{\mathcal{F}} = \mathbf{V}\mathbf{Q}\mathbf{V}^H$ and both sides are diagonal matrices. Considering the structure of $\mathcal{S}_{\mathcal{F}}$ in (14) and applying Theorem 1, we are thus led to following result.

Theorem 3. *The optimal training signals under the MMSE criterion should satisfy the following condition for a specific channel statistics (i.e., $\mathbf{\Sigma}$)*

$$\mathbf{V} \cdot \left(\tau \mathbf{I}_{N_T L} - \sigma_n^2 \mathbf{\Lambda}^{-1} \right)^+ \cdot \mathbf{V}^H = \mathbf{I}_{N_T} \otimes \left[2 \sum_{i=1}^{N_S} \mathbf{S}_{i\mathcal{F}}^H \mathbf{S}_{i\mathcal{F}} \right]. \tag{45}$$

Equation (45) specifies the essential characteristics of the optimum sequence under the MMSE criterion. It indicates that the optimal design should employ a water-filling type power allocation. Evidently, the structure of the covariance matrix $\mathbf{\Sigma}$ will have a large impact on the optimal training signal design. For example, when $\mathbf{\Sigma}$ is diagonal, then from (34), we can see that $\mathrm{E}\{\mathbf{h}^q (\mathbf{h}^q)^H\}$ can be a block circulant matrix, and the optimum condition (45) would represent a water-filling in power distribution with respect to the power spectral density samples of the CIR. For this special case, the optimal sequence may be generated through the frequency-domain water-filling. For cases where $\mathbf{\Sigma}$ is not diagonal, the optimal condition (45) may need to be jointly considered with the Kronecker product approximation in [22]. We omit further discussions for brevity.

4. CFR Estimation for Alamouti-Like Transmissions

Here, we study the CFR estimation for the special case of $N_T = 2$ and $N_R = 1$. This corresponds to the Alamouti-type transmission, where $N_S = N_c = 2$ and $\mathcal{R} = 1$. The transmission structure for the training sequences is illustrated in Table 2. The length of total training symbols from each transmit antenna, N_b, is equal to $N_b = 2(L + \nu)$, and its minimum length is $N_b = 4\nu + 2$ when L is chosen to be the minimum value $\nu + 1$. At the receiver, CPs are removed, which yields the channel input-output relationship in matrix vector form as

$$\mathbf{x}_1(k) = \mathbf{H}^{(1,1)} \mathbf{s}_1 + \mathbf{H}^{(2,1)} \mathbf{s}_2 + \mathbf{n}_1(k),$$

$$\mathbf{x}_1(k+1) = -\mathbf{H}^{(1,1)} \mathbf{P}_L^{(1)} \mathbf{s}_2^* + \mathbf{H}^{(2,1)} \mathbf{P}_L^{(1)} \mathbf{s}_1^* + \mathbf{n}_1(k+1), \tag{46}$$

Table 2: Transmission structure of training sequences ($N_T = 2$).

	time slot k	time slot $k+1$
TX1	$\overline{\mathbf{s}}_1(k) = \mathbf{s}_1$	$\overline{\mathbf{s}}_1(k+1) = -\mathbf{P}_L^{(1)} \mathbf{s}_2^*$
TX2	$\overline{\mathbf{s}}_2(k) = \mathbf{s}_2$	$\overline{\mathbf{s}}_2(k+1) = \mathbf{P}_L^{(1)} \mathbf{s}_1^*$

where $\mathbf{x}_1(k)$ and $\mathbf{x}_1(k+1)$ denote two consecutive received blocks at the single receive antenna. Applying the orthonormal FFT matrix $\mathbf{F}_L$ on (46), we obtain the frequency domain input-output relationship as shown below

$$\begin{bmatrix} \mathbf{x}_{1\mathcal{F}}(k) \\ \mathbf{x}_{1\mathcal{F}}(k+1) \end{bmatrix} = \begin{bmatrix} \mathbf{S}_{1\mathcal{F}} & \mathbf{S}_{2\mathcal{F}} \\ -\mathbf{S}_{2\mathcal{F}}^* & \mathbf{S}_{1\mathcal{F}}^* \end{bmatrix} \begin{bmatrix} \mathbf{h}_{\mathcal{F}}^{(1,1)} \\ \mathbf{h}_{\mathcal{F}}^{(2,1)} \end{bmatrix} + \begin{bmatrix} \mathbf{n}_{1\mathcal{F}}(k) \\ \mathbf{n}_{1\mathcal{F}}(k+1) \end{bmatrix}. \tag{47}$$

For this special case, the CFR estimation based on both the LS and MMSE criteria can be readily obtained by following the procedures outlined in Section 3. In this section, we further demonstrate that the CFR estimation for this special case can be implemented adaptively with block-wise recursive algorithms. Additionally, we also provide a brief convergence analysis of these algorithms.

4.1. Adaptive Implementation of CFR Estimation. It is easy to show that the CFR estimator for this special case has the following structure

$$\mathbf{G}_{\mathcal{F}} = \begin{bmatrix} \mathbf{G}_{1\mathcal{F}} & -\mathbf{G}_{2\mathcal{F}} \\ \mathbf{G}_{2\mathcal{F}}^H & \mathbf{G}_{1\mathcal{F}}^H \end{bmatrix}, \tag{48}$$

where $\mathbf{G}_{1\mathcal{F}}$ and $\mathbf{G}_{2\mathcal{F}}$ are both $L \times L$ diagonal matrices. Consider the LS estimator as an example, we have $\mathbf{G}_{1\mathcal{F}} = [\mathbf{S}_{1\mathcal{F}}^H \mathbf{S}_{1\mathcal{F}} + \mathbf{S}_{2\mathcal{F}}^H \mathbf{S}_{2\mathcal{F}}]^{-1} \mathbf{S}_{1\mathcal{F}}^H$ and $\mathbf{G}_{2\mathcal{F}} = [\mathbf{S}_{1\mathcal{F}}^H \mathbf{S}_{1\mathcal{F}} + \mathbf{S}_{2\mathcal{F}}^H \mathbf{S}_{2\mathcal{F}}]^{-1} \mathbf{S}_{2\mathcal{F}}$. Now let us define the diagonal vectors of $\mathbf{G}_{1\mathcal{F}}$ and $\mathbf{G}_{2\mathcal{F}}$ as $\mathbf{g}_{1\mathcal{F}}$ and $\mathbf{g}_{2\mathcal{F}}$, respectively, that is, $\mathbf{G}_{1\mathcal{F}} = \mathrm{diag}\{\mathbf{g}_{1\mathcal{F}}\}$ and $\mathbf{G}_{2\mathcal{F}} = \mathrm{diag}\{\mathbf{g}_{2\mathcal{F}}\}$. Then, we can write the CFR estimate as

$$\underbrace{\begin{bmatrix} \hat{\mathbf{h}}_{1\mathcal{F}} \\ \hat{\mathbf{h}}_{2\mathcal{F}} \end{bmatrix}}_{\hat{\mathbf{h}}_{\mathcal{F}}} = \underbrace{\begin{bmatrix} \mathbf{G}_{1\mathcal{F}} & -\mathbf{G}_{2\mathcal{F}} \\ \mathbf{G}_{2\mathcal{F}}^H & \mathbf{G}_{1\mathcal{F}}^H \end{bmatrix}}_{\mathbf{G}_{\mathcal{F}}} \underbrace{\begin{bmatrix} \mathbf{x}_{1\mathcal{F}}(k) \\ \mathbf{x}_{1\mathcal{F}}(k+1) \end{bmatrix}}_{\mathbf{x}_{\mathcal{F}}}. \tag{49}$$

We further define $L \times L$ diagonal matrices $\mathbf{X}_{1\mathcal{F}}(k) = \mathrm{diag}\{\mathbf{x}_{1\mathcal{F}}(k)\}$ and $\mathbf{X}_{1\mathcal{F}}(k+1) = \mathrm{diag}\{\mathbf{x}_{1\mathcal{F}}(k+1)\}$. Then, (49) can be reformulated into

$$\underbrace{\begin{bmatrix} \mathbf{g}_{1\mathcal{F}} \\ \mathbf{g}_{2\mathcal{F}} \end{bmatrix}}_{\mathbf{g}_{\mathcal{F}}} = \underbrace{\begin{bmatrix} \mathbf{\Phi} \cdot \mathbf{X}_{1\mathcal{F}}^H(k) & \mathbf{\Phi} \cdot \mathbf{X}_{1\mathcal{F}}(k+1) \\ -\mathbf{\Phi} \cdot \mathbf{X}_{1\mathcal{F}}^H(k+1) & \mathbf{\Phi} \cdot \mathbf{X}_{1\mathcal{F}}(k) \end{bmatrix}}_{\mathbf{U}_{\mathcal{F}}} \underbrace{\begin{bmatrix} \hat{\mathbf{h}}_{1\mathcal{F}} \\ \hat{\mathbf{h}}_{2\mathcal{F}}^* \end{bmatrix}}_{\check{\mathbf{h}}_{\mathcal{F}}} \tag{50}$$

or the simplified form

$$\mathbf{g}_{\mathcal{F}} = \mathbf{U}_{\mathcal{F}} \check{\mathbf{h}}_{\mathcal{F}}, \tag{51}$$

where in (50), $\mathbf{\Phi} = [\mathbf{X}_{1\mathcal{F}}^H(k)\mathbf{X}_{1\mathcal{F}}(k) + \mathbf{X}_{1\mathcal{F}}^H(k+1)\mathbf{X}_{1\mathcal{F}}(k+1)]^{-1}$; $\check{\mathbf{h}}_{\mathcal{F}}$ is a $2L \times 1$ vector; $\mathbf{U}_{\mathcal{F}}$ is an orthogonal matrix with

the size of $2L \times 2L$; $\mathbf{g}_{\mathcal{F}}$ is a $2L \times 1$ vector that contains the elements of $\mathbf{g}_{1\mathcal{F}}$ and $\mathbf{g}_{2\mathcal{F}}$.

We would like to emphasize that this reformulation from (49) to (50) is largely attributed to the benign property of Alamouti's code. This, as a result, enables the CFR estimation to be performed adaptively, and the channel to be tracked when the adaptive filter operates. To be more specific, we can view $\mathbf{U}_{\mathcal{F}}$ as the tap-input data matrix, $\mathbf{g}_{\mathcal{F}}$ as the output, and $\check{\mathbf{h}}_{\mathcal{F}}$ as the filter coefficients. The block diagram of this adaptive filter is depicted in Figure 2. We further define the error signal $\check{\mathbf{e}}_{\mathcal{F}}$, which is generated by comparing the filter output with the desired response, that is,

$$\check{\mathbf{e}}_{\mathcal{F}} = \mathbf{g}_{\mathcal{F}} - \mathbf{U}_{\mathcal{F}}\check{\mathbf{h}}_{\mathcal{F}}. \qquad (52)$$

Note that as $\mathbf{g}_{\mathcal{F}}$ is fixed and already available beforehand at the receiver, the adaptive filter can always operate at the training mode. Hence, if the channel is slowly time-varying, the adaptive method, through estimating the current channel gains based on the previous channel estimate, can achieve accuracy refinement without significantly increasing the complexity. Simulation results illustrating this can be found in Section 5. For notational convenience, we add in the time index for vectors or matrices in the ensuing description. And we summarize the recursive algorithms that are used to update the CFR estimate in Table 3, which include the block least mean square (LMS) algorithm and the block recursive least squares (RLS) algorithm.

The block RLS algorithm usually achieves a quicker convergence than the block LMS algorithm (as will be shown later by simulation results). But such a quick convergence is attained at the cost of a heavy increase in the computational complexity. To exemplify this, let us examine the computational complexity of both algorithms. At each iteration, the block LMS algorithm requires around $O(8L)$ computations, while the block RLS algorithm requires $O(24L^3 + 20L^2 + 4L)$ operations. A fast version of this block RLS algorithm, namely fast subsampled-updating RLS algorithm [23], can be used to achieve some complexity reduction, but may make this filter cumbersome. Fortunately, thanks to the special structure of the Alamouti's code, it is easy to verify that $\mathbf{U}_{\mathcal{F}}^H(k)\mathbf{U}_{\mathcal{F}}(k) = \mathbf{I}_{2L}$. Furthermore, we can induce that $\boldsymbol{\mathcal{P}}(k)$ (cf. Table 3) is a $2L \times 2L$ diagonal matrix, that is, $\boldsymbol{\mathcal{P}}(k) = \mathbf{I}_2 \otimes \mathbf{P}(k)$, where $\mathbf{P}(k)$ denotes an $L \times L$ diagonal matrix. Then, by following a similar technique used in [24, 25], we can avoid the need for matrix inversion in the block RLS algorithm and hence can eventually achieve a substantial reduction in the computational complexity but without losing the convergence advantage. For brevity, we summarize the simplified algorithm in Table 4. This simplified algorithm requires only $O(13L)$ operations for each iteration, which is much less than that of the original block RLS algorithm.

It is worthwhile to make a remark here that the above adaptive implementation of the CFR estimation is a special property owned by the Alamouti scheme with $N_T = 2$. When N_T increases beyond 2, the linear CFR estimator $\mathbf{G}_{\mathcal{F}}$, under both the LS and MMSE criteria (cf. (18) and (35)), will no longer have the simple Alamouti's structure. And so, a similar transformation as that from (49) to (50) may not necessarily

TABLE 3: Adaption algorithms for Alamouti-like transmissions.

Block LMS algorithm
Computation: for $k = 2, 4, \ldots$, compute
$\check{\mathbf{e}}_{\mathcal{F}}(k-2) = \mathbf{g}_{\mathcal{F}}(k-2) - \mathbf{U}_{\mathcal{F}}(k-2)\check{\mathbf{h}}_{\mathcal{F}}(k-2)$
$\check{\mathbf{h}}_{\mathcal{F}}(k) = \check{\mathbf{h}}_{\mathcal{F}}(k-2) + \mu\mathbf{U}_{\mathcal{F}}^H(k-2)\check{\mathbf{e}}_{\mathcal{F}}(k-2)$
where μ denotes the step size.

Block RLS algorithm
Initialize the algorithm by setting
$\check{\mathbf{h}}_{\mathcal{F}}(0) = \mathbf{0}$,
$\boldsymbol{\mathcal{P}}(0) = \delta^{-1}\mathbf{I}_{2L}$.
δ is a small positive constant and λ is the forgetting factor ($\lambda < 1$).
For each instant of time, $k = 2, 4, \ldots$, compute
$\boldsymbol{\mathcal{C}}(k) = \boldsymbol{\mathcal{P}}(k-2)\mathbf{U}_{\mathcal{F}}^H(k)$
$\boldsymbol{\mathcal{V}}(k) = \lambda\mathbf{I}_{2L} + \mathbf{U}_{\mathcal{F}}^H(k)\boldsymbol{\mathcal{C}}(k)$
$\boldsymbol{\mathcal{K}}(k) = \boldsymbol{\mathcal{C}}(k) \cdot \boldsymbol{\mathcal{V}}^{-1}(k)$
$\boldsymbol{\mathcal{P}}(k) = \lambda^{-1}[\boldsymbol{\mathcal{P}}(k-2) - \boldsymbol{\mathcal{K}}(k)\mathbf{U}_{\mathcal{F}}(k)\boldsymbol{\mathcal{P}}(k-2)]$
$\check{\mathbf{e}}_{\mathcal{F}}(k) = \mathbf{g}_{\mathcal{F}}(k) - \mathbf{U}_{\mathcal{F}}(k)\check{\mathbf{h}}_{\mathcal{F}}(k-2)$
$\check{\mathbf{h}}_{\mathcal{F}}(k) = \check{\mathbf{h}}_{\mathcal{F}}(k-2) + \boldsymbol{\mathcal{P}}(k)\mathbf{U}_{\mathcal{F}}^H(k)\check{\mathbf{e}}_{\mathcal{F}}(k)$

TABLE 4: Simplified block RLS algorithm.

Initialize the algorithm by setting
$\check{\mathbf{h}}_{\mathcal{F}}(0) = \mathbf{0}$,
$\mathbf{P}(0) = \delta^{-1}\mathbf{I}_L$.
δ is a small positive constant and λ is the forgetting factor ($\lambda < 1$).
For each instant of time, $k = 2, 4, \ldots$, compute
$\boldsymbol{\Omega}(k) = [\lambda\mathbf{I}_L + \mathbf{P}(k)]^{-1}$
$\mathbf{P}(k) = \lambda^{-1}[\mathbf{P}(k-2) - \mathbf{P}(k-2)\boldsymbol{\Omega}(k)\mathbf{P}(k-2)]$
$\check{\mathbf{e}}_{\mathcal{F}}(k) = \mathbf{g}_{\mathcal{F}}(k) - \mathbf{U}_{\mathcal{F}}(k)\check{\mathbf{h}}_{\mathcal{F}}(k-2)$
$\check{\mathbf{h}}_{\mathcal{F}}(k) = \check{\mathbf{h}}_{\mathcal{F}}(k-2) + [\mathbf{I}_2 \otimes \mathbf{P}(k)]\mathbf{U}_{\mathcal{F}}^H(k)\check{\mathbf{e}}_{\mathcal{F}}(k)$

hold. Then, the adaptive implementation for CFR estimation for cases of $N_T > 2$ requires further investigation.

4.2. Convergence Analysis. Convergence behaviors of these block-level recursive algorithms are briefly discussed as follows. We are interested in the behavior of $\xi(k) = \mathrm{E}\{\check{\mathbf{e}}_{\mathcal{F}}(k)\check{\mathbf{e}}_{\mathcal{F}}^H(k)\}$, particularly at the steady state, where $\check{\mathbf{e}}_{\mathcal{F}}(k)$ denotes the error signal, as defined in (52). For the block LMS algorithm, we define the weight-error vector as

$$\mathbf{v}_{\mathcal{F}}(k) = \check{\mathbf{h}}_{\mathcal{F}}(k) - \mathbf{h}_{\mathcal{F},0}, \qquad (53)$$

where $\mathbf{h}_{\mathcal{F},0}$ is the optimum tap-weight vector for the filter. Thus, we have

$$\mathbf{v}_{\mathcal{F}}(k) = \mathbf{v}_{\mathcal{F}}(k-2) + \mu\mathbf{U}_{\mathcal{F}}^H(k-2)\check{\mathbf{e}}_{\mathcal{F}}(k-2). \qquad (54)$$

Defining $\mathbf{e}_{\mathcal{F},0}(k) = \mathbf{g}_{\mathcal{F}}(k) - \mathbf{U}_{\mathcal{F}}(k)\mathbf{h}_{\mathcal{F},0}$, we have

$$\check{\mathbf{e}}_{\mathcal{F}}(k) = \mathbf{e}_{\mathcal{F},0}(k) - \mathbf{U}_{\mathcal{F}}^H(k-2)\mathbf{v}_{\mathcal{F}}(k). \qquad (55)$$

Let the weight-error correlation matrix be given as

$$\mathbf{R}_{\mathbf{vv}}(k) = \mathrm{E}\left\{\mathbf{v}_{\mathcal{F}}(k) \cdot \mathbf{v}_{\mathcal{F}}^H(k)\right\}. \qquad (56)$$
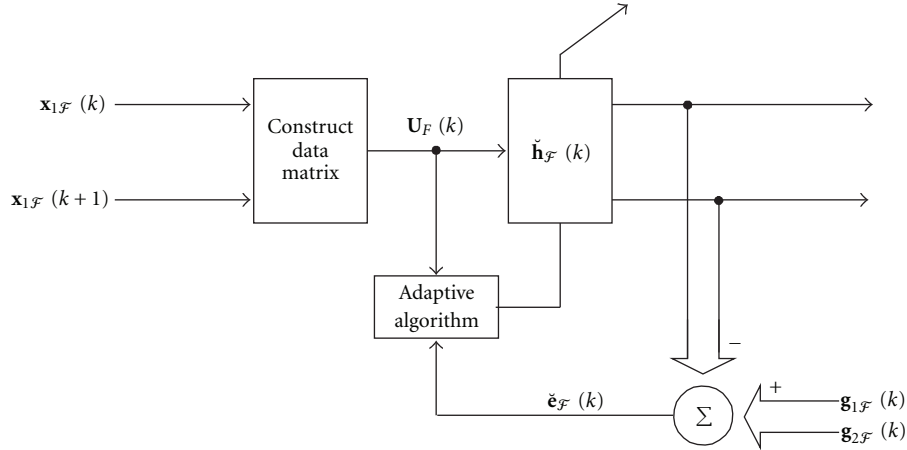
FIGURE 2: Block diagram of the adaptive filter.

Thus, the MSE of weight vector error can be obtained by simply taking the trace of $\mathbf{R_{vv}}(k)$. To facilitate the convergence analysis, we make the following assumptions.

(A3) Elements of $\mathbf{e}_{\mathcal{F},0}(k)$ are samples of a white noise process, which implies that $\mathrm{E}\{\mathbf{e}_{\mathcal{F},0}(k)\mathbf{e}_{\mathcal{F},0}^H(k)\} = \xi_{\min} \cdot \mathbf{I}_{2L}$, where $\xi_{\min}$ is the minimum MSE at the filter output.

(A4) $\mathbf{U}_{\mathcal{F}}(k)$ and $\mathbf{e}_{\mathcal{F},0}(k)$ are jointly Gaussian, and are uncorrelated with each other.

(A5) $\mathbf{v}_{\mathcal{F}}(k)$ is independent of $\mathbf{U}_{\mathcal{F}}(k)$ and $\mathbf{e}_{\mathcal{F},0}(k)$. Further, we assume $\mathbf{R_{uu}} = \mathrm{E}\{\mathbf{U}_{\mathcal{F}}^H(k)\mathbf{U}_{\mathcal{F}}(k)\}/2L$, where $\mathbf{R_{uu}}$ is the correlation matrix of the filter tap inputs.

Based on the above assumptions and following a similar procedure in [26, Appendix 8A], we can compute the *excess MSE*, which is defined as the difference between the steady-state MSE (i.e., $\xi(k = \infty)$) and the minimum MSE $\xi_{\min}$ of an adaptive filter, approximately by

$$\xi_{\text{excess}}^{\text{BLMS}} = \frac{\mu\xi_{\min}}{2}\,\text{tr}(\mathbf{R_{uu}}), \tag{57}$$

where $\text{tr}(\mathbf{R}_{uu})$ is equivalent to the sum of the powers of the signal samples at the filter tap inputs. Accordingly, the *misadjustment*, a dimension-free degradation measure that is defined as the ratio of the steady-state value of the excess MSE to the minimum MSE, can be written as

$$\mathcal{M}^{\text{BLMS}} = \frac{\mu}{2}\,\text{tr}(\mathbf{R_{uu}}). \tag{58}$$

Also, the steady-state MSE of the block LMS algorithm is given by

$$\xi_{\text{steady}}^{\text{BLMS}} = 2L\xi_{\min} + \frac{\mu\xi_{\min}}{2}\,\text{tr}(\mathbf{R_{uu}}). \tag{59}$$

It is obvious that the convergence behavior of the block LMS algorithm is governed by the eigenvalues of the correlation matrix $\mathbf{R_{uu}}$ of the filter tap input. Therefore, similar to the conventional LMS algorithm, the block LMS algorithm in nature is also a stochastic implementation of the *steepest-descent method* [26].

For the block RLS algorithm, its convergence analysis is undertaken on an adaptive identification scheme [27]. We consider a linear multiple regression model characterized by

$$\mathbf{g}_{\mathcal{F}}(k) = \mathbf{U}_{\mathcal{F}}(k)\mathbf{h}_{\mathcal{F},0} + \mathbf{e}_{\mathcal{F},0}(k), \tag{60}$$

where $\mathbf{h}_{\mathcal{F},0}$ is the regression parameter vector, $\mathbf{U}_{\mathcal{F}}(k)$ is the tap-input matrix, $\mathbf{e}_{\mathcal{F},0}(k)$ is the measurement noise, and $\mathbf{g}_{\mathcal{F}}(k)$ is the desired response. We define the weight error vector $\mathbf{v}_{\mathcal{F}}(k)$ the same as in (53) and its correlation matrix $\mathbf{R_{vv}}(k)$ the same as in (56). Further, we assume that the input signal vector is drawn from a stochastic process which is ergodic in the autocorrelation function, thus the time average can be used instead of the ensemble average [28]. Then, for $\lambda < 1$, following a similar approach as described in [27] for the analysis of RLS algorithms, the excess MSE for this block RLS algorithm at steady state can be written as

$$\xi_{\text{excess}}^{\text{BRLS}} = \frac{1-\lambda}{1+\lambda}2L\xi_{\min}, \tag{61}$$

and the misadjustment is simply

$$\mathcal{M}^{\text{BRLS}} = \frac{1-\lambda}{1+\lambda}2L. \tag{62}$$

Finally, the steady-state MSE is approximately given by

$$\xi_{\text{steady}}^{\text{BRLS}} = \frac{4L}{1+\lambda}\xi_{\min}. \tag{63}$$

5. Simulation Results

In this section, we provide some simulation results to demonstrate the efficacy of our proposed scheme. In our simulations, we employ a specific block structure for both data and training sequences, which is illustrated in Figure 3, taking the case of $N_T = 2$ as an example. This structure would be able to accommodate the proposed CFR estimation
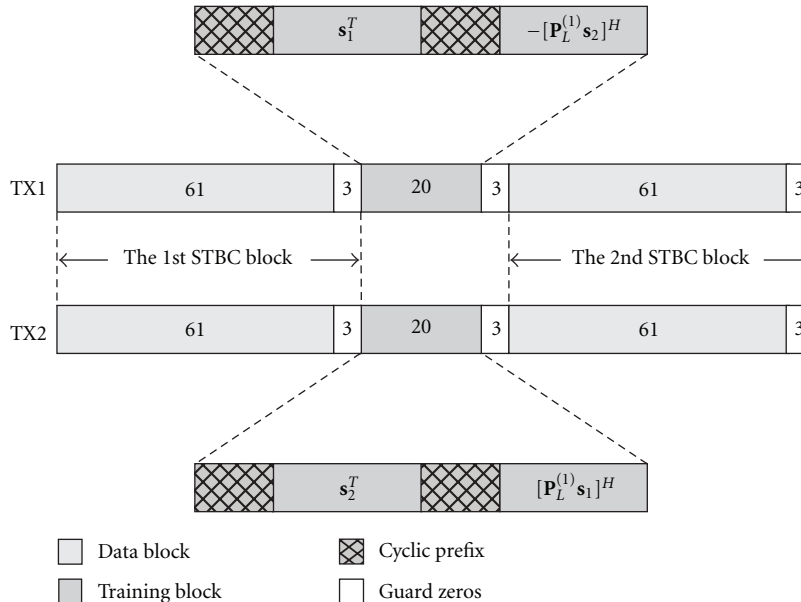
FIGURE 3: Block structure for both data and training sequences.

scheme and various FDE techniques. We assume the channel is frequency selective with channel memory $\nu = 3$, and further assume block fading, that is, the channel fading gains are constant over one ST-coded block including both data and training subblocks, but vary from block to block. For simplicity, we assume no *a priori* knowledge is available regarding the channel second-order statistics. Hence, only LS method is considered in our simulations. Chu sequences [20], a special case which satisfies the optimal condition given in (32), are chosen to be the training sequences. We use 8-PSK for data transmission without channel coding. At the receiver, channel estimation and equalization are both processed in the frequency domain. As a result, the FFT modules for FDE can be easily reused for the CFR estimation. Several different FDE approaches that are applicable to the structure shown in Figure 3 can be found in [9], and are employed in our simulations.

Figures 4(a) and 4(b) illustrate the BER performance corresponding to the frequency-domain MMSE linear equalization and MMSE decision-feedback equalization, respectively, under both CFR estimation and perfect CFR knowledge. When $L = 4$ ($N_b = 14$), that is, the minimum length to estimate the CFR, we have $\mathcal{P}_0 = 16$. The performance penalties due to inaccurate channel estimation, if evaluated at BER $= 10^{-4}$, are about 2.4 dB for the decision-feedback equalization and 2.8 dB for the linear equalization. When L extends to 7, or equivalently N_b extends to 20 as shown in Figure 3, $\mathcal{P}_0$ is accordingly increased to 28. Then, the BER performance penalties for the decision-feedback equalization and the linear equalization are reduced to 1.1 dB and 1.9 dB, respectively.

Furthermore, we also compare the performance of our approach with the method proposed in [29]. The approach reported in [29] was designed for channel estimation in MIMO systems with SC-FDE. It allows the transmitted sequence to be nulled on certain frequency tones, causing

the transmitted training sequences to be orthogonal in the frequency domain. Essentially, this approach [29] is equivalent to the on-off type estimation for each channel. To ensure a fair comparison, we apply the reference method [29] to the same structure depicted in Figure 3 for the case of $N_T = 2$. Then, both our scheme and the reference scheme [29] will achieve full rate, that is, $\mathcal{R} = 1$. Since there are 20 symbols in total allocated for the channel parameter estimation in the structure shown in Figure 3, when implementing the approach reported in [29], we allocate 16 for training sequences, and 4 (rather than $\nu = 3$) for the CP. This is because it is required in [29] that the length of training sequences must be evenly divisible by N_T. Furthermore, in the simulations, Chu sequences [20] are also adopted as the training sequences for this benchmark approach, as they as well satisfy the condition of optimality described in [29]. The BER performance of such an algorithm is depicted in Figure 4 by dash-dot lines. As illustrated by Figure 4, the system using our proposed scheme performs as well as, if not better than, the system using the approach described in [29]. However, considering the fact that implementation of the method given in [29] requires the transformation from CFR to CIR and then back to CFR (see details in [29]), our approach appears much simpler and straightforward.

Under similar simulation set-up, we also study the case of 2TX-2RX where the Alamouti-type STBC is employed at the transmitter side. At the receiver side, CFR estimation is performed based on the received signals across those two receive antennas, which is followed by FDE. In particular, we consider the equal gain diversity combining in the frequency domain. We further consider the case of 3TX-1RX, where the code design illustrated in (7) is used. BER performance of these scenarios under the frequency-domain linear equalization is depicted in Figure 5. For the purpose of comparison, we also plot in the same figure the BER
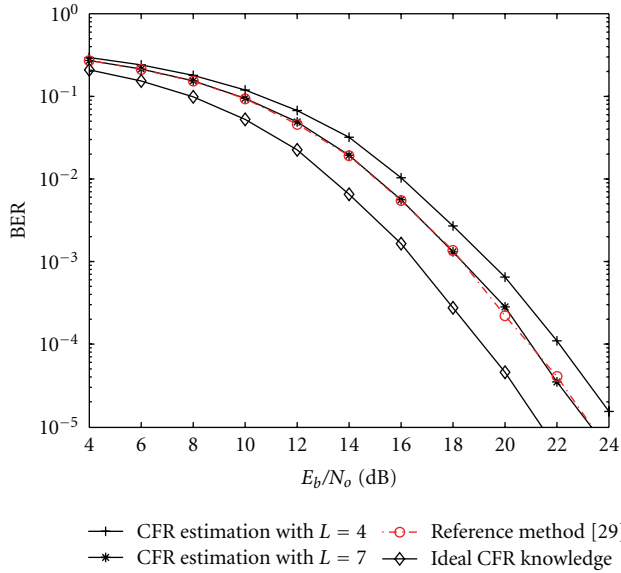
(a) Linear equalization
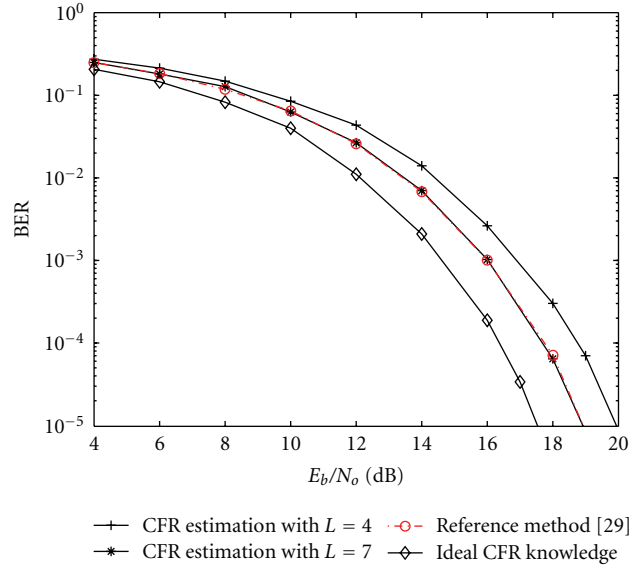
(b) Decision-feedback equalization

FIGURE 4: BER performance with FDE under CFR estimation and perfect CFR knowledge.
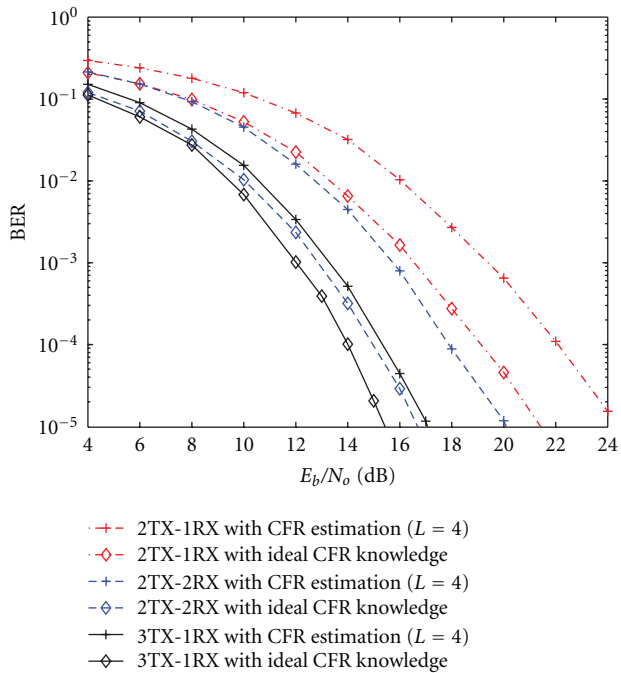

FIGURE 5: BER performance comparison for 2TX-1RX, 2TX-2RX, and 3TX-1RX with linear equalization.

performance of the 2TX-1RX case. From these curves, we notice that performance penalties due to inaccurate channel estimation are almost the same for the 2TX-1RX and 2TX-2RX cases, but are relatively smaller for the 3TX-1RX case. Furthermore, because of the addition of one more receive antenna, the BER performance of 2TX-2RX is much improved over that of the 2TX-1RX case. However, as shown

in Figure 5, the BER performance of 2TX-2RX is inferior to that of the 3TX-1RX case. This is largely due to the fact that this 3TX-1RX system we consider here is not a full-rate system (i.e., $\mathcal{R} = 1/2$), which is in contrast to those systems employing two transmit antennas and Alamouti-type code.

For the special case of $N_T = 2$, we also conduct simulations to study the behaviors of these adaptive estimation algorithms. For simplicity, Chu sequences [20] are used again in our simulations. We set $L = \nu + 1$, $\mathcal{P}_0 = 16$, and $\sigma_n^2 = 0.1$. Block fading is still adopted, but the channel fadings are further assumed to be correlated in the time domain. This means the Doppler spread is introduced, and it may affect performance of the adaptive algorithms, as will be confirmed later. The rate of fading in our simulations is determined by $f_d T$, where f_d denotes the maximum Doppler frequency shift and T denotes the duration of one whole ST-coded block. A larger value of $f_d T$ implies faster fading and vice versa. The following simulation results are obtained by setting $f_d T = 10^{-4}$, unless otherwise stated. Figure 6(a) shows a plot of the squared error $\|\check{\mathbf{e}}_{\mathcal{F}}(k)\|^2$ versus the number of iterations for a single run or trial of the block-wise LMS and RLS algorithms. Since those algorithms only iterate once for each ST-coded frame, the number of iterations also corresponds to the number of frames. As is shown by Figure 6(a), the learning curves for a single trial of both adaptive algorithms exhibit a noisy form. However, it is clearly seen that the block RLS algorithm converges much faster than the block LMS algorithm. Additionally, we are also interested in the behavior of the squared error deviation $\|\mathbf{v}_{\mathcal{F}}(k)\|^2$ for both algorithms. For the same realization, Figure 6(b) shows the transient behavior of $\|\mathbf{v}_{\mathcal{F}}(k)\|^2$ for both algorithms. As $\|\check{\mathbf{e}}_{\mathcal{F}}(k)\|^2$ converges, $\|\mathbf{v}_{\mathcal{F}}(k)\|^2$ converges accordingly. But notice that the curves in both figures are plotted at different vertical scales. It is worth noting that in our simulations, we ran
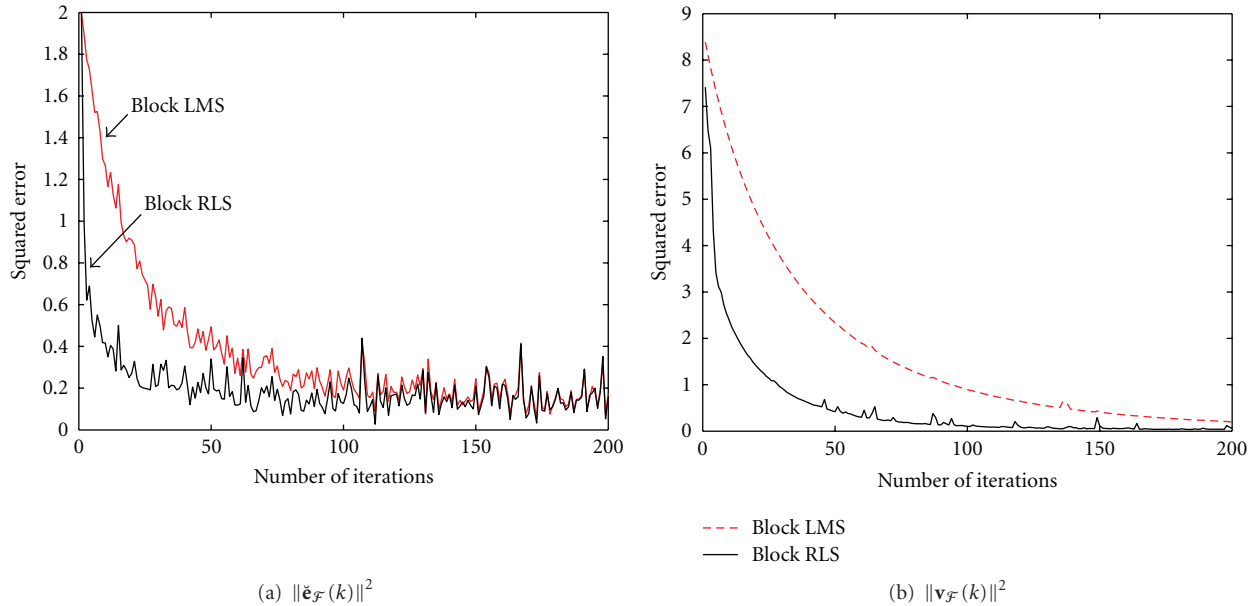
(a) $\|\breve{\mathbf{e}}_{\mathcal{F}}(k)\|^2$

(b) $\|\mathbf{v}_{\mathcal{F}}(k)\|^2$

FIGURE 6: Transient behavior of squared error $\|\breve{\mathbf{e}}_{\mathcal{F}}(k)\|^2$ and $\|\mathbf{v}_{\mathcal{F}}(k)\|^2$ of the block-wise LMS and RLS algorithms. $f_d T = 10^{-4}$. Block RLS: $\lambda = 0.8$; block LMS: $\mu = 0.08$.
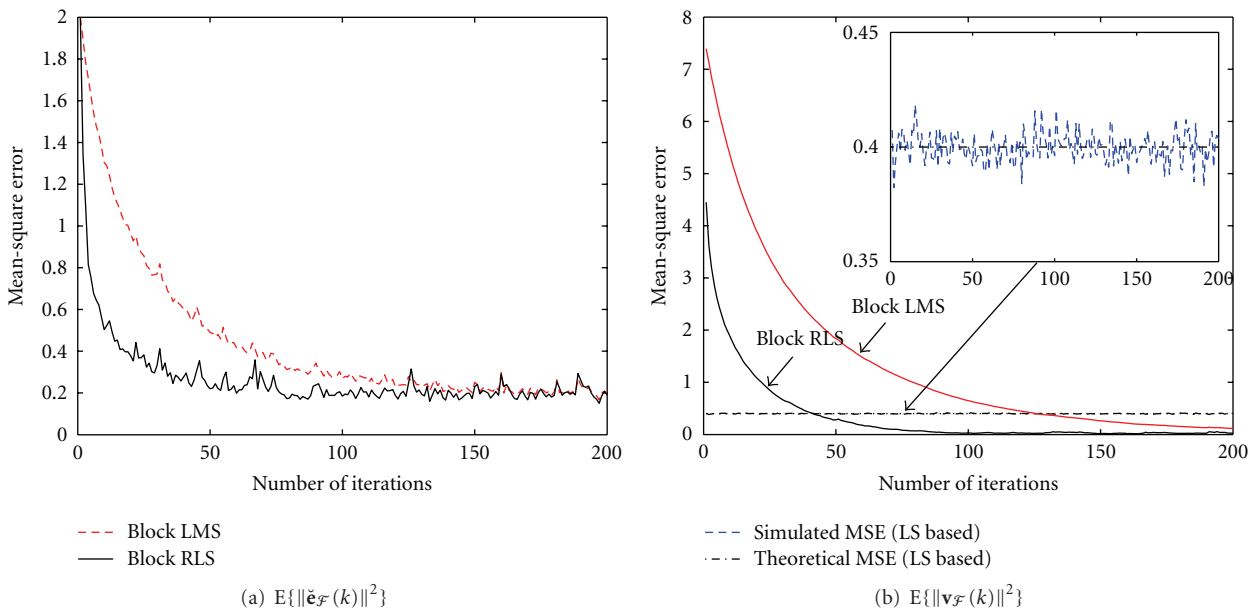

(a) $\mathrm{E}\{\|\breve{\mathbf{e}}_{\mathcal{F}}(k)\|^2\}$

(b) $\mathrm{E}\{\|\mathbf{v}_{\mathcal{F}}(k)\|^2\}$

FIGURE 7: Learning curves of the block-wise LMS and RLS algorithms for the mean-square error $\mathrm{E}\{\|\breve{\mathbf{e}}_{\mathcal{F}}(k)\|^2\}$ and $\mathrm{E}\{\|\mathbf{v}_{\mathcal{F}}(k)\|^2\}$. $f_d T = 10^{-4}$. Block RLS: $\lambda = 0.8$; block LMS: $\mu = 0.08$.

the filter from scratch by simply initializing elements of the channel estimate (i.e., the filter coefficients) all to zeros. This is to demonstrate the convergence performance of these block-wise algorithms. However, in practice, it is certainly possible to speed up the convergence process and reduce the amount of training data. For example, for the first frame, we can obtain the channel estimate by using nonadaptive approach from (49) (i.e., *hot start* initialization). Afterwards, we can apply the adaptive method.

Given the same set of parameters that lead to the results shown in Figure 6, we conduct 100 independent trials and compute the ensemble average. In Figure 7(a), we plot the learning curves of both block-wise algorithms for $\mathrm{E}\{\|\breve{\mathbf{e}}_{\mathcal{F}}(k)\|^2\}$ versus the number of iterations. It is clearly seen that ensemble averaging helps smooth out the effects of gradient noise in the learning curves. For the same set of trials, we also compute the corresponding values of $\mathrm{E}\{\|\mathbf{v}_{\mathcal{F}}(k)\|^2\}$, and plot them in Figure 7(b). In addition, for

the purpose of comparison, MSE values obtained from the nonadaptive CFR estimation experiments based on the LS method are also plotted in the same figure, together with the theoretical value. Such a theoretical MSE value can be obtained by plugging these simulation parameters into (33), and we obtain $E\{\|\mathbf{E}_{\mathcal{F}}\|^2\} = 0.4$. The subplot in Figure 7(b) indicates a very good match between the simulated values and the theoretical one, which in turn corroborates the correctness of our derivations. Moreover, it is also observed that after the learning curves converge (especially for the block RLS algorithm), the MSE values attained are much smaller than those obtained by the nonadaptive method or the one computed theoretically. This basically demonstrates the performance advantage of using this adaptive approach.

Finally, we provide some simulation results in Figure 8 to demonstrate the error performance of both adaptive estimation algorithms at higher Doppler spreads. In particular, we consider three different Doppler spreads: $f_d T = 10^{-4}$, 10^{-3}, and 10^{-2}. And we conduct 100 independent trials for each case. For simplicity, we leave the step sizes unchanged in our simulations, that is, $\lambda = 0.8$ and $\mu = 0.08$; but note that it is desirable to reduce them accordingly as the frequency dispersion or Doppler spread increases. Here, we only study the behavior of $E\{\|\mathbf{v}_{\mathcal{F}}(k)\|^2\}$, and for the ease of comparison, we also plot in Figure 8 the theoretical MSE value of the nonadaptive LS estimation approach. The results shown in Figure 8 indicate that as the Doppler spread increases moderately, for example, from $f_d T = 10^{-4}$ to $f_d T = 10^{-3}$, the estimation accuracy of both algorithms will degrade a little, but not severely. However, further increase in the Doppler spread, for example, from $f_d T = 10^{-3}$ to $f_d T = 10^{-2}$, will lead to a drastic degradation in the estimation accuracy for both algorithms. In fact, in this case, the estimation accuracy of each of these two adaptive algorithms is inferior to that of the nonadaptive estimation approach, indicating that they are unable to track faster channel variations and thus may no longer be usable in practice.

6. Conclusion

In this paper, we presented and studied a training-based CFR estimation scheme for ST-coded MIMO systems with SC-FDE. This scheme is different from the traditional one which obtains the CIR firstly then transfers it to the CFR. In this scheme, CFR estimation is jointly implemented with FDE; thus, estimate of the CFR can be obtained directly and the hardware complexity of the transceiver can also be reduced. To be more specific, training sequences are ST block encoded at the transmitter using the same encoder for data sequences. At the receiver, similar procedures are applied to both data and training sequences, including the CP removal and FFT processing. Then, estimation of the CFR is performed immediately afterwards. Conditioning on different *a priori* channel knowledge, we further studied the CFR estimation based on two criteria: LS and MMSE. A thorough analysis of the MSE in estimating the CFR
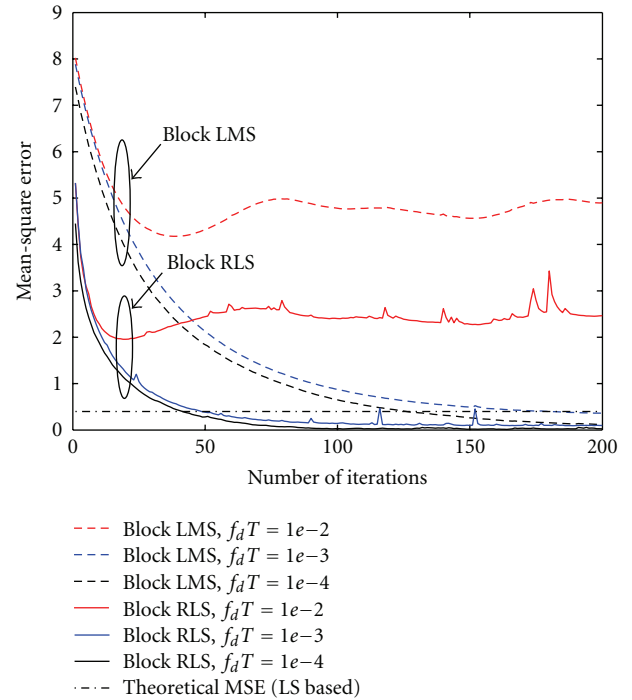

FIGURE 8: Learning curves of the block-wise LMS and RLS algorithms for $E\{\|\mathbf{v}_{\mathcal{F}}(k)\|^2\}$ under different Doppler spreads: $f_d T = 10^{-4}$, 10^{-3}, and 10^{-2}. Block RLS: $\lambda = 0.8$; block LMS: $\mu = 0.08$.

was provided under each criterion. Moreover, imposing a constraint on the transmit power of training sequences, we also investigated the optimal design of training signals. It is shown that under the LS criterion, training sequences having a constant sum magnitude at each frequency tone, such as Chu sequences, will lead to the least MSE. For the MMSE criterion, we have shown that the optimal design of training sequences features a water-filling-type power distribution. Additionally, we demonstrated that adaptive implementation of the CFR is feasible when the number of transmit antennas is equal to 2, which is due to the benign property of Alamouti's code. However, we feel that the identical property may not be possessed when N_T increases beyond 2 although it may need further investigation.

Acknowledgments

References

[1] H. Sari, G. Karam, and I. Jeanclaude, "Transmission techniques for digital terrestrial TV broadcasting," *IEEE Communications Magazine*, vol. 33, no. 2, pp. 100–109, 1995.

[2] D. Falconer, S. L. Ariyavisitakul, A. Benyamin-Seeyar, and B. Eidson, "Frequency domain equalization for single-carrier broadband wireless systems," *IEEE Communications Magazine*, vol. 40, no. 4, pp. 58–66, 2002.

[3] F. Pancaldi, G. M. Vitetta, R. Kalbasi, N. Al-Dhahir, M. Uysal, and H. Mheidat, "Single-carrier frequency domain equalization: a focus on wireless applications," *IEEE Signal Processing Magazine*, vol. 25, no. 5, pp. 37–56, 2008.

[4] E. Dahlman, S. Parkvall, J. Skold, and P. Beming, *3G Evolution: HSPA and LTE for Mobile Broadband*, Academic Press, New York, NY, USA, 2nd edition, 2008.

[5] N. Al-Dhahir, "Single-carrier frequency-domain equalization for space-time block-coded transmissions over frequency-selective fading channels," *IEEE Communications Letters*, vol. 5, no. 7, pp. 304–306, 2001.

[6] S. Zhou and G. B. Giannakis, "Single-carrier space-time block-coded transmissions over frequency-selective fading channels," *IEEE Transactions on Information Theory*, vol. 49, no. 1, pp. 164–179, 2003.

[7] H. Mheidat, M. Uysal, and N. Al-Dhahir, "Time- and frequency-domain equalization for quasi-orthogonal STBC over frequency-selective channels," in *Proceedings of the IEEE International Conference on Communications*, pp. 697–701, June 2004.

[8] K. Takeda, T. Itagaki, and F. Adachi, "Application of space-time transmit diversity to single-carrier transmission with frequency-domain equalisation and receive antenna diversity in a frequency-selective fading channel," *IEE Proceedings on Communications*, vol. 151, no. 6, pp. 627–632, 2004.

[9] Y. Yang, Y. H. Chew, and T. T. Tjhung, "Single-carrier frequency-domain equalization for space-time coded systems over multipath channels," in *Proceedings of the IEEE Vehicular Technology Conference*, pp. 2193–2197, Melbourne, Australia, May 2006.

[10] M. Biguesh and A. B. Gershman, "Training-based MIMO channel estimation: a study of estimator tradeoffs and optimal training signals," *IEEE Transactions on Signal Processing*, vol. 54, no. 3, pp. 884–893, 2006.

[11] Q. Zhang and T. Le-Ngoc, "Channel-estimate-based frequency-domain equalization (CE-FDE) for broadband single-carrier transmission," *Wireless Communications and Mobile Computing*, vol. 4, no. 4, pp. 449–461, 2004.

[12] Y. Wang and X. Dong, "Frequency-domain channel estimation for SC-FDE in UWB communications," *IEEE Transactions on Communications*, vol. 54, no. 12, pp. 2155–2163, 2006.

[13] C. Fragouli, N. Al-Dhahir, and W. Turin, "Training-based channel estimation for multiple-antenna broadband transmissions," *IEEE Transactions on Wireless Communications*, vol. 2, no. 2, pp. 384–391, 2003.

[14] C. Pietsch and J. Linder, "MIMO channel estimation for transmissions based on Alamouti's scheme," in *Proceedings of the 7th Management Committee Meeting*, Paris, France, May 2003, TD(03)100, COST273.

[15] S. M. Alamouti, "A simple transmit diversity technique for wireless communications," *IEEE Journal on Selected Areas in Communications*, vol. 16, no. 8, pp. 1451–1458, 1998.

[16] V. Tarokh, H. Jafarkhani, and A. R. Calderbank, "Space-time block codes from orthogonal designs," *IEEE Transactions on Information Theory*, vol. 45, no. 5, pp. 1456–1467, 1999.

[17] H. Wang and X. G. Xia, "Upper bounds of rates of complex orthogonal space-time block codes," *IEEE Transactions on Information Theory*, vol. 49, no. 10, pp. 2788–2796, 2003.

[18] E. G. Larsson and P. Stoica, *Space-Time Block Coding for Wireless Communications*, Cambridge University Press, Cambridge, UK, 2003.

[19] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, Cambridge, UK, 2004.

[20] D. C. Chu, "Polyphase codes with good periodic correlation properties," *IEEE Transactions on Information Theory*, vol. 18, no. 4, pp. 531–532, 1972.

[21] J. Coon, M. Beach, and J. McGeehan, "Optimal training sequences for channel estimation in cyclic-prefix-based single-carrier systems with transmit diversity," *IEEE Signal Processing Letters*, vol. 11, no. 9, pp. 729–732, 2004.

[22] Y. Yang, R. S. Blum, Z. S. He, and D. R. Fuhrmann, "MIMO radar waveform design via alternating projection," *IEEE Transactions on Signal Processing*, vol. 58, pp. 1440–1445, 2010.

[23] D. T. M. Slock and K. Maouche, "The fast subsampled-updating recursive least-squares (FSU RLS) algorithm for adaptive filtering based on displacement structure and the FFT," *Signal Processing*, vol. 40, no. 1, pp. 5–20, 1994.

[24] W. M. Younis, A. H. Sayed, and N. Al-Dhahir, "Efficient adaptive receivers for joint equalization and interference cancellation in multiuser space-time block-coded systems," *IEEE Transactions on Signal Processing*, vol. 51, no. 11, pp. 2849–2862, 2003.

[25] Y. Yang, Y. H. Chew, and T. T. Tjhung, "Adaptive frequency-domain equalization for space-time block-coded DS-CDMA downlink," in *Proceedings of the IEEE International Conference on Communications*, pp. 2343–2347, Seoul, South Korea, May 2005.

[26] B. Farhang-Boroujeny, *Adaptive Filters: Theory and Applications*, John Wiley & Sons, New York, NY, USA, 1998.

[27] M. Montazeri and P. Duhamel, "A set of algorithms linking NLMS and block RLS algorithms," *IEEE Transactions on Signal Processing*, vol. 43, no. 2, pp. 444–453, 1995.

[28] S. Haykin, *Adaptive Filter Theory*, Prentice-Hall, Upper Saddle River, NJ, USA, 4th edition, 2002.

[29] J. Siew, J. Coon, R. J. Piechocki et al., "A channel estimation algorithm for MIMO-SCFDE," *IEEE Communications Letters*, vol. 8, no. 9, pp. 555–557, 2004.