On the Typical Minimum Distance of Protograph-Based Non-Binary LDPC Codes

Dariush Divsalar  
Jet Propulsion Laboratory  
California Institute of Technology  
Pasadena, CA 91109, USA  
Email: Dariush.Divsalar@jpl.nasa.gov

Lara Dolecek  
Department of Electrical Engineering  
University of California, Los Angeles  
Los Angeles, California 90095, USA  
Email: dolecek@ee.ucla.edu

Abstract—This paper proves the existence of the typical minimum distance for certain ensembles of the protograph-based nonbinary (PB NB) LDPC codes with degree-2 variable nodes. Using recently obtained ensemble weight enumerators for the PB NB codes, we show that asymptotically in the code length $N$, the entries in the weight enumerators up to some weight $d^* = \delta^* N$ ($\delta^* > 0$) vanish under certain conditions on degree-2 nodes. As a consequence, the probability that the code minimum distance is less than $d^*$ goes to zero as the code length goes to infinity. Results of this type advance the understanding of structured non-binary LDPC codes.

I. INTRODUCTION

Low-density parity-check (LDPC) constitute an important class of linear codes. LDPC code defined over both binary and non-binary fields were proposed by Gallager in 1963, [11]. The quality of a linear code is commonly captured by its weight enumerator polynomial (WE) that specifies the number of codewords of each possible weight. A substantial amount of recent work has focused on characterizing the codeword weight enumerators of various classes of binary LDPC codes. Since it is generally difficult to characterize the WE of a given code, the ensemble-wide approach is usually employed. Ensemble weight enumerators of unstructured (ir)regular binary LDPC codes have been reported in [12], [27], [5], [6], [17], [18], and [3], and numerous references therein. Results on bounding the minimum distance, and on the linear minimum distance property of binary LDPC codes were established in [16], [21], [20], and [28].

In addition to the characterization of unstructured binary LDPC codes, recent research efforts have also been geared towards the design and analysis of binary LDPC with prescribed structures. Early work [22] and [23] introduced multi-edge type codes. Protograph-based binary LDPC codes [29] are a subclass of multi-edge LDPC codes. In [10] a method for the computation of the asymptotic (infinite block size) weight enumerators of binary LDPC codes with a protograph structure was proposed. In [7], [1], and [25] ensemble weight enumerators binary LDPC codes with a protograph structure were derived. Results were too extended to the asymptotic setting.

In contrast to the binary domain, much less is known about the ensemble properties of non-binary structured LDPC codes. Recent results in [13], [15] characterized codeword weight distributions of certain regular non-binary codes. Novel code design based on non-binary LDPC protographs was recently discussed in [4] and [19]. In our recent work [8], we studied a class of non-binary LDPC codes based on protographs, and we derived certain WEs of these codes. In the present paper, we upper bound the ensemble weight enumerators of protograph-based non-binary (PB NB) LDPC codes to prove the existence of the typical minimum distance for a class of PB NB LDPC codes with degree-2 variable nodes. The existence of a typical minimum distance implies a linear growth of the minimum distance with the code block length [11].

This paper proceeds as follows: In Section II, we summarize the relevant background on PB NB codes [8]. In Section III we compute an upper bound on weight enumerators. In Section IV, we prove the existence of a typical relative minimum distance for certain class of protograph-based nonbinary LDPC codes. In Section V, we present a class of protograph-based non-binary LDPC codes with a typical minimum distance. Finally, in Section VII we conclude the paper.

II. BACKGROUND AND PREVIOUS RESULTS

We first summarize protograph-based non-binary (PB NB) LDPC codes (simply referred by PB NB codes) and their ensembles introduced in [8].

A. PB NB Codes, Ensembles, and Their Enumerators

A protograph-based non-binary code is constructed from the constituent protograph by a copy-scale-and-permute procedure. When the protograph $G = (V,C,E)$ is copied $N$ times each variable node $v_i \in V$ (each constraint node $c_i \in C$) produces the set $V_i$ of variable nodes $\{v_{i,1}, \ldots, v_{i,N}\}$ (the set $C_i$ of constraint nodes $\{c_{i,1}, \ldots, c_{i,N}\}$) in the resultant graph $G^N$. Likewise, each edge $e_i \in E$ in the protograph produces the set $E_i$ of edges in the resultant graph where $E_i = \{e_{i,1}, \ldots, e_{i,N}\}$, and the edge $e_{i,j}$ for $1 \leq j \leq N$ connects the variable node $v_{k,j}$ and the constraint node $c_{l,j}$ if the edge $e_i$ connects the variable node $v_k$ and the constraint
node $c_i$ in the mother protograph. We denote the resultant matrix $G^N = (V^N, C^N, E^N)$.

**Definition 1 (PB NB code):** Given the mother protograph $G = (V, C, E)$, a $(G, N, \{s_k\}_{k}, \{\pi_i\}_{i})$ PB NB code is constructed from the daughter graph $G^N = (V^N, C^N, E^N)$ by scaling each edge $k$ in $G^N$ by a nonzero element $s_k$ of $GF(q)$ for $1 \leq k \leq N \cdot |E|$, followed by permuting the edges in the set $E_i$ according to $\pi_i$ for each $1 \leq i \leq |E|$.

See Figure 1 for illustration.

**Definition 2 (PB NB code ensemble):** The $(G, N, q)$ PB NB ensemble is the collection of all $(G, N, \{s_k\}_{k}, \{\pi_i\}_{i})$ PB NB codes with all possible choices of $s_k$’s as non-zero elements of $GF(q)$ (for $1 \leq k \leq N \times |E|$) and $\{\pi_i\}$’s as all possible $N$-permutations (for $1 \leq i \leq N$).

For convenience we quickly summarize the necessary notation from [8], [1] along with the main results regarding the weight enumerators: Let $C_j$ be the single parity check code induced by node $c_j$, of degree $m_j$. Let $K_j = q^{(m_j-1)}$ denote the number of codewords in $C_j$. Further, let $M^{C_j}_{v_c}$ be the $K_j \times m_j$ matrix with the codewords of $C_j$ as its rows (whose entries are by construction in $GF(q)$), and let $M^{C_j}_{b,v}$ be the $K_j \times m_j$ binary matrix obtained by converting all nonzero entries of $M^{C_j}_{v_c}$ to 1. Note that by construction, some rows of $M^{C_j}_{b,v}$ may be the same.

Let the set $M^{C_j}_{b,v}$ represent all rows $x$ of $M^{C_j}_{v_c}$, where $x = [x_1, x_2, \ldots, x_{m_j}]$, $x_i \in \{0, 1\}$, and let $K_{j,v} \times m_j$ denote the binary matrix $M^{C_j}_{b,v}$ as the submatrix of $M^{C_j}_{b,v}$ that consists of all distinct rows of $M^{C_j}_{v_c}$. Then, the number of rows in $M^{C_j}_{b,v}$ is $K_{j,v} = 1 + \sum_{i=2}^{m_j} \binom{m_j}{i}$. Lastly, let the set $M^{C_j}_{b,v_i}$ represent the rows $x_k = [x_{k,1}, x_{k,2}, \ldots, x_{m_j}, k, x_{ki}, x_{ki} \in \{0, 1\},$ for $i = 1, 2, \ldots, m_j, k = 1, 2, \ldots, K_{j,v_i}$ of $M^{C_j}_{b,v}$.

To ease the notation, let $c$ be the proxy for the constraint node $c_j$.

**Theorem 1:** The weight-vector enumerator $A^{CN}(\mathbf{w})$ of $C^N$, the code induced by $N$ copies of the constraint code $c$, is given by,

$$A^{CN}(\mathbf{w}) = \sum_{\{n\}} \frac{N!}{n_1! n_2! \cdots n_K!} \mathbf{w}^{n_1} \mathbf{w}^{n_2} \cdots \mathbf{w}^{n_K},$$

where $\{n\}$ is the set of integer-vector solutions to $\mathbf{w} = n \cdot M^{C_j}_{b,v}$, with $n_1, n_2, \ldots, n_{K_j} \geq 0$, and $\sum_{k=1}^{K_j} n_k = N$.

The vector $f_q = [f_{q,1}, f_{q,2}, \ldots, f_{q,K_j}]$ has entries $f_{q,k} = \ln g(q, |x_k|)$, where $x_k$ is the $k$-th element of $M^{C_j}_{b,v}$, $|x_k|$ is the weight of $x_k$, and $g(q, i) = \frac{q-1}{q}[(q-1)^i+(-1)^i]$.

Let $n_k$ be the number of occurrences of the $k$-th codeword among the $N$ copies of $c_j$. Collect these counts into the vector $n$, where $n = [n_1, n_2, \ldots, n_{K_j}]$. Consider the weight-vector $d = [d_1, d_2, \ldots, d_{m_j}], 0 \leq d_i \leq N$, as a pattern of $n_i$ weights corresponding to the $n_i$ length-$N$ variable node inputs satisfying the protograph constraints. Now, let $d_j = [d_{j,1}, d_{j,2}, \ldots, d_{j_m}]$ be the weight vector which describes the weights of the $N$-bit words on the edges connected to the constraint node $c_j$, produced by the variable nodes $\{v_{j1}, v_{j2}, \ldots, v_{jm}\}$ neighboring $c_j$.

**Theorem 2:** The weight-vector enumerator of the PB NB code averaged over the entire ensemble is

$$A(d) = \frac{\prod_{j=1}^{n_c} A^{CN}(d_j)}{\prod_{j=1}^{n_c} (q-1)^{d_j}} (N)^{t_i-1},$$

where $A^{CN}(d_j)$ is the weight-vector enumerator of the code $C^N_j$ induced by the $N$ copies of the constraint node $c_j$. Here, $t_i$ stands for the degree of the variable node $v_i$. The elements of $d_j$ are a subset of the elements of $d = [d_1, d_2, \ldots, d_{m_j}]$ obtained from the edge connections in the mother protograph $G$.

**Proof:** The proofs of both theorems are provided in [8].

The average number of codewords of weight $d$ in the ensemble, denoted by $A_d$, equals the sum of $A(d)$ over all $d$ for which $\sum_{d_i} d_i = d$. This can be written as

$$A_d = \sum_d A(d),$$

under the constraint $\sum_{d_i} d_i = d$.

The binary weight enumerators for the binary image of PB NB codes can also be obtained using (3) and the results by El-Khamy [9].

### III. AN UPPER BOUND ON THE WEIGHT ENUMERATOR $A_d$

In order to prove the existence of typical minimum distance we need to upper bound the $A_d$ in (3). To do so we need the
following lemmas.

**Lemma 1:** An upper bound on a multinomial $\frac{N}{(n_2)!/(n_3)! \ldots (n_{K_{j,r}})!}$ can be expressed as

$$e^{(1+\ln(K_{j,r}-1))y} + y \ln \frac{N}{y}$$

where $y = \sum_{k=2}^{K_{j,r}} n_k$.

**Proof:** First, note that

$$\frac{N!}{(N-y)!} \leq N^y.$$ (5)

Using a lower bound on the Stirling formula for factorials we have

$$\frac{1}{(n_2)!/(n_3)! \ldots (n_{K_{j,r}})!} \leq \prod_{i=1}^{K_{j,r}} n_i^{n_i} e^{-n_i} \leq e^{y} e^{-\sum_{i=2}^{K_{j,r}} n_i \ln n_i}.$$ (6)

The second exponent in (6) can be written as

$$-\sum_{i=2}^{K_{j,r}} \frac{n_i \ln n_i}{y} = -y \sum_{i=2}^{K_{j,r}} \frac{n_i}{y} \ln \frac{n_i}{y} + \ln y$$

$$= -y \sum_{i=2}^{K_{j,r}} n_i \ln \frac{n_i}{y} - y \ln y.$$ (7)

We upper bound the first term on the right-hand side of (7) as

$$-\sum_{i=2}^{K_{j,r}} \frac{n_i \ln n_i}{y} \leq \ln(K_{j,r} - 1).$$ (8)

Thus, the multinomial $\frac{N!}{(n_2)!/(n_3)! \ldots (n_{K_{j,r}})!}$ can be upper bounded as

$$\frac{N!}{(N-y)!} \leq e^{(1+\ln(K_{j,r}-1))y} + y \ln \frac{N}{y}.$$ (9)

**Lemma 2:** If $w = n \cdot M^{C_j}_{b,r}$, then $y = \sum_{i=2}^{K_{j,r}} n_i$ can be bounded as

$$\max\{w_1, w_2, \ldots, w_{m_j}\} \leq y \leq \frac{1}{2} \sum_{k=1}^{m_j} w_k.$$ (10)

**Proof:** Note that for each constraint node $c_j$ of degree $m_j$ we have $w = n \cdot M^{C_j}_{b,r}$. Post multiplying both side of this equation by $1^T$ where $1 = [1,1,1,\ldots,1]$ is the all-ones vector results in $\sum_{k=1}^{m_j} w_k = \sum_{i=2}^{K_{j,r}} n_i x_i$. This equation then implies that $2y \leq \sum_{k=1}^{m_j} w_k \leq m_j y$, where $y = \sum_{k=2}^{K_{j,r}} n_k$ ($K_{j,r} = 1 + \sum_{k=2}^{m_j} (\frac{\sqrt{m_j}}{\sqrt{m_j}}) \leq 2m_j$, where $K_{j,r}$ is the number codewords in $M^{C_j}_{b,r}$). Considering the $k$th component of $w$ in $w = n \cdot M^{C_j}_{b,r}$ we can write $w_k = \sum_{i=2}^{K_{j,r}} n_i x_{i,k}$ where $x_{i,k}$ is the $k$th component of the $i$th row in $M_{b,r}$. Since $x_{i,k}$ is either 0 or 1, we can upper bound $x_{i,k} \leq 1$. This implies that for any $k = 1, 2, \ldots, m_j$, $w_k \leq y$. Thus, for $c_j$ we have

$$\max\{w_1, w_2, \ldots, w_{m_j}\} \leq \frac{1}{2} \sum_{k=1}^{m_j} w_k.$$ (11)

**Lemma 3:** If $n$ satisfies $w = n \cdot M^{C_j}_{b,r}$, then $n \cdot f^T_q$ can be upper bounded as

$$n \cdot f^T_q \leq \left(\prod_{i=1}^{m_j} w_i - y\right) \ln(q-1)$$ (12)

**Proof:** At this point we have to upperbound $n \cdot f^T_q$. Note that $f_{q,k} = \ln g(q, |x_k|)$, where $x_k$ is the $k$-th element of $M^C_{b,r}$, $|x_k|$ is the weight of $x_k$. Then,

$$n \cdot f^T_q = \sum_{k=2}^{K_{j,r}} n_k \ln g(q, |x_k|),$$ (13)

since $|x_1| = 0$, the summation can start at $k = 2$.

$$g(q, |x_k|) = \frac{q - y}{q} \ln((q - 1) |x_k| - 1) \ln(q - 1)$$ (14)

for all $k \geq 2$. To show this relationship, note that $(q - 1)^{i-2} \geq (-1)^{i}$ for all $i \geq 2$. By algebraic manipulations, it follows that

$$n \cdot f^T_q \leq \left(\prod_{i=1}^{m_j} w_i - y\right) \ln(q-1)$$ (16)

follows.

**Lemma 4:** If $n$ satisfies $w = n \cdot M^{C_j}_{b,r}$, then the cardinality (support) of $\{n\}$, denoted by $|\{n\}|$, can be upper bounded as

$$|\{n\}| \leq e^{K_{j,r} + 2 \ln \frac{1}{\sum_{k=1}^\infty w_k}}.$$ (17)

**Proof:** We can rewrite $2y \leq \sum_{k=1}^{m_j} w_k \leq m_j y$ as

$$\frac{1}{m_j} \sum_{k=1}^{m_j} w_k \leq y \leq \frac{1}{2} \sum_{k=1}^{m_j} w_k.$$ Note that for each value of $y = y_0$, $n_1$ takes the value $N - y_0$. We get $(y_0 + K_{j,r} - 1)$ possibilities for $n_2, n_3, \ldots, n_{K_{j,r}}$. Thus

$$|\{n\}| \leq \frac{1}{m_j} \sum_{k=1}^{m_j} w_k.$$ (18)

Note that

$$\frac{|\{n\}|}{y_0} = e^{y_0 - \ln \frac{1}{m_j} \sum_{k=1}^{m_j} w_k}.$$ Using $\ln x \leq x - 1$, $[x] \leq x + 1$, and

$$\sum_{i=1}^{m_j} e^i \leq e^{i+2},$$ we obtain $|\{n\}| \leq e^{K_{j,r} + 2 \ln \sum_{k=1}^\infty w_k}$. ■
Lemma 5: If \( n \) satisfies \( w = n \cdot M_{h,r} \) then the weight-vector enumerator \( A^{CN}(w) \) given by
\[
A^{CN}(w) = \sum_{\{n\}} C(N; n_{11}, n_{22}, \ldots, n_{K_e}) e^n t_i^T
\]
in Theorem 1 can be upper bounded as
\[
A^{CN}(w) \leq \nu_j \prod_{i=1}^{m_j} \frac{1}{2 + m^\ast \ln 2 + 2(1 - \eta) \ln(q - 1) w_i + \frac{1}{2} w_i \ln \frac{N}{e^\tau}} \eta_i, \quad (20)
\]
where \( \nu_j = e^{(K_j + 2)} \), \( \eta = \frac{1}{2} \) if \( q \leq q^\ast \), and \( \eta = \frac{1}{m} \) if \( q > q^\ast \).

Proof: Apply Lemma 2 result to Lemma 1 and note that
\[
e^{\eta y \ln \frac{N}{e^\tau}} \leq \frac{1}{2} \sum_{k=1}^{m_j} w_k \ln \frac{N}{e^\tau} \leq \frac{1}{2} \sum_{k=1}^{m_j} w_k \ln \frac{N}{e^\tau}.
\]
(21)
After using (21), we further upper bound (4) in Lemma 1 as
\[
e^{(1 + \ln(K_j, r - 1) + y \ln \frac{N}{e^\tau})} \leq e^{(1 + m^\ast \ln 2 + y \ln \frac{N}{e^\tau} + \frac{1}{2} \sum_{k=1}^{m_j} w_k \ln \frac{N}{e^\tau})}. \quad (22)
\]
In the upper bound we used \( \ln(K_j, r - 1) < m_j \ln 2 < m^\ast \ln 2 \) where \( m^\ast = \max_j m_j \). Next we apply Lemma 3,
\[
e^{(1 + m^\ast \ln 2 - 2 \ln(q - 1)) y + \frac{1}{2} \sum_{k=1}^{m_j} w_k \ln(N + \frac{N}{e^\tau})} \leq \frac{1}{2} \sum_{k=1}^{m_j} w_k \ln \frac{N}{e^\tau}.
\]
(23)

If \( q \leq q^\ast \) then
\[
(1 + m^\ast \ln 2 - 2 \ln(q - 1)) y \leq (1 + m^\ast \ln 2 - 2 \ln(q - 1)) \frac{1}{2} \sum_{k=1}^{m_j} w_k k,
\]
and if \( q > q^\ast \) then we have to use the following upper bounds
\[
ln(q - 1)) y \leq \frac{1}{m^\ast} \sum_{k=1}^{m_j} w_k \ln(q - 1). \quad (26)
\]
Note that the upper bound on each term in the summation in (19) is independent of \( n \). Now, by Lemma 4, the upper bound in (20) follows.

Lemma 6: The weight-vector enumerator \( A(d) \) given by
\[
A(d) = \frac{\prod_{j=1}^{n_e} A^{CN}(d_j)}{\prod_{i=1}^{n_v} (q - 1)^{d_i(t_i - 1)} \left(\begin{array}{c} N \\ d_i \end{array}\right)^{t_i - 1}}, \quad (27)
\]
in Theorem 2 can be upper bounded as
\[
A(d) \leq \nu \prod_{i=1}^{n_v} e^{\frac{1}{2} (t_i - 2) d_i \ln \frac{d_i}{e^\tau} + \frac{1}{2} (2 + m^\ast \ln 2 - 2(t_i - 1) \ln(q - 1) d_i) \ln N/N}, \quad (28)
\]
where \( \nu = \prod_{j=1}^{n_e} \nu_j \).

Proof: Using the upper bound in Lemma 5, the numerator in (27) can be written as
\[
\prod_{j=1}^{n_e} A^{CN}(d_j) \leq \nu \prod_{i=1}^{n_v} e^{\frac{1}{2} m^\ast \ln 2 + 2(1 - \eta) \ln(q - 1) d_i + \frac{1}{2} d_i \ln N/N}, \quad (29)
\]
and the denominator in (27) can be lower bounded as
\[
\prod_{i=1}^{n_v} (q - 1)^{d_i(t_i - 1)} \left(\begin{array}{c} N \\ d_i \end{array}\right)^{t_i - 1} \geq \prod_{i=1}^{n_v} e^{(t_i - 1) d_i \ln(q - 1) e^{(t_i - 1) d_i \ln N/N}}, \quad (30)
\]
Using the above results the desired upper bound in (28) follows.

IV. A GENERIC UPPER BOUND ON \( A(d) \)

We are going to show that for certain classes of PB NB LDPC codes to be discussed shortly, the upper bound on \( A(d) \) can be represented as
\[
A(d) \leq \nu e^{\alpha d \ln \frac{N}{e^\tau} + (\alpha - 1) d} \quad (31)
\]
where \( d = \sum d_i = d1^T \), and \( \alpha > 0 \). Equipped with this upper bound, the weight enumerator \( A_d \) can be upper bounded as
\[
A_d = \sum_{\{d\}} A(d) \leq \nu |\{d\}| e^{\alpha d \ln \frac{N}{e^\tau} + (\alpha - 1) d}. \quad (32)
\]
We can show that \(|\{d\}| = \frac{d + n_e - 1)!}{d! (n_e - 1)!} \leq \frac{d^{d + n_e - 1)}{d!} \leq d^{d + n_e - 1)} \leq d^{d + n_e - 1)} \), where we used the fundamental inequality \( \ln(x) \leq x - 1 \). Now, the upper bound on \( A_d \) can be written as
\[
A_d \leq c e^{\alpha d \ln \frac{N}{e^\tau} + \beta d} \quad (33)
\]
Lemma 7: If an upper bound on \( A_d \) can be represented as
\[
A_d \leq c e^{\alpha d \ln \frac{N}{e^\tau} + \beta d} \quad (34)
\]
for some constant \( k \) independent of \( N \). Then the upper bound in (33) can represented as
\[
A_d \leq c e^{\alpha N F(\delta)}. \quad (35)
\]
We note that for \( 0 \leq \delta \leq \delta_o \) the function \( F(\delta) \) is convex, it is negative, and has a minimum at \( \delta = \delta_o/e \). The minimum value at this point is \( F(\delta_o/e) = -\delta_o/e \). For \( d = 1, \delta = 1/N \) with \( F(1/N) = -\ln(N\delta_o)/N \).

Now we upper bound \( F(\delta) \) with two straight lines, one that connects the point \((\frac{N}{e^\tau}, F(\frac{N}{e^\tau}))\) to the point \((\frac{N}{e^\tau}, F(\frac{N}{e^\tau}))\). Denote this line by \( L_1(\delta) = \alpha_1 \delta + \beta_1 \) where \( \alpha_1 = \frac{\frac{N}{e^\tau} - \frac{N}{e^\tau}}{\frac{N}{e^\tau} - \frac{N}{e^\tau}} \).
\[ \beta_1 = -\frac{\alpha_1}{N} - \frac{1}{N} \ln N \delta_0. \] Then \( F(\delta) \leq L_1(\delta) \) for \( 1/N \leq \delta \leq \delta_0 / \epsilon. \) The other line connects the point \( \left( \frac{\delta_2}{N} , F(\frac{\delta_2}{N}) \right) \) to the point \( \left( \delta_2 , F(\delta_2) \right). \) Denote this line by \( L_2(\delta) = \alpha_2 \delta + \beta_2 \) where \( \alpha_2 = \frac{\beta_2}{\delta_2} \), and \( \beta_2 = -\frac{\delta_2}{\epsilon}. \) Thus \( F(\delta) \leq L_2(\delta) \) for \( \delta_0 / \epsilon \leq \delta \leq \delta_0. \) Then
\[
\sum_{d=1}^{[N \delta_0/e]} A_d \leq c \frac{N^{-\alpha}}{\delta_0^\alpha (1-\epsilon e^{-1})} .
\] (36)

Now for any \( \delta_0 / \epsilon < \delta^* < \delta_0 \) we have
\[
\sum_{d=1}^{[N \delta_0/e]} A_d \leq c \frac{e^{\frac{\alpha \epsilon}{N}}}{e^{\frac{\alpha \epsilon}{N}} - 1} .
\] (37)

We used \[ |x| \leq x \text{ and } |x| \geq x - 1 \text{ in the above equation.} \] Choose \( \delta^* = \delta_0 - (e-1) \frac{1}{N} \ln (N \delta_0). \) Then
\[
\sum_{d=1}^{[N \delta_0/e]} A_d \leq \frac{e^{\frac{\alpha \epsilon}{N}}}{e^{\frac{\alpha \epsilon}{N}} - 1} \frac{N^{-\alpha}}{\delta_0^\alpha} .
\] (38)

Therefore the total sum is
\[
\sum_{d=1}^{[N \delta_0/e]} A_d \leq \frac{e^{\frac{\alpha \epsilon}{N}}}{e^{\frac{\alpha \epsilon}{N}} - 1} \frac{N^{-\alpha}}{\delta_0^\alpha}.
\]

This upper bound goes to zero as \( N \to \infty \).

V. PB NB LDPC codes with the typical relative minimum distance

In this section, we investigate the existence of a typical relative minimum distance for PB NB LDPC codes. The following theorem states the existence of the typical relative minimum distance as defined by Gallager [11].

**Theorem 3:** If an upper on \( A(d) \) can be represented as
\[
A(d) \leq \nu e^{a(d \ln \frac{d}{\gamma} + (\alpha \beta - 1) d)}
\] (39)

for some \( \alpha > 0 \), then there exists a \( \delta_0 > 0 \) and a small \( \epsilon > 0 \) such that \( Pr \{ d_{\min} \leq [N(\delta_0 - \epsilon)] \} \to 0 \) as \( N \to \infty \). The \( \delta_0 = e^{-\beta} \) serves as a lower bound on the typical relative minimum distance.

**Proof:** We already proved that there exists \( \delta_0 > 0 \) such that \( \sum_{d=1}^{[N \delta_0/e]} A_d \leq kN^{-\alpha} \). Using Markov’s inequality we have \( Pr \{ d_{\min} \leq [N(\delta_0 - \epsilon)] \} \leq \sum_{d=1}^{[N \delta_0/e]} A_d \leq kN^{-\alpha} \to 0 \) as \( N \to \infty \).

These results allow us to determine whether or not the typical minimum distance in the ensemble grows linearly with codeword length.

A. Ensemble of PB NB LDPC codes with degree-2 nodes

We provide conditions on the connections of degree-2 variable nodes to constraint nodes for having a typical relative minimum distance. Consider a class of PB NB LDPC codes where degree-2 subgraph of the protograph does not have any cycles. This restrictions requires the number of degree-2 variable nodes in the subgraph be strictly less than the number of checks in the subgraph. Denote the total number of degree-2 variable nodes by \( T_2 \leq n_v - 1 \). We call this class the degree-2-constrained class of PB NB LDPC codes, \( C_{T_2} \). This condition requires at least two check nodes in the subgraph induced by degree-2 variable nodes to each have only one connection to a degree-2 variable node. The PB NB LDPC codes that do not belong to this class and have cycles will have at most logarithmic minimum distance growth rate. Suppose the PB NB LDPC ensemble has \( T_2 \) variable nodes with degree 2 and \( n_v - T_2 \) nodes with degree 3 or higher. Let us rename the weights of the degree-2 nodes by \( l_k \) and the weights of the other nodes by \( u_i \). Then,
\[
A(d) \leq \nu \prod_{i=1}^{n_t} e^{(L_2 - 1) l_k} u_i + \prod_{k=1}^{T_2} e^{(2 + m^* \ln 2 - 2(\gamma - 1)) l_k} ,
\] (40)

The slope of \( \left( \frac{1}{2} - 1 \right) u_i \ln \frac{u_i}{N} + t_i (2 + m^* \ln 2 - 2(\gamma - 1)) l_k \) with respect to \( t_i \) is strictly negative when \( u_i < N e^{-(2 + m^* \ln 2 - 2(\gamma - 1)) l_k} \). This inequality hold if \( d < N \delta_0 = N e^{-\beta} \) (to be defined shortly). The \( t_m \) is defined as \( t_m = \min_{t_i > 1} t_i \). Then we can upper bound \( A(d) \) as
\[
A(d) \leq \nu \prod_{i=1}^{n_t} e^{L_2 - 1} u_i + \prod_{k=1}^{T_2} e^{(2 + m^* \ln 2 - 2(\gamma - 1)) l_k} ,
\] (41)

Let \( \sum_{k=1}^{T_2} l_k = L \) and \( \sum_{k=1}^{n_t} u_i = u \). Also note that \( \ln \frac{u}{N} \leq \ln \frac{u}{d} \) and \( d = L - \tilde{d} \). Then we obtain
\[
A(d) \leq \nu e^{E(d, L)} ,
\] (42)

where
\[
E(d, L) = \frac{t_m - 2}{2} (d - L) \ln \frac{(d - L)}{N} + \frac{t_m (2 + m^* \ln 2 - 2(\gamma - 1) \ln (q - 1))}{2} \frac{2}{(d - L)} - 2 (2 + m^* \ln 2 - 2(\gamma - 1) \ln (q - 1)) L \] (43)

At this point we want to obtain an upper bound on \( L \) before further upper bounding (42). Using (11) we can show \( L \leq \gamma u \), where \( \gamma = \frac{\alpha}{1 - t^*} \), and \( t^* = \max_{t_i} t_i \).

Note that \( \frac{L}{d} \leq \gamma u = \gamma (d - L) \) implies that \( L \leq d/L \) as we go back to (42). Note that the first derivative of the function \( E(d, L) \) with respect to \( L \) is \( \frac{t_m - 2}{2} \ln \frac{N}{(d-L)} - \frac{t_m}{2} \ln [2 + m^* \ln 2 - 2(\gamma - 1) \ln (q - 1)] \). For \( 0 \leq L \leq \frac{d}{1 + \gamma} \), the second derivative is \( \frac{t_m - 2}{2} \frac{1}{d-L} > 0 \). Thus the function is convex. At this point we need to compare both boundary values at \( L = 0 \) and \( L = \frac{d}{1 + \gamma} \). As was mentioned before we are interested in the region \( d \leq N \delta_0 \). We will show that \( E(d, \frac{d}{1 + \gamma^2}) = E(d, 0) \geq 0 \). Therefore we proved \( \max_{0 \leq L \leq \frac{d}{1 + \gamma^2}} E(d, L) \leq E(d, \frac{d}{1 + \gamma}) \) and \( E(d, L) \leq E(d, \frac{d}{1 + \gamma^2}) \). The weight vector enumerator can be upper bounded as
\[
A(d) \leq \nu e^{E(d, \frac{d}{1 + \gamma})} ,
\] (44)

where \( E(d, \frac{d}{1 + \gamma}) = \alpha d \ln \frac{d}{N} + (\alpha - 1) d \). We obtain
\[
\alpha = \frac{t_m - 2}{2(1 + \gamma)}.
\] (45)
and
\[
\beta = \alpha^{-1} + (1 + \alpha^{-1}) (2 + m^* \ln 2) - \ln (1 + \gamma) - [2\eta + (2\eta - 1)\alpha^{-1}] \ln (q - 1).
\]  

(46)

For a protograph-based NB LDPC code with no degree-2 variable nodes we can set \( \gamma = 0 \).

For \( d \leq N\delta_o = N e^{-\beta} \) we have \( u_t \leq u \leq \frac{1}{1 + \gamma} d \leq N \frac{\delta_o}{1 + \gamma} = N e^{-\beta + \ln (1 + \gamma)}. \) However, \( \beta + \ln (1 + \gamma) > (2 + m^* \ln 2) - 2\eta \ln (q - 1) \). Thus the condition \( u_t < N e^{-(2 + m^* \ln 2) - 2\eta \ln (q - 1)} \) is satisfied as it was required to show.

Now define \( \Delta E = E(d, \frac{1}{1 + \gamma} d) - E(d, 0) \).
\[
\Delta E = \gamma d\alpha\left( -\ln \frac{d}{N} - (2 + m^* \ln 2) + 2\eta \ln (q - 1) - \frac{1}{\gamma} \ln (1 + \gamma) \right).
\]

Since \( d \leq N e^{-\beta} \), then \( -\ln \frac{d}{N} \geq \beta \). Thus \( \Delta E \geq d(\gamma - \frac{1}{\gamma} \ln (1 + \gamma) + \gamma((2 + m^* \ln 2) - (2\eta - 1) \ln (q - 1))) \).

Using the fundamental inequality \( \ln (1 + \gamma) \leq \gamma \), we get \( \Delta E \geq \frac{\gamma d}{2} \{1 + 2(2 + m^* \ln 2) - 2(2\eta - 1) \ln (q - 1)\} \geq 0 \) as it was required to show.

VI. DISCUSSION

Our upper bounding technique shows that there exists \( \delta_o > 0 \) that provides a proof for existence of the typical relative minimum distance for certain class of PB NB LDPC codes. We note that as the degree of check nodes increases, (which also requires the degree of variable nodes other than degree-2 to increase), \( q^* \) also increases. The expression for \( \delta_o \) reveals that for \( q \leq q^* \), \( \delta_o \) also increases with \( q \) up to \( q^* \). However, when \( q > q^* \), \( \delta_o \) decreases as \( q \) increases beyond \( q^* \). Since we used upper bounds on enumerators, \( \delta_o \) only serves as a lower bound on the true typical relative minimum distance \( \delta_{\text{min}} \) for a particular PB NB LDPC code within the class \( \mathcal{C}_T \). The true \( \delta_{\text{min}} \) within this class can be computed numerically from the asymptotic results provided in [8]. Although \( \delta_o \) versus \( q \) shows an increase up to \( q^* \) and then decrease such behavior may not represent the true functional behavior of \( \delta_{\text{min}} \) with respect to \( q \) due to the upper bounding of \( A_d \). However, similar behavior of a typical relative minimum distance for regular and irregular non-binary LDPC codes was observed by Kasai et al. [14].

Such observations for weighted non-binary repeat multiple accumulate codes were also made by Rosnes and Graell i Amat [24]. Fig. 2 shows the typical relative minimum distance for rate 1/2 non-binary regular Gallager LDPC codes versus \( q \) based on a numerical calculation. As it is seen from this figure, there is a maximum for \( \delta_{\text{min}} \) for some \( q \). It is not exactly at \( q^* \) but it is close. For very large \( q \), \( \delta_{\text{min}} \) decreases towards zero. Andriyanova et al. also proved that the asymptotic binary weight of non-binary regular LDPC codes, when the alphabet size goes to infinity, has a large number of codewords of poor weight [2]. If we increase both degree of variable and check nodes such that the rate remains the same, then \( \delta_{\text{min}} \) approaches the Gilbert Varshamov bound as it is seen in Fig. 2.

The next item to discuss is the weight enumeration result for irregular non-binary LDPC codes. We know that if \( \lambda'(0)\rho'(1) < 1 \), then the typical relative minimum distance exists where \( \lambda(x) \) and \( \rho(x) \) are degree distributions for variables and check nodes. In this paper we have already shown that if the number \( T_2 \) of degree-2 variable nodes is less than or equal to the number of check nodes minus 1 (provided that there is no cycles in the subgraph of degree-2 nodes), then the typical relative minimum distance exists. Now consider a rate 1/2 protograph-based non-binary LDPC code with \( T_2 = n_c - 1 \) degree-2 nodes and \( T_2 + 2 \) degree-3 nodes.

So we have \( n_v = 2T_2 + 2 \) and \( n_c = T_2 + 1 \). The degree distribution of this code is \( \lambda(x) = \frac{2T_2}{5T_2 + 6} x + \frac{3T_2 + 6}{5T_2 + 6} x^2 \) and \( \rho(x) = \frac{5T_2}{5T_2 + 6} x^4 + \frac{6}{5T_2 + 6} x^5 \). Then \( \lambda'(0)\rho'(1) = \frac{40T_2}{(20T_2 + 60T_2)^2} \).

For \( T_2 = 2 \) this number is 1.09 > 1 and approaches 1.6 for very large number of \( T_2 \).

The other item to discuss is regarding some recent results on use of binary image of non-binary LDPC with degree-2 nodes. For example Savin and Declercq, have shown linear growing minimum distance for ultra-sparse non-binary cluster- LDPC codes with all degree-2 nodes [26]. We should note that this ensemble is different than regular and irregular non-binary LDPC codes where the scalings on edges are selected randomly. Binary image of these non-binary codes resemble GLDPC codes. For binary GLDPC codes, were each constraint node has the minimum distance strictly greater than 2, with use of the same technique in this paper it might be possible to show that at least for the protograph version, the typical minimum distance exists under certain conditions. For non-binary LDPC codes, if we consider a check node with attached scaling as a block code, then with a careful selection of scales for large enough \( q \), the binary image of this block code should have a minimum distance strictly greater than 2. Thus, such block code can be viewed as a constraint node in the GLDPC set up.
VII. CONCLUSION

In this paper we proved that under certain conditions the typical minimum distance for PB NB LDPC codes exists.

REFERENCES