On the Number of Errors Correctable with Codes on Graphs

Alexander Barg
Dept. of ECE and Inst. for Systems Research
University of Maryland
College Park, MD 20742, USA
IPPI RAS, Moscow, Russia
Email: abarg@umd.edu

Arya Mazumdar
Department of ECE
University of Maryland
College Park, MD 20742, USA
Email: arya@umd.edu

***Abstract*—We estimate the number of errors corrected by two different ensembles of codes on graphs (generalized LDPC codes), namely codes on regular bipartite graphs and their extension to hypergraphs.**

I. Introduction

Considerable attention in recent years was devoted to the study of error correction with codes on graphs. In this paper we are interested in establishing the number of errors correctable with codes on graphs constructed as generalizations of low-density parity-check codes. The generalization that we have in mind is concerned with replacing the repetition and single-parity-check codes as local codes at the graph's vertices with other error-correcting codes.

Error correction with codes on graphs has been studied along two lines, namely, by computing the average number of errors correctable with some decoding algorithm by codes from a certain ensemble of graph codes, or by examining explicit code families whose construction involves graphs with a large spectral gap. The first direction originates in the work of Zyablov and Pinsker [8] which showed that random LDPC codes of growing length can correct a positive fraction of errors. Recently these results have been extended by Burshtein [3] where the estimate of the number of errors was improved compared to [8], and by Zyablov et al. [7] who provided estimates of the number of correctable errors under the assumption of local single error-correcting (Hamming) codes. The second line of work, initiated in Sipser and Spielman's [5], pursues estimates of error correction with codes on graphs with a small second eigenvalue and ensuing expansion properties. An extension of this construction from graphs to hypergraphs was proposed by Bilu and Hoory [2] who showed that such codes for some code rates have minimum distance greater than many bipartite-graph constructions. Interestingly, the codes considered in [2] are a direct extension of a construction in [4], [8] in the same way as Tanner's codes extend LDPC codes.

As is well known, graphs with high expansion and random graphs share many properties that can be used to prove estimates of error correction. This equivalence was emphasized in our recent work [1] which showed that ensembles of codes on random graphs and explicit expander-like constructions share many common features such as properties of the minimum distance and weight distribution.

Turning to error correction, we note one difference between (generalized) LDPC codes on random graphs and explicit constructions based on the spectral gap. In estimating the number of errors corrected by the latter, one is forced to rely on local codes with rather large minimum distance d_0, for instance, d_0 greater than the square root of the degree n of the graph. Even though in the construction of [5] and later works, n is kept constant, this effectively rules out of consideration local codes with small minimum distance such as the Hamming codes and the like. This restriction is absent for the random ensemble of graph codes as was recently shown in [7].

In this paper we obtain new estimates of the number of correctable errors with random ensembles of bipartite-graph and hypergraph codes, employing some ideas in [1]. Our results imply that codes on bipartite graphs and hypergraphs constructed from local codes with small distance can correct a positive proportion of errors under iterative decoding.

II. Code ensembles

To construct an $[N, RN]$ binary linear code C, consider a bipartite graph $G(V = V_1 \cup V_2, E)$, where $|V_1| = |V_2| = m$ and the degree of every vertex v in V is $\deg v = n$. Let $A[n, R_0 n, d_0]$ be a linear binary code called the local code below. We identify the coordinates of C with the set E and for a $v \in V$ denote by $\boldsymbol{x}(v) \in \{0,1\}^n$ the projection of a vector $\boldsymbol{x} \in \{0,1\}^N, N = nm$, on the edges incident to v. By definition,

$$C = \{\boldsymbol{x} \in \{0,1\}^N : \forall_{v \in V} \boldsymbol{x}(v) \in A\}. \tag{1}$$

The ensemble of codes $\mathscr{G}(A, m)$ is constructed by sampling a graph from the set of graphs defined by a random permutation on N elements which establishes how the edges originating in V_1 are connected to the vertices in V_2.

Generalizing, consider an l-partite n-regular uniform hypergraph $H = (V, E)$ i.e., a finite set $V = V_1 \cup \cdots \cup V_l$, where $|V_1| = \cdots = |V_l| = m$, and a collection E of l-subsets (hyperedges) of V such that every $e \in E$ intersects each $V_i, 1 \le i \le l$ by exactly one element and each vertex $v \in V$

appears in exactly n different subsets. Aiming at constructing an $[N, RN]$ binary linear code C by imposing local constraints at the vertices, we again identify the coordinates of C with the (hyper)edges of H. By definition, the code C is formed of the vectors $\boldsymbol{x}$ that satisfy condition (1) for every vertex in V. The ensemble of codes $\mathscr{H}(A, l, m)$ in this case is constructed by sampling a random hypergraph from the set of hypergraphs defined by $l-1$ independent random permutations on N elements. Of course, $\mathscr{H}(A, l, m)$ becomes $\mathscr{G}(A, m)$ for $l = 2$. As is easy to see, the rate R of the code C satisfies $R \geq lR_0 - (l-1), l = 2, 3, \ldots$.

Remarks. 1. An equivalent description of the bipartite code ensemble is obtained by considering an edge-vertex incidence graph of the graph $G(V, E)$, i.e., a bipartite graph $D = (D_1 \cup D_2, \mathcal{E})$ where $D_1 = E, D_2 = V_1 \cup V_2$, each vertex in D_1 is connected by an edge to a vertex in V_1 and a vertex in V_2 and there are no other edges in $\mathcal{E}$. Thus, for all $v \in D_1, \deg(v) = 2$ and for all $v \in D_2$, $\deg(v) = n$. The local code constrains are imposed on the vertices in D_2. By increasing the number of parts in D_2 from two to l, we then obtain the hypergraph codes defined above. The ensemble of hypergraph codes with local constraints given by single parity-check codes was introduced by Gallager [4, p.12]. The proportion of errors correctable with these codes was estimated in [8]. Several generalizations of this ensemble were studied in [1], [2], [7].

2. The derivations of this paper are not specific to binary codes: any local linear codes such as RS codes can be substituted with no conceptual changes to the analysis and the conclusions.

III. Decoding algorithms for graph (Generalized LDPC) codes

Even though the ensemble $\mathscr{G}(A, m)$ forms a particular case of the ensemble $\mathscr{H}(A, l, m)$, in our analysis we employ different decoding algorithms for the cases $l = 2$ and $l \geq 3$. The reason for this is that edge-oriented procedures commonly used for bipartite-graph codes do not generalize well to hypergraphs.

A. Decoding for the ensemble $\mathscr{G}(A, m)$. In our estimates of the number of correctable errors for the ensemble $\mathscr{G}$ we rely upon the algorithm of [6] which iterates between decoding all the vertices in parts V_1 and V_2 in parallel using some decoding algorithm of the code A. Let $C \in \mathscr{G}(A, m)$ be a code. For the ease of analysis we assume that the local codes are decoded to correct up to t errors, where $t \geq 0$ is an integer that satisfies $2t + 1 \leq d_0$. Formally, define a mapping $\psi_{A,t} : \{0,1\}^n \to \{0,1\}^n$ such that $\psi_{A,t}(\boldsymbol{z}) = \boldsymbol{x} \in A$ if $\boldsymbol{x}$ is the unique codeword that satisfies $d(\boldsymbol{z}, \boldsymbol{x}) \leq t$ and $\psi_{A,t}(\boldsymbol{z}) = \boldsymbol{z}$ otherwise. Let $\boldsymbol{y}^{(i)}$ be the estimate of the transmitted vector before the ith iteration, $i \geq 1$. The next steps are repeated for a certain number of iterations.

Algorithm I ($\boldsymbol{y}^{(1)}$)

- i odd: for all $v \in V_1$ put $\boldsymbol{y}^{(i+1)}(v) = \psi_{A,t}(\boldsymbol{y}^{(i)}(v))$;
- i even: for all $v \in V_2$ put $\boldsymbol{y}^{(i+1)}(v) = \psi_{A,t}(\boldsymbol{y}^{(i)}(v))$.

B. Decoding for the ensemble $\mathscr{H}(A, l, m)$. For the hypergraph ensemble $\mathscr{H}(A, l, M)$ we use the decoding algorithm proposed in [1]. It proves to be the best choice in terms of the number of correctable errors among several possible algorithms for these codes such as the ones in [2], [7] as well as procedures analogous to Algorithm I above.

Let $C \in \mathscr{H}(A, l, m)$ be a code and let $H(V, E)$ be the graph associated with it. For every $i = 1, 2 \ldots, l$ we will define an i-th subprocedure that decodes the local code A on every vertex in the part V_i. Suppose that a vector $\boldsymbol{u} \in \{0,1\}^N$ is associated with the edges of H. Let $v_{i,1}, \ldots, v_{i,m}$ be the vertices in the part V_i of H and let $\boldsymbol{u}_{i,1} = \boldsymbol{u}(v_{i,1}), \ldots, \boldsymbol{u}_{i,m} = \boldsymbol{u}(v_{i,m})$ be the m subvectors obtained from $\boldsymbol{u}$ upon permuting its coordinates according to the order of edges in V_i and projecting it on the vertices $v_{i,1}, \ldots, v_{i,m}$. In other words, the vector $(\boldsymbol{u}_{i,1}, \ldots, \boldsymbol{u}_{i,m})$ is obtained from $\boldsymbol{u}$ using the permutation that establishes edge connections between parts V_1 and V_i. The ith subprocedure replaces the vector $(\boldsymbol{u}_{i,1}, \ldots, \boldsymbol{u}_{i,m})$ with the vector $(\psi_{A,t}(\boldsymbol{u}_{i,1}), \ldots, \psi_{A,t}(\boldsymbol{u}_{i,m}))$.

The algorithm proceeds in iterations. Let $\boldsymbol{y} \in \{0,1\}^N$ be the received vector and let $\boldsymbol{y}(v)$ be its projection on a vertex $v \in V$. Denote by $Y_i^{(j)} = \{\boldsymbol{y}_{i,k}^{(j)}\}$ the set of estimates of the transmitted codeword (i.e., the set of N-vectors) stored at the vertices of the component V_i before the jth iteration $j = 1, 2, \ldots$. Decoding begins with setting $Y_i^{(1)} = \{\boldsymbol{y}\}$ for all $i = 1, \ldots, l$. After the first iteration we obtain l potentially different vectors which form the current estimates of the transmitted vector. These vectors form the sets $Y_i^{(2)}, i = 1, \ldots, l$. In the next iteration each subprocedure will have to be applied to each of the l outcomes of the preceding iteration. Proceeding in this way, we observe that $|Y_i^{(j)}| \leq l^{j-1}$.

This algorithm, called Algorithm II below, will only be applied for a constant number s of iterations until we can guarantee that at least one subprocedure has reduced the number of errors to a specified proportion, say from σN to some $\sigma_1 N, \sigma_1 < \sigma$. We then let another algorithm take over and decode all the l^s candidates, concluding by choosing the codevector closest to $\boldsymbol{y}$ by the Hamming distance. Here we let this algorithm to be the decoding algorithm of bipartite-graph codes (Algorithm I), making sure that σ_1 is below the proportion of errors that are necessarily corrected by this algorithm for the ensemble $\mathscr{G}(A, m)$. This is possible because, leaving any two parts of the original hypergraph H to form a bipartite graph G, we obtain a random code from the ensemble $\mathscr{G}(A, m)$ which with high probability will remove all the residual errors from at least one candidate estimate.

Though the last step of the algorithm described is different from [1], the main idea is the same, so we refer to that paper for a more detailed description and a discussion of the algorithm.

IV. Number of correctable errors for the ensemble $\mathscr{G}(A, m)$

Let $C \in \mathscr{G}(A, m)$ be a code and let $G(V, E)$ be the graph associated with it. For a given subset of vertices $S \subset V_i, i = 1, 2$ denote by $T_r(S)$ the set of vertices $v \in V$

that are connected to S by $r+1$ or more edges, where $r \in \{0,\dots,n-1\}$ is an integer.

Let $t \geq 0$ be any integer such that $2t+1 \leq d_0$. The calculation in this section is based on the following simple observation: if for all $S \subset V_i, i=1,2, |S| \leq \sigma m$ there exists $\epsilon > 0$ such that $|T_t(S)| \leq |S| - \epsilon m$, then any $\sigma t m = \sigma t(N/n)$ errors will be corrected by Algorithm I in $O(\log m)$ iterations.

Let $\mathcal{Z}_n = \{\boldsymbol{z} \in [0,1]^{n+1} : \sum_{i=0}^n z_i = 1\}$ be the $(n+1)$-dimensional probability simplex. Define

$$F_{n,t}(\sigma) = \max_{\boldsymbol{z} \in \mathscr{M}(t,\sigma)} \Big(h(\boldsymbol{z}) + \sum_{i=1}^n z_i \log \binom{n}{i} \Big) \qquad (2)$$

where

$$\mathscr{M}(t,\sigma) = \Big\{ \boldsymbol{z} \in \mathcal{Z}_n : \sum_{i=1}^n i z_i = \sigma n,\ \sum_{i=t+1}^n z_i = \sigma \Big\},$$

and $h(\boldsymbol{z})$ is the entropy of the probability vector $\boldsymbol{z}$. In the particular case of $n=1$ we write $h(z)$ instead of $h(z, 1-z)$.

The main result of this section is given by the next theorem.

Theorem 4.1: Let $A[n, R_0 n, d_0]$ be the local code, $m \to \infty$, and let $1 \leq t < d_0/2$. All codes in the ensemble $\mathscr{G}(A,m)$ except for an exponentially small (in N) proportion of them correct any combination of errors of weight $\sigma t m$ in $O(\log m)$ iterations of Algorithm I, where $0 \leq \sigma < \sigma^*$ and σ^* is the smallest positive root of the equation

$$F_{n,t}(\sigma) = (n-1)h(\sigma).$$

Proof (outline): Suppose that the vertices in V_1 decoded incorrectly after the first iteration of the algorithm form a subset $S \subset V_1, |S| = \sigma m$. Let $m_i = |\{v \in V_2 : \deg_S(v) = i\}|, i = 0,1,\dots,n$. Clearly,

$$\sum_{i=0}^n m_i = m,\ \sum_{i=t+1}^n m_i = |T_t(S)|,\ \sum_{i=1}^n i m_i = |S| n.$$

Let us compute the probability (over the choice of G) that $|T_t(S)| \geq |S|$. Let $\boldsymbol{\mu} = (m_0, m_1, \dots, m_n) \in (\mathbb{Z}_+ \cup \{0\})^{n+1}$ be a vector with nonnegative integer components, let

$$M(t,\sigma) = \{\boldsymbol{\mu} :\ \sum_{i=0}^n m_i = m,\ \sum_{i=1}^n i m_i = \sigma N,\ \sum_{i=t+1}^n m_i \geq \sigma m\},$$

and let $\binom{m}{\boldsymbol{\mu}}$ denote the multinomial coefficient. We have

$$P(|T_t(S)| \geq |S|) = \frac{1}{\binom{N}{\sigma N}} \sum_{\boldsymbol{\mu} \in M(t,\sigma)} \binom{m}{\boldsymbol{\mu}} \prod_{i=0}^n \binom{n}{i}^{m_i}.$$

Let $\mathscr{L}_1(s)$ denote the event that V_1 contains a subset $S, |S| = s$ for which $|T_t(S)| \geq |S|$. We have

$$P(\mathscr{L}_1(\sigma m)) \leq \binom{m}{\sigma m} P(|T_t(S)| \geq |S|)$$

and

$$P\Big(\bigcup_{i=1}^{\sigma m} \mathscr{L}_1(i) \Big) \leq m P(\mathscr{L}_1(\sigma m)).$$

Denote by $\mathscr{L}_2(\sigma)$ an analogous event with respect to S_2. Then

$$P\Big(\bigcup_{i=1}^{\sigma m} (\mathscr{L}_1(i) \cup \mathscr{L}_2(i)) \Big) \leq 2m \frac{\binom{m}{\sigma m}}{\binom{N}{\sigma N}} \sum_{\boldsymbol{\mu} \in M(t,\sigma)} \binom{m}{\boldsymbol{\mu}} \prod_{i=0}^n \binom{n}{i}^{m_i}.$$

Letting L to be the logarithm of the left-hand side divided by N, we obtain (omitting $o_N(1)$ terms)

$$L \leq -(1 - 1/n) h(\sigma) + \frac{1}{n} \max_{\boldsymbol{z} \in \mathscr{M}'(t,\sigma)} \Big(h(\boldsymbol{z}) + \sum_{i=0}^n z_i \log \binom{n}{i} \Big)$$

where

$$\mathscr{M}'(t,\sigma) = \Big\{ \boldsymbol{z} \in \mathcal{Z}_n : \sum_{i=1}^n i z_i = \sigma n,\ \sum_{i=t+1}^n z_i \geq \sigma \Big\}$$

and $z_i = m_i/m$. Finally, it can be shown that the region $\mathscr{M}'$ in the optimization can be replaced with $\mathscr{M}$, i.e., the maximum is attained on the boundary. We omit the details. ∎

For computational purposes we note that the maximization in the definition of F can be performed explicitly. We obtain

$$F_{n,t}(\sigma) = h(\sigma) - \sigma n \log x + \sigma \log \sum_{i=t+1}^n \binom{n}{i} x^i + (1-\sigma) \log \sum_{i=0}^t \binom{n}{i} x^i, \quad (3)$$

where $x > 0$ is found from the equation

$$\sum_{i=0}^t \sum_{j=t+1}^n \binom{n}{i} \binom{n}{j} \Big(\frac{\sigma(n-j)}{1-\sigma} - i \Big) x^{i+j} = 0.$$

Example 1: Using Theorem 4.1 together with (3) we can compute the proportion of errors corrected by codes in the ensemble $\mathscr{G}(A,m), m \to \infty$ for several choices of the local code A. For instance, taking A to be the binary Golay code of length $n=23$ we find $\sigma^* \approx 0.0048586$ and therefore, the proportion of correctable errors is $\frac{\sigma^* t}{n} \approx 0.00063$. Similarly, for the 2-error-correcting $[n=31, k=21]$ BCH code we find $\sigma^* \approx 0.000035$ and $\frac{\sigma^* t}{n} \approx 0.0000023$.

To underscore similarities with the results obtained for expander-like codes (e.g., [6]) we compute the proportion of errors correctable with codes from the ensemble $\mathscr{G}(A,m)$ in the case of large n.

Proposition 4.2: Let $t = \tau n$. Then the ensemble $\mathscr{G}(A,m)$ contains codes that correct $\sigma\tau N$ errors for any $\sigma \leq \sigma^*$, where σ^* is given by

$$\sigma^* = \sup \Big\{ \sigma > 0 : \forall_{0<x<\sigma}\ (1-x) h\Big(\frac{x(1-\tau)}{1-x} \Big) + x h(\tau) + \varepsilon_n < h(x) \Big\}$$

where ε_n is a quantity of order $O\big(\frac{\log n}{n}\big)$ which can be found explicitly.

The proof is obtained by analyzing the estimate for L in the preceding proof in the case of large n and will be omitted.

Therefore, if $n \to \infty$, the value of σ^* approaches τ, so the ensemble $\mathscr{G}$ contains codes that correct up to a τ^2 proportion of errors, where $\tau n = d_0/2$ is the error-correcting capability of the code A. As in the case of expander codes, the proportion

of correctable errors can be improved to $d_0^2/(2n)$ by using a more powerful decoding algorithm.

V. Number of correctable errors for the ensemble $\mathscr{H}(A,l,m)$

In this section we first state a sufficient condition for the existence of at least one subprocedure within Algorithm II that reduces the number of errors, and then perform the analysis of random hypergraphs to show that with high probability this condition is satisfied. Overall this will show that the number of errors in at least one of the candidates in the list generated after a few iterations is reduced to a desired level.

Denote by $E(v)$ the set of edges incident to a vertex $v \in V$. Let $C \in \mathscr{H}(A,l,m)$ be a code and let $H(V,E)$ be its associated graph. Let $\mathcal{E} \subset E$ be the set of errors at the start of some iteration of the algorithm. The next set of arguments will refer to this iteration. Let $G_i = \{v \in V_i : |E(v) \cap \mathcal{E}| \le t\}$ be the set of vertices such that each of them is incident to no more than t edges from $\mathcal{E}$ (such errors will be corrected upon one decoding). Let $B_i = \{v \in V_i : |E(v) \cap \mathcal{E}| \ge d_0 - t\}$ be the set of vertices that can introduce errors after one decoding iteration.

The main condition for successful decoding is given in the next lemma.

Lemma 5.1: Assume that for every $\mathcal{E} \subset E, |\mathcal{E}| \le \alpha N$ there exists $i = i(\mathcal{E}), 1 \le i \le l$ such that $|\mathcal{E}(G_i)| \ge tB_i + \epsilon N$, where $\mathcal{E}(G_i)$ is the set of edges of $\mathcal{E}$ incident to the vertices of G_i and $\epsilon > 0$. Then for any $0 < \beta < \alpha$, Algorithm II will reduce any αN errors in the received vector to at most βN errors in $O_\beta(1)$ iterations.

Proof: We need to prove that at least one of the subprocedures will find a vector with no more than βN errors after a constant number of iterations. In any given iteration by the assumption of the lemma there exists a component V_i for which the ith subprocedure will decrease the count of errors by $|\mathcal{E}(G_i)| - tB_i \ge \epsilon N$. Thus, in each iteration there exists a subprocedure that reduces the number of errors by a positive fraction. ∎

Next we show that the assumption of Lemma 5.1 holds with high probability over the ensemble. We will again use the function $F_{n,t}(\gamma)$ defined in (2), except that in this section the region $\mathscr{M}(t,\gamma)$ will be as follows:

$$\mathscr{M}(t,\gamma) = \Big\{\boldsymbol{z} \in \mathcal{Z}_n : \sum_{i=1}^{n} i z_i = \gamma n, \sum_{i=1}^{t} i z_i = \sum_{i=d_0-t}^{n} t z_i\Big\}.$$

Lemma 5.2: Let $m \to \infty$ and let

$$\gamma^* = \sup\{x > 0 : \forall_{0<\gamma\le x}\, (l/n)F_{n,t}(\gamma) < (l-1)h(\gamma)\}. \quad (4)$$

A hypergraph from the ensemble of l-partite uniform n-regular hypergraphs with probability $1 - 2^{-\Omega(N)}$ has the property that for all $\mathcal{E} \subset E, |\mathcal{E}| < \gamma^* N$, the inequality $|\mathcal{E}(G_i)| > tB_i$ holds for at least one $i \in \{1,\dots,l\}$.

Proof: (outline) Let $\mathcal{E} \subset E, |\mathcal{E}| = \gamma N$. Let $m_i = |\{v \in V_1 : |E(v) \cap \mathcal{E}| = i\}|, i = 1,\dots,n$. Clearly $|\mathcal{E}(G_1)| = \sum_{i=0}^{t} i m_i$ and $|B_1| = \sum_{i=d_0-t}^{n} m_i$. We have

$$p \triangleq P(t|B_1| \ge |\mathcal{E}(G_1)|) = \frac{1}{\binom{N}{\gamma N}} \sum_{\boldsymbol{\mu} \in M(t,\gamma)} \binom{m}{\boldsymbol{\mu}} \prod_{i=0}^{n} \binom{n}{i}^{m_i},$$

where

$$M(t,\gamma) = \{\boldsymbol{\mu} \in (\mathbb{Z}_+ \cup 0)^{n+1} : \sum_{i=0}^{n} m_i = m, \sum_{i=1}^{n} i m_i = \gamma N, \sum_{i=d_0-t}^{n} t m_i \ge \sum_{i=0}^{t} i m_i\}.$$

Denote by $\mathscr{L}(\mathcal{E})$ the event that for a given subset $\mathcal{E} \subset E, |\mathcal{E}| = \gamma N$ no part V_i satisfies the assumption of Lemma 5.1. Then $P(\mathscr{L}(\mathcal{E})) = p^l$ and

$$P(\exists \mathcal{E} : |\mathcal{E}| \le \gamma N \text{ and } \mathscr{L}(\mathcal{E})) \le N \binom{N}{\gamma N} p^l.$$

Letting L to be the logarithm of the left-hand side of this inequality divided by N and omitting $o_N(1)$ terms, we obtain

$$L \le -(l-1)h(\gamma) + \frac{l}{n} \max_{\boldsymbol{z} \in \mathscr{M}'(t,\gamma)} \Big(h(\boldsymbol{z}) + \sum_{i=1}^{n} z_i \log \binom{n}{i}\Big),$$

where

$$\mathscr{M}'(t,\gamma) = \{\boldsymbol{z} \in \mathcal{Z}_n : \sum_{i=1}^{n} i z_i = \gamma n, \sum_{i=1}^{t} i z_i \le \sum_{i=d_0-t}^{n} t z_i\}$$

and $z_i = m_i/m$. Finally it can be shown that the optimization region $\mathscr{M}'$ can be replaced by $\mathscr{M}$. The details will be again omitted. ∎

This enables us to establish the main result of this section.

Theorem 5.3: Algorithm II corrects any combination of $\gamma^* N$ errors for any code $C \in \mathscr{H}(A,l,m)$ except for a proportion of codes that declines exponentially with the code length N.

Proof: With high probability over the ensemble of hypergraphs considered, for a given hypergraph $H(V_H, E_H)$ a constant number s of iterations of the algorithm will decrease the weight of error from $\gamma^* N$ to any given positive proportion β for at least one of the l^s candidates in the list $Y_1^{(s+1)}$. Take $\beta = \sigma^*$, where σ^* is the quantity given by Theorem 4.1. Next consider the bipartite graph $G(V_G = V_1 \cup V_2, E_G)$ where V_1, V_2 are the two parts of H and where $(v_1, v_2) \in E_G$ if $v_1, v_2 \in e$ for some edge $e \in E_H$. By the previous section, with high probability these $\sigma^* N$ errors can be corrected $O(\log m)$ iterations of Algorithm I. ∎

The complexity of this decoding is $O(N \log N)$ where the implicit constant depends on the code A.

To give some examples, let us evaluate $F_{n,t}(\gamma)$ for $t = 1, d_0 = 3$. Performing the maximization, we obtain

$$F_{n,1}(\gamma) = -\gamma n \log x + \log\Big(1 + 2\sqrt{n \sum_{i=2}^{n} \binom{n}{i} x^{i+1}}\Big),$$

where x is the only positive root of the equation

$$\frac{\sum_{i=2}^{n}(i+1)\binom{n}{i}x^{i+1}}{2n\sum_{i=2}^{n}\binom{n}{i}x^{i+1}+\sqrt{n\sum_{i=2}^{n}\binom{n}{i}x^{i+1}}}=\gamma.$$

This enables us to find the proportion of correctable errors for the case when A is the Hamming code of length $n=2^r-1$. In the following table $r=9$.

Example 2:

l	17	23	28	34
Rate	0.7006	0.5949	0.5069	0.4012
γ^*	0.000235	0.000401	0.000521	0.000644

l	40	45	51
Rate	0.2955	0.2074	0.1018
γ^*	0.000747	0.000821	0.000898

It is also of interest to compute the values of γ^* for code rate $R(C)\approx 0.5$.

n	127	255	511	1023
l	9	16	28	51
Rate	0.5039	0.4980	0.5068	0.5015
γ^*	0.0002012	0.0004873	0.0005207	0.0004227

Finally note that the proportion of errors corrected by LDPC codes under a "flipping" algorithm of [8] for the regular $(5,10)$ ensemble is about 0.00002 [3].

The case of large n. As in the previous section, it is interesting to examine the case of long local codes A because it reveals some parallels with the analysis of the decoding algorithm in the case of nonrandom hypergraphs [1]. We begin with the observation that the proportion γ^* of correctable errors for the ensemble $\mathscr{H}(A,t,m)$ computed above is a function of the number of errors t that each local code corrects in each iteration. Let $n\to\infty, t=\tau n, d_0=\delta_0 n$, and let $\gamma^*(\tau)$ be the proportion of errors defined in (4). Note that for $n\to\infty$ the inequality condition in (4) can be written as

$$\frac{l}{n}\max_{\boldsymbol{z}\in\mathscr{M}(t,\gamma)} z_i\log\binom{n}{i}=(l-1)h(\gamma).$$

The value of $\gamma^*(\tau)$ is made more specific in the next lemma whose proof will be omitted.

Lemma 5.4: $\gamma^*(\tau)=\min(\tau,x(\tau))$ where $x(\tau)$ is the smallest positive root of the equation

$$l\Big[\Big(1-\frac{x}{\delta_0}\Big)h\Big(\frac{x\tau}{\delta_0-x}\Big)+\frac{x}{\delta_0}h(\delta_0-\tau)\Big]=(l-1)h(x).$$

From this, the next proposition is immediate.

Proposition 5.5: The proportion of errors correctable by Algorithm II for codes in the ensemble $\mathscr{H}(A,l,m)$ with long local codes equals

$$\gamma^*=\max_{0<\tau<\delta_0/2}\gamma^*(\tau).$$

Recall that a lower bound on average distance $d_0=\delta_0 N$ for the ensemble $\mathscr{H}(A,l,m)$ in the case of large m and n is determined the next equation [1]:

$$\frac{lx}{\delta_0}h(\delta_0)=(l-1)h(x).$$

Using this bound and the last proposition, in the next few examples we compute the proportion of correctable errors for the ensemble $\mathscr{H}(A,l,m)$.

Example 3. Let $l=3$. Using local codes with $\delta_0=0.05$ we can construct hypergraph codes of rate $R\geq 0.19$. Then the ensemble-average relative distance is at least $\delta\approx 0.007$ and the proportion of errors correctable by Algorithm II is found from (5.5) to be $\gamma^*\approx 0.0035$. Similarly, for $\delta_0=0.01$ and $l=10$ we find $\gamma^*\approx 0.002198$ and $\delta\approx 0.00538, R\geq 0.14$.

VI. Conclusion

We have estimated the proportion of errors correctable by codes from ensembles defined by random l-partite graphs, $l\geq 2$. In contrast to the case of expander codes [1], [2], [5], [6] our calculations cover the case of local codes of arbitrary given length and distance including small values of the distance. The behavior of code ensembles considered here was examined from a different perspective in [1] where we computed estimates of the expected distance and weight distribution of these codes. Both in [1] and in this paper, some properties of random code ensembles were shown to be close to those of codes based on graphs with good expansion.

Acknowledgment

This work is supported in part by NSF grants CCF0830699, CCF0635271, and DMS0807411.

References

[1] A. Barg, A. Mazumdar, and G. Zémor, "Weight distribution and decoding of codes on hypergraph," Advances in Mathematics of Communication **2** (2008), no. 4, 433–450, arxiv:0808.3453.

[2] Y. Bilu and S. Hoory, "On codes from hypergraphs," European Journal of Combinatorics **25** (2004), 339–354.

[3] D. Burshtein, "On the error correction of regular LDPC codes using the flipping algorithm," IEEE Trans. Inform. Theory **54** (2008), no. 2, 517–530.

[4] R. G. Gallager, Low-Density Parity-Check Codes, MIT Press, 1963.

[5] M. Sipser and D. A. Spielman, "Expander codes," IEEE Trans. Inform. Theory **42** (1996), no. 6, 1710–1722.

[6] G. Zémor, "On expander codes," IEEE Trans. Inform. Theory **47** (2001), no. 2, 835–837.

[7] V. Zyablov, M. Lončar, R. Johannesson, and P. Rybin, "On the asymptotic performance of low-complexity decoded LDPC codes with constituent Hamming codes," Proc. 5th Int. Symp. on Turbo Codes & Related Topics, Lausanne, Switzerland, September 2008.

[8] V. V. Zyablov and M. S. Pinsker, "Estimation of the error-correcting complexity of Gallager low-density codes," Problems of Information Transmission **11** (1975), no. 1, 18–28.