On the Weight and Stopping Set Size Distributions of LDPC Codes and their Generalizations

Mark F. Flanagan
University College Dublin
Belfield, Dublin 4, Ireland
Email: mark.flanagan@ieee.org

Enrico Paolini
DEIS/WiLAB, University of Bologna,
Via Venezia 52, 47023 Cesena, Italy
Email: e.paolini@unibo.it

***Abstract*— This paper overviews and connects together recent results by the authors regarding the weight and stopping set size distributions of generalizations of low-density parity-check (LDPC) codes. We begin by reviewing a result which allows for evaluation of the growth rate of the weight distribution for codewords of *any* weight (linear in the codeword length). We then show how asymptotic analysis of this general result can lead to a result which encapsulates the behaviour of the growth rate of the weight distribution for the case of *small* linear-weight codewords. We then show that many existing results in the literature along this line may be viewed as special cases of this general result.**

I. Introduction

Recently, low-density parity-check (LDPC) codes have been intensively studied due to their near-Shannon-limit performance under low-complexity belief-propagation decoding. Regular LDPC codes were first proposed by Gallager in 1963 [1]. In the last decade, the capability of irregular LDPC codes to outperform regular ones in the waterfall region of the performance curve and to asymptotically approach (or even achieve) the communication channel capacity has been recognized and deeply investigated (see for instance [2]-[5]).

It is usual to represent an LDPC code as a bipartite graph, also known as a Tanner graph [6], where the nodes are grouped into two disjoint sets, namely, the variable nodes (VNs) and the check nodes (CNs), such that each edge may only connect a VN to a CN. In the Tanner graph of an LDPC code, a generic degree-q VN can be interpreted as a length-q repetition code, while a degree-s CN can be interpreted as a length-s single parity-check (SPC) code.

Prior to the rediscovery of LDPC codes, generalized LDPC (GLDPC) codes were introduced by Tanner in 1981 [6]. A GLDPC code generalizes the concept of an LDPC code in that a degree-s CN may in principle be any (s, h) linear block code, s being the code length and h the code dimension. Such a CN accounts for $(s - h)$ linearly independent parity-check equations. A CN associated with a linear block code which is not a SPC code is said to be a generalized CN.

Generalized LDPC codes represent a promising solution for low-rate channel coding schemes, due to an overall rate loss introduced by the generalized CNs [7]. Doubly-generalized LDPC (D-GLDPC) codes generalize the concept of GLDPC codes while facilitating greater design flexibility in terms of code rate [8]-[11]. In a D-GLDPC code, the VNs as well as

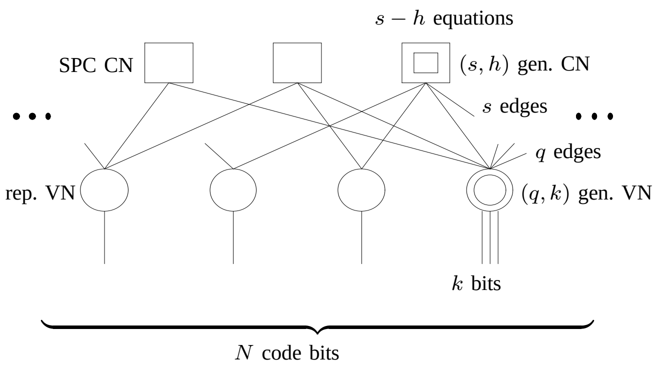

Fig. 1. Structure of a D-GLDPC code. An (s, h) generalized CN accounts for $s - h$ linearly independent parity-check equations, and contributes s edges to the Tanner graph. A (q, k) generalized VN is associated with k D-GLDPC code bits, and contributes q edges to the Tanner graph.

the CNs may be of any generic linear block code types. A degree-q VN may in principle be any (q, k) linear block code, q being the code length and k the code dimension. Such a VN is associated with k D-GLDPC code bits. It interprets these bits as its local information bits and interfaces to the CN set through its q local code bits. A VN which corresponds to a linear block code which is not a repetition code is said to be a generalized VN. A D-GLDPC code is said to be regular if all of its VNs are of the same type and all of its CNs are of the same type and is said to be irregular otherwise[1]. The structure of a D-GLDPC code is depicted in Fig. 1.

An important tool for analysis of LDPC codes and their generalizations is represented by the *growth rate of the weight distribution* (or *spectral shape*), whose definition is reviewed in Section II. It allows for analysis of the asymptotic distribution of linear-weight codewords for a code randomly chosen from a given ensemble. This concept was introduced in [1], within the context of Gallager's regular LDPC codes. More recently, the study of the growth rate of the weight distribution has been extended to irregular LDPC ensembles. Pioneering works in this area are [12]-[14]. In [14] a complete solution for

[1]Note that VNs associated with different representations of the same linear block code (i.e. with different generator matrices) are regarded as belonging to different types.

the growth rate for irregular LDPC codes was developed. The growth rate of the weight distribution of binary GLDPC codes was investigated in [15]–[18]. Related results for expander codes were developed in [19]-[21]. Also, [24] investigates the asymptotic weight enumerators of many LDPC-like codes including turbo codes and repeat-accumulate codes.

In this paper, an overview is provided of recent results by the authors regarding the growth rate of the weight distribution of D-GLDPC codes. First, a result which allows for evaluation of the growth rate for codewords of *any* weight is reviewed. It is then shown how asymptotic analysis of this general result can lead to a result which encapsulates the behaviour of the growth rate in the case of *small* linear-weight codewords. We then provide a series of corollaries which contain many existing results in the literature as special cases.

We also stress that all of the results in this paper may be extended to the growth rate of the stopping set size distribution. For the definition of stopping sets of LDPC codes, we refer the reader to [22], [23].

II. Irregular Doubly-Generalized LDPC Code Ensemble

We define a D-GLDPC code ensemble $\mathcal{M}_n$ as follows, where n denotes the number of VNs. There are n_c different CN types $t \in I_c = \{1, 2, \cdots, n_c\}$, and n_v different VN types $t \in I_v = \{1, 2, \cdots, n_v\}$. For each CN type $t \in I_c$, we denote by h_t, s_t and r_t the CN dimension, length and minimum distance, respectively. For each VN type $t \in I_v$, we denote by k_t, q_t and p_t the VN dimension, length and minimum distance, respectively. For $t \in I_c$, ρ_t denotes the fraction of edges connected to CNs of type t. Similarly, for $t \in I_v$, λ_t denotes the fraction of edges connected to VNs of type t. Note that all of these variables are independent of n.

The polynomials $\rho(x)$ and $\lambda(x)$ are defined by

$$\rho(x) = \sum_{t \in I_c} \rho_t x^{s_t - 1}$$

and

$$\lambda(x) = \sum_{t \in I_v} \lambda_t x^{q_t - 1} \ .$$

If E denotes the number of edges in the Tanner graph, the number of CNs of type $t \in I_c$ is then given by $E\rho_t/s_t$, and the number of VNs of type $t \in I_v$ is then given by $E\lambda_t/q_t$. Denoting as usual $\int_0^1 \rho(x)\,\mathrm{d}x$ and $\int_0^1 \lambda(x)\,\mathrm{d}x$ by $\int \rho$ and $\int \lambda$ respectively, we see that the number of edges in the Tanner graph is given by

$$E = \frac{n}{\int \lambda}$$

and the number of CNs is given by $m = E \int \rho$. Therefore, the fraction of CNs of type $t \in I_c$ is given by

$$\gamma_t = \frac{\rho_t}{s_t \int \rho} \tag{1}$$

and the fraction of VNs of type $t \in I_v$ is given by

$$\delta_t = \frac{\lambda_t}{q_t \int \lambda} \ . \tag{2}$$

Also the length of any D-GLDPC codeword in the ensemble is given by

$$N = \sum_{t \in I_v} \left(\frac{E\lambda_t}{q_t} \right) k_t = \frac{n}{\int \lambda} \sum_{t \in I_v} \frac{\lambda_t k_t}{q_t} \ . \tag{3}$$

Note that this is a linear function of n. Similarly, the total number of parity-check equations for any D-GLDPC code in the ensemble is given by

$$M = \frac{m}{\int \rho} \sum_{t \in I_c} \frac{\rho_t (s_t - h_t)}{s_t} \ .$$

A member of the ensemble then corresponds to a permutation on the E edges connecting CNs to VNs. The *design rate* of the D-GLDPC ensemble is given by

$$R = 1 - \frac{M}{N} = 1 - \frac{\sum_{t \in I_c} \rho_t (1 - R_t)}{\sum_{t \in I_v} \lambda_t R_t} \tag{4}$$

where for $t \in I_c$ (resp. $t \in I_v$), R_t is the local code rate of CNs (resp. VNs) of type t.

The growth rate of the weight distribution of the irregular D-GLDPC ensemble sequence $\{\mathcal{M}_n\}$ is defined by

$$G(\alpha) = \lim_{n \to \infty} \frac{1}{n} \log \mathbb{E}_{\mathcal{M}_n} \left[N_{\alpha n} \right] \tag{5}$$

where $\mathbb{E}_{\mathcal{M}_n}$ denotes the expectation operator over the ensemble $\mathcal{M}_n$, and N_w denotes the number of codewords of weight w of a randomly chosen D-GLDPC code in the ensemble $\mathcal{M}_n$. The limit in (5) assumes the inclusion of only those positive integers n for which $\alpha n \in \mathbb{Z}$ and $\mathbb{E}_{\mathcal{M}_n}[N_{\alpha n}]$ is positive (i.e., where the expression whose limit we seek is well defined). Note that the argument of the growth rate function $G(\alpha)$ is equal to the ratio of D-GLDPC codeword length to the number of VNs; by (3), this captures the behaviour of codewords linear in the block length, as in [14] for the LDPC case.

For any irregular D-GLDPC ensemble sequence with growth rate of the weight distribution $G(\alpha)$, we define the *critical ratio* $\alpha^* = \inf\{\alpha > 0 \mid G(\alpha) \geq 0\}$. A positive critical ratio α^* is an important first-order property of a D-GLDPC code ensemble.

III. Further Definitions and Notation

The weight enumerating polynomial for CN type $t \in I_c$ is given by

$$A^{(t)}(x) = \sum_{u=0}^{s_t} A_u^{(t)} x^u = 1 + \sum_{u=r_t}^{s_t} A_u^{(t)} x^u \ .$$

Here $A_u^{(t)} \geq 0$ denotes the number of weight-u codewords for CNs of type t. Note that $A_{r_t}^{(t)} > 0$ for all $t \in I_c$.

The bivariate weight enumerating polynomial for VN type $t \in I_v$ is given by

$$B^{(t)}(x, y) = \sum_{u=0}^{k_t} \sum_{v=0}^{q_t} B_{u,v}^{(t)} x^u y^v = 1 + \sum_{u=1}^{k_t} \sum_{v=p_t}^{q_t} B_{u,v}^{(t)} x^u y^v \ .$$

Here $B_{u,v}^{(t)} \geq 0$ denotes the number of weight-v codewords generated by input words of weight u, for VNs of type t. Also, for each $t \in I_v$, corresponding to the polynomial $B^{(t)}(x,y)$ we denote the sets

$$S_t = \{(i,j) \in \mathbb{Z}^2 \ : \ B_{i,j}^{(t)} > 0\} \tag{6}$$

and

$$S_t^- = S_t \backslash \{(0,0)\} \ . \tag{7}$$

We denote the smallest minimum distance over all CN types by

$$r = \min\{r_t \ : \ t \in I_c\} \geq 2$$

and the set of CN types with this minimum distance by

$$X_c = \{t \in I_c \ : \ r_t = r\} \ .$$

We define the parameter

$$\psi = r/(r-1) \tag{8}$$

and note that we have $1 < \psi \leq 2$ with equality if and only if $r = 2$. We also define the (positive) parameter

$$C = r \sum_{t \in X_c} \frac{\rho_t A_r^{(t)}}{s_t} \ . \tag{9}$$

Similarly, we denote the smallest minimum distance over all VN types by

$$p = \min\{p_t \ : \ t \in I_v\} \geq 2$$

and the set of VN types with this minimum distance by

$$X_v = \{t \in I_v \ : \ p_t = p\} \ .$$

We define the parameter

$$T = \min_{t \in I_v} \min_{(i,j) \in S_t^-} \left\{ \frac{j - \psi}{i} \right\} \tag{10}$$

and the set

$$Y_v = \left\{ t \in I_v \ : \ \min_{(i,j) \in S_t^-} \left\{ \frac{j-\psi}{i} \right\} = T \right\} \ .$$

Since $1 < \psi \leq 2$ with equality if and only if $r = 2$, and $j \geq p \geq 2$ for all $t \in I_v$, $(i,j) \in S_t^-$, it follows that $T \geq 0$ with equality if and only if $r = p = 2$. Also, for $t \in Y_v$, define

$$P_t = \left\{ (i,j) \in S_t^- \ : \ \frac{j-\psi}{i} = T \right\} \ . \tag{11}$$

Note that in the specific case $r = p = 2$, we have $T = 0$ and $Y_v = X_v$, and we may write $P_t = \{(i,2) \ : \ i \in L_t\}$ where $L_t = \{i \in \mathbb{Z} \ : \ B_{i,2}^{(t)} > 0\}$ for each $t \in X_v$ – note that these sets are nonempty.

We define the polynomials

$$Q_1(x) = \sum_{t \in Y_v} \frac{\lambda_t}{q_t} \sum_{(i,j) \in P_t} j B_{i,j}^{(t)} C^{j/r} \left(\frac{\int \lambda}{e} \right)^{iT/\psi} x^i \tag{12}$$

and

$$Q_2(x) = \sum_{t \in Y_v} \frac{\lambda_t}{q_t} \sum_{(i,j) \in P_t} i B_{i,j}^{(t)} C^{j/r} \left(\frac{\int \lambda}{e} \right)^{iT/\psi} x^i \ . \tag{13}$$

Since all of the coefficients of $Q_1(x)$ and $Q_2(x)$ are positive, these polynomials are both monotonically increasing on $[0,\infty)$ and therefore their inverses, denoted by $Q_1^{-1}(x)$ and $Q_2^{-1}(x)$ respectively, are well-defined and unique on this interval. Note that in the case $r = p = 2$, we have

$$Q_1(x) = C \cdot P(x)$$

where

$$P(x) = 2 \sum_{t \in X_v} \frac{\lambda_t}{q_t} \sum_{i \in L_t} B_{i,2}^{(t)} x^i \ . \tag{14}$$

Also note that in the case $r = p = 2$, (9) becomes

$$C = 2 \sum_{t \in X_c} \frac{\rho_t A_2^{(t)}}{s_t} \tag{15}$$

and we define the (positive) parameter

$$V = 2 \sum_{t \in X_v} \frac{\lambda_t B_2^{(t)}}{q_t} \tag{16}$$

as the counterpart of the parameter C in the variable node domain. Here $B_2^{(t)} = \sum_{i \in L_t} B_{i,2}^{(t)}$ is the total number of weight-2 codewords for VNs of type t. Note that in this case the parameter C depends only on the CNs with minimum distance 2, and the parameter V and the polynomial $P(x)$ depend only on the VNs with minimum distance 2. Also note that while the polynomial $P(x)$ given by (14) depends on the VN *representations* (i.e. generator matrices), the parameter V given by (16) does not.

Finally, throughout this paper, the notation $e = \exp(1)$ denotes Napier's number.

IV. Growth Rate for Doubly-Generalized LDPC Code Ensemble

The following theorem, proved in [26], provides an exact characterization of the growth rate of the weight distribution for a general range of α.

Theorem 4.1: The growth rate of the weight distribution of the irregular D-GLDPC ensemble sequence $\{\mathcal{M}_n\}$ is given by

$$\begin{aligned} G(\alpha) = &\sum_{t \in I_v} \delta_t \log B^{(t)}(x_0, y_0) - \alpha \log x_0 \\ &+ \left(\frac{\int \rho}{\int \lambda} \right) \sum_{s \in I_c} \gamma_s \log A^{(s)}(z_0) + \frac{\log\left(1 - \beta \int \lambda\right)}{\int \lambda} \end{aligned} \tag{17}$$

where x_0, y_0, z_0 and β are the unique positive real solutions to the 4×4 system of polynomial equations[2]

$$z_0 \left(\frac{\int \rho}{\int \lambda} \right) \sum_{t \in I_c} \gamma_t \frac{\frac{\mathrm{d}A^{(t)}}{\mathrm{d}z}(z_0)}{A^{(t)}(z_0)} = \beta \ , \tag{18}$$

$$x_0 \sum_{t \in I_v} \delta_t \frac{\frac{\partial B^{(t)}}{\partial x}(x_0, y_0)}{B^{(t)}(x_0, y_0)} = \alpha \ , \tag{19}$$

[2]Note that while (18), (19) and (20) are not polynomial as set down here, each may be made polynomial by multiplying across by an appropriate factor.

$$y_0 \sum_{t \in I_v} \delta_t \frac{\frac{\partial B^{(t)}}{\partial y}(x_0, y_0)}{B^{(t)}(x_0, y_0)} = \beta \; , \tag{20}$$

and

$$\left(\beta \int \lambda\right)(1 + y_0 z_0) = y_0 z_0 \; . \tag{21}$$

It is important to note that this always yields a system of 4 equations in 4 unknowns, *regardless* of the number of different CN and VN types.

The following theorem characterizes the asymptotic behaviour of the growth rate function of Theorem 4.1 as $\alpha \to 0$.

Theorem 4.2: Consider an irregular D-GLDPC code ensemble sequence $\mathcal{M}_n$. For sufficiently small α, the growth rate of the weight distribution is given by

$$G(\alpha) = \frac{T}{\psi}\alpha \log \alpha + \alpha \left[\log\left(\frac{1}{Q_1^{-1}(1)}\right) + \frac{T}{\psi} \log\left(\frac{1}{Q_2(Q_1^{-1}(1))}\right)\right] + O(\alpha^2) \; . \tag{22}$$

A rigorous proof of this theorem is given in [27]; however we may justify this result in the context of Theorem 4.1 through the reasoning given in the next section.

V. Solution for small linear-weight codewords

In this section we analyze Theorem 4.1 in the case $\alpha \approx 0$. We prove a weaker form of Theorem 4.2, namely

Proposition 5.1: For $\alpha \approx 0$,

$$G(\alpha) \approx \frac{T}{\psi}\alpha \log \alpha + \alpha \left[\log\left(\frac{1}{Q_1^{-1}(1)}\right) + \frac{T}{\psi} \log\left(\frac{1}{Q_2(Q_1^{-1}(1))}\right)\right] \; . \tag{23}$$

Here we prove this in a more direct, although somewhat less rigorous, manner than does the proof given in [27], using the polynomial-system solution of Theorem 4.1. We proceed as follows. First, observe from (19) that as $\alpha \to 0$, we must have $x_0^i y_0^j \to 0$ for all $(i,j) \in S_t^{(-)}$ where $t \in I_v$. From (20), this in turn implies that $\beta \to 0$. Now, consider (18) as $\beta \to 0$. The expression on the left-hand side is a rational polynomial in z_0 whose denominator tends to unity as $\beta \to 0$, and whose numerator is dominated as $\beta \to 0$ by the term corresponding to the lowest power of z_0. Therefore, (18) becomes

$$z_0 \left(\frac{\int \rho}{\int \lambda}\right) \sum_{t \in X_c} \gamma_t \left(r A_r^{(t)} z_0^{r-1}\right) = \beta \; , \tag{24}$$

or (using (1))

$$z_0^r \sum_{t \in X_c} \frac{\rho_t r A_r^{(t)}}{s_t} = \beta \int \lambda$$

which, using (9), may be expressed as

$$z_0 = \left(\frac{\beta \int \lambda}{C}\right)^{1/r} \; . \tag{25}$$

Also note that as $\alpha \to 0$, (21) becomes

$$z_0 y_0 = \beta \int \lambda \; . \tag{26}$$

Combining (26) with (25) and using (8) yields

$$y_0 = C^{1/r} \left(\beta \int \lambda\right)^{1/\psi} \; . \tag{27}$$

Next, consider (20) as $\alpha \to 0$. The expression on the left-hand side is a rational polynomial in x_0 and y_0, whose denominator tends to unity as $\alpha \to 0$. Therefore (20) becomes

$$\sum_{t \in I_v} \delta_t \sum_{(i,j) \in S_t^{(-)}} j B_{i,j}^{(t)} x_0^i y_0^j = \beta \; , \tag{28}$$

which, using (2) and substituting for y_0 from (27), may be written as

$$\sum_{t \in I_v} \frac{\lambda_t}{q_t} \sum_{(i,j) \in S_t^{(-)}} j B_{i,j}^{(t)} C^{j/r} \left(x_0 \left(\beta \int \lambda\right)^{\frac{j-\psi}{i\psi}}\right)^i = 1 \; .$$

Since x_0 is independent of i, j and t, the expression on the left-hand side is dominated as $\alpha \to 0$ by those terms such that $(j-\psi)/i$ is minimized, i.e. by terms corresponding to $t \in Y_v$, $(i,j) \in P_t$. The expression may thus be approximated as

$$\sum_{t \in Y_v} \frac{\lambda_t}{q_t} \sum_{(i,j) \in P_t} j B_{i,j}^{(t)} C^{j/r} \left(\beta \int \lambda\right)^{iT/\psi} x_0^i = 1 \; .$$

This may then be written as

$$Q_1(x_0') = 1 \; , \tag{29}$$

where

$$x_0' = x_0 \cdot (e\beta)^{T/\psi} \; . \tag{30}$$

Next consider (19) as $\alpha \to 0$. As in the previous case, the expression on the left-hand side is a rational polynomial in x_0 and y_0, whose denominator tends to unity as $\alpha \to 0$, and whose numerator is dominated as $\alpha \to 0$ by the terms corresponding to $x_0^i y_0^j$ for $t \in Y_v$, $(i,j) \in P_t$. Therefore (19) becomes

$$x_0 \sum_{t \in Y_v} \delta_t \left(\sum_{(i,j) \in P_t} i B_{i,j}^{(t)} x_0^{i-1} y_0^j \right) = \alpha \; . \tag{31}$$

Similarly to the previous case, using (2) and substituting for y_0 from (27), then using the fact that $j = iT + \psi$ for all $t \in Y_v$, $(i,j) \in P_t$, yields

$$Q_2(x_0') = \frac{\alpha}{\beta} \; , \tag{32}$$

where x_0' is given by (30). Therefore

$$\beta = \frac{\alpha}{Q_2(Q_1^{-1}(1))} \; . \tag{33}$$

Note that β is a linear function of α (asymptotically as $\alpha \to 0$). Note also that this implies (via (30)) that x_0 is proportional to $\alpha^{-T/\psi}$ and (via (27)) that y_0 is proportional to $\alpha^{1/\psi}$, and therefore $x_0^i y_0^j$ is proportional to α for all $t \in Y_v$, $(i,j) \in P_t$ (as $j = iT + \psi$ for all such (i,j)).

Finally, we consider $G(\alpha)$ given by (17) for $\alpha \approx 0$. Using (30), along with the approximation $\log(1+t) \approx t$ for $t \approx 0$, we obtain

$$G(\alpha) \approx \sum_{t \in I_v} \delta_t \sum_{(i,j) \in S_t^-} B_{i,j}^{(t)} x_0^i y_0^j - \alpha \log \left[x_0' (e\beta)^{-T/\psi} \right] + \frac{\int \rho}{\int \lambda} \sum_{t \in I_c} \gamma_t \sum_{u=r_t}^{s_t} A_u^{(t)} z_0^u - \beta \quad (34)$$

The polynomial in z_0 is dominated as $\alpha \to 0$ by the term corresponding to the lowest power of z_0. Also, the bivariate polynomial in x_0 and y_0 is dominated as $\alpha \to 0$ by the terms corresponding to $x_0^i y_0^j$ for $t \in Y_v$, $(i,j) \in P_t$. Therefore we may write

$$G(\alpha) \approx \sum_{t \in Y_v} \delta_t \sum_{(i,j) \in P_t} \left(\frac{j - iT}{\psi} \right) B_{i,j}^{(t)} x_0^i y_0^j - \alpha \log x_0' + \frac{\alpha T}{\psi} (1 + \log \beta) + \frac{\int \rho}{\int \lambda} \sum_{t \in X_c} \gamma_t A_r^{(t)} z_0^r - \beta \quad (35)$$

where we have used the fact that $j = iT + \psi$ for all $t \in Y_v$, $(i,j) \in P_t$. This then simplifies to

$$G(\alpha) \approx \frac{\beta - T\alpha}{\psi} - \alpha \log x_0' + \frac{\alpha T}{\psi} (1 + \log \beta) + \frac{\beta}{r} - \beta \quad (36)$$

where we have used (24), (28) and (31). Combining terms and using (29), (33) and (8), leads to (23).

VI. Discussion and Corollaries

Note that from Theorem 4.2, only the VN input-output weight enumerating function coefficients $B_{i,j}^{(t)}$ such that (i,j) lies in one of the sets P_t ($t \in Y_v$) make a contribution to the growth rate of the weight distribution. It is instructive to consider the following geometric construction of these 'dominant' sets in the (i,j) plane. A line $\mathcal{L}$ through the fixed point $(0,\psi)$ is rotated in an anticlockwise fashion until it comes in contact with one or more of the points $(i,j) \in S_t^{(-)}$, $t \in I_v$. The slope of the line $\mathcal{L}$ at this point is defined as T, the set of $t \in I_v$ which have points $(i,j) \in \mathcal{L}$ are defined as Y_v, and for each such t the set of such points on $\mathcal{L}$ is defined as P_t. This interpretation is illustrated in Figure 2 for an example D-GLDPC code with three VN types $I_v = \{R, G, B\}$. Note that the position of the fixed point depends on the smallest CN minimum distance, and always lies somewhere on the line segment joining $(0,1)$ and $(0,2)$ (including the latter endpoint).

We next provide a series of corollaries to Theorem 4.2; this serves to illustrate the manner in which several related results in the literature follow as special cases of this general result.

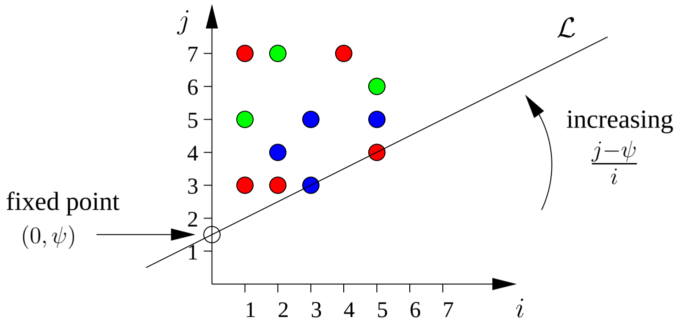

Fig. 2. Diagram of the VN input-output weight enumerating functions and growth rate dominant sets in the (i,j) plane. The illustration is for a smallest CN minimum distance of $r = 3$, so $\psi = 3/2$. The sets $S_t^{(-)}$ for the three VN types $t \in I_v = \{R, G, B\}$ are illustrated by the red, green and blue circles respectively. The line $\mathcal{L}$ has a fixed point at $(0,\psi)$ and is rotated in an anticlockwise fashion until it touches one of the red, green or blue points. In this example, this occurs at two points $(3,3) \in S_B^{(-)}$ and $(5,4) \in S_R^{(-)}$ (simultaneously). Therefore in this example $T = 1/2$, $Y_v = \{B, R\}$, $P_B = \{(3,3)\}$ and $P_R = \{(5,4)\}$.

Corollary 6.1: In the case where either $r > 2$ or $p > 2$, for sufficiently small α the growth rate of the weight distribution is given by

$$G(\alpha) = \frac{T}{\psi} \alpha \log \alpha + O(\alpha) \ , \quad (37)$$

where $T > 0$.

Thus the code has a positive critical ratio α^* in this case. This generalizes results along this line in [17], [18], [19]. A special case of Corollary 6.1, which occurs in many D-GLDPC code ensembles, is as follows.

Corollary 6.2: Suppose $r > 2$ or $p > 2$ and also $\cup_{t \in Y_v} P_t = \{(i,j)\}$ for a single point (i,j), i.e. a single point (i,j) achieves the minimum in (10) although this (i,j) may be manifest in different VN types $t \in Y_v$. Then, for sufficiently small α

$$G(\alpha) = \frac{T}{\psi} \alpha \log \alpha + K\alpha + O(\alpha^2) \ , \quad (38)$$

where K is given by

$$K = \frac{1}{i} \left[\log \left(i \sum_{t \in Y_v} B_{i,j}^{(t)} \delta_t \right) + \frac{j}{r} \log C + \frac{j}{\psi} \log \left(\frac{j \int \lambda}{i} \right) \right] - \frac{T}{\psi} \ . \quad (39)$$

Corollary 6.3: Consider a GLDPC code ensemble with irregular CN set and irregular VN set (i.e. different VN degrees). In this case, p denotes the minimum VN length. Then, for sufficiently small α

$$G(\alpha) = \left(p - \frac{p}{r} - 1 \right) \alpha \log \alpha + K\alpha + O(\alpha^2) \ , \quad (40)$$

where

$$K = \log \left(e \sum_{t \in Y_v} B_{1,p}^{(t)} \delta_t \right) + \frac{p}{r} \log C + \frac{p}{\psi} \log \left(\frac{p \int \lambda}{e} \right) \ . \quad (41)$$

This provides a generalization of the result of [17] which derived (40) for the case of GLDPC codes with *regular* CN sets and irregular VN degrees, and which did not include the result (41) regarding the evaluation of the parameter K.

Corollary 6.4: Consider a D-GLDPC code ensemble $\mathcal{M}_n$ satisfying $r = p = 2$. Then, for sufficiently small α, the growth rate of the weight distribution is given by

$$G(\alpha) = \alpha \log \left[\frac{1}{P^{-1}(1/C)} \right] + O(\alpha^2) \ . \qquad (42)$$

where the polynomial $P(x)$ and the parameter C are given by (14) and (15) respectively.

Note that in this special case the growth rate depends only on the CNs and VNs with minimum distance equal to 2. Note also that (42) is a first-order Taylor series around $\alpha = 0$ which directly generalizes the results of [14] and [18] (for irregular LDPC and GLDPC codes respectively) to the case of irregular D-GLDPC codes. Corollary 6.4 first appeared in [25].

Corollary 6.5: Consider a D-GLPDC code ensemble $\mathcal{M}_n$ satisfying $r = p = 2$. Then, a necessary and sufficient condition for $\mathcal{M}_n$ to have a positive critical ratio α^* is

$$C \cdot V < 1 \qquad (43)$$

where C and V are given by (15) and (16) respectively.

Note that Corollary 6.5 generalizes a result of [14] which states that an irregular LDPC code ensemble has positive critical ratio if and only if $\lambda'(0)\rho'(1) < 1$, where $\lambda(x)$ and $\rho(x)$ respectively denote the edge-perspective VN and CN degree distributions.

Finally, note that although we have presented our results in the context of the weight distribution, all of the results in this paper may be extended in a straightforward manner to cover the growth rate of the stopping set size distribution. Stopping sets are defined in the context of iterative decoding over the binary erasure channel (BEC), and in the case of generalized check and variable nodes the definition of stopping set depends on the decoding algorithm used to locally recover from erasures. The relevant definitions for D-GLDPC codes are given in [27, Appendix II].

ACKNOWLEDGMENT

This work was supported in part by the EC under Seventh FP grant agreement ICT OPTIMIX n. INFSO-ICT-214625 and in part by the University of Bologna (ISA-ESRF fellowship). The authors would like to thank Marco Chiani and Marc Fossorier for helpful discussions.

REFERENCES

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

[2] M. Luby, M. Mitzenmacher, M. Shokrollahi, and D. Spielman, "Improved low-density parity-check codes using irregular graphs," *IEEE Trans. Inf. Theory*, vol. 47, no. 2, pp. 585–598, Feb. 2001.

[3] M. Luby, M. Mitzenmacher, M. Shokrollahi, and D. Spielman, "Efficient erasure correcting codes," *IEEE Trans. Inf. Theory*, vol. 47, no. 2, pp. 569–584, Feb. 2001.

[4] T. Richardson, M. Shokrollahi, and R. Urbanke, "Design of capacity-approaching irregular low-density parity-check codes," *IEEE Trans. Inf. Theory*, vol. 47, no. 2, pp. 619–637, Feb. 2001.

[5] H. Pfister and I. Sason, "Accumulate-repeat-accumulate codes: Capacity-achieving ensembles of systematic codes for the erasure channel with bounded complexity," *IEEE Trans. Inf. Theory*, vol. 53, no. 6, pp. 2088–2115, June 2007.

[6] R. M. Tanner, "A recursive approach to low complexity codes," *IEEE Trans. Inf. Theory*, vol. 27, no. 5, pp. 533–547, Sept. 1981.

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

[8] Y. Wang and M. Fossorier, "Doubly Generalized LDPC codes," in *Proc. of IEEE 2006 Int. Symp. on Information Theory* July 2006, pp. 669–673.

[9] E. Paolini, M. Fossorier and M. Chiani, "Doubly-generalized LDPC codes: Stability bound over the BEC," *IEEE Trans. Inf. Theory,* vol. 55, no. 3, pp. 1027–1046, March 2009.

[10] Y. Wang and M. Fossorier, "Doubly Generalized LDPC Codes over the AWGN Channel," *IEEE Trans. Commun.*, vol. 57, no. 5, pp. 1312–1319, May 2009

[11] E. Paolini, M. Fossorier and M. Chiani, "Generalized and doubly-generalized LDPC codes with random component codes for the binary erasure channel," *IEEE Trans. Inf. Theory,* to appear.

[12] S. Litsyn and V. Shevelev, "On ensembles of low-density parity-check codes: Asymptotic distance distributions," *IEEE Trans. Inf. Theory*, vol. 48, pp. 887–908, Apr. 2002.

[13] D. Burshtein and G. Miller, "Asymptotic enumeration methods for analyzing LDPC codes," *IEEE Trans. Inf. Theory,* vol. 50, no. 6, pp. 1115–1131, June 2004.

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

[15] J. Boutros, O. Pothier, and G. Zemor, "Generalized low density (Tanner) codes," in *Proc. of 1999 IEEE Int. Conf. on Communications*, ICC 1999, vol. 1, June 1999, pp. 441–445.

[16] M. Lentmaier and K. Zigangirov, "On generalized low-density parity-check codes based on Hamming component codes," *IEEE Commun. Lett.*, vol. 3, no. 8, pp. 248–250, Aug. 1999.

[17] 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*, June/July 2004, p. 7.

[18] E. Paolini, M. Chiani and M. Fossorier, "On the growth rate of GLDPC codes weight distribution," in *Proc. of 2008 IEEE Int. Symp. on Spread Spectrum Techniques and Applications*, Aug. 2008, pp. 790–794.

[19] A. Barg and G. Zémor, "Distance properties of expander codes," *IEEE Trans. Inf. Theory,* vol. 52, no. 1, pp. 78–90, Jan. 2006.

[20] M. Sipser and D. A. Spielman, "Expander codes," *IEEE Trans. Inf. Theory,* vol. 42, no. 6, pp. 1710–1722, Nov. 1996.

[21] A. Barg, A. Mazumdar and G. Zémor, "Weight distribution and decoding of codes on hypergraphs," *Advances in Mathematics of Communications*, vol. 2, no. 4, pp. 433–450, Nov. 2008.

[22] 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. Inf. Theory,* vol. 48, no. 6, pp. 1570–1579, June 2002.

[23] A. Orlitsky, K. Viswanathan, and J. Zhang, "Stopping set distribution of LDPC code ensembles," *IEEE Trans. Inf. Theory,* vol. 51, no. 3, pp. 929–953, Mar. 2005.

[24] C-L. Wang and M. P. C. Fossorier, "On Asymptotic Weight Enumerators of LDPC-like codes," *IEEE J. Selected Areas Commun.*, vol. 27, no. 6, pp. 899–907, Aug. 2009.

[25] M. F. Flanagan, E. Paolini, M. Chiani, and M. Fossorier, "On the growth rate of the weight distribution of irregular doubly-generalized LDPC codes," in *Proc. 2008 Allerton Conf. on Communications, Control & Computing*, Sept. 2008, pp. 922–929.

[26] M. F. Flanagan, E. Paolini, M. Chiani, and M. Fossorier, "Growth rate of the weight distribution of irregular doubly-generalized LDPC codes: General case and efficient evaluation," in *Proc. IEEE Global Communications Conference*, Nov./Dec 2009.

[27] M. F. Flanagan, E. Paolini, M. Chiani, and M. Fossorier, "On the growth rate of the weight distribution of irregular doubly-generalized LDPC codes," *IEEE Trans. Inf. Theory*, 2009, submitted.