**ENEE 739C: Advanced Topics in Signal Processing: Coding Theory**
**Instructor: Alexander Barg**

Lecture 10 (**draft;** 11/12/03). Codes on graphs (continued): bipartite-graph codes. Attaining capacity with linear complexity.

http://www.enee.umd.edu/~abarg/ENEE739C/course.html

Recall that serially concatenated codes achieve capacity of a DMC under a quadratic-time decoding complexity. In this part we will explore another way of parallel-concatenating codes which will enable us to achieve capacity of a BSC with complexity $O(N)$. We develop a link between the expansion property of the Tanner graph and convergence of iterative decoding[1].

To develop some proofs, we will follow papers [5, 2]. Please **print** them out and **use** as lecture notes.

**Definition 1.** *The adjacency matrix of a graph $G = (V, E)$ is an $n \times n$ matrix $A$ in which $a_{i,j} = \chi\{(v_i, v_j) \in E\}$, i.e.,*

$$a_{ij} = \begin{cases} 1 & v_i \text{ and } v_j \text{ are adjacent} \\ 0 & \text{not adjacent} \end{cases}$$

*Observe that $A$ is symmetric.*

We consider bipartite, balanced $\Delta$-regular graphs $G = (V_0 \cup V_1, E)$. Here "balanced" means that $|V_0| = |V_1| = n$, $\Delta$-regular means that $\deg(v) = \Delta$ for every vertex $v$ of the graph. The adjacency matrix has the form

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

where $\mathbf{M}$ is the $n \times n$ matrix whose rows correspond to $V_0$ and columns to $V_1$. Let us fix an arbitrary ordering of the edges of $G$. We wish to define a code of length $N = n\Delta$ whose coordinates are in one-to-one correspondence with the edges in $E$.

Let $\mathcal{A}, \mathcal{B}$ be linear binary $[\Delta, R_0\Delta, d_0 = \delta_0\Delta]$ and $[\Delta, R_1\Delta, d_1 = \delta_1\Delta]$ codes. Let $\mathbf{x}$ be a binary $N$-vector. Its projection on the neighborhood of a vertex $v$, i.e., a $\Delta$-vector corresponding to the edges in $\mathcal{N}(v)$, will be denoted by $\mathbf{x}_v$.

**Definition 2.** *A* bipartite-graph (BG) *code $\mathcal{C} = \mathcal{C}(G; \mathcal{A}, \mathcal{B})$ is a set of $N$-vectors satisfying*

$$\left\{ \mathbf{x} \in \mathscr{H}_2^N : \begin{array}{l} \forall_{v \in V_0}\ \mathbf{x}_v \in \mathcal{A} \\ \forall_{w \in V_1}\ \mathbf{x}_w \in \mathcal{B} \end{array} \right\}$$

In later analysis we will assume that $n \to \infty, \Delta = \text{const}$.

Ramanujan graphs. First we convince ourselves that the largest eigenvalue of $A$ is $\Delta$. (In fact $-\Delta \le \lambda_i \le \Delta$, where $\lambda_i$ is any eigenvalue). We will assume that all the remaining eigenvalues are small compared to this, in particular let $\lambda$ be the second largest eigenvalue of $A$, then we assume that

$$\lambda \le 2\sqrt{\Delta - 1}.$$

Graphs with such properties are known to exist and are easliy constructible.

code parameters We need two results on the properties of Ramanujan graphs relevant in our context. Consider a subgraph $G_{ST} \subset G$ defined by the subsets of vertices $S \subset V_0$ and $T \subset V_1$ and all the edges with one end in $S$ and the other in $T$. The average degree $\bar{d}_{ST}$ of a vertex in $G_{ST}$ is the number of these edges divided by $|S| + |T|$.

**Lemma 1.** *Let $S \subset V_0$ and $T \subset V_1$. Then the average degree $\bar{d}_{ST}$ of the subgraph induced by $S \cup T$ satisfies*

$$\bar{d}_{ST} \le \frac{2|S||T|}{|S| + |T|} \frac{\Delta}{n} + \lambda.$$

*Proof:* see Lemma 4 in [5].

[1] the decoding algorithm will be closer in spirit to Flipping Algorithm of the last lecture than to any of the message passing decoders.

**Lemma 2.** *Let $|S| \subset V_0$ with $|S| = s = \sigma n$. Let $\alpha > \alpha_\sigma$, where $\alpha_\sigma = \frac{\lambda}{2\sigma\Delta}$. Let $T$ be the subset of $V_1$ defined by $T = \{w \in V_1, d_S(w) \geq (1+\alpha)\sigma\Delta\}$. Then*

$$|T| \leq \frac{\alpha}{\alpha - \alpha_\sigma}|S|.$$

*Proof :* The number of edges of the graph $G_{ST}$ is at least $|T|(1+\alpha)\sigma\Delta$, and therefore the average degree $\bar{d}_{ST}$ satisfies

$$\frac{2|T|(1+\alpha)\sigma\Delta}{|S|+|T|}.$$

By the previous lemma

$$\frac{2|T|(1+\alpha)\sigma\Delta}{|S|+|T|} \leq \frac{2|S||T|}{|S|+|T|}\frac{\Delta}{n} + \lambda,$$

which implies the claimed inequality. ▮

**Theorem 3.** *$\mathcal{C}$ is an $[N = n\Delta, NR, D = N\delta]$ code where*

$$R \geq R_0 + R_1 - 1$$

$$\delta \geq \delta_0\delta_1\Big(1 - \frac{\lambda}{d_0}\Big)\Big(1 - \frac{\lambda}{2d_1}\Big)$$

*Proof :* Let us compute the number of linear restrictions that a codeword should satisfy:

$$N(1-R) \leq n\Delta(1-R_0) + n\Delta(1-R_1).$$

This implies the inequality for the rate $R$.

To prove the statement about the distance, let $\mathbf{c} = (c_1, \ldots, c_N)$ be a nonzero codeword of $\mathcal{C}$. Let $S$ and $T$ be respectively the set of left and right vertices where the subvector $\mathbf{c}_v \neq 0$.

Consider the subgraph $G_1 \subset G$ induced by $S$ and $T$. By definition of $\mathcal{C}$, every vertex of $S$ has degree at least $d_0$ and every vertex of $T$ has degree at least $d_1$ in $G_1$, so that Lemma 1 implies

$$\begin{aligned} \frac{d_0|S| + d_1|T|}{|S|+|T|} &\leq \frac{2|S||T|}{|S|+|T|}\frac{\Delta}{n} + \lambda \\ d_0|S| + d_1|T| &\leq 2|S||T|\frac{\Delta}{n} + \lambda(|S|+|T|) \end{aligned}$$

which is rewritten as

$$\frac{d_0 - \lambda}{|T|} + \frac{d_1 - \lambda}{|S|} \leq 2\frac{\Delta}{n} \tag{1}$$

and because $d_0 - \lambda$ and $d_1 - \lambda$ are assumed to be positive quantities, (1) yields :

$$|S| \geq \frac{d_1 - \lambda}{2\Delta}n \tag{2}$$

$$|T| \geq \frac{d_0 - \lambda}{2\Delta}n. \tag{3}$$

Let $\sigma = |S|/n$, let $\alpha = d_1/\sigma\Delta - 1$, and let $\alpha_\sigma = \lambda/2\sigma\Delta$. If $\alpha \leq \alpha_\sigma$, then $(d_1 - \lambda/2) \leq \sigma\Delta$ which means

$$\delta_1\left(1 - \frac{\lambda}{2d_1}\right)n \leq |S|$$

and $D \geq |S|d_0$ implies

$$D \geq \delta_0\delta_1\left(1 - \frac{\lambda}{2d_1}\right)N$$

which proves (slightly better than) the theorem. We may therefore assume that $\alpha > \alpha_\sigma$ and Lemma 2 applies and gives

$$\begin{aligned} |T| &\leq \frac{\lambda/2\sigma\Delta}{d_1/\sigma\Delta - 1 - \lambda/2\sigma\Delta}|S| \\ |T| &\leq \frac{\lambda}{2d_1 - 2\sigma\Delta - \lambda}|S| \end{aligned}$$

which together with (3) yields

$$\frac{d_0-\lambda}{2\Delta}n \le \frac{\lambda}{2d_1-2\sigma\Delta-\lambda}|S|$$

and

$$\begin{aligned}\frac{d_0-\lambda}{\Delta}(d_1-\sigma\Delta-\lambda/2)n &\le \lambda|S|\\ \frac{d_0-\lambda}{\Delta}(d_1-\lambda/2)n &\le d_0|S|\\ \Big(1-\frac{\lambda}{d_0}\Big)\Big(1-\frac{\lambda}{2d_1}\Big)\delta_1 n &\le |S|.\end{aligned}$$

Together with $D \ge |S|d_0$ this proves the theorem. ▮

As usual, we will call $\delta N$ the *designed distance* of the code $\mathcal{C}$. Since $\lambda \approx \sqrt{\Delta}$ and both $d_0$ and $d_1$ can be chosen proportional to $\Delta$, this theorem enables us to conclude that the designed distance of BG codes *approaches the product bound* on the minimum distance. This result parallels the corresponding part[2] of Prop. 8.6, although its proof (taken from [1]) is substantially harder. Note that the construction complexity of thw code $\mathcal{C}$ is proportional to the complexity of constructing the graph $G$.

Decoding. The code $\mathcal{C}$ can be decoded in iterations as follows. Let $\mathbf{y}$ be a vector received from the BSC. Define the *left decoding round* $L(\mathbf{y})$ as parallel decoding of all the vectors $\mathbf{y}_v, v \in V_0$ with the code $\mathcal{A}$. Since this code is of constant length, we will assume that each of the projections is decoded by max likelihood. The *right decoding round* $R(\mathbf{y})$ is defined analogously, replacing $V_0$ with $V_1$ and the code $\mathcal{A}$ with $\mathcal{B}$. Iterative decoding proceeds as follows:

$$\dots R(L(R(L(\mathbf{y})))) \dots$$

Call this procedure *Expander Decoding.*

Properties of expander decoding can be summarized as follows.

**Theorem 4.** [5] *Consider an expander code $\mathcal{C}(G,\mathcal{A},\mathcal{A})$, where $\mathcal{A}$ is a $[\Delta, R_0\Delta, d_0=\delta_0\Delta]$ code. Suppose that $d_0 \ge 3\lambda$. If the weight of the error vector $\mathbf{y}$ satisfies*

$$\mathrm{wt}(\mathbf{y}) \le \alpha\frac{\delta_0}{2}\Big(\frac{\delta_0}{2}-\frac{\lambda}{\Delta}\Big)N$$

*for any $\alpha<1$, then expander decoding converges to the all-zero codeword in $O(\log n)$ decoding rounds. The algorithm can be implemented as a sequential procedure of complexity $O(N)$.*

It is interesting to note that the GMD technique has been successfully applied to expander codes [4] to show a possibility to correct half the product bound fraction of errors (as opposed to $\frac{1}{4}$ in this theorem).

**Theorem 5.** [2] *We assume transmission over a BSC($p$). For a given rate $R$, any $\epsilon>0$ and any $\alpha<1$ there exists an $[N, NR]$ expander code $\mathcal{C}$ such that the error probability of expander decoding $P_e(\mathcal{C},p) \le 2^{-\alpha N f(R,p)}$, where*

$$f(R,p) = \max_{R \le R_0 \le \mathscr{C}} \frac{1}{2}E_0(R_0,p)h_2^{-1}(R_0-R).$$

*The decoding complexity has the same growth order as in the previous theorem.*

This theorem gives a family of codes that reach capacity of a BSC (and provide an exponential decline of the error probability) under linear-complexity decoding (recall that the best result for serial concatenations is quadratic-time complecity). The results proved in this lecture fall below the Forney and Zyablov bounds of concatenated codes. It is possible to modify the code construction and the decoding procedure so that expander codes com arbitrarily close both bounds. Moreover, it is possible to correct (close to) $\frac{1}{2}\delta_{\mathrm{Z}}(R)$ proportion of errors under an $O(N)$ expander decoding algorithm. The proofs become substantially harder (see [1]).

[2] Note that the rate of $\mathcal{C}$ is below the product bound $R_0R_1$.

The error exponent results of this lecture have the same deficiency as the corresponding results of Lecture 8, namely as the rate approaches capacity (of a BSC), the constant factor in the complexity estimate grows exponentially in the gap $\epsilon = \mathscr{C} - R$.

*Comments:* Several ideas converge at this construction. First, LDPC codes viewed "locally" in the neighborhood of every vertex of the Tanner graph employ weak codes, a repetition code on the "variable" side and a single parity-check code on the "constraints" side. We have replaced these codes with more powerful local codes. The other ingredient is the link between the convergence of iterative decoding and spectral properties suggested in [3].

REPLICATED BG CODES. We briefly discuss a way to improve the parameters of the above construction. For that let us assume that every edge corresponds to $t \geq 1$ bits of the transmission, where $t$ is a constant. See the corresponding part of [2].

ASYMPTOTIC PROPERTIES OF AN ENSEMBLE OF REGULAR-GRAPH CODES Consider the ensemble of random BG codes defined as follows. Suppose the left and right codes are the same $[\Delta, R_0\Delta]$ binary linear code $\mathcal{A}$. Let $\mathbf{M}$ be a $\Delta(1-R_0)\times\Delta$ parity-check matrix of $\mathcal{A}$. The parity-check matrix of the code $C$ can be written as follows: $\mathbf{H} = [\mathbf{H}_1, \mathbf{H}_2]^t$, where

$$\mathbf{H}_1 = \begin{bmatrix} \mathbf{M} & & & \\ & \mathbf{M} & & \\ & & \cdots & \\ & & & \mathbf{M} \end{bmatrix}$$

(a band matrix with $\mathbf{M}$ repeated $n$ times) and $\mathbf{H}_2 = \pi(\mathbf{H}_1)$ is a permutation of the columns of $\mathcal{H}_1$ defined by the edges of the graph $G$.

To form an ensemble of random bipartite-graph codes assume that $\mathbf{M}$ is a random binary matrix with uniform distribution and that the permutation $\pi$ is chosen with uniform distribution from the set of all permutation on $N = n\Delta$ elements. Choose the uniform probability on $Z_N = \{0,1\}^N$ and endow the product space of couples $(\mathcal{H}, \mathbf{x})$ with the product probability.

Our goal is to prove that if $d_0 \geq 3$, the ensemble of bipartite-graph codes contains asymptotically good codes (and sometimes codes that meet the GV bound). Recall from the previous lecture that the same is true for the LDPC codes as the number of ones in the columns (rows) of $\mathbf{H}$ is allowed to grow.

**Theorem 6.** *Consider the ensemble of random BG codes of length $N = n\Delta$ and rate $R$. For $n \to infty$ The expected weight distribution is $\mathsf{E}[A_{\omega N}] \leq 2^{NF+o(N)}$, where*

$$F = \omega[R - 1 - 2\log(1-2^{R_0-1})] - h_2(\omega) \qquad \text{if } 0 < \omega \leq 1 - 2^{R_0-1} \tag{4}$$

$$F = h_2(\omega) + R - 1 \qquad \text{if } \omega \geq 1 - 2^{R_0-1} \tag{5}$$

*Proof :* The average number of codewords of weight $w$ is

$$A_w = \binom{N}{w} \Pr[\mathbf{H}\mathbf{x}^t = 0 \mid \operatorname{wt}(\mathbf{x}) = w]. \tag{6}$$

Let us compute the probability $\Pr[\mathbf{H}\mathbf{x}^t = 0 \mid \operatorname{wt}(\mathbf{x}) = w]$.

Observe that

$$\begin{aligned} \Pr[\mathbf{H}\mathbf{x}^t = 0 \mid \operatorname{wt}(\mathbf{x}) = w] &= \Pr[\mathbf{H}_1\mathbf{x}^t = 0 \mid \operatorname{wt}(\mathbf{x}) = w]\Pr[\mathbf{H}_2\mathbf{x}^t = 0 \mid \operatorname{wt}(\mathbf{x}) = w] \\ &= (\Pr[\mathbf{H}_1\mathbf{x}^t = 0 \mid \operatorname{wt}(\mathbf{x}) = w])^2. \end{aligned} \tag{7}$$

Let $w = \omega n\Delta$. Let $\mathcal{X}_{m,w} \subset \{0,1\}^N$ be the event where $\mathbf{x}$ is of weight $w$ and contains nonzero entries in exactly $m$ groups of coordinates of the form $(x_{i\Delta+j}, j = 1,\dots,\Delta; i = 0,\dots,n-1)$. Let $w_i = \omega_i\Delta$ be the number of ones in the $i$th group. We have

$$\Pr_{Z_N}[\mathcal{X}_{m,w}] = 2^{-N}\binom{n}{m}\sum_{\sum w_i = w}\prod_{i=1}^{m}\binom{\Delta}{w_i} \cong 2^{-N}\binom{n}{m}\sum_{\sum w_i = w} 2^{\Delta\sum_i h_2(\omega_i)}$$

By convexity of the entropy function (or by using Lagrange multipliers), the maximum of the last expression on $\omega_1, \dots, \omega_m$ under the restriction $\sum_i \omega_i = \omega n$ is attained when $\omega_i = \omega n/m, i = 1, \dots, m$. For large $\Delta$ we therefore have

$$\Pr_{Z_N}[\mathcal{X}_{m,w}] \cong 2^{-N+m\Delta h_2(\omega n/m)}$$

Now we have

$$\begin{aligned}
\Pr[\mathbf{H}_1\mathbf{x}^t = 0 \mid \operatorname{wt}(\mathbf{x}) = w] &= \frac{\Pr[\mathbf{H}_1\mathbf{x}^t = 0\ ,\ \operatorname{wt}(\mathbf{x}) = w]}{\Pr[\operatorname{wt}(\mathbf{x}) = w]} \qquad \text{and} \\
\Pr[\mathbf{H}_1\mathbf{x}^t = 0\ ,\ \operatorname{wt}(\mathbf{x}) = w] &= \sum_m \Pr[\mathbf{H}_1\mathbf{x}^t = 0\ , \mathcal{X}_{m,w}]
\end{aligned}$$

and clearly

$$\Pr[\mathbf{H}_1\mathbf{x}^t = 0\ ,\ \mathcal{X}_{m,w}] = 2^{m\Delta(R_0-1)} \Pr[\mathcal{X}_{m,w}]$$

so that

$$\Pr[\mathbf{H}_1\mathbf{x}^t = 0\ ,\ \operatorname{wt}(\mathbf{x}) = w] \cong 2^{-N+\max_m(m\Delta(R_0-1)+m\Delta h_2(\omega n/m))}$$

and

$$\Pr[\mathbf{H}_1\mathbf{x}^t = 0 \mid \operatorname{wt}(\mathbf{x}) = w] \cong 2^{-h_2(\omega)N+\max_m(m\Delta(R_0-1)+m\Delta h_2(\omega n/m))}.$$

Given (6) and (7), and setting $x = m/n$ we obtain therefore $A_w = 2^{NF(R_0,x)}$, where

$$\begin{aligned}
F(R_0, x) &= -h_2(\omega) + 2 \max_{\omega \le x \le 1}(x(R_0 - 1 + h_2(\omega/x))) + o(1) \\
&\le -h_2(\omega) + \max_{\omega \le x \le 1}(x(R - 1 + 2h_2(\omega/x))) + o(1) \qquad (8)
\end{aligned}$$

The unconstrained maximum on $x$ in the last expression is attained for $x = x_0 = \omega/(1-z)$, where $2\log z = R-1$. Thus, the optimizing value of $x$ equals $x_0$ if this quantity is less that 1 and 1 otherwise. Substituting $x = x_0$ into (8) and taking into account the equality $R - 1 + 2h_2(z) = 2(1-z)\log(z/(1-z))$, we obtain

$$F(R_0, x) \le -h_2(\omega) + \frac{\omega}{1-z}(R - 1 + 2h_2(z))$$

which is exactly (4). Substituting $x = 1$ we obtain the second part of the claim. ▮

This result enables us to draw conclusions about the average minimum distance of codes in the ensemble. From (4)-(5) it is clear that the relation between these expressions is

$$\omega[R - 1 - 2\log(1 - 2^{R_0-1})] - h(\omega) \ge h(\omega) + R - 1$$

(since $x_0$ in the previous proof is the only maximum point), and that this inequality is strict for $\omega < 1 - 2^{R_0-1}$ and turns into an equality for greater values of $\omega$. Thus if $1 - 2^{R_0-1} < \delta_{\mathrm{GV}}(R)$, the first time the exponent of the ensemble average weight spectrum becomes positive is $\omega = \delta_{\mathrm{GV}}(R)$. This would mean that for large $n$ there exist codes in the bipartite-graph ensemble that approach the GV bound; however, there is one obstacle for this conclusion. What is that? We will discuss the problem and skip the proof of the following theorem.

**Theorem 7.** *Consider the random ensemble of BG codes. Let $\omega^*$ be the only nonzero root of the equation*

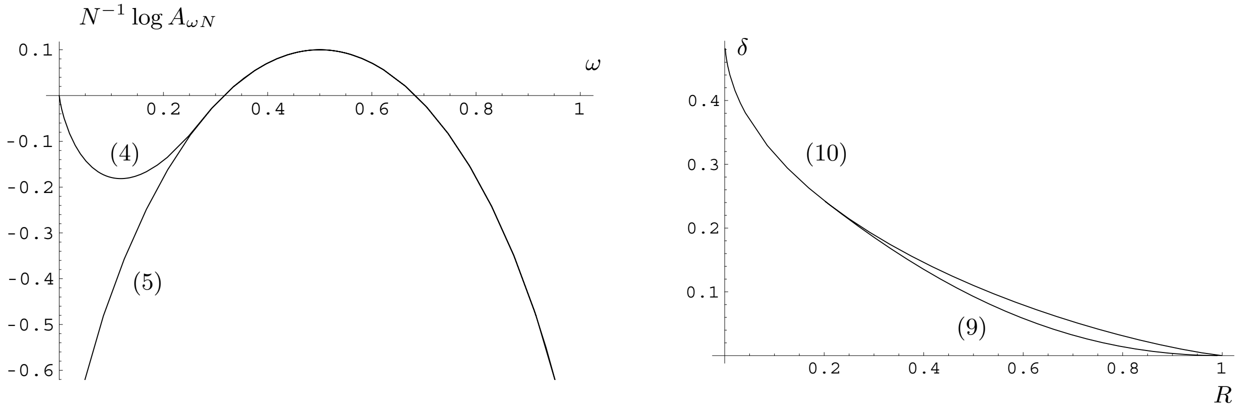
$$\omega\left(R - 1 - 2\log(1 - 2^{(1/2)(R-1)})\right) = h_2(\omega).$$

*The ensemble average relative distance behaves as*

$$\delta(R) = \omega^* \qquad \textit{if } R_0 \le \log(2(1 - \delta_{\mathrm{GV}}(R))) \qquad (9)$$

$$\delta(R) = \delta_{\mathrm{GV}}(R) \qquad \textit{if } R_0 > \log(2(1 - \delta_{\mathrm{GV}}(R))) \qquad (10)$$

*In particular, for $R \le 0.202$ the ensemble contains codes that meet the GV bound.*



(a), $R = 0.1$

(b)

Figure 1. Average weight spectrum (a) and distance (b) of the ensemble of bipartite-graph codes)

## References

1. A. Barg and G. Zémor, *Concatenated codes: Serial and parallel*, manuscript (2003).
2. ______, *Error exponents of expander codes*, IEEE Trans. Inform. Theory **48** (2002), no. 6, 1725–1729.
3. M. Sipser and D. A. Spielman, *Expander codes*, IEEE Trans. Inform. Theory **42** (1996), no. 6, 1710–1722.
4. V. Skachek and R. Roth, *Generalized minimum distance decoding of expander codes*, Proc. IEEE Information Theory Workshop, Paris, France, March 2003.
5. G. Zémor, *On expander codes*, IEEE Trans. Inform. Theory **47** (2001), no. 2, 835–837.