Weight Distribution for Non-binary Cluster LDPC Code Ensemble

Takayuki Nozaki
Kanagawa University, JAPAN
Email: nozaki@kanagawa-u.ac.jp

Masaki Maehara, Kenta Kasai, and Kohichi Sakaniwa
Tokyo Institute of Technology, JAPAN
Email: {maehara, kenta, sakaniwa}@comm.ce.titech.ac.jp

Abstract—**In this paper, we derive the average weight distributions for the irregular non-binary cluster low-density parity-check (LDPC) code ensembles. Moreover, we give the exponential growth rate of the average weight distribution in the limit of large code length. We show that there exist $(2, d_{\rm c})$-regular non-binary cluster LDPC code ensembles whose normalized typical minimum distances are strictly positive.**

I. Introduction

Gallager invented low-density parity-check (LDPC) codes [1]. Due to the sparseness of the parity check matrices, LDPC codes are efficiently decoded by the belief propagation (BP) decoder. Optimized LDPC codes exhibit performance very close to the Shannon limit [2]. Davey and MacKay [3] have found that non-binary LDPC codes can outperform binary ones.

The LDPC codes are defined by sparse parity check matrices or sparse Tanner graphs. For the non-binary LDPC codes, the Tanner graphs are represented by bipartite graph with variable nodes, check nodes and labeled edges. The LDPC codes defined by Tanner graphs with the variable nodes of degree $d_{\rm v}$ and the check nodes of degree $d_{\rm c}$ are called $(d_{\rm v}, d_{\rm c})$-regular LDPC codes. It is empirically known that $(2, d_{\rm c})$-regular non-binary LDPC codes exhibit good decoding performance among other LDPC codes for the non-binary LDPC code defined over Galois field of order greater than 32 [4].

Savin and Declercq proposed the non-binary cluster LDPC codes [5]. For the non-binary cluster LDPC code, each edge in the Tanner graphs is labeled by a *cluster* which is a full-rank $p \times r$ binary matrix, where $p \geq r$. In [5], Savin and Declercq showed that there exist $(2, d_{\rm c})$-regular non-binary cluster LDPC ensembles whose minimum distance grows linearly with the code length.

Deriving the weight distribution is important to analyze the decoding performances for the linear codes. In particular, in the case for LDPC codes, weight distribution gives a bound of decoding error probability under maximum likelihood decoding [6] and error floors under belief propagation decoding and maximum likelihood decoding [7] [8].

Studies on weight distribution for non-binary LDPC codes date back to [1]. Gallager derived the symbol-weight distribution of Gallager code ensemble defined over $\mathbb{Z}/q\mathbb{Z}$ [1]. Kasai et al. derived the average symbol and bit weight distributions and the exponential growth rates for the irregular non-binary LDPC code ensembles defined over Galois field $\mathbb{F}_q$, and showed that the normalized typical minimum distance does not monotonically grow with q [9]. Andriyanova et al. derived the bit weight distributions and the exponential growth rates for the regular non-binary LDPC code ensembles defined over Galois field and general linear groups [10].

In this paper, we derive the average symbol and bit weight distributions for the irregular non-binary cluster LDPC code ensembles. Moreover, we give the exponential growth rates of the average weight distributions in the limit of large code length.

The remainder of this paper is organized as follows: Section II defines the irregular non-binary cluster LDPC code ensembles. Section III derives the average weight distributions for the irregular non-binary LDPC code ensembles. Section IV gives the exponential growth rates of the average weight distributions in the limit of large code length and shows some numerical examples for the exponential growth rates. Because of space limitations, the proofs are given in [11].

II. Preliminaries

In this section, we review non-binary cluster LDPC codes [5] and define the irregular non-binary cluster LDPC code ensembles. We introduce some notations used throughout this paper.

A. Non-binary Cluster LDPC Code

The LDPC codes are defined by sparse parity check matrices or sparse Tanner graphs. For the non-binary LDPC codes, the Tanner graphs are represented by bipartite graphs with variable nodes, check nodes and labeled edges.

For the non-binary cluster LDPC codes, each edge in the Tanner graphs is labeled by a *cluster* which is a full-rank $p \times r$ binary matrix, where $p \geq r$. Let $\mathbb{F}_2$ be the finite field of order 2. Note that the non-binary LDPC codes defined by Tanner graphs labeled by general linear group $\mathrm{GL}(p, \mathbb{F}_2)$ are special cases for the non-binary cluster LDPC code with $p = r$.

We denote the cluster in the edge between the i-th variable node and the j-th check node, by $h_{j,i}$. For the cluster LDPC codes, r-bits are assigned to each variable node in the Tanner graphs. We refer to the r-bits assigned to the i-th variable node as *symbol* assigned to the i-th variable node, and denote it by $\boldsymbol{x}_i \in \mathbb{F}_2^r$.

For integers a, b, we denote the set of integers between a and b, as $[a;b]$. More precisely, we define

$$[a;b] := \begin{cases} \{n \in \mathbb{N} \mid a \le n \le b\}, & a \le b, \\ \emptyset = \{\}, & a > b. \end{cases}$$

The non-binary cluster LDPC code defined by a Tanner graph $\mathtt{G}$ is given as follows:

$$C(\mathtt{G}) = \{(\boldsymbol{x}_1, \dots, \boldsymbol{x}_N) \in (\mathbb{F}_2^r)^N \mid \textstyle\sum_{i \in \mathcal{N}_{\mathrm{c}}(j)} h_{j,i} \boldsymbol{x}_i^T = \boldsymbol{0}^T \in \mathbb{F}_2^p \quad \forall j \in [1;M]\},$$

where $\mathcal{N}_{\mathrm{c}}(j)$ represents the set of indexes of the variable nodes adjacent to the j-th check node. Note that N is called symbol code length and the bit code length n is given by rN.

B. Irregular Non-binary Cluster LDPC Code Ensemble

Let $\mathcal{L}$ and $\mathcal{R}$ be the sets of degrees of the variable nodes and the check nodes, respectively. Irregular non-binary cluster LDPC codes are characterized with the number of variable nodes N, the size of the clusters p, r and a pair of *degree distribution*, $\lambda(x) = \sum_{i \in \mathcal{L}} \lambda_i x^{i-1}$ and $\rho(x) = \sum_{i \in \mathcal{R}} \rho_i x^{i-1}$, where λ_i and ρ_i are the fractions of the edges connected to the variable nodes and the check nodes of degree i, respectively.

The total number of the edges in the Tanner graph is

$$E := N / \textstyle\int_0^1 \lambda(x) dx.$$

The number of check node M is given by

$$M = \left(\textstyle\int_0^1 \rho(x) dx / \int_0^1 \lambda(x) dx \right) N =: \kappa N.$$

Let L_i and R_j be the fraction of the variable nodes of degree i and the check nodes of degree j, respectively, i.e.,

$$L_i := \lambda_i / \left(i \textstyle\int_0^1 \lambda(x) dx \right), \quad R_j := \rho_j / \left(j \int_0^1 \rho(x) dx \right).$$

The design rate is given as follows:

$$1 - \kappa p / r.$$

Assume that we are given the number of variable nodes N, the size of the clusters p, r and the degree distribution pair (λ, ρ). An irregular non-binary cluster LDPC code ensemble $\mathcal{G}(N, p, r, \lambda, \rho)$ is defined as the following way. There exist $L_i N$ variable nodes of degree i and $R_j M$ check nodes of degree j. A node of degree i has i sockets for its connected edges. Consider a permutation π on the number of edges. Join the i-th socket on the variable node side to the $\pi(i)$-th socket on the check node side. The bipartite graphs are chosen with equal probability from all the permutations on the number of edges. Each cluster in the edges is chosen from a full-rank $p \times r$ binary matrix with equal probability.

III. Weight Distribution for Non-binary Cluster LDPC Code

In this section, we derive the average symbol and bit weight distributions for the irregular non-binary cluster LDPC code ensemble $\mathcal{G}(N, p, r, \lambda, \rho)$.

We denote the r-bit representation of $\boldsymbol{x}_i \in \mathbb{F}_2^r$, by $(x_{i,1}, \dots, x_{i,r})$. For a given codeword $\boldsymbol{x} = (\boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_N)$, we denote the symbol and bit weight of $\boldsymbol{x}$, by $w(\boldsymbol{x})$ and $w_{\mathrm{b}}(\boldsymbol{x})$. More precisely, we define

$$w(\boldsymbol{x}) := |\{i \in [1;N] \mid \boldsymbol{x}_i \neq \boldsymbol{0}\}|,$$
$$w_{\mathrm{b}}(\boldsymbol{x}) := |\{(i,j) \in [1;N] \times [1;r] \mid x_{i,j} \neq 0\}|.$$

For a given Tanner graph $\mathtt{G}$, let $A^{\mathtt{G}}(\ell)$ (resp. $A_{\mathrm{b}}^{\mathtt{G}}(\ell)$) be the number of codewords of symbol (resp. bit) weight ℓ in $C(\mathtt{G})$, i.e.,

$$A^{\mathtt{G}}(\ell) = |\{\boldsymbol{x} \in C(\mathtt{G}) \mid w(\boldsymbol{x}) = \ell\}|,$$
$$A_{\mathrm{b}}^{\mathtt{G}}(\ell) = |\{\boldsymbol{x} \in C(\mathtt{G}) \mid w_{\mathrm{b}}(\boldsymbol{x}) = \ell\}|.$$

For the irregular non-binary cluster LDPC code ensemble $\mathcal{G}(N, r, p, \lambda, \rho)$, we denote the average number of codewords of symbol and bit weight ℓ, by $A(\ell)$ and $A_{\mathrm{b}}(\ell)$, respectively. Since each Tanner graph in the ensemble $\mathcal{G} = \mathcal{G}(N, r, p, \lambda, \rho)$ is chosen with uniform probability, the following equations hold:

$$A(\ell) = \textstyle\sum_{\mathtt{G} \in \mathcal{G}} A^{\mathtt{G}}(\ell) / |\mathcal{G}|, \quad A_{\mathrm{b}}(\ell) = \sum_{\mathtt{G} \in \mathcal{G}} A_{\mathrm{b}}^{\mathtt{G}}(\ell) / |\mathcal{G}|.$$

First, we will derive the average symbol weight distributions for the irregular non-binary cluster LDPC code ensembles. Recall that each cluster in the edges is chosen from full-rank $p \times r$ binary matrix.

Theorem 1: The average number of codewords of symbol weight ℓ for the irregular non-binary cluster LDPC code ensemble $\mathcal{G}(N, p, r, \lambda, \rho)$ is

$$A(\ell) = \sum_{k=0}^{E} \frac{(2^r - 1)^\ell \mathrm{coef}\big((P(s,t)Q(u))^N, s^\ell t^k u^k\big)}{\binom{E}{k} (2^p - 1)^k} \tag{1}$$
$$=: \textstyle\sum_{k=0}^{E} A(\ell, k),$$
$$P(s,t) := \textstyle\prod_{i \in \mathcal{L}} \big(1 + s t^i\big)^{L_i}, \quad Q(u) := \prod_{j \in \mathcal{R}} f_j(u)^{\kappa R_j},$$
$$f_j(u) := 2^{-p} \big[\{1 + (2^p - 1)u\}^j + (2^p - 1)(1 - u)^j\big], \tag{2}$$

where $\mathrm{coef}(g(s,t,u), s^i t^j u^k)$ is the coefficient of the term $s^i t^j u^k$ of a polynomial $g(s,t,u)$.

Proof: We follow a similar way in [9, Theorem 1]. ■

In a similar way to the average symbol weight distribution, we are able to derive the average bit weight distribution for the irregular non-binary cluster LDPC code ensemble $\mathcal{G}(N, r, p, \lambda, \rho)$.

Theorem 2: Let $n = rN$ be the bit code length. Define $f_j(u)$ as in (2). The average number of codewords of bit weight ℓ for the irregular non-binary cluster LDPC code ensemble $\mathcal{G}(N, p, r, \lambda, \rho)$ is

$$A_{\mathrm{b}}(\ell) = \sum_{k=0}^{E} \frac{\mathrm{coef}\big((P_{\mathrm{b}}(s,t)Q_{\mathrm{b}}(u))^n, s^\ell t^k u^k\big)}{\binom{E}{k} (2^p - 1)^k},$$
$$P_{\mathrm{b}}(s,t) := \textstyle\prod_{i \in \mathcal{L}} [1 + \{(1+s)^r - 1\} t^i]^{L_i / r},$$
$$Q_{\mathrm{b}}(u) := \textstyle\prod_{j \in \mathcal{R}} f_j(u)^{\kappa R_j / r}.$$

IV. Asymptotic Analysis

In this section, we investigate the asymptotic behavior of the average symbol and bit weight distributions for the non-binary cluster LDPC code ensembles in the limit of large code length.

A. Growth Rate

We define

$$\gamma(\omega) := \lim_{N\to\infty} \frac{1}{N} \log_{2^r} A(\omega N) = \lim_{N\to\infty} \frac{1}{rN} \log_2 A(\omega N),$$
$$\gamma_{\mathrm{b}}(\omega_{\mathrm{b}}) := \lim_{n\to\infty} \frac{1}{n} \log_2 A_{\mathrm{b}}(\omega_{\mathrm{b}} n),$$

and refer to them as the *exponential growth rates* or simply *growth rates* of the average number of codewords in terms of symbol and bit weight, respectively. To simplify the notation, we denote $\log_2(\cdot)$ as $\log(\cdot)$.

With the growth rate, we can roughly estimate the average number of codewords of symbol weight ωN (resp. bit weight $\omega_{\mathrm{b}} n$) by

$$A(\omega N) \sim (2^r)^{\gamma(\omega)N}, \quad (\text{resp. } A_{\mathrm{b}}(\omega_{\mathrm{b}} n) \sim 2^{\gamma_{\mathrm{b}}(\omega_{\mathrm{b}})n},)$$

where $a_N \sim b_N$ means that $\lim_{N\to\infty} N^{-1} \log a_N/b_N = 0$.

1) Growth Rate of Symbol Weight Distribution: Since the number of terms in (1) is equal to $E+1$, we get

$$\max_{k\in[0;E]} A(\ell,k) \le A(\ell) \le (E+1) \max_{k\in[0;E]} A(\ell,k).$$

Therefore, we have

$$\lim_{N\to\infty} \frac{1}{rN} \log A(\ell) = \lim_{N\to\infty} \frac{1}{rN} \max_{k\in[0;E]} \log A(\ell,k)$$

From Theorem 1 and [12, Theorem 2], we obtain the following theorem.

Theorem 3: Define $\omega = \ell/N$, $\beta := k/N$ and $\epsilon := E/N$. The growth rate $\gamma(\omega)$ of the average number of codewords of normalized symbol weight ω for the irregular non-binary cluster LDPC code ensemble $\mathcal{G}(N,p,r,\lambda,\rho)$ with sufficiently large N is given by, for $0<\omega<1$,

$$\begin{aligned}\gamma(\omega) &= \sup_{\beta>0} \inf_{\substack{s>0,t>0,\\u>0}} r^{-1}\big[\log P(s,t) + \log Q(u) - \epsilon h(\beta/\epsilon) \\ &\qquad - \beta\log[tu(2^p-1)] - \omega\log[s/(2^r-1)]\big] \\ &=: \sup_{\beta>0} \inf_{s>0,t>0u>0} \gamma(\omega,\beta,s,t,u) =: \sup_{\beta>0} \gamma(\omega,\beta), \end{aligned} \tag{3}$$

where $h(x) := -x\log x - (1-x)\log(1-x)$ for $0<x<1$. A point (s,t,u) which achieves the minimum of the function $\gamma(\omega,\beta,s,t,u)$ is given in a solution of the following equations:

$$\omega = \frac{s}{P}\frac{\partial P}{\partial s} = \sum_{i\in\mathcal{L}} L_i \frac{st^i}{1+st^i}, \tag{4}$$

$$\beta = \frac{t}{P}\frac{\partial P}{\partial t} = \sum_{i\in\mathcal{L}} L_i \frac{ist^i}{1+st^i}, \tag{5}$$

$$\beta = \frac{u}{Q}\frac{\partial Q}{\partial u} = \sum_{j\in\mathcal{R}} \kappa R_j \frac{u}{f_j(u)}\frac{\partial f_j}{\partial u}(u), \tag{6}$$

where

$$\frac{\partial f_j}{\partial u}(u) = j\frac{2^p-1}{2^p}[\{1+(2^p-1)u\}^{j-1} - (1-u)^{j-1}].$$

The point β which gives the maximum of $\gamma(\omega,\beta)$ needs to satisfy the stationary condition

$$\beta = (2^p-1)tu(\epsilon-\beta). \tag{7}$$

The growth rates of average number of codewords with $\omega=0$ and $\omega=1$ are given as follows:

Corollary 1: For the irregular non-binary cluster LDPC code ensemble $\mathcal{G}(N,p,r,\lambda,\rho)$ in the limit of large symbol code length N, the following equations hold:

$$\begin{aligned}&\gamma(0) = 0,\\ &\gamma(1) = r^{-1}\big[\log(2^r-1) - \epsilon\log(2^p-1) - \kappa p \\ &\qquad + \textstyle\sum_{j\in\mathcal{R}} \kappa R_j \log\{(2^p-1)^j + (-1)^j(2^p-1)\}\big].\end{aligned}$$

Moreover, by letting p,r tend to infinity with a fixed ratio, we have

$$\gamma(1) \to 1 - \kappa p/r,$$

namely, $\gamma(1)$ tends to the design rate.

For a fixed normalized symbol weight ω, the intermediate variables s,t,u and β are derived from (4), (5), (6) and (7). Hence, the intermediate variables s,t,u and β are represented as functions of ω. Thus, we denote those intermediate variables, by $s(\omega), t(\omega), u(\omega), \beta(\omega)$.

The derivation of $\gamma(\omega)$ in terms of ω is simply expressed as the following lemma.

Lemma 1: For $s>0$ such that (4), (5), (6) and (7) hold, we have

$$\frac{d\gamma}{d\omega}(\omega) = -\frac{1}{r}\log\frac{s(\omega)}{2^r-1}.$$

Proof: We follow a similar way in [13, Lemma 2]. ∎

2) Growth Rate of Bit Weight Distribution: In a similar way to symbol weight, we can derive the growth rate for the average number of codewords of bit weight.

Theorem 4: Define $\omega_{\mathrm{b}} = \ell/n$, $\beta_{\mathrm{b}} := k/n$ and $\epsilon_{\mathrm{b}} := E/n$. The growth rate $\gamma_{\mathrm{b}}(\omega_{\mathrm{b}})$ of the average number of codewords of normalized bit weight ω_{b} for the irregular non-binary cluster LDPC code ensemble $\mathcal{G}(N,p,r,\lambda,\rho)$ with sufficiently large n is given by, for $0<\omega_{\mathrm{b}}<1$,

$$\begin{aligned}\gamma_{\mathrm{b}}(\omega_{\mathrm{b}}) &= \sup_{\beta_{\mathrm{b}}>0} \inf_{s>0,t>0,u>0} \big[\log P_{\mathrm{b}}(s,t) + \log Q_{\mathrm{b}}(u) \\ &\qquad - \epsilon_{\mathrm{b}} h(\beta_{\mathrm{b}}/\epsilon_{\mathrm{b}}) - \beta_{\mathrm{b}}\log(tu(2^p-1)) - \omega_{\mathrm{b}}\log s\big] \\ &=: \sup_{\beta_{\mathrm{b}}>0} \inf_{\substack{s>0,t>0,\\u>0}} \gamma_{\mathrm{b}}(\omega_{\mathrm{b}},\beta_{\mathrm{b}},s,t,u) =: \sup_{\beta_{\mathrm{b}}>0} \gamma_{\mathrm{b}}(\omega_{\mathrm{b}},\beta_{\mathrm{b}}).\end{aligned}$$

A point (s,t,u) which achieves the minimum of the function $\gamma_{\mathrm{b}}(\omega_{\mathrm{b}},\beta_{\mathrm{b}},s,t,u)$ is given in a solution of the following equations:

$$\omega_{\mathrm{b}} = \frac{s}{P_{\mathrm{b}}}\frac{\partial P_{\mathrm{b}}}{\partial s} = \sum_{i\in\mathcal{L}} L_i \frac{(1+s)^{r-1}st^i}{1+\{(1+s)^r-1\}t^i}, \tag{8}$$

$$\beta_{\mathrm{b}} = \frac{t}{P_{\mathrm{b}}}\frac{\partial P_{\mathrm{b}}}{\partial t} = \sum_{i\in\mathcal{L}} \frac{L_i}{r}\frac{i\{(1+s)^r-1\}t^i}{1+\{(1+s)^r-1\}t^i}, \tag{9}$$

$$\beta_{\mathrm{b}} = \frac{u}{Q_{\mathrm{b}}}\frac{\partial Q_{\mathrm{b}}}{\partial u} = \sum_{j\in\mathcal{R}} \frac{\kappa R_j}{r}\frac{u}{f_j(u)}\frac{\partial f_j(u)}{\partial u}. \tag{10}$$

The value $\beta_{\rm b}$ which gives the maximum of $\gamma_{\rm b}(\omega_{\rm b},\beta_{\rm b})$ needs to satisfy the stationary condition

$$\beta_{\rm b} = (2^p-1)tu(\epsilon_{\rm b}-\beta_{\rm b}).$$

Corollary 2: For the irregular non-binary cluster LDPC code ensemble $\mathcal{G}(N,p,r,\lambda,\rho)$ in the limit of large bit code length, the following equations hold:

$$\begin{aligned}&\gamma_{\rm b}(0)=0,\\ &\gamma_{\rm b}(1)=-\epsilon_{\rm b}\log(2^p-1)-\kappa p/r\\ &\qquad+\textstyle\sum_{j\in\mathcal{R}}(\kappa R_j/r)\log\{(2^p-1)^j+(-1)^j(2^p-1)\}.\end{aligned}$$

Moreover, by letting p,r tend to infinity with fixed ratio, we have

$$\gamma_{\rm b}(1)\to -\kappa p/r.$$

Lemma 2: For $s>0$ such that (8), (9) and (10) hold, we have

$$\frac{d\gamma_{\rm b}}{d\omega_{\rm b}}(\omega_{\rm b}) = -\log s(\omega_{\rm b}).$$

B. Analysis of Small Weight Codeword

In this section, we investigate the growth rate of the average number of codewords of symbol and bit weight with small ω.

Theorem 5: For the irregular non-binary cluster LDPC code ensemble $\mathcal{G}(N,p,r,\lambda,\rho)$ with $\lambda_2>0$, the growth rate $\gamma(\omega)$ of the average number of codewords in terms of symbol weight, in the limit of large symbol code length for small ω, is given by

$$\gamma(\omega) = -\frac{\omega}{r}\log\left[\frac{2^p-1}{(2^r-1)\lambda'(0)\rho'(1)}\right]+o(\omega), \tag{11}$$

where we denote $f(x)=o(g(x))$ if and only if $\lim_{x\searrow 0}\left|\frac{f(x)}{g(x)}\right|=0$ and where $\lambda'(0)\rho'(1)=\lambda_2\sum_{j\in\mathcal{R}}(j-1)\rho_j$.

Similarly, the growth rate of the average number of codewords of bit weight with small weight $\omega_{\rm b}$ is given in the following theorem.

Theorem 6: For the irregular non-binary cluster LDPC code ensemble $\mathcal{G}(N,p,r,\lambda,\rho)$ with $\lambda_2>0$, the growth rate $\gamma_{\rm b}(\omega_{\rm b})$ of the average number of codewords in terms of bit weight, in the limit of large bit code length for small $\omega_{\rm b}$, is given by

$$\gamma_{\rm b}(\omega_{\rm b}) = -\omega_{\rm b}\log\left[\left(\frac{2^p-1}{\lambda'(0)\rho'(1)}+1\right)^{1/r}-1\right]+o(\omega_{\rm b}).$$

We define

$$\begin{aligned}\delta^* &:= \inf\{\omega>0 \mid \gamma(\omega)\ge 0\},\\ \delta^*_{\rm b} &:= \inf\{\omega_{\rm b}>0 \mid \gamma_{\rm b}(\omega_{\rm b})\ge 0\},\end{aligned}$$

and refer to them as the *normalized typical minimum distance* in terms of symbol and bit weight, respectively. Recall that the average number of codeword of symbol weight ωN (resp. bit weight $\omega_{\rm b}n$) is approximated by $A(\omega N)\sim 2^{r\gamma(\omega)N}$ (resp. $A_{\rm b}(\omega_{\rm b}n)\sim 2^{\gamma_{\rm b}(\omega_{\rm b})n}$). Since $\gamma(\omega)<0$ (resp. $\gamma_{\rm b}(\omega_{\rm b})<0$) for $\omega\in(0,\delta^*)$ (resp. for $\omega_{\rm b}\in(0,\delta^*_{\rm b})$), there are exponentially few codewords of symbol weight ωN (resp. bit weight $\omega_{\rm b}n$) for $\omega\in(0,\delta^*)$ (resp. for $\omega_{\rm b}\in(0,\delta^*_{\rm b})$).

Theorem 5 and 6 give the following corollary.

Corollary 3: For the irregular non-binary cluster LDPC code ensemble $\mathcal{G}(N,p,r,\lambda,\rho)$ with sufficiently large N, the normalized typical minimum distances δ^* and $\delta^*_{\rm b}$ in terms of symbol and bit weight, respectively, are strictly positive if

$$\lambda'(0)\rho'(1) < \frac{2^p-1}{2^r-1}. \tag{12}$$

Remark 1: For the non-binary LDPC code ensembles defined over finite field $\mathbb{F}_{2^p}$, the normalized typical minimum distances are strictly positive if $\lambda'(0)\rho'(1)<1$ [9]. For the non-binary LDPC code ensembles defined by the parity check matrices over general linear group $\mathrm{GL}(p,\mathbb{F}_2)$, a necessary condition that the normalized typical minimum distances are strictly positive is also $\lambda'(0)\rho'(1)<1$ from Corollary 3 with $p=r$. On the other hand, in the case for the non-binary cluster LDPC code ensembles, a necessary condition that the normalized typical minimum distances are strictly positive depends on not only $\lambda'(0)\rho'(1)$ but also the size of the clusters p,r as in Corollary 3.

Therefore, for any degree distribution pair (λ,ρ) (even if the degree of all the variable nodes are two), we are able to satisfy (12) by large p,r with fixed ratio.

C. Numerical Examples

In this section, we show some numerical examples of the growth rates for the cluster non-binary LDPC code ensembles. As an example, we employ the $(2,8)$-regular non-binary cluster LDPC code ensembles. To keep the design rate at half, we fix the ratio of the cluster size as $p/r=2$.

Figures 1 and 2 give the growth rates to the average symbol weight distributions for the cluster size $(p,r)=(2,1),(4,2),\ldots,(18,9)$. As shown in Corollary 1, $\gamma(1)$ tends to the design rate 0.5. From Figure 2, we see that the slopes of the growth rates at $\omega=0$ are negative and the normalized typical minimum distance δ^* is strictly positive for $(p,r)=(6,3),(8,4),\ldots,(18,9)$. This confirms Corollary 3.

Figures 3 and 4 give the growth rates to the average bit weight distributions for the cluster size $(p,r)=(2,1),(4,2),\ldots,(18,9)$. The black solid curve in Figure 3 shows the growth rate of the binary random code ensemble of rate 0.5. As shown in Corollary 2, $\gamma_{\rm b}(1)$ tends to -0.5. Moreover, we see that the curves in $\omega_{\rm b}>1/2$ converge to the growth rate of the binary random code ensemble. From Figure 4, we see that the slopes of the growth rates at $\omega_{\rm b}=0$ are negative and the normalized typical minimum distance $\delta^*_{\rm b}$ is strictly positive for $(p,r)=(6,3),(8,4),\ldots,(18,9)$. This confirms Corollary 3.

From Figures 2 and 4, we see that the normalized typical minimum distances δ^* and $\delta^*_{\rm b}$ do not monotonically increase with the size of the clusters (p,r). In this case, the normalized typical minimum distances $\delta^*,\delta^*_{\rm b}$ have the local maximum at $(p,r)=(12,6)$.

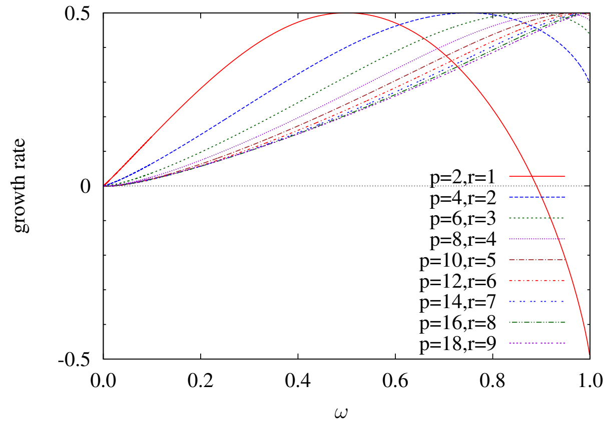

Fig. 1. Growth rates to the average symbol weight distributions for the $(2, 8)$-regular non-binary cluster LDPC code ensembles with the cluster size $(p, r) = (2, 1), (4, 2), \ldots, (18, 9)$.

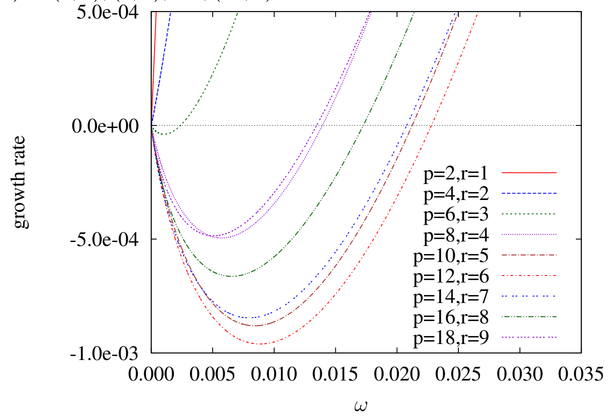

Fig. 2. Growth rates to the average symbol weight distributions for the $(2, 8)$-regular non-binary cluster LDPC code ensembles with the cluster size $(p, r) = (2, 1), (4, 2), \ldots, (18, 9)$.

V. Conclusion

In this paper, we have derived the average weight distributions for the irregular non-binary cluster LDPC code ensembles. Moreover, we have given the exponential growth rate of the average weight distribution in the limit of large code length. We have shown that there exist $(2, d_c)$-regular non-binary cluster LDPC code ensembles whose normalized typical minimum distances are strictly positive.

Acknowledgment

This work was partially supported by Grant-in-Aid for JSPS Fellows.

References

[1] R. G. Gallager, *Low Density Parity Check Codes.* in Research Monograph series, MIT Press, Cambridge, 1963.

[2] T. Richardson, M. A. 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.

[3] M. Davey and D. MacKay, "Low-density parity check codes over GF(q)," *IEEE Commun. Lett.*, vol. 2, no. 6, pp. 165–167, Jun. 1998.

[4] X.-Y. Hu, E. Eleftheriou, and D. Arnold, "Regular and irregular progressive edge-growth tanner graphs," *IEEE Trans. Inf. Theory*, vol. 51, no. 1, pp. 386–398, Jan. 2005.

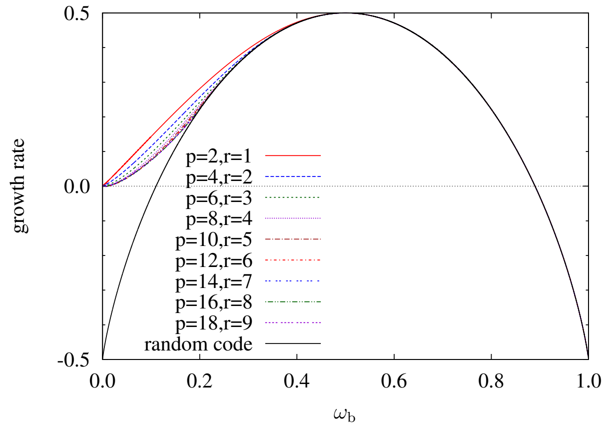

Fig. 3. Growth rates to the average bit weight distributions for the $(2, 8)$-regular non-binary cluster LDPC code ensembles with the cluster size $(p, r) = (2, 1), (4, 2), \ldots, (18, 9)$. The black solid curve (random code) gives the growth rate for the binary random code ensemble of rate 0.5.

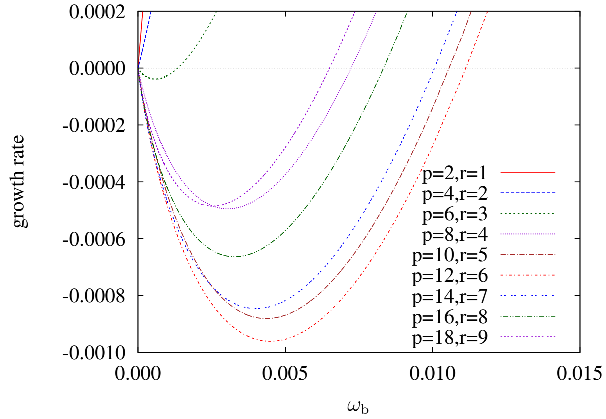

Fig. 4. Growth rates to the average bit weight distributions for the $(2, 8)$-regular non-binary cluster LDPC code ensembles with the cluster size $(p, r) = (2, 1), (4, 2), \ldots, (18, 9)$.

[5] V. Savin and D. Declercq, "Linear growing minimum distance of ultra-sparse non-binary cluster-LDPC codes," in *Proc. 2011 IEEE Int. Symp. Inf. Theory(ISIT)*, Aug. 2011, pp. 523–527.

[6] G. Miller and D. Burshtein, "Bounds on the maximum-likelihood decoding error probability of low-density parity-check codes," *IEEE Trans. Inf. Theory*, vol. 47, no. 7, pp. 2696–2710, nov 2001.

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

[8] T. Richardson and R. Urbanke, *Modern Coding Theory.* Cambridge University Press, Mar. 2008.

[9] K. Kasai, C. Poulliat, D. Declercq, and K. Sakaniwa, "Weight distribution of non-binary LDPC codes," *IEICE Trans. Fundamentals*, vol. E94-A, no. 4, pp. 1106–1115, Apr. 2011.

[10] I. Andriyanova, V. Rathi, and J.-P. Tillich, "Binary weight distribution of non-binary ldpc codes," in *Proc. 2009 IEEE Int. Symp. Inf. Theory(ISIT)*, June-July 2009, pp. 65–69.

[11] T. Nozaki, M. Maehara, K. Kasai, and K. Sakaniwa, "Weight distribution for non-binary cluster LDPC code ensemble," May 2013, arxiv Preprint arXiv:1305.2480.

[12] D. Burshtein and G. Miller, "Asymptotic enumeration methods for analyzing LDPC codes," *IEEE Transactions on Information Theory*, vol. 50, no. 6, pp. 1115–1131, Jun. 2004.

[13] K. Kasai, T. Awano, D. Declercq, C. Poulliat, and K. Sakaniwa, "Weight distribution of multi-edge type LDPC codes," *IEICE Trans. Fundamentals*, vol. E93-A, no. 11, pp. 1942–1948, Nov. 2010.