PAPER

Analysis of Error Floors of Non-binary LDPC Codes over BEC*

Takayuki NOZAKI[†a)], ***Student Member***, **Kenta KASAI**[†b)], ***Member***, ***and*** **Kohichi SAKANIWA**[†c)], ***Fellow***

SUMMARY In this paper, we investigate the error floors of the non-binary low-density parity-check codes transmitted over the binary erasure channels under belief propagation decoding. We propose a method to improve the decoding erasure rates in the error floors by optimizing labels in zigzag cycles in the Tanner graphs of codes. Furthermore, we give lower bounds on the bit and the symbol erasure rates in the error floors. The simulation results show that the presented lower bounds are tight for the codes designed by the proposed method.

key words: *non-binary LDPC codes, error floors, belief propagation*

1. Introduction

Gallager invented low-density parity-check (LDPC) codes [2]. Due to the sparseness of the parity check matrices, LDPC codes are efficiently decoded by the belief propagation (BP) decoder. Optimized LDPC codes can exhibit performance very close to the Shannon limit [3].

Davey and MacKay [4] and others [5]–[7] have found non-binary LDPC codes can outperform binary ones. In this paper, we consider the non-binary LDPC codes defined over the Galois field $\mathbb{F}_q$ with $q = 2^m$.

A non-binary LDPC code C over $\mathbb{F}_q$ is defined by the null space of a sparse $M \times N$ parity-check matrix $\mathbf{H} = (h_{i,j})$ over $\mathbb{F}_q$:

$$C = \left\{ \boldsymbol{x} \in \mathbb{F}_q^N \mid \mathbf{H}\boldsymbol{x} = \mathbf{0} \in \mathbb{F}_q^M \right\}.$$

The Tanner graph for a non-binary LDPC code is represented by a bipartite graph with variable nodes, check nodes and labeled edges. The v-th variable node and the c-th check node are connected with an edge labeled $h_{c,v} \in \mathbb{F}_q \setminus \{0\}$ iff $h_{c,v} \neq 0$. The LDPC codes defined by Tanner graphs with the variable nodes of degree j and the check nodes of degree k are called (j, k)-regular LDPC codes. It is empirically known that the $(2, k)$-regular LDPC codes exhibit good decoding performance among other LDPC codes for $q \geq 64$ [8]. However, this is not the case for $q < 64$. In this paper, we consider the irregular non-binary LDPC codes which contain variable nodes of degree two for the generality of the code ensemble.

A stopping set $\mathcal{S}$ is a set of variable nodes such that all the neighbors of $\mathcal{S}$ are connected to $\mathcal{S}$ at least twice. For the binary LDPC code, the stopping sets are the fixed point of the BP decoder. The *peeling decoder* for the non-binary LDPC codes produces the same outputs as the BP decoder [9]. The fixed points of the peeling decoder for the non-binary LDPC codes are referred to as *stopping constellations* [9]. The stopping constellations for the non-binary LDPC codes correspond to the stopping sets for the binary LDPC codes. The error floors of non-binary LDPC codes decoded by the BP decoder are mainly caused by *nonzero* codewords or stopping constellations of small weight. We focus on nonzero codewords at first. A zigzag cycle is a cycle such that the degrees of all the variable nodes in the cycle are two. A zigzag cycle of weight s consists of s variable nodes of degree two. It is known that the set of variable nodes in a zigzag cycle forms a stopping set. For the binary LDPC codes, small zigzag cycles always yield nonzero codewords which result in serious degradation of the decoding performance. On the other hand, zigzag cycles in the non-binary codes do not always yield nonzero codewords. Let $\mathbf{H}_q^{(s)}$ denote the submatrix over $\mathbb{F}_q$ corresponding to a zigzag cycle of weight s with labels $h_1, h_2, \ldots, h_{2s}$ in the Tanner graph. For example, the submatrix $\mathbf{H}_q^{(4)}$ is written as

$$\mathbf{H}_q^{(4)} = \begin{pmatrix} h_1 & h_2 & 0 & 0 \\ 0 & h_3 & h_4 & 0 \\ 0 & 0 & h_5 & h_6 \\ h_8 & 0 & 0 & h_7 \end{pmatrix}.$$

The zigzag cycle corresponding to $\mathbf{H}_q^{(s)}$ yields nonzero codeword iff $\mathbf{H}_q^{(s)}$ is singular, i.e.,

$$\det \mathbf{H}_q^{(s)} = \prod_{i=1}^{s} h_{2i} + \prod_{i=1}^{s} h_{2i-1} = 0,$$

which is equivalent to

$$\beta := \prod_{i=1}^{s} h_{2i-1}^{-1} h_{2i} = 1.$$

It can be seen that zigzag cycles in the Tanner graphs for the binary LDPC codes always yield nonzero codewords since $\det \mathbf{H}_2^{(s)} = 0$. On the other hand, for the non-binary case, zigzag cycles in the Tanner graphs do not yield nonzero codewords if the corresponding submatrices are nonsingular.

Manuscript received March 28, 2011.
Manuscript revised July 16, 2011.

[†]The authors are with the Dept. of Communications and Integrated Systems, Tokyo Institute of Technology, Tokyo, 152-8550 Japan.

*The material in this paper was presented in part at IEEE International Symposium on Information Theory (ISIT 2010) [1].

a) E-mail: nozaki@comm.ss.titech.ac.jp
b) E-mail: kenta@comm.ss.titech.ac.jp
c) E-mail: sakaniwa@comm.ss.titech.ac.jp

DOI: 10.1587/transfun.E95.A.381

To lower the error floors of codes under maximum likelihood decoding, Poulliat et al. proposed *cycle cancellation* [10]. The cycle cancellation is a method to design the edge labels in zigzag cycles so that the corresponding submatrices are nonsingular. We see that from the simulation result [10] the resulting codes have lower error floors under BP decoding. However, it is found in our analyses that some zigzag cycles, even if their submatrices are nonsingular, can cause decoding failures under BP decoding over the binary erasure channel (BEC), i.e., some zigzag cycles yield stopping constellations.

In this paper, we analyze *nonsingular* zigzag cycles which cause the decoding failures under BP decoding. We clarify that the condition for successful decoding of zigzag cycles over the BEC depends on the parameter β. More precisely, if the parameter β is not a nonzero element of proper subfields of $\mathbb{F}_q$, the zigzag cycles do not yield stopping constellations. Based on this fact, we propose a design method of selecting labels so as to eliminate small zigzag cycles which yield stopping constellations.

For the binary LDPC code ensembles over the BEC, a closed-form expression for the bit erasure rate in the error floors was given in [11, p.155]. However, for the non-binary LDPC code ensembles, no closed-form expressions or bounds for the bit and the symbol erasure rates in the error floors have been given. In this paper, we give lower bounds on the bit and the symbol erasure rates in the error floors for the non-binary LDPC code ensembles. More precisely, those lower bounds are derived from the decoding erasures caused by the zigzag cycles. Furthermore, the simulation results show that the lower bounds on the bit and the symbol erasure rates are tight for the expurgated ensemble constructed by our proposed method over the BEC.

This paper is organized as follows. In Sect. 2, we briefly review the peeling decoder for the non-binary LDPC codes, define stopping constellations and define decoding failures. In Sect. 3, we investigate BP decoding of zigzag cycles over the BEC and propose the improved cycle cancellation. In Sect. 4, we give lower bounds on the bit and the symbol erasure rates in the error floors for expurgated ensembles.

2. Preliminaries

Let α be a primitive element of $\mathbb{F}_{2^m}$. Once a primitive element α of $\mathbb{F}_{2^m}$ is fixed, each symbol is given by an m-bit representation [12, p.110]. We denote the m-bit representation of $\gamma \in \mathbb{F}_{2^m}$, by $b(\gamma)$. We denote the i-th bit of $b(\gamma)$, by $b_i(\gamma)$. For example, with a primitive element $\alpha \in \mathbb{F}_{2^3}$ such that $\alpha^3 + \alpha + 1 = 0$, each symbol is represented as $b(0) = (0,0,0)$, $b(1) = (1,0,0)$, $b(\alpha) = (0,1,0)$, $b(\alpha^2) = (0,0,1)$, $b(\alpha^3) = (1,1,0)$, $b(\alpha^4) = (0,1,1)$, $b(\alpha^5) = (1,1,1)$ and $b(\alpha^6) = (1,0,1)$.

Let N be the symbol code length. We regard the codewords in the non-binary LDPC codes as binary codewords, i.e., the codewords $\boldsymbol{x}$ are represented by $(x_{1,1}, x_{1,2}, \ldots, x_{N,m})$. We consider the transmission over the BEC. The channel output takes values in the alphabet $\{0, 1, ?\}$, where ? indicates an erasure. We denote the received word as $(y_{1,1}, y_{1,2}, \ldots, y_{N,m})$.

The BP decoder for the non-binary LDPC codes [4] exchanges messages of length 2^m. We assume that all-zero codewords are sent without loss of generality to analyze the decoding error rate [13, Lemma 1]. All the non-zero entries in a message arising in the BP decoder are equal [13, Lemma 2]. Moreover, the set of the m-bit representations for the indices corresponding to nonzero entries of a message arising in the BP decoder forms a linear subspace of $\mathbb{F}_2^m$ [13, Lemma 2]. In other words, the set of the indices corresponding to nonzero entries of a message arising in the BP decoder is closed under the addition in $\mathbb{F}_{2^m}$. Hence, each message in the BP decoder is represented by a subset in $\mathbb{F}_{2^m}$ which is closed under the addition in $\mathbb{F}_{2^m}$.

2.1 Peeling Decoder

To analyze the condition of successful decoding under BP decoding, we need to analyze the fixed points of the peeling decoder for the non-binary LDPC codes which are referred to as *stopping constellations* [9]. To understand the stopping constellation, we recall the *states* in the peeling decoder for the non-binary case [9].

The peeling decoder assigns a set of candidate symbols for the decoding result to each variable node. Such a set is referred to as state of the v-th variable node and denoted by E_v, where $E_v \subseteq \mathbb{F}_{2^m}$. Recall that we assume that the all-zero codewords are sent. Initially, for all $v \in \{1, 2, \ldots, N\}$, the peeling decoder assigns the state

$$E_v = \{\gamma \mid b_i(\gamma) = 0 \text{ (for } i \text{ s.t. } y_{v,i} = 0), \\ b_j(\gamma) \in \{0,1\} \text{ (for } j \text{ s.t. } y_{v,j} = ?)\} \quad (1)$$

to the v-th variable node. In words, the peeling decoder assigns the states corresponding to the channel outputs to the variable nodes. Let $\mathcal{N}(c)$ be the set of the position of the variable nodes connecting to the c-th check node. Let $h_{c,i}$ be the label on the edge connected to the variable node in the position $i \in \mathcal{N}(c)$ and the c-th check node. For any subsets $A_1, A_2, \ldots, A_k \subseteq \mathbb{F}_{2^m}$, we denote $\sum_{i=1}^k A_i := \{\sum_{i=1}^k a_i \mid a_j \in A_j \ (j = 1, 2, \ldots, k)\}$. To simplify the notation, for $\gamma \in \mathbb{F}_{2^m}$ and $E \subseteq \mathbb{F}_{2^m}$, we define $\gamma E := \{\gamma e \mid e \in E\}$. If $E_v \cap h_{c,v}^{-1}(\sum_{i \in \mathcal{N}(c)\setminus\{v\}} h_{c,i} E_i)$ is a proper subset of E_v, then (v, c) is said to be an *active pair*. The peeling decoder involves the following 3 steps:

1. Initially the peeling decoder assigns the states corresponding to the channel outputs to the variable nodes.
2. The peeling decoder chooses an active pair (v, c) uniformly at random. The peeling decoder assigns $E_v \leftarrow E_v \cap h_{c,v}^{-1}(\sum_{i \in \mathcal{N}(c)\setminus\{v\}} h_{c,i} E_i)$ to the v-th variable node.
3. If there is no active pair, then the peeling decoder stops. Otherwise repeat step 2.

Note that the cardinality of the states of the variable nodes do not increase as decoding proceeds.

The states are the subset in $\mathbb{F}_{2^m}$ which is closed under the addition in $\mathbb{F}_{2^m}$. The proof is similar to the proof of [13, Lemma 2] and [14, Lemma 2]. From Eq. (1), initially, the states are subset in $\mathbb{F}_{2^m}$ which is closed under the addition in $\mathbb{F}_{2^m}$. We claim that if the subset $E \subseteq \mathbb{F}_{2^m}$ is closed under the addition, the subset γE is also closed under the addition for $\gamma \in \mathbb{F}_{2^m} \setminus \{0\}$. For all $e'_1, e'_2 \in \gamma E$, there exist $e_1, e_2 \in E$ such that $e'_1 = \gamma e_1$ and $e'_2 = \gamma e_2$. For all $e'_1, e'_2 \in \gamma E$, we see that

$$e'_1 + e'_2 = \gamma e_1 + \gamma e_2 = \gamma(e_1 + e_2) \in \gamma E.$$

Hence, the subset $\gamma E \subseteq \mathbb{F}_{2^m}$ is closed under the addition if $E \subseteq \mathbb{F}_{2^m}$ is closed under the addition. We claim that the subset $E_1 \cap E_2 \subseteq \mathbb{F}_{2^m}$ is closed under the addition if the subsets $E_1, E_2 \subseteq \mathbb{F}_{2^m}$ are closed under the addition. For all $e_1, e_2 \in E_1 \cap E_2$, we see that $e_1 + e_2 \in E_1$ and $e_1 + e_2 \in E_2$ since $e_1, e_2 \in E_1$ and $e_1, e_2 \in E_2$. Since $e_1 + e_2 \in E_1 \cap E_2$, the subset $E_1 \cap E_2 \subseteq \mathbb{F}_{2^m}$ is closed under the addition if the subsets $E_1, E_2 \subseteq \mathbb{F}_{2^m}$ are closed under the addition. Obviously, if the subsets $E_1, E_2, \ldots, E_k \in \mathbb{F}_{2^m}$ are closed under the addition, $\sum_{i=1}^{k} E_1$ is closed under the addition. Hence $E_v \cap h_{c,v}^{-1}(\sum_{i \in \mathcal{N}(c)\setminus\{v\}} h_{c,i} E_i)$ is closed under the addition, if E_i is closed under the addition for $i \in \mathcal{N}(c) \setminus \{v\}$. Recall that initially the states are subset in $\mathbb{F}_{2^m}$ which is closed under the addition in $\mathbb{F}_{2^m}$. Thus, all the states are closed under the addition in $\mathbb{F}_{2^m}$.

2.2 Stopping Constellation

A stopping constellation $\{E_v\}_{v\in\{1,2,\ldots,N\}}$ is defined as an assignment of states such that

$$E_v \subseteq h_{c,v}^{-1}\left(\sum_{i\in\mathcal{N}(c)\setminus\{v\}} h_{c,i} E_i\right)$$

for any $v \in \{1, 2, \ldots, N\}$ and the check nodes in the position $c \in \mathcal{N}(v)$. In other words, stopping constellations are fixed points of the peeling decoder. In this paper, we refer to the number of states whose cardinality are not equal to 1 as the *weight* of the stopping constellation.

For the BEC and sufficiently large number of iterations, the BP decoder stops in a fixed point of decoding. In [9], Rathi et al. proved that the BP decoder and the peeling decoder stop in the largest stopping constellation contained in the subsets in $\mathbb{F}_{2^m}$ corresponding to the channel outputs. In other words, the BP decoder and the peeling decoder stop in the same fixed point of decoding. Thus, if we analyze stopping constellations, we are able to analyze the condition of successful decoding under BP decoding.

2.3 Decoding Failure

Recall that we assume that all-zero codewords are sent. The decoding failures are defined from the states of the variable nodes in the fixed point of decoding. The symbol corresponding to the v-th variable node is *correct* if and only if $E_v = \{0\}$. The block is correct if and only if $E_v = \{0\}$ for all $v \in \{1, 2, \ldots, N\}$. The i-th bit in the v-th symbol is correct if and only if $b_i(\gamma) = 0$ for all $\gamma \in E_v$. The block erasure rate, the symbol erasure rate and the bit erasure rate are defined by the fraction of the blocks, the symbols and the bits which are not correct, respectively.

3. Zigzag Cycle Code Analysis

A zigzag cycle is a cycle such that the degrees of all the variable nodes in the cycle are two. The zigzag cycle code is defined by a Tanner graph which forms a single zigzag cycle as shown in Fig. 1. In this section, we investigate the zigzag cycle codes to clarify a condition for decoding failures on the zigzag cycles in Tanner graphs. We also show the decoding performance for zigzag cycle codes under BP decoding.

3.1 Condition for Successful Decoding

In the following theorem, we clarify a necessary condition for successful decoding of the zigzag cycle codes over the BEC by the BP decoder.

Theorem 1: Consider zigzag cycle codes of length s with labels $h_1, h_2, \ldots, h_{2s} \in \mathbb{F}_{2^m} \setminus \{0\}$ over the BEC. Let α be the primitive element of $\mathbb{F}_{2^m}$. Define

$$\mathcal{H}_m := \bigcup_{r>0:r|m,r\neq m} \left\{\alpha^{i(2^m-1)/(2^r-1)} \mid i = 0, 1, \ldots, 2^r - 2\right\}. \quad (2)$$

All the symbols in a zigzag cycle code are correct unless all the bits are erased, if

$$\prod_{i=1}^{s} h_{2i-1}^{-1} h_{2i} = \beta \notin \mathcal{H}_m.$$

Specifically, $\{1\} = \mathcal{H}_m \subseteq \mathbb{F}_{2^m}$ for a prime m.

The proof of Theorem 1 shall be shown in Appendix. Note that $\{\alpha^{i(2^m-1)/(2^r-1)} \mid i = 0, 1, \ldots, 2^r - 2\}$ forms the set of the nonzero elements of the proper subfield of order 2^r for $r \mid m$. In other words, $\mathcal{H}_m$ consists of the nonzero elements of the proper subfields of $\mathbb{F}_{2^m}$.

We refer to β as the *cycle parameter* of the zigzag cycle code. Theorem 1 shows that the condition of successful decoding under BP decoding for the zigzag cycle codes over the BEC depends on the cycle parameter β. In Table 1, we

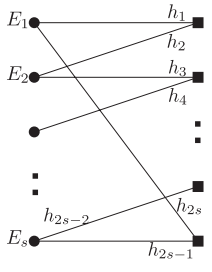

Fig. 1 A zigzag cycle of weight s with labels $h_1, h_2, \ldots, h_{2s}$.

Table 1 The elements of $\mathcal{H}_m$ over $\mathbb{F}_{2^m}$ for $m = 4, 6, 8, 9$.

Field	The elements of $\mathcal{H}_m$
$\mathbb{F}_{2^4}$	$1, \alpha^5, \alpha^{10}$
$\mathbb{F}_{2^6}$	$1, \alpha^9, \alpha^{18}, \alpha^{21}, \alpha^{27}, \alpha^{36}, \alpha^{42}, \alpha^{45}, \alpha^{54}$
$\mathbb{F}_{2^8}$	$1, \alpha^{17}, \alpha^{34}, \alpha^{51}, \alpha^{68}, \alpha^{85}, \alpha^{102}, \alpha^{119}, \alpha^{136}, \alpha^{153}, \alpha^{170}, \alpha^{187}, \alpha^{204}, \alpha^{221}, \alpha^{238}$
$\mathbb{F}_{2^9}$	$1, \alpha^{73}, \alpha^{146}, \alpha^{219}, \alpha^{292}, \alpha^{365}, \alpha^{438}$

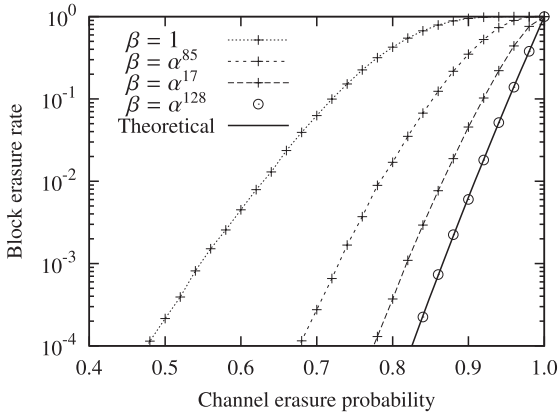

Fig. 2 The block erasure rates for zigzag cycle codes with cycle parameter $\beta = 1, \alpha^{85}, \alpha^{17}, \alpha^{128}$ over the BEC under BP decoding. The zigzag cycle codes are of weight 6 over $\mathbb{F}_{2^8}$. Let ϵ be the channel erasure probability. The solid curve shows the theoretical block erasure rate ϵ^{48} of zigzag cycle codes with cycle parameter $\beta \notin \mathcal{H}_8$.

list the cycle parameters in $\mathcal{H}_m \subseteq \mathbb{F}_{2^m}$ for $m = 4, 6, 8$ and 9. It follows from Theorem 1 that it is desired to avoid the zigzag cycle codes with cycle parameter $\beta \in \mathcal{H}_m$, since those codes can cause decoding failures even if not all the bits are erased.

We propose an *improved cycle cancellation* to get lower error floors. The improved cycle cancellation is a method to design the labels in Tanner graphs so that zigzag cycles of small weight satisfy $\beta \notin \mathcal{H}_m$. The zigzag cycles designed by the improved cycle cancellation are successfully decoded under BP decoding unless all the bits are erased. Hence, zigzag cycles designed by the improved cycle cancellation recover more erasures than those designed by the cycle cancellation [10].

We compare the block erasure rates of zigzag cycle codes designed by the improved cycle cancellation with that of zigzag cycle codes satisfying $\beta \in \mathcal{H}_m$ in Sect. 3.2.

3.2 Simulation Results

Figure 2 shows the block erasure rates of zigzag cycle codes over the BEC under BP decoding. Each curve of $\beta = \alpha^j$ in Fig. 2 shows the block erasure rate of zigzag cycle codes of weight 6 over $\mathbb{F}_{2^8}$ with cycle parameter $\beta = 1, \alpha^{17}, \alpha^{85} \in \mathcal{H}_8$. The circles in Fig. 2 show the block erasure rate of zigzag cycles with cycle parameter $\beta = \alpha^{128} \notin \mathcal{H}_8$.

The solid curve in Fig. 2 shows the theoretical block erasure rate of zigzag cycle codes with cycle parameter $\beta \notin \mathcal{H}_8$. A zigzag cycle code is *recoverable* if all the symbols in the zigzag cycle code are correct by the BP decoder. The zigzag cycle codes with cycle parameter $\beta \notin \mathcal{H}_8$ are recoverable unless all the bits are erased. All the bits are erased with probability ϵ^{48} for the BEC with channel erasure probability ϵ since the bit code length is 6 symbols or equivalently 6×8=48 bits. Hence, the theoretical block erasure rate of zigzag cycle codes designed by the improved cycle cancellation is given by ϵ^{48}.

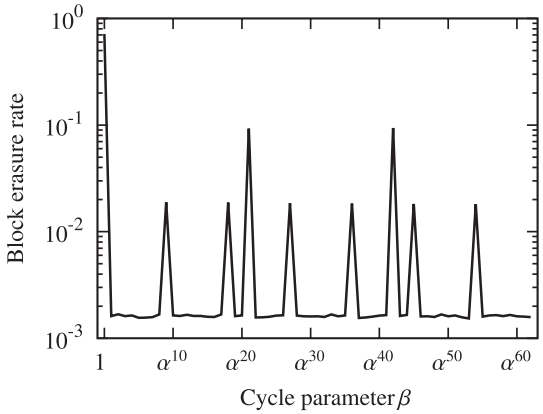

Fig. 3 The block erasure rates for zigzag cycle codes over the BEC with channel erasure probability 0.7 under BP decoding. The zigzag cycle codes are weight 3 over $\mathbb{F}_{2^6}$. We see that the zigzag cycle codes with cycle parameter $\beta \notin \mathcal{H}_6$ exhibit good decoding performance, where $\mathcal{H}_6 = \{1, \alpha^9, \alpha^{18}, \alpha^{21}, \alpha^{27}, \alpha^{36}, \alpha^{42}, \alpha^{45}, \alpha^{54}\}$.

The cycle cancellation avoids only the zigzag cycles with cycle parameter $\beta = 1$. In other words, the cycle cancellation cannot avoid the zigzag cycles with cycle parameter $\beta = \alpha^{17}$ and $\beta = \alpha^{85}$. On the other hand, the improved cycle cancellation avoids the zigzag cycles with cycle parameter not only $\beta = 1$ but also $\beta = \alpha^{17}$ and $\beta = \alpha^{85}$ since $1, \alpha^{17}, \alpha^{85} \in \mathcal{H}_8$.

The smallest stopping state is defined in Sect. A.3. The smallest stopping state containing 1 for $\beta = \alpha^{85}$ is given by $\{0, 1, \alpha^{85}, \alpha^{170}\}$. Then, the cardinality of this stopping state is 4. On the other hand, the smallest stopping state containing 1 for $\beta = \alpha^{17}$ is given by $\{0\} \cup \{\alpha^{17i} \mid i = 0, 1, \ldots, 14\}$. Then, the cardinality of this stopping state is 16. We see that from Fig. 2 the block erasure rate increases as the cardinality of the smallest stopping state decreases.

Figure 3 shows the block erasure rates of zigzag cycle codes over the BEC with channel erasure probability 0.7 under BP decoding. The zigzag cycle codes are weight 3 over $\mathbb{F}_{2^6}$. From Fig. 3, we see that the zigzag cycle codes with cycle parameter $\beta \notin \mathcal{H}_6$ exhibit good decoding performance.

4. Error Floor Analysis

From Theorem 1, we see that no zigzag cycles designed by the improved cycle cancellation are recoverable iff all the bits in the zigzag cycles are erased. From Sect. A.2, we see that all the zigzag cycles are not recoverable if all the bits are erased. By using this result, in this section, we give lower bounds on the bit and the symbol erasure rates under

BP decoding for an expurgated ensembles. More precisely, those lower bounds are derived from the decoding erasures caused by the zigzag cycles. Simulation results show that those lower bounds are tight bounds on the bit and the symbol erasure rates in the error floors for the expurgated ensembles designed by our proposed method.

4.1 Code Ensemble

Recall that a stopping set $\mathcal{S}$ is a set of variable nodes such that all the neighbors of $\mathcal{S}$ are connected to $\mathcal{S}$ at least twice. Since the stopping sets depend only on the structure of a Tanner graph, we extend the definition of the stopping set for the non-binary LDPC codes. Recall that a zigzag cycle is a cycle such that the degrees of all the variable nodes in the cycles are two. Since all the neighbors of the set $\mathcal{Z}$ of the variable nodes in a zigzag cycle are connected to $\mathcal{Z}$ exactly twice, the set $\mathcal{Z}$ of the variable nodes in a zigzag cycle forms a stopping set.

To analyze the bit and the symbol erasure rates in the error floors of the non-binary LDPC codes, we consider the following expurgated ensemble.

Definition 1: Let $\mathrm{LDPC}(N, m, \lambda, \rho)$ denote the set of LDPC codes of symbol code length N over $\mathbb{F}_{2^m}$ defined by Tanner graphs with a degree distribution pair (λ, ρ) [11] and labels of edges chosen elements from $\mathbb{F}_{2^m} \setminus \{0\}$ uniformly at random. Let $s_\mathrm{g} \in \mathbb{N}$ be an expurgation parameter. The expurgated ensemble $\mathrm{ELDPC}(N, m, \lambda, \rho, s_\mathrm{g})$ consists of the subset of codes in $\mathrm{LDPC}(N, m, \lambda, \rho)$ which contain no stopping sets of size in $\{1, \ldots, s_\mathrm{g} - 1\}$. Note that the expurgated ensemble $\mathrm{ELDPC}(N, m, \lambda, \rho, 1)$ is equivalent to $\mathrm{LDPC}(N, m, \lambda, \rho)$. Let $s_\mathrm{c} \in \mathbb{N}$ be an expurgation parameter for labeling in the Tanner graph, where $s_\mathrm{g} \le s_\mathrm{c}$. Define expurgated ensemble $\mathrm{ELDPC}(N, m, \lambda, \rho, s_\mathrm{g}, s_\mathrm{c}, \mathcal{H})$ as the subset of codes in $\mathrm{ELDPC}(N, m, \lambda, \rho, s_\mathrm{g})$ which contain no zigzag cycles of weight in $\{s_\mathrm{g}, \ldots, s_\mathrm{c} - 1\}$ with cycle parameter $\beta \in \mathcal{H}$.

Since the sets of the variable nodes in zigzag cycles form stopping sets, the codes in the expurgated ensemble $\mathrm{ELDPC}(N, m, \lambda, \rho, s_\mathrm{g})$ contain no zigzag cycles of weight in $\{1, 2, \ldots, s_\mathrm{g} - 1\}$.

Example 1: The codes in the expurgated ensemble $\mathrm{ELDPC}(N, m, \lambda, \rho, s_\mathrm{g}, s_\mathrm{c}, \{1\})$ contain no stopping sets of size in $\{1, 2, \ldots, s_\mathrm{g} - 1\}$ and no zigzag cycles with cycle parameter $\beta = 1$ of weight in $\{s_\mathrm{g}, \ldots, s_\mathrm{c} - 1\}$. In other words, the expurgated ensemble $\mathrm{ELDPC}(N, m, \lambda, \rho, s_\mathrm{g}, s_\mathrm{c}, \{1\})$ is constructed by the cycle cancellation. Since the sets of the variable nodes in zigzag cycles form stopping sets, the codes in $\mathrm{ELDPC}(N, m, \lambda, \rho, s_\mathrm{g}, s_\mathrm{c}, \{1\})$ contain no zigzag cycles of weight in $\{1, 2, \ldots, s_\mathrm{g} - 1\}$.

Recall that $\mathcal{H}_m$ is defined as in Eq. (2). Similarly, the expurgated ensemble $\mathrm{ELDPC}(N, m, \lambda, \rho, s_\mathrm{g}, s_\mathrm{c}, \mathcal{H}_m)$ is constructed by the improved cycle cancellation.

4.2 Analysis of Error Floors

The following theorem gives lower bounds on the bit and the symbol erasure rates under BP decoding for the expurgated ensemble $\mathrm{ELDPC}(N, m, \lambda, \rho, s_\mathrm{g}, s_\mathrm{c}, \mathcal{H}_m)$.

Theorem 2: Let $\mathrm{P_b}(N, m, \lambda, \rho, s_\mathrm{g}, s_\mathrm{c}, \mathcal{H}_m, \epsilon)$ and $\mathrm{P_s}(N, m, \lambda, \rho, s_\mathrm{g}, s_\mathrm{c}, \mathcal{H}_m, \epsilon)$ be the bit and the symbol erasure rates, respectively, for the expurgated ensemble $\mathrm{ELDPC}(N, m, \lambda, \rho, s_\mathrm{g}, s_\mathrm{c}, \mathcal{H}_m)$ by the BP decoder over the BEC with channel erasure probability ϵ. Define $\mu := \lambda'(0)\rho'(1)$ and

$$\epsilon_m^* := \begin{cases} 1, & \mu \le 1, \\ \mu^{-\frac{1}{m}}, & \mu > 1. \end{cases} \tag{3}$$

For sufficiently large N, the bit and the symbol erasure rates for $\mu > 0$ and $\epsilon < \epsilon_m^*$ are bounded by

$$\mathrm{P_b}(N, m, \lambda, \rho, s_\mathrm{g}, s_\mathrm{c}, \mathcal{H}_m, \epsilon) \ge \frac{1}{2N} \frac{(\mu\epsilon^m)^{s_\mathrm{g}}}{1 - \mu\epsilon^m} + o\left(\frac{1}{N}\right), \tag{4}$$

$$\mathrm{P_s}(N, m, \lambda, \rho, s_\mathrm{g}, s_\mathrm{c}, \mathcal{H}_m, \epsilon) \ge \frac{1}{2N} \frac{(\mu\epsilon^m)^{s_\mathrm{g}}}{1 - \mu\epsilon^m} + o\left(\frac{1}{N}\right). \tag{5}$$

proof: First, we will consider the symbol erasure rate. The symbol erasure rate is represented by the sum of two contributions, the symbol erasures caused by the stopping constellations from the zigzag cycles and from the stopping sets other than the zigzag cycles[†]. Let $\tilde{P}_\mathrm{z}(N, m, \lambda, \rho, s_\mathrm{g}, s_\mathrm{c}, \mathcal{H}_m, \epsilon)$ and $\tilde{P}_\mathrm{ss}(N, m, \lambda, \rho, s_\mathrm{g}, s_\mathrm{c}, \mathcal{H}_m, \epsilon)$ be the contributions of the zigzag cycles and of the stopping sets other than the zigzag cycles, respectively, for the symbol erasure rates of the ensemble $\mathrm{ELDPC}(N, m, \lambda, \rho, s_\mathrm{g}, s_\mathrm{c}, \mathcal{H}_m)$ over the BEC with channel erasure probability ϵ. Then, we have

$$\begin{aligned} &\mathrm{P_s}(N, m, \lambda, \rho, s_\mathrm{g}, s_\mathrm{c}, \mathcal{H}_m, \epsilon) \\ &\quad = \tilde{P}_\mathrm{z}(N, m, \lambda, \rho, s_\mathrm{g}, s_\mathrm{c}, \mathcal{H}_m, \epsilon) \\ &\qquad + \tilde{P}_\mathrm{ss}(N, m, \lambda, \rho, s_\mathrm{g}, s_\mathrm{c}, \mathcal{H}_m, \epsilon) \\ &\quad \ge \tilde{P}_\mathrm{z}(N, m, \lambda, \rho, s_\mathrm{g}, s_\mathrm{c}, \mathcal{H}_m, \epsilon). \end{aligned}$$

In words, the symbol erasure rate is lower bounded by the contribution of the zigzag cycles for the symbol erasure rate.

We will consider $\tilde{P}_\mathrm{z}(N, m, \lambda, \rho, s_\mathrm{g}, s_\mathrm{c}, \mathcal{H}_m, \epsilon)$. Let $\tilde{P}_1(N, s, m, \lambda, \rho, s_\mathrm{g}, s_\mathrm{c}, \mathcal{H}_m, \epsilon)$ be the symbol erasure rate caused by the stopping constellations from zigzag cycles of weight s under BP decoding over the BEC with channel erasure probability ϵ for $\mathrm{ELDPC}(N, m, \lambda, \rho, s_\mathrm{g}, s_\mathrm{c}, \mathcal{H}_m)$. From Definition 1, codes in the expurgated ensemble $\mathrm{ELDPC}(N, m, \lambda, \rho, s_\mathrm{g}, s_\mathrm{c}, \mathcal{H}_m)$ contain no zigzag cycles of weight in $\{1, 2, \ldots, s_\mathrm{g} - 1\}$. Hence, we consider the symbol erasure rate caused by stopping constellations from zigzag cycles of weight at least s_g. If we fix a finite W and let N tend to infinity, the zigzag cycles of weight at most W become asymptotically non-overlapping with high probability

[†]For a Fixed Tanner graph and a given stopping set $\mathcal{S}$, there exist at least one stopping constellation $\{E_v\}_{v\in\{1,2,\ldots,N\}}$ such that the set of variable nodes in $\{v \mid E_v \neq \{0\}\}$ is $\mathcal{S}$ [15, Lemma 2]. In this proof, we refer those stopping constellations to stopping constellations from stopping set $\mathcal{S}$.

[11, p.155]. Thus, for a fixed W and sufficiently large N we have

$$\tilde{P}_{\mathrm{z}}(N,m,\lambda,\rho,s_{\mathrm{g}},s_{\mathrm{c}},\mathcal{H}_m,\epsilon) \geq \sum_{s=s_{\mathrm{g}}}^{W} \tilde{P}_1(N,s,m,\lambda,\rho,s_{\mathrm{g}},s_{\mathrm{c}},\mathcal{H}_m,\epsilon).$$

In Sect. 3.2, zigzag cycle codes designed by the improved cycle cancellation can not be recovered iff all the bits are erased. From this result, we see that zigzag cycles with cycle parameter $\beta \notin \mathcal{H}_m$ in a Tanner graph can not be recovered iff all the bits in the cycle are erased, which happens with probability ϵ^{ms}. In other words, symbols in zigzag cycles of weight $s \in \{s_{\mathrm{g}},\ldots,s_{\mathrm{c}}-1\}$ are not recovered with probability ϵ^{ms}. From Sect. A.2, no symbols in the zigzag cycle of weight s with cycle parameter $\beta \in \mathcal{H}_m$ are correct if all the bits in the zigzag cycle are erased. Hence, all the zigzag cycles are not recovered with probability at least ϵ^{ms}. In other words, the zigzag cycles of weight $s \in \{s_{\mathrm{c}},\ldots,W\}$ are not recovered with probability at least ϵ^{ms}. By [11, C. 37] for a fixed W, the expectation of the number of zigzag cycles of weight $s \leq W$ in the expurgated ensemble ELDPC$(N,m,\lambda,\rho,s_{\mathrm{g}},s_{\mathrm{c}},\mathcal{H}_m)$ is given by

$$\frac{\mu^s}{2s},$$

for sufficiently large N. From Sect. A.3, if all the bits in the zigzag cycle are erased, no symbols in zigzag cycle are correct. Hence, if all the bits in the zigzag cycle of weight s are erased, the zigzag cycle causes s symbol erasures. Since s symbols are in the zigzag cycles of weight s, the zigzag cycles of weight s cause a symbol erasure rate of s/N if the bits in the zigzag cycles of weight s are erased. Therefore, for sufficiently large N, we have for $s \in \{s_{\mathrm{g}},\ldots,s_{\mathrm{c}}-1\}$,

$$\tilde{P}_1(N,s,m,\lambda,\rho,s_{\mathrm{g}},s_{\mathrm{c}},\mathcal{H}_m,\epsilon) = \frac{1}{2N}\mu^s\epsilon^{ms} + o\left(\frac{1}{N}\right),$$

and for $s \in \{s_{\mathrm{c}},\ldots,W\}$

$$\tilde{P}_1(N,s,m,\lambda,\rho,s_{\mathrm{g}},s_{\mathrm{c}},\mathcal{H}_m,\epsilon) \geq \frac{1}{2N}\mu^s\epsilon^{ms} + o\left(\frac{1}{N}\right).$$

Thus, we have

$$\tilde{P}(N,m,\lambda,\rho,s_{\mathrm{g}},s_{\mathrm{c}},\mathcal{H}_m,\epsilon) \geq \frac{1}{2N}\sum_{s=s_{\mathrm{g}}}^{W}\mu^s\epsilon^{ms} + o\left(\frac{1}{N}\right).$$

If $\epsilon < \epsilon_m^*$, for sufficiently large N and W, we see that

$$\tilde{P}(N,m,\lambda,\rho,s_{\mathrm{g}},s_{\mathrm{c}},\mathcal{H}_m,\epsilon) \geq \frac{1}{2N}\frac{(\mu\epsilon^m)^{s_{\mathrm{g}}}}{1-\mu\epsilon^m} + o\left(\frac{1}{N}\right).$$

Hence, for sufficiently large N, the symbol decoding erasure rate is bounded by

$$\mathrm{P_s}(N,m,\lambda,\rho,s_{\mathrm{g}},s_{\mathrm{c}},\mathcal{H}_m,\epsilon) \geq \frac{1}{2N}\frac{(\mu\epsilon^m)^{s_{\mathrm{g}}}}{1-\mu\epsilon^m} + o\left(\frac{1}{N}\right).$$

We will consider the bit erasure rate. The proof is similar to the proof for the symbol erasure rate. From Sect. A.3, if all the bits in the zigzag cycle are erased, all the states of the variable nodes in the zigzag cycle are equal to $\mathbb{F}_{2^m}$. Hence, if all the bits in the zigzag cycle are erased, no bits in the zigzag cycle are correct. Note that the bit code length is Nm. Since sm bits are in the zigzag cycles of weight s, the zigzag cycles of weight s cause a bit erasure rate of s/N if all the bits in the zigzag cycles of weight s are erased. Thus, the bit erasure rate caused by zigzag cycles is lower bounded by

$$\frac{1}{2N}\sum_{s=s_{\mathrm{g}}}^{W}\mu^s\epsilon^{ms} + o\left(\frac{1}{N}\right).$$

By using this result, we obtain a lower bound on the bit erasure rate for the expurgated ensemble similarly. □

Discussion 1: Since the symbol and the bit erasure rates of all the zigzag cycles of weight s are lower bounded by ϵ^{sm}, the bit and the symbol erasure rates are not dependent on the parameter s_{c} and the subset $\mathcal{H}_m$. Hence, Eqs. (4) and (5) do not depend on the parameter s_{c} and the subset $\mathcal{H}_m$.

4.3 Simulation Results

Figure 4 compares the symbol erasure rate for the expurgated ensemble constructed by the improved cycle cancellation ELDPC$(315,4,x,x^2,1,8,\mathcal{H}_4)$ with that for the expurgated ensemble constructed by the cycle cancellation ELDPC$(315,4,x,x^2,1,8,\{1\})$, where $\mathcal{H}_4 = \{1,\alpha^5,\alpha^{10}\}$. It can be seen that our proposed codes exhibit a better decoding performance than codes designed by the cycle cancellation. Figure 4 also shows the lower bound on the symbol erasure rate which is given by Eq. (5). We see that Eq. (5)

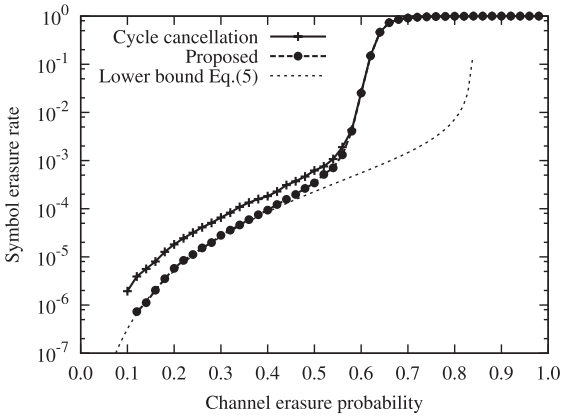

Fig. 4 Comparison of the symbol erasure rate for the expurgated ensemble ELDPC$(315,4,x,x^2,1,8,\mathcal{H}_4)$ (proposed) with that for the expurgated ensemble ELDPC$(315,4,x,x^2,1,8,\{1\})$ (cycle cancellation). The lower bound is given by Eq. (5). It can be seen that our proposed codes exhibit a better decoding performance than the cycle cancellation. It can be seen that Eq. (5) is a tight lower bound on the symbol erasure rate for the expurgated ensemble ELDPC$(315,4,x,x^2,1,8,\mathcal{H}_4)$ for small ϵ.

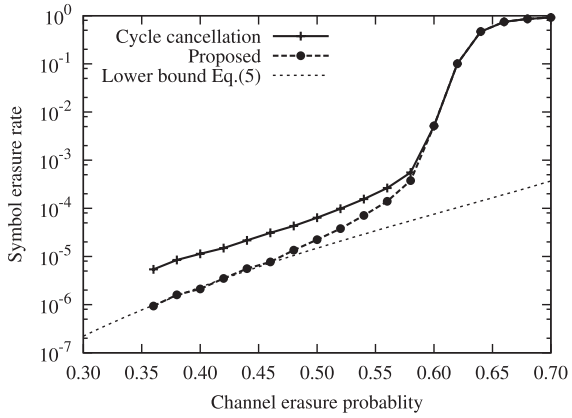

Fig. 5 Comparison of the symbol erasure rate for the expurgated ensemble ELDPC(600, 4, $x, x^2, 2, 12, \mathcal{H}_4$) (proposed) with that for the expurgated ensemble ELDPC(600, 4, $x, x^2, 2, 12, \{1\}$) (cycle cancellation). The lower bound is given by Eq. (5). This is the case for $s_g > 1$.

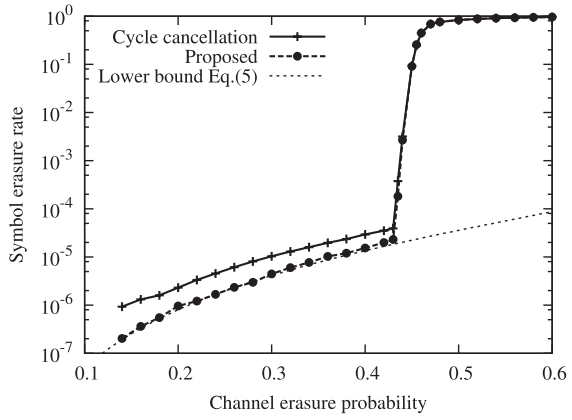

Fig. 6 Comparison of the symbol erasure rate for the expurgated ensemble ELDPC(2000, 4, $\lambda, \rho, 1, 8, \mathcal{H}_4$) (proposed) with that for the expurgated ensemble ELDPC(2000, 4, $\lambda, \rho, 1, 8, \{1\}$) (cycle cancellation), where $\lambda = 0.5x + 0.5x^2$ and $\rho = 0.5x^3 + 0.5x^5$. The lower bound is given by Eq. (5). This is the case for an irregular LDPC code ensemble case.

is a tight lower bound on the symbol erasure rate for the expurgated ensemble ELDPC(315, 4, $x, x^2, 1, 8, \mathcal{H}_4$) in the error floor.

Figure 5 compares the symbol erasure rate for the expurgated ensemble constructed by the improved cycle cancellation ELDPC(600, 4, $x, x^2, 2, 12, \mathcal{H}_4$) with that for the expurgated ensemble constructed by the cycle cancellation ELDPC(600, 4, $x, x^2, 2, 12, \{1\}$). The lower bound on the symbol erasure rate is given by Eq. (5). This is the case for $s_g \geq 2$. Figure 6 compares the symbol erasure rate for the expurgated ensemble constructed by the improved cycle cancellation ELDPC(2000, 4, $\lambda, \rho, 1, 8, \mathcal{H}_4$) with that for the expurgated ensemble constructed by the cycle cancellation ELDPC(2000, 4, $\lambda, \rho, 1, 8, \{1\}$) where $\lambda = 0.5x + 0.5x^2$ and $\rho = 0.5x^3 + 0.5x^5$. The lower bound on the symbol erasure rate is given by Eq. (5). This is the case for an irregular non-binary LDPC code ensemble. From Figs. 5 and 6, we see that Eq. (5) is a tight lower bound on the symbol erasure rate of the expurgated ensemble constructed by the improved cycle cancellation in the error floor and our proposed codes exhibit a better decoding performance than codes designed by the cycle cancellation.

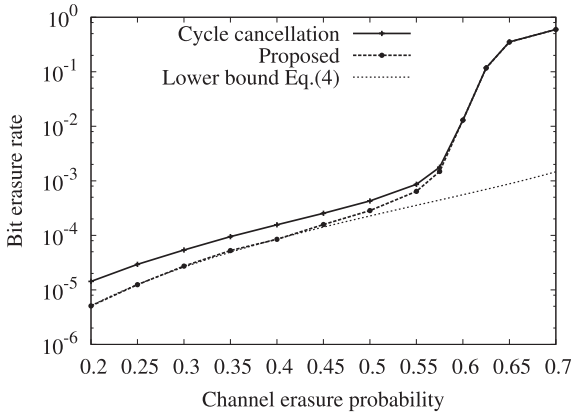

Fig. 7 Comparison of the bit erasure rate for the expurgated ensemble ELDPC(315, 4, $x, x^2, 1, 8, \mathcal{H}_4$) (proposed) with that for the expurgated ensemble ELDPC(315, 4, $x, x^2, 1, 8, \{1\}$) (cycle cancellation). The lower bound is given by Eq. (4). It can be seen that our proposed codes exhibit a better decoding performance than the cycle cancellation. It can be seen that Eq. (4) is a tight lower bound on the symbol erasure rate for the expurgated ensemble ELDPC(315, 4, $x, x^2, 1, 8, \mathcal{H}_4$) for small ϵ.

Figure 7 compares the bit erasure rate for the expurgated ensemble constructed by the improved cycle cancellation ELDPC(315, 4, $x, x^2, 1, 8, \mathcal{H}_4$) with that for the expurgated ensemble constructed by the cycle cancellation ELDPC(315, 4, $x, x^2, 1, 8, \{1\}$). It can be seen that our proposed codes exhibit a better decoding performance than codes designed by the cycle cancellation. Figure 7 also shows the lower bound on the bit erasure rate which is given by Eq. (4). We see that Eq. (4) is a tight lower bound on the bit erasure rate for the expurgated ensemble ELDPC(315, 4, $x, x^2, 1, 8, \mathcal{H}_4$) in the error floor.

4.4 Monotonicity of Error Floor

In Sect. 4.3, we see that the lower bound given by Eq. (4) is a tight lower bound on the bit erasure rate in the error floor for the expurgated ensemble constructed by the improved cycle cancellation. It is empirically known that the error floors for the non-binary LDPC codes decrease as the size of Galois field increases [8]. In this subsection, we show the monotonicity of the error floor by using the lower bound given by Eq. (4).

Let n be the bit code length, i.e., $n := mN$. From Eq. (4), we have

$$\lim_{n\to\infty} n\mathrm{P_b}(n, m, \lambda, \rho, s_\mathrm{g}, s_\mathrm{c}, \mathcal{H}_m, \epsilon) \geq \frac{m}{2}\frac{(\mu\epsilon^m)^{s_\mathrm{g}}}{1-\mu\epsilon^m} =: f(m, \epsilon, s_\mathrm{g}). \quad (6)$$

The following lemma shows that for a fixed large bit code length, the lower bound on the bit erasure rate is decreasing in m, i.e., $f(m, \epsilon, s_\mathrm{g})$ is decreasing in m.

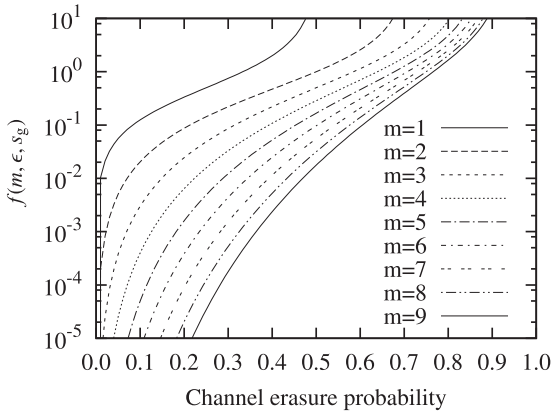

Fig. 8 Curves given by Eq. (6) for $\mu = 2$, $s_g = 1$ and $m = 1, 2, \ldots, 9$.

Lemma 1: Define $f(m, \epsilon, s_g)$ as in Eq. (6). Define ϵ^*_m as in Eq. (3). Then $f(m, \epsilon, s_g) > f(m+1, \epsilon, s_g)$ for $\mu \geq 1$ and $0 < \epsilon < \min\{\epsilon^*_m, \epsilon^*_{m+1}\} = \epsilon^*_m$.

Proof: From Eq. (6), we have

$$f(m, \epsilon, s_g) - f(m+1, \epsilon, s_g) = \frac{(\mu\epsilon^m)^{s_g} g(m, \epsilon, s_g)}{2(1-\mu\epsilon^{m+1})(1-\mu\epsilon^m)},$$

where

$$g(m, \epsilon, s_g) := m(1-\mu\epsilon^{m+1}) - (m+1)\epsilon^{s_g}(1-\mu\epsilon^m).$$

For $\epsilon < \epsilon^*_m$, $g(m, \epsilon, s_g)$ is increasing in s_g. Hence, we have $g(m, \epsilon, s_g) \geq g(m, \epsilon, 1)$. For $\epsilon < \epsilon^*_m$, $g(m, \epsilon, 1)$ is decreasing in ϵ. Note that $\min\{\epsilon^*_m, \epsilon^*_{m+1}\} < \mu^{-\frac{1}{m}}$. Thus, we see that for $\epsilon < \mu^{-\frac{1}{m}}$ and $\mu \geq 1$

$$\begin{aligned} g(m, \epsilon, s_g) \geq g(m, \epsilon, 1) &> g(m, \mu^{-\frac{1}{m}}, 1) \\ &= m(1-\mu^{-\frac{1}{m}}) > 0. \end{aligned}$$

Therefore, we have $f(m+1, \epsilon, s_g) - f(m, \epsilon, s_g) < 0$ for $\mu \geq 1$ and $0 < \epsilon < \min\{\epsilon^*_m, \epsilon^*_{m+1}\}$. □

Figure 8 shows curves given by Eq. (6) for $\mu = 2$, $s_g = 1$ and $m = 1, 2, \ldots, 9$. We see that the lower bound decreases as the order of the Galois field increases.

5. Conclusion and Future Work

In this paper, we propose a method to improve the error floors for the non-binary LDPC codes which contain the variable nodes of degree two over the BEC under BP decoding. We derive lower bounds on the bit and the symbol erasure rates in the error floors for the expurgated ensembles under BP decoding. From the simulation results, the lower bounds are tight for the bit and the symbol erasure rates for the expurgated ensembles constructed by the proposed method over the BEC under BP decoding.

We will optimize the labels in the expurgated ensembles by the method in [16]. As discuss in [10], stopping constellations for 3 imbricated cycles are not analyzed. We will clarify the stopping constellation for their cycles.

Acknowledgments

The authors are so grateful to anonymous reviewers for their valuable comments. The work of T. Nozaki was supported by Grant-in-Aid for JSPS Fellows. The work of K. Kasai was supported by the grant from the Storage Research Consortium.

References

[1] T. Nozaki, K. Kasai, and K. Sakaniwa, "Error floors of non-binary LDPC codes," Proc. 2010 IEEE Int. Symp. Inf. Theory (ISIT), pp.729–733, June 2010.
[2] R.G. Gallager, "Low Density Parity Check Codes," in Research Monograph series, MIT Press, Cambridge, 1963.
[3] T.J. Richardson, M.A. Shokrollahi, and R.L. Urbanke, "Design of capacity-approaching irregular low-density parity-check codes," IEEE Trans. Inf. Theory, vol.47, no.2, pp.619–637, Feb. 2001.
[4] M. Davey and D. MacKay, "Low-density parity check codes over GF(q)," IEEE Commun. Lett., vol.2, no.6, pp.165–167, June 1998.
[5] W. Chang and J. Cruz, "Nonbinary LDPC codes for 4-kB sectors," IEEE Trans. Magn., vol.44, no.11, pp.3781–3784, Nov. 2008.
[6] K. Kasai, T. Tsujimoto, R. Matsumoto, and K. Sakaniwa, "Rate-compatible Slepian-Wolf coding with short non-binary LDPC codes," Proc. Symp. on Inf. Theory and its Applications (SITA2009), Dec. 2009.
[7] B. Zhou, J. Kang, S. Song, S. Lin, K. Abdel-Ghaffar, and M. Xu, "Construction of non-binary quasi-cyclic LDPC codes by arrays and array dispersions," IEEE Trans. Commun., vol.57, no.6, pp.1652–1662, June 2009.
[8] X.Y. Hu, E. Eleftheriou, and D. Arnold, "Regular and irregular progressive edge-growth tanner graphs," IEEE Trans. Inf. Theory, vol.51, no.1, pp.386–398, Jan. 2005.
[9] V. Rathi, "Conditional entropy of non-binary LDPC codes over the BEC," Proc. 2008 IEEE Int. Symp. Inf. Theory (ISIT), pp.945–949, July 2008.
[10] C. Poulliat, M. Fossorier, and D. Declercq, "Design of regular ($2,d_c$)-LDPC codes over GF(q) using their binary images," IEEE Trans. Commun., vol.56, no.10, pp.1626–1635, Oct. 2008.
[11] T. Richardson and R. Urbanke, Modern Coding Theory, Cambridge University Press, March 2008.
[12] F.J. MacWilliams and N.J.A. Sloane, The Theory of Error-Correcting Codes, Elsevier, Amsterdam, 1977.
[13] V. Rathi and R. Urbanke, "Density evolution, thresholds and the stability condition for non-binary LDPC codes," IEE Proc. Commun., vol.152, no.6, pp.1069–1074, Dec. 2005.
[14] V. Rathi, Non-binary LDPC codes and EXIT like functions, Ph.D. Thesis, EPFL, Lausanne, 2008.
[15] T. Nozaki, K. Kasai, and K. Sakaniwa, "Analysis of stopping constellation distribution for irregular non-binary LDPC code ensemble," IEICE Trans. Fundamentals, vol.E94-A, no.11, pp.2153–2160, Nov. 2011.
[16] V. Savin, "Non binary LDPC codes over the binary erasure channel: Density evolution analysis," First International Symposium on Applied Sciences on Biomedical and Communication Technologies, 2008, pp.1–5, Oct. 2008.
[17] R. Lidl and H. Niederreiter, Introduction to finite fields and their applications, Cambridge University Press, New York, NY, USA, 1986.

Appendix: Proof of Theorem 1

In this section, we prove Theorem 1. To prove Theorem 1, we give several lemmas in the following sections.

A.1 Analysis of Stopping Constellation for Zigzag Cycle Codes

Consider zigzag cycle codes of weight s with labels $h_1, h_2, \ldots, h_{2s} \in \mathbb{F}_{2^m} \setminus \{0\}$ as depicted in Fig. 1. Let $E_1, \ldots, E_s \subseteq \mathbb{F}_{2^m}$ be the states of the variable nodes.

Lemma 2: For any zigzag cycles of weight s with labels $h_1, h_2, \ldots, h_{2s} \in \mathbb{F}_{2^m} \setminus \{0\}$, an assignment of states $\{E_i\}_{i=1}^s$ forms a stopping constellation if and only if for $i = 1, \ldots, s$:

$$E_i = h_{2i-1}^{-1} h_{2i} E_{i+1}, \qquad E_i = h_{2i-2}^{-1} h_{2i-3} E_{i-1},$$

where

$$E_0 := E_s, \quad E_{s+1} := E_1,$$
$$h_0 := h_{2s} \quad h_{-1} := h_{2s-1}.$$

Proof: From the definition of stopping constellation, it holds that for $i = 1, \ldots, s$

$$E_i \subseteq h_{2i-1}^{-1} h_{2i} E_{i+1}, \qquad E_i \subseteq h_{2i-2}^{-1} h_{2i-3} E_{i-1}.$$

From those equations, we have

$$E_1 \subseteq h_1^{-1} h_2 E_2 \subseteq h_1^{-1} h_2 h_3^{-1} h_4 E_3 \subseteq \cdots \subseteq \beta E_1. \qquad (\mathrm{A}\cdot 1)$$

Similarly, we have $E_1 \subseteq \beta^{-1} E_1$. Note that $E_1 \subseteq \beta^{-1} E_1$ iff $\beta E_1 \subseteq E_1$, and we have $\beta E_1 \subseteq E_1 \subseteq \beta E_1$. Thus, we have

$$E_1 = \beta E_1. \qquad (\mathrm{A}\cdot 2)$$

From Eqs. (A· 1) and (A· 2), we get $E_1 = h_1^{-1} h_2 E_2$. Similarly, we have $E_i = h_{2i-1}^{-1} h_{2i} E_{i+1}$ and $E_i = h_{2i-2}^{-1} h_{2i-3} E_{i-1}$ for $i = 1, 2, \ldots, s$. The converse is clear from the definition. □

A.2 The Condition of Successful Decoding for Zigzag Cycle Codes

From Lemma 2, for all the stopping constellations $\{E_i\}_{i=1}^s$ of zigzag cycle codes, we see that E_j for $j = 2, \ldots, s$ depends only on E_1, i.e.,

$$E_j = \prod_{i=1}^{j-1} h_{2i-1}^{-1} h_{2i} E_1$$

for $j = 2, \ldots, s$. Hence, in order to clarify the stopping constellation for zigzag cycle codes, without loss of generality, we may focus on analyzing the state E_1. From Lemma 2, we see that $E_1 = \beta E_1$. A *stopping state* for $\beta \in \mathbb{F}_{2^m} \setminus \{0\}$ is defined as a subset $E \subseteq \mathbb{F}_{2^m}$ such that

$$E = \beta E.$$

Let $\mathcal{E}_\beta$ denote the set of all the stopping states for β.

A zigzag cycle code is *recoverable* if all the symbol in the zigzag cycle code are correct by the BP decoder. From the definition, it is clear that the assignment of states such that $E_i = \mathbb{F}_{2^m}$ for $i = 1, 2, \ldots, s$ forms a stopping constellation for any zigzag cycle code of weight s. Note that $\mathbb{F}_{2^m}$ is a subset of $\mathbb{F}_{2^m}$. Thus, no zigzag cycle codes over the BEC are recoverable if all the bits are erased, i.e., $\mathbb{F}_{2^m} \in \mathcal{E}_\beta$ for all $\beta \in \mathbb{F}_{2^m} \setminus \{0\}$. More precisely, if all the bits are erased, no symbols and no bits in the zigzag cycle are correct. Similarly, the assignment of states such that $E_i = \{0\}$ for $i = 1, 2, \ldots, s$ also forms a stopping constellation for any zigzag cycle code of weight s, i.e., $\{0\} \in \mathcal{E}_\beta$ for all $\beta \in \mathbb{F}_{2^m} \setminus \{0\}$. Such a stopping constellation corresponds to the case that all the bits are correct by the BP decoder.

Hence, the zigzag cycle codes with labels $h_1, \ldots, h_{2s}$ are recoverable unless all the bits are erased if $\mathcal{E}_\beta = \{\{0\}, \mathbb{F}_{2^m}\}$. In other words, whether the zigzag cycle codes with labels $h_1, \ldots, h_{2s}$ are recoverable unless all the bits are erased, depends only on the cycle parameter $\beta = \prod_{i=1}^s h_{2i-1}^{-1} h_{2i}$.

A.3 Analysis of Stopping States

In this subsection, we clarify the condition of β such that $\mathcal{E}_\beta = \{\{0\}, \mathbb{F}_{2^m}\}$.

For $\beta \in \mathbb{F}_{2^m} \setminus \{0\}$, let $\mathcal{E}_\beta^{(\alpha^i)}$ denote the set of the stopping states containing α^i, i.e., $\alpha^i \in E$ for all $E \in \mathcal{E}_\beta^{(\alpha^i)}$. The *smallest* stopping state containing α^i for β, denoted by $E_\beta^{(\alpha^i)}$, is the stopping state for β such that $E_\beta^{(\alpha^i)} \subseteq E$ for all $E \in \mathcal{E}_\beta^{(\alpha^i)}$ and $\alpha^i \in E_\beta^{(\alpha^i)}$. It is clear $E_\beta^{(\alpha^i)}$ equals

$$\bigcap_{E \in \mathcal{E}_\beta^{(\alpha^i)}} E. \qquad (\mathrm{A}\cdot 3)$$

Since $\alpha^i \in E$ for all $E \in \mathcal{E}_\beta^{(\alpha^i)}$, we have α^i is in Eq. (A· 3). We show the closure of Eq. (A· 3) under the addition. If γ_1, γ_2 are in Eq. (A· 3), γ_1, γ_2 are in E for all $E \in \mathcal{E}_\beta^{(\alpha^i)}$. Since γ_1, γ_2 are in E for all $E \in \mathcal{E}_\beta^{(\alpha^i)}$, $\gamma_1 + \gamma_2$ is in E for all $E \in \mathcal{E}_\beta^{(\alpha^i)}$. Hence $\gamma_1 + \gamma_2$ is in Eq. (A· 3). Obviously Eq. (A· 3) is a subset of E for all $E \in \mathcal{E}_\beta^{(\alpha^i)}$. Note that

$$\beta \bigcap_{E \in \mathcal{E}_\beta^{(\alpha^i)}} E = \bigcap_{E \in \mathcal{E}_\beta^{(\alpha^i)}} \beta E = \bigcap_{E \in \mathcal{E}_\beta^{(\alpha^i)}} E.$$

Therefore, $E_\beta^{(\alpha^i)}$ is the smallest stopping state containing α^i for β.

Next, we show the uniqueness of the smallest stopping state containing α^i for β. Let E^* be another smallest stopping state for β containing α^i. The existence of a stopping state E^* contradicts the definition of Eq. (A· 3), since the intersection of E^* and Eq. (A· 3) contains α^i and is a stopping state for β.

Lemma 3: The smallest stopping state containing $\alpha^0 = 1$ for $\beta \in \mathbb{F}_{2^m} \setminus \{0\}$ is a subfield of $\mathbb{F}_{2^m}$.

Proof: For all $E \in \mathcal{E}_\beta^{(1)}$, since $1 \in E$ and $E = \beta E$, we have $\beta \in E$. Hence, we have $\beta \in E_\beta^{(1)}$. Recursively, $\beta^j \in E_\beta^{(1)}$ for $j = 0, 1, \ldots, \sigma - 1$, where σ is the order of β, i.e., σ is

the smallest positive integer such that $\beta^\sigma = 1$. Since $E_\beta^{(1)}$ is closed under the addition, we have $\sum_{j=0}^{\sigma-1} a_j\beta^j \in E_\beta^{(\alpha^i)}$, where $a_0, a_1, \ldots, a_{\sigma-1} \in \{0, 1\}$. Hence, we have

$$E_\beta^{(1)} \supseteq A := \left\{ \sum_{j=0}^{\sigma-1} a_j\beta^j \mid a_0, a_1, \ldots, a_{\sigma-1} \in \{0, 1\} \right\}.$$

Note that $A = \beta A$ and A is closed under the addition. Thus, we have $E_\beta^{(1)} = A$.

We claim that $E_\beta^{(1)}$ is a subfield of $\mathbb{F}_{2^m}$. Obviously, we have the closure of $E_\beta^{(1)}$ under addition and multiplication. The additive identity is 0 and the multiplicative identity is 1. The additive inverse for $\gamma \in E_\beta^{(1)}$ is γ. For $\gamma \in E_\beta^{(1)}$, γ^{2^m-2} is in $E_\beta^{(1)}$ since the closure of $E_\beta^{(1)}$ under multiplication. The multiplicative inverse for $\gamma \in E_\beta^{(1)} \setminus \{0\}$ is γ^{2^m-2} (Note that $\gamma \in \mathbb{F}_{2^m} \setminus \{0\}$). We are able to check that all field axioms are satisfied. Therefore, $E_\beta^{(1)}$ is a subfield of $\mathbb{F}_{2^m}$. □

Lemma 4: Define $\mathcal{H}_m$ as in Eq. (2). If $\beta \notin \mathcal{H}_m \cup \{0\}$, it holds that $E_\beta^{(1)} = \mathbb{F}_{2^m}$.

Proof: From Lemma 3, $E_\beta^{(1)}$ is a subfield of $\mathbb{F}_{2^m}$. Note that the order of proper subfield of $\mathbb{F}_{2^m}$ is 2^r [17, p.45], where r is a positive integer such that $r \mid m$ and $r \neq m$. We will prove $E_\beta^{(1)}$ is not equal to any proper subfields of order 2^r. Define $g := \frac{2^m-1}{2^r-1}$. From $\beta \notin \mathcal{H}_m \setminus \{0\}$, we have $\beta = \alpha^{ig+j}$, where $j \in \{1, 2, \ldots, g-1\}$. If β is a member of the proper subfield of order 2^r, then $\beta^{2^r} - \beta = 0$ [17, p.45]. However,

$$\beta^{2^r} - \beta = \beta(\alpha^{j(2^r-1)} - 1) \neq 0.$$

Hence, β is not a member of the proper subfield of order 2^r. Thus, we have $E_\beta^{(1)}$ is not equal to the proper subfield of order 2^r for any positive integer r such that $r \mid m$ and $r \neq m$. Therefore, we obtain $E_\beta^{(1)} = \mathbb{F}_{2^m}$. □

Lemma 5: Let $\mathcal{E}_\beta$ denote the set of stopping states for β. Define $\mathcal{H}_m$ as in Eq. (2). If $\mathcal{E}_\beta \setminus \{\{0\}, \mathbb{F}_{2^m}\} \neq \emptyset$, then $\beta \in \mathcal{H}_m$.

Proof: Let E be an element of $\mathcal{E}_\beta \setminus \{\{0\}, \mathbb{F}_{2^m}\}$. Note that $\alpha^i E \in \mathcal{E}_\beta \setminus \{\{0\}, \mathbb{F}_{2^m}\}$ for $i = 0, 1, \ldots, 2^m - 2$. If E contains α^i, then 1 is an element of $\alpha^{-i}E \in \mathcal{E}_\beta \setminus \{\{0\}, \mathbb{F}_{2^m}\}$. Hence, without loss of generality, we assume that $E \in \mathcal{E}_\beta \setminus \{\{0\}, \mathbb{F}_{2^m}\}$ and 1 is an element of E, i.e., $E \in \mathcal{E}_\beta^{(1)}$. Since $E_\beta^{(1)} \neq \mathbb{F}_{2^m}$ and $\beta \neq 0$, we have $\beta \in \mathcal{H}_m$ from Lemma 4. □

Lemma 6: Define $\mathcal{H}_m$ as in Eq. (2). If $\beta \in \mathcal{H}_m$ then $\mathcal{E}_\beta \setminus \{\{0\}, \mathbb{F}_{2^m}\} \neq \emptyset$.

Proof: If $\beta \in \mathcal{H}_m$, there exists a positive integer r such that $r \mid m$, $r \neq m$ and $\beta \in \{\alpha^{i(2^m-1)/(2^r-1)} \mid i = 0, 1, \ldots, 2^r - 2\}$. Then, a stopping state for β is written as the following:

$$E = \{0\} \cup \left\{ \alpha^{j(2^m-1)/(2^r-1)} \mid j = 0, 1, \ldots, 2^r - 2 \right\},$$

in fact $E = \beta E$ and E is a subfield of $\mathbb{F}_{2^m}$ of order 2^r. Hence, we have $E \in \mathcal{E}_\beta \setminus \{\{0\}, \mathbb{F}_{2^m}\} \neq \emptyset$. □

A.4 Proof of Theorem 1

Note that $\{\{0\}, \mathbb{F}_{2^m}\} \subseteq \mathcal{E}_\beta$ for all $\beta \in \mathbb{F}_{2^m} \setminus \{0\}$. Hence, we have $\mathcal{E}_\beta = \{\{0\}, \mathbb{F}_{2^m}\}$ iff $\mathcal{E}_\beta \setminus \{\{0\}, \mathbb{F}_{2^m}\} = \emptyset$. Define $\mathcal{H}_m$ as in Eq. (2). From Lemma 5 and 6, we have that $\beta \notin \mathcal{H}_m$ is a necessary and sufficient condition for $\mathcal{E}_\beta = \{\{0\}, \mathbb{F}_{2^m}\}$. From Sect. A.2, we see that the zigzag cycle codes with labels $h_1, h_2, \ldots, h_{2s}$ are recoverable unless all the bits are erased if $\mathcal{E}_\beta = \{\{0\}, \mathbb{F}_{2^m}\}$, where $\beta = \prod_{i=1}^{s} h_{2i-1}^{-1} h_{2i}$. Hence, we obtain that the zigzag cycle codes with labels $h_1, h_2, \ldots, h_{2s}$ are recoverable unless all the bits are erased, if $\beta \notin \mathcal{H}_m$.

Takayuki Nozaki received B.E. and M.E. degrees from Tokyo Institute of Technology in 2008 and 2010, respectively. He is currently pursuing the D.E. degree in Department of Communications and Integrated Systems at Tokyo Institute of Technology. His research interests are codes on graph and iterative decoding algorithm. He is a student member of IEEE.

Kenta Kasai received B.E., M.E. and Ph.D. degrees from Tokyo Institute of Technology in 2001, 2003 and 2006, respectively. Since April 2006, he has been an assistant professor in the Department of Communications and Integrated Systems, Graduate School of Science and Engineering, Tokyo Institute of Technology. His current research interests include codes on graphs and iterative decoding algorithms.

Kohichi Sakaniwa received B.E., M.E., and Ph.D. degrees all in electronic engineering from the Tokyo Institute of Technology, Tokyo Japan, in 1972, 1974 and 1977, respectively. He joined the Tokyo Institute of Technology in 1977 as a research associate and served as an associate professor from 1983 to 1991. Since 1991 he has been a professor in the Department of Electrical and Electronic Engineering, and since 2000 in the Department of Communication and Integrated Systems, Graduate School of Science and Engineering, both in the Tokyo Inst. of Tech. From November 1987 to July 1988, he stayed at the University of Southwestern Louisiana as a Visiting Professor. He received the Excellent Paper Award from the IEICE of Japan in 1982, 1990, 1992 and 1994. His research area includes Communication Theory, Error Correcting Coding, (Adaptive) Digital Signal Processing and so on. Dr. Sakaniwa is a member of IEEE, Information Processing Society of Japan, and Institute of Image Information and Television Engineers of Japan.