Properties of Syndrome Distribution for Blind Reconstruction of Cyclic Codes

Arti D. Yardi and Saravanan Vijayakumaran

Department of Electrical Engineering

Indian Institute of Technology Bombay, Mumbai 400076, India

Email: {arti,sarva}@ee.iitb.ac.in

Abstract—In the problem of blind reconstruction of channel codes, the receiver does not have the knowledge of the channel code used at the transmitter and the aim is to identify this unknown channel code corresponding to the given received sequence. In this paper, we study this blind reconstruction problem for binary cyclic codes. In the literature, several researchers have proposed blind reconstruction algorithms that make use of the distribution of the syndromes (remainders) of the received polynomials with respect to a candidate polynomial for the generator polynomial of the cyclic code. However, very limited analysis is available for the syndrome distribution and its properties. In this paper, we study the syndrome structure of the received polynomials. Specifically, we prove that the syndrome distribution of the noise-free sequence can either be uniform or restricted uniform. We also provide the necessary and sufficient conditions for it to be of the either type. For the noise-affected received sequence we prove that, finding the syndrome distribution is in general computationally intractable. We also apply these results to analyze the performance of the existing methods and verify some of the assumptions made in the literature for blind reconstruction.

I. Introduction

Channel codes play a vital role in the digital communication system to make the system robust to the errors introduced by the channel noise. When the channel code used at the transmitter is known at the receiver, the received data can be decoded to obtain the transmitted messages [1]. However there could be situations when the channel code used at the transmitter is not known at the receiver. For example, in military surveillance the channel code used by an adversary might not be known. In such scenarios, in order to decode the received data, one needs to first identify this unknown channel (see Fig. 1). This problem of identifying the channel corresponding to the given received data is known as *blind reconstruction of channel codes* [2]–[4].

This blind reconstruction problem is in general NP-hard [5]. While identifying a particular channel code, it is typically assumed that the family of the code, such as convolutional or linear block code, is known. The underlying structure of this particular family is then used to identify the code. In the literature, various algorithms have been proposed for blind reconstruction of convolutional codes [6], [7], turbo codes [8], [9], linear block codes [5], [10], [11], LDPC codes [12], [13], and cyclic codes [14]–[19].

Chabot [15], Lee et al. [14], and Yardi et al. [16] have studied this blind reconstruction problem for cyclic codes

Fig. 1. A system model for blind reconstruction problem of channel codes.

when the length of the code is assumed to be known at the receiver. Zhou et al. [18], [19] and Yardi et al. [17] consider the situation when the length of the cyclic code is not known. In this paper, we focus on the unknown length scenario. For the unknown length scenario, a key idea proposed in the existing methods is summarized next [17]–[19]. The unknown cyclic code $C(n_0, g_0)$ is identified by finding its length n_0 and the factors of its generator polynomial $g_0(X)$. Since the first received bit might not be the first bit of a received codeword, for blind reconstruction, one also needs to identify the location of the codeword boundaries or synchronization of the received data. The analysis begins by assuming a length n, synchronization, and a candidate polynomial $f(X)$ for the factor of the generator polynomial. Note that $f(X)$ is factor of $X^n + 1$ since for an assumed n, the generator polynomial has to be a factor of $X^n + 1$ [1]. For the assumed n, synchronization, and $f(X)$ there are the following two cases.

(a) Both n and synchronization are correct, and $f(X)$ is a factor of $g_0(X)$

(b) Either n or synchronization is not correct or $f(X)$ is not a factor of $g_0(X)$

For the chosen n, synchronization, and $f(X)$, the key step in the existing methods consists of determining which one of the above two cases holds.

In order to use the optimal likelihood ratio tests for determining whether (a) is true or (b) is true, one needs to find the probability of the received data when condition (a) is true and when condition (b) is true [20]. However, we next explain that finding this probability is, in general, computationally intractable. When condition (a) is true, let

$\mathbb{P}[\mathbf{y}]$ be the probability of receiving an n_0-bit vector $\mathbf{y}$. This probability can be computed by conditioning over all possible transmitted codewords in $C(n_0, g_0)$ as follows,

$$\mathbb{P}[\mathbf{y}] = \sum_{\mathbf{v} \in C(n_0, g_0)} \mathbb{P}\Big[\mathbf{y}\Big|\mathbf{v} \text{ is transmitted}\Big]\mathbb{P}\Big[\mathbf{v} \text{ is transmitted}\Big].$$

When $f(X) = g_0(X)$ and the true code $C(n_0, g_0)$ used at the transmitter is known at the receiver, it is shown in [21] that finding $\mathbb{P}[\mathbf{y}]$ is, in general, computationally intractable. Since in our case $C(n_0, g_0)$ is not known, obtaining $\mathbb{P}[\mathbf{y}]$ is even more computationally intractable. Hence in the literature, researchers have proposed suboptimal tests which make use of the syndromes of the received polynomials to take a decision between (a) and (b) [17]–[19]. In [17] and [19], the properties of the zero syndromes of the received polynomials are used to distinguish between (a) and (b). Whereas in [18], the marginal distribution of the coefficients of the syndromes is used for blind reconstruction.

Understanding the syndrome structure of the received polynomials is thus important to study the problem of blind reconstruction of binary cyclic codes. However, very limited analysis is available for the syndrome distribution and its properties. Due to lack of knowledge of the syndrome distribution, typically in the literature some assumptions are made to simplify the analysis [18], [19]. For example, in [18] it is assumed that when either of the assumed parameter is incorrect (case (b) mentioned above), every coefficient in the syndrome of the received polynomial is equally likely to be zero or one. In [19], the received data is assumed to behave as a random bitstream for the incorrect parameters. In this paper, we analyze the properties of the syndrome distribution and verify these assumptions. These syndrome properties can also be use to study the theoretical performance of the method proposed in [17]. The main contributions of this paper are as follows.

(1) We first characterize the syndrome distribution of the noise-free polynomials with respect to a candidate polynomial $f(X)$. We prove that when either of the assumed parameter are incorrect (case (b) mentioned above), the distribution of the syndrome can be either uniform or restricted uniform (see (7), (8), and Proposition 1). We also provide the necessary and sufficient conditions for the distribution to be restricted uniform (see Theorems 1, 2, and 3).

(2) We study the syndrome distribution of the noise-affected received polynomials. We prove that when the syndrome distribution of the noise-free polynomial is uniform, the distribution of the syndrome of the noise-affected polynomial would also be uniform (see Theorem 4). We also show that, when the distribution of the syndrome of the noise-free polynomial is restricted uniform, finding the distribution of the noise-affected polynomial is in general computationally intractable.

(3) Finally, using the syndrome analysis mentioned in (1) and (2) above, we verify the assumptions made in [19] and [18] and provide a theoretical analysis of the blind reconstruction method proposed in [17].

Organization: The system model for blind reconstruction of cyclic codes and some preliminaries are provided in Section II. We study the syndrome distribution of the noise-free sequence in Section III. This analysis is then extended to the noise-affected case in Section IV. In Section V, we provide a theoretical analysis of the existing blind reconstruction methods. Finally, we conclude in Section VI.

Notation: The set of natural numbers is denoted by $\mathbb{N}$ and $\mathbb{F}_2$ denotes the finite field with two elements 0 and 1. The polynomial ring with coefficients from $\mathbb{F}_2$ is denoted by $\mathbb{F}_2[X]$. The integer $\lfloor m \rfloor$ denotes the greatest integer less than or equal to m. We use boldface letters to denote the vectors and lower case letters for the components of a vector. For example, vector $\mathbf{y} = \begin{bmatrix} y_0 & y_1 & \cdots & y_{n-1} \end{bmatrix}$, where y_i for $i = 0, 1, \ldots, n-1$ are the components of $\mathbf{y}$. The polynomial representation of vector $\mathbf{y}$, is given by $\mathbf{y}(X) = y_0 + y_1 X + \ldots + y_{n-1} X^{n-1}$. Note that the polynomials corresponding to vectors are denoted by boldface letters. For integers l, r, $0 \le l < r < n$, we define $\mathbf{y}(l : r) := \begin{bmatrix} y_l & y_{l+1} & \cdots & y_r \end{bmatrix}$. When $l = 0$, the vector $\mathbf{y}(0 : r)$ is called as a *prefix* of $\mathbf{y}$ and when $r = n - 1$, the vector $\mathbf{y}(l : n - 1)$ is called as a *suffix* of $\mathbf{y}$.

II. SYSTEM MODEL AND PRELIMINARIES

A linear block code of length n is denoted by $C(n)$ and the cyclic code of length n and the generator polynomial $g(X)$ is denoted by $C(n, g)$. Let k be the dimension of $C(n, g)$. It is known that $k = n - \deg(g)$, where $\deg(g)$ is the degree of $g(X)$ [1]. When $k = 0$ or $k = n$, the code $C(n, g)$ is said to be a trivial cyclic code. Any codeword polynomial $\mathbf{v}(X)$ can be written as $\mathbf{v}(X) = \mathbf{u}(X)g(X)$ where $\mathbf{u}(X)$ is a message polynomial. The set of polynomials in $\mathbb{F}_2[X]$ of degrees strictly less than n is denoted by $\mathcal{P}_n$, i.e.,

$$\mathcal{P}_n = \Big\{ f(X) \in \mathbb{F}_2[X] \Big| \deg(f(X)) \le n - 1 \Big\}. \quad (1)$$

Using to this notation, $\mathbf{v}(X) \in \mathcal{P}_n$ and $\mathbf{u}(X) \in \mathcal{P}_k$.

Suppose the cyclic code $C(n_0, g_0)$ of dimension k_0 is used at the transmitter. Each transmitted codeword is independent and identically distributed (i.i.d.) according to the uniform distribution over the set of codewords of $C(n_0, g_0)$. We assume that the noise is introduced by a binary symmetric channel (BSC) of crossover probability $p < 1/2$. The received bitstream is denoted by $y_0, y_1, \ldots, y_{N-1}$. We now define the synchronization s_0 of this bitstream as follows.

Definition 1. *The synchronization s_0 of the received bitstream $y_0, y_1, \ldots, y_{N-1}$ is defined as the smallest integer such that the vector $\begin{bmatrix} y_{s_0} & \cdots & y_{s_0+n_0-1} \end{bmatrix}$ of length n_0 is the noise-affected version of the transmitted codeword of the cyclic code $C(n_0, g_0)$ used at the transmitter. Note that $0 \le s_0 < n_0$.* $\square$

Let $n \in \mathbb{N}$ be an assumed length of the code. For an assumed synchronization s, $0 \le s < n$ ignore $y_0, y_1, \ldots, y_{s-1}$ from the received bitstream and divide the remaining bitstream into vectors of length n. Thus the first n-bit vector is given by $\mathbf{y}_1(n, s) = \begin{bmatrix} y_s & \cdots & y_{s+n-1} \end{bmatrix}$. Similarly the jth n-bit vector is given by $\mathbf{y}_j(n, s) = \begin{bmatrix} y_{s+(j-1)n} & \cdots & y_{s+jn-1} \end{bmatrix}$. Suppose we have received $M = \lfloor (N - s)/n \rfloor$ vectors of length n. For the sake of simplicity we will drop parameters n and s from

$\mathbf{y}_j(n, s)$. Thus $\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_M$ is the sequence of n-bit vectors for an assumed synchronization s. Note that, the polynomial corresponding to $\mathbf{y}_j$ is given by $\mathbf{y}_j(X)$, for $j = 1, 2, \ldots, M$.

We assume that cyclic code $C(n_0, g_0)$ used at the transmitter is non-degenerate, where a degenerate code is defined as follows.

Definition 2. *Let G be a generator matrix of a linear block code $C(n)$. Then $C(n)$ is said to be degenerate if G can be written as,*

$$G = \left[\underbrace{G' \; G' \; \cdots \; G'}_{l \; times} \right], \tag{2}$$

where $l \in \mathbb{N}, l > 1$ and G' is a generator matrix of some other linear block code $C'(n')$ of length $n' = n/l$ [22, Ch. 8]. The code $C'(n')$ is said to be a component code of $C(n)$. For a cyclic code, its component code is also cyclic [17]. □

For blind reconstruction of a degenerate cyclic code, it is sufficient to identify its non-degenerate component (see [17] for details). Hence without loss of generality we consider the situation when the cyclic code used at the transmitter is not degenerate.

A. Preliminaries

In this section, we consider some preliminaries that will be required throughout the paper.

Definition 3. *The order of a polynomial $f(X)$ over $\mathbb{F}_2(X)$ is defined as the least positive integer l such that $f(X)$ divides $X^l + 1$ [23, Sec. 3.1].* □

We next recall a definition of a *linear recurring sequence* and its *period*.

Definition 4. *For a positive integer l, a sequence of bits $v_0, v_1, \cdots$ is said to be a linear recurring sequence of lth order if they follow a relation*

$$v_{r+l} = \sum_{i=0}^{l-1} h_i v_{r+i}, \;\; for \; r = 0, 1, \ldots \tag{3}$$

where $h_i \in \mathbb{F}_2$ for $i = 0, 1, \ldots, l-1$. It is known that any linear recurring sequence is ultimately periodic and its period is defined as a positive integer n such that $v_{r+n} = v_r$, for $r = 0, 1, \ldots$ [23, Sec. 6.1]. □

The *minimal polynomial* associated with a linear recurring sequence is defined next.

Definition 5. *Suppose d is the least positive integer such that the linear recurring sequence $v_0, v_1, \cdots$ satisfies the relation given in (3). Then the polynomial $h(X) := h_{d-1}X^{d-1} + h_{d-2}X^{d-2} + \ldots + h_0$ is called as the minimal polynomial associated with this sequence [23, Sec. 6.4].* □

It is known that any linear recurring sequence has a unique minimal polynomial and the order of the minimal polynomial is equal to the least period of this sequence [23, Sec. 6.4]. We next define a vector of *degenerate pattern*.

(a) Degenerate   (b) Uniform   (c) Restricted uniform

Fig. 2. An illustration of the three types of distributions defined in (6), (7), and (8) for a random variable X with the support set $\mathcal{X} = \{0, 1, \ldots, 7\}$.

Definition 6. *An n-bit vector $\mathbf{v}$ is said to be of degenerate pattern if it can be written as*

$$\mathbf{v} = \left[\underbrace{\mathbf{w} \; \mathbf{w} \; \cdots \; \mathbf{w}}_{l \; times} \right], \tag{4}$$

where $l \in \mathbb{N}$ and $\mathbf{w}$ itself is a vector of length $n' = n/l$ such that $\mathbf{w}$ is not a vector of degenerate pattern [24]. □

It is known that the sequence of bits given by $[\mathbf{w} \; \mathbf{w} \; \cdots]$ in (4) is a linear recurring sequence with least period n' [23]. We now define the *minimal generating polynomial* associated with this sequence as follows.

Definition 7. *Let $h(X)$ be the minimal polynomial of the linear recurring sequence $[\mathbf{w} \; \mathbf{w} \; \cdots]$ with least period n'. Then the minimal generating polynomial $m(X)$ associated with $[\mathbf{w} \; \mathbf{w} \; \cdots]$ is defined as*

$$m(X) := \frac{X^{n'} + 1}{h'(X)}, \tag{5}$$

where $h'(X) = X^{\deg(h)} h(X^{-1})$. It is known that the polynomial $\mathbf{w}(X)$ corresponding to $\mathbf{w}$ is a multiple of $m(X)$ [25, Sec. 7.4]. Note that $h(X)$ is the generator polynomial of the dual code of $C(n', m)$. □

We next provide a definition of the *outer direct sum* of two linear block codes.

Definition 8. *The outer direct sum $C_1(n_1) + C_2(n_2)$ of codes $C_1(n_1)$ and $C_2(n_2)$ is defined as a linear block code formed by concatenating all possible codewords of $C_1(n_1)$ with all possible codewords of $C_2(n_2)$, i.e.,*

$$C_1(n_1) + C_2(n_2) := \left\{ \left[\mathbf{v} \;\; \mathbf{w} \right] \middle| \mathbf{v} \in C_1(n_1), \mathbf{w} \in C_2(n_2) \right\}.$$
□

We now define three types of distributions for a discrete random variable X with a finite support set $\mathcal{X}$ such that the cardinality of $|\mathcal{X}|$ of set $\mathcal{X}$ is equal to 2^L for some integer $L \geq 1$. An example situation for these three types of distributions is shown in Fig. 2.

(1) Degenerate distribution (see Fig. 2(a))

Random variable X is said to follow the degenerate distribution if it takes a particular value $x_0 \in \mathcal{X}$ with probability one, i.e.,

$$\mathbb{P}[X = x] = \begin{cases} 1 & \text{if } x = x_0 \text{ for some } x_0 \in \mathcal{X}, \\ 0 & \text{otherwise.} \end{cases} \tag{6}$$

(2) Uniform distribution (see Fig. 2(b))

When random variable X follows the uniform distribution on its support set $\mathcal{X}$,

$$\mathbb{P}[X = x] = \begin{cases} 1/|\mathcal{X}| & \text{if } x \in \mathcal{X}, \\ 0 & \text{otherwise.} \end{cases} \quad (7)$$

(3) Restricted uniform distribution (see Fig. 2(c))

Consider a strict subset $\mathcal{X}_0$ of $\mathcal{X}$ such that $|\mathcal{X}_0| = 2^l$ for some integer l, where $1 \leq l < L$. Random variable X is said to follow the restricted uniform distribution on $\mathcal{X}$ if it follows the uniform distribution on set $\mathcal{X}_0$, i.e.,

$$\mathbb{P}[X = x] = \begin{cases} 1/|\mathcal{X}_0| & \text{if } x \in \mathcal{X}_0, \\ 0 & \text{otherwise.} \end{cases} \quad (8)$$

III. SYNDROME DISTRIBUTION OF THE NOISE-FREE SEQUENCE

Recall that for an assumed length n and synchronization s, the received sequence of polynomials is given by $\mathbf{y}_1(X), \mathbf{y}_2(X), \ldots, \mathbf{y}_M(X)$ (see Section II). Suppose $f(X)$ is factor of $X^n + 1$. For blind reconstruction, we need to study the distribution of $\mathbf{y}_j(X) \bmod f(X)$, for $j = 1, 2, \ldots, M$. Suppose the jth received polynomial $\mathbf{y}_j(X)$ is given by,

$$\mathbf{y}_j(X) = \mathbf{w}_j(X) + \mathbf{e}_j(X), \quad (9)$$

where $\mathbf{w}_j(X)$ is the noise-free polynomial and $\mathbf{e}_j(X)$ the error polynomial. In this section, we study the distribution of the syndrome of the noise-free polynomial, i.e., the distribution of $\mathbf{w}_j(X) \bmod f(X)$, where $1 \leq j \leq M$. The distribution of $\mathbf{y}_j(X) \bmod f(X)$ will be studied in the next section.

We first consider the case when either $n \neq ln_0$ or $s \neq s_0$, where $l \in \mathbb{N}$. The case when $n = ln_0$ and $s = s_0$ will be studied towards the end of this section. Consider a noise-free sequence of codewords of the true code $C(n_0, g_0)$ as shown in Fig. 3(a). Example situations when this sequence is divided into vectors of length n such that $n < n_0$, $s \neq s_0$ and $n > n_0$, $s \neq s_0$ are illustrated in Figures 3(b) and (c) respectively. In Fig. 3(a), $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_4 \in C(n_0, g_0)$. In Fig. 3(b), $\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_5$ are vectors of length $n < n_0$ and in Fig. 3(c), $\mathbf{w}_1$ and $\mathbf{w}_2$ are vectors of length $n > n_0$. In this section, we use the alphabets $\mathbf{v}$ and $\mathbf{w}$ to denote the vectors of lengths n_0 and n respectively.

From Fig. 3(b) and (c), it can be seen that when either $n \neq ln_0$ or $s \neq s_0$, a vector $\mathbf{w}_j$ of length n is either of the following two types.

1) Vector $\mathbf{w}_j$ is formed by the consecutive n bits of some codeword in code $C(n_0, g_0)$. For example, vectors $\mathbf{w}_1$ and $\mathbf{w}_4$ in Fig. 3(b) are formed by the consecutive n bits of codewords $\mathbf{v}_1$ and $\mathbf{v}_3$ of $C(n_0, g_0)$ respectively.

2) Vector $\mathbf{w}_j$ is formed by the concatenation of the suffix, a sequence of q codewords, and the prefix of a codeword in the true code, where $q \in \mathbb{Z}, q \geq 0$. For example, $\mathbf{w}_1$ in Fig. 3(c) is formed by the concatenation of the suffix of $\mathbf{v}_1$ of length d_1, $\mathbf{v}_2$, and the prefix of $\mathbf{v}_3$ of length d_2, where $0 \leq d_1, d_2 < n_0$ such that $n = d_1 + n_0 + d_2$. The vector $\mathbf{w}_2$ in Fig. 3(b) is formed by the concatenation of the suffix of $\mathbf{v}_2$ of length d'_1 and the prefix of $\mathbf{v}_3$ of length d'_2, such that $n = d'_1 + d'_2$.

Fig. 3. A binary cyclic code $C(n_0, g_0)$ is used at the transmitter and $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_4 \in C(n_0, g_0)$. Figures (a), (b), and (c) correspond to the situations when $n = n_0$, $s = s_0$, $n < n_0$, $s \neq s_0$, and $n > n_0$, $s \neq s_0$ respectively.

We denote the vector $\mathbf{w}_j$ of the second type by $\mathbf{w}'_j$ to distinguish between the n bits vectors of the two types mentioned above. For the simplicity of notation, we will ignore suffix j from $\mathbf{w}_j$ and $\mathbf{w}'_j$. Using this notation, $\mathbf{w}$ is an n-bit vector formed by the consecutive n bits of a codeword in $C(n_0, g_0)$. Since $C(n_0, g_0)$ is a cyclic code, it is sufficient to consider the case when $\mathbf{w}$ is formed by the initial n bits of a codeword in $C(n_0, g_0)$, i.e., $\mathbf{w}$ is given by,

$$\mathbf{w} = \mathbf{v}(0 : n - 1), \quad (10)$$

where $\mathbf{v} \in C(n_0, g_0)$. Let $\mathcal{W}(n)$ be the linear subspace obtained by puncturing the last $n_0 - n$ bits of codewords of code $C(n_0, g_0)$. It follows that $\mathbf{w} \in \mathcal{W}(n)$.

As explained in the previous paragraph, $\mathbf{w}'$ is an n-bit vector formed by the concatenation of the suffix of length d_1, a sequence of q codewords, and the prefix of length d_2, where $d_1, d_2, q \in \mathbb{N}$, such that $n = d_1 + qn_0 + d_2$, i.e., $\mathbf{w}'$ is given by,

$$\mathbf{w}' = \left[\mathbf{v}_1(n_0 - d_1 : n_0 - 1) \; \mathbf{v}_2 \; \cdots \; \mathbf{v}_{q+1} \; \mathbf{v}_{q+2}(0 : d_2 - 1) \right] \quad (11)$$

where $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{q+2} \in C(n_0, g_0)$. Let $C_1(d_1)$ and $C_2(d_2)$ be the linear block codes obtained by considering the set of suffixes and prefixes of lengths d_1 and d_2 of codewords in $C(n_0, g_0)$ respectively. Let $\mathcal{W}'(n)$ be the linear subspace obtained by concatenating all possible suffixes of length d_1, q

codewords, and prefixes of length d_2, i.e.,

$$\mathcal{W}'(n) := C_1(d_1) + \underbrace{C(n_0, g_0) + \cdots + C(n_0, g_0)}_{q \text{ times}} + C_2(d_2).$$

(12)

From (11), it can be seen that $\mathbf{w}' \in \mathcal{W}'(n)$. Note that, since every codeword in $C(n_0, g_0)$ is chosen according to the uniform distribution, any $\mathbf{w} \in \mathcal{W}(n)$ and $\mathbf{w}' \in \mathcal{W}'(n)$ occur with the uniform distribution over the set of codewords in $\mathcal{W}(n)$ and $\mathcal{W}'(n)$ respectively.

For a factor $f(X)$ of $X^n + 1$, suppose $r(X) = \mathbf{w}(X) \bmod f(X)$ and $r'(X) = \mathbf{w}'(X) \bmod f(X)$. From Fig. 3, in order to study the syndrome distribution of the noise-free sequence, we need to study the distributions $r(X)$ and $r'(X)$. In the following proposition, we first prove that the distributions of $r(X)$ and $r'(X)$ can either be uniform or restricted uniform.

Proposition 1. *For a cyclic code $C(n_0, g_0)$, let $\mathcal{W}(n)$ and $\mathcal{W}'(n)$ be the linear subspaces as defined in the previous paragraph, where n is not a multiple of n_0. For a factor $f(X)$ of $X^n + 1$, suppose $r(X) = \mathbf{w}(X) \bmod f(X)$ and $r'(X) = \mathbf{w}'(X) \bmod f(X)$, where $\mathbf{w}(X) \in \mathcal{W}(n)$ and $\mathbf{w}'(X) \in \mathcal{W}'(n)$. Then the random variables corresponding to $r(X)$ and $r'(X)$ can either follow the uniform distribution or the restricted uniform distribution. (see (7), (8), Fig. 2).*

Proof: The proof is given in Appendix B. ∎

Note that Proposition 1 is true irrespective of whether $f(X)$ is a factor of $g_0(X)$ or not. This proposition says that the distribution of $r(X)$ and $r'(X)$ can be either be uniform or restricted uniform, but it does not specify when the distribution will be of either of the type. In the next two sections we will answer this question.

A. Analyzing the distribution of $r(X)$

In this section, we characterize the distribution of $r(X) = \mathbf{w}(X) \bmod f(X)$, when $\mathbf{w}(X)$ is formed by the n consecutive bits of a codeword in $C(n_0, g_0)$. Due to the cyclic nature of the code, it is sufficient to consider the case when $\mathbf{w}(X)$ is formed by the initial n bits of a codeword in $C(n_0, g_0)$. Depending on the chosen n and the degree of $f(X)$ we have the following cases.

(a) $n \leq k_0$, where recall that k_0 is the dimension of $C(n_0, k_0)$

When $n \leq k_0$, a vector $\mathbf{w}$ formed the initial n bits of a codeword in $C(n_0, g_0)$ can take all possible 2^n values in $\mathbb{F}_2^n$ since, for a cyclic code any set of k_0 consecutive coordinate locations form an information set [26]. From our system assumption, any codeword in $C(n_0, g_0)$ is chosen i.i.d. according to the uniform distribution. Hence $\mathbf{w}(X)$ will take all possible values in $\mathcal{P}_n$ with equal probability and the random variable corresponding to $r(X)$ will follow the uniform distribution.

(b) $\deg(f) > k_0$

The syndrome $r(X) = \mathbf{w}(X) \bmod f(X)$ can take $2^{\deg(f)}$ possible values in $\mathcal{P}_{\deg(f)}$. Whereas, $\mathbf{w}(X)$ can take at most 2^{k_0} possible values. When $\deg(f) > k_0$, the number of possible syndromes are more than the number of possible $\mathbf{w}(X)$. This implies that the random variable

corresponding to $r(X)$ cannot follow the uniform distribution and from Proposition 1, $r(X)$ follows the restricted uniform distribution.

(c) $k_0 < n < n_0$ and $\deg(f) \leq k_0$

In this case, the distribution of $r(X)$ can either be uniform or restricted uniform. We characterize the conditions under which the restricted uniform distribution is possible in the following theorem.

Theorem 1. *Consider a non-degenerate cyclic code $C(n_0, g_0)$ of length n_0, dimension k_0, and generator polynomial $g_0(X)$. Let $g_0^\perp(X)$ be the generator polynomial of the dual code of $C(n_0, g_0)$. For an integer n and $\mathbf{v}(X) \in C(n_0, g_0)$, suppose $\mathbf{w}(X) = \mathbf{v}(X) \bmod X^n$ such that $k_0 < n < n_0$. Suppose $f(X)$ is a factor of $X^n + 1$ such that $\deg(f) \leq k_0$ and $r(X) = \mathbf{w}(X) \bmod f(X)$. Then the necessary condition for the random variable corresponding to $r(X)$ to follow the restricted uniform distribution is that $g_0^\perp(X)$ should have a factor of order strictly less than n_0.*

Conversely, when $g_0^\perp(X)$ has a factor $m^\perp(X)$ of order n' such that $1 \leq n' < n_0$, syndrome $r(X)$ follows the restricted uniform distribution if the chosen n and $f(X)$ satisfy the following conditions.

1) $n = bn'$ for some $b \in \mathbb{N}$.
2) $f(X)$ is a factor of $m(X)(1 + X^{n'} + X^{2n'} + \ldots + X^{(b-1)n'})$, where $m(X)$ is the minimal generating polynomial of the linear recurring sequence whose minimal polynomial is $m^\perp(X)$ such that $\deg(m^\perp) > k_0 - \deg(f)$ (see Definition 7).

Proof: The proof is given in Appendix C. ∎

We next provide an example of a cyclic code that satisfies the claim of this theorem.

Example 1. *Consider a non-degenerate cyclic code $C(15, g_0)$ with generator polynomial $g_0(X) = (X^4 + X^3 + 1)(X^4 + X^3 + X^2 + X + 1)(X + 1)$ and dimension $k_0 = 6$. The generator polynomial of the dual code of $C(15, g_0)$ is $g_0^\perp(X) = (X^2 + X + 1)(X^4 + X^3 + 1)$. Note that the factor $m^\perp(X) = X^2 + X + 1$ of $g_0^\perp(X)$ has the order $n' = 3$, which is strictly less than $n_0 = 15$. The minimal generating polynomial corresponding to $m^\perp(X) = X^2 + X + 1$ is $m(X) = X + 1$.*

For $n = 9$ and $f(X) = X^6 + X^3 + 1$, by considering all possible codewords in $C(15, g_0)$ it can be checked that the random variable corresponding to $r(X) = \mathbf{w}(X) \bmod f(X)$ follows the restricted uniform distribution[1]. Note that the chosen n and $f(X)$ satisfy the conditions of the theorem as $n = 3n'$, i.e., $b = 3$ and $f(X) = X^6 + X^3 + 1$ is a factor of $m(X)(1 + X^{n'} + X^{2n'} + \ldots + X^{(b-1)n'}) = (X + 1)(1 + X^3 + X^6)$ such that $\deg(m^\perp) > k_0 - \deg(f)$. □

B. Analyzing the distribution of $r'(X)$

In this section, we study the distribution of $r'(X) = \mathbf{w}'(X) \bmod f(X)$, where $\mathbf{w}' \in \mathcal{W}'(n)$ (see (11) and (12)). We first consider the case when $n < n_0$ and $r(X)$ follows

---

[1]For this n and f, the probability of zero syndrome is 0.0625. For the uniform distribution, the probability of zero syndrome would be $1/2^{\deg(f)} = 1/2^6 = 0.015625$.

the uniform distribution. In the following proposition, we will prove that when $r(X)$ follows the uniform distribution, $r'(X)$ also follows the uniform distribution.

Theorem 2. *Suppose assumed length n is strictly less than the true length n_0 of the code. Let $r(X) = \mathbf{w}(X) \bmod f(X)$ and $r'(X) = \mathbf{w}'(X) \bmod f(X)$, where $\mathbf{w}$ and $\mathbf{w}'$ are defined in (10) and (11) respectively. Then the random variable corresponding to $r'(X)$ follows the uniform distribution if the random variable corresponding to $r(X)$ follows the uniform distribution.*

Proof: The proof is given in Appendix D. ∎

We next consider the case when either $n > n_0$ or $r(X)$ follows the restricted uniform distribution. From Proposition 1, we know that $r'(X)$ will follow the uniform distribution or restricted uniform distribution. We now provide the conditions under which $r'(X)$ will follow the uniform and the restricted uniform distributions. From (11), $\mathbf{w}'$ is given by,

$$\mathbf{w}' = \Big[\mathbf{v}_1(n_0 - d_1 : n_0 - 1)\ \mathbf{v}_2\ \cdots\ \mathbf{v}_{q+1}\ \mathbf{v}_{q+2}(0 : d_2 - 1)\Big]$$
$$= \Big[\mathbf{c}_1\ \mathbf{v}_2\ \cdots\ \mathbf{v}_{q+1}\ \mathbf{c}_2\Big],$$
$$\tag{13}$$

where $\mathbf{v}_i \in C(n_0, g_0)$ for $i = 1, 2, \ldots, q+2$, $\mathbf{c}_1 := \mathbf{v}_1(n_0 - d_1 : n_0 - 1)$, and $\mathbf{c}_2 := \mathbf{v}_{q+2}(0 : d_2 - 1)$. From (13), $r'(X)$ is given by,

$$
\begin{aligned}
r'(X) &= \mathbf{w}'(X) \bmod f(X) \\
&= \Big[\mathbf{c}_1(X) + X^{d_1}\mathbf{v}_2(X) + \ldots + X^{d_1+(q-1)n_0}\mathbf{v}_{q+1}(X) \\
&\quad + X^{d_1+qn_0}\mathbf{c}_2(X)\Big] \bmod f(X) \\
&= t_1(X) + t_2(X) + \ldots + t_{q+2}(X),
\end{aligned}
$$
$$\tag{14}$$

where $t_1(X) = \mathbf{c}_1(X) \bmod f(X)$, $t_i(X) = X^{d_1+(i-2)n_0}\mathbf{v}_i(X) \bmod f(X)$, and $t_{q+2}(X) = X^{d_1+qn_0}\mathbf{c}_2(X) \bmod f(X)$, for $i = 2, 3, \ldots, q+1$. The distribution of $t_1(X)$ and $t_{q+2}(X)$ can be studied using Section III-A, since $\mathbf{c}_1(X)$ and $\mathbf{c}_2(X)$ are formed by the consecutive d_1 and d_2 bits of a codeword in $C(n_0, g_0)$. We now study the distribution of $t_i(X)$, for $i = 2, 3, \ldots, q+1$. First note that when $\mathbf{v}(X) \bmod f(X)$ follows the uniform distribution, $X^d\mathbf{v}(X) \bmod f(X)$ also follows the uniform distribution for any positive integer d. Similarly, when $\mathbf{v}(X) \bmod f(X)$ follows the restricted uniform distribution, $X^d\mathbf{v}(X) \bmod f(X)$ also follows the restricted uniform distribution. Hence it is sufficient to study the distribution of $\mathbf{v}(X) \bmod f(X)$.

- When $f(X)$ is a factor of $g_0(X)$, $\mathbf{v}(X) \bmod f(X)$ is zero with probability one, since $\mathbf{v}(X) = \mathbf{u}(X)g_0(X)$ for some $\mathbf{u}(X) \in \mathcal{P}_{k_0}$. Thus $\mathbf{v}(X) \bmod f(X)$ follows the degenerate distribution.
- When $f(X)$ is not a factor of $g_0(X)$ and $\deg(f) \le k_0$, from Theorem 3.2 of [16], $\mathbf{v}(X) \bmod f(X)$ follows the uniform distribution.
- When $f(X)$ is not a factor of $g_0(X)$ and $\deg(f) > k_0$, $\mathbf{v}(X) \bmod f(X)$ follows the restricted uniform distribution since the number of possible values that $r(X)$ can

take are more than the number of possible values $\mathbf{v}(X)$ can take, as explained in Section III-A.

We now study the distribution of $r'(X)$ in the following theorem.

Theorem 3. *Let $r'(X)$ be as defined in (14). Then $r'(X)$ follows the uniform distribution when every $t_i(X)$, for $i = 1, 2, \ldots, q+2$ follows the uniform distribution, otherwise it follows the restricted uniform distribution.*

Proof: The proof is given in Appendix E. ∎

Theorems 1, 2, and 3 completely characterize the distribution of syndromes of the noise-free sequence when either $n \ne ln_0$ or $s \ne s_0$. We next consider the case when $n = ln_0$ and $s = s_0$.

C. The case when $n = ln_0$ and $s = s_0$

When $n = ln_0$ and $s = s_0$, every noise-free n-bit vector $\mathbf{w}_j$ is formed by the concatenation of l codewords of the true code $C(n_0, g_0)$, i.e., any $\mathbf{w}_j$ for $1 \le j \le M$ is given by,

$$\mathbf{w}_j = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \ldots & \mathbf{v}_l \end{bmatrix}, \tag{15}$$

where $\mathbf{v}_i \in C(n_0, g_0)$ for $i = 1, 2, \ldots, l$. Depending on whether $f(X)$ is a factor of $g_0(X)$ or not and the degree of $f(X)$, we have the following cases.

(a) When $f(X)$ is a factor of $g_0(X)$, from (15) $\mathbf{w}_j(X) \bmod f(X)$ is always zero since every $\mathbf{v}_i(X)$, for $i = 1, 2, \ldots, l$ is a multiple of $g_0(X)$. This implies that $\mathbf{w}_j(X) \bmod f(X)$ follows the degenerate distribution.

(b) When $f(X)$ is not a factor of $g_0(X)$ and $\deg(f) \le k_0$, from Theorem 3.2 of [16], $\mathbf{v}_i(X) \bmod f(X)$ follows the uniform distribution, for $i = 1, 2, \ldots, l$. From Theorem 3 this implies that $\mathbf{w}_j(X) \bmod f(X)$ also follows the uniform distribution.

(c) When $f(X)$ is not a factor of $g_0(X)$ and $\deg(f) > k_0$, as explained in the previous section, each $\mathbf{v}_i(X) \bmod f(X)$ follows the restricted uniform distribution and from Theorem 3, $\mathbf{w}_j(X) \bmod f(X)$ follows the restricted uniform distribution.

D. Summary of the distribution of $\mathbf{w}_j(X) \bmod f(X)$

In this section, we summarize the results for the distribution of $\mathbf{w}_j(X) \bmod f(X)$. Depending upon the chosen n, s, and $f(X)$, we have the following cases.

- When $n = ln_0$ for some $l \in \mathbb{N}$, $s = s_0$, and $f(X)$ is factor of $g_0(X)$, $\mathbf{w}_j(X) \bmod f(X)$ follows the degenerate distribution (see Section III-C (a)).
- When either $n \ne ln_0$ or $s \ne s_0$ or $f(X)$ is not a factor of $g_0(X)$, the distribution of $\mathbf{w}_j(X) \bmod f(X)$ is either uniform or restricted uniform. Theorems 1, 2, and 3 and Section III-C (b), (c) provide the conditions when the distribution is uniform or restricted uniform.

IV. SYNDROME DISTRIBUTION OF THE NOISE-AFFECTED RECEIVED SEQUENCE

In the previous section, we studied the distribution of $\mathbf{w}_j(X) \bmod f(X)$, where $1 \le j \le M$. In this section,

we study the distribution of $\mathbf{y}_j(X) \bmod f(X)$, where recall that $\mathbf{y}_j(X)$ is the noise-affected version of $\mathbf{w}_j(X)$ (see (9)). In the previous section, we proved that the distribution of $\mathbf{w}_j(X) \bmod f(X)$ is either degenerate or uniform or restricted uniform. We consider the case when $\mathbf{w}_j(X) \bmod f(X)$ follows each type of the distribution separately and study the distribution of $\mathbf{y}_j(X) \bmod f(X)$. The case when $\mathbf{w}_j(X) \bmod f(X)$ follows the degenerate distribution, i.e., when $n = ln_0$, $s = s_0$, and $f(X)$ is a factor of $g_0(X)$ is studied in detail in [16] and [17]. In the following theorem, we consider that case when $\mathbf{w}_j(X) \bmod f(X)$ follows the uniform distribution.

Theorem 4. *Let* $\mathbf{y}_j(X)$ *and* $\mathbf{w}_j(X)$ *be the* j*th noise-affected received polynomial and error-free polynomial respectively, for* $j = 1, 2, \ldots, M$*. Then* $\mathbf{y}_j(X) \bmod f(X)$ *follows the uniform distribution if* $\mathbf{w}_j(X) \bmod f(X)$ *follows the uniform distribution.*

Proof: Since $\mathbf{w}_j(X) \bmod f(X)$ follows the uniform distribution, it takes any value in $\mathcal{P}_{\deg(f)}$ with probability $1/2^{\deg(f)}$. We now find the probability that $\mathbf{y}_j(X) \bmod f(X)$ takes a value $b(X) \in \mathcal{P}_{\deg(f)}$ as follows.

$$
\mathbb{P}\Big[\mathbf{y}_j(X) \bmod f(X) = b(X)\Big]
$$
$$
= \mathbb{P}\Big[[\mathbf{w}_j(X) + \mathbf{e}_j(X)] \bmod f(X) = b(X)\Big]
$$
$$
= \sum_{\mathbf{d}(X) \in \mathcal{P}_n} \mathbb{P}\Big[[\mathbf{w}_j(X) + \mathbf{d}(X)] \bmod f(X) = b(X)\Big]
$$
$$
\mathbb{P}\Big[\mathbf{e}_j(X) = \mathbf{d}(X)\Big]
$$
$$
= \sum_{\mathbf{d}(X) \in \mathcal{P}_n} \mathbb{P}\Big[\mathbf{w}_j(X) = \mathbf{d}(X) \bmod f(X) + b(X)\Big]
$$
$$
\mathbb{P}\Big[\mathbf{e}_j(X) = \mathbf{d}(X)\Big]
$$
$$
\stackrel{(a)}{=} \sum_{\mathbf{d}(X) \in \mathcal{P}_n} \frac{1}{2^{\deg(f)}}\mathbb{P}\Big[\mathbf{e}_j(X) = \mathbf{d}(X)\Big]
$$
$$
= \frac{1}{2^{\deg(f)}} \sum_{\mathbf{d}(X) \in \mathcal{P}_n} \mathbb{P}\Big[\mathbf{e}_j(X) = \mathbf{d}(X)\Big] = \frac{1}{2^{\deg(f)}}. \quad (16)
$$

where the equality in (a) is obtained since $[\mathbf{d}(X) \bmod f(X) + b(X)]$ is polynomial in $\mathcal{P}_{\deg(f)}$ and $\mathbf{w}_j(X) \bmod f(X)$ takes any value in $\mathcal{P}_{\deg(f)}$ with probability $1/2^{\deg(f)}$. From (16), $\mathbf{y}_j(X) \bmod f(X)$ follows the uniform distribution and the proof is complete. ∎

We now consider the case when $\mathbf{w}_j(X) \bmod f(X)$ follows the restricted uniform distribution. Let us first consider an example distribution of $\mathbf{y}_j(X) \bmod f(X)$ when $\mathbf{w}_j(X) \bmod f(X)$ follows the restricted uniform distribution.

Example 2. *Suppose code* $C(n_0, g_0)$ *with* $n_0 = 15$ *and* $g_0(X) = (X^4 + X + 1)(X^4 + X^3 + 1)$ *is used at the transmitter. For* $n = 10$*, suppose* n*-bit vector* $\mathbf{w}_j$ *is formed by the initial* $n = 10$ *bits of a codeword in* $C(n_0, g_0)$*. For a factor* $f(X) = X^4 + X^3 + X^2 + X + 1$ *of* $X^{10} + 1$*, the distributions of* $\mathbf{w}_j(X) \bmod f(X)$ *and* $\mathbf{y}_j(X) \bmod f(X)$ *are shown in Fig. 4(a) and (b) respectively. It can be seen that,* $\mathbf{w}_j(X) \bmod f(X)$ *follows the restricted uniform distribution*

(a) Distribution of $\mathbf{w}_j(X) \bmod f(X)$

(b) Distribution of $\mathbf{y}_j(X) \bmod f(X)$

Fig. 4. The distributions of $\mathbf{w}_j(X) \bmod f(X)$ and $\mathbf{y}_j(X) \bmod f(X)$ are illustrated when $\mathbf{w}_j(X)$ is formed by the initial 10-bits of a codeword in code $C(15, g_0)$ with $g_0(X) = (X^4 + X + 1)(X^4 + X^3 + 1)$ and $f(X) = X^4 + X^3 + X^2 + X + 1$.

but the distribution of $\mathbf{y}_j(X) \bmod f(X)$ *is neither uniform nor restricted uniform.* ∎

Example 2 suggests that, when $\mathbf{w}_j(X) \bmod f(X)$ follows the restricted uniform distribution, the distribution of $\mathbf{y}_j(X) \bmod f(X)$ need not be uniform or restricted uniform. Let $\mathcal{S}$ be the support set of $\mathbf{w}_j(X) \bmod f(X)$. In Example 2, the support set of $\mathbf{w}_j(X) \bmod f(X)$ is $\mathcal{S} = \{0, X^2, X^3 + 1, X^3 + X^2 + 1\}$ (see Fig. 4(a)). From the definition of the restricted uniform distribution, for any $a(X) \in \mathcal{S}$,

$$
\mathbb{P}\Big[\mathbf{w}_j(X) \bmod f(X) = a(X)\Big] = \frac{1}{|\mathcal{S}|}. \quad (17)
$$

The probability that $\mathbf{y}_j(X) \bmod f(X)$ takes the value $b(X) \in \mathcal{P}_{\deg(f)}$ is given by,

$$
\mathbb{P}\Big[\mathbf{y}_j(X) \bmod f(X) = b(X)\Big]
$$
$$
= \mathbb{P}\Big[[\mathbf{w}_j(X) + \mathbf{e}_j(X)] \bmod f(X) = b(X)\Big]
$$
$$
= \mathbb{P}\Big[\mathbf{e}_j(X) \bmod f(X) = [\mathbf{w}_j(X) \bmod f(X)] + b(X)\Big]
$$
$$
\stackrel{(a)}{=} \sum_{a(X) \in \mathcal{S}} \mathbb{P}\Big[\mathbf{e}_j(X) \bmod f(X) = a(X) + b(X)\Big]
$$
$$
\mathbb{P}\Big[\mathbf{w}_j(X) \bmod f(X) = a(X)\Big]
$$
$$
= \frac{1}{|\mathcal{S}|} \sum_{a(X) \in \mathcal{S}} \mathbb{P}\Big[\mathbf{e}_j(X) \bmod f(X) = a(X) + b(X)\Big],
$$
$$
\quad (18)
$$

where the equality in (a) is obtained by conditioning over the support set $\mathcal{S}$ of $\mathbf{w}_j(X) \bmod f(X)$ and the last equality is obtained from (17). For $a_1(X), a_2(X) \in \mathcal{S}$ since $a_1(X) + a_2(X) \in \mathcal{S}$, from (18) we get

$$
\mathbb{P}\Big[\mathbf{y}_j(X) \bmod f(X) = a_1(X)\Big]
$$
$$
= \mathbb{P}\Big[\mathbf{y}_j(X) \bmod f(X) = a_2(X)\Big]. \quad (19)
$$

From (19), $\mathbf{y}_j(X) \bmod f(X)$ takes any value in $\mathcal{S}$ with the equal probability. In Example 2, it can be seen that the probability of observing any two syndromes in $\mathcal{S}$ is the same.

In Fig. 4(b), $\mathbf{y}_j(X) \bmod f(X)$ takes the two values X^2 and $X^3 + X^2 + 1$ in $\mathcal{S}$ with equal probability. However, for calculating the value of $\mathbb{P}[\mathbf{y}_j(X) \bmod f(X) = b(X)]$ for any $b(X) \in \mathcal{P}_{\deg(f)}$ would require the knowledge of the support set $\mathcal{S}$ and the coset weight distribution of code $C(n,f)$ (see (18)). Since finding the coset weight distribution is NP-hard and the knowledge of the support set $\mathcal{S}$ would require the knowledge of the unknown true code $C(n_0, g_0)$, finding the value of $\mathbb{P}[\mathbf{y}_j(X) \bmod f(X) = b(X)]$ is in general computationally intractable. Thus finding the distribution of $\mathbf{y}_j(X) \bmod f(X)$ when $\mathbf{w}_j(X) \bmod f(X)$ follows the restricted uniform distribution is computationally intractable.

V. APPLICATION TO BLIND RECONSTRUCTION OF CYCLIC CODES

In the literature, Yardi et al. [17] and Zhou et al. [18], [19] have proposed blind reconstruction methods when both the length of the cyclic code and the synchronization of the received data are not known. In this section, we provide a theoretical analysis of these methods.

A. A theoretical analysis of the blind reconstruction method proposed in [17]

Yardi et al. have proposed the zero syndrome distribution based method for blind reconstruction [17]. In this method, authors make use of the zero syndromes of the received polynomials. Suppose $r_j(X) = \mathbf{y}_j(X) \bmod f(X)$ for $j = 1, 2, \ldots, M$. They proved that, for a given n, s, and $f(X)$ there are either of the following two cases (see Theorem 1 of [17]).

(1) When $n = ln_0$ such that $l \in \mathbb{N}$, $s = s_0$, and $f(X)$ is a factor of $g_0(X)$,
$$\mathbb{P}[r_j(X) = 0] = P(C(n,f)),$$
for $j = 1, 2, \ldots, M$ and $P(C(n,f))$ is defined as,
$$P(C(n,f)) := \sum_{i=0}^{n} A_i p^i (1-p)^{n-i} \quad (20)$$
where $\{A_0, A_1, \cdots, A_n\}$ is the weight distribution of $C(n,f)$.

(2) When either $n \neq ln_0$ or $s \neq s_0$ or $f(X)$ is a not factor of $g_0(X)$,
$$\mathbb{P}[r_j(X) = 0] < P(C(n,f)),$$
for $j = 1, 2, \ldots, M$.

Using (1) and (2), they formulated and solved the blind reconstruction problem via hypothesis testing problem given by,
$$H_0 : \mathbb{I}_{\{r_j(X)=0\}} \sim \text{Bernoulli}\Big(P(C(n,f))\Big)$$
$$H_1 : \mathbb{I}_{\{r_j(X)=0\}} \sim \text{Bernoulli}\Big(P_j\Big) \text{ s.t. } P_j < P(C(n,f)),$$
$$(21)$$
where $j = 1, 2, \ldots, M$ and $\mathbb{I}_{\{r_j(X)=0\}}$ is the indicator random variable for the event $r_j(X) = 0$. For analyzing the performance of this method, one needs to analyze the performance

of the hypothesis testing in (21). The performance of the hypothesis testing can be characterized using the Kullback-Leibler (KL) divergence between the two distributions [27, Ch. 11]. However in (21), the distribution under hypothesis H_1 is not known in general. From Theorem 4, when $\mathbf{w}_j(X) \bmod f(X)$ follows the uniform distribution, $\mathbf{y}_j(X) \bmod f(X)$ also follows the uniform distribution and hence in this case under hypothesis H_1 we have,
$$H_1 : \mathbb{I}_{\{r_j(X)=0\}} \sim \text{Bernoulli}\left(\frac{1}{2^{\deg(f)}}\right). \quad (22)$$

When $\mathbf{w}_j(X) \bmod f(X)$ follows the restricted uniform distribution, due to the reasons mentioned in Section IV, characterizing the distribution of $r_j(X) = \mathbf{y}_j(X) \bmod f(X)$ is computationally intractable. Hence in the following theorem we provide an upper bound on $\mathbb{P}[r_j(X) = 0]$ which is strictly less than $P(C(n,f))$.

Theorem 5. *Let $\mathbf{y}_1(X), \mathbf{y}_2(X), \ldots, \mathbf{y}_M(X)$ be the sequence of received polynomials for an assumed length n and synchronization (see Section II). For a factor $f(X)$ of X^n+1, suppose $r_j(X) = \mathbf{y}_j(X) \bmod f(X)$, for $j = 1, 2, \ldots, M$. When either the assumed length n is not a multiple of the correct length n_0 or synchronization is not correct or $f(X)$ is not a factor of the generator polynomial $g_0(X)$ of the code $C(n_0, g_0)$ used at the transmitter,*
$$\mathbb{P}[r_j(X) = 0] \leq \mathcal{P}\big(C(n,f)\big)\left(\frac{\lambda+1}{2}\right), \quad (23)$$
where the expression for $\mathcal{P}(C(n,f))$ is given in (20) and λ is defined as follows
$$\lambda := \frac{1 - (1-2p)^{n-\deg(f)+1}}{1 + (1-2p)^{n-\deg(f)+1}}. \quad (24)$$

Proof: The proof is given in Appendix F. ∎

Using Theorem 5, we now find a lower bound on the KL-divergence between the two distributions in (21) as follows. Let P and Q denote the pmf of $\mathbb{I}_{\{r_j(X)=0\}}$ under hypothesis H_0 and H_1 respectively. Suppose $P := \begin{bmatrix} p_0 & p_1 \end{bmatrix}$ and $Q := \begin{bmatrix} q_0 & q_1 \end{bmatrix}$, where p_0 and q_0 are the probabilities of observing the all-zero syndrome under hypotheses H_0 and H_1 respectively. A lower bound on the KL-divergence $D_{KL}(P,Q)$ between distributions P and Q is given by [27, Sec. 11.6],
$$D_{KL}(P,Q) \geq \frac{1}{2\ln 2}\Big(|p_0 - q_0| + |p_1 - q_1|\Big)^2. \quad (25)$$
Substituting $p_1 = 1 - p_0$ and $q_1 = 1 - q_0$ in (25) we get,
$$D_{KL}(P,Q) \geq \frac{1}{2\ln 2}\Big(|p_0 - q_0| + |(1-p_0) - (1-q_0)|\Big)^2 \quad (26)$$
$$= \frac{1}{2\ln 2}\Big(|p_0 - q_0| + |q_0 - p_0|\Big)^2 \quad (27)$$
$$\overset{(a)}{=} \frac{1}{2\ln 2}\Big(2|p_0 - q_0|\Big)^2 \quad (28)$$
$$= \frac{2}{\ln 2}\big(p_0 - q_0\big)^2, \quad (29)$$
where the equality in (a) is obtained since $|p_0 - q_0| = |q_0 - p_0|$ and the last equality is obtained since $|p_0 - q_0|^2 = (p_0 - q_0)^2$.

From Theorem 5 we have, $q_0 \leq p_0(\lambda+1)/2$ and substituting this in (29) we get,

$$D_{KL}(P,Q) \geq \frac{2}{\ln 2}\left[p_0 - \left(\frac{p_0(\lambda+1)}{2}\right)\right]^2, \quad (30)$$

$$= \frac{2}{\ln 2}\left[p_0\left(\frac{1-\lambda}{2}\right)\right]^2 \quad (31)$$

$$= \frac{2}{\ln 2}\left(\frac{1-\lambda}{2}\right)^2\left(P(C(n,f))\right)^2, \quad (32)$$

where the last equality is obtained by substituting $p_0 = P(C(n,f))$. From (32), we obtain a lower bound on the KL-divergence between two distributions in (21).

To summarize, depending on whether $\mathbf{w}_j(X) \bmod f(X)$ follows the uniform or the restricted uniform distribution, we can characterize the distribution of $\mathbb{I}_{\{r_j(X)=0\}}$ using (22) and (23), which is required for analyzing the performance of the hypothesis testing of (21) in the zero syndrome distribution based method.

B. A theoretical analysis of the blind reconstruction method proposed in [18]

Zhou et al. have proposed the factor-entropy based method for blind reconstruction of binary cyclic codes [18]. The basic idea of this method is as follows. Suppose H is a parity check matrix of the code $C(n,f)$ generated by a factor $f(X)$ of $X^n + 1$. They consider the sequence received vectors $\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_M$ and find the inner product of each $\mathbf{y}_j$ with H given by,

$$\mathbf{y}_j H^T = \begin{bmatrix} r_{j,0} & r_{j,1} & \ldots & r_{j,\deg(f)-1} \end{bmatrix}. \quad (33)$$

They define the *mean value of probability of zero syndrome* $P(\mathbf{y}_j, f)$ as,

$$P(\mathbf{y}_j, f) := \frac{1}{\deg(f)}\sum_{l=0}^{\deg(f)-1}\mathbb{P}[r_{j,l}=0]. \quad (34)$$

For blind reconstruction they assume that, when either n or s is incorrect,

$$P(\mathbf{y}_j, f_1) = P(\mathbf{y}_j, f_2), \quad (35)$$

where $f_1(X)$ and $f_2(X)$ are any two factors of $X^n + 1$. They further assume that when $n = n_0$ and $s = s_0$ the assumption in (35) is not valid. The correct length and synchronization are distinguished from any incorrect ones using this assumption.

In this section, we verify the validity of the assumption in (35). We next illustrate an example situation where for an incorrect s, $P(\mathbf{y}_j, f_1) \neq P(\mathbf{y}_j, f_2)$ which implies that the assumption in (35) is not correct.

Example 3. *Suppose the cyclic code $C(n_0, g_0)$ with $n_0 = 7$ and $g(X) = X^3 + X + 1$ is used for the communication. Let us assume that $s_0 = 0$ is the correct synchronization. For $n = 7$ and $s = 1$, the values of $P(\mathbf{y}_j, f)$ for all possible factors of X^7+1 are provided in Table I, for $j = 1, 2, \ldots, M$. It can be seen that, for the chosen incorrect parameters, the assumption in (35) is not valid. In this example, note that the assumed length n was correct, $f(X)$ was a factor of $g_0(X)$,*

$f(X)$	$P(\mathbf{y}_j,f)$ for $p=0$	$P(\mathbf{y}_j,f)$ for $p=0.01$	$P(\mathbf{y}_j,f)$ for $p=0.05$
$X+1$	0.5	0.5	0.5
X^3+X+1	0.8334	0.8076	0.7184
X^3+X^2+1	0.5	0.5	0.5

TABLE I
THE VALUES OF $P(\mathbf{y}_j, f)$ FOR ALL POSSIBLE FACTORS OF $X^7 + 1$ ARE ILLUSTRATED WHEN $n = 7$ AND $s = 1$.

but the assumed synchronization was not correct. In general, when $n = n_0$ but $s \neq s_0$, the assumption in (35) need not be true. □

Though the assumption in (35) is not valid always, in the following theorem we prove that, the assumption in (35) is true when $C(n_0, g_0)$ not degenerate and the assumed length $n < n_0$.

Theorem 6. *Suppose the true code $C(n_0, g_0)$ used at the transmitter is not degenerate. For an assumed length n and synchronization s, let $\mathbf{y}_j$ and $\mathbf{w}_j$ be the jth noise-affected and noise-free n-bits vectors respectively, for $j = 1, 2, \ldots, M$. For a factor $f(X)$ of $X^n + 1$, let $P(\mathbf{y}_j, f)$ be as defined in (34). If $n < n_0$, we have $P(\mathbf{y}_j, f_1) = P(\mathbf{y}_j, f_2)$, where $f_1(X)$ and $f_2(X)$ are any two factors of $X^n + 1$.*

Proof: The proof is given in Appendix G. ■

C. A theoretical analysis of blind reconstruction method proposed in [19]

In [19], Zhou et al. have proposed the root-entropy based method for blind reconstruction. In this method, for an assumed length n and synchronization s, authors consider the received polynomials $\mathbf{y}_1(X), \mathbf{y}_2(X), \ldots, \mathbf{y}_M(X)$. They find an empirical probability of each root β of $X^n + 1$ being a root of the received polynomials. They assume that, when either n or s is incorrect, all possible roots of $X^n + 1$ are equally likely to be roots of the received polynomials, i.e., for any two roots β_1 and β_2 of $X^n + 1$,

$$\mathbb{P}\Big[\beta_1 \text{ is a root of } \mathbf{y}_j(X)\Big] = \mathbb{P}\Big[\beta_2 \text{ is a root of } \mathbf{y}_j(X)\Big], \quad (36)$$

for $j = 1, 2, \ldots, M$. In Example 4, we provide an example situation when this assumption is not true.

Example 4. *Suppose code $C(n_0, g_0)$ with $n_0 = 15$ and $g_0(X) = (X^4+X^3+1)(X^4+X^3+X^2+X+1)(X+1)$ is used at the transmitter and $s_0 = 0$ is the correct synchronization of the received sequence. For an assumed length $n = 7$ and synchronization $s = 0$, the first n-bit received vector $\mathbf{y}_1$ will be $\mathbf{y}_1 = \mathbf{w}_1 + \mathbf{e}_1$, where $\mathbf{w}_1$ is formed by the initial 7 bits of a codeword in code $C(15, g_0)$. Consider two roots β_1 and β_2 of $X^7 + 1$, whose minimal polynomials are $f_1(X) = X + 1$ and $f_2(X) = X^3 + X + 1$ respectively. It is known that, β_i is a root of $\mathbf{y}_1(X)$ if and only if the minimal polynomial $f_i(X)$ of β_i is a factor of $\mathbf{y}_1(X)$, for $i = 1, 2$ [23, Sec. 2.2]. Thus the probability of a given β_i is a root of $\mathbf{y}_1(X)$ is the same as that of the probability that $f_i(X)$ is a factor of $\mathbf{y}_1(X)$. For a factor $f_i(X)$ of $X^7 + 1$, the probability that $f_i(X)$ is a*

factor of $\mathbf{y}_1(X)$ *can be found by conditioning over all possible codewords in* $C(15, g_0)$. *For two factors* $X^3 + X + 1$ *and* $X + 1$ *of* $X^7 + 1$, *it can be verified that for any value of crossover probability* p,

$$\mathbb{P}\Big[\beta_1 \text{ is a root of } \mathbf{y}_1(X)\Big] = \frac{1}{2^{\deg(f)}} = 0.125$$

$$\mathbb{P}\Big[\beta_2 \text{ is a root of } \mathbf{y}_1(X)\Big] = \frac{1}{2^{\deg(f)}} = 0.5.$$

It can be seen that, the assumption in (36) is not valid for $\mathbf{y}_1(X)$. $\qquad\square$

VI. CONCLUSION

In this paper, we analyzed the syndrome distribution of the noise-free and noise-affected received sequence. For the noise-free case, we completely characterized the syndrome distribution of the received sequence. We proved that the distribution of syndrome of any noise-free received polynomial with respect to a candidate polynomial $f(X)$ is degenerate if and only if the assumed length is an integer multiple of the correct length, assumed synchronization is correct, and $f(X)$ is a factor of the generator polynomial of the true code. We proved that, in all the remaining cases the distribution can either be uniform or restricted uniform. We also provided the conditions under which this distribution will be of either of the type. For the noise-affected situation we observed that, while the syndrome distribution could be completely characterized for some of the assumed parameters, in general finding this distribution becomes computationally intractable. Finally, we provided a theoretical analysis of the existing methods available in the literature for blind reconstruction.

APPENDIX A: SOME PROPERTIES OF LINEAR BLOCK CODES

Lemma 1. *Consider a non-trivial linear block code* $C(n)$ *of length* n *and dimension* k. *Every codeword in this code is chosen according to the uniform distribution. Consider a codeword* $\mathbf{v} \in C(n)$ *and a vector* $\mathbf{h} \in \mathbb{F}_2^n$. *Then the inner product* $\mathbf{v}\mathbf{h}^T$ *is equally likely to be zero or one if and only if* $\mathbf{h} \notin C^\perp(n)$.

Proof: Suppose the inner product $\mathbf{v}\mathbf{h}^T$ is equally likely to be zero or one. If $\mathbf{h} \in C^\perp(n)$, then the inner product $\mathbf{v}\mathbf{h}^T$ will always be zero, which is a contradiction. This implies that $\mathbf{h} \notin C^\perp(n)$ and the proof is complete.

We now prove the converse. Suppose $\mathbf{h} \notin C^\perp(n)$. Suppose $\langle C^\perp(n), \mathbf{h} \rangle$ denotes the subspace spanned by a set of linearly independent vectors of $C^\perp(n)$ and $\mathbf{h}$. In this case we have $C^\perp(n) \subseteq \langle C^\perp(n), \mathbf{h} \rangle$ and this implies that,

$$\left\langle C^\perp(n), \mathbf{h} \right\rangle^\perp \subseteq \left(C^\perp(n) \right)^\perp = C(n). \tag{37}$$

Since the dimension of $C(n)$ is k, the dimensions of $C^\perp(n)$ and $\langle C^\perp(n), \mathbf{h} \rangle$ will be $n-k$ and $n-k+1$ respectively. The dimension of $\langle C^\perp(n), \mathbf{h} \rangle^\perp$ is $n - (n-k+1) = k-1$ and hence in (37) we have,

$$\left\langle C^\perp(n), \mathbf{h} \right\rangle^\perp \subset C(n). \tag{38}$$

Since $\mathbf{h} \notin C^\perp(n)$ and the dimension of $\langle C^\perp(n), \mathbf{h} \rangle^\perp$ is exactly one less than the dimension of $C(n)$, from (38), the inner product $\mathbf{v}\mathbf{h}^T$ will be zero for exactly 2^{k-1} number of codewords in $C(n)$. Since every $\mathbf{v} \in C(n)$ is chosen according to the uniform distribution, $\mathbf{v}\mathbf{h}^T$ will be equally likely to be zero or one and the proof is complete. $\qquad\blacksquare$

Lemma 2. *Consider a linear block code* $C(n)$ *and a vector* $\mathbf{h} \notin C^\perp(n)$. *For positive integers* d_1 *and* d_2, *let* $C_1(d_1)$ *and* $C_2(d_2)$ *be the linear subspaces formed by the set of prefixes and suffixes of codewords in* $C(n)$ *of lengths* d_1 *and* d_2 *respectively, where* $1 \leq d_1, d_2 < n$ *such that* $d_1 + d_2 = n$. *Then either of the following is true.*
(1) $\mathbf{h}(0 : d_1 - 1) \notin C_1^\perp(d_1)$
(2) $\mathbf{h}(d_1 : n - 1) \notin C_2^\perp(d_2)$

Proof: When $\mathbf{h}(0 : d_1 - 1) \notin C_1^\perp(d_1)$, condition (1) of the lemma is satisfied and the lemma is trivially true. Hence we consider the case when $\mathbf{h}(0 : d_1 - 1) \in C_1^\perp(d_1)$. This implies that, for any $\mathbf{v} \in C(n)$,

$$\mathbf{v}(0 : d_1 - 1)\mathbf{h}(0 : d_1 - 1)^T = 0. \tag{39}$$

We now prove by contradiction that $\mathbf{h}(d_1 : n - 1) \notin C_2^\perp(d_2)$. When $\mathbf{h}(d_1 : n - 1) \in C_2^\perp(d_2)$, for any $\mathbf{v} \in C(n)$,

$$\mathbf{v}(d_1 : n - 1)\mathbf{h}(d_1 : n - 1)^T = 0. \tag{40}$$

From (39) and (40), $\mathbf{v}\mathbf{h}^T$ will always be zero. Since it is given that $\mathbf{h} \notin C^\perp(n)$, we get a contradiction from Lemma 1. Hence $\mathbf{h}(d_1 : n - 1) \notin C_2^\perp(d_2)$. Thus the condition (2) of the lemma is satisfied and the proof is complete. $\qquad\blacksquare$

Lemma 3. *Consider a cyclic code* $C(n, g)$ *of length* n *and generator polynomial* $g(X)$. *Let* $g^\perp(X)$ *be the generator polynomial of the dual code of* $C(n, g)$. *Then* $C(n, g)$ *contains a codeword of a degenerate pattern if and only if there exists a factor of* $g^\perp(X)$ *whose order is strictly less than* n *(see Definitions 3, 6).*

Proof: Suppose $g^\perp(X)$ has a factor $f^\perp(X)$ of order n' such that $1 \leq n' < n$. Let $f(X)$ be the generator polynomial of the dual code of $C(n, f^\perp)$. It is known that $C(n, f) \subseteq C(n, g)$ [25, Sec. 7.4]. Since the order of $f^\perp(X)$ is strictly less than n, $C(n, f)$ will be a degenerate code [22, Sec. 8.3] and the proof is complete since $C(n, f) \subseteq C(n, g)$.

We now prove the converse. Suppose there exists a codeword $\mathbf{v} \in C(n, g)$ of a degenerate pattern, i.e., $\mathbf{v}$ can be written as,

$$\mathbf{v} = \Big[\underbrace{\mathbf{w} \ \ \mathbf{w} \ \ \cdots \ \ \mathbf{w}}_{l \text{ times}}\Big], \tag{41}$$

where $l > 1$ and $\mathbf{w}$ is a vector of length $n' = n/l$ such that $\mathbf{w}$ is not a vector of a degenerate pattern. Since $\mathbf{v}$ is a codeword in a cyclic code, the vector $\mathbf{v}^{(i)}$ obtained by i right cyclic shifts of $\mathbf{v}$ will also be a codeword in $C(n, g)$ given by,

$$\mathbf{v}^{(i)} = \Big[\underbrace{\mathbf{w}^{(i)} \ \ \mathbf{w}^{(i)} \ \ \cdots \ \ \mathbf{w}^{(i)}}_{l \text{ times}}\Big], \tag{42}$$

where $\mathbf{w}^{(i)}$ is the vector obtained by i right cyclic shifts of $\mathbf{w}$ and $1 \leq i < n$. From (41), vector $\mathbf{v}^{(n')}$ obtained by n' right

cyclic shifts of $\mathbf{v}$ will be equal to $\begin{bmatrix} \mathbf{w} & \mathbf{w} & \ldots & \mathbf{w} \end{bmatrix} = \mathbf{v}$. Thus the set of codewords $\{\mathbf{v}^{(1)}, \mathbf{v}^{(2)}, \ldots, \mathbf{v}^{(n')} = \mathbf{v}\}$ will be distinct. It can be easily shown that the subspace spanned by $\{\mathbf{v}^{(1)}, \mathbf{v}^{(2)}, \ldots, \mathbf{v}^{(n')}\}$ is a cyclic code. Let $f(X)$ be the generator polynomial of this code, denoted by $C(n, f)$. Observe that every codeword in code $C(n, f)$ is of a degenerate pattern (see (42)), i.e., $C(n, f)$ is a degenerate cyclic code such that $C(n, f) \subseteq C(n, g)$. This implies that $C(n, f^\perp) \supseteq C(n, g^\perp)$ [1] and $f^\perp(X)$ is a factor of $g^\perp(X)$.

From (42), the period of the linear recurring sequence given by $\begin{bmatrix} \mathbf{w}^{(i)} & \mathbf{w}^{(i)} & \cdots \end{bmatrix}$ is n', which implies that $f(X)$ divides $X^{n'} + 1$ [23, Sec. 3.1]. Since $n' < n$, the order of $f^\perp(X)$ is strictly less than n and the proof is complete. ∎

APPENDIX B: PROOF OF PROPOSITION 1

In this appendix, we will prove Proposition 1. We first summarize some properties of syndrome $r(X)$ that will be required to prove this proposition.

Property 1. *Let $\mathcal{W}$ be a linear subspace of $\mathbb{F}_2^n$. Let $\mathbf{w}(X)$ be the polynomial corresponding to $\mathbf{w} \in \mathcal{W}$. For a polynomial $f(X) \in \mathbb{F}_2[X]$, suppose the syndrome $r(X)$ of $\mathbf{w}(X)$ with respect to $f(X)$ is given by,*

$$r(X) = \mathbf{w}(X) \bmod f(X)$$
$$= r_0 + r_1 X + \ldots + r_{\deg(f)-1} X^{\deg(f)-1}, \quad (43)$$

where each $r_l \in \mathbb{F}_2$, for $l = 0, 1, \ldots, \deg(f) - 1$. Then for every coefficient r_l of $r(X)$, there exists a vector $\mathbf{h}_l \in \mathbb{F}_2^n$ such that

$$r_l = \mathbf{w} \mathbf{h}_l^T,$$

where $l = 0, 1, \ldots, \deg(f) - 1$. □

Proof. For polynomial $f(X)$, define the map L acting on $\mathbf{w} \in \mathcal{W}$ as follows,

$$L(\mathbf{w}) := \mathbf{w}(X) \bmod f(X) = r(X). \quad (44)$$

It can be seen that L is a linear map. Let $\mathbf{r}$ be the vector corresponding to $r(X)$, where $\mathbf{r} \in \mathbb{F}_2^{\deg(f)}$. Since $\mathbf{w}$ and $\mathbf{r}$ are in one-to-one correspondence with $\mathbf{w}(X)$ and $r(X)$ respectively, the linear map L can be given by

$$L : \mathbb{F}_2^n \to \mathbb{F}_2^{\deg(f)}. \quad (45)$$

It is known that, corresponding to every linear transformation $L : \mathbb{F}_2^n \to \mathbb{F}_2^{\deg(f)}$, there exists some matrix $A \in \mathbb{F}_2^{n \times \deg(f)}$ associated to it such that

$$L(\mathbf{w}) = \mathbf{w} A = \mathbf{r}, \quad (46)$$

where $\mathbf{w} \in \mathbb{F}_2^n$ and $\mathbf{r} \in \mathbb{F}_2^{\deg(f)}$ are considered as row vectors [28, Ch. 4]. Suppose matrix A is given by,

$$A = \begin{bmatrix} \mathbf{h}_0 & \mathbf{h}_1 & \cdots & \mathbf{h}_{\deg(f)-1} \end{bmatrix}, \quad (47)$$

where $\mathbf{h}_l \in \mathbb{F}_2^n$ for $l = 0, 1, \ldots, \deg(f) - 1$ are the columns of matrix A. From (46) and (47) it can be seen that the lth coefficient r_l of $\mathbf{r}$ can be written as $\mathbf{w} \mathbf{h}_l^T$ and the proof is complete. □

Let us consider an example to explain this property.

Example 5. *For $n = 7$ and $f(X) = X^3 + X^2 + 1$, $r(X)$ is given by*

$$r(X) = \mathbf{w}(X) \bmod f(X) = r_0 + r_1 X + r_2 X^2$$
$$= (w_0 + w_1 X + \ldots + w_6 X^6) \bmod (X^3 + X^2 + 1)$$
$$= \left[w_0 + w_3 + w_4 + w_5 \right] + \left[w_1 + w_4 + w_5 + w_6 \right] X$$
$$\qquad + \left[w_2 + w_3 + w_4 + w_6 \right] X^2$$
$$= \mathbf{w} \mathbf{h}_0^T + \mathbf{w} \mathbf{h}_1^T X + \mathbf{w} \mathbf{h}_2^T X^2, \quad (48)$$

where $\mathbf{h}_0 = [1\ 0\ 0\ 1\ 1\ 1\ 0]$, $\mathbf{h}_1 = [0\ 1\ 0\ 0\ 1\ 1\ 1]$ and $\mathbf{h}_2 = [0\ 0\ 1\ 1\ 1\ 0\ 1]$. □

Property 2. *When $f(X)$ is a factor of $X^n + 1$, the set of vectors $\{\mathbf{h}_0, \mathbf{h}_1, \ldots, \mathbf{h}_{\deg(f)-1}\}$ in Property 1 form a basis for the dual code $C(n, f^\perp)$ of cyclic code $C(n, f)$.* □

We first provide an example of this property and then provide a proof.

Example 6. *In Example 5, $f(X) = X^3 + X^2 + 1$ is a factor of $X^7 + 1$. The generator polynomial of the dual code of $C(7, f)$ is $f^\perp(X) = (X + 1)(X^3 + X^2 + 1)$. In (48), the polynomials corresponding to $\mathbf{h}_0, \mathbf{h}_1$, and $\mathbf{h}_2$ are given by, $\mathbf{h}_0(X) = (X + 1) f^\perp(X)$, $\mathbf{h}_1(X) = X(X + 1) f^\perp(X)$, and $\mathbf{h}_2(X) = X^2 f^\perp(X)$ respectively. It can be seen that $\mathbf{h}_0, \mathbf{h}_1, \mathbf{h}_2$ are linearly independent and form a basis of $C(7, f^\perp)$.* □

Proof of Property 2: For $\mathbf{w}(X) \in \mathcal{P}_n$, it is known that $\mathbf{w}(X) \in C(n, f)$ if and only if $r(X) = \mathbf{w}(X) \bmod f(X) = 0$ [1]. From (44), $r(X) = 0$ implies that $r_l = \mathbf{w} \mathbf{h}_l^T = 0$, for $l = 0, 1, \ldots, \deg(f) - 1$. The inner product $\mathbf{w} \mathbf{h}_l^T = 0$ for every $\mathbf{w} \in C(n, f)$ implies that $\mathbf{h}_l \in C(n, f^\perp)$, where $C(n, f^\perp)$ is the dual code of $C(n, f)$. Using this, the code $C(n, f)$ is given by,

$$C(n, f) = \left\{ \mathbf{w} \in \mathbb{F}_2^n \middle| \mathbf{w} \mathbf{h}_l^T = 0 \text{ for } l = 0, 1, \ldots, \deg(f) - 1 \right\}. \quad (49)$$

We now prove by contradiction that the set of vectors $\{\mathbf{h}_0, \mathbf{h}_1, \ldots, \mathbf{h}_{\deg(f)-1}\}$ in (49) are independent which completes the proof of the property. Without loss of generality suppose $\mathbf{h}_0$ can be written a linear combination of $\{\mathbf{h}_1, \ldots, \mathbf{h}_{\deg(f)-1}\}$ given by,

$$\mathbf{h}_0 = a_1 \mathbf{h}_1 + a_2 \mathbf{h}_2 + \ldots + a_{\deg(f)-1} \mathbf{h}_{\deg(f)-1}, \quad (50)$$

where $a_i \in \mathbb{F}_2$ for $i = 1, 2, \ldots, \deg(f) - 1$. From (50), if $\mathbf{w} \mathbf{h}_i^T = 0$ for $i = 1, 2, \ldots, \deg(f) - 1$ we get $\mathbf{w} \mathbf{h}_0^T = 0$. Using this in (49) we get,

$$C(n, f) = \left\{ \mathbf{w} \in \mathbb{F}_2^n \middle| \mathbf{w} \mathbf{h}_i^T = 0 \text{ for } i = 1, \ldots, \deg(f) - 1 \right\}. \quad (51)$$

From (51), the dimension of $C(n, f)$ should be greater than or equal to $n - \deg(f) + 1$. This is a contradiction since the dimension of $C(n, f)$ is equal to $n - \deg(f)$ [1]. This completes the proof. ∎

Using Properties 1 and 2 we now characterize the distribution of $r(X) = \mathbf{w}(X) \bmod f(X)$, when $\mathbf{w}(X)$ lies in any linear subspace $\mathcal{W}(n)$ in the following lemma.

Lemma 4. *Consider a linear subspace $\mathcal{W}(n)$ of $\mathbb{F}_2^n$. Suppose every $\mathbf{w}(X) \in \mathcal{W}(n)$ is chosen i.i.d. according to the uniform distribution. For a factor $f(X)$ of $X^n + 1$, suppose $r(X) = \mathbf{w}(X) \bmod f(X)$. Then the distribution of the random variable corresponding to $r(X)$ can either be degenerate or uniform or restricted uniform (see (6), (7), (8), Fig. 2).*

Proof: Suppose $r(X) = \mathbf{w}(X) \bmod f(X)$ is given by,

$$
\begin{aligned}
r(X) &= r_0 + r_1 X + \ldots + r_{\deg(f)-1} X^{\deg(f)-1} \\
&= \mathbf{w}\mathbf{h}_0^T + \mathbf{w}\mathbf{h}_1^T X + \ldots + \mathbf{w}\mathbf{h}_{\deg(f)-1}^T X^{\deg(f)-1},
\end{aligned} \tag{52}
$$

where the last equality is obtained from Property 1 such that each $\mathbf{h}_l \in \mathbb{F}_2^n$ for $l = 0, 1, \ldots, \deg(f) - 1$. Let $\{R_0, R_1, \ldots, R_{\deg(f)-1}\}$ be the set of random variables corresponding to $\{r_0, r_1, \ldots, r_{\deg(f)-1}\}$. For a given $\mathbf{h}_l$ and $\mathcal{W}(n)$ there are two possibilities, either $\mathbf{h}_l \in \mathcal{W}^\perp(n)$ or $\mathbf{h}_l \notin \mathcal{W}^\perp(n)$, where $\mathcal{W}^\perp(n)$ is the dual code of $\mathcal{W}(n)$. When $\mathbf{h}_l \in \mathcal{W}^\perp(n)$, the corresponding R_l is always zero and when $\mathbf{h}_l \notin \mathcal{W}^\perp(n)$, from Lemma 1 of Appendix A, the corresponding R_l is equally likely to be zero or one. We now consider various situations for the set of random variables $\{R_0, R_1, \ldots, R_{\deg(f)-1}\}$.

(i) Each R_l for $l = 0, 1, \ldots, \deg(f) - 1$ is zero with probability one. This implies that in (52), the random variable corresponding to $r(X)$ is zero with probability one, which is the degenerate distribution (see Fig. 2 (a), (6)).

(ii) The set of random variables $\{R_0, R_1, \ldots, R_{\deg(f)-1}\}$ satisfy a linear relation given by

$$
a_0 R_0 + a_1 R_1 + \ldots + a_{\deg(f)-1} R_{\deg(f)-1} = 0, \tag{53}
$$

where each $a_l \in \mathbb{F}_2$, for $l = 0, 1, \ldots, \deg(f) - 1$. We consider the case when at least one of the R_l is equally likely to be one or zero, otherwise this case will get reduced to case (i). From (53), R_0 depends on $R_1, R_2, \ldots R_{\deg(f)-1}$. Thus the random vector $[R_0 \ R_1 \ \ldots \ R_{\deg(f)-1}]$ cannot take all possible $2^{\deg(f)}$ values. As each R_l is either zero with probability one or equally likely to be zero or one, in (52) $r(X)$ will follow the restricted uniform distribution (see Fig. 2 (c), (8)).

(iii) The set of random variables $\{R_0, R_1, \ldots, R_{\deg(f)-1}\}$ are independent. In this case, when each R_l for $l = 0, 1, \ldots, \deg(f) - 1$ is equally likely to be zero or one, the random variable corresponding to $r(X)$ will take all possible $2^{\deg(f)}$ values with equal probability, which is the uniform distribution (see Fig. 2 (b), (7)).

We now show that the set of random variables $\{R_0, R_1, \ldots, R_{\deg(f)-1}\}$ satisfies either of the above situations, which completes the proof. Let us first consider the case when $\mathcal{W}(n)$ is a nontrivial code. The case when $\mathcal{W}(n)$ is a trivial code will be considered later in this proof. For the two codes $\mathcal{W}^\perp(n)$ and $C(n, f^\perp)$ there are the following four possibilities.

1) $\mathcal{W}^\perp(n) = C(n, f^\perp)$
2) $C(n, f^\perp) \subset \mathcal{W}^\perp(n)$, where $\subset$ denotes strict subset
3) $\mathcal{W}^\perp(n) \subset C(n, f^\perp)$
4) $\mathcal{W}^\perp(n) \nsubseteq C(n, f^\perp)$ and $C(n, f^\perp) \nsubseteq \mathcal{W}^\perp(n)$

In cases 1) and 2), we have $C(n, f^\perp) \subseteq \mathcal{W}^\perp(n)$. From Property 2, every $\mathbf{h}_l \in C(n, f^\perp)$ and hence we have $\mathbf{h}_l \in \mathcal{W}^\perp(n)$, for $l = 0, 1, \ldots, \deg(f) - 1$ (see (52)). When $\mathbf{h}_l \in \mathcal{W}^\perp(n)$, the corresponding R_l is always zero which is the case (i).

In cases 2) and 4), there exists a vector $\mathbf{h} \in C(n, f^\perp) \cap \mathcal{W}^\perp(n)$, where $\cap$ denotes the intersection. When $C(n, f^\perp) \cap \mathcal{W}^\perp(n) = \mathbf{0}_n$ we have $\mathbf{h} = \mathbf{0}_n$, otherwise there exists a vector $\mathbf{h} \neq \mathbf{0}_n$ that belongs to the intersection space $C(n, f^\perp) \cap \mathcal{W}^\perp(n)$. Let us first consider the case when there exists a vector $\mathbf{h} \in C(n, f^\perp) \cap \mathcal{W}^\perp(n)$ such that $\mathbf{h} \neq \mathbf{0}_n$. From Property 2, the vector space spanned by $\{\mathbf{h}_0, \mathbf{h}_1, \ldots, \mathbf{h}_{\deg(f)-1}\}$ is equal to the code $C(n, f^\perp)$ and hence $\mathbf{h} \in C(n, f^\perp)$ can be written as

$$
\mathbf{h} = a_0 \mathbf{h}_0 + a_1 \mathbf{h}_1 + \ldots + a_{\deg(f)-1} \mathbf{h}_{\deg(f)-1}, \tag{54}
$$

where each $a_l \in \mathbb{F}_2$ for $l = 0, 1, \ldots, \deg(f) - 1$ such that for some i, $0 \leq i < \deg(f)$, $a_i \neq 0$. Since $\mathbf{h} \in \mathcal{W}^\perp(n)$, we have $\mathbf{w}\mathbf{h}^T = 0$ and from (54) we get

$$
\begin{aligned}
& \mathbf{w}\Big(a_0 \mathbf{h}_0 + a_1 \mathbf{h}_1 + \ldots + a_{\deg(f)-1} \mathbf{h}_{\deg(f)-1}\Big)^T = 0 \\
\implies & a_0 \mathbf{w}\mathbf{h}_0^T + a_1 \mathbf{w}\mathbf{h}_1^T + \ldots + a_{\deg(f)-1} \mathbf{w}\mathbf{h}_{\deg(f)-1}^T = 0 \\
\implies & a_0 r_0 + a_1 r_1 + \ldots + a_{\deg(f)-1} r_{\deg(f)-1} = 0,
\end{aligned} \tag{55}
$$

where the last equality is obtained from (52). Observe that this corresponds to the case (ii) when the set of random variables $\{R_0, R_1, \ldots, R_{\deg(f)-1}\}$ satisfy a linear relation.

We next consider the case when only the all-zero vector exists in the intersection of $C(n, f^\perp)$ and $\mathcal{W}^\perp(n)$, i.e., in (54), $\mathbf{h} = \mathbf{0}_n$. From (54) and (55) this implies that, the set of random variables $\{R_0, R_1, \ldots, R_{\deg(f)-1}\}$ do not satisfy any linear relation. Thus the set of random variables $\{R_0, R_1, \ldots, R_{\deg(f)-1}\}$ are independent. We now prove by contradiction that each R_l is equally likely to be zero or one. Suppose for some i, $0 \leq i < \deg(f)$, R_i is always zero, which implies that $\mathbf{h}_i \in \mathcal{W}^\perp(n)$. Since $\mathbf{h}_i \in C(n, f^\perp)$ we have $\mathbf{h}_i \in \mathcal{W}^\perp(n) \cap C(n, f^\perp)$ such that $\mathbf{h}_i \neq \mathbf{0}_n$, which is a contradiction. Note that this situation corresponds to case (iii) and the proof is complete.

We now consider the case when $\mathcal{W}(n)$ is a trivial code. When $\mathcal{W}(n)$ contains only the all-zero codeword, $r(X)$ will be zero with probability one and follows the degenerate distribution. When $\mathcal{W}(n) = \mathbb{F}_2^n$, since $\mathbf{w}(X)$ takes any value in $\mathcal{W}(n)$ with the uniform distribution, the random variable corresponding to $r(X) = \mathbf{w}(X) \bmod f(X)$ follows the uniform distribution and the proof is complete. ∎

We now use this lemma to prove Proposition 1.

Proof of Proposition 1:
Recall that the subspace $\mathcal{W}(n)$ is obtained by considering the initial n bits of codewords of $C(n_0, g_0)$ and $\mathcal{W}'(n)$ is defined in (12). In [17], Yardi et al. proved that there exists a codeword $\mathbf{w}_1(X) \in \mathcal{W}(n)$ and a codeword $\mathbf{w}_1'(X) \in \mathcal{W}'(n)$ such that $\mathbf{w}_1(X), \mathbf{w}_1'(X) \notin C(n, f)$, where $C(n, f)$ is the cyclic code generated by $f(X)$ (see Appendix B, Proposition 1 of [17]). For $\mathbf{w}_1(X)$ and $\mathbf{w}_1'(X)$, the corresponding syndromes $r(X) = \mathbf{w}_1(X) \bmod f(X)$ and $r'(X) = \mathbf{w}_1'(X) \bmod f(X)$ will be

nonzero polynomials. Since the all-zero vector is always a codeword in any linear block code, $r(X)$ and $r'(X)$ will be the zero polynomial for the all-zero codeword. Since $r(X)$ and $r'(X)$ can take at least two values in $\mathcal{P}_{\deg(f)}$ with a non-zero probability, $r(X)$ and $r'(X)$ cannot follow the degenerate distribution (see (6)). From Lemma 4, the distribution of the random variables corresponding to $r(X)$ and $r'(X)$ will either be uniform or restricted uniform and the proof is complete. ■

APPENDIX C: PROOF OF THEOREM 1

We first prove the necessary condition of the theorem that, if $r(X) = \mathbf{w}(X) \bmod f(X)$ follows the restricted uniform distribution, $g_0^\perp(X)$ has a factor of order strictly less than n_0. Let $\mathcal{W}(n)$ be the vector space obtained by puncturing the last $n_0 - n$ bits of codewords in code $C(n_0, g_0)$ such that $\mathbf{w}(X) \in \mathcal{W}(n)$. Since $C(n_0, g_0)$ is a cyclic code, the initial k_0 bits can be considered an information set and the assumption $k_0 < n < n_0$ implies that the dimension of $\mathcal{W}(n)$ is k_0.

Suppose $r(X) = \mathbf{w}(X) \bmod f(X)$ is given by,

$$r(X) = r_0 + r_1 X + \ldots + r_{\deg(f)-1} X^{\deg(f)-1}$$
$$= \mathbf{w}\mathbf{h}_0^T + \mathbf{w}\mathbf{h}_1^T X + \ldots + \mathbf{w}\mathbf{h}_{\deg(f)-1}^T X^{\deg(f)-1}, \quad (56)$$

where the last equality is obtained from Property 1 of Appendix B such that each $\mathbf{h}_l \in \mathbb{F}_2^n$ for $l = 0, 1, \ldots, \deg(f) - 1$. From Property 2 of Appendix B, each $\mathbf{h}_l \in C(n, f^\perp)$ where $C(n, f^\perp)$ is the dual code of the cyclic code $C(n, f)$ and the vector space spanned by $\{\mathbf{h}_0, \mathbf{h}_1, \ldots, \mathbf{h}_{\deg(f)-1}\}$ is equal to $C(n, f^\perp)$.

As explained in the proof of Lemma 4, the random variable corresponding to $r(X)$ follows the restricted uniform distribution if and only if there exists a non-zero vector $\mathbf{h} \in \mathbb{F}_2^n$ that lies in the intersection space of the codes $C(n, f^\perp)$ and $\mathcal{W}^\perp(n)$ (see Appendix B). Since $\mathbf{h} \in C(n, f^\perp)$, for some $\mathbf{u}_1(X) \in \mathcal{P}_{n-\deg(f^\perp)}$ we have

$$\mathbf{h}(X) = \mathbf{u}_1(X) f^\perp(X). \quad (57)$$

We now prove that the vector $\mathbf{h}' := [\mathbf{h} \; \mathbf{0}_{n_0-n}]$ lies in the code $C(n_0, g_0^\perp)$. For any $\mathbf{v} \in C(n_0, g_0)$, the inner product of $\mathbf{v}$ and $\mathbf{h}'$ is given by

$$\mathbf{v}(\mathbf{h}')^T = \Big[\mathbf{v}(0:n-1) \;\; \mathbf{v}(n:n_0-1)\Big] \Big[\mathbf{h} \;\; \mathbf{0}_{n_0-n}\Big]^T$$
$$= \mathbf{v}(0:n-1)\mathbf{h}^T \quad (58)$$
$$= \mathbf{w}\mathbf{h}^T$$
$$= 0,$$

where the last equality is obtained since $\mathbf{h} \in \mathcal{W}^\perp(n)$. Since $\mathbf{h}' \in C(n_0, g_0^\perp)$, for some $\mathbf{u}_2(X) \in \mathcal{P}_{n_0-k_0}$ we have

$$\mathbf{h}'(X) = \mathbf{h}(X) = \mathbf{u}_2(X) g_0^\perp(X). \quad (59)$$

Equating (57) and (59) we get

$$\mathbf{h}(X) = \mathbf{u}_1(X) f^\perp(X) = \mathbf{u}_2(X) g_0^\perp(X). \quad (60)$$

From the assumption of the theorem, $\deg(f) \leq k_0$ which implies that $n - \deg(f) \geq n - k_0$. Since $\deg(f^\perp) = n - \deg(f)$, we have $\deg(f^\perp) \geq n - k_0$. Since $\deg(g_0^\perp) = k_0$ and $\deg(h) \leq n-1$, from (60) we get $\deg(u_2) \leq n-k_0-1$. Thus

in (60) we have $\deg(u_2) \leq n-k_0-1$ and $\deg(f^\perp) \geq n-k_0$. This implies that there exists a factor $f_1(X)$ of $f^\perp(X)$ such that $f_1(X)$ is a factor of $g_0^\perp(X)$. Since $f_1(X)$ is a factor of $X^n + 1$ and $n < n_0$, this implies that $g_0^\perp(X)$ has a factor of order strictly less than n_0 and the proof of the necessary condition is complete.

We will now prove the converse. Suppose $g_0^\perp(X)$ has a factor $m^\perp(X)$ of order n' such that $1 \leq n' < n$. For a non-degenerate code $C(n_0, g_0)$, the order of $g_0^\perp(X)$ is equal to n_0 [22, Sec. 8.3] and hence $m^\perp(X) \neq g_0^\perp(X)$. From Lemma 3 of Appendix A, this implies that there exists a codeword of a degenerate pattern in $C(n_0, g_0)$, i.e., there exists $\mathbf{v} \in C(n_0, g_0)$ given by,

$$\mathbf{v} = \Big[\underbrace{\mathbf{w}' \;\; \mathbf{w}' \;\; \cdots \;\; \mathbf{w}'}_{l \text{ times}}\Big], \quad (61)$$

where $l > 1$, $\mathbf{w}'$ is vector of length n' such that $\mathbf{w}'$ is not a vector of a degenerate pattern (see Definition 6). Note that $m^\perp(X)$ is the minimal polynomial polynomial of the linear recurring sequence given by $[\mathbf{w}' \; \mathbf{w}' \; \cdots]$ [22, Sec. 8.3]. It is given that, $m(X)$ is the minimal generating polynomial of this sequence. Thus each $\mathbf{w}'(X)$ is a multiple of $m(X)$ (see Definition 7). Suppose $\mathbf{w}'(X) = \mathbf{u}'(X) m(X)$, for some $\mathbf{u}'(X) \in \mathcal{P}_{\deg(m^\perp)}$, since $\deg(m^\perp) = n' - \deg(m)$. Substituting this in (61) we get,

$$\mathbf{v}(X) = \mathbf{w}'(X) + X^{n'}\mathbf{w}'(X) + \ldots + X^{(l-1)n'}\mathbf{w}'(X) \quad (62)$$
$$= \mathbf{u}'(X)m(X) + X^{n'}\mathbf{u}'(X)m(X) + \ldots +$$
$$X^{(l-1)n'}\mathbf{u}'(X)m(X) \quad (63)$$
$$= \mathbf{u}'(X)m(X)\Big(1 + X^{n'} + \ldots + X^{(l-1)n'}\Big) \quad (64)$$

Let $C^\perp(n_0, m^\perp)$ be the dual code of $C(n_0, m^\perp)$, where $C(n_0, m^\perp)$ is the cyclic code of length n_0 generated by $m^\perp(X)$. Note that $\mathbf{v}(X) \in C^\perp(n_0, m^\perp)$ and from (64), the set of codewords in $C^\perp(n_0, m^\perp)$ are obtained considering all possible $2^{\deg(m^\perp)}$ values of $\mathbf{u}'(X) \in \mathcal{P}_{\deg(m^\perp)}$. Since $m^\perp(X)$ is a factor of $g_0^\perp(X)$ we have $C(n_0, g_0^\perp) \subset C(n_0, m^\perp)$ and this implies that $C^\perp(n_0, m^\perp) \subset C(n_0, g_0)$. Thus the codewords in $C(n_0, g_0)$ that are multiples of $m(X)$ are exactly the $2^{\deg(m^\perp)}$ codewords in $C^\perp(n_0, m^\perp)$.

From the assumptions of the converse, we have $n = bn'$ for some $b \geq 1$. Thus the vector $\mathbf{w}$ formed by the initial n bits of $\mathbf{v}$ in (61) is given by,

$$\mathbf{w} = \Big[\underbrace{\mathbf{w}' \;\; \mathbf{w}' \;\; \cdots \;\; \mathbf{w}'}_{b \text{ times}}\Big]. \quad (65)$$

Substituting $\mathbf{w}'(X) = \mathbf{u}'(X)m(X)$ we get,

$$\mathbf{w}(X) = \mathbf{w}'(X) + X^{n'}\mathbf{w}'(X) + \ldots + X^{(b-1)n'}\mathbf{w}'(X)$$
$$= \mathbf{u}'(X)m(X)\Big(1 + X^{n'} + \ldots + X^{(b-1)n'}\Big). \quad (66)$$

As explained in the first paragraph of the proof, the dimension of $\mathcal{W}(n)$ is k_0 and hence corresponding to every $\mathbf{v} \in C(n_0, g_0)$ there is a unique $\mathbf{w} \in \mathcal{W}(n)$. From (64) and (66), this implies that the number of $\mathbf{w}(X) \in \mathcal{W}(n)$ that are multiples of $m(X)$ are equal to $2^{\deg(m^\perp)}$. From (66), any $\mathbf{w}(X) \in \mathcal{W}(n)$ that is a multiple of $m(X)$ is also a

multiple of $(1 + X^{n'} + \ldots + X^{(b-1)n'})$. For a factor $f(X)$ of $m(X)(1 + X^{n'} + X^{2n'} + \ldots + X^{(b-1)n'})$, the probability that $r(X) = \mathbf{w}(X) \bmod f(X)$ is the all-zero polynomial is given by,

$$
\begin{aligned}
\mathbb{P}\big[r(X) = 0\big] &= \mathbb{P}\big[\mathbf{w}(X) \bmod f(X) = 0\big] \\
&= \frac{\text{Number of } \mathbf{w}(X) \in \mathcal{W}(n) \text{ that are multiples of } f(X)}{\text{Total number of } \mathbf{w}(X) \in \mathcal{W}(n)} \\
&= \frac{2^{\deg(m^\perp)}}{2^{k_0}} \\
&\overset{(a)}{>} \frac{2^{k_0 - \deg(f)}}{2^{k_0}} \\
&= \frac{1}{2^{\deg(f)}}
\end{aligned}
\tag{67}
$$

where the inequality in (a) is obtained since $\deg(m^\perp) > k_0 - \deg(f)$.

From Proposition 1, the random variable corresponding to $r(X)$ can either follow the uniform distribution or the restricted uniform distribution. For the uniform distribution, the probability of zero syndrome is equal to $1/2^{\deg(f)}$. From (67), the probability of zero syndrome is strictly more than $1/2^{\deg(f)}$ and hence $r(X)$ should follow the restricted uniform distribution. This completes the proof of the converse. ∎

APPENDIX D: PROOF OF THEOREM 2

Recall that the subspace $\mathcal{W}(n)$ is obtained by considering the initial n bits of codewords of $C(n_0, g_0)$. Since $n < n_0$ from (12) we have $\mathcal{W}'(n) = C_1(d_1) + C_2(d_2)$, where $C_1(d_1)$ and $C_2(d_2)$ are the linear block codes obtained by considering the set of suffixes and prefixes of lengths d_1 and d_2 of codewords in $C(n_0, g_0)$ respectively. Note that due to the cyclic nature, the subspaces spanned by the set of prefixes of length d_1 and the set of suffixes of length d_1 are identical. This implies that the code $\mathcal{W}'(n)$ consists of all possible prefixes of length d_1 concatenated with all possible suffixes of length d_2 and hence,

$$
\mathcal{W}(n) \subseteq \mathcal{W}'(n).
\tag{68}
$$

From (68) we have,

$$
\mathcal{W}'^\perp(n) \subseteq \mathcal{W}^\perp(n),
\tag{69}
$$

where $\mathcal{W}'^\perp(n)$ and $\mathcal{W}^\perp(n)$ are the dual codes of $\mathcal{W}'(n)$ and $\mathcal{W}(n)$ respectively.

In order to prove that $r'(X) = \mathbf{w}'(X) \bmod f(X)$ for $\mathbf{w}'(X) \in \mathcal{W}'(n)$ follows the uniform distribution using the arguments similar to the proof of Proposition 1, we need to prove that $\mathcal{W}'^\perp(n) \cap C(n, f^\perp) = \mathbf{0}_n$. From the assumptions of the theorem, $r(X) = \mathbf{w}(X) \bmod f(X)$ for $\mathbf{w}(X) \in \mathcal{W}(n)$ follows the uniform distribution. Using the arguments similar to the proof of Proposition 1, this is possible when $\mathcal{W}^\perp(n) \cap C(n, f^\perp) = \mathbf{0}_n$. From (69), this implies that $\mathcal{W}'^\perp(n) \cap C(n, f^\perp) = \mathbf{0}_n$ and the proof is complete. ∎

APPENDIX E: PROOF OF THEOREM 3

Let us first consider that case when $q = 0$, i.e., $r'(X)$ is given by,

$$
r'(X) = t_1(X) + t_2(X).
\tag{70}
$$

We now consider the situation when both $t_1(X)$ and $t_2(X)$ follow the uniform distribution. The probability that $r'(X)$ is a zero polynomial is given by,

$$
\begin{aligned}
\mathbb{P}\big[r'(X) = 0\big] &\overset{(a)}{=} \mathbb{P}\big[t_1(X) + t_2(X) = 0\big] \\
&= \mathbb{P}\big[t_1(X) = t_2(X)\big] \\
&= \sum_{a(X) \in \mathcal{P}_{\deg(f)}} \mathbb{P}\Big[t_1(X) = t_2(X) = a(X)\Big] \\
&\overset{(b)}{=} \sum_{a(X) \in \mathcal{P}_{\deg(f)}} \mathbb{P}\Big[t_1(X) = a(X)\Big]\mathbb{P}\Big[t_2(X) = a(X)\Big] \\
&\overset{(c)}{=} \sum_{a(X) \in \mathcal{P}_{\deg(f)}} \frac{1}{2^{\deg(f)}} \frac{1}{2^{\deg(f)}} = \frac{1}{2^{\deg(f)}}.
\end{aligned}
\tag{71}
$$

The equality in (a) is obtained from (70) and (b), (c) follow since the random variables corresponding to $t_1(X)$ and $t_2(X)$ are i.i.d. according to the uniform distribution. From Proposition 1, the random variable corresponding to $r'(X)$ can either follow the uniform distribution or the restricted uniform distribution. From (71), the random variable corresponding to $r'(X)$ follows the uniform distribution.

We next consider the case when either $t_1(X)$ or $t_2(X)$ follow the restricted uniform distribution. Without loss of generality let us consider the case when $t_1(X)$ follows the restricted uniform distribution. From the definition of the restricted uniform distribution we get, $\mathbb{P}[t_1(X) = a(X)] > 1/2^{\deg(f)}$ and in (71) we have

$$
\mathbb{P}\Big[r'(X) = 0\Big] > \frac{1}{2^{\deg(f)}}.
\tag{72}
$$

As explained earlier, the random variable corresponding to $r'(X)$ can either follow the uniform distribution or the restricted uniform distribution. For the uniform distribution, the probability of zero syndrome should be equal to $1/2^{\deg(f)}$. From (72), the probability of zero syndrome is more than $1/2^{\deg(f)}$ and hence $r'(X)$ follows the restricted uniform distribution. This completes the proof for the case when $q = 0$.

The case when $q > 0$ can be proved using similar arguments and hence we will not discuss it in detail. □

APPENDIX F: PROOF OF THEOREM 5

Since the proof is the same for any jth received polynomial $\mathbf{y}_j(X)$, for simplicity of notation we will ignore the suffix j from $\mathbf{y}_j(X)$ in this proof. Using this, the received polynomial $\mathbf{y}(X)$ is given by,

$$
\mathbf{y}(X) = \mathbf{w}(X) + \mathbf{e}(X),
\tag{73}
$$

where $\mathbf{w}(X)$ is the error-free polynomial and $\mathbf{e}(X)$ is the polynomial corresponding to the error introduced by BSC(p).

The probability of observing the all-zero syndrome is given by,

$$\mathbb{P}\Big[r(X) = 0\Big] = \mathbb{P}\Big[\mathbf{y}(X) \bmod f(X) = 0\Big]$$
$$= \mathbb{P}\Big[\mathbf{y}(X) \in C(n, f)\Big] \quad (74)$$

where the last equality is obtained since the cyclic code $C(n, f)$ consists of possible multiples of $f(X)$. For a given $\mathbf{w}(X)$ there are two possibilities, either $\mathbf{w}(X) \in C(n, f)$ or $\mathbf{w}(X) \notin C(n, f)$. Suppose,

$$\begin{aligned} Q &: \text{Event when } \mathbf{w}(X) \in C(n, f), \\ Q^c &: \text{Event when } \mathbf{w}(X) \notin C(n, f). \end{aligned} \quad (75)$$

Using total probability law in (74) we get,

$$\mathbb{P}\Big[r(X) = 0\Big] = \mathbb{P}\Big[\mathbf{y}(X) \in C(n, f)\Big|Q\Big]\mathbb{P}[Q] + $$
$$\mathbb{P}\Big[\mathbf{y}(X) \in C(n, f)\Big|Q^c\Big]\mathbb{P}[Q^c]. \quad (76)$$

From (73) and (75), when the event Q is true, we have $\mathbf{y}(X) \in C(n, f)$ if $\mathbf{e}(X) \in C(n, f)$. Similarly, when the event Q^c is true, we have $\mathbf{y}(X) \in C(n, f)$ if $\mathbf{e}(X)$ belongs to some proper coset $\mathcal{G}(n, f)$ of code $C(n, f)$. Using this in (76) we have,

$$\mathbb{P}\Big[r(X) = 0\Big] = \mathbb{P}\Big[\mathbf{e}(X) \in C(n, f)\Big]\mathbb{P}[Q] + $$
$$\mathbb{P}\Big[\mathbf{e}(X) \in \mathcal{G}(n, f))\Big]\mathbb{P}[Q^c] \quad (77)$$

From Sullivan's subgroup-coset inequality theorem [29], for any proper coset $\mathcal{G}(n, f)$ of $C(n, f)$ we have,

$$\frac{\mathbb{P}[\mathbf{e}(X) \in C(n, f)]}{\mathbb{P}[\mathbf{e}(X) \in \mathcal{G}(n, f)]} \geq \frac{1 - (1 - 2p)^{n - \deg(f) + 1}}{1 + (1 - 2p)^{n - \deg(f) + 1}} = \lambda. \quad (78)$$

We next find the probability of the event $\mathbf{e}(X) \in C(n, f)$ as follows.

$$\mathbb{P}[\mathbf{e}(X) \in C(n, f)] = \sum_{\mathbf{v}(X) \in C(n,f)} \mathbb{P}[\mathbf{e}(X) = \mathbf{v}(X)]$$
$$= \sum_{i=0}^{n} A_i p^i (1 - p)^{n-i} \quad (79)$$
$$= P(C(n, f)),$$

where $\{A_0, A_1, \cdots, A_n\}$ is the weight distribution of $C(n, f)$ and last equality is obtained from (20).

Substituting (78) in (77) we have,

$$\mathbb{P}\Big[r(X) = 0\Big] \leq \mathbb{P}\Big[\mathbf{e}(X) \in C(n, f)\Big]\mathbb{P}[Q] + $$
$$\lambda\mathbb{P}\Big[\mathbf{e}(X) \in C(n, f))\Big]\mathbb{P}[Q^c] \quad (80)$$
$$\overset{(b)}{=} \mathcal{P}(C(n, f))\mathbb{P}[Q] + \lambda\mathcal{P}(C(n, f))\big(1 - \mathbb{P}[Q]\big) \quad (81)$$
$$= \mathcal{P}(C(n, f))\Big[\mathbb{P}[Q] + \lambda\big(1 - \mathbb{P}[Q]\big)\Big] \quad (82)$$
$$= \mathcal{P}(C(n, f))\Big[\mathbb{P}[Q](1 - \lambda) + \lambda\Big] \quad (83)$$

The equality in (b) is obtained from (79) and since $\mathbb{P}[Q^c] = 1 - \mathbb{P}[Q]$ (see (75)).

From the assumption of the theorem, either $n \neq ln_0$ or assumed synchronization $s \neq s_0$ or $f(X)$ is not a factor of $g_0(X)$. When either $n \neq ln_0$ or $s \neq s_0$ or $f(X)$ is not a

factor of $g_0(X)$, from Proposition 1 and Section III-C, the distribution of $\mathbf{w}(X) \bmod f(X)$ is either uniform or restricted uniform. From the definition of the uniform and the restricted uniform distributions, $\mathbb{P}[\mathbf{w}(X) \bmod f(X) = 0]$ is less than or equal to $1/2$, i.e.,

$$\mathbb{P}[\mathbf{w}(X) \in C(n, f)] = \mathbb{P}[Q] \leq \frac{1}{2}. \quad (84)$$

Substituting (84) in (83) we get,

$$\mathbb{P}\Big[r(X) = 0\Big] \leq \mathcal{P}(C(n, f)) \left(\frac{1}{2}(1 - \lambda) + \lambda\right) \quad (85)$$
$$= \mathcal{P}(C(n, f)) \left(\frac{\lambda + 1}{2}\right) \quad (86)$$

and the proof is complete. ∎

APPENDIX G: PROOF OF THEOREM 6

Since the proof is the same for any jth received vector $\mathbf{y}_j$, we will ignore the suffix j from $\mathbf{y}_j$ for the sake of simplicity. Using this notation, an n-bit received vector is given by,

$$\mathbf{y} = \mathbf{w} + \mathbf{e}, \quad (87)$$

where $\mathbf{w}$ is an error-free vector and $\mathbf{e}$ is an error vector introduced by BSC(p). For a factor $f(X)$ of $X^n + 1$, suppose a parity check matrix H of $C(n, f)$ is given by

$$H = \begin{bmatrix} f_0^\perp & f_1^\perp & \cdot & \cdot & f_{\deg(f^\perp)}^\perp & 0 & \cdot & 0 \\ 0 & f_0^\perp & \cdot & \cdot & & f_{\deg(f^\perp)}^\perp & \cdot & 0 \\ \vdots & & \ddots & & & & & \vdots \\ 0 & \cdot & 0 & f_0^\perp & & \cdot & \cdot & f_{\deg(f^\perp)}^\perp \end{bmatrix}$$
$$= \begin{bmatrix} \mathbf{h}_0 \\ \mathbf{h}_1 \\ \vdots \\ \mathbf{h}_{\deg(f)-1} \end{bmatrix}, \quad (88)$$

where the polynomial corresponding to the first row of H is the generator polynomial $f^\perp(X)$ of the dual code of $C(n, f)$, and $\mathbf{h}_0, \mathbf{h}_1, \ldots, \mathbf{h}_{\deg(f)-1}$ are the rows of H. Suppose $\mathbf{w}H^T$ is given by,

$$\mathbf{w}H^T = \mathbf{t} = \begin{bmatrix} \mathbf{w}\mathbf{h}_1^T & \mathbf{w}\mathbf{h}_2^T & \ldots & \mathbf{w}\mathbf{h}_{\deg(f)-1}^T \end{bmatrix} \quad (89)$$
$$= \begin{bmatrix} t_0 & t_1 & \ldots & t_{\deg(f)-1} \end{bmatrix} \quad (90)$$

where $t_l = \mathbf{w}\mathbf{h}_l^T$, for $l = 0, 1, \ldots, \deg(f) - 1$. As shown in Fig. 3, an n-bit noise-free vector $\mathbf{w}$ is either of the following two types.

(i) $\mathbf{w}$ is formed by the consecutive n bits of a codeword in the true code $C(n_0, g_0)$, i.e., $\mathbf{w} \in \mathcal{W}(n)$, where $\mathcal{W}(n)$ is defined in the first paragraph of Section III.

(ii) $\mathbf{w}$ is a concatenation of the suffix of a codeword of length d_1, a sequence of q codewords, and the prefix of a codeword of length d_2, where $0 \leq d_1, d_2 < n_0$, $q \geq 1$ such that $n = d_1 + qn_0 + d_2$, i.e., $\mathbf{w} \in \mathcal{W}'(n)$, where $\mathcal{W}'(n)$ is defined in (12).

We now consider the cases when $\mathbf{w} \in \mathcal{W}(n)$ and $\mathbf{w} \in \mathcal{W}'(n)$ separately and prove that t_l in (90) is equally likely to be zero or one for $l = 0, 1, \ldots, \deg(f) - 1$.

(i) Case when $\mathbf{w} \in \mathcal{W}(n)$

From the assumptions of the theorem we have $n < n_0$. For a given $\mathbf{h}_l$ we have either $\mathbf{h}_l \in \mathcal{W}^\perp(n)$ or $\mathbf{h}_l \notin \mathcal{W}^\perp(n)$, where $\mathcal{W}^\perp(n)$ is the dual code of $\mathcal{W}(n)$. We now prove by contradiction that each $\mathbf{h}_l \notin \mathcal{W}^\perp(n)$, for $l = 0, 1, \ldots, \deg(f) - 1$. Suppose $\mathbf{h}_l \in \mathcal{W}^\perp(n)$ for some l, $0 \leq l < \deg(f)$. Using the similar steps as in (58) we can prove that, $[\mathbf{h}_l \ \mathbf{0}_{n_0-n}] \in C(n_0, g_0^\perp)$, where $C(n_0, g_0^\perp)$ is the dual code of $C(n_0, g_0)$. From (88), the polynomial corresponding to $\mathbf{h}_l$ can be written as $\mathbf{h}_l(X) = X^l f^\perp(X)$ and $[\mathbf{h}_l \ \mathbf{0}_{n_0-n}] \in C(n_0, g_0^\perp)$ implies that,

$$\mathbf{h}_l(X) = X^l f^\perp(X) = \mathbf{u}(X) g_0^\perp(X), \quad (91)$$

where $\mathbf{u}(X) \in \mathcal{P}_{n_0-\deg(g_0^\perp)}$. For a nontrivial cyclic code, $g_0^\perp(X)$ does not divide X^l for any integer l [1], and hence (91) implies that $g_0^\perp(X)$ should divide $f^\perp(X)$. Since $f^\perp(X)$ divides $X^n + 1$, $g_0^\perp(X)$ also divides $X^n + 1$. Since $n < n_0$, $C(n_0, g_0)$ will be a degenerate code [22, Sec. 8.3], which is a contradiction according to the assumptions of the theorem. This proves that $\mathbf{h}_l \notin \mathcal{W}^\perp(n)$ for $l = 0, 1, \ldots, \deg(f) - 1$. From Lemma 1, $\mathbf{h}_l \notin \mathcal{W}^\perp(n)$ implies that t_l is equally likely to be zero or one.

(ii) When $\mathbf{w} \in \mathcal{W}'(n)$

Since $n < n_0$, an n-bit vector $\mathbf{w}$ is given by,

$$\mathbf{w} = \begin{bmatrix} \mathbf{v}_1(n_0 - d_1 : n_0 - 1) & \mathbf{v}_2(0 : d_2 - 1) \end{bmatrix}, \quad (92)$$

where $\mathbf{v}_1, \mathbf{v}_2 \in C(n_0, g_0)$. For a given $\mathbf{h}_l$ the inner product $\mathbf{w}\mathbf{h}_l^T$ is given by,

$$\mathbf{w}\mathbf{h}_l^T = \mathbf{w}(0 : d_1 - 1)\mathbf{h}_l(0 : d_1 - 1)^T + \\ \mathbf{w}(d_1 : n - 1)\mathbf{h}_l(d_1 : n - 1)^T. \quad (93)$$

Recall that in part (i) we proved that $\mathbf{h}_l \notin \mathcal{W}^\perp(n)$ for $l = 0, 1, \ldots, \deg(f) - 1$. From Lemmas 1 and 2 of Appendix A, $\mathbf{h}_l \notin \mathcal{W}^\perp(n)$ implies that either $\mathbf{w}(0 : d_1 - 1)\mathbf{h}_l(0 : d_1 - 1)^T$ or $\mathbf{w}(d_1 : n - 1)\mathbf{h}_l(d_1 : n - 1)^T$ is equally likely to be zero or one. This implies that in (93), $t_l = \mathbf{w}\mathbf{h}_l^T$ is equally likely to be zero or one.

We now have that each bit in $\mathbf{w}\mathbf{h}_l$ is equally likely to be zero or one, for $l = 0, 1, \ldots, \deg(f) - 1$. Let us consider the noise-affected version of $\mathbf{y}$ of $\mathbf{w}$ (see (87)). Suppose the inner product $\mathbf{y}\mathbf{h}_l$ is given by,

$$\mathbf{y}H^T = \mathbf{r} = \begin{bmatrix} r_0 & r_1 & \ldots & r_{\deg(f)-1} \end{bmatrix} \quad (94)$$

where each r_l is given by,

$$r_l = \mathbf{y}\mathbf{h}_l^T = \begin{bmatrix} \mathbf{w} + \mathbf{e} \end{bmatrix} \mathbf{h}_l^T \quad (95)$$
$$= \mathbf{w}\mathbf{h}_l^T + \mathbf{e}\mathbf{h}_l^T \quad (96)$$

Since $\mathbf{w}\mathbf{h}_l^T$ is equally likely to be zero or one, in (96) r_l is equally likely to be zero or one. Using this we now prove that $P(\mathbf{y}_j, f_1) = P(\mathbf{y}_j, f_2)$, where $f_1(X)$ and $f_2(X)$ are any two factors of $X^n + 1$. For any factor $f(X)$ of $X^n + 1$, $P(\mathbf{y}, f)$ is given by,

$$\begin{aligned} P(\mathbf{y}, f) &= \frac{1}{\deg(f)} \sum_{l=0}^{\deg(f)-1} \mathbb{P}[r_l = 0] \\ &\overset{(a)}{=} \frac{1}{\deg(f)} \sum_{l=0}^{\deg(f)-1} \frac{1}{2} \\ &= \frac{1}{2} \frac{1}{\deg(f)} \sum_{l=0}^{\deg(f)-1} 1 = \frac{1}{2}, \end{aligned} \quad (97)$$

where the equality in (a) is obtained since each r_l is equally likely to be zero or one. It can be seen that the value of $P(\mathbf{y}, f)$ does not depend on the chosen $f(X)$. This implies that $P(\mathbf{y}, f_1) = P(\mathbf{y}, f_2)$ and the proof is complete. ∎

REFERENCES

[1] S. Lin and D. Costello, *Error Control Coding*, 2nd ed. Englewood Cliffs, New Jersey, USA: Prentice-Hall, 2004.

[2] B. Rice, "Determining the parameters of a rate 1/n convolutional encoder over GF(q)," in *Proceedings of 3rd International Conference on Finite Fields and Applications*, Glasgow, Scotland, July 1995.

[3] G. Planquette, "Identification de trains binaires codés," *Ph.D. Thesis, Universite de Rennes I, France*, 1996.

[4] E. Filiol, "Reconstruction of convolutional encoders over GF(q)," in *Crytography and Coding: Lecture Notes in Computer Science*, vol. 1335, Berlin, Heidelberg, 1997, pp. 101–109.

[5] A. Valembois, "Detection and recognition of a binary linear code," *Discrete Applied Mathematics*, vol. 111, pp. 199–218, July 2001.

[6] M. Marazin, R. Gautier, and G. Burel, "Blind recovery of k/n rate convolutional encoders in a noisy environment," *EURASIP Journal on Wireless Communications and Networking*, no. 1, pp. 1–9, 2011.

[7] J. Dingel and J. Hagenauer, "Parameter estimation of a convolutional encoder from noisy observations," in *Proceedings of IEEE International Symposium on Information Theory*, Nice, France, June 2007, pp. 1776–1780.

[8] J. Barbier, "Reconstruction of turbo-code encoders," in *Proceedings of SPIE*, vol. 5819, 2005, pp. 463–473.

[9] M. Côte and N. Sendrier, "Reconstruction of a turbo-code interleaver from noisy observation," in *Proceedings of IEEE International Symposium on Information Theory*, Austin, Texas, 2010, pp. 2003–2007.

[10] G. Sicot, S. Houcke, and J. Barbier, "Blind detection of interleaver parameters," *Signal Processing*, vol. 89, no. 4, pp. 450–462, April 2009.

[11] M. Cluzeau and M. Finiasz, "Recovering a code's length and synchronization from a noisy intercepted bitstream," in *Proceedings of IEEE International Symposium on Information Theory*, Seoul, Korea, July 2009, pp. 2737–2741.

[12] M. Cluzeau, "Block code reconstruction using iterative decoding techniques," in *Proceedings of IEEE International Symposium on Information Theory*, Seattle, USA, July 2006, pp. 2269–2273.

[13] R. Moosavi and E. Larsson, "Fast blind recognition of channel codes," *IEEE Transactions on Communications*, vol. 62, no. 5, pp. 1393–1405, 2014.

[14] H. Lee, C. Park, J. Lee, and Y. Song, "Reconstruction of BCH codes using probability compensation," in *Proceedings of IEEE APCC*, Jeju Island, Korea, October 2012, pp. 591–594.

[15] C. Chabot, "Reconnaissance de codes, structure des codes quasi-cycliques," *PhD thesis, University of Limoges*, 2009.

[16] A. Yardi, S. Vijayakumaran, and A. Kumar, "Blind reconstruction of binary cyclic codes," in *Proceedings of European Wireless*, Barcelona, Spain, May 2014, pp. 849–854.

[17] ——, "Blind reconstruction of binary cyclic codes from unsynchronized bitstream," *IEEE Transactions on Communications*, vol. 64, no. 7, pp. 2693–2706, 2016.

[18] J. Zhou, Z. Huang, S. Su, and Y. Shaowu, "Blind recognition of binary cyclic codes," *EURASIP Journal on Wireless Communications and Networking*, vol. 2013, no. 1, pp. 1–17, 2013.

[19] J. Zhou, Z. Huang, C. Liu, S. Su, and Y. Zhang, "Information-dispersion-entropy-based blind recognition of binary BCH codes in soft decision situations," *Entropy*, vol. 15, no. 5, pp. 1705–1725, 2013.

[20] H. V. Poor, *Introduction to Signal Detection and Estimation*, 2nd ed. New York, USA: Springer-Verlag, 1994.

[21] A. Yardi, A. Kumar, and S. Vijayakumaran, "Channel-code detection by a third-party receiver via the likelihood ratio test," in *Proceedings of IEEE International Symposium on Information Theory*, Honolulu, HI, USA, June 2014, pp. 1051–1055.

[22] F. MacWilliams and N. Sloane, *The Theory of Error Correcting Codes*. Amsterdam,Netherlands: North-Holland Publishing Company, 1977.

[23] R. Lidl and H. Niederreiter, *Introduction to Finite Fields and Their Applications*. Cambridge, United Kingdom: Cambridge University Press, 1986.

[24] G. Cancellieri, *Polynomial Theory of Error Correcting Codes*. Cham, Switzerland: Springer, 2015.

[25] W. Peterson and E. Weldon, *Error-Correcting Codes*, 2nd ed. Cambridge, Massachusetts, USA: MIT Press, 1996.

[26] W. Huffman and V. Pless, *Fundamentals of Error-Correcting Codes*. Cambridge, United Kingdom: Cambridge University Press, 2003.

[27] T. Cover and J. Thomas, *Elements of Information Theory*. New York, USA: Wiley, 1991.

[28] M. Artin, *Algebra*. New Jersey, USA: Prentice-Hall, 1991.

[29] D. Sullivan, "A fundamental inequality between the probabilities of binary subgroups and cosets," *IEEE Transactions on Information Theory*, vol. 13, no. 1, pp. 91–94, 1967.