Weight Distribution of Non-binary LDPC Codes

Kenta KASAI†‡, Charly POULLIAT‡, David DECLERCQ†, Tomoharu SHIBUYA* and Kohichi SAKANIWA†

†Dept. of Communications and Integrated Systems
Tokyo Institute of Technology
152-8550 Tokyo, JAPAN
{kenta, sakaniwa}@comm.ss.titech.ac.jp

‡ETIS ENSEA/UCP/CNRS
UMR-8051
95014 Cergy, FRANCE
{poullait,declercq}@ensea.fr

*Department of Mathematics,
Sophia University
102-8554 Tokyo, JAPAN
tshibuya@sophia.ac.jp

Abstract

Weight distributions of binary low-density parity-check (LDPC) codes are well studied in [1],[5],[4],[2]. We investigate the average distributions of symbol and binary weight for non-binary LDPC code ensemble. We derive the asymptotic growth rate and its linear approximation for small weight. Moreover, we show the typical minimum distance does not monotonically increase with the size of the field where the codes are defined.

1. Introduction

Studies on weight distributions for non-binary LDPC codes date back to the landmark PhD thesis [1]. Gallager derived symbol-weight distribution of Gallager code ensembles [1] defined on \( \mathbb{Z}/q\mathbb{Z} \) and showed the minimum distance grow linearly with code length with variable node degree greater than 2. Hu [3] showed asymptotic bit-weight distributions for random parity-check code ensembles and showed cycle codes can show good decoding performance with non-binary fields.

In this paper we derive the average distributions of symbol and binary weight for non-binary LDPC code ensemble defined on \( GF(q) \). We derive the asymptotic growth rate and the condition for the exponentially few average number of codewords of small linear weight. In [6], density ecolution for non-binary codes for binary erasure channels are compactly derived. Rathi also showed that the threshold is not a monotonic function of the field size. We show that this is also true for the typical minimum distance in this paper.

2. Non-binary LDPC codes

The non-binary LDPC codes also defined on bipartite graphs. The graphs have variable nodes \( v_1,v_2,\ldots,v_n \) and check nodes \( c_1,c_2,\ldots,c_m \). Each edge \( e = (v,c) \in E \) is labeled \( h_e \in GF(q)\setminus\{0\} \). Define that \( (x_1,x_2,\ldots,x_n) \) is a codeword iff

\[
\sum_{(v,c) \in E} h_{(v,c)} x_i = 0 \quad \text{for} \quad j = 1,2,\ldots,m.
\]

Assume we are given edge-perspective degree distributions for variable and check nodes \( (\lambda(x) = \sum_{x=2}^{d_v} \lambda_i x^{i-1} \text{ and } \rho(x) = \sum_{j=2}^{d_c} \rho_j x^{j-1}) \), the size \( q = 2^p \) of Galois field where the codes are defined, and the code length \( n \). In this setting, we define the code ensemble \( \mathcal{G}(n,\lambda(x),\rho(x),q) \) as an equiprobable set of the codes defined by the graphs with \( n \) variable nodes, degree distributions \( (\lambda(x),\rho(x)) \) and uniformly chosen labels from \( GF(q)\setminus\{0\} \).

It is easy to see that \#\( \mathcal{G} = E!(q-1)^E \), where \( E = \int_0^n \frac{n \lambda(x) dx}{h_0} \) is the number of edges and the design rate \( r = 1 - \int_0^n \frac{\lambda(x) dx}{h_0 \rho(x) dx} \) does not depend on \( q \).

3. Symbol Weight Distribution for Non-binary LDPC codes

Let \( N(G,\ell) \) be the number of codewords of symbol-weight \( \ell \) in a code \( G \).

For a given sequence \( \bar{x} \) of weight \( \ell \), we consider the graphs which have \( \bar{x} \) as its codeword. \( \mathcal{G}_{\bar{x}} := \{ G \in \mathcal{G} \mid \bar{x} \in G \} \)

We claim \#\( \mathcal{G}_{\bar{x}} = \#\mathcal{G}_{\bar{x}} \). To this end, we will introduce a mapping \( \mathcal{G}_{\bar{x}} \supseteq \mathcal{F} \mapsto \mathcal{G}' \in \mathcal{G}_{\bar{x}} \). Let \( h_{(v,c)} \) denote the label of the edge \( (v,c) \) in \( \mathcal{G} \). The image graph \( \mathcal{G}' \) is defined to have the same connection of \( \mathcal{G} \) and labels \( h'_{(v,c)} := x_i^{-1} h_{(v,c)} \). It is easy to verify that \( \mathcal{G}' \) accepts \( \bar{L} \). The mapping is an injection since the labels which is mapped to \( \mathcal{G}' \) should have the same connection of \( \mathcal{G} \) and labels \( h'_{(v,c)} := x_i^{-1} h_{(v,c)} \). This concludes the claim. Thus, without loss of generality,
we can assume a codeword \( \mathbf{1}_\ell = (1,1,\ldots,1,0,0,\ldots,0) \) whose first \( \ell \) symbols are 1 and the others are 0.

\[
E_G[N(G,\ell)] = \frac{1}{\#G} \sum_{G \in \mathcal{G}} \#\{E \in \mathcal{G}|W_H(E) = \ell\}
\]

\[
= \frac{1}{\#G} \sum_{E \in \mathcal{G}} \#\{G \in \mathcal{G}|E \subseteq G\}
\]

\[
= \frac{1}{\#G} \binom{n}{\ell} (q-1)^\ell \#\{G \in \mathcal{G}|\mathbf{1}_\ell \in G\}.
\]

We say an edge is active if it connects to a variable node whose first \( \ell \) symbols are 1 and the others are 0.

We consider satisfied check nodes of length \( j \) with \( k \) active edges. Let \( a(k) \) be the number of constellation for such check nodes, i.e., the way of check socket allocation and active edge labelling. For example, it holds that \( a(0) = 1, a(1) = 0 \) and \( a(2) = \binom{j}{2}(q-1) \). Let \( b(k) \) be the number of constellations for satisfied check nodes of degree \( k \) with \( k \) active edges, then it holds that \( a(k) = \binom{j}{k}b(k) \). It is easy to see that

\[
b(k) = (q-1)^{k-1} - b(k-1)
\]

Solving this equation, consequently, we obtain

\[
a(k) = \frac{(-1)^k(q-1) + (q-1)^k}{q}\binom{j}{k}.
\]

Consider a \((d_v,d_e)\)-regular LDPC code ensemble \( \mathcal{G} = \mathcal{G}(n,x^{d_v-1},x^{d_e-1},q) \). The number of ways of allocating \( d_v,m \) sockets for \( m \) check nodes of degree \( d_v \) and \( d_e \) active edge labeling which satisfy all \( m \) check nodes is given by the \( m \)-fold convolution of \( a \)

\[
a^{\otimes m}(d_v,d_e) = \sum_{k_1=1}^{d_v} \cdots \sum_{k_m=1}^{d_v} \prod_{i=1}^{m} a(k_i) = \text{coef} (f_{d_v}(u)^m, u^{d_e})
\]

where we define \( f_j(u) \) as

\[
j \sum_{\ell=0}^{j} a(\ell)u^\ell = \frac{(1+(q-1)u)^j + (q-1)(1-u)^j}{q}
\]

There are \((E-d_v,\ell)!\) and \((d_e,\ell)!\) ways of permuting non-active and active edges, respectively and \((q-1)^{E-d_v}\ell\) ways of non-active edge labeling. Thus we have that

\[
\#\{G \in \mathcal{G}|\mathbf{1}_\ell \in G\} = (E-d_v,\ell)! (d_e,\ell)! (q-1)^{E-d_v}\ell \text{coef} (f_{d_v}(u)^m, u^{d_e})
\]

**Theorem 1** Therefore, we obtain the average symbol-weight distribution of \((d_v,d_e)\)-regular LDPC code ensemble as follows.

\[
E_G[N(G,\ell)] = \binom{n}{\ell} \frac{\text{coef} (f_{d_v}(u)^m, u^{d_e})}{(d_e,\ell)(q-1)^{(d_e-1)\ell}}
\]

In a similar way as in [5, Lemma 3], we can extend this to irregular LDPC code ensembles \( \mathcal{G} = \mathcal{G}(n,\lambda(x),\rho(x)) \) and obtain the following theorem.

**Theorem 2** For irregular non-binary LDPC code defined over GF(q), the average symbol-weight distribution is given as follows.

\[
E_G[N(G,\ell)] = \sum_{\ell = 1}^{\ell_{\text{max}}} d_{\ell} \prod_{i=2}^{d_{\ell}} \binom{L_i n}{\ell_i}
\]

\[
\times \text{coef} \left( \frac{\prod_{j=1}^{d_v} f_j(u) (1-r)^n R_j, u \sum_{i=1}^{d_v} i \ell_i}{(\sum_{i=1}^{d_v} i \ell_i)(q-1)^{\sum_{i=1}^{d_v} i \ell_i - \ell}} \right)
\]

\[
\sum_{k \binom{k}{1} (q-1)^{k-1}} \text{coef} \left( \frac{((Q(s,t) P(u)^{1-r})^n s, t s^k w^k)}{(Q(s,t)^{q-1})^{k-\ell}} \right),
\]

\[
Q(s,t) := \prod_{i=2}^{d_v} \prod_{j=2}^{d_v} f_j(u) R_j,
\]

where \( L_i \) and \( R_j \) are the fraction of variable and check nodes of degree \( i \) and \( j \), respectively, i.e.

\[
L_i = \frac{\ell_i}{\ell} \mathcal{N}(x) dx, \quad R_j = \frac{\ell_j}{\ell} \mathcal{N}(x) dx.
\]

4. **Bit Weight Distribution for Non-binary LDPC codes**

In order to transmit over binary-input channels, we do not send the non-binary symbols in GF(q) but binary images of length \( \log_2(q) \) per symbols, where we assume that \( q = 2^p \). Define the code length in bit perspective as \( n_b = n_p \). The mapping \( g : \text{GF}(q) \rightarrow \{0,1\}^p \) is arbitrary except that we assume \( g(0) = 0 \) but uniquely fixed. There are \( \binom{j}{\ell} \) binary sequences of weight \( \ell \) greater than 0 of length \( p \). Define \( Q_b(s,t) := \prod_{i=2}^{d_v} \prod_{j=2}^{d_v} f_j(u) R_j \), then we have

\[
\text{coef} (Q_b(s,t), t^s k^s)
\]

constellations with \( \ell \) binary-weight and \( k \) active edges.

Let \( N_b(G,\ell_k) \) be the number of codewords of bit-weight \( \ell_k \) in a code \( G \).

**Theorem 3** For irregular non-binary LDPC code defined over GF(q), the average bit-weight distribution is
given as follows. 

\[ E_G[N_b(G,\ell)] = \sum_{\ell_2,\ell_3 \cdots \ell_{d_v}} \prod_{i=2}^{d_v} \frac{(L_i n_b)}{\ell_i x i}, \quad 0 \leq \ell_i = \sum_{p=1}^{\ell_{i-1}} \ell_i = \ell_i n, \quad \sum_{p=1}^{d_v} \sum_{i=2}^{\ell_i} z_i x i = \ell, \quad \text{cof} \left( \prod_{i=2}^{d_v} f_j(u) \left(1 - r\right)^n R_j, u, \sum_{i=2}^{d_v} \ell_i \right) \frac{\log(\omega n)}{(q - 1)^k} (q - 1)^k \]

\[ Q_b(s, t) := \sum_{i=2}^{d_v} \frac{(Q_b(s, t) \beta P_b(u) \left(1 - r\right)^n t^k s^k u^k)}{(\gamma_i) (q - 1)^k} \]

5. Asymptotic Analysis

5.1. Growth Rate on Bit-Weight

The growth rate of the average bit-weight distribution is defined as follows.

\[ \gamma_b(\omega) := \lim_{n \to \infty} \frac{1}{n} \log E_G[N_b(G,\omega n_b)]. \]

The positive (negative) \( \gamma(\omega) \) means that there are exponentially many (few) codewords of normalized bit-weight \( \omega \) in the ensemble \( G \).

**Theorem 4** The growth rate of the weight distribution for irregular non-binary LDPC code ensemble is given as follows.

\[ \gamma_b(\omega) := \lim_{n \to \infty} \frac{1}{n} \log E_G[N_b(G,\omega n_b)] = \lim_{n \to \infty} \frac{1}{n} \log E_G[N_b(G,\omega n_b)] = \frac{1}{p} \log Q_b(s, t) + \frac{1 - r}{p} \log P_b(u) - \beta \log(u) - \beta \log(s) - \omega \log(t) - \epsilon_b h(\beta \epsilon_b n_b) - \beta \log(q - 1), \]

where \((\beta, u, s, t)\) is a positive solution of the following equations,

\[ \frac{(1 - r)u}{p} \frac{\partial P_b}{\partial u} = \beta, \quad (1) \]

\[ \frac{s}{p} \frac{\partial Q_b}{\partial s} = \beta, \quad (2) \]

\[ \frac{u(q - 1)(\epsilon_b + \beta)}{\beta} = s, \quad (3) \]

\[ \frac{t}{p} \frac{\partial Q_b}{\partial t} = \omega \]

and \( \mu \) is the number of real solutions for \( P(x_1, \ldots, x_d) = P(\tilde{x}_1, \ldots, \tilde{x}_d) \).

**Proof:** We use Hayman approximation and saddle point approximation [5]. Define \( C_b(k) := \text{cof} \left( (Q_b(s, t) P_b(u) \left(1 - r\right)^n t^k s^k u^k) \right) \) and \( E/n_b := e_b \) then we have

\[ E_G[N_b(G, \ell)] = \sum_k \frac{C_b(k)}{(\gamma_i n_b) (q - 1)^k} \]

\[ C_b(k) = \frac{\mu}{\sqrt{2\pi n_b}} \frac{p}{P_b(u)} \frac{\partial Q_b}{\partial P_b} = \frac{k}{n_b} \frac{t}{P_b(s, t)} = \frac{\ell}{n_b} \]

where \((u, s, t)\) is a positive solution of

\[ (1 - r)u \frac{\partial P_b}{\partial u} = \frac{k}{n_b} s \frac{\partial Q_b}{\partial s} = \frac{t}{n_b} \frac{\partial Q_b}{\partial t} = \ell \frac{\partial P_b}{\partial P_b} = \mu \]

Assume \( \frac{C_b(k)}{(\gamma_i n_b) (q - 1)^k} \) takes its maximum at \( k = k^* \) and denote the positive solution of (??) by \((\bar{u}, \bar{s}, \bar{t})\).

\[ E_G[N(G, \ell)] = \frac{C_b(k^*)}{(\gamma_i n_b) (q - 1)^k} \sum_k (\gamma_i n_b) (q - 1)^k C_b(k) \]

\[ = \left( e^{k^* - k} \log(e^k - 1) \right) \left( 1 + o(1) \right) \]

where we use (4) in (a) and (b) is obtained by Taylor expansion at \( k = k^* \) and defining

\[ g(k^*) = \log \frac{k^*}{\bar{s} (q - 1)(e_b n_b - k^*) + \frac{e_b n_b - k^*}{k^* (e_b n_b - k^*)}} \]

\[ \frac{1}{\sigma^2} = \frac{1}{k^*} + \frac{e_b n_b - k^*}{k^* (e_b n_b - k^*)} \]

\[ \frac{1}{\bar{s} (q - 1)(e_b n_b - k^*)} + \frac{1}{n_b} \left( 1, 1, 0 \right) B_0 \left( 1, 1, 0 \right)^T \]

According to the assumption that \( \frac{C_b(k)}{(\gamma_i n_b) (q - 1)^k} \) takes its
maximum at $k = k^*$, we know $g(k^*) = 0$. Therefore

$$E_G[N_b(G, t)]$$

$$= \frac{C_0(b^*)}{(q - 1)^{k^*}} \sum_k e^{-\frac{k-k^*}{2\sigma^2}} (1 + o(1))$$

$$= \frac{C_0(b^*)}{(q - 1)^{k^*}} \left[ \int_{-\infty}^{\infty} e^{-\frac{x^2}{2\sigma^2}} dx + o(1) \right]$$

$$= \frac{C_0(b^*)}{(q - 1)^{k^*}} \sqrt{2\pi\sigma^2}$$

$$= \frac{\sigma^2 k^* \left( 1 - \frac{k^*}{e\sigma_0^2} \right) \mu(P_b(u)^{1-r} Q_b(s, t))^{2k^*}}{2\pi \sigma_x^2 \det(B_0)}$$

$$e^{-n_{\text{det}}(k^*)} = k^* \log(q^{-1}) (1 + o(1))$$

The following proposition tells us the growth rate of the all-one code words.

**Proposition 5**

$$\gamma_b(1) = \frac{(1 - r)}{p} \sum_{j=2}^{d_c} R_j \log \left( \frac{(q - 1)^j + (-1)^j (q - 1)}{q (q - 1)^j} \right)$$

$$= - (1 - r) \quad (q \to \infty)$$

**Proof:**

$$E_G[N_b(G, n_b)]$$

$$= \prod_{j=2}^{d_c} \frac{(q-1)^j + (-1)^j (q-1)}{q}^{(1-r)n_R}$$

$$= \prod_{j=2}^{d_c} \left( \frac{(q - 1)^j + (-1)^j (q - 1)}{q (q - 1)^j} \right)^{(1-r)n_R}$$

$$= O(q^{-1-r n_R})$$

$$= O(q^{-1-r n_u/p})$$

where we used the fact that $(1 - r) \sum_{j=2}^{d_c} j R_j n = E$ in (a).

The derivative of $\gamma(\alpha)$ can be simply evaluated as in the following lemma.

**Lemma 6** $\gamma_b'(\omega) = -\log(t(\omega))$

**Proof:** We denote the derivative of $X$ with respect to $w$ by $X'$. Then we have

$$\gamma_b(w) = \frac{1}{p} Q'(s, t) \left( \frac{1 - r}{P} \right) P' + \frac{1}{p} Q(s, t) \left( \frac{1 - r}{P} \right) P'$$

$$- \beta' \log u - \beta u' - \beta' \log s - \beta s'$$

$$- \log t - w t' - \log \frac{e^{1-b} - \beta}{2} - \beta' - \beta' \log(q - 1).$$

By substituting (3), it holds

$$\beta' \left( - \log u - \log s - \log \frac{e^{1-b} - \beta}{2} - \log(q - 1) \right) = 0$$

Using (1), we know that

$$\frac{1 - r}{p} \frac{P'}{P} - \beta u' = 0.$$

From (2), we know that

$$\frac{1}{p} \frac{Q'(s, t)}{Q(s, t)} - \beta s' - w t' = 0.$$

Solving (1), (2) and (3), we can see that $u, t$ and $s$ go to 0 as $w$ approaches to 0. As a consequence, what remained in (5) is $- \log t$.

The number of codewords of small weight is our interest. The following theorem tells us that whether the ensemble has exponentially few or many codewords of small linear weight.

**Theorem 7** For irregular non-binary LDPC code defined over GF(q), the growth rate of the bit-weight distribution for small linear weight is given as follows.

$$\gamma_b(\omega) = - \log \left( \frac{q - 1}{\lambda'(0)\rho(1)} + 1 \right) \frac{1}{\gamma} \omega + O(\omega^2)$$

**Proof:** From Lemma 6, it follows that

$$\gamma_b(\omega) = - \log(\lim_{\omega \to 0} f(\omega)) \omega + O(\omega^2).$$
For sufficiently small \( w \), from (2) we have
\[
\beta = \frac{1}{p} \sum_i i L_i ((1 + t)^p - 1) s^i = \frac{1}{p} \sum_i i L_i ((1 + t)^p - 1) s^i.
\]

While from (1), we have
\[
\beta = \frac{1}{p} (1 - r) \frac{u 2P}{P}.
\]

Therefore it follows that
\[
(1 + t)^p - 1 = \frac{1}{(1 - r) u \frac{2P}{P}} \sum_i i L_i s^i
\]

for \( i = 2, 3, \ldots \). For \( i = 2 \), using (1) we get
\[
2L_2 s^2 = \frac{2L_2 (1 - r) \frac{2P}{P}}{u (q - 1)^2 ((eP - (1 - r)) u \frac{2P}{P})} = \frac{\lambda'(0) p'(1)}{q - 1} (u \to 0).
\]

Similarly for \( i \geq 2 \), we get
\[
2L_2 s^2 = \frac{2L_2 (1 - r) \frac{2P}{P}}{u (q - 1)^2 ((eP - (1 - r)) u \frac{2P}{P})} = 0.
\]

Substituting (7) and (8) to (6) and solving \( t \), we obtain
\[
\lim_{u \to 0} t = \left( \frac{q - 1}{\lambda'(0) p'(1)} + 1 \right) \frac{1}{\beta} - 1.
\]

### 5.2. Growth Rate on Symbol-Weight

The growth rate of the average symbol-weight distributions can be derived in a similar way as in the proof for the growth rate of the average bit-weight distributions.

**Theorem 8** For irregular non-binary LDPC code defined over \( GF(q) \), the growth rate of the symbol-weight distribution is given as follows.

\[
\gamma(\omega) := \lim_{n \to \infty} \frac{1}{n} \log E_Q[N(G, \omega n)] = \log Q(s, t) + (1 - r) \log P(u) - \beta \log(u) - \beta \log(s)
\]

\[ - \omega \log(t) - \frac{\beta}{e} - (\beta - \omega) \log(q - 1) \]

where \((s, t, u, \beta)\) is the positive solution of

\[
(1 - r) u \frac{dp}{P} = \beta, \quad s \frac{dQ}{Q(s, t)} = \beta, \quad \beta \frac{u (q - 1)(e - \beta)}{\omega} = s, \quad t \frac{dQ}{Q(s, t)} = \omega
\]

In a similarly with Lemma 6, we can show the following lemma.

**Lemma 9** \( \gamma'(\omega) = - \log(t(\omega)) \)

**Theorem 10** For irregular non-binary LDPC code defined over \( GF(q) \), the growth rate of the symbol-weight distribution for small linear weight is given as follows.

\[
\gamma(\omega) = - \omega \log(t(\omega)) + o(\omega)
\]

\[
\lim_{\omega \to 0} \gamma'(\omega) = \lim_{\omega \to 0} - \omega \log(t(\omega)) = \lambda'(0) p'(1)
\]

The proof is similar to the proof of Theorem 7, therefore we omit the proof.

The symbol weight distribution for small weight does not depend on \( q \). On the other hand, the bit weight distribution for small weight depends on \( q \), however the condition that the growth rate goes down at 0 is given by \( \lambda'(0) p'(1) < 1 \) which does not depend on \( q \).

### 6. Numerical Examples

We define the typical minimum distance \( \alpha^* \) as follows.

\[
\alpha^* = \inf \{ \alpha > 0 \mid \gamma(\alpha) > 0 \}
\]

In Fig. 1, we show the growth rate for irregular LDPC code ensembles with degree distributions \( \lambda(x) = \frac{1}{2} x + \frac{3}{2} x^2 \), \( \rho(x) = x^4 \), defined on \( GF(q) \), \( q = 2, 3, \ldots, 9 \). Surprisingly, the typical minimum distance does not monotonically grow with \( q \). It first grows with \( q \), and then it decreases. In other words, there is the size which maximize the the typical minimum distance. The maximizing size is \( q = 2^5 \) (cyan) for symbol-weight and \( q = 2^4 \) (magenta) for bit-weight.

As shown in Theorem 7, we can see in Fig. 1(b) the tangents of growth rate of symbol-weight distributions at \( \omega = 0 \) for \( q = 2, 3, \ldots, 9 \) are identical, while in Fig. 1(d) those of bit-weight distributions varies depending on \( q \) but never get positive at \( \omega = 0 \).

### 7. Conclusions

We derived the average distributions of symbol and binary weight for non-binary LDPC code ensemble defined on \( GF(q) \). And we also derived the asymptotic growth rate and the condition for the exponentially few average number of codewords of small linear weight. We show that the typical minimum distance is not a monotonic function of the field size.
Figure 1: Average symbol and bit weight distributions of \( (\lambda(x) = \frac{1}{7} x + \frac{6}{7} x^2, \rho(x) = x^4) \)-irregular LDPC code ensembles, \( r = 0.3 \).

**References**


