On lower bounds for the capacity of deletion channels

Eleni Drinea and Michael Mitzenmacher

TR-07-04

Computer Science Group
Harvard University
Cambridge, Massachusetts

On lower bounds for the capacity of deletion channels

Eleni Drinea
Harvard University
edrinea@deas.harvard.edu

Michael Mitzenmacher
Harvard University
michaelm@eecs.harvard.edu

Abstract

We consider binary deletion channels, where bits are deleted independently with probability d. We impove upon the framework used to analyze the capacity of binary deletion channels established by Diggavi and Grossglauser [1], improving on their lower bounds. Diggavi and Grossglauser considered codebooks with codewords generated by a first order Markov chain. Our improvements arise from two considerations. First, Diggavi and Grossglauser consider typical outputs, where an output is typical if an n bit input gives an $n(1-d)(1-\epsilon)$ bit output. We use a stronger notion of a typical output from the channel, which yields better bounds even for the cases studied by Diggavi and Grossglauser. Second, we consider outputs generated by more general processes than first order Markov chains.

1 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. An upper bound for the capacity C_{del} of binary deletion channels is provided by Ullman in [6]:

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

where d is the limit of the fraction of deletion errors over the block length of the code, as the latter goes to infinity.

In this work, we are interested in lower bounds for the capacity of binary deletion channels where bits are deleted independently with probability d. Diggavi and Grossglauser [1] have shown that random codes (i.e., codes consisting of codewords chosen independently at random from the set of all possible codewords of a certain length) yield a lower bound of

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

where $H(d) = -d\log_2 d - (1-d)\log_2(1-d)$ is the binary entropy function. This bound coincides with previous bounds (as discussed in [1]) and can be generalized to stationary and ergodic deletion processes.

The insight of Diggavi and Grossglauser is to consider random codes using a more sophisticated means of choosing the codewords. Specifically, the consider codes consisting of codewords of length N generated by a symmetric first-order Markov process with transition probability p. 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 $1-p$. The decoding algorithm they consider takes a received sequence and determines if it is a subsequence of exactly 1 codeword; if this is the case, the decoder is successful, and otherwise, the decoder fails. Using this decoder, they determine for what deletion probability d the probability of error goes to 0 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:

$$C_{\text{del}} \geq \sup_{\substack{t>0 \\ 0<p<1}} \left[-(1-d)\log\{(1-q)A + qB\} - t\right] = C^1_{\text{del}} \quad \text{nats} \tag{1}$$

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

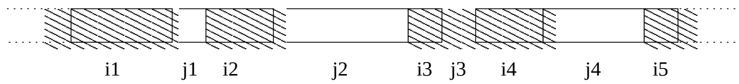

Figure 1: Blocks i_1, i_2, i_3, i_4, i_5 are **consecutive**; blocks j_1, ..., j_4 are the intermediate blocks

In this paper, we 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. Note that the probability of an atypical output is exponentially small. By using a stronger notion of a typical output, 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 integers. (The last block, strictly speaking, may need to be truncated to match the codeword length N; this consideration does not affect the asymptotics and is ignored henceforth.) Note that 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. We suggest some simple distributions for the block lengths that provide better lower bounds when the deletion probability is at least 0.4, 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.

2 Description of the model

Consider a code C with $e^{NR_\mathcal{C}}$ 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 nats. 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, for $1 \leq j \leq M_P$ and M_P being a suitable integer constant. The restriction that blocks lengths are bounded by a constant is for notational convenience only; we can also work with distributions with suitably decreasing tails, such as the geometric distribution, and in fact we give an improved result for the case of the geometric distribution subsequently. We keep generating blocks until length N is reached or exceeded. Every block is independently assigned an integer length j between 1 and M_P with probability P_j. Thus the block lengths are described by independent and identically distributed random variables, governed by the same finite distribution P. Moreover, P is symmetric in the sense that there is no discrimination in the way blocks of zeros and blocks of ones behave. No more blocks are generated when the codeword reaches length N; if N has been exceeded, the last block is properly truncated (this does not affect the asymptotics as N grows large). Applying standard results from renewal theory, we can show that for large N and $\delta = o(1)$, the number of blocks in the codeword, denoted by B, is w.h.p. $\frac{N}{\sum_i iP_i}(1 \pm \delta)$. (The proof appears in the full version of the paper.)

In the rest of the paper, we consider a random binary code C that consists of $e^{NR_\mathcal{C}}$ 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 received version of the codeword. Normal letters will be used for quantities related to X, while calligraphic letters will describe quantities related to Y. To simplify our discussion, we call m blocks of zeros (ones) **consecutive** if there are exactly $m-1$ intermediate blocks of ones (zeros) in between these blocks (see Figure 1).

Assume that the block lengths in X are generated according to some finite distribution P over the integers that is symmetric with respect to blocks of zeros and blocks of ones. Let $\mathbb{P}$ denote the set of all such distributions. Then Y also consists of blocks whose lengths are generated according to some distribution $\mathcal{P}$ over the integers. From now on, we denote by $\mathcal{P}$ the truncated at $M_{\mathcal{P}}$ version of the actual distribution; that is, $\mathcal{P}_i = 0$ for $i > M_{\mathcal{P}}$ for some integer constant $M_{\mathcal{P}}$. Again, this restriction is simply to ease notation and exposition. Like P, $\mathcal{P}$ is symmetric with respect to blocks of zeros and blocks of ones.

We can express $\mathcal{P}$ in terms of P reasoning as follows. For $i \geq 1$, consider a block in Y (other than the first or last blocks), that consists of i zeros. Call this block B_Y; then B_Y arises from some odd number of blocks in X, starting at a block of zeros and ending in a block of zeros (that are possibly the same). The reason why a block of zeros in X is considered the ending block for B_Y is that at least one bit is not deleted from the block of ones that succeeds in X this block of zeros. Following this reasoning, at least one zero is not deleted from the first block of zeros in X used for B_Y, because this block is finishing off the block of ones in Y that immediately precedes B_Y. So this first block of zeros contributes at least one zero to B_Y; yet, it might contribute as many as i zeros. This implies that, if λ bits are not deleted from this first block of zeros, then exactly $i - \lambda$ bits must not deleted from the last w blocks of zeros (which have arbitrary lengths). Of course all intermediate blocks of ones (of arbitrary length again) must be completely deleted. Formally, a block of i zeros in Y arises from $w + 1$ consecutive blocks of zeros in X, for some nonnegative integer w, if the following two conditions hold: (a) $1 \leq \lambda << \leq i$ bits are not deleted from the first block of zeros, and exactly $i - \lambda \geq 0$ bits are not deleted from the last w blocks of zeros, and (b) the intermediate w blocks of ones are completely deleted and at least one bit is not deleted from the block of ones following the $w + 1$st block of zeros.

Since the lengths of the blocks in X are independently and identically distributed, we can recursively define the length of the concatenation of *any* k blocks in X as the length of the concatenation of $k - 1$ blocks with a single block. In symbols, let $Q_{s,k}$ be the probability that k blocks concatenated have length s. Then $Q_{s,k} = \sum_{\ell=1}^{k-1} P_\ell \cdot Q_{s-\ell,k-1}$, for $s, k \geq 0$ with $Q_{0,0} = 1$, $Q_{s,1} = P_s$ and $\sum_s Q_{s,k} = 1$. Restating conditions (a) and (b) in terms of the distribution Q, the formula for the distribution of block lengths in Y is given below

$$\mathcal{P}_i = \sum_{w=0}^{\infty} \underbrace{\left\{\sum_{j\geq w} Q_{j,w} d^j\right\}}_{T_0} \underbrace{\left\{1 - \sum_{j\geq 1} P_j d^j\right\}}_{T_1} \sum_{\lambda=1}^{i} \underbrace{\frac{\sum_{\ell\geq 1} P_\ell d^\ell \left(\frac{1-d}{d}\right)^\lambda \binom{\ell}{\lambda}}{1 - \sum_{\ell\geq 1} P_\ell d^\ell}}_{T_2} \cdot \underbrace{\sum_{k\geq w} Q_{k,w} d^k \left(\frac{1-d}{d}\right)^{i-\lambda} \binom{k}{i-\lambda}}_{T_3} \quad (2)$$

$$= \left(\frac{1-d}{d}\right)^i \sum_{w=0}^{\infty} \left\{\sum_{j\geq w} Q_{j,w} d^j\right\} \sum_{\lambda=1}^{i} \left\{\sum_{\ell\geq 1} P_\ell d^\ell \binom{\ell}{\lambda} \cdot \sum_{k\geq w} Q_{k,w} d^k \binom{k}{i-\lambda}\right\} \quad (3)$$

To see (2), consider a block of i zeros in Y (the reasoning is exactly the same for a block of i ones, since P is symmetric). $2w + 1$ blocks of zeros and ones from X are used to give rise to this block. T_0 corresponds to the w intermediate blocks of ones that are completely deleted; T_1 is the probability that at least one bit is not deleted from the $2w + 2$nd block (of ones), so that the current block of zeros in Y is generated from exactly $2w + 1$ blocks from X. T_2 refers to the first block of zeros from X that we use: it expresses the probability that exactly λ (with $1 \leq \lambda \leq i$) zeros are not deleted from this block, conditioned on the event that at least 1 bit is not deleted from this block (as described previously, we are certain that at least one bit is not deleted from this first block). Then the numerator in T_2 is the probability that exactly λ bits are not deleted from the first block (with $1 \leq \lambda \leq i$, while the denominator is the probability that at least one bit is not deleted from this block. Finally T_3 corresponds to the last w blocks of zeros from which exactly $i - \lambda$ bits are not deleted. Crossing out T_1 with the denominator of T_2 and grouping together the bits that are deleted and not deleted from T_2 and T_3, yield (3).

3 A lower bound for finite distributions

Let $\mathcal{N} = N(1 - d)$, $\mathcal{K} = \frac{\mathcal{N}(1\pm\epsilon)}{\sum_i i\mathcal{P}_i}$ and $\epsilon, \delta, \gamma = o(1)$ (specifically we will think of $\epsilon, \delta, \gamma = \mathcal{N}^{1/3}$). A received sequence Y is considered a **typical output** of the channel if it satisfies all of the following: it consists of $\mathcal{N}(1 \pm \epsilon)$ bits, $\mathcal{B} = \mathcal{K}(1 \pm \delta)$ alternating blocks of zeros and ones and $\mathcal{P}_i \cdot \mathcal{B}(1 \pm \gamma)$ blocks of length i for $1 \leq i \leq M_{\mathcal{P}}$.

We denote by T the set of all typical outputs for code C, by P_T the asymptotic typical output error probability and by P_{col} the asymptotic collision error probability. If P_e is the asymptotic probability that the decoding algorithm fails, then

$$P_e \le P_T + P_{\rm col}$$

In the following subsections we provide bounds for P_T and $P_{\rm col}$ We show that P_T is exponentially small in N, and $P_{\rm col}$ is exponentially small in N for appropriate rates. This gives our lower bound on the capacity.

3.1 Typical output error probability

The following theorem states that a received sequence Y is a typical output of the channel with high probability. The proof appears in the full version of the paper, and follows from a standard application of Chernoff bounds.

Theorem 1 *Let Y be the sequence received at the end of the deletion channel when a random codeword $X \in C \subset \{0,1\}^N$ (generated as in Section 2) was transmitted. For $\epsilon, \delta, \gamma = o(1)$, the probability that Y is in the set T of the typical outputs is lower bounded by*

$$\mathbf{Pr}[Y \in T] \ge (1 - 2e^{-\frac{\mathcal{N}\epsilon^2}{3}})(1 - 2e^{-\Theta(\mathcal{N})\delta^2})(1 - M_{\mathcal{P}} \cdot 2e^{-\Theta(\mathcal{N})\gamma^2}) \tag{4}$$

Consider $\epsilon = \delta = \gamma = \mathcal{N}^{-\frac{1}{3}}$. Then Theorem 1 guarantees that $P_T \le (4 + M_{\mathcal{P}})e^{-\Theta(\mathcal{N}^{1/3})}$. Thus the asymptotic typical output error probability goes to 0 exponentially fast.

3.2 Collision error probability

We use an approach similar to the argument in [1] for computing the asymptotic collision error probability $P_{\rm col}$. For the random code C and our decoding algorithm $\mathcal{D}$, $P_{\rm col}$ is given by

$$P_{col} \quad = \quad \sum_{x \in C} \mathbf{Pr}[X = x \wedge \mathcal{D}(Y) \ne x] \le e^{NR} \cdot \mathbf{Pr}[Y \in T \text{ is a subsequence of } X\ ] \tag{5}$$

Clearly, the probability that Y is a subsequence of more than one codeword can not increase with its length. Therefore the collision error probability is upper bounded by the probability that a received sequence Y with $|Y| = \mathcal{N}(1-\epsilon)$ bits, $\mathcal{B} = \frac{\mathcal{N}(1-\epsilon)(1-\delta)}{\sum_i i\mathcal{P}_i}$ alternating blocks of zeros and ones, and $\mathcal{P}_k \cdot \mathcal{B}(1-\gamma)$ blocks of length k is a subsequence of a codeword X generated as in Section 2. To upper bound the latter quantity, we first need to compute the distribution of the number of bits from a random codeword necessary to cover a single block of length k in Y, for all k that appear with positive probability. [1] Let G be this distribution; then G is given by

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

To see (6), consider a block of k zeros (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 $i+1$-st block of zeros must have length at least 1 and such 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 $j-k$.

Consider a block of length k, $1 \le k \le M_{\mathcal{P}}$, in Y. Let J_k denote the number of bits from X needed to cover it. Then J_k is distributed according to $G_{k,j}$. There are $\mathcal{P}_k \cdot \mathcal{B}(1 \pm \gamma)$ blocks of length k in Y; the number of bits each of these needs to be covered are i.i.d. random variables. For different k's, the number of bits needed to cover the respective blocks are independent (not identically distributed) 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:

$$\mathbf{Pr}\left[\sum_{x=1}^{\mathcal{B}} J^x < N\right] \le e^{tN} \left\{ \prod_{k=1}^{M_{\mathcal{P}}} \left(E[e^{-tJ_k}]\right)^{\mathcal{P}_k} \right\}^{\frac{\mathcal{N}(1-\epsilon)(1-\delta)(1-\gamma)}{\sum_k k\mathcal{P}_k}}$$

[1] Note that a block of length i in Y may need more than one block from X to be covered. For example, a block of 5 zeros in Y may need a block of 3 zeros, its consecutive block of 7 zeros and the intermediate block of 2 ones to be covered. In this case, all 12 bits were necessary to cover the block in Y. In general, all the bits from the last block used from X will be used for the block in Y since blocks are alternating.

Hence for P_{col} to go to 0 it suffices that the rate is upper bounded by

$$R < \sup_{\substack{t>0 \\ P\in\mathbb{P}}} \left[-t - \frac{1-d}{\sum_{k=1}^{M_{\mathcal{P}}} k\mathcal{P}_k} \log \left\{ \prod_{k=1}^{M_{\mathcal{P}}} \left(E[e^{-tJ_k}] \right)^{\mathcal{P}_k} \right\} \right] \text{ nats} \tag{7}$$

Therefore we 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 is lower bounded as*

$$C_{del} \geq \sup_{\substack{t>0 \\ P\in\mathbb{P}}} \left[-t - \frac{1-d}{\sum_{k=1}^{M_{\mathcal{P}}} k\mathcal{P}_k} \sum_{k=1}^{M_{\mathcal{P}}} \mathcal{P}_k \cdot \log \left\{ \sum_{j=k}^{\infty} e^{-tj} G_{j,k} \right\} \right] \text{ nats} \tag{8}$$

for $\mathcal{P}$ given by (3), *G given by* (6).

While Theorem 2 does not yield a simple closed form under which to evaluate the capacity, given a specific distribution P, a provable lower bound for the capacity can be evaluated numerically using the theorem.

3.3 Geometric Distributions

In the special case that the block lengths in X are geometrically distributed, i.e., $P_i = (1-p)p^{i-1}$, 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]. Recall that C^1_{del} is the lower bound for capacity 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*

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

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

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

While a detailed proof appears in the full version of the paper, we will sketch an alternative proof that points out how our stronger notion of typical output yields better lower bounds than the framework in [1]. Suppose we follow the analysis in [1] step by step but restrict attention to Y's that are typical sequences in our framework. We can show that the latter correspond to Y's with $\mathcal{N}(1-\epsilon)$ bits and $\mathcal{N}(1-\epsilon)(1-q)$ transitions from 0 to 1 and vice versa, or else, to Y's with derivatives of weight $\mathcal{N}(1-\epsilon)(1-q)$. Assuming J bits needed to cover Y, we can upper bound the probability that such a Y is a subsequence of a randomly generated X by $\mathbf{Pr}[J < N] \leq e^{tN} A^{\mathcal{N}(1-\epsilon)(1-q)} B^{\mathcal{N}(1-\epsilon)q}$, for A, B as in (1); for fixed d,p, let $R_{d,p}(t)$ be the rate that this equation yields and $R^1_{d,p}(t)$ be the rate in [1]. If $t_* > 0$ and $t^1_* > 0$ maximize $R_{d,p}(t)$ and $R^1_{d,p}(t)$ respectively, then, by convexity of a^x for real x, we can show that $R_{d,p}(t_*) \geq R_{d,p}(t^1_*) \geq R^1_{d,p}(t^1_*)$.

4 Derived lower bounds for the capacity

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 Figure 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 as in the first-order Markov chain model. In trying other sequences, 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 consider the case where a block in X is assigned length 1 with probability p and length M (for

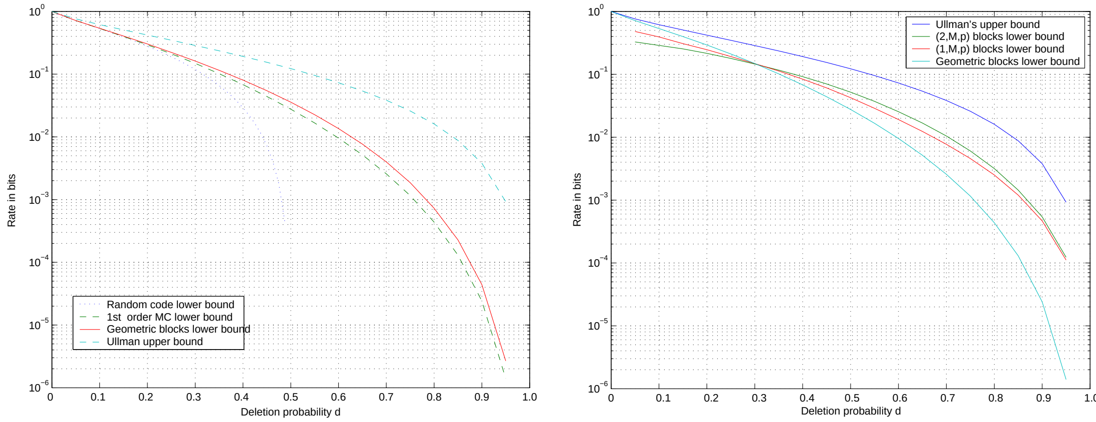

Figure 2: Improvements in rates with our framework, for geometric, $(1, M, p)$ and $(2, M, p)$ distributions.

some integer $M > 1$) with probability $1-p$. Denote this distribution by $(1, M, p)$. The right graph in Figure 2 shows the improvement such a distribution yields versus the geometric distribution when $d \geq 0.4$ (the rates in this figure are an underestimate of the real rates since the optimization of p is over only 2 decimal digits; unfortunately, no closed form was easy to compute). Considering distributions $(2, M, p)$, i.e., assign length 2 with probability p and $M > 2$ with probability $1-p$, further improves the rates when the deletion probability is larger than 0.4. However the rates for small deletion probabilities worsen.

5 Conclusions

In this paper, we 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 concluded that for small deletion probabilities, e.g., $d < 0.4$, geometrically distributed block lengths yield good rates; for $d \geq 0.4$, better rates were achieved by $(2, M, p)$ distributions.

Our work suggests two way 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. We intend to explore these issues further in future work.

References

[1] S. Diggavi and M. Grossglauser, **On Transmission over Deletion Channels**, Allerton Conference, Monticello, Illinois, October 2001.

[2] V. I. Levenshtein, **Binary codes capable of correcting deletions, insertions and reversals (in Russian)**, Doklady Akademii Nauk SSSR, 163 (No. 4, 1965), 845-848., English translation in Soviet Physics Dokl., 10 (No. 8, 1966), 707-710.

[3] V.I.Levenshtein, **Binary codes capable of correcting spurious insertions and deletions of ones (in Russian)**, Problemy Peredachi Informatsii, 1 (No. 1, 1965), 12-25., English translation in Problems of Information Transmission, 1 (No. 1,1965), 8-17.

[4] S. M. Ross, **Introduction to Probability Models**, Harcourt/Academic Press, 2000.

[5] N. J. A. Sloane, **On Single-Deletion-Correcting Codes**, appeared in Ray-Chaudhuri Festschrift 2001, URL: http://www.research.att.com/ njas/docs/dijen.ps.

[6] J. D. Ullman, **On the capabilities of codes to correct synchronization errors**, IEEE Trans. Inform. Theory, 13 (1967), 95-105.

[7] Varshamov, R. R. and Tenengolts, G. M., **Codes which correct single asymmetric errors (in Russian),** Automatika and Telemekhanika, 26 (No. 2, 1965), 288-292., English translation in Automation and Remote Control, 26 (No. 2, 1965), 286-290.