Eleventh International Workshop on Algebraic and Combinatorial Coding Theory
June 16-22, 2008, Pamporovo, Bulgaria *pp. 338-347*

On the erasure-correcting capabilities of low-complexity decoded LDPC codes with constituent Hamming codes

Victor Zyablov zyablov@iitp.ru
Institute for Information Transmission Problems, Russian Academy of Sciences,
Moscow 101447, RUSSIA

Maja Lončar maja@eit.lth.se
Rolf Johannesson rolf@eit.lth.se
Dept. of Electrical and Information Technology, Lund University,
P. O. Box 118, SE-22100 Lund, SWEDEN

Pavel Rybin* prybin@iitp.ru
Institute for Information Transmission Problems, Russian Academy of Sciences,
Moscow 101447, RUSSIA

Abstract. Low-density parity-check (LDPC) codes can be constructed using constituent block codes other than single parity-check (SPC) codes. This paper considers random LDPC codes with constituent Hamming codes and investigates their asymptotic performance over the binary erasure channel. It is shown that there exist Hamming code-based LDPC codes which, when decoded with a low-complexity iterative algorithm, are capable of correcting any erasure pattern with a number of erasures that grows linearly with the code length. The number of decoding iterations, required to correct the erasures, is a logarithmic function of the code length. The fraction of correctable erasures is computed numerically for various choices of code parameters.

1 Introduction

Gallager's low-density parity-check (LDPC) codes [1] are characterized by a sparse parity-check matrix whose rows specify single parity-check (SPC) codes over small subsets of the code symbols. LDPC codes can be represented by a bipartite Tanner graph [2], whose two disjoint sets of vertices, referred to as the variable nodes and the constraint nodes, correspond to code symbols and SPC constraints, respectively. The adjacency matrix of such a bipartite

This work was supported in part by the Royal Swedish Academy of Sciences in cooperation with the Russian Academy of Sciences, in part by the Swedish Research Council under Grant 621-2007-6281, and in part by the Program of fundamentals scientific researches of DITCS RAS (OITVS RAN), Project no. 5.1.

graph coincides with the code's parity-check matrix $\boldsymbol{H}$; an element $(\boldsymbol{H})_{ij} = 1$ indicates that the jth code symbol participates in the ith SPC code, that is, there is a branch connecting the jth variable node with the ith constraint node. For regular LDPC codes, the corresponding graph is regular: all the variable nodes have the same degree, equal to the number of ones in the each column of $\boldsymbol{H}$, and all the constraint nodes have the degree equal to the number of ones in each row of $\boldsymbol{H}$ (which is the length of the corresponding SPC code).

Alternative constructions of LDPC codes can be obtained by 'replacing' the SPC codes in the code's Tanner graph with different constituent block codes of length equal to the constraint-node degree. The so-obtained LDPC codes are often referred to as the generalized LDPC codes, *cf., e.g.,* [3], [4]. Starting from the sparse adjacency matrix of the underlying Tanner graph, the parity-check matrix of such an LDPC code is obtained by replacing every 1 in the graph's adjacency matrix with a column of the constituent code's parity-check matrix, and every 0 with an all-zero column.

This paper focuses on LDPC codes with constituent Hamming codes and investigates their performance when communicating over the binary erasure channel (BEC). The erasure-correcting capabilities of Gallager's LDPC codes for the BEC were studied in [5], where it was shown that there exist LDPC codes capable of correcting a portion of erasures that grows linearly with the code length n, with decoding complexity $\mathcal{O}(n \log n)$. Hamming code-based LDPC (H-LDPC) codes were first studied in [3]; their distance properties and iterative soft-decision decoding for the AWGN channel were further investigated in [6] and [7]. Recently, it was shown in [8] that the ensemble of H-LDPC codes contains codes with a minimum distance that asymptotically almost meets the Varshamov-Gilbert bound.

In this work, we build upon the results of [5] and we investigate the asymptotic erasure-correcting capabilities of random H-LDPC codes, when the code length n grows to infinity. We will consider a simple iterative decoder whose complexity is $\mathcal{O}(n \log n)$, and prove that there exist H-LDPC codes for which such a decoder corrects any erasure pattern with a number of erasures growing linearly with the code length. The paper is organized as follows: ensembles of H-LDPC codes and their properties are introduced in Section 2. The decoding algorithm is presented in Section 3. The main result is presented in Section 4 and supported by numerical examples. Section 5 summarizes and concludes the paper.

2 Construction and Properties of H-LDPC Codes

For any integer $m \geq 2$, there exists a Hamming code of length $n_0 = 2^m - 1$, dimension $k_0 = n_0 - m$, minimum distance $d_0 = 3$, and rate $R_0 = 1 - m/n_0$. The parity-check matrix $\boldsymbol{H}_0$ of an (n_0, k_0, d_0) Hamming code is an $m \times n_0$ matrix whose columns are all the distinct nonzero binary m-tuples. When a Hamming

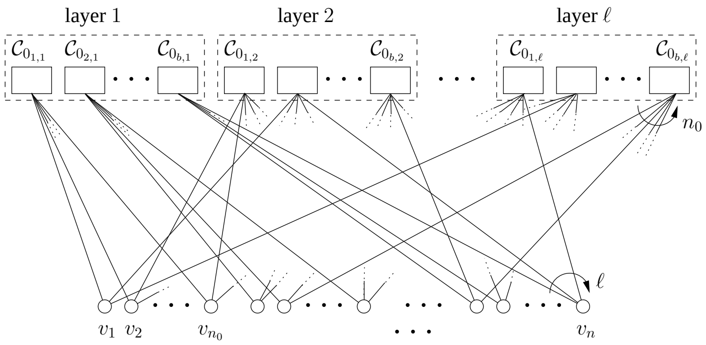

Figure 1: Tanner graph of an H-LDPC code with parity-check matrix given by (2).

code is used for communication over a BEC, it is guaranteed by the code's minimum distance that any erasure pattern with $d_0 - 1 = 2$ or fewer erasures will be corrected. Furthermore, it is also possible to correct some erasure patterns with d_0 or more (up to m) erasures, as will be discussed in detail in Section 3.

We consider H-LDPC codes whose bipartite Tanner graph is regular, with the same Hamming code associated with each constraint node. In order to construct a parity-check matrix of such an LDPC code, we start from a $bm \times bn_0$ block-diagonal matrix $\boldsymbol{H}_{\mathrm{b}}$ with the b constituent parity-check matrices $\boldsymbol{H}_0$ on the main diagonal, that is,

$$\boldsymbol{H}_{\mathrm{b}} = \begin{pmatrix} \boldsymbol{H}_0 & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{H}_0 & \mathbf{0} & \cdots & \mathbf{0} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \boldsymbol{H}_0 \end{pmatrix} \tag{1}$$

where b is very large. Let $\pi(\boldsymbol{H}_{\mathrm{b}})$ denote a random column permutation of $\boldsymbol{H}_{\mathrm{b}}$. Then the matrix constructed using $\ell \geq 2$ such permutations as *layers*,

$$\boldsymbol{H} = \begin{pmatrix} \boldsymbol{H}_1 \\ \boldsymbol{H}_2 \\ \vdots \\ \boldsymbol{H}_\ell \end{pmatrix} = \begin{pmatrix} \pi_1(\boldsymbol{H}_{\mathrm{b}}) \\ \pi_2(\boldsymbol{H}_{\mathrm{b}}) \\ \vdots \\ \pi_\ell(\boldsymbol{H}_{\mathrm{b}}) \end{pmatrix} \tag{2}$$

is a sparse $\ell bm \times bn_0$ parity-check matrix of a Hamming code-based LDPC code of length $n = bn_0$, where $n \gg n_0$. For a given constituent Hamming code with parity-check matrix $\boldsymbol{H}_0$, by sampling independently the permutations π_l, $l = 1, 2, ..., \ell$, which are all equiprobable, we obtain the ensemble of H-LDPC codes, which will be denoted by $\mathscr{C}(n_0, \ell, b)$. The rate of a code $\mathcal{C} \in \mathscr{C}(n_0, \ell, b)$

is lower-bounded by [2]

$$R \geq 1 - \frac{\ell b(n_0 - k_0)}{n} = 1 - \ell(1 - R_0) \tag{3}$$

with equality iff the matrix $\boldsymbol{H}$ has full rank. Since the rate must be positive, (3) implies a restriction on the rate of the constituent codes, namely,

$$R_0 > 1 - \frac{1}{\ell}$$

that is, the more layers there are, the higher the rate of the constituent codes must be.

Note that by replacing the Hamming constituent code with the (n_0, n_0–1, 2) SPC code, that is, by setting $\boldsymbol{H}_0 = (1\ 1\ ...\ 1)$, the construction defined by (2) reduces to Gallager's construction [1] of the (ℓ, n_0)-regular LDPC matrices.

An H-LDPC code from the ensemble $\mathscr{C}(n_0, \ell, b)$ contains ℓb constituent Hamming codes; b in each layer. The jth constituent code in the lth layer is denoted by $\mathcal{C}_{0j,l}$, and its parity-check matrix by $\boldsymbol{H}_{0j,l}$, $j = 1, 2, ..., b$, $l = 1, 2, ..., \ell$ (all matrices $\boldsymbol{H}_{0j,l}$ are equal up to column permutations). The Tanner graph representation of such an H-LDPC code is illustrated in Figure 1. There are $n = bn_0$ variable nodes and ℓb constraint nodes. The graph is regular, with the variable-node degree equal to ℓ, and the constraint-node degree equal to n_0. Each variable node is connected to exactly one constraint node in each layer.

3 Decoding Algorithm

Let $\boldsymbol{v}$ be a codeword of an H-LDPC code, transmitted over a BEC with the erasure probability δ, and let $\boldsymbol{r}$ denote the received sequence. The number of erasures in the received sequence $\boldsymbol{r}$ is denoted by W. When the code length is large, $n \to \infty$, the fraction of the erased symbols, $\omega = W/n$, converges to the erasure probability δ of the BEC, $\omega \to \delta$.

Consider an iterative erasure-correcting algorithm $\mathscr{A}$, with two variants denoted by $\mathscr{A}_1$ and $\mathscr{A}_2$, whose iterations i, $i = 1, 2, ..., i_{\max}$, consist of the following steps:

(1) For the tentative sequence $\boldsymbol{r}^{(i)}$, where $\boldsymbol{r}^{(1)}$ is the received sequence $\boldsymbol{r}$, select constituent codes $\mathcal{C}_{0j,l}$ with $\tau_{j,l}$ erasures, $j = 1, 2, ..., b$, $l = 1, 2, ..., \ell$, such that:

   (a) $\tau_{j,l} < d_0$ for algorithm $\mathscr{A}_1$

   (b) $\tau_{j,l} \leq m$ for algorithm $\mathscr{A}_2$

(2) Assuming 0s on the erased positions, compute the syndromes $\boldsymbol{s}_{j,l}$ for the constituent codes selected in the previous step.

(3) For each selected constituent code $\mathcal{C}_{0j,l}$, construct the $m \times \tau_{j,l}$ matrix $\boldsymbol{M}_{j,l}$ whose columns are the $\tau_{j,l}$ columns of $\boldsymbol{H}_{0j,l}$ which correspond to the erased positions. Note that, in general, $\text{rank}(\boldsymbol{M}_{j,l}) \leq \tau_{j,l} \leq m$.

Let $\boldsymbol{x}_{j,l}$ denote the $\tau_{j,l}$-tuple of the unknown (erased) transmitted symbols. These symbols can be recovered by solving the equation system

$$\boldsymbol{x}_{j,l}\boldsymbol{M}_{j,l}^{\mathrm{T}} = \boldsymbol{s}_{j,l}. \tag{4}$$

Clearly, the equation system has a unique solution iff the matrix $\boldsymbol{M}_{j,l}$ has full rank, that is, iff $\text{rank}(\boldsymbol{M}_{j,l}) = \tau_{j,l}$. Then the erasure pattern is correctable.

(4) For every constituent code affected by a correctable erasure pattern find the erased tuple $\boldsymbol{x}_{j,l}$ by solving (4). Replace the erasures in $\boldsymbol{r}^{(i)}$ with the so-found code symbols. This yields the updated sequence $\boldsymbol{r}^{(i+1)}$.

As mentioned earlier, when using algorithm $\mathscr{A}_2$, only some erasure patterns with more than d_0 erasures, which affect constituent codes, are correctable. The following lemma allows us to determine the exact number of the correctable patterns:

Lemma 1 *Let $\boldsymbol{M}$ be an $m \times \tau$ matrix whose columns are equal to τ columns of a parity-check matrix $\boldsymbol{H}_0$ of a Hamming code of length n_0, where $1 \leq \tau \leq m$ and $m = \log_2(n_0 + 1)$. Then the number of matrices $\boldsymbol{M}$ that have full rank,* $\text{rank}(\boldsymbol{M}) = \tau$, *is equal to*

$$M(\tau, m) = \frac{1}{\tau!}\prod_{i=0}^{\tau-1}(2^m - 2^i). \tag{5}$$

Proof: The columns of the parity-check matrix $\boldsymbol{H}_0$ of the Hamming code of length $n_0 = 2^m - 1$ are all nonzero binary m-tuples, which span the m-dimensional binary space. Thus, clearly, the number of matrices $\boldsymbol{M}$, constructed from τ columns of $\boldsymbol{H}_0$, which have $\text{rank}(\boldsymbol{M}) = \tau$, is equal to the number of different bases of τ-dimensional subspaces of the m-dimensional space. Let $\{\boldsymbol{b}_1, \boldsymbol{b}_2, ..., \boldsymbol{b}_\tau\}$ denote the set of basis vectors of a τ-dimensional subspace. The number of such sets is determined in the following way:

- First, select the vector $\boldsymbol{b}_1$ as any of the $2^m - 1$ nonzero binary m-tuples;
- Select the nonzero vector $\boldsymbol{b}_2$ different from $\boldsymbol{b}_1$, that is, $\boldsymbol{b}_2 \neq c_1\boldsymbol{b}_1$, $c_1 \in \{0, 1\}$. There are $2^m - 2$ choices.
- For $i = 3, 4, ..., \tau$, select the nonzero vector $\boldsymbol{b}_i$ such that it is not equal to a linear combination of the previously chosen $i - 1$ basis vectors, that is, $\boldsymbol{b}_i \neq c_1\boldsymbol{b}_1 + c_2\boldsymbol{b}_2 + \cdots + c_{i-1}\boldsymbol{b}_{i-1}$, where $c_1, c_2, ..., c_{i-1} \in \{0, 1\}$. Clearly, there are $2^m - 2^{i-1}$ choices for $\boldsymbol{b}_i$, $i = 3, 4, ..., \tau$.

Finally, note that the ordering of the basis vectors in the set $\{\boldsymbol{b}_1, \boldsymbol{b}_2, ..., \boldsymbol{b}_\tau\}$ is irrelevant. Thus, the total number of bases of τ-dimensional subspaces of the m-dimensional space is

$$M(\tau, m) = \frac{\prod_{i=0}^{\tau-1} (2^m - 2^i)}{\tau!}$$

■

Clearly, for a constituent Hamming code, an erasure pattern with τ erasures is correctable when the matrix $\boldsymbol{M}$, constructed from the τ columns of $\boldsymbol{H}_0$ corresponding to the erased positions, has the rank equal to τ. Thus, we have the following

Corollary 1 *The number of erasure patterns of length $n_0 = 2^m - 1$, with $\tau \le m$ erasures, which are correctable by a Hamming code of length n_0, is equal to $M(\tau, m)$ given by (5).*

Thus, the generating function for the number of correctable erasure patterns can be defined as

$$g_1(s, n_0) = \sum_{\tau=1}^{m} M(\tau, m) s^\tau = \sum_{\tau=1}^{m} \frac{\prod_{i=0}^{\tau-1} (2^m - 2^i)}{\tau!} s^\tau .$$

Note that the function $g_1(s, n_0)$ can be written as

$$g_1(s, n_0) = \widetilde{g}_1(s, n_0) + \sum_{\tau=3}^{m} \frac{\prod_{i=0}^{\tau-1} (2^m - 2^i)}{\tau!} s^\tau \tag{6}$$

where

$$\widetilde{g}_1(s, n_0) = \binom{n_0}{1} s + \binom{n_0}{2} s^2 \tag{7}$$

is the generating function of all erasure patterns with less than $d_0 = 3$ erasures, which are all correctable.

For a given erasure pattern of length n with W erasures, let a denote the number of constituent codes which are affected by correctable erasures. In general, $a = \alpha W \ell$, where $\alpha \le 1$. In the algorithm $\mathscr{A}$ it is assumed that the erasure pattern is such that there is at least one constituent code for which the erasures that affect it are correctable. In other words, we assume that $\alpha > 0$. Then, during the first iteration of the algorithm $\mathscr{A}$, all correctable erasures will be corrected, while the uncorrectable ones will result in the decoding failure. Hence, the new erasure pattern, after one decoding iteration, has fewer erasures than the initial erasure pattern. Clearly, if in each of the following iterations, the number of codes with correctable erasures is larger than zero, then the total number of erasures in $\boldsymbol{r}^{(i)}$ will decrease with the iteration number i and the algorithm $\mathscr{A}$ recovers the transmitted codeword, *i.e.*, $\boldsymbol{r}^{(i_{\max})} = \boldsymbol{v}$. Then, we can state the following

Lemma 2 *For any H-LDPC code from the ensemble $\mathscr{C}(n_0, \ell, b)$, if an erasure pattern is such that in each iteration of the algorithm $\mathscr{A}$ the number of constituent codes affected by correctable erasures is larger than zero, then the algorithm $\mathscr{A}$ recovers the transmitted codeword after $\mathcal{O}(\log n)$ iterations, where $n = bn_0$ is the code length.*

Proof: Let ε denote a lower bound on the fraction of erasures that are recovered in each iteration, $0 < \varepsilon < 1$. Then, after x iterations, the number of remaining erasures is at most $\omega n(1-\varepsilon)^x$. The final decoding iteration $i_{\max}$ is reached when

$$\omega n(1-\varepsilon)^{i_{\max}} < 1$$

that is,

$$\log(\omega n) + i_{\max} \log(1-\varepsilon) < 0$$

which yields

$$i_{\max} < \frac{1}{\log\left(\frac{1}{1-\varepsilon}\right)} \log(\omega n). \tag{8}$$

Thus, the number of iterations is a logarithmic function of the code length. ■

The complexity of each iteration of the algorithm $\mathscr{A}$ is proportional to the code length n. Thus, according to Lemma 2, the overall decoding complexity is $\mathcal{O}(n \log n)$.

4 Asymptotic Performance

As shown in the previous section, the iterative algorithm $\mathscr{A}$ corrects any erasure pattern with W or fewer erasures, if in each iteration $\alpha > 0$. The following theorem allows us to confirm the existence of H-LDPC codes for which this condition is fulfilled.

Theorem 1 *In the ensemble $\mathscr{C}(n_0, \ell, b)$ of H-LDPC codes, there exist codes (with probability p, where $\lim_{n\to\infty} p = 1$), which can correct any erasure pattern with up to $\omega_\alpha n$ erasures, with decoding complexity $\mathcal{O}(n \log n)$. The value ω_α is the largest root of the equation*

$$h(\omega) - \ell F(\alpha, \omega, n_0) = 0 \tag{9}$$

where $h(\omega)$ is the binary entropy function, $h(\omega) = -\omega \log_2 \omega - (1-\omega)\log_2(1-\omega)$, and the function $F(\alpha, \omega, n_0)$ is given by

$$F(\alpha,\omega,n_0) \triangleq h(\omega) - \frac{1}{n_0} h(\alpha\omega n_0) + \max\Bigg\{\omega \log_2 s - \\ -\frac{1}{n_0}\log_2(g_0(s,n_0)) - \alpha\omega\log_2\left(\frac{g_1(s,n_0)}{g_0(s,n_0)}\right)\Bigg\} \quad (10)$$

where $\alpha > 0$ and the maximization is performed over all s such that

$$\frac{\alpha\omega n_0}{1-\alpha\omega n_0} \le \frac{g_1(s,n_0)}{g_0(s,n_0)}.$$

The function $g_1(s,n_0)$ is the generating function of all the erasure patterns that are correctable by the constituent Hamming codes. It is equal to (7) when the constituent codes correct less than d_0 erasures (algorithm $\mathscr{A}_1$), or equal to (6) when the constituent codes correct up to m erasures (algorithm $\mathscr{A}_2$). $g_0(s,n_0)$ is the generating function of the uncorrectable erasure patterns and it equals

$$g_0(s,n_0) = (1+s)^{n_0} - g_1(s,n_0).$$

The proof of Theorem 1 is omitted here for brevity.

Theorem 1 allows us to compute numerically the fraction of the correctable erasures, ω_α, for several choices of code parameters. The computations confirm the existence of codes with a nonvanishing ω_α. First, we consider code ensembles of rates close to 1/2. Figure 2 illustrates the values of ω_α computed with $\alpha = 10^{-4}$ for algorithms $\mathscr{A}_1$ and $\mathscr{A}_2$, for several code ensembles of rates $R \approx 1/2$. Using the algorithm $\mathscr{A}_2$ up to 3.5 times more erasures can be corrected than with the algorithm $\mathscr{A}_1$. For both algorithms, with increasing n_0 (and, in order to keep the rate fixed, also with increasing ℓ) the value of ω_α increases only up to a certain point, $n_0 = 127$ for $\mathscr{A}_1$ and $n_0 = 63$ for $\mathscr{A}_2$, where it reaches its maximum. With further increase of n_0 and ℓ, ω_α decays quickly.

Next we consider code ensembles $\mathscr{C}(n_0, \ell, b)$ of different rates, $R \approx \frac{1}{4}$, $R \approx \frac{1}{2}$, and $R \approx \frac{3}{4}$, decoded with the algorithm $\mathscr{A}_2$. Figure 3 illustrates the values ω_α obtained with $\alpha = 10^{-4}$ for several code ensembles of different code rates. We have found a nonvanishing ω_α for different code lengths and rates. With increasing R, the maximum value of ω_α decreases and moves towards longer constituent codes: $n_0 = 31$ for $R \approx \frac{1}{4}$, $n_0 = 63$ for $R \approx \frac{1}{2}$ and $n_0 = 127$ for $R \approx \frac{1}{4}$.

Note that all the code ensembles considered in Figures 2 and 3 have minimum distances that almost meet the Varshamov-Gilbert bound, as shown in [8].

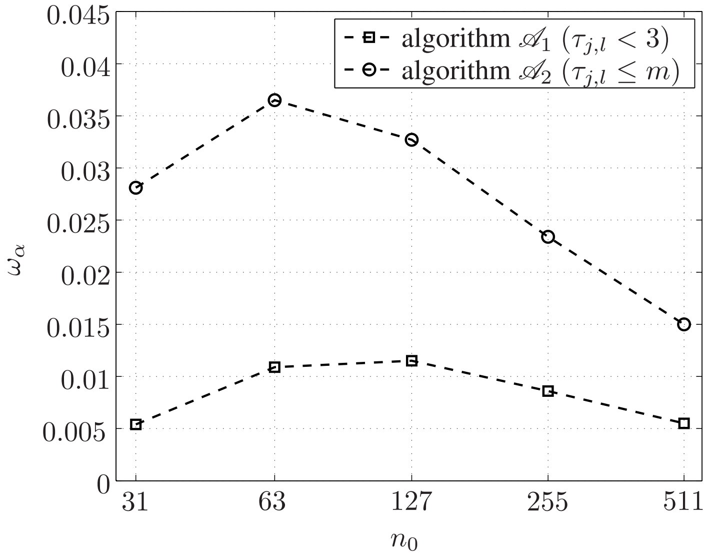

Figure 2: Values of ω_α computed according to Theorem 1 with $\alpha = 10^{-4}$ for decoding algorithms $\mathscr{A}_1$ and $\mathscr{A}_2$, for several code ensembles of rates $R \approx 1/2$.

5 Summary

We have investigated the asymptotic erasure-correcting capabilities of random LDPC codes with constituent Hamming codes, used over the binary erasure channel. A simple iterative decoding algorithm was considered, which can recover the transmitted codeword after $\mathcal{O}(\log n)$ iterations, where n is the code length. It was shown that there exist H-LDPC codes which, when decoded with such an algorithm, are capable of correcting a number of erasures that grows linearly with the code length n. The maximum fraction of correctable erasures was computed numerically for several code ensembles with different code rates and constituent-code lengths.

References

[1] R. G. Gallager, *Low-Density Parity-Check Codes*, MIT Press, Cambridge, MA, USA, 1963.

[2] M. Tanner, A recursive approach to low complexity codes, *IEEE Trans. Inform. Theory* 27, 1981, 533-547.

[3] M. Lentmaier, K. Zigangirov, Iterative decoding of generalized low-density parity-check codes, *Proc. IEEE Int. Symp. Inf. Theory*, Cambridge, MA, USA, Aug. 1998.

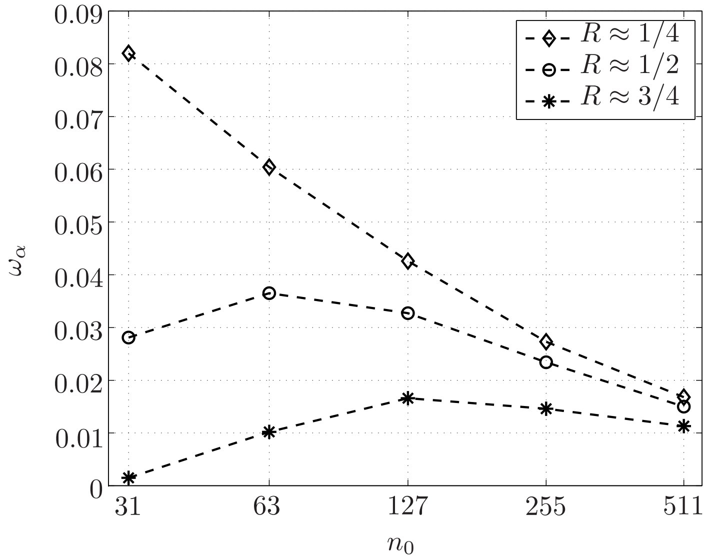

Figure 3: Values of ω_α computed according to Theorem 1 with $\alpha = 10^{-4}$ for the decoding algorithm $\mathscr{A}_2$, for several code ensembles of different rates.

[4] N. Miladinović, M. Fossorier, Generalized LDPC codes with Reed-Solomon and BCH codes as component codes for binary channels, *Proc. IEEE Global Comm. Conf. (GLOBECOM)*, St. Louis, MO, USA, 2005.

[5] V. V. Zyablov, M. S. Pinsker, Decoding complexity of low-density codes for erasure channel, *Probl. Inform. Transm.* 10, 1974, 15-28.

[6] M. Lentmaier., K. Zigangirov, On generalized low-density parity-check codes based on Hamming component codes, *IEEE Commun. Lett.* 3, 1999, 248-250.

[7] J. Boutros, O. Pothier, G. Zémor, Generalized low density (Tanner) codes, *Proc. IEEE Int. Comm. Conf.*, Vancouver, Canada, 1, 1999, 441-445.

[8] S. Stiglmayr, V. V. Zyablov, Asymptotically good low-density codes based on Hamming codes, *Proc. XI Int. Symp. Problems of Redundancy in Inform. and Control Systems, available online at http://www.k36.org/redundancy2007/proceedings.php*, Saint Petersburg, Russia, 2007, 98-103.