Ensemble Weight Enumerators for Protograph LDPC Codes

Dariush Divsalar
Jet Propulsion Laboratory
California Institute of Technology
4800 Oak Grove Drive
Pasadena, CA 91109-8099
Email: Dariush.Divsalar@jpl.nasa.gov

Abstract—Recently LDPC codes with projected graph, or protograph structures have been proposed. In this paper, finite length ensemble weight enumerators for LDPC codes with protograph structures are obtained. Asymptotic results are derived as the block size goes to infinity. In particular we are interested in obtaining ensemble average weight enumerators for protograph LDPC codes which have minimum distance that grows linearly with block size. As with irregular ensembles, linear minimum distance property is sensitive to the proportion of degree-2 variable nodes. In this paper the derived results on ensemble weight enumerators show that linear minimum distance condition on degree distribution of unstructured irregular LDPC codes is a sufficient but not a necessary condition for protograph LDPC codes.

I. INTRODUCTION

Low-density parity-check (LDPC) codes were proposed by Gallager [1] in 1962. Ensemble weight enumerators for unstructured irregular LDPC codes and turbo like codes have been reported in [12], [13], [14], [15], [16], [17], [18], [19], [21]. Recently a flurry of work has been conducted on the design of LDPC codes with imposed sub-structures, starting with the introduction of multi-edge type codes in [5] and [6]. Protograph based LDPC codes are a subclass of multi-edge LDPC codes. In [11] a different method for computation of asymptotic weight enumerators for LDPC codes with protograph structure has been proposed. The method in this paper starts with computation of ensemble weight enumerators for finite block size LDPC with a protograph structure. The results then are extended to asymptotic cases as the block size goes to infinity.

For high-speed decoding, it is advantageous for an LDPC code to be constructed from a projected graph [4], or protograph [3]. A protograph is a Tanner graph with a relatively small number of nodes. A protograph \( G = (V, C, E) \) consists of a set of variable nodes \( V \), a set of check nodes \( C \), and a set of edges \( E \). Each edge \( e \in E \) connects a variable node \( v_e \in V \) to a check node \( c_e \in C \). Parallel edges are permitted, so the mapping \( e \rightarrow (v_e, c_e) \in V \times C \) is not necessarily 1:1. As a simple example, we consider the protograph shown in Fig. 1. This graph consists of 3 variable nodes and 2 check nodes, connected by 5 edges. In this example we have 5 edge types i.e. each edge in the base protograph represents an edge type. For multi-edge LDPC codes, a group of edges (number of edges in each group can be different) represents an edge type. For unstructured irregular LDPC codes, there is only one edge type. Having the base protograph, we can obtain a larger graph by a “copy-and-permute” operation as shown in Fig. 1. This operation consists of first making \( N \) copies of the protograph, and then permuting the endpoints of each edge type among the \( N \) variable and \( N \) check nodes connected to the set of \( N \) edges copied from the same edge type in the protograph. The derived or lifted graph is the graph of a code \( N \) times as large as the code corresponding to the protograph, with the same rate and the same distribution of variable and check node degrees.

As follows in the protograph representation in the figures, those variable nodes say \( m \) nodes that are connected to the channel (transmitted nodes) will be shown as dark filled circles. Those variable node that are not connected to the channel (punctured nodes or not transmitted nodes) will be depicted by a blank circle. The check nodes will be depicted by circles with a plus sign inside. The code rate for the protograph is \( R_c = \frac{n_s - n_a}{n_s} \) provided that the parity check matrix of the derived or lifted graph is full rank.

II. ENSEMBLE WEIGHT ENUMERATOR FOR A PROTOGRAPH BASED LDPC

Consider a protograph based LDPC code as shown in Fig. 2 (for base protograph ignore the interleavers) with \( n_v \) variable and \( n_c \) check nodes.

Let \( q_{v_t} \) represent the degree of variable node \( v_t \), \( q_{c_t} \) represent the degree of check node \( c_t \) and let \( |E| \) represent the total number of edges types. Suppose we lift (copy and permute) this protograph by a factor of \( N \). This is equivalent
to using random interleavers each of size $N$ per each edge of the base protograph as shown in Fig. 2. Assign weight $d_i$ to variable node $v_i$ for $i = 1, \ldots, n_v$. Using the results for serial concatenation with interleaved codes [9] and for turbo-like codes [2] we can compute the ensemble weight enumerators for protograph LDPC codes. Consider each variable and check node in the protograph as a constituent code. These constituent codes are connected through edges in the protograph where we assume each edge has a uniform interleaver [8] of size $N$. Each variable node $v_i$ can be assumed as a constituent code with one input of weight $d_i$ and $q_{v_i}$ outputs. The input-output weight coefficient for node $v_i$ is $(N)\delta_{d_i,w_{v_i}} \ldots \delta_{d_i,w_{v_i}}$, where Kronecker Delta is defined as $\delta_{x,y} = 1$ if $x = y$, and zero otherwise. Each check node $c_i$ can be assumed as a constituent code with $q_{c_i}$ inputs and no output (or a fictitious output with weight zero). Let $A^c_{w_i}$ be the input weight enumerator for the check node $c_i$ that satisfies the check node constraint. The $A^c_{w_i}$ represents the number of sequences with input weight $w_i = (w_{v_1}, w_{v_2}, \ldots, w_{v_{m_v}})$. Let $A_d$ represent number of sequences each with weight vector $d = (d_1, d_2, \ldots, d_m)$ that is applied to the variable nodes and satisfy the protograph constraints. Then ensemble partial weight enumerators for the protograph based LDPC can be computed as:

$$A_d = \sum_{d_1, \ldots, d_m} \frac{\prod_{k=1}^{n_v} (N)\delta_{d_k,w_{v_k}} \ldots \delta_{d_k,w_{v_k}}}{\prod_{i=1}^{n_v} (N)_{w_{v_i}}} A^c_{w_i}$$

which reduces to

$$A_d = \frac{\prod_{k=1}^{n_v} A^c_{w_i}}{\prod_{i=1}^{n_v} (N)_{w_{v_i}}}$$

where $d_i = (d_{i1}, d_{i2}, \ldots, d_{im})$. Partial weights $d_{ik}$ corresponds to edge connections of a check node $c_i$ to the set of variable nodes $\{v_{i1}, v_{i2}, \ldots, v_{imi}\}$. Let $S_d$ be the set of all possible partial weights of $m$ variable nodes that are transmitted through the channel, say with indices $i_1, i_2, \ldots, i_m$ such that $d_{i1} + d_{i2} + \ldots + d_{im} = d$. Let $S_p$ be the set of all possible partial weights of the remaining variable nodes in the protograph, namely the punctured variable nodes. Then the ensemble weight enumerator for the protograph LDPC code can be written as:

$$A_d = \sum_{\{d_k\} \in S_d} \sum_{\{d_l\} \in S_p} A_d$$

A. Computation of $A^c_{w_i}$ for a check $c$ with degree 3

Define $s = \frac{w_1 + w_2 + w_3}{2}$. If $w_1 + w_2 + w_3$ is even and $\max\{w_1, w_2, w_3\} \leq s \leq N$, then

$$A_{w_1,w_2,w_3} = \frac{(N)}{(s - w_1)! (s - w_2)! (s - w_3)!}$$

otherwise $A_{w_1,w_2,w_3} = 0$.

Proof: We use the following theorem.

Multinomial theorem:

$$\sum_{i=1}^{k} \frac{k!}{n_1!n_2! \ldots n_k!}$$

where

$$C(n; n_1, n_2, \ldots, n_k) = \frac{n!}{n_1!n_2! \ldots n_k!}$$

Multidimensional z-transform for $A^c_{w_1,w_2,w_3}$ can be written as:

$$A^c_{W_1,W_2,W_3} = (1 + W_1W_2 + W_1W_3 + W_2W_3)^N$$

$$= \sum_{n_1,n_2,n_3=0}^{N} C(N; n_1, n_2, n_3, n_4) \times W_1^{n_1}W_2^{n_2}W_3^{n_3}(W_2W_3)^{n_4}$$

$$= \sum_{n_1,n_2,n_3,n_4=0}^{N} C(N; n_1, n_2, n_3, n_4) \times W_1^{n_1}W_2^{n_2}W_3^{n_3}W_3^{n_4}$$

To obtain $A^c_{w_1,w_2,w_3}$ we need to set $n_2 + n_3 = w_1$, $n_2 + n_4 = w_2$, and $n_3 + n_4 = w_3$. Solving this set of equations we get $n_1 = N - s$, $n_2 = s - w_3$, $n_3 = s - w_2$, and $n_4 = s - w_1$, where $s = \frac{w_1 + w_2 + w_3}{2}$, provided that $w_1 + w_2 + w_3$ is even. Since $n_i \geq 0$ this implies that $\max\{w_1, w_2, w_3\} \leq s \leq N$. Q.E.D.

B. Computation of $A^c_{w_i}$ for a check $c$ with degree higher than 3

The partial weight enumerators for checks with degree higher than 3 can be obtained from the result for a check with
degree 3 by concatenation. For example \( A_{w_1,w_2,w_3,w_4}^c \) can be obtained as
\[
A_{w_1,w_2,w_3,w_4}^c = \sum_{i=1}^{N} A_{w_1,w_2,w_3,w_4} A_{w_3,w_4} \binom{N}{i}
\]
(8)
The weight enumerators for higher degree checks can be obtained in a similar way.

III. ASYMMETRIC ENSEMBLE WEIGHT ENUMERATORS

We express the normalized logarithmic asymptotic weight distribution of a code as \( r(\delta) = \frac{\ln(A_d)}{n} \) where \( d \) is Hamming distance, \( \delta = \frac{d}{n} \) and \( A_d \) is the ensemble weight distribution. One application of having \( r(\delta) \) to obtain an upper bound on minimum \( E_b/N_0 \) threshold when maximum likelihood decoding is used. The bound has been derived in [7]. This bound is the tightest closed-form bound on minimum \( E_b/N_0 \) that can be written as
\[
\left( \frac{E_b}{N_0} \right)_{\text{min}} \leq \frac{1}{R_c} \max\left(1 - e^{-2r(\delta)} \right) \frac{1 - \delta}{2\delta}
\]
(9)

Let \( \gamma_{\text{ml}} \) represent the upper bound in (9) in dB. We use this maximum likelihood threshold in the examples to compare with the iterative decoding threshold for infinite block size which will be denoted by \( \gamma_{\text{iter}} \) in dB.

If the first zero crossing of \( r(\delta) \) (i.e. \( r(\delta_{\text{min}}) = 0 \) for \( \delta_{\text{min}} > 0 \)) exists, then it will indicate that the non-zero normalized minimum distance of the code grows linearly with the block size, i.e. \( \delta_{\text{min}} = \delta_{\text{min}} \times n \).

However to make the expressions simple, we first define \( \tilde{r}(\delta) = \frac{\ln(A_d)}{n} \) where \( \delta = \frac{d}{n} \) as \( N \to \infty \). Then \( r(\delta) \) can be obtained as \( r(\delta) = \frac{1}{m} \tilde{r}(m\delta) \), where \( m \) represents the number of variable nodes in the protograph that are connected to the channel and \( n = mN \).

\[
\tilde{r}(\delta) = \max_{\{d_i\} \in \delta} \max_{\{d_i\} \in \delta} \lim_{N \to \infty} \sum_{i=1}^{n} \frac{\ln(A_d)}{N} = \frac{\ln(A_d)}{N}
\]
(10)

Define \( a^{\infty}(\delta_i) = \lim_{N \to \infty} \frac{\ln(A_d)}{N} \). Vector \( \delta_i \) represents the normalized version of the vector \( d_i \), where each component is normalized by \( N \). Then
\[
\tilde{r}(\delta) = \max_{\{\delta_i\} \in \delta} \max_{\{\delta_i\} \in \delta} \sum_{i=1}^{n} a^{\infty}(\delta_i) - \sum_{i=1}^{n} (q_{\delta_i} - 1)H(\delta_i)
\]
(11)

where \( H(x) = -(1 - x) \ln(1 - x) - x \ln x \) is the entropy function. The sets \( S_\delta \) and \( S_\varepsilon \) are normalized version of the sets \( S_\delta \) and \( S_\varepsilon \) i.e. each component of the latter sets is divided by \( N \) as \( N \to \infty \).

Let \( \sigma = (\varepsilon_1 + \varepsilon_2 + \varepsilon_3)/2 \) such that \( \max\{\varepsilon_1, \varepsilon_2, \varepsilon_3\} \leq \sigma \leq 1 \), and define
\[
H_3(x_1, x_2, x_3) = -(1 - \sum_{i=1}^{3} x_i \ln(1 - \sum_{i=1}^{3} x_i) - \sum_{i=1}^{3} x_i \ln x_i
\]
(12)
then
\[
a^{\infty}(\varepsilon_1, \varepsilon_2, \varepsilon_3) = H(3(\varepsilon_1, \varepsilon_2, \varepsilon_3))
\]
(13)
\[
a^{\infty}(\varepsilon_1, \varepsilon_2, \varepsilon_3, \varepsilon_4) = \max\{a^{\infty}(\varepsilon_1, \varepsilon_2, \lambda) + a^{\infty}(\varepsilon_3, \varepsilon_4, \lambda) - H(\lambda)\}
\]
(14)
where
\[
\max\{\varepsilon_1 - \varepsilon_2, \varepsilon_3 - \varepsilon_4\} \leq \lambda \leq \min((\varepsilon_1 + \varepsilon_2), (\varepsilon_3 + \varepsilon_4), (2 - (\varepsilon_1 + \varepsilon_2)), (2 - (\varepsilon_3 + \varepsilon_4)))
\]
(15)

The asymptotic weight enumerators for higher degree checks can be obtained in a similar way.

A. Example 1

Consider a (3,6) regular LDPC code with a protograph as shown in Fig. 3(a).

\begin{figure}[h]
\centering
\includegraphics[width=0.8\linewidth]{fig3}
\caption{(a) Protograph of a (3,6) regular LDPC code, (b) equivalent protograph for ensemble weight enumeration.}
\end{figure}

Using the result in (2), it is easy to show that
\[
A_{d_1,d_2} = A_{d_1,d_1,d_1,d_1,d_1,d_1}^c \left( \frac{n}{d_1} \right)^2
\]
(16)
Using the results in previous and this section we can compute \( A_{d_1,d_1,d_1,d_1,d_1,d_1}^c \) using (4), or we can use an equivalent protograph as shown in Fig. 3(b) with all degree 3 check nodes
to compute the ensemble weight enumerators. Using either approach we get

\[ A_{d_1,d_2} = \sum_{i_1,i_2,i_3} \frac{A^{c_{1,1}}_{d_1,i_1,i_3} A^{c_{1,2}}_{d_2,i_2,i_3} A^{c_{1,3}}_{d_3,i_1,i_2} A^{c_{1,4}}_{d_4,i_3,i_1} A^{c_{1,5}}_{d_5,i_2,i_3}}{(N^2)^3 (N^2)^3 (N^2)^3 (N^2)^3 (N^2)^3} \]  

(17)

The asymptotic weight enumerators can be obtained as

\[ \tilde{f}(\delta_1, \delta_2) = \max_{\delta_1, \delta_2} \{ a^{c_{1,1}}(\delta_1, \delta_1, \lambda_1) + a^{c_{1,2}}(\delta_1, \lambda_1, \lambda_2) + a^{c_{1,3}}(\delta_2, \lambda_2, \lambda_3) + a^{c_{1,4}}(\delta_2, \delta_2, \lambda_3) + a^{c_{1,5}}(\delta_2, \delta_2, \lambda_3) \} \]

(18)

\[ \tilde{f}(\delta) = \max_{\delta_1} \tilde{f}(\delta_1, \delta - \delta_1) \]  

(19)

![Graph showing asymptotic normalized weight distributions and zero crossings for (3,6) regular LDPC, precoded (3,6) LDPC, and rate 1/2 random codes.](image)

**IV. PRECODED LDPC CODES**

LDPC codes with protograph structure can be precoded with an accumulator for better performance. Precoding places a degree 1 variable node between a constraint node and a higher degree variable (forming an accumulator) which is then optionally erased. Precoding often lowers the iterative decoding threshold of a given protograph without altering its rate [10]. Through examples we show that precoding also improves \( r(\delta) \) and \( \delta_{min} \).

Next we investigate the weight enumerators for precoded LDPC codes. Consider an LDPC code with a protograph having \( n_v \) variable nodes that are connected to the channel. Let \( A_{d_1, d_2, \ldots, d_n} \) be the ensemble weight enumerators for partial weights \( \{d_1, d_2, \ldots, d_n\} \) for this protograph. Without loss of generality suppose we precode the variable node \( v_1 \) with an accumulator as shown in the Fig. 5. The node is disconnected from the channel (punctured) and it is connected to the output of an accumulator. The input of accumulator is connected to the channel. Define the ensemble weight enumerators for the precoded version of the original protograph by \( A^\text{pre}_{d_0, d_1, d_2, \ldots, d_n} \), where \( d_0 \) is the partial weight of the variable node \( v_0 \) corresponding to the input of the accumulator.

Using the result in (2), it is easy to show that

\[ A^\text{pre}_{d_0, d_1, d_2, \ldots, d_n} = A_{d_1, d_2, \ldots, d_n} A^\text{co}_{d_0, d_1, d_2} (\frac{N}{d_1})^2 \]  

(20)

However

\[ \frac{A^\text{co}_{d_0, d_1, d_2}}{(\frac{N}{d_1})^2} = \theta(d_0, d_1) (\frac{d_0}{d_1}) \]  

(21)

Note that \( d_1 \geq \frac{d_0}{2} \), and \( d_0 \) is even. When \( d_1 = \frac{d_0}{2} \) we can show that the right hand side of (21) is less than 1. For \( d_1 > \frac{d_0}{2} \) we can upperbound the right hand side of (21) by

\[ \frac{A^\text{co}_{d_0, d_1, d_2}}{(\frac{N}{d_1})^2} \leq \theta(d_0, d_1) N^-(d_1 - \frac{d_0}{2}) \]  

(22)

where \( \theta(d_0, d_1) \) does not depend on \( N \). Thus the precoding introduces a so-called interleaving gain. Also for fixed \( d_0 \) and \( d_1 \), there exist an \( N_1 \) such that for \( N > N_1 \) we have \( \theta(d_0, d_1) N^-(d_1 - \frac{d_0}{2}) < 1 \). Thus

\[ A^\text{pre}_{d_0, d_1, d_2, \ldots, d_n} < A_{d_1, d_2, \ldots, d_n} \]  

(23)

The asymptotic result can be written as

\[ \tilde{f}^\text{pre}(\delta_0, \delta_1, \delta_2, \ldots, \delta_n) < \tilde{f}(\delta_1, \delta_2, \ldots, \delta_n) \]  

(24)

**A. Example 2**

Consider the (3,6) regular LDPC code in the example 1. This code has \( \delta_{min} = 0.023 \), \( \gamma_{ml} = 0.79 \) dB, and \( \gamma_{iter} = 1.1 \) dB. However if we precode the (3,6) LDPC code with an accumulator, then we obtain \( \delta_{min} = 0.033 \), \( \gamma_{ml} = 0.31 \) dB, and \( \gamma_{iter} = 0.87 \) dB. The asymptotic normalized weight distributions and zero crossings that specify the \( \delta_{min} \) for the precoded (3,6) LDPC in this example, are shown in Figures 4.
B. Example 3

Consider Repeat Jagged Accumulate (RJA) and its precoded version Accumulate Repeat Jagged Accumulate (ARJA) LDPC codes [20]. The RJA code has $\delta_{min}=0.013$, $\gamma_{ml}=0.83$ dB, and $\gamma_{iter}=1.0$ dB. However if we precode the RJA LDPC code with an accumulator namely ARJA, then we obtain $\delta_{min}=0.015$, $\gamma_{ml}=0.35$ dB, and $\gamma_{iter}=0.628$ dB. The asymptotic normalized weight distributions and zero crossing for the RJA and ARJA LDPC codes are shown in Fig. 6.

![Figure 6: Asymptotic normalized weight distributions for RJA LDPC code, precoded RJA (ARJA) LDPC code, and rate 1/2 random codes](image)

C. Example 4

If we increase the degree 2 variable nodes to 2/3 of the inner checks in ARJA LDPC code we can further reduce the iterative decoding threshold at the expense of lowering the linearity coefficient that represent the growth of minimum distance with respect to $n$. An example of such code is shown in Fig. 7. Note that this code does not satisfy the relation $\lambda'(0)/\rho'(1) < 1$ [14], where $\lambda(x)$, and $\rho(x)$ are the degree distributions for variable and constraint nodes. However, we note that for a protograph this condition is only a sufficient, but not a necessary, condition for minimum distance growing with $n$. Specifically, the ensemble asymptotic minimum distance over block size for this protograph is a small, but positive number $\delta_{min} = 0.004$.

![Figure 7: Protograph of rate 1/2 ARJA with 5-checks (2/3 of inner checks are degree 2). The asymptotic minimum distance over block size for this code is $\delta_{min} = 0.004$](image)

ACKNOWLEDGMENT

This research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under contract with the National Aeronautics and Space Administration.

REFERENCES