Analysis and Enumeration of Absorbing Sets for Non-Binary Graph-Based Codes

Behzad Amiri, *Student Member, IEEE,* Jörg Kliewer, *Senior Member, IEEE,* and Lara Dolecek, *Senior Member, IEEE*

***Abstract*—In this work, we first provide the definition of absorbing sets for linear channel codes over non-binary alphabets. In a graphical representation of a non-binary channel code, an absorbing set can be described by a collection of topological and edge labeling conditions. In the non-binary case, the equations relating neighboring variable and check nodes are over a non-binary field, and the edge weights are given by the non-zero elements of that non-binary field. As a consequence, it becomes more difficult for a given structure to satisfy the absorbing set constraints compared to the binary case. This observation in part explains the superior performance of non-binary codes over their binary counterparts. We show that the conditions in the non-binary absorbing set definition can be simplified in the case of non-binary elementary absorbing sets. Based on these simplified conditions, we provide design guidelines for finite-length non-binary codes free of small non-binary elementary absorbing sets. These guidelines demonstrate that even under the preserved topology, the performance of a non-binary graph-based code in the error floor region can be substantially improved by manipulating edge weights so as to avoid small absorbing sets. Our various simulation results suggest that the proposed non-binary absorbing set definition is useful for a range of code constructions and decoders. Finally, by using both insights from graph theory and combinatorial techniques, we establish the asymptotic distribution of non-binary elementary absorbing sets for regular code ensembles.**

***Index Terms*—LDPC codes, non-binary codes, absorbing sets, error floor performance.**

I. Introduction

It is well known that non-binary low-density parity-check (LDPC) codes offer performance improvements over their binary counterparts [7]. Recently, several works have been devoted to the construction of various non-binary LDPC codes. Non-binary quasi-cyclic codes [4], [17], [19], hybrid LDPC codes [33], protograph-based LDPC codes [10], cluster LDPC codes [9] and quantum LDPC codes [3] are all examples of such non-binary code constructions that offer high performance alternatives to binary LDPC codes.

In addition to code design, the development of low-complexity non-binary LDPC decoders with performance close to belief propagation decoding has also recently garnered research attention. The original implementation of the belief propagation decoder has a complexity in the order of $O(q^2)$, where q is the field size over which the non-binary code is defined. An FFT-based implementation of the message-passing decoder [8] offers a reduction of the decoding complexity down to $O(q \log q)$. The original and the FFT-based implementations of the message-passing decoder require $q-1$ values per message in the graph. In [8], [34], the authors proposed the idea of using only limited reliabilites over the propagated messages to further reduce computational complexity. Further, two low-complexity reliability-based message-passing algorithms were introduced in [6]. Moreover, the authors in [14] and [15] presented low-complexity linear programming decoding techniques for non-binary LDPC codes.

It is also well-known that under low-complexity but suboptimal message-passing algorithms, certain non-codewords are vying with codewords to be the output of the decoder. The presence of these non-codeword objects can significantly undermine the performance of iteratively decoded graph-based codes and may even result in an undesirable error floor. Given the significance of the error floor behavior for the finite block-length performance of coding schemes, extensive recent work have been devoted to studying this phenomenon for the *binary* case, such as [13], [22], and [32]. Furthermore, the design of binary codes free of these problematic objects was investigated in [24], [35], and the asymptotic enumerators of trapping/absorbing sets of different (structured and unstructured) LDPC code ensembles were studied in [1] and [21]–[23].

In contrast, for the non-binary case much less is known about how non-codeword objects and specific substructures in the Tanner graph affect the error floor performance. A key paper in this context is [30], where the authors show that small cycles in the Tanner graph of column weight two codes (i.e., the cycles that correspond to codewords with a low minimum distance) significantly degrade the code performance in the error floor region. An effective approach to improve the error floor performance of column weight two codes based on cycle manipulations is also introduced in [30].

In the case of the binary erasure channel (BEC), recent results include the introduction of the peeling decoder and the (generalized) stopping sets for non-binary LDPC codes [31]. In [28], the stopping constellation distributions for irregular non-binary LDPC code ensembles was computed. Further, zig-zag cycles and their relationship with the error floor were analyzed in [26], [30], and [27]. Ensemble enumerators

Manuscript received June 21, 2013; revised October 5 and November 27, 2013. The editor coordinating the review of this paper and approving it for publication was D. Declercq.

B. Amiri and L. Dolecek are with the Department of Electrical Engineering, University of California, Los Angeles, Los Angeles, CA 90095 USA (e-mail: amiri@ucla.edu; dolecek@ee.ucla.edu).

J. Kliewer was with the Klipsch School of Electrical and Computer Engineering, New Mexico State University, Las Cruces, NM 88003 USA. He is now with the New Jersey Institute of Technology, Newark, NJ 07103 USA (e-mail: jkliewer@njit.edu).

This paper was presented in part at the 2013 IEEE International Symposium on Information Theory [2].

Digital Object Identifier 10.1109/TCOMM.2013.122113.130465

for both stopping and trapping sets were computed in [10], [12] for protograph-based ensembles of non-binary LDPC codes. However, apart from the recent result in [29], where a simplified absorbing set definition (compared to the one in [13]) is given, no other results are known for non-binary absorbing sets to the best of our knowledge.

The main goals of this paper are multifold: 1) to define and analyze non-binary absorbing sets for codes over GF(q), $q > 2$, 2) to show that the proposed definition for non-binary absorbing set is valid over different decoders and different code constructions, 3) to highlight the difference between binary absorbing sets and non-binary absorbing sets, 4) to propose an efficient code design based on our classification of absorbing sets, and 5) to compare the distribution of non-binary absorbing sets with the one for non-binary trapping sets.

Compared to our preliminary work [2], this paper provides a more detailed discussion on the design guidelines for finite-length non-binary codes, including examples covering a broader range of code parameters, code structures, and decoders. These examples show that our proposed definition of non-binary absorbing sets appears to be universal. Further, we show that although several so-called "trapping sets"[1] (that are themselves not absorbing sets) can exist in the Tanner graph of non-binary codes, they are not part of the error profile of the codes in the error floor region. Therefore, an analysis of non-binary absorbing sets provides a more accurate estimation of the performance of non-binary codes compared to an analysis of trapping sets.

The rest of the paper is organized as follows. In Section II we summarize the well-known binary objects of interest, including the definitions of (elementary) trapping sets and (elementary) absorbing sets, along with some graph theoretic tools needed for subsequent sections. In Section III, motivated by an example, we introduce the definition of non-binary absorbing sets. Further, it is shown that the conditions in the definition of absorbing sets can be simplified in the elementary case. In Section IV we propose an algorithm to reduce the number of problematic absorbing sets in the Tanner graph of a non-binary code by carefully changing the edge weights (but not the topological structure) of the Tanner graph. The effectiveness of the proposed algorithm is shown with various examples. Section V includes the asymptotic enumeration of non-binary elementary absorbing sets in regular LDPC code ensembles. Section VI delivers the conclusions.

II. Background and Definitions

This section includes the definitions of the subgraphs known to be associated with the error floor in binary codes (both binary trapping and absorbing sets) along with some necessary graph theoretic definitions.

A. Binary trapping sets and absorbing sets

We summarize the definitions of (elementary) trapping and absorbing sets in the binary regime before proposing our new definition of non-binary absorbing sets.

[1]In this paper, "trapping sets" refer to the subgraphs as defined in [10].

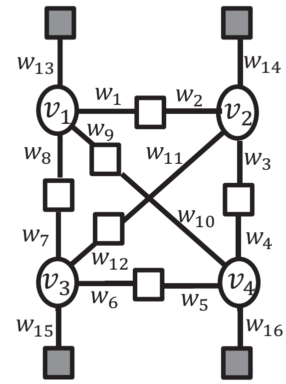


Fig. 1. Tanner graph of a $(4, 4)$ absorbing set.

Consider a subgraph of the Tanner graph of a (binary) code induced by a variable nodes, given by the node set $\mathcal{V}$. Set all variable nodes in $\mathcal{V}$ to 1, and set all other variable nodes to 0. Let $\mathcal{O}$ ($\mathcal{E}$) be the set of check nodes connected to the set $\mathcal{V}$ an odd (even) number of times. Clearly, the set $\mathcal{E}$ represents the set of satisfied check nodes, and the set $\mathcal{O}$ represents the set of unsatisfied check nodes.

Definition 1. [22] *The set $\mathcal{V}$ is an (a, b) trapping set if $|\mathcal{O}| = b$.*

Definition 2. [13] *The set $\mathcal{V}$ is an (a, b) absorbing set if $|\mathcal{O}| = b$ and if each of the a variable nodes has strictly more neighbors in $\mathcal{E}$ than in $\mathcal{O}$.*

Definition 3. [1], [22] *An elementary absorbing set (trapping set) is an absorbing set (trapping set) with each of its neighboring satisfied checks having two edges connected to the absorbing set (trapping set), and each of its neighboring unsatisfied checks having exactly one edge connected to the absorbing set (trapping set).*

Figure 1 shows a binary $(4, 4)$ elementary absorbing set. We tacitly assume that the edge weights (edge labels) are $w_i = 1$, $i = \{1, 2, \ldots, 16\}$.

B. Cycle space, cycle span, and fundamental cycles

In graph theory, a vector space is defined on the set of all cycles of an undirected graph. Let G be a finite undirected graph with edge set E. The power set of E forms a vector space by taking the symmetric difference as addition, identity function as negation, and empty set as zero. The **cycle space** of G is defined as a subspace of the edge space which includes all the cycles in the graph G. The following offers the definition of the cycle span and of fundamental cycles.

Definition 4. *A set of cycles in G denoted by $\mathcal{F}$ is called the **cycle span** of G if it forms a basis for the cycle space. In other words, $\mathcal{F}$ is a cycle span if all the cycles of G can be constructed as a combination of the cycles in $\mathcal{F}$, and this set has the fewest cycles. The cycles in the cycle span are called **fundamental cycles**.*

In the case of an elementary absorbing set where the corresponding subgraph only has degree-one and degree-two check nodes, the variable node graph is defined as follows:

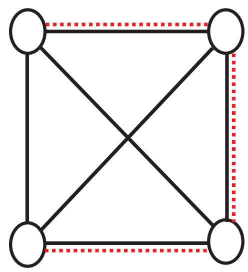


Fig. 2. Variable node (VN) graph corresponding to the Tanner graph of a $(4, 4)$ absorbing set.

Definition 5. [35] *For a bipartite graph $\mathcal{G}$ corresponding to an elementary absorbing set, a variable node (VN) graph is constructed by defining variable nodes of $\mathcal{G}$ as its vertices and degree-two check nodes connecting the variable nodes in $\mathcal{G}$ as its edges.*

Figure 2 shows the VN graph for the $(4, 4)$ absorbing set displayed in Figure 1. The edges in Figure 2 marked in dashed red form a spanning tree of the VN graph. Note that a spanning tree is not unique. From graph theory, it is also known that for an undirected connected graph with a vertices, the spanning tree has $a - 1$ edges, and that adding an edge to the spanning tree creates a fundamental cycle. Therefore, there is a one-to-one correspondence between the fundamental cycles and the set of the remaining edges not in the spanning tree. As a result, the corresponding VN graph of $\mathcal{G}$ with a vertices (variable nodes) and e edges (degree-two checks) has $e - a + 1$ fundamental cycles. As an example, the VN graph corresponding to the $(4, 4)$ absorbing set shown in Figure 2, including $a = 4$ vertices and $e = 6$ edges, has $e - a + 1 = 3$ fundamental cycles. This can easily be observed from Figure 2 (the number of edges not present in the spanning tree is 3).

III. NON-BINARY ABSORBING SETS

In contrast to binary codes, each edge in the Tanner graph of a non-binary code admits a weight taken as a non-zero element of the underlying non-binary field. Consequently, the edges of the subgraphs which correspond to the fixed points of non-binary LDPC decoders need not only be topologically connected in specific ways, but also the labels on these edges must satisfy certain conditions. In other words, suppose we consider a topology (with no edge weight assignment) that satisfies the conditions of the (binary) absorbing set, as given by Definition 2. Based on the choice of the edge labels from the Galois field GF(q), the topology may or may not cause a decoding failure.

Example 1 illustrates the difference between binary and non-binary absorbing sets and gives a motivation for the definition of non-binary absorbing sets.

Example 1. *Consider the graphical structure in Figure 1, defined over GF(q), $q = 2^p$. If there exists a set of non-zero inputs for all variable nodes that makes all degree-two check nodes satisfied, the resulting configuration will have 4 unsatisfied checks and each variable node will have strictly more satisfied than unsatisfied neighboring checks (3 vs. 1). Mathematically, the inputs v_1, v_2, v_3, v_4 and weights of the edges $w_1, \ldots, w_{12}$ then satisfy:*

$$\begin{aligned} v_1w_1 &= v_2w_2 \quad \text{over GF}(q), & v_2w_3 &= v_4w_4 \quad \text{over GF}(q),\\ v_4w_5 &= v_3w_6 \quad \text{over GF}(q), & v_3w_7 &= v_1w_8 \quad \text{over GF}(q),\\ v_2w_{11} &= v_3w_{12} \quad \text{over GF}(q), & v_1w_9 &= v_4w_{10} \quad \text{over GF}(q), \end{aligned}$$

which leads to the following conditions:

$$\begin{aligned} w_1w_7w_{11} &= w_2w_8w_{12} \quad \text{over GF}(q),\\ w_3w_5w_{12} &= w_4w_6w_{11} \quad \text{over GF}(q),\\ w_2w_4w_9 &= w_1w_3w_{10} \quad \text{over GF}(q), \end{aligned} \qquad (1)$$

where all w and v are non-zero elements of GF(q).

For example, for $q = 8$, Figure 3(a) shows a choice of weights satisfying the conditions in (1). *With these weights, there are $q - 1$ choices (out of $(q - 1)^4$) for the set (v_1, v_2, v_3, v_4) such that each variable node has exactly 3 satisfied and 1 unsatisfied checks. One example is $(1, \alpha, \alpha^2, 1)$, where α is a primitive element of GF(8) based on the primitive polynomial $p(x) = x^3 + x + 1$. In contrast, the weights in Figure 3(b) do not satisfy the conditions in* (1) *and result in a configuration that has 4 variable nodes and z unsatisfied checks. Here, $5 \leq z \leq 10$, and the value of z depends on the input values v_1 through v_4. For example, the same input $(1, \alpha, \alpha^2, 1)$ results in $z = 10$. Clearly, this configuration with all 10 neighboring checks being unsatisfied is not expected to be problematic for message-passing decoding.*

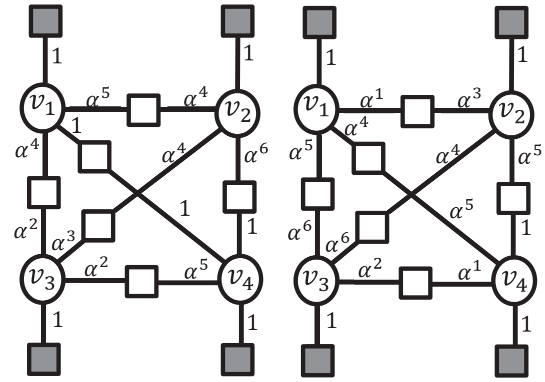


Fig. 3. (a) Non-binary elementary $(4, 4)$ absorbing set over GF(8), (b) $(4, z)$ trapping set, $5 \leq z \leq 10$, weights do not satisfy the absorbing set conditions over GF(8). Here, α is a primitive element of GF(8) based on the primitive polynomial $p(x) = x^3 + x + 1$.

Example 1 above provides a motivation for studying non-binary absorbing sets. Consider a non-binary LDPC code with an $m \times n$ parity check matrix H defined over GF(q), $q = 2^p$. The corresponding Tanner graph has n variable nodes and m check nodes. Definition 6 below states the conditions for a subset of variable nodes to be an (a, b) non-binary absorbing set. We assume transmission of the all-zero codeword. We also assume that these a decoding errors only occur in variable nodes included in this absorbing set, and that values of the variable nodes outside of the absorbing set are all 0 ($\in$ GF(q)).

Consider a subset $\mathcal{V}$ of variable nodes with $|\mathcal{V}| = a$. We form the $\ell \times a$ matrix A, a submatrix of matrix H, consisting of the columns of matrix H that correspond to variable nodes in $\mathcal{V}$ and ℓ check nodes connected to $\mathcal{V}$.

Definition 6. *The configuration $\mathcal{V}$ is an (a, b) absorbing set over GF(q) if there exists an $(\ell - b) \times a$ submatrix B of rank r_B, with elements $b_{j,i}, 1 \leq j \leq \ell - b, 1 \leq i \leq a$, in matrix A that satisfies the following conditions.*

1) *Let $N(B)$ be the null-space of matrix B and let $\boldsymbol{d}_i, 1 \leq i \leq b$ be the ith row of matrix D, where D is formed by excluding the matrix B from A. Then,*

$$\exists \boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_a \end{bmatrix} \in N(B) \text{ s.t. } x_i \neq 0 \text{ for } \forall i \in \{1, \dots, a\}$$

$$\text{and } \nexists i, \boldsymbol{d}_i \boldsymbol{x} = 0. \tag{2}$$

2) *Let $d_{j,i}$, $1 \leq j \leq b, 1 \leq i \leq a$, be the elements of the matrix D. Then,*

$$\forall i \in \{1, 2, \dots, a\} : \left(\sum_{j=1}^{\ell - b} S(b_{j,i}) \right) > \left(\sum_{j=1}^{b} S(d_{j,i}) \right), \tag{3}$$

where the function S is

$$S(x) = \begin{cases} 1 \text{ when } x \neq 0, \\ 0 \text{ when } x = 0. \end{cases} \tag{4}$$

Condition 1 in Definition 6 requires that there exists a vector in the null-space of the matrix B whose elements are all non-zero. Therefore, there exists a solution to $Bx = 0$ over GF(q) such that all components of the solution are non-zero. A consequence of this condition is that $r_B < a$. Also, condition 1 guarantees that for vector $\boldsymbol{x}$ in the null-space of matrix B, none of the check nodes associated with the rows of the matrix D are satisfied (otherwise, input $\boldsymbol{x}$ results in an $(a, \tilde{b})$, $\tilde{b} < b$, absorbing set). Condition 2 ensures that for each variable node, the number of connected satisfied checks is larger than the number of connected unsatisfied checks.

Observe that the proposed definition is in agreement with the existing definition of a binary absorbing set [13]. In particular, for $q = 2$, the condition 1 is automatically satisfied since the input vector for the variable nodes must be an all-ones vector to satisfy the checks. Also, the condition that each variable node in the absorbing set has strictly more neighbors in $\mathcal{E}$ than in $\mathcal{O}$ corresponds to condition 2.

Remark 1. *Our definition of a non-binary absorbing set is different from the definition proposed in [29] where the authors define a non-binary (primitive) absorbing set as an object which has more satisfied checks than unsatisfied checks, taken collectively over all variable nodes in this object. In contrast, our definition, similar to the original definition of binary absorbing sets [13], requires* each *variable node to be connected to more satisfied checks than unsatisfied checks.*

We also recall the definition of non-binary trapping sets from [10]:

Definition 7. *A subset $\mathcal{V}$ of variable nodes ($|\mathcal{V}| = a$) with given input $(x_1, x_2, \dots, v_a)$ is an (a, b) trapping set over GF(q) if there exist exactly b unsatisfied check nodes in its induced subgraph (with respect to the given input set).*

The above definition is a natural extension of binary trapping set definition [22]. Later, through our simulation results, we will show that non-binary absorbing sets provide a better assessment of the structures which cause decoding errors in the error floor region compared to non-binary trapping sets.

Remark 2. *Note that whether a non-binary absorbing set as defined in Definition 6 results in a decoding error depends on the choice of the input values for the variable nodes. The set of all $\boldsymbol{x}$'s which satisfy* (2) *is the set of all input values for the variable nodes that result in an (a, b) non-binary absorbing set. Other choices of variable nodes inputs can result in other $(a, \tilde{b})$, $\tilde{b} \neq b$ absorbing/trapping sets.*

As in the binary case, we say that a non-binary absorbing set $\mathcal{V}$ is **elementary** if all neighboring satisfied checks have degree 2 and all neighboring unsatisfied checks have degree 1 with respect to $\mathcal{V}$. It can be easily observed that a non-binary elementary absorbing set is necessarily a binary elementary absorbing set where all non-zero edge labels and non-zero variable node values are converted to 1 and all operations are taken modulo 2. This observation can be used in searching for non-binary elementary absorbing sets in the Tanner graph of non-binary codes. Clearly, the converse is not true as the choice of non-binary labels may violate the absorbing set constraints.

In the case of elementary absorbing sets, the absorbing set conditions can be simplified, as the following lemma shows. Consider a code $\mathcal{C}$ with a parity check matrix H over GF(q). Let $\mathcal{G}$ be its Tanner graph. Let C_p be an arbitrary cycle involving p distinct variable nodes and p distinct neighboring check nodes in the graph induced by an (a, b) non-binary elementary absorbing set in $\mathcal{G}$, $p \leq a$. We write C_p as the oriented traversal $c_1 - v_1 - c_2 - v_2 - \cdots - c_p - v_p - c_1$, where v and c denote the spanned variable and check nodes, respectively. Let w_{2i-1} denote the label on the $c_i - v_i$ edge, and let w_{2i} denote the label on the $v_i - c_{i+1}$ edge. The following lemma presents a necessary condition for a subgraph of the Tanner graph of the code $\mathcal{C}$ to be a non-binary elementary absorbing set.

Lemma 1. *In the case of elementary absorbing sets over GF(q), $q = 2^p$, for every cycle C_p (as introduced above), the weights of the edges w_i, $i \in \{1, 2, \dots, 2p\}$, satisfy the following relation:*

$$\prod_{k=1}^{p} w_{2k-1} = \prod_{k=1}^{p} w_{2k} \quad \text{over GF}(q). \tag{5}$$

Proof: For the cycle C_p (of length $2p$) we form a $p \times p$ submatrix B_{C_p} of H, corresponding to the p variable nodes and p check nodes in C_p. Since the check nodes in B_{C_p} are satisfied, there exists a non-zero solution $\boldsymbol{x}$ to $B_{C_p}\boldsymbol{x} = 0$ over GF(q). Therefore, for the square matrix B_{C_p}, we have $\det(B_{C_p}) = 0$ over GF(q). Thus,

$$\det(B_{C_p}) = \det \begin{bmatrix} w_1 & w_2 & 0 & \dots & 0 \\ 0 & w_3 & w_4 & \dots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \dots & w_{2p-3} & w_{2p-2} \\ w_{2p} & 0 & \dots & 0 & w_{2p-1} \end{bmatrix} = 0 \text{ over GF}(q). \tag{6}$$

Now, in order to satisfy (6), the condition in (5) must hold. Since every permutation of the columns of B_{C_p} results in the same determinant, the ordering in (6) can be performed without loss of generality. ■

It can be shown that if all the fundamental cycles satisfy (5), this equation also holds for all other cycles in the graph. Thus, a non-binary elementary absorbing set not only satisfies the topological conditions, i.e., the unlabeled subgraph is an elementary *binary* absorbing set, but also all of its fundamental cycles satisfy (5). Therefore, in order to verify whether a given topology (which meets the topological conditions of a binary absorbing set) is a non-binary absorbing set or not, we need to only check if the edge weights in each of the fundamental cycles satisfy (5).

Remark 3. *A condition similar to Lemma 1 is presented in [30] but only for regular codes of column weight* 2*. In the case of column weight* 2*, the smallest absorbing sets are* $(a, 0)$ *absorbing sets which correspond to the minimum weight codewords, and moreover contain only one (fundamental) cycle. The results presented in this paper apply to general column weights wherein absorbing sets may be spanned by more than one fundamental cycle.*

Example 2. *Let us interpret the configuration in Figure 1 as a* $(4, 4)$ *non-binary absorbing set. The conditions on the fundamental cycles (*3 *fundamental cycles) in* (5) *are precisely the ones given in* (1)*. If, however,* (5) *is violated for one of the three fundamental cycles, the resulting configuration will no longer be an absorbing set; it will instead become a* $(4, z), 5 \leq z \leq 10$*, trapping set (where the value of* z *depends on the input and on the number of unsatisfied checks).*

Lemma 2. *Based on the non-binary edge weights and the variable node inputs, an unlabeled* (a, b) *(i.e., binary) elementary absorbing set becomes an* (a, b^+) *absorbing set/trapping set in the non-binary case (here and in the remainder of the paper,* b^+ *denotes any integer greater than or equal to* b*).*

Proof: Based on Definition 3, unlabeled (binary) elementary absorbing sets include only degree-two and degree-one check nodes. After the assignment of non-binary edge weights, degree-one check nodes always remain unsatisfied since the multiplication of two non-zero Galois field elements is always non-zero. Based on the inputs of the variable nodes and the edge weights, degree-two check nodes may remain satisfied or may become unsatisfied. Therefore, after the assignment of non-binary edge weights, the number of unsatisfied checks either stays the same or increases. ■

Remark 4. *In Definition 6, transmission of the all-zeros codeword is assumed without loss of generality. By transmitting any other codeword, the addition of the same error vector* $\boldsymbol{x}$ *(defined in Definition 6, Condition* 1*) results in the same absorbing set error because of the linearity of the code.*

IV. Non-Binary Absorbing Sets as a Tool to Improve the Performance

In this section, equipped with techniques from graph theory, we find the ratio of all possible edge weight assignments which convert an (a, b) unlabeled elementary absorbing set to an (a, b) non-binary elementary absorbing set or to an $(a, b + 1)$ non-binary trapping set in GF(q). Further, based on the observation that proper choices of non-binary edge weight assignments result in non-problematic trapping sets, we propose an algorithm to improve the performance of non-binary codes in the error floor region.

Theorem 1. *Given an* (a, b) *unlabeled (i.e., binary) elementary absorbing set with* e *satisfied checks:*

1) *A fraction of* $(q-1)^{a-e-1}$ *out of all possible edge weight assignments taken from GF(*q*) results in* (a, b) *non-binary elementary absorbing sets.*
2) *A fraction of* $e \cdot (q-1)^{a-e-1} \cdot (q-2)$ *out of all possible edge weight assignments taken from GF(q) lead to structures resulting in* $(a, b+1)$ *non-binary trapping sets under appropriate variable node input values.*

Proof: The VN graph of the given unlabeled elementary absorbing set includes a vertices and e edges. Each edge in the spanning tree of the VN graph corresponds to two edges in the Tanner graph.

For part 1, the labels of the edges in the Tanner graph that are represented in the spanning tree can be chosen arbitrarily. Therefore, we have $(q-1)^2$ choices for the weights for each edge. Thus, there are $(q-1)^{2(a-1)}$ weight assignments for the $a-1$ edges in the spanning tree.

For each of the remaining $e-(a-1)$ edges in the VN graph that are not in the spanning tree, one of the two edges in the Tanner graph can again be chosen arbitrarily but the other edge is uniquely determined according to (5). Thus, there are $(q-1)^{e-a+1}$ weight assignments for edges not in the spanning tree. Hence, the total number of the weight assignments resulting in non-binary elementary absorbing sets is given as

$$(q-1)^{2(a-1)} \cdot (q-1)^{e-a+1} = (q-1)^{a+e-1}.$$

Since we have a total of $(q-1)^{2e}$ possible weight assignments, the resulting fraction follows.

For part 2, in addition to b degree-one unsatisfied checks, one of the degree-two check nodes is also unsatisfied. Therefore, (5) should be satisfied for all fundamental cycles except the one which includes the additional unsatisfied check node. We consider a choice of a spanning tree which does not include the edge representing the degree-two unsatisfied check node in the VN graph. Similar to part 1, the edge weights in the Tanner graph that are included in the spanning tree can be chosen arbitrarily. Thus, there are $(q-1)^{2(a-1)}$ weight assignments for the $a-1$ edges in the spanning tree.

For each of the $e-a$ edges in the VN graph which represent satisfied checks outside of the spanning tree, one of the two edges in the Tanner graph can again be chosen arbitrarily but the other edge is uniquely determined according to (5). Thus, there are $(q-1)^{e-a}$ weight assignments for edges which represent satisfied checks absent from in the spanning tree of the VN graph. For the remaining edge in the spanning tree (which corresponds to the degree-two unsatisfied check node), one of the two edges in the Tanner graph can be chosen arbitrarily ($q-1$ choices). The other edge in the Tanner graph should be chosen from $q-2$ weights, which results in a violation of (5) for the corresponding fundamental cycle. Hence, the total number of the weight assignments resulting in $(a, b+1)$ non-binary trapping sets is given as

$$\binom{e}{1}(q-1)^{2(a-1)} \cdot (q-1)^{e-a} \cdot (q-1) \cdot (q-2)$$
$$= e(q-1)^{a+e-1}(q-2).$$

Note that the multiplication by $\binom{e}{1}$ is due to the choice of one degree-two check node (from a total of e degree-two check nodes) to be unsatisfied. Since we have a total of $(q-1)^{2e}$ possible weight assignments, the resulting fraction follows. ■

Remark 5. *In Theorem 1, part 2, we enumerate all candidate subgraphs that result in $(a, b+1)$ trapping sets under appropriate variable node input values. Under other inputs, these subgraphs lead to $(a, (b+2)^+)$ trapping sets which are less problematic for the decoder due to their additional unsatisfied check node(s).*

From Theorem 1, we can conclude that compared to the number of (a, b) non-binary absorbing sets, a factor of $e(q-2)$ more $(a, b+1)$ non-binary trapping sets exists in the Tanner graph of a randomly generated code. However, our simulation results will show that the error profiles of these codes do not include any $(a, b+1)$ non-binary trapping set errors. The intuition behind this behavior can be explained as follows. In practical message passing decoders, messages and beliefs are quantized. As a result, these values are bounded between a maximum value and a minimum value which are determined by the choice of the quantization parameters. In the case of non-binary absorbing sets, where the majority of the connected check nodes to each variable node are satisfied, the incorrect beliefs at the variable nodes are reinforced after each iteration. These reinforced variable node beliefs (and messages) reach the maximum or minimum allowed values after a few iterations. Here, an incorrect value in an absorbing set cannot be corrected since the messages associated with the minority unsatisfied check nodes cannot override the incorrect messages from the majority satisfied check nodes. In other words, after saturation of messages and beliefs, a quantized message passing decoder performs roughly the same as a decoder which updates its variable node beliefs based on the majority rule (e.g., bit flipping decoding in the binary case). As a result, a quantized message passing decoder results in absorbing set errors. On the other hand, in a non-binary trapping set not being a non-binary absorbing set, the majority of the check nodes connected to each variable node are not necessarily satisfied. Therefore, under bit flipping type decoding, there is a sufficient number of unsatisfied checks to overturn the satisfied checks (that are re-enforcing the incorrect variable node values). Thus, our proposed absorbing set definition provides a better description of the problematic subgraphs for a message passing decoder in the error floor region compared to the definition of non-binary trapping sets.

In the error floor region (high SNR region), errors typically include a small number of variable nodes. Thus, in the error floor region, small absorbing sets dominate the errors of message passing decoders, and the performance of the code is determined by the absorbing set of the smallest size. Therefore, it is critical to increase the minimum absorbing set size in the code design to have a superior performance in the error floor region.

Algorithm 1 Reduction of the number of absorbing sets in the Tanner graph of a non-binary code.

1: **Input:** Tanner graph G with edge weights over GF(q).
2: Choose W, the set of non-binary absorbing sets to be eliminated.
3: Let X be the set of non-binary absorbing sets which can not be eliminated.
4: Let $X = \varnothing$.
5: Let A be the set of non-binary absorbing sets in W which have been processed by our algorithm (either successfully eliminated from G or included in X).
6: Let $A = \varnothing$.
7: For every edge $j \in T$, C_j is the set of reweighted cycles which include j.
8: $\forall j \in T$, $C_j = \varnothing$.
9: Find (a_j, b_j), the smallest non-binary absorbing set in $W \backslash A$.
10: If this absorbing set is a child of another absorbing set in A, go to 31.
11: Find U_j, the set of all (a_j, b_j) absorbing sets in the unlabeled version of the Tanner graph G (e.g., using techniques in [20]).
12: **for** $\forall u \in U_j$ **do**
13:     Find F_u, a set of fundamental cycles of u.
14:     Let E_u be the set of all edges in u.
15:     For an edge $k \in E_u$, let M_k be the set of cycles in F_u which include k.
16:     **if** (5) is satisfied for all cycles in F_u **then**
17:         Find edge $i \in E_u$ with edge weight w_i such that $|C_i|$ is minimum.
18:         **if** $\exists w'_i \neq w_i$ and $w'_i \neq 0$ such that all cycles including i do not satisfy (5) **then**
19:             Replace w_i with w'_i.
20:         **else**
21:             $E_u \leftarrow E_u \backslash i$.
22:             **if** $E_u = \varnothing$ **then**
23:                 $X \leftarrow X \cup U_j$ and go to 31.
24:             **else**
25:                 Go to 17.
26:             **end if**
27:         **end if**
28:         For every edge e in the cycles of M_i, update $C_e \leftarrow C_e \cup M_e$.
29:     **end if**
30:     **end for**
31:     Add (a, b) absorbing set to the set A.
32:     If $A \neq W$, go to 9.
33:     If $X = \varnothing$, all absorbing sets of interest are eliminated. Otherwise, it is not possible to eliminate absorbing sets in X.

In the remainder of this section, we will introduce a method to eliminate problematic non-binary absorbing sets from Tanner graphs of non-binary codes. Our simulation results will show the effectiveness of our algorithm in improving the code performance in error floor region.

We exploit the fact that non-binary edge weights enable us to reduce the number of absorbing sets by just changing the weights of edges in the Tanner graph without changing its structure. The method is stated in Algorithm 1. We say that an absorbing set A_y is defined as a child of an absorbing set A_z if A_z is a subgraph of A_y. The main steps of the method can be summarized as follows:

- Step 1: For the given Tanner graph, we first choose a set W of pairwise parameters (i.e., $W = \{(a_1, b_1), (a_2, b_2), \ldots, (a_k, b_k)\}$) corresponding to the k elementary absorbing sets which we want to eliminate. The choice of the set W depends on the code parameters (input Tanner graph), such as column weight, girth and structure [11], [35]. For example, for a Tanner graph with girth $g \geq 8$, the $(4, 4)$ absorbing set graph shown in Figure 1 does not exist since this absorbing set contains length 6 cycles. In general, $(4, 4)$ absorbing sets exist only if the column weight $c = 4$ and the girth is $g = 6$ [11].
- Step 2: We find the absorbing set (a_j, b_j) with the smallest a and b parameters in W. If the (a_j, b_j) absorbing set is a child of a previously eliminated absorbing set, we remove (a_j, b_j) from W and repeat this step for the next smallest parameters[2].
- Step 3: We find the set U_j of all binary (a_j, b_j) absorbing sets in the unlabeled Tanner graph. For each absorbing set in U_j, if all fundamental cycles satisfy (5), we change the weight of an edge to another non-zero element of GF(q). The edge and its new weight are chosen such that the previously canceled non-binary absorbing sets remain canceled. This process continues until all (a_j, b_j) absorbing sets are eliminated. Then, we remove (a_j, b_j) from W.
- Step 4: If $W \neq \varnothing$, we go to Step 2. Otherwise, the algorithm terminates.

Remark 6. *The work in [30] focuses on the case of column weight two and presents an approach to cancel all cycles of length l, $g \leq l \leq l_{max}$, where g is the girth, for this column weight choice. As stated in [30], it is impossible to cancel all cycles for all lengths l. In contrast, for codes with column weights $c \geq 2$, our approach only seeks to reweigh a selected number of cycles, i.e., one fundamental cycle per absorbing set of interest. As a result, our approach allows for further flexibility in reweighting cycles of various lengths for column weights $c \geq 2$.*

Figure 4 shows the simulation results[3] for random regular codes (denoted by 'Original') over various choices of the field order size GF(q), with block length $N \approx 2750$ bits, rate $R \approx 0.88$, column weight $c = 4$, row weight $r = 51$ for GF(2), $r = 37$ for GF(4), $r = 31$ for GF(8), and $r = 26$ for GF(16), and girth $g = 6$, transmitted over a binary-input additive white Gaussian noise (AWGN) channel, where the frame error rate (FER) versus the signal-to-noise ratio (SNR) is displayed[4]. The figure also shows the results obtained by using a modified version of the approach presented in [30] (denoted by 'P-method') which tries to cancel all the cycles of length l, $g \leq l \leq l_{max}$. The approach in [30], which was introduced for the case of $c = 2$, can easily and successfully be extended to the case of $c > 2$, as demonstrated in this paper. The figure also shows the results obtained by the code modification proposed in [25] (denoted by 'N-method') and our code modification specified in Algorithm 1 (denoted by 'A-method'). The simulations results reconfirm the superior performance of the codes modified by N-method, compared to the codes modified by P-method, as shown previously in [25]. All codes are decoded by using a Fast Fourier Transform-based q-ary SPA (FFT-QSPA) decoder [8].

TABLE I
ERROR PROFILE, $SNR = 5.2$ dB, $N = 2738$, $R = 0.891$, $c = 4$ AND GF(4), TOTAL NUMBER OF SIMULATIONS $\approx 4 \times 10^8$.

Error Type	$(4,4)$	$(5,0)$	$(5,2)$	$(6,2)$	$(6,4)$	$(6,6)$	$(7,4)$	$(8,2)$	other
Original	32	9	14	7	9	18	9	9	15
P-method	0	0	0	0	0	12	5	5	12
A-method	0	0	0	0	0	0	0	0	13

All P-method, N-method and A-method approaches enjoy the improved performance relative to the random code construction. The performance comparison for different values of q reveals that the improvement is more pronounced for smaller values of q. As we will show below in Section V, Corollary 1, for a random code construction, under higher field sizes, there are fewer non-binary elementary absorbing sets available to be canceled by using the P-method, N-method, or the A-method.

Table I includes the error profiles[5] for original, P-method and A-method codes over GF(4). Both the P-method and the A-method eliminate all $(4, 4)$ absorbing sets as well as $(5, 0)$, $(5, 2)$, $(6, 2)$ and $(6, 4)$ absorbing sets that are children of $(4, 4)$ absorbing sets. Additionally, the A-method successfully eliminates all $(6, 6)$, $(7, 4)$ and $(8, 2)$ absorbing sets, since in this approach we selectively reweigh some (but not necessarily all) length-6 cycles followed by a reweighting of some of length-8 cycles. As q increases, there are more degrees of freedom available to change the edge weights to reweigh cycles and as a result, the implementation of the cycle-elimination-only approach in [30] appears to be sufficient for larger values of q. The errors labeled as 'other' in Table I are non-converging errors. This type of error happens when the decoder does not converge to a specific object before reaching its maximum number of iterations. In other words, we declare "convergence" if the decision of the decoder does not change over 5 final iterations of decoding (out of 50 total iterations). If the decoder decision changes, the error is categorized as a non-converging error. This type of error is usually an oscillation between different trapping set errors.

[2]Note that size (a_j, b_j) absorbing sets might have different topologies. In this case, each topology of (a_j, b_j) absorbing set is considered independently of other topologies.

[3]All the simulation results presented in this paper were performed over a six month period on the Hoffman2 Cluster which is a part of High Performance Computing Resources at UCLA. The Hoffman2 cluster has more than 800 machines and about 7000 cores. The CPUs have 8, 12 or 16 cores with speed of $2.2 - 3.0$ GHz. Each core has 1GB, 4GB or 8GB of memory. The Hoffman2 cluster uses the CentOS Linux 6.2 Operating System.

[4]The exact code parameters for the simulated codes in this section can be found in Appendix B.

[5]The error profiles are calculated as follows. After the decoder reaches its maximum number of iterations, for each channel realization which results in a decoding error, we form the induced subgraph corresponding to the variable nodes in error. We then determine whether the induced subgraph is an absorbing set or not. If yes, the size of the absorbing set is also calculated. In a small minority of cases, the induced subgraph includes two or more separate subgraphs which are independently investigated.

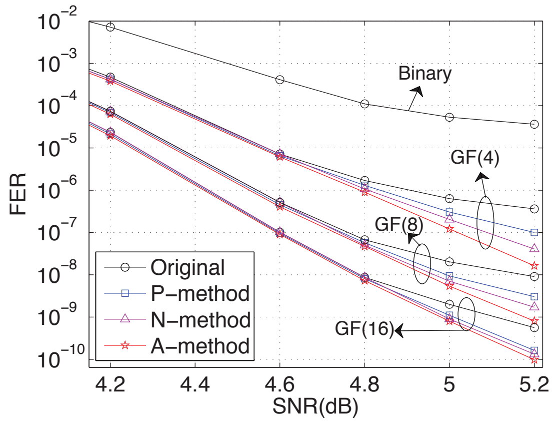


Fig. 4. FER versus SNR for the original random non-binary code and for both P- and A-method modified codes, $N \approx 2750$, $R \approx 0.88$, $c = 4$, and the QSPA-FFT decoder.

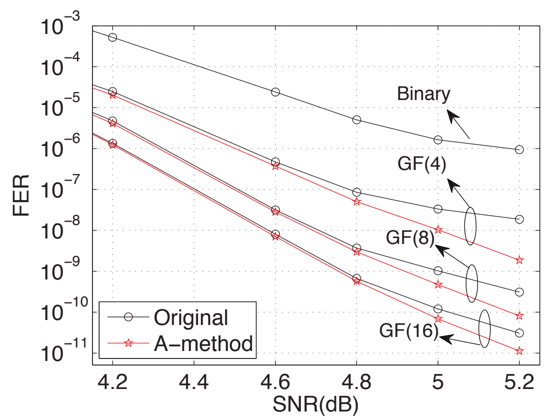


Fig. 5. FER versus SNR for the original random non-binary code and the A-method modified code, $N \approx 2350$, $R \approx 0.83$, $c = 5$, and the QSPA-FFT decoder.

Note that by performing our algorithm, (a, b) absorbing sets in the error profile of the original code are converted to structures which can turn into (a, b^+) trapping sets under appropriate variable node input values. As the error profile in Table I shows, these new structures are not problematic for the decoder. Therefore, non-binary absorbing sets provide a better definition for problematic structures in the error floor region compared to trapping sets.

The simulation results for random regular codes ('Original') and the modified codes using the A-method over GF(q), $q = 2, 4, 8, 16$ with block length $N \approx 2350$ bits, rate $R \approx 0.83$, column weight $c = 5$, and row weight $r = 50$ for GF(2), $r = 33$ for GF(4), $r = 28$ for GF(8), $r = 24$ for GF(16) are presented in Figure 5. This figure again confirms the effectiveness of our algorithm in improving the performance of non-binary codes in the error floor region. Similar improvements in performance were also observed by using the P-method, although those results have been omitted from Figure 5 for brevity. These simulations again show that the gap between the curves for the P-method and A-method decreases for larger field sizes, and that the improvement in the performance of the codes decreases as the Galois field size increases.

Figure 6 presents the simulation results for random regular codes ('Original') and modified codes using the P-method and the A-method over GF(q), $q = 2, 4, 8, 16$ with block length

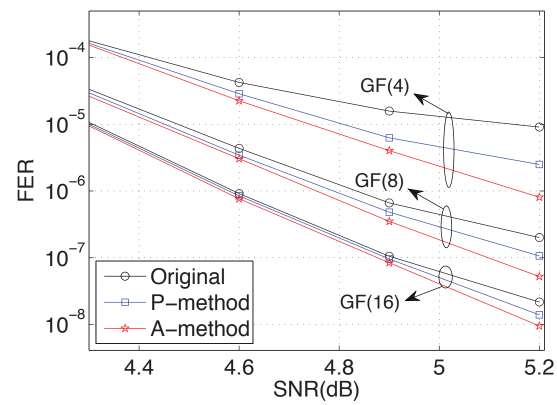


Fig. 6. FER versus SNR for the original non-binary random code and both P- and A-method modified codes, $N \approx 2750$, $R \approx 0.88$, $c = 4$, and a min-sum decoder.

TABLE II
ERROR PROFILE, $E_b/N_0 = 5.2$ dB, $N = 2738$, $R = 0.891$, $c = 4$ AND GF(4), TOTAL NUMBER OF SIMULATIONS $\approx 1.5 \times 10^7$.

Error Type	(4, 4)	(5, 0)	(5, 2)	(6, 2)	(6, 4)	(6, 6)	(7, 4)	(8, 2)	other
Original	43	9	14	8	8	12	7	6	23
P-method	0	0	0	0	0	9	5	4	14
A-method	0	0	0	0	0	0	0	0	15

$N \approx 2750$ bits, rate $R \approx 0.88$, column weight $c = 4$, row weight $r = 51$ for GF(2), $r = 37$ for GF(4), $r = 31$ for GF(8), $r = 26$ for GF(16) and girth $g = 6$ using a non-binary min-sum decoder [34]. The improvement in the performance of the code achieved by both P- and A-methods can also be observed for the min-sum decoder. Furthermore, Table II shows the error profile of these three codes over GF(4) at SNR=5.2 dB using the min-sum decoder. Although both P- and A-methods decrease the number of problematic absorbing sets in the original code, our proposed algorithm (A-method) is more effective as it cancels all small absorbing sets in the error profile.

Figure 7 shows the performance of non-binary quasi-cyclic codes (QC) [36] ('Original') with block length $N \approx 1200$, rate $R \approx 0.8$ and column weight $c = 4$ as well as modified codes using our proposed algorithm (A-method). This example shows that this algorithm works well independently of the code structure as it is effective for both structured and random codes. We again mention that the P-method also provides an improvement in the performance in error floor region. Similar to our previous examples, the gap between the curves for A- and P-methods diminishes for larger field sizes.

It is established that irregular non-binary codes over small field sizes provide better error correcting performance compared to regular non-binary codes [5], [16], and [18]. Since our proposed algorithm is not limited to any specific construction of LDPC codes, it can also be utilized for irregular codes. Figure 8 shows the performance of non-binary irregular codes constructed using the Progressive Edge-Growth (PEG) approach [16] ('Original') with block length $N \approx 2000$ bits, rate $R \approx 0.85$ and variable node degree distribution[6]

[6]Note that the chosen variable node degree distribution is obtained using the PEG approach in [16], but it is not optimal in terms of girth. The optimality of the variable node degree distribution in this example is not necessary for demostrating the effectiveness of our proposed algorithm; the example simply shows that the proposed algorithm works for irregular codes.

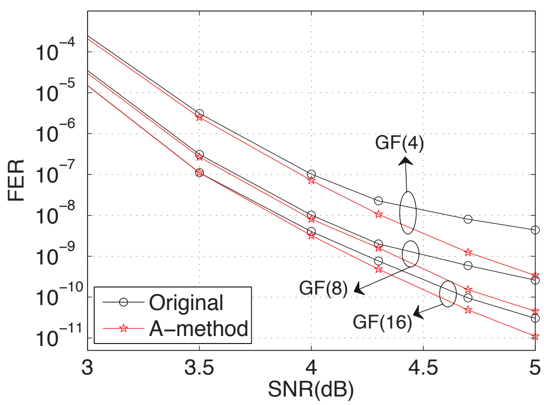


Fig. 7. FER versus SNR for non-binary QC codes and their A-method modified versions, $N \approx 1200$, $R \approx 0.8$, $c = 4$, and the QSPA-FFT decoder.

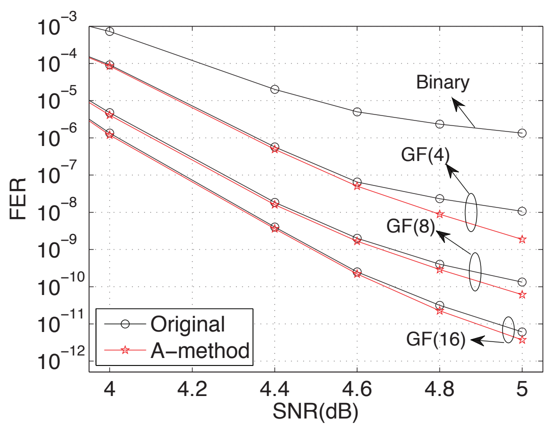


Fig. 8. FER versus SNR for non-binary irregular codes and their A-method modified versions, $N \approx 2000$, $R \approx 0.85$, $\Lambda(x) = 0.5x^4 + 0.5x^5$, and the QSPA-FFT decoder.

$\Lambda(x) = 0.5x^4 + 0.5x^5$ as well as modified codes using our proposed algorithm (A-method). This example reconfirms the effectiveness of our proposed algorithm in improving the performance of various code constructions in the error floor region.

Our various examples show that the proposed algorithm consistently improves the performances of the original codes across different choices of codes (different block lengths, column weights, and structures) and different decoder implementations (FFT-QSPA and min-sum decoders). These examples also show that our proposed definition of non-binary absorbing set appears to be applicable to any non-binary code regardless of the choice of the decoder.

V. Asymptotic Distribution of Non-Binary Absorbing Set in Regular Code Ensembles

In Section IV, we showed that non-binary absorbing sets offer a more refined finite block length characterization of errors for non-binary codes compared to trapping sets. This section extends this observation to the asymptotic regime. In particular, we find the asymptotic distribution of non-binary elementary absorbing sets in regular code ensembles by using techniques from graph theory. We also show that in the non-binary regime, as the alphabet size q gets larger, it is harder to satisfy the edge labeling conditions for non-binary absorbing sets.

Based on Theorem 1 we have the following corollary. Note that in order to distinguish unlabeled absorbing sets and non-binary absorbing sets, we refer to the weight assignments obeying the conditions in Definition 6 as *problematic*.

Corollary 1. *For regular code ensembles with column weight c and a given (a, b) unlabeled elementary absorbing set, the fraction of problematic weight assignments is given as $(q-1)^{\frac{b-(c-2)a-2}{2}}$.*

Proof: The total number of edges in an (a, b) elementary absorbing set in the regular code ensemble is ca. Since each unsatisfied check node is connected to one edge in the elementary absorbing set, $ca - b$ edges are connected to satisfied checks. Therefore the number of satisfied checks in an (a, b) absorbing set is $\frac{ca-b}{2}$. The rest of the proof is carried out by replacing e with $\frac{ca-b}{2}$ in Theorem 1. ∎

We now consider the asymptotic distribution of the non-binary absorbing sets. Let $\mathcal{G}^{c,r}_{n,m,q}$ denote the regular code ensemble with parity check matrices from the set $\Lambda^{c,r}_{n,m,q}$ which consists of all regular $m\times$n matrices of column-weight c and row weight r with elements from GF(q). The asymptotic distribution of non-binary absorbing sets in the regular code ensemble is defined as follows:

$$e^{c,r}(\theta, \lambda) \triangleq \lim_{n\to\infty} \frac{1}{n} \log \frac{z^{c,r}_{a,b,n}}{|\Lambda^{c,r}_{n,m,q}|}, \tag{7}$$

where $\theta = \frac{a}{n}$, $\lambda = \frac{b}{n}$, and $z^{c,r}_{a,b,n}$ is the number of (a, b) absorbing sets over all matrices in $\Lambda^{c,r}_{n,m,q}$. In (7) and in the following, log is taken to the base e.

The following lemma specifies the normalized asymptotic distributions for $c = 3$ and $c = 4$ cases.

Lemma 3. *The normalized logarithmic asymptotic distributions of $(a = \theta n, b = \lambda n)$ elementary absorbing sets in $\mathcal{G}^{3,r}_{n,m}$ and $\mathcal{G}^{4,r}_{n,m}$ regular code ensembles over GF(q) are given as*

$$\begin{aligned} e^{3,r}(\theta,\lambda) &= -2H_b(\theta,1-\theta) - H_b(\zeta,1-\zeta) + \lambda\log 3r \\ &+ H_b\left(1-\zeta, \lambda, \frac{3\theta-\lambda}{2}, \frac{2\zeta-3\theta-\lambda}{2}\right) \\ &+ \frac{3\theta-\lambda}{2}\log\binom{r}{2} + \theta H_b\left(\frac{\lambda}{\theta}, 1-\frac{\lambda}{\theta}\right) \\ &- 3\theta H_b\left(\frac{\lambda}{3\theta}, 1-\frac{\lambda}{3\theta}\right) - \frac{\theta-\lambda}{2} log(q-1), \end{aligned} \tag{8}$$

and

$$\begin{aligned} e^{4,r}(\theta,\lambda) &= -3H_b(\theta,1-\theta) - H_b(\zeta,1-\zeta) + \lambda\log 4r \\ &+ \frac{4\theta-\lambda}{2}\log\binom{r}{2} + \theta H_b\left(\frac{\lambda}{\theta}, 1-\frac{\lambda}{\theta}\right) \\ &+ H_b\left(1-\zeta, \lambda, \frac{4\theta-\lambda}{2}, \frac{2\zeta-4\theta-\lambda}{2}\right) \\ &- 4\theta H_b\left(\frac{\lambda}{4\theta}, 1-\frac{\lambda}{4\theta}\right) - \frac{2\theta-\lambda}{2} log(q-1), \end{aligned} \tag{9}$$

where $\zeta = \frac{m}{n}$, and $H_b(p_1, \cdots, p_N) = -\sum_{i=1}^{N} p_i \log p_i$ with $\sum_{i=1}^{N} p_i = 1$ denotes the entropy function.

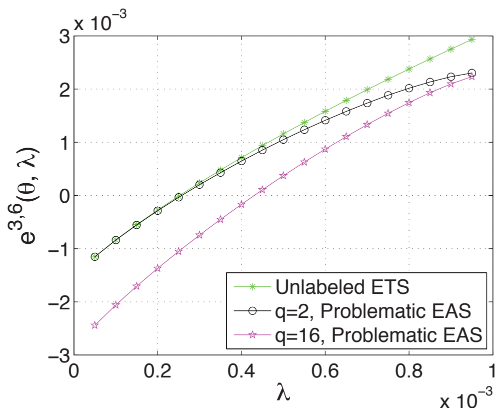


Fig. 9. Normalized logarithmic asymptotic distributions of elementary trapping sets (ETS) and absorbing sets (EAS) for $G^{3,6}_{n,m}$, $\theta = 0.001$ and $q = 2, 16$.

Proof: Using Corollary 1, from all unlabeled (binary) $(\theta n, \lambda n)$ absorbing sets in the Tanner graphs of regular code ensembles (enumerated in [1]), a fraction of $(q-1)^{\frac{(\lambda-(c-2)\theta)n-2}{2}}$ of all possible weight assignments results in $(\theta n, \lambda n)$ non-binary absorbing sets. For the details of the proof please see the Appendix A. ■

Note that (9) indicates that the normalized asymptotic number of absorbing sets in the ensemble is a *decreasing* function in the field size q for fixed θ and λ.

For a $(3, 6)$ regular ensemble, Figure 9 shows the asymptotic distribution of unlabeled $(\theta n, \lambda n)$ (binary) elementary trapping sets using the results in [22], as well as $(\theta n, \lambda n)$ problematic elementary absorbing sets for $q = 2, 16$ and fixed $\theta = 0.001$. Note that all unlabeled $(\theta n, \lambda n)$ (binary) elementary trapping sets will result in $(\theta n, (\lambda n)^+)$ non-binary trapping sets after labeling of the Tanner graph. We can observe that in particular for larger λ and for $q = 16$, the normalized asymptotic distribution of elementary absorbing sets is smaller than the one for elementary trapping sets. The reason for this behavior is that after the weight assignment in the non-binary case a smaller fraction of unlabeled trapping sets leads to problematic absorbing sets. Therefore, the resulting non-binary absorbing set enumerators provide a better assessment of the error floor compared to trapping set enumerators. Also, as the value of λ gets smaller, both curves for unlabeled elementary trapping sets and binary elementary absorbing set converge, which means that most of the unlabeled elementary trapping sets result in problematic elementary absorbing sets in this case.

Figure 10(a) shows the asymptotic distribution of absorbing sets for a regular $(3, 6)$ ensemble and fixed $\lambda/\theta \approx 1$. In this case, the distributions are similar for different values of q, since the absorbing sets include only one (fundamental) cycle and as a result, there is one constraint on the edge weights of absorbing sets. Therefore, the difference between the asymptotic distributions curves for different values of q is negligible in the logarithmic domain. On the other hand, in Figure 10(b) where $\lambda/\theta = 0.1$, the number of satisfied checks and consequently the number of fundamental cycles in the VN graph increases, which puts more constraints on edge weights to be problematic. This leads to an increasing gap between the curves for the binary and the non-binary cases. We also observe that an increase in q results in a smaller number of

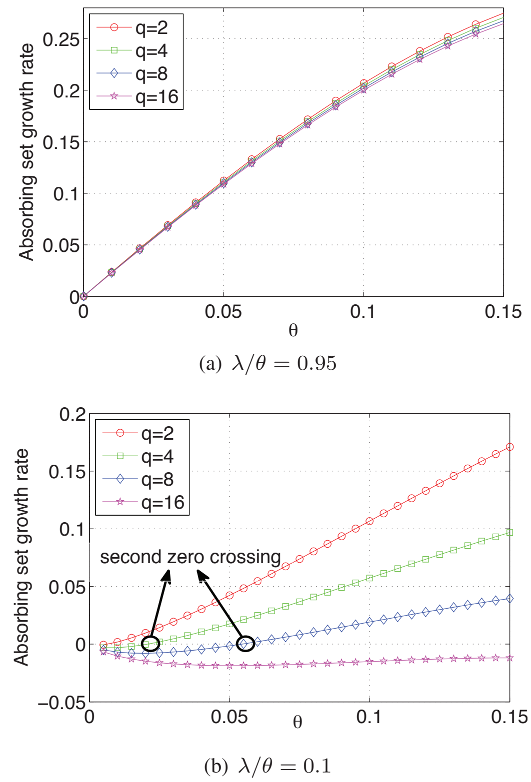


Fig. 10. Normalized logarithmic asymptotic distributions of elementary absorbing sets for regular code ensemble $G^{3,6}_{n,m}$, $q = 2, 4, 8, 16$ and (a) $\lambda/\theta = 0.95$, (b) $\lambda/\theta = 0.1$.

normalized asymptotic absorbing sets. The intuition for this behavior is that as q increases, the number of possible weight assignments also increases for an unlabeled structure and as a result, a smaller fraction of weight assignments satisfy the condition in Lemma 1. Figure 10(b) also demonstrates that as q increases, the second zero crossing of the absorbing set growth rate curve, i.e., the typical relative smallest absorbing set, increases as well. We note that the second zero crossing for GF(16) in Figure 10(b) is at $\theta \approx 0.61$, which is not shown in the figure.

VI. Conclusion

In this paper, we introduced a generalized absorbing set definition for non-binary graph based codes over GF(q). We observed that for non-binary codes where each edge in the Tanner graph has a weight chosen from non-zero elements of GF(q), not only the edges of absorbing sets are topologically connected in specific ways, but also their weights must satisfy certain conditions. We showed that in the case of non-binary elementary absorbing sets, the weight conditions can be simplified. Further, depending on the variable node input of a subgraph which satisfies the non-binary absorbing set conditions, the subgraph may or may not result in an absorbing set error (i.e., the error is input dependent). We proposed an algorithm to decrease the number of absorbing sets in the Tanner graph by only changing carefully chosen edge weights while the topology of the Tanner graph remains the same. Our simulation results using different code parameters and code structures as well as using different decoders confirmed

TABLE III
PARAMETERS OF CODES IN FIGURE 4.

	GF(2)	GF(4)	GF(8)	GF(16)
N	2601	2738	2883	2704
N_s	2601	1369	961	676
k	204	148	124	104
rate	0.921	0.891	0.870	0.846
d_v	4	4	4	4
d_c	51	37	31	26

the effectiveness of our proposed approach. Using techniques from graph theory, we showed that as q gets larger, it is harder to satisfy the edge labeling conditions for absorbing sets and as a result the number of absorbing sets decreases for larger field sizes. We also computed the normalized logarithmic asymptotic distributions of absorbing sets for regular $(3, r)$ and $(4, r)$ LDPC code ensembles. We observed that only a small fraction of unlabeled trapping sets leads to problematic absorbing sets; this results in non-binary absorbing set enumerators providing a better assessment of the error floor compared to trapping set enumerators. A potential extension of this work may include the analysis and enumeration of introduced non-binary absorbing sets in the context of various structured codes such as non-binary spatially-coupled and protograph-based codes.

APPENDIX A
PROOF OF LEMMA 3

The following proof is specifically for the $c = 4$ case; the case for $c = 3$ can be proved similarly. It is shown in Corollary 1 that for regular code ensembles with column weight l the fraction of problematic GF(q) weight assignment for an (a, b) unlabeled (binary) elementary absorbing set is $(q-1)^{\frac{b-(l-2)a-2}{2}}$. Therefore, for a normalized logarithmic asymptotic distribution of non-binary elementary $(\theta n, \lambda n)$ absorbing sets, one can find the normalized logarithmic asymptotic distribution of unlabeled (binary) elementary $(\theta n, \lambda n)$ absorbing sets, and then subtract $\lim_{n\to\infty} \frac{1}{n} \log(q-1)^{\frac{(\lambda-(l-2)\theta)n-2}{2}} = \frac{\lambda-(l-2)\theta}{2} \log(q-1)$. The normalized logarithmic asymptotic distribution of $(a = \theta n, b = \lambda n)$ unlabeled (binary) absorbing sets in $\mathcal{G}^{l,r}_{n,m}$ is calculated in our previous work as follows (please see [1] for details):

$$e^{l,r}(\theta,\lambda) = \lim_{n\to\infty} \frac{1}{n} \log \sum_{\{\delta_j,\sigma_i:1\le j\le m,1\le i\le n,S\}} \frac{\binom{n}{a}\binom{m}{b}}{\binom{nl}{al}\binom{al}{q}} \prod_{i=1}^{a}\binom{l}{\sigma_i}\prod_{j=1}^{m}\binom{r}{\delta_j},$$

where $q = \sum_{j=1}^{m-b} \delta_j$, and the summation runs over all δ_j's, and σ_i's which satisfy the following set of conditions:

$$S = \begin{cases} \sum_{j=1}^{m} \delta_j = al; \sum_{j=1}^{m-b} \delta_j = \sum_{i=1}^{a} \sigma_i; \\ \frac{l}{2} < \sigma_i \le l \quad \text{for } 1 \le i \le a; \\ \delta_j \quad \text{is even for } 1 \le j \le m-b; \\ \delta_j \quad \text{is odd for } m-b+1 \le j \le m. \end{cases}$$

In the case of elementary absorbing sets, check nodes are connected to bit nodes in an absorbing set at most twice and therefore, the δ_j's have values 0, 1, or 2. In particular, for $m-b+1 \leqslant j \leqslant m$, $\delta_j = 1$ since each unsatisfied check has only one edge connected to the absorbing set. Furthermore, the σ_i's are 2 or 3 for $1 \le i \le a$ and $l = 3$ ensembles and are 3 or 4 for $1 \le i \le a$ and $l = 4$ ensembles. As a result of these constraints in the elementary case and by using both binomial and Stirling's approximations, the normalized logarithmic asymptotic distribution of $(a = \theta n, b = \lambda n)$ unlabeled (binary) elementary absorbing sets in $\mathcal{G}^{4,r}_{n,m}$ ensemble are simplified as follows:

$$\begin{aligned} e^{4,r}(\theta,\lambda) \; = \; & -3H_b(\theta, 1-\theta) - H_b(\zeta, 1-\zeta) + \lambda \log 4r \\ & + \frac{4\theta-\lambda}{2} \log \binom{r}{2} + \theta H_b\left(\frac{\lambda}{\theta}, 1-\frac{\lambda}{\theta}\right) \\ & + H_b\left(1-\zeta, \lambda, \frac{4\theta-\lambda}{2}, \frac{2\zeta-4\theta-\lambda}{2}\right) \\ & - 4\theta H_b\left(\frac{\lambda}{4\theta}, 1-\frac{\lambda}{4\theta}\right) - \frac{2\theta-\lambda}{2} log(q-1). \end{aligned}$$

Subtracting $\frac{\lambda-(l-2)\theta}{2} \log(q-1)$ from the above equation results in Equation (9).

APPENDIX B
CODE PARAMETERS

Table III, Table IV and Table V include the code parameters for the results shown in Figures 4, 5 and 7, respectively.

TABLE IV
PARAMETERS OF CODES IN FIGURE 5.

	GF(2)	GF(4)	GF(8)	GF(16)
N	2500	2178	2352	2304
N_s	2500	1089	784	576
k	250	165	140	120
rate	0.90	0.84	0.821	0.791
d_v	5	5	5	5
d_c	50	33	28	24

TABLE V
PARAMETERS OF CODES IN FIGURE 7.

	GF(4)	GF(8)	GF(16)
N	1176	1197	1292
N_s	588	399	323
rate	0.809	0.789	0.764
p	28	21	19
d_v	4	4	4
d_c	21	19	17

ACKNOWLEDGEMENT

This work was supported in part by NSF grants CCF-0830666, CCF-1161774, and CCF-1161798 and funding support from ASTC/IDEMA.

REFERENCES

[1] B. Amiri, C. W. Lin, and L. Dolecek, "Asymptotic distribution of absorbing sets and fully absorbing sets for regular sparse code ensembles," *IEEE Trans. Commun.*, vol. 61, no. 2, pp. 455–464, Feb. 2013.

[2] B. Amiri, J. Kliewer, and L. Dolecek, "Analysis and enumeration of absorbing sets for non-binary graph-based codes," in *Proc. 2013 IEEE Int. Symp. Inf. Theory*, pp. 2815–2819.

[3] I. Andriyanova, D. Maurice, and J.-P. Tillich, "Quantum LDPC codes obtained by non-binary constructions," in *Proc. 2012 IEEE Int. Symp. Inf. Theory*, pp. 343–347.

[4] A. Bazarsky, N. Presman, and S. Litsyn, "Design of non-binary quasi-cyclic LDPC codes by ACE optimization," submitted to *IEEE Trans. Inf. Theory*, 2013. Available: http://arxiv.org/abs/1304.7487.

[5] B.-Y. Chang, D. Divsalar, and L. Dolecek, "Non-binary protograph-based LDPC codes for short blocklengths," in *Proc. 2012 IEEE Inf. Theory Workshop*, pp. 282–286.

[6] C.-Y. Chen, Q. Huang, C.-C. Chao, and S. Lin, "Two low-complexity reliability-based message-passing algorithms for decoding non-binary LDPC Codes," *IEEE Trans. Commun.*, vol. 58, no. 11, pp. 3140–3147, Nov. 2010.

[7] M. Davey and D. MacKay, "Low-density parity check codes over GF(q)," *IEEE Commun. Lett.*, vol. 2, no. 6, pp. 165–167, Jun. 1998.

[8] D. Declercq and M. Fossorier, "Decoding algorithms for nonbinary LDPC codes over GF(q)," *IEEE Trans. Commun.*, vol. 55, no. 4, pp. 633–643, Apr. 2007.

[9] D. Declercq, V. Savin, and L. Pham Sy, "Analysis and design of ultra-sparse non-binary cluster-LDPC codes," in *Proc. 2012 IEEE Int. Symp. Inf. Theory*, pp. 2531–2535.

[10] D. Divsalar and L. Dolecek, "Enumerators for protograph-based ensembles of nonbinary LDPC codes," in *Proc. 2011 IEEE Int. Symp. Inf. Theory*, pp. 913–917.

[11] L. Dolecek, "On absorbing sets of structured sparse graph codes," in *Proc. 2010 IEEE Inf. Theory Appl. Workshop.*

[12] L. Dolecek, D. Divsalar, Y. Sun, and B. Amiri, "Non-binary protograph-based LDPC codes: enumerators, analysis, and designs," submitted to *IEEE Trans. Inf. Theory*, 2013.

[13] L. Dolecek, Z. Zhang, V. Anantharam, M. J. Wainwright, and B. Nikolic, "Analysis of absorbing sets and fully absorbing sets of array-based LDPC codes," *IEEE Trans. Inf. Theory*, vol. 56, no. 1, pp. 181–201, Jan. 2010.

[14] M. F. Flanagan, V. Skachek, E. Byrne, and M. Greferath, "Linear-programming decoding of nonbinary linear codes," *IEEE Trans. Inf. Theory*, vol. 55, no. 9, pp. 4134–4154, Sept. 2009.

[15] D. Goldin and D. Burshtein, "Iterative linear programming decoding of non-binary LDPC codes with linear complexity," *IEEE Trans. Inf. Theory*, vol. 59, no. 1, pp. 282–300, Jan. 2013.

[16] 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.

[17] J. Huang, L. Liu, W. Zhou, and S. Zhou, "Large-girth nonbinary QC-codes of various lengths," *IEEE Trans. Commun.*, vol. 58, no. 12, pp. 3436–3447, Dec. 2010.

[18] J. Huang, S. Zhou, and P. Willett, "Nonbinary LDPC coding for multicarrier underwater acoustic communication," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 9, pp. 1684–1696, Dec. 2008.

[19] J. Kang, Q. Huang, L. Zhang, B. Zhou, and S. Lin, "Quasi-cyclic LDPC codes: an algebraic construction," *IEEE Trans. Commun.*, vol. 58, no. 5, pp. 1383–1396, May 2010.

[20] M. Karimi and A. H. Banihashemi, "An efficient algorithm for finding dominant trapping sets of LDPC codes," *IEEE Trans. Inf. Theory*, vol. 58, pp. 6942–6958, Nov. 2012.

[21] C. Koller, A. Graell i Amat, J. Kliewer, and D. J. Costello, "Trapping set enumerators for repeat multiple accumulate code ensembles," in *Proc. 2009 IEEE Int. Symp. Inf. Theory*, pp. 1818–1823.

[22] O. Milenkovic, E. Soljanin, and P. Whiting, "Asymptotic spectra of trapping sets in regular and irregular LDPC code ensembles," *IEEE Trans. Inf. Theory*, vol. 53, no. 1, pp. 39–55, Jan. 2007.

[23] D. G. M. Mitchell, "A mathematical approach to channel codes with a diagonal matrix structure," Ph.D. dissertation, The University of Edinburgh, 2009.

[24] D. V. Nguyen, S. K. Chilappagari, B. Vasic, and M. W. Marcellin, "On the construction of structured LDPC codes free of small trapping sets," *IEEE Trans. Inf. Theory*, vol. 58, no. 4, pp. 2280–2302, Apr. 2012.

[25] T. Nozaki, K. Kasai, and K. Sakaniwa, "Analysis of error floors of non-binary LDPC codes over MBIOS channel," in *Proc. 2011 IEEE Int. Conf. on Commun.*, pp. 1–5.

[26] T. Nozaki, K. Kasai, and K. Sakaniwa, "Error floors of non-binary LDPC codes," in *Proc. 2010 IEEE Int. Symp. Inf. Theory*, pp. 729–733.

[27] T. Nozaki, K. Kasai, and K. Sakaniwa, "Analysis of error floors of generalized non-binary LDPC codes over q-ary memoryless symmetric channels," in *Proc. 2012 IEEE Int. Symp. Inf. Theory*, pp. 2341–2345.

[28] T. Nozaki, K. Kasai, and K. Sakaniwa, "Analysis of stopping constellation distribution for irregular non-binary LDPC code ensemble," in *Proc. 2011 IEEE Int. Symp. Inf. Theory*, pp. 1101–1105.

[29] A. Poloni, S. Valle, and S. Vincenti, "NB-LDPC: absorbing set and importance sampling," in *Proc. 2012 Inter. Symp. Turbo Codes and Iter. Inf. Proc.*, pp. 101–105.

[30] 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.

[31] V. Rathi and I. Andriyanova, "Some results on MAP decoding of non-binary LDPC codes over the BEC," *IEEE Trans. Inf. Theory*, vol. 57, no. 4, pp. 2225–2242, Apr. 2011.

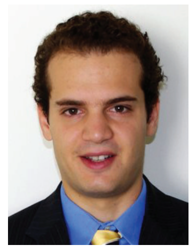
[32] T. J. Richardson, "Error floors of LDPC codes," in *Proc. 2003 Allerton Conf. Comm., Cont., and Comp.*, pp. 1426–1435.

[33] L. Sassatelli and D. Declercq, "Nonbinary hybrid LDPC codes," *IEEE Trans. Inf. Theory*, vol. 56, no. 10, pp. 5314–5334, Oct. 2010.

[34] A. Voicila, D. Declercq, F. Verdier, M. Fossorier, and P. Urard, "Low-complexity decoding for non-binary LDPC codes in high order fields," *IEEE Trans. Commun.*, vol. 58, no. 5, pp. 1365–1375, May 2010.

[35] J. Wang, L. Dolecek, and R. Wesel, "The cycle consistency matrix approach to LDPC absorbing sets in separable circulant-based codes," *IEEE Trans. Inf. Theory*, vol. 59, no. 4, pp. 2293–2314, 2013.

[36] B. Zhou, J. Kang, Y. Y. Tai, S. Lin, and Z. Ding, "High performance non-binary quasi-cyclic LDPC codes on Euclidean geometries," *IEEE Trans. Commun.*, vol. 57, no. 5, pp. 1298–1311, 2009.



Behzad Amiri (S'10) is a Ph.D. student in Electrical Engineering Department at the University of California, Los Angeles (UCLA). He received his B.S. degree from Isfahan University of Technology, Isfahan, Iran and the M.S. degree in Electrical Engineering from University of California, Los Angeles (UCLA) in 2012. He is currently a Graduate Student Researcher in Loris Lab at UCLA. His research interests include graph-based code designs and iterative algorithms.

Jörg Kliewer (S'97–M'99–SM'04) received the Dipl.-Ing. (M.Sc.) degree in electrical engineering from Hamburg University of Technology, Hamburg, Germany, in 1993 and the Dr.-Ing. degree (Ph.D.) in electrical engineering from the University of Kiel, Germany, in 1999, respectively.

From 1993 to 1998, he was a Research Assistant at the University of Kiel, and from 1999 to 2004, he was a Senior Researcher and Lecturer with the same institution. In 2004, he visited the University of Southampton, U.K., for one year, and from 2005 until 2007, he was with the University of Notre Dame, IN, as a Visiting Assistant Professor. From 2007 until 2013 he was with New Mexico State University, Las Cruces, NM, most recently as an Associate Professor. He is now with the New Jersey Institute of Technology, Newark, NJ, as an Associate Professor. His research interests include error-correcting codes, network coding, information theory, and communication networks.

Dr. Kliewer was the recipient of a Leverhulme Trust Award and a German Research Foundation Fellowship Award in 2003 and 2004, respectively. He was a Member of the Editorial Board of the EURASIP JOURNAL ON ADVANCES IN SIGNAL PROCESSING from 2005-2009 and is Associate Editor of the IEEE TRANSACTIONS ON COMMUNICATIONS since 2008.

Lara Dolecek (S'05–M'10–SM'13) is an Assistant Professor with the Electrical Engineering Department at the University of California, Los Angeles (UCLA). She holds a B.S. (with honors), M.S. and Ph.D. degrees in Electrical Engineering and Computer Sciences, as well as an M.A. degree in Statistics, all from the University of California, Berkeley. She received the 2007 David J. Sakrison Memorial Prize for the most outstanding doctoral research in the Department of Electrical Engineering and Computer Sciences at UC Berkeley. Prior to joining UCLA, she was a postdoctoral researcher with the Laboratory for Information and Decision Systems at the Massachusetts Institute of Technology. She received Northrop Grumman Excellence in Teaching Award, Intel Early Career Faculty Award, University of California Faculty Development Award, and Okawa Research Grant all in 2013, NSF CAREER Award in 2012, and Hellman Fellowship Award in 2011. She is an Associate Editor for Coding Theory for IEEE TRANSACTIONS ON COMMUNICATIONS and for IEEE COMMUNICATION LETTERS and is the lead guest editor for IEEE JSAC special issue on emerging data storage. Her research interests span coding and information theory, graphical models, statistical algorithms, and computational methods, with applications to emerging systems for data storage, processing, and communication.