Fountain Communication using Concatenated Codes

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

Abstract

This paper extends linear-complexity concatenated coding schemes to fountain communication over the discrete-time memoryless channel. Achievable fountain error exponents for one-level and multi-level concatenated fountain codes are derived. Several important properties of the concatenated coding schemes in multi-user fountain communication scenarios are demonstrated.

Index Terms

coding complexity, concatenated codes, error exponent, fountain communication

I. Introduction

Fountain communication [2][3] is a new communication model 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, which experience *arbitrary* erasures during transmission. The receiver decodes the message after the number of received symbols exceeds certain threshold. With the help of randomized coding, fountain communication achieves the same rate and error performance over different channel erasure realizations corresponding to an identical number of received symbols. Under the assumption that the erasure statistics is unknown at the transmitter, 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 [4] for erasure channels. LT codes can recover k information bits from $k+O\left(\sqrt{k}\ln^2(k/\delta)\right)$ encoded symbols with probability $1-\delta$ and a complexity of $O(k\ln(k/\delta))$, for any $\delta>0$ [4]. Shokrollahi proposed Raptor codes in [5] by combining appropriate LT codes with a pre-code. Raptor codes can recover k information bits from $k(1+\epsilon)$ encoded symbols at high probability with complexity $O\left(k\log(1/\epsilon)\right)$. LT codes and Raptor codes can achieve

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

This work was supported by the National Science Foundation under Grants CCF-0728826, CCF-1016985 and CNS-1116134. Any opinions, findings, and conclusions or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of the National Science Foundation. Part of the material in this paper were presented at the IEEE International Symposium on Information Theory, Seoul, Korea, June 2009 [1].

optimum rate with close to linear and linear complexity, respectively. However, under a fixed rate, error probabilities of the two coding schemes do not decrease exponentially in the number of received symbols. Generalization of Raptor codes from erasure channels to binary symmetric channels (BSCs) was studied by Etesami and Shokrollahi in [6]. In [7], Shamai, Telatar and Verdú systematically extended fountain communication to arbitrary channels and showed that fountain capacity [7] and Shannon capacity take the same value for stationary memoryless channels. Achievability of fountain capacity was demonstrated in [7] using a random coding scheme whose error probability decreases exponentially in the number of received symbols. Unfortunately, the random coding scheme considered in [7] is impractical due to its exponential complexity. While the question on whether exponential error probability scaling law can be achieved by linear complexity fountain codes remains open, the special properties of fountain codes have attracted many research efforts on applying fountain coding to practical communication scenarios such as relay communication [8], OFDM system [9], video broadcasting [10], data collection [11][12], error protection [13][14], and networking [15][16], etc.

In classical point-to-point communication over a discrete-time memoryless channel, it is well known that Shannon capacity can be achieved with an exponential error probability scaling law and a linear encoding/decoding complexity [17][18]. The fact that communication error probability can decrease exponentially in the codeword length at any information rate below the capacity was firstly shown by Feinstein [19]. The corresponding exponent was defined as the error exponent. Tight lower and upper bounds on error exponent were obtained by Gallager [20], and by Shannon, Gallager, Berlekamp [21], respectively. In [22], 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 the Forney's exponent, for any rate less than Shannon capacity with a polynomial complexity [22]. Forney's concatenated codes were generalized by Blokh and Zyablov [23] to multi-level concatenated codes, whose maximum achievable error exponent is known as the Blokh-Zyablov error exponent. In [17], Guruswami and Indyk introduced a class of linear complexity near maximum distance separable (MDS) error-correction codes. By using Guruswami-Indyk's codes as the outer codes in concatenated coding schemes, achievability of Forney's and Blokh-Zyablov exponents with linear coding complexity over general discrete-time memoryless channels was proved in [18].

In this paper, we show that classical concatenated coding schemes can be extended to fountain communication over the discrete-time memoryless channel to achieve positive fountain error exponent (defined in Section II) at any rate below the fountain capacity with a linear coding complexity. Note that, given the success of linear complexity concatenated coding schemes in classical communication systems [17][18], the idea of extending the concatenated coding framework to fountain communication is indeed a natural one. However, such an extension has not yet been rigorously investigated in the literature. Part of the reason is that, with the channel-dependent communication length in fountain systems, error performance analysis in the concatenated coding scheme is very challenging. The key contribution of this paper is the derivation of the achievable error exponents for one-level and multi-level concatenated fountain codes, as presented in Theorem 2 and Corollary 2. We show that these error exponents are close in value to (but are not equal to) their upper bounds, which are Forney's exponent [22] for one-level concatenation and Blokh-Zyablov exponent [23] for multi-level concatenation, respectively. Furthermore, we show that concatenated fountain codes possess several important properties useful for network applications. More specifically, when one or more transmitters send common information to multiple receivers over discrete-time memoryless channels, concatenated fountain codes can often achieve near optimal rate and error performance simultaneously for all receivers even if the receivers have different prior knowledge about the transmitted message.

The rest of the paper is organized as follows. The fountain communication model is defined in Section II. In Section III, we introduce the preliminary results on random fountain codes, which are basic components of the concatenated coding schemes. One-level and multi-level concatenated fountain codes are introduced in Section IV. Special properties of concatenated fountain codes in network communication scenarios are introduced in Sections V and VI. The conclusions are given in Section VII. We use natural logarithms throughout this paper.

II. Fountain Communication Model

Consider the fountain communication system illustrated in Figure 1. Assume that the encoder uses a fountain coding scheme [7] with W codewords to map the source message $w \in \{1, 2, \cdots, W\}$ into an infinite channel input symbol sequence $\{x_{w1}, x_{w2}, \cdots\}$. Assume that the channel is discrete-time memoryless, characterized by the conditional point mass function (PMF) or probability density function (PDF) $p_{Y|X}(y|x)$, where $x \in \mathcal{X}$ is the channel input symbol with $\mathcal{X}$ being the finite channel input alphabet, and $y \in \mathcal{Y}$ is channel output symbol with $\mathcal{Y}$ being the finite channel output alphabet, respectively. Assume that the channel information is known at both the encoder and the decoder[1]. The channel output symbols are then passed through an erasure device which generates arbitrary erasures. Define schedule $\mathcal{N} = \{i_1, i_2, \cdots, i_{|\mathcal{N}|}\}$ as a subset of positive integers, where $|\mathcal{N}|$ is its cardinality [7]. Assume that the erasure device generates erasures only at those time instances not belonging to schedule $\mathcal{N}$. In other words,

[1]The case when channel information is not available at the encoder will be investigated in Section VI.

only the channel output symbols with indices in $\mathcal{N}$, denoted by $\{y_{wi_1}, y_{wi_2}, \cdots, y_{wi_{|\mathcal{N}|}}\}$, are observed by the receiver. The schedule $\mathcal{N}$ is arbitrarily chosen and unknown at the encoder.

Rate and error performance variables of the system are defined as follows. We say the fountain rate of the system is $R = (\log W)/N$, if the decoder, after observing $|\mathcal{N}| = N$ channel symbols, outputs an estimate $\hat{w} \in \{1, 2, \cdots, W\}$ of the source message based on $\{y_{wi_1}, y_{wi_2}, \cdots, y_{wi_{|\mathcal{N}|}}\}$ and $\mathcal{N}$. Decoding error happens when $\hat{w} \neq w$. Define error probability $P_e(N)$ as,

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

We say a fountain rate R is *achievable* if there exists a fountain coding scheme with $\lim_{N\to\infty} P_e(N) = 0$ at rate R [7]. The exponential rate at which error probability vanishes is defined as the fountain error exponent, denoted by $E_F(R)$,

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

Define fountain capacity $\mathcal{C}_F$ as the supremum of all achievable fountain rates. It was shown in [7] that $\mathcal{C}_F$ equals Shannon capacity of the stationary memoryless channel. Note that the scaling law here is defined with respect to the number of *received* symbols.

III. Random Fountain Codes

In a random fountain coding scheme [7], encoder and decoder share a fountain code library $\mathcal{L} = \{C_\theta : \theta \in \Theta\}$, which is a collection of fountain codebooks C_θ indexed by a set Θ. All codebooks in the library have the same number of codewords and each codeword has an infinite number of channel input symbols. Let $C_\theta(w)_j$ be the j^{th} codeword symbol in codebook C_θ corresponding to message w, for $j \in \{1, 2, \cdots\}$. To encode the message, the encoder first selects a codebook by generating θ according to a distribution ϑ, such that the random variables $x_{w,j} : \theta \to C_\theta(w)_j$ are i.i.d. with a pre-determined input distribution p_X [7]. Then the encoder uses codebook C_θ to map the message into a codeword. We assume that the actual realization of θ is known to the decoder but is unknown to the erasure device. Therefore channel erasures, although arbitrary, are independent from the codebook generation. Maximum likelihood decoding is assumed at the decoder given the knowledge of the codebook, schedule, and channel information [7]. Due to the random codebook selection, without being conditioned on θ, the error probability experienced by each message is identical. Therefore, the error probability $P_e(N)$ defined in (1) can be written as follows [7],

$$P_e(N) = \max_{w} \sup_{\mathcal{N}, |\mathcal{N}| \geq N} Pr\{\hat{w} \neq w | w, \mathcal{N}\} = \sup_{\mathcal{N}, |\mathcal{N}| \geq N} \frac{1}{W} \sum_{w} Pr\{\hat{w} \neq w | w, \mathcal{N}\}. \tag{3}$$

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

$$E_{Fr}(R) = \max_{p_X} E_{FL}(R, p_X), \tag{4}$$

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

$$\begin{aligned} E_{FL}(R, p_X) &= \max_{0 \le \rho \le 1} \left\{ -\rho R + E_0(\rho, p_X) \right\}, \\ E_0(\rho, p_X) &= -\log \sum_y \left(\sum_x p_X(x) p_{Y|X}(y|x)^{\frac{1}{1+\rho}} \right)^{(1+\rho)}. \end{aligned} \tag{5}$$

If the channel is continuous, then summations in (5) should be replaced by integrals. ■

Theorem 1 was claimed implicitly in, and can be shown by, the proof of [7, Theorem 2].

$E_{Fr}(R)$ given in (4) equals the random-coding exponent of a classical communication system over the same channel [20]. For binary symmetric channels (BSCs), since random linear codes simultaneously achieve the random-coding exponent at high rates and the expurgated exponent at low rates [24], it can be easily shown that the same fountain error exponent is achievable by random linear fountain codes. However, it is not clear whether there exists an expurgation operation, such as the one proposed in [20], that is robust to the observation of any subset of channel outputs. Therefore, whether the expurgated exponent is achievable for fountain communication over a general discrete-time memoryless channel is unknown.

IV. Concatenated Fountain Codes

Consider a one-level concatenated fountain coding scheme illustrated in Figure 2. Assume that source message w can take $\lfloor \exp(NR) \rfloor$ possible values with equiprobability, where R is the targeted fountain information rate. Assume that the communication terminates after N channel output symbols are observed at the decoder. The one-level concatenated fountain code consists of an outer code and several inner codes. The encoder first encodes the message using the outer code into an outer codeword $\{\xi_1, \xi_2, \cdots, \xi_{N_o}\}$, with N_o outer symbols, each belonging to a finite field of appropriate size. We assume that the outer code is a linear-time encodable/decodable near MDS error-correction code of rate $r_o \in (0, 1]$. That is, at a fixed r_o and as N_o is taken to infinity, the outer code can recover the source message from a received codeword with dN_o symbol erasures and tN_o symbol errors, so long as $2t + d \le (1 - r_o - \zeta_0)$, where $\zeta_0 > 0$ is a positive constant that can be made arbitrarily small. The encoding and decoding complexities are linear in the number of outer codeword length N_o. An example of such linear complexity error-correction code was presented by Guruswami and Indyk in [17]. Each outer symbol ξ_k $(k \in \{1, \cdots, N_o\})$ can take $\left\lfloor \exp\left(\frac{N}{N_o}\frac{R}{r_o}\right)\right\rfloor$ possible values.

We use a set of random fountain codes described in Section III as the inner codes, each with $\lfloor \exp(N_i R_i) \rfloor$ codewords, where $N_i = \frac{N}{N_o}$ and $R_i = \frac{R}{r_o}$. To simplify the notations, we have assumed that N_i and N_o are both integers. We also assume that $N_o \gg N_i \gg 1$. The encoder then uses these inner codes to map each outer symbol ξ_k into an inner codeword, which is an infinite sequence of channel input symbols $\{x_{k1}, x_{k2}, \cdots\}$. The inner codewords are regarded as N_o channel input symbol queues, as shown in Figure 2. In each time unit, the encoder uses a random switch to pick one inner code and sends the first channel input symbol in the corresponding queue through the channel as modeled in Section II. The transmitted symbol is then removed from the queue. We use θ to index the realization of the compounded randomness of codebook generation and switch selection. Let $C_\theta^{(k)}(\xi_k)_j$ be the j^{th} codeword symbol of the k^{th} inner code in codebook $\mathcal{C}_\theta^{(k)}$, corresponding to ξ_k. Let $Z_{l,\theta} \in \{1, \cdots, N_o\}$ be index of the queue that the random switch chooses at the l^{th} time unit for $l \in \{1, 2, \cdots\}$. We assume that index θ is generated according to a distribution ϑ such that random variables $x_{k,\xi_k,j} : \theta \to C_\theta^{(k)}(\xi_k)_j$ are i.i.d. with a pre-determined input distribution p_X, random variables $I_l : \theta \to Z_{l,\theta}$ are i.i.d. uniform, $x_{k,\xi_k,j}$ and I_l are independent. The decoder is assumed to know the outer codebook and the code libraries of the inner codes. We also assume that the decoder knows the exact codebook used for 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 codes. Assume that, for $k \in \{1, \cdots, N_o\}$, $z_k N_i$ channel output symbols are received from the k^{th} inner code, where $z_k > 0$ and $z_k N_i$ is an integer. We term z_k the "effective codeword length parameter" of the k^{th} inner code. By definition, we have $\sum_{k=1}^{N_o} z_k = N_o$. Based on z_k, and the received channel output symbols, $\{y_{ki_1}, y_{ki_2}, \ldots, y_{ki_{z_k N_i}}\}$, 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 that, given z_k and $\{y_{ki_1}, y_{ki_2}, \cdots, y_{ki_{z_k N_i}}\}$, reliability weight α_k is computed using Forney's algorithm presented in [22, Section 4.2]. With $\{\hat{\xi}_k\}$ and $\{\alpha_k\}$ for all k, the decoder then 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 [18].

Due to random codebook selection and random switching, without being conditioned on θ, error probabilities experienced by all messages are equal, i.e., $P_e(N)$ satisfies (3). Compared with a classical concatenated code 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, given in the following theorem, 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,

$$E_{Fc}(R) = \max_{p_X, \frac{R}{\mathcal{C}_F} \le r_o \le 1, 0 \le \rho \le 1} (1 - r_o) \left(-\rho \frac{R}{r_o} + E_0(\rho, p_X) \left[1 - \frac{1 + r_o}{2} E_0(\rho, p_X) \right] \right), \quad (6)$$

where $E_0(\rho, p_X)$ is defined in (5).

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 [22], and is lower-bounded by $\tilde{E}_{Fc}(R)$, defined by

$$\tilde{E}_{Fc}(R) = \max_{p_X, \frac{R}{\mathcal{C}_F} \le r_o \le 1, 0 \le \rho \le 1} (1 - r_o) \left(-\rho \frac{R}{r_o} + E_0(\rho, p_X) \left[1 - E_0(\rho, p_X) \right] \right). \quad (7)$$

The bounds are asymptotically tight in the sense that $\lim_{R \to \mathcal{C}_F} \frac{\tilde{E}_{Fc}(R)}{E_{Fc}(R)} = 1$. ■

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 can see that $E_{Fc}(R)$ is closely approximated by $\tilde{E}_{Fc}(R)$, especially at rates close to the fountain capacity.

Extending the one-level concatenated fountain codes to the multi-level concatenated fountain codes is essentially the same as in classical communication systems [23][18] except that random fountain codes are used as inner codes in a fountain system. For a positive integer m, the 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 ($m \in \{1, 2, \cdots\}$) concatenated fountain codes,

$$E_{Fc}^{(m)}(R) = \max_{p_X, \frac{R}{\mathcal{C}_F} \le r_o \le 1} \frac{\frac{R}{r_o} - R}{\frac{R}{r_o m} \sum_{i=1}^{m} \left[E_{FL} \left(\left(\frac{i}{m} \right) \frac{R}{r_o}, p_X \right) \right]^{-1}},$$
$$E_{FL}(x, p_X) = \max_{0 \le \rho \le 1} \left(-\rho x + E_0(\rho, p_X) \left[1 - E_0(\rho, p_X) \right] \right), \quad (8)$$

where $E_0(\rho, p_X)$ is defined in (5).

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

Theorem 3 can be proved by following the analysis of m-level concatenated codes presented in [23][25] and replacing the error exponent of code in each concatenation level with the corresponding 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,

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

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 $E_{Fc}^{(\infty)}(R)$ does not deviate significantly from the Blokh-Zyablov exponent, which is the error exponent upper bound for multi-level concatenated fountain codes.

V. Rate Compatible Fountain Communication

In this section, we consider the fountain communication where the receiver already has partial knowledge about the transmitted message. Take the application of software patch distribution as an example. 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 has some of the patches, he may only want to download the new patches, rather than the whole service pack. On one hand, for the convenience of the patch server, all patches of the service pack should be encoded jointly. On the other hand, 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 performance objective to be achieved simultaneously for all users, and define such a fountain communication model as the rate compatible fountain communication. We will show next that efficient rate compatible fountain communication can be achieved using a class of extended concatenated fountain codes with linear complexity.

Assume that a source message w, which takes $\lfloor \exp(NR) \rfloor$ possible values, is partitioned into L sub-messages $[w_1, w_2, \cdots, w_L]$, where w_i $(i \in \{1, \cdots, L\})$ can take $\lfloor \exp(Nr_i) \rfloor$ possible values with $\sum_i r_i = R$. Consider the following extended one-level concatenated fountain coding scheme. For each $i \in \{1, \cdots, L\}$, 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 $\{\xi_{i1}, \cdots, \xi_{iN_o}\}$, as illustrated in Figure 5. Next, for all $k \in \{1, \cdots, N_o\}$, the encoder combines outer codeword symbols $\{\xi_{1k}, \cdots, \xi_{Lk}\}$ into a macro symbol $\xi_k = [\xi_{1k}, \cdots, \xi_{Lk}]$. A random fountain code is then used to map ξ_k into an infinite channel input sequence $\{x_{k1}, x_{k2}, \cdots\}$.

Without loss of generality, we assume that there is only one decoder (receiver) and it already has sub-messages $\{w_{l+1}, \cdots, w_L\}$, 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[2]. From the decoder's point of view, since the unknown messages $[w_1, \cdots, w_l]$ can only take $\lfloor \exp(N \sum_{i=1}^{l} r_i) \rfloor$ 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 $\{w_{l+1}, \cdots, w_L\}$, the decoder first strikes out from fountain codebooks all codewords corresponding to the wrong messages. The extended one-level concatenated fountain code is then decoded using the same procedure as described in Section IV. Assume that 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 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 (6) can still be arbitrarily approached.

Therefore, given a rate partitioning $R = [r_1, \cdots, r_L]$, 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. If the system has multiple receivers with different priori sub-messages, the rate and error performance tradeoff as characterized in Theorem 2 can be achieved simultaneously for all receivers. Extending this scheme to the 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 discrete-time memoryless channel model $p_{Y|X}$. However, such an optimization may face various challenges in practical applications. For example, suppose that 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 needs to optimize fountain codes simultaneously for multiple channels. For another example, suppose that 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 full statistical model of the virtual channel 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. In this section, we show that, even when the channel state is unknown at the transmitter, it is still possible to achieve near optimal rate and error performance using concatenated fountain codes.

[2]Assume that N_l and N_l/N_o are both integers.

Consider fountain communication over a discrete-time memoryless channel $p_{Y|X}$ using one-level concatenated fountain codes. We assume that the channel is symmetric, and hence the optimal input distribution p_X is known at the transmitter. Other than channel alphabets and the symmetry property, we assume that 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 that the transmitter and the receiver agree on achieving a fountain information rate of $\gamma I(p_X)$ where $\gamma \in [0,1]$ 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 outer code rate r_o can be predetermined at the transmitter and the following error exponent can be arbitrarily approached,

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

Without $p_{Y|X}$ at the transmitter, the optimal r_o cannot be derived. However, with the knowledge of γ, we can set a suboptimal outer code rate by letting $r_o = \frac{\sqrt{\gamma^2+8\gamma}-\gamma}{2}$ 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 as $\gamma \to 1$.

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

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

■

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. Indeed, computer simulations suggest that such optimality conclusion applies to a wide range of channels over a wide range of fountain rates. However, further investigation on this issue is outside the scope of this paper.

VII. Conclusions

We extended linear-complexity concatenated codes to 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 decoders know part of the transmitted message, with the encoder still encoding the complete message, concatenated fountain codes can achieve the same rate and error performance as if only the novel part of the message is encoded for each individual user. For one-level concatenated codes and for certain 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 that 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 $\boldsymbol{z}$ be an N_o-dimensional vector whose k^{th} element z_k is the effective codeword length parameter of the k^{th} inner code, for $k \in \{1, \cdots, N_o\}$. Note that $\boldsymbol{z}$ is a random vector. Let $dz > 0$ be a small constant. We define $\{z_g | z_g = ndz, n = 0, 1, \ldots,\}$ as the set of "grid values" each can be written as an non-negative integer multiplying dz. Define a point mass function (PMF) $f_Z^{(dz)}$ as follows. We first quantize each element of $\boldsymbol{z}$, for example z_k, to the closest grid value no larger than z_k. Denote the quantized $\boldsymbol{z}$ vector by $\boldsymbol{z}^{(q)}$, whose elements are denoted by $z_i^{(q)}$ for $i \in \{1, \cdots, N_o\}$. For any grid value z_g, we define $\mathcal{I}_{z_g} = \left\{ i \left| z_i^{(q)} = z_g \right. \right\}$ as the set of indices corresponding to which the elements of $\boldsymbol{z}^{(q)}$ vector equal the particular z_g. Given $\boldsymbol{z}$, the empirical PMF $f_Z^{(dz)}$ is a function defined for the grid values, with $f_Z^{(dz)}(z_g) = \frac{|\mathcal{I}_{z_g}|}{N_o}$, where $|\mathcal{I}_{z_g}|$ is the cardinality of $\mathcal{I}_{z_g}$. Since $f_Z^{(dz)}$ is induced from random vector $\boldsymbol{z}$, itself is random. Let $Pr\left\{ f_Z^{(dz)} \right\}$ denote the probability that the received effective inner codeword length parameter vector $\boldsymbol{z}$ gives a particular PMF $f_Z^{(dz)}$.

Let us now consider a decoding algorithm, called "dz-decoder", which is the same as the one introduced in Section IV except that the decoder, after receiving $N_i z_k$ symbols for the k^{th} inner code (for all $k \in \{1, \cdots, N_o\}$), only uses the first $N_i z_k^{(q)}$ symbols to decode the inner code. Assume that the fountain information rate R, the outer code rate r_o, and the input distribution P_X are given. Due to symmetry, it is easy to see that, without being conditioned on random variable θ (defined in Section IV), different $\boldsymbol{z}$ vectors corresponding to the same $f_Z^{(dz)}$ (which is indeed induced from $\boldsymbol{z}^{(q)}$) give the same error probability performance. Let $P_e\left(f_Z^{(dz)}\right)$ be the communication error probability of the dz-decoder given $f_Z^{(dz)}$. Communication error probability P_e of the dz-decoder *without* given $f_Z^{(dz)}$ can be written as,

$$P_e = \sum_{f_Z^{(dz)}} P_e\left(f_Z^{(dz)}\right) Pr\left\{ f_Z^{(dz)} \right\}. \tag{13}$$

For a given $f_Z^{(dz)}$, define $E_f(f_Z^{(dz)}) = -\lim_{N_i\to\infty}\lim_{N_o\to\infty}\frac{1}{N_iN_o}\log P_e\left(f_Z^{(dz)}\right)$. Consequently, we can find a constant $K_0(N_i,N_o)$, such that the following inequality holds for all $f_Z^{(dz)}$ and all N_i, N_o,

$$P_e\left(f_Z^{(dz)}\right) \le K_0(N_i,N_o)\exp\left(-N_iN_oE_f\left(f_Z^{(dz)}\right)\right), \qquad \lim_{N_i\to\infty}\lim_{N_o\to\infty}\frac{\log K_0(N_i,N_o)}{N_iN_o} = 0. \tag{14}$$

Given dz, N_i, N_o, let $K_1(N_i,N_o)$ be the total number of possible quantized $\boldsymbol{z}^{(q)}$ vectors (the quantized vector of $\boldsymbol{z}$). $K_1(N_i,N_o)$ can be upper bounded by

$$K_1(N_i,N_o) \le 2^{N_o}\frac{\left(\left\lceil\frac{N_o}{dz}\right\rceil + N_o - 1\right)!}{\left(\left\lceil\frac{N_o}{dz}\right\rceil\right)!(N_o-1)!}. \tag{15}$$

In the above bound, the term $\frac{\left(\left\lceil\frac{N_o}{dz}\right\rceil + N_o - 1\right)!}{\left(\left\lceil\frac{N_o}{dz}\right\rceil\right)!(N_o-1)!}$ represents the total number of possible outcomes of assigning $\left\lceil\frac{N_o}{dz}\right\rceil$ identical balls to N_o distinctive boxes. This is the number of possible $\boldsymbol{z}^{(q)}$ vectors we can get if the received symbols are assigned to the inner codes in groups with N_idz (assumed to be an integer) symbols per group. Let us term the assumption of assigning received symbols in groups the "symbol-grouping" assumption. To relax the symbol-grouping assumption, we note that, if the number of symbols obtained by an inner code, say the k^{th} inner code, is a little less that an integer multiplication of N_idz, then the quantization value $z_k^{(q)}$ obtained without the symbol-grouping assumption can be one unit less than the corresponding value with the symbol-grouping assumption. Therefore, the total number of possible $\boldsymbol{z}^{(q)}$ vectors we can get without the symbol-grouping assumption is upper bounded by 2^{N_o} multiplying the corresponding number with the symbol-grouping assumption. Note that, given dz, the right hand side of (15) is not a function of N_i, and it is also an upper bound on the total number of possible $f_Z^{(dz)}$ functions.

Due to Stirling's approximation [26], (15) implies that $\lim_{N_o\to\infty}\frac{\log K_1(N_i,N_o)}{N_o} < \infty$, and hence

$$\lim_{N_i\to\infty}\lim_{N_o\to\infty}\frac{\log K_1(N_i,N_o)}{N_iN_o} = 0. \tag{16}$$

Combining (13), (14) and (16), the error exponent of a dz-decoder is given by

$$E_{Fc} = -\lim_{N_i\to\infty}\lim_{N_o\to\infty}\frac{\log P_e}{N_iN_o} = \min_{f_Z^{(dz)}}\left\{E_f\left(f_Z^{(dz)}\right) - \lim_{N_i\to\infty}\lim_{N_o\to\infty}\frac{1}{N_iN_o}\log Pr\left\{f_Z^{(dz)}\right\}\right\}. \tag{17}$$

The rest of the proof contains four parts. In Part I, the expression of $\lim_{N_i\to\infty}\lim_{N_o\to\infty}\frac{1}{N_iN_o}\log Pr\left\{f_Z^{(dz)}\right\}$ is derived. In Part II, we derive the expression of $E_f\left(f_Z^{(dz)}\right)$. In Part III, we use the results of the first two parts to obtain $\lim_{dz\to 0}E_{Fc}$. Complexity and the achievable error exponent of the concatenated fountain code is obtained based on the derived results in Part IV.

Part I: Let $\boldsymbol{z}(i)$ (for all $i\in\{1,\cdots,N_o\}$) 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 $\boldsymbol{z}(i)$ equal 1 and let all other elements equal 0. Since the random switch (illustrated in Figure 2) picks inner codes uniformly, we have

$$E[\boldsymbol{z}(i)] = \frac{1}{N_o}\mathbf{1}, \quad \text{cov}[\boldsymbol{z}(i)] = \frac{1}{N_o}\boldsymbol{I}_{N_o} - \frac{1}{{N_o}^2}\mathbf{1}\mathbf{1}^T, \tag{18}$$

where $\mathbf{1}$ is an N_o-dimensional vector with all elements being one, and $\boldsymbol{I}_{N_o}$ is the identity matrix of size N_o. According to the definitions, we have $\boldsymbol{z} = \frac{1}{N_i}\sum_{i=1}^{N_iN_o}\boldsymbol{z}(i)$. Since the total number of received symbols equal N_iN_o, we must have $\mathbf{1}^T\boldsymbol{z} = N_o$.

Let $\boldsymbol{\omega}$ be a real-valued N_o-dimensional vector whose entries satisfy $-\pi\sqrt{N_iN_o} \le \omega_k < \pi\sqrt{N_iN_o}, \forall k \in \{1, \cdots, N_o\}$. Since $\boldsymbol{z}$ equals the normalized summation of N_iN_o independently distributed vectors $\boldsymbol{z}(i)$, the characteristic function of $\sqrt{\frac{N_i}{N_o}}(\boldsymbol{z}-\mathbf{1})$, denoted by $\varphi_Z(\boldsymbol{\omega}) = E\left[\exp\left(j\sqrt{\frac{N_i}{N_o}}\boldsymbol{\omega}^T(\boldsymbol{z}-\mathbf{1})\right)\right]$, can therefore be written as

$$\begin{aligned}\varphi_Z(\boldsymbol{\omega}) &= E\left[\exp\left(j\sqrt{\frac{N_i}{N_o}}\boldsymbol{\omega}^T(\boldsymbol{z}-\mathbf{1})\right)\right] = \prod_{i=1}^{N_iN_o} E\left[\exp\left(j\sqrt{\frac{1}{N_iN_o}}\boldsymbol{\omega}^T(\boldsymbol{z}(i)-\frac{1}{N_o}\mathbf{1})\right)\right] \\ &= \left\{E\left[\exp\left(j\sqrt{\frac{1}{N_iN_o}}\boldsymbol{\omega}^T(\boldsymbol{z}(i)-\frac{1}{N_o}\mathbf{1})\right)\right]\right\}^{N_iN_o} = \left[1-\frac{1}{2}\frac{\|\boldsymbol{Q}^T\boldsymbol{\omega}\|^2}{N_o^2N_i} + o\left(\frac{\|\boldsymbol{Q}^T\boldsymbol{\omega}\|^2}{N_o^2N_i}\right)\right]^{N_oN_i}, \end{aligned}\tag{19}$$

where in the last equality, $\boldsymbol{Q}$ is a real-valued $N_o \times (N_o-1)$-dimensional matrix satisfying $\boldsymbol{Q}^T\boldsymbol{Q} = \boldsymbol{I}_{N_o-1}$ and $\boldsymbol{Q}^T\mathbf{1} = \mathbf{0}$, which imply $\boldsymbol{Q}\boldsymbol{Q}^T = \boldsymbol{I}_{N_o} - \frac{1}{N_o}\mathbf{1}\mathbf{1}^T$. In other words, $\|\boldsymbol{Q}^T\boldsymbol{\omega}\|^2 = \boldsymbol{\omega}^T(\boldsymbol{I}_{N_o} - \frac{1}{N_o}\mathbf{1}\mathbf{1}^T)\boldsymbol{\omega}$.

Note that, since $\boldsymbol{z}$ is discrete-valued, $\varphi_Z(\boldsymbol{\omega})$ is similar to a multi-dimensional discrete-time Fourier transform of the PMF of $\sqrt{\frac{N_i}{N_o}}(\boldsymbol{z}-\mathbf{1})$. Because $\sqrt{N_iN_o}\left[\sqrt{\frac{N_i}{N_o}}(\boldsymbol{z}-\mathbf{1})\right] = \sum_{i=1}^{N_iN_o}\boldsymbol{z}(i) - N_i\mathbf{1}$ takes integer-valued entries, the $\varphi_Z(\boldsymbol{\omega})$ function is periodic $\boldsymbol{\omega}$ in the sense that $\varphi_Z\left(\boldsymbol{\omega} + 2\pi\sqrt{N_iN_o}\boldsymbol{e}_k\right) = \varphi_Z(\boldsymbol{\omega})$, $k \in \{1, \cdots, N_o\}$, where $\boldsymbol{e}_k$ is an N_o-dimensional vector whose k^{th} entry is one and all other entries are zeros. This is why we can focus on "frequency" vector $\boldsymbol{\omega}$ with $-\pi\sqrt{N_iN_o} \le \omega_k < \pi\sqrt{N_iN_o}, \forall k \in \{1, \cdots, N_o\}$.

Equation (19) implies that

$$\lim_{N_o\to\infty}\left\{\varphi_Z(\boldsymbol{\omega}) - \exp\left(-\frac{1}{2N_o}\boldsymbol{\omega}^T\boldsymbol{Q}\boldsymbol{Q}^T\boldsymbol{\omega}\right)\right\} = 0. \tag{20}$$

Therefore, with large enough N_o and for any $\boldsymbol{z}$, the probability $Pr\{\boldsymbol{z}\}$ is upper-bounded by

$$Pr\{\boldsymbol{z}\} \quad \le \left(\frac{1}{2\pi\sqrt{N_iN_o}}\right)^{N_o}\left(\frac{N_o}{2\pi}\right)^{\frac{N_o-1}{2}}\exp\left(-\frac{N_i}{2}\left[\|\boldsymbol{z}-\mathbf{1}\|^2 - dz\|\mathbf{1}\|^2\right]\right), \tag{21}$$

where the constant $2\pi\sqrt{N_iN_o}$ in the denominator of the first term on the right hand side of (21) is due to the range of $-\pi\sqrt{N_iN_o} \le \omega_k < \pi\sqrt{N_iN_o}, \forall k \in \{1, \cdots, N_o\}$. The constant $dz\|\mathbf{1}\|^2$ in the exponent of (21) is added to ensure the existence of a large enough N_o to satisfy the inequality, as implied by (20). Inequality (21) further implies that

$$Pr\{\boldsymbol{z}\} \quad \le \left(\frac{1}{2\pi\sqrt{N_iN_o}}\right)^{N_o}\left(\frac{N_o}{2\pi}\right)^{\frac{N_o-1}{2}}\exp\left(-\frac{N_i}{2}\left[\|\boldsymbol{z}^{(q)}-\mathbf{1}\|^2 - 3dz\|\mathbf{1}\|^2\right]\right), \tag{22}$$

where $\boldsymbol{z}^{(q)}$ is the quantized version of $\boldsymbol{z}$. Consequently, the probability of $\boldsymbol{z}^{(q)}$ is upper-bounded by

$$Pr\left\{\boldsymbol{z}^{(q)}\right\} \quad \leq \lceil N_i dz\rceil^{N_o}\left(\frac{1}{2\pi\sqrt{N_iN_o}}\right)^{N_o}\left(\frac{N_o}{2\pi}\right)^{\frac{N_o-1}{2}}\exp\left(-\frac{N_i}{2}\left[\|\boldsymbol{z}^{(q)}-\mathbf{1}\|^2-3N_odz\right]\right). \quad (23)$$

The probability of any PMF $f_Z^{(dz)}$ is upper-bounded by

$$\begin{aligned} Pr\left\{f_Z^{(dz)}\right\} &\leq K_1(N_i,N_o)Pr\left\{\boldsymbol{z}^{(q)}\right\} \\ &\leq K_1(N_i,N_o)\lceil N_i dz\rceil^{N_o}\left(\frac{1}{2\pi\sqrt{N_iN_o}}\right)^{N_o}\left(\frac{N_o}{2\pi}\right)^{\frac{N_o-1}{2}}\exp\left(-\frac{N_i}{2}\left[\|\boldsymbol{z}^{(q)}-\mathbf{1}\|^2-3N_odz\right]\right), \quad (24)\end{aligned}$$

where $K_1(N_i, N_o)$ is the total number of possible $\boldsymbol{z}^{(q)}$ vectors satisfying (16).

From (24), we can see that for all $f_Z^{(dz)}$ the following inequality holds,

$$-\lim_{N_i\to\infty}\lim_{N_o\to\infty}\frac{\log Pr\left\{f_Z^{(dz)}\right\}}{N_iN_o} \geq \frac{1}{2}\sum_{z_g}\left[(z_g-1)^2-3dz\right]f_Z^{(dz)}(z_g), \quad (25)$$

where $f_Z^{(dz)}(z_g)$ is the value of PMF $f_Z^{(dz)}$ at z_g.

Note that, because $\mathbf{1}^T\boldsymbol{z} = N_o$, for all empirical PMFs $f_Z^{(dz)}$, we have $\sum_{z_g} z_g f_Z^{(dz)}(z_g) \in [1-dz, 1]$.

Part II: Next, we will derive the expression of $E_f\left(f_Z^{(dz)}\right)$, which is the error exponent conditioned on an empirical PMF $f_Z^{(dz)}$.

Let $\boldsymbol{z}$ be a particular N_o-dimensional effective inner codeword length parameter vector following the empirical PMF $f_Z^{(dz)}$, under a given dz. Let $P_e(\boldsymbol{z})$ be the error probability given $\boldsymbol{z}$ (or $\boldsymbol{z}^{(q)}$). Let $P_e(f_Z^{(dz)})$ be the error probability given $f_Z^{(dz)}$. From the definition of the concatenated fountain codes, we can see that the inner codes are logically equivalent, so do the codeword symbols of the near MDS outer code. In other words, error probabilities corresponding to all $\boldsymbol{z}$ vectors with the same PMF $f_Z^{(dz)}$ are equal. This consequently implies that $P_e(\boldsymbol{z}) = P_e(f_Z^{(dz)})$. Therefore, when bounding $E_f\left(f_Z^{(dz)}\right)$, instead of assuming a particular $f_Z^{(dz)}$ which corresponds to multiple $\boldsymbol{z}$ vectors, we can assume a single $\boldsymbol{z}$ vector whose corresponding empirical PMF is $f_Z^{(dz)}$.

Assume that 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 satisfying $\lim_{N_i\to\infty}\lim_{N_o\to\infty}\zeta_0 = 0$. An example of such near MDS code was introduced in [17]. Assume that, for all k, the k^{th} outer codeword symbol is ξ_k, and the k^{th} inner code reports an estimate of the outer symbol $\hat{\xi}_k$ together with a reliability weight $\alpha_k \in [0,1]$. Applying Forney's GMD decoding to the outer code [18], the source message can be recovered if the following inequality holds [22, Theorem 3.1b],

$$\sum_{k=1}^{N_o}\alpha_k\mu_k > (r_o+\zeta_0)N_o, \quad (26)$$

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 $\boldsymbol{z}$ vector is bounded by

$$P_e(f_Z^{(dz)}) = P_e(\boldsymbol{z}) \leq Pr\left\{\sum_{k=1}^{N_o} \alpha_k\mu_k \leq (r_o+\zeta_0)N_o\right\} \leq \min_{s\geq 0} \frac{E\left[\exp\left(-sN_i\sum_{k=1}^{N_o}\alpha_k\mu_k\right)\right]}{\exp(-sN_i(r_o+\zeta_0)N_o)}, \quad (27)$$

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

Given the effective inner codeword length parameter vector $\boldsymbol{z}$, random variables $\alpha_k\mu_k$ for different inner codes are independent. Therefore, (27) can be further written as

$$P_e(f_Z^{(dz)}) = P_e(\boldsymbol{z}) \leq \min_{s\geq 0} \frac{\prod_{k=1}^{N_o} E\left[\exp\left(-sN_i\alpha_k\mu_k\right)\right]}{\exp(-sN_i(r_o+\zeta_0)N_o)} = \min_{s\geq 0} \frac{\exp\left(\sum_{k=1}^{N_o}\log E\left[\exp\left(-sN_i\alpha_k\mu_k\right)\right]\right)}{\exp(-sN_i(r_o+\zeta_0)N_o)} \quad (28)$$

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

Assume that the effective codeword length parameter is z_k. Given z_k, whose quantized value is $z_k^{(q)}$, depending on the received channel symbols, the decoder generates the maximum likelihood outer code estimate $\hat{\xi}_k$, and generates α_k using Forney's algorithm presented in [22, Section 4.2]. Define an adjusted error exponent function $E_z(z)$ as follows.

$$E_z(z) = \max_{0\leq\rho\leq 1} -\rho\frac{R}{r_o} + zE_0(\rho, p_X), \quad (29)$$

where $E_0(\rho, p_X)$ is defined in (5). By following Forney's error exponent analysis presented in [22, Section 4.2], we obtain

$$-\log E\left[\exp\left(-sN_i\alpha_k\mu_k\right)\right] \geq \max\left\{\min\{N_iE_z\left(z_k^{(q)}\right), N_i\left(2E_z\left(z_k^{(q)}\right)-s\right), N_is\}, 0\right\} + K_2(N_i, N_o), \quad (30)$$

where $K_2(N_i, N_o)$ is a constant satisfying $\lim_{N_i\to\infty}\lim_{N_o\to\infty}\frac{K_2(N_i,N_o)}{N_iN_o} = 0$.

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

$$\phi(z,s) = \begin{cases} -sr_o & z, E_z(z) < s/2 \\ 2E_z(z)-(1+r_o)s & z, s/2 \leq E_z(z) < s \\ (1-r_o)s & z, E_z(z) \geq s \end{cases}. \quad (31)$$

Substitute (30) into (28), and take N_i, N_o to infinity (which implies $\zeta_0 \to 0$), we get the following bound on the conditional error exponent $E_f\left(f_Z^{(dz)}\right)$,

$$E_f\left(f_Z^{(dz)}\right) \geq \max_{s\geq 0}\sum_{z_g}\phi(z_g, s)f_Z^{(dz)}(z_g). \quad (32)$$

Part III: According to (17), (25) and (32), we have

$$\begin{aligned} E_{Fc} &\geq \min_{f_Z^{(dz)}, \sum_{z_g} z_g f_Z^{(dz)}(z_g)\in[1-dz,1]} \left\{E_f(f_Z^{(dz)}) + \sum_{z_g}\frac{(z_g-1)^2}{2}f_Z^{(dz)}(z_g)\right\} - \frac{3}{2}dz \\ &\geq \min_{f_Z^{(dz)}, \sum_{z_g} z_g f_Z^{(dz)}(z_g)\in[1-dz,1]} \max_{s\geq 0}\sum_{z_g}\left(\phi(z_g,s) + \frac{(z_g-1)^2}{2}\right)f_Z^{(dz)}(z_g) - \frac{3}{2}dz. \quad (33) \end{aligned}$$

Define $E_{Fc}^{(0)} = \lim_{dz\to 0} E_{Fc}$. Let f_Z be a probability density function defined for $z \in [0,\infty)$. Inequality (33) implies that

$$\begin{aligned} E_{Fc}^{(0)} &\geq \min_{f_Z, \int_0^\infty z f_Z(z)dz=1} \max_{s\geq 0} \int_0^\infty \left(\phi(z,s) + \frac{(z-1)^2}{2}\right) f_Z(z)dz \\ &= \max_{s\geq 0} \min_{f_Z, \int_0^\infty z f_Z(z)dz=1} \int_0^\infty \left(\phi(z,s) + \frac{(z-1)^2}{2}\right) f_Z(z)dz. \end{aligned} \quad (34)$$

Assume that f_Z^* is the density function minimizing the last term in (34). If we can find $0 < \lambda < 1$, and two density functions $f_Z^{(1)}$, $f_Z^{(2)}$ with $\int_0^\infty z f_Z^{(1)}(z)dz = 1$, $\int_0^\infty z f_Z^{(2)}(z)dz = 1$, such that

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

then it is easy to show that the last term in (34) must be minimized either by $f_Z^{(1)}$ or $f_Z^{(2)}$. Since this contradicts the assumption that f_Z^* is optimum, a nontrivial decomposition like (35) must not be possible. Consequently, 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.

Let us fix γ first, and consider the following lower bound on $E_{Fc}^{(0)}(\gamma)$, which is obtained by substituting (36) into (34),

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

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 the right hand side of (37) should satisfy either $s^* = E_z(z_0)$ 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

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

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

$$E_{Fc}^{(0)} \geq \min_{0\leq z_0,\gamma\leq 1}\left[2\gamma E_z(z_0) + \frac{\gamma}{1-\gamma}\frac{(1-z_0)^2}{2} + (1-r_o-2\gamma)E_z\left(\frac{1-\gamma z_0}{1-\gamma}\right)\right]. \quad (39)$$

By substituting $E_z(z) = \max_{0\leq\rho\leq 1}[-\rho\frac{R}{r_o} + zE_0(\rho, p_X)]$ into (39), we get

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

Note that if $(1+r_o)(1-z_0)E_0(\rho, p_X) - \frac{(1-z_0)^2}{2} < 0$, then $E_{Fc}^{(0)} \geq (1-r_o)\left[-\rho\frac{R}{r_o} + E_0(\rho, p_X)\right]$ with the right hand side of the inequality equaling Forney's exponent for given p_X and 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 [22]. Therefore, we must have $(1+r_o)(1-z_0)E_0(\rho, p_X) - \frac{(1-z_0)^2}{2} \geq 0$. Consequently, the right hand sides of both (38) and (40) are minimized at the margin of $\gamma^* = \frac{1-r_o}{2}$. This gives

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

Note that if ρ is chosen to satisfy $(1+r_o)E_0(\rho, p_X) \geq 1$, the last term in (41) is minimized at $z_0^* = 0$, which gives

$$E_{Fc}^{(0)} \geq \max_{0\leq\rho\leq 1}\left\{-\rho\frac{R}{r_o}(1-r_o) + \frac{1-r_o}{1+r_o}\right\}. \tag{42}$$

The right hand side of (42) is maximized at $\rho^* = 0$. However, $\rho = 0$ implies $(1+r_o)E_0(\rho, p_X) = 0 < 1$ which contradicts the assumption $(1+r_o)E_0(\rho, p_X) \geq 1$. Therefore, we can assume that $(1+r_o)E_0(\rho, p_X) \leq 1$. Consequently, the last term in (41) is minimized at $z_0^* = 1-(1+r_o)E_0$. This gives

$$E_{Fc}^{(0)} \geq \max_{0\leq\rho\leq 1}(1-r_o)\left(-\rho\frac{R}{r_o} + E_0(\rho, p_X)\left[1 - \frac{1+r_o}{2}E_0(\rho, p_X)\right]\right). \tag{43}$$

By optimizing (43) over p_X and r_o, it can be seen that the error exponent given in (6) is achievable if we first take N_o to infinity and then take N_i to infinity.

Part IV: To achieve linear coding complexity, let us assume that N_i is fixed at a large constant while N_o is taken to infinity. According to [17], it is easy to see that the encoding complexity is linear in the number of transmitted symbols[3]. At the receiver, we keep at most $2N_i$ symbols for each inner code and drop the extra received symbols. Consequently, the effective codeword length parameter 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 bound takes the form $f_Z(z) = \frac{1-r_o}{2}\delta(z-z_0) + \frac{1+r_o}{2}\delta\left(z - \frac{2-z_0(1-r_o)}{1+r_o}\right)$, with $z_0, \frac{2-z_0(1-r_o)}{1+r_o} < 2$. Therefore, upper bounding the effective codeword length parameter 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 in the order of $O(\exp(2N_i))$. According to [18], 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 ζ_1 can be made arbitrarily small as we increase N_i, we conclude that fountain error exponent $E_{Fc}(R)$ given in (6) can be *arbitrarily approached* by one-level concatenated fountain codes with a linear coding complexity. ■

[3]In other words, we assume that no encoding complexity is spent on codeword symbols that are not transmitted.

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 $\lim_{R\to\mathcal{C}_F} \frac{\tilde{E}_{F_c}(R)}{E_{F_c}(R)} = 1$.

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

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

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

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

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 $\beta(i,j) = 0$ if $i = 0$ or $j = 0$. We also have

$$\begin{aligned}
\beta(1,1) &= \left\{\frac{R}{r_o^2} - \frac{\partial E_0(\rho, p_X)}{\partial \rho} + 2r_o E_0(\rho, p_X)\frac{\partial E_0(\rho, p_X)}{\partial \rho}\right\}\Bigg|_{r_o=1,\rho=0} = R - \mathcal{C}_F, \\
\beta(2,1) &= -2R \neq 0, \\
\beta(1,2) &= -\left\{\frac{\partial^2 E_0(\rho, p_X)}{\partial \rho^2} - 2\left(\frac{\partial E_0(\rho, p_X)}{\partial \rho}\right)^2\right\}\Bigg|_{\rho=0} \neq 0.
\end{aligned} \tag{46}$$

Similarly, define $\tilde{g}(p_X, r_o, \rho) = (1 - r_o)\left(-\rho\frac{R}{r_o} + E_0(\rho, p_X)\left[1 - E_0(\rho, p_X)\right]\right)$, such that

$$\tilde{E}_{F_c}(R) = \max_{p_X, \frac{R}{\mathcal{C}_F} \leq r_o \leq 1, 0\leq\rho\leq 1} \tilde{g}(p_X, r_o, \rho). \tag{47}$$

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

$$\tilde{g}(p_X, r_o, \rho) = \sum_{i,j} \frac{1}{(i+j)!}\tilde{\beta}(i,j)(r_o - 1)^i \rho^j. \tag{48}$$

where $\tilde{\beta}(i,j) = \frac{\partial^{(i+j)} \tilde{g}(p_X, r_o, \rho)}{\partial r_o^i \partial \rho^j}\Big|_{r_o=1,\rho=0}$. Similarly, we have $\tilde{\beta}(i,j) = 0$ if $i = 0$ or $j = 0$ and $\tilde{\beta}(1,1) = \beta(1,1) = R - \mathcal{C}_F$, $\tilde{\beta}(2,1) = \beta(2,1) \neq 0$, $\tilde{\beta}(1,2) = \beta(1,2) \neq 0$.

By L'Hospital's rule, the following equality holds,

$$\lim_{R\to\mathcal{C}_F} \frac{\tilde{E}_{Fc}(R)}{E_{Fc}(R)} = \lim_{R\to\mathcal{C}_F, r_o\to 1, \rho\to 0} \frac{\frac{1}{2}\tilde{\beta}(1,1)(r_o - 1)\rho + \frac{1}{6}\tilde{\beta}(2,1)(r_o - 1)^2\rho + \frac{1}{6}\tilde{\beta}(1,2)(r_o - 1)\rho^2}{\frac{1}{2}\beta(1,1)(r_o - 1)\rho + \frac{1}{6}\beta(2,1)(r_o - 1)^2\rho + \frac{1}{6}\beta(1,2)(r_o - 1)\rho^2} = 1. \tag{49}$$

■

C. Proof of Theorem 4

Proof: Define

$$\hat{g}(\gamma, r_o, \rho) = (1 - r_o)\left(\rho I(p_X)\left(1 - \frac{\gamma}{r_o}\right) + \frac{\rho^2}{2}\left(\left.\frac{\partial^2 E_0(\rho, p_X)}{\partial \rho^2}\right|_{\rho=0} - 2I^2(p_X)\right)\right),$$
$$\hat{E}_{Fc}(\gamma, p_X, r_o) = \max_{0\leq\rho\leq 1} \hat{g}(\gamma, r_o, \rho),$$
$$\hat{E}_{Fc}(\gamma, p_X) = \max_{0\leq r_o\leq 1} \hat{E}_{Fc}(\gamma, p_X, r_o). \quad (50)$$

We will first prove that

$$\lim_{\gamma\to 1} \frac{E_{Fcs}(\gamma, p_X)}{\hat{E}_{Fc}(\gamma, p_X)} = 1. \quad (51)$$

Note that $\hat{g}(\gamma, r_o, \rho)$ is maximized at $\rho^* = \frac{I(p_X)\left(1-\frac{\gamma}{r_o}\right)}{-\left.\frac{\partial^2 E_0(\rho,p_X)}{\partial\rho^2}\right|_{\rho=0} + 2I^2(p_X)}$, where we have assumed that $0 \leq \rho^* \leq 1$. This assumption is valid when r_o is also optimized. Consequently, $\hat{E}_{Fc}(\gamma, p_X, r_o)$ is maximized at $r_o^* = \arg\max_{0\leq r_o\leq 1}(1 - r_o)\left(1 - \frac{\gamma}{r_o}\right)^2 = \frac{\sqrt{\gamma^2+8\gamma}-\gamma}{2}$. Therefore,

$$\lim_{\gamma\to 1} \frac{E_{Fcs}(\gamma, p_X)}{\hat{E}_{Fc}(\gamma, p_X)} \geq \lim_{\gamma\to 1}\left[\left.\frac{E_{Fcs}(\gamma, p_X, \rho)}{\hat{g}(\gamma, p_X, \rho, r_o)}\right|_{\rho=\rho^*, r_o=r_o^*}\right] = 1. \quad (52)$$

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

$$\lim_{\gamma\to 1} \frac{\hat{E}_{Fc}(\gamma, p_X)}{E_{Fc}(\gamma, p_X)} = 1. \quad (53)$$

Combining (52) and (53), we get

$$\lim_{\gamma\to 1} \frac{E_{Fcs}(\gamma, p_X)}{E_{Fc}(\gamma, p_X)} = \lim_{\gamma\to 1} \frac{E_{Fcs}(\gamma, p_X)}{\hat{E}_{Fc}(\gamma, p_X)} \lim_{\gamma\to 1} \frac{\hat{E}_{Fc}(\gamma, p_X)}{E_{Fc}(\gamma, p_X)} \geq 1. \quad (54)$$

Because $E_{Fcs}(\gamma, p_X) \leq E_{Fc}(\gamma, p_X)$, (54) implies $\lim_{\gamma\to 1} \frac{E_{Fcs}(\gamma,p_X)}{E_{Fc}(\gamma,p_X)} = 1$.

■

REFERENCES

[1] Z. Wang and J. Luo, *Concatenated Fountain Codes,* IEEE ISIT'09, Seoul, Korea, Jun. 2009.

[2] J. Byers, M. Luby and A. Rege, *A Digital Fountain Approach to Reliable Distribution of Bulk Data,* ACM SIGCOMM'98, Vancouver, Canada, Sep. 1998.

[3] D. Mackay, *Fountain Codes,* IEE Proc. Commun., Vol. 152, pp. 1062-1068, Dec. 2005.

[4] M. Luby, *LT codes*, IEEE FOCS'02, Vancouver, Canada, Nov. 2002.

[5] A. Shokrollahi, *Raptor Codes,* IEEE Trans. Infom. Theory, Vol. 52, pp. 2551-2567, Jun. 2006.

[6] O. Etesami and A. Shokrollahi, *Raptor Codes on Binary Memoryless Symmetric Channels*, IEEE Trans. Inform. Theory, Vol. 52, pp. 2033-2051, May 2006.

[7] S. Shamai, I. Teletar and S. Verdú, *Fountain Capacity*, IEEE Trans. Inform. Theory, Vol. 53, pp. 4327-4376, Nov. 2007.

[8] X. Liu and T. Lim, *Fountain Codes over Fading Relay Channels*, IEEE Trans. Wireless Commun., Vol. 8, pp. 3278-3287, Jun. 2009.

[9] T. Jiang and X. Li, *Using Fountain Codes to Control the Peak-to-Average Power Ratio of OFDM Signals,* IEEE Trans. Veh. Tech., Vol. 59, pp. 3779-3785, Oct. 2010.

[10] D. Vukobratovic, V. Stankovic, D. Sejdinovic, L. Stankovic, and Z. Xiong, *Scalable Video Multicast Using Expanding Window Fountain Codes,* IEEE Trans. Multimedia, Vol. 11, pp. 1094-1104, Oct. 2009.

[11] S. Kokalj-Filipovic, P. Spasojevic, E. Soljanin, *Doped Fountain Coding for Minimum Delay Data Collection in Circular Networks,* IEEE J. Sel. A. Commun., Vol. 27, pp. 673-684, Jun. 2009.

[12] A. Oka and L. Lampe, *Data Extraction from Wireless Sensor Networks using Distributed Fountain Codes,* IEEE Trans. Commun., Vol. 57, pp. 2607 - 2618, Sep. 2009.

[13] S. Ahmad, R. Hamzaoui, M. Al-Akaidi, *Unequal Error Protection Using Fountain Codes With Applications to Video Communication,* IEEE Trans. Multimedia, Vol. 13, pp. 92-101, Feb. 2011.

[14] D. Sejdinovic, D. Vukobratovic, A. Doufexi, V. Senk, and R. Piechocki, *Expanding Window Fountain Codes for Unequal Error Protection,* IEEE Tran. Commun., Vol. 57, pp. 2510-2516, Sep. 2009.

[15] S. Chieochan, E. Hossain, *Wireless Fountain Coding with IEEE 802.11e Block ACK for Media Streaming in Wireline-cum-WiFi Networks: A Performance Study,* IEEE Trans. Mobile Comput., Vol. 10, pp. 1416-1433, Oct. 2011.

[16] H. Nguyen, T. Le-Nam, and H. Een-Kee, *On Transmission Efficiency for Wireless Broadcast Using Network Coding and Fountain Codes*, IEEE Commun. Lett., Vol. 15, pp. 569-571, May 2011.

[17] 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.

[18] Z. Wang and J. Luo, *Approaching Blokh-Zyablov Error Exponent with Linear-Time Encodable/Decodable Codes*, IEEE Communications Letters, Vol. 13, pp. 438-440, Jun. 2009.

[19] A. Feinstein, *Error Bounds in Noisy Channels Without Memory,* IEEE Trans. Inform. Theory, Vol. 1, pp. 13-14, Sep. 1955.

[20] R. Gallager, *A Simple Derivation of The Coding Theorem and Some Applications*, IEEE Trans. Inform. Theory, Vol. 11, pp. 3-18, Jan. 1965.

[21] 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.

[22] G. Forney, *Concatenated Codes*, The MIT Press, 1966.

[23] E. Blokh and V. Zyablov, *Linear Concatenated Codes*, Nauka, Moscow, 1982 (In Russian).

[24] A. Barg, G. Forney, *Random Codes: Minimum Distances and Error Exponents*, IEEE Trans. Inform. Theory Vol. 48, pp. 2568-2573, Sep. 2002.

[25] 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.

[26] D. Romik, "Stirling's Approximation for n!: The Ultimate Short Proof," The American Mathematical Monthly, Vol. 107, pp. 556-557, Jun.-Jul. 2000.

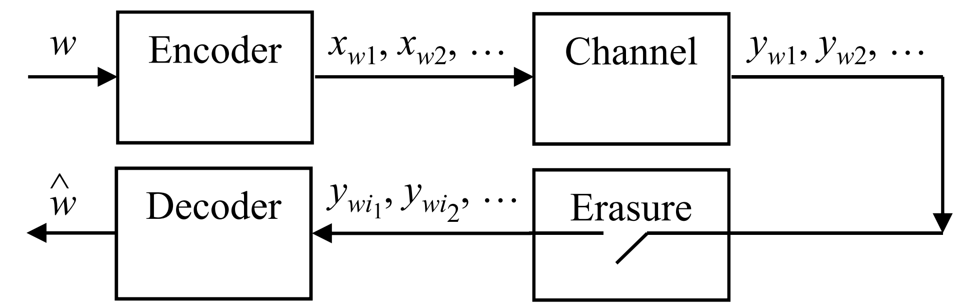

Fig. 1. Fountain communication over a memoryless channel.

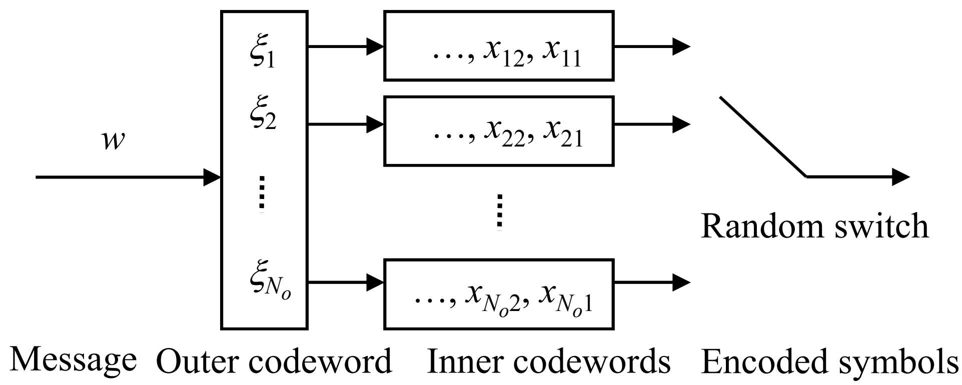

Fig. 2. One-level concatenated fountain codes.

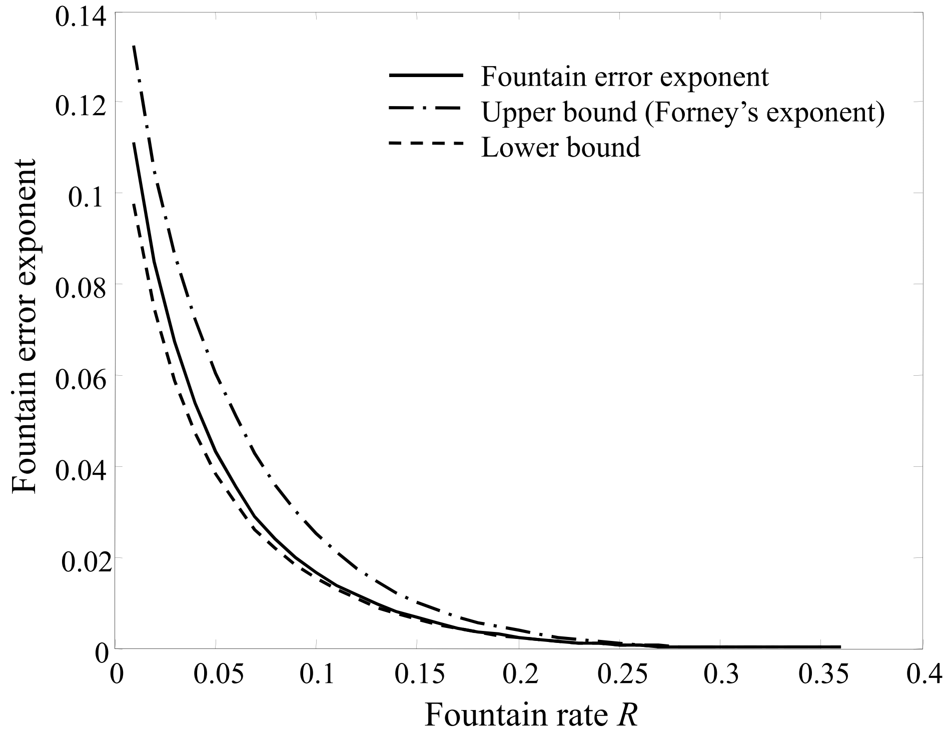

Fig. 3. Comparison of fountain error exponent $E_{Fc}(R)$, its upper bound $E_c(R)$, and its lower bound $\tilde{E}_{Fc}(R)$.

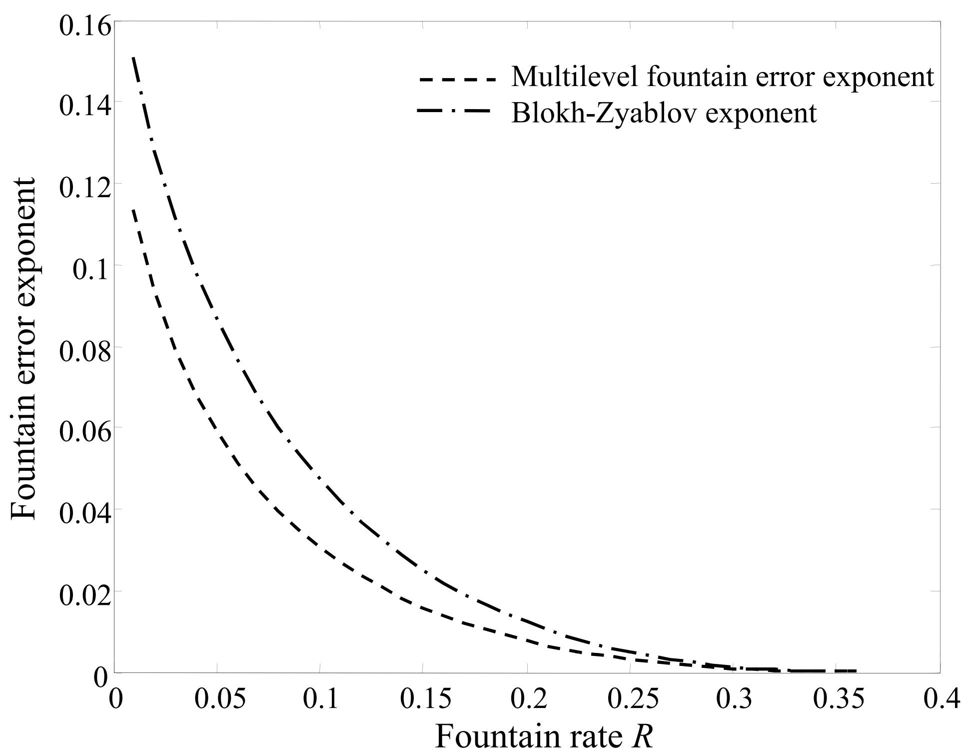

Fig. 4. Comparison of mulit-level fountain error exponent $E_{Fc}^{(\infty)}(R)$ and the Blokh-Zyablov exponent $E_c^{(\infty)}(R)$.

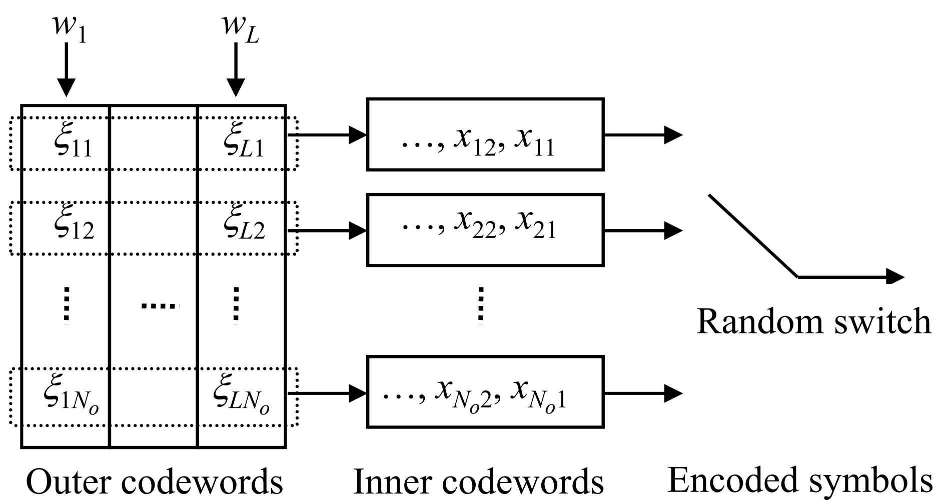

Fig. 5. Concatenated fountain codes for rate compatible communication.

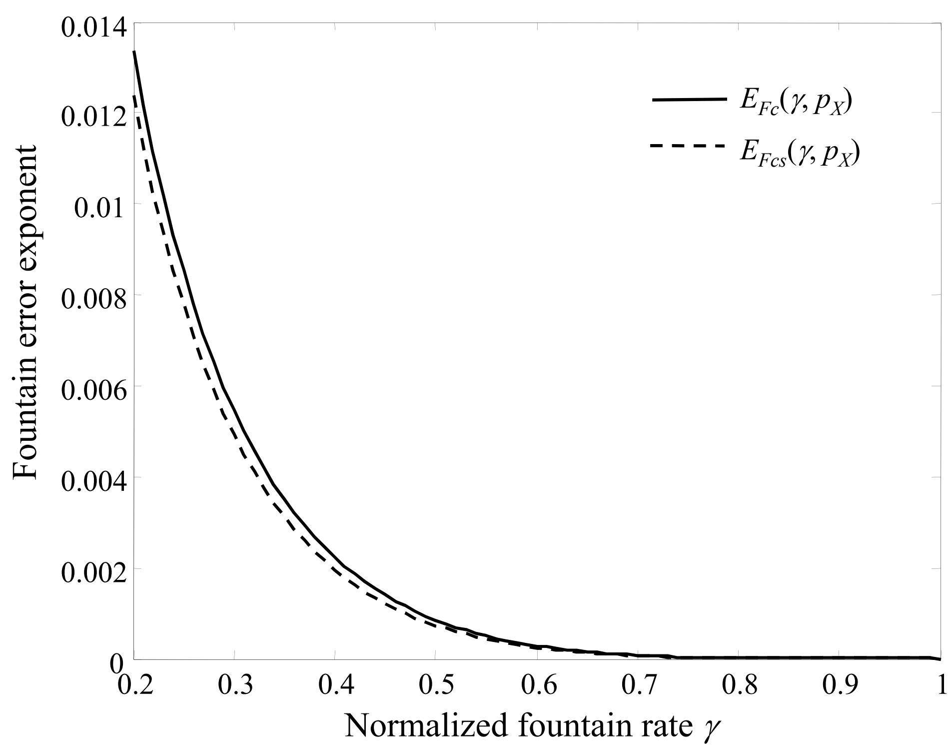

Fig. 6. Error exponents achieved by optimal r_o and suboptimal $r_o = \frac{\sqrt{\gamma^2+8\gamma}-\gamma}{2}$ versus normalized fountain rate γ.