Improved lower bounds for i.i.d. deletion channels

Eleni Drinea*
Div. of Engineering & Applied Sciences
Harvard University
Cambridge, MA 02138
edrinea@deas.harvard.edu

Michael Mitzenmacher †
Div. of Engineering & Applied Sciences
Harvard University
Cambridge, MA 02138
michaelm@eecs.harvard.edu

Abstract

We consider the capacity of binary deletion channels, where bits are deleted independently with probability d. We improve significantly upon the framework used in [1, 2] to lower bound this capacity, by utilizing a stronger definition of a typical output from the channel. In this paper, we specifically focus on codeword sequences given by a first order Markov chain. Our results give the best bounds on the capacity for all values of d; in particular, for $d \geq 0.65$, we surpass Ullman's combinatorial upper bound for channels with an asymptotic fraction of d synchronization errors. Hence our results explicitly indicate a need for new upper bounds in the case of channels with i.i.d. synchronization errors.

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.

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 *i.i.d. deletion channels.* It is known that the capacity of such channels is related to the mutual information between the codeword sent and the received sequence [3], but this does not give an effective means of proving capacity bounds. Recent work, which we describe fully below, attempts to develop Shannon-style theorems that allow computable bounds. Our work is a continuation in this vein, but yields dramatically improved bounds. For example, our bounds when $d \geq 0.65$ surpass Ullman's combinatorial upper bound for channels with synchronization errors [5]. This bound has previously been used as though it were an upper bound on the i.i.d. channel, as it seemed difficult to reach, even though it was not strictly a proven bound for this specific type of channel. Our results are therefore the first to demonstrate that under i.i.d. errors this bound can in fact be broken.

Another advantage of our approach is that it can be applied to insertion and insertion/deletion channels as well. Previous work was based on a decoding method that was successful only if the received sequence was a subsequence of exactly one codeword, and therefore did not generalize beyond deletion channels. While we do not consider more general channels here, we plan to analyze such channels in future work.

*Supported by NSF grant CCR-0118701.

†Supported by NSF grants CCR-9983832, CCR-0118701, CCR-0121154, and an Alfred P. Sloan Research Fellowship.

1.1 Previous Work

It has long been known that random codes, i.e., codes consisting of codewords chosen independently and uniformly 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 $H(d) = -d \log d - (1-d) \log (1-d)$ is the binary entropy function [1] (we denote by log the logarithm base 2 and by ln the natural logarithm throughout). Diggavi and Grossglauser had the insight to examine codewords chosen non-uniformly, in order to better cope with the memory inherent in deletion channels [1]. Specifically, they examined codes consisting of codewords of length N generated by a symmetric first-order Markov process with transition probability p. The decoding algorithm they consider takes a received sequence and determines if it is a subsequence of exactly one codeword; if this is the case, the decoder is successful, and otherwise, the decoder fails. Using this decoder, they determine for what transmission rate the probability of error goes to 0 asymptotically. This analysis yields the following lower bound for the capacity, which proves strictly better than the lower bound for random codes, and is substantially better for high deletion probabilities d:

$$C_{\text{del}} \geq \sup_{\substack{t>0 \\ 0<p<1}} [-t - (1-d) \log ((1-q)A + qB)], \tag{1}$$

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

Drinea and Mitzenmacher in [2] improve on the lower bounds in (1) by generalizing the framework above to consider codewords of length N that consist of alternating blocks of zeros and ones. The lengths of the blocks are i.i.d. random variables, according to some symmetric distribution P over the positive integers with suitably decreasing tails. For example, when the block lengths are geometrically distributed with parameter p, the resulting code has the same distribution as codes generated by the first-order Markov chain model with transition probability p. Again, the decoder is successful if and only if the received sequence is a subsequence of exactly one codeword from the randomly generated codebook. Their improvements arise from two considerations. First, the analysis of Diggavi and Grossglauser considers only *typical outputs*, which consist of at least $N(1-d)(1-\epsilon)$ bits, for some $\epsilon = o(1)$; any output that is atypical is assumed to give an error in the analysis. Note that the probability of an atypical output is exponentially small. In [2] a stronger notion of a typical output that still contributes an exponentially small error probability is used. For geometric block length distributions, this analysis yields the following improved bounds over (1):

$$C_{\text{del}} \geq \sup_{\substack{t>0 \\ 0<p<1}} \left[-t - (1-d) \log (A^{1-q} \cdot B^{q})\right], \tag{2}$$

for A, B, q as in (1). However, the more important improvement in [2] comes from allowing more general distributions for the block lengths. While obtaining a closed formula for the capacity under general distributions appears hard, specific distributions can be tested using numerical calculation. In [2] Morse-code type codes were considered; with these codes blocks are either short (i.e., length $m \geq 1$) with probability x or long (i.e., length $M > m$) with probability $1-x$. Calculations for these distributions yielded better bounds than the geometric distribution when the deletion probability was at least

0.4. Prior to this work, these are the best provable bounds we know of for i.i.d. deletion channels.

Recent work by Kavcic and Motwani attempts to bound the mutual information between the input and output of the i.i.d. deletion channel experimentally, via simulation [4]. Although their bounds are not strictly provable, their work demonstrates that the capacity of the i.i.d. deletion channel is indeed much larger than the lower bounds proven in previous theoretical work.

1.2 Our New Approach

Our work extends the approach of previous work by considering both a stronger definition of a typical output and a corresponding stronger method for decoding. Informally, the definition of a typical output in [2] requires that the received sequence consists of approximately the expected number of blocks of length k for each k (the block length distribution for the received sequence can be derived from the block length distribution P for the codewords and the deletion probability d). In this paper, our stronger notion of a typical output is motivated by the idea of mutual information. Specifically, consider a block of length k in the received sequence. Such a block arises from a group of one or more blocks from the transmitted codeword. We call the ordered sequence of lengths of this group of blocks in the codeword the *type* of a block in the received sequence; that is, a type corresponds to a compact description of the group of blocks from the codeword that generated the block in the received sequence. We now require that for a typical output with respect to a codeword, the number of blocks of length k in the received sequence arising from groups of type t is close to its expectation for every type t and length k. Our decoding algorithm checks if there is only one codeword for which the received sequence is a typical output. While this decoding algorithm is remarkably inefficient (exponential time), efficiency is not required to prove capacity bounds. Also note that the received sequence might be a subsequence of more than one codeword with this approach; we only need it to be a typical output with regard to one codeword.

In this paper, we consider only codewords with geometrically distributed block lengths. While [2] suggests other distributions might perform better, searching for such distributions remains a point for future work, and the mathematics is much simpler in the geometric case. Again, even with this restriction, our bounds are the best provable bounds for this channel.

The remainder of the paper is organized as follows. In Section 2 we review the necessary parts of the model from [2] and introduce the notion of the type of a block in the received sequence. In Section 3, a general bound for the capacity of the i.i.d. deletion channel for finite block length distributions is presented. In Section 4 we derive lower bounds in the special case of the geometric distribution; a discussion of these bounds and the upper bound provided by Ullman in [5] follows in Section 5. Section 6 concludes the paper.

2 The framework and codebook

We describe the generation of our codebook, following [2], and define the notion of types. We consider a code C with 2^{NR} binary codewords of length N, where R is the rate of the code in bits. Each codeword consists of alternating blocks of zeros and ones and is generated independently by the following stochastic process. The first block is chosen to be zeros or ones each with probability 1/2. The lengths of successive blocks are

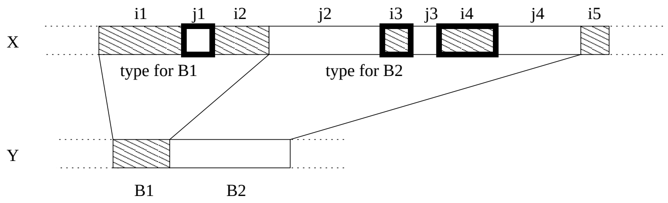

Figure 1: The blocks with lengths i_1, j_1 and i_2 from X give rise to block $B1$ in Y; the type of $B1$ is in the family $F(1, i_1, i_2, j_1)$. The blocks with lengths j_2, i_3, j_3, i_4 and j_4 give rise to block $B2$; the type of $B2$ is in the family $F(2, j_2, j_3 + j_4, i_3 + i_4)$. The thick contours of blocks $j1$, $i3$ and $i4$ indicate that these blocks were *necessarily* completely deleted.

independent identically distributed random variables given by a distribution P, so that the length is j with probability P_j for $j \geq 1$. While this approach could be extended to use different distributions P and Q for zeros and ones, we have not found this helpful so far, and hence we restrict ourselves to a single distribution. We assume throughout that our distribution P is bounded by some geometric distribution, so that $P_j \leq a^j$ for some constant $0 < a < 1$. We keep generating blocks until the codeword length N is reached or exceeded. If the last block exceeds N, it is truncated; this does not affect the asymptotics for large N. Applying standard results from renewal theory, we can show that for large N and a suitable $\delta = o(1)$, the number of blocks in the codeword is $\frac{N}{\sum_j jP_j}(1 \pm \delta)$ with high probability. (The proof appears in the full version of [2].)

Now consider a transmitted codeword X and the associated received sequence Y. The sequence Y can also be broken into alternating blocks of zeros and ones. With each block we may associate a *type* depending on the blocks in X that it was derived from. Specifically, consider a block B of $k \geq 1$ zeros in Y (everything is entirely similar for blocks of ones). We associate with B a group of consecutive blocks in X, starting with the first block in X which had an undeleted zero that is a bit in B, and including all blocks up to (but not including) the block in which the next undeleted one appears. The type is just a tuple giving the lengths of all these blocks.

More concretely, a type is a tuple of $2i+1$ numbers representing the lengths of $2i+1$ consecutive blocks in X, for $i \geq 0$. If the first block is a block of zeros, the i blocks of ones in the type must be completely deleted since the type gives rise to a single block of zeros in Y. The first block of ones in X from which at least one bit is not deleted gives rise to a new block in Y and thus begins a new type. We represent the type of a block in Y by the ordered $2i+1$-tuple $t = (z, s_1, r_1, \ldots, s_i, r_i)$. We now find the probability that a block in Y has type t. Let the random variable T be the type of the block. We have

$$\mathbf{Pr}[T = t] = \frac{P_z(1-d^z)}{1-x} \cdot \left(\prod_{\ell=1}^{i} P_{s_\ell} d^{s_\ell} P_{r_\ell} \right) \cdot (1-x) = P_z(1-d^z) \left(\prod_{\ell=1}^{i} P_{s_\ell} P_{r_\ell} \right) d^{s_1+\ldots+s_i}, \tag{3}$$

where $x = \sum_j P_j d^j$ is the probability that a block is deleted. The first term on the left hand side of (3) is the conditional probability the block in X starting the type has length z given that the block starts a type; that is, at least one bit from the block is not deleted. The second term corresponds to the remaining blocks in X, with every other block necessarily being deleted. The third term is the probability that the block after these $2i+1$ blocks has at least one undeleted bit, starting a new block in Y. Note that here and throughout the paper we ignore boundary effects, which have no effect on the asymptotics.

A more concise representation that we use henceforth is motivated by (3). Let $r = \sum_{\ell=1}^{i} r_\ell$, $s = \sum_{\ell=1}^{i} s_\ell$. For all $i \geq 0$, $z \geq 1$, $r \geq i$, $s \geq i$, we define $F(i,z,r,s)$ to be the family of types that consist of the following: $2i+1$ blocks, the first of which has length z; the lengths of the i blocks whose bits differ from the first block sum up to s; and the lengths of the i blocks whose bits are the same as the first block sum up to r. We denote by $|F(i,z,r,s)|$ the size of the family $F(i,z,r,s)$; each of the $|F(i,z,r,s)|$ members of $F(i,z,r,s)$ occurs with the same probability given by (3). In what follows, when we refer to the type of a block in Y we may instead refer to the family $F(i,z,r,s)$ of types, when we do not care which specific member of the family the type is. For examples, see Figure 1.

We introduce some additional notation. Let $Q_{n,m}$ be the probability that the total length of m blocks, each independently and identically distributed with distribution P, is n. We note that $Q_{n,m}$ is easily computed by the recursion

$$Q_{n,m} = \sum_{\ell=1}^{n-1} P_\ell Q_{n-\ell,m-1}.$$

With the same reasoning as for (3),

$$\mathbf{Pr}[T \in F(i,z,r,s)] = P_z(1-d^z) \cdot Q_{r,i} Q_{s,i} d^s, \tag{4}$$

and also, if $t \in F(i,z,r,s)$, then

$$\mathbf{Pr}[T = t] = \frac{\mathbf{Pr}[T \in F(i,z,r,s)]}{|F(i,z,r,s)|}. \tag{5}$$

With this notation, we can write an expression for the distribution of block lengths in Y. We denote this distribution by $\mathcal{P}$; like P, $\mathcal{P}$ is symmetric with respect to blocks of zeros and blocks of ones. Let K and T be random variables representing the length and type of a block in Y. Conditioned on arising from type t in family $F(i,z,r,s)$, a block of zeros in Y will have length k if exactly k of the $z+r$ zero bits of t are not deleted, with at least one arising from the first block of length z. Thus the joint probability of a block having length k and arising from type t is given by:

$$\begin{aligned}
\mathbf{Pr}[T=t, K=k] &= \mathbf{Pr}[K=k \mid T=t] \cdot \mathbf{Pr}[T=t] &(6)\\
&= \frac{\left(\binom{z+r}{k} - \binom{r}{k}\right) d^{r+z-k}(1-d)^k}{1-d^z} \cdot \frac{1}{|F(i,z,r,s)|} P_z(1-d^z) Q_{r,i} Q_{s,i} d^s \\
&= \frac{1}{|F(i,z,r,s)|} \left(\frac{1-d}{d}\right)^k \left(\binom{z+r}{k} - \binom{r}{k}\right) d^{z+r+s} \cdot P_z Q_{r,i} Q_{s,i} &(7)
\end{aligned}$$

This implies that the probability that a block in the received sequence has length $k \geq 1$ is given by

$$\begin{aligned}
\mathcal{P}_k &= \sum_{(i,z,r,s)} \sum_{t \in F(i,z,r,s)} \mathbf{Pr}[T=t, K=k] \\
&= \left(\frac{1-d}{d}\right)^k \sum_{(i,z,r,s)} \left(\binom{z+r}{k} - \binom{r}{k}\right) d^{z+r+s} \cdot P_z Q_{r,i} Q_{s,i}. &(8)
\end{aligned}$$

Similar formulas appear in [2], where types were implicitly used. Explicitly identifying the existence of types and studying their behavior proves crucial to improve the lower bounds for the capacity of i.i.d. deletion channels. In essence, we can think of the received symbols as being the lengths of the blocks, and the transmitted symbols as being the types that give rise to the blocks. Further, in effect the mutual information for these symbols gives a computable bound that we can use to bound the capacity of the deletion channel.

3 A new lower bound

We start by giving a new definition of typical outputs and show that a received sequence Y is a typical output for some codeword X with high probability. Then we show that, upon reception of a typical output Y, our decoding algorithm fails with probability exponentially small in N for appropriate rates. This yields our lower bound on the capacity.

3.1 Typical outputs

We give a somewhat less formal definition of a typical output; a complete description and analysis will appear in the full paper. The value $\mathcal{B} = \frac{N(1-d)}{\sum_k k\mathcal{P}_k}$ is approximately the expected number of blocks in the received sequence Y (it is only approximate because of boundary effects and variations in the number of bits received). A received sequence Y is a *typical output* for a codeword X if it consists of $\mathbf{Pr}[T = t, K = k] \cdot \mathcal{B}(1 \pm \beta)$ blocks of length $1 \leq k \leq c_1$ that arise from types t with at most c_2 blocks, for certain positive constants c_1 and c_2, and $\beta = \Theta(1/\sqrt{N})$.

Essentially, our choice of definition for a typical output yields that with all but vanishingly small probability, the number of blocks of Y of each length arising from each type is close to its expectation, when there are sufficiently many blocks of Y so that Chernoff bounds may hold. The case where the length of a block is greater than c_1 or the number of blocks for a type is greater than c_2 can be handled in a more explicit fashion; by choosing c_1 and c_2 sufficiently large, these cases can be made to have at most an ϵ effect on the capacity for any constant $\epsilon > 0$. We conclude that a received sequence Y fails to be a typical output from the channel with vanishingly small probability.

3.2 Decoding error probability

We now develop the main analysis of our paper. In the following, we simplify the analysis by assuming that the number of blocks of length k derived from groups of blocks of type t, denoted by $\mathcal{B}_{t,k}$, exactly equals $\mathbf{Pr}[T = t, K = k] \cdot \mathcal{B}$ for all k, t. Conditioned on the output being a typical output, the number of such blocks is really $\mathbf{Pr}[T = t, K = k]\mathcal{B}(1 + o(1))$ for sufficiently small k and types t with a sufficiently small number of blocks; a more careful analysis, to be given in an extended version of this paper, shows that the $o(1)$ terms and the effect of large blocks and types with more blocks do not affect the capacity bound derived in this way.

Fix a received sequence Y. We will use F as a shorthand for $F(i, z, r, s)$. Consider an enumeration of all families F and denote by t_j^F the j-th type in family F. For each k, the number of blocks of length k in Y is given by $\mathcal{B}_k = \sum_F \sum_{t \in F} \mathcal{B}_{t,k}$. There are

$$\binom{\mathcal{B}_k}{\mathcal{B}_{t_1^{F_1},k};\ \ldots \mathcal{B}_{t_{|F_1|}^{F_1},k};\ \mathcal{B}_{t_1^{F_2},k};\ \ldots \mathcal{B}_{t_{|F_2|}^{F_2},k};\ \ldots}$$

ways we can place the types corresponding to the blocks of length k in an attempt to reconstruct all different codewords that, when transmitted through the deletion channel, might generate these blocks according to the definition of a typical output. That is, given the received sequence Y, our decoding algorithm considers all possible $\mathcal{B}_k$ blocks of length k in Y, and considers all possible ways of choosing the type of each block in Y so that Y would have been a typical output. After doing this for all k, the decoding algorithm has an exponentially large list of all possible strings of length N for which Y would have been a typical output. If exactly one of these strings is a codeword in our codebook, then (assuming that Y was indeed a typical output, which occurs with high probability) the algorithm decodes successfully.

If T and K are again random variables denoting respectively the type of a block in Y and its length, the number of potentially transmitted codewords considered by the decoding algorithm is then

$$\begin{aligned}\prod_k \binom{\mathcal{B}_k}{\mathcal{B}_{t_1^{F_1},k};\ \ldots \mathcal{B}_{t_1^{F_2},k};\ \ldots} &\leq 2^{-\sum_k \mathcal{B}_k \sum_F \sum_{t\in F} \mathbf{Pr}[T=t \mid K=k] \log(\mathbf{Pr}[T=t \mid K=k])} \\ &= 2^{-\mathcal{B} \sum_k \sum_F \sum_{t\in F} \mathbf{Pr}[T=t, K=k] \log(\mathbf{Pr}[T=t \mid K=k])} \\ &= 2^{\frac{N(1-d)}{\sum_k k\mathcal{P}_k} H(T \mid K)}. \qquad (9)\end{aligned}$$

Here (9) is an upper bound, as a received sequence Y may correspond to a codeword X under many different segmentations into types while still having the property that Y is a typical output for X. Improving this bound may directly yield improvements on the rate.

To upper bound the probability that a fixed codeword X in our codebook could yield one of the sequences of types counted in (9), we restrict the codebook to consist only of the likely codewords. That is, standard methods give that almost all codewords arise with probability at most $2^{-\frac{N}{\sum_j jP_j} H(P) + o(N)}$, so that the probability of including a codeword with greater probability of being chosen is exponentially small. We can throw out such improbable codewords, to guarantee that all possible codewords are chosen with probability at most $2^{-\frac{N}{\sum_j jP_j} H(P) + o(N)}$. Ignoring the $o(N)$ term, which does not affect the final capacity bound, yields the following upper bound for the probability that Y is a typical output for a randomly selected codeword in our codebook:

$$2^{\frac{N(1-d)}{\sum_k k\mathcal{P}_k} H(T \mid K) - \frac{N}{\sum_j jP_j} H(P)}.$$

By a union bound, the probability that the received sequence Y is a typical output for more than one codeword is at most

$$2^{NR} \cdot 2^{\frac{N(1-d)}{\sum_k k\mathcal{P}_k} H(T \mid K) - \frac{N}{\sum_j jP_j} H(P)} = \left(2^{R + \frac{1-d}{\sum_k k\mathcal{P}_k} H(T \mid K) - \frac{1}{\sum_j jP_j} H(P)}\right)^N. \qquad (10)$$

Since all typical outputs share the same structural properties, the probability that the decoding algorithm will fail to identify a unique codeword upon reception of *any* Y that is a typical output for a codeword chosen uniformly at random is given by (10). Let $\mathbb{P}$ be the class of all distributions P such that $P_j \leq a^j$ for some constant $0 < a < 1$. For the decoder to fail with probability that goes to zero asymptotically it suffices that the rate is upper bounded by

$$R < \sup_{P\in\mathbb{P}} \left[\frac{1}{\sum_j jP_j} H(P) - \frac{1-d}{\sum_k k\mathcal{P}_k} H(T \mid K)\right] = \sup_{P\in\mathbb{P}} \left[\frac{1}{\sum_j jP_j} H(P) - \frac{1-d}{\sum_k k\mathcal{P}_k} (H(T,K) - H(K))\right].$$

Therefore we obtain the following theorem for arbitrary distributions $P \in \mathbb{P}$.

Theorem 1 *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 by*

$$C_{del} \geq \sup_{P \in \mathbb{P}} \left[\frac{1}{\sum_j jP_j} H(P) - \frac{(1-d)}{\sum_k k\mathcal{P}_k} (H(T,K) - H(\mathcal{P})) \right] \text{ bits} \tag{11}$$

for $\mathbf{Pr}[T=t, K=k]$ given by (7) and $\mathcal{P}$ given by (8).

4 Geometric distributions

In this section, we use (11) to derive a lower bound for the capacity of i.i.d. deletion channels in the special case where the block lengths in X are geometrically distributed, i.e., $P_j = (1-p)p^{j-1}$.

It is easy to show that $H(P) = \frac{H(p)}{1-p}$, where $H(\ )$ is the binary entropy function. Also, the probability that m blocks from X have length n is given by $Q_{n,m} = \binom{n-1}{m-1} p^{n-m}(1-p)^m$, since there are $n-1$ bits from which to choose the last bits of the first $m-1$ blocks (the last block ends at the n-th bit). Hence a family $F(i,z,r,s)$ consists of $\binom{s-1}{i-1} \cdot \binom{r-1}{i-1}$ members and the probability of a type (5) becomes:

$$\mathbf{Pr}[T=t] \quad = \quad p^{r+s+z} \cdot \left(\frac{1-p}{p} \right)^{2i+1} \cdot (1-d^z) \cdot d^s. \tag{12}$$

Then the joint probability of length k and type t from (7) becomes

$$\mathbf{Pr}[T=t, K=k] \quad = \quad \left(\frac{1-d}{d} \right)^k \cdot \left[\binom{z+r}{k} - \binom{r}{k} \right] \cdot (pd)^{z+r+s} \cdot \left(\frac{1-p}{p} \right)^{2i+1} . \tag{13}$$

When the block length distribution in X is geometric with parameter p, the block lengths in Y are also geometrically distributed with parameter $q = 1 - \frac{1-p}{1+d-2pd}$ (e.g., see [1, 2]). Then $H(\mathcal{P}) = \frac{H(q)}{1-q}$. The following combinatorial lemma, given in the full version, derives a formula for the joint entropy $H(T,K)$:

Lemma 1 *When the blocks in X are geometrically distributed with parameter p, the joint entropy $H(T,K)$ of the distribution of the types and the block lengths in Y is given by*

$$H(T,K) \quad = \quad \frac{H(d)+H(p)}{(1-d)(1-q)} - \sum_k \sum_{(i,z,r,s)} \sum_{t \in F(i,z,r,s)} \mathbf{Pr}[T=t, K=k] \cdot \log \left[\binom{r+z}{k} - \binom{r}{k} \right].$$

We immediately obtain the following corollary to Theorem 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*

$$\begin{aligned} C_{del} \quad \geq \quad \sup_{0<p<1} \Bigg[& (1-d)(1-q) \sum_k \sum_{(i,z,r,s)} \sum_{t \in F} \mathbf{Pr}[T=t, K=k] \cdot \log \left[\binom{r+z}{k} - \binom{r}{k} \right] \\ & + (1-d)H(q) - H(d) \Bigg], \end{aligned} \tag{14}$$

where $q = 1 - \frac{1-p}{1+d-2pd}$, $\mathbf{Pr}[T=t, K=k]$ is given by (13), and F represents $F(i,z,r,s)$.

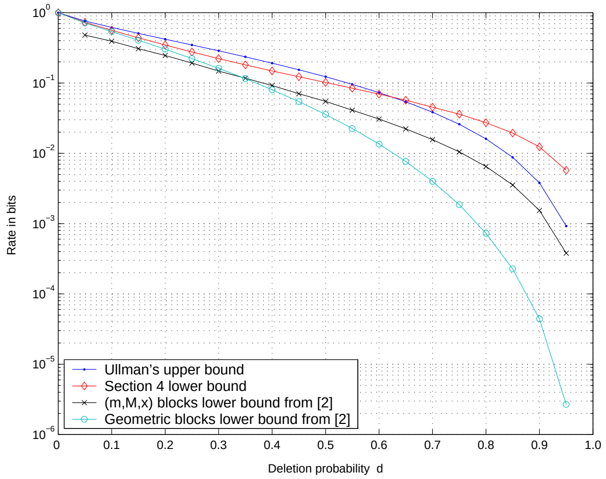

Figure 2: Improvement in rate with our framework for codewords with geometrically distributed block lengths (Section 4). Comparison with lower bounds for geometric and (m, M, x) distributions from [2], and Ullman's upper bound.

Although it seems difficult to derive a closed formula for the summation above, one can easily compute it numerically for fixed p, d. Then it is a matter of optimizing over all values of p. Our optimization was over only two decimal digits for p. Also, our numerical calculations were over a limited range of k and i, z, r, s, (which is equivalent to truncating both the distribution of the block lengths in the received sequence and the distribution of types). Hence the graph in Figure 2 presents an underestimate of the actual rates, as all terms inside the summation (14) are non-negative. To verify our results, we performed extensive simulations for codewords with geometrically distributed block lengths and $N = 2.5 \cdot 10^{10}$. The simulations verified the convergence of $H(P)$, $H(\mathcal{P})$, $H(T)$, and $H(T, K)$ to the values predicted by the theory, giving us confidence in the results of Figure 2.

5 Discussion of our results

As discussed in the introduction, an upper bound for the capacity C_{del} of channels with synchronization errors is provided by Ullman in [5]:

$$C_{\text{del}} \leq 1 - (1 + d) \log_2 (1 + d) + d \log_2 (2d) \quad \text{bits}, \tag{15}$$

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. All previous lower bounds for the capacity of i.i.d. deletion channels were strictly below (15). Our current bounds are much higher than (15) for $d \geq 0.65$. Because of this, we clarify that Ullman's upper bound does not apply to i.i.d. deletion channels, and our work does not yield any contradiction.

Ullman's bound (15) is based on a channel that introduces $d \cdot N$ insertions in the first $(1 - d) \cdot N$ bits of the codeword. Further, the insertions are restricted to be such that the number of blocks in the first N bits of the received sequence equals the number of blocks in the first $N(1 - d)$ bits of the transmitted codeword. Finally, his bound is for a codebook with zero probability of error. Obviously, the i.i.d. deletion channel is quite different: deletions instead of insertions occur in random places, and only a vanishing probability of error. We conclude that while [5] provides an upper bound for the capacity of channels with $d \cdot N$ arbitrary synchronization errors with no errors, it does not apply here. Our work is the first to demonstrate that in fact this bound can be broken, and specific upper bounds for this channel need to be developed.

6 Conclusions

We have presented new lower bounds for the capacity of i.i.d. binary deletion channels, improving on previous analysis by using a stronger definition of a typical output. For high rates, we have exceeded Ullman's upper bound for general synchronization channels.

There are many directions to advance this work that we are pursuing.

- The analysis we have given may be improvable, for example by finding a better upper bound for (9).
- There may be even stronger notions of typical output that may be useful. For example, perhaps types should be considered for several consecutive blocks in the received sequence at a time, instead of just one, to achieve a better bound. The challenge of this direction is the computation necessary to determine and evaluate the achievable rates.
- There may be better ways of selecting codewords, such as using distributions suggested in [2].
- Our approach can be applied to channels with insertions and deletions.

Finally, providing good upper bounds for the capacity of i.i.d. deletion channels is a clear challenging open question.

References

[1] S. Diggavi and M. Grossglauser, On Transmission over Deletion Channels, In *Proceedings of the 39th Annual Allerton Conference on Communication, Control, and Computing*, pp. 573-582, 2001.

[2] E. Drinea and M. Mitzenmacher, On Lower Bounds for the Capacity of Deletion Channels, In *Proceedings of the 2004 IEEE International Symposium on Information Theory*, p. 227. See also Harvard Computer Science Technical Report TR-07-04, at ftp://ftp.deas.harvard.edu/techreports/tr-2004.html.

[3] R. L. Dobrushin, Shannon's Theorems for Channels with Synchronization Errors, *Problems of Information Transmission*, 3(4):11-26, 1967. Translated from *Problemy Peredachi Informatsii*, vol. 3, no. 4, pp 18-36, 1967.

[4] A. Kavcic and R. Motwani. Insertion/deletion channels: Reduced-state lower bounds on channel capacities. In *Proceedings of the 2004 IEEE International Symposium on Information Theory*, p. 229.

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