International Symposium on Information Theory and its Applications, ISITA2008
Auckland, New Zealand, 7-10, December, 2008

On the Design of Irregular GLDPC Codes with Low Error Floor over the BEC

Enrico PAOLINI†, Marc FOSSORIER‡ and Marco CHIANI†

† DEIS/WiLAB, University of Bologna
via Venezia 52, 47023 Cesena, ITALY
E-mail: {e.paolini,marco.chiani}@unibo.it

‡ ETIS ENSEA/UCP/CNRS UMR-8051
6 avenue du Ponceau, 95014, Cergy Pontoise, FRANCE
E-mail: mfossorier@ieee.org

Abstract

The design of GLDPC codes for the binary erasure channel with low error floor under iterative decoding is investigated. Both bounded distance and *maximum a posteriori* decoding at the check nodes are considered. For both check node decoding algorithms a key parameter is identified, discriminating between an exponentially small and exponentially large expected number of small size stopping sets. A code design technique is proposed based on this theoretical investigation.

1. INTRODUCTION

Low-density parity-check (LDPC) codes [1] have attracted an increasing attention in the last decade due to their capacity-approaching performance, over a wide range of channels, under low-complexity iterative decoding. An LDPC code is graphically represented as a bipartite graph, or Tanner graph, where the nodes are grouped into two disjoint subsets, namely, variable nodes (VNs) and check nodes (CNs), such that each edge can connect only two nodes from different sets. Each VN is associated with an encoded bit. Each CN is associated with a parity-check equation and can be regarded as a single parity-check (SPC) code.

A wider class of codes is represented by generalized LDPC (GLDPC) codes. They generalize LDPC codes in that the CN set is allowed to be composed of generic (not necessarily SPC) linear block codes. GLDPC codes with a uniform VN set and a uniform CN set were first introduced in [2], where it was shown that CNs with a good minimum distance are beneficial in terms of the overall minimum distance.

In this paper we focus on *irregular* GLDPC codes, whose VN set and CN set are composed of VNs with different degrees and of a mixture of different linear block codes, respectively. Specifically, we investigate the design and construction of iteratively decoded GLDPC codes with low error floor over the binary erasure channel (BEC). Two different erasure decoding algorithms at the CNs are considered, namely, bounded distance (BD) and *maximum a posteriori* (MAP) decoding. Assuming there exist VNs and CNs with minimum distance 2, the first-order term of the Taylor expansion of the stopping set size growth rate is evaluated analytically, for both algorithms, in order to obtain guidelines for the code design. Moreover, a key parameter depending only on the minimum distance 2 VNs and CNs is obtained, discriminating between an asymptotically large or an asymptotically small number of small linear-sized stopping sets. A code design technique is proposed accordingly.

This work was supported in part by the European Commission under Project IST-214625 "Optimix".

2. PRELIMINARIES

2.1. Notation

We denote by n and m the length and the number of CNs of a GLDPC code. In its Tanner graph, the number of edges connected to any node is the node *degree*. We assume there are I_C different CN types. For each CN type $i \in \{1, \dots, I_C\}$, $\overline{d}_i$ and $\underline{d}_i$ denote the degree of a type-i CN and its minimum distance, respectively. The VN degree distribution from an edge perspective is $\lambda(x) \triangleq \sum_{i \geq 2} \lambda_i x^{i-1}$, λ_i being the fraction of edges connected to the degree-i VNs. Analogously, we define a CN degree distribution from an edge perspective as $\rho(x) \triangleq \sum_i \rho_i x^{\overline{d}_i - 1}$, ρ_i being the fraction of edges connected to the type-i CNs. Note that

$$\frac{n}{m} = \frac{\int \lambda}{\int \rho} \tag{1}$$

where $\int \lambda \triangleq \int_0^1 \lambda(x)\,\mathrm{d}x$ and $\int \rho \triangleq \int_0^1 \rho(x)\,\mathrm{d}x$. We denote by r the smallest minimum distance among the CNs and, for $r = 2$, by $\rho_2(x)$ the contribution to $\rho(x)$ of the $\underline{d} = 2$ CN types, i.e.,

$$\rho_2(x) \triangleq \sum_{i | \underline{d}_i = 2} \rho_i x^{\overline{d}_i - 1} . \tag{2}$$

Moreover, we denote by p the smallest VN degree.

Given a length-$\overline{d}$ non-zero codeword of a CN, we say that a subset of the $\overline{d}$ CN sockets *covers* the codeword if no '1' bit of the codeword corresponds to a socket not included in the subset. The $\binom{\overline{d}}{u}$ possible choices of u out of the $\overline{d}$ sockets are index from 1 to $\binom{\overline{d}}{u}$. We let $\delta_{u,j}$ be an indicator function such that $\delta_{u,j} = 1$ if the j-th choice of u sockets covers at least one non-zero codeword of the CN, and $\delta_{u,j} = 0$ otherwise.

The multiplicity of the weight-u codewords of a type-i CN is denoted by $A_u^{(i)}$. For $r = 2$ we let

$$C \triangleq \sum_{i|\underline{d}_i=2} \rho_i \frac{2\,A_2^{(i)}}{\overline{d}_i}\,. \qquad (3)$$

Finally, we denote by coeff$(p\,(x), x^t)$ the coefficient of x^t in a polynomial $p\,(x)$.

2.2. Stopping Sets of GLDPC Codes

A *stopping set* is any subset of VNs associated with an uncorrectable erasure pattern over the BEC. The number of VNs a stopping set is composed of is its size. Within the context of LDPC code iterative decoding, a stopping set is defined as any subset of VNs such that if a CN is connected to it, it is connected to it at least twice [3]. For GLDPC codes under iterative decoding we need different stopping set definitions for different erasure recovery algorithms at the CNs.

Definition 1. *A stopping set of a GLDPC code with BD erasure decoding at the CNs is any subset of VNs such that if a CN with minimum distance $\underline{d}$ is connected to it, it is connected to it at least $\underline{d}$ times* [4].

Definition 2. *A stopping set of a GLDPC code with MAP erasure decoding at the CNs is any subset of VNs such that if the value of these VNs is unknown, then no information set is known for any CN connected to it.*

Using results from [5, Appendix IV] it is possible to show that Definition 2 is equivalent to:

Definition 3. *A stopping set of a GLDPC code with MAP erasure decoding at the CNs is any subset of VNs such that if a CN is connected to it, the CN sockets connected to it cover at least one non-null CN codeword.*

2.3. Valid Assignments

Any subset of a Tanner graph edges is called an *assignment*, and the number of assignment edges is called the *assignment weight*. Calculating the Taylor expansion of GLDPC codes stopping set size distribution mostly relies on the concept of *valid assignment*. An assignment is valid from a VN perspective when for each VN, either all the VN edges belong to the assignment or no VN edge belongs to it. The definition of valid assignment from a CN perspective depends on the erasure recovery algorithm used at the CNs.

Definition 4. *In the Tanner graph of a GLDPC code with BD erasure recovery at the CNs, an assignment is valid from a CN perspective when, if an assignment edge is connected to a CN with minimum distance $\underline{d}$, then at least $\underline{d}$ assignment edges are connected to it.*

Definition 5. *In the Tanner graph of a GLDPC code with MAP erasure recovery at the CNs, an assignment is valid from a CN perspective when, if a CN is connected to the assignment, then the corresponding sockets cover at least one non-zero CN codeword.*

An assignment is said *valid* when it is valid from both a VN and a CN perspective. There is a one-to-one correspondence between stopping sets and valid assignments, for both BD and MAP decoding at the CNs.

2.4. Growth Rate of the Stopping Set Size Distribution

Let $\mathcal{G}$ be the Tanner graph of a length-n GLDPC code randomly drawn from an ensemble defined by $\lambda(x)$, by the CN type set and by $\rho(x)$. Let $N_{\mathrm{SS,BD}}(\mathcal{G}, \ell)$ and $N_{\mathrm{SS,MAP}}(\mathcal{G}, \ell)$ be the number of stopping sets of size ℓ in $\mathcal{G}$ under BD and MAP erasure recovery at the CNs, respectively. The growth rate of the stopping set size distribution is defined as

$$G_{\mathrm{SS}}(w) \triangleq \lim_{n\to\infty} \frac{1}{n} \ln \mathbb{E}\left[N_{\mathrm{SS}}(\mathcal{G}, \lfloor wn \rfloor)\right] \qquad (4)$$

where $N_{\mathrm{SS}}(\mathcal{G}, \lfloor wn \rfloor) = N_{\mathrm{SS,BD}}(\mathcal{G}, \lfloor wn \rfloor)$ for BD CN decoding and $N_{\mathrm{SS}}(\mathcal{G}, \lfloor wn \rfloor) = N_{\mathrm{SS,MAP}}(\mathcal{G}, \lfloor wn \rfloor)$ for CN MAP decoding, and where the expectation is over all the GLDPC codes of the ensemble.

3. THE MAIN RESULT

We are interested in the growth rate of the small size stopping sets, as they mainly affect the error floor performance over the BEC. Therefore, we investigate the first-order Taylor expansion of (4) around $w = 0$.

Lemma 1. *Consider γm CNs of the same type, having length $\overline{d}$ and minimum distance $\underline{d}$. The number of check-valid assignments of weight δm over these CNs is*

$$\mathrm{coeff}\left((p(x))^{\gamma m}, x^{\delta m}\right)$$

where

$$p(x) = \begin{cases} 1 + \sum_{u=\underline{d}}^{\overline{d}} \binom{\overline{d}}{u} x^u & \textit{BD CN decoding} \\ 1 + \sum_{u=\underline{d}}^{\overline{d}} \sum_{j=1}^{\binom{\overline{d}}{u}} \delta_{u,j}\, x^u & \textit{MAP CN decoding.} \end{cases}$$

Proof. See Appendix I. □

The main result of this paper is claimed next.

Theorem 1. *For an irregular GLDPC code ensemble with $r = p = 2$, the Taylor expansion of the growth rate of the stopping set size distribution is given by:*

$$\begin{aligned} G_{\mathrm{SS,BD}}(w) &= w \ln(\lambda'(0)\,\rho_2'(1)) + O(w^2) \\ G_{\mathrm{SS,MAP}}(w) &= w \ln(\lambda'(0)\,C) + O(w^2) \end{aligned} \quad (5)$$

for BD and MAP decoding at the CNs, respectively[1].

Proof. See Appendix II. □

Theorem 1 states that the parameters $\lambda'(0)\rho_2'(1)$ and $\lambda'(0)\,C$, both depending only on the degree-2 VNs and on the ($d_{\min} = 2$) CNs, discriminate between an exponentially large and an exponentially small number of small linear-sized stopping sets (under BD and MAP CN decoding, respectively) for a randomly chosen GLDPC code in the ensemble.

Remarkably, the expression of the first-order Taylor series of $G_{\mathrm{SS,MAP}}(w)$ is identical to that obtained in [7] for the growth rate of GLDPC codes weight distribution. This fact reveals that the presence of an exponentially small (resp. large) number of small linear-sized codewords is always associated with an exponentially small (resp. large) number of small linear-sized stopping sets under MAP CN decoding.

It is readily shown that we always have $\rho_2'(1) > C$. Then, the condition $\lambda'(0)\rho_2'(1) < 1$ implies $\lambda'(0)\,C < 1$, but the opposite doesn't hold. Consequently, GLDPC codes drawn from ensembles for which $\lambda'(0)\,C < 1$, even if characterized by a low error floor under MAP CN decoding, may be affected by a high error floor when switching to low-complexity BD CN decoding.

4. DESIGN AND CONSTRUCTION OF IRREGULAR GLDPC CODES WITH LOW ERROR FLOOR

As the error floor over the BEC under iterative decoding is related to the minimum stopping set size, GLDPC codes with low error floor shall be searched among ensembles with $\lambda'(0)\rho_2'(1) < 1$ (for BD) or $\lambda'(0)\,C < 1$ (for MAP).

The proposed code construction technique consists of using a constrained version of the differential evolution (DE) algorithm [8] to design a GLDPC ensemble with optimized BEC threshold under the constraints $\lambda'(0)\rho_2'(1) \le \beta < 1$ (for BD) or $\lambda'(0)\,C \le \beta < 1$ (for MAP), for some $0 < \beta < 1$. The possible presence of stopping sets whose size grows *sublinearly* with the codeword length n imposes a judicious construction of a Tanner graph matching the optimal distribution. To this aim, the PEG algorithm is used in this paper. Note that, in general, the closer β to 1, the better the threshold, the higher the error floor.

We first constructed an irregular GLDPC ensemble for which $0 < \lambda'(0)\,C < 1 < \lambda'(0)\rho_2'(1)$. We considered a hybrid CN set composed of SPC CNs of length 7 and $(7, 4)$ CNs with $\underline{d} = 2$ and $A_2 = 5$.

We ran the constrained DE GLDPC ensemble optimization for a code rate $R = 1/3$ imposing a VN degree ranging between 2 and 15, $\lambda'(0)C \le 0.9$ and $\lambda'(0)\rho_2'(1) \ge 1.2$. The returned VN degree distribution was $\lambda(x) = 0.2000\,x + 0.7118\,x^2 + 0.0882\,x^4$ while

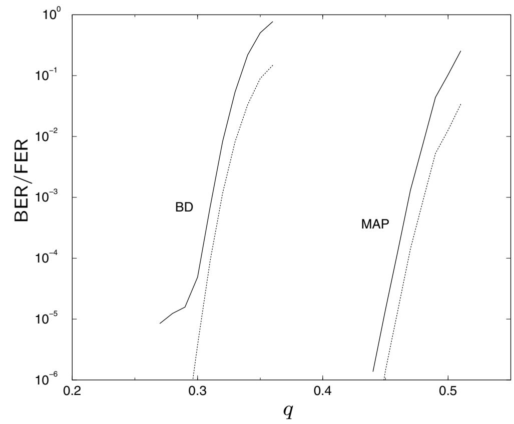

Figure 1: Performance of a $(2000, 667)$ irregular GLDPC code with $\lambda'(0)C = 0.9$ and $\lambda'(0)\rho_2'(1) = 1.2$.

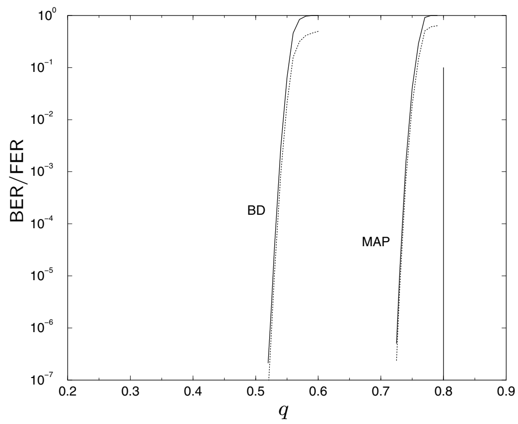

Figure 2: Performance of a $(6000, 1200)$ irregular GLDPC code with $\lambda'(0)C = \lambda'(0)\rho_2'(1) = 0.9$.

[1]Within the framework of the asymptotic stopping set distribution of irregular GLDPC codes, the role of $\lambda'(0)C$ (or $\lambda'(0)\rho_2'(1)$) is the same as that of $\lambda'(0)\rho'(1)$ for irregular LDPC code ensembles (and illustrated in [6]).

for the CN set we obtained $\rho_{SPC-7} = 0.6719$ and $\rho_{(7,4)} = 0.3281$, with threshold $q^* = 0.540389$ under MAP CN decoding. Both values of β were achieved.

A $(2000, 667)$ $R = 0.3335$ code with this distribution was constructed using the PEG algorithm. Its performance over the BEC, under iterative decoding and assuming both BD and MAP CN decodings, is depicted in Fig. 1. As expected due to the value $\lambda'(0)\rho_2'(1) = 1.2 > 1$, we observe the presence of an error floor under BD decoding at the CNs. This error floor is at $\mathsf{FER} \simeq 4 \cdot 10^{-5}$. On the other hand, no error floor is observed for the MAP counterpart down to $\mathsf{FER} \simeq 10^{-6}$.

The previous example is useful to illustrate how the same GLDPC code may have a different error floor for a different CN decoding. We then searched for good GLDPC codes with low error floor under both BD and MAP decodings at the CNs. We ran the constrained DE optimization for a code rate $R = 1/5$ requiring a VN degree between 2 and 15, a CN set composed of SPC codes with length 7 and $(7, 4)$ Hamming codes, and $\lambda'(0)\rho_2'(1) \leq 0.9$. We obtained $\lambda(x) = 0.2470\, x + 0.5164\, x^2 + 0.1666\, x^8 + 0.0700\, x^{14}$, $\rho_{SPC-7} = 0.6073$ and $\rho_{H-(7,4)} = 0.3927$, with threshold $q^* = 0.785950$ under MAP CN decoding (quite close to the Shannon limit 0.8). Again, $\beta = 0.9$ was achieved.

A $(6000, 1200)$ code was drawn from this distribution using the PEG algorithm. Its performance over the BEC is depicted in Fig. 2. This code exhibits a waterfall performance quite close to capacity under CN MAP decoding, with no error floor observed under both BD and MAP decoding down to values of FER smaller than 10^{-6}. For both CN decoding algorithms the tiny gap (in error rate) between the FER and the BER curves is an indicator of a large minimum stopping set size.

5. CONCLUSIONS

This paper investigated the design of GLDPC codes with low error floor over the BEC, under BD and MAP CN decodings. In both cases a key parameter was identified, discriminating between an exponentially small or large expected number of small size stopping sets. This analysis led to a simple code design technique.

APPENDIX I: PROOF OF LEMMA 1

Let $p_u = \binom{\bar{d}}{u}$ and $p_u = \sum_{j=1}^{\binom{\bar{d}}{u}} \delta_{u,j}$ for BD and MAP decoding at the CNs, respectively. The value of p_u represents the number of valid assignments of weight u with respect to Definition 4 and Definition 5. For a randomly chosen assignment A over the γm CNs, we consider the probability that A is valid (from a CN perspective) given that A has weight δm. We have

$$\begin{aligned} &\Pr\{\text{valid } A |\, \text{weight}(A) = \delta m\} \\ &= \frac{\Pr\{\text{weight}(A) = \delta m |\, \text{valid } A\} \cdot \Pr\{\text{valid } A\}}{\Pr\{\text{weight}(A) = \delta m\}}\,. \end{aligned} \tag{6}$$

Let $\mathcal{A}^{(\gamma m)}$ be the total number of assignments over the γm CNs. Denote by $\mathcal{A}_{\delta m}^{(\gamma m)}$ and by $\mathcal{A}_{\text{valid}}^{(\gamma m)}$ the subset of $\mathcal{A}^{(\gamma m)}$ composed of the assignments having weight δm and the subset of $\mathcal{A}^{(\gamma m)}$ composed of the assignments valid for the γm CNs, respectively. We can write

$$\Pr\{A \text{ has weight } \delta m\} = \mathcal{A}_{\delta m}^{(\gamma m)} / \mathcal{A}^{(\gamma m)}$$

and

$$\Pr\{\text{valid } A\} = \mathcal{A}_{\text{valid}}^{(\gamma m)} / \mathcal{A}^{(\gamma m)}\,.$$

Hence (6) can be written as

$$\begin{aligned} &\Pr\{\text{valid } A |\, \text{weight}(A) = \delta m\} \\ &= \frac{\Pr\{\text{weight } (A) = \delta m |\, \text{valid}(A)\} \cdot \mathcal{A}_{\text{valid}}^{(\gamma m)}}{\mathcal{A}_{\delta m}^{(\gamma m)}}\,. \end{aligned} \tag{7}$$

Furthermore, we have

$$\mathcal{A}_{\text{valid}}^{(\gamma m)} = \left[1 + \sum_{u=\underline{d}}^{\bar{d}} p_u \right]^{\gamma m}\,. \tag{8}$$

We now define X_j as the number of sockets selected in the j-th CN (out of the γm CNs) by a randomly chosen *valid* assignment over these CNs, and let $X = \sum_{j=1}^{\gamma m} X_j$ (note that the X_j's are independent). We denote by $G_j(s)$ the generating function of X_j, i.e.

$$\begin{aligned} G_j(s) &\triangleq \sum_{u=0}^{\bar{d}} \Pr\{X_j = u\} \cdot s^u \\ &= \frac{1}{1 + \sum_{u=\underline{d}}^{\bar{d}} p_u} \left(1 + \sum_{u=\underline{d}}^{\bar{d}} p_u\, s^u \right)\,. \end{aligned} \tag{9}$$

Due to the independence of the X_j's, the generating function of X, denoted by $G(s)$, is given by

$$\begin{aligned} G(s) &= \frac{1}{\left(1 + \sum_{u=\underline{d}}^{\bar{d}} p_u \right)^{\gamma m}} \left(1 + \sum_{u=\underline{d}}^{\bar{d}} p_u\, s^u \right)^{\gamma m} \\ &\triangleq \frac{1}{\mathcal{A}_{\text{valid}}^{(\gamma m)}} \left(p(s) \right)^{\gamma m} \end{aligned} \tag{10}$$

where we make use of (8). Recalling that $\Pr\{X = K\} = \frac{1}{K!} \left[\frac{\mathrm{d}^{(K)} G_X(s)}{\mathrm{d}\, s^{(K)}} \right]_{s=0}$, we can develop (7) as

$$\begin{aligned} \frac{\Pr\{X = \delta m\} \cdot \mathcal{A}_{\text{valid}}^{(\gamma m)}}{\mathcal{A}_{\delta m}^{(\gamma m)}} &= \frac{1}{\delta m!} \left[\frac{\mathrm{d}^{(\delta m)} p(s)^{\gamma m}}{\mathrm{d}\, s^{(\delta m)}} \right]_{s=0} \frac{1}{\mathcal{A}_{\delta m}^{(\gamma m)}} \\ &= \frac{\text{coeff}((p(s))^{\gamma m}, s^{\delta m})}{\mathcal{A}_{\delta m}^{(\gamma m)}}\,. \end{aligned} \tag{11}$$

As (11) represents the probability that a random assignment of weight δm is valid for the γm CNs, $\mathrm{coeff}((p(s))^{\gamma m}, s^{\delta m})$ must represent the number of assignments of weight δm valid for the γm CNs.

APPENDIX II: PROOF OF THEOREM 1

Let $p_u = \binom{\bar{d}}{u}$ and $p_u = \sum_{j=1}^{\binom{\bar{d}}{u}} \delta_{u,j}$ for BD and MAP decoding at the CNs, respectively. We make use [9, Lemma 19], recalled next for completeness.

Lemma 2. *Let $P(x) = 1 + \sum_{u=\underline{d}}^{\bar{d}} p_u x^u$ such that $p_u \in \mathbb{N}$ and $\underline{d} \geq 1$. Then*

$$\lim_{n\to\infty} \frac{1}{n} \ln \mathrm{coeff}\left((p(x))^n, x^{\alpha \bar{d} n}\right) = -\frac{\alpha\bar{d}}{\underline{d}} \ln \frac{\alpha\bar{d}}{\underline{d}} + \frac{\alpha\bar{d}}{\underline{d}} \left(1 + \ln p_{\underline{d}} + O(\alpha^2)\right).$$

Combining Lemma 1 and Lemma 2, we obtain the following asymptotic expression for the number of check-valid assignments of weight δm over γm CNs:

$$\exp\left\{m\left[-\frac{\delta}{\underline{d}} \ln \frac{\delta}{\underline{d}} + \frac{\delta}{\underline{d}} \left(1 + \ln(\gamma\, p_{\underline{d}})\right) + O(\delta^2)\right]\right\}. \tag{12}$$

Let $\boldsymbol{\omega} = [\omega_1, \ldots, \omega_{I_C}]$ be a real-valued vector such that $0 \leq \omega_i \leq 1\ \forall i$ and $\sum_i \omega_i = 1$. For an irregular GLDPC code, we wish to express the number of assignments of weight αm, valid from a CN set perspective and having $\omega_i \alpha m$ edges connected to the CNs of type i.

The number of CNs of type i is given by $m\, \frac{\rho_i}{\bar{d}_i \int \rho}$. Then, the number of valid assignments of weight $\omega_i \alpha m$ over the CNs of type i is expressed by (12) with $\gamma = \frac{\rho_i}{\bar{d}_i \int \rho}$ and $\delta = \omega_i \alpha$. By multiplying over the CN types we obtain the number of valid assignments of weight $m\alpha$ satisfying the constraint $\boldsymbol{\omega}$:

$$\begin{aligned}\prod_i \exp\Big\{m\Big[&-\frac{\omega_i\alpha}{\underline{d}_i} \ln \frac{\omega_i\alpha}{\underline{d}_i} \\ &+ \frac{\omega_i\alpha}{\underline{d}_i}\Big(1 + \ln \frac{\rho_i p_{\underline{d}}^{(i)}}{\bar{d}_i \int \rho} + O(\alpha^2)\Big)\Big]\Big\} \\ = \exp\Big\{m\sum_i\Big[&-\frac{\omega_i\alpha}{\underline{d}_i} \ln \frac{\omega_i\alpha}{\underline{d}_i} \\ &+ \frac{\omega_i\alpha}{\underline{d}_i}\Big(1 + \ln \frac{\rho_i p_{\underline{d}}^{(i)}}{\bar{d}_i \int \rho} + O(\alpha^2)\Big)\Big]\Big\}.\end{aligned} \tag{13}$$

The asymptotic expression of (13) as $m \to \infty$ is dominated by the distribution $\boldsymbol{\omega}$ which maximizes the argument of the exponential. For this reason we have to maximize (13) with respect to $\boldsymbol{\omega}$, subject to the constraint $0 \leq \omega_i \leq 1\ \forall\, i = 1, \ldots, I_C$ and $\sum_i \omega_i = 1$. By defining $\eta_i = \alpha\omega_i$, (13) can be written in the form

$$f(\boldsymbol{\eta}) = \exp\Big\{m\Big[\sum_i \frac{\eta_i}{\underline{d}_i} \ln \frac{e\rho_i p_{\underline{d}}^{(i)} \underline{d}_i}{\eta_i \bar{d}_i \int \rho} + O(\alpha^2)\Big]\Big\}, \tag{14}$$

where $\boldsymbol{\eta} = [\eta_1, \ldots, \eta_{IC}]$ and e is the Napier's number. The problem consists of maximizing $f(\boldsymbol{\eta})$ subject to $\sum_i \eta_i = \alpha$. Letting $g(\boldsymbol{\eta}) = \sum_i \eta_i - \alpha$, we have

$$\frac{\partial f}{\partial \eta_j} = \frac{1}{\underline{d}_j} \ln \frac{\rho_j p_{\underline{d}}^{(j)} \underline{d}_j}{\eta_j \bar{d}_j \int \rho} \qquad \text{and} \qquad \frac{\partial g}{\partial \eta_j} = 1$$

so that we obtain the system of $I_C + 1$ equations

$$\begin{cases} \mu + \frac{1}{\underline{d}_j} \ln \frac{\rho_j p_{\underline{d}}^{(j)} \underline{d}_j}{\eta_j \bar{d}_j \int \rho} = 0 \qquad j = 1, \ldots, I_C \\ \sum_i \eta_i = \alpha \end{cases} \tag{15}$$

in the unknowns $\boldsymbol{\eta}$ and μ, where μ is the Lagrange multiplier. From each of the first I_C equations we obtain

$$\eta_i = \frac{\rho_i K_i}{\int \rho} e^{\underline{d}_i \mu} \tag{16}$$

where $K_i \triangleq \frac{\underline{d}_i p_{\underline{d}}^{(i)}}{\bar{d}_i}$. Substituting into the last equation of (15), and letting $z = e^\mu$, we obtain

$$\sum_i \rho_i \frac{K_i}{\int \rho} z^{\underline{d}_i} = \alpha. \tag{17}$$

We are interested in valid assignments of weight $m\alpha$ with α small. The expression on the LHS of (17) is a polynomial in z with positive coefficients. For small enough α, the LHS of (17) is dominated by the lowest degree terms, i.e. by those terms associated with the ($d_{\min} = r$) CNs. For small enough α (17) is then approximated by $\sum_{i|\underline{d}_i=r} \rho_i \frac{K_i}{\int \rho} z^r = \alpha$, which can be written as $z^r \sum_{i|\underline{d}_i=r} \rho_i \frac{K_i}{\int \rho} = \alpha$, thus leading to

$$z = \left(\frac{\alpha \int \rho}{K}\right)^{1/r} \tag{18}$$

where $K \triangleq \sum_{i|\underline{d}_i=r} \rho_i K_i$. It is readily shown that for $r = 2$ we have

$$K = \begin{cases} \rho_2'(1) & \text{for BD CN decoding} \\ C & \text{for MAP CN decoding} \end{cases} \tag{19}$$

$\rho_2'(1)$ and C being defined in (2) and (3), respectively. Combining (16) and (18), we obtain the following solution to the optimization problem:

$$\omega_i = \frac{\eta_i}{\alpha} = \begin{cases} \rho_i K_i / K & \text{if} \quad \underline{d}_i = r \\ 0 & \text{if} \quad \underline{d}_i \geq r + 1. \end{cases}$$

If we now substitute the maximizing $\boldsymbol{\omega}$ into (14), after a few calculations we obtain the following asymptotic expression for the the number of assignments of weight αm, valid from a CN set perspective:

$$\exp\left\{m\left[-\frac{\alpha}{r}\ln\frac{\alpha}{r}+\frac{\alpha}{r}\left(1+\ln\frac{K}{r\int\rho}\right)+O(\alpha^2)\right]\right\}. \tag{20}$$

The probability that a randomly chosen assignment having weight αm is recognized as valid by the CN set can be computed as the ratio between (20) and the total number of assignments of weight αm. Since the total number of edges towards the CN set is equal to $m/\int\rho$, the number of assignments of weight αm is $\binom{m/\int\rho}{\alpha m}$, which for large m can be further developed as[2]

$$\binom{m/\int\rho}{\alpha m}\approx\exp\left\{m\left[-\alpha\ln\alpha+\alpha\left(1+\ln\frac{1}{\int\rho}\right)+O(\alpha^2)\right]\right\}. \tag{21}$$

Dividing (20) by (21) leads to $\exp\{m\Gamma\}$ as the probability that a randomly chosen assignment of weight αm is valid for the CN set, with

$$\Gamma=\frac{(r-1)\alpha}{r}\left(\ln\frac{\alpha}{r}-1+\ln\left(r\,K^{\frac{1}{r-1}}\int\rho\right)\right)+O(\alpha^2)$$

For $r=2$ we obtain

$$\exp\left\{m\left[\frac{\alpha}{2}\ln\frac{\alpha}{2}-\frac{\alpha}{2}\left(1-\ln\left(2K\int\rho\right)\right)+O(\alpha^2)\right]\right\}. \tag{22}$$

Analogous calculations can be developed to express the number of assignments of weight νn valid from a VN set perspective. For $p=2$, the obtained expression, formally equal to (20), is given by

$$\exp\left\{n\left[-\frac{\nu}{2}\ln\frac{\nu}{2}+\frac{\nu}{2}\left(1+\ln\frac{\lambda'(0)}{2\int\lambda}\right)+O(\nu^2)\right]\right\}. \tag{23}$$

In the limit as n tends to infinity, the average number of GLDPC stopping sets, associated with assignments of weight νn, is obtained as the product between the number of assignments of weight νn valid from a VN set perspective and the probability that an assignment of weight νn is valid from a CN set perspective. We then multiply (23) by (22), imposing the constraint $\nu\,n=\alpha\,m$. Using (1) this product is equal to

$$\exp\left\{n\left[\frac{\nu}{2}\ln\left(\lambda'(0)\,K\right)+O(\nu^2)\right]\right\}. \tag{24}$$

[2]Here we make use of the expansion $\binom{\beta m}{\alpha m}=\exp\left[m\left(-\alpha\ln\alpha+\alpha(1+\ln\beta)+O(\alpha^2)\right)\right]\cdot\left(1+O\left(1/\sqrt{n}\right)\right)$, valid for $0<\alpha<\beta<1$ (and following from Stirling's approximation). Note that, in our case, we have $\beta=1/\int\rho$.

Taking the logarithm of (24) and further dividing by n we obtain the first-order Taylor series of the growth rate of the number of valid assignments of weight νn, each one associated with a stopping set. The first order term of the Taylor series is dominated by the degree-2 VNs and ($d_{\min}=2$) CNs, that is, by subgraphs composed only of degree-2 VNs and ($d_{\min}=2$) CNs. Therefore, denoting by wn the stopping set size, we have $\nu n=2wn$ which substituted into (24) leads to (5) through (19).

References

[1] R. Gallager, *Low-Density Parity-Check Codes*. Cambridge, Massachussets: M.I.T. Press, 1963.

[2] M. Tanner, "A recursive approach to low complexity codes," *IEEE Trans. Inform. Theory*, vol. 27, pp. 533–547, Sept. 1981.

[3] C. Di, D. Proietti, I. E. Telatar, T. J. Richardson and R. 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, June 2002.

[4] N. Miladinovic and M. Fossorier, "Generalized LDPC codes and generalized stopping sets," *IEEE Trans. Commun.*, vol. 56, no. 2, pp. 201–212, Feb. 2008.

[5] E. Paolini, M. Fossorier and M. Chiani, "Doubly-generalized LDPC codes: Stability bound over the BEC," *IEEE Trans. Inform. Theory*, revised, Aug. 2008.

[6] 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, Mar. 2005.

[7] M. Flanagan, E. Paolini, M. Chiani and M. Fossorier, "On the growth rate of the weight distribution of irregular doubly-generalized LDPC codes," in *Proc. of the Forty-Sixth Allerton Conf. on Communications, Control & Computing*, Monticello, IL, USA, Sept. 2008.

[8] K. Price and R. Storn, "Differential evolution – a simple and efficient heuristic for global optimization over continuous spaces," *J. Global Optimization*, vol. 11, pp. 341–359, 1997.

[9] C. Di, T. Richardson and R. Urbanke, "Weight distribution of low-density parity-check codes," *IEEE Trans. Inform. Theory*, vol. 52, no. 11, pp. 4839–4855, Nov. 2006.