

Cross Z-Complementary Pairs (CZCPs) for Optimal Training in Spatial Modulation Over Frequency Selective Channels

Zilong Liu, Ping Yang, Yong Liang Guan, Pei Xiao

Abstract—The contributions of this paper are twofold: Firstly, we introduce a novel class of sequence pairs, called “cross Z-complementary pairs (CZCPs)”, each displaying zero-correlation zone (ZCZ) properties for *both* their aperiodic autocorrelation sums and crosscorrelation sums. Systematic constructions of *perfect* CZCPs based on selected Golay complementary pairs (GCPs) are presented. Secondly, we point out that CZCPs can be utilized as a key component in designing training sequences for *broadband* spatial modulation (SM) systems. We show that our proposed SM training sequences derived from CZCPs lead to *optimal* channel estimation performance over frequency-selective channels.

Index Terms—Golay complementary pairs (GCPs), channel estimation, training sequence design, spatial modulation, frequency-selective channels.

I. INTRODUCTION

This paper investigates a novel class of sequence pairs called “cross Z-complementary pairs (CZCPs)” and their applications for training sequence design in *broadband* spatial modulation (SM) systems. In what follows, we review the state-of-the-art on pairs of sequences and SM, then introduce our contributions in this work.

A. Pairs of Sequences

“Pairs of sequences” has been an attractive research topic since the invention of complementary pairs proposed by Marcel J. E. Golay in 1951 [1]. It was found by Golay that infrared multislit spectrometry, a device which isolates a desired radiation with a fixed single wavelength from background radiation (with many different wavelengths), can be designed with the aid of a special class of pairs of sequences, which is widely known as Golay complementary pair (GCP).

Let \mathbf{a} and \mathbf{b} be two sequences of identical length. Denote by $\rho(\mathbf{a}, \mathbf{b})(\tau)$ their aperiodic crosscorrelation (ACC) at time-shift τ (formal definition of ACC will be given in Section II). For simplicity, when $\mathbf{a} = \mathbf{b}$, $\rho(\mathbf{a}, \mathbf{b})(\tau)$ will be written as $\rho(\mathbf{a})(\tau)$. By definition, (\mathbf{a}, \mathbf{b}) is said to be a GCP if $\rho(\mathbf{a})(\tau) + \rho(\mathbf{b})(\tau) = 0$ for all $\tau \neq 0$. In principle, sequences

in a GCP work in a cooperative way so that their aperiodic autocorrelations (AAC) sums vanish for all the non-zero time-shifts [2].

For more than 60 years, huge amount of research effort has been made on the constructions of GCPs. Recursive algorithms proposed by Budišin in 1990s for polyphase and multi-level GCPs can be found in [3], [4]. Theory on para-unitary (PU) matrices has recently been applied for more efficient GCP synthesis [5]–[8]. In 1999, Davis and Jedwab developed a direct construction of polyphase GCPs through generalized Boolean functions [9]. Their construction was extended to GCPs over QAM constellations using weighted sum of several QPSK GCPs [10], [11]. Existing known binary GCPs have even lengths of the form $2^\alpha 10^\beta 26^\gamma$ only, where α, β, γ are non-negative integers satisfying $\alpha + \beta + \gamma \geq 1$ [12], [13]. For sequence pairs with odd lengths, the best known alternative to GCPs is *optimal* odd-length binary Z-complementary pairs (OB-ZCPs) [14] each displaying the largest zero-correlation zone (ZCZ) width [15] and minimum AAC sums outside the ZCZ.

GCPs and their extension, called “complementary codes (CC)” [16], [17], have found applications far beyond their initial usage in the design of multislit spectrometry. A few selected applications in wireless communications include: optimal channel estimation in multiple-input multiple-output (MIMO) frequency-selective channels [18], [19], Doppler-resilient waveform design [20], [21], interference-free multi-carrier code-division multiple-access (MC-CDMA) communications [22]–[24], and power control in MC communications [9]–[11], [25]–[27].

It should be noted that each GCP, ZCP or CC, is defined by their AAC sums only. Within each GCP, ZCP or CC, it is assumed that separate non-interfering channels are used for the transmission of the constituent sequences. Crosscorrelations between constituent sequences within each GCP, ZCP or CC, have not been considered.

B. Spatial Modulation (SM)

MIMO is a revolutionary communication paradigm which can achieve significant increase in spectral efficiency and enhanced robustness against fading through exploitation of spatial diversity [28]–[34]. SM is a special class of MIMO techniques which trades multiplexing gain with complexity and performance [35]–[39]. Unlike conventional MIMO, an SM system is equipped with multiple transmit antenna (TA)

Zilong Liu is with School of Computer Science and Electronic Engineering, University of Essex, United Kingdom (e-mail: zilong.liu@essex.ac.uk). Ping Yang is with National Key Laboratory on Communications, University of Electronic Science and Technology of China, Chengdu, China (e-mail: yang.ping@uestc.edu.cn). Pei Xiao is with Institute for Communication Systems, 5G Innovation Centre, University of Surrey, United Kingdom (e-mail: p.xiao@surrey.ac.uk). Yong Liang Guan is with the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore (E-mail: eylguan@ntu.edu.sg).

elements but only a single radio-frequency (RF) chain. During each time-slot, an SM symbol can be divided into two parts: one (called “spatial symbol”) is responsible for the selection and then activation of a TA element, the other (called “constellation symbol”) is selected from a conventional PSK/QAM constellation and transmitted from the active TA element. Such unique transmission principle of SM allows it to have at least the following advantages over conventional MIMO:

- 1) Zero inter-channel interference (ICI) over flat fading channels and thus improved average bit error probability;
- 2) Lower hardware complexity in the transmitter (due to its single RF chain) and lower detection complexity at the receiver (due to zero ICI);
- 3) Lower energy consumption which makes SM a promising candidate for large-scale antenna systems [40].

Early literature on SM mostly assume that channel state information (CSI) is perfectly known at the receiver [41], [42]. Effects of channel estimation errors to the performance of SM systems, limited to flat-fading channels, have been studied in [43]–[46]. Up to date, however, little has been understood on channel estimation of SM in frequency-selective channels. Along this research direction, two fundamental research problems arise and will be addressed in this work: 1) What is the lower bound on the channel estimation mean-squares-error (MSE)? 2) How to design efficient channel training scheme to meet this lower bound with equality?

Note that the “one-RF-chain” principle of SM prevents the transmitter from using simultaneous pilot transmission over all the TAs. Consequently, it implies that dense training sequences proposed in [48]–[50] for traditional MIMO are unapplicable in SM systems. Although an identity training matrix has been employed for joint channel estimation and data detection in SM systems [45], extension to frequency-selective channels is not straightforward. A naive scheme is to extend a perfect sequence (having zero autocorrelation sidelobes) with cyclic-prefix (CP)¹ and then send the extended sequence in turn over multiple TAs. But this training scheme would be inefficient in highly dispersive channels.

C. Contributions of This Paper

The main contributions of this work are summarized as follows.

- 1) We introduce a novel class of sequence pairs, called *CZCPs*, each displaying certain *ZCZ* properties for both their AAC sums and ACC sums. Specifically, a pair of sequences, say (\mathbf{a}, \mathbf{b}) , is called a CZCP if $\rho(\mathbf{a})(\tau) + \rho(\mathbf{b})(\tau)$ and $\rho(\mathbf{a}, \mathbf{b})(\tau) + \rho(\mathbf{b}, \mathbf{a})(\tau)$ take zero values for certain time-shifts τ . The most distinctive feature of CZCPs (compared to conventional sequence pairs) is that the two constituent sequences in a CZCP may interfere with each other during the transmission. Therefore, careful design should be carried out to force their cross-interference to zero in a certain τ region. We investigate the structural properties of CZCPs and show that a subset of GCPs are *perfect CZCPs*.

- 2) We present a generic training framework for SM training over frequency selective channels. Under this framework, we derive the lower bound on channel estimation MSE using least square (LS) estimator and conditions to meet this lower bound with equality. Then, we show that CZCPs play an instrumental role in the design of optimal SM training sequences (which are equivalent to certain sparse matrices). Numerical simulations indicate that the proposed SM training sequences lead to minimum channel estimation MSE w.r.t. the aforementioned lower bound.

D. Organization of This Paper

This paper is organized as follows. Section II introduces some notations and GCPs, followed by a sketch of the basic principle of SM. In Section III, we formally define CZCPs, give their structural properties, and then present *perfect CZCPs* obtained from computer search and systematic constructions. Section IV starts from a generic framework on the training design of SM, under which a lower bound on the channel estimation MSE is derived. Then, CZCPs are applied to design sparse training matrices for optimal channel estimation performance (w.r.t. the derived lower bound) in SM-MIMO frequency selective channels. In the end of Section IV, we evaluate the proposed SM training scheme using numerical simulations. Finally, Section V concludes this paper.

II. PRELIMINARIES

A. Notations

The following notations will be used throughout this paper.

- \mathbf{X}^T and \mathbf{X}^H denote the transpose and the Hermitian transpose of matrix \mathbf{X} , respectively;
- $\|\mathbf{X}\| = \sqrt{\sum_{m=1}^M \sum_{n=1}^N |x_{m,n}|^2}$ denotes the Frobenius norm of matrix $\mathbf{X} = [x_{m,n}]_{m,n=1}^{M,N}$;
- $\langle \mathbf{x}, \mathbf{y} \rangle$ denotes the inner-product between two complex-valued sequences $\mathbf{x} = [x_0, x_1, \dots, x_{N-1}]^T$ and $\mathbf{y} = [y_0, y_1, \dots, y_{N-1}]^T$, i.e., $\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{i=0}^{N-1} x_i y_i^*$, where N is the sequence length of \mathbf{x} (and \mathbf{y});
- $T^\tau(\mathbf{x})$ denotes the right-cyclic-shift of $\mathbf{x} = [x_0, x_1, \dots, x_{N-1}]^T$ for τ (nonnegative integer) positions, i.e.,

$$T^\tau(\mathbf{x}) = [\underbrace{x_{N-\tau}, \dots, x_{N-1}}_{\text{the last } \tau \text{ elements of } \mathbf{x}}, x_0, x_1, \dots, x_{N-\tau-1}]^T.$$

Similarly,

$$T^{-\tau}(\mathbf{x}) = [x_\tau, x_{\tau+1}, \dots, x_{N-1}, \underbrace{x_0, x_1, \dots, x_{\tau-1}}_{\text{the first } \tau \text{ elements of } \mathbf{x}}]^T.$$

- $\underline{\mathbf{x}}$ denotes the reversal of vector $\mathbf{x} = [x_0, x_1, \dots, x_{N-1}]^T$, i.e., $\underline{\mathbf{x}} = [x_{N-1}, x_{N-2}, \dots, x_0]^T$;
- $\text{Diag}[\mathbf{x}]$ denotes a diagonal matrix with the diagonal vector of \mathbf{x} and zero for all off-diagonal entries;
- $\mathbf{1}_{m \times n}$ and $\mathbf{0}_{m \times n}$ denote an all-1 matrix and an all-0 matrix, respectively, both having matrix order of $m \times n$;
- \mathbf{I}_N denotes the identity matrix of order N ;

¹whose duration should be not smaller than the largest multipath delay.

- \otimes denotes the Kronecker product operator;
- $[j]_J$ denotes the modulo J operation of integer j , e.g., $[4]_4 = 0$, $[5]_4 = 1$;
- Denote $\omega_q = \exp\left(\frac{2\pi\sqrt{-1}}{q}\right)$ (integer q);
- \mathbb{Z}_q denotes the set of integers modulo q , i.e., $\mathbb{Z}_q = \{0, 1, \dots, q-1\}$ (q integer);
- $\mathcal{A}_q = \{\omega_q^0, \omega_q^1, \dots, \omega_q^{q-1}\}$ denotes the set over q complex roots of unity.
- $\mathbb{E}(x)$ denotes the mean of random variable x ;
- $\text{Tr}(\mathbf{X})$ denotes the trace of square matrix \mathbf{X} .

For two length- N complex-valued sequences

$$\mathbf{a} = [a_0, a_1, \dots, a_{N-1}]^T, \quad \mathbf{b} = [b_0, b_1, \dots, b_{N-1}]^T,$$

denote by $\rho(\mathbf{a}, \mathbf{b})(\tau)$ the ACC between \mathbf{a} and \mathbf{b} , i.e.,

$$\rho(\mathbf{a}, \mathbf{b})(\tau) = \begin{cases} \sum_{n=0}^{N-\tau-1} a_n b_{n+\tau}^*, & 0 \leq \tau \leq N-1, \\ \sum_{n=0}^{N+\tau-1} a_{n-\tau} b_n^*, & -(N-1) \leq \tau \leq -1, \\ 0, & |\tau| \geq N. \end{cases} \quad (1)$$

Also, denote by $\phi(\mathbf{a}, \mathbf{b})(\tau)$ the *periodic cross-correlation* (PCC) between \mathbf{a} and \mathbf{b} , i.e.,

$$\phi(\mathbf{a}, \mathbf{b})(\tau) = \sum_{n=0}^{N-1} a_n b_{[n+\tau]_N}^* = \langle \mathbf{a}, T^{-\tau}(\mathbf{b}) \rangle. \quad (2)$$

Clearly, $\rho(\mathbf{a}, \mathbf{b})(\tau) = \rho^*(\mathbf{a}, \mathbf{b})(-\tau)$ and

$$\phi(\mathbf{a}, \mathbf{b})(\tau) = \rho(\mathbf{a}, \mathbf{b})(\tau) + \rho^*(\mathbf{b}, \mathbf{a})(N - \tau). \quad (3)$$

In particular, when $\mathbf{a} = \mathbf{b}$, $\rho(\mathbf{a}, \mathbf{a})(\tau)$ will be sometimes written as $\rho(\mathbf{a})(\tau)$ and called the AAC of \mathbf{a} at time-shift τ . Similarly, $\phi(\mathbf{a}, \mathbf{a})(\tau)$ will be sometimes written as $\phi(\mathbf{a})(\tau)$ and called the periodic auto-correlation (PAC) of \mathbf{a} .

B. Introduction to Golay Complementary Pair (GCP)

Next, we give a brief introduction to GCPs. We will show in Section III that a subset of GCPs are *perfect CZCPs*.

Definition 1: (\mathbf{a}, \mathbf{b}) is called a GCP if their AAC sum equals zero for any non-zero time-shift τ , i.e.,

$$\rho(\mathbf{a})(\tau) + \rho(\mathbf{b})(\tau) = 0, \quad \forall \tau \neq 0. \quad (4)$$

Furthermore, two GCPs (\mathbf{a}, \mathbf{b}) and (\mathbf{c}, \mathbf{d}) are said to be *mutually orthogonal* if they have zero ACC sums for all time-shifts, i.e.,

$$\rho(\mathbf{a}, \mathbf{c})(\tau) + \rho(\mathbf{b}, \mathbf{d})(\tau) = 0, \quad (5)$$

holds for any arbitrary τ . As a matter of fact, $(\mathbf{b}^*, -\mathbf{a}^*)$ is a GCP which is mutually orthogonal to (\mathbf{a}, \mathbf{b}) [12], where \mathbf{b} denotes the reversal of \mathbf{b} .

Definition 2: For $\mathbf{x} = [x_1, x_2, \dots, x_\mu] \in \mathbb{Z}_2^\mu$, a generalized Boolean function (GBF) $g(\mathbf{x})$ is defined as a mapping $g : \{0, 1\}^\mu \rightarrow \mathbb{Z}_q$. Each variable x_i ($i \in \{1, 2, \dots, \mu\}$) in \mathbf{x} may be regarded as a GBF (see *Example 1*). Let $[\kappa_1, \kappa_2, \dots, \kappa_\mu]$ be the binary representation of the (non-negative) integer $\kappa = \sum_{i=1}^\mu \kappa_i 2^{i-1}$, with κ_μ denoting the most significant bit and

$0 \leq \kappa \leq 2^\mu - 1$. Given $g(\mathbf{x})$, define $g_\kappa \triangleq g(\kappa_1, \kappa_2, \dots, \kappa_\mu)$ and its associated sequence

$$\mathbf{g} \triangleq [g_0, g_1, \dots, g_{2^\mu-1}]. \quad (6)$$

Furthermore, we define a complex-valued sequence associated to GBF g as follows.

$$\varphi_q(\mathbf{g}) \triangleq [\omega_q^{g_0}, \omega_q^{g_1}, \dots, \omega_q^{g_{2^\mu-1}}]. \quad (7)$$

It is stressed that each realization of g_κ is obtained by setting $x_1 = \kappa_1, x_2 = \kappa_2, \dots, x_\mu = \kappa_\mu$ into the GBF g , for a specific κ . Hence, if $g = x_t$, where $1 \leq t \leq \mu$, the associated sequence \mathbf{g} is the vector formed by $g_\kappa = \kappa_t$, when κ ranges from 0 to $2^\mu - 1$. In this case, we denote the corresponding \mathbf{g} by \mathbf{x}_t , i.e., $\mathbf{g} = \mathbf{x}_t$. Similarly, if $g = x_{t_1} x_{t_2}$ ($t_1 \neq t_2$), we write $\mathbf{g} = \mathbf{x}_{t_1} \mathbf{x}_{t_2}$, a vector obtained by element-wise multiplication of \mathbf{x}_{t_1} and \mathbf{x}_{t_2} . We need the following definition which will be used in Section III to identify a subset of GCPs as *perfect CZCPs*.

Definition 3: For two GBFs g, h over \mathbb{Z}_q , we denote by $\rho_q(g, h)(\tau)$ the aperiodic correlation of their associated complex-valued sequences, i.e.,

$$\rho_q(g, h)(\tau) \triangleq \rho(\varphi_q(\mathbf{g}), \varphi_q(\mathbf{h}))(\tau). \quad (8)$$

We present the following example to illustrate the GBFs defined above. One can find it useful in understanding the GCP construction in *Lemma 1* below.

Example 1: Let $\mu = 3$ and $q = 4$. The associated sequences of GBFs $1, x_1, x_3, 2x_1x_3 + 1$ are shown in (9). When $\kappa = 2$, for instance, its binary representation is $[0, 1, 0]$. Hence, both the third entries of \mathbf{x}_1 and \mathbf{x}_3 take identical zero. When $\kappa = 4$, its binary representation is $[0, 0, 1]$ and therefore, the fifth entries of \mathbf{x}_1 and \mathbf{x}_3 are zero and one, respectively. $2x_1x_3 + 1$ is obtained from two times the element-wise product between \mathbf{x}_1 and \mathbf{x}_3 , followed by addition with $\mathbf{1}$. As observed from (9), the sequence of GBF x_1 is in fact given by that of κ_1 , when κ ranges from 0 to 7. Similarly, the sequence of GBF x_3 is given by that of κ_3 .

Next, we present in *Lemma 1* the GCP construction proposed by Davis and Jedwab in [9]. *Lemma 1* will be used in *Construction 2* for *perfect CZCPs* having lengths of power of two.

Lemma 1: (Davis-Jedwab Construction of GCP [9]) Let

$$g(\mathbf{x}) \triangleq \frac{q}{2} \sum_{k=1}^{\mu-1} x_{\pi(k)} x_{\pi(k+1)} + \sum_{k=1}^{\mu} w_k x_k + w, \quad (10)$$

where π is a permutation of the set $\{1, 2, \dots, \mu\}$, and $w_k, w \in \mathbb{Z}_q$ (q even integer). Then, for any $w' \in \mathbb{Z}_q$, $\varphi_q(\mathbf{g})$ and $\varphi_q(\mathbf{g} + \frac{q}{2} \mathbf{x}_{\pi(1)} + w' \cdot \mathbf{1})$ form a GCP over \mathbb{Z}_q of length 2^μ .

C. Introduction to SM

We consider a single-carrier SM (SC-SM) system employing N_t TA elements and N_r receive antennas (RAs) over frequency-selective channels. Details of SC-SM and its applications in broadband large-scale antenna systems can be found in [40]. Moreover, we consider a QAM/PSK modulation

$$\begin{aligned}
\begin{bmatrix} \kappa_1 \\ \kappa_2 \\ \kappa_3 \end{bmatrix} &= \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \\
\mathbf{1} &= \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}, \\
\mathbf{x}_1 &= \begin{bmatrix} 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 1 \end{bmatrix}, \\
\mathbf{x}_3 &= \begin{bmatrix} 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}, \\
2\mathbf{x}_1\mathbf{x}_3 + \mathbf{1} &= \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 3 & 1 & 3 & 1 \end{bmatrix}.
\end{aligned} \tag{9}$$

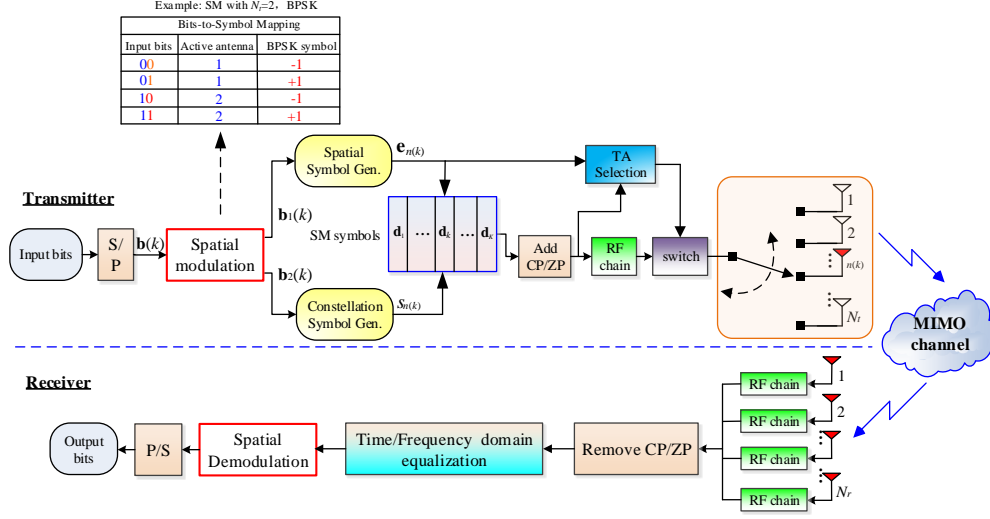


Fig. 1: A generic transceiver structure of SC-SM systems.

with constellation size of \mathcal{M}_{SM} . Fig. 1 shows the block diagram of a generic SC-SM transceiver. For simplicity, assume both $N_t, \mathcal{M}_{\text{SM}}$ take powers of two. Over each time-slot k , $\log_2(N_t \mathcal{M}_{\text{SM}}) = \log_2(N_t) + \log_2(\mathcal{M}_{\text{SM}})$ bits, denoted by vector $\mathbf{b}(k)$, are mapped to an SM symbol \mathbf{d}_k . Specifically, $\log_2(N_t)$ bits, denoted by $\mathbf{b}_1(k)$, are used to activate TA $n(k)$ through one RF chain. Here, we use $\mathbf{e}_{n(k)}$ (which is a sparse vector corresponding to the $n(k)$ -th column of the $N_t \times N_t$ identity matrix) to represent the “spatial symbol” at time-slot k . On the other hand, $\log_2(\mathcal{M}_{\text{SM}})$ bits, denoted by $\mathbf{b}_2(k)$, are used to select one “constellation symbol” $S_{n(k)}$. It is noted that

$$\mathbf{d}_k = S_{n(k)} \mathbf{e}_{n(k)}. \tag{11}$$

Suppose there are K SM symbols (i.e., $[\mathbf{d}_1, \mathbf{d}_2, \dots, \mathbf{d}_K]$) in an SC block. To combat dispersive SM channels, each SC-SM block is extended by a guard interval, which may be a CP, zero prefix (ZP), or known training sequence. Signals collected from N_r RAs are first processed at the receiver by removing the prefix part (in each SC-SM block). Then, time/frequency domain equalization is applied to suppress intersymbol interference (ISI). Subsequently, SM demodulation is carried out to extract spatial symbols and constellation symbols which are combined to retrieve output bits.

Example 2: Consider an SC-SM system with $N_t = 4$ TAs using 8-PSK modulation ($\mathcal{M}_{\text{SM}} = 8$). For illustration purpose, we consider natural mapping for TA selection and 8-PSK

modulation, i.e., each index (for a TA or an 8-PSK phase) can be obtained from its natural binary representation (with the left-most bit being the most significant bit). As defined in Subsection II-A, let $\omega_8 = \exp(\sqrt{-12\pi}/8)$. Suppose each SC-SM block constitutes $K = 4$ SM symbols. Suppose further these symbols correspond to $K \log_2(8N_t) = 20$ message bits [1001011011011000100], with $\mathbf{b}(1) = [10010]$, $\mathbf{b}(2) = [11011]$, $\mathbf{b}(3) = [01110]$, and $\mathbf{b}(4) = [00100]$. Taking the first SM symbol for example, we have $\mathbf{b}(1) = [\mathbf{b}_1(1), \mathbf{b}_2(1)]$, where $\mathbf{b}_1(1) = [10]$ and $\mathbf{b}_2(1) = [010]$. This means that during the first time-slot, only the third TA will be activated for the sending of 8-PSK symbol ω_8^2 . Therefore, the first SM symbol can be written as $\mathbf{d}_1 = [0, 0, \omega_8^2, 0]^T$. The entire SC-SM block can be expressed by the sparse matrix as follows.

$$[\mathbf{d}_1, \mathbf{d}_2, \mathbf{d}_3, \mathbf{d}_4] = \begin{bmatrix} 0 & 0 & 0 & \omega_8^4 \\ 0 & 0 & \omega_8^6 & 0 \\ \omega_8^2 & 0 & 0 & 0 \\ 0 & \omega_8^3 & 0 & 0 \end{bmatrix}. \tag{12}$$

III. CROSS Z-COMPLEMENTARY PAIRS: PROPERTIES AND CONSTRUCTIONS

This section studies the properties and constructions of CZCPs. Throughout this paper, we focus on polyphase q -ary CZCPs whose entries are drawn from \mathcal{A}_q (see Subsection II-A for its definition). Before proceeding further, we formally define CZCPs as follows.

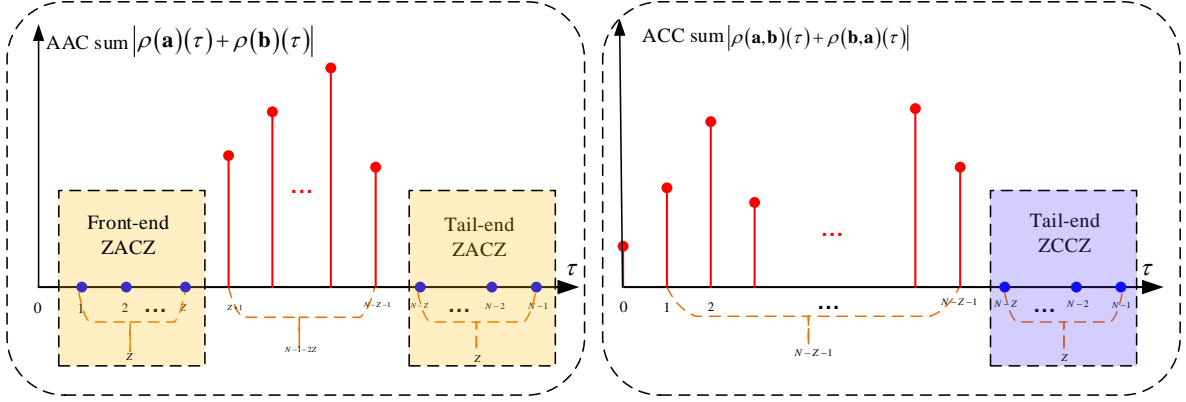


Fig. 2: Illustrative plots for the correlation properties of (N, Z) -CZCP

Definition 4: Let (\mathbf{a}, \mathbf{b}) be a pair of sequences of identical length N . For a proper integer Z , define $\mathcal{T}_1 \triangleq \{1, 2, \dots, Z\}$ and $\mathcal{T}_2 \triangleq \{N-Z, N-Z+1, \dots, N-1\}$. (\mathbf{a}, \mathbf{b}) is called an (N, Z) -CZCP if it possesses *symmetric* zero (out-of-phase) AAC sums for time-shifts over $\mathcal{T}_1 \cup \mathcal{T}_2$ and zero ACC sums for time-shifts over \mathcal{T}_2 . In short, an (N, Z) -CZCP needs to satisfy the following two conditions.

- C1 : $\rho(\mathbf{a})(\tau) + \rho(\mathbf{b})(\tau) = 0$, for all $|\tau| \in \mathcal{T}_1 \cup \mathcal{T}_2$;
 C2 : $\rho(\mathbf{a}, \mathbf{b})(\tau) + \rho(\mathbf{b}, \mathbf{a})(\tau) = 0$, for all $|\tau| \in \mathcal{T}_2$.

From *Definition 4*, C1 shows that each CZCP needs to have two zero autocorrelation zones (ZACZs) when its AAC sums are considered. In this paper, we call them “front-end ZACZ” and “tail-end ZACZ” for time-shifts over \mathcal{T}_1 and \mathcal{T}_2 , respectively. On the other hand, C2 shows that each CZCP needs to have “tail-end zero crosscorrelation zone (ZCCZ)” when its ACC sums are considered. We illustrate the correlation properties of (N, Z) -CZCP in Fig. 2. An example of quaternary $(9, 3)$ -CZCP is given below.

Example 3: Consider the length-9 quaternary pair (\mathbf{a}, \mathbf{b}) below.

$$\mathbf{a} = \omega_4^{[0,1,1,2,0,2,1,1,3]}, \quad \mathbf{b} = \omega_4^{[0,1,1,0,1,0,3,3,1]}. \quad (14)$$

(\mathbf{a}, \mathbf{b}) is a $(9, 3)$ -CZCP because

$$\begin{aligned} & \left(\left| \rho(\mathbf{a})(\tau) + \rho(\mathbf{b})(\tau) \right| \right)_{\tau=0}^8 \\ &= (18, \mathbf{0}_{1 \times 3}, 2\sqrt{2}, 2, \mathbf{0}_{1 \times 3}), \\ & \left(\left| \rho(\mathbf{a}, \mathbf{b})(\tau) + \rho(\mathbf{b}, \mathbf{a})(\tau) \right| \right)_{\tau=0}^8 \\ &= (4, 4\sqrt{2}, 2\sqrt{2}, 2\sqrt{2}, 4, 2, \mathbf{0}_{1 \times 3}). \end{aligned} \quad (15)$$

Several main properties of CZCPs are presented below².

P1: Every q -ary (N, Z) -CZCP (\mathbf{a}, \mathbf{b}) is equivalent to (N, Z) -CZCP (\mathbf{c}, \mathbf{d}) by dividing \mathbf{a} by a_0 and \mathbf{b} by b_0 , respectively, i.e., $\mathbf{c} = \mathbf{a}/a_0, \mathbf{d} = \mathbf{b}/b_0$, where the latter CZCP satisfies

$$c_i = d_i, c_{N-1-i} = -d_{N-1-i}, \quad \text{for all } i \in \{0, 1, \dots, Z-1\}. \quad (16)$$

²Their proofs can be found in the Appendix of this work.

By utilizing (16), one can readily show that $Z \leq N/2$.

P2: If (\mathbf{a}, \mathbf{b}) is an (N, Z) -CZCP, so are $(c_1 \mathbf{b}, c_2 \mathbf{a})$, $(c_1 \underline{\mathbf{b}}, c_2 \underline{\mathbf{a}})$, $(c_1 \underline{\mathbf{b}}, c_2 \underline{\mathbf{a}})$, where $c_1, c_2 \in \mathcal{A}_q$. Moreover, if (16) is satisfied by (\mathbf{a}, \mathbf{b}) , i.e., $a_i = b_i, a_{N-1-i} = -b_{N-1-i}$, for all $i \in \{0, 1, \dots, Z-1\}$, we have

$$\begin{aligned} \rho(\mathbf{a}, \underline{\mathbf{b}})(\tau) + \rho(\mathbf{b}, -\underline{\mathbf{a}})(\tau) &= 0, \quad \text{for all } \tau. \\ \rho(\mathbf{b}, \underline{\mathbf{b}})(\tau) + \rho(\mathbf{a}, -\underline{\mathbf{a}})(\tau) &= 0, \quad \text{for all } |\tau| \in \mathcal{T}_2. \end{aligned} \quad (17)$$

P3: A binary (N, Z) -CZCP (\mathbf{a}, \mathbf{b}) over $\{-1, 1\}$ should have even sequence length N and satisfy the following equation.

$$a_i + a_{N-1-i} + b_i + b_{N-1-i} = \pm 2, \quad \text{for all } i \in \{0, 1, \dots, Z-1\}. \quad (18)$$

Definition 5: By P1, every q -ary CZCP (\mathbf{a}, \mathbf{b}) is called *perfect* if $Z = N/2$ (N even). In this case, a perfect $(N, N/2)$ -CZCP reduces to a sequence pair, called *strengthened GCP*, whose equivalent CZCP (\mathbf{c}, \mathbf{d}) (see P1) is given in (19).

We illustrate P2 by the following example.

Example 4: In the context of the quaternary $(9, 3)$ -CZCP (\mathbf{a}, \mathbf{b}) in *Example 3*, we have

$$\underline{\mathbf{b}}^* = \omega_4^{[3,1,1,0,3,0,3,3,0]}, \quad -\underline{\mathbf{a}}^* = \omega_4^{[3,1,1,0,2,0,1,1,2]}. \quad (20)$$

In addition to $\rho(\mathbf{a})(\tau) + \rho(\mathbf{b})(\tau)$ and $\rho(\mathbf{a}, \mathbf{b})(\tau) + \rho(\mathbf{b}, \mathbf{a})(\tau)$ shown in *Example 3*, we have

$$\begin{aligned} & \left(\left| \rho(\underline{\mathbf{b}}^*)(\tau) + \rho(-\underline{\mathbf{a}}^*)(\tau) \right| \right)_{\tau=0}^8 \\ &= (18, \mathbf{0}_{1 \times 3}, 2\sqrt{2}, 2, \mathbf{0}_{1 \times 3}), \\ & \left(\left| \rho(\mathbf{a}, \underline{\mathbf{b}}^*)(\tau) + \rho(\mathbf{b}, -\underline{\mathbf{a}}^*)(\tau) \right| \right)_{\tau=0}^8 \\ &= (\mathbf{0}_{1 \times 9}), \\ & \left(\left| \rho(\mathbf{b}, \underline{\mathbf{b}}^*)(\tau) + \rho(\mathbf{a}, -\underline{\mathbf{a}}^*)(\tau) \right| \right)_{\tau=0}^8 \\ &= (2\sqrt{13}, 2\sqrt{26}, 2\sqrt{2}, 2\sqrt{2}, 2\sqrt{10}, 4, \mathbf{0}_{1 \times 3}). \end{aligned} \quad (21)$$

Note that the *even* value constraint on sequence length N in P3 may not be necessary for q -ary CZCPs with $q > 2$. In other words, there exist q -ary (N, Z) -CZCPs with odd N provided that $Z \leq N/2$ (see *Example 3*). One can verify (16) in P1 by

Construction 2: In the context of *Lemma 1*, let $\pi(1) = \mu$ and $w' \in \{0, q/2\} \pmod{q}$. Then, $\varphi_q(\mathbf{g})$ and $\varphi_q(\mathbf{g} + \frac{q}{2}\mathbf{x}_{\pi(1)} + w' \cdot \mathbf{1})$ form a perfect q -ary CZCP of length 2^μ .

Proof 2: Let $\mathbf{g}_1, \mathbf{g}_2$, each having length $2^{\mu-1}$, be the first- and second- halves of sequence \mathbf{g} , respectively. Similarly, we define $\mathbf{g}'_1, \mathbf{g}'_2$ for that of sequence $\mathbf{g} + \frac{q}{2}\mathbf{x}_{\pi(1)} + w' \cdot \mathbf{1}$. Hence,

$$\mathbf{g} = [\mathbf{g}_1, \mathbf{g}_2], \quad \mathbf{g} + \frac{q}{2}\mathbf{x}_{\pi(1)} + w' \cdot \mathbf{1} = [\mathbf{g}'_1, \mathbf{g}'_2]. \quad (26)$$

Denote by g_1, g_2, g'_1, g'_2 the GBFs corresponding to $\mathbf{g}_1, \mathbf{g}_2, \mathbf{g}'_1, \mathbf{g}'_2$, respectively. By setting $x_{\pi(1)} = 0$ and 1 into (10), and noting that $\pi(1) = \mu$, we obtain the GBFs corresponding to the first- and second- halves of \mathbf{g} , respectively. Specifically, we have

$$\begin{aligned} g_1 &= \frac{q}{2} \sum_{k=2}^{\mu-1} x_{\pi(k)} x_{\pi(k+1)} + \sum_{k=1}^{\mu-1} w_k x_k + w, \\ g_2 &= g_1 + \frac{q}{2} x_{\pi(2)} + w_\mu. \end{aligned} \quad (27)$$

In a similar way, we have

$$\begin{aligned} g'_1 &= g_1 + w', \\ g'_2 &= g_1 + \frac{q}{2} x_{\pi(2)} + w_\mu + w' + q/2. \end{aligned} \quad (28)$$

To show that $\varphi_q(\mathbf{g})$ and $\varphi_q(\mathbf{g} + \frac{q}{2}\mathbf{x}_{\pi(1)} + w' \cdot \mathbf{1})$ form a perfect CZCP, we only need to check if the second condition of (13) can be satisfied. Let us recall *Definition 3* for aperiodic correlation of complex-valued sequences associated to GBFs. Consider $2^{\mu-1} \leq \tau < 2^\mu$ and then carry out the calculation below.

$$\begin{aligned} & \rho \left(\varphi_q(\mathbf{g}), \varphi_q(\mathbf{g} + \frac{q}{2}\mathbf{x}_{\pi(1)} + w' \cdot \mathbf{1}) \right) (\tau) \\ & + \rho \left(\varphi_q(\mathbf{g} + \frac{q}{2}\mathbf{x}_{\pi(1)} + w' \cdot \mathbf{1}), \varphi_q(\mathbf{g}) \right) (\tau) \\ & = \rho_q \left(g, g + \frac{q}{2}x_{\pi(1)} + w' \right) (\tau) + \rho_q \left(g + \frac{q}{2}x_{\pi(1)} + w', g \right) (\tau) \\ & = \rho_q \left(\underbrace{g_1, g_1 + \frac{q}{2}x_{\pi(2)} + w_\mu + w' + q/2}_{=g'_2} \right) (\tau - 2^{\mu-1}) \\ & \quad + \rho_q \left(\underbrace{g_1 + w', g_1 + \frac{q}{2}x_{\pi(2)} + w_\mu}_{=g'_1} \right) (\tau - 2^{\mu-1}) \\ & = -\omega_q^{-w'} \cdot \rho_q \left(g_1, g_1 + \frac{q}{2}x_{\pi(2)} + w_\mu \right) (\tau - 2^{\mu-1}) \\ & \quad + \omega_q^{w'} \cdot \rho_q \left(g_1, g_1 + \frac{q}{2}x_{\pi(2)} + w_\mu \right) (\tau - 2^{\mu-1}) \\ & = \left[-\omega_q^{-w'} + \omega_q^{w'} \right] \rho_q \left(g_1, g_1 + \frac{q}{2}x_{\pi(2)} + w_\mu \right) (\tau - 2^{\mu-1}). \end{aligned} \quad (29)$$

Clearly, the above equation reduces to zero as $\left[-\omega_q^{-w'} + \omega_q^{w'} \right] = 0$ holds for $w' \in \{0, q/2\}$.

We illustrate *Construction 2* by the example below.

Example 6: Consider an quaternary CZCP of length-16, i.e., $q = 4, \mu = 4$. Furthermore, let us set $\pi =$

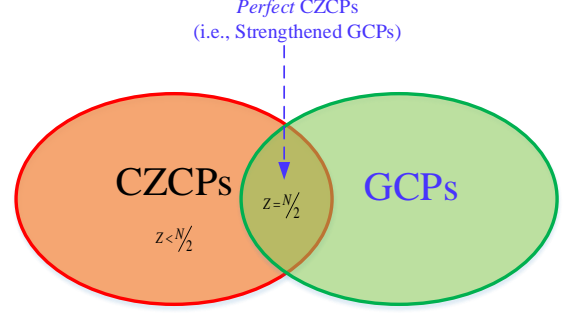


Fig. 3: Relationship between (N, Z) -CZCPs and GCPs.

$[4, 2, 3, 1], [w_1, w_2, w_3, w_4] = [3, 2, 0, 1], w = 0, w' = 2$. By *Lemma 1*, we obtain a GCP below.

$$\begin{aligned} \mathbf{a} &= \omega_4^{[0, 3, 2, 1, 0, 1, 0, 1, 1, 0, 1, 0, 1, 2, 3, 0]}, \\ \mathbf{b} &= \omega_4^{[2, 1, 0, 3, 2, 3, 2, 3, 1, 0, 1, 0, 1, 2, 3, 0]}. \end{aligned} \quad (30)$$

We assert that the above GCP is a perfect CZCP by checking the magnitudes of their AAC sum and ACC sum as follows:

$$\begin{aligned} & \left| \rho(\mathbf{a})(\tau) + \rho(\mathbf{b})(\tau) \right|_{\tau=0}^{15} \\ & = (32, \mathbf{0}_{1 \times 15}), \\ & \left| \rho(\mathbf{b}, \mathbf{a})(\tau) + \rho(\mathbf{a}, \mathbf{b})(\tau) \right|_{\tau=0}^{15} \\ & = (0, 12, 0, 4, 0, 4, 0, 4, 0, 4, \mathbf{0}_{1 \times 8}). \end{aligned} \quad (31)$$

Remark 1: Note that binary GCPs are only known to exist for lengths $2^{\alpha_1} 10^{\alpha_2} 26^{\alpha_3}$ [12], [13], where $\alpha_1, \alpha_2, \alpha_3$ are non-negative integers satisfying $\alpha_1 + \alpha_2 + \alpha_3 \geq 1$. By *Construction 2*, one can see that *perfect CZCPs* (i.e., strengthened GCPs) exist for lengths of $2^{(\alpha_1+1)} 10^{\alpha_2} 26^{\alpha_3}$ where $\alpha_1 + \alpha_2 + \alpha_3 \geq 0$.

Remark 2: *Construction 2* may be viewed as a GBF interpretation of *Construction 1* for perfect CZCPs having lengths of power of two. By counting all the possible permutations and linear coefficients of the GBF in (10), one can readily show that *Construction 2* produces $2 \cdot \frac{(\mu-1)!}{2} \cdot q^{\mu+1} = (\mu-1)! \cdot q^{\mu+1}$ perfect CZCPs.

We illustrate the relationship between CZCPs and GCPs by Fig. 3. It can be seen that CZCPs and GCPs are two different sets of sequence pairs whose intersection is given by *perfect CZCPs* (i.e., strengthened GCPs) with $Z = N/2$. Both CZCPs and GCPs are defined by their aperiodic correlation sums. But CZCPs are different from GCPs in that the former may not necessarily have zero AAC sums for *all* the non-zero time-shifts; instead, they have zero ACC sums for certain time-shifts away from the in-phase position.

We point it out that the concept of CZCP can be extended to “cross Z-complementary set (CZCS)” as follows.

Definition 6: Let $\mathcal{A} = \{\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_M\}$ be a set of sequences of identical length N . Recall $\mathcal{T}_1, \mathcal{T}_2$ which are defined in *Definition 4* based on N and an integer $Z \leq N/2$. \mathcal{A} is called an (N, Z) -CZCS if it possesses *symmetric zero* (out-of-phase) AAC sums for time-shifts over $\mathcal{T}_1 \cup \mathcal{T}_2$ and zero

ACC sums for time-shifts over \mathcal{T}_2 . In short, an (N, Z) -CZCS needs to satisfy the following two conditions.

$$\begin{aligned} \text{C1 : } \sum_{m=1}^M \rho(\mathbf{a}_m)(\tau) &= 0, \text{ for all } |\tau| \in \mathcal{T}_1 \cup \mathcal{T}_2; \\ \text{C2 : } \sum_{m=1}^M \rho(\mathbf{a}_m, \mathbf{a}_{[m+1]_M})(\tau) &= 0, \text{ for all } |\tau| \in \mathcal{T}_2. \end{aligned} \quad (32)$$

When $M = 2$, a CZCS reduces to a CZCP.

It is noted that an (N, Z) -CZCS $\{\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_M\}$ can be readily constructed through an (N, Z) -CZCP (\mathbf{a}, \mathbf{b}) . As an example, consider even M and let

$$\mathbf{a}_m = \begin{cases} \mathbf{a}, & m \text{ odd}, \\ \mathbf{b}, & m \text{ even}. \end{cases} \quad (33)$$

It is straightforward to see that the first condition of (32) can be met by $\{\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_M\}$. To show the achievability of the second condition of (32), let us calculate $\sum_{m=1}^M \rho(\mathbf{a}_m, \mathbf{a}_{[m+1]_M})(\tau)$ as follows:

$$\begin{aligned} & \sum_{m=1}^M \rho(\mathbf{a}_m, \mathbf{a}_{[m+1]_M})(\tau) \\ &= \sum_{m=1}^{M/2} \rho(\mathbf{a}_{2m-1}, \mathbf{a}_{2m})(\tau) + \sum_{m=1}^{M/2} \rho(\mathbf{a}_{2m}, \mathbf{a}_{[2m+1]_M})(\tau) \\ &= \frac{M}{2} \cdot [\rho(\mathbf{a}, \mathbf{b})(\tau) + \rho(\mathbf{b}, \mathbf{a})(\tau)] \end{aligned} \quad (34)$$

By recalling the second condition of (13) associated to (\mathbf{a}, \mathbf{b}) , we complete the proof.

IV. OPTIMAL TRAINING DESIGN USING CZCPs FOR BROADBAND SM SYSTEMS

In this section, we study optimal training design for broadband SM systems. Based on a generic training-based SC-MIMO transmission structure, we present an SM training framework using regular sparse matrices and derive the correlation properties of the row sequences of such a sparse matrix. Then, we show that the CZCPs (proposed in Section III) can be utilized as a key component in optimal SM training design.

A. Problem Formulation

We consider a generic training-based SC-MIMO transmission structure with N_t TAs as shown in Fig. 4. It subsumes the SC-SM transmission scheme in Fig. 1 as a special case but with emphasis on the training matrix design. We assume a quasi-static frequency-selective channel where the channel impulse response (CIR) from the n -th ($1 \leq n \leq N_t$) transmit antenna to the receiver is denoted by $\mathbf{h}_n = [h_{n,0}, h_{n,1}, \dots, h_{n,\lambda}]^T$, i.e., a discrete length- $(\lambda + 1)$ vector with $h_{n,l}$ ($0 \leq l \leq \lambda$) being the channel coefficient of the l -th path. Every block at a TA is divided into two parts: training sequence followed by data payload. Let $\mathbf{x}_n = [x_{n,0}, x_{n,1}, \dots, x_{n,L-1}]$ be the training sequence transmitted

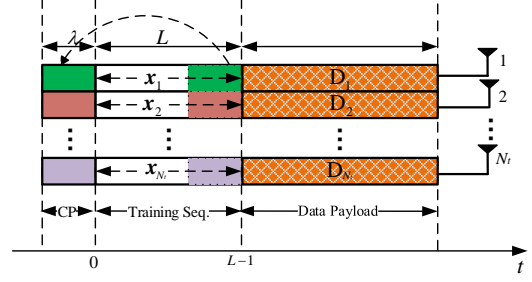


Fig. 4: Generic training-based SC-MIMO transmission structure with N_t transmit antennas.

over the n -th TA ($1 \leq n \leq N_t$). All the training sequences are assumed to have identical energy of E , i.e.,

$$\sum_{l=0}^{L-1} |x_{n,l}|^2 = E, \text{ for all } 1 \leq n \leq N_t. \quad (35)$$

To combat ISI in a frequency-selective channel, a length- λ cyclic prefix (CP)³ is placed at the front of \mathbf{x}_n which is comprised of the last λ elements of \mathbf{x}_n , i.e., $[x_{n,L-\lambda+2}, x_{n,L-\lambda+3}, \dots, x_{n,L-1}]$.

The k -th received signal at a RA can be written as

$$y_k = \sum_{n=1}^{N_t} \sum_{l=0}^{\lambda} h_{n,l} x_{n,k-l} + w_k, \quad (36)$$

where w_k a discrete uncorrelated white complex Gaussian noise sample with zero-mean and variance $\sigma_w^2/2$ per dimension. To proceed, let

$$\mathbf{X}_n = \begin{bmatrix} x_{n,0} & x_{n,L-1} & \cdots & x_{n,L-\lambda} \\ x_{n,1} & x_{n,0} & \cdots & x_{n,L-\lambda+1} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n,L-1} & x_{n,L-2} & \cdots & x_{n,L-\lambda-1} \end{bmatrix}_{L \times (\lambda+1)}, \quad (37)$$

$$\mathbf{h} = \begin{bmatrix} \mathbf{h}_1 \\ \mathbf{h}_2 \\ \vdots \\ \mathbf{h}_{N_t} \end{bmatrix}_{N_t(\lambda+1) \times 1}$$

and

$$\begin{aligned} \mathbf{y} &= [y_0, y_1, \dots, y_{L-1}]^T, \\ \mathbf{X} &= [\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_{N_t}]_{L \times (N_t(\lambda+1))}, \\ \mathbf{w} &= [w_0, w_1, \dots, w_{L-1}]. \end{aligned} \quad (38)$$

Note that \mathbf{X}_n in (37) is a Toeplitz matrix in which every column is a cyclically-shifted version of \mathbf{x}_n and thus has identical energy of E . It can be readily shown that (36) is equivalent to the following matrix equation.

$$\mathbf{y} = \mathbf{X}\mathbf{h} + \mathbf{w}. \quad (39)$$

³CP-free MIMO training may be possible with certain sophisticated signal processing algorithm; however, optimal channel estimation performance in frequency-selective channel may not be straightforward due to ISI.

Applying the least-squares (LS) channel estimator (unbiased), the estimated CIR vector is given by

$$\hat{\mathbf{h}} = (\mathbf{X}^H \mathbf{X})^{-1} \mathbf{X}^H \mathbf{y}, \quad (40)$$

and the normalized MSE can be derived as

$$\begin{aligned} \text{MSE} &= \frac{1}{(N_t \lambda + N_t)} \text{Tr} \left(\mathbb{E} \left\{ (\mathbf{h} - \hat{\mathbf{h}}) (\mathbf{h} - \hat{\mathbf{h}})^H \right\} \right) \\ &= \frac{\sigma_w^2}{(N_t \lambda + N_t)} \text{Tr} \left((\mathbf{X}^H \mathbf{X})^{-1} \right). \end{aligned} \quad (41)$$

(41) indicates that the minimum MSE is achieved if and only if $\mathbf{X}^H \mathbf{X}$ is a diagonal matrix whose elements on the diagonal are identical⁴. Note that

$$\begin{aligned} \mathbf{X}^H \mathbf{X} &= \begin{bmatrix} \mathbf{X}_1^H \mathbf{X}_1 & \mathbf{X}_1^H \mathbf{X}_2 & \cdots & \mathbf{X}_1^H \mathbf{X}_{N_t} \\ \mathbf{X}_2^H \mathbf{X}_1 & \mathbf{X}_2^H \mathbf{X}_2 & \cdots & \mathbf{X}_2^H \mathbf{X}_{N_t} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{X}_{N_t}^H \mathbf{X}_1 & \mathbf{X}_{N_t}^H \mathbf{X}_2 & \cdots & \mathbf{X}_{N_t}^H \mathbf{X}_{N_t} \end{bmatrix}_{(N_t \lambda + N_t) \times (N_t \lambda + N_t)} \\ &= \mathbf{I}_{N_t \lambda + N_t}. \end{aligned} \quad (42)$$

Thus, the minimum MSE is achieved if and only if

$$\mathbf{X}_i^H \mathbf{X}_j = \begin{cases} E \mathbf{I}_{\lambda+1}, & \text{if } i = j, \\ \mathbf{0}_{(\lambda+1) \times (\lambda+1)}, & \text{if } i \neq j, \end{cases} \quad (43)$$

with which we have

$$\text{minimum MSE} = \frac{\sigma_w^2}{E}. \quad (44)$$

Relating (43) to the PCC between \mathbf{x}_i and \mathbf{x}_j , we assert that the minimum MSE is achieved if and only if

$$\phi(\mathbf{x}_i, \mathbf{x}_j)(\tau) = \begin{cases} E, & \text{if } i = j, \tau = 0, \\ 0, & \text{if } i \neq j, 0 \leq \tau \leq \lambda, \\ & \text{or } i = j, 1 \leq \tau \leq \lambda. \end{cases} \quad (45)$$

Remark 3: From here onwards, (45) is referred to as the optimal condition of SM training sequences under LS channel estimator.

B. Proposed Training Framework For Broadband SM Systems

To uncover the optimal training sequence criteria for broadband SM systems, we define the following training matrix Ω .

$$\Omega = \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \vdots \\ \mathbf{x}_{N_t} \end{bmatrix} = \begin{bmatrix} x_{1,0} & x_{1,1} & \cdots & x_{1,L-1} \\ x_{2,0} & x_{2,1} & \cdots & x_{2,L-1} \\ \vdots & \vdots & \ddots & \vdots \\ x_{N_t,0} & x_{N_t,1} & \cdots & x_{N_t,L-1} \end{bmatrix}_{N_t \times L}. \quad (46)$$

Remark 4: In SM system, only one TA is activated over every time-slot, hence Ω should be a sparse matrix in which each column has one non-zero entry only (see *Example 1*). This implies that each training sequence \mathbf{x}_n ($1 \leq n \leq N_t$) has $Q \equiv L/N_t$ non-zero entries and $(N_t - 1)Q$ zeros. In this paper, we are concerned with non-zero entries having identical

magnitude of 1 (i.e., unimodular) and therefore, each training sequence has energy of $E = Q$.

Suppose that Ω has the following *sparse* structure.

$$\Omega = \begin{bmatrix} T^0 (\mathbf{a}_1 \mathbf{0}_{1 \times (N_t-1)Q}) \\ T^Q (\mathbf{a}_2 \mathbf{0}_{1 \times (N_t-1)Q}) \\ \vdots \\ T^{(N_t-1)Q} (\mathbf{a}_{N_t} \mathbf{0}_{1 \times (N_t-1)Q}) \end{bmatrix}_{N_t \times L}, \quad (47)$$

where $\mathbf{a}_n = [a_{n,0}, a_{n,1}, \dots, a_{n,Q}]$ ($1 \leq n \leq N_t$) are row vectors having identical length of Q and energy of E . Noted that Ω is obtained by vertical concatenation of N_t sparse sequences, each of which is a shifted version of a non-zero sequence in $\{\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_{N_t}\}$ padded with $(N_t - 1)Q$ zeros.

An example of training matrix Ω having $N_t = 4$ and $L = 4Q$ is shown below.

$$\Omega = \begin{bmatrix} \mathbf{a}_1 & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{a}_2 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{a}_3 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{a}_4 \end{bmatrix}_{4 \times 4Q}, \quad (48)$$

where $\mathbf{0}$ in (48) denotes $\mathbf{0}_{1 \times Q}$. One can easily show that $\phi(\mathbf{x}_i, \mathbf{x}_j)(1)$ is non-zero and therefore, the condition in (45) can only be satisfied if $\lambda = 0$, i.e., flat-fading channels. This means that the training matrix in (48) cannot achieve minimum channel estimation MSE [see (44)] in a frequency-selective channel.

To circumvent the above problem, we propose training matrix Ω which takes the following structure.

$$\begin{aligned} \Omega &= [\Omega_1, \Omega_2, \dots, \Omega_J], \\ \Omega_j &= \begin{bmatrix} T^0 (\mathbf{a}_1^j \mathbf{0}_{1 \times (N_t-1)\theta}) \\ T^\theta (\mathbf{a}_2^j \mathbf{0}_{1 \times (N_t-1)\theta}) \\ \vdots \\ T^{(N_t-1)\theta} (\mathbf{a}_M^j \mathbf{0}_{1 \times (N_t-1)\theta}) \end{bmatrix}_{N_t \times N_t \theta}, \end{aligned} \quad (49)$$

where $1 \leq j \leq J$, $J \geq 2$, each unimodular row vector $\mathbf{a}_n^j = [a_{n,0}^j, a_{n,1}^j, \dots, a_{n,\theta-1}^j]$ ($1 \leq j \leq J, 1 \leq n \leq N_t$) has length of θ . Clearly $N_t J \theta = L$ and $Q = J \theta = L/N_t$. The rationale is that each TA will send Q non-zero entries over J sub-blocks (where $J \geq 2$) so that they can work in a cooperative way to enable the resultant training matrix to meet the optimal condition specified in (45).

We also define matrix Ψ , which is obtained by removing all the zero vectors and cyclic-shift operators in Ω , as follows.

$$\Psi = \begin{bmatrix} \mathbf{a}_1^1 & \mathbf{a}_1^2 & \cdots & \mathbf{a}_1^J \\ \mathbf{a}_2^1 & \mathbf{a}_2^2 & \cdots & \mathbf{a}_2^J \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{a}_{N_t}^1 & \mathbf{a}_{N_t}^2 & \cdots & \mathbf{a}_{N_t}^J \end{bmatrix}_{N_t \times Q}. \quad (50)$$

One can see that Ω can be one-to-one mapped to Ψ and vice versa. As such, Ψ is called the *characteristic matrix* of the training matrix Ω . In the sequel, we sometimes write the training matrix and characteristic matrix as (N_t, J, θ) - Ω and (N_t, J, θ) - Ψ , respectively.

⁴The proof of this statement can be found in the Appendix of [48].

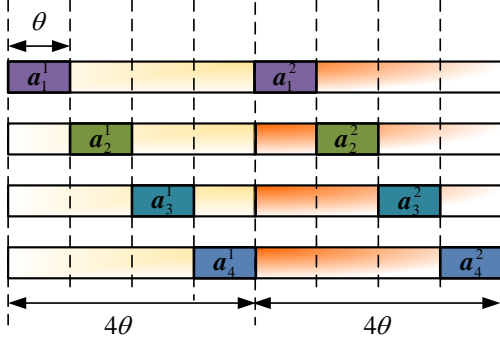


Fig. 5: Training matrix Ω with $N_t = 4, J = 2, L = 2N_t\theta$.

Example 7: Training matrix $(4, 2, \theta)$ - Ω for channel estimation in SM with 4 TAs is shown in Fig. 5. This training matrix and its characteristic matrix can be written as follows.

$$\Omega = \begin{bmatrix} \mathbf{a}_1^1 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{a}_1^2 & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{a}_2^1 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{a}_2^2 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{a}_3^1 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{a}_3^2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{a}_4^1 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{a}_4^2 \end{bmatrix}_{4 \times 8\theta}, \quad (51)$$

$$\Psi = \begin{bmatrix} \mathbf{a}_1^1 & \mathbf{a}_1^2 \\ \mathbf{a}_2^1 & \mathbf{a}_2^2 \\ \mathbf{a}_3^1 & \mathbf{a}_3^2 \\ \mathbf{a}_4^1 & \mathbf{a}_4^2 \end{bmatrix}_{4 \times 2\theta},$$

where $\mathbf{0}$ in Ω stands for $\mathbf{0}_{1 \times \theta}$. Note that in frequency-selective channels, the non-zero component sequences in row 1 and row 2 tend to interfere due to multi-path propagation, likewise for row 2 and row 3, row 3 and row 4, row 4 and row 1, so on and so forth. We will show later that such interference can be mitigated by the “tail-end ZACZ” and “tail-end ZCCZ” properties of a CZCP (see Fig. 2). Also, the ISI of each row can be mitigated by the “front-end ZACZ” property of CZCP. This makes CZCP a key component in the SM training design.

Optimal SM Training Design Criteria: Assume that $\theta \geq \lambda$. Recalling (45), we assert that the optimal SM training sequences need to satisfy the following three equations⁵.

$$\phi(\mathbf{x}_{i+1}, \mathbf{x}_i)(\tau) = \sum_{j=1}^J \rho(\mathbf{a}_{i+1}^j, \mathbf{a}_i^j)(\theta - \tau) = 0, \quad (52)$$

$$1 \leq \tau \leq \lambda, 1 \leq i < N_t,$$

$$\phi(\mathbf{x}_1, \mathbf{x}_{N_t})(\tau) = \sum_{j=1}^J \rho(\mathbf{a}_1^{[j]J+1}, \mathbf{a}_{N_t}^j)(\theta - \tau) = 0, \quad 1 \leq \tau \leq \lambda, \quad (53)$$

and

$$\phi(\mathbf{x}_i)(\tau) = \sum_{j=1}^J \rho(\mathbf{a}_i^j)(\tau) = 0, \quad 1 \leq \tau \leq \lambda, 1 \leq i \leq N_t. \quad (54)$$

⁵Note that $\phi(\mathbf{x}_{i+1}, \mathbf{x}_i)(0) = 0, \forall 1 \leq i < N_t$ and $\phi(\mathbf{x}_1, \mathbf{x}_{N_t})(0) = 0$ can always be ensured by the proposed training matrix structure in (49).

Note that (52)-(53) and (54) specify the PCC conditions and PAC condition for the rows of Ω , respectively. It is assumed that all the non-zero entries of characteristic matrix Ψ are drawn from alphabet set \mathcal{A}_q (see Subsection II-A for its definition). In particular, we are interested in binary ($q = 2$) and quaternary ($q = 4$) characteristic matrices owing to their low implementation complexity in practice. We have the following remark.

Remark 5: For any q -ary characteristic matrix with even q , J should be even. This implies that any binary ($q = 2$) or quaternary ($q = 4$) characteristic matrix should have even J .

Proof 3: Setting $\tau = 1$ in (52), we obtain

$$\phi(\mathbf{x}_{i+1}, \mathbf{x}_i)(1) = \sum_{j=1}^J a_{i+1,0}^j (a_{i,\theta-1}^j)^*. \quad (55)$$

It is noted that the right-hand-side of (55) is a summation of J entries which are drawn from \mathcal{A}_q . When q is even, every entry $a_{i+1,0}^j (a_{i,\theta-1}^j)^*$ can only be cancelled by its negation (one of the J entries in \mathcal{A}_q) to ensure $\phi(\mathbf{x}_{i+1}, \mathbf{x}_i)(1) = 0$. Therefore, J must be even.

Proposed Construction: Consider a $(2, 2, \theta)$ - Ψ “seed” characteristic matrix as follows.

$$\Psi = \begin{bmatrix} \mathbf{a}_1^1 & \mathbf{a}_1^2 \\ \mathbf{a}_2^1 & \mathbf{a}_2^2 \end{bmatrix}_{2 \times 2\theta}. \quad (56)$$

Suppose the *optimal* conditions specified in (52)-(54) are satisfied, i.e.,

$$\rho(\mathbf{a}_2^1, \mathbf{a}_1^1)(\theta - \tau) + \rho(\mathbf{a}_2^2, \mathbf{a}_1^2)(\theta - \tau) = 0, \quad 1 \leq \tau \leq \lambda, \quad (57)$$

$$\rho(\mathbf{a}_1^2, \mathbf{a}_2^1)(\theta - \tau) + \rho(\mathbf{a}_1^1, \mathbf{a}_2^2)(\theta - \tau) = 0, \quad 1 \leq \tau \leq \lambda, \quad (58)$$

and

$$\rho(\mathbf{a}_i^1)(\tau) + \rho(\mathbf{a}_i^2)(\tau) = 0, \quad 1 \leq \tau \leq \lambda, i \in \{1, 2\}. \quad (59)$$

Based on the $(2, 2, \theta)$ - Ψ given in (56), we obtain an enlarged $(N_t, 2, \theta)$ -characteristic matrix ($N_t \geq 2$ even) as follows.

$$\begin{aligned} \tilde{\Psi} &= \mathbf{1}_{N_t/2 \times 1} \otimes \Psi \\ &= \begin{bmatrix} \mathbf{1}_{N_t/2 \times 1} \otimes \mathbf{a}_1^1 & \mathbf{1}_{N_t/2 \times 1} \otimes \mathbf{a}_1^2 \\ \mathbf{1}_{N_t/2 \times 1} \otimes \mathbf{a}_2^1 & \mathbf{1}_{N_t/2 \times 1} \otimes \mathbf{a}_2^2 \end{bmatrix}_{N_t \times 2\theta}. \end{aligned} \quad (60)$$

The corresponding training matrix of $\tilde{\Psi}$ can be expressed as

$$\tilde{\Omega} = \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \vdots \\ \mathbf{x}_{N_t/2} \\ \mathbf{x}_{N_t/2+1} \\ \vdots \\ \mathbf{x}_{N_t} \end{bmatrix}_{N_t \times 2N_t\theta} = \begin{bmatrix} T^0(\mathbf{a}_1^1 \mathbf{0}) & T^\theta(\mathbf{a}_1^1 \mathbf{0}) \\ T^\theta(\mathbf{a}_1^1 \mathbf{0}) & T^{2\theta}(\mathbf{a}_1^1 \mathbf{0}) \\ \vdots & \vdots \\ T^{(N_t/2-1)\theta}(\mathbf{a}_1^1 \mathbf{0}) & T^{(N_t/2)\theta}(\mathbf{a}_1^1 \mathbf{0}) \\ T^{(N_t/2)\theta}(\mathbf{a}_1^1 \mathbf{0}) & T^{(N_t/2+1)\theta}(\mathbf{a}_1^1 \mathbf{0}) \\ \vdots & \vdots \\ T^{(N_t-1)\theta}(\mathbf{a}_1^1 \mathbf{0}) & T^{(N_t)\theta}(\mathbf{a}_1^1 \mathbf{0}) \end{bmatrix}_{N_t \times 2N_t\theta}, \quad (61)$$

where $\mathbf{0}$ stands for $\mathbf{0}_{1 \times (N_t-1)\theta}$. One can see that the *optimal* conditions in (53) and (54) can be satisfied by (58) and (59), respectively. Furthermore, by (57), we have

$$\begin{aligned} & \phi(\mathbf{x}_{N_t/2+1}, \mathbf{x}_{N_t/2})(\tau) \\ &= \rho(\mathbf{a}_1^1, \mathbf{a}_1^1)(\theta - \tau) + \rho(\mathbf{a}_2^2, \mathbf{a}_1^1)(\theta - \tau) \\ &= 0, \quad 1 \leq \tau \leq \lambda. \end{aligned} \quad (62)$$

To ensure that (52) is satisfied for all the i , we also require

$$\phi(\mathbf{x}_{i+1}, \mathbf{x}_i)(\tau) = 0, \quad \forall i \in \{1, 2, \dots, N_t-1\} \setminus \{N_t/2\}, \quad 1 \leq \tau \leq \lambda.$$

This leads to the following *additional* conditions which should be met by the $(2, 2, \theta)$ - Ψ “seed” characteristic matrix:

$$\begin{aligned} & \phi(\mathbf{x}_{i+1}, \mathbf{x}_i)(\tau) = \rho(\mathbf{a}_1^1)(\theta - \tau) + \rho(\mathbf{a}_1^2)(\theta - \tau) \\ &= 0, \quad 1 \leq i \leq N_t/2 - 1, \quad 1 \leq \tau \leq \lambda. \end{aligned} \quad (63)$$

$$\begin{aligned} & \phi(\mathbf{x}_{i+1}, \mathbf{x}_i)(\tau) = \rho(\mathbf{a}_2^1)(\theta - \tau) + \rho(\mathbf{a}_2^2)(\theta - \tau) \\ &= 0, \quad N_t/2 + 1 \leq i \leq N_t - 1, \quad 1 \leq \tau \leq \lambda. \end{aligned} \quad (64)$$

Let (\mathbf{a}, \mathbf{b}) be an $(N = \theta, Z = \lambda)$ -CZCP which is proposed in Section III. Consider two “seed” characteristic matrices taking the following structures:

$$\Psi_1 = \begin{bmatrix} \mathbf{a} & \mathbf{b} \\ \mathbf{a} & \mathbf{b} \end{bmatrix}_{2 \times 2\theta}, \quad \Psi_2 = \begin{bmatrix} \mathbf{a} & \mathbf{b} \\ \mathbf{b}^* & -\mathbf{a}^* \end{bmatrix}_{2 \times 2\theta}. \quad (65)$$

One can readily show that (57)-(59) and (63)-(64) can be satisfied by both Ψ_1 and Ψ_2 . The resultant “seed” characteristic matrices will allow us to design training matrices $(N_t, 2, \theta)$ - $\tilde{\Omega}$ [see (61)] for *optimal* channel estimation in SM system with N_t TAs.

Based on $(N_t, 2, \theta)$ - $\tilde{\Omega}$, we can also construct a training matrix (N_t, J, θ) - $\tilde{\Omega}$ (with longer row sequences and hence larger value of E and enhanced channel estimation performance) by the following expansion rule.

$$\tilde{\Omega} = \mathbf{1}_{1 \times J/2} \otimes \tilde{\Omega}, \quad \text{for } J > 2. \quad (66)$$

Finally, based on (56), (60), (61), (66), we summarize the proposed SM training design as follows.

$$\begin{aligned} & (\theta, \lambda)\text{-CZCP} \mapsto (2, 2, \theta)\text{-}\Psi \mapsto (N_t, 2, \theta)\text{-}\tilde{\Psi} \\ & \mapsto (N_t, 2, \theta)\text{-}\tilde{\Omega} \mapsto (N_t, J, \theta)\text{-}\tilde{\Omega}. \end{aligned} \quad (67)$$

One may choose $(N_t, 2, \theta)$ - $\tilde{\Omega}$ or (N_t, J, θ) - $\tilde{\Omega}$ ($J > 2$) as the SM training matrix depends on the channel estimation MSE requirement which is determined by $E = J\theta$ [see (44)].

C. Numerical Evaluation

In this subsection, we evaluate the proposed SM training sequences over frequency-selective channels. Throughout this subsection, we assume $N_t = 4, N_r = 1$. Consider a $(\lambda + 1)$ -path channel (separated by integer symbol durations) having uniform power delay profile as follows:

$$h[t] = \sum_{n=0}^{\lambda} h_i \delta[t - nT], \quad (68)$$

where h_i 's are complex-valued Gaussian random variables with zero mean and $\mathbb{E}(|h_i|^2) = 1$.

Let us consider the *perfect* ($N = 8, Z = 4$)-CZCP given in Table I. As we will see, (\mathbf{a}, \mathbf{b}) leads to optimal SM training matrix provided that the number of multi-paths is not greater than $Z + 1$ (i.e., 5). Based on the training matrix expansion rule in (61) and applying seed characteristic matrices Ψ_1 and Ψ_2 given in (65), respectively, we obtain training matrices $(4, 2, 8)$ - $\tilde{\Omega}_1$ and $(4, 2, 8)$ - $\tilde{\Omega}_2$ as follows.

$$\begin{aligned} \tilde{\Omega}_1 &= \begin{bmatrix} \mathbf{a} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{b} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{a} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{b} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{a} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{b} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{a} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{b} \end{bmatrix}, \\ \tilde{\Omega}_2 &= \begin{bmatrix} \mathbf{a} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{b} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{a} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{b} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{b}^* & \mathbf{0} & \mathbf{0} & \mathbf{0} & -\mathbf{a}^* & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{b}^* & \mathbf{0} & \mathbf{0} & \mathbf{0} & -\mathbf{a}^* \end{bmatrix}, \end{aligned} \quad (69)$$

where $\mathbf{0}$ in (69) stands for $\mathbf{0}_{1 \times 8}$. Taking one SM training matrix in (69) and following the expansion rule in (66), we can obtain $(4, 6, 8)$ - $\tilde{\Omega}$ and $(4, 18, 8)$ - $\tilde{\Omega}$. As a comparison, we consider binary *random* (but regular and sparse) training matrices in which each row (of a training matrix) contains $2\lambda J$, where $J \in \{2, 6, 18\}$, non-zero entries and each column consists of one non-zero entry only. Note that a larger value of J leads to higher value of training sequence energy E and hence better channel estimation performance⁶. A new random training matrix is generated for every different SM channel estimation (i.e., random on-the-fly). Following Section IV, LS channel estimator is employed. Fig. 6-a compares the channel estimation MSEs versus “EbNo (dB) per TA” for the above two types of training sequences (where $J = 2, 6, 18$) with frequency-selective fading channel consisting of 5 multi-paths. It is seen that the MSE curves obtained from the proposed

⁶ In practice, however, a longer training sequence will give rise to a higher training overhead. Hence, selection of the training length (determined by J) is a trade-off between channel estimation performance and training overhead.

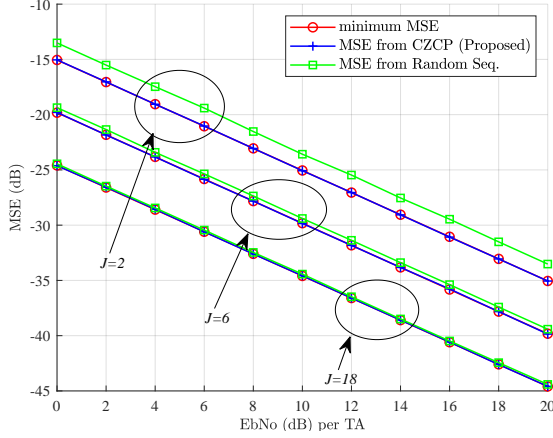
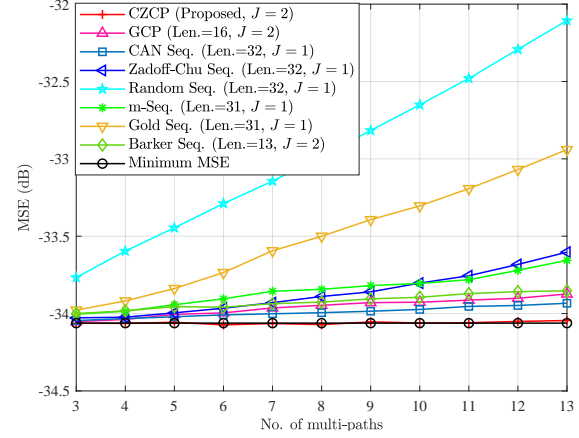
(a) No. of multi-paths = 5, $L = 32J$, $J \in \{2, 6, 18\}$, $Z = 4$.(b) EbNo=16 dB and $E = 32$.

Fig. 6: MSE comparison with training matrices from other sequences.

SM training sequences match with the minimum MSE curves very well for different values of J . Compared to the MSE curve obtained from random training sequences, about 1.5 dB gain is achieved for $J = 2$. When J increases, the MSE gap between these types of sequences diminishes. This is because the diagonal elements (i.e., taking identical value of $E = 8J$) of $\mathbf{X}^H \mathbf{X}$ in (41) become significantly larger than the off-diagonal ones. Consequently, $\mathbf{X}^H \mathbf{X}$ approaches to an identity matrix as the proposed SM training sequences can achieve.

Next, we evaluate the channel estimation MSE performances under different values of multi-paths at EbNo of 16 dB. We employ the *perfect* ($N = 16$, $Z = 8$)-CZCP given in Table I to generate our proposed SM training matrix with the same structure as $\tilde{\Omega}_1$ in (69), i.e., $J = 2$. We compare its channel estimation performance with SM training matrices from the length-16 GCP (which is not a CZCP)

$$\begin{bmatrix} 1, 1, 1, 1, 1, -1, -1, 1, 1, 1, -1, -1, 1, -1, 1, -1 \\ 1, -1, 1, -1, 1, 1, -1, -1, 1, -1, -1, 1, 1, 1, 1, 1 \end{bmatrix},$$

the length-31 m-sequence⁷

$$\mathbf{a}_m = [1, -1, -1, -1, -1, 1, -1, -1, 1, -1, 1, 1, -1, -1, 1, 1, 1, 1, -1, -1, -1, 1, 1, -1, 1, 1, -1, 1, -1, 1],$$

one length-32 CAN sequence⁸ with low aperiodic autocorrelations, the length-13 Barker sequence⁹

$$\mathbf{a}_B = [1, 1, 1, 1, 1, -1, -1, 1, 1, -1, 1, -1, 1],$$

four length-31 Gold sequences [60] with low periodic cross-correlations, four length-32 Zadoff-Chu sequences [61] with

low periodic cross-correlations, and four “random-on-the-fly” binary sequences of length-32. For the m-sequence, Gold sequences, Zadoff-Chu sequences, CAN sequence and random sequences, we adopt the training matrix structure in (48). In particular, when a single sequence (i.e., m-sequence or CAN sequence) is used as the “seed”, we set $\mathbf{a}_1 = \mathbf{a}_2 = \mathbf{a}_3 = \mathbf{a}_4$ in (48). For the length-13 Barker sequence, we adopt the following training matrix

$$\begin{bmatrix} \mathbf{a}_B & 0 & 0 & 0 & \mathbf{a}_B & 0 & 0 & 0 \\ 0 & \mathbf{a}_B & 0 & 0 & 0 & \mathbf{a}_B & 0 & 0 \\ 0 & 0 & \mathbf{a}_B & 0 & 0 & 0 & \mathbf{a}_B & 0 \\ 0 & 0 & 0 & \mathbf{a}_B & 0 & 0 & 0 & \mathbf{a}_B \end{bmatrix}_{4 \times 104}.$$

For fair comparison, we normalize the energies of all the training sequences to $E = 32$. When the number of multi-paths is 9 (or less), $\mathbf{X}^H \mathbf{X}$ corresponding to our proposed SM training matrix is an identity matrix multiplied by the training matrix energy and therefore, our proposed SM training matrix achieves the minimum MSEs, as shown in Fig. 6-b. The same may not be achieved by training matrices from other seed sequences. As a result, their channel estimation MSEs display certain distances to the minimum MSE regardless the number of multi-paths. It is interesting to note that the second best channel estimation performance is achieved by the SM training matrix from the length-32 CAN sequence (even better than that from the length-13 Barker sequence) whose aperiodic ISL is minimized. This makes sense because a sequence with lower aperiodic ISL tends to give rise to smaller magnitudes for the off-diagonal elements of $\mathbf{X}^H \mathbf{X}$ and hence better channel estimation MSEs. The worst channel estimation performance is achieved by the SM training matrices from random sequences as it is in general hard to ensure small off-diagonal elements of $\mathbf{X}^H \mathbf{X}$. The second worst channel estimation performance is achieved by the SM training matrices from Gold sequences. Moreover, the channel estimation MSEs corresponding to our proposed SM training matrix start to deviate from, but very close to, the minimum MSE when the number of multi-paths increases to beyond 9. In summary, our proposed SM training matrices give rise to the minimum channel estimation MSEs

⁷An m-sequence is a pseudo-random sequence whose periodic autocorrelation sidelobes take identical value of -1 .

⁸CAN sequences are obtained by minimizing a quadratic function associated with the integrated sidelobe level (ISL) of sequences [52]. With the aid of fast Fourier transform (FFT), CAN sequences can be generated for lengths up to 10^6 or even longer. More sequence synthesis algorithms based on numerical optimization can be found in [53]–[59].

⁹A Barker sequence has aperiodic autocorrelation sidelobes not exceeding 1. It is widely conjectured that no binary Barker sequences exist for lengths larger than 13 [51] and hence the application of Barker sequences may not be straightforward when a large sequence length is required.

provided that the number of multi-paths is no greater than $Z + 1$; when greater than $Z + 1$, our proposed SM training matrices may still give rise to good channel estimation MSEs in the condition that 1) $\mathbf{X}^H \mathbf{X}$ associated to the SM training matrix has full rank and 2) the off-diagonal elements of $\mathbf{X}^H \mathbf{X}$ have small magnitudes.

V. CONCLUSIONS

In this paper, we have proposed a new class of sequence pairs called cross-complementary pairs (CZCPs) in which every CZCP displays zero (symmetrical) aperiodic auto-correlation (AAC) sums and zero aperiodic cross-correlation (ACC) sums for certain time-shifts. Unlike Golay complementary pairs (GCPs) whose two constituent sequences must be sent over two non-interfering channels (thus only AAC sums matter in GCPs), *perfect* CZCPs should be designed to minimize both ISI (determined by AAC sums) and cross-interference (by ACC sums). By investigating the structural properties of CZCPs, we have found that *perfect* CZCPs are equivalent to a subset of GCPs whose first halves are identical and second halves have opposite polarities (See P2 in Section III). Based on this finding, we have shown by systematic constructions that *perfect* binary CZCPs exist for lengths $2^{\alpha+1}10^\beta 26^\gamma$, where α, β, γ are non-negative integers. In *Definition 6*, we have extended the concept of CZCP to cross Z-complementary set (CZCS) which consists of constituent sequences of two or more.

Secondly, we have shown that CZCPs play an important role in channel training sequence design of broadband spatial modulation (SM) systems. It should be noted that the existing dense training sequences for conventional MIMO systems are unapplicable in SM systems as only one transmit antenna (TA) is activated at each time-slot. By employing CZCP, we have presented a generic framework for the design of *optimal* SM training matrix. We have shown that these training matrices lead to minimum channel estimation mean-squared-error in quasi-static frequency-selective channels.

As a future work, more systematic constructions (e.g., recursive expansion algorithms) for CZCPs/CZCSs (perfect or non-perfect) may be designed. When a *perfect* CZCP/CZCS cannot be obtained, it will be interesting to know what are the almost-perfect CZCPs/CZCSs with respect to different sequence lengths and alphabet sizes. Efficient training design for generalized SM [62]–[64] in dispersive channels is also an interesting and challenging issue to be explored. In particular, it is worthy to investigate efficient training sequence design when the generalized SM channels are correlated [65]–[69].

ACKNOWLEDGMENT

The authors are deeply indebted to Prof. Pingzhi Fan at Southwest Jiaotong University for suggesting the terminology of “CZCP”. They also would like to thank anonymous reviewers for their invaluable suggestions which greatly help improve the quality of this work.

APPENDIX A PROOFS OF CZCP PROPERTIES

Proof of P1:

Proof 4: Since (\mathbf{c}, \mathbf{d}) is obtained by respectively dividing \mathbf{a} by a_0 and \mathbf{b} by b_0 , we have $\mathbf{c} = \mathbf{a}/a_0, \mathbf{d} = \mathbf{b}/b_0$. Therefore,

$$\rho(\mathbf{c})(\tau) + \rho(\mathbf{d})(\tau) = \frac{1}{|a_0|^2} \rho(\mathbf{a})(\tau) + \frac{1}{|b_0|^2} \rho(\mathbf{b})(\tau). \quad (70)$$

For q -ary CZCP (\mathbf{a}, \mathbf{b}) , $a_0, b_0 \in \mathcal{A}_q$ and hence $|a_0|^2 = |b_0|^2 = 1$. Thus,

$$\rho(\mathbf{c})(\tau) + \rho(\mathbf{d})(\tau) = \rho(\mathbf{a})(\tau) + \rho(\mathbf{b})(\tau). \quad (71)$$

Furthermore, since $\rho(\mathbf{a})(N-1) + \rho(\mathbf{b})(N-1) = \rho(\mathbf{a}, \mathbf{b})(N-1) + \rho(\mathbf{a}, \mathbf{b})(N-1) = 0$, we have

$$\begin{aligned} a_0 a_{N-1}^* + b_0 b_{N-1}^* &= 0, \\ a_0 b_{N-1}^* + b_0 a_{N-1}^* &= 0. \end{aligned} \quad (72)$$

Multiplying the above two identities by b_0^*, a_0^* , respectively, we assert that $a_0 b_0^* = a_0^* b_0$ which is a real number. As a result, we have

$$\begin{aligned} \rho(\mathbf{c}, \mathbf{d})(\tau) + \rho(\mathbf{d}, \mathbf{c})(\tau) &= \frac{1}{b_0^* a_0} \rho(\mathbf{a}, \mathbf{b})(\tau) + \frac{1}{b_0 a_0^*} \rho(\mathbf{b}, \mathbf{a})(\tau) \\ &= \frac{1}{b_0 a_0^*} [\rho(\mathbf{a}, \mathbf{b})(\tau) + \rho(\mathbf{b}, \mathbf{a})(\tau)]. \end{aligned} \quad (73)$$

Recalling the two conditions specified in (13), (71) and (73) indicate that (\mathbf{c}, \mathbf{d}) is also q -ary (N, λ) -CZCP (with $c_0 = d_0 = 1$).

By $\rho(\mathbf{c})(N-1) + \rho(\mathbf{d})(N-1) = 0$, we have

$$c_{N-1} + d_{N-1} = 0. \quad (74)$$

Based on this (i.e., $c_0 = d_0, c_{N-1} = -d_{N-1}$), we next carry out an induction for the proof of (16). Suppose that we have $c_i = d_i, c_{N-1-i} = -d_{N-1-i}$ for all $i \in \{0, 1, \dots, k-1\}$ where $k \leq Z-1$. Next, we calculate the aperiodic correlation sums at $\tau = N-k-1$ which involves the correlations of the following vectors.

$$\begin{bmatrix} \mathbf{c}_1 \\ \mathbf{c}_2 \\ \mathbf{d}_1 \\ \mathbf{d}_2 \end{bmatrix} = \begin{bmatrix} 1 & c_1 & c_2 & \cdots & c_{k-1} & c_k \\ c_{N-k-1} & c_{N-k} & c_{N-k+1} & \cdots & c_{N-2} & c_{N-1} \\ 1 & d_1 & d_2 & \cdots & d_{k-1} & d_k \\ d_{N-k-1} & d_{N-k} & d_{N-k+1} & \cdots & d_{N-2} & d_{N-1} \end{bmatrix}. \quad (75)$$

Note that

$$\begin{aligned} \rho(\mathbf{c})(N-k-1) + \rho(\mathbf{d})(N-k-1) \\ = \langle \mathbf{c}_1, \mathbf{c}_2 \rangle + \langle \mathbf{d}_1, \mathbf{d}_2 \rangle = 0, \end{aligned} \quad (76)$$

$$\begin{aligned} \rho(\mathbf{c}, \mathbf{d})(N-k-1) + \rho(\mathbf{d}, \mathbf{c})(N-k-1) \\ = \langle \mathbf{c}_1, \mathbf{d}_2 \rangle + \langle \mathbf{d}_1, \mathbf{c}_2 \rangle = 0. \end{aligned} \quad (77)$$

By (75), we have

$$c_{N-k-1}^* + \sum_{i=1}^k c_i c_{N-k-1+i}^* + d_{N-k-1}^* + \sum_{i=1}^k d_i d_{N-k-1+i}^* = 0, \quad (78)$$

and

$$d_{N-k-1}^* + \sum_{i=1}^k c_i d_{N-k-1+i}^* + c_{N-k-1}^* + \sum_{i=1}^k d_i c_{N-k-1+i}^* = 0. \quad (79)$$

Summing up (78) and (79), we have

$$\begin{aligned} & 2(c_{N-k-1}^* + d_{N-k-1}^*) + \sum_{i=1}^k c_i \left(\underbrace{c_{N-k-1+i}^* + d_{N-k-1+i}^*}_{=0} \right) \\ & + \sum_{i=1}^k d_i \left(\underbrace{c_{N-k-1+i}^* + d_{N-k-1+i}^*}_{=0} \right) = 0. \end{aligned} \quad (80)$$

Subtracting (78) by (79), we have

$$\begin{aligned} & (c_k - d_k) c_{N-1}^* + \sum_{i=1}^{k-1} \left(\underbrace{c_i - d_i}_{=0} \right) c_{N-k-1+i}^* - (c_k - d_k) d_{N-1}^* \\ & + \sum_{i=1}^{k-1} \left(\underbrace{d_i - c_i}_{i=0} \right) d_{N-k-1+i}^* = 0. \end{aligned} \quad (81)$$

By (80), we have $c_{N-k-1} = -d_{N-k-1}$. By (81) and considering $c_{N-1} \neq d_{N-1}$ ¹⁰, we assert that $c_k = d_k$. Continuing this induction, the proof of (16) follows.

Next, we assume $Z > N/2$ and N is even. Setting $i = N/2 - 1$ into (16), we have $c_{N/2} + d_{N/2} = 0$. Then, setting $i = N/2$ into (16), we obtain $c_{N/2} = d_{N/2}$. This requires that $c_{N/2} = d_{N/2} = 0$ which contradicts with the precondition that (\mathbf{c}, \mathbf{d}) is a q -ary CZCP. When N is odd, another contradiction is reached for $c_{(N-1)/2} = d_{(N-1)/2} = 0$. In either case, we assert that $Z \leq N/2$ should be met.

Proof of P2:

Proof 5: In Section II, we have shown that if (\mathbf{a}, \mathbf{b}) is a GCP, $(\mathbf{b}^*, -\mathbf{a}^*)$ will also be a GCP which is mutually orthogonal to (\mathbf{a}, \mathbf{b}) . The first identity of (32) is similar to the mutually orthogonal property of GCPs, although with some slight changes for its proof. Here, we only prove the second equality of (32). For $1 \leq \tau \leq Z$, we have

$$\begin{aligned} & \rho(\mathbf{b}, \mathbf{b}^*)(N - \tau) + \rho(\mathbf{a}, -\mathbf{a}^*)(N - \tau) \\ & = \sum_{i=0}^{\tau-1} (b_i b_{\tau-1-i} - a_i a_{\tau-1-i}). \end{aligned} \quad (82)$$

By assumption, we have $a_i = b_i, a_{N-1-i} = -b_{N-1-i}$, for all $i \in (\mathcal{T}_1 - 1)$. Thus, the above equation reduces to

$$\begin{aligned} & \rho(\mathbf{b}, \mathbf{b}^*)(N - \tau) + \rho(\mathbf{a}, -\mathbf{a}^*)(N - \tau) \\ & = \sum_{i=0}^{\tau-1} a_i \left(\underbrace{b_{\tau-1-i} - a_{\tau-1-i}}_{=0} \right) = 0. \end{aligned} \quad (83)$$

¹⁰Otherwise, we have $c_{N-1} = d_{N-1} = 0$ by $c_{N-1} + d_{N-1} = 0$. This contradicts with the precondition that (\mathbf{c}, \mathbf{d}) is a q -ary CZCP.

On the other hand, when $\tau = Z + 1$, we have

$$\begin{aligned} & \rho(\mathbf{b}, \mathbf{b}^*)(N - \tau) + \rho(\mathbf{a}, -\mathbf{a}^*)(N - \tau) \\ & = \sum_{i=1}^{\lambda-1} a_i \left(\underbrace{b_{Z-i} - a_{Z-i}}_{=0} \right) + 2a_0 \left(\underbrace{b_Z - a_Z}_{\neq 0} \right) \neq 0. \end{aligned} \quad (84)$$

This proves the second equality of (32).

Proof of P3:

Proof 6: For binary sequence pair (\mathbf{a}, \mathbf{b}) over $\{-1, 1\}$, by the first condition specified in (13), we have

$$\rho(\mathbf{a})(N-1) + \rho(\mathbf{b})(N-1) = a_0 a_{N-1} + b_0 b_{N-1} = 0. \quad (85)$$

Also, we require

$$\rho(\mathbf{a})(1) + \rho(\mathbf{b})(1) = \sum_{i=0}^{N-2} (a_i a_{i+1} + b_i b_{i+1}) = 0. \quad (86)$$

Let $a_i = 1 - 2\tilde{a}_i, b_i = 1 - 2\tilde{b}_i$, where $1 \leq i \leq N$ and $\tilde{a}_i, \tilde{b}_i \in \mathbb{Z}_2$. Consequently, (85) and (86), respectively, imply that

$$\tilde{a}_0 + \tilde{a}_{N-1} + \tilde{b}_0 + \tilde{b}_{N-1} = 1 \pmod{2}, \quad (87)$$

and

$$\begin{aligned} N-1 &= \sum_{i=0}^{N-2} (a_i + a_{i+1} + b_i + b_{i+1}) \pmod{2} \\ &= \tilde{a}_0 + \tilde{a}_{N-1} + \tilde{b}_0 + \tilde{b}_{N-1} \pmod{2}. \end{aligned} \quad (88)$$

Combining (87) and (88), we assert that N should be an even positive integer.

In the end, (18) of P3 can be easily obtained by an induction exploiting $\rho(\mathbf{a})(\tau) + \rho(\mathbf{b})(\tau)$ starting from $\tau = N - 1$ to $\tau = N - Z$. This induction is similar to that in [2], [14] and therefore detailed proof of (18) is omitted.

REFERENCES

- [1] M. J. E. Golay, "Statatic multislit spectroscopy and its application to the panoramic display of infrared spectra," *J. Opt. Soc. Amer.*, vol. 41, pp. 468-472, 1951.
- [2] M. J. E. Golay, "Complementary series," *IRE Trans. Inf. Theory*, vol. IT-7, pp. 82-87, Apr. 1961.
- [3] S. Budišin, "New complementary pairs of sequences," *Electron. Lett.*, vol. 26, no. 13, pp. 881-883, Jun. 1990.
- [4] S. Budišin, "New multilevel complementary pairs of sequences," *Electron. Lett.*, vol. 26, no. 22, pp. 1861-1863, Oct. 1990.
- [5] S. Budišin and P. Spasojević, "Paraunitary generation/correlation of QAM complementary sequence pairs," in *Proc. Cryptography Commun.*, vol. 6, no. 1, pp. 59-102, Oct. 2014.
- [6] Z. Wang, G. Wu, and D. Ma, "A new method to construct Golay complementary set by paraunitary matrices and Hadamard matrices," in *Proc. 9th International Conference on Sequences and Their Applications (SETA-2016)*, Sep. 2016, pp. 1-12.
- [7] S. Das, S. Budišin, S. Majhi, Z. Liu, and Y. L. Guan, "A multiplier-free generator for polyphase complete complementary codes," *IEEE Trans. Signal Process.*, vol. 66, no. 5, pp. 1184-1192, Mar. 2018.
- [8] S. Budišin and P. Spasojević, "Paraunitary-based Boolean generator for QAM complementary sequences of length 2^K ," *IEEE Trans. Inf. Theory*, vol. 64, no. 8, pp. 5938-5956, Aug. 2018.
- [9] J. A. Davis and J. Jedwab, "Peak-to-mean power control in OFDM, Golay complementary sequences, and Reed-Muller codes," *IEEE Trans. Inf. Theory*, vol. 45, no. 7, pp. 2397-2417, Nov. 1999.
- [10] Y. Li, "A construction of general QAM Golay complementary sequences," *IEEE Trans. Inf. Theory*, vol. 56, no. 11, pp. 5765-5771, Nov. 2010.

- [11] Z. Liu, Y. Li, and Y. L. Guan, "New constructions of general QAM Golay complementary sequences," *IEEE Trans. Inf. Theory*, vol. 59, no. 11, pp. 7684-7692, Nov. 2013.
- [12] P. Fan and M. Darnell, *Sequence Design for Communications Applications*. New York: Wiley, 1996.
- [13] M. G. Parker, K. G. Paterson, and C. Tellambura, "Golay complementary sequences," *Wiley Encyclopedia of Telecommunications*, J. G. Proakis, Ed. New York: Wiley Interscience, 2002.
- [14] Z. Liu, U. Parampalli, and Y. L. Guan, "Optimal odd-length binary Z-complementary pairs," *IEEE Trans. Inf. Theory*, vol. 60, no. 9, pp. 5768-5781, July 2014.
- [15] P. Fan, W. Yuan, and Y. Tu, "Z-complementary binary sequences," *IEEE Signal Process. Lett.*, vol. 14, pp. 401-404, Aug. 2007.
- [16] C. Tseng and C. Liu, "Complementary sets of sequences," *IEEE Trans. Inf. Theory*, vol. IT-18, pp. 644-665, Sept. 1972.
- [17] A. Rathinakumar and A. K. Chaturvedi, "Complete mutually orthogonal Golay complementary sets from Reed-Muller codes," *IEEE Trans. Inf. Theory*, vol. 54, pp. 1339-1346, Mar. 2008.
- [18] S. Wang and A. Abli, "Aperiodic complementary sets of sequences-based MIMO frequency selective channel estimation," *IEEE Commun. Lett.*, vol. 9, no. 10, pp. 891-893, Oct. 2005.
- [19] S. Wang and A. Abdi, "MIMO ISI channel estimation using uncorrelated Golay complementary sets of polyphase sequences," *IEEE Trans. Veh. Tech.*, vol. 56, pp. 3024-3040, Sep. 2007.
- [20] S. Z. Budišin, "Efficient pulse compressor for Golay complementary sequences," *IEE Electron. Lett.*, vol. 27, pp. 219-220, Jan. 1991.
- [21] A. Pezeshki, A. R. Calderbank, W. Moran, and S. D. Howard, "Doppler resilient Golay complementary waveforms," *IEEE Trans. Inf. Theory*, vol. 54, pp. 4254-4266, Sep. 2008.
- [22] H. H. Chen, J. F. Yeh, and N. Seuhiro, "A multicarrier CDMA architecture based on orthogonal complementary codes for new generations of wideband wireless communications," *IEEE Commun. Magazine*, vol. 39, pp. 126-135, Oct. 2001.
- [23] H. H. Chen, *The Next Generation CDMA Technologies*, John Wiley & Sons, Jul. 2007.
- [24] Z. Liu, Y. L. Guan, and H. H. Chen, "Fractional-delay-resilient receiver for interference-free MC-CDMA communications based on complete complementary codes," *IEEE Trans. Wireless Commun.*, vol. 14, no. 3, pp. 1226-1237, Mar. 2015.
- [25] B. M. Popović, "Synthesis of power efficient multitone signals with flat amplitude spectrum," *IEEE Trans. Commun.*, vol. 39, pp. 1031-1033, Jul. 1991.
- [26] K. G. Paterson, "Generalized Reed-Muller codes and power control in OFDM modulation," *IEEE Trans. Inf. Theory*, vol. 46, pp. 104-120, Jan. 2000.
- [27] Z. Liu, Y. L. Guan, and U. Parampalli, "New complete complementary codes for the peak-to-mean power control in MC-CDMA," *IEEE Trans. Commun.*, vol. 62, pp. 1105-1113, Mar. 2014.
- [28] E. Telatar, "Capacity of multi-antenna Gaussian channels," *Eur. Trans. Telecommun.*, vol. 10, no. 6, pp. 585-595, Nov./Dec. 1999.
- [29] V. Tarokh, H. Jafarkhani, and A. R. Calderbank, "Space-time block codes from orthogonal designs," *IEEE Trans. Inf. Theory*, vol. 45, no. 5, pp. 1456-1467, Jul. 1999.
- [30] B. Hassibi and B. M. Hochwald, "High-rate codes that are linear in space and time," *IEEE Trans. Inf. Theory*, vol. 48, no. 7, pp. 1804-1824, Jul. 2002.
- [31] L. Zheng and D. N. C. Tse, "Diversity and multiplexing: A fundamental tradeoff in multiple-antenna channels," *IEEE Trans. Inf. Theory*, vol. 49, no. 5, pp. 1073-1096, May 2003.
- [32] J. Mietzner, R. Schober, L. Lampe, W. H. Gerstacker, and P. A. Höher, "Multiple-antenna techniques for wireless communications-A comprehensive literature survey," *IEEE Commun. Surveys Tuts.*, vol. 11, no. 2, pp. 87-105, Second Quart., 2009.
- [33] E. Basar, "On multiple-input multiple-output OFDM with index modulation for next generation wireless networks," *IEEE Trans. Signal Process.*, vol. 64, no. 15, pp. 3868-3878, Aug. 2016.
- [34] B. Zheng, M. Wen, E. Basar, and F. Chen, "Multiple-input multiple-output OFDM with index modulation: Low-complexity detector design," *IEEE Trans. Signal Process.*, vol. 65, no. 11, pp. 2758-2772, Jun. 2017.
- [35] R. Mesleh, H. Haas, Y. Lee, and S. Yun, "Interchannel interference avoidance in MIMO transmission by exploiting spatial information," in *Proc. 16th IEEE Int. Symp. PIMRC*, Berlin, Germany, 2005, vol. 1, pp. 141-145.
- [36] R. Mesleh, H. Haas, S. Sinanovi, C. W. Ahn, and S. Yun, "Spatial modulation," *IEEE Trans. Veh. Technol.*, vol. 57, no. 4, pp. 2228-2241, Jul. 2008.
- [37] M. Di Renzo, H. Haas, and P. M. Grant, "Spatial modulation for multiple antenna wireless systems: A survey," *IEEE Commun. Mag.*, vol. 49, no. 12, pp. 182-191, Dec. 2011.
- [38] P. Yang, M. D. Renzo, Y. Xiao, S. Li, and L. Hanzo, "Design guidelines for spatial modulation," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 1, pp. 6-26, First Quart., 2015.
- [39] M. Wen, B. Zheng, K. Kim, M. Renzo, T. Tsiftsis, K.-C. Chen, and N. Al-Dhahir, "A survey on spatial modulation in emerging wireless systems: research progresses and applications," *IEEE J. Sel. Areas Commun.*, vol. 37, no. 9, pp. 1949-1972, Sep. 2019.
- [40] P. Yang, Y. Xiao, Y. L. Guan, K. V. S. Hari, A. Chockalingam, S. Sugiura, H. Haas, M. D. Renzo, C. Masouros, Z. Liu, L. Xiao, S. Li, and L. Hanzo, "Single-carrier SM-MIMO: a promising design for broadband large-scale antenna systems," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 3, pp. 1687-1716, Third Quart., 2016.
- [41] J. Jeganathan, A. Ghrayeb, and L. Szczecinski, "Spatial modulation: Optimal detection and performance analysis," *IEEE Commun. Lett.*, vol. 12, no. 8, pp. 545-547, Aug. 2008.
- [42] M. D. Renzo and H. Haas, "Bit error probability of SM-MIMO over generalized fading channels," *IEEE Trans. Veh. Technol.*, vol. 61, no. 3, pp. 1124-1144, Mar. 2012.
- [43] E. Basar, U. Aygolu, E. Panayirci, and H. Poor, "Performance of spatial modulation in the presence of channel estimation errors," *IEEE Commun. Lett.*, vol. 16, no. 2, pp. 176-179, Feb. 2012.
- [44] M. Di Renzo, D. De Leonardi, F. Graziosi, and H. Haas, "Space shift keying (SSK-) MIMO with practical channel estimates," *IEEE Trans. Commun.*, vol. 60, no. 4, pp. 998-1012, Apr. 2012.
- [45] S. Sugiura and L. Hanzo, "Effects of channel estimation on spatial modulation," *IEEE Signal Process. Lett.*, vol. 19, no. 12, pp. 805-808, Dec. 2012.
- [46] X. Wu, H. Claussen, M. Renzo, and H. Haas, "Channel estimation for spatial modulation," *IEEE Trans. Commun.*, vol. 62, no. 12, pp. 4362-4372, Dec. 2014.
- [47] C. Fragouli, N. Al-Dhahir, and W. Turin, "Training-based channel estimation for multiple-antenna broadband transmissions," *IEEE Trans. Wireless Commun.*, vol. 2, no. 2, pp. 384-391, Mar. 2003.
- [48] 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.
- [49] C. Fragouli, N. Al-Dhahir, and W. Turin, "Training-based channel estimation for multiple-antenna broadband transmissions," *IEEE Trans. Wireless Commun.*, vol. 2, no. 2, pp. 384-391, Mar. 2003.
- [50] 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, no. 5, pp. 1623-1626, Sep. 2004.
- [51] P. Borwein and M. J. Mossinghoff, "Barker sequences and flat polynomials," in *Number Theory and Polynomials*, Cambridge University Press, pp. 7188, 2008.
- [52] P. Stoica, H. He, and J. Li, "New algorithms for designing unimodular sequences with good correlation properties," *IEEE Trans. Signal Process.*, vol. 57, no. 4, pp. 1415-1425, Apr. 2009.
- [53] M. Soltanalian and P. Stoica, "Computational design of sequences with good correlation properties," *IEEE Trans. Signal Process.*, vol. 60, no. 5, pp. 2180-2193, May 2012.
- [54] M. Soltanalian and P. Stoica, "Designing unimodular codes via quadratic optimization," *IEEE Trans. Signal Process.*, vol. 62, no. 5, pp. 1221-1234, Mar. 2014.
- [55] J. Song, P. Babu, and D. Palomar, "Optimization methods for designing sequences with low autocorrelation sidelobes," *IEEE Trans. Signal Process.*, vol. 63, no. 15, pp. 3998-4009, Aug. 2015.
- [56] Y. Sun, P. Babu, and D. P. Palomar, "Majorization-minimization algorithms in signal processing, communications, and machine learning," *IEEE Trans. Signal Process.*, vol. 65, no. 3, pp. 794-816, Feb. 2017.
- [57] M. A. Kerafroudi, A. Aubry, A. De Maio, M. M. Naghsh, and M. Modarres-Hashemi, "A coordinate-descent framework to design low PSL/ISL sequences," *IEEE Trans. Signal Process.*, vol. 65, no. 22, pp. 5942-5956, Nov. 2017.
- [58] M. Alaee-Kerafroudi, M. R. Bhavani Shankar, K. V. Mishra, and B. Ottersten, "Meeting the lower bound on designing set of unimodular sequences with small aperiodic/periodic ISL," in *2019 20th International Radar Symposium (IRS)*, pp. 1-13, Jun. 2019.
- [59] M. Alaee-Kerafroudi, M. Modarres-Hashemi, and M. M. Naghsh, "Designing sets of binary sequences for MIMO radar systems," *IEEE Trans. Signal Process.*, vol. 67, no. 13, pp. 3347-3360, Jul. 2019.

- [60] R. Gold, "Maximal recursive sequences with 3-valued recursive cross-correlation functions," *IEEE Trans. Inf. Theory*, vol. 14, pp. 154-156, Jan. 1968.
- [61] D. Chu, "Polyphase codes with good periodic correlation properties (Corresp.)," *IEEE Trans. Inf. Theory*, vol. IT-18, no. 4, pp. 531-532, Jul. 1972.
- [62] J. Fu, C. Hou, W. Xiang, L. Yan, and Y. Hou, "Generalised spatial modulation with multiple active transmit antennas," in *Proc. IEEE GLOBECOM Workshops*, Dec. 2010, pp. 839-844.
- [63] T. Datta and A. Chockalingam, "On generalized spatial modulation," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC'2013)*, Apr. 2013, pp. 2716-2721.
- [64] J. Wang, S. Jia, and J. Song, "Generalised spatial modulation system with multiple active transmit antennas and low complexity detection scheme," *IEEE Trans. Wireless Commun.*, vol. 11, no. 4, pp. 1605-1615, Apr. 2012.
- [65] Y. Liu, T. Wong, and W. Hager, "Training signal design for estimation of correlated MIMO channels with colored interference," *IEEE Trans. Signal Process.*, vol. 55, no. 4, pp. 1486-1497, Apr. 2007.
- [66] D. Katselis, E. Kofidis, and S. Theodoridis, "On training optimization for estimation of correlated MIMO channels in the presence of multiuser interference," *IEEE Trans. Signal Process.*, vol. 56, no. 10, pp. 4892-4904, Oct. 2008.
- [67] M. Biguesh, S. Gazor, and M. Shariat, "Optimal training sequence for MIMO wireless systems in colored environments," *IEEE Trans. Signal Process.*, vol. 57, no. 8, pp. 3144-3153, Aug. 2009.
- [68] E. Bjornson and B. Ottersten, "A framework for training-based estimation in arbitrarily correlated Rician MIMO channels with Rician disturbance," *IEEE Trans. Signal Process.*, vol. 58, no. 3, pp. 1807-1820, Mar. 2010.
- [69] N. Shariati, J. Wang and M. Bengtsson, "Robust training sequence design for correlated MIMO channel estimation," *IEEE Trans. Signal Process.*, vol. 62, no. 1, pp. 107-120, Jan. 2014.