Eleventh International Workshop on Algebraic and Combinatorial Coding Theory
June 16-22, 2008, Pamporovo, Bulgaria *pp. 348-353*

Low-complexity error correction in LDPC codes with constituent RS codes[1]

Victor Zyablov zyablov@iitp.ru
Vladimir Potapov potapov@iitp.ru
Fedor Groshev groshev@iitp.ru
Institute for Information Transmission Problems, Russian Academy of Sciences, Moscow 101447, Russia

Abstract.
Reed-Solomon code-based LDPC (RS-LDPC) block codes are obtained by replacing single parity-check codes in Gallager's LDPC codes with Reed-Solomon constituent codes. This paper investigates asymptotic error correcting capabilities of ensembles of random RS-LDPC codes, used over the binary symmetric channel and decoded with a low-complexity harddecision iterative decoding algorithm. The number of required decoding iterations is a logarithmic function of the code length. It is shown that there exist RS-LDPC codes for which such iterative decoding corrects any error pattern with a number of errors that grows linearly with the code length. The results are supported by numerical examples, for various choices of code parameters.

1 Introduction

Long block codes can be obtained by combining one or more simpler codes in various types of concatenated structures. Such constructions are of interest since they can yield powerful codes with good error-correcting capabilities, which are decodable with low complexity, using simple constituent decoders as separate modules.

A method for constructing long codes from short constituent codes, based on bipartite graphs, was introduced by Tanner in [1]. In this method, one of the two sets of nodes in a bipartite graph is associated with code symbols, while the other set is associated with constituent block codes of length equal to the node degree. These two sets of nodes are hereinafter referred to as variable nodes and constraint nodes, respectively. Tanner's general code construction unifies many known code families that can be obtained by choosing different underlying bipartite graphs and associating different constituent codes with its constraint nodes. For example, Gallager's Low-Density Parity-Check (LDPC) codes [2], graph-based approach.

[1]This work was supported by the Program of fundamental scientific reseaches of ONIT RAN (the Department of Nanotechnologies and Information Technologies)

For Gallager's LDPC codes [2], each constraint node in the corresponding bipartite graph represents a single parity-check (SPC) code over the variable nodes connected to it.

The error-correcting capabilities of LDPC codes for the binary symmetric channel (BSC) were studied in [3], where it was shown that there exist LDPC codes capable of correcting a portion of errors that grows linearly with the code length n, with decoding complexity $\mathcal{O}(n \log n)$.

The codes associated with constraint nodes in the Tanner graph of an LDPC code can be replaced with other constituent block codes (*e.g.* Reed-Solomon codes [4]).

In this paper, we consider the asymptotic performance of random RS-LDPC codes, when the code length n grows to infinity. We will prove that there exist RS-LDPC codes which, when decoded with a simple iterative decoder of complexity $\mathcal{O}(n \log n)$, can correct any error pattern with a number of errors growing linearly with the code length. The work presented here, with constituent Reed-Solomon codes of minimum distance $d_0 = 3$.

2 Construction and properties of RS-LDPC codes

An (n_0, k_0, d_0) extended Reed-Solomon code has length $n_0 = 2^q$, dimension $k_0 = n_0 - d_0 - 1$, code rate $R_0 = 1 - (d_0 - 1)/n_{0.}$. We will consider single-error correcting extended RS code with minimum distance $d_0 = 3$,

A parity-check matrix H_0 of a Reed-Solomon code is an $(d_0 - 1) \times n_0$ matrix whose columns are all nonzero q-nary $(d_0 - 1)$-tuples. We will consider RS-LDPC codes with identical constituent codes. Let H denote a block-diagonal matrix with the b constituent parity-check matrices H_0 on the main diagonal, that is,

$$H = \begin{pmatrix} H_0 & 0 & 0 & \cdots & 0 \\ 0 & H_0 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & H_0 \end{pmatrix} \tag{2}$$

where b is very large. The matrix H is of size $b(d_0 - 1) \times bn_0$. Let $\pi(H)$ denote a random column permutation of H. Then the matrix constructed using $\ell \geq 2$ such permutations as *layers*,

$$H = \begin{pmatrix} H_1 \\ H_2 \\ \vdots \\ H_\ell \end{pmatrix} = \begin{pmatrix} \pi_1(H) \\ \pi_2(H) \\ \vdots \\ \pi_\ell(H) \end{pmatrix} \tag{3}$$

is a sparse $\ell b(d_0 - 1) \times bn_0$ parity-check matrix which characterizes the ensemble of Reed-Solomon code-based LDPC codes of length $n = bn_0$, where $n \gg n_0$.

Let $C(n_0, \ell, b)$ denote this ensemble. For a given constituent Reed-Solomon code with parity-check matrix H_0, the elements of the ensemble $C(n_0, \ell, b)$ are obtained by sampling independently the permutations π_l, $l = 1, 2, ..., \ell$, which are all equiprobable. The rate of a code $\mathcal{C} \in C(n_0, \ell, b)$ is lower-bounded by [1]

$$R \geq 1 - \frac{\ell b(n_0 - k_0)}{n} = 1 - \ell(1 - R_0) \tag{4}$$

with equality if the matrix H has full rank.

The RS-LDPC codes from the ensemble $C(n_0, \ell, b)$ contain ℓb constituent Reed-Solomon codes; b in each layer. Such RS-LDPC codes can be represented by a Tanner graph [1] with $n = bn_0$ variable nodes, and ℓb constraint nodes. Each constraint node comprises $n_0 - k_0$ parity-check constraints specified by the rows of the corresponding constituent parity-check matrix. If a code symbol is checked by a constituent code (that is, by at least one row of its parity-check matrix), there is a branch connecting the corresponding variable node and the constraint node. Each code symbol is checked by exactly one Reed-Solomon code in each layer. The graph is regular, with the variable-node degree equal to ℓ, and the constraint-node degree equal to n_0.

Let $\vec{v}$ be the transmitted codeword and $\vec{e}$ be the error pattern. Then the received sequence is given by $\vec{r} = \vec{v} + \vec{e}$. The weight of the error sequence is $W = |\vec{e}|$ and the fraction of erroneous symbols is $\omega = W/n$ for code length $n \to \infty$.

For a given error pattern with W errors, we introduce the ℓ-tuple $\vec{a} = (a_1\ a_2\ ...\ a_\ell)$, where a_l, $l = 1, 2, ..., \ell$, denotes the number of constituent codes at the lth layer whose codewords are affected by errors. Note that $\vec{a}$ contains realizations of ℓ independent random variables that are integer-valued in the range $0 \leq a_l \leq b$, $l = 1, 2, ..., \ell$. Furthermore, let a denote the total number of constituent codes affected by errors, that is,

$$a = |\vec{a}| = \sum_{l=1}^{\ell} a_l.$$

In other words, a is the number of constraint nodes in the Tanner graph that are connected to at least one variable node with an erroneously received value.

3 Decoding algorithm

Consider an iterative hard-decision decoding algorithm A, whose decoding iterations i, $i = 1, 2, ..., i_{\max}$, consist of the following two steps:

1) For the tentative sequence $r^{(i)}$, where $r^{(1)}$ is the received sequence r, decode independently ℓb constituent Reed-Solomon codes (that is, compute their syndromes $s_{j,l}$, $j = 1, 2, ..., b$, $l = 1, 2, ..., \ell$, and if the value

is nonzero, output the n_0-tuple where the position indicated by the syndrome is flipped). This yields ℓ independent decisions for each of the n symbols.

2) Flip every symbol $r_k^{(i)}$, $k = 1, 2, ..., n$, in the sequence $r^{(i)}$, for which *at least one* of the ℓ decisions requires that. This yields the updated sequence $r^{(i+1)}$.

Assume that the error pattern e is such that the number of errors that can be corrected by the constituent codes is larger than the number of uncorrectable errors. Then, during the first iteration of the algorithm A, all correctable errors will be corrected. Since, in our case, Reed-Solomon codes are single-error correcting codes, each erroneous decoding will added one new error. Hence, the new error pattern, resulting from one decoding iteration has fewer errors than the initial error pattern. Clearly, if in each of the following iterations, the number of correctable errors is larger than the number of uncorrectable ones, then the total number of errors in $r^{(i)}$ will decrease with the iteration number i and the algorithm yields the correct decision, *i.e.*, $r^{(i_{\max})} = v$. Then, we can state the following

Lemma 1 For any RS-LDPC code from the ensemble $C(n_0, \ell, b)$, if an error pattern is such that in each iteration of algorithm A the number of errors correctable by the constituent codes is larger in $(1 + \varepsilon)$ times than the number of added errors, then algorithm A yields a correct decision after $\mathcal{O}(\log n)$ iterations, where $n = bn_0$ is the code length.

The complexity of each decoding iteration of the algorithm A is proportional to the code length n. Thus, according to Theorem 1, the overall decoding complexity is $\mathcal{O}(n \log n)$, given that the number of correctable errors in the error pattern is larger than the number of the uncorrectable ones. The following lemma formulates a condition under which this holds.

Lemma 2 If for any error pattern with $w \leq W$ errors, the number of constituent Reed-Solomon codes of an RS-LDPC code from the ensemble $C(n_0, \ell, b)$ that are affected by errors is $a = \alpha w \ell$ with $\alpha \geq 2/3 + \varepsilon$, then the number of correctable errors in any such error pattern is always larger than the number of uncorrectable errors.

In other words, $\alpha \geq 2/3 + \varepsilon$ specifies the necessary expansion of the Tanner (expander) graph of the code, which ensures that the number of errors decreases in each iteration of algorithm A.

4 Asymptotic performance

As shown in the previous section, the iterative algorithm A corrects any error pattern with W or fewer errors, if the code's Tanner graph has the expansion

coefficient $\alpha \geq 2/3 + \varepsilon$. The question that arises, however, is whether such a code exists in the ensemble $C(n_0, \ell, b)$. The following theorem allows us to receive the positive answer.

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

$$h(\omega) + \omega log_2(q-1) - \ell F(\alpha, \omega, n_0) = 0 \tag{5}$$

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

$$F(\alpha, \omega, n_0) \triangleq h(\omega) + \omega \log_2(q-1) - \frac{1}{n_0} h(\alpha\omega n_0) + \max\Big\{\omega \log_2 s - \alpha\omega \log_2\Big((1 + s(q-1))^{n_0} - 1\Big)\Big\} \tag{6}$$

where $\alpha \geq 2/3 + \varepsilon$ and the maximization is performed over all s such that

$$(1 + s(q-1))^{n_0} < 1 + \frac{1 - \alpha\omega n_0}{\alpha\omega n_0}$$

Theorem 1 allows us to compute ω_α numerically for several choices of code parameters. The computations confirm the existence of codes with a nonvanishing ω_α. We use $\alpha = 0.67$, which is slightly above the limit value of $2/3$. First, we consider code ensembles of a rate close to $1/2$. Figure 1 illustrates the values of ω_α computed for several code ensembles $C(n_0, \ell, b)$ of rates approximately $1/2$. 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, $n_0 = 128$, where it reaches its maximum. With further increase of n_0 and ℓ, ω_α decays quickly.

Next we consider code ensembles of different rates, but with a fixed constituent code. Figure 2 illustrates the values ω_α for RS-LDPC codes with the constituent $(128, 126, 3)$ Reed-Solomon code and with different code rates R, obtained by varying the choice of ℓ. We have found a nonvanishing ω_α for a wide range of code rates, and its value decreases with increasing code rate.

References

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

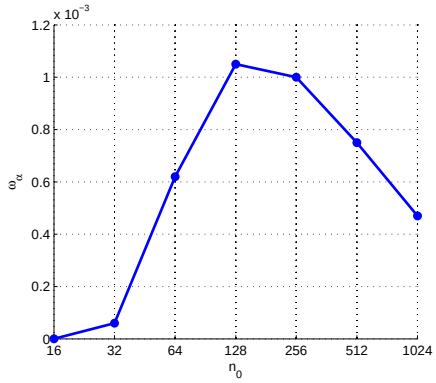

Figure 1: Values of ω_α computed for $\alpha = 0.67$ according to Theorem 1 for several code ensembles of rates approximately $R \approx 1/2$. The maximum is achieved with the constituent code length $n_0 = 128$.

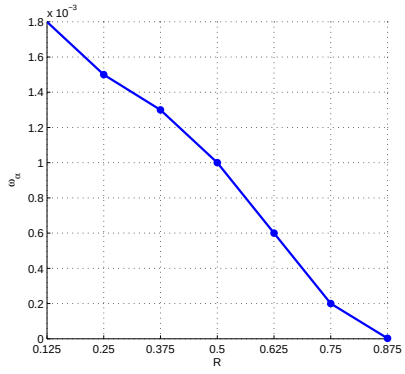

Figure 2: Values of ω_α computed for $\alpha = 0.67$ according to Theorem 1 for several code ensembles of different rates with the fixed constituent code length $n_0 = 128$.

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

[3] V. V. Zyablov, M. S. Pinsker, Estimation of the error-correction complexity for Gallager low-density codes, *Problems of Inform. Transmission* 11, 1975, 23-26.

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