On the Erasure Correction Capability of LDPC Codes

D. K. Zigangirov Institute for Information Transmission Problems Moscow, B. Karetny per. 19, Russia Email: zig@iitp.ru

Abstract— The erasure correcting capability of LDPC codes with parity-check matrices composed of permutation matrices is considered. A new lower bound for number of the iteratively correcting erasures for the Zyablov-Pinsker algorithm is given.

I. INTRODUCTION

Iterative decoding of LDPC codes for a binary erasure channel (BEC) was first studied by Zyablov and Pinsker [1]. They suggested a simple iterative algorithm for erasures correction and proved a lower bound on the number of corrected erasures. Many researchers have studied iterative decoding of LDPC codes for the erasure channel (see, for example, [2],[4]) and obtaind interesting results in this area.

Zyablov and Pinsker have studied a special random ensemble of irregular LDPC codes. In this paper, we consider the same algorithm of erasures correction as in [1], but a different ensemble of LDPC codes. We study the ensemble of regular LDPC codes with parity-check matrices composed of permutation matrices. This ensemble was introduced in [3] where a lower bound on the minimum distance of LDPC codes was derived , analogously to the Gallager bound on the minimum distance of conventional regular LDPC codes [5].

A combinatorial approach which differs from both Zyablov-Pinsker and Gallager approaches is used to derive of the bound. Our bound is stronger than the Zyablov-Pinsker bound for low rates and weaker for larger rates.

II. BASIC IDEA OF BOUND PROOF

Let $\mathbf{v} = (v_0, v_1, ..., v_{N-1}), v_n \in \{0, 1\}, n = 0, 1, ..., N - 1$, be a codeword of length n. It should satisfy the equality $\mathbf{v}\mathbf{H}^{\mathsf{T}} = \mathbf{0}$ where \mathbf{H}^{T} is transposed $N \times L$ parity-check matrix. This equality assigns L parity-check equations defining the codeword \mathbf{v} .

Suppose a codeword **v** is transmitted over a BEC. Let $\tilde{\mathbf{v}} = (\tilde{v}_0, \tilde{v}_1, ..., \tilde{v}_{N-1})$ be the received codeword such that $\tilde{v}_n \in \{0, 1, e\}$ where *e* denotes the erasure symbol. We call the number of erased symbols *e*-weight.

Suppose that the number of erased symbols in a word is equal to t and that there exists at least one parity-check equation with exactly one erased symbol. Since this symbol is equal to sum of the other symbols in this equation, it can be restored. Then, t - 1 erased symbols is left in the codeword and we can look for another parity-check equation with exactly one erased symbol. If such an equation exists, we restore one K. Sh. Zigangirov Institute for Information Transmission Problems Moscow, B. Karetny per. 19, Russia Email: kamil@it.lth.se

more erased symbol etc. The Zyablov-Pinsker algorithm is based on the same idea. The decoder looks over the paritycheck equations until it finds an equation having exactly one erased symbol. Then it restore the erased symbol and repeats procedure.

Suppose there exists an LDPC code such that for any t, $0 < t < t_0$, and for any combination of t erasures there exists a parity-check equation having exactly one erased symbol. Obviously the Zyablov-Pinsker decoder corrects all combinations of erasures which number does not exceed t_0 . If all parity-check equations either does not include erased symbols or include two or more erased symbols, the corresponding combinations of erasures are called a *stop set*.

We used this principle for derivation of our lower bound on the number of iteratively correctable erasures. In fact, consider an ensemble of LDPC codes. Let us upperbound the mathematical expectation of the number of stop sets with *e*-weight *t*. If for any $t < t_0 = \alpha N$ this expectation is upperbounded by a function exponentially decreasing with block length *N*, then there exists an LDPC code of sufficiently large block length *N* such that all combinations of less than t_0 erasures are iteratively correctable.

We have considered the ensemble $C_{(KM,JK)}$, introduced in [3]. Parity-check matrices of codes in the ensemble are composed of JK permutation $M \times M$ matrices \mathbf{P}_{jk} :

$$\mathbf{H} = (\mathbf{P}_{jk}), \ j = 1, 2, \dots, J, \ k = 1, 2, \dots, K.$$
(1)

In symbol representation, these parity-check matrices are of size $L \times N$ where L = JM, N = KM. We assume that matrices \mathbf{P}_{jk} are chosen in the ensemble $\mathcal{C}_{(KM,JK)}$ independently and take on any of the M! values with equal probability.

III. MAIN RESULT

Theorem 1: If $J \geq 3$, almost all codes of the ensemble $C_{(KM,JK)}$ iteratively correct (with complexity $O(N \log N)$, N is the block length) all combinations of t erasures, $t \leq t_0$, where

$$t_0 = \min \{ \alpha KM, 2M \} + o(M).$$
 (2)



Fig. 1. Lower bound α on the fraction of iteratively correctable erasures and Zyablov-Pinsker lower bound α_0 on the fraction of iteratively correctable erasures as functions of R^* for K = 5. For comparison the lower bound [3] ρ^* on the normalized minimum distance of the code ensemble $C_{(KM,JK)}$ is presented.

Here $o(M)/M \to 0$ if $M \to \infty$, α is the smallest $\rho \in [0, 1/2]$ such that

$$F_{J,K}(\rho,\lambda) \stackrel{\Delta}{=} \frac{1}{K} f(\lambda) - \lambda\rho - \frac{J-1}{J} H(\rho) < 0$$
(3)

for some $\lambda < 0$. The functions $f(\lambda)$ and $H(\rho)$ are

$$f(\lambda) = \ln\left[(1+e^{\lambda})^K - Ke^{\lambda_k}\right]$$
(4)

and

$$H(\rho) = -\rho \ln \rho - (1 - \rho) \ln(1 - \rho).$$
 (5)

Corollary 1: If $J \geq 3$, almost all codes of the ensemble $C_{(KM,JK)}$ iteratively (with complexity $O(N \log N)$) correct all combinations of erasures which number does not exceed some linear function αN of block length N, conditioned that $\alpha \leq 2/K$.

The parameter α (conditioned that $\alpha \leq 2/K$) can be considered as a lower bound on the *fraction of iteratively correctable erasures*. In Figures 1-3, α is presented as a function of the lower bound $R^* = 1 - J/K$ of the code rate R. For comparison, the Zyablov-Pinsker [1] lower bound α_0 on the fraction of iteratively correctable erasures and lower bound ρ^* on the the normalized minimal distance of the code ensemble $C_{(KM,JK)}$ are given.

We can see that for low rates the lower bound on the number of correctable erasures for the ensemble $C_{(KM,JK)}$ is better than the Zyablov-Pinsker bound but for large rates Zyablov-Pinsker bound is better.

IV. ENSEMBLE ANALYSIS

Let us represent the received vector \tilde{v} as the sequence of K vectors of length M, i.e. $\tilde{v} = (\tilde{v}_1, \tilde{v}_2, \dots, \tilde{v}_K)$, where $\tilde{v}_k = (\tilde{v}_{k1}, \tilde{v}_{k2}, \dots, \tilde{v}_{kM}), k = 1, 2, \dots, K$. Let t_1, t_2, \dots, t_K be *e*-weights of vectors $\tilde{v}_1, \tilde{v}_2, \dots, \tilde{v}_K, t_1 + t_2 + \dots + t_K = t$ respectively. We say that the vector \tilde{v} has *e*-weight composition



Fig. 2. Lower bound α on the fraction of iteratively correctable erasures and Zyablov-Pinsker lower bound α_0 on the fraction of iteratively correctable erasures as functions of R^* for K = 7. For comparison the lower bound [3] ρ^* on the normalized minimum distance of the code ensemble $C_{(KM,JK)}$ is presented.



Fig. 3. Lower bound α on the fraction of iteratively correctable erasures and Zyablov-Pinsker lower bound α_0 on the fraction of iteratively correctable erasures as functions of R^* for K = 11. For comparison lower bound [3] ρ^* on the normalized minimum distance of the code ensemble $C_{(KM,JK)}$ is presented.

 $\boldsymbol{t} = (t_1, t_2, \dots, t_K)$. There exist $\prod_{k=1}^K \binom{M}{t_k}$ vectors $\tilde{\boldsymbol{v}}$ with *e*-weight composition \boldsymbol{t} .

The codeword v satisfies L equations given by the paritycheck matrix H. These parity-check equations can be divided into J subsets. The subset $S^{(j)}$, j = 1, 2, ..., J, consists of M parity-check equations determined by permutation matrices P_{jk} , k = 1, 2, ..., K.

Consider the word \tilde{v} received after transmission of a codeword v over the BEC. Let *i*-th equation, i = 1, 2, ..., M, of subset $S^{(j)}$ is associated with the *K*-dimensional binary vector $p_i^{(j)}$. The *k*-th component, k = 1, 2, ..., K of $p_i^{(j)}$ is equal to one if the k-th component of the vector $\tilde{\boldsymbol{v}}_k$ is erased and it is equal to zero otherwise. Let us introduce the vector $\boldsymbol{c}^{(j)} = (\boldsymbol{p}_1^{(j)}, \boldsymbol{p}_2^{(j)}, \dots, \boldsymbol{p}_M^{(j)}).$

The derivation of the upper bound on the mathematical expectation of the number of stop sets with e-weight composition t is analogous to the derivation of the upper bound on the mathematical expectation of the number of codewords with weight composition d, described in [1]. We focus on the key elements of the proof of our bound.

Let $\gamma_0^{(j)}$ be a number of equations of subset $\mathcal{S}^{(j)}$, $j = 1, 2, \ldots, J$, not including the erased symbols of any of the vectors \tilde{v}_k , $k = 1, 2, \ldots, K$. Furthermore, let $\gamma_2^{(j)}(k_1, k_2)$ be a number of the equations of subset $\mathcal{S}^{(j)}$, including two erased symbols, one from vector \tilde{v}_{k_1} and one from \tilde{v}_{k_2} ; the other K-2 vector components of \tilde{v} included in equations of this subset are not erased. Generally, let $\gamma_i^{(j)}(k_1, k_2, \ldots, k_i)$, $i = 2, 3, \ldots, K$ be the number of the equations of subset $\mathcal{S}^{(j)}$, including *i* erased symbols, one symbol from each of vectors $\tilde{v}_{k_1}, \tilde{v}_{k_2}, \ldots, \tilde{v}_{k_i}$; the other K - i components of a vector \tilde{v} included in equations of this subset are not erased. Note that arguments of the function $\gamma_i^{(j)}(k_1, k_2, \ldots, k_i)$ are different and that this function is invariant to permutation of its arguments. To emphasize this fact, we henceforth write $\gamma_i^{(j)}(\{k_1, k_2, \ldots, k_i\})$ for $\gamma_i^{(j)}(k_1, k_2, \ldots, k_i)$.

Analogous to equations (3), (5) of [3] we get

$$\gamma_{0}^{(j)} + \sum_{\{k_{1},k_{2}\}} \gamma_{2}^{(j)}(\{k_{1},k_{2}\}) +$$

$$+ \sum_{\{k_{1},k_{2},k_{3}\}} \gamma_{3}^{(j)}(\{k_{1},k_{2},k_{3}\}) + \dots +$$

$$+ \gamma_{K}^{(j)}(\{1,2,\dots,K\}) = M,$$

$$\sum_{k_{2}\}} \gamma_{2}^{(j)}(\{k_{1},k_{2}\}) + \sum_{\{k_{2},k_{3}\}} \gamma_{3}^{(j)}(\{k_{1},k_{2},k_{3}\}) +$$

$$+ \dots + \gamma_{K}^{(j)}(\{1,2,\dots,K\}) = t_{k_{1}},$$
(6)
(7)

for $k_1 = 1, 2, \ldots, K$.

Note that sums (7) and (8) do not contain $\gamma_1^{(j)}(k_1)$, the number of equations with exactly one erased symbol, but in contrast to the sums (3) and (5) of [1], they contain the numbers of the equations with three, five, etc. erased symbols.

Consider a vector

$$\boldsymbol{\gamma}^{(j)} = (\gamma_0^{(j)}, \{\gamma_2^{(j)}(\{k_1, k_2\})\}, \dots, \gamma_K^{(j)}(\{1, 2, \dots, K\})),$$

with components satisfying (7) and (8). The total number of components of the vector $\gamma^{(j)}$ is equal to $2^K - K$. The vector $\gamma^{(j)}$ is called the *j*-composition of constraints. Then

$$\frac{1}{\gamma_{0}^{(j)}!\prod_{\{k_{1},k_{2}\}}\gamma_{2}^{(j)}(\{k_{1},k_{2}\})!} \times (8)$$

$$\times \frac{M!}{\prod_{\{k_{1},k_{2},k_{3}\}}(\{k_{1},k_{2},k_{3}\})!\cdots\gamma_{k}^{(j)}(\{1,2,\ldots,K\})!}$$

various vectors $oldsymbol{c}^{(j)}$ correspond to a vector $oldsymbol{\gamma}^{(j)}$

Next, consider a set $\Gamma^{(j)}(t)$ of vectors $\gamma^{(j)}$ having *e*-weight composition *t*. The number of vectors $c^{(j)}$, corresponding to this set is equal to

$$\phi^{(j)}(t) = \sum_{\Gamma^{(j)}(t)} \frac{1}{\gamma_0^{(j)}! \prod_{\{k_1,k_2\}} \gamma_2^{(j)}(\{k_1,k_2\})!} \times \qquad (9)$$

$$\frac{M!}{\prod_{\{k_1,k_2,k_3\}} \gamma_3^{(j)}(\{k_1,k_2,k_3\})! \cdots \gamma_K^{(j)}(\{1,2,\ldots,K\})!},$$

where summation is over all $\gamma^{(j)}$, included in set $\Gamma^{(j)}(t)$. The following two lemmas are analogous to lemmas 1 and 2 from [1].

Lemma 1: For the ensemble $C_{(KM,JK)}$, the probability $\psi^{(j)}(t)$ that an arbitrary chosen vector $\tilde{\boldsymbol{v}} = (\tilde{\boldsymbol{v}}_1, \tilde{\boldsymbol{v}}_2, \dots, \tilde{\boldsymbol{v}}_K)$ with a *e*-weight composition $\boldsymbol{t} = (t_1, t_2, \dots, t_k)$ has no correcting checks in $S^{(j)}$ is equal to

$$\psi^{(j)}(t) = \frac{\phi^{(j)}(t)}{\prod_{k=1}^{K} \binom{M}{t_k}}.$$
(10)

Lemma 2: The function $\phi^{(j)}(t)$ is upperbounded by inequality

$$\phi^{(j)}(t) \le \exp\left[M\left(g(\lambda_1, \lambda_2, \dots, \lambda_K) \sum_{k=1}^K \lambda_k \rho_k\right)\right], \quad (11)$$
$$j = 1, 2, \dots, K,$$

where

Х

$$g(\lambda_1, \lambda_2, \dots, \lambda_K) = \ln \left[\prod_{k=1}^K (1 + e^{\lambda_k}) \sum_{k=1}^K e^{\lambda_k}\right]$$
(12)

and $\rho_k = t_k/M$, $\rho_k \ge 0$ is the normalized *e*-weight of a vector \tilde{v}_k .

Note that the only essential difference between the bounds (11),(12) and the corresponding bounds of [3] is the different definition of the function $g(\lambda_1, \lambda_2, ..., \lambda_K)$.

The following lemma is similar to lemma 3 of [3].

Lemma 3: In the ensemble $C_{(KM,JK)}$, the average number E(t) of stop sets with *e*-weight composition t is upperbounded by

$$E(t) \leq \prod_{k \in \mathcal{K}^+} \sigma(\rho_k, M) \exp\left[JMF_K(\lambda_1, \dots, \lambda_K, \rho_1, \dots, \rho_K)\right]$$
(13)

where

$$F_K(\lambda_1,\ldots,\lambda_K,\rho_1,\ldots,\rho_K)) =$$
(14)

$$= g(\lambda_1, \lambda_2, \dots, \lambda_K) \sum_{k=1} \lambda_k \rho_k \frac{J-1}{J} \sum_{k=1} H(\rho_k),$$

$$\sigma(\rho, M) = \sqrt{2\pi M \rho(1-\rho)} \exp\left[\frac{1}{12M\rho(1-\rho)}\right], \quad (15)$$

 $H(\rho)$ is defined by (5) and $\mathcal{K}^+ = \{k | \rho_k > 0\}.$

The further derivation of the bound is reduced to an

analysis of function $F_K(\lambda_1, \ldots, \lambda_K, \rho_1, \ldots, \rho_K)$. First, we find variables $\lambda_1, \ldots, \lambda_K$ for which this function has a minimum. Substituting these values into function $F_K(\lambda_1, \ldots, \lambda_K, \rho_1, \ldots, \rho_K)$ we get

$$G_K(\rho_1, \rho_2, \dots, \rho_K) \stackrel{\Delta}{=} \min_{\lambda_1, \dots, \lambda_K} F_K(\lambda_1, \dots, \lambda_K, \rho_1, \dots, \rho_K).$$

Let us introduce the function

$$\theta(\rho) \stackrel{\Delta}{=} \max_{\rho_1, \rho_2, \dots, \rho_K : \frac{1}{K} \sum_{k=1}^K \rho_k = \rho} G_K(\rho_1, \rho_2 \rho_K)$$
(16)

Then, consider the set

$$\mathcal{A}(\rho) = \{ \boldsymbol{t} | \frac{1}{K} \sum_{k=1}^{K} \rho_k = \rho \}$$

of e-sets with normalized e-weight ρ . From lemma 3 follows

$$\sum_{\mathcal{A}(\rho)} E(\mathbf{t}) \leq (17)$$

$$\leq \sum_{\mathcal{A}(\rho)} \exp\left\{ JM\left[\theta(\rho) + \frac{J-1}{JM}\ln\left(\prod_{k\in\mathcal{K}^+} \sigma(\rho_k, M)\right)\right] \right\}.$$

Obviously, if all symbols in two vector components v_{k_1} and v_{k_2} of a vector v are erased, then the vector \tilde{v} define stop set. In this case, there exists a set of 2M erasures which it cannot correct. Hence, we can limit ourself by consideration of vectors v with *e*-weights t < 2M, i.e. vectors, for which $\rho < 2/K$. The following lemma similar to lemma 4 in [3] determines a point $(\rho_1, \rho_1, \ldots, \rho_K)$ in the *K*-dimensional space, corresponding to a maximum of $G_K(\rho_1, \rho_1, \ldots, \rho_K)$ for *e*-weight sets with the normalized *e*-weight $\rho < 2/K$.

Lemma 4: If
$$J \ge 3$$
 and $\rho = \frac{1}{K} \sum_{k=1}^{K} \rho_k$, $\rho < 2/K$, is fixed

the conditional maximum of the function $G_K(\rho_1, \ldots, \rho_K)$ is attained at the unique point $\rho_1 = \cdots = \rho_K = \rho$, i.e.

$$\theta(\rho) = G_K(\rho, \dots, \rho) = F_K(\lambda, \dots, \lambda, \rho, \dots, \rho), \quad (18)$$

where λ satisfies

$$\frac{d}{d\lambda}[f(\lambda)] - K\rho = 0, \tag{19}$$

and

$$f(\lambda) = g(\lambda_1 = \lambda, \dots, \lambda_k = \lambda).$$
 (20)

The proof of this lemma almost exactly coincides with the proof of the lemma 4 [3], which uses Appendices 1 and 2 of [3]. The only differences are caused by the different definition of function $g(\lambda_1, \lambda_2, \ldots, \lambda_K)$ in our work and in the paper [3]. In particular, instead of the ratio (A.1.1 [3]) the following equality should be used

$$\frac{\rho_k}{1 - \rho_k} = e^{\lambda_k} \frac{\prod_{k' \neq k} (1 + e^{\lambda_{k'}}) - 1}{\prod_{k' \neq k} (1 + e^{\lambda_{k'}}) - \sum_{k' \neq k} e^{\lambda_{k'}}},$$
 (21)

and instead of system of the equations (A.1.3 [3]) the follow-

ing system should be used:

$$e^{-\frac{J_{\mu}}{K(J-1)}} = e^{-\frac{\lambda_k}{J-1}} \frac{\prod_{k' \neq k} (1 + e^{\lambda_{k'}}) - 1}{\prod_{k' \neq k} (1 + e^{\lambda_{k'}}) - \sum_{k' \neq k} e^{\lambda_{k'}}},$$
 (22)

where μ is Lagrange multiplier.

To prove that the system of the equations (22) has only the solution $\lambda_1 = \lambda_2 = \cdots = \lambda_K$ and $\rho_1 = \rho_2 = \cdots = \rho_K = \rho$, we use the following equality

$$X \left[(1+Y^{J-1})A - Y^{J-1} - B \right] \left[(1+X^{J-1})A - 1 \right] = (23)$$

= $Y \left[(1+X^{J-1})A - X^{J-1} - B \right] \left[(1+Y^{J-1})A - 1 \right],$

where

$$X = e^{\frac{\lambda_{k_1}}{J-1}},$$

$$Y = e^{\frac{\lambda_{k_2}}{J-1}},$$

$$A = \prod_{k' \neq k_1, k_2} (1 + e^{\lambda_{k'}}),$$

$$B = \sum_{k' \neq k_1, k_2} e^{\lambda_{k'}}.$$

From uniqueness of the real solution of (23), it follows uniqueness of the solution of the system of the equations (22).

The second difference in the proof of lemma 4 and lemma 4 of [3] is because of the different definitions of function $\Delta(\rho)$ used for the proof that

$$G_{K}(\rho_{1} = \rho, \dots, \rho_{K} = \rho) \geq$$
(24)
$$\geq G_{K-1}\left(\rho_{1} = \frac{K}{K-1}\rho, \dots, \rho_{K-1} = \frac{K}{K-1}\rho\right).$$

In our case we use definition (compare with (35) [1])

$$\Delta(\rho) = \ln \frac{(1+e^{\lambda})^{K} - Ke^{\lambda}}{(1+e^{\lambda})^{K-1} + (K-1)e^{\lambda}} + (25)$$
$$+ \left(\frac{J-1}{J}\right) \left[(K-1)H\left(\frac{K}{K-1}\rho\right) - KH(\rho) \right] = \\= \ln \left(\frac{1}{1-\rho}\right) + \\+ \left(\frac{J-1}{J}\right) \left[(K-1)H\left(\frac{K}{K-1}\rho\right) - KH(\rho) \right],$$

where $H(\rho)$ is defined by (5).

From Lemmas 3 and 4 follows the main result formulated in Section III as the Theorem 1.

V. CONCLUSIONS

Most papers on iterative decoding of LDPC codes deal with symbolwise decoding. Usually, upper bounds on the probability of decoding error or failure is studied and for upperbounding the block (or frame) error probability the union bound is applied. In this paper, we are interested in iterative decoding of LDPC codes for the erasure channel "as a whole". We got a new bound on the fraction of iteratively decodable erasures and compared this bound with the Zyablov-Pinsker bound.

REFERENCES

- V. V. Zyablov and M. S. Pinsker, "Decoding complexity of low-density codes under transmission over erasure channels," *Probl. Pered. Inform.*, v. 10, pp. 15–28, Jan. 1974.
- [2] M. G. Luby, M. Mitzenmacher, M. A. Shokrollahi, and D. A. Spielman, "Efficient Erasure Correcting Codes," *IEEE Trans. Inform. Theory*, IT-47(2): 569–584, Feb. 2001.
- [3] A. Sridharan, M. Lentmaier, D. V. Truhachev, D. J. Costello, Jr., and K. Sh. Zigangirov, "On minimal distance of low-density codes with parity-check matrices composed of permutation matrices," *Probl. Pered. Inform.*, v. 41, pp. 39–52, Jan. 2005.
- [4] G.Di, D.Proietti, I.E.Teletar, 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, IT-48(6): 1570-1579, June 2002.
- [5] R. G. Gallager, *Low-Density Parity-Check Codes*. Cambridge, MA:MIT Press, 1963.