Triangulation codes: a family of non-linear codes with graceful degradation

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

***Abstract*—This paper introduces the notion of triangulation codes, a family of non-linear codes that 1) admit efficient encoding/decoding 2) their bit error rate deteriorates gracefully as the quality of the erasure channel degrades. Some coding theoretic properties of these codes are established. In the case of transmitting data over the erasure channel, it is shown that, even with sub-optimal decoding, they can achieve lower bit error rate than uncoded transmission for any number of received output symbols.**

I. Introduction

Consider the problem of transmitting data over a memory-less erasure channel. A message of length k is encoded into a codeword of length n and sent through the channel. The channel randomly removes some of the codeword bits and returns the rest. When the erasure probability is known, one can optimize n and k so that most input bits, in the limit of large k, be recovered without error. This can be done, for instance, with the use of capacity achieving codes. However, if the capacity of the channels falls slightly below the rate for such codes, then the error in recovering the input grows significantly as can be seen in Fig.1. The bit error rate (BER) is thus too sensitive to noise variations in the channel for such codes. Roughly speaking, there is a phase transition in the performance of capacity achieving codes. When the noise level is below a certain threshold the input can be recovered with small error. When the noise level exceeds that threshold the input cannot be recovered with good fidelity.

There are various applications where either the noise level is not known or long delays cannot be tolerated. Problems of sequential data recovery in video/audio streaming [1] or information dispersal are often of this nature [2]. Another application where delay becomes relevant is control over a communication channel [3]–[6]. In these applications it is important for the controller to have access to a recent history of the state estimates. Thus the messages transmitted over the channel cannot be too long. It is useful to have a coding scheme that progressively improves on state estimates as the observed data streams into the controller.

As another application, consider the problem of one to many communication. In this case one sender needs to transmit data to multiple users. Each receiver has access to data through a different channel depending on the distance, location, time, etc. In addition, the users may need to access the data with varying levels of fidelity as well. Rateless codes have been popular in such settings [7]–[9], but they suffer from a delay problem. To perfectly decode a k-bit message, they need to observe $k + o(k)$ output bits. They have poor recovery if fewer than k output bits are received. Their behavior thus resembles that of the LDPC codes as in Fig.1. The number of excess output bits (beyond k-bits) required to recover (most of) the input data is known as the overhead. In general, known ratelss codes with small overhead do not have the graceful degradation propety [10]. To alleviate this issue, some authors have considered rateless codes with the unequal protection property [11]. In these constructions some input bits are given higher priority and can be retrieved from fewer output observations. However, this approach merely shifts the phase transition point of the more important bits at cost of further delay for the less important bits. Other authors have considered the use of partial feedback in the encoding process [12]–[18]. The general result in this area is that, under partial feedback, it is possible to achieve graceful degradation at the cost of some increase in the overhead. The reliance on feedback, however, makes the encoding process more complicated and harder to apply to certain settings such as that of the multi-cast transmission disccused above. Furthermore, these codes do not seem to perform any better than uncoded transmission when the ratio of received symbols to data size is below $\frac{3}{4}$ (see the empirical results in [12], [18]).

Graceful degradation is a long sought goal in coding theory [19]. It was reported in [19] that this property does not exist in usual coding/decoding schemes designed on the basis of minimum probability of block or bit error. Hence [19]–[23] considered coding/decoding schemes based on the mean square error criterion. It was observed in [21] that mean square decoding can result in graceful degradation in the bit error sense as well over AWGN channels.

In this work we formulate a version of the graceful degradation problem for the erasure channel. Consider the schemes shown in Fig.1. One is the repetition code and the other is an LDPC code. There is a distinct jump in the performance of the LDPC code as the capacity of the channel degrades below the rate of the code. This is known as the waterfall phenomenon in the litarature [24] and is not particular to LDPC codes. For instance, all LT-type codes that have small overhead suffer from the same problem [10]. As the rate approaches zero, the error curve for the capacity achieving code approaches a step function, while the repetition coding can achieves the curve shown in Fig.1. Of course, it is possible to juxtapose the two codes to interpolate between the two curves. This would result

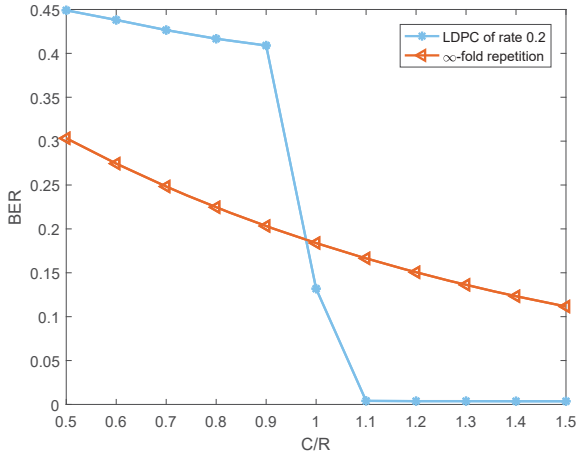


Fig. 1: Input BER for an LDPC code with 200 data bits and rate 0.2, and the repetition code with rate 0. The LDPC code is decoded using maximum likelihood decoding. Here C is the capacity of the channel and R is the rate of the code.

in a code that can outperform repetition on one side of the critical point $\frac{C}{R} = 1$. Here C is the capacity of the channel and R is the rate of the code. Note that for large k, $\frac{C}{R}$ measures the ratio of successful transmissions relative to the length of the input. We ask the following question: do there exist binary codes with efficient encoding/decoding that can outperform repetition (or, equivalently, uncoded transmission) for all $\frac{C}{R}$ as the rate approaches zero? For our purposes, a code with this property is said to have the graceful degradation property. In the language of rateless codes, we look for codes that do better than uncoded transmission for any number of received symbols.

We introduce the notion of triangulation codes to give an affirmative answer to the above question. These are non-linear codes that have smooth input-output distance properties (in a precise sense defined in section II), admit efficient decoding, and outperform repetition code even with simple sub-optimal decoding. To our knowledge, this is the first example of a code that has the graceful degradation property in the above sense. Compared to the previous work on rateless codes [12]–[18], our scheme 1) admits linear time encoding/decoding 2) does not require feedback 3) outperforms uncoded transmission for any number of received symbols.

II. Background

The (α, β)-property was introduced in [27] as a generalization of minimum distance and is closely realted to the combinatorial Joint Source Channel Coding (JSCC) problem [28]–[30].

Definition 1. *A mapping $f : \mathbb{F}_q^k \to \mathbb{F}_q^n$ is said to be (α, β) if*

$$d_H(x, y) > \alpha k \implies d_H(f(x), f(y)) > \beta n$$

where d_H denotes the Hamming distance.

We also define

$$\beta(\alpha) := \inf_{d_H(x,x') > \alpha k} d_H(f(x), f(x')) - \frac{1}{n}$$

and

$$\beta^* := \sup_{\alpha < 1} \beta(\alpha). \tag{1}$$

The (α, β)-property is discussed in [27], [28], [31]–[33]. To achieve graceful degradation, it is desirable to have $\beta(\alpha)$ increase monotonically with α and $\beta^* = 1$ [33]. Indeed if a code has this property, then a small variation in the erasure noise cannot cause significant distortion in the input. Suppose x is mapped to $f(x)$ and transmitted through an erasure channel, where βn coordinates are erased. If f has the (α, β)-property, then any point in the pre-image of channel output is distance at most αk away from x. It is thus desirable to have $\beta(\alpha)$ increase monotonically with α.

The codes that we construct in this paper can indeed achieve $\beta^* = 1$, have monotone $\beta(\alpha)$, and admit efficient decoding. We remark that these codes do not have positive minimum distance. It is shown in [33] that, over large alphabets, linear codes achieving $\beta^* = 1$ cannot have positive minimum distance.

In the next section, we introduce several constructions for such maps. We first study their (α, β)-properties and further prove an upper bound on their BER performance, which is rate independent as is our construction. In this sense our codes are rateless. It turns out that these codes have the graceful degradation property with linear time encoding/decoding complexity. We will prove in section IV that for any number of received output bits these codes recover a higher fraction of input bits compared with uncoded transmission,

III. Construction

We start with a motivation to explain the origin of the term triangulation in our codes. We first select a set M consisting of m points $p_1, \cdots, p_m$ in the Hamming cube $\{0,1\}^k$. Then given a point $x \in \{0,1\}^k$, we encode it by computing distance $f : x \to d_H(x, p_i)$. Alternatively, one can compute inner products $f : x \to (x, p_1)$ over $\mathbb{R}$. In either way, we obtain a map $f : \{0,1\}^k \to ([0,k] \cap \mathbb{Z})^n$. This is the motivation for the term triangulation codes as we try to triangulate a point by computing its distances to (or inner products with) a set of fixed points. Then we can compose every coordinate of f with a map $g : \{0, \cdots, k\} \to \{0,1\}^l$ to obtain a binary code. It is easy to construct the inner code g to be distance preserving. Consider for instance the map $g : \{1, ..., k\} \to \{0,1\}^k$ sending $x \to (1_x, 0_{k-x})$, where 1_x is a sequence of x 1's and 0_{k-x} is a sequences of $k - x$ 0's. It is easy to see that this map preserves the L_1 distance, i.e., two points that have L_1 distance d will be send to two points that have Hamming distance d. Another distance preserving map, which is used below in our constructions, is to send x to the most significant bit of its binary representation. This is a one bit quantization of x and is distance preserving in the sense that two points with distance larger than $k/2$ will be mapped to two points that are maximally apart. Thus, to obtain a binary triangulation map

with good (α, β)-properties, the main step is to make f to be distance preserving in the L_1 sense, that is, we want the L_1 distance $d_1(f(x), f(x'))$ to grow smoothly with the Hamming distance $d_H(x, x')$. It is easy to see that when $d_H(x, x')$ is small then $d_1(f(x), f(x'))$ is small as well. Indeed if x' is close to x, then distances $d_H(x', p_i)$ (resp. inner products (x, p_i)) cannot be too different from the distance $d_H(x, p_i)$ (resp. (x, p_i)). The question is if the triangulation points can be picked in a way that preserve large distances as well. That is we want $d_1(f(x), f(x'))$ to be large whenever $d_H(x, x')$ is large.

We digress at this point to discuss other problems where triangulation matrices have appeared. When the collection M is such that the map f sending $x \mapsto xM$ is injective we call M to be a binary detection matrix[1]. It is easy to see that for such matrices to exist one needs $n \geq \frac{k}{\log k}$. Erdos and Rényi [25] prove $n \geq \frac{2k}{\log k}$ and that random matrices can achieve $n = \frac{\log_2 9k}{\log k}$ asymptotically. Lindstrom [26] showed that the constant 2 is optimal and gave an explicit construction with n approaching $2\frac{k}{\log k}$ asymptotically. It is thus natural to ask if these matrices are distance preserving in the above sense?

The answer is, unfortunately, no. Consider for instance a random triangulation matrices M and let M_i denote its i-th column. Consider $f_i = xM_i$. Two input vectors x, x' of weight $\frac{k}{2}$ that are maximally apart, i.e., have Hamming distance $d_H(x, x') = k$ will be mapped to two integers that are at most $\sqrt{k}$ apart with high probability for large k (see also the discussion in section III.A below). In other words, the maps obtained in this way have poor (α, β)-properties and turn out to have poor BER performance as well. The same problem exists with Lindstrom matrices. However, it turns out that one can instead use low weight triangulation points together with 1 bit quantization to obtain good binary codes. This is equivalent to the following construction.

Given a point $x \in \{0, 1\}^k$ and a subset $S \subset \{1, \cdots, k\}$ of coordinates, define $\mathrm{maj}_S(x)$ to be the binary function that takes the value 1 if x has more 1's along S than 0's and takes value 0 otherwise. For instance, if $x = (1, 0, 1, 0)$ and $S = \{1, 2, 3\}$, then $\mathrm{maj}_S(x) = 1$. Note that as discussed above, $\mathrm{maj}_S(x)$ can equivalently be defined as the 1 bit quantization of the inner product between x and a "triangulation"vector y such that $\mathrm{supp}(y) = S$. Given a collection Σ of subsets define

$$f_\Sigma(x) : x \mapsto (\mathrm{maj}_S(x))_{S \in \Sigma}$$

We consider various collection of Σ that give rise to triangulation codes and study their properties.

A. Bernoulli sampling

Here we construct Σ according to a sequence of $\mathrm{Ber}(q)$ processes. That is, we associate an arrival process to each set S and place the i-th coordinate in S if there is an arrival at the i-th step. Set $|x| = w$. Then $\mathrm{maj}_S(x)$ is a random variable whose law is determined as follows:

$$\mathbf{P}[\mathrm{maj}_S(x) = 1] = \mathbf{P}[\mathrm{Bin}(w, q) > \mathrm{Bin}(k - w, q)]$$

[1]Binary detection matrices are used in code-division multiple access (CDMA)

The above probability is asymptotically 0 if $w < 2k$ and 1 if $w > 2k$. In other words, it does not vary smoothly with the weight of x. Thus the corresponding map has poor (α, β)-properties. Indeed there are vectors with $d(x, y) = k - \epsilon$ for which $d(f_\Sigma(x), f_\Sigma(y)) = 0$. This problem can however be fixed by considering sparse subsets S as is done in the next section.

B. Poisson sampling

We consider a similar construction as above but this time the arrival process is Bernoulli of rate l/k where l is fixed. Asymptotically, this process is equivalent to a Poisson process X of rate l. However, the majority function now has a law that varies smoothly with the weight $|x| = \omega k$:

$$\mathbf{P}[\mathrm{maj}_S(x) = 1] = \mathbf{P}[X^1 > X^0]$$

where $X^1 \sim \mathrm{Poi}(l\omega)$ and $X^0 \sim \mathrm{Poi}(l(1-\omega))$. In other words, we are splitting the arrival process into two processes: one on $\mathrm{supp}(x)$ and the other on its compliment. Alternatively we can describe the law in the form

$$\mathbf{P}[\mathrm{maj}_S(x) = 1] = \mathbf{P}[Z > 0]$$

where Z is Skellam distributed with parameters $l\omega$ and $l(1 - \omega)$. In constructing S, one may condition on the event that $X > 0$ to rule out the empty subset and make the encoding more efficient. In this case

$$\mathbf{P}[\mathrm{maj}_S(x) = 1] = \mathbf{P}[X^1 > X^0 | X^0 + X^1 > 0]$$

In the rest of this section all probabilities are conditioned on the event that there is at least one arrival in the underlying process. To analyze the (α, β)-properties of f we need to determine the join distribution of $f_\Sigma(x)$ and $f_\Sigma(x')$. For fixed x, x' we can split X into four processes $X^{00}, X^{01}, X^{10}, X^{11}$ where X^{ij} is the process supported on the susbet where x takes value i and x' takes value j. Then

$$\mathbf{P}[\mathrm{maj}_S(x) = 1, \mathrm{maj}_S(x) = 0] =$$
$$\mathbf{P}[X^{11} + X^{10} > X^{00} + X^{01}, X^{11} + X^{01} \leq X^{10} + X^{00}]$$

We can rewrite this as

$$p_{01} := \mathbf{P}[\mathrm{maj}_S(x) = 1, \mathrm{maj}_S(x) = 0]$$
$$= \mathbf{P}[X^{11} - X^{00} > X^{01} - X^{10}, X^{11} - X^{00} \leq -X^{01} + X^{10}]$$

Similarly, we have that

$$p_{10} := \mathbf{P}[\mathrm{maj}_S(x) = 0, \mathrm{maj}_S(x) = 1]$$
$$= \mathbf{P}[X^{11} - X^{00} \leq X^{01} - X^{10}, X^{11} - X^{00} > -X^{01} + X^{10}]$$

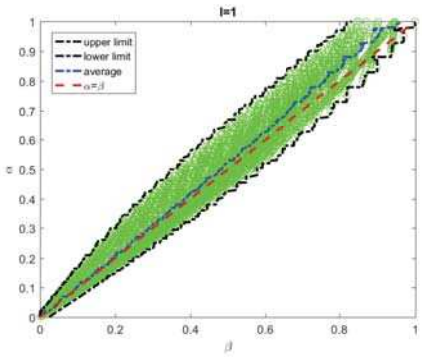
Then

$$\mathbb{E}[d(f_\Sigma(x), f_\Sigma(y))] = (p_{01} + p_{10})n$$

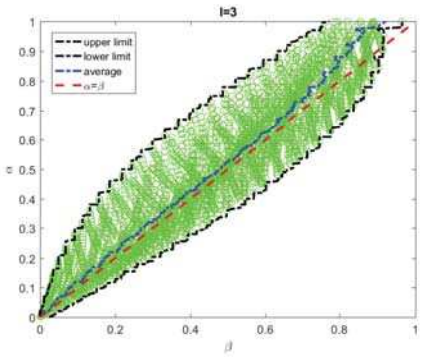
In other words

$$\mathbb{E}[\beta_f(\alpha)] = p_{01} + p_{10} \tag{2}$$

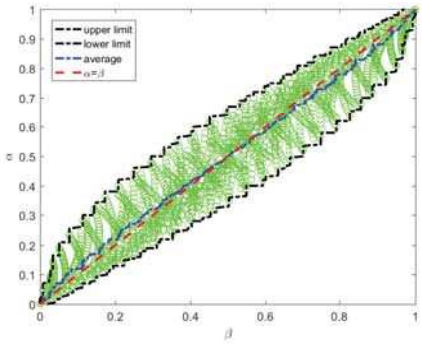
where the expectation is taken over all possible realizations of f. Let ω_1 be the relative weight of x' on $\mathrm{supp}(x)$ and let ω_2 be the relative weight of x' on the complement of $\mathrm{supp}(x)$. Clearly, the $d(f(x), f(x'))$ depends on $\omega, \omega_1, \omega_2$ ad the arrival rate l. We can vary these parameters and compute the above integral numerically to determine the set of achievable (α, β)-pairs. This is done for some special cases in Fig.2.



(a) $l = 1$ section B



(b) $l = 3$ section B



(c) $l = 3$ section C

Fig. 2: The set of achievable (α, β)-pairs for various arrival rates l. The dots represent the pairs $d(x, x')$ as a function of $d(f(x), f(x'))$ as $\omega, \omega^1, \omega^2$ vary. The alpha-beta curve is thus the upper envelop of the plotted dots. The construction of section C can achieve $\beta^* = 1$.

We remark that as $l \to 0$ the (α, β)-spectrum concentrated around the $\alpha = \beta$ line. This is indeed the expected (α, β)-curve. It can be achieved with high probability for small rates as the following proposition shows.

Proposition 1. *For every ϵ, there exists a triangulation code f with $n = \frac{k}{\epsilon^2}$ such that $d(f(x), f(y)) - \mathbb{E}[d(f(x), f(y)] \geq -\epsilon n$ for all pairs x, y. In other words,*

$$\beta_f(\alpha) \geq \mathbb{E}[\beta_f(\alpha)] - \epsilon$$

for all α.

Proof. Let f_i denote the i-th coordinate of f. Clearly

$$d(f(x), f(y)) = \sum_i d(f_i(x), f_i(y)),$$

where each $d(f_i(x), f_i(y))$ is a Bernoulli $p_{01} + p_{10}$ variable (see Eq.2). By the Hoeffding's inequality

$$\mathbb{P}(d(f(x), f(y)) - \mathbb{E}[d(f(x), f(y))] \leq -\epsilon n) \leq e^{-2n\epsilon^2}$$

There are at most 2^{2k} pairs of messages at a given distance. Hence, by the union bound, if

$$2^{2k} e^{-2n\epsilon^2} < 1 \tag{3}$$

there must exist some f for which $\frac{d(f(x), f(y))}{n}$ is not less than $\frac{\mathbb{E}[d(f(x), f(y))]}{n} - \epsilon$. Note that Eq.3 can be achieved if $n > \frac{k}{\epsilon^2}$ ■

C. Uniform sampling from a Hamming sphere

One problem with the previous construction, from an (α, β)-perspective, is that some fraction of the subsets in Σ may have even cardinality. This prevents the code from achieving $\beta^* = 1$. It is easy to check that if all $S \in \Sigma$ has odd cardinality then $\beta^* = 1$.

Here we consider a collection of subsets Σ where each $S \subset \{1, \cdots, k\}$ is the support of a point sampled uniformly at random from the Hamming sphere of some fixed odd radius l around the origin. The (α, β)-curves for this collection are similar to that of the previous section as shown in Fig.2. They slightly improve on the $\beta(\alpha)$ and, in particular, can achieve $\beta^* = 1$. They also seem to have better BER performance than the codes of section B based on our empirical results. Their BER curve for $l = 3$ under linear programming decoding is shown in Fig.3.

D. Mixing on Hamming spheres

We also consider triangulation codes obtained by sampling points on spheres of radius l where l is uniformly distributed on $\{1, 3\}$. Each subset $S \in \Sigma$ is the support of such sampled point. The purpose of this construction is to enable the use of a peeling type decoder. The weight 1 columns are introduced to get the decoding started. This is by no means an optimized construction. However it makes the analysis easier and suffices to theoretically prove the desired graceful degradation property of these codes. We obtain an analytic upper bound on the BER for these codes in the next section through a density evolution argument. The empirical performance as well as the analytic bound are shown in Fig.3. It can be seen that these codes outperform the ∞-fold repetition code for all values of $\frac{C}{R}$.

E. Linear programming decoding

To (approximately) decode the triangulation codes one can solve the following linear program

$$\begin{aligned} \text{minimize} \quad & |x| \\ & 0 \leq x_i \leq 1 \quad \forall i, \\ & (x, g_j) > \frac{|g_j|}{2} \quad \forall j \text{ s.t } y_j = 1, \\ & (x, g_j) \leq \frac{|g_j|}{2} \quad \forall j \text{ s.t } y_j = 0 \end{aligned} \tag{4}$$

The results for the construction in section C with $l = 3$ and $k = 3000$ are demonstrated in Fig.3.

F. Why non-linearity is needed

We discuss briefly here why it is useful to work with non-linear maps in the context of the problems in hand. Given a binary vector x denote by $\mathrm{supp}(x)$ the set of its non-zero coordinates. Given two vectors x, y in the Hamming cube, we say $x < y$ if $\mathrm{supp}(x) \subset \mathrm{supp}(y)$.

Definition 2. *A function $f : \{0,1\}^k \to \{0,1\}^n$ is said to be monotone if $x < y \implies f(x) < f(y)$*

Such function are good candidates to produce maps with monotonically increasing $\beta(\alpha)$. For instance, any linear repetition like map is monotone[2]. A simple observation is that

Proposition 2. *Repetition like maps are the unique binary linear maps that are monotone.*

Proof. Every vector in the generator matrix must be of weight 1 for a linear map to be monotone. ∎

It is therefore natural to look for monotone maps that are non-linear.

IV. BER UNDER ITERATIVE MAJORITY LOGIC DECODING

In this section, we introduce an iterative decoding algorithm with linear time complexity (in k) and find the corresponding BER for the triangulation codes of construction III.D . For an input bit x_i, let Δ_i be the collection of subsets in Σ containing the i-th coordinate. The decoding consists of two phases:

- A "peeling"phase where the decoder determines some of the input bits with no error. Let Σ be collection of subsets associated to the output bits returned by the channel. For each S, we denote the corresponding output bit by y_S. Initially, we set the set of decoded inputs $D := \emptyset$. For all $S \in \Sigma$ with $|S| = 1$ we set $\hat{x_S} := y_S$ and update $D = D \cup S$ and $\Sigma = \Sigma \backslash S$. Then, we iteratively go through all $S \in \Sigma$ and set $\hat{x_i} := y_S$ for $i \notin D$ if there exists $j, j' \in S \cap D$ with $\hat{x}_j \neq \hat{x}_{j'}$ and update $D = D \cup \{i\}, \Sigma = \Sigma \backslash S$. If there exists $j, j' \in S \cap D$ with $\hat{x_j} = \hat{x_j'}$ then simply remove S $\Sigma = \Sigma \backslash S$. We repeat this step until no further updates are possible in D.
- A second phase where a majority logic decoder estimates the remaining bits from Σ:

$$\hat{x}_i = \begin{cases} 0 & \sum_{S \in \Delta_i} y_S \leq \frac{|\Delta_i|}{2} \\ 1 & \text{otherwise} \end{cases}$$

  where Δ_i is the collection of subsets that contain i.

We can analyze the performance of this decoding strategy as follows. The erasure channel with $C = xR$ will return roughly xk random output bits. Each output bit corresponds to a set S with $|S| = 1$ with probability $\frac{1}{2}$. Thus the initial fraction of the determined bits is expected to be $1 - e^{-\frac{1}{2}x}$. Let q_t be the fraction of bits that are not determined after t steps. We have $q_0 = e^{-\frac{1}{2}x}$. Then Consider an undiscovered bit x_i. There are $\mathrm{Poi}(\frac{3}{2}x)$ subsets S with $|S| = 3$ that contain i. We say that a subset involving the (unknown) i-th bit is:

- very good: if the two other bits are known and are not equal.
- good: if at least one of the other two bits is not known.
- bad: if the other two bits are known and have the same value.

The subsets involving i-th bit arrive at a rate of $\frac{3}{2}x$ and split into three processes as described above. At the first step, the very good subsets for the i-th bit arrive according to a Poisson process of rate $\lambda_1 := \frac{3}{4}(1 - q_0)^2$. Thus

$$q_1 = q_0 \mathbb{P}(\mathrm{Poi}(\lambda_1) = 0) = q_0 e^{-\lambda_1}.$$

At iteration t, an undecoded bit remains undecoded if no new very good subset is generated. The probability that a subset involving an unknown bit becomes very good at the t-th step is $\frac{dq_t^2 - dq_t(1-q_{t-1})}{2}$, where $dq_t = q_{t-1} - q_{t-2}$. Indeed a given subset can be turned into a very good one at step t if it either involves two bits of opposite values that are discovered at step $t-1$, or involves a bit discovered at step $t-1$ and a previously discovered bit, again of opposite values. Thus $\lambda_t = \frac{3x(dq_t^2 - 2dq_t(1-q_{t-1}))}{4}$ and

$$q_t = q_{t-1} e^{-\lambda_t}.$$

Once this process converges, we use the good subsets that are not very good to estimate the remaining bits. Let Δ_i be the set of good subsets involving an undecoded bit. Then for each observed bit y_S for $S \in \Delta_i$ we have

$$\mathbb{P}(x_i \neq y_S) = \frac{1}{4}$$

On the other hand, the events $\{x_i \neq y_S\}_{S \in \Delta_i}$ are asymptotically independent. Indeed for two such events to be dependent we need for some coordinate $j \neq i$ to appear in at least two different subsets S, S'. A simple counting argument shows that this event happens with vanishing probability.

Let $q_t \to q_\infty$ and set $p_\infty = 1 - q_\infty$. Then a subset involving the i-th bit is either bad with probability $p_\infty^2/2$ or good with probability $q_\infty^2 + 2q_\infty p_\infty$. Hence, the good subsets in the limit arrive at a rate $\lambda_\infty^g := \frac{3x(q_\infty^2 + 2q_\infty p_\infty)}{2(q_\infty^2 + 2q_\infty p_\infty + p_\infty^2/2)}$.

Now we can write the probability of error under majority decoding in the form

$$p_i = \frac{1}{2}(\mathbb{P}(\hat{x_i} = 1 | x_i = 0) + \mathbb{P}(\hat{x_i} = 0 | x_i = 1))$$

We can upper bound each term as follows

$$\mathbb{P}(\hat{x_i} = 1 | x_i = 0) \leq \frac{1}{2}\mathbb{P}(|\Delta_i| = 0) +$$
$$\mathbb{P}(|\Delta_i| > 20) + \sum_{k=1}^{20} \mathbb{P}(|\Delta_i| = k)\mathbb{P}(\mathrm{Bin(k, 1/4)} > \mathrm{k/2})$$

and

$$\mathbb{P}(\hat{x_i} = 0 | x_i = 1) \leq \frac{1}{2}\mathbb{P}(|\Delta_i| = 0) +$$
$$\mathbb{P}(|\Delta_i| > 20) + \sum_{k=1}^{20} \mathbb{P}(|\Delta_i| = k)\mathbb{P}(\mathrm{Bin(k, 1/4)} \geq \mathrm{k/2})$$

where $\mathbb{P}(|\Delta_i| = k) = \mathbb{P}(\mathrm{Poi}(\lambda_\infty^g)) = k)$. Putting the two decoding phases together, we get the results shown in Fig.3. We note that in our analysis we only used the fact that

[2]A linear map is said to be repetition like if its generator matrix consists entirely of vectors with Hamming weight 1.

$n, k \to \infty$, but our analysis did not depend on the rate. Hence, the BER curve of Fig.3 serves as a rate-independent upper bound on the performance of triangulation codes considered in this section. We also remark that the uniform sampling on Hamming sphere under LP decoding has lower BER than the construction discussed here as can be seen in the figure. We merely used this construction due to its amenability for analysis. The empirical performance of the iterative decoding scheme discussed here agrees with two steps of the density evolution approach.

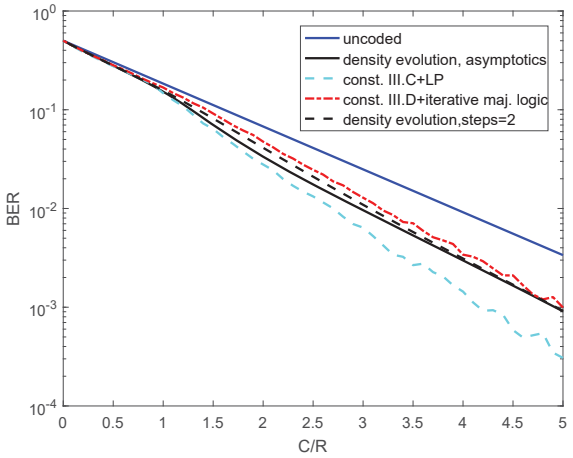


Fig. 3: The BER performance of triangulation code of sections III.C and III.D for k=3000. The BER curve is rate independent.

REFERENCES

[1] L. Xu, "Resource-efficient delivery of on-demand streaming data using uep codes," *IEEE Transactions on Communications*, vol. 51, no. 1, pp. 63–71, 2003.

[2] P. Maymounkov and D. Mazieres, "Rateless codes and big downloads," *Peer-to-Peer Systems II*, pp. 247–255, 2003.

[3] V. Kostina and B. Hassibi, "Rate-cost tradeoffs in control," in *Communication, Control, and Computing (Allerton), 2016 54th Annual Allerton Conference on*. IEEE, 2016, pp. 1157–1164.

[4] A. Khina, V. Kostina, A. Khisti, and B. Hassibi, "Sequential coding of gauss–markov sources with packet erasures and feedback," in *Proc. IEEE Info. Theory Workshop (ITW), Kaohsiung, Taiwan*, 2017.

[5] S. Tatikonda, A. Sahai, and S. Mitter, "Stochastic linear control over a communication channel," *IEEE transactions on Automatic Control*, vol. 49, no. 9, pp. 1549–1561, 2004.

[6] A. Sahai and S. Mitter, "The necessity and sufficiency of anytime capacity for stabilization of a linear system over a noisy communication link—part i: Scalar systems," *IEEE transactions on Information Theory*, vol. 52, no. 8, pp. 3369–3395, 2006.

[7] P. Maymounkov, "Online codes," Technical report, New York University, Tech. Rep., 2002.

[8] M. Luby, "Lt codes," in *Foundations of Computer Science, 2002. Proceedings. The 43rd Annual IEEE Symposium on*. IEEE, 2002, pp. 271–280.

[9] A. Shokrollahi, "Raptor codes," *IEEE transactions on information theory*, vol. 52, no. 6, pp. 2551–2567, 2006.

[10] S. Sanghavi, "Intermediate performance of rateless codes," in *Information Theory Workshop, 2007. ITW'07. IEEE*. IEEE, 2007, pp. 478–482.

[11] N. Rahnavard, B. N. Vellambi, and F. Fekri, "Rateless codes with unequal error protection property," *IEEE Transactions on Information Theory*, vol. 53, no. 4, pp. 1521–1532, 2007.

[12] A. Kamra, V. Misra, J. Feldman, and D. Rubenstein, "Growth codes: Maximizing sensor network data persistence," in *ACM SIGCOMM Computer Communication Review*, vol. 36, no. 4. ACM, 2006, pp. 255–266.

[13] A. Beimel, S. Dolev, and N. Singer, "Rt oblivious erasure correcting," *IEEE/ACM Transactions on Networking (TON)*, vol. 15, no. 6, pp. 1321–1332, 2007.

[14] A. Hagedorn, S. Agarwal, D. Starobinski, and A. Trachtenberg, "Rateless coding with feedback," in *INFOCOM 2009, IEEE*. IEEE, 2009, pp. 1791–1799.

[15] M. Hashemi, A. Trachtenberg, and Y. Cassuto, "Delete-and-conquer: Rateless coding with constrained feedback," in *Communication, Control, and Computing (Allerton), 2013 51st Annual Allerton Conference on*. IEEE, 2013, pp. 350–357.

[16] A. Talari and N. Rahnavard, "Lt-af codes: Lt codes with alternating feedback," in *Information Theory Proceedings (ISIT), 2013 IEEE International Symposium on*. IEEE, 2013, pp. 2646–2650.

[17] M. Hashemi and A. Trachtenberg, "Near real-time rateless coding with a constrained feedback budget," in *Communication, Control, and Computing (Allerton), 2014 52nd Annual Allerton Conference on*. IEEE, 2014, pp. 529–536.

[18] Y. Cassuto and A. Shokrollahi, "Online fountain codes with low overhead," *IEEE Transactions on Information Theory*, vol. 61, no. 6, pp. 3137–3149, 2015.

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

[20] G. Wolf and G. Redinbo, "The optimum mean-square estimate for decoding binary block codes," *IEEE Transactions on Information Theory*, vol. 20, no. 3, pp. 344–351, 1974.

[21] G. Redinbo and W. Cheung, "The design and implementation of unequal error-correcting coding systems," *IEEE Transactions on Communications*, vol. 30, no. 5, pp. 1125–1135, 1982.

[22] G. Redinbo, "Optimum symbol-by-symbol mean-square error channel coding," *IEEE Transactions on Information Theory*, vol. 25, no. 4, pp. 387–405, 1979.

[23] L. A. Dunning and W. E. Robbins, "Optimal encodings of linear block codes for unequal error protection," *Information and control*, vol. 37, no. 2, pp. 150–177, 1978.

[24] T. Richardson, "Error floors of ldpc codes," in *Proceedings of the annual Allerton conference on communication control and computing*, vol. 41, no. 3. The University; 1998, 2003, pp. 1426–1435.

[25] P. Erdos and A. Rényi, "On two problems of information theory," *Magyar Tud. Akad. Mat. Kutató Int. Közl*, vol. 8, pp. 229–243, 1963.

[26] B. Lindström, "On a combinatorial problem in number theory," *Canad. Math. Bull*, vol. 8, no. 4, pp. 477–490, 1965.

[27] Y. Polyanskiy, "On metric properties of maps between hamming spaces and related graph homomorphisms," *arXiv preprint arXiv:1503.02779*, 2015.

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

[29] A. J. Young and Y. Polyanskiy, "Converse and duality results for combinatorial source-channel coding in binary hamming spaces," in *Information Theory (ISIT), 2015 IEEE International Symposium on*. IEEE, 2015, pp. 261–265.

[30] Y. Kochman, A. Mazumdar, and Y. Polyanskiy, "Results on combinatorial joint source-channel coding," in *Information Theory Workshop (ITW), 2012 IEEE*. IEEE, 2012, pp. 10–14.

[31] A. Mazumdar, Y. Polyanskiy, A. S. Rawat, and H. Roozbehani, "Distance preserving maps and combinatorial joint source-channel coding for large alphabets," in *Information Theory (ISIT), 2016 IEEE International Symposium on*. IEEE, 2016, pp. 3067–3071.

[32] Y. Polyanskiy and A. Samorodnitsky, "Improved log-sobolev inequalities, hypercontractivity and uncertainty principle on the hypercube," *arXiv preprint arXiv:1606.07491*, 2016.

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