Minimum distance and pseudodistance lower bounds for generalized LDPC codes

Christine A. Kelley*

Department of Mathematics
University of Nebraska-Lincoln
Lincoln, NE 68588, USA
Email: ckelley2@math.unl.edu
*Corresponding author

Abstract: Two different ways of obtaining generalized low-density parity-check codes are considered. Lower bounds on the minimum distance, stopping distance, and pseudodistance are derived for these codes using graph based analysis. These bounds are generalizations of Tanner's bit-oriented and parity-oriented bound for simple LDPC codes. The new bounds are useful in predicting the performance of generalized LDPC codes under maximum-likelihood decoding, graph-based iterative decoding, and linear programming decoding, and rely on the connectivity of the Tanner graph.

Keywords: LDPC codes, generalized-LDPC codes, Tanner graph, bit-oriented bound, parity-oriented bound, constraint-oriented bound, iterative decoding, code graph, eigenvalues, minimum distance, stopping set, pseudoweight.

Reference to this paper should be made as follows: Kelley, C. A., 'Minimum distance and pseudodistance lower bounds for generalized LDPC codes', *Int. J. Inform. and Coding Theory*, Vol. x, No. x, pp.xxx–xxx.

Biographical notes: Christine A. Kelley is an assistant professor in the Department of Mathematics at the University of Nebraska-Lincoln. She received her Ph.D. in Mathematics in 2006 from the University of Notre Dame. Prior to her current position, she was a postdoctoral fellow at the Fields Institute in Toronto as well as a VIGRE Arnold Ross assistant professor at the Ohio State University. Her research interests include coding theory, information theory, and applied discrete mathematics.

1 Introduction

Low-density parity-check (LDPC) codes are a class of error-correcting codes that can be represented on sparse graphs and have been shown to achieve record-breaking performances with graph-based message-passing decoders (Tanner (1981);

Richardson et. al (2001)). The popularity of LDPC codes is that they can be decoded with linear time complexity using graph-based message-passing decoders, thereby allowing for the use of large block length codes in several practical applications. In contrast, maximum-likelihood (ML) decoding a generic error-correcting code is known to be NP hard.

A parameter that dominates the performance of a graph-based message passing decoder is the minimum pseudocodeword weight, in contrast to the minimum distance for an optimal (or, ML) decoder. The minimum pseudocodeword weight (or, pseudodistance) (Forney et. al (2001)) of the graph has been found to be reasonable predictor of the performance of a finite-length LDPC code under graph-based message-passing decoding and also linear programming decoding (Koetter and Vontobel (2003); Kelley and Sridhara (2007); Feldman et. al (2005)).

In this paper, we consider two different ways of obtaining generalized LDPC codes from graphs. For each case, we first derive lower bounds on their minimum distance based on the connection properties of the underlying LDPC graph. We then extend the results to lower bound the stopping distance (Di et. al (2002)), which is essentially the pseudodistance on the binary erasure channel (BEC), and finally, we lower bound the pseudodistance on the additive white Gaussian noise (AWGN) channel.

The bounds in this paper are extensions of Tanner's bit-oriented and parity-oriented bounds on the minimum distance of simple LDPC codes (Tanner (2001)). The main difference is that we consider generalized LDPC (GLDPC) codes wherein the constraints imposed by the constraint nodes in the LDPC graph (or, Tanner graph) are not simple parity-check constraints but rather constraints of a sub-code of appropriate length and rate. Generalized LDPC codes were originally introduced in (Tanner (1981)). The resulting lower bounds on the minimum distance and pseudodistance indicate that the GLDPC codes may have stronger error-correction capabilities than the corresponding simple LDPC codes represented by the same Tanner graphs.

As a final motivation, generalized LDPC codes are of interest since they form the basis for most constructions of expander codes which use Tanner graphs with good expansion properties (see for example, (Sipser and Spielman (1996))). Graphs with good expansion yield codes with good minimum distances, and are particularly suited for the message-passing decoder in dispersing messages to all nodes in the graph quickly, thus improving the decoder performance. Bounds on the minimum distance and pseudodistance of GLDPC codes were also derived in (Kelley and Sridhara (2007)) where we exploited the expansion properties of the graph. In contrast, the bounds in this paper are derived using the connectivity of the graph. Finally, it is worth noting that an extension of Tanner's bit and parity-oriented bounds for block-wise (simple) irregular LDPC codes is given in (Shin et. al (2005)).

The paper is organized as follows. In Section 2, the notation and background on LDPC codes is presented. Section 3 presents the bit-oriented and constraint-oriented bounds on the minimum distance and pseudodistance for the most common class of generalized LDPC codes. The expander codes construction by Sipser and Spielman is a special case of this class and the lower bounds derived here are compared with those derived for their construction (Sipser and Spielman (1996)). In Section 4, analogous lower bounds are derived for a second class of generalized LDPC codes that were made popular by the construction by Janwa and Lal in (Janwa and Lal (2003)). The results are summarized along with conclusions in Section 5.

2 Preliminaries

In this section we review some basic terminology and notation regarding LDPC codes, stopping sets, and pseudocodewords.

Definition 2.1: A graph $G = (X, Y; E)$ is (c, d)*-regular bipartite* if the set $X \cup Y$ of vertices in G can be partitioned into two disjoint sets X and Y such that all vertices in X have degree c and all vertices in Y have degree d and each edge $e \in E$ of G is incident with one vertex in X and one vertex in Y, i.e., $\forall e \in E$, $e = (x, y)$ with $x \in X, y \in Y$.

We will refer to the vertices of degree c as the *left* vertices, and to vertices of degree d as the *right* vertices.

Definition 2.2: A *simple* LDPC code is defined by a bipartite graph G (called a Tanner graph) whose left vertices are called *variable* nodes and whose right vertices are called *constraint* nodes. The set of codewords consists of all binary assignments to the variable nodes such that at each constraint node, the modulo two sum of the variable node assignments connected to the constraint node is zero, i.e., the parity-check constraint involving the neighboring variable nodes is satisfied.

Variable nodes for binary codes are often referred to as *bit* nodes and constraint nodes for simple LDPC codes are often referred to as *check* nodes. Definition 2.2 can be generalized by introducing more complex constraints instead of simple parity-check constraints at each constraint node, and the resulting LDPC code will be called a *generalized* LDPC (or, GLDPC) code. In this case, the set of codewords consists of all binary assignments to the variable nodes such that at each constraint node, the adjacent variable node assignments (according to some ordering) form a codeword of the subcode at that constraint node. Figure 3 in Section 3 gives an example of a GLDPC code. Figures 4 and 5 give alternative ways of obtaining GLDPC codes.

The incidence matrix H of the Tanner graph G for a simple LDPC code forms the parity-check matrix of the corresponding code. However, the incidence matrix H of the Tanner graph G for a GLDPC code is not always the parity-check matrix of the corresponding code; in particular, each row of H represents the set of variable nodes in G that must satisfy the constraints imposed by the subcode in a GLDPC code.

To analyze the performance of graph-based message passing decoding, certain combinatorial properties of the Tanner graph have been identified that control the performance of the decoder. When the channel is characterized by the binary erasure channel (BEC), it has been shown that stopping sets in the Tanner graph control the performance of the message-passing decoder.

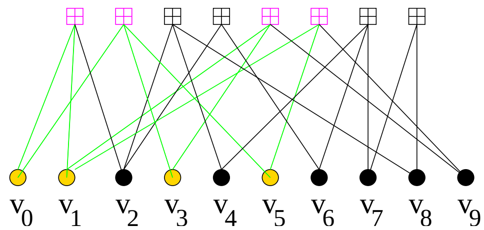

Figure 1 A stopping set $S = \{v_0, v_1, v_3, v_5\}$ in G.

Definition 2.3: (Di et. al (2002)) A *stopping set* is a subset S of the variable node set such that every constraint node that is a neighbor of some node $s \in S$ is connected to S at least twice.

Note that the above definition of a stopping set is for simple LDPC codes. The size of a stopping set S is equal to the number of elements in S. The smallest non-empty stopping set is called a *minimum* stopping set, and its size, the *stopping distance*, is denoted by $s_{\min}$.

Figure 1 shows an example of a stopping set in a graph. Observe that $\{v_4, v_7, v_8\}$ and $\{v_3, v_5, v_9\}$ are two minimum stopping sets of size $s_{\min} = 3$. On the binary erasure channel, if all of the nodes of a stopping set are erased, then the graph-based iterative decoder will not be able to recover the erased symbols associated with the nodes of the stopping set (Di et. al (2002)). Therefore, it is advantageous to design LDPC codes with large minimum stopping distance $s_{\min}$.

For other channels, it has been observed that *pseudocodewords* dominate the performance of the iterative decoder (Koetter and Vontobel (2003); Kelley and Sridhara (2007)). (In fact, pseudocodewords are a generalization of stopping sets for other channels.) Before we introduce the formal definition of graph-cover pseudocodewords of an LDPC Tanner graph G (Koetter and Vontobel (2003); Kelley and Sridhara (2007)), we first need to introduce the definition of a graph cover. A degree ℓ cover (or, lift) $\hat{G}$ of G is defined in the following manner:

Definition 2.4: A finite degree ℓ *cover* of $G = (V, W; E)$ is a bipartite graph $\hat{G}$ where for each vertex $x_i \in V \cup W$, there is a *cloud* $\hat{X}_i = \{\hat{x}_{i,1}, \hat{x}_{i,2}, \dots, \hat{x}_{i,\ell}\}$ of vertices in $\hat{G}$, with $deg(\hat{x}_{i,j}) = deg(x_i)$ for all $1 \le j \le \ell$, and for every $(x_i, x_j) \in E$, there are ℓ edges from $\hat{X}_i$ to $\hat{X}_j$ in $\hat{G}$ connected in a one-to-one manner.

Definition 2.5: Suppose that

$$\hat{\mathbf{c}} = (\hat{c}_{1,1}, \hat{c}_{1,2}, \dots, \hat{c}_{1,\ell}, \hat{c}_{2,1}, \dots, \hat{c}_{2,\ell}, \dots)$$

is a codeword in the Tanner graph $\hat{G}$ representing a degree ℓ cover of G. A *pseudocodeword* $\mathbf{p}$ *of* G is a vector $(p_1, p_2, \dots, p_n)$ obtained by reducing a codeword $\hat{\mathbf{c}}$, of the code in the graph cover $\hat{G}$, in the following way:

$$\hat{\mathbf{c}} = (\hat{c}_{1,1}, \dots, \hat{c}_{1,\ell}, \hat{c}_{2,1}, \dots, \hat{c}_{2,\ell}, \dots) \rightarrow \left(\frac{\hat{c}_{1,1}+\hat{c}_{1,2}+\cdots+\hat{c}_{1,\ell}}{\ell}, \frac{\hat{c}_{2,1}+\hat{c}_{2,2}+\cdots+\hat{c}_{2,\ell}}{\ell}, \dots\right) = (p_1, p_2, \dots, p_n) = \mathbf{p},$$

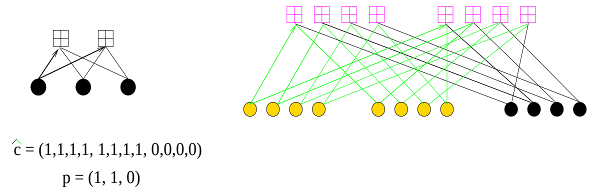

Figure 2 A pseudocodeword in the base graph (or a valid codeword in a cover).

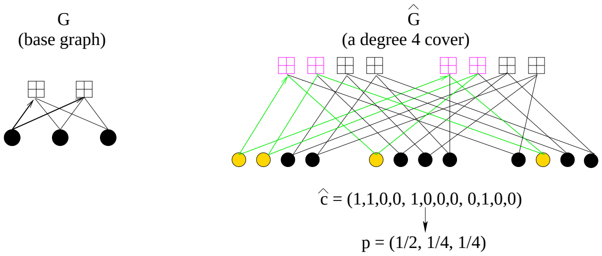

Figure 3 A pseudocodeword in the base graph (or a valid codeword in a cover).

where $p_i = \frac{\hat{c}_{i,1}+\hat{c}_{i,2}+\cdots+\hat{c}_{i,\ell}}{\ell}$.

Figures 2 and 3 show a base graph G and a degree four cover of G. In Figure 2, the vector $\hat{c}$ is a codeword in the degree four cover, that reduces to a pseudocodeword which is also a codeword in the base graph, whereas in Figure 3, the codeword $\hat{c}$ in the degree four cover reduces to a pseudocodeword that is not a codeword in the base graph.

The set of graph-cover pseudocodewords can also be described by means of a polytope, called the fundamental polytope (Koetter and Vontobel (2003); Feldman et. al (2005); Koetter et. al (2007)). In particular, graph-cover pseudocodewords are dense in the fundamental polytope.

From Definition 2.5, it is easy to show that for a simple LDPC Tanner graph G, a pseudocodeword $\mathbf{p} = (p_1, p_2, \ldots, p_n)$ is a vector that satisfies the following set of inhomogeneous inequalities:

$$0 \le p_i \le 1, \quad \text{for } i = 1, 2, \ldots, n \tag{1}$$

and, if variable nodes $i_1, i_2, \ldots, i_d$ participate in a check node of degree d, then the pseudocodeword components satisfy the following set of homogeneous inequalities

$$p_{i_j} \le \sum_{k=1,2,..d,k\neq j} p_{i_k}, \text{ for } j = 1, 2, .., d. \tag{2}$$

We note here that the above are only some of the (fundamental polytope facet-defining) inequalities satisfied by the pseudocodeword vector $\mathbf{p}$ (Koetter and Vontobel (2003); Feldman et. al (2005)).

Extending the above for GLDPC codes, it can similarly be shown that on a GLDPC Tanner graph G, a pseudocodeword $\mathbf{p} = (p_1, p_2, \ldots, p_n)$ is a vector that satisfies the following set of inequalities:

$$0 \le p_i \le 1, \quad \text{for } i = 1, 2, \ldots, n \tag{3}$$

and, if variable nodes i_1, i_2, $\ldots$, i_d participate in a constraint node of degree d and that constraint node represents a $[d, rd, \epsilon d]$ subcode (where an $[n, k, t]$ linear code is a code of blocklength n, dimension k and minimum distance t), then the pseudocodeword components satisfy

$$(d\epsilon - 1)p_{i_j} \leq \sum_{k=1,2,..d,k\neq j} p_{i_k}, \text{ for } j = 1, 2, .., d. \tag{4}$$

Note that the inequalities in equation (4) need not be facet-defining inequalities for the pseudocodeword polytope in (Koetter and Vontobel (2003); Feldman et. al (2005)).

Remark 2.6: Note that (4) implies that the pseudocodeword components of the GLDPC Tanner graph G also satisfy the following set of inequalities at the degree d constraint node representing a $[d, rd, \epsilon d]$ subcode

$$\sum_{\text{any } \lfloor \frac{d\epsilon}{2} \rfloor j\text{'s}} p_{i_j} \leq \sum_{\text{remaining terms}} p_{i_k}. \tag{5}$$

It was shown in (Koetter and Vontobel (2003); Kelley and Sridhara (2007)) that a stopping set in a simple LDPC Tanner graph is the support of a pseudocodeword as defined above. Thus, generalizing the definition of stopping sets to GLDPC codes, we have:

Definition 2.7: A *stopping set* in a GLDPC Tanner graph G is the support of a pseudocodeword $\mathbf{p}$ of G.

In Sections 3 and 4, we will consider pseudocodewords and their behavior on the binary input memoryless additive white Gaussian noise (AWGN) channel. Following the definition in (Forney et. al (2001); Wiberg (1996)), we have:

Definition 2.8: *The weight of a pseudocodeword* $\mathbf{p} = (p_1, p_2, \ldots, p_n)$ *of an LDPC Tanner graph* G *on the AWGN channel is defined as*

$$w^{AWGN}(\mathbf{p}) = \begin{cases} \frac{(\sum_{i=1}^{n} p_i)^2}{(\sum_{i=1}^{n} p_i^2)} & \mathbf{p} \neq \mathbf{0} \\ 0 & \mathbf{p} = \mathbf{0} \end{cases}$$

Definition 2.9: The *pseudodistance* (or, *minimum pseudocodeword weight*) of an LDPC Tanner graph G is the minimum weight among all non-zero pseudocodewords obtainable from all finite-degree covers of G. This parameter is denoted by $w_{\min}^{AWGN}$ for the AWGN channel.

Finally, it should be noted that while the minimum distance of a code $\mathcal{C}$ is invariant under different representations of $\mathcal{C}$, the stopping distance and pseudodistance depend on the graph representation of $\mathcal{C}$ (or equivalently, the parity-check matrix used to represent the code). Regardless of representation, however, the stopping distance and pseudodistance are always at most the minimum distance.

3 Case 1: Generalized LDPC codes

Let H be an $m \times n$ incidence matrix for a binary linear code $\mathcal{C}$ with a fixed column weight j and a fixed row weight d, and let the rows of H represent constraints of a $[d, rd, \epsilon d]$ subcode. That is, in the Tanner graph representation of H the variable node neighbors of each degree d constraint node must satisfy all the parity-check constraints of a $[d, rd, \epsilon d]$ subcode in every valid codeword configuration among the variable nodes. Let G be the corresponding Tanner graph representing H. Then G has two sets of vertices: n variable nodes and m constraint nodes and the edges in G correspond to the non-zero entries in H.

The adjacency matrix of G is an $(m+n) \times (m+n)$ matrix A having the form

$$A = \begin{bmatrix} \mathbf{0} & H \\ H^T & \mathbf{0} \end{bmatrix}$$

As A is a real symmetric matrix, it has real eigenvalues. We will assume that G is connected and therefore A has a unique largest eigenvalue λ_1. Let $\lambda_1 > \lambda_2 > \cdots > \lambda_t$ be the ordered distinct eigenvalues of A.

Furthermore, let $\mu_1 > \mu_2 > \cdots > \mu_s$ be the ordered distinct eigenvalues of $H^T H$, where H is the code's incidence matrix interpreted as a real-valued matrix with zero and one entries. Observe that $H^T H$ is an $n \times n$ matrix and has n eigenvalues in total. In particular, $H^T H$ represents the connectivity among the variable nodes via their constraint node neighbors in G. It was shown in (Tanner (2001)) that the eigenvalues of A are $\pm\sqrt{\mu_i}$ and possibly 0, for $1 \leq i \leq s$. The eigenvalues of $H^T H$ are real-valued and symmetric.

3.1 Bit-oriented bound on distance

Theorem 3.1: *Let G be a regular connected graph with n bit nodes of uniform degree c and m constraint nodes of uniform degree d, and let each constraint node represent constraints from a $[d, rd, \epsilon d]$ subcode. Then the minimum distance of the GLDPC code $\mathcal{C}$ with incidence matrix H represented by the Tanner graph G satisfies*

$$d_{\min} \geq \frac{n(cd\epsilon - \mu_2)}{cd - \mu_2},$$

where μ_2 is the second largest eigenvalue of $H^T H$.

Figure 4 shows the Tanner graph of the generalized LDPC code described in Theorem 3.1. The left set of vertices represent variable nodes of degree c and the right set of vertices represent constraint nodes of degree d and they represent constraints from a $[d, rd, \epsilon d]$ sub-code. The variable node neighbors of every constraint node satisfy all the parity-check constraints of the $[d, rd, \epsilon d]$ subcode in every valid configuration.

Proof: Let $\mathbf{x}$ be a real-valued vector of length n corresponding to the support of a minimum-weight codeword ($d_{\min}$) of $\mathcal{C}$, i.e. $\mathbf{x}$ has a one in every position where

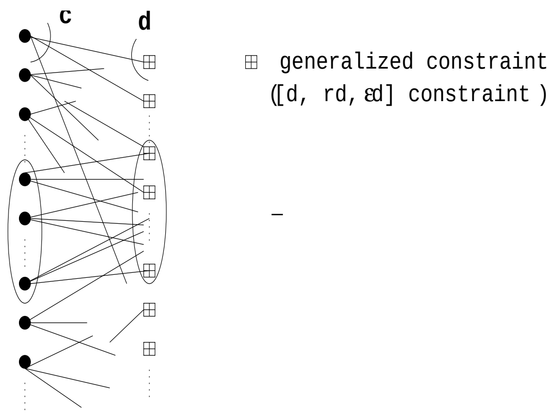

Figure 4 Generalized LDPC code with $[d, rd, \epsilon d]$ subcode constraints on the constraint nodes.

the minimum-weight codeword is non-zero and zeros elsewhere. Since $H^T H$ is a regular matrix with each row and column sum equal to cd, the first eigenvector of $H^T H$ can be taken to be $\mathbf{e}_1 = (1, 1, \ldots, 1)^T/\sqrt{n}$ and the corresponding eigenvalue is $\mu_1 = cd$. Since G is connected, $\mathbf{e}_1$ is the unique eigenvector with eigenvalue cd. Let $\mathbf{x}_i$ denote the projection of $\mathbf{x}$ onto the ith eigenspace. Then we have,

$$\| \mathbf{x} \|^2 = d_{\min}$$

$$\| \mathbf{x}_1 \|^2 = d_{\min}^2/n.$$

$H\mathbf{x} = (w_1, \ldots, w_m)^T$ is an integer-valued vector of length m where for each $i = 1, \ldots, m$, the weight w_i represents the number of active variable nodes adjacent to the ith constraint node in G. Each nonzero w_i must have value at least ϵd since each constraint represents a subcode with minimum distance ϵd and $\mathbf{x}$ is a valid codeword. (Here, "active" refers to the variable nodes that assume a non-zero value in the corresponding vector $\mathbf{x}$.) Thus,

$$\| H\mathbf{x} \|^2 = \sum_{i=1}^{m} w_i^2 \geq \epsilon d \sum_{i=1}^{m} w_i = \epsilon d(cd_{\min}). \quad (*)$$

The last term in the above inequality is obtained by observing that $d_{\min}$ of the components of $\mathbf{x}$ are one, and each is adjacent to c constraint nodes.

Converting the above expression to eigenspace representation, we have

$$\| H\mathbf{x} \|^2 = \sum_{i=1}^{s} \mu_i \| \mathbf{x}_i \|^2 = \mu_1(d_{\min}^2)/n + \sum_{i=2}^{s} \mu_i \| \mathbf{x}_i \|^2$$

$$\leq \mu_1(d_{\min}^2)/n + \mu_2 \sum_{i=2}^{s} \| \mathbf{x}_i \|^2$$

$= \mu_1(d_{\min}^2)/n + \mu_2(\| \mathbf{x} \|^2 - \| \mathbf{x}_1 \|^2)$

Substituting $\| \mathbf{x}_1 \|^2 = d_{\min}^2/n$ and $\mu_1 = cd$, and combining with expression $(*)$, we obtain

$$cd(d_{\min}^2)/n + \mu_2(d_{\min} - d_{\min}^2/n) \geq \epsilon d(cd_{\min})$$

which gives the desired bound for $d_{\min}$. □

Remark 3.2: Note that the result in (Tanner (2001)) is obtained from Theorem 3.1 by setting $\epsilon = 2/d$. The result in Theorem 3.1 is useful when the second largest eigenvalue μ_2 is as small as possible, meaning the graph is a very good expander. While there are cases of GLDPC codes for which the lower bound could be negative and hence, not meaningful, graphs with good expansion and a good choice of subcodes (in terms of their rate r and relative distance ϵ) will yield a useful lower bound on the minimum distance of the GLDPC code.

3.2 Constraint-oriented bound on distance

Let $\mu_1 > \mu_2 > \cdots > \mu_s$ be the ordered distinct eigenvalues of HH^T, where H is the code's incidence matrix interpreted as a real-valued matrix with zero and one entries. Observe that HH^T is an $m \times m$ matrix and has m eigenvalues in total. In particular, HH^T represents the connectivity among the constraint nodes via their variable node neighbors in G. It was shown in (Tanner (2001)) that $H^T H$ and HH^T in fact have the same non-zero eigenvalues. The following bound considers the connectivity from the constraint node perspective.

Theorem 3.3: *Let G be a regular connected graph with n bit nodes of uniform degree c and m constraint nodes of uniform degree d, and let each constraint node represent constraints from a $[d, rd, \epsilon d]$ subcode. Then the minimum distance of the code $\mathcal{C}$ with a incidence matrix H represented by the Tanner graph G satisfies*

$$d_{\min} \geq \frac{n\epsilon(\epsilon cd + d - \epsilon d - \mu_2)}{(cd - \mu_2)},$$

where μ_2 is the second largest eigenvalue of HH^T.

Proof: Let $\mathbf{y}$ be a real-valued vector of length m that has a one in every active constraint node position and zeros elsewhere. (The i^{th} entry in the vector $\mathbf{y}$ to correspond to the value of the i^{th} constraint node. Thus, the phrase "active constraint node" refers to the constraint nodes that assume a non-zero value in the corresponding vector $\mathbf{y}$.) Let t be the number of ones in $\mathbf{y}$. Since each row and column sum in HH^T is cd, the largest eigenvalue of HH^T is $\mu_1 = cd$ and has corresponding unique eigenvector $\mathbf{e}_1 = (1, 1, \ldots, 1)^T/\sqrt{m}$. Let $\mathbf{y}_i$ be the projection of $\mathbf{y}$ onto the ith eigenspace. Then, $\| \mathbf{y} \|^2 = t$ and $\| \mathbf{y}_1 \|^2 = \frac{t^2}{m}$. $H^T\mathbf{y} =$

$(w_1, \ldots, w_n)^T$ is an integer-valued vector of length n where for each $i = 1, \ldots, n$, the weight w_i represents the number of active constraints nodes adjacent to the ith variable node in G. Note that if the ith bit is nonzero, then $w_i = c$. Now,

$$\| H^T\mathbf{y} \|^2 = \sum_{i=1}^{n} w_i^2. \qquad (**)$$

To determine w_i^2, note that each active constraint node is adjacent to at least ϵd nonzero bit nodes in order for the constraint to be satisfied. Let $u_j(\ell)$ be the number of variable nodes with weight ℓ in $H^T\mathbf{y}$ that are adjacent to the jth active constraint node, for $1 \leq \ell \leq c$. Each active constraint node is adjacent to at least ϵd nonzero variable nodes, thus $u_j(c) \geq \epsilon d$.

The squared weight w_i^2 in $(**)$ is found by summing over the active constraint nodes. The contribution of the ith constraint node to the sum in $(**)$ is $\sum_{\ell=1}^{c} u_j(\ell)(\ell^2)(1/\ell)$. This may be seen by observing that for each of the $u_j(\ell)$ adjacent variable nodes of weight ℓ, the contribution ℓ^2 is counted ℓ times at the ℓ adjacent active constraint nodes. Thus, in summing over all active constraint nodes, we undo the overcounting by $(1/\ell)$.

Rewriting this expression and incorporating the observations we have for each active constraint node, we obtain

$$\sum_{\ell=1}^{c} u_j(\ell)(\ell^2)(1/\ell) = u_j(c)c + \sum_{\ell=1}^{c-1} u_j(\ell)\ell.$$

$$\geq \epsilon dc + d - \epsilon d.$$

Note that the inequality follows since the second summation is lower-bounded by $d - \epsilon d$ since at least ϵd of the neighbors have weight c. Since there are t active constraint nodes in $\mathbf{y}$,

$$\sum_{i=1}^{n} w_i^2 \geq t(\epsilon dc + d - \epsilon d).$$

Converting to eigenspace representation,

$$\| H^T\mathbf{y} \|^2 = \mu_1 t^2/m + \sum_{i=2}^{s} \mu_i \| \mathbf{y}_i \|^2$$

$$\leq \mu_1 t^2/m + \mu_2 \sum_{i=2}^{s} \| \mathbf{y}_i \|^2$$

$$= \mu_1 t^2/m + \mu_2(\| \mathbf{y} \|^2 - \| \mathbf{y}_1 \|^2).$$

Substituting from above,

$$\mu_1 t^2/m + \mu_2(t - t^2/m) \geq t(\epsilon dc + d - \epsilon d)$$

giving

$$t \geq m(\epsilon cd + d - \epsilon d - \mu_2)/(\mu_1 - \mu_2).$$

Observe that the graph G has t active constraint nodes and at least $d_{\min}$ active variable nodes. Moreover, $d_{\min} c \geq \epsilon dt$ since t active constraint nodes must have at least ϵd active edges going to active variable nodes of a valid configuration in the graph, and each active variable node has c active edges incident to it. Combining the above expressions and observing $nc = md$, the bound follows. □

3.3 Minimum stopping set size

A generalized stopping set is as defined in Definition 2.7 in Section 2. Under the assumption that the $[d, rd, \epsilon d]$ subcode has no idle components, meaning that there are no components that are zero in all of the codewords of the subcode, Definition 2.7 reduces to

Lemma 3.4: A stopping set in a GLDPC code is a set of variable nodes such that every node that is a neighbor of some node $s \in S$ is connected to S at least ϵd times.

Note that the condition in Lemma 3.4 is a necessary but not always sufficient condition for S to be a stopping set. Following the same proof techniques for the bit and constraint-oriented bounds on the minimum distance and using the above definition on the stopping set size, the following result can be easily shown:

Corollary 3.5: *Suppose G is a regular connected graph with n bit nodes of uniform degree c and m constraint nodes of uniform degree d, and suppose each constraint node represent constraints from a $[d, rd, \epsilon d]$ sub-code. Then the stopping distance, $s_{\min}$, of G as a Tanner graph representing the parity check matrix H is given by*

$$s_{\min} \geq \max\left\{\frac{n(\epsilon cd - \mu_2)}{cd - \mu_2}, \quad \frac{n\epsilon(\epsilon cd + d - \epsilon d - \mu_2)}{(cd - \mu_2)}\right\},$$

where μ_2 is the second largest eigenvalue (in absolute value) of $H^T H$ (and also HH^T).

3.4 Minimum pseudoweight

For simple LDPC codes, a bit-oriented bound on the pseudodistance is derived in (Vontobel and Koetter (2004)), and a parity-oriented bound on the pseudodistance is derived in (Kelley and Sridhara (2007)). In this subsection, we provide a bound on the (AWGN) pseudodistance for generalized LDPC codes.

Theorem 3.6: *Let G be a regular connected graph with n bit nodes of uniform degree c and m constraint nodes of uniform degree d, and let each constraint node represent constraints from a $[d, rd, \epsilon d]$ subcode. Then the pseudodistance on the AWGN channel of the GLDPC code $\mathcal{C}$ with incidence matrix H represented by the Tanner graph G satisfies*

$$w_{\min}^{AWGN} \geq \frac{n(cd\epsilon - \mu_2)}{cd - \mu_2},$$

where μ_2 is the second largest eigenvalue of $H^T H$ in absolute value.

Proof: The proof here is very similar to the proof of the bit-oriented bound on the minimum distance. Let $\mathbf{x}$ be a real-valued vector of length n corresponding to a pseudocodeword as defined in Section 2. Then the pseudocodeword weight of $\mathbf{x}$ on the AWGN channel is given by

$$w^{AWGN}(\mathbf{x}) = \frac{(x_1 + x_2 + \cdots + x_n)^2}{(x_1^2 + x_2^2 + \cdots + x_n^2)} = \frac{\| \mathbf{x} \|_1^2}{\| \mathbf{x} \|_2^2}, \quad (\dagger)$$

where $\| \mathbf{x} \|_1$ and $\| \mathbf{x} \|_2$ are the first and second norm of $\mathbf{x}$, respectively. Since $H^T H$ is a regular matrix with each row and column sum equal to cd, the first eigenvector of $H^T H$ can be taken to be $\mathbf{e}_1 = (1, 1, \ldots, 1)^T / \sqrt{n}$ and the corresponding eigenvalue is $\mu_1 = cd$. Since G is connected, $\mathbf{e}_1$ is the unique eigenvector with eigenvalue cd. Let $\mathbf{x}_i$ denote the projection of $\mathbf{x}$ onto the ith eigenspace. Then

$$\| \mathbf{x}_1 \|^2 = \frac{1}{n} \| \mathbf{x} \|_1^2 .$$

$H\mathbf{x} = \mathbf{w}^T = (w_1, \ldots, w_m)^T$ is a real-valued vector of length m where for each $i = 1, \ldots, m$, the weight w_i represents the sum of the active pseudocodeword components at the variable nodes adjacent to the ith constraint node in G. We may lower bound $\| \mathbf{w} \|_2^2$ as follows:

$$\| \mathbf{w} \|_2^2 = \sum_{i=1}^{m} w_i^2 = \sum_{i=1}^{m} \Big(\sum_{j:H_{i,j}=1} x_j \Big)^2 \geq \epsilon d \sum_{i=1}^{m} \sum_{j:H_{i,j}=1} x_j^2$$

$$= \epsilon d \sum_{j} \sum_{i:H_{i,j}=1} x_j^2 = c\epsilon d \cdot \| \mathbf{x} \|_2^2 \qquad (\dagger\dagger)$$

Observe that the first inequality in the above follows from the inequalities that the pseudocodeword must satisfy in equation (4). The last equality follows by observing that in the sum $\sum_{i=1}^{m}\sum_{j:H_{j,i}} x_j^2$, each term x_j^2 is counted c times, since each variable node has degree c.

Now, to upper bound $\| H\mathbf{x} \|^2$, we use the eigenspace representation to obtain

$$\| H\mathbf{x} \|^2 = \sum_{i=1}^{s} \mu_i \| \mathbf{x}_i \|^2 = \mu_1 \| \mathbf{x} \|_1^2 /n + \sum_{i=2}^{s} \mu_i \| \mathbf{x}_i \|^2$$

$$\leq \mu_1 \| \mathbf{x} \|_1^2 /n + \mu_2 \sum_{i=2}^{s} \| \mathbf{x}_i \|^2$$

$$= \mu_1 \| \mathbf{x} \|_1^2 /n + \mu_2 (\| \mathbf{x} \|_2^2 - \| \mathbf{x}_1 \|^2)$$

$$= \mu_1 \| \mathbf{x} \|_1^2 /n + \mu_2 (\| \mathbf{x} \|_2^2 - \| \mathbf{x} \|_1^2 /n)$$

Combining this expression with (††), we obtain

$$\mu_1 \| \mathbf{x} \|_1^2 /n + \mu_2 (\| \mathbf{x} \|_2^2 - \| \mathbf{x} \|_1^2 /n) \geq c\epsilon d \| \mathbf{x} \|_2^2$$

Thus, $\| \mathbf{x} \|_1^2 (\mu_1 - \mu_2)/n \geq (cd\epsilon - \mu_2) \| \mathbf{x} \|_2^2$.

Substituting $\mu_1 = cd$ and using the expression for the weight of the pseudocodeword $\mathbf{x}$ in (†), we see that the weight of any non-zero pseudocodeword $\mathbf{x}$ is lower bounded by $\frac{n(cd\epsilon - \mu_2)}{cd - \mu_2}$. This yields the desired bound for $w_{\min}^{AWGN}$. □

Remark 3.7: Note that for any code $\mathcal{C}$ and a corresponding parity check matrix representation H of this code, the following relation holds:

$$w_{\min}^{AWGN} \leq s_{\min} \leq d_{\min}.$$

Thus, the result in Theorem 3.6 immediately implies the result in Theorem 3.1 and the first part of Corollary 3.5. Also note that the result in (Vontobel and Koetter (2004)) is obtained from Theorem 3.6 by setting $\epsilon = 2/d$.

3.5 Special case of expander codes

The construction of expander codes by Sipser and Spielman in (Sipser and Spielman (1996)) is as follows. Consider a d-regular graph G with N vertices. It can be transformed to a $(2, d)$-regular bipartite graph G' by representing every edge in G by an *edge node* in G' and every vertex in G by a *vertex node* in G' and connecting the edge nodes to the vertex nodes in G' according to their incidence in G. (See Figure 5.) The graph G' is called the *edge-vertex incidence graph of* G. Thus, there are $Nd/2$ edge nodes each having degree two and there are N vertex

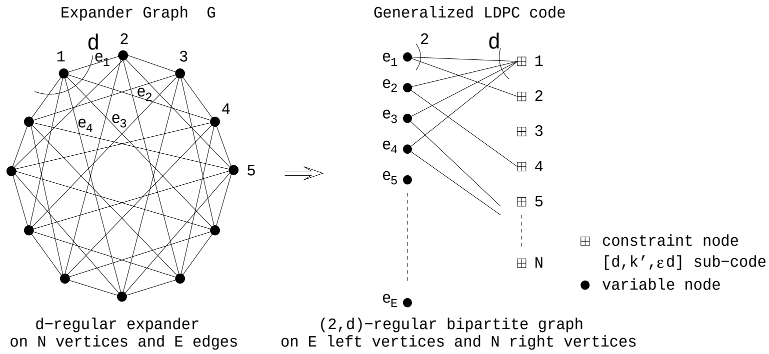

Figure 5 A d-regular graph with edges representing variable nodes and vertices representing constraints of a $[d, rd, \epsilon d]$ subcode. The graph on the right is the edge-vertex incidence graph of G.

nodes each having degree d in G'. A GLDPC code $\mathcal{C}$ is obtained by letting the edge nodes represent variable nodes and the vertex nodes each represent constraints of a $[d, rd, \epsilon d]$ subcode. Let H represent the incidence matrix for the corresponding GLDPC code $\mathcal{C}$. Then, by applying the result from the bit-oriented bound in Theorem 3.1, we get

Corollary 3.8: *The minimum distance* $d_{\min}$ *for a code represented by a d-regular graph where the edges correspond to the bit/variable nodes and the vertices correspond to a* $[d, rd, \epsilon d]$ *subcode constraints is given by*

$$d_{\min} \geq \frac{Nd}{2} \frac{(2d\epsilon - \mu_2)}{(2d - \mu_2)},$$

where μ_2 *is the second largest eigenvalue of* $H^T H$.

The above bound compares well with the lower bound of $\frac{Nd}{2}\left(\frac{\epsilon - \frac{\lambda}{d}}{1 - \frac{\lambda}{d}}\right)^2$ that is derived in (Sipser and Spielman (1996)). This bound can be improved to $\frac{Nd}{2}\epsilon\left(\frac{\epsilon - \frac{\lambda}{d}}{1 - \frac{\lambda}{d}}\right)$ by improving the lower bound in the last step of the result in Sipser and Spielman (1996). Further, λ, the second largest eigenvalue of the graph G, is related to the second eigenvalue μ_2 of $H^T H$ by the relation $\mu_2 = \lambda + d$. Substituting $\mu_2 = \lambda + d$ in the bound in Theorem 3.3 (with $c = 2$ and $n = Nd/2$) yields exactly the same bound as in Sipser and Spielman (1996).

4 Case 2: Two sub-code GLDPC codes

Let G be a (c, d)-regular bipartite graph with n left vertices of degree c and m right vertices of degree d. In (Janwa and Lal (2003)), an LDPC code was obtained by considering the edge-vertex incidence graph G' of G and using two subcodes: the edges in G are interpreted as variable nodes, the degree c left vertices as subcode constraints of a $[c, r_1 c, \epsilon_1 c]$ linear block code, and the degree d vertices as

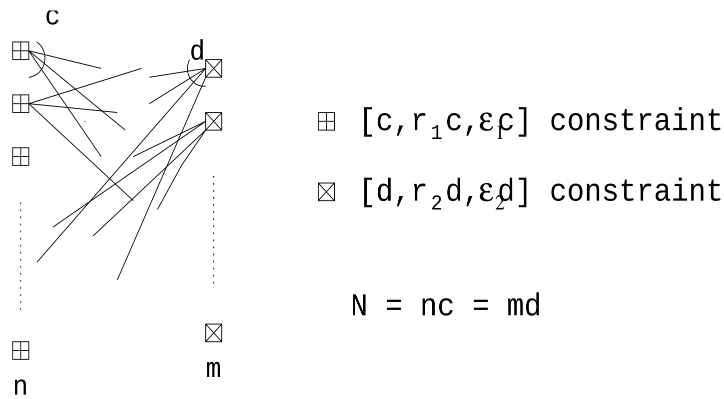

Figure 6 A (c, d)-regular graph with edges representing variable nodes, degree c nodes representing constraints of a $[c, r_1c, \epsilon_1 c]$ subcode, and degree d nodes representing constraints of a $[d, r_2d, \epsilon_2 d]$ subcode.

subcode constraints of a $[d, r_2d, \epsilon_2 d]$ linear block code. (See Figure 6.) The resulting GLDPC code has block length $N = mc = nd$ and rate $R \geq r_1 + r_2 - 1$.

Let H be the incidence matrix of G'. That is, H is a incidence matrix for the GLDPC code and has $m + n$ rows and $N = nc = md$ columns, where the first n rows have a row weight of c and correspond to the constraints of the $[c, r_1c, \epsilon_1 c]$ subcode, and the last m rows have a row weight of d and correspond to the constraints of the $[d, r_2d, \epsilon_2 d]$ subcode.

4.1 Bit-oriented bound on distance

Theorem 4.1: *Let G be a connected (c, d)-regular bipartite graph with n left nodes of uniform degree c and m right nodes of uniform degree d, and let each left node represent constraints from a $[c, r_1c, \epsilon_1 c]$ subcode, and each right node represent constraints from a $[d, r_2d, \epsilon_2 d]$ subcode. Let G' be the edge-vertex incidence graph of G with $N = nc = md$ variable nodes and $m + n$ constraint nodes. Then the minimum distance of the GLDPC code $\mathcal{C}$ with a incidence matrix H represented by the Tanner graph G' satisfies*

$$d_{\min} \geq \frac{N(\epsilon_1 c + \epsilon_2 d - \mu_2)}{(c + d) - \mu_2},$$

where μ_2 is the second largest eigenvalue of $H^T H$.

Proof: We first observe that the sum of each row and each column in the matrix $H^T H$ is $c + d$, and so $H^T H$ has largest eigenvalue $(c + d)$ with corresponding unique eigenvector $\mathbf{e}_1 = (1, 1, \ldots, 1)^T/\sqrt{N}$. As in Theorem 3.1, let $\mathbf{x}$ be a real-valued vector of length N corresponding to a minimum-weight codeword $(d_{\min})$ of the code $\mathcal{C}$, with a one in every position where the minimum-weight codeword is non-zero and zeros elsewhere. Let $\mathbf{x}_i$ denote the projection of $\mathbf{x}$ onto the ith eigenspace. Then

$$\| \mathbf{x} \|^2 = d_{\min}$$

$\| \mathbf{x}_1 \|^2 = d_{\min}^2/N$.

$H\mathbf{x} = (w_1, w_2, \ldots, w_{m+n})^T$ is an integer-valued vector of length $n+m$ where for each $i = 1, \ldots, n+m$, the weight w_i represents the number of active variable nodes adjacent to the ith constraint node in G'. Each nonzero w_i with $1 \leq i \leq n$ must be at least $\epsilon_1 c$ since each of the first n constraint nodes represents a subcode with minimum distance $\epsilon_1 c$ and $\mathbf{x}$ is a valid codeword. Similarly, each nonzero w_i with $n+1 \leq i \leq n+m$ must be at least $\epsilon_2 d$. Thus,

$$\| H\mathbf{x} \|^2 = \sum_{i=1}^{n+m} w_i^2 = \sum_{i=1}^{n} w_i^2 + \sum_{i=n+1}^{n+m} w_i^2$$

$$\geq \epsilon_1 c \sum_{i=1}^{n} w_i + \epsilon_2 d \sum_{i=n+1}^{n+m} w_i = d_{\min}(\epsilon_1 c + \epsilon_2 d). \qquad (*)$$

The last term in the above inequality is obtained by observing that each variable node connects exactly once to each type of constraint node. Thus, since $\mathbf{x}$ corresponds to a codeword of weight $d_{\min}$, we have $\sum_{i=1}^{n} w_i = d_{\min}$ and each non-zero w_i is at least $\epsilon_1 c$ for $1 \leq i \leq n$. Similarly, $\sum_{i=n+1}^{n+m} w_i = d_{\min}$ and each non-zero w_i is at least $\epsilon_2 d$ for $n+1 \leq i \leq n+m$.

Converting the above expression to eigenspace representation, we have

$$\| H\mathbf{x} \|^2 = \sum_{i=1}^{s} \mu_i \| \mathbf{x}_i \|^2 = \mu_1 (d_{\min}^2)/N + \sum_{i=2}^{s} \mu_i \| \mathbf{x}_i \|^2$$

$$\leq \mu_1 (d_{\min}^2)/N + \mu_2 \sum_{i=2}^{s} \| \mathbf{x}_i \|^2$$

$$= \mu_1 (d_{\min}^2)/N + \mu_2 (\| \mathbf{x} \|^2 - \| \mathbf{x}_1 \|^2)$$

Substituting $\| \mathbf{x}_1 \|^2 = d_{\min}^2/N$ and $\mu_1 = c + d$, and combining with expression $(*)$, we obtain

$$(c+d)(d_{\min}^2)/N + \mu_2 (d_{\min} - d_{\min}^2/N) \geq d_{\min}(\epsilon_1 c + \epsilon_2 d)$$

which gives the desired bound for $d_{\min}$. □

Remark 4.2: Note that the result in Theorem 4.1 compares well with the eigenvalue based lower bound derived in (Janwa and Lal (2003)): $d_{\min} \geq N\Big(\epsilon_1\epsilon_2 - \frac{\lambda}{2\sqrt{cd}}(\epsilon_1\sqrt{\frac{c}{d}} + \epsilon_2\sqrt{\frac{d}{c}})\Big)$, where λ is the second largest eigenvalue of G. In particular, we believe $\mu_2 = \lambda + d$, when $c = d$. Substituting $c = d$ and $\mu_2 = \lambda + d$ in the result of Theorem 4.1 yields $d_{\min} \geq N(\epsilon_1 + \epsilon_2 - 1 - \lambda/d)/(1 - \lambda/d)$. Once again, as in Remark 3.2, we note that the bound in Theorem 4.1 is useful when the graph G has good expansion and when the choice of subcodes (in terms of rates and relative minimum distances) are good.

4.2 Constraint-oriented bound on distance

Theorem 4.3: *Let G be a connected (c,d)-regular bipartite graph with n left nodes of uniform degree c and m right nodes of uniform degree d, and let each left node represent constraints from a $[c, r_1c, \epsilon_1 c]$ sub-code, and each right node represent constraints from a $[d, r_2d, \epsilon_2 d]$ linear block code. Let G' be the edge-vertex incidence graph of G with $N = nc = md$ variable nodes and $m+n$ constraint nodes. Then the minimum distance of the GLDPC code $\mathcal{C}$ with a incidence matrix H represented by the Tanner graph G' satisfies*

$$d_{\min} \geq N\epsilon_1 \frac{(c(\epsilon_1+1)-\mu_2)}{(c+d-\mu_2)}$$

where μ_2 is the second largest eigenvalue of HH^T.

Proof: Since HH^T has the same eigenvalues as H^TH (besides potentially some zero eigenvalues), the unique largest eigenvalue is $c+d$. However, due to the form of HH^T, the corresponding eigenvector is $\mathbf{e}_1 = \frac{1}{\sqrt{c^2n+d^2m}}(c,\ldots,c,d,\ldots,d)^T$, where the first n components are c and the last m components are d.

Let $\mathbf{y}$ be a real-valued vector of length $n+m$ that has a one in every active constraint node position and zeros elsewhere. Let $t = t_1 + t_2$ be the number of ones in $\mathbf{y}$, where t_1 ones occur at degree c constraint nodes and t_2 ones occur at degree d constraint nodes. Let $\mathbf{y}_i$ be the projection of $\mathbf{y}$ onto the ith eigenspace. Then, $\| \mathbf{y} \|^2 = t = t_1 + t_2$ and $\| \mathbf{y}_1 \|^2 = \frac{(t_1c+t_2d)^2}{(c^2n+d^2m)}$.

$H^T\mathbf{y} = (w_1, \ldots, w_N)^T$ is an integer-valued vector of length N where for each $i = 1, \ldots, N$, the weight w_i represents the number of active constraints nodes adjacent to the ith variable node in G. Note that if the ith bit is nonzero, then $w_i = 2$. Now,

$$\| H^T\mathbf{y} \|^2 = \sum_{i=1}^{N} w_i^2. \qquad (**)$$

To determine w_i^2, note that each active constraint node of degree c (resp., degree d) is adjacent to at least $\epsilon_1 c$ (resp., $\epsilon_2 d$) nonzero bit nodes in order for the constraint to be satisfied.

Let $u_j(\ell)$ be the number of variable nodes with weight ℓ in $H^T\mathbf{y}$ that are adjacent to the jth active constraint node, for $1 \leq \ell \leq 2$. Each active constraint node j where $1 \leq j \leq n$ is adjacent to at least $\epsilon_1 c$ nonzero variable nodes, thus $u_j(2) \geq \epsilon_1 c$. Similarly, each active constraint node j where $n+1 \leq j \leq n+m$ is adjacent to at least $\epsilon_2 d$ nonzero variable nodes, thus $u_j(2) \geq \epsilon_2 d$.

The squared weight w_i^2 in $(**)$ is found by summing over the active constraint nodes. As in Theorem 3.3, the contribution of the jth constraint node to the sum in $(**)$ is $\sum_{\ell=1}^{2} u_j(\ell)(\ell^2)(1/\ell)$.

Rewriting this expression and incorporating the observations we have for each type of active constraint node, we obtain

$$\sum_{\ell=1}^{2} u_j(\ell)(\ell^2)(1/\ell) = 2u_j(2) + u_j(1)$$

$$\geq \begin{cases} 2\epsilon_1 c + c - \epsilon_1 c = c(\epsilon_1 + 1) & \text{for } 1 \leq j \leq n \\ 2\epsilon_2 d + d - \epsilon_2 d = d(\epsilon_2 + 1) & \text{for } n+1 \leq j \leq n+m \end{cases}$$

Note that the inequality follows since for $j = 1, \ldots, n$, the expression is smallest when the jth constraint has exactly $\epsilon_1 c$ active variable node neighbors yielding $u_j(2) = \epsilon_1 c$ and $u_j(1) = c - \epsilon_1 c$. When more than $\epsilon_1 c$ variable node neighbors are active, the expression has $u_j(1) = c - \epsilon_1 c - \delta$ and $u_j(2) = \epsilon_1 c + \delta$ for $\delta > 0$ and may be lower bounded by the above. Similarly, for the case $j = n+1, \ldots, n+m$.

Since $\mathbf{y}$ has t_1 active nodes among the degree c constraint nodes and t_2 active nodes among the degree d constraint nodes,

$$\sum_{i=1}^{N} w_i^2 \geq t_1 c(\epsilon_1 + 1) + t_2 d(\epsilon_2 + 1).$$

Converting to eigenspace representation and summing over the s distinct eigenvalues of HH^T as before,

$$\| H^T \mathbf{y} \|^2 = \mu_1 \frac{(t_1 c + t_2 d)^2}{(c^2 n + d^2 m)} + \sum_{i=2}^{s} \mu_i \| \mathbf{y}_i \|^2$$

$$\leq \mu_1 \frac{(t_1 c + t_2 d)^2}{(c^2 n + d^2 m)} + \mu_2 \sum_{i=2}^{s} \| \mathbf{y}_i \|^2$$

$$= \mu_1 \frac{(t_1 c + t_2 d)^2}{(c^2 n + d^2 m)} + \mu_2 (\| \mathbf{y} \|^2 - \| \mathbf{y}_1 \|^2).$$

Substituting from above,

$$\mu_1 \frac{(t_1 c + t_2 d)^2}{(c^2 n + d^2 m)} + \mu_2 (t_1 + t_2 - \frac{(t_1 c + t_2 d)^2}{(c^2 n + d^2 m)})$$

$$\geq t_1 c(\epsilon_1 + 1) + t_2 d(\epsilon_2 + 1). \quad (***)$$

A loose bound may be obtained by considering the worst case scenarios where $t_2 = 0$ and $t_1 = t$.

For the case $t_2 = 0, t_1 = t$, the inequality in $(***)$ reduces to

$$t \geq \frac{(c(\epsilon_1 + 1) - \mu_2)}{(c + d - \mu_2)}(n + \frac{d^2}{c^2}m).$$

Since we have $nc + md = 2N$, we have $n + \frac{d}{c}m = \frac{2N}{c}$. Assuming $c \leq d$, this implies that $n + \frac{d^2}{c^2}m \geq \frac{2N}{c}$. Thus,

$$t \geq \frac{2N}{c}\frac{(c(\epsilon_1 + 1) - \mu_2)}{(c + d - \mu_2)}.$$

Furthermore, observe that for the case $t_1 = t, t_2 = 0$, the graph G has t active constraint nodes of degree c and at least $d_{\min}$ active variable nodes. Moreover, $2d_{\min} \geq \epsilon_1 ct$ since the t active constraint nodes must have at least $\epsilon_1 c$ active edges going to active variable nodes of a valid configuration in the graph, and each active variable node has d active edges incident to it. Observe that this worst case cannot happen since each active variable node connects to each type of constraint node.

Combining the above expressions, we get

$$d_{\min} \geq \frac{tc\epsilon_1}{2} \geq N\epsilon_1\frac{(c(\epsilon_1 + 1) - \mu_2)}{(c + d - \mu_2)}. \qquad \square$$

Remark 4.4: Note that for another worst case scenario where $t_1 = 0, t_2 = t$, the resulting bound on $d_{\min}$ is $d_{\min} \geq N\epsilon_2\frac{(d(\epsilon_2+1)-\mu_2)}{(c+d-\mu_2)}$. However, the above two worst case scenarios can never happen since for a non-zero value of t, both t_1 and t_2 are non-zero. Thus, a better bound may be obtained by setting $t_1 c\epsilon_1 = t_2 d\epsilon_2$. This is indeed a more judicious choice since this distribution of active constraint nodes among the degree c and degree d node is one of the extreme cases. A minimum weight codeword with weight $d_{\min}$ active variable nodes will be connected to at most $d_{\min}/(c\epsilon_1)$ degree c constraint nodes and to at most $d_{\min}/(d\epsilon_2)$ degree d constraint nodes. Hence, we set $d_{\min} \geq t_1 c\epsilon_1$ and $d_{\min} \geq t_2 d\epsilon_2$. For this case, the resulting bound is

$$d_{\min} \geq \frac{2N\epsilon_1}{(1 + \frac{\epsilon_1}{\epsilon_2})^2}\Big(\frac{c(2\epsilon_1 + 1 + \frac{\epsilon_1}{\epsilon_2}) - \mu_2(1 + \frac{c\epsilon_1}{d\epsilon_2})}{c + d - \mu_2}\Big).$$

4.3 Minimum stopping set size

We again use the generalized definition of stopping set in Definition 2.7. Under the assumption that the $[d, r_2 d, \epsilon_2 d]$ and $[c, r_1 c, \epsilon_1 c]$ subcodes have no idle components, meaning that there are no components that are zero in all of the codewords of either of the subcodes, Definition 2.7 reduces to the following:

Lemma 4.5: A stopping set in a GLDPC code with $[c, r_1 c, \epsilon_1 c]$ and $[d, r_2 d, \epsilon_2 d]$ subcodes is a set of variable nodes such that every node that is a degree c neighbor of some node $s \in S$ is connected to S at least $\epsilon_1 c$ times and every node that is a degree d neighbor of some node $s \in S$ is connected to S at least $\epsilon_2 d$ times.

Following the same proof techniques for the bit and constraint-oriented bounds on the minimum distance and using the above definition on the stopping set size, the following result can be easily shown:

Corollary 4.6: *Suppose G is a connected (c,d)-regular bipartite graph with n left nodes of uniform degree c and m right nodes of uniform degree d and suppose a GLDPC code $\mathcal{C}$ is obtained as in the construction described above. Then the stopping distance, $s_{\min}$, of G' as a Tanner graph representing the parity check matrix H is given by*

$$s_{\min} \geq \max\left\{\frac{N(\epsilon_1 c + \epsilon_2 d - \mu_2)}{c+d-\mu_2}, N\epsilon_1 \frac{(c(\epsilon_1+1)-\mu_2)}{(c+d-\mu_2)}\right\},$$

where μ_2 is the second largest eigenvalue (in absolute value) of $H^T H$ (and also HH^T).

4.4 Minimum pseudoweight

In this sub-section, a bit-oriented bound on the pseudodistance over the AWGN channel for the GLDPC with two subcodes is derived.

Theorem 4.7: *Let G be a connected (c,d)-regular bipartite graph with n left nodes of uniform degree c and m right nodes of uniform degree d, and let each left node represent constraints from a $[c, r_1 c, \epsilon_1 c]$ subcode, and each right node represent constraints from a $[d, r_2 d, \epsilon_2 d]$ subcode. Let G' be the edge-vertex incidence graph of G with $N = nc = md$ variable nodes and $m+n$ constraint nodes. Then the pseudodistance over the AWGN channel of the GLDPC code $\mathcal{C}$ with a incidence matrix H represented by the Tanner graph G' satisfies*

$$w_{\min}^{AWGN} \geq \frac{N(\epsilon_1 c + \epsilon_2 d - \mu_2)}{(c+d)-\mu_2},$$

where μ_2 is the second largest eigenvalue of $H^T H$ in absolute value.

Proof: The proof here is very similar to the proof of the bit-oriented bound on the minimum distance. Let $\mathbf{x}$ be a real-valued vector of length N corresponding to a pseudocodeword as defined in Section 2. Then the pseudocodeword weight of $\mathbf{x}$ on the AWGN channel is given by

$$w^{AWGN}(\mathbf{x}) = \frac{(x_1+x_2+\cdots+x_n)^2}{(x_1^2+x_2^2+\cdots+x_n^2)} = \frac{\|\mathbf{x}\|_1^2}{\|\mathbf{x}\|_2^2}, \quad (\dagger)$$

where $\| \mathbf{x} \|_1$ and $\| \mathbf{x} \|_2$ are the first and second norms of $\mathbf{x}$, respectively. Since $H^T H$ is a regular matrix with each row and column sum equal to $c+d$ (as in the proof of Theorem 4.1), the first eigenvector of $H^T H$ can be taken to be $\mathbf{e}_1 = (1, 1, \ldots, 1)^T / \sqrt{N}$ and the corresponding eigenvalue is $\mu_1 = c + d$. Since G is connected, $\mathbf{e}_1$ is the unique eigenvector with eigenvalue $c + d$. Let $\mathbf{x}_i$ denote the projection of $\mathbf{x}$ onto the ith eigenspace. Then

$$\| \mathbf{x}_1 \|^2 = \frac{1}{n} \| \mathbf{x} \|_1^2 .$$

$H\mathbf{x} = \mathbf{w}^T = (w_1, \ldots, w_{n+m})^T$ is a real-valued vector of length $n+m$ where for each $i = 1, \ldots, n+m$, the weight w_i represents the sum of the active pseudocodeword components at the variable nodes adjacent to the ith constraint node in G. We may lower bound $\| \mathbf{w} \|_2^2$ as follows:

$$\| \mathbf{w} \|_2^2 = \sum_{i=1}^{n+m} w_i^2 = \sum_{i=1}^{n+m} \Big(\sum_{j:H_{i,j}=1} x_j \Big)^2$$

$$= \sum_{i=1}^{n} \Big(\sum_{j:H_{i,j}=1} x_j \Big)^2 + \sum_{i=n+1}^{n+m} \Big(\sum_{j:H_{i,j}=1} x_j \Big)^2$$

$$\geq \epsilon_1 c \sum_{i=1}^{n} \Big(\sum_{j:H_{i,j}=1} x_j^2 \Big) + \epsilon_2 d \sum_{i=n+1}^{n+m} \Big(\sum_{j:H_{i,j}=1} x_j^2 \Big)$$

$$= (\epsilon_1 c + \epsilon_2 d) \cdot \| \mathbf{x} \|_2^2 \qquad (\dagger\dagger)$$

Observe that the above inequality follows from the inequalities that the pseudocodeword components must satisfy in equation (4) at the degree c constraint nodes and degree d constraint nodes. The last equality follows by observing that in the sum $\epsilon_1 c \sum_{i=1}^{n} \sum_{j:H_{j,i}} x_j^2 + \epsilon_2 d \sum_{i=n+1}^{n+m} \sum_{j:H_{j,i}} x_j^2$, each term x_j^2 is counted twice, once with weight $c\epsilon_1$ and once with weight $d\epsilon_2$ since each variable node is connected once to each type of constraint node.

Now, to upper bound $\| H\mathbf{x} \|^2$, we use the eigenspace representation to obtain

$$\| H\mathbf{x} \|^2 = \sum_{i=1}^{s} \mu_i \| \mathbf{x}_i \|^2 = \mu_1 \| \mathbf{x} \|_1^2 / N + \sum_{i=2}^{s} \mu_i \| \mathbf{x}_i \|^2$$

$$\leq \mu_1 \| \mathbf{x} \|_1^2 / N + \mu_2 \sum_{i=2}^{s} \| \mathbf{x}_i \|^2$$

$$= \mu_1 \| \mathbf{x} \|_1^2 / N + \mu_2 (\| \mathbf{x} \|_2^2 - \| \mathbf{x}_1 \|^2)$$

$$= \mu_1 \parallel \mathbf{x} \parallel_1^2 /N + \mu_2(\parallel \mathbf{x} \parallel_2^2 - \parallel \mathbf{x} \parallel_1^2 /N).$$

Combining this expression with (††), we obtain

$$\mu_1 \parallel \mathbf{x} \parallel_1^2 /N + \mu_2(\parallel \mathbf{x} \parallel_2^2 - \parallel \mathbf{x} \parallel_1^2 /N) \geq (c\epsilon_1 + d\epsilon_2) \parallel \mathbf{x} \parallel_2^2 .$$

Thus, $\parallel \mathbf{x} \parallel_1^2 (\mu_1 - \mu_2)/N \geq (c\epsilon_1 + d\epsilon_2 - \mu_2) \parallel \mathbf{x} \parallel_2^2$.

Substituting $\mu_1 = c + d$ and using the expression for the weight of the pseudocodeword $\mathbf{x}$ in (†), we see that the weight of any non-zero pseudocodeword $\mathbf{x}$ is lower bounded by $\frac{N(c\epsilon_1+d\epsilon_2-\mu_2)}{c+d-\mu_2}$. This yields the desired bound. □

5 Summary

Generalized LDPC codes form the basis of several important code constructions including expander codes and asymptotically good codes. In this paper, we derived lower bounds on the minimum distance, stopping distance, and pseudodistance for two classes of generalized LDPC codes. These bounds rely on the connectivity of the underlying Tanner graph and are extensions of Tanner's bit and parity-oriented bounds to the generalized LDPC code case. The bounds presented in this paper complement existing bounds on minimum distance and pseudodistance that require additional knowledge of the expansion properties of the graph. The results in this paper show how the use of strong subcode constraints may improve the error-correction capabilities of the codes.

Acknowledgement

This work was supported in part by NSF EPSCoR EPS-0701892. We would also like to thank the two reviewers for their detailed comments that have considerably improved the quality of the paper.

References

Tanner R. M. (1981) 'A recursive approach to low complexity codes,' *IEEE-IT*, Vol. 27, pp. 533-547.

Richardson T. J., Shokrollahi M. A., and Urbanke R. L. (2001), 'Design of capacity-approaching irregular low-density parity-check codes,' *IEEE-IT*, Vol. 47, No. 2, pp. 619-637.

Forney, Jr., G. D., Koetter R., Kschischang F., and Reznik A. (2001) 'On the effective weights of pseudocodewords for codes defined on graphs with cycles', in *"Codes, Systems and Graphical Models" (eds. B. Marcus and J. Rosenthal)*, Springer-Verlag, Vol. 123, pp. 101-112.

Koetter R. and Vontobel P. O. (2003) 'Graph-covers and iterative decoding of finite length codes', *Proceedings of the 2003 Intl. Symposium on Turbo codes*, Brest, France.

Kelley C. A. and Sridhara D. (2007) 'Pseudocodewords of Tanner graphs', *IEEE-IT*, Vol. 53, No. 11, pp. 4013-4038.

Feldman J., Wainwright M. J., and Karger D. R. (2005) 'Using linear programming to decode binary linear codes', *IEEE-IT*, Vol. 51, pp. 954-972.

Di C., Proietti D., Richardson T. J., Telatar E., and Urbanke R. L. (2002) 'Finite-length analysis of low-density parity-check codes on the binary erasure channel', *IEEE-IT*, Vol. 48, pp. 1570-1579.

Wiberg N. (1996) 'Codes and decoding on general graphs', *Ph.D. thesis*, Linköping University, Linköping, Sweden.

Tanner R. M. (2001) 'Minimum distance bounds by graph analysis', *IEEE-IT*, Vol. 47, pp. 808-821.

Sipser M. and Spielman D. A. (1996) 'Expander codes', *IEEE-IT*, Vol. 42, pp. 1710-1722.

Kelley C. A. and Sridhara D. (2007) 'Eigenvalue bounds on the pseudocodeword weight of expander codes', *Journal of Advances in Mathematics of Communications*, Vol. 1, No. 3, pp. 287-307.

Shin M-H., Kim J-S., Song H-Y. (2005) 'Generalization of Tanner's minimum distance bounds for LDPC codes', *IEEE Communications Letters*, Vol. 9, No. 3, pp. 240-242.

Janwa H. and Lal A. K. (2003) 'On Tanner codes: Minimum distance and decoding', in *Proceedings of AAECC*, Vol. 13, pp. 335-347.

Koetter R., Li W-C. W., Vontobel P. O., and Walker J. L. (2007) 'Characterizations of pseudo-codewords of LDPC codes', *Advances in Mathematics*, Vol. 217, pp. 205-229.

Vontobel P. O. and Koetter R. (2004) 'Lower bounds on the minimum pseudo-weight of linear codes', in *Proceedings of the 2004 IEEE International Symposium on Information Theory*, Chicago, USA.

van Lint J. H. and Wilson R. M. (2001), 'A Course in Combinatorics', *Cambridge University Press*, 2nd edition.