DIMACS Technical Report 2005-32
August 2005

Decoding of Expander Codes at Rates Close to Capacity

by

Alexei Ashikhmin [1]
Mathematics of Communicatons Dept.
Bell Laboratories, Lucent Technologies
600 Mountain Av., Rm. 2C-355, Murray Hill, NJ 07974, USA
e-mail: aea@research.bell-labs.com

Vitaly Skachek [2]
Computer Science Department
Technion — Israel Institute of Technology
Haifa 32000, Israel
e-mail: vitalys@cs.technion.ac.il

[1]Permanent Member

[2]The work of this author was done in part when he was visiting DIMACS/Bell Laboratories in July-August 2004. The support of DIMACS is deeply acknowledged.

DIMACS is a collaborative project of Rutgers University, Princeton University, AT&T Labs–Research, Bell Labs, NEC Laboratories America and Telcordia Technologies, as well as affiliate members Avaya Labs, HP Labs, IBM Research, Microsoft Research, Stevens Institute of Technology, Georgia Institute of Technology and Rensselaer Polytechnic Institute. DIMACS was founded as an NSF Science and Technology Center.

ABSTRACT

The concatenation of nearly-MDS expander codes of Roth and Skachek, "On Nearly-MDS Expander Codes," *Proc. IEEE ISIT'04,* with 'typical' LDPC codes is investigated. It is shown that for the rates $R = (1 - \varepsilon)C$ (C is the capacity of the binary symmetric channel (BSC)), under certain condition on the parameters of LDPC codes, these concatenated codes have decoding time linear in their length and polynomial in $1/\varepsilon$, and the decoding error probability decays exponentially. These codes are compared to the recently presented codes of Barg and Zémor, "Error Exponents of Expander Codes," *IEEE Trans. Inform. Theory,* 2002, and "Concatenated Codes: Serial and Parallel," *IEEE Trans. Inform. Theory,* 2005. It is shown that the latter families can not be tuned to have all the aforementioned properties.

1 Introduction

It is known that several families of LDPC codes [5] can attain the capacity of the binary erasure channel (BEC) [10], [13]. It is generally believed that LDPC codes can approach capacity of a variety of other communication channels. However, it is also believed that the decoding error probability decreases only polynomially with the code length. When using message-passing decoding algorithms [5], [12], the time complexity of their decoding is linear in a code length and polynomial in $1/\varepsilon$, where the code rate is $R = (1-\varepsilon)C$ and C is the capacity of BEC (see [8], [16]).

In contrast, the modifications of expander codes presented in [1], [2], [3], [14], [15] also attain the capacity of the memoryless q-ary symmetric channel, and the error probability decreases exponentially with the code length. In this work, we investigate time complexity of decoding algorithms of expander codes from [1], [3]. We show that these algorithms result in time complexity that is exponential in $1/\varepsilon^2$. Further, we propose a concatenated construction based on the expander codes that yields (under certain condition) the decoding complexity linear in code length and polynomial in $1/\varepsilon$, while having exponentially decreasing probability of decoding error.

Recall the construction in [14], [15]. Let $\mathcal{G} = (A : B, E)$ be a bipartite Δ-regular undirected connected graph with a vertex set $V = A \cup B$ such that $A \cap B = \emptyset$ and $|A| = |B| = n$, and an edge set E of size $N = \Delta n$ such that every edge in E has one endpoint in A and one endpoint in B. For every vertex $u \in V$, denote by $E(u)$ the set of edges incident with u, and assume some ordering on $E(u)$, for every $u \in V$. Let $\mathbb{F} = \mathrm{GF}(q)$ be some finite field, and $q > \Delta$.

Take $\mathcal{C}_A$ and $\mathcal{C}_B$ to be Generalized Reed-Solomon codes with parameters $[\Delta, r_A\Delta, \delta_A\Delta]$ and $[\Delta, r_B\Delta, \delta_B\Delta]$ over $\mathbb{F}$, respectively. We define the code $\mathbb{C} = (\mathcal{G}, \mathcal{C}_A : \mathcal{C}_B)$ as in [15], namely

$$\begin{aligned} \mathbb{C} \;=\; \big\{ \boldsymbol{c} \in \mathbb{F}^N \,:\, (\boldsymbol{c})_{E(u)} \in \mathcal{C}_A \text{ for every } u \in A \\ \text{and } \; (\boldsymbol{c})_{E(u)} \in \mathcal{C}_B \text{ for every } \; u \in B \big\} \;, \end{aligned} \tag{1}$$

where $(\boldsymbol{x})_{E(u)}$ denotes the sub-word of $\boldsymbol{x} = (x_e)_{e\in E} \in \mathbb{F}^N$ that is indexed by $E(u)$. The produced code $\mathbb{C}$ is a linear code of length N over $\mathbb{F}$.

Let Φ denote the alphabet $\mathbb{F}^{r_A\Delta}$. Taking some linear one-to-one mapping $\mathcal{E}_A : \Phi \to \mathcal{C}_A$ over $\mathbb{F}$, and the mapping $\psi : \mathbb{C} \to \Phi^n$ given by

$$\psi(\boldsymbol{c}) = \left(\mathcal{E}_A^{-1}\left((\boldsymbol{c})_{E(u)}\right)\right)_{u\in A} \;, \quad \boldsymbol{c} \in \mathbb{C} \;,$$

the authors of [15] define the code $\mathbb{C}_\Phi$ of length n over Φ by

$$\mathbb{C}_\Phi = \{\psi(\boldsymbol{c}) \;:\; \boldsymbol{c} \in \mathbb{C}\} \;. \tag{2}$$

The rate and the relative minimum distance of $\mathbb{C}_\Phi$ are denoted by R_Φ and δ_Φ, respectively.

Let $\lambda_\mathcal{G}$ be the second largest eigenvalue of the adjacency matrix of $\mathcal{G}$ and denote by $\gamma_\mathcal{G}$ the value $\lambda_\mathcal{G}/\Delta$. When $\mathcal{G}$ is taken from a family Δ-regular bipartite Ramanujan graphs

(e.g. [9]), we have

$$\lambda_{\mathcal{G}} \leq 2\sqrt{\Delta - 1}\ . \tag{3}$$

It was shown in [15], that the code $\mathbb{C}_\Phi$ has the relative minimum distance

$$\delta_\Phi \geq \frac{\delta_B - \gamma_{\mathcal{G}}\sqrt{\delta_B/\delta_A}}{1 - \gamma_{\mathcal{G}}}\ .$$

The linear-time decoding algorithm in Figure 1 was proposed in [15] that corrects any pattern of μ errors and ρ erasures such that $\mu + \frac{1}{2}\rho < \beta n$, where β is given by

$$\beta = \frac{(\delta_B/2) - \gamma_{\mathcal{G}}\sqrt{\delta_B/\delta_A}}{1 - \gamma_{\mathcal{G}}}\ . \tag{4}$$

The value of m in the algorithm was established in [15] such that $m = O(\log n)$. The notation "?" is used for erasures, and the notations $\mathcal{D}_A$ and $\mathcal{D}_B$ are used for decoders of the codes $\mathcal{C}_A$ and $\mathcal{C}_B$, respectively.

Input: *received word* $\boldsymbol{y} = (\boldsymbol{y}_u)_{u\in A}$ *in* $(\Phi \cup \{?\})^n$.

For $u \in A$ ***do*** $\quad (\boldsymbol{z})_{E(u)} \leftarrow \begin{cases} \mathcal{E}_A(\boldsymbol{y}_u) & \text{if } \boldsymbol{y}_u \in \Phi \\ ?? \ldots ? & \text{if } \boldsymbol{y}_u = ? \end{cases}$.

For $i \leftarrow 1, 2, \ldots, m$ ***do*** {

If i *is even* ***then*** $X \equiv A$, $\mathcal{D} \equiv \mathcal{D}_A$,
else $X \equiv B$, $\mathcal{D} \equiv \mathcal{D}_B$.

For $u \in X$ ***do*** $(\boldsymbol{z})_{E(u)} \leftarrow \mathcal{D}((\boldsymbol{z})_{E(u)})$.

}

Output: $\psi(\boldsymbol{z})$ *if* $\boldsymbol{z} \in \mathbb{C}$ *(and declare 'error' otherwise).*

Figure 1: Decoder of Roth and Skachek for the code $\mathbb{C}_\Phi$.

The proof in [15] requires that the decoder $\mathcal{D}_A$ is a mapping $\mathbb{F}^\Delta \to \mathcal{C}_A$ that recovers correctly any pattern of less than $\delta_A\Delta/2$ errors over $\mathbb{F}$, and the decoder $\mathcal{D}_B$ is a mapping $(\mathbb{F}\cup\{?\})^\Delta \to \mathcal{C}_B$ that recovers correctly any pattern of θ errors and ν erasures, provided that $2\theta+\nu < \delta_B\Delta$. The decoders $\mathcal{D}_A$ and $\mathcal{D}_B$ are polynomial-time, for example Berlekamp-Massey decoder can be used for both of them. It can be implemented then in $O(\Delta^2)$ time.

2 Proposed construction and analysis

General settings

Consider the memoryless binary symmetric channel with crossover probability p. Its capacity is given by $C = 1 - \mathsf{H}_2(p)$, where $\mathsf{H}_2(.)$ is the binary entropy function. Let $R = C(1-\varepsilon)$ be a designed rate. Take $\mathbb{F}$ to be a power of 2. A family of the codes $\mathbb{C}_\Phi$ described above will be used as an outer codes in a concatenated construction, whose parameters will be discussed in the sequel.

A 'typical' binary LDPC code of rate R_{in} and (constant for a fixed ε) length n_{in} will be used as an inner code $\mathcal{C}_{in}$ in this concatenated construction. Let us examine characteristics of a 'typical' LDPC code.

Decoding complexity: it was conjectured in [8] that per-bit complexity of message-passing decoding of LDPC or irregular repeat accumulative (IRA) codes over any 'typical' channel is $O\left(\log \frac{1}{\pi}\right) + O\left(\frac{1}{\varepsilon}\log\frac{1}{\varepsilon}\right)$, where π is a decoded error probability. For LDPC codes over BEC it was shown in [10] that the decoding complexity per bit behaves as $O(\log(1/\varepsilon))$. In this work we assume that the decoding complexity of LDPC (or some other polynomial-time in n_{in} decodable) codes over the binary symmetric channel is given by

$$O\left(n_{in} \cdot \frac{1}{\varepsilon^r}\right) , \tag{5}$$

where $r \geq 1$ is some constant. Recently, in [11], IRA codes with bounded decoding complexity per bit were constructed, but for our purposes the assumption (5) will be enough. It is possible to assume here that the inner code is polynomial-time decodable (rather than linear-time), still the resulting decoding time complexity is linear in overall length of a resulting concatenated code.

Decoding error probability: as of yet, there are no satisfying results on asymptotical behavior of the decoding error probability of LDPC codes over the binary symmetric channel under the message-passing decoding, for rates near the capacity of BEC. The behavior of the decoding error probability of LDPC codes over other channels is even less investigated. In this work, we obtain a sufficient condition on the probability of decoding error $\mathsf{Prob}_e(\mathcal{C}_{in})$ of an inner code to guarantee that the error decoding probability of the overall concatenated code decreases exponentially. In the sequel, we provide several examples of several possible decoding error probability functions for which this condition holds.

Let $\mathbb{C}_{cont}$ be a resulting concatenated code of rate $R_{cont} \geq R$ and length N_{cont}. Denote by $\mathsf{Prob}_e(\mathbb{C}_{cont})$ its error decoding probability. The following lemma is based on the result of Forney [4, Chapter 4.2].

Lemma 1 *The decoding error probability of the code $\mathbb{C}_{cont}$ under the combined decoding, when probability of decoding error for the inner code is $\mathsf{Prob}_e(\mathcal{C}_{in})$, and the outer code is decoded by the expander decoder in Figure 1, is bounded by*

$$\mathsf{Prob}_e(\mathbb{C}_{cont}) \leq \exp\{-n \cdot E\} = \exp\{-N_{cont} \cdot \frac{E}{n_{in}}\} ,$$

where $E > 0$ is a constant given by

$$\begin{aligned} E &= -\beta \log\left(\mathsf{Prob}_e(\mathcal{C}_{in})\right) - (1-\beta)\log\left(1-\mathsf{Prob}_e(\mathcal{C}_{in})\right) \\ &\quad +\beta\log(\beta) + (1-\beta)\log(1-\beta)\,, \end{aligned} \tag{6}$$

and β is defined in (4).

Proof. We analyze the error exponent, following the guidelines of the analysis of Forney [4, Chapter 4.2]. Let ς_i, $i = 1, \cdots, n$, be a random variable which equals 1 if no inner decoding error is made while decoding i-th inner codeword, and -1 otherwise. The outer code will fail to decode correctly if and only if

$$\varsigma \stackrel{\triangle}{=} \frac{1}{n}\sum_{i=1}^{n}\varsigma_i < (1-2\beta)\,.$$

Denote

$$\mu(-s) \stackrel{\triangle}{=} \log\left(\mathsf{Prob}_e(\mathcal{C}_{in})\cdot \mathsf{e}^{s} + (1-\mathsf{Prob}_e(\mathcal{C}_{in}))\cdot \mathsf{e}^{-s}\right)\,.$$

Using the Chernov bound, we obtain

$$\mathsf{Prob}_e(\mathbb{C}_\Phi) = \mathsf{Prob}\left(\frac{1}{n}\sum_{i=1}^{n}\varsigma_i < (1-2\beta)\right) < \mathsf{e}^{-n(s(2\beta-1)-\mu(-s))}\,.$$

Optimization of the exponent over values of s yields that the maximum of the expression

$$s(2\beta-1) - \mu(-s)$$

is achieved when

$$s = \tfrac{1}{2}\log\frac{(1-\mathsf{Prob}_e(\mathcal{C}_{in}))\cdot 2\beta}{\mathsf{Prob}_e(\mathcal{C}_{in})\cdot(2-2\beta)}\,,$$

and the maximum is

$$\begin{aligned} s(2\beta-1)-\mu(-s) &= -\beta\log\left(\mathsf{Prob}_e(\mathcal{C}_{in})\right) - (1-\beta)\log\left(1-\mathsf{Prob}_e(\mathcal{C}_{in})\right) \\ &+ \beta\log(\beta) + (1-\beta)\log(1-\beta)\,, \end{aligned}$$

thus completing the proof. □

Parameter selection

In this section we make a selection of parameters for the code $\mathbb{C}_{cont}$. This parametrization allows us to estimate a decoding error exponent as a function of ε.

Pick the rate of $\mathcal{C}_{in}$ to be $R_{in} = (1-\kappa\,\varepsilon)C$, where $\kappa \in (0,1)$ is a constant. Then, we can write

$$\frac{R}{R_{in}} = \frac{C(1-\varepsilon)}{C(1-\kappa\,\varepsilon)} \geq 1 - (1-\kappa)\varepsilon + O(\varepsilon^2)\,.$$

We fix $\delta_B = 1 - R/R_{in} - \delta_A = \eta(1 - R/R_{in})$, where $\eta \in (0,1)$ (and thus, $\delta_A = (1-\eta)(1 - R/R_{in})$), and select the degree Δ of the graph $\mathcal{G}$ as $\Delta = \varrho/\varepsilon^h$, where ϱ and $h \geq 2$ are constants. If $h = 2$, we require in addition that

$$\varrho > \frac{16}{\eta(1-\eta)(1-\kappa)^2} \; . \tag{7}$$

Since $\mathcal{C}_A$ and $\mathcal{C}_B$ are GRS codes,

$$R_\Phi \geq r_A + r_B - 1 = 1 - \delta_A - \delta_B = R/R_{in} \; .$$

Then, $\lambda_\mathcal{G} = 2\sqrt{\varrho/\varepsilon^h - 1}$ (see (3)) and

$$\gamma_\mathcal{G} = \frac{2\sqrt{\varrho/\varepsilon^h - 1}}{\varrho/\varepsilon^h} < \frac{2\varepsilon^{h/2}}{\sqrt{\varrho}} \; .$$

It holds

$$\beta \;\; = \;\; \frac{(\delta_B/2) - \gamma_\mathcal{G}\sqrt{\delta_B/\delta_A}}{1 - \gamma_\mathcal{G}} > (\delta_B/2) - \gamma_\mathcal{G}\sqrt{\delta_B/\delta_A} \; .$$

The selection of ϱ guarantees that

$$\beta > \vartheta\varepsilon + o(\varepsilon) \; , \tag{8}$$

where

$$0 < \vartheta = \begin{cases} \frac{\eta(1-\kappa)}{2} - 2\sqrt{\frac{\eta}{\varrho(1-\eta)}} & \text{if } h = 2 \\ \frac{\eta(1-\kappa)}{2} & \text{if } h > 2 \end{cases} \; ,$$

is a constant which depends only on κ, η and ϱ.

Example

Suppose that the decoding error probability for some inner code $\mathcal{C}_{in}$ over the binary symmetric channel with crossover probability $p < \mathsf{H}_2^{-1}(1 - R_{in})$ and some polynomial decoder is given by:

$$\mathsf{Prob}_e(\mathcal{C}_{in}) \leq \frac{1}{n_{in}^t},$$

where t is a constant, $t \geq 1$.

Set $h = 2$ and $q = \Delta$. Note that $r_A = 1 - O(\varepsilon)$. Since $r_A \Delta \log_2(q)$ bits are needed to represent each symbol of Φ, the length n_{in} of the binary inner code $\mathcal{C}_{in}$ is given by

$$\begin{aligned} n_{in} &= \frac{r_A \Delta}{R_{in}} \cdot \log_2(\Delta) & (9) \\ &= \frac{(1 - O(\varepsilon))\varrho}{R_{in}\,\varepsilon^2} \cdot \log_2\left(\frac{\varrho}{\varepsilon^2}\right) \\ &= \frac{\varrho \log_2(\varrho/\varepsilon^2)}{R_{in}\,\varepsilon^2} + o\left(\frac{\varrho \log_2(\varrho/\varepsilon^2)}{R_{in}\,\varepsilon^2}\right) \; , & (10) \end{aligned}$$

and thus, by ignoring the small term, the decoding error probability of $\mathcal{C}_{in}$ is

$$\mathsf{Prob}_e(\mathcal{C}_{in}) \leq \left(\frac{\varepsilon^2 R_{in}}{\varrho \log_2(\varrho/\varepsilon^2)}\right)^t . \tag{11}$$

We substitute expressions from (8) (only the main term) and (11) into the result of Lemma 1 to obtain

$$\begin{aligned}\mathsf{Prob}_e(\mathbb{C}_{cont}) \;<\; \exp\Bigg\{ -n\Bigg(&-\vartheta\varepsilon \cdot t \log\left(\frac{\varepsilon^2 R_{in}}{\varrho \log_2(\varrho/\varepsilon^2)}\right) \\ &- (1-\vartheta\varepsilon)\log\left(1-\left(\frac{\varepsilon^2 R_{in}}{\varrho \log_2(\varrho/\varepsilon^2)}\right)^t\right) \\ &+ \vartheta\varepsilon \log(\vartheta\varepsilon) + (1-\vartheta\varepsilon)\log(1-\vartheta\varepsilon)\Bigg)\Bigg\} . \end{aligned} \tag{12}$$

Note that for small $\varepsilon > 0$,

$$\log(1-\vartheta\varepsilon) = -\vartheta\varepsilon + O(\varepsilon^2) ,$$

and

$$\log\left(1-\left(\frac{\varepsilon^2 R_{in}}{\varrho \log_2(\varrho/\varepsilon^2)}\right)^t\right) = -o(\varepsilon^{2t}) .$$

Hence, the equation (12) (when neglecting $o(\varepsilon)$ terms) becomes

$$\begin{aligned}\mathsf{Prob}_e(\mathbb{C}_{cont}) \;&<\; \exp\left\{ -n\vartheta\varepsilon\left(-t\log\left(\frac{\varepsilon^2 R_{in}}{\varrho \log_2(\varrho/\varepsilon^2)}\right) + \log(\vartheta\varepsilon) - 1\right)\right\} \\ &=\; \exp\left\{ -\frac{N_{cont}\vartheta\varepsilon}{n_{in}} \cdot \log\left(\frac{\vartheta\varepsilon \cdot \varrho^t (\log_2(\varrho/\varepsilon^2))^t}{\mathsf{e}\cdot\varepsilon^{2t} R_{in}^t}\right)\right\} . \end{aligned}$$

Using substitution of the expression (10) for n_{in}, the latter equation can be rewritten as

$$\begin{aligned}\mathsf{Prob}_e(\mathbb{C}_{cont}) \;<\; \exp\Bigg\{ &-\frac{N_{cont}\vartheta\varepsilon\cdot\varepsilon^2 R_{in}}{2\varrho\,(\log_2(1/\varepsilon)+\Theta(1))} \\ &\cdot\left((2t-1)\log(1/\varepsilon) + t\log(1/R_{in}) + t\log\log(1/\varepsilon) + \Theta(1)\right)\Bigg\} . \end{aligned} \tag{13}$$

The dominating term in the expression

$$(2t-1)\log(1/\varepsilon) + t\log(1/R_{in}) + t\log\log(1/\varepsilon) + \Theta(1)$$

is $(2t-1)\log(1/\varepsilon)$. By taking into account that $R_{in} = C(1-O(\varepsilon))$, the equation (13) can be rewritten, when ignoring all but the main term, as

$$\mathsf{Prob}_e(\mathbb{C}_{cont}) \;<\; \exp\left\{ -N_{cont}\cdot\left(\frac{(2t-1)\,\vartheta\,\varepsilon^3\,C}{2\varrho\log 2} + o(\varepsilon^3)\right)\right\} .$$

Thus, the decoding error probability is given by

$$\mathsf{Prob}_e(\mathbb{C}_{cont}) < \exp\{-N_{cont} \cdot E(C,\varepsilon)\} \ ,$$

where

$$\begin{aligned} E(C,\varepsilon) &= \max_{\varrho,\vartheta} \left\{ \frac{\vartheta}{\varrho} \right\} \cdot \frac{(2t-1)\,C}{2\log 2} \cdot \varepsilon^3 \\ &= \max_{\kappa,\,\eta,\,\varrho} \left\{ \frac{\eta(1-\kappa)}{2\varrho} - 2\sqrt{\frac{\eta}{\varrho^3(1-\eta)}} \right\} \cdot \frac{(2t-1)\,C}{2\log 2} \cdot \varepsilon^3 \ , \end{aligned} \tag{14}$$

and the parameters (κ, η, ϱ) are taken over

$$\kappa \in (0,1) \ ; \ \eta \in (0,1) \ ; \ \varrho > \frac{16}{\eta(1-\eta)(1-\kappa)^2} \ . \tag{15}$$

Next, we optimize the value of the constant

$$\Upsilon = \max_{\kappa,\,\eta,\,\varrho} \left\{ \frac{\eta(1-\kappa)}{2\varrho} - 2\sqrt{\frac{\eta}{\varrho^3(1-\eta)}} \right\} \ .$$

It is easy to see that the maximum is received for $\kappa \to 0$. We substitute $\kappa = 0$ in expression (14) to obtain

$$\Upsilon = \max_{\eta,\,\varrho} \left\{ \frac{\eta}{2\varrho} - 2\sqrt{\frac{\eta}{\varrho^3(1-\eta)}} \right\} \ . \tag{16}$$

By taking a derivative of Υ over ϱ and comparing it to zero, we obtain that

$$\varrho = \frac{36}{\eta(1-\eta)} \ .$$

By substituting it back to expression (16) and finding point of maximum, we have $\eta = 2/3$ and $\varrho = 162$. These values obviously satisfy condition (15). The appropriate value of Υ is then

$$\begin{aligned} \Upsilon &= \frac{\eta}{2\varrho} - 2\sqrt{\frac{\eta}{\varrho^3(1-\eta)}} = \frac{2/3}{2 \cdot 162} - 2\sqrt{\frac{2/3}{162^3 \cdot (1/3)}} \\ &= \frac{1}{1458} = 6.8587 \cdot 10^{-4} \ . \end{aligned}$$

Finally, we have

$$E(C,\varepsilon) = \frac{(2t-1)\,C}{2916 \cdot \log 2} \cdot \varepsilon^3 \ .$$

Figure 2 shows value of error exponent $E(C,\varepsilon)$ in the example for $t = 1$, 2 and 3.

Remark. It is possible to improve an error exponent by a constant factor if allowing the decoder for the code $\mathcal{C}_{in}$ to put out an "erasure" message in a case of unreliable decoding of the code $\mathcal{C}_{in}$. See [4, Chapter 4.2] for details.

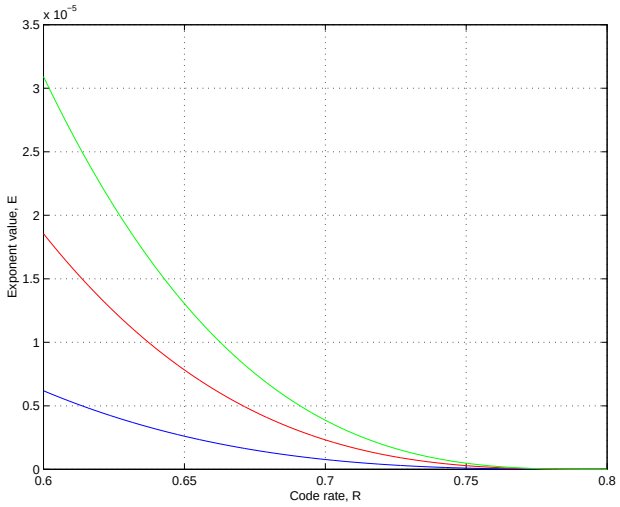


Figure 2: Error exponent $E(C, \varepsilon)$ for the code $\mathbb{C}_{cont}$.

Selection: $\mathsf{Prob}_e(\mathcal{C}_{in}) = 1/n_{in}^t$; $C = 0.8$; $t = 1, 2, 3$ (bottom to top).

Sufficient condition

In this section we derive a sufficient condition on the probability of decoding error of the inner code for providing a positive error exponent for $\mathbb{C}_{cont}$ under assumption that the outer code is an expander code. We use the notation $\mathcal{C}_{in}\,[R_{in}, n_{in}]$ for the code $\mathcal{C}_{in}$ of rate R_{in} and length n_{in}.

Theorem 2 *Suppose, there exist constants $\varepsilon_0 \in (0, 1)$, $\mathsf{a} > 1$ and $h_0 > 2$, such that for any $0 < \epsilon < \varepsilon_0$, the decoding error probability of a family of codes $\mathcal{C}_{in}$ satisfies*

$$\mathsf{Prob}_e\left(\mathcal{C}_{in}\left[(1-\epsilon)C,\ \frac{1}{\epsilon^{h_0}}\right]\right) < \frac{\epsilon}{2\mathsf{e}^{\mathsf{a}}}\ .$$

Then for any rate $R < C$ there exists a family of the codes $\mathbb{C}_{cont}$ with exponentially decaying error probability.

Proof. Let $R = (1-\varepsilon)C$ be a design rate of the code $\mathbb{C}_{cont}$. In the sequel, we select $0 < \kappa < \min\{1, \varepsilon_0/\varepsilon\}$ such that the rate of the code $\mathcal{C}_{in}$ is $R_{in} = (1 - \kappa\,\varepsilon)C = (1-\epsilon)C > (1-\varepsilon_0)C$. We aim to show for this selection of κ that the decoding of the code $\mathbb{C}_{cont}$ has an exponentially decaying error probability.

We select $R_{in} > R$, $h > h_0$, and $\delta_B = 1 - R/R_{in} - \delta_A = \eta(1 - R/R_{in})$, where $\eta \in (0, 1)$ is a constant. We also take $\varrho > 0$ and $\Delta = \varrho/(\kappa \cdot \varepsilon)^h$.

As before, $R_\Phi \geq 1 - \delta_A - \delta_B = R/R_{in}$, and

$$\beta > \vartheta\varepsilon + o(\varepsilon)\ , \tag{17}$$

where $\vartheta \in (0, \frac{1}{2})$ is a constant, which depends only on κ and η. Moreover, by an appropriate selection of κ and η it is possible to make ϑ as close to $\frac{1}{2}$ as desired.

We substitute expressions for R_{in} and Δ into (9). Then, the length n_{in} of the code $\mathcal{C}_{in}$ is given by

$$n_{in} = \frac{\varrho\, r_A\, \log_2(\varrho/(\kappa\varepsilon)^h)}{(\kappa\varepsilon)^h(1-\kappa\varepsilon)C} \ .$$

We substitute the main term of the expression (17) into the result of Lemma 1 to obtain (for small enough ε) that

$$\begin{aligned} \mathsf{Prob}_e(\mathbb{C}_{cont}) \ &< \ \exp\Bigg\{ - n\Bigg(- \vartheta\varepsilon \cdot \log\left(\mathsf{Prob}_e(\mathcal{C}_{in})\right) - \ (1-\vartheta\varepsilon)\log\left(1 - \mathsf{Prob}_e(\mathcal{C}_{in})\right) \\ &\qquad + \vartheta\varepsilon\log\left(\vartheta\varepsilon\right) + (1-\vartheta\varepsilon)\log\left(1-\vartheta\varepsilon\right)\Bigg)\Bigg\} . \end{aligned} \tag{18}$$

For small enough ε,

$$(1-\vartheta\varepsilon)\log(1-\vartheta\varepsilon) > -\mathsf{a}\cdot\vartheta\varepsilon \ ,$$

and, by ignoring the positive term in the exponent, the equation (18) can be rewritten as

$$\begin{aligned} \mathsf{Prob}_e(\mathbb{C}_{cont}) \ &< \ \exp\Big\{ - n\Big(- \vartheta\varepsilon \cdot \log\left(\mathsf{Prob}_e(\mathcal{C}_{in})\right) + \ \vartheta\varepsilon\log\left(\vartheta\varepsilon\right) - \mathsf{a}\vartheta\varepsilon\Big)\Big\} \\ &= \ \exp\left\{-N_{cont}\cdot\frac{\vartheta\varepsilon}{n_{in}}\left(\log\left(\frac{\vartheta\varepsilon}{\mathsf{Prob}_e(\mathcal{C}_{in})}\right) - \mathsf{a}\right)\right\} \ . \end{aligned}$$

In order to have a positive exponent it is sufficient that there exist some $\epsilon = \kappa\varepsilon$ such that for the length $n_{in}(\varepsilon)$ and the rate $R_{in} = (1-\kappa\varepsilon)C$, it holds

$$\frac{\vartheta\varepsilon}{\mathsf{Prob}_e(\mathcal{C}_{in}[R_{in},\ n_{in}])} > \mathsf{e}^{\mathsf{a}} \ .$$

Since by an appropriate selection of κ and η it is possible to make ϑ as close to $\frac{1}{2}$ as desired, the latter condition can be rewritten as

$$\frac{\frac{1}{2}\varepsilon}{\mathsf{Prob}_e(\mathcal{C}_{in}[R_{in},\ n_{in}])} > \mathsf{e}^{\mathsf{a}} \ . \tag{19}$$

By taking $\epsilon = \kappa\varepsilon < \varepsilon_1$ (for small enough ε_1) it is possible to have

$$n_{in} = \frac{\varrho\, r_A\, \log(\varrho/(\kappa\varepsilon)^h)}{(\kappa\varepsilon)^h(1-\kappa\varepsilon)C} > \frac{1}{(\kappa\varepsilon)^{h_0}}$$

(note that $r_A \to 1$ as $\varepsilon \to 0$). Thus, by selection

$$\kappa < \min\{1, \varepsilon_0/\varepsilon,\ \varepsilon_1/\varepsilon\} \ , \tag{20}$$

we obtain

$$\mathsf{Prob}_e(\mathcal{C}_{in}[R_{in}, n_{in}]) < \mathsf{Prob}_e\left(\mathcal{C}_{in}\left[(1-\kappa\varepsilon)C, \frac{1}{(\kappa\varepsilon)^{h_0}}\right]\right) < \frac{\kappa\varepsilon}{2\mathsf{e}^{\mathsf{a}}} < \frac{\varepsilon}{2\mathsf{e}^{\mathsf{a}}} .$$

It follows that for selection as in (20), the error exponent is strictly positive. □

Example. Suppose that the decoding error probability of the code $\mathcal{C}_{in}$ of rate $R_{in} = (1-\varepsilon)C$ and length n_{in} (for some polynomial decoder) is bounded by

$$\mathsf{Prob}_e(\mathcal{C}_{in}) < \frac{1}{n_{in}} \cdot \frac{1}{\varepsilon^4} .$$

For $h_0 > 5$ and $\mathsf{a} > 1$ there obviously exists ε_0 such that for every $0 < \epsilon < \varepsilon_0$,

$$\mathsf{Prob}_e(\mathcal{C}_{in}) < \frac{1}{n_{in}} \cdot \frac{1}{\epsilon^4} = \epsilon^{h_0} \cdot \frac{1}{\epsilon^4} = \epsilon^{h_0-4} < \frac{\epsilon}{2\mathsf{e}^{\mathsf{a}}} ,$$

and therefore the conditions of Theorem 2 satisfied. This selection guarantees the existence of a positive error exponent.

Example. Suppose that the decoding error probability of the code $\mathcal{C}_{in}$ (of rate $R_{in} = (1-\varepsilon)C$ and length n_{in}) is bounded by

$$\mathsf{Prob}_e(\mathcal{C}_{in}) < \mathsf{e}^{-n_{in}\varepsilon^2}.$$

For $h_0 = 3$ and $\mathsf{a} > 1$ there obviously exists ε_0 such that for every $0 < \epsilon < \varepsilon_0$,

$$\mathsf{Prob}_e(\mathcal{C}_{in}) < \mathsf{e}^{-n_{in}\epsilon^2} = \mathsf{e}^{-(\epsilon^2/\epsilon^3)} = \mathsf{e}^{-(1/\epsilon)} < \frac{\epsilon}{2\mathsf{e}^{\mathsf{a}}} ,$$

and therefore Theorem 2 yields the existence of a positive error exponent.

Decoding complexity

Theorem 3 *The time complexity of decoding algorithm of the code $\mathbb{C}_{cont}$ under the combined decoding, when the decoding complexity of an inner code is as in (5) and the outer code is an expander code, is linear in N_{cont} and is polynomial in $1/\varepsilon$.*

Proof. First, we estimate the decoding time complexity of the outer code $\mathbb{C}_\Phi$. We show that the total number of applications of one of the decoders $\mathcal{D}_A$ and $\mathcal{D}_B$ is upper-bounded by $O(n)$, and does not depend on ε. Our analysis is closely related to the analysis in [15, Section 4].

Denote by σ ($\sigma < \beta$) the number of erroneous symbols in the code $\mathbb{C}_\Phi$. Denote by σ_i the relative number of erroneous vertices in the set A and the set B after i-th iteration of the decoder in Figure 1, for odd and even i, respectively. In particular, $\sigma_1 = \sigma$, since each erroneous symbol $\boldsymbol{y}_u$ of $\boldsymbol{y}$ is translated into erroneous vertex $u \in A$.

It was shown in [15, Theorem 4.1] that for even $i \geq 2$, σ_i and σ are related as

$$\frac{1}{\sqrt{\sigma_{i+1}}} \;\geq\; \left(\left(\frac{\delta_A\delta_B}{4\gamma_{\mathcal{G}}^2}\right)^{i/2}\left(1-\frac{\sigma}{\beta}\right)+\frac{\sigma}{\beta}\right)\frac{1}{\sqrt{\sigma}} \tag{21}$$

Consider the case $h > 2$ (the proof can be easily adjusted for the case $h = 2$). The inequality (21) can be rewritten as

$$\frac{1}{\sqrt{\sigma_{i+1}}} \geq \left(\left(\Theta(\varepsilon^{2-h})\right)^{i/2}\left(1-\frac{\sigma}{\beta}\right)+\frac{\sigma}{\beta}\right)\frac{1}{\sqrt{\sigma}}\,, \tag{22}$$

and obviously the number of erroneous vertices decreases exponentially, and for $\varepsilon \to 0$ the number of iterations approaches some constant.

The total number of applications of one of the decoders $\mathcal{D}_A$ and $\mathcal{D}_B$ is bounded by

$$\Gamma = (\sigma_1 + \sigma_2 + \sigma_3 + \sigma_4 + \cdots + \sigma_m)\cdot n\,,$$

where m is a total number of iterations, and $\sigma_m < 1/n$. Write number of erroneous vertices for odd and even iterations separately,

$$\Gamma_1 = (\sigma_1 + \sigma_3 + \sigma_5 + \cdots)\cdot n \text{ and } \Gamma_2 = (\sigma_2 + \sigma_4 + \sigma_6 + \cdots)\cdot n\,.$$

Using (22) we see that $\Gamma_1 \leq \sigma_1 n \cdot c_\omega$, where c_ω is a constant that does not depend on n and ε. It was shown in [15, Theorem 4.1] that for even $i > 1$

$$\sigma_{i-1}/\sigma_i \geq \delta_A/\delta_B\,,$$

and thus $\Gamma_2 \leq \sigma_1 n \cdot c_\omega \cdot \delta_B/\delta_A$. Finally, we conclude that $\Gamma = \Gamma_1 + \Gamma_2 = O(n)$, and the sum does not depend on ε.

Using techniques, similar to that of Sipser and Spielman [17], one can maintain on each iteration of the algorithm a list of erroneous vertices. Both decoders $\mathcal{D}_A$ and $\mathcal{D}_B$ can be implemented to work in time $O(\Delta^2) = O(1/\varepsilon^{2h})$. Each application of decoder $\mathcal{D}_A$ ($\mathcal{D}_B$) changes at most $\frac{1}{2}\delta_A\Delta$ ($\frac{1}{2}\delta_B\Delta$) $= O(\varepsilon\Delta) = O(1/\varepsilon^{h-1})$ symbols, — only this number of neighbors u should be checked to satisfy $(\boldsymbol{z})_{E(u)} \in \mathcal{C}_B$ ($(\boldsymbol{z})_{E(u)} \in \mathcal{C}_A$). If some of equations are not satisfied, the vertex is mentioned as "erroneous". Each such check requires to compute $O(\varepsilon\Delta) = O(1/\varepsilon^{h-1})$ parity check equations of length $\Delta = O(1/\varepsilon^h)$. The complexity of operations involved in i-th iteration of the decoder is thus $O(\sigma_i n/\varepsilon^{3h-2})$. The total complexity of decoding of $\mathbb{C}_\Phi$ is, in turn, $O(n/\varepsilon^{3h-2})$.

We assume that each inner code has the decoding complexity bounded by $O(n_{in}/\varepsilon^r)$, for example a 'typical' LDPC code can be chosen as an inner code [1]. Each code of length n_{in}

[1] *The decoding complexity of $\mathbb{C}_{cont}$ will be linear in N_{cont} even if $\mathcal{C}_{in}$ has a polynomial decoder. However, in this case the power of $1/\varepsilon$ in the complexity expression will be higher.*

forms the sub-block of the word of $\mathbb{C}_{cont}$ and there are n such blocks in total. Therefore, the overall complexity of decoding all inner codes is

$$O\left(\frac{n_{in}}{\varepsilon^r}\cdot n\right) = O\left(\frac{N_{cont}}{\varepsilon^r}\right) .$$

Recall that using the equation (9), and substituting the selection of Δ, the length n_{in} of the inner code is given by

$$n_{in} = O(\log(1/\varepsilon)/\varepsilon^h).$$

Summing the decoding complexities of the inner code $\mathcal{C}_{in}$ and the outer code $\mathbb{C}_\Phi$, we obtain

$$O\left(N_{cont}\cdot\left(\frac{1}{\varepsilon^{2h-2}\log(1/\varepsilon)}+\frac{1}{\varepsilon^r}\right)\right).$$

□

3 Time complexity of decoder in [1]

The purpose of this and the next sections is to compare the parameters of the codes from Section 2 with the codes presented by Barg and Zémor in [1] and [3]. Similarly to the previous sections, we assume that the design rate is $R = (1-\varepsilon)C$. In the sequel we show that the parameters of codes from [1] and [3] cannot be modified such that the time complexity would be only sub-exponential in $1/\varepsilon$ while keeping a non-zero error exponent.

Construction

We consider codes in [1]. We briefly describe the construction and the decoding algorithm, following the presentation in [1].

Let $\mathcal{G} = (A : B, E)$ be a bipartite Δ-regular undirected connected graph with a vertex set $V = A \cup B$ such that $A \cap B = \emptyset$ and $|A| = |B| = n$, and an edge set E of size $N = \Delta n$ such that every edge in E has one endpoint in A and one endpoint in B.

Let the size of $\mathbb{F}$ be a power of 2. Let $\mathcal{C}_A$ and $\mathcal{C}_B$ be two random codes of length Δ over $\mathbb{F}$. The code $\mathbb{C}_{BZ2} = (\mathcal{G}, \mathcal{C}_A : \mathcal{C}_B)$ is defined similarly to the definition of $\mathbb{C}$ in (1), with respect to $\mathcal{C}_A$ and $\mathcal{C}_B$ as defined in this paragraph.

Decoding

Formal definition of the decoder appears in Figure 3. The number of iterations m is taken to be $O(\log n)$. The decoders $\mathcal{D}_A$ and $\mathcal{D}_B$ are the maximum-likelihood decoders for the codes $\mathcal{C}_A$ and $\mathcal{C}_B$, respectively.

Input: *Received word* $\boldsymbol{y} = (\boldsymbol{y}_e)_{e\in E}$ *in* $\mathbb{F}^N$.

Let $\boldsymbol{z} \leftarrow \boldsymbol{y}$.

For $i \leftarrow 1, 2, \ldots, m$ ***do*** {

If i *is odd* ***then*** $X \equiv A$, $\mathcal{D} \equiv \mathcal{D}_A$,
else $X \equiv B$, $\mathcal{D} \equiv \mathcal{D}_B$.

For $u \in X$ ***do*** $(\boldsymbol{z})_{E(u)} \leftarrow \mathcal{D}((\boldsymbol{z})_{E(u)})$.

}

Output: $\boldsymbol{z}$ *if* $\boldsymbol{z} \in \mathbb{C}_{BZ2}$ *(and declare 'error' otherwise).*

Figure 3: Decoder of Barg and Zémor for the code $\mathbb{C}_{BZ2}$.

Analysis

Following the analysis of [1] it is possible to show that for the code $\mathbb{C}_{BZ2}$ over $\mathbb{F} = \mathrm{GF}(2)$, the error probability, $\mathsf{Prob}_e(\mathbb{C}_{BZ2})$, is bounded by

$$P_e(\mathbb{C}_{BZ2}, p) \leq \exp\{-\alpha N f_3(R, p)\} \,,$$

where $0 < \alpha < 1$, and the main term of $f_3(R, p)$ is less or equal to

$$\max_{R\leq R_0<C} \quad \left\{ E_0(R_0, p) \left(\frac{\mathsf{H}_2^{-1}(R_0-R)}{2} - \Theta\left(\frac{1}{\sqrt{\Delta}}\right)\right)\right\}, \tag{23}$$

where $E_0(R_0, p)$ is the *random coding exponent* for rate R_0 over the BSC with the crossover probability p.

Proposition 4 *If the codes* $\mathbb{C}_{BZ2}$ *over* $\mathbb{F} = \mathrm{GF}(2)$ *have a positive error exponent, then* $\Delta = \Omega\left(1/(\mathsf{H}_2^{-1}(\varepsilon))^2\right)$.

Proof. In order to have a positive error exponent it is needed that

$$\frac{\mathsf{H}_2^{-1}(R_0 - R)}{2} - \Theta\left(\frac{1}{\sqrt{\Delta}}\right) > 0 \,.$$

Observe that $R_0 - R \leq C - R = C\varepsilon \leq \varepsilon$. Recall the error exponent in (23) and conclude that

$$\tfrac{1}{2}\mathsf{H}_2^{-1}(\varepsilon) \geq \tfrac{1}{2}\mathsf{H}_2^{-1}(R_0 - R) > \Theta\left(1/\sqrt{\Delta}\right),$$

and thus $\Delta = \Omega\left(1/(\mathsf{H}_2^{-1}(\varepsilon))^2\right)$. □

It is suggested in [1] to use the maximum-likelihood decoding for *random* codes $\mathcal{C}_A$ and $\mathcal{C}_B$. This decoding, however, has time complexity at least

$$\exp\{\Omega(\Delta)\} = \exp\{\Omega\left(1/(\mathsf{H}_2^{-1}(\varepsilon))^2\right)\} .$$

In the analysis of [1] it is possible to take $\mathbb{C}_{BZ2}$ over large $\mathbb{F}$ (see [1, Section IV]). In that case, the error probability, $\mathsf{Prob}_e(\mathbb{C}_{BZ2})$, is bounded by

$$P_e(\mathbb{C}_{BZ2}, p) \leq \exp\{-\alpha N f_2(R,p)\} ,$$

and the main term of $f_2(R,p)$ is less or equal to

$$\max_{R \leq R_0 < C} \left\{ E_0(R_0,p) \left(\tfrac{R_0 - R}{2} - \Theta\left(\tfrac{1}{\sqrt{\Delta}}\right)\right)\right\} . \tag{24}$$

In this case, Proposition 4 can be rewritten as

Proposition 5 *If the codes $\mathbb{C}_{BZ2}$ have a positive error exponent, then $\Delta = \Omega\left(1/\varepsilon^2\right)$.*

The proof is very similar to that of Proposition 4.

When using the maximum-likelihood decoder for *random* codes $\mathcal{C}_A$ and $\mathcal{C}_B$, the decoding time complexity is at least

$$\exp\{\Omega(\Delta)\} = \exp\{\Omega\left(1/\varepsilon^2\right)\} .$$

4 Time complexity of decoder in [3]

Construction

Recall the construction of expander codes presented in [3]. Let $\mathcal{G} = (V, E)$ be a bipartite graph with $V = V_0 \cup (V_1 \cup V_2)$, such that each edge has one endpoint in V_0 and one endpoint in either V_1 or V_2. Let $|V_i| = n$ for $i = 0, 1, 2$. Let the degree of each vertex in V_0, V_1, and V_2 be Δ, Δ_1, and $\Delta_2 = \Delta - \Delta_1$, respectively. In addition, let the subgraph $\mathcal{G}_1$ induced by $V_0 \cup V_1$ be a regular bipartite Ramanujan graph and denote by E_1 its edge set. Let λ_1 be a second largest eigenvalue of the adjacency matrix of $\mathcal{G}_1$.

Let $\mathcal{C}_A$ be a $[l\Delta, R_0 l\Delta, d_0 = l\Delta\delta_0]$ linear binary code of rate $R_0 = \Delta_1/\Delta$. Let $\mathcal{C}_B$ be q-ary $[\Delta_1, R_1\Delta_1, d_1 = \Delta_1\delta_1]$ additive code, and let $q = 2^l$. Let $\mathcal{C}_{aux}$ be q-ary code of length Δ_1. The code $\mathbb{C}_{BZ3}$ is defined as the set of vectors $\boldsymbol{x} = \{x_1, x_2, \cdots, x_N\}$, indexed by the set E of size $N = \Delta n$, such that

1. For every vertex $v \in V_0$, the subvector $(x_j)_{j \in E(v)}$ is a q-ary codeword of $\mathcal{C}_A$ and the set of coordinates $E_1(v)$ is an information set for the code $\mathcal{C}_A$.

2. For every vertex $v \in V_1$, the subvector $(x_j)_{j \in E(v)}$ is a q-ary codeword of $\mathcal{C}_B$.

3. For every vertex $v \in V_0$, the subvector $(x_j)_{j \in E_1(v)}$ is a codeword of $\mathcal{C}_{aux}$.

Decoding

The authors of [3] proposed decoding algorithm for the code $\mathbb{C}_{BZ3}$. In the first iteration, each subvector $\boldsymbol{z}(v)$, $v \in V_0$, is treated as following: the decoder computes, for every symbol b of the q-ary alphabet, and for every edge $e \in E_1$ incident to v, the weight of the edge as follows:

$$d_{e,b}(\boldsymbol{z}) = \min_{\boldsymbol{a} \in \mathcal{C}_A : a(e)=b} \mathsf{d}(\boldsymbol{a}, \boldsymbol{z}(v)),$$

where $a(e)$ denotes the q-ary coordinate of the codeword $\boldsymbol{a}$ that corresponds to the edge e, and where $\mathsf{d}(\cdot,\cdot)$ is the binary Hamming distance. This information is passed along the edge e to the corresponding decoder on the right-hand side of the bipartite graph. In the second iteration, for every vertex $w \in V_1$ the right decoder associated to it finds a q-ary codeword $\boldsymbol{b} = (b_1, \ldots, b_{\Delta_1}) \in \mathcal{C}_B$ that satisfies

$$\boldsymbol{b} = \arg \min_{\boldsymbol{x}=(x_1,\ldots,x_{\Delta_1}) \in \mathcal{C}_B} \sum_{i=1}^{\Delta_1} d_{w(i),x_i}(\boldsymbol{z}) \ ,$$

and writes b_i on the edge $w(i)$, $i = 1, \ldots, \Delta_1$.

Then, the decoder continues similarly to the decoder in [1].

Analysis

Lemma 6 *Consider the binary symmetric channel with crossover probability p. Let $0 < \varepsilon \ll p$. Then,*

$$\mathsf{H}_2^{-1}\left(\mathsf{H}_2(p) + \varepsilon(1 - \mathsf{H}_2(p))\right) = p + \frac{\varepsilon(1-\mathsf{H}_2(p))}{\log_2\left((1-p)/p\right)} - \frac{\varepsilon^2(1-\mathsf{H}_2(p))^2 \log_2 \mathsf{e}}{2p(p-1)\left(\log_2\left((1-p)/p\right)\right)^3} + O(\varepsilon^3).$$

The proof of this lemma appears in the Appendix.

Proposition 7 *Let C be the capacity of a BSC and let $\mathcal{C}_A$ be a random code with rate $R = (1-\varepsilon)C$. The decoding error probability of the code $\mathcal{C}_A$, under the maximum-likelihood decoding, behaves as $\exp\{-\Theta(\varepsilon^2)\}$ when $\varepsilon \to 0$.*

Proof. We start with the well-known expression for the exponent of the probability of the decoding error at the rates close to the channel capacity [6].

$$E_0(R,p) = \begin{cases} T(\delta,p) + R - 1 & \text{if } R_{crit} \le R < C \\ 1 - \log_2\left(1 + \sqrt{4p(1-p)}\right) - R & \text{if } R_{min} \le R < R_{crit} \\ -\delta \log_2 \sqrt{4p(1-p)} & \text{if } 0 \le R < R_{min} \ , \end{cases}$$

where R_{min} and R_{crit} are some threshold rates,

$$\delta = \delta_{GV}(R) = \mathsf{H}_2^{-1}(1-R) \ ,$$

and

$$T(x,y) = -x\log_2 y - (1-x)\log_2(1-y) \ .$$

At the code rates R which are close to C, the relevant expression for random coding exponent becomes

$$E_0(R,p) = T(\delta,p) + R - 1 \ . \tag{25}$$

Next, we express all terms of the relevant part of (25) in terms of ε. We recall, that $R = (1-\varepsilon)(1-\mathsf{H}_2(p))$ and, thus,

$$\mathsf{H}_2^{-1}(1-R) = \mathsf{H}_2^{-1}(\varepsilon + \mathsf{H}_2(p) - \varepsilon\mathsf{H}_2(p)) \ .$$

Thus, when disregarding $O(\varepsilon^3)$ term, the equation (25) becomes

$$\begin{aligned}
E_0(R,p) &= (1-\varepsilon)(1-\mathsf{H}_2(p)) - 1 + T\left(\mathsf{H}_2^{-1}(\varepsilon + \mathsf{H}_2(p) - \varepsilon\mathsf{H}_2(p)),\ p\right) \\
&\overset{(*)}{=} -\varepsilon - (1-\varepsilon)\mathsf{H}_2(p) + T\Bigg(p + \frac{\varepsilon(1-\mathsf{H}_2(p))}{\log_2((1-p)/p)} \\
&\qquad -\frac{\varepsilon^2(1-\mathsf{H}_2(p))^2\log_2\mathsf{e}}{2p(p-1)\left(\log_2((1-p)/p)\right)^3},\ p\Bigg) \\
&= -\varepsilon - (1-\varepsilon)\mathsf{H}_2(p) \\
&\qquad -\left(p + \frac{\varepsilon(1-\mathsf{H}_2(p))}{\log_2((1-p)/p)} - \frac{\varepsilon^2(1-\mathsf{H}_2(p))^2\log_2\mathsf{e}}{2p(p-1)\left(\log_2((1-p)/p)\right)^3}\right)\log_2 p \\
&\qquad -\left(1-p - \frac{\varepsilon(1-\mathsf{H}_2(p))}{\log_2((1-p)/p)} + \frac{\varepsilon^2(1-\mathsf{H}_2(p))^2\log_2\mathsf{e}}{2p(p-1)\left(\log_2((1-p)/p)\right)^3}\right)\log_2(1-p) \\
&= -\varepsilon(1-\mathsf{H}_2(p)) + \frac{\varepsilon(1-\mathsf{H}_2(p))(-\log_2 p + \log_2(1-p))}{\log_2((1-p)/p)} \\
&\qquad +\frac{\varepsilon^2(1-\mathsf{H}_2(p))^2\log_2\mathsf{e}(\log_2 p - \log_2(1-p))}{2p(p-1)\left(\log_2((1-p)/p)\right)^3} \\
&= \frac{\varepsilon^2(1-\mathsf{H}_2(p))^2\log_2\mathsf{e}}{2p(1-p)\left(\log_2((1-p)/p)\right)^2} = \varepsilon^2\cdot c_p \ ,
\end{aligned} \tag{26}$$

where $c_p > 0$ is a constant that depends only on the crossover probability p of the channel. Note that the transition $(*)$ follows from Lemma 6. □

Proposition 8 *If the codes $\mathbb{C}_{BZ3}$ have a positive error exponent, then $\Delta = \Omega(1/\varepsilon^2)$.*

Proof. It is shown in [3] that the decoding error probability of the code $\mathbb{C}_{BZ3}$, $\mathsf{Prob}_e(\mathbb{C}_{BZ3})$, satisfies

$$\mathsf{Prob}_e(\mathbb{C}_{BZ3}) \leq \exp\left\{-n\Delta l\delta_1(1+\alpha)^{-1}\cdot(E_0(R_0,p)-M\alpha)(1-o(1))\right\},$$

where α is a constant defined in [3] as $1>\alpha>2\lambda_1/d_1$, and

$$M = M(R,p) = \begin{cases} \frac{1}{2}\log_2((1-p)/p) & \text{if } R \leq R_{crit} \\ \log_2\left(\frac{\delta_{GV}(R)(1-p)}{(1-\delta_{GV}(R))p}\right) & \text{if } R \geq R_{crit} \end{cases},$$

$\delta_{GV} = \mathsf{H}_2^{-1}(1-R)$ is the Gilbert-Varshamov relative distance for the rate R, and $R_{crit} = 1-\mathsf{H}_2(\rho_0)$ is a so-called *critical rate*, where $\rho_0 = \sqrt{p}/(\sqrt{p}+\sqrt{1-p})$ (see [3] for details).

We are interested in small values of ε, i.e. $R \geq R_{crit}$. In this case, the value of $M(R,p)$ can be rewritten as

$$\begin{aligned} M(R,p) &= \log_2\left(\frac{\delta_{GV}(R)(1-p)}{(1-\delta_{GV}(R))p}\right) \\ &= \log_2\left(\frac{\mathsf{H}_2^{-1}(1-R)(1-p)}{(1-\mathsf{H}_2^{-1}(1-R))p}\right) \\ &= \log_2\left(\frac{\mathsf{H}_2^{-1}(\mathsf{H}_2(p)+\varepsilon-\varepsilon\mathsf{H}_2(p))(1-p)}{(1-\mathsf{H}_2^{-1}(\mathsf{H}_2(p)+\varepsilon-\varepsilon\mathsf{H}_2(p)))p}\right), \end{aligned}$$

where the last transition is due to $R=(1-\mathsf{H}_2(p))(1-\varepsilon)$. Using Lemma 6, the later relation becomes

$$M(R,p) = \log_2 \frac{\left(p+\frac{\varepsilon(1-\mathsf{H}_2(p))}{\log_2((1-p)/p)} - \frac{1}{2}\cdot\frac{\varepsilon^2(1-\mathsf{H}_2(p))^2\log_2\mathsf{e}}{p(p-1)(\log_2((1-p)/p))^3}\right)(1-p)}{\left(1-p-\frac{\varepsilon(1-\mathsf{H}_2(p))}{\log_2((1-p)/p)} + \frac{1}{2}\cdot\frac{\varepsilon^2(1-\mathsf{H}_2(p))^2\log_2\mathsf{e}}{p(p-1)(\log_2((1-p)/p))^3}\right)p} + O(\varepsilon^3).$$

When ignoring the terms of ε^2 and highest powers of ε, and denoting $\theta = \frac{\varepsilon(1-\mathsf{H}_2(p))}{\log_2((1-p)/p)}$, the later equation becomes

$$\begin{aligned} M(R,p) &= \log_2\left(\frac{p+\theta}{1-p-\theta}\cdot\frac{(1-p)}{p}\right) + O(\theta^2) \\ &= \log_2\left(\frac{1+\theta/p}{1-\theta/(1-p)}\right) + O(\theta^2) \\ &= \log_2\left((1+\theta/p)(1+\theta/(1-p))\right) + O(\theta^2) \\ &= \log_2\left(1+\theta/p+\theta/(1-p)\right) + O(\theta^2). \end{aligned}$$

Using Taylor's series for $\log(\cdot)$ around 1 we obtain

$$\begin{aligned} M(R,p) &= \log_2\mathsf{e}\cdot\left(\frac{\theta}{p}+\frac{\theta}{(1-p)}\right) + O(\theta^2) \\ &= \frac{\log_2\mathsf{e}}{p(1-p)}\cdot\theta + O(\theta^2), \end{aligned}$$

and switching back to ε notation this becomes

$$M(R,p) = \frac{\log_2 \mathsf{e}}{p(1-p)} \cdot \frac{\varepsilon(1-\mathsf{H}_2(p))}{\log_2\left((1-p)/p\right)} + O(\varepsilon^2) = \Theta(\varepsilon) \ . \tag{27}$$

Next, we evaluate the value of α. Recall that $\alpha > 2\lambda_1/d_1$. Since d_1 is linear in Δ_1, and $\Delta_1 < \Delta$, we have

$$\alpha = \Omega\left(4\sqrt{\Delta_1 - 1}/\Delta_1\right) = \Omega(1/\sqrt{\Delta_1}) = \Omega(1/\sqrt{\Delta}) \ .$$

In order to have a positive error exponent it is necessary that

$$\begin{aligned} E_0(R_0,p) - M\alpha > 0 \quad &\Rightarrow \quad \frac{E_0(R_0,p)}{M} > \alpha \\ &\Rightarrow \quad \frac{E_0(R_0,p)}{M} = \Omega(1/\sqrt{\Delta}) \ . \end{aligned}$$

Using Proposition 7, $E_0(R_0,p) = \Theta(\varepsilon^2)$, and thus from (27)

$$\varepsilon = \Omega(1/\sqrt{\Delta}) \qquad \Rightarrow \qquad \Delta = \Omega(1/\varepsilon^2) \ .$$

□

Assuming that the first two decoding iterations are as suggested in [3], we conclude that the time complexity of the decoding is $\exp\{\Omega(\Delta)\} = \exp\{\Omega(1/\varepsilon^2)\}$.

Appendix

Proof of Lemma 6.

Consider the value of the binary entropy function at the point $p + x$ for some small x. Using Taylor series around point p,

$$\mathsf{H}_2(p+x) = \mathsf{H}_2(p) + \mathsf{H}_2'(p) \cdot x + \frac{1}{2}\mathsf{H}_2''(p) \cdot x^2 + O(x^3) \ .$$

By calculation of the derivatives of the entropy function, one gets

$$\begin{aligned} \mathsf{H}_2'(\chi) &= -\log_2 \chi - \chi \cdot \frac{1}{\chi} \cdot \log_2 \mathsf{e} + \log_2(1-\chi) + (1-\chi) \cdot \frac{1}{\chi} \cdot \log_2 \mathsf{e} \\ &= \log_2\left(\frac{1-\chi}{\chi}\right) \ ; \end{aligned}$$

and

$$\mathsf{H}_2''(\chi) = \log_2 \mathsf{e} \cdot \left(-\frac{1}{1-\chi} - \frac{1}{\chi}\right) = \frac{\log_2 \mathsf{e}}{\chi(\chi-1)} \ .$$

Therefore,

$$\mathsf{H}_2(p+x) = \mathsf{H}_2(p) + \log_2\left(\frac{1-p}{p}\right)\cdot x + \frac{\log_2 \mathsf{e}}{p(p-1)}\frac{x^2}{2} + O(x^3)\ .$$

By applying the opposite of the binary entropy function on both sides of the equation,

$$\begin{aligned} p+x &= \mathsf{H}_2^{-1}\left(\mathsf{H}_2(p+x)\right) \\ &= \mathsf{H}_2^{-1}\left(\mathsf{H}_2(p) + \log_2\left(\frac{1-p}{p}\right)\cdot x + \frac{\log_2 \mathsf{e}}{p(p-1)}\cdot\frac{x^2}{2} + O(x^3)\right)\ . \end{aligned} \tag{28}$$

Denote by θ the value of $\log_2\left(\frac{1-p}{p}\right)\cdot x + \frac{\log_2 \mathsf{e}}{p(p-1)}\cdot\frac{x^2}{2}$, thus obtaining

$$p+x = \mathsf{H}_2^{-1}\left(\mathsf{H}_2(p) + \theta + O(x^3)\right)\ .$$

By solving the quadratic equation

$$\theta = \left(\log\left(\frac{1-p}{p}\right)\cdot x + \frac{1}{p(p-1)}\cdot\frac{x^2}{2}\right)\cdot\log_2 \mathsf{e}\ ,$$

or equivalently

$$x^2 + 2p(p-1)\log\left(\frac{1-p}{p}\right)x - \frac{2\theta p(p-1)}{\log_2 \mathsf{e}} = 0\ ,$$

we obtain two solutions for the intermediate x, namely

$$\begin{aligned} x &= \frac{1}{2}\left(-2p(p-1)\log\left(\frac{1-p}{p}\right) \pm \sqrt{4p^2(p-1)^2\log^2\left(\frac{1-p}{p}\right) + \frac{8\theta p(p-1)}{\log_2 \mathsf{e}}}\right) \\ &= -p(p-1)\log\left(\frac{1-p}{p}\right) \pm \sqrt{\left(p(p-1)\log\left(\frac{1-p}{p}\right)\right)^2 + \frac{2\theta p(p-1)}{\log_2 \mathsf{e}}}\ ; \end{aligned}$$

however, only one of these solutions is positive:

$$x = -p(p-1)\log\left(\frac{1-p}{p}\right) + \sqrt{\left(p(p-1)\log\left(\frac{1-p}{p}\right)\right)^2 + \frac{2\theta p(p-1)}{\log_2 \mathsf{e}}}\ .$$

The later equality can be rewritten as

$$x = p(p-1)\log\left(\frac{1-p}{p}\right)\cdot\left(-1 + \sqrt{1 + \frac{2\theta}{p(p-1)\left(\log\left((1-p)/p\right)\right)^2\log_2 \mathsf{e}}}\right)\ .$$

Using Taylor series approximation

$$\sqrt{1+\chi} = 1 + \frac{1}{2}\chi - \frac{1}{8}\chi^2 + O(\chi^3)\ ,$$

for small values of χ, this becomes

$$\begin{aligned} x &= p(p-1)\log\left(\frac{1-p}{p}\right)\cdot\left(-1+1+\frac{\theta}{p(p-1)\left(\log\left((1-p)/p\right)\right)^2\log_2\mathsf{e}}\right. \\ &\quad \left.-\frac{1}{2}\cdot\frac{\theta^2}{p^2(p-1)^2\left(\log\left((1-p)/p\right)\right)^4\left(\log_2\mathsf{e}\right)^2}+O(\theta^3)\right) \\ &= \frac{\theta}{\log_2\left((1-p)/p\right)}-\frac{1}{2}\cdot\frac{\theta^2\log_2\mathsf{e}}{p(p-1)\left(\log_2\left((1-p)/p\right)\right)^3}+O(\theta^3)\ . \end{aligned}$$

From (28) we have

$$\begin{aligned} \mathsf{H}_2^{-1}\left(\mathsf{H}_2(p)+\theta+O(\theta^3)\right) &= p+\frac{\theta}{\log_2\left((1-p)/p\right)} \\ &\quad -\frac{1}{2}\cdot\frac{\theta^2\log_2\mathsf{e}}{p(p-1)\left(\log_2\left((1-p)/p\right)\right)^3}+O(\theta^3)\ . \end{aligned}$$

For fixed value of $\chi<1$, the derivative of $\mathsf{H}_2^{-1}(\chi)$ is bounded and thus it is possible to take $O(\theta^3)$ term out of $\mathsf{H}_2^{-1}(\chi+O(\theta^3))$. Therefore we have

$$\mathsf{H}_2^{-1}\left(\mathsf{H}_2(p)+\theta\right)=p+\frac{\theta}{\log_2\left((1-p)/p\right)}-\frac{1}{2}\cdot\frac{\theta^2\log_2\mathsf{e}}{p(p-1)\left(\log_2\left((1-p)/p\right)\right)^3}+O(\theta^3)\ .$$

Finally, we substitute $\theta=\varepsilon(1-\mathsf{H}_2(p))$ and receive that

$$\begin{aligned} \mathsf{H}_2^{-1}\left(\mathsf{H}_2(p)+\varepsilon(1-\mathsf{H}_2(p))\right) &= p+\frac{\varepsilon(1-\mathsf{H}_2(p))}{\log_2\left((1-p)/p\right)} \\ &\quad -\frac{1}{2}\cdot\frac{\varepsilon^2(1-\mathsf{H}_2(p))^2\log_2\mathsf{e}}{p(p-1)\left(\log_2\left((1-p)/p\right)\right)^3}+O(\varepsilon^3)\ , \end{aligned}$$

thus completing the proof of the lemma. □

Acknowledgment

Authors are thankful to Ronny Roth and Tom Richardson for several helpful suggestions.

References

[1] A. Barg, G. Zémor, "Error Exponents of Expander Codes," *IEEE Trans. Inform. Theory,* 48 (2002), 1725–1729.

[2] A. Barg, G. Zémor, "Error Exponents of Expander Codes under Linear-Complexity Decoding," *SIAM J. Discrete Math.* 17(3) (2004), 426–445.

[3] A. Barg, G. Zémor, "Concatenated Codes: Serial and Parallel," *IEEE Trans. Inform. Theory,* 51 (2005), 1625–1634.

[4] G.D. Forney, Jr., *Concatenated Codes,* M.I.T. Press, Cambridge, Massachusetts, 1966.

[5] R.G. Gallager, *Low-Density Parity-Check Codes,* M.I.T. Press, Cambridge, Massachusetts, 1963.

[6] R.G. Gallager, *Information Theory and Reliable Communications,* John Wiley and Sons, New York, 1968.

[7] V. Guruswami, P. Indyk, "Near-Optimal Linear-Time Codes for Unique Decoding and New List-Decodable Codes Over Smaller Alphabets," *Proc. 34th Annual ACM Symposium on Theory of Computing (STOC),* Montréal, Quebec, Canada, 2002, 812–821.

[8] A. Khandekar, R.J. McEliece, "On the Complexity of Reliable Communication on the Erasure Channel," *Proc. IEEE International Symposium on Information Theory (ISIT),* Washington, DC, USA, 2001, p.1.

[9] A. Lubotsky, R. Philips, P. Sarnak, "Ramanujan Graphs," *Combinatorica,* 8 (1988), 261–277.

[10] P. Oswald, A. Shokrollahi, "Capacity-Achieving Sequences for the Erasure Channel," *IEEE Trans. Inform. Theory,* 48 (2002), 3017–3028.

[11] H. Pfister, I. Sason, R. Urbanke "Capacity-Achieving Ensembles for the Binary Erasure Channel With Bounded Complexity," *Proc. IEEE International Symposium on Information Theory (ISIT),* Chicago, USA, 2004, p.207.

[12] T. Richardson, R. Urbanke, "The Capacity of Low-Density Parity-Check Codes under Message-Passing Decoding," *IEEE Trans. Inform. Theory,* 47 (2001), 599–618.

[13] T. Richardson, A. Shokrollahi, R. Urbanke, "Design of Capacity-Approaching Low-Density Parity-Check Codes," *IEEE Trans. Inform. Theory,* 47 (2001), 619–637.

[14] R.M. Roth, V. Skachek, "On Nearly-MDS Expander Codes," *Proc. IEEE International Symposium on Information Theory (ISIT),* Chicago, USA, 2004, p.8.

[15] R.M. Roth, V. Skachek, "Improved Nearly-MDS Expander Codes," submitted to *IEEE Trans. Inform. Theory.* Appears also as Technical Report CS-2005-1, Computer Science Department, Technion, Israel, 2005.

[16] I. Sason, R. Urbanke, "Complexity Versus Performance of Capacity-Achieving Irregular Repeat-Accumulate Codes on the Binary Erasure Channel," *IEEE Trans. Inform. Theory,* 50 (2004), 1247–1256.

[17] M. Sipser, D.A. Spielman, "Expander Codes," *IEEE Trans. Inform. Theory,* 42 (1996), 1710–1722.

[18] G. Zémor, "On Expander Codes," *IEEE Trans. Inform. Theory,* 47 (2001), 835–837.