

On Tanner Codes: Minimum Distance and Decoding

H. Janwa, A. K. Lal

Department of Mathematics and Computer Science, University of Puerto Rico, Rio Piedras Campus, PO Box: 23355, San Juan, PR 00931-3355, USA
(email: {hjanwa,arlal}@rrpac.upr.clu.edu)

Received: March 2, 2001; revised version: November 28, 2001

Abstract. A bound on the minimum distance of Tanner codes / expander codes of Sipser and Spielman is obtained. Furthermore, a generalization of a decoding algorithm of Zémor to Tanner codes is presented. The algorithm can be implemented using the same complexity as that of Zémor and has a similar error-correction capability. Explicit families of Tanner codes are presented for which the decoding algorithm is applicable.

Keywords: Tanner codes, Expander codes, LDPC codes, Decoding, Minimum distance, Expander graphs, Ramanujan graphs, N-gons, Multi-partite graphs.

1. Introduction

Sipser and Spielman [13] construct a new family of asymptotically good, linear error-correcting codes using Ramanujan graphs. Furthermore, they give a logarithmic time parallel decoding algorithm that uses a linear number of processors. They called the above family of codes *expander codes*. These codes are low density parity check (LDPC) codes as defined by Gallager [5].

The expander codes are obtained using a (c, d) -regular bipartite graph and a code C of smaller length giving check conditions for the bits. As pointed out by Sipser and Spielman [13, Remark 8] such a construction first appeared in the work of Tanner [16]. Tanner's construction is a generalization of the product codes of Elias [4] and LDPC codes of Gallager [5]. The decoding algorithm Tanner proposed is a generalization and unification of the decoding schemes presented by Elias and Gallager. For brevity in the sequel, we call these codes

The first author has his Research supported in part by NSF Grant No. CISE-9986985.
The second author is on leave from Indian Institute of Technology, Kanpur, India, and is currently visiting UPR.

Tanner codes. Sipser and Spielman [13] gave a new approach and direction to these codes. This fundamental work has motivated in construction of codes that attain near Shannon capacity performance (see for example [10] and [18]).

For a graph G , let B be its edge-vertex incidence matrix. Then the bipartite graph with adjacency matrix $\begin{bmatrix} 0 & B \\ B^t & 0 \end{bmatrix}$ is called an *edge-vertex graph*. Sipser and Spielman [13] presented a decoding algorithm (for the special case of expander codes from edge-vertex graphs) that uses a linear number of processors in logarithmic parallel steps. This decoding algorithm can correct approximately up to $\frac{\epsilon^2}{48}$ fraction of errors, where ϵ is the relative minimum distance of the original code C .

For the special case when edge-vertex graphs come from d -regular bipartite graphs, Zémor [17] presents a decoding algorithm that has the same complexity as that of Sipser and Spielman [13] but can correct 12 times more errors. Zémor generalizes a result of Alon and Chung [3] on the edge density of a subgraph of a d -regular graph and uses it in the analysis of his decoding algorithm. We first generalize this result of Zémor to give a bound on the minimum distance of Tanner codes. This is the only bound on the minimum distance of Tanner codes obtained from arbitrary (c, d) -regular bipartite graphs.

We also present an algorithm for decoding Tanner codes and analyze its complexity and error-correcting capabilities using the generalized result. This algorithm is a generalization of the algorithm of Zémor mentioned earlier.

If L is the length of the Tanner code then the generalized algorithm can be implemented using a circuit of size $\mathcal{O}(L \log L)$ and depth $\mathcal{O}(\log L)$. The algorithm can correct up to $\frac{d_1}{2c} \left(\frac{d_2}{2d} - \frac{\mu}{d} \right)$ errors in the received word, where $d_1 \geq d_2$ are respectively, the minimum distances of the codes C_1, C_2 , specifying the checks at the disjoint vertex sets of the (c, d) -regular bipartite graph. Thus the general decoding algorithm is able to correct a similar fraction of errors of the minimum distance bound (see Theorem 3.1) as that of Zémor (which is for the case $c = d$ and $C_1 = C_2$).

The outline of the paper is as follows. In Section I we introduce Tanner codes [16]. To derive results on the minimum distance of Tanner codes and their decoding, we generalize a result of Zémor on the edge density of a subgraph in Section 2. A bound on the minimum distance is presented in Section 3. A decoding algorithm for Tanner codes is presented in Section 4. In Section 5 an analysis, error-capability, and decoding complexity of the algorithm is presented. Finally in Section 6, asymptotic results and applications are discussed.

1.A. Expander Graphs

Let G be a (c, d) -regular bipartite graph with vertex set $M \cup N$ and edge set E . Let $|M| = m$ and $|N| = n$, and therefore, $mc = nd$. For $S \subset M$, define $\partial^*S = \{u \in N : (v, u) \in E \text{ for some } v \in S\}$. Then the graph G is called an

$(n, c, d, \alpha, \delta)$ -*expander* (see Tanner [15]) if every subset of at most αn vertices of M “expands” by a factor of at least δ , *i.e.*

$$\forall S \subset M, \text{ with } |S| \leq \alpha n, \quad |\partial^* S| \geq \delta |S|.$$

Let A be the adjacency matrix of a (c, d) -regular connected bipartite graph G . Then A is symmetric and has $n + m$ eigenvalues, with \sqrt{cd} as the largest and $-\sqrt{cd}$ as the smallest eigenvalue. Thus, we denote the eigenvalues of A by $\sqrt{cd} = \lambda_0 > \lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_{n+m-2} = -\lambda_1 > \lambda_{n+m-1} = -\sqrt{cd}$. We define,

$$\mu(G) = \max\{|\lambda_i| : |\lambda_i| \neq \sqrt{cd}\} = \lambda_1.$$

Throughout this paper, a graph G will always denote a *connected* graph and $\mu := \mu(G)$ if G is understood from the context.

Remark 1.1. A bound on the expansion coefficient of bipartite graphs was obtained by Tanner [15] in terms of μ .

1.B. Tanner Codes

Tanner’s construction of codes for bit degree 2 (a generalization of the edge-vertex case) follows (see Tanner [16] or Sipser and Spielman [13]): Let G be a (c, d) -regular bipartite graph with vertex set $M \cup N$ and edge set E with $|M| = m$ and $|N| = n$, and thus $mc = nd$. For each $v \in M$ we associate a linear $[c, r_1c, d_1 = \delta_1c]$ code C_1 , and for each $u \in N$ we associate a linear $[d, r_2d, d_2 = \delta_2d]$ code C_2 . For a particular enumeration of the edges, associate the edges of the graph G with the symbols \mathbf{x}_i , $1 \leq i \leq mc$. Let the symbols incident with a vertex $v \in M$ be $\mathbf{x}_{v(1)}, \mathbf{x}_{v(2)}, \dots, \mathbf{x}_{v(c)}$ and that with vertex $u \in N$ be $\mathbf{x}_{u(1)}, \mathbf{x}_{u(2)}, \dots, \mathbf{x}_{u(d)}$. The new code $\mathcal{C}(G, C_1, C_2)$ of length mc is constructed by requiring that $(\mathbf{x}_{v(1)}, \mathbf{x}_{v(2)}, \dots, \mathbf{x}_{v(c)})$ be a codeword of the code C_1 for each $v \in M$ and $(\mathbf{x}_{u(1)}, \mathbf{x}_{u(2)}, \dots, \mathbf{x}_{u(d)})$ be a codeword of the code C_2 for each $u \in N$.

Tanner [16] showed that the rate of the code is at least $r_1 + r_2 - 1$. It was observed that the rate and the minimum distance of $\mathcal{C}(G, C_1, C_2)$ are dependent on the way the edges are enumerated. We also use the notation $\mathcal{C}(G, C)$, if there is only one initial code C .

2. A Result on the Edge Density of a Subgraph

The following result is based on the results of Zémor [17] and Alon and Chung [3].

Lemma 2.1. *Let G be a (c, d) -regular bipartite graph with vertex set $M \cup N$ with $|M| = m$ and $|N| = n$. For $S \subset M$ and $T \subset N$ let $e(S, T)$ denote the number of edges from the set S to the set T . Then*

$$\left| |e(S, T)| - \frac{d |S| |T|}{m} \right| \leq \frac{\mu}{2} \left(|S| + |T| - \frac{|S|^2}{m} - \frac{|T|^2}{n} \right) \quad (1)$$

Therefore, if \bar{d}_{ST} denotes the average degree of the subgraph induced by $S \cup T$, then

$$\left| \bar{d}_{ST} - \frac{2d |S| |T|}{(|S| + |T|)m} \right| \leq \mu \left(1 - \frac{1}{nm} \left(\frac{n|S|^2 + m|T|^2}{|S| + |T|} \right) \right). \quad (2)$$

Proof. Let A be the $(n+m) \times (n+m)$ adjacency matrix of the (c, d) -regular bipartite graph G . Let $\tau = \sqrt{d/c}$. The eigenvectors corresponding to the eigenvalues \sqrt{cd} and $-\sqrt{cd}$ of A are respectively, $\Phi_0 = \frac{1}{\sqrt{2m}}(1, \dots, 1, \tau, \dots, \tau)^t$ and $\Phi_{n+m-1} = \frac{1}{\sqrt{2m}}(1, \dots, 1, -\tau, \dots, -\tau)^t$. Let $\Phi_1, \Phi_2, \dots, \Phi_{n+m-2}$ be the other orthonormal eigenvectors of the symmetric matrix A .

Let f be the indicator vector of the set $S \cup T$. That is, if v_i is the i -th vertex of the bipartite graph G then $f = (f_0, f_1, \dots, f_{n+m-2})^t$ with

$$f_i = \begin{cases} 1 & \text{if } v_i \in S \cup T \\ 0 & \text{otherwise} \end{cases}.$$

Therefore

$$\langle f, \Phi_0 \rangle = \frac{|S| + \tau|T|}{\sqrt{2m}} \quad \text{and} \quad \langle f, \Phi_{n+m-1} \rangle = \frac{|S| - \tau|T|}{\sqrt{2m}}. \quad (3)$$

Let $g = f - \langle f, \Phi_0 \rangle \Phi_0 - \langle f, \Phi_{n+m-1} \rangle \Phi_{n+m-1}$. Since g is orthogonal to both Φ_0 and Φ_{n+m-1} , using Equation (3) and $f^t A f = 2 |e(S, T)|$, we get

$$\begin{aligned} 2 |e(S, T)| &= f^t A f \\ &= g^t A g + \langle f, \Phi_0 \rangle^2 \Phi_0^t A \Phi_0 + \langle f, \Phi_{n+m-1} \rangle^2 \Phi_{n+m-1}^t A \Phi_{n+m-1} \\ &= g^t A g + \sqrt{cd} \left(\langle f, \Phi_0 \rangle^2 - \langle f, \Phi_{n+m-1} \rangle^2 \right) \\ &= g^t A g + \sqrt{cd} \frac{4|S| |T| \tau}{2m} \\ &= g^t A g + \frac{2|S| |T| d}{m}. \end{aligned} \quad (4)$$

We now establish a bound on $g^t A g$. Since $\{\Phi_i\}_{i=0}^{n+m-1}$ is an orthonormal basis of \mathbb{R}^{n+m} and g is orthogonal to Φ_0 and Φ_{n+m-1} , we have $g = \sum_{i=1}^{n+m-2} \gamma_i \Phi_i$.

Recall that $\mu = \lambda_1 = -\lambda_{n+m-2}$,

$$g^t A g = \sum_{i=1}^{n+m-2} \lambda_i \gamma_i^2 \leq \mu \sum_{i=1}^{n+m-2} \gamma_i^2 = \mu \langle g, g \rangle \quad (5)$$

and $g^t A g = \sum_{i=1}^{n+m-2} \lambda_i \gamma_i^2 \geq -\mu \sum_{i=1}^{n+m-2} \gamma_i^2 = -\mu \langle g, g \rangle$.

Since $g = f - \langle f, \Phi_0 \rangle \Phi_0 - \langle f, \Phi_{n+m-1} \rangle \Phi_{n+m-1}$,

$$g_i = \begin{cases} 1 - \frac{|S|}{m} & \text{if } v_i \in S \\ -\frac{|S|}{m} & \text{if } v_i \in M \setminus S \\ 1 - \frac{|T|}{n} & \text{if } v_i \in T \\ -\frac{|T|}{n} & \text{if } v_i \in N \setminus T \end{cases}$$

and

$$\langle g, g \rangle = |S| + |T| - \frac{|S|^2}{m} - \frac{|T|^2}{n}. \quad (6)$$

Hence Equation (1) follows. Observing that $(|S| + |T|)\bar{d}_{ST} = 2|e(S, T)|$ we get inequality (2) of the Lemma.

In case $c = d = \Delta$, we get the following result of Zémor [17, Lemma 4].

Corollary 2.1. *Let G be a Δ -regular bipartite graph with vertex set $M \cup N$ where $|M| = |N| = n$. Let $S \subset M$ and $T \subset N$. Then the average degree \bar{d}_{ST} of the subgraph induced by $S \cup T$ satisfies:*

$$\bar{d}_{ST} \leq \frac{2\Delta}{n} \frac{|S||T|}{|S| + |T|} + \mu - \frac{\mu}{n} \left(\frac{|S|^2 + |T|^2}{|S| + |T|} \right).$$

3. Bound on the Minimum Distance of the Code

Theorem 3.1. *Suppose $d_1 \geq d_2 > \mu/2$. Then the minimum distance of the Tanner code \mathcal{C} is at least*

$$\frac{m}{d} \left\{ d_1 d_2 - \frac{\mu}{2} (d_1 + d_2) \right\}. \quad (7)$$

Proof. Let $\mathbf{x} \in \mathcal{C}$ be any codeword and $X = \{i \in E : \mathbf{x}_i = 1\}$ be its support. Also let, $S = \{u \in M : (u, v) \in X \text{ for some } v \in N\}$ and $T = \{u \in M : (u, v) \in X \text{ for some } v \in N\}$; where we recall that M and N are disjoint vertex sets of the bipartite graph G . Then $|e(S, T)| \geq |X|$. For \mathbf{x} to be a codeword, its restrictions at vertices S and T must be codewords of the codes C_1 and C_2 , respectively.

If $\frac{|X|}{|S|} < d_1$, then there will be at least one vertex $v \in S$ which will have less than d_1 non-zero bits and hence it will not be a codeword in C_1 . Therefore let us assume that $\frac{|X|}{|S|} \geq d_1$. That is, $|S| \leq \frac{|X|}{d_1}$.

Now the inequality (1) in Section 2 and $|S| \leq \frac{|X|}{d_1}$ implies that

$$|X| \leq |e(S, T)| \leq \frac{d|S||T|}{m} + \frac{\mu}{2}(|S| + |T|) \leq \left(\frac{|X|d}{d_1 m} + \frac{\mu}{2} \right) |T| + \frac{|X|\mu}{d_1 2}.$$

Hence

$$|T| \geq \frac{|X|(2d_1 - \mu)m}{\mu d_1 m + 2d|X|}.$$

Thus $|X|$ non-zero bits goes to at least $\frac{|X|(2d_1 - \mu)m}{\mu d_1 m + 2d|X|}$ vertices. Therefore the result follows as \mathbf{x} is a codeword implies that

$$|X| / \frac{|X|(2d_1 - \mu)m}{\mu d_1 m + 2d|X|} = \frac{\mu d_1 m + 2d|X|}{(2d_1 - \mu)m} \geq d_2.$$

Remark 3.2. Let G be a d -regular graph on n vertices. Then the edge vertex graph \mathcal{H}_G of G is a bipartite $(2, d)$ -regular graph. Lubotzky, Phillips and Sarnak [9] and independently Margulis [11] gave explicit constructions of infinite families of expander graphs with fixed regularity. The expander code families of Sipser and Spielman [13] which give asymptotically good error-correcting capabilities are based on edge-vertex graphs of the above families of expander graphs.

Let $[d, rd, \delta d]$ be the initial code C . Then Sipser and Spielman [13] showed that the relative minimum distance of $\mathcal{C}(\mathcal{H}_G, C)$ is at least $\left(\frac{\delta d - \mu}{d - \mu} \right)^2$. In [6], we showed that the relative minimum distance of $\mathcal{C}(\mathcal{H}_G, C)$ is at least $\delta \frac{\delta d - \mu}{d - \mu}$ whenever $\delta d > \mu(G)$. In certain cases, this bound gives quite an improvement as compared to the bound given in [13].

For $c = d$ and $C_1 = C_2$, our new bound is better than that of Sipser and Spielman [13] by a fraction of $\frac{\mu(1 + \frac{\mu}{d}\epsilon - 2\epsilon)}{d_1 - \mu}$. This gives an approximate improvement of $\frac{\mu}{d_1}$ since ϵ is small. However, it is weaker than the bound in Janwa and Lal [6] by a small fraction $\left(\frac{\mu}{d} \right)$, which is approximately equal to $\frac{2}{\sqrt{d}}$ for Ramanujan graphs. (The bound in the case $c = d$ and $C_1 = C_2$ was found independently by G. Zémor and A. Barg [18]. We thank a referee for pointing this out.)

4. The Decoding Algorithm

In this section, we present a decoding algorithm for Tanner codes. As we have mentioned in the introduction this algorithm is a generalization of a decoding algorithm of Zémor [17] (see also Zémor and Barg [18]). The codes Zémor considers are based on d -regular bipartite graphs. To generalize his algorithm to Tanner codes based on (c, d) -regular bipartite graphs, we need to overcome some technical difficulties and at the same time it gives us a lot of choice for choosing different types of codes and graphs.

Since G is a (c, d) -regular bipartite graph with vertex set $M \cup N$ with $|M| = m$ and $|N| = n$, we can partition the edge set as

$$E = \bigcup_{i=1}^m \{E_v : v \in M\} = \bigcup_{j=1}^n \{F_u : u \in N\},$$

where for each $v \in M$, the set E_v is the collection of c edges that are incident with the vertex v ; and for each $u \in N$, the set F_u is the collection of d edges that are incident with the vertex u .

Let $\mathbf{y} \in \mathbb{F}_2^{mc}$ be the received vector and let $d_1 \geq d_2$. The decoding algorithm consists of the following steps:

Step I

- Use the sets E_v for $v \in M$ to get the words $(\mathbf{y}_{v(1)}, \mathbf{y}_{v(2)}, \dots, \mathbf{y}_{v(c)})$ at each vertex.
- Use a complete decoding algorithm for the code C_1 to decode $(\mathbf{y}_{v(1)}, \mathbf{y}_{v(2)}, \dots, \mathbf{y}_{v(c)})$. Since the sets E_v are disjoint, this can be done in parallel for each $v \in M$.

This step may result in getting wrong codewords if the distance of the received word at some vertex $v \in M$ is $\geq \frac{d_1}{2}$. This iteration results in getting new bits for the edges and thus we get a new vector $\mathbf{z} \in \mathbb{F}_2^{mc}$.

Step II

- Look at the words formed at each F_u for $u \in N$ and decode the received word \mathbf{z} by applying in parallel a complete decoding algorithm for the code C_2 . This again results in a new vector $\mathbf{w} \in \mathbb{F}_2^{mc}$.

Step III

- **IF** $\mathbf{w} = \mathbf{z}$, **THEN** output \mathbf{w} as the decoded codeword.
ELSE, decode $(\mathbf{w}_{v(1)}, \mathbf{w}_{v(2)}, \dots, \mathbf{w}_{v(c)})$ by applying in parallel a complete decoding algorithm for the code C_1 to get \mathbf{v} .

Step IV

- **IF** $\mathbf{v} = \mathbf{w}$, **THEN** *output \mathbf{v} as the decoded codeword.*
 ELSE, *go to Step II with \mathbf{v} replaced by \mathbf{z} .*

We show in the next section that this algorithm converges to a correct codeword in $\mathcal{O}(\log L)$ parallel steps (where $L = mc$), provided the number of errors in the received word $\leq \frac{m}{d} \frac{d_1}{2} \cdot \left(\frac{d_2}{2} - \mu\right) = \frac{\delta_1}{2} \left(\frac{\delta_2}{2} - \frac{\mu}{d}\right) L$.

5. Analysis of the Decoding Algorithm

In this section, we first generalize a result of Zémor [17, cf. Lemma 5] that is crucial in determining the convergence of the decoding algorithm and then give a bound on the error-correction capability of the decoding algorithm and also its complexity.

Lemma 5.2. *For any $\eta > 0$, suppose $d_2 \geq (2 + \eta)\mu$. Let $S \subset M$ such that*

$$|S| \leq \frac{\alpha m}{d} \left(\frac{d_2}{2} - \mu\right)$$

where $\alpha < 1$. Let $T \subset N$ and suppose that there exists a set $Z \subset E$ of edges such that

- (1) Every edge of Z has one of its endpoints in S ;
- (2) Every vertex of T is incident to at least $\frac{d_2}{2}$ edges of Z .

Then

$$|T| < \beta |S|$$

where $\beta = \frac{1}{1 + \eta(1 - \alpha)} < 1$.

Proof. Since every vertex of T is incident to at least $\frac{d_2}{2}$ edges of Z and every edge of Z has one of its endpoints in S , Lemma 2.1 gives

$$2 \cdot |T| \frac{d_2}{2} \leq 2|e(S, T)| \leq \frac{2d|S| |T|}{m} + \mu(|S| + |T|).$$

Hence

$$|T| \left(d_2 - \mu - \frac{2d|S|}{m} \right) \leq \mu |S|. \quad (8)$$

But

$$\begin{aligned}
d_2 - \mu - \frac{2d|S|}{m} &\geq d_2 - \mu - \frac{2d}{m} \frac{\alpha m}{d} \left(\frac{d_2}{2} - \mu \right) \\
&= d_2(1 - \alpha) - \mu(1 - 2\alpha) \tag{9}
\end{aligned}$$

$$\begin{aligned}
&\geq (2 + \eta)\mu(1 - \alpha) - \mu(1 - 2\alpha) \\
&= \mu(1 + \eta(1 - \alpha)). \tag{10}
\end{aligned}$$

Hence the result follows from Equation (8)

We also need the following lemma, whose proof is similar to the proof of Lemma 5.2

Lemma 5.3. *For any $\eta > 0$, suppose $d_1 \geq (2 + \eta)\mu$. Let $T \subset N$, $\alpha < 1$ and $|T| \leq \frac{\alpha m}{d} \left(\frac{d_1}{2} - \mu \right) = \frac{\alpha n}{c} \left(\frac{d_1}{2} - \mu \right)$. Let $S \subset M$ and suppose that there exists a set $Z \subset E$ of edges such that*

- (i) *Every edge of Z has one of its endpoints in T ;*
- (ii) *Every vertex of S is incident to at least $\frac{d_1}{2}$ edges of Z .*

Then $|S| < \beta|T|$ where $\beta = \frac{1}{1 + \eta(1 - \alpha)} < 1$.

We now prove a bound on the error-correction capabilities of the decoding algorithm.

Theorem 5.2. *For $\eta > 0$, suppose $d_1 \geq d_2 \geq (2 + \eta)\mu$ and $L = mc$. If the weight of the error vector \mathbf{y} satisfies*

$$|\mathbf{y}| \leq \frac{\alpha m}{d} \cdot \frac{d_1}{2} \left(\frac{d_2}{2} - \mu \right) = \alpha \frac{\delta_1}{2} \left(\frac{\delta_2}{2} - \frac{\mu}{d} \right) \cdot L \tag{11}$$

for some $\alpha < 1$, then the algorithm of Section 4 converges to the initial codeword in $\mathcal{O}(\log L)$ parallel steps using a circuit of size $\mathcal{O}(L \log L)$.

Proof. Recall that the vertex set of the graph is $M \cup N$. Since the code is linear, without loss of generality, for analysis purpose, we can assume that the initial uncorrupted codeword is the zero codeword. Let \mathbf{y} be the error vector, which now equals the received vector, and let Y be the corresponding set of edges, i.e., $Y = \{i \in E : \mathbf{y}_i = 1\}$. Let \mathbf{z} be the vector obtained from \mathbf{y} after Step I of the decoding algorithm (induced at M), and let $Z = \{i \in E : \mathbf{z}_i = 1\}$ be the corresponding set of edges. Let \mathbf{w} be the vector obtained after Step II of the decoding algorithm (induced at N) and let W be the corresponding set of edges.

We look at the partitions of Y and Z induced by $\{E_v, v \in M\}$. If v is incident to fewer than $\frac{d_1}{2}$ edges of Y , then these will be totally corrected by the decoding procedure, i.e., $E_v \cap Z = \emptyset$. Let S_0 be the set of vertices $v \in M$ such that $E_v \cap Z \neq \emptyset$. Then

$$v \in S_0 \text{ implies } |E_v \cap Y| \geq \frac{d_1}{2}. \quad (12)$$

Similarly, let T_0 be the set of vertices $u \in N$ such that $E_u \cap W \neq \emptyset$. Then a similar argument gives

$$u \in T_0 \text{ implies } |E_u \cap Z| \geq \frac{d_2}{2}. \quad (13)$$

We now check that the sets S_0 , T_0 and Z just defined satisfy the conditions of Lemma 5.2 Equation (12) implies that $|Y| \geq |S_0| \frac{d_1}{2}$. Hence

$$|S_0| \leq \frac{2}{d_1} |Y| \leq \frac{2}{d_1} \frac{\alpha m d_1}{d} \frac{d_2}{2} \left(\frac{d_2}{2} - \mu \right) = \frac{\alpha m}{d} \left(\frac{d_2}{2} - \mu \right)$$

and hence by the definition of S_0 and using Equation (13), we see that the conditions of Lemma 5.2 are satisfied. Therefore, Lemma 5.2 implies

$$|T_0| \leq \beta |S_0| \text{ where } \beta < 1. \quad (14)$$

Consider Step II of the decoding algorithm. The condition $d_1 \geq d_2$ and Equation (14) gives

$$|T_0| \leq \beta |S_0| \leq |S_0| \leq \frac{\alpha m}{d} \left(\frac{d_2}{2} - \mu \right) \leq \frac{\alpha m}{d} \left(\frac{d_1}{2} - \mu \right).$$

Hence $|T_0|$ satisfies the condition of Lemma 5.3. If $\mathbf{w} = \mathbf{z}$, then the algorithm terminates. Otherwise $\mathbf{w} \neq \mathbf{z}$ and the decoding algorithm in Step III will give a vector \mathbf{v} with V being the corresponding set of edges.

Let S_1 be the set of vertices $v \in M$ such that $E_v \cap V \neq \emptyset$. Then a similar reasoning gives

$$v \in S_1 \text{ implies } |E_v \cap W| \geq \frac{d_1}{2}.$$

Thus the conditions of Lemma 5.3 are satisfied and hence

$$|S_1| \leq \beta |T_0| \leq \beta^2 |S_0|.$$

In the next step, we will get T_1 and so on. In general, for $i \geq 1$, we will have the recurrences

$$|T_i| \leq \beta |S_i|, \quad |S_{i+1}| \leq \beta |T_i|$$

with $\beta < 1$ and $|S_0| \leq \frac{\alpha m}{d} \left(\frac{d_2}{2} - \mu \right)$.

The proof of convergence now follows by repeating the arguments and noting that $\beta < 1$, implying that the sizes of S_i 's and T_i 's decrease at each step. Hence the decoding is complete when for some i either $S_i = \emptyset$ or $T_i = \emptyset$. Therefore, the algorithm terminates after t steps where $\beta^t |S_0| < 1$.

Thus an iteration of $\log_{\frac{1}{\beta}} \left[\alpha m c \cdot \frac{d_1}{2c} \left(\frac{d_2}{2d} - \frac{\mu}{d} \right) \right]$ parallel decoding rounds will correct $\alpha m c \cdot \frac{d_1}{2c} \left(\frac{d_2}{2d} - \frac{\mu}{d} \right)$ errors in Tanner codes. Since the codes that gave the check bits are fixed codes, this decoding algorithm can be implemented by a circuit of size $\mathcal{O}(L \log L)$ and depth $\mathcal{O}(\log L)$.

Remark 5.3. The decoding algorithm is able to correct a similar fraction of errors of the minimum distance bound (see Theorem 3.1) as that of Zémor for the case $c = d$ and $C_1 = C_2$.

Remark 5.4. Our algorithm in Section 4 can be generalized for expander codes obtained using a class of multi-partite graphs. An analysis of such an algorithm is in progress. Excellent multi-partite Ramanujan graphs exist (see, for example Janwa and Moreno [7]).

6. Applications and Asymptotic Results

There are many infinite families of (c, d) -regular bipartite graphs that are good expanders. Several of these families have $c \neq d$, and hence the techniques used by Sipser and Spielman or Zémor or Barg and Zémor are not applicable. For example, the incidence structures of the generalized N -gons define (c, d) -regular bipartite graphs that are excellent expanders as can be seen from their eigenvalues (see Tanner [15]).

Also, the infinite families of Tanner graphs corresponding to the incidence structures of Euclidean geometries $EG(m, 2^s)$ and projective geometries $PG(m, 2^s)$ provide us with example of (c, d) -regular bipartite graphs (see Kou, Lin, and Fossorier [8]). One of these families yield $(2^s, \frac{2^{ms}-1}{2^s-1} - 1)$ -regular bipartite Tanner graphs.

As observed by Tanner, one can tensor product these graphs with other Δ -regular good expander graphs to obtain $(c\Delta, d\Delta)$ -regular graphs that are good expander graphs in many cases.

For asymptotic results, let us take, the family of Ramanujan graphs $\{X^{p,q}\}$ of Lubotzky, Phillips, and Sarnak [9], where p, q varies over the prime pair $p, q \equiv 1 \pmod{4}$. We tensor this family with the $(p, p+2)$ -regular bipartite 4-gon, thus obtaining the family $\{Y^{p,q}\}$ of $(p(p+1), (p+2)(p+1))$ -regular bipartite graphs.

Theorem 6.3. *For all $\delta_1, \delta_2 > 0$ such that $1 - H(\delta_1) - H(\delta_2) > 0$, where $H(\cdot)$ is the binary entropy function, there exists a polynomial-time constructible family of expander codes of rate $1 - H(\delta_1) - H(\delta_2) > 0$ and minimum relative distance arbitrarily close to $\delta_1\delta_2$ in which any $\alpha < \frac{\delta_1\delta_2}{4}$ fraction of error can be corrected by a circuit of size $O(n \log n)$ and depth $O(\log n)$.*

Proof. By the Gilbert-Varshamov bound, we know that for all sufficiently large block lengths there always exists linear codes with minimum relative distance ϵ and rate $1 - H(\epsilon)$. We will use two such codes in our asymptotic construction, one with parameter $[c, (1 - H(\delta_1))c, \delta_1c]$ and the other with parameter $[d, (1 - H(\delta_2))d, \delta_2d]$.

We have $\mu(Y^{p,q}) = \sqrt{p(p+2)}\mu(X^{p,q}) \leq \sqrt{p(p+2)}2\sqrt{p}$. Therefore, so long as $\delta_1 p(p+1) \geq \delta_2 (p+2)(p+1) \geq (2 + \eta)[\sqrt{p(p+2)}][2\sqrt{p}] \geq$

$(2 + \eta)\mu(Y^{p^q})$, the conditions of Theorem 5.2 are satisfied, and therefore, the result follows.

Remark 6.5. The above conditions imply that $p \geq \frac{4(2 + \eta)^2}{\delta_2^2}$. In fact, we may replace p , a prime, with p^m by using a recent result of Morgenstern [12] on the existence of $(p^m + 1)$ -regular Ramanujan graphs.

Theorem 6.3 is a generalization of the theorem of Zémor [17] when $c = d$ and $C_1 = C_2$. Our theorem has some advantages over the results of Zémor [17] and the earlier result of Sipser and Spielman [13]. In particular,

- We need not confine ourselves to families of regular graphs.
- To show the existence of codes corresponding to Theorem 6.3, both [17] and [13] also choose the Ramanujan families we consider here. However, the sequence of codes by Zémor have lengths $q(q^2 - 1)$, with $q \equiv 1 \pmod{4}$ and $\left(\frac{p}{q}\right) = -1$. Therefore, for a fixed p , this sequence gets sparser and sparser, as q varies with the above conditions. However, we have a multitude of sequences of codes by combining the Ramanujan family with different n -gons, or other (c, d) -regular graphs.
- The articles [13] and [17] are confined to choosing initial codes of the type $[p + 1, (1 - H(\delta))(p + 1), \delta(p + 1)]$. However, here we may select codes of the type $[m, (1 - H(\delta))m, \delta m]$, where, as we have mentioned, we have many more choices of m . This gives us much more flexibility in choosing codes that might have good decoding algorithms. This helps us in decreasing the constants involved in the complexity of decoding.


Acknowledgements. We are thankful to G. Zémor [17] for making his paper available online. We would like to thank the reviewers whose constructive suggestions led to several improvements in the paper.

The second author would like to thank the Department of Mathematics and Computer Science, University of Puerto Rico, Rio Piedras Campus, San Juan, for their hospitality and a visiting appointment during August 2000 to July 2001, when the work was done.

References

1. Alon, N.: Eigenvalues and expanders. *Combinatorica*, **6**, 83–96 (1986)
2. Alon, N., Bruck, J., Naor, J., Naor, M., Roth, R.M.: Construction of asymptotically good low-rate error-correcting codes through pseudo-random graphs. *IEEE Trans. Inform. Theory*, **38**(2), 509–516 (1992)
3. Alon, N., Chung, F.R.K.: Explicit construction of linear sized tolerant networks. *Discrete Math.*, **72**, 15–19 (1988)
4. Elias, P.: Error-free coding. *IRE Trans. Inform. Theory*, vol. PGIT-4, pp. 29–37, Sept. 1954
5. Gallager, R.G.: *Low Density Parity Check Codes*, M.I.T. Press, 1963

6. Janwa, H.L., Lal, A.K.: On Expander Graphs: Parameters and Applications. submitted to Siam J. of Disc. Math., February 02, 2001
7. Janwa, H.L., Moreno, O.: Ramanujan Graphs, Codes, Exponential Sums, and Sequence, to be submitted to *IEEE Trans. Inform. Theory*
8. Kou, Y., Lin, S., Fossorier, M.P.C.: Low-Density Parity-Check Codes Based on Finite Geometries: A Rediscovery and New Results. *IEEE Trans. Inform. Theory*, **47**, 2711–2736 (2001)
9. Lubotzky, A., Phillips, R., Sarnak, P.: Ramanujan graphs. *Combinatorica*, **8**(3), 261–277 (1988)
10. Mackay, D.J.C.: Good Error-Correcting Codes based on very Sparse Matrices. *IEEE Trans. Inform. Theory*, **45**, 399–431 (1999)
11. Margulis, G.A.: Explicit group theoretical constructions of combinatorial schemes and their application to design of expanders and concentrators. *Probl. Inform. Transm.*, **24**(1), 39–46 (1988)
12. Morgenstern, M.: Existence and explicit constructions of $(q + 1)$ -regular Ramanujan graphs for every prime power q . *J. Comb. Theory, Ser. B*, **62**, 44–62 (1994)
13. Sipser, M., Spielman, D.A.: Expander Codes. *IEEE Trans. Inform. Theory*, **42**(6), 1710–1722 (1996)
14. Spielman, D.A.: Linear-time encodable and decodable error-correcting codes. *IEEE Trans. Inform. Theory*, **42**(6), 1723–1731 (1996)
15. Tanner, R.M.: Explicit concentrators from generalized n-gons. *SIAM J. Alg. Disc. Math.*, **5**(3), 287–293 (1984)
16. Tanner, R.M.: A Recursive Approach to Low Complexity Codes. *IEEE Trans. Inform. Theory*, **IT-27**(5), 533–547 (1981)
17. Zémor, G.: On Expander Codes. *IEEE Trans. Inform. Theory*, **IT-47**(2), 835–837 (2001)
18. Zémor, G., Barg, A.: Error exponents of expander codes, to appear in *IEEE Trans. Inform. Theory*

	200	0098	Despatch : 8/7/2002	Journal : AAECC
	Journal No.	Article No.	Author Received	No. of Pages : 13
	Disk Received : Yes		Disk Used : Yes	