Channels with both Random Errors and Burst Erasures: Capacities, LDPC Code Thresholds, and Code Performances

Kan Li
University of Hawaii
Honolulu, HI 96822 USA
Email: likan@hawaii.edu

Aleksandar Kavčić
University of Hawaii
Honolulu, HI 96822 USA
Email: kavcic@hawaii.edu

Raman Venkataramani
Seagate Technology
Colorado
Email: ramanv@ieee.org

M. Fatih Erden
Seagate Technology
Bloomington, MN 55435 USA
Email: Fatih.Erden@seagate.com

***Abstract*—We derive the capacities of a class of channels, either memoryless or indecomposable finite-state, that also suffer from bursts of erasures. For such channels, we analyze the performances of low-density parity-check (LDPC) codes and code ensembles under belief propagation (BP) decoding, using density evolution (DE) techniques. Although known LDPC codes perform well in non-erasure-affected channels, their performances are far from the capacities when both random errors and erasures are present. We show that enhancing the codes' erasure handling using published methods beats, in some instances, the BP thresholds. However, to achieve capacity, codes must be constructed to tackle both effects simultaneously.**

I. Introduction

Aside from the primary impairment by random noise, a secondary effect in the magnetic data storage channel is the occasional occurrence of intense transient noise in the read-back signal, when the read-back head collides with disk surface contaminants (e.g., dust particles) and heats up for a short period of time. This friction-induced thermal effect can be modeled, information-theoretically, as a burst of erasures.

In this paper we consider a channel that can be represented as a concatenation of two channels:

1) The first channel is either a memoryless channel or an indecomposable finite-state channel (IFSC), and
2) the second channel is a burst erasure channel (BuEC).

We call the resulting channel the IFSC-BuEC concatenation. We derive the capacity for such a concatenation to be the product of the capacities of each channel.

In the second part of the paper, we consider the performances of low-density parity-check (LDPC) codes for the IFSC-BuEC concatenated channel. LDPC codes were first introduced by Gallager [1] and have been shown either by code construction [2] or density evolution [3] to approach the capacities of many memoryless and finite-state channels [2]-[4]. In separate developments, several researchers [5]-[7] have shown how to construct LDPC codes that enhance their burst-erasure correcting capabilities under belief propagation (BP) decoding. However, we are not aware of any studies of LDPC codes conducted for an IFSC-BuEC concatenation, and we conduct such a study here.

First we develop the density evolution (DE) algorithm to compute noise tolerance thresholds for the IFSC-BuEC concatenation, and then we compare the performances of existing LDPC codes to the thresholds. We show that LDPC codes constructed specifically for the IFSC are not necessarily good for the IFSC-BuEC concatenation. Next, we demonstrate that the methods in [5]-[7] for enhancing LDPC codes over BuECs are also effective for the IFSC-BuEC concatenation, and that, in some instances, they result in codes that beat the DE thresholds. However, we demonstrate that despite this enhancement, the performances of existing methods are still far from the capacities of the IFSC-BuEC concatenations. This suggests that to achieve the capacities of the IFSC-BuEC concatenation, we must construct LDPC codes specifically for the IFSC-BuEC concatenation.

Organization: Section II defines the channels under consideration and presents their capacities with proofs. Section III introduces the density evolution algorithm and thresholds computation for LDPC codes over this class of channels. In Section IV, we compare the performances of two types of LDPC codes, designed with random error and erasure handling separately. Section V concludes the paper.

II. Channels and Capacities

First we define the channel under consideration. The channel $(X \to Y)$ is a concatenation of the channels $(X \to Z)$ and $(Z \to Y)$. The channel $(X \to Z)$ is assumed to be either a memoryless or an indecomposable finite-state machine channel. The channel $(Z \to Y)$ is considered to be a burst erasure channel with parameter q. The BuEC $(Z \to Y)$ behaves as follows: when n symbols $Z_1^n = [Z_1, \cdots, Z_n]$ are transmitted through the channel (here we can let $n \to \infty$), exactly $n \cdot q$ consecutive symbols $Z_i^{i+nq-1} = [Z_i, \cdots, Z_{i+nq-1}]$ are erased, where i is a variable denoting the start of the burst. The start position i could be deterministic or random. We shall prove that the capacity of the channel $(X \to Y)$, denoted by $C_{X\to Y}$, equals

$$C_{X\to Y} = (1-q)C_{X\to Z} \tag{1}$$

where $C_{X\to Z}$ is the capacity of the channel $(X \to Z)$.

We shall first consider n to be finite, and q to be a rational number $q = \frac{\ell}{n}$, and prove (1) for the case that $(X \to Z)$ is a memoryless channel. The proof for the general case $0 \leq q \leq 1$ and $(X \to Y)$ an indecomposable channel is a simple (though

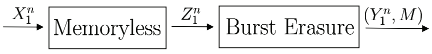

Fig. 1. Model of Memoryless Channel with Burst Erasure

tedious) modification of this basic proof, and we give its sketch in the Appendix.

A. Capacity of a Memoryless Channel with Burst Erasure

Let the channel $(X \to Z)$ be memoryless. The channel $(Z \to Y)$ is a burst erasure channel with input-output law

$$Y_i = \begin{cases} e & i \in [M, M+\ell) \\ Z_i & i \notin [M, M+\ell) \end{cases} \tag{2}$$

where $M \in \{1, 2, \cdots, n-\ell+1\}$ is a random variable, independent of X or Z, indicating the starting index of the burst erasure of length ℓ. Here, $\frac{\ell}{n} = q$. Fig. 1 depicts the channel $(X \to Y)$.

Theorem 1: The capacity for the channel described by Fig. 1 and the channel law (2) is

$$C_{X\to Y} = (1-q)C_{X\to Z}. \tag{3}$$

Proof: The proof proceeds as follows. First we will show that $\frac{1}{n}I(X_1^n;Y_1^n,M) \le (1-q)C_{X\to Z}$. Then we will show that the upper bound is attainable, i.e. that there exists a distribution $p(x_1^n) = p^*(x_1^n)$ that induces the information rate $\frac{1}{n}I(X_1^n;Y_1^n,M)|_{p(x_1^n)=p^*(x_1^n)} = (1-q)C_{X\to Z}$.

Let us first prove $\frac{1}{n}I(X_1^n;Y_1^n,M) \le (1-q)C_{X\to Z}$. Since M and X_1^n are independent, we have

$$\begin{aligned} I&\left(X_1^n;Y_1^n,M\right) \\ &= H(Y_1^n,M) - H(Y_1^n,M|X_1^n) \\ &= [H(Y_1^n|M)+H(M)] - [H(M|X_1^n)+H(Y_1^n|M,X_1^n)] \\ &= H(Y_1^n|M) - H(Y_1^n|X_1^n,M) \qquad (4) \\ &= I(X_1^n;Y_1^n|M). \qquad (5) \end{aligned}$$

The first term in (4) is upperbounded as follows:

$$\begin{aligned} H(Y_1^n|M) &\le \sum_{i=1}^{n} H(Y_i|M) \\ &= \sum_{i=1}^{n}\sum_{m=1}^{n-\ell+1} \Pr(M=m)\cdot H(Y_i|M=m). \end{aligned} \tag{6}$$

Notice

$$\begin{aligned} H(Y_i|M=m) &= \begin{cases} 0 & i \in [m, m+\ell) \\ H(Z_i) & i \notin [m, m+\ell) \end{cases} \qquad (7) \\ &= \begin{cases} 0 & m \in (i-\ell, i] \\ H(Z_i) & m \notin (i-\ell, i] \end{cases}. \qquad (8) \end{aligned}$$

Using (7) and (8), we further manipulate (6) as

$$\begin{aligned} H(Y_1^n|M) &\le \sum_{i=1}^{n} H(Y_i|M) = \sum_{m=1}^{n-\ell+1}\Pr(M=m)\sum_{i\notin[m,m+l)} H(Z_i) \\ &= \sum_{i=1}^{n} H(Z_i)\sum_{m\notin(i-\ell,i]}\Pr(M=m) = \sum_{i=1}^{n} b(i)\cdot H(Z_i) \qquad (9) \end{aligned}$$

where

$$b(i) = \sum_{m\notin(i-\ell,i]} \Pr(M=m) \tag{10}$$

and

$$\sum_{i=1}^{n} b(i) = n-\ell. \tag{11}$$

Similarly, the second term on the right-hand side of (4) can be expanded as follows:

$$\begin{aligned} H(Y_1^n|X_1^n,M) &= \sum_{i=1}^{n} H(Y_i|X_1^n,M,Y_1^{i-1}) \\ &\overset{(a)}{=} \sum_{i=1}^{n} H(Y_i|X_i,M) = \sum_{i=1}^{n} b(i)\cdot H(Z_i|X_i) \end{aligned} \tag{12}$$

where equality (a) follows from the memorylessness of the channel ($X \to Z$). Now, combine (4), (9), and (12) to get

$$\begin{aligned} \frac{1}{n}I(X_1^n;Y_1^n,M) &\le \frac{1}{n}\sum_{i=1}^{n} b(i)\left[H(Z_i)-H(Z_i|X_i)\right] \\ &= \frac{1}{n}\sum_{i=1}^{n} b(i)\cdot I(X_i;Z_i) \\ &\le \frac{1}{n}\max_{p(x_1^n)}\sum_{i=1}^{n} b(i)\cdot I(X_i;Z_i) \\ &\overset{(b)}{=} \frac{1}{n}\sum_{i=1}^{n} b(i)\cdot\left[\max_{p(x_i)} I(X_i;Z_i)\right] \\ &\overset{(c)}{=} \frac{1}{n}\sum_{i=1}^{n} b(i)\cdot\left[I(X_i;Z_i)|_{p(x_i)=p^*(x_i)}\right] \\ &= \frac{1}{n}\sum_{i=1}^{n} b(i)\cdot C_{X\to Z} \overset{(d)}{=} (1-\frac{\ell}{n})C_{X\to Z}. \end{aligned} \tag{13}$$

Here, equality (b) follows from the memorylessness of the channel ($X \to Z$), equality (c) uses $p^*(x_i) = \arg\max_{p(x_i)} I(X_i;Z_i)$, and (d) follows from (11).

Now, we will prove that there exists a distribution $p(x_1^n) = p^*(x_1^n)$ for which $\frac{1}{n}I(X_1^n;Y_1^n,M)$ meets (13) with equality. Let $p^*(x_1^n) = \prod_{i=1}^{n} p^*(x_i)$ be the distribution that achieves the capacity $C_{X\to Z}$, i.e., $p^*(\cdot)$ is the distribution of the independent and identically-distributed (i.i.d.) capacity-achieving input for the memoryless channel ($X \to Z$). Since $p^*(x_1^n) = \prod_{i=1}^{n} p^*(x_i)$, we have

$$\begin{aligned} p_{Y_1^n|M}(y_1^n|m) &= \sum_{x_1^n} p_{Y_1^n|X_1^n,M}(y_1^n|x_1^n,m)\cdot p_{X_1^n|M}(x_1^n|m) \\ &\overset{(e)}{=} \sum_{x_1^n} p_{Y_1^n|X_1^n,M}(y_1^n|x_1^n,m)\cdot p^*(x_1^n) \\ &\overset{(f)}{=} \sum_{x_1^n} p_{Y_1^n|X_1^n,M}(y_1^n|x_1^n,m)\cdot \prod_{i}^{n} p^*(x_i) \\ &\overset{(g)}{=} \sum_{x_1^n}\prod_{i=1}^{n} p_{Y_i|X_i,M}(y_i|x_i,m)\cdot p^*(x_i) \\ &= \prod_{i=1}^{n} p_{Y_i|M}(y_i|m) \end{aligned} \tag{14}$$

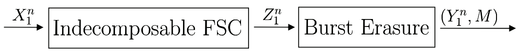

Fig. 2. Channel Model for Indecomposable FSC with Burst Erasure

where equality (e) follows from the fact that X_1^n and M are independent; equality (f) uses $p^*(x_1^n) = \prod_i^n p^*(x_i)$; and equality (g) follows from the memorylessness of the channel ($X \to Z$). From (14) we conclude that if $p(x_1^n) = \prod_{i=1}^n p^*(x_i)$, then $H(Y_1^n|M) = \sum_{i=1}^n H(Y_i|M)$. It follows that if $p(x_1^n) = \prod_{i=1}^n p^*(x_i)$, we get

$$\begin{aligned}\frac{1}{n} I(X_1^n; Y_1^n, M)|_{p(x_i^n)=p^*(x_i^n)} &= \frac{1}{n}\sum_{i=1}^n I(X_i; Y_i, M)|_{p(x_i)=p^*(x_i)} \\ &= \frac{1}{n}\sum_{i=1}^n b(i) I(X_i; Z_i)|_{p(x_i)=p^*(x_i)} \\ &= (1-q)C_{X\to Z}\end{aligned}$$

which ends the proof. ■

B. Capacity of a Channel with Memory and Burst Erasure

Here, we shall consider only discrete indecomposable finite-state channels as defined by Gallager [8]. The overall channel under consideration is a concatenation of an IFSC ($X \to Z$) and a burst erasure channel ($Z \to Y$) whose channel input-output law is

$$Y_i = \begin{cases} e & i \in [M, M+\ell) \\ Z_i & i \notin [M, M+\ell) \end{cases} \tag{15}$$

where $M \in \{1, 2, \cdots, n-\ell+1\}$ is a random variable, independent of X or Z, indicating the starting index of the burst erasure of length ℓ. The burst erasure channel is parametrized by the factor q, where exactly $q \cdot n$ symbols of the IFSC outputs are erased and $n \to \infty$. Fig. 2 depicts the channel ($X \to Y$).

Theorem 2: The capacity for the channel described by Fig. 2 and the channel law (15) is

$$C_{X\to Y} = (1-q)C_{X\to Z}. \tag{16}$$

A sketch of the proof is given in the Appendix.

III. LDPC Codes and Density Evolution Thresholds

The asymptotic belief propagation (BP) decoding performance for an ensemble of random LDPC codes with a given degree distribution can be computed using a technique called density evolution (DE) [9]. Specifically, the average asymptotic behavior for a given class of LDPC codes can be evaluated from the probability density functions (pdfs) of the messages in the iterative decoding process. Denote the decoding error probability in the ℓ-th iteration by $p_e^{(\ell)}$. In the limit as the code length $n \to \infty$ and decoding iteration $\ell \to \infty$, the "zero-error" noise standard deviation *threshold* σ^* is defined as

$$\sigma^* = \sup \sigma \tag{17}$$

where the supremum is taken over all noise standard deviations σ for which

$$\lim_{\ell\to\infty} p_e^{(\ell)} = 0. \tag{18}$$

For memoryless binary-input symmetric-output channels, numerical evaluation of the thresholds can be easily obtained from the discretized density evolution introduced in [10]. For IFSCs, specifically intersymbol interference (ISI) channels, thresholds can be computed using the method outlined in [4].

When random errors are combined with burst erasures in the channel, the density evolution algorithm remains unchanged except for the pdf of the channel output. It was shown in [11] for memoryless additive white Gaussian noise (AWGN) channels with erasures and in [12] for partial response channels with erasures that the pdf of the log-likelihood ratio (LLR) of the channel output $\mathcal{O}$ is

$$f'_{\mathcal{O}}(\xi) = (1-q)\cdot f_{\mathcal{O}}(\xi) + q\cdot\delta(\xi) \tag{19}$$

where $f_{\mathcal{O}}(\xi)$ is the output message pdf of the noise-only channel and $\delta(\cdot)$ is the Dirac delta function. If we utilize $f'_{\mathcal{O}}(\xi)$ as the density of the message from the channel to the decoder, we can conduct the DE algorithm for either memoryless channels or IFSCs with burst erasures. The threshold computation then follows [10] or [4], respectively.

IV. LDPC Code Performance

On one hand, the degree distribution of a given LDPC code determines its performance threshold in the presence of random errors. On the other hand, the burst erasure performance under BP decoding is characterized by the burst erasure capability or efficiency, as defined in [5]. However, it is not clear how a code performs if the channel has *both* random errors and erasure bursts. Here, we examine such a scenario. In particular, we want to examine 2 types of codes.

1) LDPC codes and code ensembles that perform well in memoryless or finite-state channels are well known [9],[4]. We want to examine how such codes perform when the channel also has burst erasures.
2) There exist several methods in the literature to construct LDPC codes that handle well bursts of erasures when no other channel impairment (such as additive noise) is present in the channel [5]-[7].

The simulation results presented in this section will demonstrate that applying methods to construct codes of type 2) can produce codes that beat the thresholds for codes of type 1), i.e., the methods in [5]-[7] can enhance the performances of LDPC codes over channels with both random errors and burst erasures. However, we will also see that while in some instances the BP thresholds are beaten, the performances of either codes of type 1) or those enhanced by methods in 2) are far from the capacities given by Theorem 1 and Theorem 2. Hence, we will conclude that in order to achieve the capacity of a channel that has both random errors and erasures, we must construct codes that tackle both effects *simultaneously*.

Here we generate several common codes of regular degrees and compare their bit-error rate (BER) performances with the capacities and threshold bounds, $\mathrm{SNR}_c(q)$ and $\mathrm{SNR}_{\sigma^*}(q)$

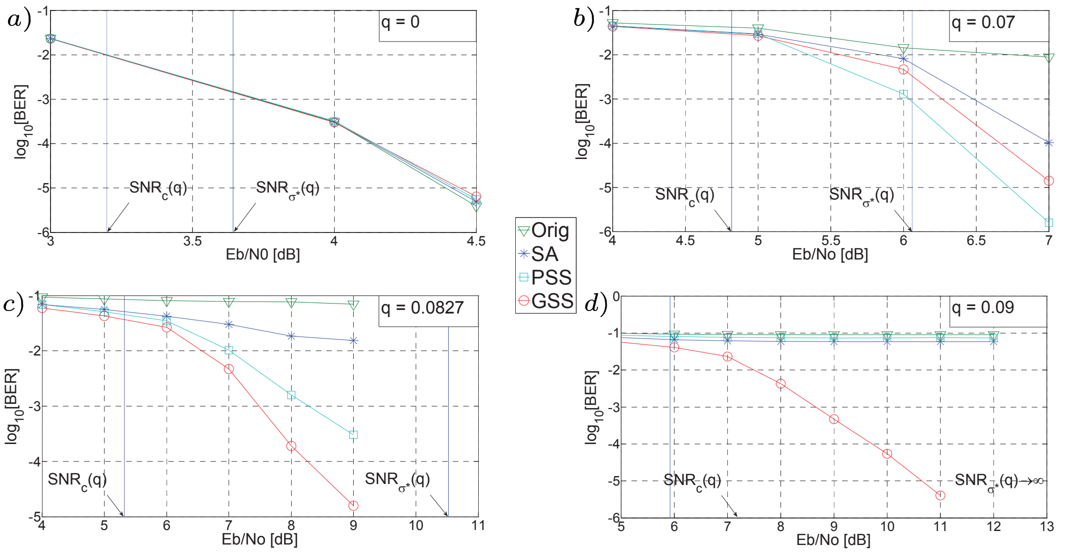

Fig. 3. Comparison of finite-length LDPC codes to capacity $\mathrm{SNR}_c(q)$ and BP threshold $\mathrm{SNR}_{\sigma^*}(q)$ for AWGN-BuEC concatenations with erasure ratio q equaling 0, 0.07, 0.0827, and 0.009.

respectively, for the AWGN channel and BuEC concatenation. The parity-check matrices used are for a rate $r = 0.9$, length $n = 4560$, (3, 30) regular quasi-cyclic (QC) LDPC code. Our goal is to manipulate the original finite-length QC code (denoted by Orig in Fig. 3) using simple column swaps or variable-node permutations (swaps) to achieve or approach the concatenated channel capacity or threshold bounds. The enhanced codes are constructed from the following permutation methods: the greedy search and swap (GSS) algorithm [5], the pivot searching and swapping (PSS) method in [6], and the simulated annealing (SA) approach in [7]. We plot the codes performances on the concatenated channels with increasing q values.

We see from Fig. 3a that for short bursts ($q \approx 0$), the BP thresholds are reasonably close to capacity. However, as the burst ratios q get larger (Fig. 3b-d), the thresholds diverge from capacity by an increasingly large margin. It is clear that existing LDPC codes are not adequate solutions for achieving capacity. Construction of special codes that simultaneously addresses erasure bursts and random errors is necessary. However, permuted suboptimal codes do perform better than non-permuted ones. In some scenarios, judiciously permuted finite-length LDPC codes beat the random LDPC code thresholds. For example, in Fig. 3c,d, as q rises, the code produced by GSS [5] consistently beats the corresponding threshold values $\mathrm{SNR}_{\sigma^*}(q)$. Since for a moderate length code with regular (i.e. suboptimal) degree distribution, the capacity $\mathrm{SNR}_c(q)$ is not within reach using simple permutations, to achieve capacity, a combination of two techniques is needed. First, degree distribution optimization to guide finite-length code construction. Second, judicious permutation of a finite length code to enhance performance.

V. Conclusion

We presented and proved the capacities of a class of channels, either memoryless or indecomposable finite-state, with both random errors and burst erasures. Enhancing the erasure handling of LDPC codes over these channels using published methods beats, in some cases, the BP thresholds. However, to achieve capacity, codes must be constructed to tackle both effects simultaneously.

Appendix
Sketch of the Proof of Theorem 2

A sketch of the proof proceeds as follows. We consider using the channel in blocks of ρ symbols. Let us introduce the following notation:

ρ	–	block size
n'	–	number of blocks
M	–	index of the first erased symbol

where $n = n' \cdot \rho$.

Let the i-th block be defined as

$$\underline{X}_i \triangleq \left[X_{(i-1)\rho+1}, X_{(i-1)\rho+2}, \cdots, X_{i\rho}\right] = X_{(i-1)\rho+1}^{i\rho}.$$

We denote contiguous blocks by $\underline{X}_i^j = [\underline{X}_i, \underline{X}_{i+1}, ..., \underline{X}_j]$.

We emulate the proof used for Theorem 1 (the memoryless channel $X \rightarrow Z$). First we show that the IFSC ($X \rightarrow Z$) behaves asymptotically as a block-memoryless channel and as a consequence the capacity of the channel ($X \rightarrow Y$) is upper bounded by $(1-q)C_{X\rightarrow Z}$. Then we will show that there exists a block distribution $p(\underline{x}_1^{n'}) = p^*(\underline{x}_1^{n'})$ that induces the information rate $\lim_{n\rightarrow\infty} \frac{1}{n} I(X_1^n; Y_1^n, M)|_{p(\underline{x}_1^n)=p^*(\underline{x}_1^n)} = (1-q)C_{X\rightarrow Z}$.

Let $\underline{e} \triangleq \underbrace{[e, \cdots, e]}_{\rho}$ be a block of erasures. We define a new block-channel whose channel input-output law is

$$\underline{Y}_i = \begin{cases} \underline{e} & \text{if any symbol in } \underline{Z}_i \text{ equals } e \\ Z_{(i-1)\rho+1}^{i\rho} & \text{otherwise.} \end{cases} \tag{20}$$

Let ℓ', where $\left\lceil \frac{\ell}{\rho} \right\rceil \leq \ell' \leq \left\lceil \frac{\ell}{\rho} \right\rceil + 1$, be the number of erased blocks and let M' be the index of the first erased block, $M' = \min\{i | \underline{Y}_i = \underline{e}\}$. Clearly, the number of erased symbols in the sequence $\underline{Y}_1^{n'} = [\underline{Y}_1, \underline{Y}_2, \cdots, \underline{Y}_{n'}]$ is $\ell' \cdot \rho$, while the number of erased symbols in the sequence $Y_1^n = [Y_1, Y_2, \cdots, Y_n]$ is $\ell \leq \ell' \cdot \rho$. Notice, however, that $\lim_{\substack{n' \to \infty \\ \rho \to \infty}} \frac{\ell'}{n'} = \lim_{n \to \infty} \frac{\ell}{n} = q$. Denote by $n_e' = \ell' \cdot \rho - \ell$ the extra erased symbols introduced by (20). The number n_e' cannot be higher than 2ρ. So, the ratio of number of extra erased symbols n_e' to the total number of symbols n satisfies

$$\frac{n_e'}{n} = \frac{\ell' \cdot \rho - \ell}{n' \cdot \rho} < \frac{2\rho}{n'\rho} = \frac{2}{n'}. \tag{21}$$

When both $n' \to \infty$ and $\rho \to \infty$, the fraction of extra erased symbols goes to zero, i.e. $\lim_{\substack{n' \to \infty \\ \rho \to \infty}} \frac{n_e'}{n} \to 0$.

Since the fraction of erased symobls goes to 0 as $n' \to \infty$,

$$\begin{aligned}\lim_{\substack{n' \to \infty \\ \rho \to \infty}} \frac{1}{n} I(X_1^n; Y_1^n, M | S_0) &= \lim_{n' \to \infty} \left[\lim_{\rho \to \infty} \frac{1}{n'\rho} I(\underline{X}_1^{n'}; \underline{Y}_1^{n'}, M' | S_0) \right] \\ &= \lim_{n' \to \infty} \frac{1}{n'} \left[\lim_{\rho \to \infty} \frac{1}{\rho} I(\underline{X}_1^{n'}; \underline{Y}_1^{n'}, M') \right] \end{aligned} \tag{22}$$

where S_0 is the state of the IFSC $(X \to Z)$ at time 0. Note that (22) is valid because the channel $(X \to Z)$ is indecomposable so that the information rate does not depend on the initial state.

Let

$$b'(i) = \sum_{\substack{m=1 \\ m \notin (i-\ell', i]}}^{n'} Pr(M' = m) \tag{23}$$

and $\sum_{i=1}^{n'} b'(i) = (n' - \ell')$. Next, we manipulate (22).

$$\begin{aligned}
&\lim_{\substack{n' \to \infty \\ \rho \to \infty}} \frac{1}{n'\rho} I(\underline{X}_1^{n'}; \underline{Y}_1^{n'}, M') \\
&\stackrel{(i)}{=} \lim_{\substack{n' \to \infty \\ \rho \to \infty}} \frac{1}{n'\rho} \Big\{ \left[H(\underline{Y}_1^{n'} | M') + H(M') \right] \\
&\qquad - \left[H(\underline{Y}_1^{n'} | \underline{X}_1^{n'}, M') + H(M' | \underline{X}_1^{n'}) \right] \Big\} \\
&\stackrel{(ii)}{=} \lim_{\substack{n' \to \infty \\ \rho \to \infty}} \frac{1}{n'\rho} \left[H(\underline{Y}_1^{n'} | M') - H(\underline{Y}_1^{n'} | \underline{X}_1^{n'}, M') \right] \\
&\stackrel{(iii)}{=} \lim_{\substack{n' \to \infty \\ \rho \to \infty}} \frac{1}{n'\rho} \left\{ \sum_{i=1}^{n'} \left[H(\underline{Y}_i | \underline{Y}_1^{i-1}, M') - H(\underline{Y}_i | \underline{X}_1^{n'}, \underline{Y}_1^{i-1}, M') \right] \right\} \\
&\stackrel{(iv)}{=} \lim_{\substack{n' \to \infty \\ \rho \to \infty}} \frac{1}{n'\rho} \left\{ \sum_{i=1}^{n'} \left[H(\underline{Y}_i | \underline{Y}_1^{i-1}, M') - H(\underline{Y}_i | \underline{X}_1^{i}, \underline{Y}_1^{i-1}, M') \right] \right\} \\
&\stackrel{(v)}{\leq} \lim_{\substack{n' \to \infty \\ \rho \to \infty}} \frac{1}{n'\rho} \left\{ \sum_{i=1}^{n'} \left[H(\underline{Y}_i | M') - H(\underline{Y}_i | \underline{X}_1^{i}, \underline{Y}_1^{i-1}, M', S_{(i-1)\rho}) \right] \right\} \\
&\stackrel{(vi)}{=} \lim_{\substack{n' \to \infty \\ \rho \to \infty}} \frac{1}{n'\rho} \left\{ \sum_{i=1}^{n'} \left[H(\underline{Y}_i | M') - H(\underline{Y}_i | \underline{X}_i, M', S_{(i-1)\rho}) \right] \right\} \\
&\stackrel{(vii)}{=} \lim_{n' \to \infty} \frac{1}{n'} \sum_{i=1}^{n'} \left[\lim_{\rho \to \infty} \frac{1}{\rho} I(\underline{X}_i; \underline{Y}_i, M') \right] \\
&\stackrel{(viii)}{=} \lim_{n' \to \infty} \frac{1}{n'} \sum_{i=1}^{n'} b'(i) \left[\lim_{\rho \to \infty} \frac{1}{\rho} I(\underline{X}_i; \underline{Z}_i) \right] \\
&\stackrel{(ix)}{\leq} \lim_{n' \to \infty} \frac{1}{n'} \sum_{i=1}^{n'} b'(i) C_{X \to Z} \\
&\stackrel{(x)}{=} C_{X \to Z} \lim_{n' \to \infty} \frac{n' - \ell'}{n'} = (1-q) C_{X \to Z}.
\end{aligned}$$

Equality (i) follows from the definition of mutual information and the chain rule for entropy; (ii) follows from the fact that M' is independent of $\underline{X}_1^{n'}$; (iii) follows from the chain rule for entropy; (iv) is valid because the channel is used without feedback; (v) is valid because conditioning does not increase entropy; (vi) follows because the channel is finite-state (knowledge of state $S_{(i-1)\rho}$ decouples the past $\underline{X}_1^{i-1}, \underline{Y}_1^{i-1}$); (vii) holds because the channel is indecomposable and $\rho \to \infty$; $(viii)$ follows from arguments similar to those in the proof of Theorem 1; (ix) uses the fact that $C_{X \to Z} = \max \lim_{\rho \to \infty} \frac{1}{\rho} I(\underline{X}_i; \underline{Z}_i)$; and (x) uses the definition for $b'(i)$. Now, we can finally combine

$$\lim_{n' \to \infty} \frac{1}{n'} \left[\lim_{\rho \to \infty} \frac{1}{\rho} I(\underline{X}_1^{n'}; \underline{Y}_1^{n'}, M') \right] \leq \lim_{n' \to \infty} \left(\frac{n' - \ell'}{n'} \right) \cdot C_{X \to Z}$$

i.e.

$$\lim_{n \to \infty} I(X_1^n; Y_1^n, M) \leq (1-q) \cdot C_{X \to Z}. \tag{24}$$

Equality in (24) can be achieved by choosing $p(\underline{x}_i^{n'}) = \prod_{i=1}^{n'} p^*(\underline{x}_i)$, where $p^*(\underline{x}_i)$ is the capacity achieving distribution for the channel $(X \to Z)$. ■

REFERENCES

[1] R. G. Gallager, "Low-density parity-check codes," *IRE Trans. Inform. Theory*, vol. IT-8, no. 1, pp. 21-28, Jan. 1962.

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

[3] T. Richardson, M. A. Shokrollahi, and R. Urbanke, "Design of capacity-approaching irregular low-density parity-check codes," *IEEE Trans. Inform. Theory*, vol. 47, pp. 619-637, Feb. 2001.

[4] A. Kavčić, X. Ma, and M. Mitzenmacher, "Binary intersymbol interference channels: Gallager codes, density evolution, and code performance bounds," *IEEE Trans. Inform. Theory*, vol. 49, no. 7, pp. 1636-1652, July 2003.

[5] K. Li, A. Kavčić, and M. F. Erden, "Construction of burst-erasure efficient LDPC codes for use with belief propagation decoding," *2010 IEEE Int. Conf. Commun.*, accepted.

[6] E. Paolini and M. Chiani, "Construction of near-optimum burst erasure correcting low-density parity-check codes," *IEEE Trans. Commun.*, vol. 57, no. 5, pp. 1320-1328, May 2009.

[7] G. Sridharan, A. Kumarasubramanian, A. Thangaraj, and S. Bhashyam "Optimizing burst erasure correction of LDPC codes by interleaving," *IEEE Int. Symp. Inf. Theory, Toronto, Canada*, pp. 1143-1147, July 2008.

[8] R. G. Gallager, *Information Theory and Reliable Communication.* New York: John Wiley and Sons, Inc., 1968.

[9] T. Richardson and R. Urbanke, "The capacity of low-density parity check codes under message-passing decoding," *IEEE Trans. Inform. Theory*, vol. 47, pp. 599-618, Feb. 2001.

[10] S.-Y. Chung, "On the construction of some capacity-approaching coding schemes," Ph.D. dissertation, MIT, Cambridge, MA, Sept. 2000.

[11] J. Ha and S. W. McLaughlin, "Low density parity check codes over Gaussian channels with erasures," in *IEEE Trans. Inform. Theory*, vol. 49, no. 7, pp. 1801 - 1809, July 2003.

[12] W. Tan and J. R. Cruz' "Analyzing LDPC codes on partial response channels with erasures using density evolution," *IEEE Trans. Magn.*, vol. 40, no. 5, pp 3411-3418, Sept. 2004.