[8] B. C. Arnold, N. Balakrishnan, and H. N. Nagaraja, *A First Course in Order Statistics*. New York: Wiley, 1992.

[9] E. T. Whittaker and G. N. Watson, *A Course of Modern Analysis*, 4th ed. Cambridge, U.K.: Cambridge University Press, 1996.

[10] R. J. Serfling, *Approximation Theorems of Mathematical Statistics*. New York: Wiley, 1980.

[11] A. Jalali, R. Padovani, and R. Pankaj, "Data throughput of CDMA-HDR: A high efficiency-high data rate personal communication wireless system," in *Proc. Veh. Technol. Conf.*, Tokyo, Japan, May 2000, vol. 3, pp. 1854–1858.

[12] P. Svedman, S. K. Wilson, L. Cimini, and B. Ottersten, "A simplified opportunistic feedback and scheduling scheme for OFDM," in *Proc. Veh. Technol. Conf.*, Milan, Italy, May 2004.

[13] W. Rhee and J. M. Cioffi, "Increase in capacity of multiuser OFDM system using dynamic subchannel allocation," in *Proc. Veh. Technol. Conf.*, May 2000, vol. 2, pp. 1085–1089.

Fountain Capacity

Shlomo Shamai (Shitz), *Fellow, IEEE*,
İ. Emre Telatar, *Member, IEEE*, and Sergio Verdú, *Fellow, IEEE*

***Abstract*—Fountain codes are currently employed for reliable and efficient transmission of information via erasure channels with unknown erasure rates. This correspondence introduces the notion of fountain capacity for arbitrary channels. In contrast to the conventional definition of rate, in the fountain setup the definition of rate penalizes the reception of symbols by the receiver rather than their transmission. Fountain capacity measures the maximum rate compatible with reliable reception regardless of the erasure pattern. We show that fountain capacity and Shannon capacity are equal for stationary memoryless channels. In contrast, Shannon capacity may exceed fountain capacity if the channel has memory or is not stationary.**

***Index Terms*—Arbitrarily varying channels, channel capacity, content distribution, erasure channels, fountain codes.**

I. Introduction

Fountain codes are a class of sparse-graph codes that have received considerable attention in the last few years. The first *fountain codes* were the LT erasure-correcting codes introduced by Luby in [1]. The LT codes are linear rateless codes that encode a vector of k symbols of information with an infinite sequence of parity-check bits. The parity-check equations (known to the decoder) are chosen equiprobably from a random ensemble: The cardinality of the parity-check equations has a histogram given by the so-called robust soliton distribution and all k information symbols have identical probability to participate in any given parity-check equation. The infinite sequence is transmitted through an erasure channel. The decoder runs a belief propagation algorithm observing only as many channel outputs as necessary to recover the k transmitted bits.

Better performance can be obtained with the fountain codes known as *raptor codes* introduced by Shokrollahi in [2] for erasure correction. Raptor codes have been applied to other channels such as binary channels in [3]–[5] and Gaussian channels [6].

A typical application of fountain codes is a system where the same message is to be broadcast simultaneously to several receivers, served by erasure channels with different erasure rates. The conventional Shannon-theoretic approach to this scenario is the compound channel (see, e.g., [7]), where the actual channel is unknown to the encoder and chosen from a given uncertainty set. Ensuring reliable communication for all receivers, the compound capacity is upper-bounded by the smallest capacity among those channels in the uncertainty set. This bound is tight in those cases, such as the compound erasure channel, in which the mutual information of all channels in the uncertainty class is maximized by the same input distribution. This setup not only requires the transmitter to cater to the worst channel conditions but it incurs a considerable waste of channel resources for those receivers that enjoy better erasure rates than the worst. The use of fountain codes enables receivers to stop listening to the channel once the information is decoded reliably. Thus, receivers only need to obtain from the channel a number of symbols that is a small multiple (close to 1) of the number of information symbols. This happens sooner for those receivers that experience favorable channel conditions. As customary in the information theory of channels with nonprobabilistic description of the uncertainty, we adopt a worst case approach in order to capture the robustness of the fountain codes with respect to the patterns of erasures.

Fountain codes have been adopted in the 3GPP wireless standard for Multimedia Broadcast/Multicast [8], [9] and they have been used in lossless data compression in [10].

In addition to their appealing conceptual structure, the commercial success and excellent efficiency achieved by fountain codes are incentives to investigate their Shannon-theoretic limits. The main difference from the standard Shannon setup is in the definition of rate: a fountain code is rateless (or zero-rate) in that it adds an infinite amount of redundancy to the information vector. Instead of defining the rate from the perspective of the encoder, in the fountain setup we define it from the perspective of the decoder: ratio of information symbols transmitted to channel symbols received. So while the classical definition of rate penalizes the use of the channel by the transmitter ("pay-per-use"), in the fountain setup the definition of rate penalizes the reception of (non-erased) symbols by the receiver ("pay-per-view"). Independent of the fountain code setting and within the context of broadcast channels, it has been recognized in [11]–[13] that the classical definition of rate is overly pessimistic for asynchronous broadcast where a common message is transmitted to several receivers which are "turned on" at not necessarily identical times. In [11]–[13], the individual rates in the capacity region are normalized by the time until the corresponding receiver is switched off. Recent works that deal with the conventional Shannon capacity of the concatenation of noisy channels and erasure channels include [14], [15].

This correspondence is organized as follows. In Section II, we give the definition of fountain capacity for an arbitrary channel, along with the associated notions of reliability and allowable encoding strategies. We show that fountain capacity is upper-bounded by Shannon capacity.

Manuscript received October 8, 2006; revised July 20, 2007. This work was supported in part by the National Science Foundation under Grants CCR-0312879 and CCF-0635154 and by the U.S.–Israel Binational Science Foundation under Grant 2004140. The material in this correspondence was presented at the IEEE International Symposium on Information Theory, Seattle, WA, July 2006.

S. Shamai (Shitz) is with the Department of Electrical Engineering, Technion–Israel Institute of Technology, Technion City, Haifa, 32000, Israel (e-mail: sshlomo@ee.technion.ac.il).

İ. E. Telatar is with the Information Theory Laboratory (LTHI), School of Computer and Communication Sciences (I &C), Ecole Polytechnique Federale de Lausanne (EPFL), CH-1015, Lausanne, Switzerland (e-mail: Emre.Telatar@epfl.ch).

S. Verdú is with the Department of Electrical Engineering, Princeton University, Princeton NJ 08544 USA (e-mail: verdu@princeton.edu).

Communicated by G. Kramer, Associate Editor for Shannon Theory.

Digital Object Identifier 10.1109/TIT.2007.907495

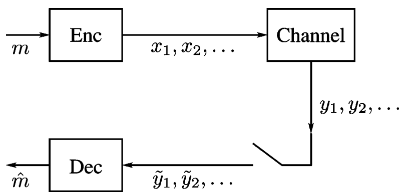

Fig. 1. Concatenation of noisy and erasure channel.

In Section III, we show that the fountain capacity of any discrete memoryless channel is equal to its Shannon capacity, a result which no longer holds for nonstationary channels. Moreover, we also consider memoryless compound and arbitrarily varying channels (AVC), and show that the compound/AVC capacities of those channels are equal to the corresponding compound/AVC fountain capacities.[1]

In Section IV, we give bounds on fountain capacity of channels with memory and we show examples of channels with memory whose Shannon capacity is larger than their fountain capacity.

II. Definition of Fountain Capacity

In Fig. 1, we depict the basic fountain setup where the receiver, but not the transmitter, is aware of the schedule of times at which the switch is on. In the definition of fountain capacity it is desirable to guarantee reliability regardless of the schedule, while at the same time defining rate by the ratio of received information bits to observed channel symbols.

For the purpose of defining fountain capacity, we consider a general channel $\{P_{Y^n|X^n}\}_{n=1}^{\infty}$ with input and output alphabets $\mathcal{X}$, $\mathcal{Y}$, respectively.

A fountain codebook with M codewords is a mapping

$$\mathcal{C} : \{1, \ldots, M\} \to \mathcal{X}^{\mathbb{Z}^+}$$

that associates to each message m in $\{1, \ldots, M\}$ an infinite sequence $(X_{m1}, X_{m2}, \ldots)$ of channel input symbols.

A *fountain code library with M codewords,* $\mathcal{L}$, is a collection of fountain codebooks with M codewords, $\mathcal{L} = \{\mathcal{C}_\theta : \theta \in \Theta\}$, indexed by a set Θ.

A *schedule* $\aleph$ is a subset of the positive integers, whose cardinality we denote by $|\aleph|$. The receiver is only allowed to see the channel outputs $(Y_i,\ i \in \aleph)$ at those times belonging to the schedule $\aleph$. The schedule is unknown to the encoder, and the schedule is chosen without knowledge of either the message, codebook, or the channel output.

A *fountain decoder* maps $(Y_i,\ i \in \aleph)$ to a message in $\{1, \ldots, M\}$, knowing the codebook used at the encoder. This captures the common practice in fountain coding where the actual sequence of parity-check equations is known at the decoder.

Assuming that the maximum-likelihood decoder is used, and therefore that the decoder chooses the most likely message upon knowledge of the codebook, schedule, and channel law, the *error probability* achieved by a codebook $\mathcal{C}$ and a schedule $\aleph$ (averaged over equiprobable messages) is denoted by $e(\mathcal{C}, \aleph)$.

In a fountain communication system, the transmitter and receiver are equipped with a fountain code library $\mathcal{L} = \{\mathcal{C}_\theta : \theta \in \Theta\}$ with M codewords, and a $\theta \in \Theta$ drawn according to a probability distribution γ. Observe that θ is known to both the transmitter and receiver and thus "random coding" is allowed as a communication technique, not just as a method to prove the existence of good codes. To communicate message m, the transmitter sends the infinite sequence $\mathcal{C}_\theta(m)$; the receiver, upon observing the channel output $\{Y_i, i \in \aleph\}$, declares the maximum-likelihood estimate $\hat{m}$ of m.

Definition 1: A fountain rate R is said to be achievable if there exists a sequence of fountain code libraries $\mathcal{L}_1, \mathcal{L}_2, \ldots$, where

$$\mathcal{L}_n = \{\mathcal{C}_{n,\theta} : \theta \in \Theta_n\}$$

has $\lceil 2^{nR} \rceil$ codewords, and a sequence of distributions γ_n on Θ_n such that

$$\lim_{n\to\infty} \sup_{\aleph : |\aleph| \geq n} \int_{\Theta_n} e(\mathcal{C}_{n,\theta}, \aleph)\, d\gamma_n(\theta) = 0. \tag{1}$$

The channel *fountain capacity* C_F is the supremum of all the achievable fountain rates.

Note that in the above definition of achievable rate, the choice of the schedule is performed by an adversary who knows the code library and the probability law by which a codebook in this library is chosen, but is unaware of which codebook is actually chosen. This adversary chooses the schedule with the aim of maximizing the ensemble average probability of error under the constraint that a sufficient number of channel symbols are observed by the receiver.

An easy consequence of the definition above is as follows.

Proposition 1: The fountain capacity is upper bounded by the Shannon capacity.

Proof: We can lower-bound the left side of (1) by taking the contiguous schedule $\aleph = \{1, \ldots, n\}$ in which case the setup boils down to the conventional setup [7], in which rates above Shannon capacity are not achievable even with random coding. □

It is straightforward to incorporate the ingredient of compound or AVCs into the capacity by taking the supremum in (1) to be with respect to not only the schedule but the channel uncertainty class. The same reasoning as in the proof above shows that the fountain capacity in these settings is upper-bounded by the corresponding *random-coding* Shannon capacities.

To see why we need to consider random codes to arrive at a nontrivial definition of fountain capacity, suppose that the scheduler knows which codebook is used. We can view the codebook as M infinitely long rows. Since there are $|\mathcal{X}|^M$ possibilities (at most) for each column, the scheduler can always find an infinite subsequence in which the columns are all equal, in which case the decoder sees a repetition code which cannot achieve any positive rate with vanishing error probability.

One can state a more general conclusion along these lines.

Theorem 1: If $\mathcal{L}_1, \mathcal{L}_2, \ldots$, is a sequence of codelibraries with $\mathcal{L}_n = \{C_{n,\theta} : \theta \in \Theta_n\}$, and if each Θ_n is a finite set, then $\mathcal{L}_1, \mathcal{L}_2, \ldots$ cannot achieve any positive fountain rate.

Proof: Since $K_n = |\Theta_n|$ is finite, we can view the code library $\mathcal{L}_n$ as a collection of $\lceil 2^{nR} \rceil K_n$ rows. As there are $|\mathcal{X}|^{\lceil 2^{nR} \rceil K_n}$ possibilities for each column, there exist a schedule for which all columns are identical for time indices in the schedule. Thus, no matter which codebook is used, it still looks like a repetition code to the decoder. □

The finiteness of the input alphabet is crucial for the foregoing proof. For an additive Gaussian noise channel with transmitter subject a power constraint the same conclusion holds. Although now we have the freedom to choose columns that never repeat, the power constraint on the codebook enables the scheduler to find columns that are sufficiently close in Euclidean distance so as to render the variance of the encoded sequence at the times of the schedule to be as as small as desired.

[1]For the AVC, this only holds in the so-called random coding setting, when the "jammer" is not informed of the actual code, only of the ensemble from where it is chosen.

III. Memoryless Channels

A. Stationary Channels

Theorem 2: For a stationary memoryless channel, the fountain capacity C_F equals the Shannon capacity C_S.

Proof: Given a rate $R < C_S$, find an input distribution P_X on the input alphabet $\mathcal{X}$ of the channel so that $R < I(X;Y)$. Consider now choosing $\mathcal{L}_n$ to contain all codebooks with $\lceil 2^{nR} \rceil$ codewords, and choose the probability distribution γ to make the random variables $X_{m,j} : \theta \mapsto \mathcal{C}_\theta(m)_j$ independent and identically distributed (i.i.d.) with distribution P_X. Observe now, that for any $\aleph$, the integral in (1) is nothing but the ensemble average error probability of the i.i.d. random coding ensemble of rate $Rn/|\aleph|$ over the memoryless channel $P_{Y|X}$, and thus depends on $\aleph$ only through its cardinality $|\aleph|$. For $|\aleph| \geq n$, the rate of the random coding ensemble is less than R, and by [17, Theorem 5.6.2], this ensemble average error probability approaches zero as n gets large. □

The same argument as in the proof of Theorem 2 also establishes that if we have a compound *memoryless* channel, its "compound fountain capacity" equals its usual compound channel capacity. Note that we can view this memoryless compound channel model as one where a single schedule affects all possible channels in the uncertainty class and there is a single rate defined based on the cardinality of that common schedule. Note that this setting is different from the erasure compound channel example in the Section I, where the fountain capacity of each individual channel is achieved and is equal to its Shannon capacity.

For a memoryless AVC, the channel law is also a function of a state s under the control of an adversary; formally, the channel is described by $P_{Y|X,S}$. The adversary is completely free in his choice of state sequence, and he does so with the full knowledge of the mechanism employed by the transmitter and receiver, but without knowing which message is being communicated. If random coding is allowed, and if the adversary knows the random coding ensemble (but not the code in use), the capacity of an AVC is given by (see, e.g., [18])

$$C_R = \max_P \min_\zeta I(P, W_\zeta) = \min_\zeta \max_P I(P, W_\zeta) \tag{2}$$

where the maximization is over probability distributions P on the channel input, minimization is over all probability distributions ζ on the state, W_ζ denotes the channel $W_\zeta(y|x) = \sum_s \zeta(s) P_{Y|X,S}(y|x,s)$, and $I(P,W)$ denotes the mutual information between two random variables with distribution $P(x)W(y|x)$. The proof of this result establishes that if the random coding ensemble is the one that chooses the codewords by making each letter of each codeword i.i.d. with distribution P and the code has rate less than $\min_\zeta I(P, W_\zeta)$, then the error probability for any choice of the state sequence approaches zero as the block length increases.

If we use the same code library as in the proof of Theorem 2, with $R < C_R$ we see that the integral in (1) again depends on $\aleph$ only through its size, and for $|\aleph| \geq n$ it is the error probability of a random code of block length $|\aleph|$ of rate less than R. By the above discussion, this error probability approaches zero for any state sequence as n gets large, and so we see that the fountain capacity for a memoryless AVC equals (2), the random coding Shannon capacity of the AVC.

B. Nonstationary Channels

For a nonstationary discrete memoryless channel with

$$P_{Y^n|X^n}(y^n|x^n) = \prod_{i=1}^n W_i(y_i|x_i) \tag{3}$$

where $W_i \in G$ and G is a finite set, the Shannon capacity is [19]

$$C_S = \liminf_{n\to\infty} \frac{1}{n} \sum_{i=1}^n C_i \tag{4}$$

where

$$C_j = \max_P I(P, W_j). \tag{5}$$

In contrast, the fountain capacity is

$$C_F = \min_{j: W_j \in G^*} C_j \tag{6}$$

where G^* is the subset of G containing the channels that appear infinitely often in the sequence $W_1, W_2, \ldots$. To show that the right-hand side of (6) is an upper bound to fountain capacity we can simply consider a schedule $\aleph$ containing only those times at which W_j is a minimizer of (6). To show achievability of the right-hand side of (6), we only need to generalize the proof of Theorem 2 by replacing i.i.d. random coding by independent nonidentically distributed random coding, where the distribution at time j is given by the maximizer of (5).

IV. Channels With Memory

A general formula for the fountain capacity of channels with memory that would generalize the formula for Shannon capacity in [19] remains an open problem. One of the central difficulties in determining the fountain capacity of channels with memory is that the least favorable schedule is heavily dependent on the channel.

A general upper bound on fountain capacity is

$$C_F \leq \liminf_{n\to\infty} \sup_{\boldsymbol{X}} \inf_{|\aleph_n|=n} \frac{1}{n} I(\boldsymbol{X}; Y[\aleph_n]) \tag{7}$$

where the supremum is over the set of sequences of finite-dimensional input distributions $\{P_{X^m}\}_{m=1}^\infty$, and $Y[\{i_1,\ldots,i_n\}] = \{Y_{i_1},\ldots,Y_{i_n}\}$. The standard argument based on Fano's inequality is easy to modify to normalize the rate not by the number of channel uses but by the number of outputs revealed to the decoder. Furthermore, since reliability is required regardless of the location of the n received symbols, any achievable rate must be upper-bounded by the right-hand side of (7).

Another upper bound to the fountain capacity is

$$C_F \leq \inf_{0<\mathsf{e}<1} \frac{C(\mathsf{e})}{1-\mathsf{e}} \tag{8}$$

where $C(\mathsf{e})$ is the capacity of the concatenation of the noisy channel with a memoryless erasure channel with erasure rate e. To justify (8), note that for any e, $\frac{C(\mathsf{e})}{1-\mathsf{e}}$ would be an achievable fountain rate if the reliability in (1) were modified to average over i.i.d. erasures instead of minimizing with respect to the schedule. The upper bound in (8) is useful in those cases, such as the Gaussian dispersive channel [15], where $C(\mathsf{e})$ is known.

We can show a simple lower bound on fountain capacity for power-constrained additive stationary Gaussian noise channels

q

$$Y_i = X_i + Z_i \tag{9}$$

where $E\left[Z_i^2\right] = \sigma^2$ and the input power is constrained not to exceed P.

The fountain capacity of (9) is lower-bounded by

$$C_F \geq \frac{1}{2}\log\left(1+\frac{P}{\sigma^2}\right). \tag{10}$$

To show (10), we first recall that because of the asymptotic equipartition property for arbitrary Gaussian processes [16, eq. (46)], we can claim, invoking standard information-theoretic arguments, that if X_i are independent Gaussian with variance P, then

$$\liminf_{n\to\infty}\inf_{\aleph}\frac{1}{n}I(X[\aleph];Y[\aleph])$$

is an achievable fountain rate where the infimization is over all schedules of size n. Using the independence of the inputs, we can write for any $\aleph$

$$\frac{1}{n}I(X[\aleph];Y[\aleph]) \geq \frac{1}{n}\sum_{i\in\aleph} I(X_i;Y_i) \tag{11}$$

$$= \frac{1}{2}\log\left(1+\frac{P}{\sigma^2}\right). \tag{12}$$

In the presence of memory, fountain capacity can be quite a bit smaller than Shannon capacity. We provide two examples.

Example 1: The input and output alphabets are equal to the interval $[0,1]$. The channel described by

$$Y_i = X_i \oplus W_i \tag{13}$$

where the noise process $\{W_i\}$ is defined as

$$(\ldots,W_{-1},W_0,W_1,W_2,W_3,\ldots) = \begin{cases}(\ldots,N_{-1},N_0,N_0,N_1,N_1,\ldots) & \text{with prob. } 1/2\\ (\ldots,N_0,N_0,N_1,N_1,N_2,\ldots) & \text{with prob. } 1/2\end{cases} \tag{14}$$

where $\ldots,N_{-1},N_0,N_1,\ldots$ is an i.i.d. sequence of random variables uniformly distributed on $[0,1]$.

The zero-error capacity (and thus, the Shannon capacity) of the channel above is infinity. To see this, note that regardless of the cardinality of $\mathcal{S}\subset[0,1]$, if we let $X_0 = X_1 = 0$ and

$$X_{2i} = 0, \quad X_{2i+1} = S_i, \qquad i = 1,2,\ldots \tag{15}$$

with $S_i \in \mathcal{S}$, we can recover $\{S_i\}$ noiselessly, with probability 1, via the equations

$$\hat{S}_i = \begin{cases} Y_{2i+1}\ominus Y_{2i}, & \text{if } Y_0 = Y_1,\\ Y_{2i+1}\ominus Y_{2i+2}, & \text{otherwise}\end{cases} \tag{16}$$

where $\ominus$ stands for subtraction modulo the unit interval.

However, if the schedule $\aleph$ contains only the even integers, then the channel output observed at the receiver is

$$Y_{2k} = X_{2k}\oplus W_{2k}, \qquad k = 1,2,\ldots.$$

Noting that the sequence $\{W_{2k} : k = 1,2,\ldots\}$ is i.i.d. and that each W_{2k} is uniformly distributed in the interval $[0,1]$, the capacity of this channel is zero, and thus so is the fountain capacity.

Note that a simpler example of zero fountain capacity and infinite capacity can be given by not attempting to stationarize the noise; we do that in order to explicitly show that it is memory (rather than nonstationarity as in (3)), that accounts for the discrepancy.

The discrete counterpart of the channel in this example, where the input/output alphabet is binary, can be similarly be shown to have zero fountain capacity and Shannon capacity equal to 1/2 bit. To that end, a "silent" period taking negligible rate at the beginning of the codeword enables the decoder to ascertain the phase of the noise with arbitrary reliability.

The next example is perhaps somewhat more familiar.

Example 2: Consider an additive Gaussian noise channel with colored noise

$$Y_k = X_k + Z_k \tag{17}$$

where the channel input $X_1, X_2, \ldots$ is power-constrained to have average (over messages and time) power P and $\{Z_k\}$ is zero mean, stationary additive Gaussian noise, whose law is independent of the channel input. Consider a "low-pass" noise whose power spectral density $S_z(\theta)$ is confined to half the bandwidth

$$S_z(\theta) = \begin{cases} 2N, & -\pi/2 \leq \theta < \pi/2\\ 0, & \text{else}\end{cases} \tag{18}$$

where $N = (2\pi)^{-1}\int_{[-\pi,\pi)} S_z(\theta)\,d\theta$ is the variance of Z_k.

It is clear that the Shannon capacity in this case is infinite for any nonzero P, as there are noise-free frequency bands. However, the fountain capacity is equal to $\frac{1}{2}\log(1+P/N)$. According to (10), we only need to show that $\frac{1}{2}\log(1+P/N)$ is an upper bound to fountain capacity. This follows from the observation that $\{Z_{2k} : k = 1,2,\ldots\}$ form an i.i.d. sequence of zero-mean Gaussian random variables of variance N, and thus for the schedule that lets the receiver see only the outputs at even times, the equivalent channel is an additive *white* Gaussian channel whose capacity is $\frac{1}{2}\log(1+P/N)$.

For those who are unhappy with noise processes that are not regular [20] (in the sense that past samples of the noise determine the future samples) one can modify the example by taking an $0<\epsilon<N$ and letting

$$S_z(\theta) = \begin{cases} 2(N-\epsilon), & -\pi/2 \leq \theta < \pi/2\\ 2\epsilon, & \text{else.}\end{cases}$$

It is easy to check that $\{Z_{2k} : k = 1,2,\ldots\}$ still form an i.i.d. sequence of zero-mean Gaussian random variables of variance N, so that the fountain capacity is still equal to $\frac{1}{2}\log(1+P/N)$. The Shannon capacity will now be given by the water-pouring solution, but in any case is larger than $\frac{1}{4}\log(1+P/\epsilon)$ (by allocating power to only to high frequencies). We again see that the discrepancy between Shannon and fountain capacities can be made arbitrarily large by taking ϵ arbitrarily small; indeed, for any given nonzero $a<A$, Shannon capacity can be made larger than A and fountain capacity smaller than a by a proper choice of P and ϵ.

V. Concluding Remarks

The setting proposed here to formalize the notion of fountain capacity is reminiscent of the random-coding setting used for the arbitrarily varying channel in which the "jammer" does not know the code used by the communicator. Indeed, given a channel with input alphabet $\mathcal{X}$ and output alphabet $\mathcal{Y}$, we can define a new channel by equipping the original channel with a state chosen from alphabet $\mathcal{S} = \{0,1\}$, and augmenting the output alphabet with an erasure symbol such that when the state is 0, the output of the new channel equals the output of the original channel, and when the state is 1, the output of the new channel equals the erasure symbol.

Recall that in an AVC setting, the state sequence is controlled by an adversary who knows the communication mechanism used by the transmitter and receiver, but not the message being sent. If randomized coding used, then the adversary knows—just as in the fountain setting above—the code library, but not which codebook is actually used. Thus, the role of the adversary for this AVC is of determining the schedule. However, there are important differences. The AVC setting defines the rate as a property of the transmission code; it does not allow, as is done here, to define the rate from the perspective of the receiver. The AVC setting does allow one to consider an average cost constrained adversary. Through this, one can insist that a guaranteed fraction of channel outputs are received unerased. The AVC formulation would charge the adversary a unit cost for erasing an output, compute the AVC capacity $C(\lambda)$ under an average cost constraint λ, convert this "pay-per-use" rate to a "pay-per-view" rate $C(\lambda)/(1-\lambda)$, and then infimize over all $0 < \lambda < 1$. This leads to a stronger, albeit harder to compute, bound than (8). The following example illustrates the fact that such an AVC capacity under cost constraints does not adequately capture the notion of fountain rate.

Example 3: Consider a memoryless but time-varying binary-symmetric channel, whose crossover probability p_j at time j is given by

$$p_j = \begin{cases} 1/2, & j \text{ is prime} \\ 0, & \text{else.} \end{cases}$$

Since the channel is noiseless at an asymptotically unit fraction of the time, and thus, for any $\lambda < 1$, the adversary is forced to allow noiseless transmissions at a fraction of the time equal to $1 - \lambda$. It follows that $C(\lambda) = 1 - \lambda$. Thus, the capacity of the AVC as defined above is equal to 1, while from (6) it follows that fountain capacity is zero.

Another difference is that (recall Theorem 1) if the code library is allowed to contain a finite number of codes, then no positive fountain rate is achievable; in contrast, the AVC capacity with a finite number of codes or even deterministic coding can be nonzero [18].

For stationary memoryless channels, feedback fountain capacity is upper-bounded by Shannon capacity since causal feedback does not increase the ordinary capacity and one of the options of the scheduler is to adopt a contiguous schedule; on the other hand, since feedback can just be ignored, feedback fountain capacity is lower-bounded by fountain capacity which is equal to Shannon capacity. Note that this holds whether the feedback is of the channel output (prior to the scheduler) or of the decoder input (after the scheduler). However, if the encoder obtains the schedule causally, i.e., it knows whether $i \in \aleph$ prior to the transmission of $i + 1$, random coding can be dispensed with (at least if the channel is memoryless). To see this, note that for any given code we can obtain the same fountain rate and error probability for any schedule, by simply repeating the transmission of the last symbol if it has been erased.

Several open problems are worth exploring: a) the fountain capacity of some channels with memory may yield to analysis if the scheduler is subject to constraints such as run-length or cost; b) fountain capacity per unit cost (cf. [21]), where the cost of receiving a symbol need not be equal for all symbols; c) generalizing the lower bound in (10) to other stationary channels; d) fountain capacity regions for multiterminal systems ranging from challenging setups with multiple schedulers to a simple memoryless multiple-access channel with a single scheduler at the receiver (for which the fountain capacity region equals the conventional one).

References

[1] M. G. Luby, "LT codes," in *Proc. 43rd IEEE Symp. Foundations of Computer Science*, Vancouver, BC, Canada, Nov. 2002, pp. 271–280.

[2] A. Shokrollahi, "Raptor codes," *IEEE Trans. Inf. Theory*, vol. 52, no. 6, pp. 2551–2567, Jun. 2006.

[3] P. Pakzad and A. Shokrollahi, "Design principles for raptor codes," in *Proc. 2006 IEEE Information Theory Workshop*, Punta del Este, Uruguay, Mar. 2006, pp. 165–169.

[4] O. Etesami and A. Shokrollahi, "Raptor codes on binary memoryless symmetric channels," *IEEE Trans. Inf. Theory*, vol. 52, no. 5, pp. 2033–2051, May 2006.

[5] R. Palanki and J. Yedidia, "Rateless codes on noisy channels," in *Proc. 2004 IEEE Int. Symp. Information Theory*, Chicago, IL, Jun./Jul. 2004, p. 37.

[6] U. Erez, M. D. Trott, and G. W. Wornell, "Rateless coding for Gaussian channels and perfect incremental redundancy," in *Proc. Information Theory and Applications Workshop*, San Diego, Feb. 2006.

[7] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. New York: Academic, 1981.

[8] Press Release: Digital Fountain, Inc., 2005 [Online]. Available: http://www.digitalfountain.com/technology/standards/index.cfm

[9] Multimedia Broadcast/Multicast Service (MBMS); Protocols and Codecs. [Online]. Available: http://www.3gpp.org/ftp/ Specs/html-info/ 26346.htm

[10] G. Caire, S. Shamai (Shitz), A. Shokrollahi, and S. Verdú, "Fountain codes for lossless compression of binary sources," in *Proc. 2004 IEEE Workshop on Information Theory*, San Antonio, TX, Oct. 2004.

[11] T. M. Cover, "Comments on broadcast channels," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2524–2530, Oct. 1998.

[12] N. Shulman and M. Feder, "Static broadcasting," in *Proc. 2000 IEEE Int. Symp. Information Theory*, Sorrento, Italy, Jun. 2000, p. 23.

[13] M. Feder and N. Shulman, "Source broadcasting with unknown amount of receiver side information," in *Proc. 2002 IEEE Information Theory Workshop*, Bangalore, India, Oct. 2002, pp. 127–130.

[14] S. Verdú and T. Weissman, The Information Lost in Erasures, 2007, submitted for publication.

[15] A. M. Tulino, S. Verdú, G. Caire, and S. Shamai, "The Gaussian-erasure channel," in *Proc. IEEE 2007 Int. Symp. Information Theory*, Nice, France, Jun. 2007, pp. 1721–1725.

[16] T. Cover and S. Pombra, "Gaussian feedback capacity," *IEEE Trans. Inf. Theory*, vol. 35, no. 1, pp. 37–43, Jan. 1989.

[17] R. G. Gallager, *Information Theory and Reliable Communication*. New York: Wiley, 1968.

[18] A. Lapidoth and P. Narayan, "Reliable communication under channel uncertainty," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2148–2177, Oct. 1998.

[19] S. Verdú and T. S. Han, "A general formula for channel capacity," *IEEE Trans. Inf. Theory*, vol. 40, no. 4, pp. 1147–1157, Jul. 1994.

[20] J. L. Doob, *Stochastic Processes*. New York: Wiley, 1953.

[21] S. Verdú, "On channel capacity per unit cost," *IEEE Trans. Inf. Theory*, vol. 36, no. 5, pp. 1019–1030, Sep. 1990.