Graceful degradation over the BEC via non-linear codes

Hajir Roozbehani and Yury Polyanskiy
Email: {hajir,yp}@mit.edu

***Abstract*—We study a problem of constructing codes that transform a channel with high bit error rate (BER) into one with low BER (at the expense of rate). Our focus is on obtaining codes with smooth ("graceful") input-output BER curves (as opposed to threshold-like curves typical for long error-correcting codes).**

This paper restricts attention to binary erasure channels (BEC) and contains two contributions. First, we introduce the notion of Low Density Majority Codes (LDMCs). These codes are non-linear sparse-graph codes, which output majority function evaluated on randomly chosen small subsets of the data bits. This is similar to Low Density Generator Matrix codes (LDGMs), except that the XOR function is replaced with the majority. We show that even with a few iterations of belief propagation (BP) the attained input-output curves provably improve upon performance of any linear systematic code. The effect of non-linearity bootstraping the initial iterations of BP, suggests that LDMCs should improve performance in various applications where LDGMs have been used traditionally.

Second, we establish several *two-point converse bounds* that lower bound the BER achievable at one erasure probability as a function of BER achieved at another one. The novel nature of our bounds is that they are specific to subclasses of codes (linear systematic and non-linear systematic) and outperform similar bounds implied by the area theorem for the EXIT function.

I. Introduction

In this paper we study a case of joint-source channel coding (JSCC) for a binary source and binary erasure channel (BEC). Let $S^k = (S_1, S_2, \cdots, S_k) \sim \mathrm{Ber}(1/2)^{\otimes k}$ be *information bits*. An *encoder* $f : \{0,1\}^k \to \{0,1\}^n$ maps S^k to a (possibly longer) sequence $X^n = (X_1, \cdots, X_n)$ where each X_i is called a *coded bit* and X^n a *codeword*. The rate of the code f is denoted by $R = k/n$ and its bandwidth expansion by $\rho = n/k$. The channel BEC_ϵ takes X^n and produces $Y^n = (Y_1, \ldots, Y_n)$ where each $Y_j = X_j$ with probability $(1-\epsilon)$ and $Y_j = ?$ otherwise. Here we will be interested in performance of the code simultaneously for multiple values of ϵ, and for this reason we denote Y^n by $Y^n(\epsilon)$ to emphasize the value of the erasure probability. Upon observing the distorted information $Y^n(\epsilon)$, *decoder*[1] g produces $\hat{S}^k(\epsilon) = g(Y^n(\epsilon))$. We measure quality of the decoder by the data bit error rate (BER):

$$\mathrm{BER}_{f,g}(\epsilon) \triangleq \frac{1}{k}\sum_{i=1}^{k} \mathbb{P}[S_i \neq \hat{S}_i(\epsilon)] = \frac{1}{k}\mathbb{E}[d_H(S^k, \hat{S}^k(\epsilon))]\,,$$

where d_H stands for the Hamming distance.

The motivation behind this work is thoroughly discussed in [1]. In short, many modern coding systems, such as the one for example being applied in optical communication [2], [3], consist of a layered design: an inner code, which is used to reduce the overall noise; and an outer code, which is used to correct the residual error left from the inner code. Thus, the goal of the inner code is to match the variable real-world channel conditions to a prescribed lower level noise that falls within the error correcting capability of a pre-selected (and highly optimized) outer code.

In other words, the source bits $S_1, \ldots, S_k \in \{0,1\}$ that appear at the input of the inner code need not be reconstructed perfectly, but only approximately. Here we assume that the inner code passes upstream a hard-decision about each bit (that is, an estimate $\hat{S}_i \in \{0,1\}$ of the true value of S_i). Thus, the figure of merit is a curve relating the channel's noise to the probability of bit-flip error of the induced channel $S_i \mapsto \hat{S}_i$ that the outer code is facing.[2] What types of curves are desirable? First of all, we do not want the loss to drop to zero at some finite noise level (since this will then be an overkill: the outer code has nothing to do). Second, it is desirable that the loss decrease with channel improvement, rather than staying flat in a range of parameters (which would be achievable by a separated compress-then-code scheme). These two suggest that the resulting curve should be a smooth and monotone one. With such a requirement the problem is known as *graceful degradation* and has been attracting attention since the early days of channel coding [4], [5]. Despite this, no widely accepted solution is available. This paper's main purpose is to advocate the usage of sparse-graph *non-linear* codes for the problems of graceful degradation and channel matching.

Consider now Fig. 1 which plots the BER functions for some codes. We can see that using LDPC code as an inner code is rather undesirable: if the channel noise is below the BP threshold, then the BER is almost zero and hence the outer code has nothing to do (i.e., its redundancy is being wasted). While if the channel noise is only slightly above the BP threshold the BER sharply rises, this time making outer

This work was supported in part by the National Science Foundation award under grant agreement CCF-17-17842 and by the Center for Science of Information (CSoI), an NSF Science and Technology Center, under grant agreement CCF-09-39370.

[1] The decoder may or may not use the knowledge of ϵ, but for the BEC this is irrelevant.

[2] Note that a more complete description would be that of the "multi-letter" induced channel $S^k \mapsto \hat{S}^k$. However, in this paper we tacitly assume that performance of the outer code is unchanged if we replace the multi-letter channel with the parallel independent channels $\{S_i \mapsto \hat{S}_i\}_{i\in[k]}$. This is justified, for example, if the outer code uses a large interleaver (with length much larger than k), or employs an iterative (sparse-graph) decoder.

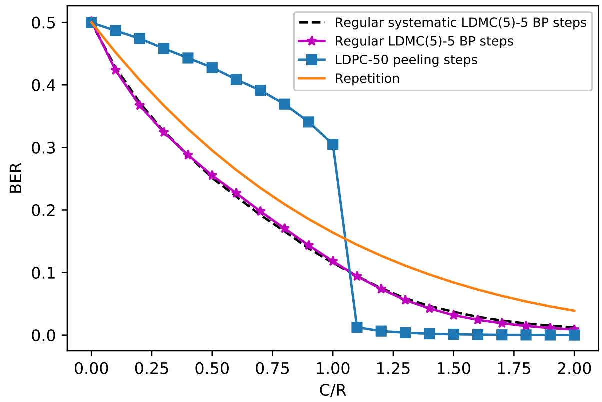


Fig. 1: BER performance vs. capacity-to-rate-ratio of the erasure channel for four codes with rate $R = \frac{1}{5}$ and $k = 3 \times 10^5$ data bits: an LDPC code using 50 iterations of peeling decoder, the repetition code, a regular (non-systematic) LDMC(5) using 5 iterations of BP, and a regular systematic LDMC(5) using 5 iterations of BP. The LDPC code is the dual of a systematic LDGM(16) code, i.e., the variable nodes have degree 17 (and check nodes have approximately Poisson distributed degrees). The LDPC code suffers from the cliff effect, while the LDMCs and the repetition codes degrade gracefully. The LDMC codes uniformly dominates the repetition code for all levels of the channel noise.

code's job impossible. In this sense, a simple (blocklength 5) repetition code might be preferable. However, the *low-density majority codes* (LDMCs) introduced in this paper universally improve upon the repetition. This fact (universal domination) should be surprising for several reasons. First, there is a famous area theorem in coding theory [6], which seems to suggest that BER curves for any two codes of the same rate should crossover[3]. Second, a line of research in *combinatorial JSCC* [7] has demonstrated that the repetition code is optimal [8], [9] in a certain sense, that loosely speaking translates into a high-noise regime here. Third, Prop. 6 below shows that no linear code of rate 1/2 can uniformly dominate the repetition code. Thus, the universal domination of the repetition code by the LDMC observed in Fig. 1 came as a surprise to us. An experienced coding theorist, however, will object to Fig. 1 on the grounds that LDGMs, not LDPCs, should be used for the problem of error-reduction – this was in fact the approach in [2], [3]. So can LDMCs claim superior performance to LDGMs too? Yes, and in fact in a certain sense we claim that LDMCs are superior to all linear systematic codes. This is the message of Fig. 2 and we explain the details next.

What kind of (asymptotic) fundamental limits can we define for this problem? Let us fix the rate $R = \frac{k}{n}$ of a code. The lowest possible BER $\delta^*(R, W)$ achievable over a (memoryless) channel W is found from comparing the source rate-distortion function with the capacity $C(W)$ of the channel:

$$R(1 - h_b(\delta^*(R, W))) = C(W), \tag{1}$$

where $h_b(x) = -x \log x - (1 - x) \log(1 - x)$. Below we call $\delta^*(R, W)$ a *Shannon single-point bound*. Single-point here means that this is a fundamental limit for communicating over a single fixed channel noise level. As we emphasized, graceful degradation is all about looking at a multitude of noise levels. A curious lesson from multi-user information theory shows that it is not possible for a single code to be simultaneously optimal for two channels (for the BSC this was shown in [10] and Prop. 1 shows it for the BEC).

Correspondingly, we introduce a *two-point fundamental limit*:

$$\delta_2^*(R, \delta_a, W_a, W) = \limsup_{k\to\infty} \inf \frac{1}{k} \mathbb{E}_{Y^n \sim W(\cdot|X^n)}[d_H(S^k, \hat{S}^k)] , \tag{2}$$

where the infimum is over all encoders $f : S^k \to X^n$ and decoders $g : Y^n \to \hat{S}^k$ satisfying

$$\frac{1}{k}\mathbb{E}_{Y^n \sim W_a(\cdot|X^n)}[d_H(S^k, \hat{S}^k)] \le \delta_a ,$$

where δ_a, W_a are called, respectively, the anchor distortion and anchor channel. In other words, the value of δ_2^* shows the lowest distortion achievable over the channel W among codes that are already sufficiently good for a channel W_a. Clearly, the two-point performance is related to a two-user broadcast channel [11, Chapter 5].

Similarly, we can make definitions of δ^* and δ_2^* but for a restricted class of encoders, namely linear systematic ones. We bound δ_2^* in Prop. 1 for general codes, in Prop. 7 for general systematic codes, and in Theorem 1 for the subclass of linear systematic codes.

Armed with these definitions, we can return to Fig. 2. What it clearly demonstrates is two things. On one hand, the single-point (top subplot) comparison shows that this specific LDMC is far from Shannon optimality at any value of the channel noise. On the other hand, the two-point (bottom subplot) comparison shows LDMC outperforming any rate-$\frac{1}{2}$ linear systematic code among the class of those which have comparable performance at the anchor point $C(W_a) = 1/4$.

We hope that this short discussion, along with Fig. 1-2, convinces the reader that indeed the LDMCs (and in general adding non-linearity to the encoding process) appear to be the step in the right direction for the problem of *graceful degradation*. However, perhaps even more excitingly, the performance of available LDPC/LDGM codes can be improved by adding some fraction of the LDMC nodes as shown in [1, Section V].

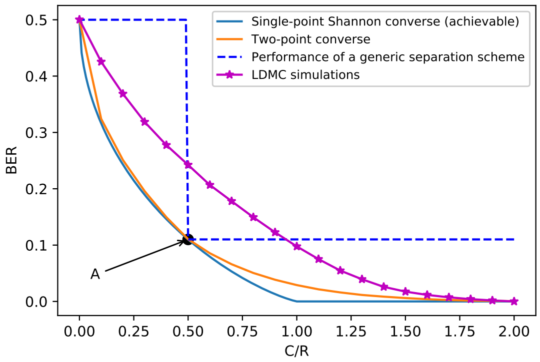
II. Main results

A. The LDMC ensemble

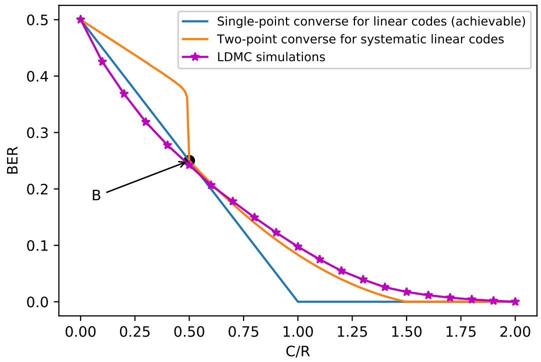
We first define the notion of a check regular code ensemble generated by a Boolean function.

Definition 1. *Let* $\mathbf{P}_\triangle$ *be a joint distribution on* d*-subsets of* $[k] = \{1, \cdots, k\}$*. Given a Boolean function* $f : \{0, 1\}^d \to \{0, 1\}$*, the (check regular) ensemble of codes on*

[3]The caveat here is that the theorem talks about BER evaluated for coded bits, while here we are only interested in the data (or systematic) bits.



(a) General codes



(b) Linear codes

Fig. 2: BER performance vs. capacity-to-noise ratio of the binary erasure channel for rate $R = 1/2$ codes. The starred magenta curve corresponds to simulation results of the systematic regular LDMC(9) with $k = 10^5$ data bits using 3 iterations of BP. Top figure (a) compares LDMC against lower bounds for general codes: Shannon's single-point bound from (1) and a two-point from Prop. 1. While every point A on the Shannon converse curve is achievable by a separation compress-then-code architecture, the latter suffers from non-gracefulness (dashed blue curve). Bottom figure (b) compares LDMC against lower bounds for linear codes. The single-point line is from [1, Proposition 3]. While any point B arbitrarily close to the single-point line is achievable by a linear systematic code, any such code will suffer from non-gracefulness. This is shown by the two-point converse curve on the right. It plots evaluation of Theorem 1, which lower bounds BER of any linear systematic code achieving $\mathrm{BER} \le 0.2501$ over $\mathrm{BEC}(0.75)$. In all, LDMCs are superior to any other linear systematic code whose curve passes through point B.

$\{0,1\}^k$ *generated by* $(f, \mathbf{P}_\Delta)$ *is the family of random codes* $f_\Delta : s \mapsto (f(s_T))_{T\in\Delta}$ *obtained by sampling* $\Delta \sim \mathbf{P}_\Delta$. *Here* s_T *is the restriction of* s *to the coordinates indexed by* T.

Given $x \in \{0,1\}^d$, we consider the d-majority function

$$\text{d-maj}(x) = \mathbb{1}_{\{\sum_i x_i > \frac{d}{2}\}}.$$

We have the following definition:

Definition 2. *Let* $\mathbf{U}_\Delta = \mathrm{Unif}^{\otimes n}(\{\text{d-subsets of } [k]\})$ *be the uniform product distribution on the* d*-subsets of* $[k]$*. The ensemble of codes generated by* $(\text{d-maj}, \mathbf{U}_\Delta)$ *is called the Low Density Majority Code (LDMC) ensemble of degree* d *and denoted by LDMC(*d*). Furthermore, define the event* $A := \cup_{ij}\{\sum_{T\in\Delta} \mathbb{1}_{\{i\in T\}} = \sum_{T\in\Delta} \mathbb{1}_{\{j\in T\}}\}$*, i.e., the event that each* i *appears in the same number of* d*-subsets* T*. Then the ensemble generated by* $(\text{d-maj}, \mathbf{U}_{\Delta|A})$ *is called a regular LDMC(*d*) ensemble.*

B. Two point converses for graceful degradation

1) General codes: The results on broadcast channels [12, Theorem 1], [13, Section V.B] give the following bound shown in [14, Section II.C2]:

Proposition 1 ([12]–[14]). *Consider a sequence of codes of rate* R *encoding iid* $\mathrm{Ber}(1/2)$ *bits and achieving the per-letter Hamming distortion* $\delta^*(R, W)$*, cf.* (1)*, when* $W = \mathrm{BEC}_\epsilon$*. Then over* $W = \mathrm{BEC}_\tau$ *their distortion is lower bounded by*

$$\delta \ge \mathbb{1}_{\{\tau\ge\epsilon\}}\eta(\delta^*, \epsilon, \tau) + \mathbb{1}_{\{\tau<\epsilon\}} \inf_y \{y : \eta(y, \tau, \epsilon) \le \delta^*\},$$

where

$$\eta(\delta^*, \epsilon, \tau) \triangleq \sup_{q\in[0,1/2]} \frac{h_b^{-1}\left((1-(1-\tau)/R) + \frac{1-\tau}{1-\epsilon}(h_b(q*\delta^*) - h_b(\delta^*))\right) - q}{1-2q}.$$

Remark 1. *We note that the above result is rate independent in the sense that it can be re-parametrized in terms of the capacity-to-rate ratios only. Proposition 7 below gives a new (rate dependent) bound for general systematic codes that in some regimes improves the above.*

Systematic linear codes form the vast majority of the codes that are used in practice. In this section, we work towards proving a two-point converse bound for this class of codes.

In the following, by $\ker(A)$ we refer to the left kernel of A, that is the subspace of vectors x satisfying $xA = 0$.

Definition 3. *Given a matrix* A *define* [4]

$$\mathrm{hrank}(A) \triangleq |\{j : \ker(A) \subset \{x : x_j = 0\}\}|.$$

Definition 4. *Given a matrix* A*, define* $\tilde{A}(p, q)$ *to be a random sub-matrix of* A *that is obtained by sampling each row of* A *with probability* p *and each column of* A *with probability* q *independently of other rows/columns.*

The following proposition is well known (cf. [15]).

Proposition 2. *Consider a system of equations* $xG = y$ *over* $\mathbb{F}_2$*. If* $\ker(G) \subset \{x : x_i = 0\}$*, then* x_i *is uniquely determined from solving* $xG = y$*. Otherwise, there is a bijection between the set of solutions* $\{x : xG = y, x_i = 0\}$ *and* $\{x : xG = y, x_i = 1\}$*. In particular, if exactly* t *coordinates are uniquely determined by the above equations, then* $\mathrm{hrank}(G) = t$*.*

Our next proposition relates BER and hrank.

Proposition 3. *Let* $G = [I\ A]$ *be the generator matrix of a systematic linear code* f*. Then* $\mathrm{BER}_f(\epsilon) \le \delta$ *if and only if*

$$\mathbb{E}[\mathrm{hrank}\left(\tilde{A}(\epsilon, 1-\epsilon)\right)] \ge (\epsilon - 2\delta)k.$$

[4] Equivalently, $\mathrm{hrank}(A) \triangleq k - |\{j : \exists x \in \ker(A) \text{ s.t } x_j \ne 0\}|$.

Proof. If BER is bounded by δ, there are, on average, at most $2\delta k$ bits that are not uniquely determined by solving $x\tilde{G}(1, 1-\epsilon) = y$. For a systematic code, the channel returns $\mathrm{Bin}(k, 1-\epsilon)$ systematic bits. The remaining systematic bits s_r are to be determined from solving $s_r\tilde{A}(\epsilon, 1-\epsilon) = \tilde{y}$ where $\tilde{y}$ is some vector that depends on the channel output y and the returned systematic bits. If t additional systematic bits are recovered, then $\mathrm{hrank}(\tilde{A}(\epsilon, 1-\epsilon)) = t$ by Proposition 2. Since on average at least $(\epsilon - 2\delta)k$ additional systematic bits are recovered, the claim on the average hrank follows. ■

The next proposition shows how matrices with positive hrank behave under row sub-sampling. Our main observation is that row sub-sampled matrices of a (thin) matrix with large hrank have bounded rank. In particular, if a (thin) matrix has full hrank, its sub-sampled matrices cannot have full rank.

Proposition 4. *Consider an arbitrary field $\mathbb{F}$ and let $\epsilon_1 > \epsilon_2$. Given a $k \times m$ matrix A,*

$$\mathbb{E}[\mathrm{rank}\left(\tilde{A}(\epsilon_2, 1)\right)] \le \mathrm{rank}(A) - (1 - \frac{\epsilon_2}{\epsilon_1})\mathbb{E}[\mathrm{hrank}\left(\tilde{A}(\epsilon_1, 1)\right)],$$

and

$$\mathbb{E}[\mathrm{hrank}\left(\tilde{A}(\epsilon_2, 1)\right)] \ge \frac{\epsilon_2}{\epsilon_1}\mathbb{E}[\mathrm{hrank}\left(\tilde{A}(\epsilon_1, 1)\right)].$$

Therefore, if $\mathbb{E}[\mathrm{rank}\left(\tilde{A}(\epsilon_2, 1)\right)] = \mathrm{rank}(A) - o(k)$, then $\mathbb{E}[\mathrm{rank}\left(\tilde{A}(\epsilon_1, 1)\right)] = o(k)$.

Proof. Suppose that $\mathrm{hrank}\left(\tilde{A}(\epsilon_1, q)\right) = t$. This means that there are at least t rows a_j in $\tilde{A}(\epsilon_1, q)$ such that a_j is not in the span of $\{a_i : i \neq j\}$. Let B be the row-submatrix of $\tilde{A}(\epsilon_1, q)$ associated to these t rows, and B^c be its complement, i.e., the matrix with rows $\{a_j : a_j \in \tilde{A}(\epsilon_1, q), a_j \notin B\}$. We claim that the complement of B is a matrix of rank $\mathrm{rank}(A) - t$. To see this, note that $\mathrm{Im}(B) \cap \mathrm{Im}(B^c) = \{0\}$, for otherwise we get linear dependencies of the form $h = \sum_i \alpha_i b_i \neq 0$ where $b_i \in B$ and $h \in \mathrm{Im}(B^c)$, which contradicts the construction of B. This means that $\mathrm{rank}(B^c) + \mathrm{rank}(B) = \mathrm{rank}(A)$. The claim now follows since $\mathrm{rank}(B) = t$. Under row sub-sampling, each row of B is selected with probability ϵ_2/ϵ_1 independently of other rows. Thus,

$$\mathbb{E}[\mathrm{hrank}\left(\tilde{A}(\epsilon_2, q)\right) | \mathrm{hrank}\left(\tilde{A}(\epsilon_1, q)\right) = t] \ge \frac{\epsilon_2}{\epsilon_1}t$$

The rows selected from B^c can contribute at most $\mathrm{rank}(A) - t$ to the rank of $\tilde{A}(\epsilon_2, q)$. Hence

$$\mathbb{E}[\mathrm{rank}\left(\tilde{A}(\epsilon_2, q)\right) | \mathrm{hrank}(\tilde{A}(\epsilon_1, q)) = t] \le \frac{\epsilon_2}{\epsilon_1}t + \mathrm{rank}(A) - t$$

Taking the average over the hrank of $\tilde{A}(\epsilon_1, q)$ proves the first two results. The last inequality follows by re-arranging the terms.

■

The next Proposition shows that rank is well behaved under column sub-sampling.

Proposition 5. *Consider an arbitrary field $\mathbb{F}$ and let $p > q$. Given a $k \times m$ matrix A over $\mathbb{F}$,*

$$\mathbb{E}[\mathrm{rank}\left(\tilde{A}(1, p)\right)] \le \min\{pm, \frac{p}{q}\mathbb{E}[\mathrm{rank}\left(\tilde{A}(1, q)\right)]\}.$$

Proof. Pick a column basis for $\tilde{A}(1, p)$. We can realize $\tilde{A}(1, q)$ by sub-sampling columns of $\tilde{A}(1, p)$. In this way, each column in the basis of $\tilde{A}(1, p)$ is selected with probability q/p independently of other columns. In other words,

$$\mathbb{E}[\mathrm{rank}\left(\tilde{A}(1, q)\right)] \ge \frac{q}{p}\mathbb{E}[\mathrm{rank}\left(\tilde{A}(1, p)\right)].$$

The desired result follows. ■

We are now ready to prove our main result.

Theorem 1 ([1, Theorem 1]). *Let $f : s \mapsto sG$ be a systematic linear code of rate $1/\rho$ with generator matrix $G = [I\ A]$ over $\mathbb{F}_2$. Fix $\epsilon_1 > \epsilon_2$ and $\delta_1 \le \frac{\epsilon_1}{2}$. If $\mathrm{BER}_f(\epsilon_1) \le \delta_1$, then*

$$\mathrm{BER}_f(\epsilon_2) \ge \kappa(\rho, \delta_1, \epsilon_2, \epsilon_1) \triangleq \frac{\epsilon_2 - \frac{1-\epsilon_2}{1-\epsilon_1}\left[\frac{\epsilon_2}{\epsilon_1}\gamma + (\rho-1)(1-\epsilon_1) - \gamma\right]}{2} \quad (3)$$

with $\gamma = \epsilon_1 - 2\delta_1$. If $\epsilon_2 > \epsilon_1$ then

$$\mathrm{BER}_f(\epsilon_2) \ge \frac{\epsilon_2}{2} - \frac{\epsilon_2}{\epsilon_2 - \epsilon_1}\frac{1}{1-\epsilon_1}\left(\delta_1 - \frac{1}{2}(1 - \rho(1-\epsilon_1))\right). \quad (4)$$

In particular, if $\mathrm{BER}(\epsilon_1) = \frac{1}{2}(1 - \rho(1-\epsilon_1)) + o(1)$, then $\mathrm{BER}(\epsilon_2) = \frac{\epsilon_2}{2} - o(1)$ for all $\epsilon_2 > \epsilon_1$.

Proof. By Proposition 3, we have $\mathbb{E}[\mathrm{hrank}\left(\tilde{A}(\epsilon_1, 1-\epsilon_1)\right)] \ge \gamma k$. By Proposition 4, we have

$$\mathbb{E}[\mathrm{rank}\left(\tilde{A}(\epsilon_2, 1-\epsilon_1)\right)] \le (\frac{\epsilon_2}{\epsilon_1}\gamma + (\rho-1)(1-\epsilon_1) - \gamma)k.$$

By Proposition 5, we have

$$\mathbb{E}[\mathrm{rank}\left(\tilde{A}(\epsilon_2, 1-\epsilon_2)\right)] \le \frac{1-\epsilon_2}{1-\epsilon_1}(\frac{\epsilon_2}{\epsilon_1}\gamma + (\rho-1)(1-\epsilon_1) - \gamma)k.$$

The first result now follows from Proposition 3 upon observing that $\mathrm{hrank}(\tilde{A}) \le \mathrm{rank}(\tilde{A})$.

For the second case, we trivially have the following estimate (by interchanging the roles of ϵ_1 and ϵ_2 and applying the first part):

$$\mathrm{BER}_f(\epsilon_2) \ge \inf_{\delta_2}\{\delta_2 : \kappa(\rho, \delta_2, \epsilon_1, \epsilon_2) \le \delta_1\}.$$

The estimate (4) then follows by evaluating this infimum (which is a minimization of a linear function).

■

One simple application of Theorem 1 demonstrates that no linear systematic code can uniformly dominate 2-fold repetition.

Proposition 6 ([1, Proposition 7]). *Let g be the 2-fold repetition code with bit-MAP decoder and f be a linear systematic code of rate $1/2$. If there exists ϵ_2 such that $\mathrm{BER}_f(\epsilon_2) < \mathrm{BER}_g(\epsilon_2)$, then there exists some $\epsilon^* > \epsilon_2$ such that for all $\epsilon_1 \in (\epsilon^*, 1)$ we have $\mathrm{BER}_f(\epsilon_1) > \mathrm{BER}_g(\epsilon_1)$. Moreover, if $\mathrm{BER}_f(\epsilon_2) = t\mathrm{BER}_g(\epsilon_2)$ for some $t < 1$, then we can pick $\epsilon^* = \max(\epsilon_2, 1 - \frac{(1-t)\epsilon_2^2}{1-\epsilon_2})$.*

C. Bounds via area theorem

The lower bound of Theorem 1 states that a linear systematic code cannot have small BER for all erasure probabilities. In this sense, it has the flavor of a "conservation law". In coding theory, it is often important to understand how a code behaves over a family of parametrized channels. The main existing tool in the literature to study such questions is the so called area theorem. Here we introduce the theorem and study its consequences for two point bounds on BER. It is shown in [1, Section III] that the bound in Theorem 1 is tighter than what can be inferred from the area theorem for linear codes. However, the area theorem gives rise to new bounds for general systematic codes that are in some regimes tighter than the best previously known bounds (see [1, Section III]).

Following [15], we define the notion of an *extrinsic information transfer* (EXIT) function.

Definition 5. *Let X be a codeword chosen from an (n,k) code C according to the uniform distribution. Let $Y(\epsilon)$ be obtained by transmitting X through a BEC(ϵ). Let*

$$Y_{\sim i}(\epsilon) = (Y_1(\epsilon), \cdots, Y_{i-1}(\epsilon), ?, Y_{i+1}(\epsilon), \cdots, Y_n(\epsilon))$$

be obtained by erasing the i-th bit from $Y(\epsilon)$. The i-th EXIT function of C is defined as

$$h_i(\epsilon) = H(X_i|Y_{\sim i}(\epsilon))$$

The average EXIT function is

$$h(\epsilon) = \frac{1}{n}\sum_{i=1}^{n} h_i(\epsilon)$$

The area theorem states that for a binary code of rate R we have $R = \int_0^1 h(\epsilon)d\epsilon$. Let us now find the implications of the area theorem for the input BER of linear systematic codes. To this end we define the average systematic EXIT function

$$h^{\text{sys}}(\epsilon) = \frac{1}{k}\sum_{i=1}^{k} h_i(\epsilon).$$

Likewise we can define the non-systematic EXIT function as follows:

$$h^{\text{non-sys}}(\epsilon) = \frac{1}{n-k}\sum_{i=k+1}^{n} h_i(\epsilon).$$

We first give a lemma to show that the coded bit error rate converges to 0 continuously as the input bit error rate vanishes.

Lemma 1 (Data BER vs EXIT function [1, Lemma1]). *Fix $\epsilon < \epsilon_0$. Let f be a binary code of rate R.*

(a) If f is linear, then

$$h(\epsilon) \le \frac{2R}{\epsilon_0 - \epsilon}\text{BER}_f(\epsilon_0). \tag{5}$$

(b) If f is general we have

$$h(\epsilon) \le \frac{R}{\epsilon_0 - \epsilon} h_b(\text{BER}_f(\epsilon_0)),$$

where h_b is the binary entropy function.

Proposition 7 ([1, Proposition 8]). *Let $\epsilon_2 < \epsilon_1$. Let f be a binary code of rate R with $\text{BER}(\epsilon_2) \le \delta_2$. Define*

$$\zeta(x, \epsilon_2, \epsilon_1) \triangleq \sup_{\{\epsilon_0 : \epsilon_0 < \epsilon_2\}} \frac{1}{R}\Bigg(\frac{1}{(\epsilon_1 - \epsilon_0)}(R - (1-\epsilon_1) - \epsilon_0\frac{xR}{\epsilon_2 - \epsilon_0}) - 1 + R\Bigg). \tag{6}$$

The following hold:

(a) If f is linear systematic then

$$\text{BER}(\epsilon_1) \ge \frac{\epsilon_1}{2}\zeta(2\delta_2, \epsilon_2, \epsilon_1).$$

In particular, if $\text{BER}(\epsilon_2) = o(1)$, then

$$\text{BER}(\epsilon_1) \ge \frac{\epsilon_1}{2R}\left(\frac{1}{(\epsilon_1 - \epsilon_2)}(R - (1-\epsilon_1)) - 1 + R\right) + o(1)$$

(b) If f is systematic (but possibly non-linear), then

$$\text{BER}(\epsilon_1) \ge \epsilon_1 h_b^{-1}\left(\zeta(h_b(\delta_2), \epsilon_2, \epsilon_1)\right).$$

Proof. To prove the lower bound on $h(\epsilon_2)$, we may approximate $h(\epsilon_1)$ in a worst-cast fashion as a piece-wise constant function. To do this, note that $h(\epsilon) \le h(\epsilon_2)$ for all $\epsilon \le \epsilon_2$, and $h(\epsilon) \le h(\epsilon_1)$ for all $\epsilon \in (\epsilon_2, \epsilon_1]$, and $h(\epsilon) \le 1$ for all $\epsilon > \epsilon_1$. Then the area theorem gives that

$$1 - \epsilon_1 + h(\epsilon_1)(\epsilon_1 - \epsilon_2) + h(\epsilon_2)\epsilon_2 \ge R$$

We note that

$$h(\epsilon) = Rh^{\text{sys}}(\epsilon) + (1-R)h^{\text{non-sys}}(\epsilon) \tag{7}$$

Using the above two relations, we have

$$Rh^{\text{sys}}(\epsilon_1) \ge \frac{R - (1-\epsilon_1) - h(\epsilon_2)\epsilon_2}{\epsilon_1 - \epsilon_2} - (1-R)h^{\text{non-sys}}(\epsilon_1)$$

Using $h^{\text{non-sys}} \le 1$, we get

$$h^{\text{sys}}(\epsilon_1) \ge \frac{1}{R(\epsilon_1 - \epsilon_2)}(R - (1-\epsilon_1) - h(\epsilon_2)\epsilon_2) - (\frac{1}{R} - 1). \tag{8}$$

(a) Applying Lemma 1a to bound $h(\epsilon_2)$ we get from (8)

$$h^{\text{sys}}(\epsilon_1) \ge \zeta(2\delta_2, \epsilon_1, \epsilon_2).$$

The bounds on BER follow from noticing that for a linear systematic code

$$\text{BER}(\epsilon) = \frac{\epsilon h^{\text{sys}}(\epsilon)}{2}.$$

(b) Applying Lemma 1b to bound $h(\epsilon_2)$ in (8) gives

$$h^{\text{sys}}(\epsilon_1) \ge \zeta(h_b(\delta_2), \epsilon_1, \epsilon_2). \tag{9}$$

Let $\tilde{X}_i = \tilde{X}_i(Y_{\sim i}(\epsilon_1))$ be the MAP decoder of X_i given $Y_{\sim i}(\epsilon_1)$. From Fano's and Jensen's inequalities we have

$$h^{\text{sys}}(\epsilon_1) = \frac{1}{k}\sum_{i=1}^{k} H(X_i|Y_{\sim i}(\epsilon_1)) \le \frac{1}{k}\sum_{i=1}^{k} h_b(\mathbb{P}[X_i \ne \tilde{X}_i]) \le h_b\left(\frac{1}{k}\sum_{i=1}^{k}\mathbb{P}[X_i \ne \tilde{X}_i]\right).$$

Now, notice that $\mathbb{P}[X_i \ne \tilde{X}_i] = \epsilon_1\mathbb{P}[X_i \ne \tilde{X}_i]$ and, thus,

$$h^{\text{sys}}(\epsilon_1) \le h_b(\frac{1}{\epsilon_1}\text{BER}(\epsilon_1)).$$

The proof is concluded by applying (9).

■

We end with a brief discussion to advocate for use of non-linear codes. Fig. 2b shows evaluation of Theorem 1. It also shows that in the higher noise regime LDMC codes can outperform any linear code that has comparable performance at some lower values of C/R. Fig. 1 shows that LDMC codes can uniformly dominate repetition. By Proposition 6, no linear systematic code can dominate repetition. It should thus be expected that LDMCs (and non-linear codes in general) can achieve certain perfomance levels that are fundamentally unattainable by linear codes. It is shown in [1] that indeed using non-linearities in certain settings can uniformly improve the performance and convergence rate of LDGMs. It is also shown (see [1, Fig. 3]) that any systematic code of high rate that achieves low BER at rates closed to capacity will not degrade gracefully and hence is not suitable for error reduction. However, we cannot rule out the existence of non-systematic nonlinear codes that would simultaneously be graceful and almost capacity-achieving.

REFERENCES

[1] H. Roozbehani and Y. Polyanskiy, "Low density majority codes and the problem of graceful degradation," *arXiv preprint arXiv:1911.12263*, 2019.

[2] L. M. Zhang and F. R. Kschischang, "Low-complexity soft-decision concatenated ldgm-staircase fec for high-bit-rate fiber-optic communication," *Journal of Lightwave Technology*, vol. 35, no. 18, pp. 3991–3999, 2017.

[3] M. Barakatain and F. R. Kschischang, "Low-complexity concatenated ldpc-staircase codes," *Journal of Lightwave Technology*, vol. 36, no. 12, pp. 2443–2449, 2018.

[4] J. Ziv, "The behavior of analog communication systems," *IEEE Transactions on Information Theory*, vol. 16, no. 5, pp. 587–594, 1970.

[5] G. R. Redinbo, "Optimum soft decision decoding with graceful degradation," *Information and Control*, vol. 41, no. 2, pp. 165–185, 1979.

[6] A. Ashikhmin, G. Kramer, and S. ten Brink, "Extrinsic information transfer functions: model and erasure channel properties," *IEEE Transactions on Information Theory*, vol. 50, no. 11, pp. 2657–2673, 2004.

[7] Y. Kochman, A. Mazumdar, and Y. Polyanskiy, "The adversarial joint source-channel problem," in *2012 IEEE International Symposium on Information Theory Proceedings*. IEEE, 2012, pp. 2112–2116.

[8] Y. Polyanskiy, "On metric properties of maps between Hamming spaces and related graph homomorphisms," *J. Combin. Theory Ser. A*, vol. 145, pp. 227–251, 2017.

[9] H. Roozbehani and Y. Polyanskiy, "Input-output distance properties of good linear codes," in *2018 IEEE International Symposium on Information Theory (ISIT)*. IEEE, 2018, pp. 2510–2514.

[10] Y. Kochman, O. Ordentlich, and Y. Polyanskiy, "Ozarow-type outer bounds for memoryless sources and channels," in *2018 IEEE International Symposium on Information Theory (ISIT)*. IEEE, 2018, pp. 1774–1778.

[11] A. El Gamal and Y. Kim, *Network information theory*. Cambridge University Press, 2011.

[12] L. Tan, A. Khisti, and E. Soljanin, "Distortion bounds for broadcasting a binary source over binary erasure channels," in *2013 13th Canadian Workshop on Information Theory*. IEEE, 2013, pp. 49–54.

[13] K. Khezeli and J. Chen, "A source-channel separation theorem with application to the source broadcast problem," *IEEE Transactions on Information Theory*, vol. 62, no. 4, pp. 1764–1781, 2016.

[14] Y. Kochman, O. Ordentlich, and Y. Polyanskiy, "A lower bound on the expected distortion of joint source-channel coding," *IEEE Transactions on Information Theory*, 2020.

[15] T. Richardson and R. Urbanke, *Modern coding theory*. Cambridge university press, 2008.