and

$$G_K^2 \leq 1 + 2\int_0^\infty e^{-x^2}dx = 1 + \sqrt{2\pi}.$$

By the Poisson summation formula, we know that

$$\hat{K}^* = 1/\sqrt{2\pi}\sum_k e^{-(\omega+2k\pi)^2/2}.$$

Then we have $e^{-2\pi}/\sqrt{2\pi} + 2/e \leq \hat{K}^* \leq 1/\sqrt{2\pi} + 2$ and

$$\| K \|_{L_0^1[-1,1]} \leq 1/\sqrt{2\pi}(\sqrt{2}e^{-1/4} + 4e^{-1/2}).$$

By Theorem 4, we derive an estimate of the upper bound

$$r_K < \frac{e^{-2\pi}/\sqrt{2\pi} + 2/e}{1/\sqrt{2\pi}(\sqrt{2}e^{-1/4} + 4e^{-1/2})} \approx 0.35.$$

3) Example 4: Let $\hat{\varphi}(\omega) = \chi_{[-2a\pi,2a\pi)}(\omega), 0 < a < \frac{1}{2}$. Then

$$G_\varphi(\omega) = \chi_{[-2a\pi,2a\pi)}(\omega) \text{ on } [-\pi,\pi).$$

Since $\hat{\varphi}^*(\omega) = \sum_k \hat{\varphi}(\omega + 2k\pi)$ in $L^2[0,2\pi]$ (see [4]), we derive

$$\hat{\varphi}^*(\omega)\chi_{[-\pi,\pi)} = \chi_{[-2a\pi,2a\pi)}(\omega).$$

Then $C_1 = C_2 = 1$. It is easy to verify that $\{\varphi(\cdot - n) | n \in \mathbb{Z}\}$ is a frame for $V_0(\varphi)$ (see [13]). By [4], $\{\varphi(\cdot - n) | n \in \mathbb{Z}\}$ is not a Riesz basis for the subspace $V_0(\varphi)$. Since $\hat{\varphi'}(\omega) = i\omega\hat{\varphi}(\omega)$, $M = \sup_\omega \sum_k |\hat{\varphi'}(\omega + 2k\pi)|^2 = (2a\pi)^2, 0 < a < \frac{1}{2}$. By Theorem 5, we derive an estimate

$$\sup_n |\delta_n| < \frac{1}{(2a\pi)^2}.$$

Acknowledgment

The authors would like to express our deep thanks to Gérard Battail and the referees for their very helpful suggestions, which greatly improved the quality of this paper.

References

[1] G. G. Walter, “A sampling theorem for wavelet subspace,” *IEEE Trans. Inf. Theory*, vol. 38, pp. 881–884, Apr. 1992.

[2] W. Sun and X. Zhou, “Frames and sampling theorem,” *Sci.China*, vol. 41, no. 2, pp. 606–612, 1998.

[3] A. J. E. M. Janssen, “The Zak transform and sampling theorem for wavelet subspaces,” *IEEE Trans.Signal Process.*, vol. 41, pp. 3360–3364, 1993.

[4] W. Chen and S. Itoh, “A sampling theorem for shift-invariant subspace,” *IEEE Trans.Signal Processing*, vol. 46, pp. 2822–2824, 1998.

[5] W. Sun and X. Zhou, “Sampling theorem for wavelet subspaces: Error estimate and irregular sampling,” *IEEE Trans.Signal Process.*, vol. 48, pp. 223–226, 2000.

[6] Y. Liu and G. Walter, “Irregular sampling in wavelet subspaces,” *J. Fourier Anal. Appl.*, vol. 2, no. 2, pp. 181–189, 1995.

[7] W. Chen, S. Itoh, and J. Shiki, “Irregular sampling theorems for wavelet subspaces,” *IEEE Trans. Inf. Theory*, vol. 44, pp. 1131–1142, 1998.

[8] ——, “On sampling in shift invariant spaces,” *IEEE Trans. Inf. Theory*, vol. 48, pp. 2802–2810, 2002.

[9] O. Christensen, “A Paley-Wiener theorem for frames,” *Proc. Amer. Math.Soc.*, vol. 123, no. 7, pp. 2199–2202, 1995.

[10] ——, “Frame perturbations,” *Proc. Amer. Math.Soc.*, vol. 123, no. 4, pp. 1217–1220, 1995.

[11] P. G. Casazza and O. Christensen, “Perturbation of operators and applications to frame theory,” *J. Fourier Anal. Appl.*, vol. 3, no. 5, pp. 543–558, 1997.

[12] C. K. Chui, “An introduction to wavelets,” in *Wavelet Analysis and Its Applications*. New York: Academic, 1992, vol. 1.

[13] J. J. Benedetto and S. Li, “Multiresolution analysis frames with applications,” in *ICASSP’93*, Minneapolis, Apr. 26–30, 1993, vol. III, pp. 304–307.

[14] O. Christensen, “Frames, Riesz bases, and discrete Gabor wavelet expansions,” *Bull. Amer. Math. Soc.*, vol. 38, no. 3, pp. 273–291, 2001.

[15] P. Zhao, G. Liu, and C. Zhao, “Frames and sampling theorems for translation-invariant subspaces,” *IEEE Signal Process. Lett.*, vol. 11, pp. 8–11, 2004.

[16] C. Zhao and P. Zhao, “Sampling theorem and irregular sampling theorem for multiwavelet subspaces,” *IEEE Trans. Signal Process.*, vol. 53, pp. 705–713, 2005.

[17] R. Young, *An Introduction to Non-Harmonic Fourier Series*. New York: Academic, 1980.

On Lower Bounds for the Capacity of Deletion Channels

Eleni Drinea and Michael Mitzenmacher, *Member, IEEE*

***Abstract*—This correspondence considers binary deletion channels, where bits are deleted independently with probability d; it improves upon the framework used to analyze the capacity of binary deletion channels established by Diggavi and Grossglauser, improving on their lower bounds. Diggavi and Grossglauser considered codebooks with codewords generated by a first-order Markov chain. They only consider typical outputs, where an output is typical if an N bit input gives an $N(1-d)(1-\epsilon)$ bit output. The improvements in this correspondence arise from two considerations. First, a stronger notion of a typical output from the channel is used, which yields better bounds even for the codebooks studied by Diggavi and Grossglauser. Second, codewords generated by more general processes than first-order Markov chains are considered.**

***Index Terms*—Binary deletion channel, channel capacity, channels with memory.**

I. Introduction

Deletion channels are a special case of channels with synchronization errors. A synchronization error is an error due either to the omission of a bit from a sequence or to the insertion into a sequence of a bit which does not belong; in both cases, all subsequent bits remain intact, but are shifted left or right, respectively.

In this work, we are interested in lower bounds for the capacity of binary deletion channels where bits are deleted independently with probability d, or *independent and identically distributed (i.i.d.) deletion*

Manuscript received November 21, 2004; revised May 11, 2005. The work of E. Drinea is supported by the National Sccience Foundation under Grant CCR-0118701. The work of M. Mitzenmacher is supported by the National Science Foundation under Grants CCR-9983832, CCR-0118701, CCR-0121154, and an Alfred P. Sloan Research fellowship. The material in this correspondence was presented in part at the IEEE International Symposium on Information Theory, Chicago, IL, June/July 2004.

The authors are with the Division of Engineering and Applied Sciences, Harvard University, Cambridge, MA 02138 USA (e-mail: edrinea@deas.harvard.edu; michaelm@eecs.harvard.edu).

Communicated by R. J. McEliece, Associate Editor for Coding Theory.

Color version of Fig. 2 is available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIT.2006.881832

channels. It is known that the capacity of such channels is related to the mutual information between the codeword sent and the received sequence [4], but this does not give an effective means of proving capacity bounds. Diggavi and Grossglauser [1] have shown that random codes where codewords are chosen independently and uniformly at random from the set of all possible codewords of a certain length can provide a lower bound of

$$C_{\rm del} \geq 1 - H(d) \text{ bits}, \qquad \text{for } d < 0.5$$

where $H(d) = -d \log d - (1-d) \log (1-d)$ is the binary entropy function [1] (we denote by $\log$ the logarithm base 2 and by $\ln$ the natural logarithm throughout). This bound coincides with previous bounds (as discussed in [1]), and can be generalized to stationary and ergodic deletion processes.

Diggavi and Grossglauser then go on to give much improved lower bounds. Their insight revolves around using random codes, but with more sophisticated means of choosing the codewords and a more sophisticated analysis. In particular, they consider codes consisting of codewords of length N generated by a symmetric first-order Markov process with transition probability p. More specifically, the first bit in the codeword is 0 with probability $1/2$; every bit after the first one is the same as its previous one with probability p, while it is flipped with probability $1-p$. It can be shown that the sequence after passing through the i.i.d. deletion channel also obeys a first-order Markov process with transition probability q (a formula for q will be given later). The decoding algorithm they consider takes a received sequence and determines whether this is a subsequence of exactly one codeword from the randomly generated codebook; if this is the case, the decoder is successful, and otherwise, the decoder fails. To analyze this decoder, they use the fact that a simple greedy algorithm can be used to determine if a sequence Y is a subsequence of another sequence X. Specifically, reading Y and X from left to right, the greedy algorithm matches the first character of Y to the leftmost matching character of X, the second character of Y to the subsequent leftmost matching character of X, and so on. By analyzing this greedy algorithm, they determine for what transmission rate the probability of error goes to zero asymptotically. This analysis yields the following lower bound for the capacity, which proves strictly better than the previous lower bound (for random codes), and is substantially better for high deletion probabilities d:

$$C_{\rm del} \geq \sup_{\substack{t>0 \\ 0<p<1}} \left[-t \cdot \log e - (1-d) \log \{(1-q)A + qB\}\right] \text{ bits} \qquad (1)$$

where $A = \frac{(1-p)e^{-t}}{1-pe^{-t}}$, $B = \frac{(1-p)^2 e^{-2t}}{1-pe^{-t}} + pe^{-t}$, and $q = 1 - \frac{1-p}{1+d-2pd}$. In this correspondence, we further improve on the bound in (1). Our improvement arises from two considerations. First, in the analysis of Diggavi and Grossglauser, they consider only *typical outputs*, which consist of at least $N(1-d)(1-\epsilon)$ bits, for some $\epsilon = o(1)$. In their analysis, any output that is atypical is considered an error. The probability of an atypical output is exponentially small. By using a stronger notion of a typical output, and declaring successful decoding if and only if such a typical output is the subsequence of exactly one codeword, we can improve the analysis while keeping an error rate that goes to 0 asymptotically. This technique improves the bound even for the case of codewords generated by first-order Markov chains considered by Diggavi and Grossglauser.

Modifying the definition of a typical output sequence improves the capacity bound rather mildly. Our more important improvement comes by considering the following generalization of the framework described above: we encode the messages by codewords of length N that consist of alternating blocks of zeros and ones. The lengths of the blocks are determined sequentially and are i.i.d. random variables, according to some distribution P over the positive integers. The first-order Markov chains used by Diggavi and Grossglauser give block lengths that are geometrically distributed. There is no reason *a priori* why the geometric distribution is the right choice, either in terms of reaching capacity or in terms of proving lower bounds in this fashion. To consider other distributions, we extend the analysis of [1] to this more general case. We suggest some simple distributions for the block lengths that provide better lower bounds when the deletion probability is at least 0.35, and report initial results for when the deletion probability is smaller.

Of course, the ultimate goal would be to determine the optimal block-length distribution for every value of d, and prove these distributions meet some upper bound. Although this is beyond our current understanding, our extensions to the Diggavi and Grossglauser framework moves us further in this direction.

Before beginning, we review some of the previous work. Prior to this work, the best provable lower bounds for the i.i.d. deletion channel are given by (1). Dolgopolov [5] also obtains upper and lower bounds on channels with i.i.d. deletions and insertions, considering codebooks with codewords chosen uniformly at random. Such codewords appear to perform worse than codebooks chosen by more general Markovian processes.

Vvedenskaya and Dobrushin in [10] attempt to bound the mutual information between the input and output of the i.i.d. deletion channel experimentally, via simulation. They estimate lower bounds for the capacity of the i.i.d. deletion channel for $d \leq 0.3$ using codewords generated by a low-order Markov chain (up to order 2). However, it is not clear that their results give true bounds. The quantities used for the estimation of the mutual information are computed from codewords at most 13 bits long and received sequences at most 8 bits long, but the estimates for the capacity obtained in this manner appear to decrease with the sequence length. Recent work by Kavčić and Motwani [7] also employs the Monte Carlo method for estimating information rates, using much larger simulations and codeword lengths. The lower bounds for the i.i.d. deletion channel reported in [7] are significantly lower than those reported in [10]. Although the bounds in [7] and [10] are not strictly provable (as they rely on Monte Carlo simulation), they strongly suggest that the capacity of the i.i.d. deletion channel is indeed much larger than the lower bounds proven in past theoretical work.

The remainder of the correspondence is organized as follows. In Section II, we describe our framework. In Section III, a general lower bound for the capacity of the i.i.d. deletion channel is presented. In Sections IV and V, we derive specific lower bounds, arising from considering geometric and our suggested alternative (m, M, x) block length distributions for our codebooks, respectively. A discussion of these bounds and the upper bounds provided by Ullman in [9] and Dolgopolov in [5], follows in Section VI. Section VII concludes the correspondence.

II. Description of the Model

Consider a code C with $2^{NR_{\mathcal{C}}}$ binary codewords of length N, where $\mathcal{C}$ refers to the corresponding family of codes and $R_{\mathcal{C}}$ is the rate of $\mathcal{C}$ in bits. Each codeword consists of alternating blocks of zeros and ones and is generated independently by the following stochastic process. The first block of the codeword is a block of zeros with probability $1/2$; it consists of j zeros with probability P_j, where P is a probability distribution over the positive integers with geometrically decreasing tails. More specifically, for real constants $0 < c \leq 1, 0 \leq \alpha < 1$ and an integer constant $U \geq 1$, we require that $P_j \leq c$ for all $1 \leq j \leq U$, and $P_j \leq (1-\alpha) \cdot \alpha^{j-1}$ for all $j > U$. We let $\mathbb{P}$ denote the set of all such valid distributions. Clearly, P has finite mean and variance; in fact it has a well-defined moment generating function on an interval around 0, a fact which we will make use of subsequently.

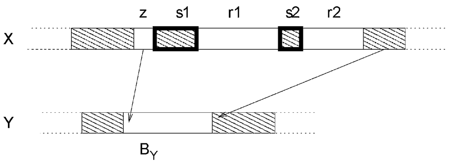

Fig. 1. B_Y in Y arises from blocks z, s_1, r_1, s_2, and r_2 from X. Blank rectangles denote blocks of zeros, while filled rectangles denote blocks of ones. Thick contours indicate that the corresponding blocks are completely deleted. Arrows indicate that at least one bit is not deleted from the blocks they originate from.

We keep generating blocks so that every block is independently assigned an integer length j with probability P_j. Thus, the block lengths are described by i.i.d. random variables, governed by the same distribution P. Moreover, P is symmetric in the sense that there is no difference in the way blocks of zeros and blocks of ones behave. Blocks are generated until the codeword reaches length N; the last block is properly truncated so that the length is N. This truncation does not affect the asymptotic behavior as N grows large, and is ignored henceforth. Applying a standard large deviations bound, we can show that for large N and a specific $\delta = O(N^{-1/3})$, the number of blocks in the codeword is $\frac{N}{\sum_{j\geq 1} jP_j}(1\pm\delta)$ with probability all but super-polynomially small in N (see Proposition 1 in the Appendix). Note that here and throughout the correspondence we use the notation $T(1\pm\tau)$ to refer to a number that is meant to be between $T(1-\tau)$ and $T(1+\tau)$ where the meaning is clear.

In the remainder of the correspondence, we consider a random binary code C that consists of 2^{NR_C} codewords generated as above. We will omit the subscript $\mathcal{C}$ from R where the meaning is clear. We denote by X the transmitted codeword and by Y the sequence received at the end of the channel. Standard letters will be used for quantities related to X, while calligraphic letters will describe quantities related to Y.

When X is transmitted over the deletion channel, the received sequence Y also consists of alternating blocks of zeros and ones. The lengths of these blocks are again i.i.d. random variables according to some distribution $\mathcal{P}$ over the integers. Like P, $\mathcal{P}$ is symmetric with respect to blocks of zeros and ones.

We can express $\mathcal{P}$ in terms of P by reasoning as follows. Consider a block B_Y of $k\geq 1$ zeros in Y, other than the first or the last block[1] (the reasoning is exactly the same for a block of k ones since $\mathcal{P}$ is symmetric). Then B_Y arises from an odd number of blocks in X, starting at a block of zeros and ending in a block of zeros, that are possibly the same. Any blocks of ones between these blocks of zeros in X must be completely deleted (see Fig. 1). We consider a block of zeros in X to be the ending block for B_Y (block r_2 in Fig. 1) if at least one bit is not deleted from its succeeding block of ones in X (therefore, that block of ones in X starts a new block of ones in Y). Following this reasoning, we know that at least one zero is not deleted from the first block of zeros in X used for B_Y (block z in Fig. 1), because this first block is finishing off the block of ones that immediately precedes B_Y in Y.

More formally, a block of $k\geq 1$ zeros in Y arises from $2i+1$ (for some $i\geq 0$) contiguous blocks in X starting at a block of zeros, if the following three conditions hold: a) the i intermediate blocks of ones are completely deleted, b) at least one bit is not deleted from the block of ones in X following the $2i+1$th block, and c) exactly k zeros are not deleted from the $i+1$ blocks of zeros, with at least one of these zeros arising from the first block of zeros. Since the lengths of the blocks in X are i.i.d., we can recursively define the length of the concatenation of *any* m blocks in X as the length of the concatenation of $m-1$ blocks with a single block. In symbols, let $Q_{n,m}$ be the probability that m blocks concatenated have length n. Then

$$Q_{n,m} = \sum_{\ell=1}^{n-m+1} P_\ell \cdot Q_{n-\ell,m-1}, \qquad \text{for } n,m\geq 0$$

with $Q_{0,0}=1$ and $Q_{n,1}=P_n$. Also, let $D=\sum_{j\geq 1} P_j\cdot d^j$ be the probability that a block in X is completely deleted. Restating conditions a)–c) in terms of D and $Q_{n,m}$, the probability that a block in Y consists of $k\geq 1$ bits is given as follows:

$$\mathcal{P}_k = \sum_{i=0}^{\infty}\underbrace{\left\{\sum_{s\geq i} Q_{s,i}d^s\right\}}_{T_0}\underbrace{\left\{1-\sum_{j\geq 1}P_jd^j\right\}}_{T_1} \cdot\underbrace{\sum_{z\geq 1}\sum_{r\geq i} P_z\cdot Q_{r,i}\cdot\frac{d^{z+r-k}\cdot(1-d)^k}{1-\sum_{j\geq 1}P_j\cdot d^j}}_{T_2}\underbrace{\left[\binom{z+r}{k}-\binom{r}{k}\right]}_{T_3} \qquad (2)$$

$$= \left(\frac{1-d}{d}\right)^k\sum_{i=0}^{\infty}D^i\cdot\left\{\sum_{z\geq 1}\sum_{r\geq i}P_z\cdot Q_{r,i}\cdot d^{z+r} \cdot\left[\binom{z+r}{k}-\binom{r}{k}\right]\right\}. \qquad (3)$$

Here and throughout this correspondence, we use the convention that $\binom{n}{k}$ is zero for $k>n$.

The first factor in (2), T_0, corresponds to condition a): it expresses the probability that the i intermediate blocks of ones are completely deleted. The following factor, T_1, corresponds to condition b) and expresses the probability that at least one bit is not deleted from the $2i+2$th block (of ones) in X. The last two factors, T_2 and T_3, refer to the $i+1$ blocks of zeros in X and correspond to condition c). First, T_2 expresses the joint probability that the first block has length z, the other i blocks together have length r, and exactly k of these $z+r$ zeros are not deleted, given that at least one zero (from the k zeros) comes from the first block. Then T_3 counts the number of ways in which these k zeros may be chosen from the $z+r$: any way that includes at least one zero from the first z zeros is acceptable. Finally, (3) is obtained from (2) by observing that T_0 simplifies to D^i, since it is just the probability that i independently generated blocks are deleted.

Just as we require large deviation bounds on the number of blocks in the original codeword, we will want large deviation bounds on the number of blocks in the received sequence as well. From our analysis above, we can now compute the moment generating function of $\mathcal{P}$, denoted by $L(t)$.

Lemma 1: The moment generating function $L(t)$ of the block length distribution $\mathcal{P}$ in the received sequence is given by

$$L(t) = \frac{H(t)-D}{1-H(t)\cdot D}$$

where $H(t)=\sum_{z=1}^{\infty}P_z\cdot(d+(1-d)e^t)^z$ and $D=\sum_{z=1}^{\infty}P_z\cdot d^z$. Moreover, the average block length in the received sequence is given by

$$\sum_k k\mathcal{P}_k = (1-d)\cdot\frac{1+D}{1-D}\cdot\sum_z zP_z. \qquad (4)$$

[1] The first and the last block in Y may not follow $\mathcal{P}$; this does not affect the asymptotic analysis.

Proof: Let $h(t) = d + (1-d)\cdot e^t$ and $H(t) = \sum_{z=1}^{\infty} P_z \cdot h(t)^z$. Let $V_i = \sum_{r=i}^{\infty} Q_{r,i} \cdot h(t)^r, i \geq 0$. Then

$$\begin{aligned} V_i &= \sum_{r=i}^{\infty} Q_{r,i} \cdot h(t)^r \\ &= \sum_{r=i}^{\infty} \sum_{s=i-1}^{r-1} Q_{s,i-1} \cdot h(t)^s \cdot P_{r-s} \cdot h(t)^{r-s} \\ &= \sum_{s=i-1}^{\infty} Q_{s,i-1} \cdot h(t)^s \cdot \sum_{r=s+1}^{\infty} P_{r-s} \cdot h(t)^{r-s} \\ &= \sum_{s=i-1}^{\infty} Q_{s,i-1} \cdot h(t)^s \cdot \sum_{z=1}^{\infty} P_z \cdot h(t)^z \\ &= V_{i-1} \cdot H(t). \end{aligned}$$

It follows that $V_i = H(t)^i$.

Using (3), we obtain for the moment generating function $L(t)$

$$\begin{aligned} \sum_k \mathcal{P}_k \cdot e^{t\cdot k} &= \sum_{i=0}^{\infty} D^i \sum_{z=1}^{\infty} \sum_{r=i}^{\infty} P_z Q_{r,i} \cdot d^{z+r} \\ &\quad \cdot \sum_{k=1}^{\infty} \left(\frac{(1-d)\cdot e^t}{d} \right)^k \left[\binom{z+r}{k} - \binom{r}{k} \right] \\ &= \sum_{i=0}^{\infty} D^i \cdot \sum_{z=1}^{\infty} \sum_{r=i}^{\infty} P_z Q_{r,i} \cdot (h(t)^{r+z} - h(t)^r \cdot d^z) \\ &= \sum_{i=0}^{\infty} D^i \cdot \left(\sum_{r=i}^{\infty} Q_{r,i} h(t)^r \sum_{z=1}^{\infty} P_z h(t)^z \right. \\ &\quad \left. - \sum_{r=i}^{\infty} Q_{r,i} h(t)^r \cdot \sum_{z=1}^{\infty} P_z d^z \right) \\ &= \sum_{i=0}^{\infty} D^i \cdot (H(t)^i \cdot H(t) - H(t)^i \cdot D) \\ &= \sum_{i=0}^{\infty} D^i \cdot [H(t)^{i+1} - D \cdot H(t)^i) \\ &= \frac{H(t) - D}{1 - D \cdot H(t)} \end{aligned}$$

where going from line 1 to line 2 in the preceding equation, we used the fact that the two summations terminate at $z + r$ and r, respectively (since $\binom{n}{k} = 0$ for $k > n$).

To calculate the average block length in Y, we first observe that

$$\begin{aligned} H'(t) &= \left(\sum_{z=1}^{\infty} P_z \cdot h(t)^z \right)' = \sum_{z=1}^{\infty} (P_z \cdot h(t)^z)' \\ &= \sum_{z=1}^{\infty} P_z \cdot z \cdot h(t)^{z-1} \cdot h'(t) \\ &= (1-d)e^t \cdot \sum_{z=1}^{\infty} z \cdot P_z \cdot h(t)^{z-1}. \end{aligned}$$

Hence $H'(0) = (1-d) \sum_{z=1}^{\infty} z \cdot P_z$, and we obtain for $\sum_k k\mathcal{P}_k = L'(0)$

$$\begin{aligned} L'(0) &= \frac{H'(0)(1 - D\cdot H(0)) + D \cdot H'(0) \cdot (H(0) - D)}{(1 - D \cdot H(0))^2} \\ &= \frac{H'(0)(1-D^2)}{(1 - D\cdot H(0))^2} = (1-d) \cdot \frac{1+D}{1-D} \cdot \sum_z z P_z. \qquad \square \end{aligned}$$

III. A Lower Bound for Distributions With Geometrically Decreasing Tails

Before beginning our formalization of the lower bound, we introduce some notation here. Let $\mathcal{N} = N \cdot (1-d)$ and $\mathcal{B} = \frac{\mathcal{N}}{\sum_{k \geq 1} k\mathcal{P}_k}$. We denote by $\mathcal{B}_k$ the number of blocks of length k in Y. Let K be the set of block lengths k such that $\mathcal{P}_k \geq \mathcal{N}^{-1/3}$. A received sequence Y is considered a *typical output* of the channel if for each $k \in K$, it consists of $\mathcal{P}_k \mathcal{B} \cdot (1 \pm \gamma)(1 \pm \delta)(1 \pm \epsilon)$ blocks of length k, for $\epsilon = \delta = \mathcal{N}^{-1/3}$, and $\gamma = \mathcal{N}^{-1/6}$. The choices for K and γ, δ, ϵ are made so that appropriate strong concentration results (to be discussed shortly) hold for each $\mathcal{B}_k$ with $k \in K$; other choices with $\gamma = \delta = \epsilon = o(1)$ and $k \in K$ if and only if $\mathcal{P}_k \cdot \mathcal{B} = \Omega(N^{1-\zeta})$ for a small constant $0 < \zeta < 1$ could guarantee similar results as well. Such concentration results are essential for proving that for appropriate rates, our decoding algorithm fails with exponentially small probability upon reception of a typical output. Finally, we denote by T the set of all typical outputs for code C.

As mentioned in the Introduction, our decoding algorithm is successful if and only if the received sequence is in the set of typical outputs *and* it is the subsequence of exactly one codeword.[2] In the following subsections, we provide upper bounds for the probabilities of the negations of these two events. Specifically, we first show that a received sequence is atypical with probability vanishingly small in N. Then we show that our decoding algorithm fails with probability exponentially small in N for appropriate rates. This gives our lower bound on the capacity.

A. Typical Outputs

The following theorem states that a received sequence Y fails to be a typical output of the channel with probability that goes to zero as N grows large.

Theorem 1: Let Y be the sequence received at the end of the deletion channel when a random codeword X generated as in Section II is transmitted. The probability that Y is not in the set T of the typical outputs is upper-bounded by

$$P_T < e^{-\Theta(N^{1/3})}. \tag{5}$$

Proof: A standard application of Chernoff bounds shows that the received sequence consists of $\mathcal{N} \cdot (1 \pm \epsilon)$ bits, for $\epsilon = \mathcal{N}^{-1/3}$, with probability at least $1 - 2e^{-\frac{\mathcal{N}^{1/3}}{3}}$. Then Proposition 1 in the Appendix guarantees that, conditioned on $\mathcal{N}(1 \pm \epsilon)$ bits in Y and for $\delta = \mathcal{N}^{-1/3}$, the number of blocks in the received sequence Y is $\frac{\mathcal{N}(1\pm\epsilon)}{\sum_k k\mathcal{P}_k}(1 \pm \delta)$ with probability at least $1 - e^{-\Theta(N^{1/3})}$. Finally, for every $k \in K$, a simple application of Chernoff bounds shows that, conditioned on there being $\mathcal{B} \cdot (1 \pm \delta)(1 \pm \epsilon)$ blocks in Y, $\mathcal{B}_k$ is strongly concentrated around its expectation $\mathcal{P}_k \mathcal{B} \cdot (1 \pm \epsilon)(1 \pm \delta)$

$$\begin{aligned} &\mathbf{Pr}\left[|\mathcal{B}_k - \mathcal{P}_k \mathcal{B}(1 \pm \delta)(1 \pm \epsilon)| > \gamma \cdot \mathcal{P}_k \mathcal{B}(1 \pm \delta)(1 \pm \epsilon) \right] \\ &\qquad < 2e^{-\frac{\mathcal{P}_k \mathcal{B}(1\pm\delta)(1\pm\epsilon)\gamma^2}{3}}. \end{aligned}$$

Let $\gamma = \mathcal{N}^{-1/6}$. Since $\mathcal{P}_k \geq \mathcal{N}^{-1/3}$ for all $k \in K$, the probability that there exists at least one $\mathcal{B}_k$ which fails to be as described in the definition of the typical output (conditioned on $\mathcal{B}(1 \pm \delta)(1 \pm \epsilon)$ blocks in Y) is upper-bounded by

$$|K| \cdot 2e^{-\Omega(\mathcal{N}^{1/3})} < \mathcal{N} \cdot 2e^{-\Omega(\mathcal{N}^{1/3})}.$$

The theorem follows. □

[2] Strictly speaking, a received sequence that is atypical does not necessarily constitute a decoding error, since even such a sequence might allow for successful decoding. Hence, declaring an error in this case only yields underestimates for the rate.

B. Decoding Error Probability

We shall now show that upon reception of a typical output (which is the case with all but vanishingly small probability), our decoder fails with probability exponentially small in N for appropriate rates. To this end, we need to upper-bound the probability that any typical output is a subsequence of more than one codeword. We denote this probability by P_S and use an approach similar to [1] for computing it. More specifically, we will first upper-bound the probability that a fixed typical sequence Y (arising from a codeword X) is a subsequence of another random codeword $X' \neq X$ generated as in Section II. As in [1], this argument will be based on the fact that a greedy algorithm matching bits from the left in the received sequence Y to the successively earliest matching bits from the left in X' can determine whether Y is a subsequence of X' or not. A slight difference in our analysis is that we will work with blocks in the received sequence, instead of individual bits in the received sequence as done in [1], since our received sequence is not governed by a first-order Markov chain but by a distribution on block sizes. Since all typical outputs and all codewords share the same structural properties, this will also be the probability that *any* typical output is a subsequence of *any* other random codeword. Then the desired probability P_S follows by a union bound over all codewords.

Let $G_{j,k}$ be the distribution of the number of bits j from a random codeword necessary to cover a single block of length k in Y using this greedy strategy. To be clear, a block of length k in Y may need more than one block from X' to be covered. For example, a block of five zeros in Y may be covered by a block of three zeros followed by an intermediate block of two ones and the another block of seven zeros. In this case, we say that all 12 bits were necessary to cover the block in Y, and the next block of ones in Y will start being covered by the subsequent block of ones in X'. In general, we say that all the bits from the last block used from X' will be used for the block in Y since blocks are alternating. Then G is given by

$$G_{j,k} = \sum_{i=0}^{k-1} \sum_{\substack{i \le r \le k-1 \\ i \le s \le j-k}} Q_{r,i} Q_{s,i} P_{j-r-s}. \tag{6}$$

To see (6), consider a block of k zeros (without loss of generality (w.l.o.g.), since $\mathcal{P}$ is symmetric) in Y. This block will be covered with j bits belonging to $2i+1$ blocks in X', starting at a block of zeros. All together, the first i consecutive blocks of zeros may have length at most $k-1$; otherwise, they would suffice to cover the block in Y. The $i+1$th block of zeros must have length at least 1 and be sufficiently long so that the total number of zeros from all the $i+1$ blocks of zeros is at least k. The concatenation of the i intermediate blocks of ones may have any length between i and $j-k$.

Fix a typical output Y and consider a block of length k in Y. Let J_k denote the number of bits from X' needed to cover it. Then J_k is distributed according to $G_{j,k}$. There are $\mathcal{P}_k\mathcal{B} \cdot (1 \pm \gamma)(1 \pm \delta)(1 \pm \epsilon)$ blocks of length k in Y, for every $k \in K$. The number of bits each of these blocks needs to be covered are i.i.d. random variables. If J^x is the number of bits needed to cover block x in Y, we can use the Chernoff bounds to bound the probability that a randomly generated codeword contains Y as a subsequence as follows:

$$\mathbf{Pr}\left(\sum_{x=1}^{\mathcal{B}} J^x < N\right) \le e^{tN} \left[\prod_{k \in K} \left(E\left[e^{-tJ_k}\right]\right)^{\mathcal{B}_k}\right]. \tag{7}$$

Since Y is a typical sequence

$$\begin{aligned}\mathcal{B}_k &= \frac{\mathcal{P}_k \cdot \mathcal{N}}{\sum_{k \ge 1} k\mathcal{P}_k}(1 \pm \epsilon)(1 \pm \gamma)(1 \pm \delta) \\ &\ge \frac{\mathcal{P}_k \cdot N(1-d)}{\sum_{k \ge 1} k\mathcal{P}_k}(1 - o(1)).\end{aligned}$$

For $k \ge 1$ and $t > 0$

$$E[e^{-tJ_k}] = \sum_{j=k}^{\infty} e^{-tj} G_{j,k} < e^{-t} \cdot \sum_{j=k}^{\infty} G_{j,k} < 1.$$

Then, by a union bound, the probability that Y is a subsequence of more than one codeword is at most

$$\begin{aligned}P_S &< 2^{NR} \cdot e^{tN} \cdot \left[\prod_{k \in K} \left(E\left[e^{-tJ_k}\right]\right)^{\mathcal{P}_k}\right]^{\frac{N(1-d)}{\sum_{k \ge 1} k\mathcal{P}_k}(1-o(1))} \\ &= \left(2^R e^t \left[\prod_{k \in K} \left(E\left[e^{-tJ_k}\right]\right)^{\mathcal{P}_k}\right]^{\frac{(1-d)}{\sum_{k \ge 1} k\mathcal{P}_k}(1-o(1))}\right)^N.\end{aligned}$$

Since the $o(1)$ term in the exponent does not affect the asymptotics, it can be ignored. Hence, for the probability that the decoder fails to go to zero asymptotically it suffices that the expression raised to the Nth power is less than 1. Therefore, we can achieve any rate R (in bits) satisfying

$$R < \sup_{\substack{t>0 \\ P \in \mathbb{P}}} \left[-t \cdot \log e - \frac{1-d}{\sum_{k \ge 1} k\mathcal{P}_k} \log\left\{\prod_{k \in K} \left(E\left[e^{-tJ_k}\right]\right)^{\mathcal{P}_k}\right\}\right] \tag{8}$$

We thus obtain the following theorem for arbitrary distributions $P \in \mathbb{P}$.

Theorem 2: Consider a channel that deletes every transmitted bit independently and with probability d and a binary input alphabet. The capacity of this channel in bits is lower-bounded by

$$C_{\mathrm{del}} \ge \sup_{\substack{t>0 \\ P \in \mathbb{P}}} \left[-t \cdot \log e - \frac{1-d}{\sum_{k \ge 1} k\mathcal{P}_k} \cdot \sum_{k \in K} \mathcal{P}_k \cdot \log\left\{\sum_{j=k}^{\infty} e^{-tj} G_{j,k}\right\}\right] \tag{9}$$

for $\mathcal{P}$ given by (3), $\sum_k k \cdot \mathcal{P}_k$ given by (4), and G given by (6).

While Theorem 2 does not yield a simple closed form, given a specific distribution P, a provable lower bound for the capacity can be evaluated numerically using the theorem. We remark that since $\sum_{j=k}^{\infty} e^{-tj} G_{j,k} < 1$, every logarithm inside the summation over K is negative. This implies that summing over a finite number of k's strictly underestimates the final capacity bounds, and hence K in (9) may be replaced by any subset of finite cardinality. This observation allows numerical calculations to be performed over a finite number of k's, while still providing a provable lower bound.

IV. Geometric Distributions

In the special case where the block lengths in X are geometrically distributed, i.e., $P_j = (1-p)p^{j-1}$, the following corollary to Theorem 2 shows that the lower bounds to the capacity achieved by our framework are always better than the lower bounds obtained in [1].

Corollary 1: Consider a channel that deletes every transmitted bit independently and with probability $0 < d < 1$, a binary input alphabet and geometric block length distribution P. The capacity of this channel is lower-bounded by

$$C_{\mathrm{del}} \ge \sup_{\substack{t>0 \\ 0<p<1}} \left[-t \cdot \log e - (1-d)\log\left(A^{1-q} \cdot B^q\right)\right]$$

where $A = \frac{(1-p)e^{-t}}{1-pe^{-t}}$, $B = \frac{(1-p)^2e^{-2t}}{1-pe^{-t}} + pe^{-t}$ and $q = 1 - \frac{1-p}{1+d-2pd}$. Moreover

$$\begin{aligned}&\sup_{\substack{t>0 \\ 0<p<1}} \left[-t \cdot \log e - (1-d)\log\left(A^{1-q} \cdot B^q\right)\right] \\ &\qquad \ge \sup_{\substack{t>0 \\ 0<p<1}} \left[-(1-d)\log\left\{(1-q)A + qB\right\} - t \cdot \log e\right].\end{aligned}$$

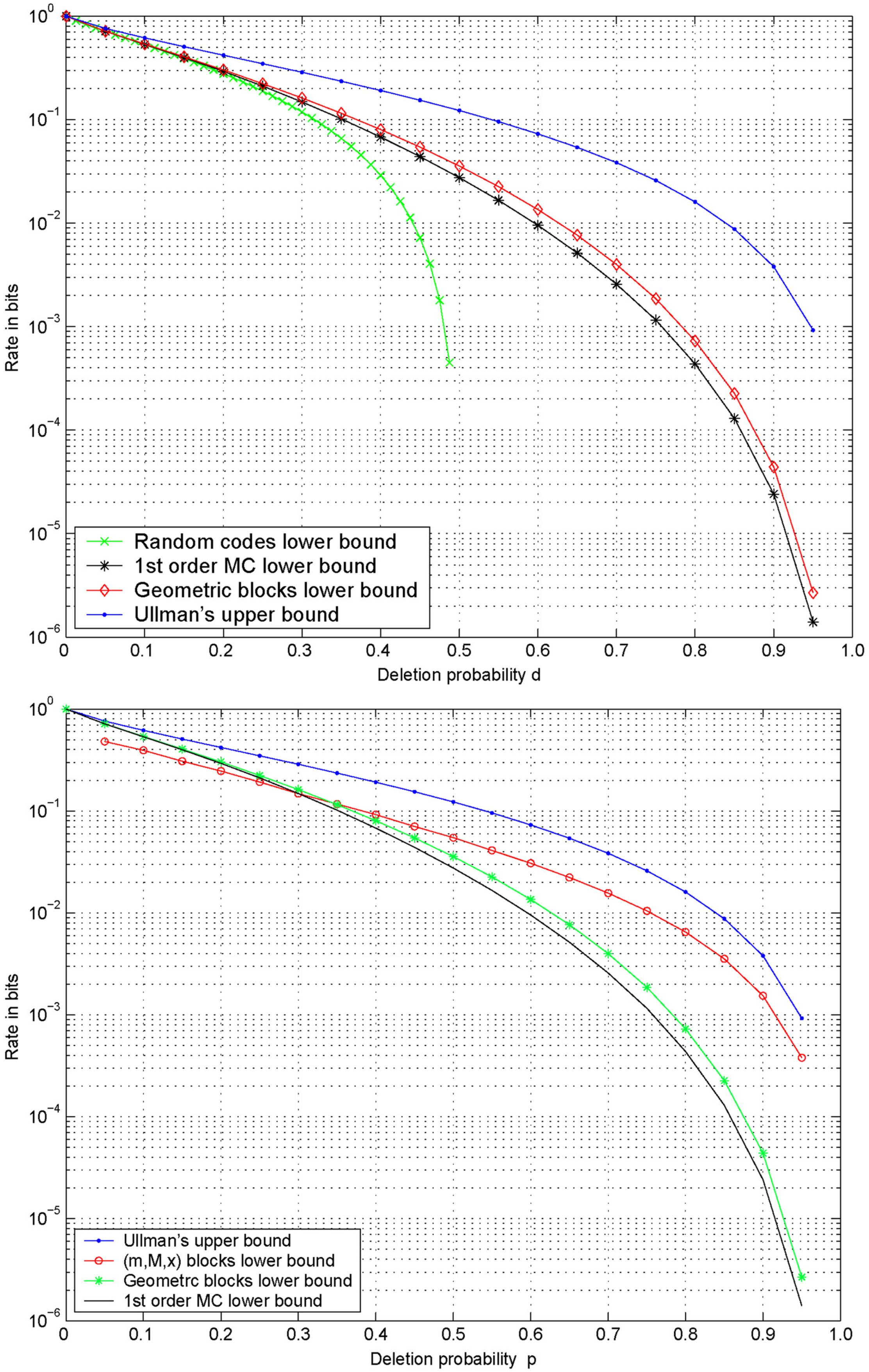

Fig. 2. Improvements in rates with our framework, for geometric and (m, M, x) distributions.

A detailed proof appears in the Appendix. We point out that our proof can also be derived by following the corresponding proof in [1] while using our stronger notion of a typical output. The left graph in Fig. 2 and the numerical results of Table I show resulting underestimates of rates obtained using the corollary. (The results are underestimates as we optimized p up to only two decimal digits for $d < 0.96$ and three decimal digits for $g \geq 0.9$.)

V. Numerically Derived Lower Bounds for the Capacity

We also considered codebooks consisting of codewords with more general block length distributions $P \in \mathbb{P}$. As Theorem 2 does not generally yield bounds as simply expressible as for the case of geometric distributions, our results arise from numerical calculations. In trying alternative block length distrubutions, we first tried to improve on geometric distributions using local-search-based approaches to repeatedly modify the distribution toward a better rate. This approach did appear to lead to minor improvements across the board for all rates, but was extremely slow.

We suspected that for high deletion rates, greater variability in block lengths would lead to improved code rates. We considered the case where a block in X is either assigned a short integer length m with probability x or a larger integer length $M > m$ with probability $1 - x$. We denote such distributions by (m, M, x). The intuition behind this choice is that is provides an alphabet similar to a Morse code for the deletion channel, with short blocks being the equivalent of dots and long blocks being the equivalent of dashes. Of course there is still the possibility that short blocks may be deleted, or short and long blocks may be confused, which limits the posible sending rate.

The upper graph in Fig. 2 shows the improvement (m, M, x) distributions yield over the geometric distribution when $d \geq 0.35$. Because we found no easily computable closed form for the capacity using these distributions, the calculated rates (given in Table II) are underestimates of the best achievable rates, derived as follows.

Let R^d be the rate achieved by the best possible distribution (m, M, x) for the fixed deletion probability d. Similarly, let $R^{d,m}$ be the rate achieved by the best pair (M, x) when m, d are fixed; and $R^{d,m,M}$ be the rate achieved by the best x for fixed M, m, and d. We compute local maxima that approximate these quantities; computation becomes fairly time consuming even for moderate deletion probabilities. Let $\hat{R}^d, \hat{R}^{d,m}$, and $\hat{R}^{d,m,M}$ denote our approximations to $R^d, R^{d,m}$ and $R^{d,m,M}$, respectively. For each d, we only consider a limited number of triplets (m, M, x). Let $\hat{R}^{d,m,M}(x)$ be the rate we compute according to Theorem 2 for the distribution (m, M, x) when the deletion probability is d. Starting at $m = M = 1$ and $x = 0.01$, and successively incrementing x by 0.01, we set $\hat{R}^{d,m,M} = \hat{R}^{d,m,M}(x)$ for the first x that satisfies $\hat{R}^{d,m,M}(x) > \hat{R}^{d,m,M}(x + 0.01)$; we only optimize x over two decimal digits. Similarly, we set $\hat{R}^{d,m} = \hat{R}^{d,m,M}$ for the first M such that $\hat{R}^{d,m,M} > \hat{R}^{d,m,M+1}$. Finally, we set $\hat{R}^d = \hat{R}^{d,m}$ for the first m that satisfies $\hat{R}^{d,m} > \hat{R}^{d,m+1}$. The lower graph in Fig. 2 shows $\hat{R}^d$; clearly, $R^d \geq \hat{R}^d$. Tables I and II explicitly specify the input distribution that achieves $\hat{R}^d$ for geometric and (m, M, x) distributions respectively, for each deletion probability d between 0.05 and 0.95 in increments of 0.05.

Table II suggests that for (m, M, x) distributions, the m value should be chosen so that short blocks are deleted fairly infrequently, and M should be chosen so that small numbers of consecutive small blocks are unlikely to be confused with a long block.

TABLE I
Lower Bounds Based on Codebooks Derived From Geometric Distributions

d	p	Rate R in bits
0.05	0.53	0.715374
0.10	0.56	0.537944
0.15	0.60	0.404950
0.20	0.64	0.302442
0.25	0.70	0.222211
0.30	0.72	0.161947
0.35	0.76	0.115322
0.40	0.79	0.080279
0.45	0.83	0.054412
0.50	0.86	0.035694
0.55	0.89	0.022525
0.60	0.92	0.013545
0.65	0.94	0.007662
0.70	0.96	0.003990
0.75	0.97	0.0018602
0.80	0.98	0.00072753
0.85	0.99	0.00022696
0.90	0.996	0.00004406
0.95	0.999	0.000002678

TABLE II
Lower Bounds Based on Codebooks Derived From (m, M, x) Distributions

d	m	M	x	Rate R in bits
0.05	1	3	0.65	0.479821
0.10	1	3	0.61	0.393384
0.15	1	3	0.57	0.307949
0.20	1	4	0.59	0.246441
0.25	1	5	0.61	0.191807
0.30	1	5	0.56	0.147945
0.35	2	8	0.55	0.116820
0.40	2	10	0.56	0.091924
0.45	3	14	0.54	0.070492
0.50	3	16	0.53	0.054772
0.55	3	19	0.52	0.041009
0.60	4	27	0.51	0.030715
0.65	5	37	0.48	0.022297
0.70	5	46	0.48	0.015642
0.75	7	73	0.45	0.010469
0.80	8	107	0.44	0.006490
0.85	11	191	0.42	0.003543
0.90	16	405	0.40	0.001541
0.95	26	739	0.24	0.000318

VI. Discussion of Our Results

As discussed in the Introduction, the improvement in our bounds as compared to the bounds in [1] is due to two reasons. The left graph in Fig. 2 shows the improvement in the rates due to the stronger definition of the typical output sequence. Here the lengths of the blocks are still geometrically distributed. The right graph in Fig. 2 shows the improvement due to using Morse-code-like block-length distributions. As already discussed, both curves are underestimates of the actual rates achievable by the technique described in Section II. The graph also shows a combinatorial upper bound for the capacity of channels with synchronization errors derived by Ullman in [9]:

$$C_{\text{del}} \leq 1 - (1 + d) \log_2 (1 + d) + d \log_2 (2d) \text{ bits} \qquad (10)$$

where d in his notation is the limit of the fraction of synchronization errors over the block length of the code, as the latter goes to infinity. However, Ullman's bound is based on a channel that introduces $d \cdot N$ insertions only in the first $(1 - d) \cdot N$ bits of the codeword and it is for a codebook with zero probability of error. Hence it does not necessarily constitute an upper bound for the i.i.d. deletion channel, although it has been used as an upper bound for comparison purposes in previous work [1]. In fact, using different techniques, we have recently shown [3] that this bound can be broken in the case of the i.i.d. deletion channel for deletion probability larger than 0.65 and codewords following some first-order Markov chain.

Dolgopolov [5] relies on a theorem by Dobrushin [4] relating the capacity of the i.i.d. channel with synchronization errors to the mutual information between the transmitted codeword and the received sequence to derive upper bounds for the binary i.i.d. deletion channel:

$$C_{\text{del}} \leq \left(1 - \frac{d}{2}\right) \log(2 - d) + \frac{d}{2} \log d \text{ bits.} \tag{11}$$

These bounds hold for codebooks with nonzero probability of error and therefore are closer to the nature of our bounds than Ullman's bounds. However, they rely on an unproven assumption, and arise from considering codebooks consisting of uniformly at random chosen codewords. Hence, it is not surprising that in [3], we show lower bounds on the i.i.d. deletion channel that exceed (11) for $d > 0.8$.

VII. Conclusion

We have presented lower bounds for the capacity of binary deletion channels that delete every transmitted bit independently and with probability d. We suggested using codes that consist of codewords with alternating blocks of zeros and ones; the lengths of these blocks are independently distributed according to the same distribution P over the integers. We both improved the previous lower bound argument for geometrically distributed block lengths and showed better lower bounds using (m, M, x) distributions for $d \geq 0.35$. Our work suggests two ways to continue improving the lower bound for the capacity of the deletion channel. First, we might introduce even more powerful notions of typical outputs that would allow for better analysis. Second, determining better distributions for blocks as a function of d could yield improved results.

Appendix

In this appendix, we provide additional technical details. We first prove a proposition which is key to showing that received sequences are typical with high probability (see Theorem 1).

Proposition 1: Consider a random codeword generated as in Section II. Let $\mu = \sum_j jP_j$ and $\sigma^2 = \sum_j j^2 P_j - \mu^2$. Then for $\delta = \sigma\mu N^{-1/3}$, the number of blocks in X is $\frac{N}{\mu}(1 \pm \delta)$ with probability at least $1 - e^{-\Theta(N^{1/3})}$.

Similarly, for $\mu_Y = \sum_k k\mathcal{P}_k$, $\sigma_Y^2 = \sum_k k^2 \mathcal{P}_k - \mu_Y^2$, $\epsilon = \mathcal{N}^{-1/3}$, and $\delta = \mathcal{N}^{-1/3}$, the number of blocks in the received sequence Y is $\frac{\mathcal{N}(1 \pm \epsilon)}{\mu_Y}(1 \pm \delta)$ with probability at least $1 - e^{-\Theta(N^{1/3})}$.

Proof: Let $Z_i, 1 \leq i \leq \frac{N}{\mu}$ be i.i.d. random variables, each distributed according to P, with $E[Z_i] = \mu$ and $\text{Var}(Z_i) = \sigma^2$. Let $W_i, 1 \leq i \leq \frac{N}{\mu}$ be i.i.d. random variables, such that $W_i = Z_i - \mu$. Then $E[W_i] = 0$ and $\text{Var}(W_i) = \sigma^2$. Recall by our definitions that since $P \in \mathbb{P}$, there exist constants U, α, and c such that $P_j \leq c$ for all $1 \leq j \leq U$, and $P_j \leq (1 - \alpha) \cdot \alpha^{j-1}$ for all $j > U$. A simple calculation yields that the moment generating function of W_i is well defined in an interval around 0; specifically

$$\begin{aligned} E\left[e^{tW_i}\right] &\leq e^{-t\mu} \left(\sum_{j=1}^{U} ce^{tj} + \sum_{j>U} (1 - \alpha)\alpha^{j-1} e^{tj} \right) \\ &= e^{-t\mu} \left(ce^t \frac{e^{tU} - 1}{e^t - 1} + \frac{(1 - \alpha)e^t (\alpha e^t)^U}{1 - \alpha e^t} \right). \end{aligned}$$

We can therefore apply standard large deviation bounds; specifically, we apply [6, p. 553, eq. (7.28)], or alternatively the form corresponding to Theorem 5.23 in [8, p. 178] (and the corresponding equations in [8, p. 183]) with the value $x = N^{1/6}$. This immediately yields

$$\begin{aligned} &1 - \mathbf{Pr}\left[\frac{\sum_{i=1}^{N/\mu} W_i}{\sigma\sqrt{N}} < N^{1/6} \right] \\ &\quad = \left(1 - \Phi(N^{1/6})\right) \cdot e^{\frac{E[W_i^3]}{6\sigma^3}} \left(1 + O(N^{-1/3})\right) \\ &\quad \Rightarrow \mathbf{Pr}\left[\sum_{i=1}^{N/\mu} Z_i - N < \sigma N^{2/3} \right] > 1 - e^{-\Theta(N^{1/3})} \end{aligned}$$

where $\Phi(x)$ is the standard normal distribution and the result follows from the inequality $1 - \Phi(x) < \frac{1}{x\sqrt{2\pi}} e^{-x^2/2}$ and the fact that $E[W_i^3]$ and σ^2 are finite. An entirely symmetric bound for $x = -N^{1/6}$ from the same equation gives

$$\mathbf{Pr}\left[\sum_{i=1}^{N/\mu} Z_i - N > -\sigma N^{2/3} \right] > 1 - e^{-\Theta(N^{1/3})}.$$

We conclude that asymptotically

$$\mathbf{Pr}\left[\left| \sum_{i=1}^{N/\mu} Z_i - N \right| \geq \sigma N^{2/3} \right] < e^{-\Theta(N^{1/3})}. \tag{12}$$

Since each block has length at least 1, (12) implies that with probability at least $1 - e^{-\Theta(N^{1/3})}$, $N/\mu - \sigma N^{2/3}$ blocks result in total length less than N while $N/\mu + \sigma N^{2/3}$ blocks result in total length greater than N. Since μ and σ are both finite, we conclude that for $\delta = \mu\sigma N^{-1/3}$, the number of blocks in X is $\frac{N}{\mu}(1 \pm \delta)$ with probability at least $1 - e^{-\Theta(N^{1/3})}$.

The second part of the proposition is quite similar. First, we note that $|Y| = \mathcal{N}(1 \pm \epsilon)$ with probability at least $1 - e^{-\Theta(N^{1/3})}$, for $\epsilon = \mathcal{N}^{-1/3}$, by standard Chernoff bounds. With this we need only show that the random variables $W_i = Z_i - \sum_k k\mathcal{P}_k$ have a well-defined moment-generating function in an interval around 0, where the Z_i's are distributed according to the block length distribution $\mathcal{P}$ in the received sequence. Lemma 1 gives the moment-generating function $L(t)$ of $\mathcal{P}$. Again, since $P \in \mathbb{P}$, there exist constants U, α, and c such that $P_j \leq c$ for all $1 \leq j \leq U$, and $P_j \leq (1 - \alpha) \cdot \alpha^{j-1}$ for all $j > U$. Hence,

$$\begin{aligned} D = \sum_{z=1}^{\infty} P_z d^z &\leq \sum_{z=1}^{U} cd^z + \sum_{z>U} (1 - \alpha)\alpha^{z-1} d^z \\ &\leq cd \frac{d^U - 1}{d - 1} + \frac{(1 - \alpha)d(\alpha d)^U}{1 - \alpha d}. \end{aligned}$$

Similarly

$$H(t) = \sum_{z=1}^{\infty} P_z h(t)^z < c \cdot h(t) \cdot \frac{h(t)^U - 1}{h(t) - 1} + \frac{(1 - \alpha)h(t)(\alpha h(t))^U}{1 - \alpha h(t)}$$

where $h(t) = (1 - d) \cdot e^t + d$. (Note that $H(0) = 1$.) It follows that the moment-generating function $L(t)$ is finite in a neighborhood around 0. A similar argument to that used for the codeword X now shows the number of blocks in the received sequence is $\frac{\mathcal{N}(1 \pm \epsilon)}{\mu_Y}(1 \pm \delta)$ with probability at least $1 - e^{-\Theta(N^{1/3})}$, and the lemma easily follows. □

We now provide the necessary technical arguments behind Corollary 1, which we repeat below.

Corollary 1: Consider a channel that deletes every transmitted bit independently and with probability $0 < d < 1$, a binary input alphabet, and geometric block length distribution P. The capacity of this channel is lower-bounded by

$$C_{\text{del}} \geq \sup_{\substack{t>0 \\ 0<p<1}} \left[-t \cdot \log e - (1-d)\log\left(A^{q} \cdot B^{1-q}\right)\right]$$

where $A = \frac{(1-p)e^{-t}}{1-pe^{-t}}$, $B = \frac{(1-p)^2 e^{-2t}}{1-pe^{-t}} + pe^{-t}$ and $q = 1 - \frac{1-p}{1+d-2pd}$. Moreover

$$\sup_{\substack{t>0 \\ 0<p<1}} \left[-t \cdot \log e - (1-d)\log\left(A^{1-q} \cdot B^{q}\right)\right] \geq \sup_{\substack{t>0 \\ 0<p<1}} \left[-(1-d)\log\{(1-q)A + qB\} - t \cdot \log e\right].$$

Proof: Let $P_j = (1-p)p^{j-1}$. We start by deriving closed formulas for distributions Q and G when the block lengths in X are distributed according to P. We present the intuitive explanations of the corresponding formulas whenever possible instead of their mathematical derivations.

Fact 1: The probability that i blocks concatenated have length r is given by $Q_{r,i} = (\begin{smallmatrix} r-1 \\ i-1 \end{smallmatrix})p^{r-i}(1-p)^i$.

Proof: If $r = i = 0$, then $Q_{0,0} = 1$, since zero blocks concatenated have length 0. For $r < i$, $Q_{r,i} = 0$, since a block consists of at least one bit. For $i \geq 1$ and $r \geq i$, it is easy to observe that $Q_{r,i}$ is indeed given by Fact 2: from a total of r bits, there are $r-1$ choices to place the last bits of $i-1$ blocks (the ith block is fixed to end at the rth bit). The last factor $(p^{r-i}(1-p)^i)$ corresponds to the individual probability of each of the aforementioned $(\begin{smallmatrix} r-1 \\ i-1 \end{smallmatrix})$ equiprobable arrangements. □

Fact 2: The probability that it required j bits from X to cover a block of k bits in Y is given by

$$G_{j,k} = (1-p)p^{j-1} \sum_{i=0}^{k-1} \binom{k-1}{i} \binom{j-k}{i} \left(\frac{1-p}{p}\right)^{2i}. \tag{13}$$

Proof: Consider a block of k zeros in Y, denoted by B_Y (the reasoning is exactly the same for a block of k ones, since $\mathcal{P}$ is symmetric). In order to cover B_Y, $2i+1$ blocks from X are used, for $0 \leq i \leq k-1$, starting at a block of zeros and ending in a block of zeros. The total length of the first i blocks of zeros is at most $k-1$ (otherwise, they would suffice to cover B_Y). Then the number of ways to choose the lengths of these i blocks is given by the number of nonnegative integer solutions to the inequality $l_1 + l_2 + \cdots + l_i < k$, with $l_j \geq 1$ for $1 \leq j \leq i$. The latter equals $\binom{k-1}{i}$.

Similarly, there are $\binom{j-k}{i}$ ways to choose the lengths for the i blocks of ones (their total length may range from i up to $j-k$). Finally, $p^{j-2i-1}(1-p)^{2i+1}$ yields the probability of an individual arrangement of $2i+1$ blocks with total length j. □

The following combinatorial fact will be directly applied to (8) to upper-bound the rate.

Fact 3: Let $\mathcal{P}_k = (1-q) \cdot q^{k-1}$ for $q = 1 - \frac{1-p}{1+d-2pd}$, G given by (13), $\kappa = \frac{1-p}{p+(1-2p)e^{-t}}$, and $\lambda = \frac{e^{-t}}{1-pe^{-t}} \cdot (p + (1-2p)e^{-t})$. Then

$$\prod_{k \in K} \left\{\sum_{j=k}^{\infty} G_{j,k} \cdot e^{-tj}\right\}^{\mathcal{P}_k} = \kappa^{1-o(1)} \cdot \lambda^{\frac{1}{1-q}-o(1)}.$$

Proof: We first derive a closed form for $\sum_{j=k}^{\infty} G_{j,k} \cdot e^{-tj}$ in terms of κ and λ

$$\begin{aligned}
\sum_{j=k}^{\infty} G_{j,k} \cdot e^{-tj} &= \sum_{j=k}^{\infty} p^{j-1}(1-p) \\
&\quad \cdot \sum_{i=0}^{k-1} \binom{k-1}{i} \binom{j-k}{i} \left(\frac{1}{p}-1\right)^{2i} \cdot e^{-tj} \\
&= \frac{1-p}{p} \sum_{i=0}^{k-1} \binom{k-1}{i} \left(\frac{1}{p}-1\right)^{2i} \\
&\quad \cdot \sum_{j=k}^{\infty} \binom{j-k}{i} (pe^{-t})^j \\
&= \frac{1-p}{p} \frac{(pe^{-t})^k}{1-pe^{-t}} \\
&\quad \cdot \sum_{i=0}^{k-1} \binom{k-1}{i} \left(\left(\frac{1}{p}-1\right)^2 \cdot \frac{pe^{-t}}{1-pe^{-t}}\right)^i \\
&= \frac{1-p}{p} \frac{(pe^{-t})^k}{1-pe^{-t}} \\
&\quad \cdot \left(1 + \left(\frac{1}{p}-1\right)^2 \cdot \frac{pe^{-t}}{1-pe^{-t}}\right)^{k-1} \\
&= \frac{1-p}{p} \left(\frac{pe^{-t}}{1-pe^{-t}}\right)^k \\
&\quad \cdot \left(1 + \left(\frac{1}{p}-2\right)e^{-t}\right)^{k-1} \\
&= \kappa \cdot \lambda^k.
\end{aligned}$$

It follows that

$$\begin{aligned}
\prod_{k \in K} (\kappa \cdot \lambda^k)^{q^{k-1} \cdot (1-q)} &= \left\{\kappa^{\sum_{k \in K} q^k} \lambda^{\sum_{k \in K} k \cdot q^k}\right\}^{\frac{1-q}{q}} \\
&= \Bigg\{\kappa^{q\frac{1-q^{O(\log \mathcal{N})}}{1-q}} \\
&\qquad \cdot \lambda^{\frac{q}{(1-q)^2} - \frac{O(\log \mathcal{N}) \cdot q^{O(\log \mathcal{N})}}{(1-q)^2}}\Bigg\}^{\frac{1-q}{q}} \\
&= \kappa^{1-o(1)} \cdot \lambda^{\frac{1}{1-q}-o(1)}
\end{aligned}$$

since for the geometric distribution, $\mathcal{P}_k \geq \mathcal{N}^{-1/3}$ implies that K consists of all block lengths up to $\frac{\log \mathcal{N}}{-3 \log q}$. □

Plugging Fact 3 into (8) (and ignoring the $o(1)$ terms which do not affect the asymptotic nature of the latter), we get

$$\begin{aligned}
R &< \sup_{\substack{t>0 \\ 0<p<1}} \left[-t \cdot \log e - \frac{1-d}{1/(1-q)} \log\left(\kappa \cdot \lambda^{\frac{1}{1-q}}\right)\right] \\
&= \sup_{\substack{t>0 \\ 0<p<1}} \left[-t \cdot \log e - (1-d)\log\left(\kappa^{1-q} \cdot \lambda\right)\right].
\end{aligned}$$

It is easy to check that

$$\begin{aligned}
\kappa^{1-q} \cdot \lambda &= \frac{(1-p)^{1-q} e^{-t} (p + (1-2p)e^{-t})^q}{1-pe^{-t}} \\
&= A^{1-q} \cdot B^q
\end{aligned}$$

Thus,

$$R < \sup_{\substack{t>0 \\ 0<p<1}} \left[-t \cdot \log e - (1-d)\log\left(A^{1-q} \cdot B^q\right)\right]. \tag{14}$$

Recall that the rate R^1 computed in [1] is bounded by

$$R^1 < \sup_{\substack{t>0\\0<p<1}} \left[-t\cdot\log e - (1-d)\log\left((1-q)A+qB\right)\right]$$

(see (1)). Let $t^* > 0$ be such that for fixed $d, p, R^1(t^*)$ is maximized. Then for all $0 \leq q \leq 1, A^q \cdot B^{1-q} \leq qA + (1-q)B$ by convexity. Hence, we conclude that

$$R \geq R(t^*) \geq R^1(t^*).$$

In fact, the optimization of (14) for fixed d, p has a closed form since it results in a quadratic equation in t (similar to (1)). □

Acknowledgment

The authors would like to thank the anonymous referees for valuable comments and pointers to relevant prior work.

References

[1] S. Diggavi and M. Grossglauser, "On information transmission over a finite buffer channel," *IEEE Trans. Inf. Theory*, vol. 52, no. 3, pp. 1226–1237, Mar. 2006.

[2] E. Drinea and M. Mitzenmacher, "On lower bounds for the capacity of deletion channels," in *Proc. IEEE Int. Symp. Information Theory*, Chicago, IL, Jun./Jul. 2004, p. 227.

[3] E. Drinea and M. Mitzenmacher, "Improved lower bounds for i.i.d. deletion channels," in *Proc. 42nd Annu. Allerton Conf. Communication, Control and Computing*, Monticello, IL, Oct. 2004.

[4] R. L. Dobrushin, "Shannon's theorems for channels with synchronization errors," *Probl. Inf. Transm.*, vol. 3, no. 4, pp. 11–26, 1967, Translated from *Probl. Pered. Inf.*, vol.3, no. 4, pp. 18–36, 1967.

[5] A. S. Dolgopolov, "Capacity bounds for a channel with synchronization errors," *Probl. Inf. Transm.*, vol. 26, no. 2, pp. 111–120, 1990, Translated from *Probl. Pered. Inform.*, vol. 26, no. 2, pp. 27–37, Apr./Jun. 1990.

[6] W. Feller, *An Introduction to Probability Theory and its Applications*, 2nd ed. New York: Wiley, 1971, vol. 2.

[7] A. Kavčić and R. Motwani, "Insertion/deletion channels: Reduced-state lower bounds on channel capacities," in *Proc. IEEE Int. Symp. Information Theory*, Chicago, IL, Jun./Jul. 2004, p. 229.

[8] V. V. Petrov, *Limit Theorems of Probability Theory*. Oxford, U.K.: Clarendon, 1995.

[9] J. D. Ullman, "On the capabilities of codes to correct synchronization errors," *IEEE Trans. Inf. Theory*, vol. IT-13, no. 1, pp. 95–105, Jan. 1967.

[10] N. D. Vvedenskaya and R. L. Dobrushin, "The computation on a computer of the channel capacity of a line with symbol drop-out," *Probl. Inf. Transm.*, vol. 4, no. 3, pp. 76–79, 1968, Translated from *Probl. Pered. Inform.*, vol. 4, pp. 92–95, 1968..

A Simple Lower Bound for the Capacity of the Deletion Channel

Michael Mitzenmacher, *Member, IEEE*, and Eleni Drinea

***Abstract*—We present a simple proof that the capacity of the binary independent and identically distributed (i.i.d.) deletion channel, where each bit is deleted independently with probability d, is at least $(1-d)/9$, by developing a correspondence between the deletion channel and an insertion/deletion channel that we call a *Poisson-repeat channel*.**

***Index Terms*—Binary deletion channel, channel capacity, insertion and/or deletion channels.**

I. Introduction

In this work, we consider a natural correspondence between the binary independent and identically distributed (i.i.d.) deletion channel (referred to henceforth simply as the *deletion channel*), where a fixed number of bits n are transmitted and each is deleted independently with probability d, and a simple insertion/deletion channel that we call a *Poisson-repeat channel*. Based on this correspondence, we are able to conclude that the capacity of the deletion channel in bits, which we denote here by C_d, is at least $0.1185 \cdot (1-d)$ for every $d, 0 < d < 1$. We prefer to write this in the simpler form

$$C_d \geq (1-d)/9$$

to emphasize that this bound is within a constant factor of the trivial upper bound on the capacity of $(1-d)$ (based on the capacity of the binary erasure channel) for all d. As far as we can tell, no previous work has given a capacity lower bound that is within a fixed constant factor of $(1-d)$. Our approach also naturally generalizes to larger alphabets, but for this work we focus on the binary case.

The deletion channel has been the subject of recent study. The best lower bounds known for the capacity arise from an argument of Drinea and Mitzenmacher [2], [3], which we apply here to lower-bound the capacity of the Poisson-repeat channel. For deletion channels with larger alphabets, the work of Diggavi and Grossglauser [1] gives the best general capacity bounds. For more information and background, see [2], [3].

II. The Poisson-Repeat Channel

We define a Poisson-repeat channel with parameter λ as follows: the input is a binary string of length n. As each bit passes through the channel, it is replaced by a discrete Poisson number of copies of that bit, where the number of copies has mean λ and is independent for each bit. Notice that some bits will be replaced by 0 copies. The receiver obtains the concatenation of the bits output by the channel.

We use basic facts about the Poisson distribution that can be found in standard texts (see, e.g., [4]). For example, the sum of a constant number of independent random variables with a Poisson distribution also has a Poisson distribution; similarly, if we have a number of items

Manuscript received February 14, 2006; revised June 13, 2006. The work of E. Drinea is supported by the National Science Foundation under Grant CCR-0118701. The work of M. Mitzenmacher is supported by the National Science Foundation under Grants CCR-9983832, CCR-0118701, CCR-0121154, and an Alfred P. Sloan Research fellowship.

The authors are with the Division of Engineering and Applied Sciences, Harvard University, Cambridge, MA 02138 USA (e-mail: michaelm@eecs.harvard.edu; edrinea@deas.harvard.edu).

Communicated by R. J. McEliece, Associate Editor for Coding Theory.
Digital Object Identifier 10.1109/TIT.2006.881844