

# On the Growth Rate of Irregular GLDPC Codes Weight Distribution

Enrico Paolini and Marco Chiani  
DEIS, WiLAB, University of Bologna  
via Venezia 52, 47023 Cesena (FC), Italy  
{e.paolini,marco.chiani}@unibo.it

Marc Fossorier  
ETIS ENSEA / UCP / CNRS UMR-8051  
6, avenue du Ponceau, 95014, Cergy Pontoise, France  
mfossorier@ieee.org

**Abstract**—In this paper the exponential growth rate of irregular generalized low-density parity-check (GLDPC) codes weight distribution is considered. Specifically, the Taylor series of the growth rate is expanded to the first order with the purpose of studying its behavior in correspondence with the small weight codewords. It is proved that the linear term of the Taylor series, and then the expected number of small linear-sized weight codewords of a randomly chosen GLDPC code in the irregular ensemble, is dominated by the degree-2 variable nodes and by the check nodes with minimum distance 2. A parameter is introduced, only depending on such variable and check nodes, discriminating between an exponentially small and exponentially large expected number of small weight codewords.

## I. INTRODUCTION

Recently, generalized LDPC (GLDPC) codes have gained an increasing interest as a promising solution for low-rate channel coding schemes, offering a good compromise between waterfall performance and error floor under belief-propagation decoding. Introduced in [1], GLDPC codes represent a generalization of low-density parity-check (LDPC) codes [2]. Using the same nomenclature as for LDPC codes, a GLDPC code is said regular when all the variable nodes (VNs) have the same degree and all the check nodes (CNs) are of the same type, and irregular otherwise. In the bipartite graph of a traditional irregular LDPC code, the variable node decoder (VND) is composed of a mixture of repetition codes (variable nodes) and the check node decoder (CND) of a mixture of single parity-check (SPC) codes (check nodes). In an irregular GLDPC code, the VND is still composed of a mixture of repetition codes, while the CND is composed of a mixture of generic linear block codes. Examples of effective irregular GLDPC code constructions can be found, for instance, in [3] (CND composed of a mixture of Hamming and SPC CNs) or [4] (quasi-cyclic construction based on protographs leading to low error floors).

It is commonly accepted that using generalized CNs with good distance properties provides benefits from a minimum distance, and thus error floor, point of view. Even though results have been developed for regular GLDPC codes (e.g., [5, Theorem 1]) or for GLDPC codes with an irregular VND and a uniform CND, composed of CNs all of the same type, to the best of the authors' knowledge a study of the minimum distance properties of irregular GLDPC code ensembles has not been developed yet.

An analysis of the weight distribution of irregular LDPC codes has been proposed in [6]–[8]. For large values of the codeword length  $n$ , it is common to refer to the concept of asymptotic exponential growth rate of the weight distribution [2], which expresses the dominant asymptotic behavior. Following [8], the growth rate is defined in this paper as

$$G(w) \triangleq \lim_{n \rightarrow \infty} \frac{1}{n} \ln \mathbb{E} [N(\mathcal{G}), \lfloor wn \rfloor], \quad (1)$$

where  $\mathbb{E} [N(\mathcal{G}), \lfloor wn \rfloor]$  denotes the average number of codewords of weight  $\lfloor wn \rfloor$  associated with the bipartite graph  $\mathcal{G}$  randomly chosen within the ensemble.

In [8], a complete solution for the growth rate of irregular LDPC codes is given considering both the unstructured standard ensemble, that is the ensemble of all the possible LDPC codes with a given degree distribution, and the expurgated ensemble where catastrophic zig-zag cycles involving degree-2 VNs are removed. The first step towards a complete solution for the growth rate consisted in the evaluation of the growth rate of the small and linear-sized weight codewords, that is (1) with  $w$  close to zero, obtained by expanding the Taylor series of  $G(w)$  with starting point  $w_0 = 0$  and truncating it at the first order. The obtained result

$$G(w) = w \ln (\lambda'(0)\rho'(1)) + O(w^2), \quad (2)$$

makes evident the fundamental role played by the parameter  $\lambda'(0)\rho'(1)$  from the point of view of the growth rate of small linear-sized weight codewords. Starting from this result, a probabilistic analysis of the minimum distance of a randomly chosen LDPC codes from unstructured ensembles has been developed in [8].

In this paper, (2) is extended to unstructured irregular GLDPC code ensembles, assuming that each node in the bipartite graph is associated with a linear block code whose minimum distance is at least 2 (for the repetition VNs this is equivalent to requiring a degree at least 2). The main result of this work is expressed by Theorem 1 of Section III. It provides an insight about the distance properties of irregular GLDPC code ensembles. By comparing (2) and (4) we see that, when moving from LDPC codes to the more general class of GLDPC codes, the parameter  $\lambda'(0)\rho'(1)$  is generalized by  $\lambda'(0)C$ , where  $C$  is defined in Section III

only depending on the CNs with minimum distance 2<sup>1</sup>. This work represents an extension of [10], where the growth rate  $G(w)$  was investigated, for small  $w$ , for GLDPC codes with a uniform CN. We show here that, when a mixture of check component codes is considered, the growth rate of the number of small weight codewords is dominated by the CN types with the smallest minimum distance.

In Section II some notation is introduced. Our main result is claimed and proved in Section III. Concluding remarks are given in Section IV.

## II. NOTATION

We assume that the CN is composed by  $I_C$  different linear block code types. For each CN we denote by  $h$ ,  $\bar{d}$  and  $\underline{d}$  the CN dimension, the CN length and the CN minimum distance, respectively. The length  $\bar{d}$  of a CN is equal to the number of edges connected to it and is also referred to as the CN degree. The symbol  $\sum^{[w]}$  is used to denote the summation over those CN or the VN types having minimum distance  $w$ .

The unstructured GLDPC code ensemble is defined from an edge perspective by specifying the GLDPC codeword length  $n$ , the  $I_C$  different CN types and the VN and CN degree distributions  $\lambda(x)$  and  $\rho(x)$ , where  $\lambda(x) = \sum_{i \geq 2} \lambda_i x^{i-1}$  is the standard edge-oriented LDPC degree distribution, while  $\rho(x)$  is defined as

$$\rho(x) = \sum_{i=1}^{I_C} \rho_i x^{\bar{d}_i - 1},$$

where  $\rho_i$  is the fraction of edges connected to the CNs of type  $i$ . Denoting as usual  $\int_0^1 \lambda(x) dx$  and  $\int_0^1 \rho(x) dx$  by  $\int \lambda$  and  $\int \rho$ , respectively, we have that

$$\frac{n}{m} = \frac{\int \lambda}{\int \rho}, \quad (3)$$

since  $m/\int \rho = n/\int \lambda$  is the total number of edges. We denote by  $p$  and  $r$  the smallest minimum distance among the VNs and CNs, respectively.

Most of the calculations in Section III rely on the concept of *assignment* that was introduced in [8] for LDPC codes, here extended to GLDPC codes. Given the bipartite graph of a GLDPC code, an assignment is by definition any subset of the graph edges. The number of edges composing an assignment is known as the assignment weight. An assignment is said valid from the CN perspective when the following condition holds: supposing that each edge of the assignment carries a “1” and that each of the other edges carries a “0”, each CN recognizes a valid codeword. Analogously, an assignment is said valid from the VND perspective when the following condition holds: supposing that each edge of the assignment carries a “1” and that each of the other edges carries a “0”, each VN recognizes a valid codeword. An assignment is said

<sup>1</sup>It is worthwhile remarking that  $\lambda'(0)C$  generalizes  $\lambda'(0)\rho'(1)$  also in the context of the stability condition over the binary erasure channel. The stability condition can be expressed as  $q^* \leq [\lambda'(0)C]^{-1}$  ( $q^*$  being the asymptotic threshold) [9], where for LDPC codes we have  $q^* \leq [\lambda'(0)\rho'(1)]^{-1}$ .

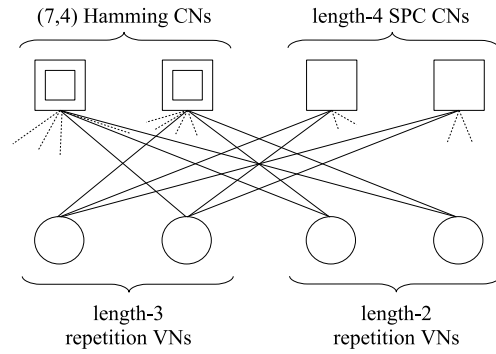


Fig. 1. Example of valid assignment of weight 10. The solid edges belong to the assignments, the dashed edges do not belong to it.

valid when it is recognized as valid by both the CN and the VND. Each valid assignment is associated with one and only one GLDPC code codeword, and vice-versa.

*Example 1:* In Fig. 1 a valid assignment of weight 10 for a GLDPC code is depicted. This assignment involves two (7, 4) Hamming CNs with parity-check matrix

$$\begin{bmatrix} 1 & 1 & 0 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 1 \end{bmatrix}$$

and two length-4 SPC CNs on the CN side; it involves two length-3 repetition VNs and two length-2 repetition VNs on the VND side. The dashed sockets correspond to edges which do not belong to the assignment (so that they carry a “0”). The assignment is valid, since it is recognized as valid by both the CN and the VND. In fact, the leftmost Hamming CN recognizes the valid codeword  $[0, 0, 0, 1, 1, 0, 1]$ , the other Hamming CN recognizes the valid codeword  $[0, 0, 1, 0, 0, 1, 1]$ , and each of the two SPC CNs recognizes a valid weight-2 codeword. Moreover, each of the repetition VNs recognizes its only non-null codeword. Any other CN or VN not involved in the assignment recognizes its all-zero codeword. Since each repetition VN is associated with one GLDPC encoded bit, this valid assignment corresponds to a weight-4 GLDPC codeword.

## III. GROWTH RATE TAYLOR SERIES FOR GLDPC CODES

For the CNs of type  $i$  we define

$$C_i \triangleq \frac{d_i A_d^{(i)}}{\bar{d}_i},$$

and

$$C \triangleq \sum_i^{[r]} \rho_i C_i$$

where  $A_u^{(i)}$  is the multiplicity of the weight- $u$  codewords for the type- $i$  CNs. For the particular case of a non-generalized LDPC code it is readily shown that  $C = \rho'(1)$ . Then we have the following result.

*Theorem 1:* For an irregular GLDPC code ensemble with  $p = r = 2$ , the Taylor expansion of the weight distribution

growth rate, computed assuming  $w_0 = 0$  as starting point, is given by

$$G(w) = w \ln(\lambda'(0) C) + O(w^2). \quad (4)$$

The theorem is proved next. For clearness, the proof is here organized into five steps. With respect to the analogous proof given in [8] for LDPC codes, the main contributions are represented by Subsection III-A (where the calculation of the number of valid assignments, from a CN perspective, of some weight over a certain number of generalized CNs of the same type is extended to GLDPC codes using the generating function of random variables) and by Subsection III-C (where it is proved that the first order term of the Taylor series is dominated by the minimum distance 2 CNs).

#### A. Number of valid assignments of weight $\delta m$ over $\gamma m$ generalized CNs

Let us consider  $\gamma m$  CNs of the same type, and let us calculate the number of assignments of weight  $\delta m$  which are recognized as valid by the  $\gamma m$  CNs. For a randomly chosen assignment  $A$  over the  $\gamma m$  CNs, we consider the probability that  $A$  is valid given that  $A$  has weight  $\delta m$ . We have

$$\begin{aligned} & \Pr\{A \text{ is valid} \mid A \text{ has weight } \delta m\} \\ &= \frac{\Pr\{A \text{ has weight } \delta m \mid A \text{ is valid}\} \cdot \Pr\{A \text{ is valid}\}}{\Pr\{A \text{ has weight } \delta m\}}. \end{aligned} \quad (5)$$

Let  $\mathcal{A}^{(\gamma m)}$  be the total number of assignments over the  $\gamma m$  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  $\delta m$  and the subset of  $\mathcal{A}^{(\gamma m)}$  composed of the assignments recognized as valid by the  $\gamma m$  CNs, respectively. Then we can write

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

and

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

Hence, (5) can be written as

$$\begin{aligned} & \Pr\{A \text{ is valid} \mid A \text{ has weight } \delta m\} \\ &= \frac{\Pr\{A \text{ has weight } \delta m \mid A \text{ is valid}\} \cdot \mathcal{A}_{\text{valid}}^{(\gamma m)}}{\mathcal{A}_{\delta m}^{(\gamma m)}}. \end{aligned} \quad (6)$$

Since each CN has  $2^h$  codewords and has  $\bar{d}$  sockets, we have

$$\mathcal{A}_{\text{valid}}^{(\gamma m)} = 2^{\gamma m h} \quad \text{and} \quad \mathcal{A}_{\delta m}^{(\gamma m)} = \binom{\gamma m \bar{d}}{\delta m}. \quad (7)$$

We now define the random variable (r.v.)  $X_j$  as the number of sockets selected by a *valid* assignment in the  $j$ -th CN out of the  $\gamma m$  CNs, and the r.v.  $X$  as

$$X = \sum_{j=1}^{\gamma m} X_j,$$

where the  $X_j$ 's are independent. We denote by  $G_j(s)$  the generating function of  $X_j$  (that is the probability mass function in polynomial form), which for all  $j$  is given by

$$\begin{aligned} G_j(s) &= \sum_{u=0}^{\bar{d}} \Pr\{X_j = u\} \cdot s^u \\ &= \frac{1}{2^h} \left( 1 + \sum_{u=\underline{d}}^{\bar{d}} A_u s^u \right), \end{aligned} \quad (8)$$

being  $\Pr\{X_j = u\} = A_u/2^h$ . Due to the independence of the  $X_j$ 's, the generating function of  $X$ , denoted by  $G(s)$ , is given by

$$\begin{aligned} G(s) &= \frac{1}{2^{\gamma m h}} \left[ 1 + \sum_{u=\underline{d}}^{\bar{d}} A_u s^u \right]^{\gamma m} \\ &= \frac{1}{2^{\gamma m h}} P(s)^{\gamma m} \end{aligned} \quad (9)$$

where, by definition,  $P(s) = 1 + \sum_{u=\underline{d}}^{\bar{d}} A_u s^u$ .

Recalling that  $\Pr\{X = K\} = \frac{1}{K!} \left[ \frac{d^{(K)} G_X(s)}{d s^{(K)}} \right]_{s=0}$  and using (7), we can develop (6) as

$$\begin{aligned} & \frac{\Pr\{X = \delta m\} \cdot 2^{\gamma m h}}{\binom{\gamma m \bar{d}}{\delta m}} \\ &= \frac{1}{\delta m!} \left[ \frac{d^{(\delta m)} P(s)^{\gamma m}}{d s^{(\delta m)}} \right]_{s=0} \frac{1}{\binom{\gamma m \bar{d}}{\delta m}} \\ &= \frac{\text{coeff}(P(s)^{\gamma m}, s^{\delta m})}{\binom{\gamma m \bar{d}}{\delta m}}, \end{aligned} \quad (10)$$

where we use the notation  $\text{coeff}(Q(s), s^u)$  to denote the coefficient of  $s^u$  in a polynomial  $Q(s)$ . In fact, the derivative of order  $\delta m$  of the polynomial  $Q(s) = P(s)^{\gamma m}$  is equal to  $(\delta m!)$  times the coefficient of  $s^{\delta m}$  in  $P(s)^{\gamma m}$ . By comparing (10) with (5), we recognize that  $\text{coeff}(P(s)^{\gamma m}, s^{\delta m})$  is the number of valid assignments of weight  $\delta m$  over the  $\gamma m$  CNs having the same type. In order to compute the growth rate of this quantity for  $\delta$  close to 0, we exploit [8, Lemma 19], reported next.

*Lemma 1:* 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

$$\begin{aligned} & \lim_{n \rightarrow \infty} \frac{1}{n} \ln \text{coeff}(P(x)^n, x^{\alpha \bar{d} n}) \\ &= -\frac{\alpha \bar{d}}{\underline{d}} \ln \frac{\alpha \bar{d}}{\underline{d}} + \frac{\alpha \bar{d}}{\underline{d}} (1 + \ln p_{\underline{d}} + O(\alpha^2)). \end{aligned}$$

By applying Lemma 1, we obtain

$$\begin{aligned} & \text{coeff}(P(s)^{\gamma m}, s^{\delta m}) \\ &= \exp \left\{ m \left[ -\frac{\delta}{\underline{d}} \ln \frac{\delta}{\underline{d}} + \frac{\delta}{\underline{d}} (1 + \ln(\gamma A_{\underline{d}})) + O(\delta^2) \right] \right\} \end{aligned} \quad (11)$$

as an expression (valid asymptotically) for the number of valid assignments of weight  $\delta m$  over the  $\gamma m$  CNs. So, (11)

generalizes the analogous expression obtained in [8] for non-generalized LDPC codes. This generalization is not straightforward, since the number of weight  $\delta m$  assignments for non-generalized LDPC codes where only SPC CNs are involved can be easily expressed, but it is not the case in general.

### B. Number of *CND* valid assignments of weight $\alpha m$ with constrained edges repartition

Let  $\omega = [\omega_1, \dots, \omega_{I_C}]$  be a real vector such that  $0 \leq \omega_i \leq 1 \forall i$  and  $\sum_i \omega_i = 1$ . We now consider the *CND* of an irregular GLDPC code, with the aim of expressing the number of assignments of weight  $\alpha m$ , recognized as valid by the *CND*, and having  $\omega_i \alpha m$  edges connected to the CNs of type  $i$ .

The number of CNs of type  $i$  is given by  $m \frac{\rho_i}{d_i \int \rho}$ . Then, the number of valid assignments of weight  $\omega_i \alpha m$  over the CNs of type  $i$  is expressed by (11) with  $\gamma = \frac{\rho_i}{d_i \int \rho}$  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  $\omega$ :

$$\begin{aligned} & \prod_i \exp \left\{ m \left[ -\frac{\omega_i \alpha}{d_i} \ln \frac{\omega_i \alpha}{d_i} \right. \right. \\ & \quad \left. \left. + \frac{\omega_i \alpha}{d_i} \left( 1 + \ln \left( \frac{\rho_i}{d_i \int \rho} A_d^{(i)} \right) + O(\alpha^2) \right) \right] \right\} \\ &= \exp \left\{ m \sum_i \left[ -\frac{\omega_i \alpha}{d_i} \ln \frac{\omega_i \alpha}{d_i} \right. \right. \\ & \quad \left. \left. + \frac{\omega_i \alpha}{d_i} \left( 1 + \ln \left( \frac{\rho_i}{d_i \int \rho} A_d^{(i)} \right) + O(\alpha^2) \right) \right] \right\}. \end{aligned} \quad (12)$$

### C. Dominant term evaluation

When evaluating the growth rate of (12), we are interested in the dominant term. For this reason we have to maximize (12) subject to the constraint  $\sum_i \omega_i = 1$  and  $0 \leq \omega_i \leq 1 \forall i$ . By defining  $\eta_i = \alpha \omega_i$ , (12) can be written in the form

$$\exp \left\{ m \left[ \sum_i \frac{\eta_i}{d_i} \ln \frac{e \rho_i A_d^{(i)} d_i}{\eta_i d_i \int \rho} + O(\alpha^2) \right] \right\}, \quad (13)$$

where  $e$  is the Neper's number. The problem then consists in maximizing  $f(\eta_1, \dots, \eta_{I_C})$  given by (13) subject to  $\sum_i \eta_i = \alpha$ . Defining  $g(\eta_1, \dots, \eta_{I_C}) = \sum_i \eta_i - \alpha$ , we have

$$\frac{\partial f}{\partial \eta_j} = \frac{1}{d_j} \ln \frac{\rho_j A_d^{(j)} d_j}{\eta_j d_j \int \rho} \quad \text{and} \quad \frac{\partial g}{\partial \eta_j} = 1,$$

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

$$\begin{cases} \mu + (1/d_j) \cdot \ln \frac{\rho_j A_d^{(j)} d_j}{\eta_j d_j \int \rho} = 0 & j = 1, \dots, I_C \\ \sum_i \eta_i = \alpha \end{cases} \quad (14)$$

in the unknowns  $\eta_i$  ( $i = 1, \dots, I_C$ ) and  $\mu$ , where  $\mu$  is the Lagrange multiplier. From each of the first  $I_C$  equations

$$\begin{aligned} \eta_j &= \frac{\rho_j A_d^{(j)} d_j}{d_j \int \rho} \cdot e^{d_j \mu} \\ &= \frac{\rho_j C_j}{\int \rho} \cdot e^{d_j \mu}, \end{aligned} \quad (15)$$

which substituted into the last equation of (14) leads to

$$\sum_i \rho_i \frac{C_i}{\int \rho} \cdot z^{d_i} = \alpha \quad (16)$$

where  $z = e^\mu$  by definition.

Since the aim is to evaluate the Taylor expansion of the growth rate  $G(w)$  truncated at the first order, we are interested in valid assignments of weight  $m \alpha$  with  $\alpha$  small. The expression at the LHS of (16) is a polynomial in  $z$  with positive coefficients. For small enough  $\alpha$ , the LHS of (16) is dominated by the lowest degree terms, i.e. by those terms which are associated with the CNs having  $d_{\min} = r$ . For small enough  $\alpha$  (16) is then approximated by

$$\sum_i^{[r]} \rho_i \frac{C_i}{\int \rho} \cdot z^r = \alpha, \quad (17)$$

which can be written as  $z^r \cdot \sum_i^{[r]} \rho_i \frac{C_i}{\int \rho} = \alpha$ , thus leading to

$$\begin{aligned} z &= \left( \frac{\alpha \int \rho}{\sum_i^{[r]} \rho_i C_i} \right)^{1/r} \\ &= \left( \frac{\alpha \int \rho}{C} \right)^{1/r}. \end{aligned} \quad (18)$$

Combining (15) and (18), we obtain the following solution to the optimization problem:

$$\omega_i = \frac{\eta_i}{\alpha} = \begin{cases} \rho_i C_i / C & \text{if } d_i = r \\ 0 & \text{if } d_i \geq r + 1. \end{cases}$$

If we now substitute the obtained expression of the  $\omega_i$ 's into (13), after a few calculations we finally obtain the following expression for the dominant term of the number of valid assignments of weight  $\alpha m$ :

$$\exp \left\{ m \left[ -\frac{\alpha}{r} \ln \frac{\alpha}{r} + \frac{\alpha}{r} \left( 1 + \ln \frac{C}{r \int \rho} \right) + O(\alpha^2) \right] \right\}. \quad (19)$$

### D. Probability that a randomly chosen assignment of weight $\alpha m$ is valid from a *CND* perspective

The probability that a randomly chosen assignment having weight  $\alpha m$  is recognized as valid by the *CND* can be computed as the ratio between the number of valid assignments of weight  $\alpha m$ , expressed by (19), and the total number of assignments of weight  $\alpha m$ . Since the total number of edges connected to the *CND* is equal to  $m / \int \rho$ , the number of assignments of weight  $\alpha m$  is given by  $\binom{m / \int \rho}{\alpha m}$ , which can be further developed using the expansion (which follows from Stirling's approximation)

$$\begin{aligned} \binom{\beta m}{\alpha m} &= \exp [m (-\alpha \ln \alpha + \alpha (1 + \ln \beta) + O(\alpha^2))] \\ &\quad \cdot \left( 1 + O \left( n^{-\frac{1}{2}} \right) \right), \end{aligned}$$

valid for  $0 < \alpha < \beta < 1$ . By applying this result to  $\binom{m / \int \rho}{\alpha m}$  with  $\beta = 1 / \int \rho$ , we obtain for large  $m$

$$\binom{m / \int \rho}{\alpha m} \approx \exp \left\{ m \left[ -\alpha \ln \alpha + \alpha \left( 1 + \ln \frac{1}{\int \rho} \right) + O(\alpha^2) \right] \right\}. \quad (20)$$

Dividing (19) by (20) leads to the following expression of the probability that a randomly chosen assignment of weight  $\alpha m$  is valid for the VND:

$$\frac{(r-1)\alpha}{r} \ln \frac{\alpha}{r} - \frac{(r-1)\alpha}{r} \left( 1 - \ln \left( r C^{\frac{1}{r-1}} \int \rho \right) \right) + O(\alpha^2), \quad (21)$$

which for  $r = 2$  is specialized

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

Expression (22) generalizes [8, Lemma 20] to irregular GLDPC code ensembles.

#### E. Number of VND valid assignments of weight $\nu n$

The first three steps of the proof can be applied also to the VND, in order to express the number of assignments of weight  $\nu n$  which are recognized as valid by the VND. The obtained expression is analogous to (19) and is given by

$$\exp \left\{ n \left[ -\frac{\nu}{p} \ln \frac{\nu}{p} + \frac{\nu}{p} \left( 1 + \ln \frac{V}{p \int \lambda} \right) + O(\nu^2) \right] \right\}, \quad (23)$$

where

$$V \triangleq \sum_i^{[p]} \lambda_i \frac{2A_2^{(i)}}{\bar{d}_i}$$

the summation being over the VN types with minimum distance  $p$ . Since for GLDPC codes all the VNs are repetition codes, the only  $d_{\min} = p$  VNs are the degree- $p$  VNs, for which  $A_p = 1$  and  $\bar{d} = p$ . For  $p = 2$  we simply have  $V = \lambda'(0)$ , through which (23) becomes

$$\exp \left\{ n \left[ -\frac{\nu}{2} \ln \frac{\nu}{2} + \frac{\nu}{2} \left( 1 + \ln \frac{\lambda'(0)}{2 \int \lambda} \right) + O(\nu^2) \right] \right\}. \quad (24)$$

In the limit where the codeword length  $n$  tends to infinity, the average number of GLDPC codewords is obtained as the product between the number of assignments of a certain weight recognized as valid by the VND and the probability that an assignment of the same weight is recognized as valid by the VND. We then multiply (24) by (22), imposing the constraint

$$\nu n = \alpha m. \quad (25)$$

Recalling (3) this product is equal to

$$\exp \left\{ n \left[ \frac{\nu}{2} \ln (\lambda'(0) C) + O(\nu^2) \right] \right\}. \quad (26)$$

Taking the logarithm of (26) and further dividing by  $n$  we obtain the Taylor expansion (to the first order) for the growth rate of the number of assignments having weight  $\nu n$  and recognized as valid by both the VND and VND, and thus associated each one with a valid GLDPC codeword. As evident from the domination argument of the proof, the first order term of the Taylor series is dominated by the variable and check nodes having minimum distance 2, that is, by subgraphs composed only of  $d_{\min} = 2$  VNs (degree-2 VNs) and  $d_{\min} = 2$  CNs. The codewords associated with

these assignments are characterized by all the “1” positions corresponding to degree-2 VNs. Therefore, denoting by  $wn$  the codeword weight, we have  $\nu n = 2wn$  which substituted into (26) leads to (4).

We observe that in the general case where either  $p > 2$  or  $r > 2$  the expression  $G(w)$  for small  $w$  is given by

$$G(w) = wp \ln \left[ \left( \frac{wp \int \lambda}{e} \right)^{1 - \frac{1}{p} - \frac{1}{r}} V^{\frac{1}{p}} C^{\frac{1}{r}} \right], \quad (27)$$

which can be obtained by multiplying (21) by (23), and using (3) and (25). Equation (27) generalizes the analogous result from [10] to the case of a hybrid VND. For small enough  $w$ ,  $G(w)$  is always negative so that an exponentially small number of small linear-sized weight codewords is expected.

#### IV. CONCLUSION

The growth rate  $G(w)$  of the small linear-sized weight codewords has been investigated for unstructured irregular GLDPC code ensembles, by calculating the linear term of its Taylor series around  $w_0 = 0$ . It has been proved that, if both  $d_{\min} = 2$  CNs and degree-2 VNs are present, the expected number of small weight codewords of a randomly chosen GLDPC code is dominated by subgraphs, namely valid assignments, only involving such CNs and VNs. The parameter  $\lambda'(0)C$  has been introduced, discriminating between irregular GLDPC code ensembles where the expected number of small weight codewords is exponentially small ( $\lambda'(0)C < 1$ ) or exponentially large ( $\lambda'(0)C > 1$ ). This result generalizes as well the connection with the stability condition over the BEC.

#### ACKNOWLEDGMENT

The work was supported in part by the EC under Seventh FP grant agreement ICT OPTIMIX n.INFSO-ICT-214625.

#### REFERENCES

- [1] R. M. Tanner, “A recursive approach to low complexity codes,” *IEEE Trans. Inform. Theory*, vol. 27, pp. 533–547, Sept. 1981.
- [2] R. Gallager, *Low-Density Parity-Check Codes*. Cambridge, MA: M.I.T. Press, 1963.
- [3] J. Chen and R. Tanner, “A hybrid coding scheme for the Gilbert-Elliott channel,” *IEEE Trans. Commun.*, vol. 54, pp. 564–571, Oct. 2006.
- [4] G. Liva, W. E. Ryan, and M. Chiani, “Quasi-cyclic generalized LDPC codes with low error floors,” *IEEE Trans. Commun.*, vol. 56, pp. 49–57, Jan. 2008.
- [5] M. Lentmaier and K. Zigangirov, “On generalized low-density parity-check codes based on Hamming component codes,” *IEEE Commun. Lett.*, vol. 3, pp. 248–250, Aug. 1999.
- [6] S. Litsyn and V. Shevelev, “On ensembles of low-density parity-check codes: Asymptotic distance distributions,” *IEEE Trans. Inform. Theory*, vol. 48, pp. 887–908, Apr. 2002.
- [7] G. Miller and D. Burshtein, “Asymptotic enumeration method for analyzing LDPC codes,” *IEEE Trans. Inform. Theory*, vol. 50, pp. 1115–1131, June 2004.
- [8] C. Di, T. Richardson, and R. Urbanke, “Weight distribution of low-density parity-check codes,” *IEEE Trans. Inform. Theory*, vol. 52, pp. 4839–4855, Nov. 2006.
- [9] E. Paolini, M. Fossorier, and M. Chiani, “Doubly-generalized LDPC codes: Stability bound over the BEC,” *IEEE Trans. Inform. Theory*, submitted.
- [10] J. P. Tillich, “The average weight distribution of Tanner code ensembles and a way to modify them to improve their weight distribution,” in *Proc. of 2004 IEEE Int. Symp. on Information Theory*, (Chicago, USA), June/July 2004.