On a class of optimal rateless codes

Vijay G. Subramanian and Douglas J. Leith
Hamilton Insitute, National University of Ireland, Maynooth, Co. Kildare, Ireland.
Email: [Vijay.Subramanian,Doug.leith]@nuim.ie.

***Abstract*—In this paper we analyze a class of systematic fountain/rateless codes constructed using Bernoulli$(1/2)$ random variables. Using simple bounds we then show that this class of codes stochastically minimizes the number of coded packets receptions needed to successfully decode all the information packets. This optimality holds over a large class of random codes that includes Bernoulli(q) random codes with $q \leq 1/2$ and LT codes. We then conclude by demonstrating asymptotic optimality for intermediate decoding of the same codes.**

I. Introduction

Since the introduction of LT codes by Luby [1] there has been considerable interest in rateless/fountain codes; the reader is referred to articles by MacKay [2], [3] and Mitzenmacher [4] for a good introduction to this topic. These codes have been proposed for streaming applications over a packet erasure channel so as to achieve good application layer performance without the encoder and decoder being aware of the packet erasure rate. In the operation of fountain codes, the encoder first has to buffer all the packets to construct a block that will subsequently be coded and then transmitted across the packet erasure channel. The receiver has to then wait for more receptions than the size of a block to successfully recover all the original packets in the block. As shown in Hyytia, *et al.* [5], for small block lengths the excess over a block can be quite large relative to the block length. From the results of [1] this excess becomes negligible relative to the block length for large block lengths. The delay performance of LT codes is particularly bad because the decoding process is such that none of the original packets can be recovered before the entire block can be decoded successfully. From the application's perspective waiting for all packets to be decoded can introduce a large delay before packets are transferred to the application layer. In this paper we explore the question of whether there exist codes such that subject to re-ordering constraints, packets can be released to the application layer well before the whole block can be decoded.

A. Background Material and Motivation

Rateless codes as proposed by [1] are generated as follows. Given a set of n packets $\{u_i\}_{i=1}^n$ of equal length to be relayed from a source to a destination, the encoder at the source first picks a probability distribution $\mathbb{P}_n$ on the set of all binary sequences of length n, equivalently a subset $\mathcal{X}$ of $\{1, 2, \ldots, n\}$. We will use the binary sequence representation of $\mathcal{X}$ as a row vector and the subset representation interchangeably. Each sequence leads to a coded packet $x(\mathcal{X})$ generated by a binary addition (bit-wise XOR) of the packets corresponding to the locations of the 1s in the sequence. The identity of the specific combination of packets that make up a specific coded packet is conveyed in the header. The coded packet corresponding to the all-zero sequence is a dummy packet that is distinguished by the all-zero sequence in the header; in practice, such a sequence would be suppressed and the performance would improve. Thereafter, the source generates an *i.i.d.* sequence of subsets $\{\mathcal{X}_i\}_{i=1}^{+\infty}$ of $\{1, 2, \ldots, n\}$ with distribution $\mathbb{P}_n$. At every transmission opportunity $i \in \mathbb{N}$ the source transmits the coded packet corresponding to $\mathcal{X}_i$, i.e., $x(\mathcal{X}_i)$ or x_i for brevity. This procedure is repeated up until the time that the decoder at the destination can reconstruct all the n packets. In general the probability distributions $\mathbb{P}_n$ are chosen such that equal probability is given to sequences of the same type/degree, i.e., having the same number of 1s. With such a restriction in place it is sufficient for the encoder to use a probability distribution on the possible degrees (with an abuse of notation also called $\mathbb{P}_n$) to generate a sequence of degrees $\{d_i\}_{i=1}^{+\infty}$. The specific binary sequence used for generating the coded sequence is then picked uniformly among all possible binary sequences of the given type.

Each coded packet is assumed to be transmitted across an erasure channel, i.e., with some probability $\epsilon \in [0, 1)$ the transmitted packet is lost, and otherwise it is successfully received at the decoder. We assume that neither the encoder nor the decoder have knowledge of ϵ or of any bounds on ϵ; in other words, we are interested in universal schemes. At any given time i the decoder generates a received matrix $G(i)$ by filling up the rows with the binary sequences received up until time i. All the coded packets can be recovered at the first time i when $G(i)$ has rank n in $\mathcal{F} := \mathbf{GF}(2)$. The decoder determines this by attempting to invert $G(i)$ using Gaussian elimination in $\mathcal{F}$. Often a simplification of the Gaussian elimination procedure that inverts $G(i)$ by starting with the received sequences of degree 1 is followed. The sequences of degree 1, if any, are subtracted from the other received sequences. If any new sequences of degree 1 emerge, then the procedure is repeated up until either all n packets have been reconstructed or no new sequences of degree 1 emerge. Using a graphical representation of $G(i)$ this simpler procedure can be implemented in a very efficient manner using belief-propagation/message-passing algorithms.

The analysis of rateless codes is carried out in the

This work is supported by SFI grant 07/IN.1/I901. The authors would like thank David Malone, NUIM for his comments and suggestions.

asymptotic regime of large n under the assumption that the sequence $\{\mathbb{P}_n\}_{n=1}^{+\infty}$ converges point-wise (weak convergence) to a probability distribution on $\{0\} \cup \mathbb{N}$ as n increases without bound. With the belief-propagation decoding algorithm [1] showed that the soliton distribution optimized in expectation the number of coded packets that need to be received so as to fully recover all the uncoded packets. This ideal distribution, however, works very poorly in practice in conjunction with the belief-propagation decoding algorithm. Instead the sequence of distributions $\mathbb{P}_n$ used in practice are from the robust soliton distribution that are chosen so as to converge point-wise to the soliton distribution. Using precoding Shokrollahi [6] and Maymounkov [7] proposed linear encoding and decoding schemes that only need a small but linear in n excess coded packets over n to recover all the uncoded packets using the belief propagation algorithm. Sanghavi [8] considered a related question wherein only a fixed fraction packets need to be decoded with high probability but using the belief propagation algorithm, and posed the problem of determining the optimal degree distribution for each intermediate decoding choice. Using the results of Darling and Norris [9] which were shown to be applicable to this problem by Maneva and Shokrollahi [10], it was shown in [8] that the optimal distributions are constant in intervals $[1-1/k, 1-1/(k+1)]$ for $k \in \mathbb{N}$ with support only in $\{1, 2, \ldots, k\}$. While they can be determined exactly for $k = 1$ and 2, tight bounds using linear programs were provided for the rest of the region. A key finding was that using the class of degree distributions described above and the belief propagation algorithm, to decode $\lfloor rn \rfloor$ packets for $r \in [0, 1]$ one always needs to receive an asymptotically linear in n (with positive slope) excess of coded packets over $\lfloor rn \rfloor$ except for $r \in \{0, 1\}$. In fact the excess can be quite severe: for example, with $r = 1/2$ one gets $\log(2) - 0.5 = 0.1931$ to be the lowest asymptotic excess. Even though the soliton distribution does not satisfy the requirements of the theory in [9], using approximations to the soliton distribution [8] showed that it would be the worst degree distribution, i.e., intermediate decoding at fraction r is not possible up until the time that all uncoded packets can be recovered with any linear excess less than $1 - r$.

Since rateless codes are often considered for streaming applications, coding-decoding delay is of great importance and intermediate decoding would be beneficial in reducing the playback buffer subject to re-ordering issues. With this in mind the impact of the asymptotic excess can be best appreciated when one considers average packet delays. Specifically, we are interested in the delay from when the encoder has all the information packets to when each packet can be decoded. For $r \in [0, 1]$ if an encoding scheme needs a $f(r)n$ normalized number of coded packets to be transmitted for rn information packets to be decoded (for large n), then the normalized average delay is given by $\int_0^1 f(r)dr$; asymptotically the real average delay scales like $n \int_0^1 f(r)dr$. Note that $f(r) \geq r$ and therefore a lower bound on the normalized average delay is 0.5. Using the results of [8] one can find a tighter lower bound for encoding schemes with sparse degree distributions when the belief-propagation decoder is used; using numerical integration a better lower bound that results is 0.7003. As mentioned in [8] even this lower bound is not achievable by any one degree distribution, and therefore this metric could be a lot worse for any particular degree distribution. In this context it is noteworthy that the soliton distribution incurs 1 unit of normalized average delay. Thus one pays a significant delay price for any chosen sparse degree distribution.

The performance issues described above lead to the question of whether it is possible to get better intermediate decoding performance, and hence better delay performance, by changing some of the constraints. In Studholme and Blake [11] it was shown that by changing the class of degree distributions allowed and the decoding algorithm, it is possible to decode all the uncoded packets with almost constant overhead irrespective of the block length. In this paper we complement this result by showing, with a modification of a known code construction, that changing the class of degree distributions allowed and the decoding algorithm, allows one to also achieve optimal intermediate decoding universally, i.e., for any given $r \in [0, 1]$, (asymptotically) the decoder does not need to receive a linear excess of coded packets over $\lfloor rn \rfloor$. We also show optimality in terms of decoding performance for the code for finite block lengths, and in particular, for small block lengths where full Gaussian elimination is computationally feasible to implement. In addition to making the case for full Gaussian elimination, using the specific code construction we also propose expanding the search of encoding schemes to asymptotically defective degree distributions, i.e., by allowing the degree random variable to be infinite. In a graphical sense we seek to move away from the asymptotically sparse connectivity of the graphs in [9] and, instead, expand the search to graphs that are more connected. We then make a few comments on the minimum connectivity required in terms of the average degree to achieve (asymptotically) optimal performance.

This paper is structured as follows. We start by describing the systematic code construction in Section II and demonstrate how the systematic part improves the decoding performance of the code. In the same section we also present some useful bounds on the performance of the proposed code. Thereafter we contrast our codes with a large class of random codes that contains Bernoulli(q) random codes with $q \leq 1/2$ in Section III to demonstrate optimality of our proposal. This is extended to asymptotic optimality in terms of the zero excess needed to recover any given rank in Section IV and we conclude in Section V.

II. Code Construction

Our code proposal is a modification of a known random code and is systematic in nature. Assume that we have n information packets ($u_i,\ i = 1, \ldots, n$). The code construction can then be described as follows: if $i \leq n$, then we pick

$\mathcal{X}_i = \{i\}$; and for $i > n$ we pick $\mathcal{X}_i$ uniformly from among the subsets of $\{1, 2, \ldots, n\}$. The systematic part allows the encoder to start transmitting packets as soon as they arrive. This will necessarily improve the delay performance. From the description above it is clear that if $i > n$ we generate the coded packets by an XOR of a random number of information packets, i.e., $e_i = \bigoplus_{j=1}^{n} \tilde{q}_{i,j} u_j$ where $\tilde{q}_{i,j}$ is chosen in $\{0, 1\}$ equally likely. Thus the coded packets can be represented by row vector $\tilde{q}^i \in \mathcal{F}^n$.

During the decoding process, as new coded packets arrive, we check to see if the rank increases or not. As soon at the rank of the received packets reaches n, we perform inversion to decode the packets. The analysis is most easily understood if one takes a column-wise view; the rank calculation will be the same. Assume that we receive i of the uncoded packets. Thereafter assume that we receive l coded packets. Then we can collect all the received $\mathcal{X}$s in a matrix as follows:

$$G = \begin{bmatrix} I_{i\times i} & 0_{i\times n-i} \\ C_{l\times i} & D_{l\times n-i} \end{bmatrix},$$

where the dimensions of the sub-matrices are also listed. The matrix I is the $i \times i$ identity matrix with rank i since it is associated with reception of i unencoded packets. Both C and D have *i.i.d.* Bernoulli$(1/2)$ entries. Now it is clear that G is of full rank (n) if and only if D has rank $n - i$. Now each column of D is picked uniformly from one of 2^l strings of length l. We would like each of these to increase the rank of D. Say we have finished processing up until column j under the restriction that $j = 0, 1, \ldots, n - i - 1$ with $j = 0$ representing the initial state where no columns have been processed. Assuming that the rank is j, then the probability that the next column $((j+1)^{\text{th}})$ does not increase the rank is $p_j = \frac{2^j}{2^l}$; the rank j subspace can cover 2^j values while the total number of values possible is 2^l. When $j = 0$ the rank does not increase if the all-zero vector is produced. Thus the probability of decoding success is given by

$$S_{i,l} = \prod_{j=0}^{n-i-1} (1 - p_j) = \prod_{j=0}^{n-i-1} (1 - 2^{j-l}), \tag{1}$$

and the probability of decoding failure starting with rank i and having received l coded packets is

$$F_{i,l} = 1 - S_{i,l} = 1 - \prod_{j=0}^{n-i-1} (1 - 2^{j-l}). \tag{2}$$

Basically the uniform nature of the randomness and the specific transmission pattern that we have assumed, allows us to analyse the rank recovered as if from the very beginning.

Now assuming that the packet drop/erasure probability is ϵ, the probability of failure of decoding all the uncoded packets after receiving $n + \delta$ packets at the receiver is given by

$$F(n, \delta, \epsilon) := \sum_{i=0}^{n-1} \binom{n}{i} \epsilon^{n-i} (1 - \epsilon)^i F_{i,n+\delta-i}. \tag{3}$$

Before proceeding further we make one key observation with the help of the following inequalities. Note for $i \in \{0, 1, \ldots, n-1\}$ we have the following bound

$$\begin{aligned} S_{i,n+\delta-i} &= \prod_{j=0}^{n-i-1} (1 - 2^{j-n-\delta+i}) = \prod_{j=1}^{n-i} (1 - 2^{-j-\delta}) \\ &\geq \prod_{j=1}^{n} (1 - 2^{-j-\delta}) = \prod_{j=0}^{n-1} (1 - 2^{j-n-\delta}) = S_{0,n+\delta}. \end{aligned}$$

This then implies for $i \in \{0, 1, \ldots, n-1\}$ that $F_{i,n+\delta-i} \leq F_{0,n+\delta}$, and therefore also that

$$F(n, \delta, \epsilon) \leq \big(1 - (1 - \epsilon)^n\big) F_{0,n+\delta} \leq F_{0,n+\delta}. \tag{4}$$

In other words, the inequality in (4) says that directly adding the systematic part improves the decoding performance of our code. This is very different from what happens to the LT codes generated by the robust soliton distribution as observed in [6] where the performance degrades if the systematic part is directly added.

We now explore some bounds on $F_{i,l}$ that help in our analysis. A simple upper bound on $S_{i,l}$ is the following that uses $0 \leq 1 - x \leq e^{-x}$ for $x \in [0, 1]$, namely, $S_{i,l} \leq \exp\big(-2^{-l}(2^{n-i} - 1)\big)$. This then directly implies the following two lower bounds

$$\begin{aligned} F_{i,n+\delta-i} &\geq 1 - \exp\big(-2^{-n-\delta+i}(2^{n-i} - 1)\big) \\ &= 1 - \exp\big(-2^{-\delta}(1 - 2^{-n+i})\big) \\ F(n, \delta, \epsilon) &\geq 1 - (1 - \epsilon)^n \\ &\quad - \sum_{i=0}^{n-1} \binom{n}{i} \epsilon^{n-i} (1 - \epsilon)^i e^{2^{-\delta-(n-i)} - 2^{-\delta}}. \end{aligned} \tag{5}$$

Using the inequality $\prod_{j=1}^{k} (1 - x_j) \geq 1 - \sum_{j=1}^{k} x_j$ for $x_j \in [0, 1]$ for all $j = 1, 2, \ldots, k$ (this can be easily proved using induction), we get the following lower bound on $S_{i,l}$, namely, $S_{i,l} \geq 1 - 2^{n-i-l} + 2^{-l} \geq 1 - 2^{n-i-l}$. This then implies the following upper bounds

$$\begin{aligned} &F_{i,n+\delta-i} \leq 2^{-\delta} \text{ and} \\ &F(n, \delta, \epsilon) \leq (1 - (1 - \epsilon)^n) 2^{-\delta} \leq 2^{-\delta}. \end{aligned} \tag{6}$$

We now present an alternate upper bound using the classical Ky-Fan inequality [12]. Let $0 \leq y_i \leq 1/2$ be given for $i = 1, 2, \ldots, m$. Define $\bar{y}_i := 1 - y_i$. Then the Ky-Fan inequality states that

$$\left(\frac{\prod_{i=1}^{m} \bar{y}_i}{\prod_{i=1}^{m} y_i} \right)^{1/m} \geq \frac{\sum_{i=1}^{m} \bar{y}_i}{\sum_{i=1}^{m} y_i};$$

in other words, the ratio of the geometric means of the $\bar{y}_i$s and the y_is is greater than the ratio of their arithmetic means.

We can apply this to lower bound $S_{i,l}$ if we assume that l is chosen such that $n - i - 1 - l \leq -1$, i.e., if $l \geq n - i$. For the application in mind this is true since $\delta \geq 0$ implies l is at least $n - i$. Therefore assuming $l \geq n - i$ and using

the Ky-Fan inequality we get

$$S_{i,l} \geq \left(\frac{2^{\frac{n-i-1}{2}} \left((n-i) - \frac{2^{n-i}-1}{2^l} \right)}{2^{n-i}-1} \right)^{n-1} \quad \text{and}$$

$$F_{i,l} \leq 1 - \left(\frac{2^{\frac{n-i-1}{2}} \left((n-i) - \frac{2^{n-i}-1}{2^l} \right)}{2^{n-i}-1} \right)^{n-i}.$$

The upper bound this implies is then

$$F(n,\delta,\epsilon) \leq \sum_{i=0}^{n-1} \binom{n}{i} \epsilon^{n-i}(1-\epsilon)^i \times \left[1 - \left(\frac{2^{\frac{n-i-1}{2}} \left((n-i) - \frac{2^{n-i}-1}{2^{\delta+n-i}} \right)}{2^{n-i}-1} \right)^{n-i} \right]. \quad (7)$$

This is tight for small ϵ; in fact, for small enough ϵ it is better than the $2^{-\delta}$ upper bound. The tightest of upper bounds in (6) and (7), and the lower bound in (5) are compared against simulated performance of the code for a block-length $n = 5$ at different packet erasure rates in Figure 1.Note that in all cases the lower bound well approximates the real performance. However, the upper bound is typically tight only when the bound developed using the Ky-Fan inequality is better than the $2^{-\delta}$ upper bound.

III. Additional Properties of Equiprobable Codes

Before commenting upon the asymptotic performance of our code we prove a few more non-asymptotic properties for equiprobable codes. These show an optimality for choosing equiprobable binary sequences over a large class of random codes.

A. Performance of nonsystematic codes

Consider a random $l \times n$ matrix G over field $\mathcal{F}$. Assume the columns $G_1, ..., G_n$ are independent and

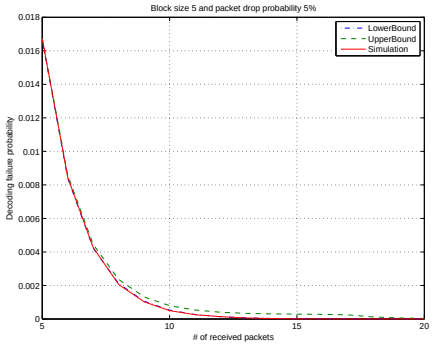
$$\mathbb{P}(G_i = v) = q_{n,i}(v), \quad v \in \mathcal{F}^l, \quad i = 1, \ldots, n.$$

For two vectors $u, v \in \mathcal{F}^l$ let $u \otimes v$ denote the vector obtained by bitwise AND of u and v. Let $1(u)$ denote the index set to the non-zero elements of vector u and $|1(u)|$ denote the number of non-zero elements in u. Then one defines the inner product uv of vectors u and v over field $\mathcal{F}$ as $uv = |1(u \otimes v)| \mod 2$. The finite-field Fourier transform [13] of the probability measure $q_{n,i}(u)$ is then

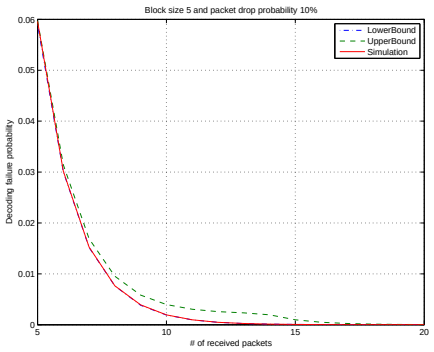
$$\hat{q}_{n,i}(u) = \sum_{v \in \mathcal{F}^n} q_{n,i}(v)(-1)^{uv}, \quad u \in \mathcal{F}^l, \quad (8)$$

since -1 is the unique additive character of $\mathcal{F}$.

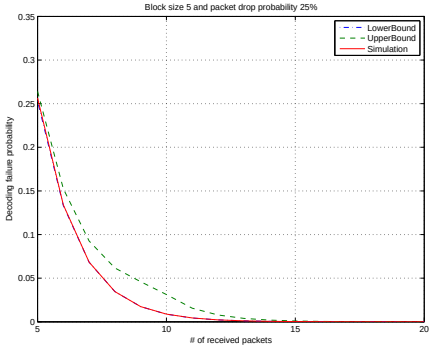
Following Sloane [14, Lemma 2.1], Alekseychuk [15], and [13], we have a key relationship for any k dimensional



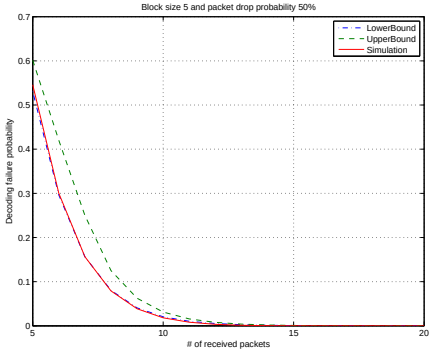
(a) Packet erasure probability: 5%



(b) Packet erasure probability: 10%



(c) Packet erasure probability: 25%



(d) Packet erasure probability: 50%

Fig. 1. Probability of decoding failure at receiver versus excess coded packets received with a systematic equiprobable code. At present resolution, the lower bound (dashes-and-dots) is indistinguishable from simulated performance (solid line).

subspace $\mathcal{L}_k$ of $\mathcal{F}^l$ that

$$\begin{aligned}\mathbb{P}(G_i \in \mathcal{L}_k) &= \sum_{v\in\mathcal{L}_k} q_{n,i}(v) \\ &= 1/|\mathcal{L}_k^{\perp}| \sum_{u\in\mathcal{L}_k^{\perp}} \hat{q}_{n,i}(u) \\ &= 2^{-(l-k)} \sum_{u\in\mathcal{L}_k^{\perp}} \hat{q}_{n,i}(u),\end{aligned} \tag{9}$$

where $\mathcal{L}_k^{\perp}$ is the orthogonal component of $\mathcal{L}_k$, i.e., the set of vectors in $\mathcal{F}^l$ that are orthogonal to every vector in $\mathcal{L}_k$, and $|\mathcal{L}_k^{\perp}|$ is the cardinality of $\mathcal{L}_k^{\perp}$.

Consider the simplest case where the elements of G are *i.i.d.* Bernoulli(q) random variables where $0 < q < 1$. By assuming (without loss of generality) that the non-zero elements of u occur first and by using the binomial formula, we can explicitly calculate

$$\hat{q}_{n,i}(u) = \begin{cases}(1-2q)^{|1(u)|} & \text{if } |1(u)| \neq 0; \\ 1 & \text{otherwise.}\end{cases} \tag{10}$$

Therefore if one can enumerate the sequences in $\mathcal{L}_k^{\perp}$ and tally the number of sequences of the same type, then from (10) and (9) one can obtain $\mathbb{P}(G_i \in \mathcal{L}_k)$ by evaluating the weight enumerator [14] of $\mathcal{L}_k^{\perp}$ at $(1, 1-2q)$; a manifestation of the MacWilliams identity. In the equiprobable situation, where $q = 1/2$, we recover the usual expression $\mathbb{P}(G_i \in \mathcal{L}_k) = 2^{-(l-k)}$ and $\pi_n(1/2) := \mathbb{P}(rank(G) = n) = \prod_{j=0}^{n-1}(1 - 2^{-(l-k)})$. For $q < 1/2$, $\hat{q}_{n,i} > 0$ and $\mathbb{P}(G_i \in \mathcal{L}_k) > 2^{-(l-k)}$. Following the argument in Section II it is easy to derive $\pi_n(q) := \mathbb{P}(rank(G) = n) = \prod_{k=0}^{n-1}\big(1 - \mathbb{P}(G_k \in \mathcal{L}_k)\big)$. Hence, $\pi_n(q) < \pi_n(1/2)$, i.e., any choice of $q < 1/2$ yields a strictly lower probability of G being a full rank matrix than when $q = 1/2$.

We can immediately generalise this observation to any probability distribution for which

$$\hat{q}_{n,i}(u) \geq 0 \ \forall u \neq 0 \text{ and } \exists u \neq 0 : \hat{q}_{n,i}(u) > 0 \tag{11}$$

which we refer to as *Condition 1*. From the definition of the inner-product and (8) we have

$$\begin{aligned}\hat{q}_{n,i}(u) = &\sum_{s=0}^{\lfloor |1(u)|/2 \rfloor} \mathbb{P}(v : |1(u\otimes v)| = 2s) \\ &- 1_{\{|1(u)| \text{ is odd}\}} \sum_{s=0}^{\lfloor |1(u)|/2 \rfloor} \mathbb{P}(v : |1(u\otimes v)| = 2s+1) \\ &- 1_{\{|1(u)| \text{ is even}\}} \sum_{s=0}^{\lfloor |1(u)|/2 \rfloor - 1} \mathbb{P}(v : |1(u\otimes v)| = 2*s+1).\end{aligned}$$

For ease of exposition we write this as follows

$$\begin{aligned}\hat{q}_{n,i}(u) = &\sum_{s=0,2,4,..} \mathbb{P}(v : |1(u\otimes v)| = s) \\ &- \sum_{s=1,3,5,..} \mathbb{P}(v : |1(u\otimes v)| = s).\end{aligned}$$

Thus *Condition 1* holds if and only if

$$\begin{aligned}&\sum_{s=0,2,4,..} \mathbb{P}(v : |1(u\otimes v)| = s) \geq \\ &\sum_{s=1,3,5,..} \mathbb{P}(v : |1(u\otimes v)| = s) \ \forall\ u \neq 0 \text{ and} \\ &\exists\ u : \sum_{s=0,2,4,..} \mathbb{P}(v : |1(u\otimes v)| = s) > \\ &\sum_{s=1,3,5,..} \mathbb{P}(v : |1(u\otimes v)| = s).\end{aligned} \tag{12}$$

In particular, *Condition 1* (11) is satisfied by any distribution that generates $(0,1)$ matrices that have more 0s than 1s. That is, matrices where $q_l^k < 1/2$, where q_l^k is the probability that element l of column k is non-zero. To see this, observe that for a given u

$$\mathbb{P}(v : |1(u\otimes v)| = s) = \sum_{V\in V_s} \prod_{i\in V} q_i^k \prod_{j\in 1(u)\setminus V} (1-q_j^k),$$

where V_s is the set of all subsets of 1(u) with s elements. From this we get the following

$$\begin{aligned}\hat{q}_{n,i}(u) &= \sum_{s=0,2,4,..} \mathbb{P}(v : |1(u\otimes v)| = s) \\ &\quad - \sum_{s=1,3,5,..} \mathbb{P}(v : |1(u\otimes v)| = s) \\ &= \sum_{s=0,2,4,..} \sum_{V\in V_s} \prod_{i\in V} q_i^k \prod_{j\in 1(u)\setminus V} (1-q_j^k) \\ &\quad - \sum_{s=1,3,5,..} \sum_{V\in V_s} \prod_{i\in V} q_i^k \prod_{j\in 1(u)\setminus V} (1-q_j^k).\end{aligned}$$

Therefore we can simplify the expression to get

$$\begin{aligned}\hat{q}_{n,i}(u) &= \sum_{s=0,2,4,..} \sum_{V\in V_s} \prod_{i\in V} (-q_i^k) \prod_{j\in 1(u)\setminus V} (1-q_j^k) \\ &\quad + \sum_{s=1,3,5,..} \sum_{V\in V_s} \prod_{i\in V} (-q_i^k) \prod_{j\in 1(u)\setminus V} (1-q_j^k) \\ &= \sum_{s=0}^{|1(u)|} \sum_{V\in V_s} \prod_{i\in V} (-q_i^k) \prod_{j\in 1(u)\setminus V} (1-q_j^k) \\ &= \prod_{i\in 1(u)} (1-2q_i^k),\end{aligned}$$

from which the generalization follows.

We can summarize the results of this section in the following manner. For a given rateless code let τ be the random variable that denotes the first time that all the uncoded packets can be recovered from the received coded packets (received through an erasure channel). For a big class of random codes that includes the class of Bernoulli(q) random codes with $q \leq 1/2$, the results above show that choosing Bernoulli($1/2$) random codes is the optimal choice in terms of stochastically minimizing τ for any finite n. Our simulation results indicate that the above property is also true when $q \geq 1/2$ but only for large enough n. The stochastic or-

dering result is surprising since the results of Kovalenko [16], Masol [17] and others found in Levitskaya[18] only prove that if q is not too close to either 0 or 1, then the distribution of the rank converges for very large n in the variation norm to the case of $q = 1/2$.

B. Performance of systematic codes

Turning to the case of systematic codes we can make the following important observation when $q_l^k = q$ for all l and k. If $q < 1/2$ and $|1(u)| \neq 0$, we find that $\hat{q}_{n,i}(u) = (1-2q)^{|1(u)|}$ decreases with q. Therefore $\mathbb{P}(G_i \in \mathcal{L}_k)$ decreases with q and $\mathbb{P}(rank(G) = n)$ increases with q. As in the case of $q = 1/2$ we can redefine

$$F(n,\delta,\epsilon,q) := \sum_{i=0}^{n-1} \binom{n}{i} \epsilon^{n-i}(1-\epsilon)^i F_{i,n+\delta-i}(q), \quad (13)$$

where $F(n,\delta,\epsilon,q)$ is the probability of failure of decoding all uncoded packets after receiving $n+\delta$ packets at the receiver with an initial uncoded set of transmissions followed by randomly generated binary sequences of length n generated by *i.i.d.* Bernoulli(q) random variables. Above $F_{i,l}(q)$ is the probability of decoding failure starting rank i and having received l coded packets. Note that $F_{i,l}(q)$ is nothing but $1 - \pi_{n-i}(q)$. Therefore $F(n,\delta,\epsilon,q)$ decreases with $q \in [0,1/2]$. Therefore our construction of systematic Bernoulli($1/2$) random codes still demonstrate optimal performance in terms of stochastically minimizing τ. As mentioned in Section II the systematic part helps improve the performance of Bernoulli($1/2$) codes whereas it could hurt the performance when $q < 1/2$.

IV. Asymptotically Optimal Intermediate Decoding

We now analyze the asymptotic behaviour of our code construction and will show that it yields the best trade-off for intermediate decoding. As done previously assume that the block-size is n packets and the packet loss probability is ϵ, but assume that we receive a total of l packets at the decoder. Instead of recovering all the packets, we will analyse the rank of G once l packets are present at the decoder.

Denote the decoded rank random variable by R_l^n when n is the block-size and l the number of received packets. Then for $0 \leq i \leq \min(l,n)$ we have

$$\begin{aligned}\mathbb{P}(R_l^n \geq i) &= \sum_{j=i}^{n} \binom{n}{j}(1-\epsilon)^j \epsilon^{n-j} \\ &+ \sum_{s=0}^{i-1} \binom{n}{s} \epsilon^{n-s}(1-\epsilon)^s P_U(n-s, i-s, l-s),\end{aligned} \quad (14)$$

where $P_U(n,i,l)$ is the probability of recovering rank of at least i from l received coded packets (U in the definition refers to Bernoulli($1/2$) random codes) when the block-size is n. The first term in the equation accounts for the possibility that at least i packets are uncoded packets, and the second considers the (disjoint from first term) possibility that all the uncoded packets have been transmitted and at most $i-1$ uncoded packets arrive at the receiver. Obviously we have $P_U(n,0,l) \equiv 1$, $P_U(n,i,l) \equiv 0$ if $\min(n,l) < i$, and monotonicity of $P_U(n,i,l)$ in l, i.e., $P_U(n,i,l+1) \geq P_U(n,i,l)$.

By conditioning on whether the first column is an all zero column or not ,we can derive the following recursive relationship to compute these probabilities, namely,

$$\begin{aligned}P_U(n,i,l) &= 2^{-l} P_U(n-1,i,l) \\ &+ (1-2^{-l}) P_U(n-1,i-1,l-1).\end{aligned} \quad (15)$$

Since $i \leq n$, the rank of G can at most be i. Therefore one can view the problem such that i becomes the block-length and n the number of received coded packets. Thus taking a transpose of the matrix representation of the received packets and using the analysis from the Section II we have

$$P_U(n,i,i) = \prod_{j=0}^{i-1}(1-2^{j-n}) \quad \forall\, i = 1,2,\ldots,n. \quad (16)$$

From this discussion and Section III, it is also clear that for every finite n (and $i \leq n$) choosing Bernoulli(q) random codes with $q < 1/2$ will only result in a smaller value for this probability. Therefore the random variable R_l^n is stochastically maximized by the systematic equiprobable code. The analysis in Section II summarized in (1) also gives one more constraint, namely, if $l \leq n$, then $P_U(n,n,l) = \prod_{j=0}^{n-1}(1-2^{j-l})$.

Pick $r \in (0,1)$. Then we will show that $\lim_{n\to+\infty} P_U(n,\lfloor rn\rfloor,\lfloor rn\rfloor) = 1$. Using $0 \leq 1-x \leq e^{-x}$ for $x \in [0,1]$ we get $P_U(n,i,i) \leq \exp\left(-2^{i-n}+2^{-n}\right)$ and $P_U(n,n,l) \leq \exp\left(-2^{n-l}+2^{-l}\right)$. Using the inequality $\prod_{j=1}^{k}(1-x_j) \geq 1 - \sum_{j=1}^{k} x_j$ for $x_j \in [0,1]$ for all $j = 1,2,\ldots,k$ we have $P_U(n,i,i) \geq 1-2^{i-n}$ and $P_U(n,n,l) \geq 1-2^{n-l}$. Using these bounds one gets

$$\begin{aligned}1 - 2^{-n\left(1-\frac{\lfloor rn\rfloor}{n}\right)} &\leq P_U(n,\lfloor rn\rfloor,\lfloor rn\rfloor) \\ &\leq \exp\left(-2^{-n\left(1-\frac{\lfloor rn\rfloor}{n}\right)} + 2^{-n}\right).\end{aligned} \quad (17)$$

Now it is immediate that $\lim_{n\to+\infty} P_U(n,\lfloor rn\rfloor,\lfloor rn\rfloor) = 1$ for all $r \in (0,1)$.

For the case of $r = 0$ we only need to prove that $\lim_{n\to+\infty} P_U(n,\delta(n),\delta(n)) = 1$ for any sequence $\delta(n) \geq 1$ such that $\lim_{n\to+\infty} \delta(n)/n = 0$. This readily follows by using bounds similar to (17). Essentially this result follows from the fact that the probability of the all-zero codeword decays to 0 exponentially fast in n.

For $r = 1$ define $\delta = l - n$, and then we get

$$\begin{aligned}1 - 2^{-\delta} &\leq P_U(n,\lfloor rn\rfloor,\lfloor rn\rfloor+\delta) \\ &\leq \exp\left(-2^{-\delta}+2^{-n-\delta}\right).\end{aligned} \quad (18)$$

Therefore if $\lim_{n\to+\infty}\delta(n) = +\infty$, then $\lim_{n\to+\infty} P_U(n,n,n+\delta(n)) = 1$. Note that we do not need $\delta(n)$ to be $O(n)$ and it is fine if the sequence $\delta(n)$ is $o(n)$, as long as $\lim_{n\to+\infty}\delta(n) = +\infty$.

Now we present a simple bound that will help prove

optimality properties for our systematic coding scheme that hold for all $\epsilon \in [0,1)$ when n is large. For $0 \leq s \leq i-1$ where $i < n$ we have

$$\begin{aligned} & P_U(n-s, i-s, i-s) \\ &= \prod_{j=0}^{i-s-1} (1-2^{j-n+s}) \\ &= \prod_{j=s}^{i-1} (1-2^{j-n}) \\ &\geq \prod_{j=0}^{i-1} (1-2^{j-n}) = P_U(n,i,i). \end{aligned} \tag{19}$$

Note that the inequality above is strict if $s > 0$. Also, by definition $P_U(n,i,i) \leq 1$. Therefore we have

$$\begin{aligned} & \mathbb{P}(R_i^n = i) \geq P_U(n,i,i) \\ & \quad \times \left(\sum_{j=i}^{n} \binom{n}{j} (1-\epsilon)^j \epsilon^{n-j} + \sum_{s=0}^{i-1} \binom{n}{s} \epsilon^{n-s} (1-\epsilon)^s \right) \\ & \quad = P_U(n,i,i). \end{aligned} \tag{20}$$

From the earlier result that $\lim_{n\to+\infty} P_U(n, \lfloor rn \rfloor, \lfloor rn \rfloor) = 1$ for all $r \in (0,1)$ we have $\frac{R^n_{\lfloor rn \rfloor}}{n}$ converging in probability to r. In other words for large n with high probability (tending to 1 as $n \to +\infty$) any intermediate rank can be recovered from the reception of the same number of packets.

Consider the case of $i = n$ and $l = n + \delta$ where $\delta \geq 0$. Here for $0 \leq s \leq n-1$ we have

$$\begin{aligned} P_U(n-s, n-s, n+\delta-s) &= \prod_{j=0}^{n-s-1} (1-2^{j-n-\delta+s}) \\ &= \prod_{j=s}^{n-1} (1-2^{j-n-\delta}) \\ &\geq \prod_{j=0}^{n-1} (1-2^{j-n-\delta}), \end{aligned} \tag{21}$$

where the last term is $P_U(n,n,n+\delta)$. This then yields the following bound

$$\begin{aligned} & \mathbb{P}(R^n_{n+\delta} = n) = (1-\epsilon)^n \\ & \quad + \sum_{s=0}^{n-1} \binom{n}{s} (1-\epsilon)^s \epsilon^{n-s} P_U(n-s, n-s, l-s) \\ & \quad \geq P_U(n,n,n+\delta). \end{aligned} \tag{22}$$

Again bounds similar to (17) show that $\lim_{n\to+\infty} \mathbb{P}(R^n_{n+\delta(n)} = n) = 1$, as long as $\lim_{n\to+\infty} \delta(n) = +\infty$.

We have thus shown that the performance of our coding scheme, when only the coded packets are received, is asymptotically optimal. Since this lower bounds the performance of our systematic scheme (independent of the erasure probability), we have demonstrated that it is universally (in ϵ) optimal.

We once again point out that strictly speaking in our result r is the fractional rank of the $G(\cdot)$ random matrix which could be very different from the actual fraction of uncoded packets that can be successfully reconstructed.

Using results from [16], [17] and others in [18], even if the probability of subsets of $\{1,2,\ldots,n\}$ are not chosen uniformly, it is still possible to get the same asymptotic (in n) decoding performance if the subsets are chosen by generating n Bernoulli($q(n)$) random variables so long as $\frac{\log(n)+c}{n} \leq q(n) \leq 1 - \frac{\log(n)+c}{n}$. Thus even though, for every finite n, the codes with $q = 1/2$ perform better than when $q < 1/2$, asymptotically there no linear excess with properly chosen values of $q(n)$. There are a few properties that one can state for this class of codes: 1) asymptotically the all-ones or all-zeros vector has negligible probability; 2) the mean degree, i.e., $nq(n)$, is at least $\log(n) + c$ which tends to infinity as $n \to +\infty$ while it is at the most $n - \log(n) - c$ so that different sequences will have at least a few uncoded packets being disjoint; and 3) asymptotically any finite degree sequence has negligible probability. One should contrast this with the class of codes that have a finite asymptotic degree as in the analysis of [1], [9] and [8]. For this class one has $nq(n)$ asymptotically being a finite constant, except for the robust soliton codes where the mean degree grows without bound to the mean of the soliton distribution, i.e., infinity. Even with matrix inversion as opposed to just the belief-propagation algorithm, from [16], [17] one can predict that it would not be possible to obtain the performance of equiprobable codes in terms of asymptotically no linear excess. The one major drawback with using highly connected graphs is that matrix inversion will be very expensive if the block length is large. However, it is still unknown as to whether graphs generated with $q(n) \approx \frac{\log(n)+c}{n}$ will be sparse enough that matrix inversion will not be very expensive. Exploring this as well as the performance of belief-propagation for such codes is a topic for future research.

V. Conclusion

We analyzed a class of systematic rateless codes constructed using Bernoulli($1/2$) random variables. Over a large class of random codes we showed that the systematic equiprobable code construction has optimal performance in terms of the number of received packets necessary for recovery of coded packets. Thereafter we also showed that our code construction is also asymptotically optimal in terms of intermediate decoding.

References

[1] M. Luby, "LT codes," in *Proc. of the* 43rd *Annual IEEE Symposium on FOCS*, Nov 2002, pp. 271–280.

[2] D. MacKay, *Information theory, inference and learning algorithms*. New York: Cambridge University Press, 2003.

[3] ——, "Fountain codes," in *The IEE Proc.*, vol. 152. IEE, Dec 2005, pp. 1062–1068.

[4] M. Mitzenmacher, "Digital fountains: A survey and look forward," Notes, 2006.

[5] E. Hyytia, T. Tirronen, and J. Virtamo, "Optimal degree distribution for LT codes with small message length," in *Proc. of* 26^{th} *IEEE Infocom*, Anchorage, AK, May 2007, pp. 2576–2580.
[6] A. Shokrollahi, "Raptor codes," *IEEE Trans. Inform. Theory*, vol. 52, no. 6, pp. 2551–2567, 2006.
[7] P. Maymounkov, "Online codes," NYU, Tech. Rep. TR2002-833, 2002.
[8] S. Sanghavi, "Intermediate performance of rateless codes," in *Proc. of IEEE ITW*, Tahoe City, CA, Sept 2007, pp. 478–482.
[9] R. Darling and J. Norris, "Structure of large random hypergraphs," *Ann. Appl. Probab.*, vol. 15, no. 1A, pp. 125–152, 2005.
[10] E. Maneva and A. Shokrollahi, "New model for rigourous analysis fo LT-codes," in *Proc. of IEEE ISIT*, Seattle, WA, July 2006, pp. 2677–2679.
[11] C. Studholme and I. Blake, "Random matrices and codes for the erasure channel," *Algorithmica*, April 2008, published online.
[12] E. Neuman and J. Sándor, "On the Ky Fan inequality and related inequalities. II," *Bull. Austral. Math. Soc.*, vol. 72, no. 1, pp. 87–107, 2005.
[13] R. Lidl and H. Niederreiter, *Finite fields*, 2nd ed. Cambridge: Cambridge University Press, 1997, vol. 20.
[14] N. Sloane, "Weight enumerators of codes," in *Combinatorics, Part 1: Theory of designs, finite geometry and coding theory (Proc. Advanced Study Inst., Breukelen, 1974)*. Amsterdam: Math. Centrum, 1974, pp. 111–138. Math. Centre Tracts, No. 55.
[15] A. Alekseychuk, "Non-asymptotic bounds for probabilities of the rank of a random matrix over a finite field," *Discrete Math. Appl.*, vol. 17, no. 3, pp. 269–278, 2007.
[16] I. Kovalenko, "Invariance theorems for random Boolean matrices," *Kibernetika (Kiev)*, no. 5, pp. 138–152, 1975, English translation: Cybernetics 11 (1975), no. 5, 818–834 (1976).
[17] V. Masol, "On the probability of uniqueness of the solution of a system of linear random Boolean equations," *Vīsnik Kiīv. Unīv. Ser. Mat. Mekh.*, no. 30, pp. 58–62, 116, 1988.
[18] A. Levit·skaya, "Systems of random equations over finite algebraic structures," *Kibernet. Sistem. Anal.*, vol. 41, no. 1, pp. 82–116, 190, 2005, English translation: Cybernet. Systems Anal. 41 (2005), no. 1, 67–93.