

# Union bound evaluation for non-binary turbo coded modulations

Rami Klaimi, Charbel Abdel Nour, Catherine Douillard, Joumana Farah

► **To cite this version:**

Rami Klaimi, Charbel Abdel Nour, Catherine Douillard, Joumana Farah. Union bound evaluation for non-binary turbo coded modulations. IEEE Communications Letters, Institute of Electrical and Electronics Engineers, In press, 10.1109/LCOMM.2020.2980268 . hal-02502303

**HAL Id: hal-02502303**

**<https://hal-imt-atlantique.archives-ouvertes.fr/hal-02502303>**

Submitted on 9 Mar 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Union Bound Evaluation for Non-Binary Turbo Coded Modulations

Rami Klaimi, *Member, IEEE*, Charbel Abdel Nour, *Senior Member, IEEE*,  
Catherine Douillard, *Senior Member, IEEE*, and Joumana Farah, *Member, IEEE*

**Abstract**—A method to compute the truncated union bound of non-binary turbo codes mapped to high order modulations is proposed in this letter. It calls for the estimate of the truncated Euclidean distance spectrum. To this end, we identify the error events that limit the Euclidean distance of non-binary turbo codes based on memory-1 convolutional codes defined over  $\text{GF}(q)$ ,  $q > 2$ . Application examples are elaborated for codes over  $\text{GF}(64)$ . The resulting bounds are found to accurately predict the asymptotic performance of the coded modulation scheme.

**Index Terms**—Non-binary turbo codes, Euclidean distance spectrum, union bound, coded modulation.

## I. INTRODUCTION

NON-BINARY (NB) forward error correction (FEC) codes represent strong candidates for next generation communication standards. First, coded modulation (CM) capacity is always greater than or equal to bit-interleaved coded modulation (BICM) capacity [1], [2]. This gain can be up to 1.0 dB in additive white Gaussian noise (AWGN) channel. Second, NB codes have been shown to perform better than binary codes for small block lengths [3], [4]. The transmission of short frames is of high importance for ultra reliable low-latency communication (uRLLC) scenarios [5] in the upcoming standards. On another note, probabilistic amplitude shaping (PAS) [6] was recently proposed to close the shaping gap between constrained capacity and unconstrained capacity, by adjusting the channel input distribution. BICM schemes based on binary FEC codes and using PAS were then able to approach the unconstrained channel capacity limits. Further performance improvements of PAS with NB FEC codes were reported in [7], [8] where coding gains between 0.35 dB and 0.63 dB were reported for short block sizes, compared to binary FEC codes designed with the same PAS settings.

The first attempt to approach CM capacity dates back to the turbo trellis-coded modulation (TTCM) schemes [9], based on Ungerboeck's TCMs. More recently, several studies have targeted the design of NB turbo codes (NB-TC). In [3], a NB-TC structure was derived from a protograph sub-ensemble of a NB-LDPC code. In [10], NB-TCs over  $\text{GF}(q)$  were designed based on their EXIT charts and an estimate of their truncated union bounds (TUB). Both studies report significant gains in the short-to-moderate block size regime compared to state-of-the-art binary FEC codes. Recently, new NB recursive systematic convolutional codes over  $\text{GF}(q)$ ,  $q > 2$ , were proposed in [11]. The study of the interleaver and the introduction of a low-complexity decoder were then

presented in [12] and [13], respectively, to complement [11] towards the construction of practical NB-TCs outperforming previously published NB FEC codes.

The TUB on the frame error probability is a useful tool to predict the asymptotic performance of FEC codes, which requires the computation of their truncated distance spectrum. The estimation of the truncated Hamming distance spectrum for binary TCs mapped to low- and high-order modulations, as well as for NB-TCs mapped to binary modulations has already been considered in the literature [14], [15]. However, to the best of our knowledge, this question has not yet fully been addressed for NB-TCs mapped to high-order constellations. In this case, the computation of the bound calls for the estimate of the truncated Euclidean distance (ED) spectrum of the code. When the cardinality of the code alphabet  $q$  is high, a large number of transitions connect the states in the code trellis, making the enumeration of the sequences very complex. In [10], the  $M$ -algorithm is proposed to calculate TUBs for NB convolutional codes (NB-CC): only  $M$  branches, having the least reliable metrics, are considered at each trellis section. The union bound analysis is then only used to select the constituent convolutional codes and no TUB is derived for the actual resulting NB-TC. In this work, we propose a different and low-complexity method to enumerate error events and to compute TUBs for the NB-TCs studied in [11] and [12].

This letter is structured as follows. Section II introduces the structure of the NB-TCs under study and the main features of their constituent convolutional codes. It also provides the formulation of the union bound for NB codes. In Section III, we describe the different types of error events with low EDs. A method is proposed in Section IV to calculate the resulting ED, as well as their corresponding multiplicities in order to compute the TUB on frame error probability. Simulation results are shown in Section V, and compared to the computed TUBs for several NB-TCs. Section VI concludes the paper.

## II. PRELIMINARIES ON NB-TCs

### A. Non-binary turbo codes: structure and constituent codes

The NB-TCs studied in this letter consist of a parallel concatenation of two memory-1 NB-CC (defined in [11]) and a symbol-wise interleaver as shown in the left-hand side of Fig. 1. The proposed method applies when the output  $u$  of the code defined over  $\text{GF}(q)$  is directly mapped to a  $q$ -QAM constellation or to a smaller constellation, such as a 4-QAM. In the latter case,  $q$  has to be a power of 4 and the  $q$ -ary symbols are decomposed into 4-ary symbols. Since high-order QAM constellations do not have the uniform error property [16], the distance spectrum is not the same for each codeword and one cannot assume the transmission of the all-zero codeword to determine the distance spectrum of the whole code. Therefore,

R. Klaimi, C. Abdel Nour and C. Douillard are with IMT Atlantique, Lab-STICC, UBL, 29238 Brest, France (e-mail: rami.klaimi@imt-atlantique.fr; charbel.abdelnour@imt-atlantique.fr; catherine.douillard@imt-atlantique.fr).

J. Farah is with the Faculty of Engineering, Lebanese University, Roumieh, Lebanon (e-mail: joumanafarah@ul.edu.lb).

all possible pairs of sequences diverging from a state in the trellis diagram and then converging back to a state have to be considered. Such sequence pairs are called diverging and converging (DC) sequences in this letter. It was shown in [11] that, since the trellises of these codes are fully connected (as shown in Fig. 1(c) for  $q = 4$ ), finding the length-2 and length-3 DC sequences is sufficient to compute the first two distance terms of the ED spectrum with their exact multiplicities.

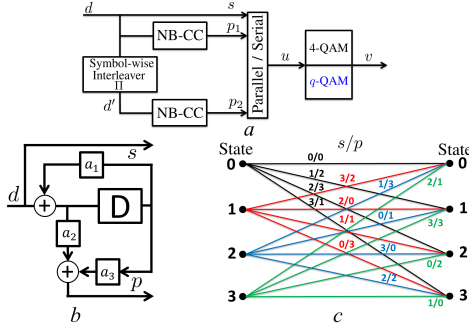


Fig. 1. (a) Transmitter block diagram. (b) Structure of the used NB-CC with one memory element. (c) Example of a fully connected trellis for a code defined over  $GF(4)$ , with generator polynomial  $P_{GF(4)}(x) = 1 + x + x^2$  and  $(a_1, a_2, a_3) = (3, 2, 2)$ .

### B. Union bound formulation for NB codes

The frame error probability,  $P_f$ , of an NB code defined over  $GF(q)$  is upper bounded by:

$$P_f \leq \frac{1}{|C|} \sum_{m=1}^{|C|} \sum_{k=1, k \neq m}^{|C|} P(X^m \rightarrow X^k) \quad (1)$$

where  $C$  represents the set of codewords and  $|C|$  is its cardinality:  $|C| = q^K$  if  $K$  is the message length.  $P(X^m \rightarrow X^k)$  is the probability of transmitting codeword  $X^m$  and decoding  $X^k$ . For a transmission over the additive white Gaussian noise (AWGN) channel, this expression becomes:

$$P_f \leq \frac{2}{|C|} \sum_{i=1}^{i_{max}} n(d_i) Q\left(\frac{d_i}{2\sigma}\right) = 2 \sum_{i=1}^{i_{max}} r(d_i) Q\left(\frac{d_i}{2\sigma}\right) \quad (2)$$

where  $d_i, i = 1, \dots, i_{max}$  are the EDs between the codewords, sorted in ascending order,  $d_{min} = d_1 < \dots < d_{i_{max}}$ ,  $n(d_i)$  and  $r(d_i) = \frac{n(d_i)}{|C|}$  represent the multiplicity and the normalized multiplicity of distance  $d_i$ . In practice, the squared ED (SED)  $d_i^2$  between two codewords is computed as the sum of the SEDs between the constellation signals the codewords are mapped to.  $\sigma$  is the noise standard deviation and  $Q(x)$  is the tail distribution function of the standard normal distribution. The factor 2 in (2) comes from the fact that each distance  $d_i$  has to be accounted for twice: if  $d_i$  is the ED between codewords  $X^m$  and  $X^k$ , it appears in both terms  $P(X^m \rightarrow X^k)$  and  $P(X^k \rightarrow X^m)$  in (1).

### III. IDENTIFICATION OF LOW-DISTANCE DC SEQUENCES FOR NB-TCs

The computation of the exact union bound requires the enumeration of the complete ED spectrum of the code, which is not feasible for large values of  $q$  and/or  $K$ . Therefore, in

practice, for a given code, one has to identify the most frequent error events, in order to limit the search process. Thanks to the fully connected structure of the trellis of the NB-CCs adopted in this work [11], low-distance error events at the constituent code level can be obtained by enumerating the short – i.e. length-2 and length-3 – DC sequences in the corresponding trellis, as already mentioned in II-A. As a consequence, the low-distance error events for the resulting NB-TC are obtained when short DC sequences for the first NB-CC are combined with short DC sequences for the second NB-CC, due to the effect of interleaving. We will refer to these sequences as turbo-DC sequences.

To the best of our knowledge, no method currently exists to achieve an accurate enumeration of low-distance turbo-DC sequences for a specific interleaver. To address this problem, we have observed the structure of low-distance error events at the output of the turbo decoder and we have identified the underlying reasons for their occurrence, regardless of the interleaver structure: the presence of short cycles in the Tanner graph of the code and the presence of competing sequences differing only in two input symbols.

In what follows,  $X^1$  denotes the transmitted sequence and  $X^2$  is a competitor of  $X^1$ .  $\tilde{X}^1$  and  $\tilde{X}^2$  are their interleaved versions. In the figures, we assume the transmission of the all-zero sequence ( $X_j^1 = 0 \forall j \in [0, K-1]$ ) starting from state  $\mathbf{0}$  for illustrative purposes only, since all sequences and initial states are to be considered for the estimation of the ED spectrum.

#### A. Effect of short cycles in the Tanner graph of NB-TCs: type-A turbo-DC sequences

Due to iterative decoding, short cycles in the Tanner graph of a turbo code [17] induce a high level of correlation between the extrinsic information sequences exchanged by the constituent decoders. This was shown to hinder the effectiveness of iterative processing for error correction [18], [12]. In the case of the NB-TCs under study, short cycles have a greater effect on performance than for conventional binary TCs, due to the fully connected trellises. Fig. 2 shows an example of a cycle and the resulting turbo-DC sequences in the trellises of the NB-TC constituent codes. The interleaver simultaneously creates a short cycle of length 6 and a pair of turbo-DC sequences, involving the all-zero sequence  $X^1$  and its competitor  $X^2$ . The turbo-DC sequences created by the cycle are composed of 4 DC sub-sequences numbered from I to IV. The competing sequence differs from the transmitted all-zero sequence in positions  $\{3, 4, 7, 8, 9\}$  in the natural order and  $\{2, 3, 4, 8, 9\}$  in the interleaved order. In general, we can always find a sequence  $X^2$  (respectively  $\tilde{X}^2$ ), competing with  $X^1$  (respectively  $\tilde{X}^1$ ) such that  $X^1$  and  $X^2$  (respectively  $\tilde{X}^1$  and  $\tilde{X}^2$ ) share the same systematic symbols in all positions except in the positions belonging to a short cycle. When such a scenario occurs, since sequences  $X^1$  and  $X^2$  (respectively, their interleaved versions  $\tilde{X}^1$  and  $\tilde{X}^2$ ) only differ in a low number of trellis sections, they are highly likely to provide a low ED. These sequences caused by short cycles are called type-A turbo-DC sequences in the remainder of the paper.

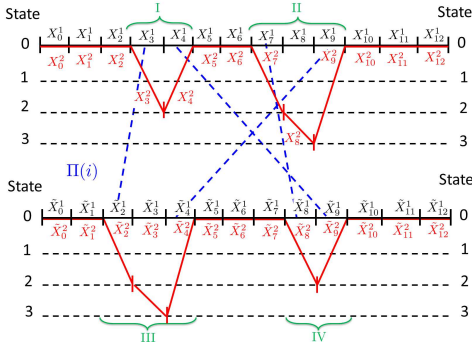


Fig. 2. Example of type-A turbo-DC sequences created by a short cycle for a  $q=4$  NB-TC. The all-zero sequence is assumed to be transmitted. The length of the cycle is  $g=6$ .

### B. Effect of DC sequences differing in two input symbols: type-B turbo-DC sequences

Another type of turbo-DC sequences that may result in low EDs is depicted in Fig. 3. They consist of a pair of DC sequences of length  $(l_1 + 2)$  in the trellis diagram of the first constituent code (natural order) and a pair of DC sequences of length  $(l_2 + 2)$  in the trellis diagram of the second constituent code (interleaved order). If  $S$  is the minimum spread value of the interleaver, such sequences can only exist if  $l_1 + l_2 + 4 \geq S$ . Due to the structure of the constituent codes, it is ensured that  $X^1$  and  $X^2$  (respectively  $\tilde{X}^1$  and  $\tilde{X}^2$ ) differ by at least the systematic symbols in the trellis sections where they diverge and converge: in Fig. 3, they differ at positions  $\{3, 5\}$  in the natural order, and at positions  $\{2, 10\}$  in the interleaved order. Due to the fully connected structure of the trellis, there exist some sequences  $X^1$  and  $X^2$  (respectively  $\tilde{X}^1$  and  $\tilde{X}^2$ ) with the same systematic symbols in the  $l_1$  (respectively  $l_2$ ) remaining positions. In this case, these positions only contribute to the ED through their parity symbols. For low values of  $L=l_1+l_2$ , these turbo-DC sequences have low EDs even though they meet the spread constraint of the interleaver: they are called type-B turbo-DC sequences in the remainder of the paper.

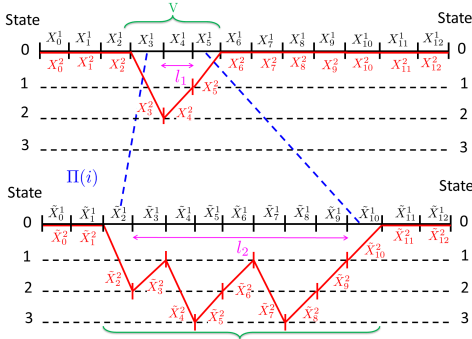


Fig. 3. Example of type-B turbo-DC sequences for a  $q=4$  NB-TC. A length- $(l_1 + 2)$  DC sequence is shown in the natural order and a length- $(l_2 + 2)$  DC sequence is shown in the interleaved order.

## IV. ESTIMATION OF THE TRUNCATED DISTANCE SPECTRUM OF NB-TCs

The proposed method first identifies all the type-A sequences corresponding to the shortest cycles created by the interleaver. Then, according to the minimum ED obtained in the first step, we set an upper limit on the length of the type-B sequences that will be searched for in a second step. The

proposed method ensures that all turbo-DC sequences with the lowest ED value are identified, as well as some of the turbo-DC sequences with higher distances.

### A. Computing the Euclidean distance of type-A sequences

For type-A turbo-DC sequences, the first task is to determine cycles having the shortest lengths, in order to calculate the corresponding minimum achievable ED and then estimate the corresponding multiplicity. The computation of the minimum achievable ED for a given cycle involves finding the different possible turbo-DC sequence shapes for the considered cycle and splitting them into elementary DC sequences, which can be classified into two groups: length-2 DC sequences and length- $l$  DC sequences with  $l > 2$ . Then, the minimum SED can be computed for each turbo-DC sequence shape by simply adding up the SEDs of the constituent DC sequences.

For DC sequences longer than 2 (II and III in Fig. 2), the minimum ED values can be obtained by splitting up the sequences into three types of trellis sections: a diverging section  $D$ , a converging section  $C$  and  $l - 2$  intermediate sections  $I$ . The diverging section  $D$  involves competing transitions that diverge from the same state in the trellis with different systematic and parity values ( $X_7^1$  and  $X_7^2$  for II in Fig. 2), the converging section  $C$  involves the competing trellis transitions that converge to the same state ( $X_9^1$  and  $X_9^2$  for II in Fig. 2), and the competing transitions for intermediate sections  $I$  do not share the same starting state nor the same ending state ( $X_8^1$  and  $X_8^2$  for II in Fig. 2). For a given NB-CC, the minimum values of the SEDs related to each section type,  $D$ ,  $I$  and  $C$ , can be easily computed from the trellis. The minimum SED for the DC sequences of length  $l > 2$ ,  $d^2(\text{DC} - l)$ , is given by:

$$d^2(\text{DC} - l) = d^2(D) + (l - 2)d^2(I) + d^2(C) \quad (3)$$

where  $d^2(D)$ ,  $d^2(C)$  and  $d^2(I)$  are the minimum SED values for each section type. For a rate-1/2 NB-CC code, each term in (3) is the sum of two distance terms related to the systematic and parity symbols  $s$  and  $p$ :  $d^2(\cdot) = d_s^2(\cdot) + d_p^2(\cdot)$ . Then, each term is calculated as the SED between the constellation signals on which the symbols have been mapped. In fact, due to the design rules of the NB-CC [11] and, especially due to the fully-connected trellis, the minimum distance value for a section  $I$  is achieved when the systematic symbols of both sequences are identical in the considered trellis section. Then  $d_s(I)$  is equal to 0 and the minimum achievable SED for length- $l$  ( $l > 2$ ) DC sequences is equal to:

$$d^2(\text{DC} - l) = (d_s^2(D) + d_p^2(D)) + (l - 2)d_p^2(I) + (d_s^2(C) + d_p^2(C)) \quad (4)$$

When the NB code is directly mapped to a  $q$ -QAM constellation,  $d_p(I)$  is equal to the minimum Euclidean distance between two signals in the constellation,  $d_{q-QAM}$ .

For length-2 DC sequences such as I and IV in Fig. 2, this decomposition is not valid since the NB-CCs are designed to guarantee that the minimum SED of a length-2 sequence is greater than the sum of the minimum SEDs of sections  $D$  and  $C$  taken separately [11]. Therefore, length-2 DC sequences have to be handled apart and  $d^2(\text{DC} - 2)$  is known for a given

NB-CC. It can also be written as:

$$d^2(\text{DC} - 2) = d_s^2(\text{DC} - 2) + d_p^2(\text{DC} - 2) \quad (5)$$

In the example of Fig. 2, the cycle of length 6 results in pairs of turbo-DC sequences composed of length-2 and length-3 DC sequences for both constituent codes. The minimum achievable SED  $d_{g=6}^2$  for this shape of turbo-DC sequences is obtained by adding up the minimum achievable SEDs of the constituent DC sequences (I to IV). However, since the systematic symbols are transmitted once for both constituent codes, they should not be counted twice. Therefore, only the parity terms are considered for sequences I and II or for sequences III and IV. Then:

$$d_{g=6}^2 = \underbrace{d^2(\text{DC} - 2)}_{\text{I}} + \underbrace{d^2(\text{DC} - 3)}_{\text{II}} + \underbrace{d_p^2(\text{DC} - 3)}_{\text{III}} + \underbrace{d_p^2(\text{DC} - 2)}_{\text{IV}} \quad (6)$$

The SED of a type-A turbo-DC sequence increases with the length of the cycle. In practice, we enumerate all the type-A sequences with the shortest cycle length  $g_{min}$  (i. e. the *girth*) and their corresponding SED is computed using (4) and/or (5). In order to get a more accurate value of the bound, the proposed approach can be extended to higher cycle lengths, but the search complexity is then significantly increased.

### B. Computing the Euclidean distance of type-B sequences

Type-B turbo-DC sequences are composed of two pairs of DC sequences of length  $l_1 + 2$  and  $l_2 + 2$ . As mentioned earlier, for this type of sequences, the lowest SED values are obtained when all the systematic symbols in the competing sequences are identical, except in the  $D$  and  $C$  trellis sections. Therefore, using the same notations as in the previous section and counting the systematic terms only once, according to (4), the minimum achievable SED for the type-B turbo-DC sequences of Fig. 3 is expressed as:

$$d_{l_1+l_2}^2 = \underbrace{\left( d_s^2(D) + d_s^2(C) \right)}_{\text{V or VI}} + \underbrace{\left( d_p^2(D) + l_1 d_p^2(I) + d_p^2(C) \right)}_{\text{V}} + \underbrace{\left( d_p^2(D) + l_2 d_p^2(I) + d_p^2(C) \right)}_{\text{VI}} \quad (7)$$

In order to limit the complexity of the enumeration process for type-B turbo-DC sequences, an upper limit has been set on the value of  $L = l_1 + l_2$  to be considered. Targeting sequences with low SED, only type-B sequences with distances lower than or equal to the minimum SED achieved by the enumerated type-A turbo-DC sequences,  $d_{gmin}^2$ , are considered. The value of  $L$  is chosen so that  $d_{gmin} \geq d_{l_1+l_2}$ . Rearranging (7), we obtain:

$$L \leq L_{max} = \frac{d_{gmin}^2 - \left( d_s^2(D) + d_s^2(C) \right) - 2 \left( d_p^2(D) + d_p^2(C) \right)}{d_p^2(I)} \quad (8)$$

Type-B turbo-DC sequences with length up to  $L_{max}$  are enumerated and their corresponding SEDs are computed using (7). In order to get a more accurate value of the bound, the enumeration can be extended to longer type-B sequences, but the search complexity is then significantly increased.

### C. Computation of the truncated union bound

After the enumeration of type-A and type-B turbo-DC sequences, the TUB is computed following (2). However, the computation of the term  $r(d_i)$  in (2) cannot be directly derived because all the enumerated types of turbo-DC sequences do not have the same weight in the bound formula. In fact, each considered pair of turbo-DC sequences represents a larger subset of codewords, defined by the possible values of the  $K-l$  information symbols that do not participate in the considered turbo-DC sequence of length  $l$ . These  $K-l$  systematic symbols can take any value in  $\text{GF}(q)$ . For type-A sequences,  $l = g$  where  $g$  is the length of the correlation cycle, and for type-B sequences  $l = 2 + l_1 + l_2$ . In both cases, each turbo-DC sequence accounts for a subset of  $q^{K-l}$  pairs of codewords.

The enumeration of each subset of type-A and type-B sequences provides an ED sub-spectrum, with distance values  $d_i, i \geq 1$  and their multiplicities. Let  $\mathcal{S}(l, d_i)$  denote the set of the enumerated type-A and type-B turbo-DC sequences of length  $l$  having Euclidean distance  $d_i$  and let  $|\mathcal{S}(l, d_i)|$  be the cardinality of this set. Since  $|\mathcal{C}| = q^K$ , which is the total number of codewords, the normalized multiplicity is computed as:

$$r(d_i) = \frac{\sum_l |\mathcal{S}(l, d_i)| q^{K-l}}{q^K} \quad (9)$$

Depending on the cycle length considered for enumerating type-A sequences and the value  $L_{max}$  considered for the enumeration of type-B sequences, the multiplicities of larger EDs may be underestimated. Despite this simplification and since the  $Q(\cdot)$  function decreases rapidly, limiting the study to the lowest EDs is often sufficient to predict the asymptotic performance of the considered codes.

## V. NUMERICAL RESULTS

Following the proposed method, the TUBs of two NB-TCs over  $\text{GF}(64)$  mapped to 64-QAM and 4-QAM are estimated. Two NB-CCs taken from [11] are used:  $c_1$  and  $c_2$  are NB-CCs with the worst and best distance spectra reported in [11], respectively. The polynomial used for the generation in  $\text{GF}(64)$  is  $P_{\text{GF}(64)}(x) = 1 + x^2 + x^3 + x^5 + x^6$ . The binary images of the 64-QAM and 4-QAM symbols are denoted by  $b_5 b_4 b_3 b_2 b_1 b_0$  and  $b_1 b_0$ , where  $b_5$  and  $b_1$  represent the most significant bits (MSB), respectively. The mapping to the constellations is given in Table I.

TABLE I  
BINARY MAPPING OF THE QUADRATURE/IN-PHASE AXIS FOR 64-QAM AND 4-QAM CONSTELLATIONS.

64-QAM					4-QAM		
Value	+7	+5	+3	+1	Value	+1	-1
$b_5 b_4 b_3 b_1$	000	001	011	010	$b_1$	0	1
Value	-1	-3	-5	-7			
$b_5 b_3 b_1$	110	111	101	100			

No puncturing is applied, and two different frame sizes are considered:  $K = 160$   $\text{GF}(64)$  symbols, i.e. equivalent to 960 bits, for the codes mapped to 64-QAM and  $K = 30$   $\text{GF}(64)$  symbols, i.e. equivalent to 180 bits, for the codes mapped to 4-QAM. The symbol interleaver is based on almost regular permutation (ARP) [19] designed according to minimum



spread and girth criteria as in [12] and whose parameters are given in Table II.

TABLE II  
ARP INTERLEAVER PARAMETERS FOR THE SIMULATED NB-TCs (ARP EQUATION:  $\Pi(i) = (Pi + S(i \text{ MOD } Q)) \text{ MOD } K$ ).

	Girth	$P$	$Q$	$(S(0), \dots, S(Q-1))$
64-QAM	8	143	8	(0,150,62,10,110,22,20,98)
4-QAM	4	19	3	(0,25,14)

To estimate the ED spectrum, only type-A sequences corresponding to the minimum girth achieved by the interleaver (given in Table II) are enumerated. When mapped to a 64-QAM constellation, the resulting values of  $L_{max}$  used to enumerate type-B sequences are:  $L_{max} = 9$  for code  $c_1$ , and  $L_{max} = 32$  for code  $c_2$ . Table III provides the values of the first ED terms with their normalized multiplicities. TUBs were calculated according to (2) and (9).

TABLE III  
TRUNCATED DISTANCE SPECTRA FOR NB-TCs OVER GF(64) MAPPED TO A 64-QAM CONSTELLATION.  $R=1/3$  AND  $K = 160$  SYMBOLS.  $d_i, i = 1, 2, 3$  ARE THE 3 FIRST ED TERMS OF THE NB-TC SPECTRUM, AND  $r(d_i)$  THE NORMALIZED MULTIPLICITIES.

Code	$d_1$	$d_2$	$d_3$
$(a_1, a_2, a_3)$	$r(d_1)$	$r(d_2)$	$r(d_3)$
$c_1$	1.023	1.113	1.155
(41,2,0)	8.5e-04	7.3e-06	1.4e-08
$c_2$	1.603	1.662	1.718
(31,5,18)	1.4e-08	1.4e-08	5.8e-05

Fig. 4 shows the corresponding error correction performance in terms of frame error rate (FER) as a function of  $E_s/N_0$ : the rightmost set of curves corresponds to the case of a 64-QAM constellation, and the leftmost set of curves corresponds to the case of a 4-QAM constellation. TUBs are shown to approach the NB-TCs simulated performance very closely in the error floor region. The results of Fig. 4 validate the proposed approach for different component codes, constellation orders and frame sizes.

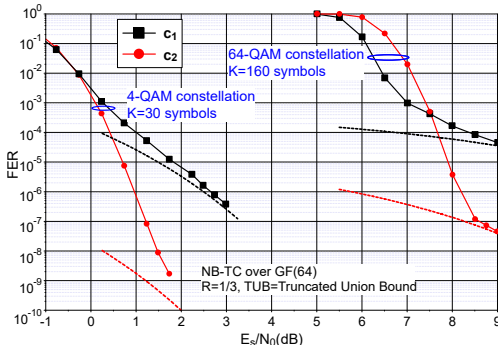


Fig. 4. FER performance of the NB-TCs defined over GF(64) with their corresponding TUBs. AWGN channel, 8 decoding iterations with the scaled Max-Log-MAP algorithm.

## VI. CONCLUSION

This letter describes a practical method for the computation of the TUB of NB-TCs associated with high-order modulations. The shapes of most frequent error events in the code trellises are identified. They are in practice caused by

short cycles in the Tanner graph of the NB-TC and by input sequences differing in only two symbols. The corresponding lowest ED values are calculated and TUBs can be estimated. Resulting bounds are found to approach simulated results in several considered cases, validating the proposed approach. They can therefore be a useful tool for the design of efficient memory-1 NB-TCs. The same method can be applied for NB-CCs with higher constraint lengths. However, in the last case, longer DC sequences should be considered and the enumeration complexity is then significantly increased.

## ACKNOWLEDGMENT

This work was partially funded by the EPIC project of the EU's Horizon 2020 research and innovation programme under grant agreement No. 760150, by Orange Labs and by the Pracom cluster. It has also received support from the Brittany region.

## REFERENCES

- [1] A. Alvarado, "On bit-interleaved coded modulation with QAM constellations," Master's thesis, Chalmers Univ. of Technology, Sweden, 2008.
- [2] G. Caire, G. Taricco, and E. Biglieri, "Bit-interleaved coded modulation," *IEEE Trans. Inform. Theory*, vol. 44, no. 3, pp. 927–946, 1998.
- [3] G. Liva, E. Paolini, B. Matuz, S. Scalise, and M. Chiani, "Short turbo codes over high order fields," *IEEE Trans. Commun.*, vol. 61, no. 6, pp. 2201–2211, June 2013.
- [4] L. Dolecek, D. Divsalar, Y. Sun, and B. Amiri, "Non-binary protograph based LDPC codes: enumerators, analysis, and designs," *IEEE Trans. on Inf. Theory*, vol. 60, no. 7, pp. 3913–3941, July 2014.
- [5] G. Durisi, T. Koch, and P. Popovski, "Toward massive, ultrareliable, and low-latency wireless communication with short packets," *Proceedings of the IEEE*, vol. 104, no. 9, pp. 1711–1726, 2016.
- [6] G. Böcherer, F. Steiner, and P. Schulte, "Bandwidth efficient and rate-matched low-density parity-check coded modulation," *IEEE Trans. on Commun.*, vol. 63, no. 12, pp. 4651–4665, 2015.
- [7] J. J. Boutros, F. Jandel, and C. Méasson, "Probabilistic shaping and non-binary codes," in *IEEE Int. Symp. Inform. Th. (ISIT)*, Aachen, Germany, Jun 2017, pp. 2308–2312.
- [8] F. Steiner, G. Liva, and G. Böcherer, "Ultra-sparse non-binary LDPC codes for probabilistic amplitude shaping," in *IEEE Global Commun. Conf.*, Singapore, Dec 2017, pp. 1–5.
- [9] P. Robertson and T. Würz, "Bandwidth-efficient turbo trellis-coded modulation using punctured component codes," *IEEE J. Sel. Areas Commun.*, vol. 16, no. 2, pp. 206–218, Feb 1998.
- [10] T. Matsumine and H. Ochiai, "Capacity-approaching non-binary turbo codes: A hybrid design based on exit charts and union bounds," *IEEE Access*, vol. 6, pp. 70952–70963, 2018.
- [11] R. Klaimi, C. Abdel Nour, C. Douillard, and J. Farah, "Design of Low-complexity convolutional codes over GF(q)," in *IEEE Global Commun. Conf.*, Abu Dhabi, U.A.E., Dec. 2018.
- [12] R. Garzón-Bohórquez, R. Klaimi, C. Abdel Nour, and C. Douillard, "Mitigating correlation problems in turbo decoders," in *Intern. Symp. on Turbo Codes & Iterative Inf. Process.*, Hong Kong, China, Dec. 2018.
- [13] R. Klaimi, C. Abdel Nour, C. Douillard, and J. Farah, "Low-complexity decoders for non-binary turbo codes," in *Intern. Symp. on Turbo Codes & Iterative Inf. Process.*, Hong Kong, China, Dec. 2018.
- [14] C. Berrou, S. Vaton, M. Jezequel, and C. Douillard, "Computing the minimum distance of linear codes by the error impulse method," in *IEEE Global Commun. Conf.*, Taipei, Taiwan, Nov 2002, pp. 1017–1020.
- [15] S. Crozier, P. Guinand, and A. Hunt, "Estimating the minimum distance of turbo-codes using double and triple impulse methods," *IEEE Commun. Lett.*, vol. 9, no. 7, pp. 631–633, 2005.
- [16] S. Benedetto and E. Biglieri, *Principles of digital transmission: with wireless applications*. Springer Science & Business Media, 1999.
- [17] N. Wiberg, "Codes and Decoding on General Graphs," Ph.D. dissertation, Linköping University, Sweden, April 1996.
- [18] J. Hokfelt, O. Edfors, and T. Maseng, "Turbo codes: Correlated extrinsic information and its impact on iterative decoding performance," in *IEEE Veh. Techn. Conf.*, vol. 3, Houston, TX, USA, 1999, pp. 1871–1875.
- [19] C. Berrou, Y. Saouter, C. Douillard, S. Kerouédan, and M. Jézéquel, "Designing good permutations for turbo codes: towards a single model," in *IEEE Int. Conf. Commun.*, vol. 1, Paris, France, 2004, pp. 341–345.