Average Spectra for Ensembles of LDPC Codes and Applications

Irina E. Bocharova[1,2], **Boris D. Kudryashov**[1], **Vitaly Skachek**[2], **and Yauhen Yakimenka**[2]

[1] Department of Information Systems
St. Petersburg University of Information Technologies, Mechanics and Optics
St. Petersburg 197101, Russia
Email: irinaboc@ut.ee, kudryashov_boris@bk.ru

[2] Institute of Computer Science
University of Tartu
Tartu 50409, Estonia
Email: { vitaly, yauhen }@ut.ee

***Abstract*—The exact values of finite length average weight distributions for both binary ensembles and binary images of nonbinary ensembles of regular LDPC codes are computed. The exact average stopping set size distribution for the binary ensemble is also obtained. The computed spectra are applied in order to bound from above the average stopping redundancy of the ensemble of binary regular LDPC codes. The asymptotic typical normalized minimum distances for the binary image of the ensemble of nonbinary regular LDPC codes and the typical minimum stopping distances for the binary regular LDPC codes are also presented.**

I. Introduction

It is well known [1] that long binary (J, K)-LDPC codes approach the Gilbert-Varshamov (GV) bound if their column weight J is large enough. A detailed analysis of asymptotic behavior of average spectra, also known as weight distributions, for several ensembles of regular LDPC codes can be found in [2]. The average stopping set distributions for binary regular and irregular LDPC codes were studied in [3].

The analysis of the average weight distributions for binary images of nonbinary LDPC codes over extensions of the binary Galois field (called in the sequel *nonbinary LDPC codes*) is presented in [4], and their excellent performance is demonstrated. A recent overview of the known results in this area can be found in [5,6].

Regardless of that, the problem of computing the average spectra for ensembles of LDPC codes for finite lengths was studied in a large number of papers, and it is still considered as computationally complex. The main difficulty is, typically, related to finding coefficients of the corresponding weight generating functions. In this paper we introduce a low-complexity recurrent procedure to computing precise average weight distributions for binary and nonbinary regular LDPC codes, as well as average stopping set distributions [3] for binary regular LDPC codes. The obtained numerical results for short codes lead to some useful conclusions. The average spectra estimates could help in solving many interesting problems. For example, they are used in [7] in order to bound the error probability of the maximum-likelihood (ML) decoding and for computing asymptotic ML decoding thresholds on AWGN and BSC channels. In combination with stopping redundancy-based approach [8], the new results allow for estimating the potential efficiency of near ML decoding for short LDPC codes.

The asymptotic version of the new estimates yields a new bound on the normalized typical minimum distances of regular nonbinary LDPC codes. The estimates show that, unlike for the binary regular LDPC codes, the nonbinary regular LDPC codes with column weight $J = 2$ can approach the GV bound as the alphabet size increases.

This work is supported in part by the Norwegian-Estonian Research Cooperation Programme under the grant EMP133 and by the Estonian Research Council under the grant PUT405.

II. Generating functions of spectra for ensembles of linear codes

A. Finite length spectra via weight generating functions

In this section, we study the average weight distribution for different ensembles of binary linear $[n, k, d]$ codes, where n and k are the code length and dimension, respectively, and d denotes the minimum Hamming distance of the code. The weight distribution of a linear code from a random ensemble can be represented via its weight generating function

$$G_n(s) = \sum_{w=0}^{n} A_{n,w} s^w \ ,$$

where $A_{n,w}$ is the random variable representing the number of binary codewords of weight w and length n. Our goal is to find $\mathrm{E}\{A_{n,w}\}$, where $\mathrm{E}\{\cdot\}$ is the expected value over the code ensemble. In general, computing the coefficients $A_{n,w}$ is a rather difficult task. However, if a weight generating function can be expressed in terms of a degree of another weight generating function then the coefficients $A_{n,w}$ can be computed recursively. The next lemma (see e.g. [9]), presents a recurrent equation for coefficients $A_{n,w}$ in a case where $G_n(s)$ is a degree of another weight generating function.

Lemma 1: Let $f(s) = \sum_{l \geq 0} f_l s^l$ be a generating function. Then the coefficients in series expansion of the generating function $F_L(s) = (f(s))^L = \sum_{l \geq 0} F_{l,L} s^l$ satisfy the following recurrent equation

$$F_{l,L} = \begin{cases} f_l, & L = 1 \\ \sum_{i=0}^{l} f_i F_{l-i,L-1}, & L > 1 \end{cases} \ . \qquad (1)$$

Notice that Lemma 1 is also valid if $f(s)$ is a function of another generating function.

B. General linear codes

For completeness, we present the average spectrum for the ensemble of random linear codes determined by the equiprobable $r \times n$ parity-check matrices H, where $r = n - k$. The weight generating function of all binary sequences of length n is $G_n(s) = (1+s)^n$. Then the average spectrum coefficients are

$$\mathrm{E}\{A_{n,w}\} = \binom{n}{w} 2^{-r} , \tag{2}$$

where 2^{-r} is the probability that a binary sequence $\boldsymbol{x}$ of length n and weight w satisfies $\boldsymbol{x}H^{\mathrm{T}} = \mathbf{0}$.

It follows from Lemma 1 that coefficients $A_{n,w}$ can be computed recursively as

$$A_{n,w} = A_{n-1,w} + A_{n-1,w-1} .$$

C. Binary (J, K)-regular LDPC codes

1) Average weight distribution: We consider the Gallager ensemble of regular (J, K)-LDPC codes [1]. In this ensemble a random parity-check matrix H consists of the strips H_i of width $M = r/J$ rows each, $i = 1, 2, \ldots, J$. All strips are random column permutations of the strip where the jth row contains K ones in positions $(j-1)K+1, (j-1)K+2, \ldots, jK$, for $j = 1, 2, \ldots, n/K$.

The weight generating function of binary sequences of length n and weight w satisfying the equality

$$\boldsymbol{x}H_i^{\mathrm{T}} = \mathbf{0} \tag{3}$$

is the same for all strips $i \in \{1, 2, \ldots, J\}$, and it has the form [1]:

$$G(s) = \sum_{w=0}^{n} N_{n,w} s^w = \big(g(s)\big)^M , \tag{4}$$

where $N_{n,w}$ is the number of binary sequences $\boldsymbol{x}$ of weight w and length n satisfying the equality (3), and

$$g(s) = \sum_{i \text{ even}} \binom{K}{i} s^i = \frac{(1+s)^K + (1-s)^K}{2}$$

is the weight generating function of binary sequences satisfying the nonzero part of one parity-check equation. The probability that (3) is valid for a random $\boldsymbol{x}$ of length n and weight w is equal to

$$p(w) = \frac{N_{n,w}}{\binom{n}{w}} \tag{5}$$

for each of the strips. The average spectrum coefficients are given by

$$\mathrm{E}\{A_{n,w}\} = \binom{n}{w} \big(p(w)\big)^J = \binom{n}{w}^{1-J} N_{n,w}^J , \tag{6}$$

where $\big(p(w)\big)^J$ is the probability that $\boldsymbol{x}$ satisfies (3) for all $i = 1, \ldots, J$ simultaneously. Observe that by applying Lemma 1 the coefficients $N_{n,w}$ can be computed recursively.

2) Average stopping set distribution: A subset of symbol positions of an LDPC code is called a stopping set if there are no parity checks with one unknown involving these symbol positions. The smallest size of a stopping set d_{stop} is called the minimum stopping distance of the code. It depends on the parity-check matrix of the code, rather than on the code itself. It is shown in [10] that stopping sets play an important role in iterative decoding over binary erasure channel (BEC), and they cause failure of iterative decoding algorithms.

The average spectrum of stopping distances for the Gallager ensemble can be found similarly to the average weight spectrum. The weight generating function of binary sequences of length K and weight either zero or larger than or equal to 2 has the form [3]

$$g(s) = \sum_{w=0,2,3,\ldots,K} \binom{K}{w} s^w = (1+s)^K - Ks . \tag{7}$$

The stopping weight generating function for each of the strips can be obtained analogous to (4). By applying (5) and (6), we obtain the average stopping weight spectrum, where the coefficients could similarly be computed recursively by using Lemma 1.

D. Binary images of nonbinary LDPC codes

Consider the Gallager ensemble of q-ary LDPC codes, where $q = 2^m$, $m \geq 1$ is an integer. The weight generating function of q-ary sequences $\boldsymbol{x}$ of length n satisfying the nonzero part of one q-ary parity-check equation is given in [1] as

$$g(s) = \frac{(1+(q-1)s)^K + (q-1)(1-s)^K}{q} . \tag{8}$$

Each q-ary symbol can be represented as a binary sequence (image) of length m. It is easy to see that different representations of a finite field of characteristic two will lead to different generating functions of binary images for the same ensemble of nonbinary LDPC codes. Following the techniques in [11], we study an average binary weight spectrum for the ensemble of m-dimensional binary images. By assuming a binomial distribution of zeros and ones in the m-dimensional binary image of the q-ary symbol, we obtain the generating function of the average binary weights of a q-ary symbol in the form

$$\phi(s) = \frac{1}{q-1} \sum_{w=1}^{m} \binom{m}{w} s^w = \frac{(1+s)^m - 1}{q-1} . \tag{9}$$

Then, the average binary weight generating function for one strip is given by

$$G(s) = \big(g(\phi(s))\big)^M = \sum_{w=0}^{nm} N_{nm,w} s^w ,$$

where $N_{nm,w}$ denotes the average number of binary sequences $\boldsymbol{\beta}$ of weight w and of length nm satisfying $\boldsymbol{\beta}\mathcal{B}_i^{\mathrm{T}} = \mathbf{0}$. Here $\mathcal{B}_i$ denotes the average binary image of H_i. Then, analogous

to (6), we obtain the average binary weight enumerator of nonbinary regular LDPC code as

$$\mathrm{E}\{A_{nm,w}\} = \binom{nm}{w}\big(p(w)\big)^J = \binom{nm}{w}^{1-J} N_{nm,w}^J, \quad (10)$$

where $p(w) = \binom{nm}{w}^{-1} N_{nm,w}$. We proceed by computing $N_{nm,w}$ recursively.

The recursive procedure for computing the spectra is presented in [7] in more detail.

III. Numerical results and applications

A. Computational results

In Figures 1 and 2, we present the average minimum distance and the average minimum stopping distance, respectively, computed for the Gallager ensemble of rate $R = 1/3$ and $1/2$ binary regular (J, K)-LDPC codes of finite lengths (24...256). In Figure 3, the average minimum distance of the binary images for the ensemble of rate $1/3$ and $1/2$ nonbinary regular LDPC codes over $q = 2^m$, $m = 2$, are shown. The minimum distance (minimum stopping distance) is optimized over J. For each code length, an optimal value of J yielding the largest minimum distance (minimum stopping distance) is shown in the same figure. The optimized minimum distance (minimum stopping distance) is compared with the upper d_{U} and lower d_{L} bounds on the minimum distance of general linear codes [12] as well as with the lower GV bound on d for the ensemble of general linear codes

$$\sum_{w=0}^{d_{\mathrm{GV}}-1} \binom{n}{w} 2^{-r} < 1\,.$$

Notice that for binary (J, K)-regular LDPC codes, rate R is larger than $1-J/K$. In numerical calculations we did not take this fact into account, that is, the gap between bounds and the numerical calculations is less than it can be seen in Figures 1 and 2.

The numerical results lead to the following observations.

- Random regular LDPC codes, especially nonbinary LDPC codes over extensions of the binary Galois field, with optimized value of J have the minimum distance close to the minimum distance of the random linear codes.
- The gap between the minimum distances of the random LDPC codes and the random linear codes decreases as the code rate decreases.
- Although for the best LDPC codes, the stopping distance typically coincides with the minimum distance, for the random regular LDPC codes, the stopping distances are about two times smaller than their minimum distances.

In Table I, we present examples of short $(J, 2J)$-regular LDPC codes selected among 100 codes with the optimum values of J, which have the largest minimum distances. For comparison, in the same table, d_{L}, d_{GV}, and the ensemble average minimum distance $\hat{d}$ are presented.

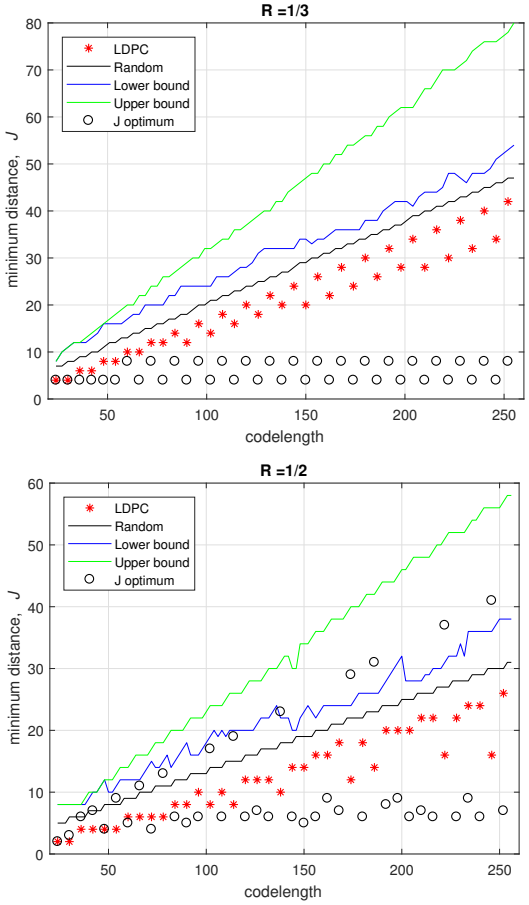

Fig. 1. Minimum distances for binary codes

TABLE I
Examples of codes from the Gallager $(J, 2J)$ ensemble

(n,k,d)	J	$\hat{d}$	d_{GV}	d_{L}	d_{stop}	$\hat{d}_{\mathrm{stop}}$	ρ	$\hat{\rho}$
(40,24,6)	5	4	5	7	6	4	73	20
(60,35,8)	6	6	7	10	6	6	289	76
(90,49,10)	5	8	11	14	8	8	1118	86

B. Asymptotic analysis

In this subsection, we apply the average weight distribution of binary images of the Gallager ensemble of nonbinary LDPC codes in order to analyze the asymptotic behavior of their normalized minimum distance. Let $\delta = w/(nm)$ denote the normalized weight. We follow the approach in [1, Ch. 5] to bound from above the average weight distribution (10)

$$\mathrm{E}\{A_{nm,w}\} \le \binom{nm}{w}^{1-J} \big(g(\phi(s))\big)^{MJ} s^{-wJ}\,.$$

By replacing s by e^{ρ}, we find a critical value δ, where the asymptotic exponent

$$\lim_{n\to\infty} \frac{\ln \mathrm{E}\{A_{nm,w}\}}{nm}$$
$$= \min_{\rho} \left\{ (1-J)h(\delta) + \frac{J}{Km} \ln\left(g(\phi(e^{\rho}))\right) - \rho\delta J \right\} \quad (11)$$

is equal to zero, $h(\delta) = -\delta\ln(\delta) - (1-\delta)\ln(\delta)$ is the binary entropy function in nats. The corresponding values δ represent-

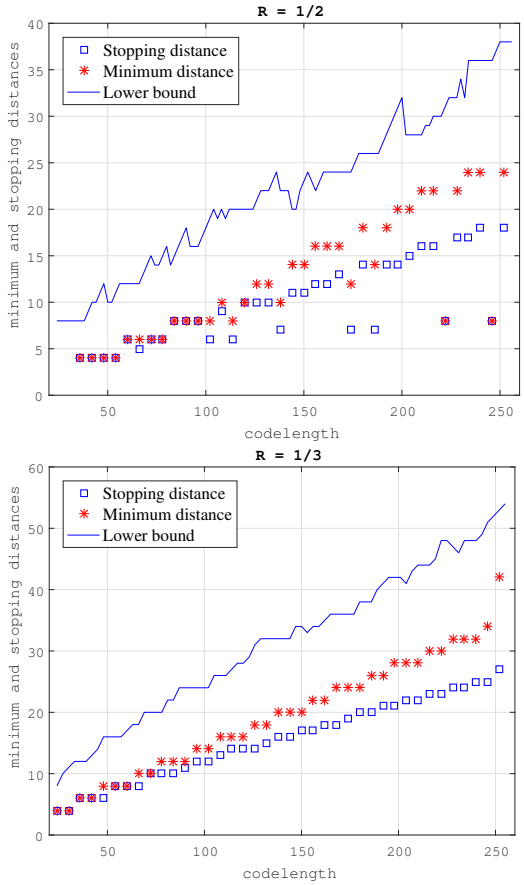

Fig. 2. Stopping distances for binary codes

TABLE II
NORMALIZED MINIMUM DISTANCES FOR BINARY AND NONBINARY LDPC CODE ENSEMBLES. NUMBERS IN PARENTHESES ARE TYPICAL ASYMPTOTIC NORMALIZED STOPPING DISTANCES.

m	J				
	2	3	4	5	10
1	0	0.0227	0.0627	0.0843	0.1083
		(0.0180)	(0.0454)	(0.0580)	(0.0619)
2	0	0.0398	0.0774	0.0943	0.1095
3	0	0.0559	0.0880	0.1007	0.1099
4	0.0018	0.0686	0.0952	0.1046	0.1100
8	0.0396	0.0938	0.1066	0.1094	0.1100
16	0.0688	0.1049	0.1097	0.1100	0.1100

ing typical normalized minimum distances for different m and J are shown in Table II. If instead of composition of (8) and (9) we use g from (7) and $m = 1$, we obtain asymptotic typical normalized stopping distances (see also Table II). It is easy to verify that if $m \to \infty$ then (11) tends to $h(\delta) - (1 - R)\ln 2$ for $J \geq 2$, i.e. δ tends to the GV bound. Notice that, unlike for binary LDPC codes, the normalized minimum distance of nonbinary LDPC codes differs from zero already for $J = 2$ if $m \geq 4$. This explains good performance of nonbinary LDPC codes with two nonzero elements in each column.

C. Stopping redundancy

In this subsection, we aim at modifying the parity-check matrix, such that the performance of the iterative decoders on the BEC becomes close to that of the ML-decoding. This can be achieved by adding redundant rows to a parity-check matrix of a code. It is noted in [8] that adding linearly dependent rows to the parity-check matrix of the code can increase its stopping distance. Moreover, after adding a sufficient number of redundant rows, it is always possible to make stopping distance equal to the minimum distance of the code. *Stopping redundancy* is defined as the minimum number of rows in a (redundant) parity-check matrix, such that there are no stopping sets of size smaller than the minimum distance. In order to calculate upper bounds on the stopping redundancy, the knowledge of the minimum distance of a given code is necessary.

There are several results on upper bounds of stopping redundancy ([8], [13]–[17]). To the best of our knowledge, the sharpest bounds for general codes are presented in [16,17]. To calculate these bounds, one needs to know n, k, d, and structural properties of the parity-check matrix.

Let us briefly describe (one of) the bounds in [17]. We construct a $\tau \times n$ matrix consisting of τ rows from the linear space spanned by the rows of H. If there are not more than u_i stopping sets of size i in this $\tau \times n$ matrix, then the stopping redundancy is upper-bounded as follows:

$$\rho \leq \tau + \min_{t \in \mathbb{N}} \{t + \kappa_t\} + \Delta\,, \tag{12}$$

where

$$\mathcal{D}_t = \sum_{i=d_{\text{stop}}}^{d-1} u_i \prod_{j=\tau+1}^{\tau+t} \left(1 - \frac{i \cdot 2^{r-i}}{2^r - j}\right)\,,$$

$$P_{t,j}(x) = \left\lfloor x \left(1 - \frac{(d-1)\cdot 2^{r-d+1}}{2^r - (\tau + t + j)}\right) \right\rfloor\,,$$

$$\kappa_t = \min\left\{j \in \mathbb{N} : P_{t,j}(P_{t,j-1}(\ldots P_{t,1}(\lfloor \mathcal{D}_t \rfloor)\ldots)) = 0\right\}\,,$$

and Δ is a term, which guarantees that the obtained extended parity-check matrix has a required rank $n - k$. If d_{stop} is not known one can use $d_{\text{stop}} = 1$. It does not make the bound much weaker because the most dominant term is usually u_{d-1}.

We can apply this bound to the (J, K)-regular codes from the Gallager ensemble. First we calculate the exact number of stopping sets of size i ($i = 1, 2, \ldots, d-1$) in the first $M \times n$ strip of the parity-check matrix H. In what follows, we term i-*set* any set of i coordinates of a codeword. An i-set is *covered* by a matrix if in this matrix this i-set does not form a stopping set.

Denote by $T_i(\sigma)$ a number of i-sets covered by each of the first σ rows of the strip (or any σ rows, due to symmetry). Then, it is easy to see that

$$T_i(\sigma) = \binom{n - K\sigma}{i - \sigma} K^\sigma\,.$$

Further, denote by A_i a number of i-sets covered by any of the M rows of the strip. From the principle of inclusion-exclusion we obtain:

$$A_i = \sum_{\sigma=1}^{M} (-1)^{\sigma-1} \binom{M}{\sigma} T_i(\sigma)\,.$$

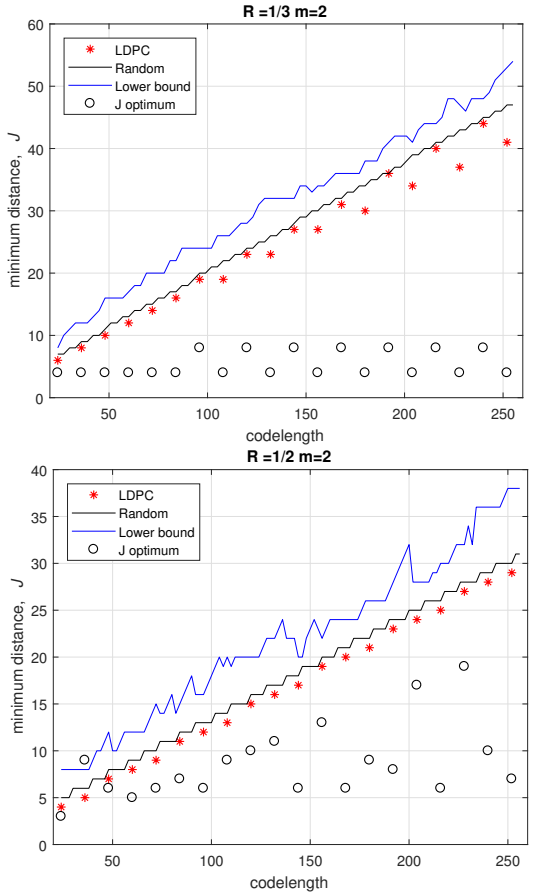

Fig. 3. Minimum distances for binary images of nonbinary codes

Since the number of all i-sets is $B_i = \binom{n}{i}$, there are $B_i - A_i$ stopping sets in the first strip. We substitute the values $\tau = M$, $u_i = B_i - A_i$, and $\Delta = n - k - \max(M, d-1)$ into (12) in order to compute ρ (see Table I). Note that this bound holds for *every* code in the ensemble.

However, we can also calculate the average number of stopping sets of size i in the canonical $(MJ) \times n$ parity-check matrix of a code. For an i-set, the probability *not* to be covered by one strip of the matrix is $1 - A_i/B_i$, and hence the probability not to be covered by any of the J strips is $(1 - A_i/B_i)^J$. Therefore, the average number of the stopping sets of size i in H is

$$\hat{u}_i = B_i \left(1 - \frac{A_i}{B_i}\right)^J .$$

For an $[n, k, d]$ code with this amount of the stopping sets and the stopping distance d_{stop}, we assign $\tau = MJ$, $u_i = \hat{u}_i$, $\Delta = 0$. Then, we obtain the bound $\hat{\rho}$ on the stopping redundancy (see Table I).

IV. Conclusions

In this paper the exact coefficients of the average distance and stopping distance spectra for ensembles of regular LDPC codes are computed. In particular, a recurrent low-complexity approach to computing the average distance spectrum coefficients for the average binary image of the ensemble of nonbinary regular LDPC codes is suggested.

The computed spectra show that binary images of nonbinary LDPC codes can have minimum distances close to the minimum distances of the best binary linear codes.

The computed minimum distance and stopping distance allow for analyzing the stopping redundancy of the LDPC ensembles. Based on the obtained estimates on the stopping redundancy one can expect that on the BEC channel any number of erasures up to the code minimum distance can be corrected with only moderate increase in the complexity compared to the complexity of the iterative decoding.

References

[1] R. G. Gallager, *Low-density parity-check codes*. M.I.T. Press: Cambridge, MA, 1963.

[2] S. Litsyn and V. Shevelev, "On ensembles of low-density parity-check codes: asymptotic distance distributions," *IEEE Trans. Inform. Theory*, vol. 48, no. 4, pp. 887–908, 2002.

[3] A. Orlitsky, K. Viswanathan, and J. Zhang, "Stopping set distribution of LDPC code ensembles," *IEEE Trans. Inform. Theory*, vol. 51, no. 3, pp. 929–953, 2005.

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

[5] L. Dolecek, D. Divsalar, Y. Sun, and B. Amiri, "Non-binary protograph-based LDPC codes: Enumerators, analysis, and designs," *IEEE Trans. Inform. Theory*, vol. 60, no. 7, pp. 3913–3941, 2014.

[6] K. Kasai, C. Poulliat, D. Declercq, and K. Sakaniwa, "Weight distributions of non-binary LDPC codes," *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, vol. 94, no. 4, pp. 1106–1115, 2011.

[7] I. E. Bocharova, B. D. Kudryashov, V. Skachek, and Y. Yakimenka, "Performance of ML decoding for ensembles of binary and nonbinary regular LDPC codes of finite lengths," in *Proc. IEEE Int. Symp. on Inf. Theory (ISIT)*, 2017, submitted.

[8] M. Schwartz and A. Vardy, "On the stopping distance and the stopping redundancy of codes," *IEEE Trans. Inform. Theory*, vol. 52, no. 3, pp. 922–932, 2006.

[9] D. V. Kruchinin and V. V. Kruchinin, "Application of a composition of generating functions for obtaining explicit formulas of polynomials," *Journal of Mathematical Analysis and Applications*, vol. 404, no. 1, pp. 161–171, 2013.

[10] C. Di, D. Proietti, I. E. Telatar, T. J. Richardson, and R. L. Urbanke, "Finite-length analysis of low-density parity-check codes on the binary erasure channel," *IEEE Trans. Inform. Theory*, vol. 48, no. 6, pp. 1570–1579, 2002.

[11] M. El-Khamy and R. J. McEliece, "Bounds on the average binary minimum distance and the maximum likelihood performance of Reed Solomon codes," in *42nd Allerton Conf. on Communication, Control and Computing*, 2004.

[12] M. Grassl, "Bounds on the minimum distance of linear codes and quantum codes," Online available at http://www.codetables.de, 2007, accessed on 2017-01-06.

[13] H. D. Hollmann and L. M. Tolhuizen, "On parity-check collections for iterative erasure decoding that correct all correctable erasure patterns of a given size," *IEEE Trans. Inform. Theory*, vol. 53, no. 2, pp. 823–828, 2007.

[14] T. Hehn, O. Milenkovic, S. Laendner, and J. B. Huber, "Permutation decoding and the stopping redundancy hierarchy of cyclic and extended cyclic codes," *IEEE Trans. Inform. Theory*, vol. 54, no. 12, pp. 5308–5331, 2008.

[15] J. Han and P. H. Siegel, "Improved upper bounds on stopping redundancy," *IEEE Trans. Inform. Theory*, vol. 53, no. 1, pp. 90–104, 2007.

[16] J. Han, P. H. Siegel, and A. Vardy, "Improved probabilistic bounds on stopping redundancy," *IEEE Trans. Inform. Theory*, vol. 54, no. 4, pp. 1749–1753, 2008.

[17] Y. Yakimenka and V. Skachek, "Refined upper bounds on stopping redundancy of binary linear codes," in *IEEE Information Theory Workshop (ITW)*, 2015, pp. 1–5.