

# MIMO ISI Channel Estimation Using Uncorrelated Golay Complementary Sets of Polyphase Sequences

Shuangquan Wang<sup>†</sup>, *Member, IEEE*, Ali Abdi<sup>‡</sup>, *Senior Member, IEEE*

**Abstract**—In this paper, optimal training sequence design for multiple input multiple output (MIMO) intersymbol interference (ISI) channels is addressed and several novel low-complexity channel estimators are proposed, using uncorrelated Golay complementary sets of polyphase sequences<sup>1</sup>. The theoretical analysis and simulation show that when the additive noise is Gaussian, the proposed best linear unbiased estimator (BLUE) achieves the minimum possible classical Cramér-Rao lower bound (CRLB), if the channel coefficients are regarded as unknown deterministic. On the other hand, the proposed linear minimum mean square error (LMMSE) estimator attains the minimum possible Bayesian CRLB, when the underlying channel coefficients are Gaussian and independent of the additive Gaussian noise. The proposed channel estimators not only achieve the best estimation performance, but also can be implemented with low complexity, via DSP or ASIC/FPGA. This has been possible due to the special structures intrinsic to uncorrelated Golay complementary sets of polyphase sequences, which makes the proposed channel estimators ready to use in the practical MIMO systems.

**Index Terms**—MIMO, Intersymbol Interference, Golay Sequences, Complementary Sets of Sequences, Channel Estimation, Training Sequences, and Frequency Selective.

## I. INTRODUCTION

THE utilization of antenna arrays at the base station and the mobile station in a wireless communication system increases the capacity linearly with  $\min(N_T, N_R)$ , where  $N_T$  and  $N_R$  are the numbers of transmit and receive antennas, respectively, provided that the environment is sufficiently rich in multi-path components [3][4]. The early analyses and simulations of the multiple-input multiple-output (MIMO) systems relied on the assumption of perfect channel state information (CSI) [3][5]. However, in practice one needs to estimate CSI for efficient implementation of MIMO systems.

In general, there are two classes of methods for CSI estimation: blind identification (see, e.g., [6]) and training-based estimation. For quasi-static or slowly-varying fading channels, training-based channel estimation at the receiver is widely

used in practice [7], as it has much less complexity and better performance, compared to the blind approach. The insertion of training symbols consumes some transmission bandwidth, but this overhead is negligible if they are only a small fraction of the transmitted frame. Moreover, the training-based scheme can be optimal in the high signal-to-noise ratio (SNR) region [8]–[10]. When the training-based approach is used, the goal of channel estimator design is to make the estimation error as small as possible, with low implementation complexity, for fixed training sequence length and power.

Optimal training sequence design for MIMO systems has been studied in the literature (see [9][11]–[18] and references therein). Some information theoretical guidelines for sequence design over MIMO flat fading channels and single-input single-output (SISO) frequency-selective channels are given in [9] and [10], respectively. The optimal sequence design for MIMO flat fading channels in the presence of colored interference is studied [15]. A lower bound on training-based channel estimation error for MIMO intersymbol interference (ISI)<sup>2</sup> channels, based on a parameterized channel model, is presented in [19]. An optimal training design for MIMO ISI channels is given in [11], where training symbols are superimposed on data in the system model. For both SISO and MIMO ISI channels, optimal training sequences are delta impulses [10, (13)] [11, Table I], which maximize the lower bound of the ergodic channel capacity. For MIMO-OFDM systems, optimal training sequence design is discussed in [16]–[18] via minimizing the mean square error of the channel estimator.

For MIMO flat fading channels, it is straightforward to design training sequences to satisfy the semi-unitary condition for the minimum mean square error (MMSE) of the estimator, given in [9, (18)], [20, (9.4.16)], and [11, C1]. For example, a Hadamard matrix or a part of it can be used. However, for the MIMO ISI scenario, the training sequences, which satisfy the semi-unitary condition, must have impulse-like auto-correlations and zero cross-correlations within a given correlation window, whose length depends on the delay spread of the channel. One way to achieve zero cross-correlation is to transmit training symbols only from one antenna at a time [11], where each training sequence is a delta sequence. However, this approach may result in high peak-to-average power ratio (PAPR), and low energy efficiency, which are important concerns in practice.

In the literature, there are some PAPR-friendly optimal sequences, which satisfy the semi-unitary condition, and can

Manuscript received February 25, 2006; revised November 11, 2006; accepted December 09, 2006. This paper was presented in part at the *IEEE Global Telecommunications Conference*, St. Louis, MO, 2005 [1], and the *IEEE Sarnoff Symposium*, Princeton, NJ, 2006 [2]. The review and approval of this paper was coordinated by Prof. Zhengdao Wang.

<sup>†</sup>S. Wang was with the Department of Electrical and Computer Engineering, New Jersey Institute of Technology, Newark, NJ 07102, USA. He is now with NEC Laboratories America, Inc., Princeton, NJ 08540, USA. (e-mail: shuangquan.wang@njit.edu).

<sup>‡</sup>A. Abdi is with the Department of Electrical and Computer Engineering, New Jersey Institute of Technology, Newark, NJ 07102, USA (e-mail: ali.abdi@njit.edu).

Digital Object Identifier xx.xxxx/TVT.200x.xxxxxx

<sup>1</sup>A polyphase sequence is a sequence of complex numbers, each of unit magnitude.

<sup>2</sup>In this paper, we use ISI and frequency-selective channels interchangeably.

be used for training. Examples include a set of sequences with a zero correlation zone (ZCZ) [12][13][21][22] and different phases of a perfect<sup>3</sup> polyphase sequence such as the Frank sequence [23] or Chu sequence [24]. However, both ZCZ and perfect polyphase sequences are developed based on the periodic correlation properties. Therefore, they can not be applied to scenarios where aperiodic correlations are needed, for example, when using a zero-padding guard period to separate the transmitted data and training symbols, as discussed in Sec. III-A. Moreover, the implementation complexity of the corresponding channel estimators could be an issue, due to the lack of proper structures.

Golay complementary sequences [25] have been widely used in infrared spectrometry [26], radar [27], synchronization [28], PAPR control [29], MC-CDMA [30], channel identification [31], and SISO ISI channel estimation [32]–[34], due to their good structure and perfect correlation properties. Recently, uncorrelated *aperiodic* Golay complementary sets of *binary* sequences are used for optimal training in a zero-padding block transmission system [35], and uncorrelated *periodic* Golay complementary sets of *binary* sequences in a cyclic-prefix padding block transmission system [36]. In this paper, we extend the *binary* sequences in [35] and [36] to the *polyphase* ones for optimal training. Obviously, the results in this paper will include those in [35] and [36] as special cases. After the system model and the criterion for optimal training are developed, several optimal low-complexity channel estimation schemes using uncorrelated Golay complementary sets of *polyphase* sequences are proposed. The theoretical analysis and simulation show that when the additive noise is Gaussian, the proposed best linear unbiased estimator (BLUE) achieves the minimum possible classical Cramér-Rao lower bound (CRLB), if the channel coefficients are regarded as unknown deterministics. On the other hand, the proposed linear minimum mean square error (LMMSE) estimator attains the minimum possible Bayesian CRLB<sup>4</sup>, when the underlying channel coefficients are Gaussian and independent of the additive Gaussian noise. Last but not the least, the proposed schemes can be implemented with low complexity, via DSP or ASIC/FPGA<sup>5</sup>, due to the special structure intrinsic to Golay complementary *polyphase* sequences.

The rest of this paper is organized as follows. Definition and construction of uncorrelated aperiodic and periodic Golay complementary sets of *polyphase* sequences are given in Sec. II, and MIMO system and channel models are presented in Sec. III. Sec. IV and V deal with the optimal training criterion and construction of optimal training sequences using uncorrelated Golay complementary sets of *polyphase* sequences, respectively. The channel estimation algorithm and its fast DSP and ASIC/FPGA implementations are studied in Sec. VI and VII, respectively. Comparison with other existing optimal sequences and the simulation results are presented in

Sec. VIII and IX, respectively, and the concluding remarks are summarized in Sec. X.

*Notation:*  $(\cdot)^*$  is reserved for the complex conjugate,  $(\cdot)'$  for the matrix transpose,  $(\cdot)^\dagger$  for the matrix Hermitian,  $(\cdot)^{-1}$  for the matrix inverse,  $\text{tr}[\cdot]$  for the trace of a matrix,  $\propto$  for proportional to,  $\text{diag}(\sigma_1, \sigma_2, \dots, \sigma_n)$  denotes a diagonal matrix with  $\sigma_1, \sigma_2, \dots, \sigma_n$  on the main diagonal,  $\text{dg}(\mathbf{A})$  is a diagonal matrix which contains the main diagonal elements of matrix  $\mathbf{A}$ ,  $\text{vec}(\cdot)$  stacks all the columns of its matrix argument into one tall column vector,  $[\mathbf{A}]_{m,n}$ ,  $m, n \geq 0$ , is the  $(m, n)^{\text{th}}$  element of the matrix  $\mathbf{A}$ ,  $\mathbb{E}[\cdot]$  is the mathematical expectation,  $\overline{(\cdot)}$  is the sample average,  $\mathbf{0}_{m \times n}$  is an  $m \times n$  zero matrix,  $\mathbf{1}_{m \times n}$  is an  $m \times n$  matrix whose entries are all 1,  $\mathbf{I}_m$  denotes the  $m \times m$  identity matrix,  $t \in [m, n]$  implies that  $t$  is an integer such that  $m \leq t \leq n$ ,  $\otimes$  represents the Kronecker product,  $\odot$  stands for the (elementwise) Hadamard product,  $\oplus$  represents the elementwise complex addition of two vectors,  $\boxtimes$  and  $\boxplus$  represent the multiplication and addition of two complex numbers, respectively,  $\lceil x \rceil$  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  $m$  [40, p. 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 circular convolution “mod  $N$ ”, and  $*$  for linear convolution. All the lower-case bold letters represent row and column vectors, whereas upper-case bold letters are used for matrices.

## II. DEFINITION AND GENERATION OF UNCORRELATED GOLAY COMPLEMENTARY SETS OF POLYPHASE SEQUENCES

For convenience and completeness, we give a brief description of uncorrelated aperiodic and periodic Golay complementary sets of *polyphase* sequences. More discussion can be found in [41]–[44].

### A. Aperiodic Case

Let  $\mathbf{a}_i = [a_{i,0}, a_{i,1}, \dots, a_{i,(N-1)}]$  and  $\mathbf{b}_i = [b_{i,0}, b_{i,1}, \dots, b_{i,(N-1)}]$  be sequences of *complex* numbers with unit amplitudes. The aperiodic cross-correlation function (ACCF) between  $\mathbf{a}_i$  and  $\mathbf{b}_i$  is defined by

$$\varphi_{\mathbf{a}_i, \mathbf{b}_i}(k) = \begin{cases} \sum_{j=0}^{N-1-k} a_{i,j} b_{i,(j+k)}^* & k \in [0, N-1], \\ \sum_{j=|k|}^{N-1} a_{i,j} b_{i,(j-k)}^* & k \in [-N+1, -1]. \end{cases} \quad (1)$$

A set of  $p$  sequences, each with  $N$  elements,  $\{\mathbf{a}_i\}_{i=0}^{p-1}$  is aperiodic complementary if and only if (iff)  $\sum_{i=0}^{p-1} \varphi_{\mathbf{a}_i, \mathbf{a}_i}(k) = 0$ ,  $k \neq 0$ , and  $\sum_{i=0}^{p-1} \varphi_{\mathbf{a}_i, \mathbf{a}_i}(0) = pN$ , where  $\varphi_{\mathbf{a}_i, \mathbf{a}_i}(k)$ ,  $|k| \leq N-1$  is the aperiodic autocorrelation function (AACF) of  $\mathbf{a}_i$ .

If another set of sequences  $\{\mathbf{b}_i\}_{i=0}^{p-1}$  is aperiodic complementary and  $\sum_{i=0}^{p-1} \varphi_{\mathbf{a}_i, \mathbf{b}_i}(k) = 0$ ,  $|k| \leq N-1$ , 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 aperiodic Golay complementary sets of *polyphase* sequences  $\{\mathbf{a}_i\}_{i=0}^{p-1}$ ,  $\{\mathbf{b}_i\}_{i=0}^{p-1}$ ,  $\dots$ ,  $\{\mathbf{z}_i\}_{i=0}^{p-1}$  are *mutually uncorrelated* if every two aperiodic Golay complementary sets of *polyphase* sequences in the collection are mates of each other [44].

<sup>3</sup>Its periodic autocorrelation function is a delta impulse.

<sup>4</sup>The concept of CRLB for random parameter estimation was introduced in [37], and named Bayesian CRLB later in [38] and [39].

<sup>5</sup>The abbreviations DSP, ASIC and FPGA stand for digital signal processor, application-specific integrated circuit and field-programmable gate array, respectively.

### B. Periodic Case

The periodic cross-correlation function (PCCF) between  $\mathbf{a}_i$  and  $\mathbf{b}_i$  is defined as

$$\tilde{\varphi}_{\mathbf{a}_i, \mathbf{b}_i}(k) = \sum_{j=0}^{N-1} a_{i,j} b_{i,(j+k)_N}^*, |k| \leq N-1. \quad (2)$$

A set of sequences  $\{\mathbf{a}_i\}_{i=0}^{p-1}$ , each with  $N$  elements, is periodic complementary *iff*  $\sum_{i=0}^{p-1} \tilde{\varphi}_{\mathbf{a}_i, \mathbf{a}_i}(k) = 0, k \neq 0$ , and  $\sum_{i=0}^{p-1} \tilde{\varphi}_{\mathbf{a}_i, \mathbf{a}_i}(0) = pN$ , where  $\tilde{\varphi}_{\mathbf{a}_i, \mathbf{a}_i}(k), |k| \leq N-1$  is the periodic autocorrelation function (PACF) of  $\mathbf{a}_i$ .

If another set of sequences  $\{\mathbf{b}_i\}_{i=0}^{p-1}$  is periodic complementary and  $\sum_{i=0}^{p-1} \tilde{\varphi}_{\mathbf{a}_i, \mathbf{b}_i}(k) = 0, |k| \leq N-1$ , 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 Golay complementary sets of *polyphase* sequences  $\{\mathbf{a}_i\}_{i=0}^{p-1}, \{\mathbf{b}_i\}_{i=0}^{p-1}, \dots, \{\mathbf{z}_i\}_{i=0}^{p-1}$  are *mutually uncorrelated* if every two periodic Golay complementary sets of *polyphase* sequences in the collection are mates of each other [44].

### C. Sequence Generation

Based on (1) and (2), it is easy to verify

$$\tilde{\varphi}_{\mathbf{a}_i, \mathbf{b}_i}(k) = \begin{cases} \varphi_{\mathbf{a}_i, \mathbf{b}_i}(k) + \varphi_{\mathbf{a}_i, \mathbf{b}_i}(k-N) & k \in [0, N-1], \\ \varphi_{\mathbf{a}_i, \mathbf{b}_i}(k) + \varphi_{\mathbf{a}_i, \mathbf{b}_i}(k+N) & k \in [-N+1, -1], \end{cases} \quad (3)$$

which shows that aperiodic Golay complementary sets of *polyphase* sequences are also periodic Golay complementary [43][45]. Therefore, we only explain how to construct aperiodic Golay complementary sets of *polyphase* sequences.

In this paper, we focus on the case of  $p = 2$ , and  $p > 2$  can be easily addressed, based on the methods described in [41]. According to the complex extension of property 9) in [25], the Golay complementary pair  $\{\mathbf{a}_0, \mathbf{a}_1\}$  with  $N = 2^M$  elements,  $M \geq 1$ , can be constructed by the following recursive equation [27]

$$\begin{aligned} a_{0,k}^{(m)} &= a_{0,k}^{(m-1)} + w_m a_{1,(k-d_m)}^{(m-1)}, \\ a_{1,k}^{(m)} &= a_{0,k}^{(m-1)} - w_m a_{1,(k-d_m)}^{(m-1)}, \end{aligned} \quad m \in [1, M], \quad (4)$$

with  $a_{0,k}^{(0)} = a_{1,k}^{(0)} = \delta_k$ , where  $\delta_0 = 1, \delta_k = 0, k \neq 0$  and  $w_k$  is the complex number with unit amplitude, i.e.,  $|w_m| = 1, \forall m$ . In (4),  $d_m$  is the  $m^{\text{th}}$  element of the delay vector  $\mathbf{d}$ , defined by

$$\mathbf{d} = [d_1, d_2, \dots, d_M], \quad (5)$$

which is a permutation of  $[2^0, 2^1, \dots, 2^{M-1}]$ . After  $M$  iterations, we get a pair of Golay complementary *polyphase* sequences  $\mathbf{a}_0^{(M)}$  and  $\mathbf{a}_1^{(M)}$ , each of length  $N$ . To simplify notation, we write  $\mathbf{a}_0^{(M)}$  as  $\mathbf{a}_0$ , and  $\mathbf{a}_1^{(M)}$  as  $\mathbf{a}_1$ , i.e.,  $\mathbf{a}_0 = \mathbf{a}_0^{(M)}$  and  $\mathbf{a}_1 = \mathbf{a}_1^{(M)}$ .

By the complex extension of property 3) in [25],  $\mathbf{a}_0$  and  $\overleftarrow{\mathbf{a}}_1^*$  are complementary too, where  $\overleftarrow{\mathbf{b}}$  is the reverse of the sequence  $\mathbf{b}$ , i.e.  $\overleftarrow{b}_i = b_{N-1-i}, i \in [0, N-1]$ .

According to the complex version of Theorem 11 in [41], if  $\{\mathbf{a}, \mathbf{b}\}$  are a Golay complementary set of two *polyphase* sequences with length  $N$ , then  $\{\overleftarrow{\mathbf{b}}^*, -\overleftarrow{\mathbf{a}}^*\}$  is its mate. Therefore,  $\{\mathbf{a}_1, -\overleftarrow{\mathbf{a}}_0^*\}$  is the mate of  $\{\mathbf{a}_0, \overleftarrow{\mathbf{a}}_1^*\}$ , vice versa.

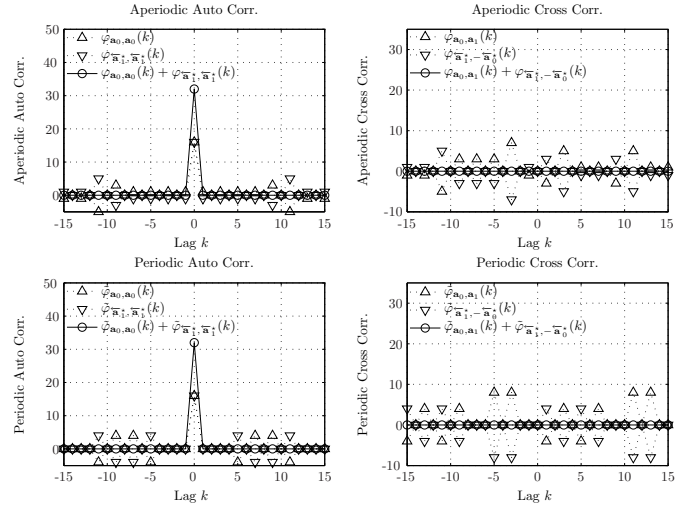


Fig. 1. The aperiodic and periodic auto and cross-correlation functions of the two uncorrelated Golay complementary sets of *polyphase* sequences, with  $N = 16$ .

For example, if  $N = 16, w_k = 1, \forall k$ , and  $\mathbf{d} = [1, 2, 4, 8]$ , i.e.,  $d_m = 2^{m-1}$ , we obtain a Golay complementary binary pair  $\{\mathbf{a}_0, \mathbf{a}_1\}$  from (4) as  $\mathbf{a}_0 = [+++-+ - + + + - - - + -]$  and  $\mathbf{a}_1 = [+++-+ - + - - - + + - +]$ , where “+” denotes “+1” and “-” represents “-1”. Various correlations of  $\{\mathbf{a}_0, \overleftarrow{\mathbf{a}}_1^*\}$  and its mate  $\{\mathbf{a}_1, -\overleftarrow{\mathbf{a}}_0^*\}$  are shown in Fig. 1. Clearly, the individual AACFs of the first set, i.e.,  $\varphi_{\mathbf{a}_0, \mathbf{a}_0}(k)$  and  $\varphi_{\overleftarrow{\mathbf{a}}_1^*, \overleftarrow{\mathbf{a}}_1^*}(k)$ , are not impulse, but their sum  $\varphi_{\mathbf{a}_0, \mathbf{a}_0}(k) + \varphi_{\overleftarrow{\mathbf{a}}_1^*, \overleftarrow{\mathbf{a}}_1^*}(k)$  is an impulse. The same property applies to the second set. The individual ACCFs  $\varphi_{\mathbf{a}_0, \mathbf{a}_1}(k)$  and  $\varphi_{\overleftarrow{\mathbf{a}}_1^*, -\overleftarrow{\mathbf{a}}_0^*}(k)$  between the two sets are not zero for all lags, but their sum  $\varphi_{\mathbf{a}_0, \mathbf{a}_1}(k) + \varphi_{\overleftarrow{\mathbf{a}}_1^*, -\overleftarrow{\mathbf{a}}_0^*}(k)$  is zero everywhere. The same observations apply to the PACFs and PCCFs.

The family size of uncorrelated Golay complementary sets with two *polyphase* sequences per set ( $p = 2$ ), i.e., the maximum number of *mutually uncorrelated* Golay complementary sets, is a parameter that indicates how many transmit antennas uncorrelated Golay sets can support, without any modification to the original sequences. From the above description, the family size is at least two, and they are given by  $\{\mathbf{a}_0, \overleftarrow{\mathbf{a}}_1^*\}$  and  $\{\mathbf{a}_1, -\overleftarrow{\mathbf{a}}_0^*\}$ <sup>6</sup>. For binary sequences with aperiodic correlations, the family size is two if each set only has two sequences [41][46]. However, for both unit-amplitude *polyphase* sequences with aperiodic or periodic correlations and binary sequences with periodic correlations, the family size is unknown, to the best of our knowledge. Studying the family size of the mentioned cases is out of the scope of this paper, and hence we assume it is equal to two in this paper, as it is enough for us to consider two uncorrelated Golay complementary sets.

### III. SYSTEM AND CHANNEL MODELS

Similarly to [35] and [36], block transmission over block fading channels is assumed here. In what follows, we consider

<sup>6</sup>Given a Golay complementary pair  $\{\mathbf{a}_0, \mathbf{a}_1\}$ , the uncorrelated sets are not unique, for example,  $\{\mathbf{a}_0, \mathbf{a}_1\}$  and  $\{\overleftarrow{\mathbf{a}}_1^*, -\overleftarrow{\mathbf{a}}_0^*\}$  are another possibility.

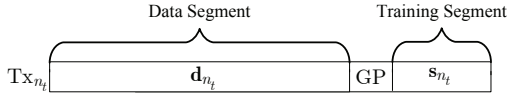


Fig. 2. A frame structure in a MIMO system.

an  $L + 1$  tap channel impulse response (CIR) with the tap index  $l \in [0, L]$ , the transmit antenna index  $n_t \in [1, N_T]$ , the receive antenna index  $n_r \in [1, N_R]$ , and the training symbol index  $n \in [0, N_s - 1]$ . Note that  $L$  can be interpreted as the delay spread of the channel in units of symbol intervals, and  $L = 0$  corresponds to a frequency-flat or ISI-free channel.

Let  $\mathbf{H} = [\mathbf{H}_0, \mathbf{H}_1, \dots, \mathbf{H}_L]$  be the discrete-time equivalent baseband CIR of the MIMO frequency-selective channel, where  $\mathbf{H}_l$  is the  $l^{\text{th}}$  tap of the MIMO CIRs and defined by the following matrix [35, Eq. (2)]

$$\mathbf{H}_l = \begin{bmatrix} h_{1,1}(l) & \cdots & h_{1,N_T}(l) \\ \vdots & \ddots & \vdots \\ h_{N_R,1}(l) & \cdots & h_{N_R,N_T}(l) \end{bmatrix}, \quad (6)$$

where  $h_{n_r,n_t}(l)$  is the  $l^{\text{th}}$  tap of the CIR between the  $n_r^{\text{th}}$  receive antenna and the  $n_t^{\text{th}}$  transmit antenna, with the combined effect of the transceiver filters and the multipath propagation environment.

In order to decouple channel estimation from data detection, time-division multiplexing (TDM) of training and data symbols is the only approach in single-carrier systems [11, Lemma 1]. For the training-aided block transmission with decoupled training from data, a typical TDM frame structure for a MIMO system is shown in Fig. 2, where the frame structure is the same for all the transmit antennas,  $\mathbf{d}_{n_t}$  and  $\mathbf{s}_{n_t}$  are the data and training symbols transmitted by the  $n_t^{\text{th}}$  transmit antenna, respectively, the “GP”, i.e., the guard period, between the data and training symbols, is used to separate the data and training symbols, and its length is at least  $L$  for perfect separation in multipath. In this paper, we set the length of “GP” to  $L$ .

As stated in [11], the “GP” can be filled with either  $L$  0’s or  $L$  known symbols. In this paper, the cyclic-prefix (CP) with length  $L$  of the training sequence is used as  $L$  known symbols. In what follows, the former is called the zero-padding (ZP) based guard period, and the latter is called the CP based guard period. For both approaches, the system equation corresponding to the training part has the same linear form and is given by

$$\mathbf{Y} = \sqrt{\frac{\rho_s}{N_T}} \mathbf{H} \mathbf{S} + \mathbf{E}, \quad (7)$$

where  $\rho_s$  is the average received SNR at each receive antenna over the training phase,  $\mathbf{S}$  is the training matrix,  $\mathbf{E}$  is the additive noise matrix, and  $\mathbf{Y}$  is the received signal matrix. The only difference between the ZP- and CP-based guard periods is the training matrix  $\mathbf{S}$ . In what follows, the training matrix is presented for both scenarios.

#### A. ZP-Based Guard Period

In this case, “GP” in Fig. 2 is filled with  $L$  0’s. Using matrix notation, the signals received by  $N_R$  antennas, corresponding

to the training symbols transmitted from  $N_T$  antennas, can be written as  $\mathbf{Y}^{\text{ZP}} = \sqrt{\rho_s/N_T} \mathbf{H} \mathbf{S}^{\text{ZP}} + \mathbf{E}^{\text{ZP}}$ , where the training matrix  $\mathbf{S}^{\text{ZP}}$ , whose dimension is  $N_T(L + 1) \times (N_s + L)$ , is given by

$$\mathbf{S}^{\text{ZP}} = \begin{bmatrix} \mathbf{s}(0) & \mathbf{s}(1) & \cdots & \mathbf{s}(N_s-1) & \mathbf{0}_{N_T \times 1} & \cdots & \mathbf{0}_{N_T \times 1} \\ \mathbf{0}_{N_T \times 1} & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \mathbf{0}_{N_T \times 1} \\ \mathbf{0}_{N_T \times 1} & \cdots & \mathbf{0}_{N_T \times 1} & \mathbf{s}(0) & \mathbf{s}(1) & \cdots & \mathbf{s}(N_s-1) \end{bmatrix}, \quad (8)$$

in which we have the following  $N_T \times 1$  vector

$$\mathbf{s}(n) = [s_1(n), s_2(n), \dots, s_{N_T}(n)]'. \quad (9)$$

Note that  $\mathbf{Y}^{\text{ZP}} = [\mathbf{y}(0), \mathbf{y}(1), \dots, \mathbf{y}(N_s + L - 1)]$ , where  $\mathbf{y}(n) = [y_1(n), y_2(n), \dots, y_{N_R}(n)]'$ . The noise matrix is defined as  $\mathbf{E}^{\text{ZP}} = [\mathbf{e}(0), \mathbf{e}(1), \dots, \mathbf{e}(N_s + L - 1)]$ , in which  $\mathbf{e}(n) = [e_1(n), e_2(n), \dots, e_{N_R}(n)]'$ .

#### B. CP-Based Guard Period

In this case, “GP” in Fig. 2 for the  $n_t^{\text{th}}$  transmit antenna is filled with the CP of  $\mathbf{s}_{n_t}$ , i.e.,  $\mathbf{CP}_{n_t} = [s_{n_t}(N_s - L), \dots, s_{n_t}(N_s - 1)]$ . The received training signal on  $N_R$  receive antennas, after discarding those affected by the data due to ISI, can be written as  $\mathbf{Y}^{\text{CP}} = \sqrt{\rho_s/N_T} \mathbf{H} \mathbf{S}^{\text{CP}} + \mathbf{E}^{\text{CP}}$ , where the training matrix  $\mathbf{S}^{\text{CP}}$ , whose dimension is  $N_T(L + 1) \times N_s$ , is given by

$$\mathbf{S}^{\text{CP}} = \begin{bmatrix} \mathbf{s}(0) & \mathbf{s}(1) & \cdots & \mathbf{s}(N_s - 1) \\ \mathbf{s}(N_s - 1) & \mathbf{s}(0) & \ddots & \mathbf{s}(N_s - 2) \\ \vdots & \ddots & \ddots & \vdots \\ \mathbf{s}(N_s - L) & \mathbf{s}(N_s - L + 1) & \cdots & \mathbf{s}(N_s - L - 1) \end{bmatrix}, \quad (10)$$

where  $\mathbf{s}(n)$  is given in (9). Furthermore,  $\mathbf{Y}^{\text{CP}} = [\mathbf{y}(0), \mathbf{y}(1), \dots, \mathbf{y}(N_s - 1)]$  and  $\mathbf{E}^{\text{CP}} = [\mathbf{e}(0), \mathbf{e}(1), \dots, \mathbf{e}(N_s - 1)]$ .

Clearly, for both cases,  $s_{n_t}(n)$  is the training symbol transmitted by the  $n_t^{\text{th}}$  transmit antenna at time  $n$ ,  $y_{n_r}(n)$  is the signal received by the  $n_r^{\text{th}}$  receive antenna at time  $n$ ,  $e_{n_r}(n)$  is the additive noise component in  $y_{n_r}(n)$ .

### IV. THE CRITERION FOR OPTIMAL TRAINING

Both the ZP- and CP-based guard period choices provide the same linear form of system equations given in (7). Therefore, we do not distinguish between them, when developing the criterion for optimal training, and the conclusions of this section apply to both cases.

For channel estimation, we assume the elements of the additive noise matrix  $\mathbf{E}$  are independent with zero mean and unit variance (not necessarily Gaussian). In addition, we treat the channel matrix  $\mathbf{H}$  in two different ways. In the first approach,  $\mathbf{H}$  is an unknown deterministic matrix. In the second setup,  $\mathbf{H}$  is random and independent of the additive noise  $\mathbf{E}$ , elements of  $\mathbf{H}$  are independent with zero mean, and each subchannel  $\mathbf{h}_{n_r,n_t}$  has unit power, i.e.,  $\|\mathbf{h}_{n_r,n_t}\|_F^2 = 1$ , where

$$\mathbf{h}_{n_r,n_t} = [h_{n_r,n_t}(0), h_{n_r,n_t}(1), \dots, h_{n_r,n_t}(L)]. \quad (11)$$

Moreover, the  $l^{\text{th}}$  tap of the subchannel  $\mathbf{h}_{n_r, n_t}$  has the power  $\sigma_{n_r, n_t, l}$ , i.e.,  $\mathbb{E}[|h_{n_r, n_t}(l)|^2] = \sigma_{n_r, n_t, l}$ , and the covariance matrix of  $\mathbf{H}$  is defined as  $\mathbf{C}_h = \mathbb{E}[\mathbf{h}\mathbf{h}^\dagger]$ , where  $\mathbf{h} = \text{vec}(\mathbf{H})$ . Since elements of  $\mathbf{H}$  are independent with zero mean,  $\mathbf{C}_h$  is an  $N_R N_T (L+1) \times N_R N_T (L+1)$  diagonal matrix, whose  $\kappa^{\text{th}}$  diagonal element is given by

$$[\mathbf{C}_h]_{\kappa, \kappa} = \sigma_{n_r, n_t, l}, \quad \kappa \in [0, N_R N_T (L+1) - 1], \quad (12)$$

where  $\kappa$  and  $l$  are related according to  $\kappa = lN_R N_T + (n_t - 1)N_R + n_r - 1$ , for a given  $n_t$  and  $n_r$ .

Following the terminology of [47], the best linear estimators for deterministic and random channel representations are the BLUE and the LMMSE estimator, respectively, which are presented in the following proposition.

*Proposition 1:* For the linear system model in (7), the BLUE (best linear unbiased estimator) and LMMSE (linear minimum mean square error) estimator of  $\mathbf{H}$  are given by<sup>7</sup>

$$\hat{\mathbf{h}} = \sqrt{\frac{N_T}{\rho_s}} \left[ (\mathbf{S}^* \mathbf{S}') \otimes \mathbf{I}_{N_R} + \alpha \frac{N_T}{\rho_s} \mathbf{C}_h^{-1} \right]^{-1} (\mathbf{S}^* \otimes \mathbf{I}_{N_R}) \mathbf{y}, \quad (13)$$

with the following total mean square error (TMSE)

$$\varepsilon_\alpha = \frac{N_T}{\rho_s} \text{tr} \left\{ \left[ (\mathbf{S} \mathbf{S}^\dagger) \otimes \mathbf{I}_{N_R} + \alpha \frac{N_T}{\rho_s} \mathbf{C}_h^{-1} \right]^{-1} \right\}, \quad (14)$$

where  $\alpha = 0$  and  $1$  correspond to the BLUE and LMMSE estimator, respectively, and  $\mathbf{y} = \text{vec}(\mathbf{Y})$ . Furthermore, when  $\mathbf{E}$  is Gaussian in BLUE,  $\varepsilon_0$  is the classical CRLB, whereas  $\mathbf{E}$  and  $\mathbf{H}$  are Gaussian and independent in the LMMSE estimator,  $\varepsilon_1$  is the Bayesian CRLB.

*Proof:* See Appendix I.  $\square$

From (14), we can conclude that the TMSE depends on the training symbol matrix  $\mathbf{S}$ , when the channel estimation approach, the number of transmit and receive antennas, SNR, and the fading covariance matrix  $\mathbf{C}_h$  are fixed. Under the transmit power constraint of training symbols, the minimization of  $\varepsilon_\alpha$  through  $\mathbf{S}$  is presented in the following proposition.

*Proposition 2:* Suppose the transmit power is constrained by  $\frac{1}{N_s(L+1)} \text{tr}[\mathbf{S}\mathbf{S}^\dagger] \leq N_T$ , and the channel statistical information  $\mathbf{C}_h$  is known at the transmitter. For BLUE, the TMSE  $\varepsilon_0$  is minimized *iff* the training sequences satisfy the semi-unitary condition

$$\mathbf{S}\mathbf{S}^\dagger = N_s \mathbf{I}_{N_T(L+1)}. \quad (15)$$

For LMMSE estimator, the minimum of the TMSE  $\varepsilon_1$  is achieved *iff* the training sequences satisfy the following condition

$$\mathbf{S}\mathbf{S}^\dagger = N_s \mathbf{I}_{L+1} \otimes \mathbf{C}_s, \quad (16)$$

$$= \underbrace{(\mathbf{I}_{L+1} \otimes \mathbf{C}_s)}_{\text{Power allocation}} \underbrace{[N_s \mathbf{I}_{N_T(L+1)}]}_{\text{Condition in (15)}}, \quad (17)$$

where  $\mathbf{C}_s = \text{diag}(P_1, P_2, \dots, P_{N_T})$ , and  $P_{n_t} = \frac{1}{N_s} \|\mathbf{s}_{n_t}\|_F^2$ ,  $\forall n_t$ , is the optimal training power emitted from the  $n_t^{\text{th}}$

<sup>7</sup>To obtain a meaningful estimate of  $\mathbf{H}$ , we need at least as many measurements as unknowns [9], which implies  $N_s + L \geq N_T(L+1)$  and  $N_s \geq N_T(L+1)$  for the ZP- and CP-based guard periods, respectively.

transmit antenna, obtained by solving

$$\sum_{n_r=1}^{N_R} \sum_{l=0}^L \frac{\frac{N_s \rho_s}{N_T}}{(\sigma_{n_r, n_t, l}^{-1} + \frac{N_s \rho_s}{N_T} P_{n_t})^2} = \lambda, \quad \forall n_t \in [1, N_T], \quad (18)$$

under the training power constraint  $\sum_{n_t=1}^{N_T} P_{n_t} = N_T$ , where  $\lambda$  is the Lagrangian multiplier. Since  $P_{n_t}$ 's must be non-negative, it may not always be possible to find a solution to satisfy all of  $N_T$  equations in (18), simultaneously. In this case, the Kuhn-Tucker conditions [48] need to be used to solve (18) and obtain the waterfilling-like solutions.

*Proof:* See Appendix II.  $\square$

When  $\mathbf{H}$  is deterministic and  $\mathbf{E}$  is Gaussian, (15) substituted in (14), along with  $\alpha = 0$ , provide the minimum possible classical CRLB  $\varepsilon_0$ . On the other hand, when both  $\mathbf{H}$  and  $\mathbf{E}$  are Gaussian and independent, the combination of (14) and (17), together with  $\alpha = 1$ , give the minimum possible Bayesian CRLB  $\varepsilon_1$ . Note that according to (17), each transmit antenna has to send a different power level, under the total transmit power constraint, to minimize TMSE if the LMMSE estimator is used.

For the ZP-based guard period and from the structure of  $\mathbf{S}^{\text{ZP}}$  in (8), (15) and (17) imply that

- The aperiodic autocorrelation of the training sequence from each transmit antenna is zero within  $L$  tap shifts, except for the zero shift, i.e.,  $\varphi_{\mathbf{s}_{n_t}, \mathbf{s}_{n_t}}(k) = 0, 1 \leq |k| \leq L, \forall n_t$ ,
- The aperiodic crosscorrelation of any two training sequences from two different transmit antennas is zero within  $L$  tap shifts, i.e.,  $\varphi_{\mathbf{s}_{n_t}, \mathbf{s}_{\tilde{n}_t}}(k) = 0, |k| \leq L, \forall n_t \neq \tilde{n}_t$ .

The same observations apply to the CP-based guard period case, except for replacing aperiodic auto- and cross-correlations with periodic ones. Equations similar to (15) are also derived in [9] and [11] for MIMO flat and MIMO ISI channels, respectively, by maximizing the ergodic capacity lower bound.

As shown at the end of Appendix II-A.1, if  $\sigma_{n_r, n_t, l}$  is independent of  $n_t$ , the optimal training power allocation is given by  $P_{n_t} = 1, \forall n_t$ , and the orthogonal condition in (17) reduces to the semi-unitary condition in (15). In what follows, without loss of generality, we consider the special but widely used case where  $\sigma_{n_r, n_t, l}$  is independent of  $n_r$  and  $n_t$ , i.e.,  $\sigma_{n_r, n_t, l} = \sigma_l, \forall n_r, n_t$ , for simplicity. Hence, the covariance matrix of the channel is given by  $\mathbf{C}_h = \mathbf{C}_\Sigma \otimes \mathbf{I}_{N_T N_R}$ , where  $\mathbf{C}_\Sigma = \text{diag}(\sigma_0, \sigma_1, \dots, \sigma_L)$  is the covariance matrix of each subchannel.

## V. CONSTRUCTION OF OPTIMAL TRAINING SEQUENCES

Since it is difficult to find training matrices, with the structures given in (8) and (10), to satisfy the semi-unitary condition [20, p. 179]<sup>8</sup>, i.e.,  $\mathbf{S}\mathbf{S}^\dagger = N_s \mathbf{I}_{N_T(L+1)}$ , we came up with the idea of reformulating  $\mathbf{S}$  into two parts, i.e.,  $\mathbf{S} = [\mathbf{S}_1, \mathbf{S}_2]$ ,

<sup>8</sup>As stated in the introduction, there are some sequences such as ZCZ, Frank and Chu sequences, which satisfy the condition in the case of CP-based guard period. But for the ZP-based guard period, there is no report on such sequences, except for the delta impulse sequences.

where  $\mathbf{S}_1$  and  $\mathbf{S}_2$  are two individual training matrices, which have the same block-circulant structure as (8) and (10), for the ZP- and CP-based cases, respectively. Certainly, it is still hard to find individual  $\mathbf{S}_1$  and  $\mathbf{S}_2$  such that  $\mathbf{S}_1 \mathbf{S}_1^\dagger \propto \mathbf{I}_{N_T(L+1)}$  and  $\mathbf{S}_2 \mathbf{S}_2^\dagger \propto \mathbf{I}_{N_T(L+1)}$ , due to the above argument. However, by using uncorrelated Golay complementary sets of *polyphase* sequences of Sec. II, it is possible to design  $\mathbf{S}_1$  and  $\mathbf{S}_2$  such that

$$\mathbf{S}_1 \mathbf{S}_1^\dagger + \mathbf{S}_2 \mathbf{S}_2^\dagger = N_s \mathbf{I}_{N_T(L+1)}, \quad (19)$$

which satisfies (15), hence achieves the minimum possible estimation error according to Proposition 2.

Based on the discussion in Sec. II, we know the family size of mutually uncorrelated Golay complementary sets is two if  $p = 2$ , which implies that, without any other effort such as zero insertion or cyclic phase shift, it can support two transmit antennas only. For  $N_T > 2$ , more training sequences need to be constructed. In the following two subsections, we will design the training sequences for ZP- and CP-based guard period transmission. Specifically, Subsection V-A inserts zeros to the basic Golay complementary pair  $\{\mathbf{a}_0, \mathbf{a}_1\}$  to generate training sequences, and Subsection V-B shifts the phase of the basic Golay complementary pair  $\{\mathbf{a}_0, \mathbf{a}_1\}$  to construct training sequences.

#### A. ZP-Based Guard Period

Let  $\{\mathbf{a}_0, \mathbf{a}_1\}$  be a pair of Golay complementary *polyphase* sequences generated by (4), each of length  $N$ . Note that the complementary property will not change if we append some 0's to the beginning or end or both sides of each sequence. We define four row vector of length  $\check{N}^9$  as

$$\mathbf{u} = \sqrt{\check{N}/N} [\mathbf{a}_0, \mathbf{o}], \quad \mathbf{v} = \sqrt{\check{N}/N} [\mathbf{a}_1, \mathbf{o}], \quad (20)$$

$$\check{\mathbf{u}} = \sqrt{\check{N}/N} [\check{\mathbf{a}}_0^*, \mathbf{o}], \quad \check{\mathbf{v}} = \sqrt{\check{N}/N} [\check{\mathbf{a}}_1^*, \mathbf{o}], \quad (21)$$

where  $\check{N} = N + \ell$  is the sequence length after the insertion of  $\ell$  0's into the original one,  $\ell = (\check{N}_T/2 - 1)(L + 1)$ ,  $\check{N}_T = 2 \lceil N_T/2 \rceil^{10}$  and  $\mathbf{o} = \mathbf{0}_{1 \times \ell}$ . Note that, in order to keep the expected received SNR  $\rho_s$  at each receive antenna unchanged over the training phase, the factor  $\sqrt{\check{N}/N}$  is added to (20) and (21), for the power compensation due to the insertion of  $\ell$  0's into the original Golay complementary *polyphase* sequences. Based on the properties discussed in Sec. II,  $\{\mathbf{u}, \check{\mathbf{v}}\}$  and  $\{\mathbf{v}, -\check{\mathbf{u}}\}$  are mutually uncorrelated.

#### B. CP-Based Guard Period

From Fig. 1, it is clear to see that the PACF  $\tilde{\varphi}_{\mathbf{a}_0, \mathbf{a}_0}(k) + \tilde{\varphi}_{\check{\mathbf{a}}_1^*, \check{\mathbf{a}}_1^*}(k)$  has an impulse-like shape within all shifts, and the PCCF  $\tilde{\varphi}_{\mathbf{a}_0, \mathbf{a}_1}(k) + \tilde{\varphi}_{\check{\mathbf{a}}_1^*, -\check{\mathbf{a}}_0^*}(k)$  is zero everywhere, which verify (3). Therefore, we can use different phases of uncorrelated Golay complementary sets of *polyphase* sequences as the training symbols for different transmit antennas. For this

<sup>9</sup>Since we use two training sequences for one frame of each transmit antenna, the condition for a meaningful estimation of  $\mathbf{H}$  should be changed to  $\check{N} + L \geq \lceil N_T/2 \rceil (L + 1)$  for the ZP-based guard period, and  $\check{N} \geq \lceil N_T/2 \rceil (L + 1)$  for the CP-based guard period.

<sup>10</sup>For example, when  $N_T = 4$ , or 3, we have  $\check{N}_T = 4$ .

TABLE I  
ASSIGNMENT OF TRAINING SYMBOLS TO TRANSMIT ANTENNAS FOR BOTH ZP- AND CP-BASED GUARD PERIODS

$T_x$	$\mathbf{s}_{n_t,1}$	$\mathbf{s}_{n_t,2}$
1	$\mathbf{u}$	$\check{\mathbf{v}}$
2	$\mathbf{v}$	$-\check{\mathbf{u}}$
3	$\mathbf{u} \Pi_{\check{N}}^{L+1}$	$\check{\mathbf{v}} \Pi_{\check{N}}^{L+1}$
4	$\mathbf{v} \Pi_{\check{N}}^{L+1}$	$-\check{\mathbf{u}} \Pi_{\check{N}}^{L+1}$
$\vdots$	$\vdots$	$\vdots$
$\check{N}_T - 1$	$\mathbf{u} \Pi_{\check{N}}^\ell$	$\check{\mathbf{v}} \Pi_{\check{N}}^\ell$
$\check{N}_T$	$\mathbf{v} \Pi_{\check{N}}^\ell$	$-\check{\mathbf{u}} \Pi_{\check{N}}^\ell$

case, we do not need to insert 0's into the original Golay complementary pair, so,  $\check{N} = N$ , and

$$\mathbf{u} = \mathbf{a}_0, \quad \mathbf{v} = \mathbf{a}_1, \quad (22)$$

$$\check{\mathbf{u}} = \check{\mathbf{a}}_0^*, \quad \check{\mathbf{v}} = \check{\mathbf{a}}_1^*. \quad (23)$$

For concise notations, we use the same letters in Table I for both ZP and CP cases, where the training symbol assignment for all the transmit antennas are described. In the ZP case,  $\mathbf{u}$ ,  $\mathbf{v}$ ,  $\check{\mathbf{u}}$  and  $\check{\mathbf{v}}$  are given by (20) and (21), and  $\check{N} = N + \ell$ , whereas in the CP case, (22) and (23) give the definitions for  $\mathbf{u}$ ,  $\mathbf{v}$ ,  $\check{\mathbf{u}}$  and  $\check{\mathbf{v}}$ , and  $\check{N} = N$ . As a simple example, for  $N_T = 4$ ,  $N = 4$ , and  $L = 1$  (two taps in each subchannel), Table I is reproduced in Tables II and III, for ZP and CP cases, respectively, using  $\mathbf{a}_0 = [+++-]$  and  $\mathbf{a}_1 = [++-+]$ , which can be obtained from (4) by  $d_1 = 1$ ,  $d_2 = 2$ , and  $w_1 = w_2 = 1$ . Based on the training symbols in Tables II and III,  $\mathbf{S}_\nu^{\text{ZP}}$  and  $\mathbf{S}_\nu^{\text{CP}}$ ,  $\nu \in [1, 2]$ , can be generated according to (8) and (10), respectively. Therefore  $\mathbf{S} = [\mathbf{S}_1, \mathbf{S}_2]$  for ZP and CP cases are respectively given by

TABLE II  
TRAINING EXAMPLE FOR THE ZP CASE, WHERE  $N_T = 4$ ,  $N = 4$ ,  $L = 1$ , AND  $\check{N} = N + \ell = 6$ .

$T_x$	$\mathbf{s}_{n_t,1}$	$\mathbf{s}_{n_t,2}$
1	$\sqrt{1.5}[+++-00]$	$\sqrt{1.5}[+-++00]$
2	$\sqrt{1.5}[+-+000]$	$\sqrt{1.5}[+---00]$
3	$\sqrt{1.5}[00++++]$	$\sqrt{1.5}[00++-+]$
4	$\sqrt{1.5}[00++-+]$	$\sqrt{1.5}[00+- - -]$

TABLE III  
TRAINING EXAMPLE FOR THE CP CASE, WHERE  $N_T = 4$ ,  $N = 4$ ,  $L = 1$ , AND  $\check{N} = N = 4$ .

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



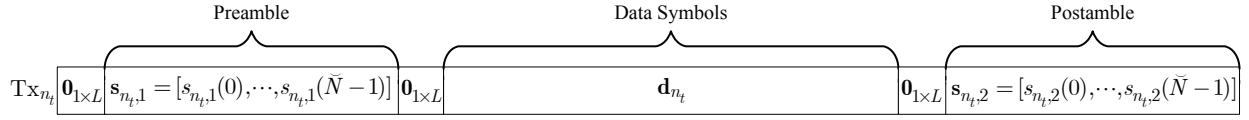


Fig. 3. The frame structure for the ZP-based guard period.

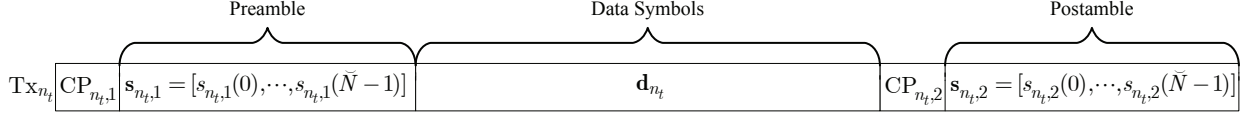


Fig. 4. The frame structure for the CP-based guard period.

$$\mathbf{S}^{\text{ZP}} = \sqrt{1.5} \begin{bmatrix} \begin{matrix} + & + & + & - & 0 & 0 & 0 \\ + & + & - & + & 0 & 0 & 0 \\ 0 & 0 & + & + & + & - & 0 \\ 0 & 0 & + & - & + & - & 0 \\ 0 & + & + & + & - & 0 & 0 \\ 0 & + & - & + & + & 0 & 0 \\ 0 & 0 & 0 & + & + & + & - \\ 0 & 0 & 0 & + & - & - & + \end{matrix} & \begin{matrix} + & - & + & + & 0 & 0 & 0 \\ + & - & - & - & 0 & 0 & 0 \\ 0 & 0 & + & - & + & + & 0 \\ 0 & 0 & + & - & - & - & 0 \\ 0 & + & - & + & + & 0 & 0 \\ 0 & + & - & - & - & 0 & 0 \\ 0 & 0 & 0 & + & - & + & + \\ 0 & 0 & 0 & + & - & - & - \end{matrix} \end{bmatrix}, \quad (24)$$

$\underbrace{\hspace{10em}}_{\mathbf{S}_1^{\text{ZP}}} \quad \underbrace{\hspace{10em}}_{\mathbf{S}_2^{\text{ZP}}}$

and

$$\mathbf{S}^{\text{CP}} = \begin{bmatrix} \begin{matrix} + & + & + & - \\ + & + & - & + \\ + & - & + & + \\ - & + & + & + \\ - & + & + & + \\ + & + & + & - \\ + & + & - & + \\ + & - & + & + \end{matrix} & \begin{matrix} + & - & + & + \\ + & - & - & - \\ + & + & + & - \\ - & - & + & - \\ + & + & - & + \\ - & + & - & - \\ - & + & + & + \\ - & - & - & + \end{matrix} \end{bmatrix}. \quad (25)$$

$\underbrace{\hspace{10em}}_{\mathbf{S}_1^{\text{CP}}} \quad \underbrace{\hspace{10em}}_{\mathbf{S}_2^{\text{CP}}}$

From (24) and (25), it is easy to check that  $\mathbf{S}^{\text{ZP}} (\mathbf{S}^{\text{ZP}})^T = 2\tilde{N}\mathbf{I}_{N_T(L+1)} = 12\mathbf{I}_8$  and  $\mathbf{S}^{\text{CP}} (\mathbf{S}^{\text{CP}})^T = 2\tilde{N}\mathbf{I}_{N_T(L+1)} = 8\mathbf{I}_8$ .

Based on the above discussion, we propose the frame structures shown in Figs. 3 and 4 for the ZP- and CP-based guard periods, respectively, where  $s_{n_t,1}$  and  $s_{n_t,2}$ ,  $n_t \in [1, N_T]$ , are given in Table I, and  $\text{CP}_{n_t,1} = [s_{n_t,1}(\tilde{N}-L), \dots, s_{n_t,1}(\tilde{N}-1)]$  and  $\text{CP}_{n_t,2} = [s_{n_t,2}(\tilde{N}-L), \dots, s_{n_t,2}(\tilde{N}-1)]$  are the cyclic prefixes of  $s_{n_t,1}$  and  $s_{n_t,2}$ , respectively.

Note that for both ZP and CP cases and starting from a given Golay complementary pair  $\{\mathbf{a}_0, \mathbf{a}_1\}$ , the training assignment is not unique. We have only shown one possible assignment in Table I, to save space. In addition, when the number of transmit antennas  $N_T$  is odd, the last row of Table I will not be used.

## VI. CHANNEL ESTIMATION ALGORITHM

It is easy to check that the training sequences given in Table I satisfy (19) with  $N_s = 2\tilde{N}$ . Hence the condition in (15) with  $\mathbf{S} = [\mathbf{S}_1, \mathbf{S}_2]$  is satisfied, which demonstrates the optimality of the training symbols given in Table I for the ZP- and CP-based guard periods, respectively.

For the proposed scheme that each transmit antenna uses two training sequences<sup>11</sup>, i.e.,  $p = 2$ , per frame, as shown in Fig. 3 and 4, the system model (7) can be rewritten as  $[\mathbf{Y}_1, \mathbf{Y}_2] = \sqrt{\rho_s/N_T} \mathbf{H}[\mathbf{S}_1, \mathbf{S}_2] + [\mathbf{E}_1, \mathbf{E}_2]$ . By replacing  $\mathbf{y}$ ,  $\mathbf{S}$ ,  $\mathbf{C}_h$  and  $\alpha$  with  $\text{vec}([\mathbf{Y}_1, \mathbf{Y}_2])$ ,  $[\mathbf{S}_1, \mathbf{S}_2]$ ,  $\mathbf{C}_\Sigma \otimes \mathbf{I}_{N_T N_R}$  and 0, respectively, into (13), with  $\hat{\mathbf{h}}^{\text{BLUE}} = \text{vec}(\hat{\mathbf{H}}^{\text{BLUE}})$ , we obtain

$$\hat{\mathbf{H}}^{\text{BLUE}} = \frac{\sqrt{N_T}}{2\tilde{N}\sqrt{\rho_s}} \sum_{\nu=1}^2 \mathbf{Y}_\nu \mathbf{S}_\nu^\dagger, \quad (26)$$

with  $\hat{\mathbf{h}}^{\text{LMMSE}} = \text{vec}(\hat{\mathbf{H}}^{\text{LMMSE}})$ , one can also verify that

$$\hat{\mathbf{H}}^{\text{LMMSE}} = \sqrt{\frac{N_T}{\rho_s}} \left( \sum_{\nu=1}^2 \mathbf{Y}_\nu \mathbf{S}_\nu^\dagger \right) \times \left( \sum_{\nu=1}^2 \mathbf{S}_\nu \mathbf{S}_\nu^\dagger + \frac{N_T}{\rho_s} \mathbf{C}_\Sigma^{-1} \otimes \mathbf{I}_{N_T} \right)^{-1}, \quad (27)$$

$$= \hat{\mathbf{H}}^{\text{BLUE}} \left[ \left( \mathbf{I}_{L+1} + \frac{N_T}{2\tilde{N}\rho_s} \mathbf{C}_\Sigma^{-1} \right) \otimes \mathbf{I}_{N_T} \right]^{-1}, \quad (28)$$

where the last line comes from (19) with  $N_s = 2\tilde{N}$ . For the  $l^{\text{th}}$  tap of the  $n_r$ - $n_t$  subchannel, the BLUE and LMMSE estimator are given by

$$\hat{h}_{n_r, n_t}(l) = \frac{\sigma_l \sqrt{\rho_s N_T}}{2\tilde{N}\rho_s \sigma_l + \alpha N_T} \left[ \sum_{\nu=1}^2 \mathbf{Y}_\nu \mathbf{S}_\nu^\dagger \right]_{n_r-1, lN_T+n_t-1}, \quad (29)$$

where  $n_r \in [1, N_R]$ ,  $n_t \in [1, N_T]$ , and  $l \in [0, L]$ . With (19) and  $N_s = 2\tilde{N}$ , the TMSE for both estimators can also be shown to be

$$\varepsilon_\alpha = \frac{N_R N_T}{\rho_s} \text{tr} \left[ \left( \sum_{\nu=1}^2 \mathbf{S}_\nu \mathbf{S}_\nu^\dagger + \alpha \frac{N_T}{\rho_s} \mathbf{C}_\Sigma^{-1} \otimes \mathbf{I}_{N_T} \right)^{-1} \right], \quad (30)$$

$$= \sum_{l=0}^L \frac{N_R N_T^2 \sigma_l}{2\tilde{N}\rho_s \sigma_l + \alpha N_T}. \quad (31)$$

## VII. FAST SOFTWARE AND HARDWARE IMPLEMENTATION OF THE CHANNEL ESTIMATOR

In this section, we develop low-complexity software and hardware implementations for the channel estimator in (29),

<sup>11</sup>For  $p > 2$ , the channel estimation algorithm is discussed in Appendix III.

via DSP and ASIC/FPGA, respectively. Due to the different structures of training matrices in (8) and (10), we treat the ZP- and CP-based guard periods separately, in the following subsections. For convenience, we define the new scaled SNR  $\gamma$  as

$$\gamma = \frac{\check{N}}{N} \rho_s. \quad (32)$$

Note that  $\check{N} = N + \ell$  for the ZP-based guard period, and  $\check{N} = N$  for the CP-based guard period.

#### A. Software (DSP) Implementation

Nowadays the DSP has enough power to implement complex algorithms, and are widely used in both base stations and mobile terminals. In this subsection, we derive the fast DSP implementation of (29).

1) *ZP-Based Guard Period*: For demonstrating the idea only, we assume the training sequences generated by (4) are real and binary, i.e.,  $w_m = \pm 1, \forall m$ , and the number of transmit antennas  $N_T$  is even. In a similar way to [35], let us define

$$\mathbf{H}_{n_r}^t = \begin{bmatrix} h_{n_r, 2t-1}(0) & h_{n_r, 2t-1}(1) & \cdots & h_{n_r, 2t-1}(L) \\ h_{n_r, 2t}(0) & h_{n_r, 2t}(1) & \cdots & h_{n_r, 2t}(L) \end{bmatrix}, \quad (33)$$

where  $t = \lceil \frac{n_r}{2} \rceil$ , and

$$\mathbf{H}_{n_r} = \begin{bmatrix} \mathbf{H}_{n_r}^1 & \mathbf{H}_{n_r}^2 & \cdots & \mathbf{H}_{n_r}^{\frac{N_T}{2}} \end{bmatrix}, \quad (34)$$

whose dimension is  $2 \times \frac{N_T}{2} (L+1)$ . Based on (29), (33) and (34), estimates for all the channel coefficients that correspond to the  $n_r^{\text{th}}$  receive antenna, can be compactly written as<sup>12</sup>

$$\hat{\mathbf{H}}_{n_r} = \left( \sum_{\nu=1}^2 \mathbf{X}_\nu \mathbf{Y}_{n_r, \nu} \right) \left( \mathbf{I}_{\frac{N_T}{2}} \otimes \mathbf{\Lambda} \right), \quad (35)$$

where  $\mathbf{X}_1 = \begin{bmatrix} \mathbf{a}_0 \\ \mathbf{a}_1 \end{bmatrix}$ ,  $\mathbf{X}_2 = \begin{bmatrix} \check{\mathbf{a}}_1 \\ -\check{\mathbf{a}}_0 \end{bmatrix}$ ,  $\mathbf{\Lambda} = \text{diag}(\mathbf{f})$ , and  $\mathbf{f}$  is a row vector of length  $N_T$ , defined by

$$\mathbf{f} = \left[ \frac{\sigma_0 \sqrt{\gamma N_T}}{2\gamma N \sigma_0 + \alpha N_T}, \frac{\sigma_1 \sqrt{\gamma N_T}}{2\gamma N \sigma_1 + \alpha N_T}, \cdots, \frac{\sigma_L \sqrt{\gamma N_T}}{2\gamma N \sigma_L + \alpha N_T} \right]. \quad (36)$$

Furthermore,  $\mathbf{Y}_{n_r, \nu}$  is a  $N \times \frac{N_T}{2} (L+1)$  Hankel matrix with the  $(i, j)^{\text{th}}$  element,  $[\mathbf{Y}_{n_r, \nu}]_{i, j}$ , given by  $y_{n_r, \nu}(i+j)$ , i.e.

$$\mathbf{Y}_{n_r, \nu} = \begin{bmatrix} y_{n_r, \nu}(0) & y_{n_r, \nu}(1) & \cdots & y_{n_r, \nu}(\check{N}+L-N) \\ y_{n_r, \nu}(1) & \ddots & \ddots & y_{n_r, \nu}(\check{N}+L-N+1) \\ \vdots & \ddots & \ddots & \vdots \\ y_{n_r, \nu}(N-1) & y_{n_r, \nu}(N) & \cdots & y_{n_r, \nu}(\check{N}+L-1) \end{bmatrix}, \quad (37)$$

where  $\nu \in [1, 2]$ ,  $y_{n_r, \nu}(n)$  is the signal received by the  $n_r^{\text{th}}$  receive antenna at time  $n$ , corresponding to the  $\nu^{\text{th}}$  training sequence. According to (35), clearly the structure of the estimator is identical for all the receive antennas, so we focus on the  $n_r^{\text{th}}$  antenna in the sequel.

First, we consider how to compute  $\mathbf{X}_1 \mathbf{Y}_{n_r, 1}$  in an efficient way. Without loss of the generality, we assume  $w_m = 1$  and

<sup>12</sup>Since we assume  $\mathbf{a}_0$  and  $\mathbf{a}_1$  are binary and real, the conjugate operation on  $\mathbf{X}_\nu$ ,  $\nu \in [1, 2]$ , is not necessary.

$d_m = 2^{m-1}$ ,  $m \in [1, M]$ , in (4) of Sec. II, therefore we have  $N = 2^M$  according to (4). In addition, we split  $\mathbf{Y}_{n_r, 1}$  into two parts as  $\mathbf{Y}_{n_r, 1} = \begin{bmatrix} \mathbf{P}^{(M)} \\ \mathbf{Q}^{(M)} \end{bmatrix}$ , where  $\mathbf{P}^{(M)}$  contains the first  $2^{M-1}$  rows of  $\mathbf{Y}_{n_r, 1}$ , and  $\mathbf{Q}^{(M)}$  contains the last  $2^{M-1}$  rows of  $\mathbf{Y}_{n_r, 1}$ . We further split  $\mathbf{P}^{(M)}$  and  $\mathbf{Q}^{(M)}$  as  $\mathbf{P}^{(M)} = \begin{bmatrix} \mathbf{P}_1^{(M)} \\ \mathbf{P}_2^{(M)} \end{bmatrix}$  and  $\mathbf{Q}^{(M)} = \begin{bmatrix} \mathbf{Q}_1^{(M)} \\ \mathbf{Q}_2^{(M)} \end{bmatrix}$ , respectively, where  $\mathbf{P}_1^{(M)}$  and  $\mathbf{P}_2^{(M)}$  contain the first and last  $2^{M-2}$  rows of  $\mathbf{P}^{(M)}$ , respectively, and  $\mathbf{Q}_1^{(M)}$  and  $\mathbf{Q}_2^{(M)}$  contain the first and last  $2^{M-2}$  rows of  $\mathbf{Q}^{(M)}$ , respectively. Since  $\mathbf{a}_0 = \mathbf{a}_0^{(M)}$  and  $\mathbf{a}_1 = \mathbf{a}_1^{(M)}$ , as defined right after (5), then  $\mathbf{X}_1 \mathbf{Y}_{n_r, 1}$  in (35) can be written as

$$\begin{aligned} \mathbf{X}_1 \mathbf{Y}_{n_r, 1} &= \begin{bmatrix} \mathbf{a}_0^{(M)} \\ \mathbf{a}_1^{(M)} \end{bmatrix} \begin{bmatrix} \mathbf{P}^{(M)} \\ \mathbf{Q}^{(M)} \end{bmatrix}, \\ &= \begin{bmatrix} \mathbf{a}_0^{(M-1)} & \mathbf{a}_1^{(M-1)} \\ \mathbf{a}_0^{(M-1)} & -\mathbf{a}_1^{(M-1)} \end{bmatrix} \begin{bmatrix} \mathbf{P}^{(M)} \\ \mathbf{Q}^{(M)} \end{bmatrix}, \\ &= \begin{bmatrix} \mathbf{a}_0^{(M-2)} & \mathbf{a}_1^{(M-2)} & \mathbf{a}_0^{(M-2)} & -\mathbf{a}_1^{(M-2)} \\ \mathbf{a}_0^{(M-2)} & \mathbf{a}_1^{(M-2)} & -\mathbf{a}_0^{(M-2)} & \mathbf{a}_1^{(M-2)} \end{bmatrix} \begin{bmatrix} \mathbf{P}_1^{(M)} \\ \mathbf{P}_2^{(M)} \\ \mathbf{Q}_1^{(M)} \\ \mathbf{Q}_2^{(M)} \end{bmatrix}, \\ &= \begin{bmatrix} \mathbf{a}_0^{(M-2)} \mathbf{P}^{(M-1)+} + \mathbf{a}_1^{(M-2)} \mathbf{Q}^{(M-1)-} \\ \mathbf{a}_0^{(M-2)} \mathbf{P}^{(M-1)-} + \mathbf{a}_1^{(M-2)} \mathbf{Q}^{(M-1)+} \end{bmatrix}, \end{aligned} \quad (38)$$

where  $\mathbf{a}_0^{(m)} = \begin{bmatrix} \mathbf{a}_0^{(m-1)} & \mathbf{a}_1^{(m-1)} \end{bmatrix}$  and  $\mathbf{a}_1^{(m)} = \begin{bmatrix} \mathbf{a}_0^{(m-1)} & -\mathbf{a}_1^{(m-1)} \end{bmatrix}$ ,  $m = M-1, M$ , derived from (4), are used in the second and third lines of (38). Furthermore, by definition we have

$$\begin{aligned} \mathbf{P}^{(M-1)+} &= \mathbf{P}_1^{(M)} + \mathbf{Q}_1^{(M)}, \\ \mathbf{P}^{(M-1)-} &= \mathbf{P}_1^{(M)} - \mathbf{Q}_1^{(M)}, \\ \mathbf{Q}^{(M-1)+} &= \mathbf{P}_2^{(M)} + \mathbf{Q}_2^{(M)}, \\ \mathbf{Q}^{(M-1)-} &= \mathbf{P}_2^{(M)} - \mathbf{Q}_2^{(M)}. \end{aligned} \quad (39)$$

To show the computational efficiency of the last term in (38) for calculating  $\mathbf{X}_1 \mathbf{Y}_{n_r, 1}$ , first we focus on (39). The number of additions (we take each subtraction as an addition) in (39) is  $2(\check{N} + L - 2^{M-1}) = N + 2(\check{L} - 1)$ , where  $\check{L} = \frac{N_T}{2} (L+1)$ . Furthermore,  $\mathbf{a}_0^{(M-2)} \mathbf{P}^{(M-1)+} + \mathbf{a}_1^{(M-2)} \mathbf{Q}^{(M-1)-}$  and  $\mathbf{a}_0^{(M-2)} \mathbf{P}^{(M-1)-} + \mathbf{a}_1^{(M-2)} \mathbf{Q}^{(M-1)+}$  in the last line of (38) can be calculated in a recursive way [34], with the number of additions equal to  $2 \left[ 2 \sum_{m=0}^{M-3} (2^m + \check{L} - 1) + \check{L} \right]$ .

Therefore, the total number of additions for  $\mathbf{X}_1 \mathbf{Y}_{n_r, 1}$  in (38) is  $(4 \log_2 N - 5) \check{L} + 2(N - 2 \log_2 N + 1)$ . Note that since the elements of vectors  $\mathbf{a}_0^{(M)}$  and  $\mathbf{a}_1^{(M)}$  are all  $\pm 1$ 's, there is no multiplication in (38).  $\mathbf{X}_2 \mathbf{Y}_{n_r, 2}$  will take the same number of operations as  $\mathbf{X}_1 \mathbf{Y}_{n_r, 1}$ , so the total number of additions for  $\sum_{\nu=1}^2 \mathbf{X}_\nu \mathbf{Y}_{n_r, \nu}$  in (35) is  $2(4 \log_2 N - 5) \check{L} + 4(N - 2 \log_2 N + 1) + 2 \check{L} = 8(\log_2 N - 1) \check{L} + 4(N - 2 \log_2 N + 1)$ . However, if we calculate  $\sum_{\nu=1}^2 \mathbf{X}_\nu \mathbf{Y}_{n_r, \nu}$  in (35) directly, the number of additions is  $2 \check{L} (2N - 1)$ , which could be much bigger. For example, if  $N = 128$ ,  $L = 31$ , and  $N_T = 4$ , then the total



number of additions by the fast recursive method proposed in (38) is 3532, but 32640 for the direct computation of (35).

2) *CP-Based Guard Period*: Here  $N_T$  can be even or odd, and sequences can be real or complex. The basic idea is to use the fact that a matrix-vector product can be efficiently implemented by the fast Fourier transform (FFT), if the matrix is circulant.

In a similar manner to [36], 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 transmit antennas and the  $n_r^{\text{th}}$  receive antenna, respectively, as follows

$$\begin{aligned} \mathbf{h}_{n_r,o} &= [\mathbf{h}_{n_r,1} \ \mathbf{h}_{n_r,3} \ \cdots \ \mathbf{h}_{n_r,2\lceil N_T/2 \rceil-1}], \\ \mathbf{h}_{n_r,e} &= [\mathbf{h}_{n_r,2} \ \mathbf{h}_{n_r,4} \ \cdots \ \mathbf{h}_{n_r,2\lfloor N_T/2 \rfloor}], \end{aligned} \quad (40)$$

where  $\mathbf{h}_{n_r,n_t}$  is defined in (11). Moreover, we define four circulant matrices

$$\begin{aligned} \mathbf{C}_{0,1} &= \text{circ}(\mathbf{a}_0, N_o), \quad \mathbf{C}_{1,1} = \text{circ}(\mathbf{a}_1, N_e), \\ \mathbf{C}_{0,2} &= \text{circ}(\mathbf{a}_1^*, N_o), \quad \mathbf{C}_{1,2} = \text{circ}(-\mathbf{a}_0^*, N_e), \end{aligned} \quad (41)$$

with  $\text{circ}(\mathbf{c})$ ,  $\mathbf{c} = [c_0, c_1, \dots, c_{n-1}]$ , as an  $n \times n$  circulant matrix, whose  $(j, k)^{\text{th}}$  element<sup>13</sup> is  $c_{(k-j)_n}$ , and  $\text{circ}(\mathbf{c}, m)$ ,  $m \leq n$ , includes the first  $m$  rows of  $\text{circ}(\mathbf{c})$ . Based on Table I, (29), (40) and (41), estimators for the channel coefficients specified in (40) can be written as

$$\begin{aligned} \hat{\mathbf{h}}_{n_r,o} &= \left( \sum_{\nu=1}^2 \mathbf{y}_{n_r,\nu} \mathbf{C}_{0,\nu}^\dagger \right) (\mathbf{I}_{\lceil N_T/2 \rceil} \otimes \mathbf{\Lambda}), \\ \hat{\mathbf{h}}_{n_r,e} &= \left( \sum_{\nu=1}^2 \mathbf{y}_{n_r,\nu} \mathbf{C}_{1,\nu}^\dagger \right) (\mathbf{I}_{\lfloor N_T/2 \rfloor} \otimes \mathbf{\Lambda}), \end{aligned} \quad (42)$$

where  $\mathbf{y}_{n_r,\nu} = [y_{n_r,\nu}(0), y_{n_r,\nu}(1), \dots, y_{n_r,\nu}(N-1)]$ ,  $\nu \in [1, 2]$ , and  $\mathbf{\Lambda} = \text{diag}(\mathbf{f})$ , with  $\mathbf{f}$  given in (36). For fast implementation of the vector-matrix products in (42), we need the following result.

*Proposition 3:* [40, Chap. 3] If  $\mathbf{C}$  is an  $N \times N$  circulant matrix, then it can be diagonalized by the Fourier matrix  $\mathbf{F}$  of order  $N$  such that  $[\mathbf{F}^\dagger]_{i,j} = \frac{1}{\sqrt{N}} \omega^{(i-1)(j-1)}$  [40, (2.5.3)],  $\omega = \exp(j\frac{2\pi}{N})$ , and  $j = \sqrt{-1}$ . Therefore

$$\mathbf{C} = \mathbf{F}^\dagger \Delta_{\mathbf{c}} \mathbf{F}, \quad (43)$$

where  $\Delta_{\mathbf{c}} = \sqrt{N} \text{diag}(\mathbf{c} \mathbf{F}^\dagger)$  and  $\mathbf{c}$  is the first row of  $\mathbf{C}$  such that  $\mathbf{C} = \text{circ}(\mathbf{c})$ .

Note that  $\mathbf{y}_{n_r,1} \mathbf{C}_{0,1}^\dagger$  is the first  $N_o$  elements of  $\mathbf{y}_{n_r,1} [\text{circ}(\mathbf{a}_0)]^\dagger$ . Therefore, it can be calculated very efficiently, according to Proposition 3, as follows

$$\begin{aligned} \mathbf{y}_{n_r,1} [\text{circ}(\mathbf{a}_0)]^\dagger &= \mathbf{y}_{n_r,1} (\mathbf{F}^\dagger \Delta_{\mathbf{a}_0} \mathbf{F})^\dagger, \\ &= [(\mathbf{y}_{n_r,1} \mathbf{F}^\dagger) \Delta_{\mathbf{a}_0}^\dagger] \mathbf{F}, \\ &= \sqrt{N} \text{FFT}(\text{FFT}(\mathbf{a}_0^*) \odot \text{IFFT}(\mathbf{y}_{n_r,1})), \end{aligned} \quad (44)$$

<sup>13</sup>Note that  $(\cdot)_n$  is the modulus operator, as defined in Sec. I. For example, if  $k = 1$ ,  $j = 3$  and  $n = 8$ , then  $c_{(k-j)_n} = c_6$  denotes the 7<sup>th</sup> element of the vector  $\mathbf{c}$ .

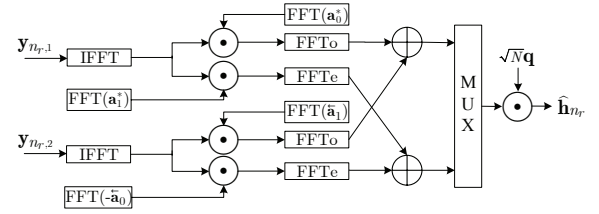


Fig. 5. The FFT-based estimator on the  $n_r^{\text{th}}$  receive antenna for the CP-based guard period.

where  $\text{FFT}(\mathbf{x}) = \mathbf{x} \mathbf{F}$  and  $\text{IFFT}(\mathbf{x}) = \mathbf{x} \mathbf{F}^\dagger$ , with  $\mathbf{x}$  as a  $1 \times N$  vector. With the same reasoning we obtain

$$\mathbf{y}_{n_r,1} [\text{circ}(\mathbf{a}_1)]^\dagger = \sqrt{N} \text{FFT}(\text{FFT}(\mathbf{a}_1^*) \odot \text{IFFT}(\mathbf{y}_{n_r,1})), \quad (45)$$

$$\mathbf{y}_{n_r,2} [\text{circ}(\mathbf{a}_1^*)]^\dagger = \sqrt{N} \text{FFT}(\text{FFT}(\mathbf{a}_1) \odot \text{IFFT}(\mathbf{y}_{n_r,2})), \quad (46)$$

$$\mathbf{y}_{n_r,2} [\text{circ}(-\mathbf{a}_0^*)]^\dagger = \sqrt{N} \text{FFT}(\text{FFT}(-\mathbf{a}_0) \odot \text{IFFT}(\mathbf{y}_{n_r,2})). \quad (47)$$

Based on (29), (42), and (44)-(47), we propose the structure shown in Fig. 5, where  $\hat{\mathbf{h}}_{n_r} = [\hat{\mathbf{h}}_{n_r,1}, \hat{\mathbf{h}}_{n_r,2}, \dots, \hat{\mathbf{h}}_{n_r,N_T}]$ , the number of points for each FFT and IFFT operation is  $N$ , “FFTo” only generates the first  $N_o$  values, “FFTe” generates the first  $N_e$  values, “MUX” multiplexes inputs in groups of  $L+1$  elements, with the first group coming from the upper branch, and  $\sqrt{N} \mathbf{q} = \sqrt{N} \mathbf{1}_{1 \times N_T} \otimes \mathbf{f}$  is the scaling vector of  $\hat{\mathbf{h}}_{n_r}$ , whose elements reflect the scalar coefficient in (29).

Regarding the computational complexity per receive antenna, there are  $N(3 \log_2 N + 4)$  complex multiplications,  $6N \log_2 N + N_T(L+1)$  complex additions and  $2N_T(L+1)$  real multiplications<sup>14</sup>, as there are 6 FFT/IFFT blocks, each with  $\frac{N}{2} \log_2 N$  multiplications and  $N \log_2 N$  additions [49], 4 “ $\odot$ ” units, each with  $N$  multiplications, one “ $\oplus$ ” unit with  $N_T(L+1)$  additions, and one scaling unit with  $2N_T(L+1)$  real multiplications. On the other hand, for each receive antenna, there are  $2NN_T(L+1)$  complex multiplications,  $(2N-1)N_T(L+1)$  complex additions and  $2N_T(L+1)$  real multiplications, if (29) is directly implemented via matrix multiplication. The proposed method offers significant computational saving, specially when the number of unknowns per receive antenna,  $N_T(L+1)$ , is large. For example, with  $N_T = 8$ ,  $L = 15$  and  $N = 64$ , the complexity is reduced by 94%, upon the proposed method.

## B. Hardware (ASIC/FPGA) Implementation

Although existing DSP’s can calculate the channel estimates fast, using the proposed fast algorithms, the cost and power consumption of such powerful DSP’s are not low. Compared to DSP’s, ASIC/FPGA designs have several advantages over DSP’s, as they provide intrinsic parallelism, high performance, and low cost. In the following subsections, algorithms and filter structures suitable for ASIC/FPGA implementations are developed for the ZP- and CP-based guard periods, respectively.

<sup>14</sup>The scaling vector  $\mathbf{q}$ ,  $\text{FFT}(\mathbf{a}_0^*)$ ,  $\text{FFT}(\mathbf{a}_1^*)$ ,  $\text{FFT}(-\mathbf{a}_0)$ , and  $\text{FFT}(\mathbf{a}_1)$  are not considered as they can be calculated once and stored for all the  $N_R$  receive antennas.

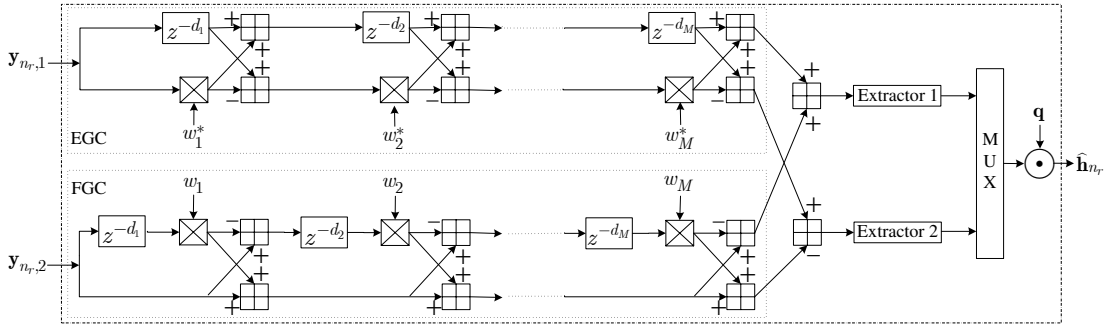


Fig. 6. The fast hardware channel estimator on the  $n_r^{\text{th}}$  receive antenna for the ZP-based guard period: The EGC/FGC structure.

1) *ZP-Based Guard Period*: The goal is to develop fast hardware designs, to calculate the first parenthesis in (35), the main part of the channel estimator of the  $n_r^{\text{th}}$  receive antenna. Here we consider complex  $\mathbf{a}_0$  and  $\mathbf{a}_1$  which are generated such that the delay vector  $\mathbf{d}$  is any permutation of  $[1, 2, \dots, 2^{M-1}]$ , and the weights are complex with unit amplitude, i.e.,  $|w_m| = 1, \forall m$ . Note that  $\mathbf{X}_\nu$  should be replaced with  $\mathbf{X}_\nu^*$  in (35) since it is complex now.

According to [35, Sec. V] and by defining  $\mathbf{y}_{n_r,\nu} = [y_{n_r,\nu}(0), y_{n_r,\nu}(1), \dots, y_{n_r,\nu}(N+L-1)]$ ,  $\nu \in [1, 2]$ , we can feed the received vector  $\mathbf{y}_{n_r,1}$  into a finite impulse response (FIR) filter bank (consisting of two parallel FIR filters), with impulse responses  $\bar{\mathbf{a}}_0^*$  and  $\bar{\mathbf{a}}_1^*$ , to calculate the 1<sup>st</sup> and 2<sup>nd</sup> rows of  $\mathbf{X}_1^* \mathbf{Y}_{n_r,1}$ , respectively. Also the elements of  $\mathbf{X}_2^* \mathbf{Y}_{n_r,2}$  can be computed by feeding  $\mathbf{y}_{n_r,2}$  into an FIR filter bank whose impulse responses are  $\mathbf{a}_1$  and  $-\mathbf{a}_0$ . However, the conventional implementation of each filter needs  $N-1$  complex adders and  $N$  complex multipliers. In what follows, we develop efficient filters by utilizing the special structure of our proposed training sequences.

If we take the  $Z$  transform of (4) with respect to  $k$ , we obtain [27]

$$\begin{aligned} A_0^{(m)}(z) &= A_0^{(m-1)}(z) + w_m A_1^{(m-1)}(z) z^{-d_m} \\ A_1^{(m)}(z) &= A_0^{(m-1)}(z) - w_m A_1^{(m-1)}(z) z^{-d_m}, \quad m \in [1, M], \end{aligned} \quad (48)$$

with  $A_0^{(0)}(z) = A_1^{(0)}(z) = 1$ . The filter structure corresponding to (48) is the fast Golay correlator (FGC) [27, Fig. 1], which is the matched filter for the sequences  $\bar{\mathbf{a}}_1^*$  and  $\bar{\mathbf{a}}_0^*$  [50]. As stated in [50], the matched filter for the sequences  $\mathbf{a}_0$  and  $\mathbf{a}_1$  is the efficient Golay correlator (EGC) [50, Fig. 2].

In order to convolve the inputs  $\mathbf{y}_{n_r,1}$  and  $\mathbf{y}_{n_r,2}$  with  $\{\bar{\mathbf{a}}_0^*, \bar{\mathbf{a}}_1^*\}$  and  $\{\mathbf{a}_1, -\mathbf{a}_0\}$ , to obtain the elements of  $\mathbf{X}_1^* \mathbf{Y}_{n_r,1}$  and  $\mathbf{X}_2^* \mathbf{Y}_{n_r,2}$  in (35), respectively, we propose to utilize EGC and FGC for the implementation of (35). The efficient hardware is given in Fig. 6, where “ $z^{-D}$ ” is the delay unit of length  $D$ , “Extractor 1” discards the first  $N-1$  values, takes the following  $N_o = \lfloor N_T/2 \rfloor (L+1)$  elements and drops the remaining, “Extractor 2” throws the first  $N-1$  values away, keeps the next  $N_e = \lfloor N_T/2 \rfloor (L+1)$  elements, and discards the rest. Moreover, “MUX” has the same function as that in Fig. 5. Note that Fig. 6 includes Fig. 2 of [35] as a special case. This can be easily shown by setting  $\alpha = 0$  in

(36),  $d_m = 2^{M-m}$ ,  $m \in [1, M]$  in (4), and  $w_k = 1, \forall k$ , i.e., removing all the “ $\boxtimes$ ” units.

Other hardware can be thought of, which use either EGC or FGC, but not both. The FGC-only structure is shown in Fig. 7, and similar structure can be derived with EGC only. The key idea relies on the fact that  $\mathbf{a} * \bar{\mathbf{b}} = \overleftarrow{\mathbf{a}} * \mathbf{b}$ . For hardware implementation, the reversing operation  $\overleftarrow{[\cdot]}$  is implemented by the last-in-first-out (LIFO) unit. In this scheme, the three switches are connected to the nodes labeled as ①, to calculate  $\mathbf{X}_1^* \mathbf{Y}_{n_r,1}$ . After finishing the computation of  $\mathbf{X}_1^* \mathbf{Y}_{n_r,1}$ , FGC is reset and all the switches are connected to the nodes labeled as ②, to compute  $\mathbf{X}_2^* \mathbf{Y}_{n_r,2}$ .

Regarding the hardware complexity, according to Fig. 6, both EGC and FGC have  $M$  stages, each stage has two “ $\boxplus$ ” units, one “ $\boxtimes$ ” unit, and a buffer of length  $d_m$  for the  $m^{\text{th}}$  stage. In addition, there are two “ $\boxplus$ ” units and one “ $\odot$ ” unit for post processing. Therefore, there are  $4M+2 = 4\log_2 N + 2$  “ $\boxplus$ ” units and  $2M+1 = 2\log_2 N + 1$  “ $\boxtimes$ ” units<sup>15</sup> on each receive antenna. However, there are  $4(N-1)+2$  “ $\boxplus$ ” units and  $4N+1$  “ $\boxtimes$ ” units, if (29) is implemented by the conventional method (four FIR filters mentioned at the beginning). For not so small  $N$ , the ratio is close to  $\log_2 N/N$  for the “ $\boxplus$ ” unit, and  $\log_2 N/(2N)$  for the “ $\boxtimes$ ” unit, which demonstrate the efficiency of Fig. 6. Similar order of magnitude hardware efficiency holds for Fig. 7.

If binary sequences are used for training, there are only  $4\log_2 N + 2$  “ $\boxplus$ ” units on each receive antenna, and one “ $\boxtimes$ ” unit is required for scaling [35].

2) *CP-Based Guard Period*: Similarly to the ZP case, we develop the fast hardware implementation of (42). The derivation in [36, Sec. IV] can be easily extended to the complex case by replacing the transpose with the conjugate transpose. Therefore, by using EGC and FGC developed above, the efficient filter structure for the  $n_r^{\text{th}}$  receive antenna

<sup>15</sup>Note that “ $\odot$ ” counts as one “ $\boxtimes$ ” unit.

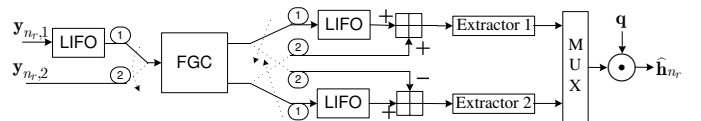


Fig. 7. The fast hardware channel estimator on the  $n_r^{\text{th}}$  receive antenna for the ZP-based guard period: The FGC only structure.

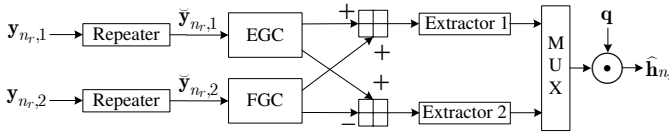


Fig. 8. The fast hardware channel estimator on the  $n_r^{\text{th}}$  receive antenna for the CP-based guard period: The EGC/FGC structure.

is given in Fig. 8, where  $\check{y}_{n_r,\nu} = [y_{n_r,\nu}(0), \dots, y_{n_r,\nu}(N-1), y_{n_r,\nu}(0), \dots, y_{n_r,\nu}(N_o-2)]$ , i.e.,  $\check{y}_{n_r,\nu}(n) = y_{n_r,\nu}((n)_N)$ ,  $\nu \in [1, 2]$ ,  $n \in [0, N + N_o - 2]$ , “Repeater” generates  $\check{y}_{n_r,\nu}$  from  $y_{n_r,\nu}$ , and other parts have the same functions as those in Fig. 6. In addition, similarly to the ZP case, EGC-only and FGC-only structures can be easily derived for the CP case as well.

By comparing Figs. 6 and 8, one can see that the two filter structures are very similar, except that Fig. 8 has two extra “Repeater” units. Therefore, they have the same hardware complexity in terms of the number of “ $\boxplus$ ” and “ $\boxminus$ ” units. In addition, it is clear to see that Fig. 8 includes Fig. 3 of [36] as a special case by setting  $\alpha = 1$  in (36) and  $w_k = 1, \forall k$ .

## VIII. COMPARISON WITH OTHER TRAINING SEQUENCES

### A. Comparison with the ZCZ Sequences

ZCZ sequences are considered in [12][13]. They satisfy the condition of (15) with  $\mathbf{S}$  given in (10), therefore, can be used for optimal training in the CP case. Compared to the scenario where each transmit antenna uses one ZCZ sequence per transmission frame, our proposed scheme contains an extra guard period of length  $L$  (the second CP in Fig. 4)<sup>16</sup>. However, this overhead is negligible if  $L$  is much smaller than frame length, which is the case in, for example, time invariant indoor MIMO systems. The benefit of our scheme is that the channel estimator can be implemented with low software and hardware complexity, as described in Sec. VII, using the special intrinsic structure of the sequences generated by (4). For ZCZ sequences we are unaware of such a feature, which is desirable in practice. Moreover, ZCZ sequences can not be used in the ZP-based guard period case, where aperiodic correlations are needed for optimal training design. To the best of our knowledge, there is no report on such ZCZ sequences.

### B. Comparison with Perfect Polyphase Sequences

For the CP-based guard period case, as stated in the introduction, we can use different phases of a perfect polyphase sequence, e.g., Frank sequence [23] or Chu sequence [24], as training symbols on different transmit antennas. It can save  $L$  separation symbols compared to our proposed method, as mentioned in the previous subsection. However, the resolution of the phase, which is important in a practical implementation,

is small when the sequence length is large [51]. Furthermore, all the perfect polyphase sequences are complex<sup>17</sup>, whereas in practice, binary training symbols are preferred since they are easier to generate and make the corresponding channel estimator simpler to implement. And similar to ZCZ sequences, they lack an efficient structure for fast hardware implementation. Finally, both Frank and Chu sequences are based on periodic correlations and can not be used in the ZP-based guard period case.

### C. Comparison with the Training Sequences of [11]

Delta impulse sequences are rarely used in practical communication systems due to their high PAPR. To calculate the PAPR of the training part, let us define the PAPR of a sequence  $\mathbf{x} = [x_1, x_2, \dots, x_N]$  as

$$\text{PAPR}(\mathbf{x}) = \frac{\max_{1 \leq n \leq N} |x_n|^2}{\frac{1}{N} \sum_{n=1}^N |x_n|^2}. \quad (49)$$

For the proposed ZP-based training, the PAPR is  $1 + \ell/N$  according to (20), (21) and (49), which is less than 2 while taking  $N \geq \ell$ . For CP-based training, the PAPR is 1 according to (22), (23) and (49), which is the lowest one can achieve. However, the PAPR of the impulse sequences given in [11, Table I] is  $N_T(L+1)$ , since there is only one non-zero entry in the sequence. This large PAPR results in low energy efficiency and is undesirable in practical applications.

Overall, the scheme proposed in this paper is very flexible since it can generate both binary and complex training sequences for both ZP- and CP-based guard periods. The most important aspect of the proposed training scheme is that the channel estimator can be implemented by DSP or ASIC/FPGA, with low complexity, which is of high interest in practice, at the negligible cost of a small overhead (extra  $L$  separation symbols).

## IX. SIMULATION RESULTS

### A. Comparison Between the Optimal and Suboptimal Training Power Allocations in the LMMSE Estimator

As mentioned in Sec. IV, in order to minimize the TMSE of the LMMSE estimator in (14), with  $\alpha = 1$ , one needs to choose  $\mathbf{S}$  such that (17) is satisfied, where the optimal power allocation matrix  $\mathbf{C}_s$  should be obtained according to (18). Now it is important to know how much we may lose by simply taking  $\mathbf{C}_s = \mathbf{I}_{N_T}$ , when using (13) with  $\alpha = 1$ . As an example, a  $2 \times 3$  MIMO ISI channel with  $L = 3$  is used, where  $\sigma_{n_r, n_t, l}$ 's are given in Table IV such that  $\sum_{l=0}^L \sigma_{n_r, n_t, l} = 1, \forall n_r, n_t$ . The number of training symbols  $N_s$  is chosen to be 16.

In Fig. 9, “Optimal” implies that the training power allocation matrix  $\mathbf{C}_s$  over all the transmit antennas is computed according to (18), combined with the Kuhn-Tucker conditions for each SNR, whereas “Suboptimal” indicates that all the transmit antennas have equal training power, i.e.,  $\mathbf{C}_s = \mathbf{I}_{N_T}$ . From Fig. 9, we can see that, at low SNR, different training powers are allocated to different transmit antennas, and at

<sup>16</sup>Note that it is not the case if training sequences are binary. To satisfy (15), the minimum length for binary ZCZ sequences (if they exist) is  $N_{s, \text{ZCZ}}^{\min} = 2N_T L$  according to (2) of [22]. However, our proposed scheme has  $N_s^{\min} = N_T(L+1) + L$ , which is smaller than  $N_{s, \text{ZCZ}}^{\min}$ . This suggests that the proposed scheme outperforms ZCZ sequences in terms of spectrum efficiency and implementation complexity, if binary training sequences are used.

<sup>17</sup>Only one perfect polyphase binary sequence exists, which is  $[+ + + -]$  [52].

TABLE IV

POWERS OF DIFFERENT SUBCHANNELS AND TAPS IN THE  $2 \times 3$  MIMO  
EXAMPLE

$\sigma_{n_r, n_t, l}$	$l = 0$	$l = 1$	$l = 2$	$l = 3$
$(n_r, n_t) = (1, 1)$	0.6	0.1	0.1	0.2
$(n_r, n_t) = (1, 2)$	0.2	0.3	0.1	0.4
$(n_r, n_t) = (1, 3)$	0.05	0.55	0.2	0.2
$(n_r, n_t) = (2, 1)$	0.3	0.3	0.2	0.2
$(n_r, n_t) = (2, 2)$	0.2	0.2	0.35	0.25
$(n_r, n_t) = (2, 3)$	0.15	0.3	0.2	0.35

high SNR, the power allocation converges to the equal-power scheme, i.e., each transmit antenna has the same training power. In addition, over all practical SNR's, there is almost no difference between the optimal and suboptimal schemes in terms of TMSE. It can be explained by the following argument: at low SNR, channel estimation error introduced by the additive noise is dominant and the benefit from the optimal power allocation is negligible, compared to that large error. At high SNR, on the other hand, all the CIR's can be estimated with very small error and the suboptimal scheme converges to the optimal scheme. Therefore, in practice, the estimation accuracy loss due to the equal-power scheme is negligible, in terms of TMSE.

### B. Comparison Between BLUE and LMMSE Estimator

In the simulation we take  $L = 7$ , i.e., there are 8 taps in each subchannel  $\mathbf{h}_{n_r, n_t}$ ,  $N_T = 8$ ,  $N_R = 8$ , and  $N = 32, 64$ . For the additive noise, we assume elements of  $\mathbf{E}$  are white complex Gaussian with unit variance. For the underlying fading channel, elements of  $\mathbf{H}$  are independent complex Gaussian, and each subchannel has the same exponential power delay profile such that  $\sigma_l = \frac{(1-e^{-1})e^{-l}}{1-e^{-L-1}}$ ,  $l \in [0, L]$ . Note that  $\sum_{l=0}^L \sigma_l = 1$ . Moreover,  $\mathbf{H}$  and  $\mathbf{E}$  are independent. Fig. 10 and 11 show the normalized theoretical minimum classical CRLB and Bayesian CRLB for BLUE and LMMSE estimator, given by  $\frac{N_T(L+1)}{2\gamma N}$  and  $\sum_{l=0}^L \frac{N_T \sigma_l}{2\gamma N \sigma_l + N_T}$ , respectively, derived from (31) and normalized by  $N_T N_R$ . Note that, although the

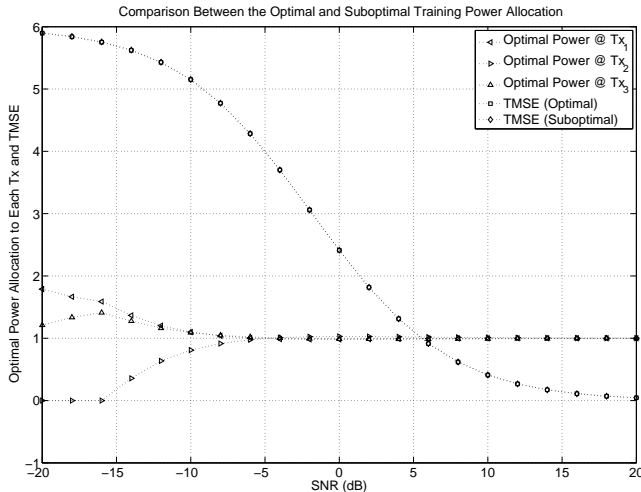


Fig. 9. Comparison between the optimal and suboptimal training power allocation in a  $2 \times 3$  MIMO ISI system with  $L = 3$  and the LMMSE estimator.

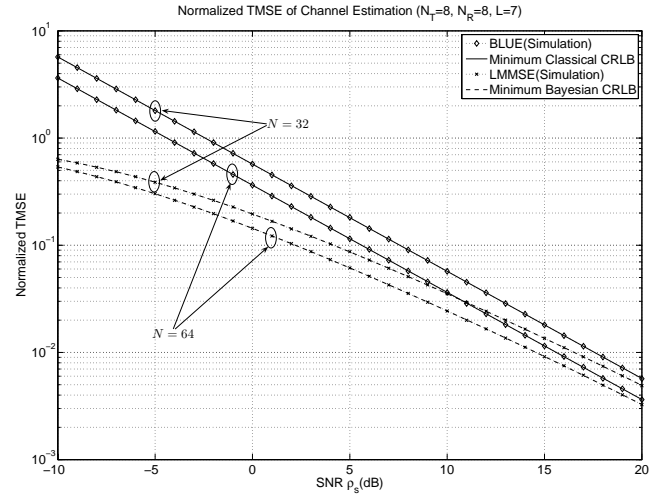


Fig. 10. The normalized TMSE of the proposed estimators for different number of training symbols: ZP-based guard period.

formula of TMSE for both ZP- and CP-based guard periods is the same for a specific estimator, the value of  $\gamma$ , defined in (32), is different. The simulated normalized TMSE of both estimators,  $\|\hat{\mathbf{H}} - \mathbf{H}\|_F^2 / \|\mathbf{H}\|_F^2$ , are plotted as well, which match the theoretical values perfectly.

From Fig. 10 and 11 it is obvious that each TMSE decreases, as  $N$  increases. The same conclusion holds for TMSE versus SNR. Furthermore, the LMMSE estimator has a better estimation performance than the BLUE because it utilizes the statistical information of the channel, while BLUE does not. At high SNRs, this difference becomes negligible, i.e., asymptotically, BLUE and LMMSE have the same performance in the high-SNR regime.

### C. BER Performance

Although the end-to-end bit-error-rate (BER) performance is an important factor, it is generally true that the less the channel estimation error, the better the BER performance

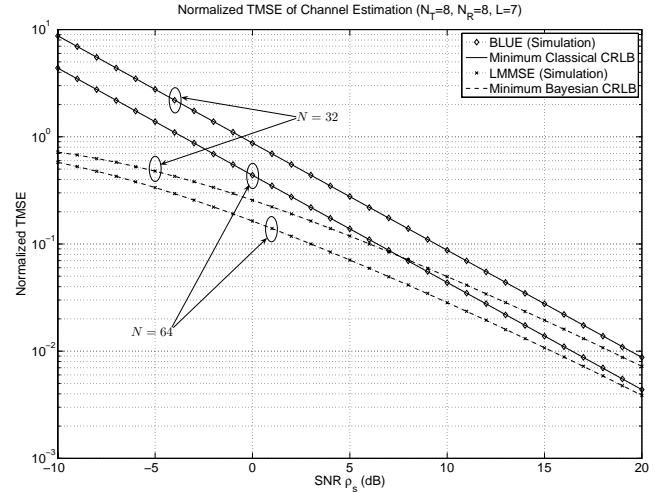


Fig. 11. The normalized total MSE of the proposed estimators for different number of training symbols: CP-based guard period.

for a fixed data detection scheme. Fig. 10 and 11 show the proposed schemes achieve the minimum possible CRLB, which is the best one can do for channel estimation. Therefore, the proposed schemes guarantee the best BER performance of a given detection method as well.

## X. CONCLUSION

In this paper we have shown how to construct optimal low-complexity training sequences using uncorrelated Golay complementary sets of *polyphase* sequences for MIMO frequency-selective block-fading channels, by extending the results in [35][36] to the complex sequences. Both zero padding- and cyclic prefix-based guard periods are considered. The optimality of the proposed estimators is due to the fact that they achieve the minimum possible Cramér-Rao lower bound, also verified by Monte Carlo simulations. A variety of fast DSP algorithms and hardware structures are presented to implement the channel estimators. Complexity of each algorithm or structure is analyzed as well.

The optimal training power allocation to the transmit antennas is also studied in this paper, assuming that the statistical information of the MIMO ISI channel is known at the transmit side. The theoretical analysis and simulation results show that there is a small performance improvement (in terms of TMSE) over the practical range of SNR's, when comparing optimal power allocation with the equal-power scheme.

Comparison with other existing training sequences is also carried out in this paper. The result shows the proposed schemes can be implemented with low complexity, by DSP and AISC/FPGA, much needed in practice for low manufacturing cost and long battery life. The low-complexity implementation combined with optimal estimation performance makes the proposed training-based channel estimators suitable and ready for real-world MIMO systems such as, but not limited to, MIMO-CDMA, MIMO-OFDM and MIMO-UWB systems.

## ACKNOWLEDGEMENT

The authors thank anonymous reviewers for their careful reviews of the paper. Their comments have certainly improved the quality of the paper.

## APPENDIX I

### PROOF OF PROPOSITION 1

#### A. ZP-Based Guard Period

1) *LMMSE Estimator* ( $\alpha = 1$ ): Based on  $\text{vec}(\mathbf{AZB}) = (\mathbf{B}' \otimes \mathbf{A})\text{vec}(\mathbf{Z})$  [53, p. 61], we can rewrite (7) as  $\mathbf{y} = \mathbf{\Phi}\mathbf{h} + \mathbf{e}$ , where  $\mathbf{\Phi} = \sqrt{\frac{\rho_s}{N_T}}(\mathbf{S}' \otimes \mathbf{I}_{N_R})$ ,  $\mathbf{S}$  is  $\mathbf{S}^{\text{ZP}}$  in (8),  $\mathbf{y} = \text{vec}(\mathbf{Y})$ ,  $\mathbf{h} = \text{vec}(\mathbf{H})$ , and  $\mathbf{e} = \text{vec}(\mathbf{E})$ .

According to the Bayesian Gauss-Markov Theorem [47, p. 391], with the covariance matrix of the additive noise  $\mathbf{e}$  as  $\mathbf{C}_e = \mathbb{E}[\mathbf{e}\mathbf{e}^\dagger] = \mathbf{I}_{(N_s+L)N_R}$ , and the covariance matrix of the fading channel  $\mathbf{h}$  as given in (12), the LMMSE estimator of  $\mathbf{h}$  is given by

$$\begin{aligned} \hat{\mathbf{h}} &= (\mathbf{\Phi}^\dagger \mathbf{\Phi} + \mathbf{C}_h^{-1})^{-1} \mathbf{\Phi}^\dagger \mathbf{y}, \\ &= \sqrt{\frac{N_T}{\rho_s}} \left[ (\mathbf{S}^* \mathbf{S}') \otimes \mathbf{I}_{N_R} + \frac{N_T}{\rho_s} \mathbf{C}_h^{-1} \right]^{-1} (\mathbf{S}^* \otimes \mathbf{I}_{N_R}) \mathbf{y}, \end{aligned} \quad (50)$$

which is equal to (13) while  $\alpha = 1$ .

For the error vector  $\boldsymbol{\epsilon} = \mathbf{h} - \hat{\mathbf{h}}$ , its covariance matrix,  $\mathbf{C}_\epsilon = \mathbb{E}[\boldsymbol{\epsilon}\boldsymbol{\epsilon}^\dagger]$ , can be expressed as [47, p. 391]

$$\begin{aligned} \mathbf{C}_\epsilon &= (\mathbf{\Phi}^\dagger \mathbf{\Phi} + \mathbf{C}_h^{-1})^{-1}, \\ &= \frac{N_T}{\rho_s} \left[ (\mathbf{S}^* \mathbf{S}') \otimes \mathbf{I}_{N_R} + \frac{N_T}{\rho_s} \mathbf{C}_h^{-1} \right]^{-1}. \end{aligned} \quad (51)$$

Therefore the TMSE, defined by  $\varepsilon_1 = \text{tr}[\mathbf{C}_\epsilon]$ , which is the same as  $\text{tr}[\mathbf{C}_\epsilon^*]$  since  $\mathbf{C}_\epsilon$  is Hermitian, is equal to (14), with  $\alpha = 1$ .

If both  $\mathbf{H}$  and  $\mathbf{E}$  are Gaussian and independent, the posterior probability density function (PDF)  $p(\mathbf{h}|\mathbf{y})$  is also Gaussian [47, p. 324], and LMMSE estimator has the same performance as the MMSE estimator [47, p. 391]. On the other hand, according to property #3 in [37, p. 84], the MMSE estimator attains the Bayesian CRLB, when the posterior PDF is Gaussian. Therefore, the TMSE  $\varepsilon_1$  in (14) is the Bayesian CRLB for Gaussian and independent  $\mathbf{H}$  and  $\mathbf{E}$ .

2) *BLUE* ( $\alpha = 0$ ): For BLUE, using the Gauss-Markov Theorem [47, p. 141], we can follow the same procedure as shown in Appendix I-A.1, to prove (13) and (14), when  $\alpha = 0$ , and also show that  $\varepsilon_0$  is the classical CRLB when  $\mathbf{H}$  is deterministic and  $\mathbf{E}$  is Gaussian. For this case, (13) and (14) can be further simplified to

$$\hat{\mathbf{H}} = \sqrt{\frac{N_T}{\rho_s}} (\mathbf{S}\mathbf{S}^\dagger)^{-1} \mathbf{S}^\dagger \mathbf{Y}, \quad (52)$$

and

$$\varepsilon_0 = \frac{N_R N_T}{\rho_s} \text{tr}[(\mathbf{S}\mathbf{S}^\dagger)^{-1}]. \quad (53)$$

#### B. CP-Based Guard Period

The proposition can be proved using the same procedure as in Appendix I-A, except that  $\mathbf{S}$  denotes  $\mathbf{S}^{\text{CP}}$  in (10) and the covariance matrix of the additive noise  $\mathbf{e}$  is given by  $\mathbf{C}_e = \mathbb{E}[\mathbf{e}\mathbf{e}^\dagger] = \mathbf{I}_{N_s N_R}$ .

## APPENDIX II

### PROOF OF PROPOSITION 2

#### A. ZP-Based Guard Period

1) *LMMSE Estimator* ( $\alpha = 1$ ): For an  $N_s \times N_s$  positive definite matrix  $\mathbf{A}$ , we have  $\text{tr}(\mathbf{A}^{-1}) \geq \text{tr}\{\text{dg}(\mathbf{A})\}^{-1}$ , where the equality holds iff  $\mathbf{A}$  is diagonal [54, Lemma 1]. This implies that  $\varepsilon_1$  in (14) is minimized iff  $\mathbf{S}\mathbf{S}^\dagger$  is diagonal, since  $\mathbf{C}_h^{-1}$  itself is diagonal. This particularly justifies (16). Now we need to determine the diagonal elements of  $\mathbf{S}\mathbf{S}^\dagger$  such that  $\varepsilon_1$  is minimized further. According to the definition of  $\mathbf{S}$  in (8), we have  $\text{dg}(\mathbf{S}\mathbf{S}^\dagger) = N_s \mathbf{I}_{L+1} \otimes \mathbf{C}_s$ ,<sup>18</sup> where  $\mathbf{C}_s = \text{diag}(P_1, P_2, \dots, P_{N_T})$ ,  $P_{n_t} = \frac{1}{N_s} \|\mathbf{s}_{n_t}\|_F^2$ , and  $\sum_{n_t=1}^{N_T} P_{n_t} \leq N_T$  due to the transmit power constraint  $\frac{1}{N_s(L+1)} \text{tr}[\mathbf{S}\mathbf{S}^\dagger] \leq N_T$ . Based on  $\text{tr}(\mathbf{A}^{-1}) \geq \text{tr}\{\text{dg}(\mathbf{A})\}^{-1}$ ,  $\text{dg}(\mathbf{A} + \mathbf{B}) = \text{dg}(\mathbf{A}) +$

<sup>18</sup>This completes the proof of (16) since  $\text{dg}(\mathbf{S}\mathbf{S}^\dagger) = \mathbf{S}\mathbf{S}^\dagger$  when  $\mathbf{S}\mathbf{S}^\dagger$  is diagonal.

$\text{dg}(\mathbf{B})$ , and  $\text{dg}(\mathbf{A} \otimes \mathbf{I}_{N_R}) = \text{dg}(\mathbf{A}) \otimes \mathbf{I}_{N_R}$ ,  $\varepsilon_1$  in (14) can be lower bounded by

$$\begin{aligned} \varepsilon_1 &\geq \frac{N_T}{\rho_s} \text{tr} \left[ \left( N_s \mathbf{I}_{L+1} \otimes \mathbf{C}_s \otimes \mathbf{I}_{N_R} + \frac{N_T}{\rho_s} \mathbf{C}_h^{-1} \right)^{-1} \right], \\ &= \sum_{n_r=1}^{N_R} \sum_{n_t=1}^{N_T} \sum_{l=0}^L \frac{1}{\sigma_{n_r, n_t, l}^{-1} + \frac{N_s \rho_s}{N_T} P_{n_t}}. \end{aligned} \quad (54)$$

To minimize the lower bound in (54), we use Lagrangian multiplier  $\lambda$ , take the derivative of  $\sum_{n_r=1}^{N_R} \sum_{n_t=1}^{N_T} \sum_{l=0}^L \frac{1}{\sigma_{n_r, n_t, l}^{-1} + \frac{N_s \rho_s}{N_T} P_{n_t}} + \lambda \left( \sum_{n_t=1}^{N_T} P_{n_t} - N_T \right)$  with respect to  $P_{n_t}$ ,  $\forall n_t$ , and set it to 0, which yields (18). Since  $P_{n_t}$ 's must be non-negative, it may not always be possible to find a solution to satisfy (18). In this case, the Kuhn-Tucker conditions [48] need to be used to solve (18), and the waterfilling-like power allocation is derived.

If  $\sigma_{n_r, n_t, l}$  is independent of  $n_t$ , i.e., the set  $\{\sigma_{n_r, n_t, l}\}_{(l=0, n_r=1)}^{(L, N_R)}$  has the same elements,  $\forall n_t$ , then clearly  $P_{n_t} = 1, \forall n_t$ , satisfies (18) and the power constraint  $\sum_{n_t=1}^{N_T} P_{n_t} \leq N_T$ . Since the lower bound in (54) is strictly convex with respect to  $P_{n_t}, \forall n_t$ , its unique minimum is attained at  $P_{n_t} = 1, \forall n_t$ , which gives  $\mathbf{C}_s = \mathbf{I}_{N_T}$ , and hence simplify (17) to (15).

2) *BLUE* ( $\alpha = 0$ ): For  $\alpha = 0$ , we can follow the same procedure in Appendix II-A.1 to prove the minimum of  $\varepsilon_0$  is achieved iff  $\mathbf{S}\mathbf{S}^\dagger = N_s \mathbf{I}_{N_T(L+1)}$ . This result is also reported in [20, pp. 178-179].

### B. CP-Based Guard Period

The proposition can be proved using the same procedure as in Appendix II-A.

## APPENDIX III

### THE CASE OF $p > 2$ TRAINING SEQUENCES PER FRAME

For the case of  $p > 2$ , i.e., when each transmit antenna uses  $p$  training sequences per frame, the received signal corresponding the  $\nu^{\text{th}}$  training sequence can be written as

$$\mathbf{Y}_\nu = \sqrt{\frac{\rho_s}{N_T}} \mathbf{H} \mathbf{S}_\nu + \mathbf{E}_\nu, \quad (55)$$

where  $\mathbf{E}_\nu$  is the additive noise component.

With  $\mathbf{C}_h = \mathbf{C}_{\Sigma} \otimes \mathbf{I}_{N_T N_R}$ , the BLUE and LMMSE estimators of  $\mathbf{H}$  are given by

$$\hat{\mathbf{H}} = \sqrt{\frac{N_T}{\rho_s}} \left( \sum_{\nu=1}^p \mathbf{Y}_\nu \mathbf{S}_\nu^\dagger \right) \left( \sum_{\nu=1}^p \mathbf{S}_\nu \mathbf{S}_\nu^\dagger + \alpha \frac{N_T}{\rho_s} \mathbf{C}_{\Sigma}^{-1} \otimes \mathbf{I}_{N_T} \right)^{-1}, \quad (56)$$

with TMSE

$$\varepsilon_\alpha = \frac{N_R N_T}{\rho_s} \text{tr} \left[ \left( \sum_{\nu=1}^p \mathbf{S}_\nu \mathbf{S}_\nu^\dagger + \alpha \frac{N_T}{\rho_s} \mathbf{C}_{\Sigma}^{-1} \otimes \mathbf{I}_{N_T} \right)^{-1} \right]. \quad (57)$$

For  $p = 2$ , (56) and (57) reduce to (26), (27) and (30), respectively.

If uncorrelated Golay complementary sets, each with  $p$  sequences, are used, i.e.,  $\sum_{\nu=1}^p \mathbf{S}_\nu \mathbf{S}_\nu^\dagger = p \check{N} \mathbf{I}_{N_T(L+1)}$ , which is the extension of (19), then (56) reduces to

$$\hat{\mathbf{H}} = \left( \sum_{\nu=1}^p \mathbf{Y}_\nu \mathbf{S}_\nu^\dagger \right) \left[ p \check{N} \sqrt{\frac{\rho_s}{N_T}} \mathbf{I}_{L+1} + \alpha \sqrt{\frac{N_T}{\rho_s}} \mathbf{C}_{\Sigma}^{-1} \right]^{-1} \otimes \mathbf{I}_{N_T}, \quad (58)$$

whose elements are

$$\hat{h}_{n_r, n_t}(l) = \frac{\sigma_l \sqrt{\rho_s N_T}}{p \check{N} \rho_s \sigma_l + \alpha N_T} \left[ \sum_{\nu=1}^p \mathbf{Y}_\nu \mathbf{S}_\nu^\dagger \right]_{n_r-1, l N_T + n_t - 1}, \quad (59)$$

and TMSE is given by

$$\varepsilon_\alpha = \sum_{l=0}^L \frac{N_R N_T^2 \sigma_l}{p \check{N} \rho_s \sigma_l + \alpha N_T}. \quad (60)$$

For  $p = 2$ , (58), (59) and (60) reduce to (28), (29) and (31), respectively.

## REFERENCES

- [1] S. Wang and A. Abdi, "MIMO frequency selective channel estimation using aperiodic complementary sets of sequences," in *Proc. IEEE Global Telecommun. Conf.*, St. Louis, MO, 2005, pp. 2239-2243.
- [2] —, "Optimal training sequences for efficient MIMO frequency selective fading channel estimation," in *Proc. IEEE Samoff Symp.*, Princeton, NJ, 2006.
- [3] G. J. Foschini and M. J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Personal Commun.*, vol. 6, pp. 311-335, 1998.
- [4] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," *European Trans. Telecommun.*, vol. 10, pp. 585-595, 1999.
- [5] S. M. Alamouti, "A simple transmit diversity technique for wireless communications," *IEEE J. Select. Areas Commun.*, vol. 16, pp. 1451-1458, Oct. 1998.
- [6] H. Bölcskei, R. W. Heath Jr., and A. J. Paulraj, "Blind channel identification and equalization in OFDM-based multiantenna systems," *IEEE Trans. Signal Processing*, vol. 50, pp. 96-109, Jan. 2002.
- [7] 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.
- [8] L. Zheng and D. N. C. Tse, "Communication on the Grassmann manifold: a geometric approach to the noncoherent multiple-antenna channel," *IEEE Trans. Inform. Theory*, vol. 48, pp. 359-383, Feb. 2002.
- [9] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links?" *IEEE Trans. Inform. Theory*, vol. 49, pp. 951-963, Apr. 2003.
- [10] H. Vikalo, B. Hassibi, B. M. Hochwald, and T. Kailath, "On the capacity of frequency-selective channels in training-based transmission schemes," *IEEE Trans. Signal Processing*, vol. 52, pp. 2572-2583, Sept. 2004.
- [11] 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.
- [12] 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.
- [13] 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.
- [14] C. Fragouli, N. Al-Dhahir, and W. Turin, "Training-based channel estimation for multiple-antenna broadband transmissions," *IEEE Trans. Wireless Commun.*, vol. 2, pp. 384-391, Mar. 2003.
- [15] T. F. Wong and B. Park, "Training sequence optimization in MIMO systems with colored interference," *IEEE Trans. Commun.*, vol. 52, pp. 1939-1947, Nov. 2004.
- [16] I. Barhumi, G. Leus, and M. Moonen, "Optimal training design for MIMO OFDM systems in mobile wireless channels," *IEEE Trans. Signal Processing*, vol. 51, pp. 1615-1624, June 2003.



- [17] H. Minn and N. Al-Dahir, "Optimal training signals for MIMO OFDM channel estimation," *IEEE Trans. Wireless Commun.*, vol. 5, pp. 1158–1168, May 2006.
- [18] A. Vosoughi and A. Scaglione, "On the effect of receiver estimation error upon channel mutual information," *IEEE Trans. Signal Processing*, vol. 54, pp. 459–472, Feb. 2006.
- [19] O. Simeone and U. Spagnolini, "Lower bound on training-based channel estimation error for frequency-selective block-fading Rayleigh MIMO channels," *IEEE Trans. Signal Processing*, vol. 52, pp. 3265–3277, Nov. 2004.
- [20] E. G. Larsson and P. Stoica, *Space-Time Block Coding for Wireless Communications*. Cambridge, UK: Cambridge University Press, 2003.
- [21] P. Z. Fan, N. Suehiro, N. Kuroyanagi, and X. M. Deng, "Class of binary sequences with zero correlation zone," *Electron. Lett.*, vol. 35, pp. 777–779, May 1999.
- [22] H. Torii, M. Nakamura, and N. Suehiro, "A new class of zero-correlation zone sequences," *IEEE Trans. Inform. Theory*, vol. 50, pp. 559–565, Mar. 2004.
- [23] R. Frank, S. Zadoff, and R. Heimiller, "Phase shift pulse codes with good periodic correlation properties," *IEEE Trans. Inform. Theory*, vol. 8, pp. 381–382, Oct. 1962.
- [24] D. Chu, "Polyphase codes with good periodic correlation properties," *IEEE Trans. Inform. Theory*, vol. 18, pp. 531–532, July 1972.
- [25] M. J. E. Golay, "Complementary series," *IEEE Trans. Inform. Theory*, vol. 7, pp. 82–87, Apr. 1961.
- [26] —, "Multislit spectrometry," *J. Opt. Soc. Amer.*, vol. 39, p. 437, 1949.
- [27] S. Z. Budišin, "Efficient pulse compressor for Golay complementary sequences," *Electron. Lett.*, vol. 27, pp. 219–220, Jan. 1991.
- [28] K. Niu, S. Wang, Q.-F. He, and W.-L. Wu, "A novel matched filter for primary synchronization channel in W-CDMA," in *Proc. IEEE Veh. Technol. Conf.*, Birmingham, AL, 2002, pp. 2052–2055.
- [29] J. A. Davis and J. Jedwab, "Peak-to-mean power control in OFDM, Golay complementary sequences, and Reed-Muller codes," *IEEE Trans. Inform. Theory*, vol. 45, pp. 2397–2417, Nov. 1999.
- [30] H.-H. Chen, J.-F. Yeh, and N. Suehiro, "A multicarrier CDMA architecture based on orthogonal complementary codes for new generations of wideband wireless communications," *IEEE Commun. Mag.*, vol. 39, no. 10, pp. 126–135, Oct. 2001.
- [31] S. Grob and P. D. J. Clark, "Enhanced channel impulse response identification for the ITU HF measurement campaign," *Electron. Lett.*, vol. 34, pp. 1022–1023, May 1998.
- [32] C. Tellambura, Y. J. Guo, and S. K. Barton, "Channel estimation using aperiodic binary sequences," *IEEE Commun. Lett.*, vol. 2, pp. 140–142, May 1998.
- [33] P. Spasojevic and C. N. Georgiades, "Complementary sequences for ISI channel estimation," *IEEE Trans. Inform. Theory*, vol. 47, pp. 1145–1152, Mar. 2001.
- [34] B. Xu and G. Bi, "Channel estimation using complementary sequence pairs for UWB/OFDM systems," *Electron. Lett.*, vol. 40, pp. 1196–1197, Sept. 2004.
- [35] 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.
- [36] —, "Low-complexity optimal estimation of MIMO ISI channels with binary training sequences," *IEEE Signal Processing Lett.*, vol. 13, pp. 657–660, Nov. 2006.
- [37] H. L. Van Trees, *Detection, Estimation, and Modulation Theory, Vol. I*. New York: Wiley, 1968.
- [38] B. Z. Bobrovsky, E. Mayer-Wolf, and M. Zakai, "Some classes of global Cramér-Rao bounds," *Ann. Statist.*, vol. 15, pp. 1221–1438, 1987.
- [39] 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.
- [40] P. J. Davis, *Circulant Matrices*. New York: Wiley, 1979.
- [41] C. C. Tseng and C. L. Liu, "Complementary set of sequences," *IEEE Trans. Inform. Theory*, vol. 18, pp. 644–652, Sept. 1972.
- [42] R. Sivaswamy, "Multiphase complementary codes," *IEEE Trans. Inform. Theory*, vol. 24, pp. 546–552, Sept. 1978.
- [43] L. Bomer and M. Antweiler, "Periodic complementary binary sequences," *IEEE Trans. Inform. Theory*, vol. 36, pp. 1487–1494, Nov. 1990.
- [44] P. Fan and M. Darnell, *Sequence Design for Communications Applications*. Hertfordshire, UK: Research Studies Press, 1996.
- [45] D. Ž. Doković, "Note on periodic complementary sets of binary sequences," *Designs, Codes and Cryptography*, vol. 13, pp. 251–256, 1998.
- [46] Y. Taki, H. Miyakawa, M. Hatori, and S. Namba, "Even-shift orthogonal sequences," *IEEE Trans. Inform. Theory*, vol. 15, pp. 295–300, Mar. 1969.
- [47] S. M. Kay, *Fundamentals of Statistical Signal Processing, Volume I: Estimation Theory*. Upper Saddle River, NJ: Prentice Hall PTR, 1993.
- [48] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, UK: Cambridge University Press, 2004.
- [49] A. V. Oppenheim, R. W. Schaffer, and J. R. Buck, *Discrete-Time Signal Processing*, 2nd ed. Upper Saddle River, NJ: Prentice Hall, 1999.
- [50] B. M. Popović, "Efficient Golay correlator," *Electron. Lett.*, vol. 35, pp. 1427–1428, Aug. 1999.
- [51] M. Antweiler and L. Bomer, "Merit factor of Chu and Frank sequences," *Electron. Lett.*, vol. 26, pp. 2068–2070, Dec. 1990.
- [52] H. D. Luke, H. D. Schotten, and H. Hadinejad-Mahram, "Binary and quadriphase sequences with optimal autocorrelation properties: A survey," *IEEE Trans. Inform. Theory*, vol. 49, pp. 3271–3282, Dec. 2003.
- [53] W.-H. Steeb, *Kronecker Product of Matrices and Applications*. Mannheim, Germany: B. I. Wissenschaftsverlag, 1991.
- [54] S. Ohno and G. B. Giannakis, "Capacity maximizing MMSE-optimal pilots for wireless OFDM over frequency-selective block Rayleigh-fading channels," *IEEE Trans. Inform. Theory*, vol. 50, pp. 2138–2145, Sept. 2004.



**Shuangquan Wang** (S'00-M'06) received his B.S. degree in Computer Engineering from Xi'an Institute of Posts and Telecommunications, Xi'an, China, in 1998, M.S. degree in Electrical Engineering from Beijing University of Posts and Telecommunications (BUPT), Beijing, China, in 2001, and Ph.D. degree in Electrical Engineering from New Jersey Institute of Technology, Newark, USA, in 2006. From 2001 to 2002, he worked on TD-SCDMA in Siemens Ltd. China, Beijing. Now he is with NEC Laboratories America, Inc., Princeton. His current research interests include MIMO channel modeling, estimation and its 2<sup>nd</sup>-order statistics, random matrix theory and its application in wireless MIMO communications, multiuser communications, and space-time transceiver design in MIMO-OFDM systems.

Dr. Wang has won several awards for his research contributions. These include the BUPT EI index prize in 2005, the 2005 New Jersey Inventors Hall of Fame Graduate Student Award, and the 1<sup>st</sup> Place in Graduate Poster Contest, IEEE GLOBECOM 2005.

PLACE  
PHOTO  
HERE

**Ali Abdi** (S'98-M'01-SM'06) received the Ph.D. degree in electrical engineering from the University of Minnesota, Minneapolis, in 2001. He joined the Department of Electrical and Computer Engineering of New Jersey Institute of Technology (NJIT), Newark, in 2001, as an Assistant Professor. His current research interests include estimation and characterization of wireless channels, digital communication in terrestrial and underwater channels, blind modulation recognition and parameter estimation, space-time processing and interference cancellation.

Dr. Abdi is an Associate Editor for the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY.