

On the existence of typical minimum distance for protograph-based LDPC Codes

Shadi Abu-Surra
Samsung Telecommunications America
Email: sasurra@sta.samsung.com

Dariush Divsalar
Jet Propulsion Laboratory
California Institute of Technology
Email: Dariush.Divsalar@jpl.nasa.gov

William E. Ryan
University of Arizona
Email: ryan@ece.arizona.edu

Abstract—In this paper we prove that, for a certain class of protograph-based LDPC codes with degree-two variable nodes, a typical minimum distance exists. We obtain a tight bound on the sum of weight enumerators, up to some weight d^* , for the ensemble of finite-length protograph LDPC codes. Then we prove that this sum goes to zero as the block length goes to infinity. Finally, we prove that $\Pr(d < d^*)$ goes to zero as the block length goes to infinity. This typical minimum distance exists if degree-two nodes have certain connections to the check nodes. This is also important in practice since it identifies a certain class of protograph LDPC codes that have typical minimum distances.

I. INTRODUCTION

Low-density parity-check (LDPC) codes were proposed by Gallager [1] in 1963. Ensemble weight enumerators for unstructured irregular LDPC codes and turbo-like codes have been reported in [10], [11], [12], [13], [14], [15], [16], [17], [18], [2], [3]. Recently, researchers became interested in the design of LDPC codes with imposed sub-structures, starting with the introduction of multi-edge type codes in [7] and [8]. Protograph-based LDPC codes are a subclass of multi-edge LDPC codes. In [9] a method for the computation of asymptotic (infinite block size) weight enumerators for LDPC codes with protograph structure has been proposed. In [4] ensemble weight enumerators for finite block size LDPC codes with a protograph structure was obtained. The results then were extended to the asymptotic case as the block size goes to infinity. Weight enumerators for specific codes are useful for bounding or estimating the decoding error probability of channel codes. As noted by Gallager [1], it is generally impractical to calculate the weight enumerator for a given code. Given this, Gallager and others have calculated the average performance for ensembles of codes. Gallager derived asymptotic weight enumerators for the ensembles of regular LDPC codes. This result was extended to the irregular LDPC ensembles (see the above references). In this present paper, we upper bound the ensemble weight enumerators of protograph-based LDPC codes to prove the existence of a typical minimum distance for a class of protograph LDPC codes with degree-2 variable nodes. The existence of a typical minimum distance implies linear growth of the minimum distance with the code block length [1]. For any protograph LDPC code, we use a random permutation per each edge. Then we obtain the weight enumerators by averaging over all

possible permutations. This is equivalent to using a uniform interleaver [6] per each edge of the protograph. There was few research work on bounding the minimum distance, and linear minimum distance property for example see [22], [23], [24], [25] and references there. Recently in [20] and later in [21], the minimum distance was upper bounded using circulant permutations. The results show that if circulant permutations are used, the minimum distance will not grow linearly with the code block length.

This paper proceeds as follows: In Section II, we define a protograph LDPC code. In Section III, we define our notation and provide the ensemble weight enumerators for protograph-based LDPC codes as background. In Section IV, we define a class of protograph LDPC codes with typical minimum distance. Finally, in Section V, we prove the existence of typical minimum distance for this class of protograph LDPC codes.

II. PROTOGRAPH-BASED LDPC CODES

A protograph is a Tanner graph with a relatively small number of variable nodes (VNs) and check nodes (CNs) [5], [19], [26]. A protograph $G = (V, C, E)$ consists of a set of variable nodes $V = \{v_1, v_2, \dots, v_{n_v}\}$, a set of check nodes $C = \{c_1, c_2, \dots, c_{n_c}\}$, 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. 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. 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. Denote by q_{v_i} the degree of variable node v_i . Denote by q_{c_j} the degree of check node c_j . The code rate for the protograph is $R_c = \frac{n_v - n_c}{n_t}$ provided that the parity check matrix of the derived or lifted graph is full

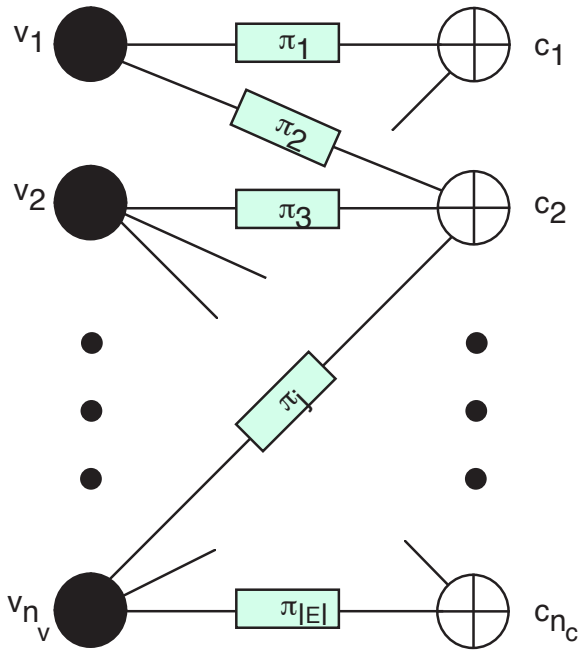


Fig. 1. Vectorized protograph.

rank; n_t is the number of transmitted variable nodes, $n_t \leq n_v$. Here for simplicity we assume $n_t = n_v$. The copy-and-permute process can be simply represented by replacing each node with a vector of nodes of the same type and replacing each edge with a bundle of (permuted) edges of the same type. This “vectorized” protograph is depicted in Fig. 1.

III. ENSEMBLE WEIGHT ENUMERATORS FOR FINITE-LENGTH PROTOGRAPH-BASED LDPC CODES

Now consider the LDPC code constructed from a protograph G by making N replicas of G and using uniform interleavers, each of size N , to permute the edges among the replicas of the protograph. We treat the VNs and CNs as constituent codes in a concatenated coding scheme. More specifically, the group of N VNs of type v_i is considered to be a constituent (repetition) code with a weight- d_i input of length N and q_{v_i} length- N outputs. Also, the group of N CNs of type c_j is considered to be a constituent code with q_{c_j} inputs, each of length N , and a fictitious output of weight zero. Let $A(\mathbf{d})$ be the average (over the ensemble) number of codewords having weight vector $\mathbf{d} = [d_1, d_2, \dots, d_{n_v}]$ corresponding to the n_v VN N -groups and satisfying the protograph constraints. $A(\mathbf{d})$ is the *weight vector enumerator* for the ensemble of codes of length $N \cdot n_v$ described by the protograph. Let us further define

$A^{v_i}(\mathbf{w}_i) = \binom{N}{d_i} \delta_{d_i, w_{i,1}} \cdots \delta_{d_i, w_{i, q_{v_i}}} =$ the weight vector enumerator for the type- v_i (VN) constituent code for a weight- d_i input, where $\mathbf{w}_i = [w_{i,1}, w_{i,2}, \dots, w_{i, q_{v_i}}]$ is a weight vector describing the constituent code’s output, and

$A^{c_j}(\mathbf{z}_j) =$ the weight vector enumerator for the type- c_j (CN) constituent code and $\mathbf{z}_j = [z_{j,1}, z_{j,2}, \dots, z_{j, q_{c_j}}]$, where

$z_{j,l} = w_{i,k}$ if the l^{th} edge of CN c_j is the k^{th} edge of VN v_i .

As shown in [4, Eq. 2], by exploiting the uniform interleaver [6] property, we may write

$$\begin{aligned} A(\mathbf{d}) &= \sum_{w_{m,u}} \frac{\prod_{i=1}^{n_v} A^{v_i}(\mathbf{w}_i) \prod_{j=1}^{n_c} A^{c_j}(\mathbf{z}_j)}{\prod_{s=1}^{n_v} \prod_{r=1}^{q_{v_s}} \binom{N}{w_{s,r}}} \\ &= \frac{\prod_{j=1}^{n_c} A^{c_j}(\mathbf{d}_j)}{\prod_{i=1}^{n_v} \binom{N}{d_i}^{q_{v_i}-1}} \end{aligned} \quad (1)$$

where the summation in the first line is over all weights $w_{m,u}$, $m = 1, \dots, n_v$ and $u = 1, \dots, q_{v_m}$, and $\mathbf{d}_j = [d_{j,1}, d_{j,2}, \dots, d_{j, q_{c_j}}]$ is a weight vector which describes the weights of the N -bit words on the edges connected to CN c_j , produced by the VNs neighboring c_j . The elements of \mathbf{d}_j comprise a subset of the elements of \mathbf{d} .

Then the average number of codewords of weight d in the ensemble, denoted by A_d , equals the sum of $A(\mathbf{d})$ over all \mathbf{d} for which $\sum_{\{d_i: v_i \in V\}} d_i = d$. Notationally,

$$A_d = \sum_{\{d_i: v_i \in V\}} A(\mathbf{d}) \quad (2)$$

under the constraint $\sum_{\{d_i: v_i \in V\}} d_i = d$. To evaluate A_d in (2), one first needs to compute the weight vector enumerators, $A^{c_j}(\mathbf{d}_j)$, for the check nodes c_j , as seen in (1).

Consider a check node c with degree 3. We need to find its weight vector enumerator $A^c(\mathbf{w})$, where $\mathbf{w} = [w_1, w_2, w_3]$ is the weight vector at the input to a degree-3 check node. Following [4], the $A^c(\mathbf{w})$ may be easily found as the coefficients of the multi-dimensional z-transform of $\{A^c(\mathbf{w})\}$ as

$$A^c(w_1, w_2, w_3) = \binom{N}{s} \frac{s!}{(s-w_1)!(s-w_2)!(s-w_3)!} \quad (3)$$

where $s = \frac{w_1+w_2+w_3}{2}$. This is true if $w_1+w_2+w_3$ is even and $\max\{w_1, w_2, w_3\} \leq s \leq N$, otherwise $A^c(w_1, w_2, w_3) = 0$. 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 $A^c(w_1, w_2, w_3, w_4)$ can be obtained as

$$A^c(w_1, w_2, w_3, w_4) = \sum_{l=1}^N \frac{A(w_1, w_2, l)A(w_3, w_4, l)}{\binom{N}{l}} \quad (4)$$

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

IV. A CLASS OF PROTOGRAPH LDPC CODES WITH TYPICAL MINIMUM DISTANCE

Consider a class of protograph-based LDPC codes where the connections between degree-2 VNs and CNs in the protograph have certain restrictions to be defined.

Any protograph with degree-2 nodes that satisfies the following criterion belongs to this class. First consider the set of degree-2 VN nodes for which each node in this set is only connected to a CN node through a single edge connection. Remove from the set each degree-2 node and the two edges connected to this degree-2 node. Repeat this process for the

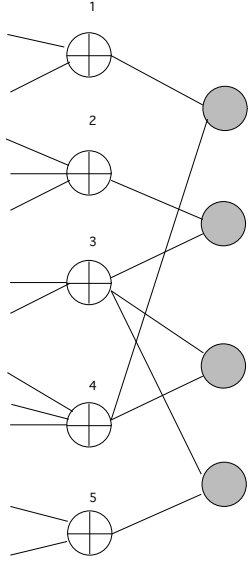


Fig. 2. A member of the class in which no degree-2 loops exist (Only degree-2 VNs and check nodes connected to them are shown)

remaining degree-2 VN nodes. If at the end of this process no degree-2 node is left, then the protograph-based LDPC code belongs to the class of protograph-based LDPC code ensembles with typical minimum distance. In this class, no degree-2 node is allowed to be connected to a CN node through double edges. This constraint also implies that no loop should exist in the graph between degree-2 nodes and the checks connected to these nodes. A member of this class is shown in Fig. 2

This class of protograph LDPC codes with degree-2 variable nodes is also similar to the class of protograph-based LDPC codes that are derived from a protograph with variable node degrees at least 3 using the check split operation as described in [4] and [19].

V. EXISTENCE OF TYPICAL MINIMUM DISTANCE

Now we will prove that class of protograph-based LDPC code ensemble with degree-2 nodes that was described in the previous section has a typical minimum distance. We first prove that there exists a $\tilde{\delta}^* > 0$ such that $\sum_{d=1}^{\lfloor N\tilde{\delta}^* \rfloor} A_d \rightarrow 0$ as $N \rightarrow \infty$, where the block size $n = n_v N$. Then, using Markov's inequality we can show $Pr\{d_{min} < \lfloor N\tilde{\delta}^* \rfloor\} \rightarrow 0$ as $N \rightarrow \infty$.

Proof: We have already computed $A^c(\mathbf{w})$ for a check c (SPC) with degree 3. If $w_1 + w_2 + w_3$ is even and

$\max\{w_1, w_2, w_3\} \leq u \leq N$, then

$$A^c(w_1, w_2, w_3) = \frac{N!}{(N-s)!(s-w_1)!(s-w_2)!(s-w_3)!} \quad (5)$$

otherwise $A^c(w_1, w_2, w_3) = 0$, where $s = \frac{w_1+w_2+w_3}{2}$. First we obtain an upper bound to $A^c(w_1, w_2, w_3)$. Note that

$$\frac{N!}{(N-s)!} \leq N^s \quad (6)$$

$$\begin{aligned} & \frac{1}{(s-w_1)!(s-w_2)!(s-w_3)!} \\ &= \frac{1}{s!} \frac{1}{s!} \frac{1}{s!} \frac{1}{(s!)^3} \end{aligned} \quad (7)$$

$$\frac{s!}{(s-w_i)!} \leq s^{w_i} \quad ; i = 1, 2, 3. \quad (8)$$

$$s! \geq s^{s+\frac{1}{2}} e^{-s} \quad (9)$$

Further, $s \geq w_i$, $i = 1, 2, 3$, implies $-\ln s \leq -\ln w_i$.

$$A^c(w_1, w_2, w_3) \leq \prod_{i=1}^3 N^{\frac{w_i}{2}} e^{\frac{3}{2}w_i - \frac{1}{2}w_i \ln(w_i) - \frac{1}{2}U(w_i) \ln(w_i)} \quad (10)$$

where $U(w_i)$ is a unit step function, i.e., for $w_i \neq 0$, $U(w_i)=1$, otherwise it is zero. The unit step function was introduced to cover the cases when $w_i=0$. An upper bound on $A^c(\mathbf{w})$ for a check c with degree 4 can be obtained using the check split method discussed in [4] and [19]. If we split the check node into two checks and connect them with a degree-2 variable node, we get

$$A^c(w_1, w_2, w_3, w_4, l) = \frac{A(w_1, w_2, l)A(w_3, w_4, l)}{\binom{N}{l}} \quad (11)$$

Using the upper bound (10) for a degree-3 check node, we get

$$\begin{aligned} & A^c(w_1, w_2, w_3, w_4, l) \leq \\ & \prod_{i=1}^4 N^{\frac{w_i}{2}} e^{\frac{3}{2}w_i - \frac{1}{2}w_i \ln(w_i) - \frac{1}{2}U(w_i) \ln(w_i)} \\ & \times e^{3l - U(l) \ln(l)} \end{aligned} \quad (12)$$

We can obtain $A^c(w_1, w_2, w_3, w_4)$ by summing over l as

$$A^c(w_1, w_2, w_3, w_4) = \sum_{l=0}^N A^c(w_1, w_2, w_3, w_4, l) \quad (13)$$

Note that

$$l \leq \min\{(w_1+w_2), (w_3+w_4)\} \leq \frac{w_1 + w_2 + w_3 + w_4}{2} \triangleq l_{max}$$

Then

$$\begin{aligned} & A^c(w_1, w_2, w_3, w_4) \leq \\ & \prod_{i=1}^4 N^{\frac{w_i}{2}} e^{\frac{3}{2}w_i - \frac{1}{2}w_i \ln(w_i) - \frac{1}{2}U(w_i) \ln(w_i)} \\ & \times \sum_{l \leq \frac{w_1+w_2+w_3+w_4}{2}} e^{3l - U(l) \ln(l)} \end{aligned} \quad (14)$$

But the summation can be upper bounded as

$$\sum_{l \leq \frac{w_1 + w_2 + w_3 + w_4}{2}} e^{3l - U(l) \ln(l)} \leq l_{max} e^{3l_{max} - \ln(l_{max})} = e^{3l_{max}} \quad (15)$$

The upper bound can be written as

$$A^c(w_1, w_2, w_3, w_4) \leq \prod_{i=1}^4 N^{\frac{w_i}{2}} e^{\frac{3}{2} 2w_i - \frac{1}{2} w_i \ln(w_i) - \frac{1}{2} U(w_i) \ln(w_i)} \quad (16)$$

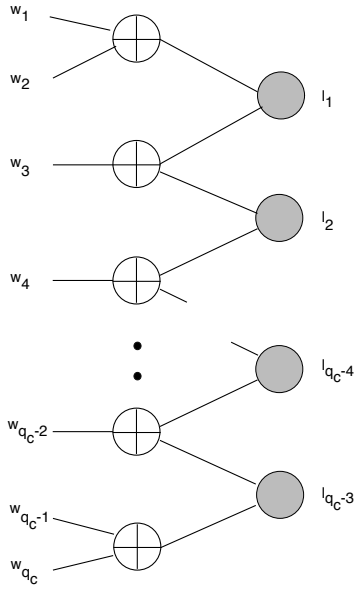


Fig. 3. Check split method to compute enumerators for a check node with degree q_c

An upper bound on $A^c(\mathbf{w})$ for a check c with degree q_c can be obtained either by using mathematical induction or directly applying the check split method to a check with degree q_c . Here we use the latter. Now suppose we use the check split method for a check node of degree q_c and we generate $q_c - 3$ degree-2 nodes. Then we have

$$A^c(w_1, w_2, w_3, \dots, w_{q_c}, l_1, l_2, \dots, l_{q_c-3}) = \frac{A(w_1, w_2, l_1) A(l_1, w_3, l_2) \dots A(l_{q_c-3}, w_{q_c-1}, w_{q_c})}{\binom{N}{l_1} \binom{N}{l_2} \dots \binom{N}{l_{q_c-3}}} \quad (17)$$

Using the upper bound for a degree-3 check, we have

$$A^c(w_1, w_2, w_3, \dots, w_{q_c}, l_1, l_2, \dots, l_{q_c-3}) \leq \prod_{i=1}^{q_c} N^{\frac{w_i}{2}} e^{\frac{3}{2} w_i - \frac{1}{2} w_i \ln(w_i) - \frac{1}{2} u(w_i) \ln(w_i) - \frac{1}{2} U(w_i) \ln(w_i)} \times \prod_{j=1}^{q_c-3} e^{3l_j - U(l_j) \ln(l_j)} \quad (18)$$

Summing over $l_j, j = 1, \dots, (q_c - 3)$, we get

$$A^c(w_1, w_2, w_3, \dots, w_{q_c}) = \sum_{l_1} \sum_{l_2} \dots \sum_{l_{q_c-3}} A^c(w_1, w_2, w_3, \dots, w_{q_c}, l_1, l_2, \dots, l_{q_c-3}) \quad (19)$$

For each check in the Figure 3 we have $l_1 \leq w_1 + w_2$, $l_j \leq l_{j-1} + w_{j+1}$ for $j = 2, \dots, q_c - 3$, $l_j \leq l_{j+1} + w_{j+2}$ for $j = 1, \dots, q_c - 4$, and $l_{q_c-3} \leq w_{q_c-1} + w_{q_c}$. These inequalities imply $l_j \leq \sum_{k=1}^{j+1} w_k$ and $l_j \leq \sum_{k=j+2}^{q_c} w_k$. Thus, $l_j \leq \min\{\sum_{k=1}^{j+1} w_k, \sum_{k=j+2}^{q_c} w_k\} \leq \frac{1}{2} \sum_{k=1}^{q_c} w_k \triangleq l_{max}$. From these inequalities we can also conclude that for any check of degree q_c we should have $w_i < \sum_{j=1, j \neq i}^{q_c} w_j$ or $\max\{w_1, w_2, \dots, w_{q_c}\} \leq \frac{1}{2} \sum_{j=1}^{q_c} w_j$. This bounding techniques for the weight of degree-2 nodes can be used to prove that the total weight L of degree-2 nodes in the class of protograph codes with typical minimum distance can be upper bounded as

$$L \leq \gamma u, \quad (20)$$

where u is the total weight of the other variable nodes with degree at least 3.

Now using the above results

$$\sum_{l_j \leq l_{max}} e^{3l_j - u(l_j) \ln(l_j)} \leq l_{max} e^{3l_{max} - \ln(l_{max})} = e^{3l_{max}} \quad (21)$$

The upper bound can be written as

$$A^c(w_1, w_2, \dots, w_{q_c}) \leq \prod_{i=1}^{q_c} N^{\frac{w_i}{2}} e^{\frac{3}{2} (q_c-2) w_i - \frac{1}{2} w_i \ln(w_i) - \frac{1}{2} u(w_i) \ln(w_i)} \quad (22)$$

Using the above results, the weight vector enumerator

$$A(\mathbf{d}) = \frac{\prod_{j=1}^{n_c} A^{c_j}(\mathbf{d}_j)}{\prod_{i=1}^{n_v} \binom{N}{d_i}^{q_{v_i}-1}} \quad (23)$$

can be upper bounded. The numerator $\prod_{j=1}^{n_c} A^{c_j}(\mathbf{d}_j)$, which is a product over the check nodes, with some manipulation can be upper bounded using the above result for check nodes as

$$\prod_{i=1}^{n_v} N^{\frac{1}{2} q_{v_i} d_i} e^{\frac{3}{2} (\sum_{j=1}^{q_{v_i}} (q_{c_j^{(i)}} - 2)) d_i - \frac{1}{2} q_{v_i} d_i \ln(d_i) - \frac{1}{2} q_{v_i} u(d_i) \ln(d_i)}$$

In the above result the product now is over the variable nodes and the checks $c_j^{(i)}$ ($j = 1, \dots, q_{v_i}$) are adjacent to variable node v_i (neighbors of v_i). Now we lower bound the denominator $\prod_{i=1}^{n_v} \binom{N}{d_i}^{q_{v_i}-1}$ in the weight vector equation by

$$\prod_{i=1}^{n_v} N^{(q_{v_i}-1)d_i} e^{-(q_{v_i}-1)d_i \ln(d_i)}.$$

Using $-\frac{1}{2}q_{v_i}u(d_i) \ln(d_i) \leq 0$ the weight vector enumerator now can be upper bounded as

$$A(\mathbf{d}) \leq \prod_{i=1}^{n_v} e^{\frac{1}{2}(q_{v_i}-2)d_i \ln(\frac{d_i}{N}) + \frac{3}{2}(\sum_{j=1}^{q_{v_i}} (q_{c_j^{(i)}}-2))d_i} \quad (24)$$

Let q_c^{max} be the maximum check degree in the protograph. Let t_2 be the number of degree-2 variable nodes. We separate the degree-2 nodes. The upper bound can be written as

$$A(\mathbf{d}) \leq \prod_{i=1}^{n_v-t_2} e^{\frac{1}{2}(q_{v_i}-2)d_i \ln(\frac{d_i}{N}) + \frac{3}{2}q_{v_i}(q_c^{max}-2)d_i} \prod_{k=1}^{t_2} e^{3(q_c^{max}-2)d_k} \quad (25)$$

In the above bound $q_{v_i} \geq 3$. Consider the following function of q_{v_i}

$$\frac{1}{2}(q_{v_i}-2)d_i \ln(\frac{d_i}{N}) + \frac{3}{2}q_{v_i}(q_c^{max}-2)d_i$$

This function is decreasing in q_{v_i} if $\frac{d_i}{N} \leq e^{-3(q_c^{max}-2)}$ is true. We can prove that this inequality is true for a distance region of interest, namely, $d \leq d_o$ (d_o will be defined shortly). Thus, the above function is maximum when $q_{v_i} = 3$. Define the sum of weights of all variable nodes with at least degree 3 as $u = \sum_{i=1}^{n_v-t_2} d_i$, and the sum of weights of all degree-2 variable nodes as $L = \sum_{k=1}^{t_2} d_k$. The total weight of all nodes is $d = u + L$. In (21) we had $L \leq \gamma u$ where γ depends on the particular connections of degree-2 nodes to checks in our class of protograph codes. However, we can obtain the worst-case upper bound per weight of a degree-2 node as $l \leq \frac{1}{2} \sum_{i=1}^{n_v-t_2} q_{v_i} d_i$. With further upper bounding we get a worst-case value for γ as $\frac{1}{2}q_v^{max}t_2$, where q_v^{max} is the maximum variable node degree. Then $L \leq \gamma u = \gamma(d-L)$. This implies that $L \leq \frac{\gamma}{1+\gamma}d$. With the above results, we have

$$A(\mathbf{d}) \leq e^{3(q_c^{max}-2)L} \prod_{i=1}^{n_v-t_2} e^{\frac{1}{2}d_i \ln(\frac{d_i}{N}) + \frac{3}{2}(q_c^{max}-2)d_i} \quad (26)$$

Using $\ln d_i \leq \ln u$, we get

$$A(\mathbf{d}) \leq e^{3(q_c^{max}-2)L} e^{\frac{1}{2}u \ln(\frac{u}{N}) + \frac{3}{2}(q_c^{max}-2)u} \quad (27)$$

or

$$A(\mathbf{d}) \leq e^{E(d,L)} \quad (28)$$

where

$$E(d,L) \triangleq \frac{1}{2}(d-L) \ln(\frac{d-L}{N}) + \frac{9}{2}(q_c^{max}-2)(d-\frac{1}{3}L) \quad (29)$$

We further upper bound the above as

$$A(\mathbf{d}) \leq e^{E(d,L)} \leq e^{\max_L E(d,L)} \quad (30)$$

The slope of the function $E(d,L)$ with respect to L is $-\frac{1}{2} \ln \frac{d-L}{N} - \frac{3}{2}q_c^{max} + \frac{5}{2}$. This slope is positive for $d \leq d_o$ (d_o to be defined shortly) and $0 \leq L \leq \frac{\gamma}{1+\gamma}d$. The weight vector enumerator can thus be upper bounded as

$$A(\mathbf{d}) \leq e^{E(d, \frac{\gamma}{1+\gamma}d)} \quad (31)$$

The weight enumerator now can be computed as

$$A_d = \sum_{\{\mathbf{d}\}} A(\mathbf{d}) \leq |\{\mathbf{d}\}| e^{E(d, \frac{\gamma}{1+\gamma}d)} \quad (32)$$

We can show that $|\{\mathbf{d}\}| = \binom{d+n_v-1}{n_v-1} = \binom{d+n_v-1}{d}$. We have $\binom{d+n_v-1}{d} \leq (\frac{e(d+n_v-1)}{d})^d = e^{d+d \ln(1+\frac{n_v-1}{d})} \leq e^{d+n_v-1}$. We used the fundamental inequality $\ln(x) \leq x-1$. Now the upper bound on A_d can be written as

$$A_d \leq e^{n_v-1} e^{d+E(d, \frac{\gamma}{1+\gamma}d)} \quad (33)$$

The exponent in the upper bound $d + E(d, \frac{\gamma}{1+\gamma}d) = \alpha(d \ln \frac{d}{N} + \beta d)$, where $\alpha = \frac{1}{2(1+\gamma)}$, and $\beta = 3(q_c^{max}-2)(3+2\gamma) + 2(1+\gamma) - \ln(1+\gamma)$. For the protograph LDPC codes with degree-3 variable nodes and higher, we should set $\gamma = 0$. Define $\tilde{\delta} = d/N$, and $F(\tilde{\delta}) = \tilde{\delta} \ln \frac{\tilde{\delta}}{\tilde{\delta}_o}$ where $\tilde{\delta}_o = e^{-\beta}$. Also define $d_o = \lfloor N\tilde{\delta}_o \rfloor$. Then

$$A_d \leq e^{n_v-1} e^{\alpha N F(\tilde{\delta})} \quad (34)$$

We note that for $0 \leq \tilde{\delta} \leq \tilde{\delta}_o$ the function $F(\tilde{\delta})$ is convex, it is negative, and has a minimum at $\tilde{\delta} = \tilde{\delta}_o/e$. The minimum value at this point is $F(\tilde{\delta}_o/e) = -\tilde{\delta}_o/e$. For $d = 2$, $\tilde{\delta} = 2/N$ with $F(2/N) = -2 \ln(N\tilde{\delta}_o/2)/N$. We assume $N\tilde{\delta}_o \gg 1$.

For $\tilde{\delta} \leq \tilde{\delta}_o$, or equivalently $d \leq d_o$, we have $d_i \leq u \leq d \leq N\tilde{\delta}_o = Ne^{-\beta}$. However, $\beta > 3(q_c^{max}-2)$. Thus the condition $d_i < Ne^{-3(q_c^{max}-2)}$ is satisfied as it was required to show. It is also easy to show that when $d \leq d_o$ the slope of $E(d,L)$ with respect to L is positive.

Now we upper bound $F(\tilde{\delta})$ with two lines, namely, one that connects the point $(\frac{2}{N}, F(\frac{2}{N}))$ to the point $(\frac{\tilde{\delta}_o}{e}, F(\frac{\tilde{\delta}_o}{e}))$. Denote this line by

$$L_1(\tilde{\delta}) = \alpha_1 \tilde{\delta} + \beta_1$$

where

$$\alpha_1 = \frac{-\frac{2}{N} \ln \frac{N\tilde{\delta}_o}{2} + \tilde{\delta}_o/e}{\frac{2}{N} - \tilde{\delta}_o/e} = -1 + \frac{\ln \frac{N\tilde{\delta}_o}{2e}}{\frac{N\tilde{\delta}_o}{2e} - 1}$$

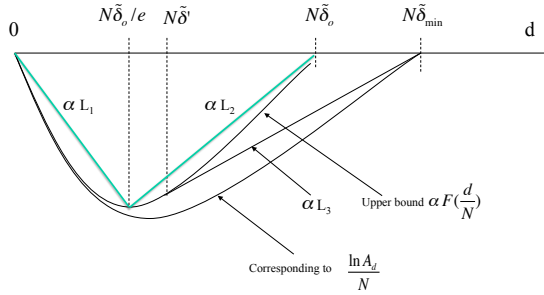


Fig. 4. upper bounding by lines

and

$$\beta_1 = -\frac{2\alpha_1}{N} - \frac{2}{N} \ln \frac{N\tilde{\delta}_o}{2}$$

Then $F(\tilde{\delta}) \leq L_1(\tilde{\delta})$ for $2/N \leq \tilde{\delta} \leq \tilde{\delta}_o/e$. The other line connects the point $(\frac{\tilde{\delta}_o}{e}, F(\frac{\tilde{\delta}_o}{e}))$ to the point $(\tilde{\delta}_o, F(\tilde{\delta}_o))$. Denote this line by

$$L_2(\tilde{\delta}) = \alpha_2\tilde{\delta} + \beta_2$$

where

$$\alpha_2 = \frac{1}{e-1}$$

and

$$\beta_2 = -\frac{\tilde{\delta}_o}{e-1}$$

Thus, $F(\tilde{\delta}) \leq L_2(\tilde{\delta})$ for $\tilde{\delta}_o/e \leq \tilde{\delta} \leq \tilde{\delta}_o$. Then

$$\begin{aligned} \sum_{d=2}^{\lfloor N\tilde{\delta}_o/e \rfloor - 1} A_d &\leq e^{n_v-1} \frac{e^{\alpha(2\alpha_1+N\beta_1)} - e^{\alpha(\alpha_1 N\tilde{\delta}_o/e + N\beta_1)}}{1 - e^{\alpha\alpha_1}} \\ &\leq e^{n_v-1} \frac{e^{\alpha(2\alpha_1+N\beta_1)}}{1 - e^{\alpha\alpha_1}} \leq e^{n_v-1} \frac{N^{-2\alpha}}{(\frac{\tilde{\delta}_o}{2})^{2\alpha}(1 - e^{\alpha\alpha_1})} \end{aligned} \quad (35)$$

Now for any $\tilde{\delta}_o/e < \tilde{\delta}^* < \tilde{\delta}_o$, we have

$$\begin{aligned} \sum_{d=\lfloor N\tilde{\delta}_o/e \rfloor}^{\lfloor N\tilde{\delta}^* \rfloor} A_d &\leq \sum_{d=\lfloor N\tilde{\delta}_o/e \rfloor}^{\lfloor N\tilde{\delta}^* \rfloor} e^{n_v-1} e^{\alpha \frac{d - N\tilde{\delta}_o}{e-1}} \\ &= e^{n_v-1} \frac{e^{\frac{\alpha}{e-1}} e^{-\frac{\alpha N(\tilde{\delta}_o - \tilde{\delta}^*)}{e-1}} - e^{-\frac{\alpha N\tilde{\delta}_o}{e}}}{e^{\frac{\alpha}{e-1}} - 1} \end{aligned} \quad (36)$$

Choose $\tilde{\delta}^* = \tilde{\delta}_o - (e-1)\frac{2}{N} \ln(\frac{N\tilde{\delta}_o}{2})$. Then

$$\begin{aligned} \sum_{d=\lfloor N\tilde{\delta}_o/e \rfloor}^{\lfloor N\tilde{\delta}^* \rfloor} A_d &\leq e^{n_v-1} \frac{e^{\frac{\alpha}{e-1}} e^{-\frac{\alpha N(\tilde{\delta}_o - \tilde{\delta}^*)}{e-1}}}{e^{\frac{\alpha}{e-1}} - 1} \\ &\leq e^{n_v-1} \frac{e^{\frac{\alpha}{e-1}}}{e^{\frac{\alpha}{e-1}} - 1} \frac{N^{-2\alpha}}{(\frac{\tilde{\delta}_o}{2})^{2\alpha}} \end{aligned} \quad (37)$$

Therefore the total sum

$$\sum_{d=2}^{\lfloor N\tilde{\delta}^* \rfloor} A_d \leq \frac{e^{n_v-1}}{(\frac{\tilde{\delta}_o}{2})^{2\alpha}} \left(\frac{1}{1 - e^{\alpha\alpha_1}} + \frac{e^{\frac{\alpha}{e-1}}}{e^{\frac{\alpha}{e-1}} - 1} \right) N^{-2\alpha}$$

This upper bound goes to zero as $N \rightarrow \infty$. Using Markov's inequality we can show $Pr\{d_{min} < \lfloor N\tilde{\delta}^* \rfloor\} \rightarrow 0$ as $N \rightarrow \infty$. Here we have shown existence of a typical minimum distance. However, due to the upper bounding technique, $\tilde{\delta}_o$ might be smaller than the typical minimum distance $\tilde{\delta}_{min}$ that can be obtained through numerical calculation. Note that $\tilde{\delta}_{min} > 0$ is the zero crossing of $\limsup \frac{\ln A_d}{N}$ as N becomes very large. If $\frac{\ln A_d}{N}$ is in fact a negative and convex function for $0 \leq d \leq N\tilde{\delta}_{min}$, then by upper bounding $\frac{\ln A_d}{N}$ by the line αL_3 as shown in Fig. 4, we can show $\sum_{d=\lfloor N\tilde{\delta}' \rfloor}^{\lfloor N\tilde{\delta}_{min} \rfloor - \epsilon} A_d$ also goes to zero as $N \rightarrow \infty$, where $\tilde{\delta}' \leq \tilde{\delta}_o$ corresponds to the tangent point. This says that not only a typical minimum distance exists, but also $\tilde{\delta}_{min}$ is the accurate value for such typical minimum distance. Note that the non-normalized $\tilde{\delta}_{min} = \tilde{\delta}_{min}/n_v$.

Sometimes we take the liberty of saying that if $\frac{\ln A_d}{N}$ as $N \rightarrow \infty$ is negative for $0 \leq \tilde{\delta} \leq \tilde{\delta}_{min}$, then $\tilde{\delta}_{min}$ is a typical minimum distance. Having a negative function, say $G(\tilde{\delta})$, there is no guaranty that $\sum_{d=2}^{N\tilde{\delta}^*} e^{NG(\tilde{\delta})}$ goes to zero as $N \rightarrow \infty$. For example, take the function $G(\tilde{\delta}) = \tilde{\delta}(\tilde{\delta} - \tilde{\delta}^*)$. It is negative between $0 \leq \tilde{\delta} \leq \tilde{\delta}^*$. It is convex and has a minimum at $\tilde{\delta} = \tilde{\delta}^*/2$, but the sum will not go to zero as $N \rightarrow \infty$.

VI. CONCLUSION

In this paper we proved that, for a certain class of protograph-based LDPC codes with degree-2 variable nodes, a typical minimum distance exists.

ACKNOWLEDGMENT

This research was supported in part by grant NNX09AL75G from NASA Goddard Space Flight Center. This research in part was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with NASA.

REFERENCES

- [1] R. G. Gallager, *Low-density parity-check codes*. Cambridge, MA: MIT Press, 1963.
- [2] S. Litsyn and V. Shevelev, "Distance distributions in ensembles of irregular low-density parity-check codes," *IEEE Trans. on Inform. Theory*, vol. 49, pp. 3140–3159, December 2003.
- [3] D. Burshtein and G. Miller, "Asymptotic enumeration methods for analyzing LDPC codes," *IEEE Trans. on Inform. Theory*, vol. 50, pp. 1115–1131, June 2004.
- [4] D. Divsalar, "Ensemble weight enumerators for protograph LDPC codes," *IEEE Int. Symp. on Inform. Theory*, pp. 1554–1558, July 2006.
- [5] J. Thorpe, "Low-density parity-check (LDPC) codes constructed from protographs," Tech. Rep. 42-154, IPN Progress Report, August 2003.
- [6] S. Benedetto, D. Divsalar, G. Montorsi, and F. Pollara, "Serial concatenation of interleaved codes: Performance analysis, design, and iterative decoding," *IEEE Trans. on Inform. Theory*, vol. 44, pp. 909–926, May 1998.

- [7] T. Richardson, "Multi-Edge Type LDPC Codes," presented at the Workshop honoring Prof. Bob McEliece on his 60th birthday, California Institute of Technology, Pasadena, California, May 24-25, 2002.
- [8] T. Richardson and R. Urbanke, "The Renaissance of Gallager's Low-Density Parity-Check Codes," *IEEE Communications Magazine*, pages 126-131, August 2003.
- [9] S.L. Fogal, Robert McEliece, Jeremy Thorpe "Enumerators for Protograph Ensembles of LDPC Codes," ISIT 2005.
- [10] Ikegaya, R.; Kasai, K.; Shibuya, T.; Sakaniwa, K.; Asymptotic weight and stopping set distributions for detailedly represented irregular LDPC code ensembles, ISIT 2004. *Proceedings, International Symposium on Information Theory*, 2004. 27 June-2 July 2004 Page(s):208
- [11] Tillich, J.-P.; The average weight distribution of Tanner code ensembles and a way to modify them to improve their weight distribution, ISIT 2004. *Proceedings of International Symposium on Information Theory*, 2004. 27 June-2 July 2004 Page(s):7
- [12] Changyan Di; Urbanke, R.; Richardson, T.; Weight distributions: how deviant can you be? *Proceedings, 2001 IEEE International Symposium on Information Theory*, 2001. 24-29 June 2001 Page(s):50
- [13] Di, C.; Montanari, A.; Urbanke, R.; Weight distributions of LDPC code ensembles: combinatorics meets statistical physics, ISIT 2004. *Proceedings of International Symposium on Information Theory*, 2004. 27 June-2 July 2004 Page(s):102
- [14] Burshtein, D.; Miller, G.; Asymptotic enumeration methods for analyzing LDPC codes, *IEEE Transactions on Information Theory*, Volume 50, Issue 6, June 2004 Page(s):1115 - 1131
- [15] Litsyn, S.; Shevelev, V.; Distance distributions in ensembles of irregular low-density parity-check codes, *IEEE Transactions on Information Theory*, Volume 49, Issue 12, Dec. 2003 Page(s):3140 - 3159
- [16] Litsyn, S.; Shevelev, V.; On ensembles of low-density parity-check codes: asymptotic distance distributions, *IEEE Transactions on Information Theory*, Volume 48, Issue 4, April 2002 Page(s):887 - 908
- [17] A. Orlitsky, K. Viswanathan, and J. Zhang, Stopping set distribution of LDPC code ensembles, *IEEE Trans. Information Theory*, vol. 51, pp. 929953, Mar. 2005.
- [18] I. Sason, E. Telatar and R. Urbanke, "Asymptotic input-output weight distributions and thresholds of convolutional and turbo-like codes," *IEEE Trans. on Information Theory*, vol. 48, no. 12, pp. 3052 -3061, December 2002.
- [19] D. Divsalar, S. Dolinar, C. Jones, K. Andrews, Capacity-Approaching Protograph Codes, *IEEE J. Selected Areas in Communications*, vol. 27, Aug. 2009, pp. 876-888.
- [20] R. Smarandache, P. Vontobel, Quasi Cyclic LDPC Codes: Influence of Proto- and Tanner-Graph Structure on Minimum Hamming Distance Upper Bounds, submitted to *IEEE Trans. Inf. Theory*, and available on arxiv.org as of Jan 26, 2009.
- [21] Brian K. Butler and Paul H. Siegel, "On Distance Properties of Quasi-Cyclic Protograph-Based LDPC Codes," submitted to ISIT 2010.
- [22] C. Koller, J. Kliewer, K. S. Zigangirov, D. J. Costello, Jr., "Minimum Distance Bounds for Multiple-Serially Concatenated Code Ensembles," ISIT 2008.
- [23] H. Pishro-Nik, and F. Fekri, "Performance of Low-Density Parity-Check Codes With," *IEEE Trans. on IT* January 2006.
- [24] A. Otmani, J-P Tillich, I. Andriyanova, "On the Minimum Distance of Generalized LDPC Codes," ISIT 2007.
- [25] J-P Tillich, G. Zmor, "On the minimum distance of structured LDPC codes with two variable nodes of degree 2 per parity-check equation," ISIT 2006.
- [26] W. E. Ryan and S. Lin, *Channel Codes: Classical and Modern*, Cambridge University Press, 2009.