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 low 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 performance of any linear systematic code. The effect of 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 performance of any linear systematic code.

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 $S^k = (S_1, S_2, \ldots, S_k) \sim \text{Ber}(1/2)$ be information bits. An encoder $f : \{0,1\}^k \to \{0,1\}^n$ maps $S^k$ to a (possibly longer) sequence $X^n = (X_1, \ldots, X_n)$ 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$_\rho$ 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 $\epsilon$, 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 $Y^n(\epsilon)$, decoder1 $g$ produces $\hat{S}^k(\epsilon) = g(Y^n(\epsilon))$. We measure quality of the decoder by the data bit error rate (BER):

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

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

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.

1The decoder may or may not use the knowledge of $\epsilon$, but for the BEC this is irrelevant.
channel noise. On the other hand, the two-point (bottom subplot) comparison shows LDMC outperforming any rate-
subset of [k] = {1, · · · , k}. Given a Boolean function f : {0, 1}^d \rightarrow {0, 1}, the (check regular) ensemble of codes on
channel W is found from comparing the source rate-distortion function with the capacity C(W) of the channel:
\[ R(1 - h_b(\delta^*(R, W))) = C(W), \]
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:
\[ \delta^*_2(R, \delta_a, W_a, W) = \limsup_{k \to \infty} \frac{1}{k} E_{Y^n \sim W_a(X^n)} [d_H(S^k, \hat{S}^k)], \]
where the infimum is over all encoders \( f : S^k \to X^n \) and decoders \( g : Y^n \to \hat{S}^k \) satisfying
\[ \frac{1}{k} E_{Y^n \sim W_a(X^n)} [d_Y(S^k, \hat{S}^k)] \leq \delta_a, \]
where \( \delta_a, W_a \) are called, respectively, the anchor distortion and anchor channel. In other words, the value of \( \delta^*_2 \)
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 \( \delta^* \) but for a restricted class of encoders, namely linear systematic ones. We
bound \( \delta^*_2 \) 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 degrada-
tion. 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 \( P_\Delta \) be a joint distribution on d-subsets of \([k] = \{1, \cdots, k\} \). Given a Boolean function
\( f : \{0, 1\}^d \rightarrow \{0, 1\} \), the (check regular) ensemble of codes on

---

Fig. 1: BER performance vs. capacity-to-rate-ratio of the era-
sure channel for four codes with rate \( R = \frac{1}{2} \) 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.

3The 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.
Consider a system of equations

Given a matrix

Fig. 2: BER performance vs. capacity-to-noise ratio of the linear systematic code whose curve passes through point B. Furthermore, define the event

while the right. It plots evaluation of Theorem 1, which lower bounds the BER the event that each i appears in the same number of d-subsets of T. Then the ensemble generated by (d-maj, Uₐ|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 Ber(1/2) bits and achieving the per-letter Hamming distortion δ*(R,W), cf. (1), when W = BECγ. Then over W = BECγ their distortion is lower bounded by

where

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

hrank(A) = \{ |{ j : ker(A) \cap { x \in \mathbb{F}_2 : x_j = 0 } } | \}.

Definition 4. Given a matrix A, define \( \hat{A}(p,q) \) 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) \( \subset \{ x \in \mathbb{F}_2 : x_1 = 0 \} \), then \( x_1 \) is uniquely determined from solving \( xG = y \). Otherwise, there is a bijection between the set of solutions \( \{ x : xG = y, x_1 = 0 \} \) and \( \{ x : xG = y, x_1 = 1 \} \). In particular, if exactly \( t \) coordinates are uniquely determined by the above equations, then 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 BERγ(ε) ≤ δ if and only if

\[ E[\text{hrank}\left( \hat{A}(ε, 1 - ε) \right)] ≥ (ε - 2δ)k. \]

\[ \text{hrank}(A) \triangleq k - |\{ j : \exists x \in \text{ker}(A) \text{ s.t. } x_j \neq 0 \}|. \]
Proof. If BER is bounded by $\delta$, there are, on average, at most $2\delta k$ bits that are not uniquely determined by solving $xG(1, 1-\epsilon) = y$. For a systematic code, the channel returns 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 $\text{hrank}(\tilde{A}(\epsilon, 1-\epsilon)) = t$ 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 $\epsilon > \epsilon_2$. Given a $k \times m$ matrix $A$, $\mathbb{E}[\text{rank}(\tilde{A}(\epsilon_1, 1))] \leq \text{rank}(A) - (1 - \frac{\epsilon_2}{\epsilon_1})\mathbb{E}[\text{hrank}(\tilde{A}(\epsilon_1, 1))]$, and $\mathbb{E}[\text{hrank}(\tilde{A}(\epsilon_1, 1))] \geq \frac{\epsilon_2}{\epsilon_1} \mathbb{E}[\text{hrank}(\tilde{A}(\epsilon_1, 1))]$.

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

**Proof.** Suppose that $\text{hrank}(\tilde{A}(\epsilon_1, 1)) = t$. This means that there are at least $t$ rows $a_j$ in $\tilde{A}(\epsilon_1, 1)$ such that $a_j$ is not in the span of $\{a_i : i \neq j\}$. Let $B$ be the row-submatrix of $\tilde{A}(\epsilon_1, 1)$ associated to these $t$ rows, and $B^c$ be its complement, i.e., the matrix with rows $\{a_j : a_j \notin \tilde{A}(\epsilon_1, 1), a_j \notin B\}$. We claim that the complement of $B$ is a matrix of rank $\text{rank}(A) - t$. To see this, note that $\text{Im}(B) \cap \text{Im}(B^c) = \{0\}$, for otherwise we get linear dependencies of the form $h = \sum a_i b_i \neq 0$ where $b_i \in B$ and $h \in \text{Im}(B^c)$, which contradicts the construction of $B$. This means that $\text{rank}(B^c) + \text{rank}(B) = \text{rank}(A)$. The claim now follows since $\text{rank}(B) = t$. Under row sub-sampling, each row of $B$ is selected with probability $\epsilon_2/\epsilon_1$ independently of other rows. Thus, $\mathbb{E}[\text{hrank}(\tilde{A}(\epsilon_2, 1))] = t \geq \frac{\epsilon_2}{\epsilon_1} t$.

The rows selected from $B^c$ can contribute at most $\text{rank}(A) - t$ to the rank of $\tilde{A}(\epsilon_2, 1)$. Hence $\mathbb{E}[\text{rank}(\tilde{A}(\epsilon_1, 1))] = t \leq \frac{\epsilon_2}{\epsilon_1} t + \text{rank}(A) - t$.

Taking the average over the hrank of $\tilde{A}(\epsilon_1, 1)$ 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 $\mathbb{F}$, $\mathbb{E}[\text{rank}(\tilde{A}(1, p))] \leq \min\{pm, p\mathbb{E}[\text{rank}(\tilde{A}(1, q))]\}$.

**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, $\mathbb{E}[\text{rank}(\tilde{A}(1, q))] \geq \frac{q}{p} \mathbb{E}[\text{rank}(\tilde{A}(1, p))]$.

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 $1/p$ with generator matrix $G = [I \ A]$ over $\mathbb{F}_2$. Fix $\epsilon_1 > \epsilon_2$ and $\Delta_1 \leq \frac{\epsilon_2}{\epsilon_1}$. If $\text{BER}_f(\epsilon_1) \leq \Delta_1$, then $\text{BER}_f(\epsilon_2) \geq \kappa(\rho, \Delta_1, \epsilon_2, \epsilon_1) \geq \Delta_2 - \frac{\epsilon_2}{\epsilon_1}(1 - \epsilon_2)(1 - \epsilon_1) - \gamma$ (3)

with $\gamma = \epsilon_1 - 2\delta_1$. If $\epsilon_2 > \epsilon_1$ then $\Delta_2^2 - \Delta_2 - o(1)$ for all $\epsilon_2 > \epsilon_1$.

**Proof.** By Proposition 3, we have $\mathbb{E}[\text{rank}(\tilde{A}(\epsilon_1, 1))] \leq \gamma k$. By Proposition 4, we have $\mathbb{E}[\text{hrank}(\tilde{A}(\epsilon_1, 1))] \leq \frac{\epsilon_2}{\epsilon_1} \gamma + (\rho - 1)(1 - \epsilon_1) - \gamma k$. By Proposition 5, we have $\mathbb{E}[\text{hrank}(\tilde{A}(\epsilon_2, 2))] \leq \frac{1 - \epsilon_1}{1 - \epsilon_2} \gamma + (\rho - 1)(1 - \epsilon_1) - \gamma k$.

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

For the second case, we trivially have the following estimate (by interchanging the roles of $\epsilon_1$ and $\epsilon_2$ and applying the first part):

$$\text{BER}_f(\epsilon_2) \geq \inf_{\delta_2} \{\delta_2 : \kappa(\rho, \delta_2, \epsilon_1, \epsilon_2) \leq \delta_1\}.$$  

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 $\epsilon_2$ such that $\text{BER}_f(\epsilon_2) < \text{BER}_g(\epsilon_2)$, then there exists some $\epsilon^* > \epsilon_2$ such that for all $\epsilon_1 \in (\epsilon^*, 1)$ we have $\text{BER}_f(\epsilon_1) > \text{BER}_g(\epsilon_1)$. Moreover, if $\text{BER}_f(\epsilon_2) = \text{BER}_g(\epsilon_2)$ for some $t < 1$, then we can pick $\epsilon^* = \max(\epsilon_2, 1 - (1 - \epsilon_2)^{1/2})$.

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 infor-
mation transfer (EXIT) function.

**Definition 5.** Let $X$ be a codeword chosen from an $(n, k)$
code $C$ according to the uniform distribution. Let $Y(e)$ be
obtained by transmitting $X$ through a BEC($e$). Let
$Y_{-i}(e) = (Y_1(e), \ldots, Y_{-i-1}(e), ? Y_{i+1}(e), \ldots, Y_n(e))$
be obtained by erasing the $i$-th bit from $Y(e)$. The $i$-th EXIT
function of $C$ is defined as
$$h_i(e) = H(X_i | Y_{-i}(e))$$
The average EXIT function is
$$h(e) = \frac{1}{n} \sum_{i=1}^{n} h_i(e)$$

The area theorem states that for a binary code of rate $R$ we have
$R = \int_0^1 h(e) de$. 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
$$\bar{h}_{sys}(e) = \frac{1}{k} \sum_{i=1}^{k} h_i(e)$$

Likewise we can define the non-systematic EXIT function as
follows:
$$\bar{h}_{non-sys}(e) = \frac{1}{n-k} \sum_{i=k+1}^{n} h_i(e)$$

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, Lemma 1]).** Fix
$\epsilon < \epsilon_0$. Let $f$ be a binary code of rate $R$.

(a) If $f$ is linear, then
$$h(e) \leq \frac{2R}{\epsilon_0 - \epsilon} \text{BER}_f(\epsilon_0).$$

(b) If $f$ is general we have
$$h(e) \leq \frac{R}{\epsilon_0 - \epsilon} \bar{h}_b(\text{BER}_f(\epsilon_0)),$$
where $\bar{h}_b$ is the binary entropy function.

**Proposition 7 ([1, Proposition 8]).** Let $e_2 < \epsilon_1$. Let $f$ be a
binary code of rate $R$ with $\text{BER}(e_2) \leq \delta_2$. Define
$$\zeta(x, \epsilon_2, \epsilon_1) \triangleq \sup_{\epsilon_1 < \epsilon < e_2} \frac{1}{R} \left( 1 \left( 1 \frac{1}{\epsilon_1 - e_0} \right) \left( (1 - \epsilon_1) \right) - 1 + R \right)$$

The following hold:

(a) If $f$ is linear systematic then
$$\text{BER}(\epsilon_1) \geq \frac{\epsilon_1}{2} \zeta(2\delta_2, e_2, \epsilon_1).$$

In particular, if $\text{BER}(e_2) = o(1)$, then
$$\text{BER}(\epsilon_1) \geq \frac{\epsilon_1}{2} \left( \frac{1}{R(1 - (1 - \epsilon_1))} - 1 + R \right) + o(1)$$

(b) If $f$ is systematic (but possibly non-linear), then
$$\text{BER}(\epsilon_1) \geq \epsilon_1 \frac{1}{2} \zeta(2\delta_2, e_2, \epsilon_1).$$

**Proof.** To prove the lower bound on $h(e_2)$, we may approx-
imate $h(\epsilon_1)$ in a worst-case fashion as a piece-wise constant
function. To do this, note that $h(\epsilon) \leq h(e_2)$ for all $\epsilon \leq e_2$, and
$h(\epsilon) \leq h(\epsilon_1)$ for all $\epsilon \in (e_2, \epsilon_1)$, and $h(\epsilon) \leq 1$ for all $\epsilon > \epsilon_1$. Then the area theorem gives that
$$1 - \epsilon_1 + h(\epsilon_1) (1 - e_2) + h(e_2) e_2 \geq R$$
We note that
$$h(\epsilon) = R h_{sys}(\epsilon) + (1 - R) h_{non-sys}(\epsilon)$$
Using the above two relations, we have
$$R h_{sys}(\epsilon_1) \geq R - (1 - \epsilon_1) - h(e_2) e_2 - (1 - R) h_{non-sys}(\epsilon_1)$$
Using $h_{non-sys} \leq 1$, we get
$$h_{sys}(\epsilon_1) \leq \frac{1}{R(1 - e_2)} (R - (1 - \epsilon_1) - h(e_2) e_2) - (1 - R).$$
(a) Applying Lemma 1a to bound $h(e_2)$ we get from (8)
$$\bar{h}_{sys}(\epsilon_1) \geq \zeta(2\delta_2, e_2, \epsilon_1).$$

The bounds on BER follow from noticing that for a linear
systematic code
$$\text{BER}(\epsilon) \geq \frac{2}{2} \bar{h}_{sys}(\epsilon)$$
(b) Applying Lemma 1b to bound $h(e_2)$ in (9) gives
$$h_{sys}(\epsilon_1) \geq \zeta(\bar{h}_b(\text{BER}(\epsilon_1), e_2).$$
Let $X = \hat{X}_i(Y_{-i}(\epsilon_1))$ be the MAP decoder of $X_i$ given
$Y_{-i}(\epsilon_1)$. From Fano’s and Jensen’s inequalities we have
$$h_{sys}(\epsilon_1) \leq \frac{1}{k} \sum_{i=1}^{k} H(X_i | Y_{-i}(\epsilon_1)) \leq \frac{1}{k} \sum_{i=1}^{k} h_b(1 | X_i \neq \hat{X}_i)$$
$$\leq h_b \left( \frac{1}{k} \sum_{i=1}^{k} 1 | X_i \neq \hat{X}_i \right) \text{.}$$
Now, notice that $1 | X_i \neq \hat{X}_i = \epsilon_i | X_i \neq \hat{X}_i]$ and, thus,
$$h_{sys}(\epsilon_1) \leq h_b(1 | \text{BER}(\epsilon) \text{)}.$$

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 LDLC codes can
outperform any linear code that has comparable performance
at some lower values of $C/R$. Fig. 1 shows that LDLC codes
can uniformly dominate repetition. By Proposition 6, no
linear systematic code can dominate repetition. It should thus
be expected that LDLCs (and non-linear codes in general)
can achieve certain performance 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


