

Fountain Communication using Concatenated Codes

Zheng Wang, *Student Member, IEEE*, Jie Luo, *Member, IEEE*

Abstract

This paper considers fountain communication over discrete-time memoryless channels. We extend concatenated coding schemes to fountain systems and derive the achievable fountain error exponents for one-level and multi-level concatenated fountain codes. Encoding and decoding complexities of the concatenated fountain codes are linear in the number of transmitted symbols and the number of received symbols, respectively. Performances of concatenated fountain codes in rate compatible fountain communication and fountain communication over an unknown channel are discussed.

Index Terms

coding complexity, concatenated codes, error exponent, fountain communication

I. INTRODUCTION

Fountain communication [1] is a new classical communication model originally proposed for reliable data transmission over erasure channels. In a point-to-point fountain communication system, the transmitter maps a message into an infinite sequence of channel symbols and sends them to the receiver. The receiver decodes the message after the number of received symbols exceeds certain threshold. Due to random symbol erasures, communication duration in a fountain system is determined by the receiver, rather than by the transmitter. The first realization of fountain codes was LT codes introduced by Luby [2] for erasure channels. LT codes can recover k information symbols from $k(1 + \epsilon)$ encoded symbols at high probability with a complexity of $O(k^{5/4}\epsilon^{1/2})$, for any $\epsilon > 0$ [2]. Shokrollahi proposed Raptor codes [3] by combining appropriate LT codes with a pre-code. Raptor codes can recover k information symbols from $k(1 + \epsilon)$ encoded symbols at high probability with complexity $O(k \log(1/\epsilon))$. For erasure channels, both LT codes and Raptor codes can achieve optimum rate irrespective of the erasure statistics. Generalization of Raptor codes from erasure channels to binary symmetric channels (BSCs) was studied by Etesami and Shokrollahi in [4]. In [5], Shamai, Telatar and Verdú systematically extended fountain communication to arbitrary channels and showed that fountain capacity [5] and Shannon capacity take the same value

The authors are with the Electrical and Computer Engineering Department, Colorado State University, Fort Collins, CO 80523. E-mail: {zhwang, rocky}@engr.colostate.edu.

This work was supported by the National Science Foundation under Grant CCF-0728826.

for stationary memoryless channels. Achievability of fountain capacity was demonstrated in [5] using a random coding scheme whose coding complexity is exponential in the number of received symbols. This consequently motivated the question whether fountain capacity of a stationary memoryless channel is achievable with a linear coding complexity.

In classical point-to-point communication over discrete-time memoryless channels, Feinstein [6] demonstrated that communication error probability can be made to decrease exponentially in the codeword length. The corresponding exponent is known as the error exponent. Tight lower and upper bounds on error exponent were obtained by Gallager [7], and by Shannon, Gallager, Berlekamp [8], respectively. In [9], Forney proposed a one-level concatenated coding scheme that combines a Hamming-sense error correction outer code with Shannon-sense random inner channel codes. One-level concatenated codes can achieve a positive error exponent, known as Forney's exponent, for any rate less than Shannon capacity with a polynomial complexity [9]. Forney's concatenated codes were generalized by Blokh and Zyablov [10] to multi-level concatenated codes, whose maximum achievable error exponent is known as the Blokh-Zyablov error exponent. In [11], Guruswami and Indyk introduced a class of linear complexity near maximum distance separable (MDS) error-correction codes. By using Guruswami-Indyk's codes as outer codes in concatenated coding schemes, achievability of Forney's and Blokh-Zyablov exponents with linear coding complexity was proved in [12].

In this paper, we extend concatenated coding schemes to fountain communication over discrete-time memoryless channels, as modeled in Section II. Random fountain codes are briefly introduced in Section III. By defining error probability scaling law with respect to the number of received symbols, we derive in Section IV the error exponents achievable by one-level and multi-level concatenated fountain codes, and show that their encoding and decoding complexities are linear in the number of transmitted symbols and the number of received symbols, respectively. In Section V, we consider rate compatible fountain communication where part of the source message is known at the receiver. With the transmitter still encoding the complete message, we show that concatenated fountain codes can achieve the same rate and error performance as if only the unknown part of the message is encoded. We briefly discuss fountain communication over an unknown channel in Section VI.

All logarithms in this paper are natural based.

II. THE FOUNTAIN COMMUNICATION MODEL

Consider the fountain communication system illustrated in Figure 1. Assume the encoder uses a fountain coding scheme [5] with W codewords to map the source message $w \in \{1, 2, \dots, W\}$ to an infinite channel input symbol sequence $\{x_{w1}, x_{w2}, \dots\}$. Assume the channel is discrete-time memoryless, characterized by the conditional point mess function (PMF) or probability density function (PDF) $p_{Y|X}(y|x)$, where

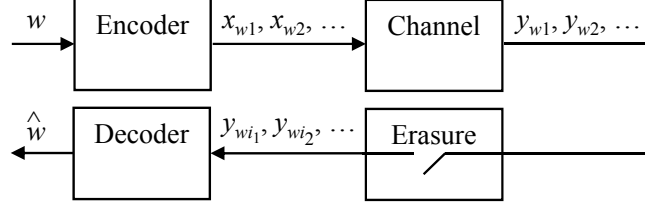


Fig. 1. Fountain communication over a memoryless channel.

$x \in X$ and $y \in Y$ are the input and output symbols, X and Y are the channel input and output alphabets, respectively. Define schedule $\mathcal{N} = \{i_1, i_2, \dots, i_{|\mathcal{N}|}\}$ as a subset of positive integers, where $|\mathcal{N}|$ is the cardinality of \mathcal{N} [5]. Assume the erasure device generates an arbitrary schedule \mathcal{N} , whose elements are indices of the received symbols $\{y_{wi_1}, y_{wi_2}, \dots, y_{wi_N}\}$, where $N = |\mathcal{N}|$. We say fountain rate of the system is $R = (\log W)/N$, if the decoder outputs an estimate \hat{w} of the source message after observing N channel symbols, based on $\{y_{wi_1}, y_{wi_2}, \dots, y_{wi_N}\}$ and \mathcal{N} . Decoding error happens when $\hat{w} \neq w$. Define error probability $P_e(N)$ as in [5],

$$P_e(N) = \sup_{\mathcal{N}, |\mathcal{N}| \geq N} \Pr\{\hat{w} \neq w | \mathcal{N}\}. \quad (1)$$

We say a fountain rate R is *achievable* if there exists a fountain coding scheme with $\lim_{N \rightarrow \infty} P_e(N) = 0$ at rate R [5]. The exponent rate at which error probability vanishes is defined as the fountain error exponent, $E_F(R)$,

$$E_F(R) = \lim_{N \rightarrow \infty} -\frac{1}{N} \log P_e(N). \quad (2)$$

Define fountain capacity \mathcal{C}_F as the supremum of all achievable fountain rates. It was shown in [5] that \mathcal{C}_F equals the Shannon capacity of a stationary memoryless channel.

III. RANDOM FOUNTAIN CODES

In a random fountain coding scheme [5], encoder and decoder share a fountain code library $\mathcal{L} = \{C_\theta : \theta \in \Theta\}$, which is a collection of fountain code books C_θ with θ being the index. All code books in the library have the same number of codewords and each codeword has an infinite number of channel input symbols. Let $C_\theta(m)_j$ be the j^{th} codeword symbol of message m in C_θ . To encode the message, the encoder first generates θ according to a distribution ϑ , such that the random variables $x_{m,j} : \theta \rightarrow C_\theta(m)_j$ are i.i.d. with a pre-determined input distribution p_X [5]. Then the encoder uses codebook C_θ to map the message into a codeword. We assume the actual realization of θ is known to the decoder but is unknown to the erasure device. Maximum likelihood decoding is assumed.

Theorem 1: Consider fountain communication over a discrete-time memoryless channel $p_{Y|X}$. Let \mathcal{C}_F be the fountain capacity. For any fountain rate $R < \mathcal{C}_F$, random fountain codes achieve the following

random-coding fountain error exponent, $E_{Fr}(R)$.

$$E_{Fr}(R) = \max_{p_X} E_{FL}(R, p_X), \quad (3)$$

where $E_{FL}(R, p_X)$ is defined as follows

$$E_{FL}(R, p_X) = \max_{0 \leq \rho \leq 1} \{-\rho R + E_0(\rho, p_X)\},$$

$$E_0(\rho, p_X) = -\log \sum_y \left(\sum_x p_X(x) p_{Y|X}(y|x)^{\frac{1}{1+\rho}} \right)^{(1+\rho)}. \quad (4)$$

If the channel is continuous, then summations in (4) should be replaced by integrals. ■

Theorem 1 was claimed implicitly in, and can be shown by, the proof of [5, Theorem 2].

$E_{Fr}(R)$ given in (3) equals the random-coding exponent of a classical communication system over the same channel [7]. For binary symmetric channels (BSCs), since random linear codes simultaneously achieves the random-coding exponent at high rates and the expurgated exponent at low rates [13], it can be easily shown that the same fountain error exponent is achievable by random linear fountain codes. However, because it is not clear whether there exists an expurgation operation, such as the one proposed in [7], that is robust to the observation of any subset of channel outputs, whether expurgated exponent is achievable for fountain communication over a general discrete-time memoryless channel is therefore unknown.

IV. CONCATENATED FOUNTAIN CODES

Consider a one-level concatenated fountain coding scheme illustrated in Figure 2. Assume source

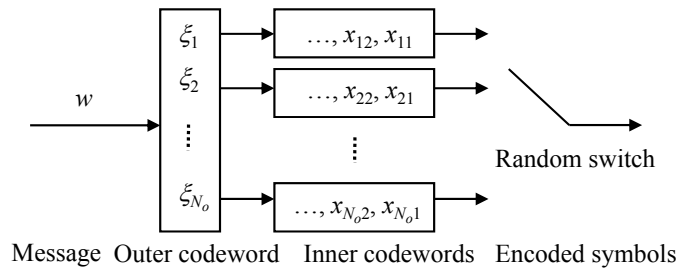


Fig. 2. One-level concatenated fountain codes.

message w can take $\exp(NR)$ possible values with an equal probability, where R is the targeted fountain information rate, and decoder decodes the source message after receiving N channel symbols. The encoder first encodes the message using an outer code into an outer codeword, $\{\xi_1, \xi_2, \dots, \xi_{N_o}\}$, with N_o outer symbols. We assume the outer code is a linear-time encodable/decodable near MDS error-correction code of rate $r_o \in [0, 1]$. That is, the outer code can recover the source message from a codeword with d symbol

erasures and t symbol errors, so long as $2t + d \leq (1 - r_o - \zeta_0)N_o$, where $\zeta_0 > 0$ is a positive constant that can be made arbitrarily small. An example of such linear complexity error-correction code was presented by Guruswami and Indyk in [11]. Each outer symbol ξ_k can take $\exp\left(\frac{N}{N_o} \frac{R}{r_o}\right)$ possible values. Define $N_i = \frac{N}{N_o}$, $R_i = \frac{R}{r_o}$. The encoder then uses a set of random fountain codes each with $\exp(N_i R_i)$ codewords to map each outer symbol ξ_k into an inner codeword, which is an infinite sequence of channel input symbols $\{x_{k1}, x_{k2}, \dots\}$. Let $C_\theta^{(k)}(\xi_k)_j$ be the j^{th} codeword symbol of the k^{th} inner code in codebook $C_\theta^{(k)}$, where θ is the codebook index as introduced in Section II. We assume θ is generated according to a distribution ϑ such that random variables $x_{k,\xi,j} : \theta \rightarrow C_\theta^{(k)}(\xi_k)_j$ are i.i.d. with a pre-determined input distribution p_X . To simplify the notations, we have assumed N_i , N_o , NR , and $N_i R_i$ should all be integers. We also assume $N_o \gg N_i \gg 1$.

After encoding, the inner codewords are regarded as N_o channel symbol queues, as illustrated in Figure 2. In the l^{th} time unit, the encoder uses a random switch to pick one inner code with index $k_l(\theta)$ uniformly, and sends the first channel input symbol in the corresponding queue through the channel. The transmitted symbol is then removed from the queue. We assume random variables $k_l : \theta \rightarrow \{1, 2, \dots, N_o\}$ are i.i.d. uniform. We assume the decoder knows the outer codebook and the code libraries of the inner codes. We also assume the encoder and the decoder share the realization of θ such that the decoder knows the exact codebook used in each inner code and the exact order in which channel input symbols are transmitted.

Decoding starts after $N = N_o N_i$ channel output symbols are received. The decoder first distributes the received symbols to the corresponding inner codewords. Assume $z_k N_i$ channel output symbols are received from the k^{th} inner codeword, where $z_k > 0$ and $z_k N_i$ is an integer. We term z_k the normalized effective codeword length of the k^{th} inner code. Based on z_k , and the received channel output symbols, $\{y_{ki_1}, y_{ki_2}, \dots, y_{ki_{z_k N_i}}\}$, the decoder computes the maximum likelihood estimate $\hat{\xi}_k$ of the outer symbol ξ_k together with an optimized reliability weight $\alpha_k \in [0, 1]$. We assume, given z_k and $\{y_{ki_1}, y_{ki_2}, \dots, y_{ki_{z_k N_i}}\}$, reliability weight α_k is computed using Forney's algorithm presented in [9, Section 4.2]. After that, the decoder carries out a generalized minimum distance (GMD) decoding of the outer code and outputs an estimate \hat{w} of the source message. GMD decoding of the outer code here is the same as that in a classical communication system, the detail of which can be found in [12].

Compared to a classical communication system where all inner codes have the same length, in a concatenated fountain coding scheme, numbers of received symbols from different inner codes may be different. Consequently, error exponent achievable by one-level concatenated fountain codes is less than Forney's exponent.

Theorem 2: Consider fountain communication over a discrete-time memoryless channel $p_{Y|X}$ with fountain capacity \mathcal{C}_F . For any fountain rate $R < \mathcal{C}_F$, the following fountain error exponent can be

arbitrarily approached by one-level concatenated fountain codes.

$$E_{Fc}(R) = \max_{p_X, \frac{R}{C_F} \leq r_o \leq 1, 0 \leq \rho \leq 1} (1 - r_o) \left(-\rho \frac{R}{r_o} + E_0(\rho, p_X) \left[1 - \frac{1 + r_o}{2} E_0(\rho, p_X) \right] \right). \quad (5)$$

where $E_0(\rho, p_X)$ is defined in (4).

Encoding and decoding complexities of the one-level concatenated codes are linear in the number of transmitted symbols and the number of received symbols, respectively. ■

The proof of Theorem 2 is given in Appendix A.

Corollary 1: $E_{Fc}(R)$ is upper-bounded by Forney's error exponent $E_c(R)$ given in [9]. $E_{Fc}(R)$ is lower bounded by $\tilde{E}_{Fc}(R)$, defined by

$$\tilde{E}_{Fc}(R) = \max_{p_X, \frac{R}{C_F} \leq r_o \leq 1, 0 \leq \rho \leq 1} (1 - r_o) \left(-\rho \frac{R}{r_o} + E_0(\rho, p_X) [1 - E_0(\rho, p_X)] \right). \quad (6)$$

The lower bounds is asymptotically tight in the sense that

$$\lim_{R \rightarrow C_F} \frac{\tilde{E}_{Fc}(R)}{E_{Fc}(R)} = 1. \quad (7)$$

■

The proof of Corollary 1 is given in Appendix B.

In Figure 3, we illustrate $E_{Fc}(R)$, $E_c(R)$, and $\tilde{E}_{Fc}(R)$ for a BSC with crossover probability 0.1. We

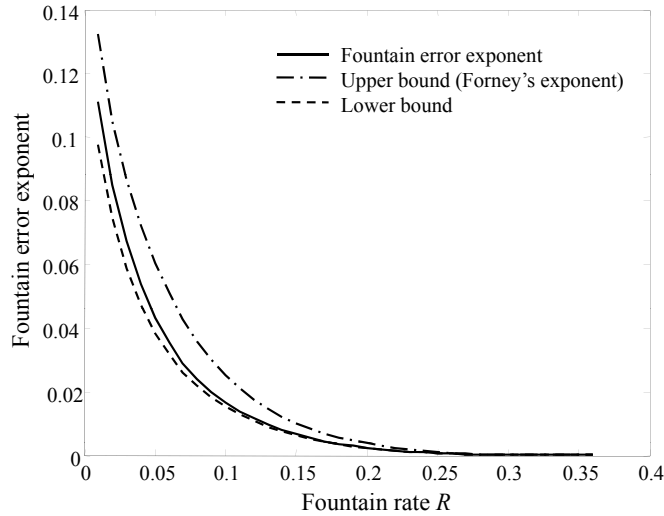


Fig. 3. Comparison of fountain error exponent $E_{Fc}(R)$, its upper bound $E_c(R)$, and its lower bound $\tilde{E}_{Fc}(R)$.

can see that $E_{Fc}(R)$ is closely approximated by $\tilde{E}_{Fc}(R)$.

Extending the one-level concatenated fountain codes to the multi-level concatenated fountain codes is essentially the same as in classical communication systems [10][12] except random fountain codes are used as inner codes in a fountain system. Achievable error exponent of an m -level concatenated fountain codes is given in the following Theorem.

Theorem 3: Consider fountain communication over a discrete-time memoryless channel $p_{Y|X}$ with fountain capacity \mathcal{C}_F . For any fountain rate $R < \mathcal{C}_F$, the following fountain error exponent can be arbitrarily approached by an m -level concatenated fountain codes,

$$E_{Fc}^{(m)}(R) = \max_{p_X, \frac{R}{\mathcal{C}_F} \leq r_o \leq 1} \frac{\frac{R}{r_o} - R}{\frac{R}{r_o m} \sum_{i=1}^m \left[E_{FL} \left(\left(\frac{i}{m} \right) \frac{R}{r_o}, p_X \right) \right]^{-1}},$$

$$E_{FL}(x, p_X) = \max_{0 \leq \rho \leq 1} (-\rho x + E_0(\rho, p_X) [1 - E_0(\rho, p_X)]). \quad (8)$$

where $E_0(\rho, p_X)$ is defined in (4).

For a given m , encoding and decoding complexities of the m -level concatenated codes are linear in the number of transmitted symbols and the number of received symbols. ■

Theorem 3 can be proved by following the analysis of m -level concatenated codes presented in [10][14] and replacing the inner code error exponent in the analysis with the error exponent lower bound given in Corollary 1.

Corollary 2: The following fountain error exponent can be arbitrarily approached by multi-level concatenated fountain codes with linear encoding/decoding complexity.

$$E_{Fc}^{(\infty)}(R) = \max_{p_X, \frac{R}{\mathcal{C}_F} \leq r_o \leq 1} \left(\frac{R}{r_o} - R \right) \left[\int_0^{\frac{R}{r_o}} \frac{dx}{E_{FL}(x, p_X)} \right]^{-1}, \quad (9)$$

where $E_{FL}(x, p_X)$ is defined in (8). ■

In Figure 4, we illustrate $E_{Fc}^{(\infty)}(R)$ and the Blokh-Zyablov exponent $E_c^{(\infty)}(R)$ for a BSC with crossover probability 0.1. It can be seen that $E_{Fc}^{(\infty)}(R)$ is not far away from the Blokh-Zyablov exponent.

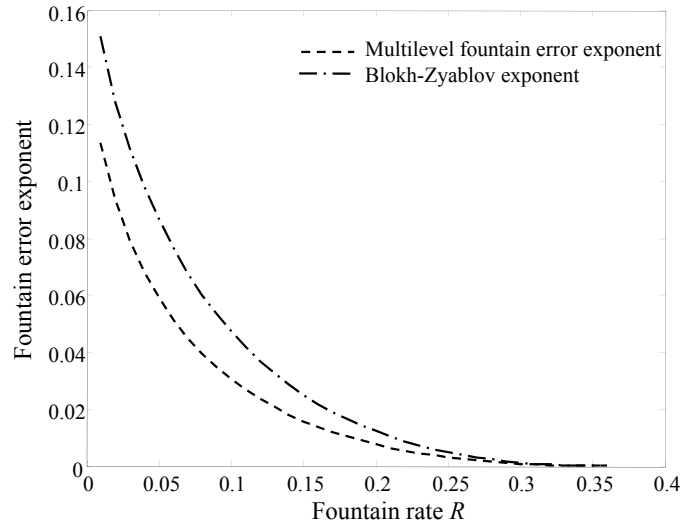


Fig. 4. Comparison of multilevel fountain error exponent $E_{Fc}^{(\infty)}(R)$ and the Blokh-Zyablov exponent $E_c^{(\infty)}(R)$.

V. RATE COMPATIBLE FOUNTAIN COMMUNICATION

Consider the application of software patch distribution. When a significant number of patches are released, the software company may want to combine the patches together as a service pack. However, if a user already have some of the patches, he may only want to download the new patches, rather than the whole service pack. For the convenience of the patch server, all patches of the service pack should be encoded jointly. But for the communication efficiency of each particular user, we also want the fountain system to achieve the same rate and error performance as if only the novel part of the service pack is transmitted. We require such optimality be achieved simultaneously for all users, and define such a fountain communication model the rate compatible fountain communication.

Assume a source message w , which takes $\exp(NR)$ possible values, can be partitioned into L sub-messages $w = [w_1, w_2, \dots, w_L]$, where $w_i, \forall i$, can take $\exp(Nr_i)$ possible values, $\sum_i r_i = R$. Consider the following extended one-level concatenated fountain coding scheme. For all $i \in \{1, \dots, L\}$, the encoder first uses a near MDS outer code with length N_o and rate r_o to encode sub-message w_i into an outer codeword $\{\xi_{i1}, \dots, \xi_{iN_o}\}$, as illustrated in Figure 5. Next, for all $k \in \{1, \dots, N_o\}$, the encoder combines

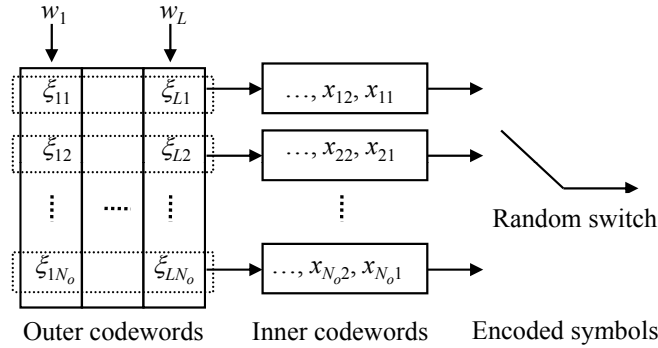


Fig. 5. Concatenated fountain codes for rate compatible communication.

outer codeword symbols $\{\xi_{1k}, \dots, \xi_{Lk}\}$ into a macro symbol $\xi_k = [\xi_{1k}, \dots, \xi_{Lk}]$. A random fountain code is then used to map ξ_k into an infinite channel input sequence $\{x_{k1}, x_{k2}, \dots\}$.

Without loss of generality, we assume the decoder already has sub-messages $\{w_{l+1}, \dots, w_L\}$, where $l \in [1, L - 1]$ is an integer. The decoder estimates the source message after $N_l = N \frac{\sum_{i=1}^l r_i}{R}$ channel output symbols are received. From the decoder point of view, since the unknown messages $[w_1, \dots, w_l]$ can only take $\exp(N \sum_{i=1}^l r_i)$ possible values, the *effective* fountain information rate of the system is $R_{ef} = \frac{N \sum_{i=1}^l r_i}{N_l} = R$. According to the known messages $[w_{l+1}, \dots, w_L]$, the decoder first strikes out from fountain codebooks all codewords corresponding to the wrong messages. The one-level concatenated fountain code is then decoded using the same procedure as described in Section IV. Assume the average number of symbols received by each inner codeword $\tilde{N}_i = \frac{N_l}{N_o} = \frac{N}{N_o} \frac{\sum_{i=1}^l r_i}{R}$ is large enough to enable

asymptotic error performance analysis. By following a similar analysis given in the proof of Theorem 2, it can be seen that error exponent $E_{Fc}(R)$ given in (5) can still be arbitrarily approached, irrespective of the value of l .

Therefore, given a rate partitioning $R = [r_1, \dots, r_L]$, the encoder can encode the complete message irrespective of the sub-messages known at the decoder. The fountain system can achieve the same rate and error performance as if only the unknown sub-messages are encoded and transmitted. Extending the scheme to multi-level concatenated codes is straightforward.

VI. FOUNTAIN COMMUNICATION OVER AN UNKNOWN CHANNEL

In previous sections, we have assumed that concatenated fountain codes should be optimized based on a known memoryless channel model $p_{Y|X}$. However, such an optimization may face various challenges in practical applications. For example, suppose a transmitter broadcasts encoded symbols to multiple receivers simultaneously. Channels experienced by different receivers may be different. Even if the channels are known, the transmitter still faces the problem of optimizing fountain codes simultaneously for multiple channels. For another example, suppose the source message (e.g., a software patch) is available at multiple servers. A user may collect encoded symbols from multiple servers separately over different channels and use these symbols to jointly decode the message. By regarding the symbols as received over a virtual channel, we want the fountain system to achieve good rate and error performance without requiring the virtual channel model at the transmitter. We term the communication model in the latter example the rate combining fountain communication. In both examples, the research question is whether key coding parameters can be determined without full channel knowledge at the transmitter.

Consider fountain communication over a memoryless channel $p_{Y|X}$ using one-level concatenated fountain codes. We assume the channel is symmetric, and hence the optimal input distribution p_X is known at the transmitter. Other than its symmetry, we assume channel information $p_{Y|X}$ is unknown at the transmitter, but known at the receiver. Given p_X , define $I(p_X) = I(X; Y)$ as the mutual information between the input and output of the memoryless channel. We assume the transmitter and the receiver agree on achieving a fountain information rate of $\gamma I(p_X)$ where γ is termed the normalized fountain rate, known at the transmitter.

Recall from the proof of Theorem 2 that, if $p_{Y|X}$ is known at the transmitter, the following error exponent can be arbitrarily approached.

$$E_{Fc}(\gamma, p_X) = \max_{0 \leq r_o \leq 1} E_{Fc}(\gamma, p_X, r_o)$$

$$E_{Fc}(\gamma, p_X, r_o) = \max_{0 \leq \rho \leq 1} (1 - r_o) I(p_X) \left(-\rho \frac{\gamma}{r_o} + \frac{E_0(\rho, p_X)}{I(p_X)} \left[1 - \frac{1 + r_o}{2} E_0(\rho, p_X) \right] \right). \quad (10)$$

Without $p_{Y|X}$ at the transmitter, we set the outer code rate r_o at $r_o = \frac{\sqrt{\gamma^2 + 8\gamma} - \gamma}{2}$ and define the corresponding error exponent by

$$E_{Fcs}(\gamma, p_X) = E_{Fc} \left(\gamma, p_X, r_o = \frac{\sqrt{\gamma^2 + 8\gamma} - \gamma}{2} \right). \quad (11)$$

The following theorem indicates that $E_{Fcs}(\gamma, p_X)$ approaches $E_{Fc}(\gamma, p_X)$ asymptotically.

Theorem 4: Given the memoryless channel $p_{Y|X}$ and a source distribution p_X , the following limit holds,

$$\lim_{\gamma \rightarrow 1} \frac{E_{Fcs}(\gamma, p_X)}{E_{Fc}(\gamma, p_X)} = 1. \quad (12)$$

■

The proof of Theorem 4 is given in Appendix C.

In Figure 6, we plot $E_{Fcs}(\gamma, p_X)$ and $E_{Fc}(\gamma, p_X)$ for BSC with crossover probability 0.1. It can be seen that setting r_o at $r_o = \frac{\sqrt{\gamma^2 + 8\gamma} - \gamma}{2}$ is near optimal for all normalized fountain rate values. Further

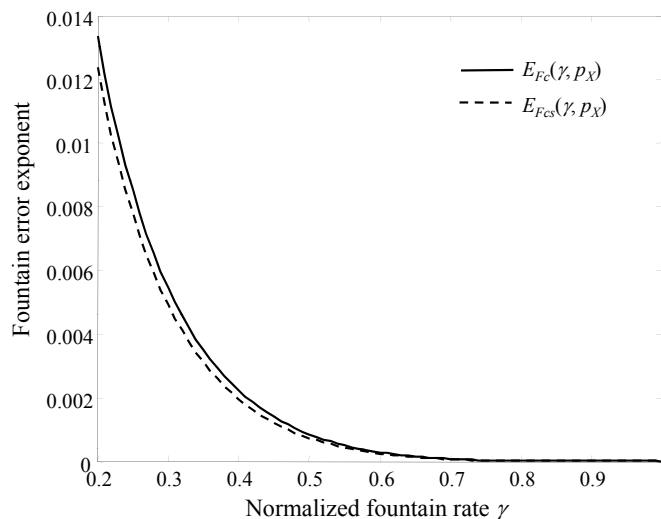


Fig. 6. Error exponents achieved by optimal r_o and suboptimal $r_o = \frac{\sqrt{\gamma^2 + 8\gamma} - \gamma}{2}$ versus normalized fountain rate γ .

discussions on fountain communication over unknown channels is outside the scope of this paper.

VII. CONCLUSIONS

We proposed concatenated fountain codes with linear coding complexity for fountain communication over a discrete-time memoryless channel. Fountain error exponents achievable by one-level and multi-level concatenated codes were derived. It was shown that the fountain error exponents are less than but close to Forney's and Blokh-Zyablov exponents. In rate compatible communication where decoder knows part of the message, with the encoder still encoding the complete message, concatenated fountain codes can achieve the same rate and error performance as if only the unknown part of the message is encoded. For

one-level concatenated codes and for some channels, it was also shown that near optimal error exponent can be achieved with an outer code rate independent of the channel statistics.

APPENDIX

A. Proof of Theorem 2

Proof: We first introduce the basic idea of the proof.

Assume the decoder starts decoding after receiving $N = N_o N_i$ symbols, where N_o is the length of the outer codeword, N_i is the *expected* number of received symbols from each inner code. In the following error exponent analysis, we will obtain asymptotic results by first taking N_o to infinity and then taking N_i to infinity.

Let \mathbf{z} be an N_o -dimensional vector whose k th element z_k is the normalized effective codeword length of the k th inner code, from which the conditional empirical distribution function $F_{Z|\theta}$ can be induced, as a function of variable $z \geq 0$, given the random variable θ specified in Section III. Let the conditional density function of $F_{Z|\theta}$ be $f_{Z|\theta}$. Note that the empirical density function $f_{Z|\theta}$ itself is a random variable, whose distribution is denoted by G_F , as a function of $f_{Z|\theta}$. Assume, given θ , the conditional error probability of the concatenated code can be written as $P_{e|\theta}(f_{Z|\theta}) = \exp(-N_i N_o E_f(f_{Z|\theta}, R))$, where the conditional error exponent $E_f(f_{Z|\theta}, R)$ is a function of $f_{Z|\theta}$. The overall error probability can therefore be written as

$$P_e = \int_{\theta} \exp(-N_i N_o E_f(f_{Z|\theta}, R)) dG_F(f_{Z|\theta}). \quad (13)$$

Consequently, error exponent of the concatenated code is given by

$$\begin{aligned} E_{Fc}(R) &= \lim_{N_i \rightarrow \infty} \lim_{N_o \rightarrow \infty} -\frac{1}{N_i N_o} \log \int_{\theta} \exp(-N_i N_o E_f(f_{Z|\theta}, R)) dG_F(f_{Z|\theta}) \\ &= \min_{f_Z} \left\{ E_f(f_Z, R) - \lim_{N_i, N_o \rightarrow \infty} \frac{1}{N_i N_o} \log dG_F(f_Z) \right\}, \end{aligned} \quad (14)$$

where in the second equality we wrote $f_{Z|\theta}$ as f_Z to simplify the notation.

The rest of the proof contains three parts. In Part I, we derive the expression of $\lim_{N_i, N_o \rightarrow \infty} \frac{1}{N_i N_o} \log dG_F(f_Z)$. In Part II, we derive the expression of $E_f(f_Z, R)$. In Part III, we use the results of the first two parts to optimize (14) and to obtain $E_{Fc}(R)$.

Part I: Let $dz > 0$ be a small constant. We define $\{z_g | z_g = ndz, n = 0, 1, \dots\}$ as the set of “grid values” each can be written as an non-negative integer multiplying dz . Given a normalized effective inner codeword length vector \mathbf{z} , the empirical density f_Z is induced as follows. We first quantize the elements of \mathbf{z} , say z_k , to the closest grid value no larger than z_k , i.e., $z_g \leq z_k$. Denote the quantized \mathbf{z} vector by $\mathbf{z}^{(q)}$. For any grid value z_g , we define $\mathcal{I}_{z_g} = \{i | z_i^{(q)} = z_g\}$ as the set of indices corresponding to which the

elements of $\mathbf{z}^{(q)}$ vector equal the particular z_g . Given \mathbf{z} , the empirical density f_Z is a discrete function defined on the grid values, with $f_Z(z_g) = \frac{|T_{z_g}|}{N_o dz}$. According to the f_Z definition, we have

$$\sum_{z_g} f_Z(z_g) dz = 1. \quad (15)$$

Let $\mathbf{z}(i)$ be an N_o -dimensional vector with only one non-zero element corresponding to the i th received symbol. If the i th received symbol belongs to the k th inner code, then we let the k th element of $\mathbf{z}(i)$ equal 1 and let all other elements equal 0. Since the random switch (illustrated in Figure 2) picks inner codes uniformly, we have

$$E[\mathbf{z}(i)] = \frac{1}{N_o} \mathbf{1}, \quad \text{cov}[\mathbf{z}(i)] = \frac{1}{N_o} \mathbf{I}_{N_o} - \frac{1}{N_o^2} \mathbf{1} \mathbf{1}^T, \quad (16)$$

where $\mathbf{1}$ is an N_o -dimensional vector with all elements being one. According to the definitions, we have $\mathbf{z} = \frac{1}{N_i} \sum_{i=1}^{N_i N_o} \mathbf{z}(i)$. Since the total number of received symbols equal $N_i N_o$, we must have $\mathbf{1}^T \mathbf{z} = N_o$. This implies that for all empirical density functions f_Z , we have

$$\sum_{z_g} z_g f_Z(z_g) dz \in [1 - dz, 1]. \quad (17)$$

Since \mathbf{z} equals the summation of $N_i N_o$ independently distributed vectors $\mathbf{z}(i)$, the characteristic function of $\sqrt{\frac{N_i}{N_o}}(\mathbf{z} - \mathbf{1})$, denoted by $\varphi_Z(\mathbf{t})$, can be written as

$$\varphi_Z(\mathbf{t}) = \left[1 - \frac{1}{2} \mathbf{t}^T \frac{1}{N_o^2 N_i} \mathbf{I}_{N_o-1} \mathbf{t} + o\left(\frac{\|\mathbf{t}\|^2}{N_o^2 N_i}\right) \right]^{N_o N_i}, \quad (18)$$

where $\mathbf{t} = [t_1, \dots, t_{N_o-1}]^T$ is an (N_o-1) -dimensional vector¹ with bounded elements, i.e., $\max_{1 \leq k \leq N_o-1} |t_k| \leq \alpha$, for some constant $\alpha > 0$. (18) implies that

$$\lim_{N_o \rightarrow \infty} \left\{ \varphi_Z(\mathbf{t}) - \exp\left(-\frac{1}{2N_o} \mathbf{t}^T \mathbf{t}\right) \right\} = 0. \quad (19)$$

Consequently, for large $N_i, N_o, N_o \gg N_i$, the probability that \mathbf{z} gives a particular quantized vector $\mathbf{z}^{(q)}$ is upper bounded by

$$Pr\{\mathbf{z}^{(q)}\} \leq \left(dz \sqrt{\frac{N_i}{N_o}}\right)^{N_o-1} \left(\frac{N_o}{2\pi}\right)^{\frac{N_o-1}{2}} \exp\left(-\frac{N_i}{2} [\|\mathbf{z}^{(q)} - \mathbf{1}\|^2 - 3N_o(dz)^2]\right). \quad (20)$$

The first term on the right hand side of (20) is the volume of the neighborhood of $\sqrt{\frac{N_i}{N_o}}(\mathbf{z} - \mathbf{1})$ in which the quantized codeword length vector of \mathbf{z} equals $\mathbf{z}^{(q)}$. The second term is a revised Gaussian density

¹Note that because $\mathbf{1}^T \mathbf{z} = N_o$, \mathbf{z} has only $N_o - 1$ linearly independent elements. The characteristic function is obtained by first projecting \mathbf{z} to an $(N_o - 1)$ -dimensional space. The detailed derivation is skipped.

derived from the characteristic equation $\exp\left(-\frac{1}{2N_o}t^T t\right)$. Note that the offset $-3N_o(dz)^2$ in the second term is necessary to ensure the validity of the upper bound.

Let f_Z be the empirical density induced from a particular quantized codeword length vector $\mathbf{z}^{(q)}$. It can be shown that the probability for \mathbf{z} to follow an empirical density f_Z is upper bounded by

$$\begin{aligned} Pr\{f_Z\} &= K_0(N_i, N_o)Pr\{\mathbf{z}^{(q)}\} \\ &\leq K_0(N_i, N_o) \left(dz\sqrt{\frac{N_i}{2\pi}}\right)^{N_o-1} \exp\left(-\frac{N_i}{2} [\|\mathbf{z}^{(q)} - \mathbf{1}\|^2 - 3N_o(dz)^2]\right) \\ &= K_0(N_i, N_o) \left(dz\sqrt{\frac{N_i}{2\pi}}\right)^{N_o-1} \exp\left(\frac{3}{2}N_iN_o(dz)^2\right) \exp\left(-\frac{N_iN_o}{2} \sum_{z_g} (z_g - 1)^2 f_Z(z_g) dz\right), \end{aligned} \quad (21)$$

where $K_0(N_i, N_o)$ is a permutation term that satisfied $\lim_{N_i, N_o \rightarrow \infty} \frac{\log K_0(N_i, N_o)}{N_i N_o} = 0$.²

From (21), we can see that for all f_Z the following inequality holds,

$$-\lim_{N_i, N_o \rightarrow \infty} \frac{\log Pr\{f_Z\}}{N_i N_o} \geq \frac{1}{2} \sum_{z_g} (z_g - 1)^2 f_Z(z_g) dz - \frac{3}{2} (dz)^2. \quad (22)$$

Part II: Next, we will derive the expression of $E_f(f_Z, R)$, which is the error exponent conditioned on an empirical inner codeword length density f_Z .

Let \mathbf{z} be a particular N_o -dimensional inner codewords length vector, which follows the density function f_Z . Given a finite dz , since error probability conditioned on f_Z can be written as $P_e(f_Z) = \exp(-N_i N_o E_f(f_Z, R))$, error probability given \mathbf{z} can be written as

$$P_e(\mathbf{z}) = \frac{\exp(-N_i N_o E_f(f_Z, R))}{K_1(N_i, N_o)}, \quad \lim_{N_i, N_o \rightarrow \infty} K_1(N_i, N_o) = 0, \quad (23)$$

where $K_1(N_i, N_o)$ is a permutation term. Consequently, we can obtain $E_f(f_Z, R)$ by assuming a particular inner codeword length vector \mathbf{z} , whose corresponding inner codeword length density is f_Z .

A strict error exponent derivation should proceed by first assuming a fixed dz . For each grid value z_g , error performances of inner codes whose effective codeword lengths belong to $(z_g, z_g + dz]$ should be bounded as function of R and z_g . After obtaining the overall error exponent of the concatenated fountain code based on a fixed dz , asymptotic result can then be obtained by taking $dz \rightarrow 0$. The order that $dz \rightarrow 0$ should be taken at the end is necessary since otherwise the previous derivations such as (20), (21), (23) are no longer valid. However, under the assumptions that the previous derivations are valid, taking $dz \rightarrow 0$ first does not affect the validity of the rest of the proof. Therefore, in order to simplify the notations, from now on, we will first take $dz \rightarrow 0$.

²Validity of this limit can be shown by jointly consider the two terms in (14). The detail is skipped.

By taking $dz \rightarrow 0$, (22) becomes

$$-\lim_{dz \rightarrow 0, N_i, N_o \rightarrow \infty} \frac{\log Pr\{f_Z\}}{N_i N_o} = \int_0^\infty \frac{(z-1)^2}{2} f_Z(z) dz. \quad (24)$$

According to the definition of f_Z and the property that $\mathbf{1}^T \mathbf{z} = N_o$, we also have

$$\int_0^\infty f_Z(z) dz = 1, \quad \int_0^\infty z f_Z(z) dz = 1. \quad (25)$$

Consequently, error exponent (14) becomes

$$E_{Fc}(R) = \min_{f_Z, \int_0^\infty z f_Z(z) dz = 1} \left\{ E_f(f_Z, R) + \int_0^\infty \frac{(z-1)^2}{2} f_Z(z) dz \right\}. \quad (26)$$

Assume the outer code has rate r_o , and is able to recover the source message from dN_o outer symbol erasures and tN_o outer symbol errors so long as $d + 2t \leq (1 - r_o - \zeta_0)$, where $\zeta_0 > 0$ is a constant that can be made arbitrarily small. To simplify the notations, we take $\zeta_0 \rightarrow 0$ first³. Assume, for all k , the k th inner code reports an estimate of the outer symbol $\hat{\xi}_k$ together with a reliability weight $\alpha_k \in [0, 1]$. Apply Forney's GMD decoding to the outer code [12], the source message can be recovered if the following inequality holds [9, Theorem 3.1b].

$$\sum_{k=1}^{N_o} \alpha_k \mu_k > r_o N_o, \quad (27)$$

where $\mu_k = 1$ if $\hat{\xi}_k = \xi_k$, and $\mu_k = -1$ if $\hat{\xi}_k \neq \xi_k$. Consequently, error probability conditioned on the given \mathbf{z} vector is bounded by

$$\begin{aligned} P_e(R, r_o, \mathbf{z}) &\leq Pr \left\{ \sum_{k=1}^{N_o} \alpha_k \mu_k \leq r_o N_o \right\} \\ &\leq \min_{s \geq 0} \frac{E \left[\exp \left(-s N_o \sum_{k=1}^{N_o} \alpha_k \mu_k \right) \right]}{\exp(-s N_o r_o N_o)}. \end{aligned} \quad (28)$$

where the last inequality is due to Chernoff's bound.

Given the inner codeword lengths \mathbf{z} , random variables $\alpha_k \mu_k$ for different inner codes are independent. Therefore, (28) can be further written as

$$\begin{aligned} P_e(R, r_o, \mathbf{z}) &\leq \min_{s \geq 0} \frac{\prod_{k=1}^{N_o} E \left[\exp(-s N_i \alpha_k \mu_k) \right]}{\exp(-s N_i r_o N_o)} \\ &= \min_{s \geq 0} \frac{\exp \left(\sum_{k=1}^{N_o} \log E \left[\exp(-s N_i \alpha_k \mu_k) \right] \right)}{\exp(-s N_i r_o N_o)}. \end{aligned} \quad (29)$$

Now we will derive the expression of $\log E \left[\exp(-s N_i \alpha_k \mu_k) \right]$ for the k th inner code.

Assume the normalized effective codeword length is z_k . Given z_k , depending on the received channel symbols, the decoder generates the maximum likelihood outer code estimate $\hat{\xi}_k$, and generates α_k using

³Taking $\zeta_0 \rightarrow 0$ requires a significant increase of N_i [11]. Although the linear complexity argument requires $\zeta_0 \rightarrow 0$ be taken after $N_o \rightarrow \infty$, switching the order of these asymptotic operations does not affect the validity of the error exponent result.

Forney's algorithm presented in [9, Section 4.2]. Define an adjusted error exponent function $E_z(z)$ as follows.

$$E_z(z) = \max_{0 \leq \rho \leq 1} -\rho \frac{R}{r_o} + z E_0(\rho, p_X), \quad (30)$$

where $E_0(\rho, p_X)$ is defined in (4). By following Forney's error exponent analysis presented in [9, Section 4.2], we obtain

$$-\log E[\exp(-s N_i \alpha_k \mu_k)] = \max[\min\{N_i E_z(z_k), N_i(2E_z(z_k) - s), N_i s\}, 0]. \quad (31)$$

Define a function $\phi(z, s)$ as follows,

$$\phi(z, s) = \begin{cases} -s r_o & z, E_z(z) < s/2 \\ 2E_z(z) - (1 + r_o)s & z, s/2 \leq E_z(z) < s \\ (1 - r_o)s & z, E_z(z) \geq s \end{cases} \quad (32)$$

Substitute (31) into (29), we get the expression of the conditional error exponent $E_f(f_Z, R)$ as

$$E_f(f_Z, R) = \max_{p_X, \frac{R}{C_F} \leq r_o \leq 1, s \geq 0} \int \phi(z, s) f_Z(z) dz. \quad (33)$$

Part III: Combining (33) with (26), fountain error exponent of the concatenated code is therefore given by

$$E_{Fc}(R) = \max_{p_X, \frac{R}{C_F} \leq r_o \leq 1, s \geq 0} \min_{f_Z, \int_0^\infty z f_Z(z) dz = 1} \int \left[\phi(z, s) + \frac{(1 - z)^2}{2} \right] f_Z(z) dz. \quad (34)$$

Assume f_Z^* is the inner codeword length density that minimizes $E_{Fc}(R)$ in (34). Assume we can find $0 < \lambda < 1$, and two density functions $f_Z^{(1)}, f_Z^{(2)}$, satisfying $\int_0^\infty z f_Z^{(1)}(z) dz = 1$, $\int_0^\infty z f_Z^{(2)}(z) dz = 1$, such that

$$f_Z^* = \lambda f_Z^{(1)} + (1 - \lambda) f_Z^{(2)}. \quad (35)$$

It is easily seen that $E_{Fc}(R)$ should be minimized either by $f_Z^{(1)}$ or $f_Z^{(2)}$, which contradicts the assumption that f_Z^* is optimum. In other words, if f_Z^* is indeed optimum, then a decomposition like (35) must not be possible. This implies that f_Z^* can take non-zero values on at most two different z values. Therefore, we can carry out the optimization in (34) only over the following class of f_Z functions, characterized by two variables $0 \leq z_0 \leq 1$ and $0 \leq \gamma \leq 1$.

$$f_Z(z) = \gamma \delta(z - z_0) + (1 - \gamma) \delta\left(z - \frac{1 - z_0 \gamma}{1 - \gamma}\right). \quad (36)$$

where $\delta()$ is the impulse function.

Now let us fix p_X , r_o , γ , and consider the following optimization of $E_{Fc}(R, p_X, r_o, \gamma)$.

$$E_{Fc}(R, p_X, r_o, \gamma) = \min_{0 \leq z_0 \leq 1} \max_{s \geq 0} \gamma \phi(z_0, s) + (1 - \gamma) \phi\left(\frac{1 - z_0 \gamma}{1 - \gamma}, s\right) + \frac{\gamma}{1 - \gamma} \frac{(1 - z_0)^2}{2}. \quad (37)$$

Since given z_0 , $\gamma \phi(z_0, s) + (1 - \gamma) \phi\left(\frac{1 - z_0 \gamma}{1 - \gamma}, s\right)$ is a linear function of s , depending on the value of γ , the optimum s^* that maximizes (37) should either satisfy $s^* = E_z(z_0)$ or $s^* = E_z\left(\frac{1 - z_0 \gamma}{1 - \gamma}\right)$.

When $\gamma \geq \frac{1 - r_o}{2}$, we have $s^* = E_z(z_0)$. This yields

$$E_{Fc}(R, p_X, r_o) = \min_{0 \leq z_0, \gamma \leq 1} \left[\frac{\gamma}{1 - \gamma} \frac{(1 - z_0)^2}{2} + (1 - r_o) E_z(z_0) \right]. \quad (38)$$

When $\gamma \leq \frac{1 - r_o}{2}$, we have $s^* = E_z\left(\frac{1 - z_0 \gamma}{1 - \gamma}\right)$, we have

$$\begin{aligned} E_{Fc}(R, p_X, r_o) &= \min_{0 \leq z_0, \gamma \leq 1} \left[\gamma \phi(z_0, s) + \frac{\gamma}{1 - \gamma} \frac{(1 - z_0)^2}{2} + (1 - \gamma)(1 - r_o) E_z\left(\frac{1 - \gamma z_0}{1 - \gamma}\right) \right] \\ &\geq \min_{0 \leq z_0, \gamma \leq 1} \left[2\gamma E_z(z_0) + \frac{\gamma}{1 - \gamma} \frac{(1 - z_0)^2}{2} + (1 - r_o - 2\gamma) E_z\left(\frac{1 - \gamma z_0}{1 - \gamma}\right) \right]. \end{aligned} \quad (39)$$

By substituting $E_z(z) = \max_{0 \leq \rho \leq 1} [-\rho \frac{R}{r_o} + z E_0(\rho, p_X)]$ into (39), we have

$$\begin{aligned} E_{Fc}(R, p_X, r_o) &\geq \min_{0 \leq z_0, \gamma \leq 1} \max_{0 \leq \rho \leq 1} \left\{ (1 - r_o) \left[-\rho \frac{R}{r_o} + E_0(\rho, p_X) \right] - \right. \\ &\quad \left. \frac{\gamma}{1 - \gamma} \left[(1 + r_o)(1 - z_0) E_0(\rho, p_X) - \frac{(1 - z_0)^2}{2} \right] \right\}. \end{aligned} \quad (40)$$

Note that if $(1 + r_o)(1 - z_0) E_0(\rho, p_X) - \frac{(1 - z_0)^2}{2} < 0$, we have $E_{Fc}(R, p_X, r_o) \geq (1 - r_o) \left[-\rho \frac{R}{r_o} + E_0(\rho, p_X) \right]$ which is Forney's exponent given p_X, r_o . This contradicts with the fact that Forney's exponent is the maximum achievable exponent for one-level concatenated codes in a classical system [9]. Therefore, we must have $(1 + r_o)(1 - z_0) E_0(\rho, p_X) - \frac{(1 - z_0)^2}{2} \geq 0$. Consequently, both (38) and (40) are minimized at $\gamma^* = \frac{1 - r_o}{2}$ and

$$\begin{aligned} E_{Fc}(R) &= \max_{\frac{R}{C_F} \leq r_o \leq 1, p_X} \min_{0 \leq z_0 \leq 1} \left\{ (1 - r_o) E_z(z_0) + \frac{1 - r_o}{1 + r_o} \frac{(1 - z_0)^2}{2} \right\} \\ &= \max_{\frac{R}{C_F} \leq r_o \leq 1, p_X, 0 \leq \rho \leq 1} \min_{0 \leq z_0 \leq 1} \left\{ (1 - r_o) \left(-\rho \frac{R}{r_o} + E_0(\rho, p_X) \right) \right. \\ &\quad \left. + \frac{1 - r_o}{1 + r_o} \frac{(1 - z_0)^2}{2} [(1 - z_0) - 2(1 + r_o) E_0(\rho, p_X)] \right\}. \end{aligned} \quad (41)$$

The last step is to optimize (41) over z_0 . Note that if $(1 + r_o) E_0 \geq 1$, for a fixed r_o , we have

$$E_{Fc}(R, r_o) \leq \max_{p_X, 0 \leq \rho \leq 1} \left\{ -\rho \frac{R}{r_o} (1 - r_o) + \frac{1 - r_o}{1 + r_o} \right\}, \quad (42)$$

which implies $\rho = 0$. But $\rho = 0$ implies $(1 + r_o) E_0 = 0 < 1$ which contradicts the assumption $(1 + r_o) E_0 \geq 1$. Therefore, we can assume $(1 + r_o) E_0 \leq 1$. Consequently, substituting $z_0^* = 1 - (1 + r_o) E_0$ into (41) gives the desired result (5).

To achieve linear coding complexity, we fix N_i at a large number and only take N_o to infinity. According to [11], it is easy to see that the encoding complexity is linear in the number of transmitted symbols. At the receiver, we keep at most $2N_i$ symbols for each inner code and drop the extra received symbols. Consequently, the normalized effective codeword length of any inner code is upper-bounded by 2. Because (38) and (40) are both minimized at $\gamma^* = \frac{1-r_o}{2}$, according to (36), the empirical density function $f_Z(z)$ that minimizes the error exponent takes the form $f_Z(z) = \frac{1-r_o}{2}\delta(z-z_0) + \frac{1+r_o}{2}\delta\left(z - \frac{2-z_0(1-r_o)}{1+r_o}\right)$. The second term implies $z = \frac{2-z_0(1-r_o)}{1+r_o} < 2$. Therefore, upper bounding the effective codeword length by 2 does not change the error exponent result. However, with $z_k \leq 2, \forall k$, the decoding complexity of any inner code is upper-bounded by a constant $O(\exp(2N_i))$. According to [12], the overall decoding complexity of the concatenated code is therefore linear in N_o , and hence is linear in N . Since fixing N_i causes a reduction of $\zeta_1 > 0$ in the achievable error exponent, and both ζ_0, ζ_1 can be made arbitrarily small as we increase N_i , we conclude that fountain error exponent $E_{Fc}(R)$ given in (5) can be *arbitrarily approached* by one-level concatenated fountain codes with a linear coding complexity. ■

B. Proof of Corollary 1

Proof: Because $0 \leq r_o \leq 1$, it is easy to see $\tilde{E}_{Fc}(R) \leq E_{Fc}(R) \leq E_c(R)$. We will next prove $\lim_{R \rightarrow \mathcal{C}_F} \frac{\tilde{E}_{Fc}(R)}{E_{Fc}(R)} = 1$.

Define

$$g(p_X, r_o, \rho) = (1 - r_o) \left(-\rho \frac{R}{r_o} + E_0(\rho, p_X) \left[1 - \frac{1 + r_o}{2} E_0(\rho, p_X) \right] \right), \quad (43)$$

such that

$$E_{Fc}(R) = \max_{p_X, \frac{R}{\mathcal{C}_F} \leq r_o \leq 1, 0 \leq \rho \leq 1} g(p_X, r_o, \rho). \quad (44)$$

Use Taylor's expansion to expand $g(p_X, r_o, \rho)$ at $r_o = 1$ and $\rho = 0$, we get

$$g(p_X, r_o, \rho) = \sum_{i,j} \frac{1}{(i+j)!} \beta(i, j) (r_o - 1)^i \rho^j, \quad (45)$$

where $\beta(i, j) = \frac{\partial^{(i+j)} g(p_X, r_o, \rho)}{\partial r_o^i \partial \rho^j} \Big|_{r_o=1, \rho=0}$, with i and j being nonnegative integers. It can be verified that

$$\begin{aligned} \beta(1, 0) &= \left\{ \rho \frac{R}{r_o^2} - E_0(\rho, p_X) + r_o E_0^2(\rho, p_X) \right\} \Big|_{r_o=1, \rho=0} = 0, \\ \beta(2, 0) &= \left\{ -\rho \frac{2R}{r_o^3} + E_0^2(\rho, p_X) \right\} \Big|_{r_o=1, \rho=0} = 0, \\ \beta(i, 0) &= \left\{ -\rho \frac{i!(-1)^i R}{r_o^{i+1}} \right\} \Big|_{r_o=1, \rho=0} = 0, \quad \forall i \geq 3. \end{aligned} \quad (46)$$

It can also be verified that

$$\begin{aligned}\beta(0,1) &= (1-r_o) \left(-\frac{R}{r_o} + \frac{\partial E_0(\rho, p_X)}{\partial \rho} - \frac{\partial E_0(\rho, p_X)}{\partial \rho} E_0(\rho, p_X)(1+r_o) \right) \Big|_{r_o=1, \rho=0} = 0, \\ \beta(0,j) &= (1-r_o) h_j(R, \rho, r_o) \Big|_{r_o=1, \rho=0} = 0, \quad \forall j \geq 2,\end{aligned}\tag{47}$$

where $h_j(R, \rho, r_o)$ is a function of R, ρ, r_o . We also have

$$\begin{aligned}\beta(1,1) &= \left\{ \frac{R}{r_o^2} - \frac{\partial E_0(\rho, p_X)}{\partial \rho} + 2r_o E_0(\rho, p_X) \frac{\partial E_0(\rho, p_X)}{\partial \rho} \right\} \Big|_{r_o=1, \rho=0} = R - C_F, \\ \beta(2,1) &= -2R \neq 0, \\ \beta(1,2) &= - \left\{ \frac{\partial^2 E_0(\rho, p_X)}{\partial \rho^2} - 2 \left(\frac{\partial E_0(\rho, p_X)}{\partial \rho} \right)^2 \right\} \Big|_{\rho=0} \neq 0.\end{aligned}\tag{48}$$

Similarly, define

$$\tilde{g}(p_X, r_o, \rho) = (1-r_o) \left(-\rho \frac{R}{r_o} + E_0(\rho, p_X) [1 - E_0(\rho, p_X)] \right),\tag{49}$$

such that

$$\tilde{E}_{Fc}(R) = \max_{p_X, \frac{R}{C_F} \leq r_o \leq 1, 0 \leq \rho \leq 1} \tilde{g}(p_X, r_o, \rho).\tag{50}$$

Use Taylor's expansion to expand $\tilde{g}(p_X, r_o, \rho)$ at $r_o = 1$ and $\rho = 0$, we get

$$\tilde{g}(p_X, r_o, \rho) = \sum_{i,j} \frac{1}{(i+j)!} \tilde{\beta}(i,j) (r_o - 1)^i \rho^j.\tag{51}$$

where $\tilde{\beta}(i,j) = \frac{\partial^{(i+j)} \tilde{g}(p_X, r_o, \rho)}{\partial r_o^i \partial \rho^j} \Big|_{r_o=1, \rho=0}$. It can be verified that

$$\begin{aligned}\tilde{\beta}(1,0) &= \left\{ \rho \frac{R}{r_o^2} - E_0(\rho, p_X) + E_0^2(\rho, p_X) \right\} \Big|_{r_o=1, \rho=0} = 0, \\ \tilde{\beta}(i,0) &= \left\{ -\rho \frac{i!(-1)^i R}{r_o^{i+1}} \right\} \Big|_{r_o=1, \rho=0} = 0, \quad \forall i \geq 2.\end{aligned}\tag{52}$$

It can also be verified that

$$\begin{aligned}\tilde{\beta}(0,1) &= (1-r_o) \left(-\frac{R}{r_o} + \frac{\partial E_0(\rho, p_X)}{\partial \rho} - \frac{\partial E_0(\rho, p_X)}{\partial \rho} 2E_0(\rho, p_X) \right) \Big|_{r_o=1, \rho=0} = 0, \\ \tilde{\beta}(0,j) &= (1-r_o) \tilde{h}_j(R, \rho, r_o) \Big|_{r_o=1, \rho=0} = 0, \quad \forall j \geq 2,\end{aligned}\tag{53}$$

where $\tilde{h}_j(R, \rho, r_o)$ is a function of R, ρ, r_o . We also have

$$\begin{aligned}\tilde{\beta}(1,1) &= \left\{ \frac{R}{r_o^2} - \frac{\partial E_0(\rho, p_X)}{\partial \rho} + 2E_0(\rho, p_X) \frac{\partial E_0(\rho, p_X)}{\partial \rho} \right\} \Big|_{r_o=1, \rho=0} = R - C_F, \\ \tilde{\beta}(2,1) &= -2R \neq 0, \\ \tilde{\beta}(1,2) &= - \left\{ \frac{\partial^2 E_0(\rho, p_X)}{\partial \rho^2} - 2 \left(\frac{\partial E_0(\rho, p_X)}{\partial \rho} \right)^2 \right\} \Big|_{\rho=0} \neq 0,\end{aligned}\tag{54}$$

Because $\beta(1, 1) = \tilde{\beta}(1, 1)$, $\beta(2, 1) = \tilde{\beta}(2, 1) \neq 0$, $\beta(1, 2) = \tilde{\beta}(1, 2) \neq 0$, by L'Hospital's rule, we have

$$\lim_{R \rightarrow \mathcal{C}_F} \frac{\tilde{E}_{Fc}(R)}{E_{Fc}(R)} = \lim_{R \rightarrow \mathcal{C}_F, r_o \rightarrow 1, \rho \rightarrow 0} \frac{\frac{1}{2}\tilde{\beta}(1, 1)(r_o - 1)\rho + \frac{1}{6}\tilde{\beta}(2, 1)(r_o - 1)^2\rho + \frac{1}{6}\tilde{\beta}(1, 2)(r_o - 1)\rho^2}{\frac{1}{2}\beta(1, 1)(r_o - 1)\rho + \frac{1}{6}\beta(2, 1)(r_o - 1)^2\rho + \frac{1}{6}\beta(1, 2)(r_o - 1)\rho^2} = 1. \quad (55)$$

■

C. Proof of Theorem 4

Proof: Define

$$\begin{aligned} \hat{g}(\gamma, r_o, \rho) &= (1 - r_o) \left(\rho I(p_X) \left(1 - \frac{\gamma}{r_o} \right) + \frac{\rho^2}{2} \left(\frac{\partial^2 E_0(\rho, p_X)}{\partial \rho^2} \Big|_{\rho=0} - 2I^2(p_X) \right) \right), \\ \hat{E}_{Fc}(\gamma, p_X, r_o) &= \max_{0 \leq \rho \leq 1} \hat{g}(\gamma, r_o, \rho), \\ \hat{E}_{Fc}(\gamma, p_X) &= \max_{0 \leq r_o \leq 1} \hat{E}_{Fc}(\gamma, p_X, r_o). \end{aligned} \quad (56)$$

We will first prove that

$$\lim_{\gamma \rightarrow 1} \frac{E_{Fcs}(\gamma, p_X)}{\hat{E}_{Fc}(\gamma, p_X)} = 1. \quad (57)$$

Note that $\hat{g}(\gamma, r_o, \rho)$ is maximized at $\rho = \rho^*$, with

$$\rho^* = \frac{I(p_X) \left(1 - \frac{\gamma}{r_o} \right)}{-\frac{\partial^2 E_0(\rho, p_X)}{\partial \rho^2} \Big|_{\rho=0} + 2I^2(p_X)}, \quad (58)$$

where we have assumed $0 \leq \rho^* \leq 1$. This assumption is valid when r_o is also optimized. Consequently,

$\hat{E}_{Fc}(\gamma, p_X, r_o)$ is maximized at $r_o = r_o^*$, with

$$r_o^* = \arg \max_{0 \leq r_o \leq 1} (1 - r_o) \left(1 - \frac{\gamma}{r_o} \right)^2 = \frac{\sqrt{\gamma^2 + 8\gamma} - \gamma}{2}. \quad (59)$$

Therefore,

$$\lim_{\gamma \rightarrow 1} \frac{E_{Fcs}(\gamma, p_X)}{\hat{E}_{Fc}(\gamma, p_X)} \geq \lim_{\gamma \rightarrow 1} \left[\frac{E_{Fcs}(\gamma, p_X, \rho)}{\hat{g}(\gamma, p_X, \rho, r_o)} \Big|_{\rho=\rho^*, r_o=r_o^*} \right] = 1. \quad (60)$$

Following a similar idea as the proof of Corollary 1, it can be shown that

$$\lim_{\gamma \rightarrow 1} \frac{\hat{E}_{Fc}(\gamma, p_X)}{E_{Fc}(\gamma, p_X)} = 1. \quad (61)$$

Combining (60) and (61), we get

$$\lim_{\gamma \rightarrow 1} \frac{E_{Fcs}(\gamma, p_X)}{E_{Fc}(\gamma, p_X)} = \lim_{\gamma \rightarrow 1} \frac{E_{Fcs}(\gamma, p_X)}{\hat{E}_{Fc}(\gamma, p_X)} \lim_{\gamma \rightarrow 1} \frac{\hat{E}_{Fc}(\gamma, p_X)}{E_{Fc}(\gamma, p_X)} \geq 1. \quad (62)$$

Because $E_{Fcs}(\gamma, p_X) \leq E_{Fc}(\gamma, p_X)$, (62) implies $\lim_{\gamma \rightarrow 1} \frac{E_{Fcs}(\gamma, p_X)}{E_{Fc}(\gamma, p_X)} = 1$.

■

REFERENCES

- [1] J. Byers, M. Luby and A. Rege, *A Digital Fountain Approach to Reliable Distribution of Bulk Data*, ACM SIGCOMM'98, Vancouver, Canada, Sep. 1998.
- [2] M. Luby, *LT codes*, IEEE FOCS'02, Vancouver, Canada, Nov. 2002.
- [3] A. Shokrollahi, *Raptor Codes*, IEEE Trans. Inform. Theory, Vol. 52, pp. 2551-2567, Jun. 2006.
- [4] O. Etesami and A. Shokrollahi, *Raptor Codes on Binary Memoryless Symmetric Channels*, IEEE Trans. Inform. Theory, Vol. 52, pp. 2033-2051, May 2006.
- [5] S. Shamai, I. Teletar and S. Verdú, *Fountain Capacity*, IEEE Trans. Inform. Theory, Vol. 53, pp. 4327-4376, Nov. 2007.
- [6] A. Feinstein, *Error Bounds in Noisy Channels Without Memory*, IEEE Trans. Inform. Theory, Vol. 1, pp. 13-14, Sep. 1955.
- [7] R. Gallager, *A Simple Derivation of The Coding Theorem and Some Applications*, IEEE Trans. Inform. Theory, Vol. 11, pp. 3-18, Jan. 1965.
- [8] C. Shannon, R. Gallager, and E. Berlekamp, *Lower Bounds to Error Probability for Coding on Discrete Memoryless Channels*, Information and Control, Vol. 10, pp. 65-103, 522-552, 1967.
- [9] G. Forney, *Concatenated Codes*, The MIT Press, 1966.
- [10] E. Blokh and V. Zyablov, *Linear Concatenated Codes*, Nauka, Moscow, 1982 (In Russian).
- [11] V. Guruswami and P. Indyk, *Linear-Time Encodable/Decodable Codes With Near-Optimal Rate*, IEEE Trans. Inform. Theory, Vol. 51, pp. 3393-3400, Oct. 2005.
- [12] Z. Wang and J. Luo, *Approaching Blokh-Zyablov Error Exponent with Linear-Time Encodable/Decodable Codes*, to appear in IEEE Comm. Lettr.
- [13] A. Barg, G. Forney, *Random Codes: Minimum Distances and Error Exponents*, IEEE Trans. Inform. Theory Vol. 48, pp. 2568-2573, Sep. 2002.
- [14] A. Barg and G. Zémor, *Multilevel Expander Codes*, Algebraic Coding Theory and Information Theory, American Math. Soc. Vol. 68, AMS-DIMACS series, pp. 69-83, 2005.