Decoding of Expander Codes at Rates Close to Capacity

Alexei Ashikhmin and Vitaly Skachek

Abstract— **The decoding error probability of codes is studied as a function of their block length. It is shown that the existence of codes with a polynomially small decoding error probability implies the existence of codes with an exponentially small decoding error probability. Specifically, it is assumed that there exists a family of codes of length N and rate $R = (1-\varepsilon)C$ (C is a capacity of a binary symmetric channel), whose decoding probability decreases polynomially in $1/N$. It is shown that if the decoding probability decreases sufficiently fast, but still only polynomially fast in $1/N$, then there exists another such family of codes whose decoding error probability decreases exponentially fast in N. Moreover, if the decoding time complexity of the assumed family of codes is polynomial in N and $1/\varepsilon$, then the decoding time complexity of the presented family is linear in N and polynomial in $1/\varepsilon$. 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 exponentially decaying (in N) error probability, and at the same time to have decoding time complexity linear in N and polynomial in $1/\varepsilon$.**

Index Terms— **Concatenated codes, decoding complexity, decoding error probability, error exponent, expander codes, IRA codes, iterative decoding, LDPC codes, linear-time decoding.**

I. Introduction

A classical work of Shannon states that reliable communications over a communication channel can be achieved for all information rates which are less than the certain threshold rate, capacity, which is a function of the channel characteristics. Codes and decoding algorithms that attain the channel capacity were extensively studied over the last decades. For such codes with respective decoding algorithms, at rates less than the capacity, the probability of decoding error approaches zero, as the code length grows.

Fastness of decrease of the decoding error probability as a function of the code length, N, is a characteristic of capacity-approaching codes, which was widely studied for many code families. However, this probability depends also on ratio between the channel capacity and an actual code rate. Namely, let the code rate be $R = (1-\varepsilon)C$, where C is the channel capacity. It is an interesting question to ask is how the decoding error probability depends on ε.

Another characteristic of (decoding algorithms of) codes is a time complexity of decoding. As of yet, there are known families of capacity-achieving codes (over various channels) with decoding algorithm time complexity only linear in N. However, one might look onto the decoding time complexity of code families in terms of ε. In the next two paragraphs we discuss these characteristics for two code families.

It is known that LDPC-type codes can attain a capacity of a binary erasure channel (BEC), the reader can refer to [11], [13], [15]. It is generally believed that LDPC-type 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. As to the decoding time complexity, it was conjectured in [9] that per-bit complexity of message-passing decoding (e.g. [6], [16]) 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. Lately, for LDPC-type codes with message-passing decoding over the BEC, the time complexity was shown to be linear in a code length and sub-linear in $1/\varepsilon$. More specifically, it was shown in [11] and [13] that the decoding complexity per bit for some sub-families of LDPC-type codes behaves as $O(\log(1/\varepsilon))$. Recently, in [14], IRA codes with bounded decoding complexity per bit were constructed.

In contrast, modifications of expander codes presented in [1], [2], [3], [17], [18] also attain the capacity of the memoryless q-ary symmetric channel, and the error probability decreases exponentially with the code length. Several recent works were devoted to analysis of fraction of errors that expander codes can correct (e.g. [4], [20], [21], [22]) and their rate-distance trade-offs (see [3], [8], [18]). While it is well known that there are decoders for expander codes having linear-time (in the code length) complexity, the dependence of this complexity on $1/\varepsilon$ was not studied. In the present work, we aim at studying this dependence. We investigate time complexity of decoding algorithms of expander codes in terms of ε, in particular for the codes in [1], [3]. We show that these specific codes have time complexity that is exponential in $1/\varepsilon^2$.

In this work, we study capacity-achieving codes over a binary symmetric channel (BSC). We show that if there exists a family of codes $\mathcal{C}_{in}$ of length N and rate $R = (1-\varepsilon)C$ (C is a BSC capacity), with the decoding probability vanishing inverse polynomially in N and ε (under conditions of our theorem), then there exists another such family of codes $\mathbb{C}_{cont}$ with the decoding error probability vanishing exponentially in N. Moreover, if the decoding time complexity of the codes $\mathcal{C}_{in}$ is polynomial in N and $1/\varepsilon$, then the decoding time complexity of the codes $\mathbb{C}_{cont}$ is linear in N and polynomial in $1/\varepsilon$.

The structure of this paper is as follows. In Section II, we describe the basic ingredients in our construction. The main result of our paper appears in Section III: we present a sufficient condition for existence of a family of codes with the decoding error probability vanishing exponentially fast. We also analyze the decoding time complexity of the presented codes. Finally, in Sections IV and V, we show that the codes in [1], [3] with their respective algorithms cannot be tuned to have decoding error probability that decreases exponentially fast (in terms of N), while the respective decoding algorithms have time complexity linear in N and polynomial in $1/\varepsilon$.

II. Preliminaries

A. Capacity-achieving codes with fast decoding

In this subsection we assume existence of some (family of) linear code $\mathcal{C}_{in}$, which achieves the capacity C of the BSC, and which has fast decoding algorithm. We denote its rate $R_{in} = (1-\varepsilon)C$, and its length n_{in} (constant for a fixed ε). Below, we discuss the parameters of this code.

Decoding complexity: we assume that the decoding complexity of $\mathcal{C}_{in}$ over the BSC is given by

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

where $\mathsf{s}, \mathsf{r} \geq 1$ are some constants. Let $\mathcal{D}_{in}$ be a decoder that have a time complexity as in (1).

The work of V. Skachek was supported in part by the DIMACS Special Focus Program in Computational Information Theory and Coding. The material in this paper was presented in part at the 2005 IEEE International Symposium on Information Theory, Adelaide, Australia, September 2005.

A. Ashikhmin is with the Mathematics of Communicatons Department, Bell Laboratories, Lucent Technologies, 600 Mountain Av., NJ 07974, USA (e-mail: aea@research.bell-labs.com).

V. Skachek is with the Computer Science Department, Technion – Israel Institute of Technology, Haifa 32000, Israel (e-mail: vitalys@cs.technion.ac.il). The work of this author was done in part when he was visiting DIMACS/Bell Laboratories.

Based on the results in [11], [13], [14], several LDPC-type code families (with respective message-passing decoding algorithms) do have such decoding complexity over the BEC (for $\mathsf{s} = 1$). There are no such results known for the BSC, although in the light of the surveyed works, this assumption sounds reasonable for LDPC-type codes over the BSC.

Decoding error probability: as of yet, there are no satisfying results on asymptotical behavior of the decoding error probability of LDPC-type codes over the binary erasure channel under the message-passing decoding, for rates near capacity of the BEC. The behavior of the decoding error probability of LDPC-type codes over other channels is even less investigated. In this work, we obtain a sufficient condition on the probability of the decoding error $\mathsf{Prob}_e(\mathcal{C}_{in})$ of the decoder $\mathcal{D}_{in}$ (for the $\mathcal{C}_{in}$) to guarantee the existence of a code with an exponentially-fast decreasing error probability.

Note: the results presented in the sequel are valid for any code $\mathcal{C}_{in}$ whose decoding time complexity and error probability are as stated above. However, LDPC-type codes are very promising candidates to meet these conditions, and in fact we do not see any other candidate at the present moment. Since there is no such candidate, it makes sense to speak about LDPC-type codes in this context.

B. Nearly-MDS expander codes

In this section, we consider linear-time decodable codes of rate $1 - \epsilon$ (for small $\epsilon > 0$) that can correct a fraction $\vartheta \epsilon^{\mathsf{b}}$ of errors, where $\vartheta > 0, \mathsf{b} > 0$ are constants. There are several code families known to date that can be shown to have the above property, and at the same time allow a linear-time (in a code length) decoding. In this connection, the reader can refer to [1], [3], [8], [20], [22]. However, as of yet, the codes in [17], [18] have the best relations between their rate, distance and alphabet size among all known expander-based linear-time decodable codes. Moreover, unlike the codes in [17], [18], not all aforementioned codes have decoding time complexity, which is polynomial in $1/\epsilon$.

Below, we recall the construction in [17], [18]. 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 $q \geq \Delta$.

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 use notation $[n, k, d]$ for a linear code of length n, dimension k, and minimum distance d.) We define the code $\mathbb{C} = (\mathcal{G}, \mathcal{C}_A : \mathcal{C}_B)$ as in [18], namely

$$\mathbb{C} \;=\; \Big\{ \boldsymbol{c} \in \mathbb{F}^N \,:\, (\boldsymbol{c})_{E(u)} \in \mathcal{C}_A \text{ for every } u \in A \text{ and } (\boldsymbol{c})_{E(u)} \in \mathcal{C}_B \text{ for every } u \in B \Big\} \;, \qquad (2)$$

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

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

Definition. An infinite sequence $\{a_i\}_{i=1}^{\infty}$, $a_i \stackrel{i\to\infty}{\longrightarrow} +\infty$, $a_i \in \mathbb{R}$, is called *a dense sequence of values* if $a_1 \leq 100$ and $a_{i+1} - a_i = o(a_i)$ (for $i \to \infty$). (The number 100 is a large absolute constant, the condition $a_1 \leq 100$ ensures that not all elements in the sequence are exponentially large.)

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 of Δ-regular bipartite Ramanujan graphs (e.g. [10], [12]), we have

$$\lambda_\mathcal{G} \leq 2\sqrt{\Delta - 1} \;. \qquad (3)$$

There are explicit constructions for such Δ-regular Ramanujan graph families for dense sequences of values Δ ([10], [12]).

It was shown in [18], that the code $\mathbb{C}_\Phi$ has the relative minimum distance

$$\delta_\Phi \geq \frac{\delta_B - \gamma_\mathcal{G}\sqrt{\delta_B/\delta_A}}{1 - \gamma_\mathcal{G}} \;. \qquad (4)$$

It is also known that the rate of $\mathbb{C}_\Phi$ is

$$R_\Phi \geq r_A + r_B - 1 \;.$$

The linear-time decoding algorithm $\mathcal{D}_\Phi$ in Figure 1 was proposed in [18]. It corrects any pattern of μ errors and ρ erasures such that $\mu + \frac{1}{2}\rho < \beta n$, where β is given by

$$\beta = \frac{(\delta_B/2) - \gamma_\mathcal{G}\sqrt{\delta_B/\delta_A}}{1 - \gamma_\mathcal{G}} \;. \qquad (5)$$

The number of iterations m in the algorithm was established in [18] 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.

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

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

<u>**For**</u> $i \leftarrow 1, 2, \ldots, m$ <u>**do**</u> {

  <u>**If**</u> i is even <u>**then**</u> $X \equiv A$, $\mathcal{D} \equiv \mathcal{D}_A$,

  <u>**else**</u> $X \equiv B$, $\mathcal{D} \equiv \mathcal{D}_B$.

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

}

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

Fig. 1. Decoder $\mathcal{D}_\Phi$ of Roth and Skachek for the code $\mathbb{C}_\Phi$.

The proof in [18] 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 (or less).

In the next proposition, we show that the parameters of the codes in [18] of rate $1 - \epsilon$ can be tuned to correct $\vartheta\epsilon$ errors for a constant $\vartheta > 0$.

Proposition 1: For any $\epsilon \in (0, 1)$, and for a sequence of alphabets $\{\Phi_i\}_{i=1}^{\infty}$ such that the sequence $\{\log_2 |\Phi_i|\}_{i=1}^{\infty}$ is dense, the codes $\mathbb{C}_\Phi$ (as above) of rate $R_\Phi \geq 1 - \epsilon$ (with decoder $\mathcal{D}_\Phi$) can correct a fraction $\vartheta\epsilon$ of errors, where $\vartheta > 0$ is some constant.

Proof. There is a dense sequence of values $\Delta \in \{\Delta_i\}_{i=1}^{\infty}$ such that there exists a family of Δ-regular bipartite Ramanujan graphs $\mathcal{G}$ (see [10], [12]). For any such value Δ, we can take both codes $\mathcal{C}_A$ and $\mathcal{C}_B$ to be GRS codes of length Δ over alphabet of size Δ, rate $r_A = r_B = 1 - \epsilon/2$ and relative minimum distance $\delta_A = \delta_B = \epsilon/2$.

Consider a code $\mathbb{C}_\Phi$ defined with respect to these $\mathcal{C}_A$ and $\mathcal{C}_B$. The rate R_Φ of $\mathbb{C}_\Phi$ satisfies $R_\Phi \geq r_A + r_B - 1 = 1 - \epsilon$. From (5), the fraction of errors that the decoder $\mathcal{D}_\Phi$ can correct is given by

$$\begin{aligned}\beta &= \frac{\delta_B/2 - \gamma_{\mathcal{G}}\sqrt{\delta_B/\delta_A}}{1-\gamma_{\mathcal{G}}} \\ &\geq \epsilon/4 - \gamma_{\mathcal{G}} \\ &= \epsilon/4 - 2\sqrt{\Delta-1}/\Delta \\ &\geq \epsilon/4 - 2/\sqrt{\Delta}\ .\end{aligned}$$

Take any Δ such that $\Delta > (16/\epsilon)^2$: for such Δ,

$$\beta > \vartheta\epsilon\ , \text{ where } \vartheta = 1/8\ .$$

Next, we observe that $|\Phi_i| = \Delta_i^{\Delta_i r_A}$. Based on the density of $\{\Delta_i\}_{i=1}^\infty$, we show the density of the sequence $\{\log_2|\Phi_i|\}_{i=1}^\infty$. Indeed, for any $i \in \mathbb{N}$,

$$\begin{aligned}&\lim_{i\to\infty}\frac{\log_2|\Phi_{i+1}| - \log_2|\Phi_i|}{\log_2|\Phi_i|} \\ &= \lim_{i\to\infty}\frac{\Delta_{i+1}\log_2\Delta_{i+1} - \Delta_i\log_2\Delta_i}{\Delta_i\log_2\Delta_i} \\ &= \lim_{i\to\infty}\left(\frac{\Delta_{i+1}\log_2\Delta_{i+1}}{\Delta_i\log_2\Delta_i}\right) - 1 \\ &= \lim_{i\to\infty}\left(\frac{\Delta_i + o(\Delta_i)}{\Delta_i}\cdot\frac{\log_2(\Delta_i+o(\Delta_i))}{\log_2\Delta_i}\right) - 1 \\ &= 1\ -\ 1\ =\ 0\ .\end{aligned}$$

Finally, from [10] and [12], Δ_1 can be taken small enough, such that $\log_2|\Phi_1| < 100$, as required. □

C. Concatenated codes

In this subsection, we revisit the definition of concatenated codes. The following ingredients will be used:

- A linear $[\mathsf{n}_{in}, k_{in}{=}R_{in}\mathsf{n}_{in}]$ code $\mathcal{C}_{in}$ over $\mathbb{F}$ (inner code).
- A linear code $\mathbb{C}_\Phi$ of length n and rate R_Φ over $\Phi = \mathbb{F}^{k_{in}}$ (outer code).
- A linear one-to-one mapping $\mathcal{E}_0\ :\ \Phi \to \mathcal{C}_{in}$.

The respective concatenated code $\mathbb{C}_{cont}$ is defined as

$$\begin{aligned}\mathbb{C}_{cont} = \Big\{&(\boldsymbol{c}_1|\boldsymbol{c}_2|\cdots|\boldsymbol{c}_n) \in \mathbb{F}^{n\cdot\mathsf{n}_{in}} : \boldsymbol{c}_i = \mathcal{E}_0(\Xi_i)\ , \\ &\text{for } i \in 1,2,\cdots,n, \text{ and } (\Xi_1\Xi_2\cdots\Xi_n) \in \mathbb{C}_\Phi\Big\}\ .\end{aligned}$$

The rate of $\mathbb{C}_{cont}$ is known to be $R_{cont} = R_{in}\cdot R_\Phi$.

Let $\mathcal{D}_{in} : \mathbb{F}^{\mathsf{n}_{in}} \to \mathcal{C}_{in}$ and $\mathcal{D}_\Phi : \Phi^n \to \mathbb{C}_\Phi$ be decoders for the codes $\mathcal{C}_{in}$ and $\mathbb{C}_\Phi$, respectively. A simple decoder $\mathcal{D}_{cont}$ for the code $\mathbb{C}_{cont}$ is presented in Figure 2. There exist more advanced decoders for the code $\mathbb{C}_{cont}$ (e.g. GMD decoding, [5]) that can correct more errors, but we consider the decoder $\mathcal{D}_{cont}$ due to its simplicity.

Input: received word $\boldsymbol{y} = (y_1\, y_2\, \cdots\, y_{n\cdot\mathsf{n}_{in}})$ in $\mathbb{F}^{n\cdot\mathsf{n}_{in}}$.
For $i \in 1,2,\cdots,n$ **do**
$u_i \leftarrow \mathcal{E}_0^{-1}\left(\mathcal{D}_{in}\left(\,(y_{j+(i-1)\cdot\mathsf{n}_{in}})_{j=1}^{\mathsf{n}_{in}}\,\right)\right)$.
Let $(z_1z_2\cdots z_n) \leftarrow \mathcal{D}_\Phi\left((u_1u_2\cdots u_n)\right)$.
Output: $(\mathcal{E}_0(z_1)|\mathcal{E}_0(z_2)|\cdots|\mathcal{E}_0(z_n))$.

Fig. 2. Decoder $\mathcal{D}_{cont}$ for the code $\mathbb{C}_{cont}$.

III. Main results

A. General settings

Consider a memoryless binary symmetric channel with crossover probability p. Its capacity is given by $C = 1 - \mathsf{H}_2(p)$, where $\mathsf{H}_2(\chi) = -\chi\log_2\chi - (1-\chi)\log_2(1-\chi)$ is the binary entropy function. Let $R = C(1-\varepsilon)$ be a design rate.

Take $\mathbb{F}$ to be GF(q), $q = 2^\ell$, $\ell \in \mathbb{N}$. Let $\mathcal{C}_{in}$ be a binary code of length n_{in} assumed in Section II-A. It can also be seen as an additive linear code of length $\mathsf{n}_{in} = n_{in}/\ell$ over $\mathbb{F}$. Let $\mathbb{C}_\Phi$ be a linear code of length n and rate R_Φ over an alphabet $\Phi = \mathbb{F}^{R_{in}\mathsf{n}_{in}}$. Pick some linear one-to-one mapping $\mathcal{E}_0\ :\ \Phi \to \mathcal{C}_{in}$. Let $\mathbb{C}_{cont}$ be a code, corresponding to a concatenation of the code $\mathcal{C}_{in}$ (as an inner code) with the code $\mathbb{C}_\Phi$ (as an outer code), as defined in Section II-C. Suppose $R_{cont} \geq R$ is a rate of the (binary) code $\mathbb{C}_{cont}$ and $N_{cont} = n\cdot n_{in}$ is its length. Denote by $\mathsf{Prob}_e(\mathbb{C}_{cont})$ its error probability, under the decoding by $\mathcal{D}_{cont}$.

The following lemma is based on the result in [5, Chapter 4.2].

Lemma 2: The error probability of the code $\mathbb{C}_{cont}$ (as defined in this section) under the decoding by $\mathcal{D}_{cont}$, when the error probability of the decoder $\mathcal{D}_{in}$ for the code $\mathcal{C}_{in}$ is $\mathsf{Prob}_e(\mathcal{C}_{in})$, and the decoder $\mathcal{D}_\Phi$ corrects any pattern of less than βn errors, is bounded by

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

where E is a constant given by

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

If a right-hand side of (6) is negative, we assume that E is zero.

The proof of this lemma appears in Appendix A.

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 [5, Chapter 4.2] for details. We omit this analysis for the sake of simplicity.

B. Sufficient condition

In this subsection, we derive a sufficient condition on the probability of decoding error of the code $\mathcal{C}_{in}$ for providing a positive error exponent for the code $\mathbb{C}_{cont}$ as defined in subsection III-A. Below, 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 3: Consider the BSC, and let C be its capacity. Suppose that the following two conditions hold:

(i) There exist constants $\mathsf{b} > 0$, $\vartheta > 0$, $\varepsilon_1 \in (0,1)$, such that for any ϵ, $0 < \epsilon < \varepsilon_1$, and for a sequence of alphabets $\{\Phi_i\}_{i=1}^\infty$ where the sequence $\{\log_2|\Phi_i|\}_{i=1}^\infty$ is dense, there exists a family of codes $\mathbb{C}_\Phi$ of rate $1-\epsilon$ (with their respective decoders) that can correct a fraction $\vartheta\epsilon^{\mathsf{b}}$ of errors.

(ii) There exist constants $\varepsilon_2 \in (0,1)$ and $h_0 > 0$, such that for any ϵ, $0 < \epsilon < \varepsilon_2$, the decoding error probability of a family of codes $\mathcal{C}_{in}$ satisfies

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

Then, for any rate $R < C$, there exist a family of the codes $\mathbb{C}_{cont}$ as defined in subsection III-A (with respective decoder) that has an exponentially decaying (in N_{cont}) error probability.

Proof. Let $R = (1-\varepsilon)C$ be a design rate of the code $\mathbb{C}_{cont}$, and $\varepsilon > 0$ be small (namely, $\varepsilon < \min\{\varepsilon_1, \varepsilon_2\}$). Let κ be a constant,

$0 < \kappa < 1$, which will be defined later, and let the rate of the code $\mathcal{C}_{in}$ be $R_{in} = (1 - \kappa\,\varepsilon)C$. We set the rate of $\mathbb{C}_\Phi$ as

$$R_\Phi = \frac{R}{R_{in}} = \frac{1-\varepsilon}{1-\kappa\varepsilon} = 1 - (1-\kappa)\varepsilon - \Theta(\varepsilon^2) \ .$$

Then, by condition (i), the fraction β of errors correctable by the code $\mathbb{C}_\Phi$ is at least $\beta \geq \vartheta((1-\kappa)\cdot\varepsilon)^{\mathsf{b}}$.

For an alphabet Φ, the length n_{in} of the code $\mathcal{C}_{in}$ is given by

$$n_{in} = \frac{\log_2 |\Phi|}{R_{in}} \ .$$

We select the smallest $\Phi \in \{\Phi_i\}_{i=1}^\infty$ such that

$$\log_2 |\Phi| \geq \frac{1}{(\kappa\varepsilon)^{h_0}} \ ,$$

and, so,

$$n_{in} > \frac{1}{(\kappa\varepsilon)^{h_0}} \ , \quad (7)$$

Next, we use Lemma 2 to evaluate the decoding error probability of the code $\mathbb{C}_{cont}$. It holds for small positive values of β that

$$(1-\beta)\ln(1-\beta) > -\beta \ ,$$

and thus, from Lemma 2 we obtain (by ignoring the positive term $-(1-\beta)\ln(1-\mathsf{Prob}_e(\mathcal{C}_{in}))$ in (6)),

$$\begin{aligned}
&\mathsf{Prob}_e\left(\mathbb{C}_{cont}\right) \\
&\quad < \quad \exp\left\{-n\cdot\left(-\beta\ln\left(\mathsf{Prob}_e(\mathcal{C}_{in})\right) + \beta\ln\beta - \beta\right)\right\} \\
&\quad = \quad \exp\left\{-N_{cont}\frac{\beta}{n_{in}}\left(\ln\beta - \ln\left(\mathsf{Prob}_e(\mathcal{C}_{in})\right) - 1\right)\right\} \ .
\end{aligned}$$

In order to have a positive error exponent, we require that

$$\ln\beta - \ln\left(\mathsf{Prob}_e(\mathcal{C}_{in})\right) - 1 > 0 \ ,$$

or, equivalently,

$$\beta > \mathsf{e}\cdot\mathsf{Prob}_e(\mathcal{C}_{in}) \ . \quad (8)$$

The decoding error probability of the selected code $\mathcal{C}_{in}$ satisfies:

$$\begin{aligned}
&\mathsf{Prob}_e\left(\mathcal{C}_{in}\left[(1-\kappa\varepsilon)C,\ n_{in}\right]\right) \\
&\quad < \quad \mathsf{Prob}_e\left(\mathcal{C}_{in}\left[(1-\kappa\varepsilon)C,\ \frac{1}{(\kappa\varepsilon)^{h_0}}\right]\right) \\
&\quad < \quad (\kappa\varepsilon)^{\mathsf{b}} \ \leq \ \frac{\vartheta((1-\kappa)\varepsilon)^{\mathsf{b}}}{\mathsf{e}} \ , \quad (9)
\end{aligned}$$

where the first inequality is due to (7), the second inequality follows from condition (ii), and the third inequality can be satisfied by a selection of a small constant κ such that $\kappa^{\mathsf{b}} \leq \vartheta(1-\kappa)^{\mathsf{b}}/\mathsf{e}$.

The inequality (9) implies (8), as required. □

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 decoder) is bounded by

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

We choose $h_0 = \mathsf{b}+5$ (where b is as in condition (i) of Theorem 3). There obviously exists ε_2 such that for every $0 < \epsilon < \varepsilon_2$, for the code $\mathcal{C}_{in}$ of length $n_{in} = 1/\epsilon^{h_0}$ and rate $R_{in} = (1-\epsilon)C$,

$$\mathsf{Prob}_e(\mathcal{C}_{in}) < \frac{1}{n_{in}}\cdot\frac{1}{\epsilon^4} = \epsilon^{h_0}\cdot\frac{1}{\epsilon^4} = \epsilon^{\mathsf{b}+1} < \epsilon^{\mathsf{b}} \ . \quad (10)$$

From the expression (10) we see that condition (ii) of Theorem 3 is satisfied. This selection guarantees existence of a positive error exponent for the code $\mathbb{C}_{cont}$.

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

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

We choose $h_0 = 3$. There obviously exists ε_2 such that for every $0 < \epsilon < \varepsilon_2$, for the code $\mathcal{C}_{in}$ of length $n_{in} = 1/\epsilon^{h_0}$ and rate $R_{in} = (1-\epsilon)C$, and for every $\mathsf{b} > 0$,

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

and therefore Theorem 3 yields existence of a positive error exponent for the code $\mathbb{C}_{cont}$.

C. Example

In this subsection, we consider a specific case of decoding error probability for the code $\mathcal{C}_{in}$. Theorem 3 can be directly applied in this case. However, we conduct a direct minimization of the decoding error probability of the code $\mathbb{C}_{cont}$, which is obtained by concatenation of the code $\mathbb{C}_\Phi$ in [18] with the assumed code $\mathcal{C}_{in}$, and obtain an analytical expression on the error exponent. We show that the overall decoding error probability for this code $\mathbb{C}_{cont}$ has a positive error exponent.

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, $t \geq 1$.

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

Let $R = (1-\varepsilon)C$ be a design code rate. 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

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

Next, we select the parameters of the code $\mathbb{C}_\Phi$ in [18], which serves as an outer code. Take $\mathcal{C}_A$ and $\mathcal{C}_B$ as GRS codes over $\mathbb{F}$, with $|\mathbb{F}| = \Delta$. 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 $\Delta = \varrho/\varepsilon^2$, where ϱ is a constant, such that

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

We have,

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

By our selection (see (3)),

$$\gamma_{\mathcal{G}} \leq \frac{2}{\sqrt{\Delta}} = \frac{2\varepsilon}{\sqrt{\varrho}} \ .$$

We obtain from (4),

$$\beta > (\delta_B/2) - \gamma_{\mathcal{G}}\sqrt{\delta_B/\delta_A} > \vartheta\varepsilon + o(\varepsilon) \ , \quad (11)$$

where

$$0 < \vartheta = \frac{\eta(1-\kappa)}{2} - 2\sqrt{\frac{\eta}{\varrho(1-\eta)}}$$

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

The number of bits needed to represent each symbol of Φ is $\log_2|\Phi| = r_A\Delta\cdot\log_2|\mathbb{F}|$. Recall that $r_A = 1 - O(\varepsilon)$. Therefore, the length n_{in} of the binary code $\mathcal{C}_{in}$ is given by

$$\begin{aligned}
n_{in} \quad &= \quad \frac{r_A\Delta}{R_{in}}\cdot\log_2(\Delta) \\
&= \quad \frac{(1-O(\varepsilon))\varrho}{R_{in}\,\varepsilon^2}\ \cdot\ \log_2\left(\frac{\varrho}{\varepsilon^2}\right) \\
&= \quad \frac{\varrho\log_2(\varrho/\varepsilon^2)}{R_{in}\,\varepsilon^2}\ +\ o\left(\frac{\varrho\log_2(\varrho/\varepsilon^2)}{R_{in}\,\varepsilon^2}\right) \ , \quad (12)
\end{aligned}$$

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

$$\mathsf{Prob}_e(\mathcal{C}_{in}) \leq \left(\frac{\varepsilon^2 R_{in}}{\varrho \log_2(\varrho/\varepsilon^2)} \right)^t . \qquad (13)$$

We substitute the expressions in (11) (only the main term) and (13) into the result of Lemma 2 to obtain

$$\begin{aligned} \mathsf{Prob}_e(\mathbb{C}_{cont}) \ &< \\ \exp\Bigg\{ - n\Bigg(&- \vartheta\varepsilon \cdot t \ln\left(\frac{\varepsilon^2 R_{in}}{\varrho \log_2(\varrho/\varepsilon^2)} \right) \\ &- (1-\vartheta\varepsilon) \ln\left(1 - \left(\frac{\varepsilon^2 R_{in}}{\varrho \log_2(\varrho/\varepsilon^2)} \right)^t \right) \\ &+ \vartheta\varepsilon \ln(\vartheta\varepsilon) + (1-\vartheta\varepsilon)\ln(1-\vartheta\varepsilon) \Bigg) \Bigg\} . \qquad (14) \end{aligned}$$

Note that for small $\varepsilon > 0$,

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

and

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

Hence, the equation (14) (when neglecting $o(\varepsilon)$ terms) becomes

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

Using substitution of the expression (12) for n_{in}, the latter equation can be rewritten as

$$\begin{aligned} \mathsf{Prob}_e(\mathbb{C}_{cont}) \ &< \\ \exp\Bigg\{ &- \frac{N_{cont}\vartheta\varepsilon \cdot \varepsilon^2 R_{in}}{2\varrho \, (\log_2(1/\varepsilon) + \Theta(1))} \\ &\cdot \Bigg((2t-1)\ln(1/\varepsilon) + t\ln(1/R_{in}) \\ &\qquad + \ t \ln\ln(1/\varepsilon) + \Theta(1) \Bigg) \Bigg\} . \qquad (15) \end{aligned}$$

The dominating term in the expression

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

is $(2t-1)\ln(1/\varepsilon)$. By taking into account that $R_{in} = C(1-O(\varepsilon))$, the equation (15) can be rewritten, when ignoring all but the main term, as

$$\begin{aligned} &\mathsf{Prob}_e(\mathbb{C}_{cont}) < \\ &\exp\left\{ - N_{cont} \cdot \left(\frac{(2t-1)\,\vartheta\,\varepsilon^3\, C}{2\varrho \cdot \log_2 \mathsf{e}} + o(\varepsilon^3) \right) \right\} . \end{aligned}$$

Thus, the decoding error probability is given by

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

where

$$\begin{aligned} E(C,\varepsilon) \ &= \ \max_{\varrho,\vartheta} \left\{ \frac{\vartheta}{\varrho} \right\} \cdot \frac{(2t-1)\,C}{2 \cdot \log_2 \mathsf{e}} \cdot \varepsilon^3 \\ &= \ \max_{\kappa,\,\eta,\,\varrho} \left\{ \frac{\eta(1-\kappa)}{2\varrho} - 2\sqrt{\frac{\eta}{\varrho^3(1-\eta)}} \right\} \\ &\qquad \cdot \frac{(2t-1)\,C}{2 \cdot \log_2 \mathsf{e}} \cdot \varepsilon^3 , \qquad (16) \end{aligned}$$

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

$$\kappa \in (0,1) \ ; \ \eta \in (0,1) \ ; \ \varrho > \frac{16}{\eta(1-\eta)(1-\kappa)^2} . \qquad (17)$$

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 (16) to obtain

$$\Upsilon = \max_{\eta,\,\varrho} \left\{ \frac{\eta}{2\varrho} - 2\sqrt{\frac{\eta}{\varrho^3(1-\eta)}} \right\} . \qquad (18)$$

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 the expression (18) and finding its maximum, we have $\eta = 2/3$ and $\varrho = 162$. These values obviously satisfy condition (17). 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

$$E(C,\varepsilon) = \frac{(2t-1)\,C}{2916 \cdot \log_2 \mathsf{e}} \cdot \varepsilon^3 .$$

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

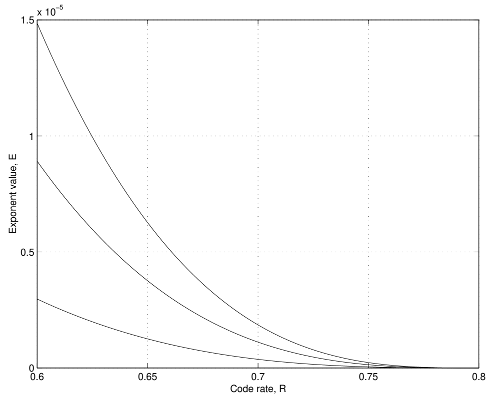

Fig. 3. Error exponent $E(C,\varepsilon)$ for the code $\mathbb{C}_{cont}$.
Selection: $\mathsf{Prob}_e(\mathcal{C}_{in}) = 1/n_{in}^t$; $C = 0.8$; $t = 1, 2, 3$ (bottom to top).

D. Decoding complexity

In this subsection, we show that under the assumption in Section II-A on the decoding time complexity of the code $\mathcal{C}_{in}$, and if the parameters of the codes are selected as in the proof of Theorem 3, then the decoding time complexity of the respective code $\mathbb{C}_{cont}$ is linear in the overall length N_{cont} and inverse polynomial in the gap from capacity ε.

Theorem 4: Consider the BSC, and let C be its capacity. Let $R = (1-\varepsilon)C$ be a design rate. Suppose that the following two conditions hold:

(i) Let $\mathbb{C}_\Phi$ be a (family of) code defined in Section II-B of rate $R_\Phi = (1-\varepsilon)/(1-\kappa\varepsilon)$, $\kappa \in (0,1)$ is a constant, over a smallest alphabet Φ satisfying $\log_2|\Phi| \geq 1/(\kappa\varepsilon)^{h_0}$ from a dense sequence $\{\log_2|\Phi_i|\}_{i=1}^{\infty}$, and $h_0 > 0$ is a constant.

(ii) Let $\mathcal{C}_{in}$ be a code of rate $R_{in} = (1-\kappa\varepsilon)C$ with the decoding complexity over the BSC of capacity C given by

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

where $\mathsf{s}, \mathsf{r} \geq 1$ are some constants.

Then, the time complexity of the respective code $\mathbb{C}_{cont}$, when decoded by $\mathcal{D}_{cont}$, is given by

$$N_{cont} \cdot \textsc{Poly}(1/\varepsilon) .$$

Proof. Below we count the total number of operations when decoding the code $\mathbb{C}_{cont}$ by the decoder $\mathcal{D}_{cont}$. There are two main steps.

- Step 1: n applications of the decoder $\mathcal{D}_{in}$ on the binary word of length n_{in}.
- Step 2: one application of the decoder $\mathcal{D}_{\Phi}$ on the word of length n over Φ.

In addition, there are n applications of each of the mappings $\mathcal{E}_0$ and $\mathcal{E}_0^{-1}$.

We separately count the number of operations during each step.

- Step 1: By the assumption on the decoding complexity of $\mathcal{D}_{in}$, n applications of this decoder result in time

$$O\left(n \cdot n_{in}^{\mathsf{s}} \cdot \frac{1}{\varepsilon^{\mathsf{r}}}\right) = O\left(N_{cont} \cdot n_{in}^{\mathsf{s}-1} \cdot \frac{1}{\varepsilon^{\mathsf{r}}}\right) . \qquad (19)$$

From the definition of $\mathbb{C}_{cont}$, $n_{in} = \log_2|\Phi| \,/\, R_{in}$, so, we have

$$n_{in} = \frac{\log_2|\Phi|}{(1-\kappa\varepsilon)C} .$$

By using the density of values of $\log_2|\Phi|$, we have $\log_2|\Phi| \in \textsc{Poly}(1/\varepsilon)$, thus yielding $n_{in} \in \textsc{Poly}(1/\varepsilon)$. By substitution into (19), we obtain that the time complexity of Step 1 is $N_{cont} \cdot \textsc{Poly}(1/\varepsilon)$.

- Step 2: it is shown in [18] that the number of applications of decoders $\mathcal{D}_A$ and $\mathcal{D}_B$ on the word of $\mathbb{C}_{\Phi}$ of length n over Φ is bounded by $\omega \cdot n$, where

$$\omega = 2 \cdot \left\lceil \frac{\ln\left(\dfrac{\Delta\beta\sqrt{\sigma}}{\beta-\sigma}\right)}{\ln\left(\dfrac{\delta_A\delta_B}{4\gamma_{\mathcal{G}}^2}\right)} \right\rceil + \frac{1+\dfrac{\delta_A}{\delta_B}}{1-\left(\dfrac{4\gamma_{\mathcal{G}}^2}{\delta_A\delta_B}\right)^2} ,$$

and σ is an actual number of errors in the word. Thus, if the ratio σ/β is bounded away from 1, and $\mathcal{G}$ is a Ramanujan graph, then the value of ω is bounded from above by an absolute constant (independent of Δ).

The decoders $\mathcal{D}_A$ and $\mathcal{D}_B$ are applied on the words of length $\Delta \in \textsc{Poly}(1/\varepsilon)$. When half minimum distance decoders for GRS codes are used, their complexity is polynomial in $1/\varepsilon$. Therefore, the decoding complexity in Step 2 is bounded by

$$n \cdot \textsc{Poly}(1/\varepsilon) \leq N_{cont} \cdot \textsc{Poly}(1/\varepsilon) .$$

Each application of mapping $\mathcal{E}_0$ or $\mathcal{E}_0^{-1}$ is equivalent to multiplication of a vector by a matrix, where the number of rows and columns in the matrix is $\textsc{Poly}(1/\varepsilon)$. This can be done in time $\textsc{Poly}(1/\varepsilon)$.

Summing up the decoding complexities of all steps of the decoder, we obtain that the total number of operations is bounded by

$$N_{cont} \cdot \textsc{Poly}(1/\varepsilon) .$$

$\square$

Note. The result in Theorem 4 is still valid if the outer code $\mathbb{C}_{\Phi}$ be replaced by any other code of rate $1-\Theta(\varepsilon)$, whose decoding time complexity is linear in n and polynomial in $1/\varepsilon$, for a log-dense sequence of alphabet sizes.

IV. Time complexity of decoder in [1]

Similarly to Section III, assume in this and the next sections that C is the capacity of the BSC with crossover probability p, and the design code rate is $R = (1-\varepsilon)C$. Our purpose is to compare the parameters of the codes from Section III with codes presented by Barg and Zémor in [1] and [3] (with their respective decoding algorithms). In the sequel we show that the parameters of the codes from [1] and [3] cannot be modified such that the decoding time complexity would be only sub-exponential in $1/\varepsilon$ while keeping a non-zero error exponent. The reason is this: both decoding algorithms in [1] and [3] make use of sub-routines (decoders for small constituent codes) that have time complexity exponential in a degree of underlying expander graph. This degree, in turn, depends (at least) polynomially on $1/\varepsilon$.

A. Construction

We briefly recall the construction and the decoder 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 the finite field $\mathbb{F}$ 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 (2), with respect to $\mathcal{C}_A$ and $\mathcal{C}_B$ as defined in this paragraph.

B. Decoding

Let us submit a word $\boldsymbol{c} = (c_e)_{e\in E} \in \mathbb{C}_{BZ2}$ to the BSC. Assume that $\boldsymbol{y} = (y_e)_{e\in E}$ is a received (erroneous) word. A formal definition of the decoder $\mathcal{D}_{BZ2}$ appears in Figure 4. The number of iterations m

Input: Received word $\boldsymbol{y} = (y_e)_{e\in E}$ 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).

Fig. 4. Decoder $\mathcal{D}_{BZ2}$ of Barg and Zémor for the code $\mathbb{C}_{BZ2}$.

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.

The analysis of codes in [1] is divided into two cases. In the first case, the codes $\mathcal{C}_A$ and $\mathcal{C}_B$ over $\mathbb{F} = \mathrm{GF}(2)$ are considered. In the second case, the analysis is generalized toward field sizes, which are large powers of 2. We analyze these two cases separately.

C. Analysis: binary codes

In the binary case, following the analysis of [1] it is possible to show that for the code $\mathbb{C}_{BZ2}$ with the decoder $\mathcal{D}_{BZ2}$, the decoding error probability, $\mathsf{Prob}_e(\mathbb{C}_{BZ2})$, is bounded by

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

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

$$\max_{R \le R_0 < C} \left\{ E_0(R_0,p) \left(\frac{\mathsf{H}_2^{-1}(R_0-R)}{2} - \Theta\left(\frac{1}{\sqrt{\Delta}}\right) \right) \right\} , \qquad (20)$$

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

Proposition 5: If the codes $\mathbb{C}_{BZ2}$ (binary, as assumed in this subsection), have a positive error exponent under the decoding by $\mathcal{D}_{BZ2}$, 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 $R_0 - R \le C - R = C\varepsilon \le \varepsilon$. It follows from (20) that

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

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

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

D. Analysis: codes over large fields

Suppose that the size of the field $\mathbb{F}$ is a large power of 2. In this case, for the code $\mathbb{C}_{BZ2}$ under the decoding by $\mathcal{D}_{BZ2}$, the decoding error probability $\mathsf{Prob}_e(\mathbb{C}_{BZ2})$ is bounded by

$$\mathsf{Prob}_e(\mathbb{C}_{BZ2},p) \le \exp\{-\alpha N f_2(R,p)\} \ ,$$

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

$$\max_{R \le R_0 < C} \left\{ E_0(R_0,p) \left(\frac{R_0-R}{2} - \Theta\left(\frac{1}{\sqrt{\Delta}}\right) \right) \right\} \ .$$

In this case, Proposition 5 can be rewritten as

Proposition 6: If the codes $\mathbb{C}_{BZ2}$ (over large $\mathbb{F}$, as assumed in this subsection) have a positive error exponent under the decoding by $\mathcal{D}_{BZ2}$, then $\Delta = \Omega\left(1/\varepsilon^2\right)$.

The proof is very similar to that of Proposition 5.

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

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

V. Time complexity of decoder in [3]

A. 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 $\mathcal{C}_{aux}$.

B. 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 $\mathsf{d}(\cdot,\cdot)$ 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].

C. Analysis

Lemma 7: Let p satisfy $0 < p < \frac{1}{2}$, and let $0 < \varepsilon \ll p$. Then,

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

The proof of this lemma appears in the Appendix B.

Proposition 8: Let C be the capacity of the BSC. The decoding error probability of a random code of rate $R = (1-\varepsilon)C$, 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 probability exponent of the decoding error of a random code under the maximum-likelihood decoding [6], [7].

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

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

$$E_0(R,p) = T(\delta,p) + R - 1 \ . \qquad (21)$$

Next, we express all terms of the relevant part of (21) 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 (21) becomes

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

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 7. □

Proposition 9: 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

$$
\begin{aligned}
\mathsf{Prob}_e(\mathbb{C}_{BZ3}) \le \exp\big\{&-n\Delta l \delta_1 (1+\alpha)^{-1} \\
&\cdot (E_0(R_0,p) - M\alpha)(1-o(1))\big\} \ ,
\end{aligned}
$$

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

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

$\delta_{GV}(R) = \mathsf{H}_2^{-1}(1-R)$ 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. $R \ge R_{crit}$. In this case, the value of $M(R,p)$ can be rewritten as

$$
\begin{aligned}
M(R,p) &= \log_2\left(\frac{\delta_{GV}(R)(1-p)}{(1-\delta_{GV}(R))p}\right) \\
&= \log_2\left(\frac{\mathsf{H}_2^{-1}(1-R)(1-p)}{(1-\mathsf{H}_2^{-1}(1-R))p}\right) \\
&= \log_2\left(\frac{\mathsf{H}_2^{-1}(\mathsf{H}_2(p)+\varepsilon-\varepsilon\mathsf{H}_2(p))(1-p)}{(1-\mathsf{H}_2^{-1}(\mathsf{H}_2(p)+\varepsilon-\varepsilon\mathsf{H}_2(p)))p}\right) \ , \qquad (22)
\end{aligned}
$$

where the last transition is due to $R = (1-\mathsf{H}_2(p))(1-\varepsilon)$. Using Lemma 7, the equality (22) becomes

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

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

$$
M(R,p) = \frac{\log_2 \mathsf{e}}{p(1-p)}\cdot\frac{\varepsilon(1-\mathsf{H}_2(p))}{\log_2((1-p)/p)} + O(\varepsilon^2) = \Theta(\varepsilon) \ . \qquad (23)
$$

Next, we evaluate the value of α. Recall that $\alpha > 2\lambda_1/d_1$, and $d_1 \le \Delta_1 \le \Delta$. We have

$$
\alpha > \frac{2\lambda_1}{d_1} \ge \frac{4\sqrt{\Delta_1 - 1}}{\Delta_1} \ge \frac{4\sqrt{\Delta - 1}}{\Delta} = \Theta\left(\frac{1}{\sqrt{\Delta}}\right) \ .
$$

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

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

Using Proposition 8, $E_0(R_0,p) = \Theta(\varepsilon^2)$, and thus from (23)

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

□

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 A

Proof of Lemma 2.

We analyze the error exponent, following the guidelines of the analysis of Forney [5, 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

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

Denote

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

Using the Chernoff bound, we obtain

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

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

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

is achieved when

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

and the maximum is

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

thus completing the proof. □

APPENDIX B

Proof of Lemma 7.

Consider the value of the binary entropy function at the point $p+x$ for small $x>0$. 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 obtains

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

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,

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

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

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

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

$$p+x = \mathsf{H}_2^{-1}\left(\mathsf{H}_2(p)+\theta+O(x^3)\right) . \qquad (24)$$

By solving the quadratic equation

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

or equivalently

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

we obtain two solutions for the intermediate x, namely

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

however, only one of these solutions is positive:

$$\begin{aligned}x &= -p(p-1)\ln\left(\frac{1-p}{p}\right) \\ &+ \sqrt{\left(p(p-1)\ln\left(\frac{1-p}{p}\right)\right)^2+\frac{2\theta p(p-1)}{\log_2\mathsf{e}}} .\end{aligned}$$

The later equality can be rewritten as

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

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

$$\begin{aligned}x &= p(p-1)\ln\left(\frac{1-p}{p}\right)\cdot \\ &\quad \Bigg(-1+1+\frac{\theta}{p(p-1)\left(\ln((1-p)/p)\right)^2\log_2\mathsf{e}} \\ &\quad -\frac{1}{2}\cdot\frac{\theta^2}{p^2(p-1)^2\left(\ln((1-p)/p)\right)^4(\log_2\mathsf{e})^2}+O(\theta^3)\Bigg) \\ &= \frac{\theta}{\log_2((1-p)/p)} \\ &\quad -\frac{1}{2}\cdot\frac{\theta^2\log_2\mathsf{e}}{p(p-1)\left(\log_2((1-p)/p)\right)^3}+O(\theta^3) . \qquad (25)\end{aligned}$$

We substitute the evaluation of value of x in (25) into the equation (24). Thus, we obtain

$$\begin{aligned}\mathsf{H}_2^{-1}\left(\mathsf{H}_2(p)+\theta+O(\theta^3)\right) &= p+\frac{\theta}{\log_2((1-p)/p)} \\ &-\frac{1}{2}\cdot\frac{\theta^2\log_2\mathsf{e}}{p(p-1)\left(\log_2((1-p)/p)\right)^3}+O(\theta^3) . \qquad (26)\end{aligned}$$

If $p<\frac{1}{2}$ is fixed and θ is small, then the value of $\mathsf{H}_2(p)+\theta$ is bounded away from 1. In this case, the derivative of $\mathsf{H}_2^{-1}(\chi)$ at point $\chi=\mathsf{H}_2(p)+\theta$ is bounded, and, therefore

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

Then, the equality (26) becomes

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

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

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

thus completing the proof of the lemma. □

Acknowledgment

The authors are thankful to Ronny Roth and Tom Richardson for several helpful suggestions. The authors would also like to thank the anonymous Reviewer C and Amin Shokrollahi for helpful comments that substantially improved the manuscript. The support of DIMACS is gratefully acknowledged.

References

[1] A. Barg and G. Zémor, "Error exponents of expander codes," *IEEE Trans. Inf. Theory,* vol. 48, no. 6, pp. 1725–1729, June 2002.

[2] A. Barg and G. Zémor, "Error exponents of expander codes under linear-complexity decoding," *SIAM J. Discrete Math.,* vol. 17, no. 3, pp. 426–445, 2004.

[3] A. Barg and G. Zémor, "Concatenated codes: serial and parallel," *IEEE Trans. Inform. Theory,* vol. 51, no. 5, pp. 1625–1634, May 2005.

[4] J. Feldman, T. Malkin, C. Stein, R.A. Servedio, and M.J. Wainwright, "LP decoding corrects a constant fraction of errors," *Proc. IEEE International Symposium on Information Theory (ISIT),* Chicago, IL, p.68, June 2004.

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

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

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

[8] V. Guruswami and P. Indyk, "Near-optimal linear-time codes for unique decoding and new list-decodable codes over smaller alphabets," in *Proc. 34th Annual ACM Symposium on Theory of Computing (STOC),* Montréal, Quebec, Canada, pp. 812–821, May 2002.

[9] A. Khandekar and R.J. McEliece, "On the complexity of reliable communication on the erasure channel," in *Proc. IEEE International Symposium on Information Theory (ISIT),* Washington, DC, p.1, June 2001.

[10] A. Lubotsky, R. Philips, and P. Sarnak, "Ramanujan graphs," *Combinatorica,* vol. 8, pp. 261–277, 1988.

[11] M. Luby, M. Mitzenmacher, A. Shokrollahi, and D. Spielman, "Efficient erasure correcting codes," *IEEE Trans. Inform. Theory,* vol. 47, no. 2, pp. 569–584, Feb. 2001.

[12] G.A. Margulis, "Explicit group theoretical constructions of combinatorial schemes and their applications to the design of expanders and concentrators," *Probl. Inf. Transm.,* vol. 24, no. 1, pp. 39–46, July 1988.

[13] P. Oswald and A. Shokrollahi, "Capacity-achieving sequences for the erasure channel," *IEEE Trans. Inform. Theory,* vol. 48, no. 12, pp. 3017–3028, Dec. 2002.

[14] H. Pfister, I. Sason, and R. Urbanke "Capacity-achieving ensembles for the binary erasure channel with bounded complexity," in *Proc. IEEE International Symposium on Information Theory (ISIT),* Chicago, IL, p.207, June 2004.

[15] T. Richardson, A. Shokrollahi, and R. Urbanke, "Design of capacity-approaching low-density parity-check codes," *IEEE Trans. Inform. Theory,* vol. 47, no. 2, pp. 619–637, Feb. 2001.

[16] T. Richardson and R. Urbanke, "The capacity of low-density parity-check codes under message-passing decoding," *IEEE Trans. Inform. Theory,* vol. 47, no. 2, pp. 599–618, Feb. 2001.

[17] R.M. Roth and V. Skachek, "On nearly-MDS expander codes," *Proc. IEEE International Symposium on Information Theory (ISIT),* Chicago, IL, p.8, June 2004.

[18] R.M. Roth and V. Skachek, "Improved nearly-MDS expander codes," to appear in *IEEE Trans. Inform. Theory,* vol. 52, no. 8, Aug. 2006. Also available at http://arXiv.org/abs/cs.IT/0601090.

[19] I. Sason and R. Urbanke, "Complexity versus performance of capacity-achieving irregular repeat-accumulate codes on the binary erasure channel," *IEEE Trans. Inform. Theory,* vol. 50, no. 6, pp. 1247–1256, June 2004.

[20] M. Sipser and D.A. Spielman, "Expander codes," *IEEE Trans. Inform. Theory,* vol. 42, no. 6, pp. 1710–1722, Nov. 1996.

[21] V. Skachek and R.M. Roth, "Generalized minimum distance iterative decoding of expander codes," in *Proc. IEEE Information Theory Workshop (ITW),* Paris, France, pp. 245–248, Mar.-Apr. 2003.

[22] G. Zémor, "On expander codes," *IEEE Trans. Inform. Theory,* vol. 47, no. 2, pp. 835–837, Feb. 2001.