Low-Complexity Optimal Estimation of MIMO ISI Channels with Binary Training Sequences

Shuangquan Wang, *Student Member, IEEE,* and Ali Abdi, *Member, IEEE*

Abstract

In this letter, a novel low-complexity optimal channel estimator using uncorrelated *periodic* complementary sets of binary sequences is proposed for multiple-input multiple-output (MIMO) intersymbol interference (ISI) channels. The estimator is optimal since it attains the minimum possible Cramér-Rao lower bound (CRLB). Moreover, it can be implemented with very low complexity via ASIC/FPGA, which makes it suitable and ready for practical MIMO systems.

Index Terms

MIMO, Uncorrelated Complementary Sets, ISI, Channel Estimation, Frequency-Selective Channel.

I. INTRODUCTION

For quasi-static or slowly-varying fading channels, training-based channel estimation has been widely used [1]. Some information theoretical guidelines for training sequence design over MIMO intersymbol interference (ISI)[1] channels are given in [2]. However, the sequences given in [2] may result in high peak-to-average power ratios (PAPR), which normally should be avoided in practice.

To overcome the above PAPR difficulty, an optimal training-based channel estimator was proposed in [3] using uncorrelated *aperiodic* complementary sets of binary sequences, where the guard period between data and training symbols are padded with zeros. Nevertheless, the approach of [3] is not applicable to systems where the guard period is filled in with some known symbols such as cyclic prefix (CP) of training sequences, shown in Fig. 1, e.g., the midamble in TD-SCDMA [4], the 3G standard proposed by China.

The authors are with the Center for Wireless Communications and Signal Processing Research (CWCSPR), Department of Electrical and Computer Engineering, New Jersey Institute of Technology, Newark, NJ 07102 USA (e-mail:sw27@njit.edu, ali.abdi@njit.edu).

[1] We use ISI and frequency-selective interchangeably in this letter.

In order to address channel estimation in such systems, in this letter, we propose to use uncorrelated *periodic*[2] complementary sets of binary sequences for optimal training. At the end, comparison with existing binary training sequences such as ZCZ sequences [5][6] and impulse sequences [2] are provided as well.

Notation: $(\cdot)^\top$ is reserved for the matrix transpose, $(\cdot)^{-1}$ for the matrix inverse, $(\cdot)^\dagger$ for the matrix Hermitian, $\mathrm{tr}[\cdot]$ for the trace of a matrix, $\mathrm{diag}(\sigma_1, \sigma_2, \cdots, \sigma_n)$ denotes the diagonal matrix with $\sigma_1, \sigma_2, \cdots, \sigma_n$ on the main diagonal, $\mathrm{vec}(\cdot)$ stacks all the columns of its matrix argument into one tall column vector, $[\mathbf{A}]_{m,n}$ is the $(m, n)^{\mathrm{th}}$ element of the matrix $\mathbf{A}$, $\mathbb{E}[\cdot]$ is the mathematical expectation, $\overline{(\cdot)}$ is the sample average, $\mathbf{I}_m$ denotes the $m \times m$ identity matrix, $\mathbf{1}_{m\times n}$ is an $m \times n$ matrix whose entries are 1, $t{\in}[m, n]$ implies that t is an integer such that $m \leq t \leq n$, $\otimes$ represents the Kronecker product, $\lceil x\rceil$ is the smallest integer not less than x, $\lfloor x\rfloor$ is the largest integer not greater than x, $\|\cdot\|_F$ denotes the Frobenius norm, $(\cdot)_N$ is the modulus N operator, $\mathbf{\Pi}_m$ is the forward shift permutation matrix of order m [7, pp. 27], and $\mathbf{A}\mathbf{\Pi}_m^l$ shifts the matrix $\mathbf{A}$, which has m columns, cyclically to the right by l columns. We also have $\circledast_N$ for $\mod N$ circular convolution and $*$ for linear convolution. Unless otherwise mentioned, lower-case bold letters represent row vectors, whereas upper-case bold letters are used for matrices.

II. SYSTEM AND CHANNEL MODELS

We consider a frequency-selective block fading MIMO channel. Let $\mathbf{H} = [\mathbf{H}_0, \mathbf{H}_1, \cdots, \mathbf{H}_L]$ be the $L + 1$ tap discrete-time channel impulse response (CIR) of the MIMO frequency-selective channel with N_T transmitters and N_R receivers, where $\mathbf{H}_l = \begin{bmatrix} h_{1,1}(l) & \cdots & h_{1,N_T}(l) \\ \vdots & \ddots & \vdots \\ h_{N_R,1}(l) & \cdots & h_{N_R,N_T}(l) \end{bmatrix}, l \in [0, L]$. In addition, we assume that elements of $\mathbf{H}$ are independent Gaussian random variable with zero mean, and each subchannel $\mathbf{h}_{n_r,n_t} = [h_{n_r,n_t}(0), h_{n_r,n_t}(1), \cdots, h_{n_r,n_t}(L)]$ has unit power, i.e., $\sum_{l=0}^{L} \mathbb{E}\left[|h_{n_r,n_t}(l)|^2\right] = 1$. Moreover, the l^{th} taps of all the subchannels have the same power σ_l, i.e., $\mathbb{E}\left[|h_{n_r,n_t}(l)|^2\right]{=}\sigma_l, l{\in}[0, L], \forall n_r, n_t$. It follows that $\sum_{l=0}^{L}\sigma_l{=}1$. We also define $\mathbf{C}_\Sigma = \mathbb{E}\left[\mathbf{h}_{n_r,n_t}^\top \mathbf{h}_{n_r,n_t}^\star\right]$ as the covariance matrix among the $L{+}1$ taps between the n_t^{th} Tx and n_r^{th} Rx antennas, given by $\mathbf{C}_\Sigma = \mathrm{diag}(\sigma_0, \sigma_1, \cdots, \sigma_L), \forall n_r, n_t$. With $\mathbf{h} = \mathrm{vec}(\mathbf{H})$, we have $\mathbf{C}_\mathbf{h} = \mathbb{E}[\mathbf{h}\mathbf{h}^\dagger] = \mathbf{C}_\Sigma{\otimes}\mathbf{I}_{N_T N_R}$.

A typical frame structure for the signal transmitted by the n_t^{th} Tx antenna of a MIMO system is shown in Fig. 1, where the vector $\mathbf{s}_{n_t}$ contains the training sequence of length N_s, $\mathrm{CP}_{n_t} = [s_{n_t}(N_s - L), \cdots, s_{n_t}(N_s - 1)]$, the cyclic prefix of $\mathbf{s}_{n_t}$, is used to separate data and training symbols, and $\mathbf{0}_{1\times L}$,

---

[2] The periodicity is due to the presence of the CP.

which denotes L 0's, is used to avoid the interframe interference [2]. The received training signal on N_R Rx antennas, after discarding those polluted by the data due to ISI, can be written as in the matrix form [8]

$$\mathbf{Y}=\sqrt{\gamma/N_T}\mathbf{H}\mathbf{S}+\mathbf{E}, \tag{1}$$

where $\mathbf{S}$, the $N_T(L+1)\times N_s$ training matrix, is given by

$$\mathbf{S}=\begin{bmatrix}\mathbf{s}(0) & \mathbf{s}(1) & \cdots & \mathbf{s}(N_s-1)\\ \mathbf{s}(N_s-1) & \mathbf{s}(0) & \ddots & \mathbf{s}(N_s-2)\\ \vdots & \ddots & \ddots & \vdots\\ \mathbf{s}(N_s-L) & \mathbf{s}(N_s-L+1) & \cdots & \mathbf{s}(N_s-L-1)\end{bmatrix}, \tag{2}$$

in which $\mathbf{s}(n)=[s_1(n),s_2(n),\cdots,s_{N_T}(n)]^\top$, and $\mathbf{Y}=[\mathbf{y}(0),\mathbf{y}(1),\cdots,\mathbf{y}(N_s-1)]$, in which $\mathbf{y}(n)=[y_1(n),y_2(n),\cdots,y_{N_R}(n)]^\top$. The complex noise matrix in (1) is defined as $\mathbf{E}=[\mathbf{e}(0),\mathbf{e}(1),\cdots,\mathbf{e}(N_s–1)]$, where $\mathbf{e}(n)=[e_1(n),e_2(n),\cdots,e_{N_R}(n)]^\top$. Note that $\mathbf{s}(n)$, $\mathbf{y}(n)$ and $\mathbf{e}(n)$, $\forall n$, are column vectors, $s_{n_t}(n)$ is the training symbol transmitted by the n_t^{th} Tx antenna at time n, $y_{n_r}(n)$ is the signal received by the n_r^{th} Rx antenna at time n, $e_{n_r}(n)$ is the additive noise component in $y_{n_r}(n)$, and finally γ is the expected signal-to-noise ratio (SNR) on each Rx antenna.

III. OPTIMAL TRAINING SEQUENCES

A. *Criterion for Optimal Training*

For channel estimation, we assume the elements of the additive noise matrix $\mathbf{E}$ are spatio-temporally white Gaussian, and independent of the elements of $\mathbf{H}$. Following the terminology of [9], the best estimator for the random channel $\mathbf{H}$ is the minimum mean square error (MMSE) estimator, presented by the following proposition.

Proposition 1: For the system model in (1), the MMSE estimator of $\mathbf{H}$ is given by[3]

$$\widehat{\mathbf{H}}=\sqrt{\gamma^{-1}N_T}\mathbf{Y}\mathbf{S}^\top\Big(\mathbf{S}\mathbf{S}^\top+\gamma^{-1}N_T\mathbf{C}_\Sigma^{-1}\otimes\mathbf{I}_{N_T}\Big)^{-1}, \tag{3}$$

with the following total mean square error (TMSE)

$$\varepsilon=\gamma^{-1}N_RN_T\,\mathrm{tr}\Big[\Big(\mathbf{S}\mathbf{S}^\top+\gamma^{-1}N_T\mathbf{C}_\Sigma^{-1}\otimes\mathbf{I}_{N_T}\Big)^{-1}\Big]. \tag{4}$$

---

[3] To obtain a meaningful estimate, we need $N_T(L+1)\leq N_s$.

Note that ε coincides with the Bayesian CRLB[4]. This proposition can be easily proved by using Theorem 11.1 of [9], combined with the definition of $\mathbf{C_h}$ and the equality $\mathrm{vec}(\mathbf{AZB}) = (\mathbf{B}^\top \otimes \mathbf{A})\mathrm{vec}(\mathbf{Z})$, upon rewriting (1) as $\mathrm{vec}(\mathbf{Y}) = \sqrt{\gamma/N_T}(\mathbf{S}^\top \otimes \mathbf{I}_{N_R})\mathrm{vec}(\mathbf{H}) + \mathrm{vec}(\mathbf{E})$.

From (4), we can conclude that the Bayesian CRLB depends on the training symbol matrix $\mathbf{S}$, when the number of Tx and Rx antennas, SNR, and the fading covariance matrix $\mathbf{C}_\Sigma$ are fixed. Under the Tx power constraint of training symbols, minimization of ε through $\mathbf{S}$ is achieved by the following proposition.

Proposition 2: If each training sequence has unit power, i.e., $\frac{1}{N_s}\sum_{n=0}^{N_s-1}|s_{n_t}(n)|^2 = 1$, $\forall n_t$, the minimum Bayesian CRLB is obtained if and only if the training sequences satisfy

$$\mathbf{S}\mathbf{S}^\top = N_s \mathbf{I}_{N_T(L+1)}. \tag{5}$$

The semi-unitary condition in (5) is derived in [2] and [8]. In general, it is hard to find such training sequences to satisfy (5) [8, pp. 179], since it requires that all the training sequences have perfect periodic autocorrelations and cross-correlations within L temporal shifts. However, by using uncorrelated complementary sequences, we can handle this problem. In fact, if we use more than one, say, two training sequences [3], each of length N per frame per Tx antenna[5], the optimality condition in (5) reduces to

$$\mathbf{S}_A\mathbf{S}_A^\top + \mathbf{S}_B\mathbf{S}_B^\top = 2N\mathbf{I}_{N_T(L+1)}, \tag{6}$$

by plugging $\mathbf{S} = [\mathbf{S}_A, \mathbf{S}_B]$ into (5).

The new condition in (6) implies that $\mathbf{S}_A$ and $\mathbf{S}_B$ need to be uncorrelated complementary in terms of periodic correlations, which is the property owned by the uncorrelated *periodic* complementary sets binary sequences. Therefore we can use them for optimal training. In what follows, a brief description for them is given first, then the optimal training sequences are constructed accordingly.

B. Uncorrelated Periodic Complementary Sets of Sequences

Let $\mathbf{a}_i = [a_{i,0}, a_{i,1}, \cdots, a_{i,(N-1)}]$ be a sequence of ± 1's, and $\psi_{\mathbf{a}_i,\mathbf{a}_i}(k) = \sum_{j=0}^{N-1} a_{i,j}a_{i,(j+k)_N}, |k| \leq N–1$, is the *periodic* auto-correlation of $\mathbf{a}_i$. A set of sequences $\{\mathbf{a}_i\}_{i=0}^{p-1}$, each with N elements, is *periodic* complementary if and only if $\sum_{i=0}^{p-1}\psi_{\mathbf{a}_i,\mathbf{a}_i}(k) = 0, k \neq 0$ [12]. If another set of sequences $\{\mathbf{b}_i\}_{i=0}^{p-1}$ is *periodic* complementary and $\sum_{i=0}^{p-1}\psi_{\mathbf{a}_i,\mathbf{b}_i}(k) = 0, |k| \leq N–1$, where $\psi_{\mathbf{a}_i,\mathbf{b}_i}(k) = \sum_{j=0}^{N-1} a_{i,j}b_{i,(j+k)_N}, |k| \leq N-1$, then we call $\{\mathbf{b}_i\}_{i=0}^{p-1}$ a mate of $\{\mathbf{a}_i\}_{i=0}^{p-1}$, and vice versa. A collection of *periodic* complementary

---

[4] The concept of CRLB for random parameter estimation was introduced in [10], and named Bayesian CRLB later, say, in [11].

[5] Now the condition in footnote 3 can be written as $N_T(L+1) \leq 2N$.

sets of sequences $\{\mathbf{a}_i\}_{i=0}^{p-1}$, $\{\mathbf{b}_i\}_{i=0}^{p-1}$, $\cdots$, $\{\mathbf{z}_i\}_{i=0}^{p-1}$ are *mutually uncorrelated* if every two *periodic* complementary sets of sequences in the collection are mates of each other [13].

Since the *aperiodic* complementary pair $\{\mathbf{a}_0, \mathbf{a}_1\}$ is also *periodic* complementary [12], the recursive method used to construct an *aperiodic* complementary pair [3][14] can be utilized to generate a *periodic* complementary pair, i.e.,

$$\begin{aligned} a_{0,k}^{(m)} &= a_{0,k}^{(m-1)} + a_{1,(k-d_m)}^{(m-1)} \\ a_{1,k}^{(m)} &= a_{0,k}^{(m-1)} - a_{1,(k-d_m)}^{(m-1)} \end{aligned}, \quad m \in [1, M], \tag{7}$$

with $a_{0,k}^{(0)} = a_{1,k}^{(0)} = \delta_k$, where $\delta_0 = 1, \delta_k = 0, k \neq 0$, and $\{d_1, d_2, \cdots, d_M\}$ is any permutation of $\{2^0, 2^1, \cdots, 2^{M-1}\}$. After M iterations, we get a pair of complementary sequences $\mathbf{a}_0$ and $\mathbf{a}_1$, each of length $N = 2^M$.

C. Construction of Optimal Training Sequences

Based on property 3) in [15], $\mathbf{a}_0$ and $\overleftarrow{\mathbf{a}_1}$ are also complementary, where $\overleftarrow{\mathbf{b}}$ is the reverse of the sequence $\mathbf{b}$, i.e., $\overleftarrow{b}_k = b_{N-1-k}, k \in [0, N-1]$. Moreover, according to Theorem 11 in [16], $\{\mathbf{a}_1, -\overleftarrow{\mathbf{a}_0}\}$ is the mate of $\{\mathbf{a}_0, \overleftarrow{\mathbf{a}_1}\}$. For $p = 2$, two uncorrelated sets of complementary sequences can be generated by the procedure described above, which are not enough when we have $N_T > 2$. However, we can use different shifts of the two sequences for different antennas. One possible training assignment for all the Tx antennas is given in Table I, where $\check{N}_T = 2\lceil N_T/2 \rceil$ and the Π's represent the shifts. Based on Table I, we have the midamble, which contains $\mathbf{s}_{n_t,A}$ and $\mathbf{s}_{n_t,B}$, shown in Fig. 2, where $\mathrm{CP}_{n_t,A} = [s_{n_t,A}(N-L), \cdots, s_{n_t,A}(N-1)]$ and $\mathrm{CP}_{n_t,B} = [s_{n_t,B}(N-L), \cdots, s_{n_t,B}(N-1)]$. Of course the training symbols will interfere with the data. However, the interference can be mitigated after channel estimation using the known training symbols.

As a simple example, for $N_T = 4$, $N = 4$, and $L = 1$ (two taps in each subchannel), Table I is reproduced in Table II, using $\mathbf{a}_0 = [+++-]$ and $\mathbf{a}_1 = [++-+]$, where "$+$" and "$-$" denotes "$+1$" and "-1", respectively. $\mathbf{a}_0$ and $\mathbf{a}_1$ can be obtained from (7) by $d_1 = 1$ and $d_2 = 2$. Based on the training symbols in Table II, $\mathbf{S}_\nu$, $\nu \in \{A, B\}$, can be generated according to (2). Therefore $\mathbf{S} = [\mathbf{S}_A, \mathbf{S}_B]$ is given by

$$\mathbf{S} = \left[\underbrace{\begin{matrix} + + + - \\ + + - + \\ + - + + \\ - + + + \\ - + + + \\ + + + - \\ + + - + \\ + - + + \end{matrix}}_{\mathbf{S}_A} \middle| \underbrace{\begin{matrix} + - + + \\ + - - - \\ + + + - \\ - - + - \\ + + - + \\ - + - - \\ - + + + \\ - - - + \end{matrix}}_{\mathbf{S}_B} \right]. \tag{8}$$

From (8), it is easy to check that $\mathbf{S}\mathbf{S}^\top = 2N\mathbf{I}_{N_T(L+1)} = 8\mathbf{I}_8$.

D. MIMO ISI Channel Estimator

For the two training sequences of Fig. 2, (1) can be rewritten as $[\mathbf{Y}_A, \mathbf{Y}_B] = \sqrt{\gamma/N_T}\mathbf{H}[\mathbf{S}_A, \mathbf{S}_B] + [\mathbf{E}_A, \mathbf{E}_B]$. Based on (2) and Table I, it is easy to check $\mathbf{S}_A\mathbf{S}_A^\top + \mathbf{S}_B\mathbf{S}_B^\top = 2N\mathbf{I}_{N_T(L+1)}$, which satisfies (6) and demonstrates the optimality of the training symbols given in Table I. Furthermore, (3) reduces to $\widehat{\mathbf{H}} = \left(\sum_{\nu\in\{A,B\}}\mathbf{Y}_\nu\mathbf{S}_\nu^\top\right)\left[\left(2N\sqrt{\frac{\gamma}{N_T}}\mathbf{I}_{L+1} + \sqrt{\frac{N_T}{\gamma}}\mathbf{C}_\Sigma^{-1}\right)^{-1} \otimes \mathbf{I}_{N_T}\right]$, whose elements are

$$\widehat{h}_{n_r,n_t}(l) = \frac{\sigma_l\sqrt{\gamma N_T}}{2N\gamma\sigma_l + N_T}\left[\sum_{\nu\in\{A,B\}}\mathbf{Y}_\nu\mathbf{S}_\nu^\top\right]_{n_r, lN_T+n_t}, \tag{9}$$

where $n_r \in [1, N_R]$, $n_t \in [1, N_T]$, and $l \in [0, L]$.

IV. FAST IMPLEMENTATION OF THE CHANNEL ESTIMATOR

In order to take the advantages of the filter structure in [3], we convert the circular convolution into the linear one and propose a similar efficient implementation, detailed as follows.

From (9), it is clear that the computations for all the N_R Rx antennas take the same procedure. Therefore, in what follows we focus on the implementation of the channel estimator on the n_r^{th} Rx antenna. First we define two row vectors $\mathbf{h}_{n_r,o}$ of length $N_o = \lceil N_T/2 \rceil (L+1)$ and $\mathbf{h}_{n_r,e}$ of length $N_e = \lfloor N_T/2 \rfloor (L+1)$. These row vectors include the CIRs of all the subchannels between the odd- and even-numbered Tx antennas and the n_r^{th} Rx antenna, respectively, as follows

$$\begin{aligned} \mathbf{h}_{n_r,o} &= \left[\mathbf{h}_{n_r,1}\ \mathbf{h}_{n_r,3}\ \cdots\ \mathbf{h}_{n_r,2\lceil N_T/2\rceil-1}\right], \\ \mathbf{h}_{n_r,e} &= \left[\mathbf{h}_{n_r,2}\ \mathbf{h}_{n_r,4}\ \cdots\ \mathbf{h}_{n_r,2\lfloor N_T/2\rfloor}\right], \end{aligned} \tag{10}$$

where $\mathbf{h}_{n_r,n_t} = [h_{n_r,n_t}(0)\ h_{n_r,n_t}(1)\ \cdots\ h_{n_r,n_t}(L)]$, as defined previously. Moreover, we define four circulant matrices

$$\begin{aligned} \mathbf{C}_{0,A} &= \mathrm{circ}(\mathbf{a}_0, N_o), \quad \mathbf{C}_{1,A} = \mathrm{circ}(\mathbf{a}_1, N_e), \\ \mathbf{C}_{0,B} &= \mathrm{circ}(\overleftarrow{\mathbf{a}_1}, N_o), \quad \mathbf{C}_{1,B} = \mathrm{circ}(-\overleftarrow{\mathbf{a}_0}, N_e), \end{aligned} \tag{11}$$

with $\mathrm{circ}(\mathbf{c})$, $\mathbf{c} = [c_0, c_1, \cdots, c_{n-1}]$, as an $n \times n$ circulant matrix, whose $(i,j)^{\text{th}}$ element is $c_{(j-i)_n}$, and $\mathrm{circ}(\mathbf{c}, m)$ includes the first m rows of $\mathrm{circ}(\mathbf{c})$ such that $m \le n$. Based on Table I and (9)-(11), estimators for the channel coefficients specified in (10) can be written as

$$\begin{aligned} \widehat{\mathbf{h}}_{n_r,o} &= \Big(\textstyle\sum_{\nu\in\{A,B\}} \mathbf{y}_{n_r,\nu} \mathbf{C}_{0,\nu}^{\top}\Big)\Big(\mathbf{I}_{\lceil N_T/2 \rceil} \otimes \mathbf{Q}\Big), \\ \widehat{\mathbf{h}}_{n_r,e} &= \Big(\textstyle\sum_{\nu\in\{A,B\}} \mathbf{y}_{n_r,\nu} \mathbf{C}_{1,\nu}^{\top}\Big)\Big(\mathbf{I}_{\lfloor N_T/2 \rfloor} \otimes \mathbf{Q}\Big), \end{aligned} \tag{12}$$

where $\mathbf{y}_{n_r,\nu} = [y_{n_r,\nu}(0), y_{n_r,\nu}(1), \cdots, y_{n_r,\nu}(N-1)]$, $\nu \in \{A,B\}$, $\mathbf{Q} = \mathrm{diag}(\mathbf{f})$ and $\mathbf{f}$ is a $1 \times (L+1)$ row vector, and defined as $\mathbf{f} = [\frac{\sigma_0\sqrt{\gamma N_T}}{2N\gamma\sigma_0 + N_T}, \frac{\sigma_1\sqrt{\gamma N_T}}{2N\gamma\sigma_1 + N_T}, \cdots, \frac{\sigma_L\sqrt{\gamma N_T}}{2N\gamma\sigma_L + N_T}]$.

Since $\Big(\mathbf{I}_{\lceil N_T/2 \rceil} \otimes \mathbf{Q}\Big)$ and $\Big(\mathbf{I}_{\lfloor N_T/2 \rfloor} \otimes \mathbf{Q}\Big)$ in (12) serve as scaling factors only for channel estimation, we ignore them in the following discussion. Let us focus the first vector-matrix multiplication in (12), where the k^{th} element of the resulting row vector can be written as $[\mathbf{y}_{n_r,A}\mathbf{C}_{0,A}^{\top}]_{1,k+1} = \sum_{j=0}^{N-1} [y_{n_r,A}(j) a_{0,(j-k)_N}]$ $= \sum_{n=0}^{N-1} [y_{n_r,A}((k-1-n)_N) \overleftarrow{a}_{0,n}]$, $k \in [0, N_o - 1]$, which implies that $\mathbf{y}_{n_r,A}\mathbf{C}_{0,A}^{\top}$ is the circular convolution of $\mathbf{y}_{n_r,A}$ and $\overleftarrow{\mathbf{a}_0}$, i.e., $\mathbf{y}_{n_r,A}\mathbf{C}_{0,A}^{\top} = \mathbf{y}_{n_r,A} \circledast_N \overleftarrow{\mathbf{a}_0}$. With the same reasoning we get $\mathbf{y}_{n_r,B}\mathbf{C}_{0,B}^{\top} = \mathbf{y}_{n_r,B} \circledast_N \mathbf{a}_1$, $\mathbf{y}_{n_r,A}\mathbf{C}_{1,A}^{\top} = \mathbf{y}_{n_r,A} \circledast_N \overleftarrow{\mathbf{a}_1}$ and $\mathbf{y}_{n_r,B}\mathbf{C}_{1,B}^{\top} = \mathbf{y}_{n_r,B} \circledast_N (-\mathbf{a}_0)$.

We define two new row vectors $\breve{\mathbf{y}}_{n_r,A}$ and $\breve{\mathbf{y}}_{n_r,B}$, each of length $N + N_o - 1$, as $\breve{\mathbf{y}}_{n_r,\nu} = [y_{n_r,\nu}(0), \cdots,$ $y_{n_r,\nu}(N-1), y_{n_r,\nu}(0), \cdots, y_{n_r,\nu}(N_o - 2)]$, i.e., $\breve{y}_{n_r,\nu}(n) = y_{n_r,\nu}((n)_N)$, where $\nu \in \{A,B\}$ and $n \in [0, N + N_o - 2]$. Using $\breve{\mathbf{y}}_{n_r,A}$ and $\breve{\mathbf{y}}_{n_r,B}$, we can convert four circular convolutions of (12) into linear convolutions. For example, $[\mathbf{y}_{n_r,A}\mathbf{C}_{0,A}^{\top}]_{1,k+1} = \sum_{n=0}^{N-1} [\breve{y}_{n_r,A}(N + k - 1 - n) \overleftarrow{a}_{0,n}]$, $k \in [0, N_o - 1]$, and hence all the elements of $\mathbf{y}_{n_r,A}\mathbf{C}_{0,A}^{\top}$ can be computed by the linear convolution of $\breve{\mathbf{y}}_{n_r,A}$ and $\overleftarrow{\mathbf{a}_0}$, i.e., $\breve{\mathbf{y}}_{n_r,A} * \overleftarrow{\mathbf{a}_0}$.

Based on the linear convolutions and according to [3, Sec. V], (12) can be implemented using the efficient filter structure shown in Fig. 3 on the n_r^{th} Rx antenna, where "Rep" generates $\breve{\mathbf{y}}_{n_r,\nu}$ from $\mathbf{y}_{n_r,\nu}$, "$\bigoplus$" is the complex adder, "$\bigotimes$" represents the complex multiplier with one real input and one complex input, "z^{-D}" is the delay unit of length D, $\{d_1, d_2, \cdots, d_M\}$ is defined in Sec. III-B, "Ext 1" discards the first $N-1$ values, keeps the next N_o elements and throws away the rest, whereas "Ext 2" discards the first $N-1$ values, keeps the next N_e elements and discards the rest. "MUX" multiplexes inputs in groups of $L+1$ elements, with the first group coming from the upper branch, $\mathbf{q} = \mathbf{1}_{1\times N_T} \otimes \mathbf{f}$ is the scaling vector, derived from $\mathbf{Q}$ according to (12), and $\widehat{\mathbf{h}}_{n_r} = \Big[\widehat{\mathbf{h}}_{n_r,1}, \cdots, \widehat{\mathbf{h}}_{n_r,N_T}\Big]$ is the estimated CIR.

EGC stands for efficient Golay correlator and serves as the matched filters of $\{\mathbf{a}_0, \mathbf{a}_1\}$, FGC denotes fast Golay correlator and is functioned as the matched filter of $\{\overleftarrow{\mathbf{a}_1}, -\overleftarrow{\mathbf{a}_0}\}$ [3].

V. COMPARISON WITH EXISTING SEQUENCES

By the following comparisons with ZCZ sequences [5][6] and delta sequences [2, Table I], we can conclude that the proposed scheme outperforms them in terms of both low PAPR and low implementation complexity.

A. Comparison with ZCZ Sequences

ZCZ sequences [5][6] satisfy (5), therefore, can be used for optimal training. Compared to the scenario where each Tx antenna uses one ZCZ sequence per transmission frame, our scheme contains only an extra guard period of length L (see Fig. 2, where the 2^{nd} CP is used to separate the two training sequences). However, this loss in spectrum efficiency is negligible, due to the large frame length.

Regarding to the hardware complexity, according to Fig. 3, on each Rx antenna, only $4\log_2 N + 2$ $\bigoplus$'s [3] and one $\bigotimes$ are needed. However, for ZCZ sequences, they need $N_T(2N-1)$ $\bigoplus$'s and one $\bigotimes$, since N_T finite impulse response filters are required, each with $N_s = 2N$ coefficients. This is more complex than the proposed scheme of Fig. 3.

The PAPR of a sequence $\mathbf{x} = [x_1, x_2, \cdots, x_N]$ is defined as $\mathrm{PAPR}(\mathbf{x}) = \frac{\max_{1 \le n \le N} |x_n|^2}{\frac{1}{N}\sum_{n=1}^{N} |x_n|^2}$. Hence the PAPR is 1 for the training sequences given in Table I and ZCZ sequences.

B. Comparison with Impulse Sequences

Similar to the above comparison, when one compares our method with the impulse sequences [2, Table I], our scheme only needs an extra guard period of length L.

Regarding to the complexity, the impulse sequences only require one $\bigotimes$ for scaling, which has less complexity. However, the PAPR of the impulse sequences is $N_T(L+1)$, much greater than 1, and not desired in practice.

VI. SIMULATION RESULTS AND CONCLUSION

In the simulation we take $L = 7$, $N_T = 4$, $N_R = 4$, and $N = 16, 32, 64$. Moreover, each subchannel has the same exponential power delay profile such that $\sigma_l = \frac{(1-e^{-1})e^{-l}}{1-e^{-L-1}}, l \in [0, L]$. Fig. 4 shows the normalized theoretical minimum Bayesian CRLB, given by $\sum_{l=0}^{L} \frac{N_T \sigma_l}{2N\gamma\sigma_l + N_T}$, derived from (4) and (5) with $N_s = 2N$, normalized by $N_T N_R$. The simulated normalized TMSE, $\overline{\|\widehat{\mathbf{H}} - \mathbf{H}\|_F^2} / \overline{\|\mathbf{H}\|_F^2}$, is plotted as well, which perfectly matches the minimum Bayesian CRLB.

The numerical results of ZCZ sequences, $G(32,4,4)$, $G(64,4,8)$ and $G(128,4,16)$ [6], and impulse sequences of length $N_s = 2N, N = 16, 32, 64$ [2], are also plotted in Fig. 4, where we can see that, for all the considered cases, all of them attain the minimum Bayesian CRLB since they satisfy the condition in (5).

REFERENCES

[1] L. Tong, B. M. Sadler, and M. Dong, "Pilot-assisted wireless transmissions: General model, design criteria, and signal processing," *IEEE Signal Processing Mag.*, vol. 21, no. 6, pp. 12–25, 2004.

[2] X. Ma, L. Yang, and G. B. Giannakis, "Optimal training for MIMO frequency-selective fading channels," *IEEE Trans. Wireless Commun.*, vol. 4, pp. 453–466, Mar. 2005.

[3] S. Wang and A. Abdi, "Aperiodic complementary sets of sequences-based MIMO frequency selective channel estimation," *IEEE Commun. Lett.*, vol. 9, pp. 891–893, Oct. 2005.

[4] *Physical Channels and Mapping of Transport Channels onto Physical Channels*, CWTS Std. TS C102, 1999.

[5] S.-A. Yang and J. Wu, "Optimal binary training sequence design for multiple-antenna systems over dispersive fading channels," *IEEE Trans. Veh. Technol.*, vol. 51, pp. 1271–1276, Sept. 2002.

[6] P. Fan and W. H. Mow, "On optimal training sequence design for multiple-antenna systems over dispersive fading channels and its extensions," *IEEE Trans. Veh. Technol.*, vol. 53, pp. 1623–1626, Sept. 2004.

[7] P. J. Davis, *Circulant Matrices*. New York: Wiley, 1979.

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

[9] S. M. Kay, *Fundamentals of Statistical Signal Processing, Volume I: Estimation Theory*. Upper Saddle River, NJ: Prentice Hall PTR, 1993.

[10] H. L. Van Trees, *Detection, Estimation, and Modulation Theory, Vol. I*. New York: Wiley, 1968.

[11] R. D. Gill and B. Y. Levit, "Applications of the van Trees inequality: A Bayesian Cramér-Rao bound," *Bernoulli*, vol. 1, pp. 59–79, 1995.

[12] L. Bomer and M. Antweiler, "Periodic complementary binary sequences," *IEEE Trans. Inform. Theory*, vol. 36, pp. 1487–1494, Nov. 1990.

[13] P. Fan and M. Darnell, *Sequence Design for Communications Applications*. Hertfordshire, UK: Research Studies Press, 1996.

[14] S. Z. Budišin, "Efficient pulse compressor for Golay complementary sequences," *Electron. Lett.*, vol. 27, pp. 219–220, Jan. 1991.

[15] M. J. E. Golay, "Complementary series," *IEEE Trans. Inform. Theory*, vol. 7, pp. 82–87, Apr. 1961.

[16] C. C. Tseng and C. L. Liu, "Complementary set of sequences," *IEEE Trans. Inform. Theory*, vol. 18, pp. 644–652, Sept. 1972.

TABLE I

ASSIGNMENT OF TRAINING SEQUENCES TO TX ANTENNAS

Tx	$\mathbf{s}_{n_t,A}$	$\mathbf{s}_{n_t,B}$
1	$\mathbf{a}_0$	$\overleftarrow{\mathbf{a}_1}$
2	$\mathbf{a}_1$	$-\overleftarrow{\mathbf{a}_0}$
3	$\mathbf{a}_0\mathbf{\Pi}_N^{L+1}$	$\overleftarrow{\mathbf{a}_1}\mathbf{\Pi}_N^{L+1}$
4	$\mathbf{a}_1\mathbf{\Pi}_N^{L+1}$	$-\overleftarrow{\mathbf{a}_0}\mathbf{\Pi}_N^{L+1}$
$\vdots$	$\vdots$	$\vdots$
$\check{N}_T-1$	$\mathbf{a}_0\mathbf{\Pi}_N^{(\lceil N_T/2\rceil-1)(L+1)}$	$\overleftarrow{\mathbf{a}_1}\mathbf{\Pi}_N^{(\lceil N_T/2\rceil-1)(L+1)}$
$\check{N}_T$	$\mathbf{a}_1\mathbf{\Pi}_N^{(\lceil N_T/2\rceil-1)(L+1)}$	$-\overleftarrow{\mathbf{a}_0}\mathbf{\Pi}_N^{(\lceil N_T/2\rceil-1)(L+1)}$

TABLE II

TRAINING EXAMPLE WITH $N_T = 4$, $N = 4$ AND $L = 1$.

Tx	$\mathbf{s}_{n_t,A}$	$\mathbf{s}_{n_t,B}$
1	$[+++-]$	$[+-++]$
2	$[++-+]$	$[+---]$
3	$[+-++]$	$[+++-]$
4	$[-+++]$	$[--+-]$

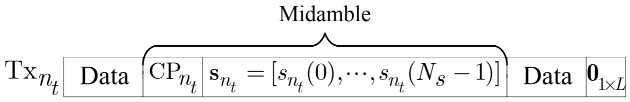

Fig. 1. A typical frame structure on the n_t^{th} Tx antenna using midamble.

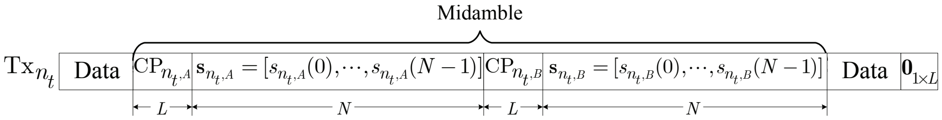

Fig. 2. The proposed entire frame structure on the n_t^{th} Tx antenna.

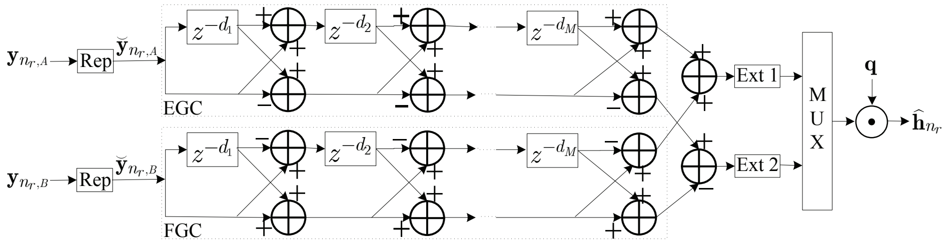

Fig. 3. The proposed fast hardware implementation of the channel estimator on the n_r^{th} Rx antenna.

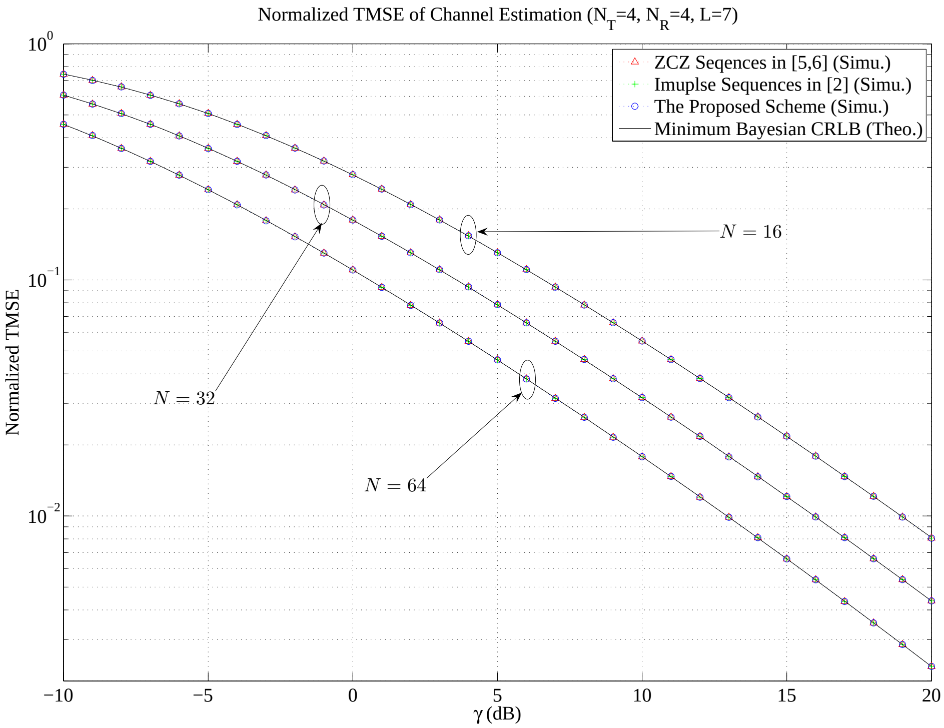

Fig. 4. The normalized TMSE of the proposed estimator for different N's (Note that $N = 16$ is the minimum length of our scheme for a meaningful estimate[5]) and comparison with ZCZ and impulse sequences.