Blind Reconstruction of Binary Cyclic Codes over Binary Erasure Channel

Arti D Yardi

Blind Reconstruction of Binary Cyclic Codes over Binary Erasure Channel

Arti D. Yardi
IRIT/INP-ENSEEIHT, University of Toulouse, France
Email: artidilip.yardi@enseeiht.fr

***Abstract*—Given a sequence of noise-affected codewords of an unknown channel code, the problem of *blind reconstruction of channel codes* consists of identifying this unknown channel code. This problem has many applications in military surveillance and cognitive radios. In this paper, we study this problem for the case when the noise is introduced by the binary erasure channel (BEC) and the unknown channel code is a binary cyclic code of known length. We provide an algorithm to find the generator polynomial of the unknown cyclic code. We also provide an analysis of our algorithm where we provide a lower bound on the probability of correctly identifying the factors of the generator polynomial.**

I. Introduction

In the problem of *blind reconstruction of channel codes*, the channel code that is used at the transmitter is not known at the receiver. Such situations may arise in military surveillance applications or in cognitive radios [1], where the channel code corresponding to the received data is not known. The receiver has access to the noise-affected sequence of this unknown channel code and the aim is to identify the channel code corresponding to it. This blind reconstruction problem is in general known to be NP-hard [2]. 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. This problem has been studied for several families of channel codes such as convolutional codes [3], [4], linear block codes [2], [5], LDPC codes [1], [6], and cyclic codes [7]–[11].

For the case of binary cyclic codes, this problem has been studied by Chabot [7], Lee et al. [8], Yardi et al. [9], [10], and Zhou et al. [11]. Chabot, Lee et al., and Yardi et al. consider the situation when the noise is introduced by the binary symmetric channel and the situation of additive white Gaussian noise is studied by Zhou et al. However, for several communication systems of practical interests, there are situations when the noise is modeled by BEC. In this paper, we study this blind reconstruction problem for BEC under the assumption that the unknown channel code is a binary cyclic code of known length.

For BEC, one can provide a simple blind reconstruction algorithm as follows. Consider the set of received vectors that have no erasures and take the greatest common divisor (gcd) of these received polynomials. For a cyclic code, since every codeword is a multiple of the generator polynomial $g(X)$ of the code [12], this gcd will be equal to $g(X)$ with high probability. However in the presence of erasures and with limited number of received polynomials, one might not be able to find the generator polynomial with this naive algorithm.

For BEC, while a variety of such blind reconstruction algorithms can be proposed, it may not be possible to analyze the performance of all such algorithms or analytically characterize metrics such as probability of correct identification of the unknown code. The main contributions of this paper are:

(1) We first provide a blind reconstruction algorithm to identify the factors of $g(X)$ (see Theorem 1).
(2) We provide a theoretical analysis of our algorithm, where we provide a lower bound on correctly identifying the factors of $g(X)$ (see Theorems 2 and 3).
(3) Finally, we provide simulation results for the proposed algorithm for a variety of cyclic codes and also compare the lower bound versus its true value.

Organization: In Section II, we describe the system model for the problem of blind reconstruction of cyclic codes over BEC. The proposed blind reconstruction method is given in Section III and a theoretical analysis of this method is discussed in Section IV. In Section V, we provide simulation results of the proposed algorithm for various cyclic codes and finally provide some future directions in Section VI.

Notation: The finite field with two elements 0 and 1 is denoted by $\mathbb{F}_2$ and $\mathbb{F}_2[X]$ denotes the polynomial ring with coefficients from $\mathbb{F}_2$. The set of natural numbers is denoted by $\mathcal{N}$. We use boldface letters to denote the vectors and lower case letters for the components of a vector. For example, vector $\mathbf{w} = \begin{bmatrix} w_0 & w_1 & \dots & w_{n-1} \end{bmatrix}$, where w_i for $i = 0, 1, \dots, n-1$ are the components of $\mathbf{w}$. The polynomial representation of vector $\mathbf{w}$, is given by $\mathbf{w}(X) = w_0 + w_1 X + \ldots + w_{n-1} X^{n-1}$. The sequence of M vectors $\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_M$ is denoted by $\mathbf{w}_1^M$. The set of polynomials in $\mathbb{F}_2[X]$ of degree strictly less than l is denoted by $\mathcal{P}_l$. The product of two polynomials $f(X)$ and $f'(X)$ is denoted by ff'. Random variables are denoted by capital letters and their realizations are indicated with lowercase letters. For example, the realization of a random variable X is denoted by x. The cyclic code of length n and generator polynomial $g(X)$ is denoted by $C(n, g)$.

II. System Model and notation

Suppose the binary cyclic code $C(n, g)$ of length n, dimension k, and generator polynomial $g(X)$ is used at the transmitter. We assume that the length n of the code is known at the receiver but $g(X)$ and k are not known. We assume that $C(n, g)$ corrects at least one error, which implies

that its minimum distance $d_{\min}(C(n,g)) \geq 3$ [12]. Let $\mathbf{v}_1^M = \begin{bmatrix}\mathbf{v}_1 & \mathbf{v}_2 & \dots & \mathbf{v}_M\end{bmatrix}$ be the sequence of transmitted codewords, where $M \in \mathbb{N}$. Each transmitted codeword is independent and identically distributed (i.i.d.) according to the uniform distribution over the set of codewords of $C(n,g)$. The transmitted codewords are affected by the noise introduced by a BEC of erasure probability p to get the received sequence $\mathbf{y}_1^M = \begin{bmatrix}\mathbf{y}_1 & \mathbf{y}_2 & \dots & \mathbf{y}_M\end{bmatrix}$. For BEC($p$), each bit in the received sequence is erased with probability p. Let e_j be the number of erasures in $\mathbf{y}_j$, for $j = 1, 2, \dots, M$. Since the transmitted codewords are i.i.d., the received vectors are also i.i.d. Hence, for the sake of simplicity we will drop parameters j from $\mathbf{v}_j$, $\mathbf{y}_j$, and e_j whenever the arguments are applicable for any jth received vector. Using this notation, $\mathbf{y}$ is an erased version of $\mathbf{v}$ with e the number of erasures. For the blind reconstruction problem of the paper, the aim of the receiver is to identify $g(X)$ of the code using the received sequence $\mathbf{y}_1^M$.

III. Proposed blind reconstruction method

In this section, we propose a simple blind reconstruction method to find the generator polynomial $g(X)$ of the code. The set of factors of X^n+1 is the candidate set of polynomials for the factors of $g(X)$ [12]. Suppose $g(X)$ is factorized as

$$g(X) = \prod_i [f_i(X)]^{m_i}, \tag{1}$$

where $f_i(X)$ are irreducible factors of $g(X)$ and m_i are their respective multiplicities. We identify $g(X)$ by identifying its irreducible factors and their multiplicities.

Recall that, $\mathbf{y}$ is an erased version of a transmitted codeword $\mathbf{v}$ with e number of erasures. For the cyclic codes, it is known that each bit is equally likely to be zero or one [12]. Hence we substitute 0 and 1 in all the erased entries in $\mathbf{y}$ and obtain a set of 2^e vectors denoted by $\mathbf{w}_1^{2^e} = \begin{bmatrix}\mathbf{w}_1 & \mathbf{w}_2 & \dots & \mathbf{w}_{2^e}\end{bmatrix}$. In order to identify whether a given candidate factor $f(X)$ of X^n+1 is a factor of $g(X)$ or not, we study the properties of the set $\mathbf{w}_1^{2^e}$ in the following theorem.

Theorem 1. *Suppose $\mathbf{y}$ is the erased version of a transmitted codeword of $C(n,g)$ with e number of erasures. Let $\mathbf{w}_1^{2^e} = \begin{bmatrix}\mathbf{w}_1 & \mathbf{w}_2 & \dots & \mathbf{w}_{2^e}\end{bmatrix}$ be the set of vectors obtained by substituting all possible 2^e values in the erased bits of $\mathbf{y}$. Let $\mathbf{Y}$ be the random vector corresponding to $\mathbf{y}$. Then for a factor $f(X)$ of X^n+1 we have the following two cases.*

(a) If $f(X)$ is a factor of $g(X)$, then for every realization $\mathbf{y}$ of $\mathbf{Y}$, there exists $\mathbf{w}_i$, such that $f(X)$ divides $\mathbf{w}_i(X)$ for some i, $1 \leq i \leq 2^e$.

(b) If $f(X)$ is not a factor of $g(X)$, then there exists a realization $\mathbf{y}$ of $\mathbf{Y}$ such that $f(X)$ does not divide $\mathbf{w}_i(X)$, for $i = 1, 2, \dots, 2^e$.

Proof: We first consider the case when $f(X)$ is a factor of $g(X)$. For any received vector $\mathbf{y}$, the correct substitution will always be a multiple of $g(X)$ and $f(X)$ will divide this correct substitution. This completes the proof of part (a).

Let now consider the case when $f(X)$ is not a factor of $g(X)$. In this case, we need to show the existence of a received vector $\mathbf{y}$ that satisfies the condition of part (b). Note that the random vector $\mathbf{Y}$ is obtained by choosing a codeword $\mathbf{v} \in C(n,g)$ with probability $1/2^k$ and then erasing the coefficients of $\mathbf{v}$ independently with probability p. Thus with probability $1/2^k$, codeword $\mathbf{g}$ corresponding to $g(X)$ will be transmitted and with probability $(1-p)^n$ none of the coefficient of $\mathbf{g}$ will get erased. The corresponding received vector $\mathbf{y}$ will be equal to $\mathbf{g}$. Since $f(X)$ is not a factor of $g(X)$, this $\mathbf{y}$ satisfies the conditions of part (b) and the proof is complete. ■

Theorem 1 can be used to distinguish between the factors and non-factors of $g(X)$. The basic idea of the algorithm consists of substituting all possible values in the erased bits of the received vectors $\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_M$ and find the factors of $g(X)$ using Theorem 1. When the number of received vectors M tend to infinity, Theorem 1 guarantees that the true generator polynomial $g(X)$ will be identified with probability one. However with limited amount of data, we observe that some non-factors of $g(X)$ may satisfy the condition (a) of Theorem 1 and may get wrongly decided as factors of $g(X)$. Consider the following example to explain this observation.

Example 1. *Suppose $C(n,g)$, with $n = 7$ and $g(X) = X^3 + X^2 + 1$ is used at the transmitter. Suppose the number of received vectors M is equal to one. Suppose $\mathbf{v} = [1\ 0\ 1\ 1\ 0\ 0\ 0]$ and $\mathbf{y} = [1\ 0\ ?\ ?\ 0\ 0\ 0]$, where ? denotes the erasure. The set of polynomials obtained by substituting $[0\ 0], [0\ 1], [1\ 0], [1\ 1]$ at the erased locations is $\{1, (X+1)(X^2+X+1), (X+1)^2, (X^3+X^2+1)\}$. The condition (a) of Theorem 1 is satisfied by the factors $X+1$ and X^3+X^2+1 of X^7+1.* □

In Example 1, since $g(X)$ is a factor of X^7+1, we can remove the non-factors of X^7+1 from the set of substituted polynomials. The corresponding set will be $\{1, (X+1), (X+1), (X^3+X^2+1)\}$. Observe that the cyclic code corresponding to all the wrong substitutions has a minimum distance strictly less than three. Since for the true code $d_{\min}(C(n,g)) \geq 3$, we can discard the substituted entry if the cyclic corresponding to it has the minimum distance strictly less than three. We use this idea to propose a blind reconstruction method in Algorithm 1.

IV. A theoretical performance of the proposed algorithm

In Algorithm 1, all factors of $g(X)$ are always correctly identified. In order to study the performance of this algorithm, the key step is to find the probability that a given non-factor is correctly decided as a non-factor of $g(X)$. Suppose $f(X)$ is a factor of X^n+1 but it is not a factor of $g(X)$. This $f(X)$ is correctly declared as a non-factor if there exists a received vector $\mathbf{y}$ that satisfies the condition (b) of Theorem 1 in step (ii) of Algorithm 1. To find the probability of receiving such a $\mathbf{y}$, one needs to condition over all possible 2^k codewords in $C(n,g)$ and all possible 2^n erasure patterns. Since the number of codewords and erasure patterns are exponential in number, finding this probability is in general computationally intractable. Hence in this section, we find an lower bound on receiving such a $\mathbf{y}$. Since all the received vectors are i.i.d., we find this lower bound when the number of received

Algorithm 1 Proposed blind reconstruction algorithm

(i) **Preprocessing step:** For the received vector $\mathbf{y}_j$ with e_j erasures, substitute 0 and 1 at each erased location to obtain the set of vectors $\mathbf{w}_{j,1}^{2^{e_j}}$, for $j = 1, 2, \ldots, M$. Perform the following operations on the set $\mathbf{w}_{j,1}^{2^{e_j}}$.
   (a) Factorize each $\mathbf{w}_{j,i}(X) = u_{j,i}(X)h_{j,i}(X)$ such that $u_{j,i}(X)$ is not a factor of $X^n + 1$ and $h_{j,i}(X)$ is a factor of $X^n + 1$.
   (b) Discard $\mathbf{w}_{j,i}(X)$ if the cyclic code generated by $h_{j,i}(X)$ has minimum distance strictly less than three. Let $\mathbf{w}_{j,1}(X), \mathbf{w}_{j,2}(X), \ldots, \mathbf{w}_{j,s_j}(X)$ be the set of non-discarded polynomials.

(ii) **Decision step:** Declare $f(X)$ as a non-factor of $g(X)$ if there exists $\mathbf{y}_j$ for some $1 \leq j \leq M$ such that $f(X)$ does not divide $\mathbf{w}_{j,l}(X)$ for $l = 1, 2, \ldots, s_j$, otherwise $f(X)$ is declared as a factor of $g(X)$ (see Theorem 1).

(iii) **Find multiplicity:** If $f(X)$ is declared as a factor of $g(X)$ in step (ii), then we need to find its multiplicity (see (1)). Similar to step (ii), find an integer m such that $[f(X)]^m$ is a factor of $g(X)$ but $[f(X)]^{m+1}$ is not. This m will be the multiplicity of $f(X)$.

(iv) Repeat steps (ii) and (iii) for all irreducible factors of $X^n + 1$ and obtain $g(X)$ from equation (1).

vectors is equal to one, i.e., $M = 1$. For $M > 1$ received vectors, the desired $\mathbf{y}$ can be any one of the received vector. In Algorithm 1, we remove the substituted entry if the cyclic code corresponding to it has minimum distance strictly less than three. For a given non-factor $f(X)$, depending on whether its minimum distance $d_{\min}(C(n, f)) \geq 3$ or $d_{\min}(C(n, f)) < 3$, we provide bounds in Sections IV-A and IV-B respectively.

A. Lower Bound for non-factor $f(X)$ with $d_{\min}(C(n, f)) \geq 3$

Theorem 2. *Suppose $f(X)$ is an irreducible factor of X^n+1 such that $f(X)$ not a factor of $g(X)$ and $d_{\min}(C(n, f)) \geq 3$. Let $A(f)$ be the event that $f(X)$ is declared as a non-factor in step (ii) of Algorithm 1, when the number of received vectors M is equal to one. Then a lower bound on $\mathbb{P}\big[A(f)\big]$ is,*

$$\mathbb{P}\big[A(f)\big] \geq \sum_{e=0}^{\deg(f)-1} \left(1 - \frac{2^e}{2^{\deg(f)}}\right) \binom{n}{e} p^e (1-p)^{n-e}.$$

Proof: We condition on the number of erasures and find a lower bound on $\mathbb{P}[A(f)]$. Suppose the number of erasures in received $\mathbf{y}$ is equal to e. The probability of this event is equal to $\binom{n}{e}p^e(1-p)^{n-e}$. Let $\{l_1, l_2, \ldots, l_e\}$ be the set of erasure locations in $\mathbf{y}$, where $0 \leq l_1 < l_2 < \ldots < l_e < n$. The set of vectors obtained by substituting all possible 2^e values in these erasure locations is given by $\mathbf{w}_1^{2^e} = \begin{bmatrix}\mathbf{w}_1 & \mathbf{w}_2 & \ldots & \mathbf{w}_{2^e}\end{bmatrix}$. We first prove that for every $\mathbf{w}_i$, there exists a vector $\mathbf{d}_i = \begin{bmatrix}d_{i,1} & d_{i,2} & \ldots & d_{i,e}\end{bmatrix} \in \mathbb{F}_2^e$ such that

$$\begin{aligned}\mathbf{w}_i(X) &= \mathbf{v}(X) + d_{i,1}X^{l_1} + d_{i,2}X^{l_2} + \ldots + d_{i,e}X^{l_e} \quad (2)\\ &= \mathbf{v}(X) + \mathbf{d}_i(X), \quad (3)\end{aligned}$$

where $\mathbf{d}_i(X) \coloneqq d_{i,1}X^{l_1} + d_{i,2}X^{l_2} + \ldots + d_{i,e}X^{l_e}$, for $i = 1, 2, \ldots, 2^e$ and $\mathbf{v}(X)$ is the transmitted codeword. Suppose $\mathbf{w}_i$ is obtained by substituting $\mathbf{c}_i = \begin{bmatrix}c_{i,1} & c_{i,2} & \ldots & c_{i,e}\end{bmatrix} \in \mathbb{F}_2^e$ to the erased entries of $\mathbf{y}$. For the transmitted codeword $\mathbf{v}$, $\mathbf{c} = \begin{bmatrix}v_{l_1} & v_{l_2} & \ldots & v_{l_e}\end{bmatrix} \in \mathbb{F}_2^e$ will be the correct substitution. By choosing a vector $\mathbf{d}_i \in \mathbb{F}_2^e$ such that $\mathbf{d}_i = \mathbf{c}_i + \mathbf{c}$, we get (3). Since all the substituted entries take all possible values in $\mathbb{F}_2^e$, the corresponding set of vectors $\mathbf{d}_1, \mathbf{d}_2 \ldots, \mathbf{d}_{2^e}$ will also take all possible values in $\mathbb{F}_2^e$.

Let $A(f)^c$ be the complement of event $A(f)$. We find an upper bound on $\mathbb{P}[A(f)^c]$ which will provide a lower bound on $\mathbb{P}[A(f)]$. The probability of event $A(f)^c$ is given by,

$$\begin{aligned}\mathbb{P}[A(f)^c] &= \mathbb{P}\Big[\mathbf{w}_i(X) \bmod f(X) = 0 \text{ for some } \mathbf{w}_i \in \mathbf{w}_1^{2^e}\Big]\\ &\overset{(a)}{=} \mathbb{P}\Big[\mathbf{v}(X) + \mathbf{d}_i(X) \bmod f(X) = 0 \text{ for some } \mathbf{d}_i \in \mathbb{F}_2^e\Big]\\ &= \mathbb{P}\Big[\mathbf{v}(X) \bmod f(X) = \mathbf{d}_i(X) \bmod f(X)\\ &\qquad\qquad \text{for some } \mathbf{d}_i \in \mathbb{F}_2^e\Big], \quad (4)\end{aligned}$$

where equality in (a) is obtained from (3). From Proposition 3.1 of [9], $\mathbf{v}(X) \bmod f(X)$ takes any value in $\mathcal{P}_{\deg(f)}$ with probability $1/2^{\deg(f)}$ and using this in (4) we get,

$$\mathbb{P}[A(f)^c] = \frac{1}{2^{\deg(f)}} \sum_{b \in \mathcal{P}_{\deg(f)}} \mathbb{P}\Big[\mathbf{d}_i(X) \bmod f(X) = b(X) \text{ for some } \mathbf{d}_i \in \mathbb{F}_2^e\Big]. \quad (5)$$

Let $\mathcal{D}$ be the support set of $\mathbf{d}_i(X) \bmod f(X)$ and $|\mathcal{D}|$ denotes the cardinality of set $\mathcal{D}$. Depending on the number of erasures e, we now have the following two cases.

(1) Case when $e < \deg(f)$:
In this case, $\mathbf{d}_i(X) \bmod f(X)$ can take at most 2^e distinct values since $\mathbf{d}_i$ can take at most 2^e distinct values in $\mathbb{F}_2^e$, i.e., $|\mathcal{D}| \leq 2^e$. In (5), $\mathbb{P}[\mathbf{d}_i(X) \bmod f(X) = a(X)$ for some $\mathbf{d}_i \in \mathbb{F}_2^e]$ is equal to one if $a(X) \in \mathcal{D}$ and it is zero otherwise. This implies that,

$$\mathbb{P}[A(f)^c] = \frac{1}{2^{\deg(f)}}|\mathcal{D}| \leq \frac{2^e}{2^{\deg(f)}}. \quad (6)$$

(2) Case when $e \geq \deg(f)$:
In this case, $\mathbf{d}_i(X) \bmod f(X)$ can take at most $2^{\deg(f)}$ possible values and hence $|\mathcal{D}| \leq 2^{\deg(f)}$. Using this in (5) we have,

$$\mathbb{P}[A(f)^c] = \frac{1}{2^{\deg(f)}}|\mathcal{D}| \leq \frac{2^{\deg(f)}}{2^{\deg(f)}} = 1. \quad (7)$$

The required lower bound is obtained from (6) and (7) by conditioning over $E = e$ since $\mathbb{P}(A(f)) = 1 - \mathbb{P}(A(f)^c)$. ∎

B. Lower Bound for non-factor $f(X)$ with $d_{\min}(C(n, f)) < 3$

In this section, we consider the case when $d_{\min}(C(n, f)) < 3$. We first introduce some notation that will be required in this section. Let $\mathcal{F}$ be the set of irreducible factors of $X^n + 1$. For

a given $f(X) \in \mathcal{F}$ that is not a factor of $g(X)$, we define two sets $\mathcal{F}_1$ and $\mathcal{F}_2$ as follows,

$$\mathcal{F}_1 \coloneqq \{f'(X) \in \mathcal{F} \text{ such that } d_{\min}(C(n, ff')) \geq 3 \text{ and } f'(X) \text{ does not divide } g(X)\} \quad (8)$$

$$\mathcal{F}_2 \coloneqq \{f'(X) \in \mathcal{F} \text{ such that } d_{\min}(C(n, ff')) \geq 3 \text{ and } f'(X) \text{ divides } g(X)\}, \quad (9)$$

where $C(n, ff')$ is the cyclic code generated by $f(X)f'(X)$. Let $\mathbf{w}_1(X), \mathbf{w}_2(X), \ldots, \mathbf{w}_s(X)$ be the set of non-discarded polynomials in step (i)-(b) of Algorithm 1. Let $B(f)$ be the event that $f(X)$ is declared as a non-factor in step (ii) of Algorithm 1. The probability of event $B(f)$ is given by,

$$\mathbb{P}[B(f)] = \mathbb{P}\big[\mathbf{w}_i(X) \bmod f(X) \neq 0 \text{ for } i = 1, \ldots, s\big] \quad (10)$$

Remark 1. *Without loss of generality suppose $\mathbf{w}_1(X)$ in (10) corresponds to the transmitted codeword of $C(n, g)$. When $f(X)$ divides $\mathbf{w}_1(X)$, we have $\mathbb{P}[B(f)] = 0$. Thus event $B(f)$ can occur only when $f(X)$ does not divide $\mathbf{w}_1(X)$.* □

In the following lemma, we find an upper bound on $B(f)^c$ (complement of $B(f)$) when $f(X)$ does not divide $\mathbf{w}_1(X)$.

Lemma 1. *Let $B(f)$, $f(X)$, $\mathcal{F}_1$, and, $\mathcal{F}_2$ be as defined above. Suppose $f(X)$ does not divide $\mathbf{w}_1(X)$ in (10). Then under the condition the number of erasures in received vector $\mathbf{y}$ is equal to e, an upper bound on $\mathbb{P}\big[B(f)^c \big| E = e\big]$ is given by*

$$\mathbb{P}\big[B(f)^c \big| E = e\big] \leq \sum_{f' \in \mathcal{F}_1} \frac{2^e}{2^{\deg(f')}} + \sum_{f' \in \mathcal{F}_2} \beta(f'),$$

where $\beta(f')$ is given by

$$\beta(f') \coloneqq \sum_{i=d_{\min}(C(n,f'))}^{e} A_i(f') \binom{n-i}{e-i} \Big/ \binom{n}{e},$$

where the sequence of integers $\{A_0(f'), A_1(f'), \ldots, A_n(f')\}$ is the weight distribution of $C(n, f')$.

Proof: In Algorithm 1, each $\mathbf{w}_i(X)$ in (10) belongs to some cyclic code of length n with minimum distance greater than or equal to three. Since $d_{\min}(C(n, f)) < 3$, $\mathbf{w}_i(X) \bmod f(X) = 0$ implies that $\mathbf{w}_i(X) \in C(n, ff')$, such that $d_{\min}(C(n, ff')) \geq 3$. Note that it is sufficient to consider the case when $f'(X) \in \mathcal{F}$ since when $f'(X) \notin \mathcal{F}$, $f'(X)$ can be written as $f'(X) = f_1(X)f_2(X)$ such that $f_1(X) \in \mathcal{F}$ and $\mathbf{w}'_i(X) \in C(n, ff_1)$. Using this, $\mathbb{P}[B(f)^c \big| E = e]$ given by,

$$\begin{aligned}
\mathbb{P}[B(f)^c \big| E = e] &= \mathbb{P}\big[\mathbf{w}_i(X) \bmod f(X)f'(X) = 0 \text{ for }, \\
&\qquad \text{some } i, 2 \leq i \leq s \text{ and some } f'(X) \in \mathcal{F}_1 \cup \mathcal{F}_2\big] \\
&\overset{(a)}{\leq} \sum_{f' \in \mathcal{F}_1 \cup \mathcal{F}_2} \mathbb{P}\big[\mathbf{w}_i(X) \in C(n, ff') \text{ for some } i, 2 \leq i \leq s\big] \\
&\coloneqq \sum_{f' \in \mathcal{F}_1} \mathbb{P}[D(ff') \big| E = e] + \sum_{f' \in \mathcal{F}_2} \mathbb{P}[D(ff') \big| E = e], \quad (11)
\end{aligned}$$

where the upper bound in (a) follows from the union bound. The event $D(ff' \big| E = e)$ in (11) correspond to $\mathbf{w}_i(X) \in C(n, ff')$ for some $i, 2 \leq i \leq s$. The equality in (11) follows since the sets $\mathcal{F}_1$ and $\mathcal{F}_2$ are disjoint (see (8), (9)). We find an upper bound on $\mathbb{P}[B(f)^c \big| E = e]$ by finding an upper bound on each $\mathbb{P}[D(ff') \big| E = e]$ in (11).

(i) Case when $f' \in \mathcal{F}_1$:
In this case, the event $\mathbf{w}_i(X) \in C(n, ff')$ implies $\mathbf{w}_i(X) \in C(n, f')$. From (6) and (7) we have $\mathbb{P}[D(ff') \big| E = e] \leq 2^e / 2^{\deg(f')}$ (see proof of Thm. 2).

(ii) Case when $f' \in \mathcal{F}_2$:
From (3) we have, $\mathbf{w}_i(X) = \mathbf{v}(X) + \mathbf{d}_i(X)$, where $\mathbf{v}(X)$ is the transmitted codeword, for $i = 2, 3, \ldots, s$. Using this $\mathbb{P}[D(ff') \big| E = e]$ is given by,

$$\begin{aligned}
\mathbb{P}[D(ff') \big| E = e] &= \mathbb{P}\Big[[\mathbf{v}(X) + \mathbf{d}_i(X)] \in C(n, ff') \\
&\qquad \text{for some } i, 2 \leq i \leq s\Big] \\
&\overset{(a)}{\leq} \mathbb{P}\Big[[\mathbf{v}(X) + \mathbf{d}_i(X)] \in C(n, f') \text{ for some } 2 \leq i \leq s\Big], \\
&= \mathbb{P}\Big[\mathbf{d}_i(X) \in C(n, f') \text{ for some } i, 2 \leq i \leq s\Big], \quad (12)
\end{aligned}$$

where the upper bound in (a) is obtained since every codeword in $C(n, ff')$ also belongs to $C(n, f')$ and the equality in the last step follows since $\mathbf{v}(X) \in C(n, f')$. As explained in the proof of Theorem 2, for the correct substitution $\mathbf{d}_i(X) = 0$ and for any incorrect substitution $\mathbf{d}_i(X) \neq 0$. Let $\{l_1, l_2, \ldots, l_e\}$ be the set of erasure locations in $\mathbf{y}$. From (3), each $\mathbf{d}_i(X) = d_{i,1}X^{l_1} + d_{i,2}X^{l_2} + \ldots + d_{i,e}X^{l_e}$ is obtained by substituting 0 and 1 in each $d_{i,j}$, for $j = 1, 2, \ldots, e$. If there exists a codeword $\mathbf{v}(X)$ in $C(n, f')$ whose support set is contained in the locations $\{l_1, l_2, \ldots, l_e\}$ then the coefficients of $\mathbf{d}_i(X)$ can be chosen to be equal to $\mathbf{v}(X)$ and in (12) we have $\mathbb{P}\big[D(ff') \big| E = e\big] = 1$. Note that, the weight of this $\mathbf{v}(X)$ should be less than or equal to e. Using this the probability in (12) is given by

$$\mathbb{P}\big[D(ff') \big| E = e\big] \leq \sum_{i=d_{\min}(C(n,f'))}^{e} A_i \binom{n-i}{e-i} \Big/ \binom{n}{e}.$$

The upper bound of the theorem is obtained from (11). ■

Using Lemma 1 and Remark 1, we now obtain a lower bound on $\mathbb{P}[B(f)]$ in the following theorem.

Theorem 3. *Suppose $f(X)$ is an irreducible factor of $X^n + 1$ such that $f(X)$ not a factor of $g(X)$ and $d_{\min}(C(n, f)) < 3$. Let $B(f)$ be the event that $f(X)$ is declared as a non-factor in step (ii) of Algorithm 1, when the number of received vectors M is equal to one. When $E = e$, let $\alpha(e)$ be the upper bound obtained in Lemma 1. Let e' be the largest number of erasures such that $\alpha(e) < 1$. Then a lower bound on $\mathbb{P}[B(f)]$ is*

$$\mathbb{P}[B(f)] \geq \left(1 - \frac{1}{2^{\deg(f)}}\right) \sum_{e=0}^{e'} \big[1 - \alpha(e)\big] \binom{n}{e} p^e (1-p)^{n-e}.$$

Proof: From Remark 1, event $B(f)$ can happen only when $f(X)$ does not divide correct substitution $\mathbf{w}_1(X)$ and its probability is equal to $1 - (1/2^{\deg(f)})$ (Proposition 3.1 of [9]). The lower bound of the theorem follows Lemma 1 by conditioning over event $E = e$ from $e = 0, 1, \ldots, e'$. ■

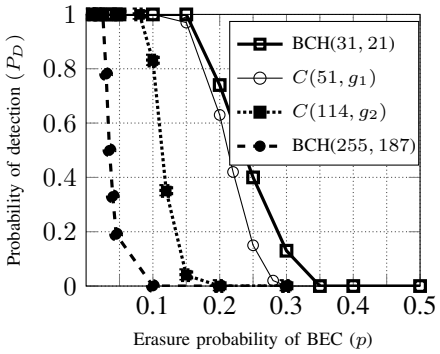

Fig. 1. Plot of P_D versus p for BCH$(31, 21)$, $C(51, g_1)$, $C(102, g_2)$, and BCH$(255, 187)$ where $g_1(X) = (X^8 + X^4 + X^3 + X + 1)(X^8 + X^5 + X^4 + X^3 + 1)$ and $g_2(X) = (X^2 + X + 1)(X^{18} + X^{16} + X^{15} + X^{14} + X^9 + X^4 + X^3 + X^2 + 1)^2$ when $M = 30$ vectors of true code are received.

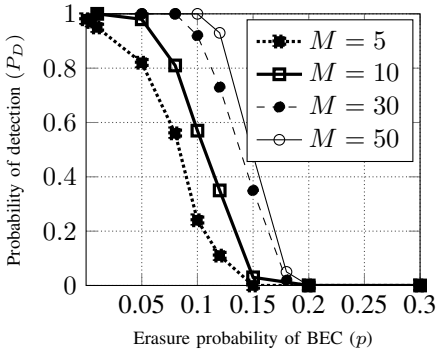

Fig. 2. Performance of Algorithm 1 for BCH$(63, 39)$ code when $M = 5$, 10, 30, and 50 vectors of the true code are received.

V. Simulation results

For simulations, we generate a sequence of codewords $\mathbf{v}_1^M$ and every bit in this sequence is erased independently with probability p to get $\mathbf{y}_1^M$. Algorithm 1 is applied to this $\mathbf{y}_1^M$. Since Algorithm 1 consists of substituting all possible values at the erased locations, the cost of implementation increases with increase in the number of erasures in the received vectors. Hence we discard the received vector if the number of erasures are more than $e_{\max}$. For the simulations we have chosen $e_{\max} = 7$. Fig. 1 shows the plot of probability of detection (P_D) versus erasure probability p for various cyclic codes. Fig. 2 shows the plot of P_D versus p for BCH$(63, 39)$ for various M. Simulations are performed using SageMath [13].

We next compare the lower bound on correctly identifying the non-factors of the generator polynomial obtained using Theorems 2 or 3 versus its true value in Fig. 3. As explained in Section IV, finding the true value is computationally intractable and hence we find approximate true value via Monte Carlo simulations of Algorithm 1 with $M = 1$ and $e_{\max} = 7$.

VI. Conclusion and future work

In this paper, we studied the problem of blind reconstruction of cyclic codes of known lengths, when the noise is introduced by BEC. We proposed a blind reconstruction algorithm and provided a lower bound on correctly identifying the factors of the generator polynomial. To the best of our knowledge, this is the first time a blind reconstruction problem has been analyzed for the binary erasure channel. As part of future work, we plan to study this problem for the case when the length of cyclic code is not known. Studying this blind reconstruction problem over BEC for other families of channel codes such as convolutional, Turbo, linear block codes is also of interest.

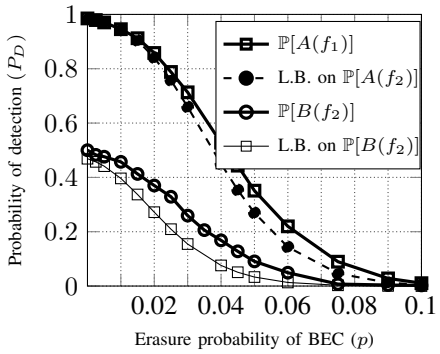

Fig. 3. For BCH$(63, 39)$, lower bound on correctly identifying the non-factors and their corresponding true values are plotted for $f_1(X) = X^6 + X^4 + X^3 + X + 1$ and $f_2(X) = X + 1$. For $f_1(X)$, lower bound (L.B.) obtained via Theorem 2 and $f_2(X)$ L.B. is obtained using Theorem 3.

Acknowledgments

This work is supported by ANR-11-LABEX-0040-CIMI within the program ANR-11-IDEX-0002-02 of Centre International de Mathématiques et Informatique de Toulouse, France.

References

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

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

[3] J. Dingel and J. Hagenauer, "Parameter estimation of a convolutional encoder from noisy observations," in *Proceedings of IEEE ISIT*, Nice, France, June 2007, pp. 1776–1780.

[4] 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.

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

[6] M. Cluzeau, "Block code reconstruction using iterative decoding techniques," in *Proc. of IEEE ISIT*, Seattle, USA, July 2006, pp. 2269–2273.

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

[8] H. Lee, C. Park, J. Lee, and Y. Song, "Reconstruction of BCH codes using probability compensation," in *Proc. of APCC*, Korea, 2012.

[9] 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.

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

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

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

[13] The Sage Developers, *Sage Mathematics Software (Version 6.2)*, 2015, `http://www.sagemath.org`.