# Generalized Minimum Distance Iterative Decoding of Tanner Codes

Saejoon Kim, *Member, IEEE*

***Abstract*— In this Letter, we present a linear-time decoding algorithm for Tanner codes that can correct errors up to close to half the minimum distance of the code. The algorithm is a simple generalization to the iterative decoding algorithm of expander codes proposed by Skachek and Roth [3].**

***Index Terms*— Expander codes, Tanner codes, generalized minimum distance decoding.**

## I. Introduction

EXPANDER codes have attracted attention in the coding community recently with their guaranteed fraction of correctable errors and efficient decoding algorithms. Decoding algorithms for expander codes have improved steadily, and the latest improvement is a generalized minimum distance (GMD) iterative decoding by Skachek and Roth [3] which can correct errors up to close to half the minimum distance of the code. Note that this is the largest number of errors that is guaranteed correctable by any decoding algorithm for a given minimum distance. In this Letter, we propose a simple generalization to Skachek and Roth's algorithm applicable to Tanner codes which are generalizations to expander codes. Our decoding algorithm can correct errors up to close to half the minimum distance of Tanner codes and outperforms Janwa and Lal's decoding algorithm [2].

## II. Preliminaries

Let $G = (V_1 \cup V_2, E)$ be a $\triangle$-regular bipartite graph between the sets of vertices $V_1$ and $V_2$ where $|V_1| = |V_2| = n$, and associate an error-correcting code $\tilde{\mathcal{C}}$ over $\mathbb{F}_2$ of length $\triangle$ and relative minimum distance $\delta = d/\triangle$ with each vertex in $G$. Assume an ordering on the edges $E$ from 1 to $N = \triangle n$, and let $E(v)$ be the set of edges incident on the vertex $v$. For a word $\mathbf{x} = (x_e)_{e \in E}$ of length $N$ whose entries are indexed by the ordering on $E$, let $(\mathbf{x})_v$ represent the part of $\mathbf{x}$ that is indexed by $E(v)$. Note that $\mathbf{x}$ can be partitioned into $n$ sub-words $\mathbf{x}_v$ for $v \in V_1$, or for $v \in V_2$. Let us define an error-correcting code of length $N$ over $\mathbb{F}_2$ from the graph $G$ and error-correcting code $\tilde{\mathcal{C}}$ by the set of codewords $\mathbf{c}$ such that $(\mathbf{c})_v$ is a codeword of $\tilde{\mathcal{C}}$ for each $v \in V_1 \cup V_2$.

If the second largest eigenvalue (in absolute value) of $G$, $\lambda$, is much smaller than $\triangle$, then $G$ is called an expander graph [5], and the corresponding code an *expander code*. Moreover, it has been shown that relative minimum distance of this code is greater than or equal to $\delta^2 - O(\frac{\lambda}{\triangle})$. Assume that $G$ is an expander graph.

Hereafter, let $\mathsf{d}(\cdot,\cdot)$ denote the Hamming distance, and $\mathsf{dec}$ denote the decoding that corrects errors up to less than half the minimum distance of a code. Furthermore let $\hat{\mathsf{dec}}$ denote the decoding that recovers any $e$ errors and $s$ erasures where $2e + s$ is less than the minimum distance of a code. We know that such $\mathsf{dec}$ and $\hat{\mathsf{dec}}$ always exist for any code.

Next theorem from [3] gives the error-correcting capability of the GMD iterative decoding for expander codes, where the number of iterations $m = \lfloor \log n / \log(2-\alpha) \rfloor + 1$. The value $\epsilon$ in the algorithm denotes a vector of erasures of length $\triangle$.

*Theorem 1:* Suppose that $d \geq 3\lambda$ and let

$$J_{exp}(\alpha) = \frac{1}{2}\alpha N\Big(\delta^2 - O(\frac{\lambda}{\triangle})\Big)$$

for some $0 < \alpha < 1$. Then the *GMD Iterative Decoding Algorithm for Expander Code* corrects up to $\lfloor J_{exp}(\alpha) \rfloor$ errors.

---

*GMD Iterative Decoding Algorithm for Expander Code*

For every $v \in V_1$, let $(\mathbf{x})_v = \mathsf{dec}((\mathbf{y})_v)$ for received word $\mathbf{y}$.
For $\theta = 1$ to $\lceil \frac{d}{2} \rceil$ do {
  Iteration 1: for every $v \in V_1$, let

$$(\mathbf{z})_v = \begin{cases} (\mathbf{x})_v & \text{if } \mathsf{d}((\mathbf{y})_v, (\mathbf{x})_v) < \theta \\ \epsilon & \text{otherwise} \end{cases}$$

  Iteration 2: for every $v \in V_2$, let $(\mathbf{z})_v = \hat{\mathsf{dec}}((\mathbf{z})_v)$.
  For $i = 3$ to $m$ do {
    Let $W$ stand for $V_1$ if $i$ odd, and for $V_2$ otherwise.
    Iteration $i$: for every $v \in W$, let $(\mathbf{z})_v = \mathsf{dec}((\mathbf{z})_v)$.
  }
  If $\mathsf{d}(\mathbf{y}, \mathbf{z}) \leq J_{exp}(\alpha)$ break.
}

---

The goal of this Letter is to generalize the above lower bound on the number of errors that can be corrected for codes generated from arbitrary $(\triangle_1, \triangle_2)$-regular bipartite graphs which have nodes all of degree $\triangle_1$ on one side and of degree $\triangle_2$ on the other side.

## III. Code Construction

Let $G = (V_1 \cup V_2, E)$ now be a $(\triangle_1, \triangle_2)$-regular bipartite graph between the sets of vertices $V_1$ and $V_2$ where $|V_1| = n_1$

Manuscript received January 24, 2005. The associate editor coordinating the review of this letter and approving it for publication was Prof. Jing Li.

S. Kim is with the Center for Information Technology, Yonsei University, Seoul, Korea (email: saejoon@yonsei.ac.kr).

Digital Object Identifier 10.1109/LCOMM.2005.08022.

and $|V_2| = n_2$. Let $\lambda$ be the second largest eigenvalue in absolute value of the graph. Throughout, we will be speaking of the described graph $G$ unless explicitly specified otherwise. A lower bound on the edges included in a subgraph of $G$ was proved in [2].

*Lemma 1:* For $G = (V_1 \cup V_2, E)$, $S \subset V_1$ and $T \subset V_2$, let $e(S,T)$ denote the number of edges between the sets $S$ and $T$. Then,

$$\Big||e(S,T)| - \frac{\triangle_2|S||T|}{n_1}\Big| \leq \frac{\lambda}{2}\Big(|S| + |T| - \frac{|S|^2}{n_1} - \frac{|T|^2}{n_2}\Big).$$

As before, assume an ordering on the edges $E$ from 1 to $N = \triangle_1 n_1 = \triangle_2 n_2$, and let $E(v)$ be the set of edges incident on the vertex $v$. Note that a word $\mathbf{x} = (x_e)_{e\in E}$ of length $N$ whose entries are indexed by the ordering on $E$ can be partitioned into $n_1$ sub-words $\mathbf{x}_v$ for $v \in V_1$, or into $n_2$ sub-words $\mathbf{x}_v$ for $v \in V_2$.

Now, let $\mathcal{C}_i$ be a linear $[\triangle_i, k_i = r_i\triangle_i, d_i = \delta_i\triangle_i]$ code over $\mathbb{F}_2$, $i = 1, 2$. Our error-correcting code $\mathcal{C}$ is explicitly constructed from the graph $G$ which can be constructed in polynomial-time, and constant block length error-correcting codes $\mathcal{C}_1$ and $\mathcal{C}_2$. It is a Tanner code by construction.

*Definition 1:* The Tanner code $\mathcal{C} = (G, \mathcal{C}_1, \mathcal{C}_2)$ is a linear code defined as

$$\mathcal{C} = \{\mathbf{c} \in \mathbb{F}_2^N : (\mathbf{c})_v \in \mathcal{C}_i,\ \forall v \in V_i, i = 1, 2\}$$

where $N = \triangle_1 n_1 = \triangle_2 n_2$.

The rate of the code is clearly at least $r_1 + r_2 - 1$ [4], and to get the best code in terms of rate and minimum distance, pick $\mathcal{C}_i$ that meets the Gilbert-Varshamov bound of $r_i = 1 - H(\delta_i)$ for $i = 1, 2$. The minimum relative distance of the code is shown to be at least

$$\Big(\delta_1\delta_2 - \frac{\lambda}{2}\big(\frac{\delta_1}{\triangle_2} + \frac{\delta_2}{\triangle_1}\big)\Big)$$

for $d_2 \geq d_1 \geq \lambda$ [2].

The following is a modified version of the GMD iterative decoding algorithm for expander codes that also decodes Tanner codes.

---

*GMD Iterative Decoding Algorithm for Tanner Code*

Replace $d$ and $J_{exp}(\alpha)$ by $d_1$ and $J(\alpha)$, respectively, in *GMD Iterative Decoding Algorithm for Expander Code*.

---

$J(\alpha)$ and the number of iteration $m$ for this decoding algorithm will be derived in the next section.

## IV. Decoding Analysis

Analysis of the decoding algorithm for Tanner codes is very similar to that for expander codes which is clear from construction. We shall henceforth refer the details to [3] in places where the analysis is very close. Assume, without loss of generality, that all-zero codeword $\mathbf{c}$ is sent and $\mathbf{y}$ is received.

For *Iteration* 1, let $\mathbf{z}(\theta) = (z_e)_{e\in E}$ for $\theta = 1, \cdots, \lceil\frac{d_1}{2}\rceil$ be the result of Iteration 1, and let

$$\begin{aligned} Z(\theta) &= \{e \in E : z_e \in \mathbb{F}_2, z_e \neq c_e\} \\ S(\theta) &= \{v_1 \in V_1 : E(v_1) \cap Z(\theta) \neq \emptyset\} \\ R(\theta) &= \{v_1 \in V_1 : z_e = ?, e \in E(v_1)\} \end{aligned}$$

where ? denotes the erasure symbol. We have the following lemma from [3] that we need in proving the result of this Letter.

*Lemma 2:* There exists a threshold $\theta \in \{1, 2, \cdots, \lceil\frac{d_1}{2}\rceil\}$ such that $2|S(\theta)| + |R(\theta)| \leq 2\mathsf{d}(\mathbf{y}, \mathbf{c})/d_1$.

For *Iteration* 2, let $\mathbf{w}(\theta) = (w_e)_{e\in E}$ for $\theta = 1, \cdots, \lceil\frac{d_1}{2}\rceil$ be the result of Iteration 2, and let

$$\begin{aligned} W(\theta) &= \{e \in E : w_e \neq c_e\} \\ T(\theta) &= \{v_2 \in V_2 : E(v_2) \cap W(\theta) \neq \emptyset\}. \end{aligned}$$

To get a bound on the size of $T(\theta)$, we make use of the *split graph* of $G$ as in [3]. The split graph of $G$, $G' = (V_1' \cup V_2', E')$, has, for $i = 1, 2$, corresponding to one vertex $v_i \in V_i$ in $G$, two vertices $v_i, v_i' \in V_i'$ in $G'$, and corresponding to one edge $e = (v_1, v_2)$ in $G$, four edges $e_1 = (v_1, v_2), e_2 = (v_1', v_2)$, $e_1' = (v_1, v_2')$, and $e_2' = (v_1', v_2')$ in $G'$. The second largest eigenvalue in absolute value of the split graph is known to be $2\lambda$.

Define a function $\phi : (\mathbb{F}_2 \cup ?)^N \to F^{4N}$ from $(x_e)_{e\in E}$ to $(x'_{e'})_{e'\in E'}$ such that if $x_e \neq ?$ then let $x'_{e'} = x_e$ for all $e' \in \{e_1, e_1', e_2, e_2'\}$, and else let $x'_{e'} = c_e$ for $e' \in \{e_1, e_1'\}$ and $x'_{e'} \neq c_e$ for $e' \in \{e_2, e_2'\}$. Writing $\phi(\mathbf{c}) = \mathbf{c}' = (c'_{e'})_{e'\in E'}$, $\phi(\mathbf{z}(\theta)) = \mathbf{z}'(\theta) = (z'_{e'})_{e'\in E'}$, and

$$\begin{aligned} Z'(\theta) &= \{e' \in E' : z'_{e'} \neq c'_{e'}\} \\ S'(\theta) &= \{v_1' \in V_1' : E'(v_1') \cap Z'(\theta) \neq \emptyset\}, \end{aligned}$$

Lemma 2 is converted to $|S'(\theta)| \leq 2\mathsf{d}(\mathbf{y}, \mathbf{c})/d_1$ for some $\theta \in \{1, 2, \cdots, \lceil\frac{d_1}{2}\rceil\}$. Now, let $\mathcal{C}'$ be a blocklength $2\triangle_2$ error-correcting code over $\mathbb{F}_2$ such that $\mathcal{C}' = \{(\mathbf{c}\ \mathbf{c}) : \mathbf{c} \in \mathcal{C}_2\}$, and let $\mathsf{dec}'$ be the decoder for $\mathcal{C}'$ that corrects up to less than $d_2$ errors. Define $\mathbf{w}'(\theta)$ such that $(\mathbf{w}'(\theta))_v = \mathsf{dec}'((\mathbf{z}'(\theta))_v)\ \forall\ v \in V_2'$, and let

$$\begin{aligned} W'(\theta) &= \{e' \in E' : w'_{e'} \neq c'_{e'}\} \\ T'(\theta) &= \{v_2' \in V_2' : E'(v_2) \cap W'(\theta) \neq \emptyset\}, \end{aligned}$$

Then, [3] shows that for every $\theta \in \{1, \cdots, \lceil\frac{d_2}{2}\rceil\}$, $2|T(\theta)| \leq |T'(\theta)|$. To get a relationship between the sizes of subsets $S'$ and $T'$, we have the next lemma.

*Lemma 3:* Suppose $d_2 \geq d_1 \geq 2\lambda$, and consider subsets $S' \subset V_1', T' \subset V_2'$, and $Z' \subset E'$ such that

- $|S'| \leq n_2\Big(\delta_1 - \frac{2\lambda}{\triangle_1}\Big)$
- every edge in $Z'$ has one of its endpoints in $S'$
- every vertex of $T'$ is incident to at least $d_2$ edges of $Z'$.

Then $|T'| \leq |S'|$.

**Proof:** Lemma 1 tells us that $d_{S'T'}$, the average degree of the subgraph induced by $S' \cup T'$, is

$$d_{S'T'} \leq 2\lambda + \frac{2\triangle_2|S'||T'|}{(|S'| + |T'|)n_1}.$$

However by the second and third conditions in the Lemma, $d_{S'T'} \geq \frac{2d_2|T'|}{|S'|+|T'|}$ which gives us

$$2d_2|T'| \leq 2\lambda(|S'| + |T'|) + \frac{2\triangle_2}{n_1}|S'||T'|.$$

By plugging in the value for $|S'|$ from the first condition in the Lemma, the proof is completed. □

Next lemma gives a bound on the size of $T(\theta)$ for some $\theta$ given a received word that differs from a codeword in a constant fraction of bits.

*Lemma 4:* Suppose that $d_2 \geq d_1 \geq 2\lambda$, $d_1 = \frac{d_2}{\eta}$, and $\mathsf{d}(\mathbf{y}, \mathbf{c}) \leq \frac{1}{2}\alpha N\delta_2(\delta_1 - \frac{2\lambda}{\triangle_1})$ for some $\alpha < \frac{1}{\eta}$. Then, there exists a threshold $\theta \in \{1, \cdots, \lceil \frac{d_1}{2} \rceil\}$ such that

$$|T(\theta)| \leq \alpha\eta n_2\Big(\frac{\delta_1}{2} - \frac{\lambda}{\triangle_1}\Big).$$

**Proof:** We know that there exists $\theta \in \{1, \cdots, \lceil \frac{d_1}{2} \rceil\}$ such that

$$\begin{aligned} |S'(\theta)| &\leq \frac{2\eta\mathsf{d}(\mathbf{y}, \mathbf{c})}{d_2} \\ &\leq \alpha\eta n_2(\delta_1 - \frac{2\lambda}{\triangle_1}). \end{aligned}$$

Hence by Lemma 3, $|T'(\theta)| \leq \alpha\eta n_2(\delta_1 - \frac{2\lambda}{\triangle_1})$. Noting that $2|T(\theta)| \leq |T'(\theta)|$ finishes the proof. □

For *Iterations* $3, \cdots, m$, we have the main result of this Letter.

*Theorem 2:* Suppose $d_2 \geq d_1 \geq (2+\zeta)\lambda$, $d_1 = \frac{d_2}{\eta}$, and let

$$J(\alpha) = \frac{1}{2}\alpha N\delta_2(\delta_1 - \frac{2\lambda}{\triangle_1})$$

for some $0 < \alpha < \frac{1}{\eta}$ and $\zeta > 0$. The the *GMD Iterative Decoding Algorithm for Tanner Code* corrects up to $\lfloor J(\alpha) \rfloor$ errors.

**Proof:** Iterations 1 and 2 follow from previous argument. Consider the end of Iteration 2 at which for $\mathsf{d}(\mathbf{y}, \mathbf{c}) \leq J(\alpha)$, we have $|T(\theta)| \leq \alpha\eta n_2(\frac{\delta_1}{2} - \frac{\lambda}{\triangle_1})$ for some $\theta$ by Lemma 4. Assume this value of $\theta$ in the subsequent. Denote $\mathbf{z}^{(l)} = (z_e^{(l)})_{e \in E}$ to be the result of Iteration $l$, and let $Z^{(l)} = \{e \in E : z_e^{(l)} \in \mathbb{F}_2, z_e^{(l)} \neq c_e\}$ for $l \geq 3$.

Consider $S^{(3)} \subset V_1$ where $E(v) \cap Z^{(3)} \neq \emptyset$ for $v \in S^{(3)}$. Then by Lemma 5.3 in [2], $|S^{(3)}| \leq \frac{1}{1+\zeta(1-\alpha\eta)}|T(\theta)|$. But

$$|S^{(3)}| < |T(\theta)| \leq \alpha\eta\frac{n_2}{\triangle_1}(\frac{d_1}{2} - \lambda) \leq \alpha\eta\frac{n_2}{\triangle_1}(\frac{d_2}{2} - \lambda).$$

Therefore for $T^{(4)} \subset V_2$ where $E(v) \cap Z^{(4)} \neq \emptyset$ for $v \in T^{(4)}$, by Lemma 5.2 in [2], we know that $|T^{(4)}| \leq \frac{1}{1+\zeta(1-\alpha\eta)}|S^{(3)}|$. Repeating this procedure, for $S^{(l)}$ and $T^{(l)}$ where $E(v) \cap Z^{(l)} \neq \emptyset$ for $v \in S^{(l)}$ and $l$ odd, and $v \in T^{(l)}$ and $l$ even, we obtain $|S^{(3)}| > |T^{(4)}| > |S^{(5)}| > \cdots$. As a result, after at most

$$\log_{1+\zeta(1-\alpha\eta)} \frac{1}{2}\alpha N\delta_2(\delta_1 - \frac{2\lambda}{\triangle_1})$$

iterations, up to $\lfloor J(\alpha) \rfloor$ errors are corrected. □

As mentioned in [2], there exist many infinite families of excellent $(\triangle_1, \triangle_2)$-regular bipartite expanders that can be explicitly constructed in polynomial-time. Using such bipartite graphs and codes $\mathcal{C}_i$, $i = 1, 2$, that meet the Gilbert-Varshamov bound, next theorem is established.

*Theorem 3:* For all $\delta_1, \delta_2 > 0$, $\delta_2 = \eta\delta_1\frac{\triangle_1}{\triangle_2}$, $\eta \geq 1$, such that $1 - H(\delta_1) - H(\delta_2) > 0$, there exists an infinite family of polynomial-time explicitly constructible Tanner codes of rate $1 - H(\delta_1) - H(\delta_2)$ and minimum relative distance arbitrarily close to $\delta_1\delta_2$ in which any $\frac{\alpha}{2}\delta_1\delta_2$ fraction of errors can be corrected for any $\alpha < \frac{1}{\eta}$ by a circuit of size $O(N \log N)$ and logarithmic depth.

As noted in [1], *GMD Iterative Decoding Algorithm for Tanner Code* can be trivially modified to yield a sequential decoding algorithm with complexity $O(N)$.

## V. Conclusion

In this Letter, we presented a natural generalization to Skachek and Roth's decoding algorithm applicable to Tanner codes. The generalization can correct errors up to close to half the minimum distance of the code in logarithmic time using a linear number of processors, or in linear time.

## References


[1] A. Barg and G. Zémor, "Error exponents of expander codes," *IEEE Trans. Inform. Theory*, vol. 48, pp. 1725-1729, June 2002.

[2] H. Janwa and A. K. Lal, "On Tanner codes: minimum distance and decoding," *AAECC*, 13, 2002.

[3] V. Skachek and R. M. Roth, "Generalized minimum distance iterative decoding of expander codes," in *Proc. Information Theory Workshop*, 2003.

[4] R. M. Tanner, "A recursive approach to low complexity codes," *IEEE Trans. Inform. Theory*, vol. 27, pp. 533-547, Sept. 1981.

[5] R. M. Tanner, "Explicit construction of concentrators from generalized N-gons," *SIAM J. Alg. Disc. Meth.*, vol. 5, pp. 287-293, Sept. 1984.

[6] G. Zémor, "On expander codes," *IEEE Trans. Inform. Theory*, vol. 47, pp. 835-837, Feb. 2001.