

Fountain Capacity

Shlomo Shamai

Technion

Haifa, Israel

sshlo@ee.technion.ac.il

Emre Telatar

EPFL — I&C — LTHI

Lausanne, Switzerland

emre.telatar@epfl.ch

Sergio Verdú

Princeton University

Princeton, NJ, USA

verdu@princeton.edu

Abstract—Fountain codes have been successfully employed for reliable and efficient transmission of information via erasure channels with unknown erasure rates. This paper introduces the notion of fountain capacity for arbitrary channels, and shows that it is equal to the conventional Shannon capacity for stationary memoryless channels. In contrast, when the channel is not stationary or has memory, Shannon capacity and fountain capacity need not be equal.

I. FOUNTAIN CODES

The first *fountain codes* were the LT codes introduced by M. 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.

Incurring only a slight increase in encoding/decoding complexity, better performance can be obtained with the *Raptor codes* introduced by A. Shokrollahi in [2]. The same codes have been applied to other channels such as binary symmetric channels in [3] and [4].

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., [5]), where the actual channel is unknown to the encoder and chosen from a given uncertainty set. The resulting capacity, which ensures reliable communication for all receivers, boils down (in the case in which, like in the compound erasure channel, the mutual information of all channels in the uncertainty class is maximized by the same input distribution) to the worst-case capacity. 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 listen to the channel only for a sufficiently long period of time to ensure that their information is decoded reliably. Thus, receivers that face favorable channel conditions

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.

Fountain codes have been adopted in the 3GPP wireless standard for Multimedia Broadcast/Multicast [6], [7] and they have been used in lossless data compression in [8].

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.

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. Fountain capacity is upper bounded by Shannon capacity. In Section III we show that the fountain capacity of any discrete memoryless channel is equal to its Shannon capacity. 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 capacities¹.

In Section IV we show examples of channels with memory whose Shannon capacity is much larger than their fountain capacity.

II. DEFINITION OF FOUNTAIN CAPACITY

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

$$\mathcal{C} : \{1, \dots, M\} \rightarrow \mathcal{X}^{\mathbb{Z}^+}$$

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

A *fountain codebook* 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 .

¹For 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.

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.

A *fountain decoder* maps $(Y_i, i \in \aleph)$ to a message in $\{1, \dots, M\}$, knowing the codebook used at the encoder.

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 codebook $\mathcal{L} = \{\mathcal{C}_\theta : \theta \in \Theta\}$ 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 codebooks $\mathcal{L}_1, \mathcal{L}_2, \dots$, 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

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

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 codebook 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:

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 $\aleph = \{1, \dots, n\}$ in which case the setup boils down to the conventional setup [5], in which rates above Shannon capacity are not achievable even with random coding. ■

It is straightforward to incorporate the ingredient of compound or arbitrarily-varying channels (AVC) 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 $\mathcal{L}_1, \mathcal{L}_2, \dots$, is a sequence of codebooks with $\mathcal{L}_n = \{\mathcal{C}_{n,\theta} : \theta \in \Theta_n\}$, and if each Θ_n is a finite set, then $\mathcal{L}_1, \mathcal{L}_2, \dots$ cannot achieve any positive fountain rate.

Proof: Since $K_n = |\Theta_n|$ is finite, we can view the codebook \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. ■

III. MEMORYLESS 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 $X_{m,j} : \theta \mapsto \mathcal{C}_\theta(m)_j$ 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 [9, Theorem 5.6.2], this ensemble average error probability approaches zero as n gets large. ■

The same argument as in the proof just above also establishes that if we have a compound *memoryless* channel, its ‘compound fountain capacity’ equals its usual compound channel capacity.

For a memoryless arbitrarily varying channel (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 the 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., [10])

$$C_R = \max_P \min_\zeta I(P, W_\zeta) \quad (2)$$

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 blocklength increases.

If we use the same codebook 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.

It is easy to see that for a non-stationary discrete memoryless channel with

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

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

$$C_S = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n C_i \quad (4)$$

whereas the fountain capacity is

$$C_F = \min_{j: W_j \in G^*} C_j \quad (5)$$

where

$$C_j = \sup_P I(P, W_j)$$

and G^* is the subset of G containing the channels that appear infinitely often in the sequence W_1, W_2, \dots

IV. CHANNELS WITH MEMORY

We have not been able to find a general formula for the fountain capacity of channels with memory (that would generalize the formula for Shannon capacity in [11]). However, 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 the real interval $[0, 1]$ and modulo-1 addition in that interval is denoted by \oplus . The channel described by

$$Y_i = X_i \oplus W_i \quad (6)$$

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

$$\begin{aligned} (\dots, W_{-1}, W_0, W_1, W_2, W_3, \dots) = \\ \begin{cases} (\dots, N_{-1}, N_0, N_0, N_1, N_1, \dots) & \text{with prob. } 1/2, \\ (\dots, N_0, N_0, N_1, N_1, N_2, \dots) & \text{with prob. } 1/2. \end{cases} \end{aligned} \quad (7)$$

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

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

$$X_{2i} = 0, \quad X_{2i+1} = S_i, \quad i = 1, 2, \dots \quad (8)$$

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

$$\hat{S}_i = \begin{cases} Y_{2i+1} \ominus Y_{2i} & \text{if } Y_0 = Y_1, \\ Y_{2i+1} \ominus Y_{2i+2} & \text{otherwise} \end{cases} \quad (9)$$

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

$$Y_{2k} = X_{2k} \oplus W_{2k}, \quad k = 1, 2, \dots$$

Noting that the sequence $\{W_{2k} : k = 1, 2, \dots\}$ 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 next example is perhaps somewhat more familiar:

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

$$Y_k = X_k + Z_k \quad (10)$$

where the channel input X_1, X_2, \dots 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:

$$S_z(\theta) = \begin{cases} 2N & -\pi/2 \leq \theta < \pi/2 \\ 0 & \text{else,} \end{cases} \quad (11)$$

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 non-zero P , as there are noise-free frequency bands.

However, the fountain capacity is upper bounded by $\frac{1}{2} \log(1 + P/N)$. This follows from the observation that $\{Z_{2k} : k = 1, 2, \dots\}$ 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 (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 $\{Z_{2k} : k = 1, 2, \dots\}$ still form an i.i.d. sequence of zero mean Gaussian random variables of variance N , so that the fountain capacity is still bounded by $\frac{1}{2} \log(1 + P/N)$. 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 .

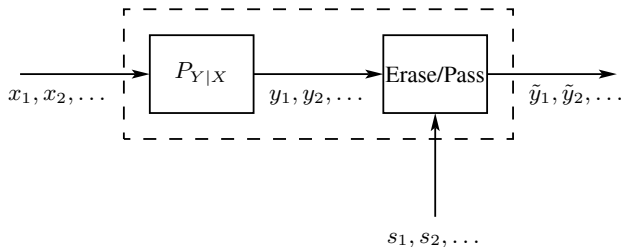
Considering causal channels for which the output sequence up to time n , $Y_{-\infty}^n$, is independent of the future inputs X_{n+1}^∞ when conditioned on past inputs $X_{-\infty}^n$, Fano's inequality establishes that the fountain capacity is upper bounded by

$$C_F \leq \lim_{n \rightarrow \infty} \sup_{X_{-\infty}^n} \inf_{B_n} \frac{1}{|B_n|} I(X_{-\infty}^n; B_n) \quad (12)$$

where the infimum is taken over all subsets $B_n \subset \{Y_1, \dots, Y_n\}$. Under some additional assumptions on the ergodic behavior of such channels, we conjecture that (12) holds with equality.

V. CONCLUDING REMARKS

The setting proposed here to formalize the notion of fountain capacity is very 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 E 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 symbol E .



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 codebook, 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 with respect to the actions of the adversary, or equivalently, from the perspective of the receiver. The AVC setting does allow one to consider an average cost constrained adversary. Through this, one can, for example, insist that a guaranteed fraction of channel outputs are received unerased. However, even with this AVC capacity under cost constraints, one cannot capture the notion of a fountain rate as defined here.

Another difference is apparent by recalling that if the deterministic coding average-error-probability capacity of an AVC is nonzero, then it does not increase further if random coding is allowed [12]. However, in the fountain setting, knowledge of the codebook by the scheduler renders the fountain capacity trivially zero (Theorem 1).

REFERENCES

- [1] M. G. Luby, “LT codes,” *Proc. 43rd IEEE Symposium on Foundations of Computer Science*, pp. 271–280, 2002.
- [2] A. Shokrollahi, “Raptor codes,” Preprint, <http://algo.epfl.ch/contents/output/pubs/raptor.pdf>, Jan. 2004. Also, *Proc. 2004 IEEE International Symposium on Information Theory*, p. 36, June 2004.
- [3] O. Etesami, M. Molkaiaie, and A. Shokrollahi, “Raptor codes on symmetric channels,” *Proc. 2004 IEEE International Symposium on Information Theory*, p. 38, June 2004.
- [4] R. Palanki and J. Yedidia, “Rateless codes on noisy channels,” *Proc. 2004 IEEE International Symposium on Information Theory*, p. 37, June 2004.
- [5] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. New York: Academic, 1981.
- [6] Digital Fountain, Inc. Press release, June 2005. Available at: <http://www.digitalfountain.com/technology/standards/index.cfm>.
- [7] 3GPP Specification detail, *Multimedia Broadcast/Multicast Service (MBMS); Protocols and codecs*, Available at: <http://www.3gpp.org/ftp/Specs/html-info/26346.htm>
- [8] G. Caire, S. Shamai, A. Shokrollahi, and S. Verdú, “Fountain codes for lossless compression of binary sources,” *2004 IEEE Workshop on Information Theory*, San Antonio, TX, Oct. 24–29, 2004.
- [9] R. G. Gallager, *Information Theory and Reliable Communication*. New York: Wiley, 1968.
- [10] A. Lapidoth and P. Narayan, “Reliable communication under channel uncertainty,” *IEEE Transactions on Information Theory*, vol. 44, no. 6, pp. 2148–2177, Oct. 1998.
- [11] S. Verdú and T. S. Han, “A General Formula for Channel Capacity,” *IEEE Transactions on Information Theory*, vol. 40, no. 4, pp. 1147–1157, July 1994.
- [12] R. Ahlswede, “Elimination of correlation in random codes for arbitrarily varying channels,” *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, vol. 44, no. 2, pp. 159–175, 1978.