Improved Lower Bounds for the Capacity of i.i.d. Deletion and Duplication Channels

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

***Abstract*—This paper considers the capacity of binary deletion channels, where bits are deleted independently with probability d. It improves significantly upon the best previous framework used to obtain provable lower bounds on this capacity by utilizing a stronger definition of a typical output from the channel. The new results give the best known provable bounds on the capacity for all values of d. Moreover, the techniques presented here extend to yield lower bounds for channels with certain types of random insertions, namely, duplications, or combinations of duplications and deletions. To demonstrate these techniques in this context, two binary channels are analyzed: a channel where each transmitted bit is copied with probability v and a channel where each transmitted bit is copied a geometrically distributed number of times.**

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

I. Introduction

THIS work is motivated by the goal of finding the capacity of deletion channels, following on the previous work of Diggavi and Grossglauser [2] and Drinea and Mitzenmacher [4]. Deleted symbols do not arrive at the receiver; undeleted symbols remain intact, but might be shifted left, so the receiver does not know what symbols have been deleted. Specifically, we are interested in lower bounds for the capacity of the binary *independent and identically distributed (i.i.d.) deletion channel*. In this model, N bits are sent, and each bit is deleted independently with probability d; the capacity corresponds to the limiting case as N goes to infinity.

It is known that the capacity of such channels is related to the mutual information between the codeword sent and the received sequence [5], 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 lower bounds. Our work is a continuation in this vein, but yields dramatically improved lower bounds; as far as we know, they prove the best provable lower bounds for all values of d. As a further comparison point, Ullman provides a combinatorial upper bound for channels with synchronization errors [11]. This bound has previously been used as though it were an upper bound on the i.i.d. channel [2], as it seemed difficult to reach, even though it was not strictly a proven bound for this specific type of channel. Our lower bounds when $d \geq 0.65$ surpass Ullman's upper bound, and are first to demonstrate that under i.i.d. errors this bound can in fact be broken.

A further advantage of our approach is that it can be applied to handle channels that (randomly) insert bits via duplications and channels that both duplicate and delete bits as well. Specifically, we show how our argument generalizes to provide lower bounds for a binary channel that sends a number of copies of each bit, where the number of copies is governed by a fixed distribution and is independent for each bit. Sending zero copies of a bit is equivalent to deleting it. Previous arguments in this line of research were based on a decoding method that is successful only if the received sequence is a subsequence of exactly one codeword, and therefore only applied to deletion channels. As a demonstration, we provide lower bounds for the capacity of two channels which introduce i.i.d. insertions only. The first channel, or *elementary i.i.d. duplication channel*, duplicates each transmitted bit independently with probability d. The second channel, or *i.i.d. geometric duplication channel*, introduces a geometrically distributed number of copies independently for each transmitted bit.

A. Previous Work

Let C_{del} be the capacity of the i.i.d. deletion channel with deletion probability d, where the dependence on d is implied throughout. 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 on C_{del}

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

where

$$H(d) = -d \log d - (1-d) \log (1-d)$$

is the binary entropy function [2]. (We denote the logarithm base 2 by $\log$ and the natural logarithm by $\ln$ throughout.)

Diggavi and Grossglauser had the insight to examine codewords chosen nonuniformly, in order to better cope with the memory inherent in the output of the deletion channels [2]. 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 analyze takes a received sequence and determines if it is a subsequence of exactly one codeword, using a greedy algorithm; if

Manuscript received January 9, 2006; revised October 11, 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 material in this paper was presented in part at the 42nd Annual Allerton Conference on Communication, Control and Computing, Monticello, IL, September 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.
Digital Object Identifier 10.1109/TIT.2007.901221

this is the case, the decoder is successful, and otherwise, the decoder fails. Their 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:

$$C_{\mathrm{del}} \geq \sup_{\substack{t>0\\0<p<1}} \left[-\frac{t}{\ln 2} - (1-d)\log\left((1-q)A + qB\right)\right] \quad (1)$$

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 [4] 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, determined by a distribution P over the positive integers with geometrically 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 [4], a stronger notion of a typical output that contributes a super-polynomially small error probability is used. For geometric block length distributions, this analysis yields the following improved bounds over (1):

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

for A, B, q as in (1). A more important improvement in [4] comes from allowing more general distributions for the block lengths. While obtaining a closed formula for the capacity under general distributions does not appear possible, specific distributions can be tested using numerical calculation. In [4], 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, denoted henceforth as (m, M, x) distributions, yielded better bounds than the geometric distribution when the deletion probability was greater than 0.35.

Prior to this work, the best provable lower bounds for the i.i.d. deletion channel are given by the methods of Drinea and Mitzenmacher in [4]. There has also been work bounding the capacity via simulation techniques. For example, Vvedenskaya and Dobrushin [12] attempt to bound the mutual information between the input and output of the i.i.d. deletion channel via simulation. They estimate lower bounds for the capacity of the i.i.d. deletion channel using codewords generated by a low-order Markov chain (up to order 2). However, because at the time they were only able to experiment with very short codewords, it is not clear that their results give true bounds. Recent work by Kavčić and Motwani [8] also employs the Monte Carlo method for estimating information rates, using much larger simulations and codeword lengths. Although these bounds are not strictly provable, they both suggest that the capacity of the i.i.d. deletion channel is indeed much larger than the lower bounds proven in previous theoretical work.

Subsequent to this work, the authors use these techniques to obtain a simple lower bound for the i.i.d. deletion channel: the capacity is at least $(1-d)/9$ for *any* value of d [9]. This result shows that the capacity of the i.i.d. deletion channel is within a constant factor of the corresponding erasure channel (where deletions become erasures); it also proves the capacity is bounded away from 0 for all constant $d < 1$. This result stems from showing the capacity of the deletion channel can be bounded by bounding the capacity of a single specific channel dubbed a Poisson-repeat channel, which falls into the class of channels where bits are duplicated or deleted considered here. This channel is analyzed numerically using the techniques presented in this paper; the fact that the capacity of a Poisson-repeat channel with parameter 1 has capacity at least $1/9$ translates directly into the corresponding result for the deletion channel. See [9] for more details.

B. 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. In [4], codewords are generated by laying out alternating blocks of zeros and ones; the received sequence, too, can be thought of in terms of alternating blocks, and 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. Informally, the definition of a typical output in [4] requires that the received sequence consists of approximately the expected number of blocks of length k for each k. 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 pair (t, k) that appears sufficiently often. This will be described in more detail below.

In conjunction with this stronger notion of a typical sequence, we also make use of a stronger decoding algorithm. In [2], [4], a greedy algorithm is used to determine if the received sequence is the subsequence of just one codeword. This decoding approach is limited to deletion channels. Our new decoding algorithm checks if there is only one corresponding 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.

We introduce a useful way of thinking about this decoding process. Suppose for the received sequence, we knew for each block length k exactly how many blocks were derived from each possible type. (We know this approximately if the received sequence is typical.) We could then think of there being a specific number of *jigsaw puzzle* pieces, with each piece corresponding to a type/block length pair (t,k); a piece would cover a block in the received sequence, and give the relative position of the blocks corresponding to the type from which the received block was derived. Covering the entire received sequence with jigsaw puzzle pieces would then give a potential codeword; considering all possible ways of laying out the pieces consistent with the received sequence would give all possible codewords from which the received sequence could have been derived. If just one actual codeword appears with this algorithm, the decoding is successful. We analyze this decoding algorithm, although some additional work beyond what we have just described is necessary since we do not know the exact number of pieces corresponding to each pair (t,k).

In our analysis, we allow the input block-length distribution P to either be the geometric distribution or a distribution with finite support. We specifically consider codewords with geometrically distributed block lengths as well as Morse-type codes suggested in [4]. Geometrically distributed codewords yield the highest rates under our analysis for all values of deletion probability d; Morse-type codes perform almost as well under our analysis when d is very large (larger than 0.9). While other distributions might perform better, searching for such distributions remains a point for future work. 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 II, we review the necessary parts of the model from [4] and introduce the notion of the type of a block in the received sequence. In Section III, a general bound (under permissible input distributions) for the capacity of the i.i.d. deletion channel is presented. In Sections IV and V, we derive specific lower bounds in the special cases of the geometric and the (m,M,x) distribution, respectively; a discussion of these bounds and the upper bounds provided by Ullman in [11] and Dolgopolov in [6] follows in Section VI. Section VII extends our new approach to a more general class of channels with i.i.d. deletions and duplications. In Section VIII, we derive specific lower bounds for some simple channels with duplications to demonstrate our approach in this context. We conclude with a discussion of further directions for the challenging problems related to deletion and insertion channels.

II. Codebooks and Types

We describe the generation of our codebook, following [4], 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 i.i.d. 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 gives larger lower bounds, and hence we restrict ourselves to a single distribution. We assume throughout that P is either the geometric distribution or has finite support. 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 a standard large-deviations bound, we can show that for large N and $\delta = O(N^{-1/3})$, the number of blocks in the codeword is $\frac{N}{\sum_j jP_j}(1\pm\delta)$ with probability at least $1-e^{-\Theta(N^{1/3})}$ (see [4, Proposition 1] for a proof). Note that here and throughout the paper 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.

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_Y of $k \geq 1$ zeros in Y (everything is entirely similar for blocks of ones). We associate with B_Y a group of consecutive blocks in X, starting with the first block in X which had an undeleted zero that was received as a bit in B_Y, 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. Also, let $D=\sum_j P_j d^j$ be the probability that a block is deleted. We have

$$\begin{aligned}\mathbf{Pr}[T=t] &= \frac{P_z(1-d^z)}{1-D}\cdot\left(\prod_{\ell=1}^{i} P_{s_\ell} d^{s_\ell} P_{r_\ell}\right)\cdot(1-D)\\ &= P_z(1-d^z)\left(\prod_{\ell=1}^{i} P_{s_\ell} P_{r_\ell}\right) d^{s_1+\cdots+s_i}. \qquad (3)\end{aligned}$$

The first term on the middle expression in (3) is the conditional probability that the block in X starting the type has length z given that 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.

For notational convenience, we introduce a more concise representation motivated by (3), which we use henceforth. Let $s=\sum_{\ell=1}^{i} s_\ell$, $r=\sum_{\ell=1}^{i} r_\ell$. For all $i \geq 0$, $z \geq 1$, $r \geq i$, $s \geq i$,

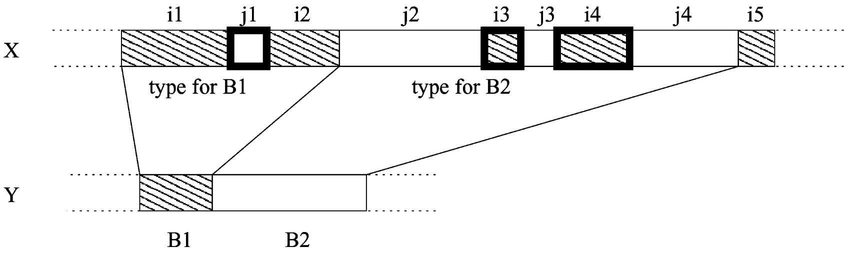

Fig. 1. The blocks with lengths i_1, j_1, and i_2 from X give rise to block $B1$ in Y; the type of $B1$ is in the family $F(1, i_1, i_2, j_1)$. The blocks with lengths j_2, i_3, j_3, i_4, and j_4 give rise to block $B2$; the type of $B2$ is in the family $F(2, j_2, j_3 + j_4, i_3 + i_4)$. The thick contours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

we define $F(i, z, r, s)$ to be the family of types that consist of $2i + 1$ blocks, where the first block has length z, the i blocks whose bits are the same as the first block have total length r, and the i blocks whose bits differ from the first block have total length s. For examples, see Fig. 1.

We introduce some additional notation. Let $Q_{n,m}$ be the probability that the total length of $m \geq 1$ blocks, each i.i.d. with distribution P, is $n \geq m$. For convenience, we may also take $Q_{0,0} = 1$. We note that $Q_{n,m}$ is easily computed by the recursion

$$Q_{n,m} = \sum_{\ell=1}^{n-m+1} P_\ell Q_{n-\ell,m-1}. \tag{4}$$

With the same reasoning as for (3)

$$\mathbf{Pr}[T \in F(i,z,r,s)] = P_z(1-d^z) \cdot Q_{r,i} Q_{s,i} d^s. \tag{5}$$

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 $t = (z, s_1, r_1, \ldots, s_i, r_i)$ 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

$$\begin{aligned}\mathbf{Pr}[T = t, K = k] &= \mathbf{Pr}[K = k \,|\, T = t] \cdot \mathbf{Pr}[T = t] \\ &= \frac{\left(\binom{z+r}{k} - \binom{r}{k}\right) d^{z+r-k}(1-d)^k}{1-d^z} \\ &\quad \cdot P_z(1-d^z)\left(\prod_{\ell=1}^{i} P_{s_\ell} P_{r_\ell}\right) d^s \\ &= \left(\frac{1-d}{d}\right)^k \left(\binom{z+r}{k} - \binom{r}{k}\right) \\ &\quad \cdot d^{z+r+s} P_z \prod_{\ell=1}^{i} P_{s_\ell} P_{r_\ell}\end{aligned} \tag{6}$$

and similarly, by (5)

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

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

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

since $\sum_s Q_{s,i} d^s$ is simply the probability that i blocks from X are deleted, and this probability equals D^i. Similar formulas appear in [4], 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 the i.i.d. deletion channel. 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.

III. 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 probability all but super-polynomially small in N. 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.

A. Typical Outputs

We introduce some notation that will be used throughout this section. Let $\mathbb{P}$ be the class of all distributions P such that P is either the geometric distribution, i.e., $P_j = (1-\alpha)\alpha^{j-1}$ for some constant $0 < \alpha < 1$, or $P_j = 0$ for $j > U$, where $U \geq 1$ is an integer constant. Here U corresponds to the largest block length assumed by any of the finite distributions in $\mathbb{P}$. (We note that any constant U can be chosen; it is just that some of our arguments will require that we have an upper bound U available. Any distribution with finite support can be handled with this approach, using the appropriate constant U.) If not otherwise stated, it should be assumed that the lengths of the blocks in the input X are from a distribution $P \in \mathbb{P}$.

Let $\gamma_1 = \frac{1}{\log(1/\alpha)}$ if P is geometric with parameter α and let $\gamma_1 = \frac{1}{2}$ if P has finite support. (As will soon be clear, the choice of $\frac{1}{2}$ here is arbitrary. Any constant $\gamma_1 > 0$ would do.) Also, let $\gamma_2 = \frac{1}{\log[1/D]}$ and $\gamma_3 = \gamma_1 \cdot (2\gamma_2 + 1)$. It is straightforward to show via a union bound that with probability at least $1 - N^{-\Omega(\log N)}$, both of the following happen: no block in X is longer than $\gamma_1 \log^2 N$, and, since the probability of a type is at most D^i, no type consists of more than $2\gamma_2 \log^2 N + 1$ blocks. We conclude that with probability at least $1 - N^{-\Omega(\log N)}$, no type is longer than $\gamma_1 \cdot (2\gamma_2 + 1) \cdot \log^4 N = \gamma_3 \log^4 N$. This immediately implies that no block in Y is longer than $\gamma_1(\gamma_2 + 1) \cdot \log^4 N < \gamma_3 \log^4 N$ with probability at least $1 - N^{-\Omega(\log N)}$.

Although types as long as $\Omega(\log^4 N)$ appear with nonnegligible probability (given that we want our results to hold with probability at least $1 - N^{-\Omega(\log N)}$), for distributions $P \in \mathbb{P}$, there may be types that appear sufficiently frequently that we must consider them, but not sufficiently frequently that we can apply standard Chernoff bounds to them. Fortunately, (3) implies that the vast majority of types appearing in our codeword are relatively short (a formal proof of this statement appears in Section III-C, Proposition 1, which shows that the total appearances of "long" types for a suitable definition of long are $o(N)$). Therefore, our subsequent analysis will mainly focus on the large set of "short" types as long types will be easier to handle due to their small number of occurrences. We define the set of short types $\mathbb{T}$ to consist of all types with at most $(\log N)^{2/3}$ bits.[1] Similarly, we define $\mathbb{K}$ to be the set of block lengths that are no longer than $(\log N)^{2/3}$; hence $\mathbb{K}$ is the set of all possible block lengths arising from types in $\mathbb{T}$. We also define $\mathbb{S}$ to be the set of pairs (t, k) such that $t \in \mathbb{T}$, $k \in \mathbb{K}$, and $\mathbf{Pr}[T = t, K = k] \geq N^{-1/3}$.[2] Let $\epsilon = \delta = N^{-1/3}$, and $\gamma = N^{-1/6}$. A standard application of Chernoff bounds (e.g., [10, Theorems 4.2 and 4.3]) shows that the received sequence consists of $N(1 - d) \cdot (1 \pm \epsilon)$ bits, with probability at least $1 - e^{-\Theta(N^{1/3})}$. For

$$\mathcal{B} = \frac{N(1-d)}{\sum_k k\mathcal{P}_k}$$

conditioned on $N(1-d)(1 \pm \epsilon)$ bits in Y, the number of blocks in Y is $\mathcal{B}(1 \pm \epsilon)(1 \pm \delta)$ with probability at least $1 - e^{-\Theta(N^{1/3})}$ (again, see [4, Proposition 1] for a proof). A received sequence Y is a typical output for some codeword X if it consists of $\mathbf{Pr}[T = t, K = k] \cdot \mathcal{B} \cdot (1 \pm \epsilon)(1 \pm \delta)(1 \pm \gamma)$ blocks of length k arising from type t, for all pairs (t, k) in $\mathbb{S}$. The following theorem shows that a received sequence Y is a typical output for some codeword X with all but vanishingly small probability.

Theorem 1: A received sequence Y fails to be a typical output for some codeword X with probability at most $e^{-\Theta(N^{1/3})}$.

Proof: Consider a pair (t, k) in $\mathbb{S}$. Using a standard application of Chernoff bounds (e.g., [10, Theorems 4.2 and 4.3]), conditioned on the number of blocks in Y being $\mathcal{B} \cdot (1 \pm \epsilon)(1 \pm \delta)$, the probability that such a pair fails to comply with the definition of a typical output is at most $e^{-\Omega(N^{1/3})}$. For any distribution P, a fixed family $F(i, z, r, s)$ with $i > 0$ may consist of at most $\binom{r-1}{i-1} \cdot \binom{s-1}{i-1}$ types. (When $i = 0$ there is just one type in the family.) Then the size of $\mathbb{T}$ is at most

$$\begin{aligned}|\mathbb{T}| &< (\log N)^{2/3} + \sum_{i=1}^{(\log N)^{2/3}} \sum_{z=1}^{(\log N)^{2/3}} \sum_{r=i}^{(\log N)^{2/3}} \sum_{s=i}^{(\log N)^{2/3}} \binom{r-1}{i-1} \cdot \binom{s-1}{i-1} \\ &\leq (\log N)^{2/3} \cdot \sum_{i=0}^{(\log N)^{2/3}} \binom{(\log N)^{2/3}}{i} \cdot \binom{(\log N)^{2/3}}{i} \\ &< (\log N)^{2/3} \cdot \left(2^{(\log N)^{2/3}}\right)^2\end{aligned}$$

and we obtain the following upper bound for the size of $\mathbb{T} \times \mathbb{K}$:

$$|\mathbb{T} \times \mathbb{K}| < (\log N)^{4/3} \cdot 2^{2(\log N)^{2/3}}. \tag{9}$$

Conditioned on $\mathcal{B}(1 \pm \epsilon)(1 \pm \delta)$ blocks in Y, the probability that there exists a pair in $\mathbb{S}$ that causes Y to be an atypical output is at most

$$\begin{aligned}|\mathbb{S}| \cdot e^{-\Omega(N^{1/3})} &< (\log N)^{4/3} \cdot 2^{2(\log N)^{2/3}} \cdot e^{-\Omega(N^{1/3})} \\ &= e^{-\Omega(N^{1/3})}.\end{aligned}$$

Since Y consists of $\mathcal{B}(1 \pm \epsilon)(1 \pm \delta)$ blocks with probability at least $1 - e^{-\Theta(N^{1/3})}$, the theorem follows. □

B. Decoding Error Probability: A Simplified Analysis

We now develop the main analysis of our paper. In the following, we temporarily 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 t, k. Conditioned on the output being a typical output, the number of such blocks is really $\mathbf{Pr}[T = t, K = k] \cdot \mathcal{B} \cdot (1 \pm o(1))$ for pairs in $\mathbb{S}$, i.e., pairs that occur with sufficiently high probability. A more careful analysis, given in Section III-C, shows that the $o(1)$ terms and the effect of pairs outside $\mathbb{S}$ affect the asymptotic capacity bounds we derive by an $o(1)$ term, which can be ignored.

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; let F_ℓ be the ℓth family in that enumeration and denote by $t_j^{F_\ell}$ the jth

[1] The choice of $(\log N)^{2/3}$ is by no means unique: it is simply a choice that allows for our analysis to go through. Other $o(\log N)$ values might work as well; no attempt has been made to optimize for this quantity hence $\mathbb{T}$ should not be considered as *the* typical set of types but rather as a convenient set that includes the vast majority of the most frequently occurring types.

[2] The choice of the threshold probability $N^{-1/3}$ is only to keep the analysis clean; Section III-C concludes that other choices for this probability work as well.

type in family F_ℓ. 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. In intuitive terms, we have a jigsaw puzzle, with each piece corresponding to a (t, k) pair of a type and a block length, and initially we have the right number of pieces for each pair. We consider all possible ways of putting the jigsaw puzzle together consistent with the received sequence and the pieces we begin with. After doing this, 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

$$\begin{aligned}
\prod_k &\binom{\mathcal{B}_k}{\mathcal{B}_{t_1^{F_1},k}; \ldots \mathcal{B}_{t_1^{F_2},k}; \ldots} \\
&\leq 2^{-\sum_k \mathcal{B}_k \sum_F \sum_{t \in F} \mathbf{Pr}[T=t \mid K=k] \log(\mathbf{Pr}[T=t \mid K=k])} \\
&= 2^{-\mathcal{B} \sum_k \sum_F \sum_{t \in F} \mathbf{Pr}[T=t, K=k] \log(\mathbf{Pr}[T=t \mid K=k])} \qquad (10) \\
&= 2^{\frac{N(1-d)}{\sum_k k\mathcal{P}_k} H(T \mid K)}. \qquad (11)
\end{aligned}$$

The first inequality follows from using Stirling's formula (e.g., see [10]) to bound the factorials. Also, the expression of (11) 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, and is an open problem.

To upper bound the probability that a fixed codeword X in our codebook could yield one of the sequences of types counted in (11), we restrict the codebook to consist only of the likely codewords. That is, standard methods (e.g., see [1, Ch. 3]) 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

$$\begin{aligned}
2^{NR} \cdot 2&^{\frac{N(1-d)}{\sum_k k\mathcal{P}_k} H(T \mid K) - \frac{N}{\sum_j jP_j} H(P)} \\
&= \left(2^{R + \frac{1-d}{\sum_k k\mathcal{P}_k} H(T \mid K) - \frac{1}{\sum_j jP_j} H(P)} \right)^N. \qquad (12)
\end{aligned}$$

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 the right-hand side of (12). For the decoder to fail with probability that goes to zero asymptotically it suffices that the rate is upper-bounded by

$$\begin{aligned}
R &< \frac{1}{\sum_j jP_j} H(P) - \frac{1-d}{\sum_k k\mathcal{P}_k} H(T \mid K) \\
&= \frac{1}{\sum_j jP_j} H(P) - \frac{1-d}{\sum_k k\mathcal{P}_k} (H(T, K) - H(K)).
\end{aligned}$$

Therefore, we obtain the following theorem.

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

$$C_{\text{del}} \geq \sup_{P \in \mathbb{P}} \left[\frac{1}{\sum_j jP_j} H(P) - \frac{(1-d)}{\sum_k k\mathcal{P}_k} (H(T, K) - H(\mathcal{P})) \right] \qquad (13)$$

for $\mathbf{Pr}[T = t, K = k]$ given by (6) and $\mathcal{P}$ given by (8).

Although the above argument only provides a random codebook to be sent through the deletion channel, via standard arguments it implies the existence of a fixed set of codewords of length n for the deletion channel that could be used in place of the random generating process.

The following lemma provides a simplified formula for $H(T, K)$ for arbitrary P (the proof appears in the Appendix).

Lemma 1: The joint entropy $H(T, K)$ of the joint distribution of the types in X and the block lengths in Y is given by

$$\begin{aligned}
H(T, K) = -\sum_k \sum_{(i,z,r,s)} \sum_{t \in F(i,z,r,s)} &\mathbf{Pr}[T = t, K = k] \\
&\cdot \log \left[\binom{r+z}{k} - \binom{r}{k} \right] \qquad (14) \\
&+ \left(H(d) + \frac{H(P)}{\sum_j jP_j} \right) \cdot \frac{\sum_k k\mathcal{P}_k}{1-d}. \qquad (15)
\end{aligned}$$

Although it might be difficult to derive a closed formula for the summation term in (14) for arbitrary distributions P, one can easily compute it numerically for fixed P, d over a limited range of t and k. Since each term inside the summation is positive and $H(T, K)$ appears negated in (13), summing over a finite number of these terms yields strict underestimates of the actual lower bounds derivable by our approach. We immediately obtain the following simplification of Theorem 2.

Theorem 3: 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

$$C_{\text{del}} \geq \sup_{P \in \mathbb{P}} \left[-H(d) + \frac{(1-d)}{\sum_k k\mathcal{P}_k} \left(H(\mathcal{P}) + \sum_k \sum_{(i,z,r,s)} \sum_{t \in F} \mathbf{Pr}[T=t, K=k] \cdot \log\left[\binom{r+z}{k} - \binom{r}{k} \right] \right) \right] \tag{16}$$

for $\mathbf{Pr}[T=t, K=k]$ given by (6), $\mathcal{P}$ given by (8), and F standing for $F(i,z,r,s)$.

C. Analysis Revisited

As mentioned early in the previous section, the capacity bound given by Theorem 2 considers outputs satisfying more properties than the typical outputs defined in Section III-A. In particular, in our simplified analysis, we assume that the number of occurrences of each (t,k) pair matches its expectation. However, the definition of typical outputs allows for deviations from the expectation for pairs in $\mathbb{S}$, and the deviations for all other pairs may be significantly larger. In this section, we show that due to the asymptotic nature of our bounds, the effect of these deviations can be ignored. That is, the new decoding algorithm, which we describe below, implied by these deviations considers a number of potentially transmitted codewords which is within a $2^{o(N)}$ factor of the upper bound of (11), with probability at least $1 - N^{-\Omega(\log N)}$. Hence, the subsequent analysis in Section III-B still holds under the new decoding algorithm.

The new decoding algorithm, to cope with the deviations in the number of occurrences of every pair (t,k), can be thought of in the following way. First, instead of solving one jigsaw puzzle, the algorithm will try to solve many different jigsaw puzzles. Each puzzle will correspond to a set of values for $\mathcal{B}_{t,k}$, giving the number of pieces for each (t,k) pair. Second, since for certain rarely occurring (t,k) pairs, the number of such pairs may deviate largely from their expectation, we develop special jigsaw puzzle pieces to more easily cope with such pairs.

We therefore start by explicitly specifying the range of values $\mathcal{B}_{t,k}$ may take on for every (t,k). Consider a fixed received sequence Y; Section III-A guarantees that Y is a typical output for some codeword X with super-polynomially high probability. In the previous section, we assumed that in a typical output Y, $\mathcal{B}_k$ exactly equals $\mathcal{P}_k\mathcal{B}$ for all k, and $\mathcal{B}_{t,k}$ exactly equals $\mathbf{Pr}[T=t, K=k]\mathcal{B}$ for all (t,k). However, the definition of a typical output Y in Section III-B allows $\mathcal{B}_k$ and $\mathcal{B}_{t,k}$ to assume values within a permitted range. More specifically, $\mathcal{B}_k$ may be any number between $\mathcal{P}_k\mathcal{B}(1-\epsilon)(1-\delta)(1-\gamma_k)$ and $\mathcal{P}_k\mathcal{B}(1+\epsilon)(1+\delta)(1+\gamma_k)$, where $\epsilon = \delta = N^{-1/3}$ denote variations from the expectation in the number of bits and the number of blocks in Y, respectively, while γ_k denotes variations in the number of blocks of length k. Similarly, $\mathcal{B}_{t,k}$ may be any number between $\mathbf{Pr}[T=t, K=k]\mathcal{B}(1-\epsilon)(1-\delta)(1-\gamma_{t,k})$ and $\mathbf{Pr}[T=t, K=k]\mathcal{B}(1+\epsilon)(1+\delta)(1+\gamma_{t,k})$, where $\gamma_{t,k}$ denotes variations in the number of blocks of length k arising from type t. In Section III-A, $\gamma_{t,k}$ (and, therefore, the permitted range of $\mathcal{B}_{t,k}$) was determined for a certain subset of pairs (t,k). The following proposition determines the permitted range of $\mathcal{B}_{t,k}$ for all (t,k).

Proposition 1: Let $|t|$ denote the length of type t. With probability at least $1 - N^{-\Omega(\log N)}$, a typical output Y with respect to some codeword X satisfies all of the following.

- For pairs $(t,k) \in \mathbb{T} \times \mathbb{K}$
$$\mathcal{B}_{t,k} = \mathbf{Pr}[T=t, K=k]\mathcal{B}(1 \pm \Theta(N^{-1/3} + \gamma_{t,k}))$$
where
—$\gamma_{t,k} = N^{-1/6}$ for $(t,k) \in \mathbb{S}$; hence
$$\mathcal{B}_{t,k} = \mathbf{Pr}[T=t, K=k]\mathcal{B}(1 \pm \Theta(N^{-1/6})).$$
—$\gamma_{t,k} = (\frac{O(\log^2 N)}{\mathbf{Pr}[T=t,K=k]\cdot N})^{1/2}$ if
$$\mathbf{Pr}[T=t, K=k] > \frac{\log^2 N}{N}$$
and $\gamma_{t,k} = \frac{O(\log^2 N)}{\mathbf{Pr}[T=t,K=k]\cdot N}$ otherwise, for $(t,k) \in (\mathbb{T} \times \mathbb{K}) - \mathbb{S}$; hence $\mathcal{B}_{t,k}$ is $O(N^{2/3})$.
- For pairs (t,k) such that $\log^{2/3} N < |t| \leq \gamma_3 \log^4 N$ and $k \geq 1$, let V_* denote their total number of appearances. Then for constant $0 < \kappa < 1$, and any positive constant M, V_* is at most
$$V_* \leq \Theta(\kappa^{(\log N)^{2/3}} \cdot N) < \Theta\left(\frac{N}{\log^M N}\right).$$
- No pair (t,k) such that $|t| > \gamma_3 \log^4 N$ occurs.

Proof: In Section III-A, we showed that for each $(t,k) \in \mathbb{S}$ and $\gamma_{t,k} = N^{-1/6}$

$$\mathcal{B}_{t,k} = \mathbf{Pr}[T=t, K=k]\mathcal{B}(1 \pm \epsilon)(1 \pm \delta)(1 \pm \gamma_{t,k})$$

with probability at least $1 - e^{-\Omega(N^{1/3})}$. For $\epsilon = \delta = N^{-1/3}$, we obtain

$$\mathcal{B}_{t,k} = \mathbf{Pr}[T=t, K=k]\mathcal{B}(1 \pm \Theta(N^{-1/6})).$$

Now consider a pair $(t,k) \in (\mathbb{T} \times \mathbb{K}) - \mathbb{S}$. Since $\mathbf{Pr}[T=t, K=k] = N^{-1/3-f(N)}$, $f(N) > 0$, the expected number of appearances of this pair may be too small to allow for as small $\gamma_{t,k}$ and as strong probability bounds as the pairs in $\mathbb{S}$ do. To this effect, let

$$\gamma_{t,k} = \frac{O(\log N)}{\sqrt{\mathbf{Pr}[T=t, K=k] \cdot N}}, \quad \text{if } \mathbf{Pr}[T=t, K=k] > \frac{\log^2 N}{N}.$$

By standard Chernoff bounds (e.g., [10, Theorems 4.2 and 4.3]), the probability that such a pair exceeds its expectation by more than $O(\log N \cdot \sqrt{\mathbf{Pr}[T=t, K=k]N})$ is at most $2^{-\Omega(\log^2 N)} = N^{-\Omega(\log N)}$. Since the expectation of $\mathcal{B}_{t,k}$ is at most $O(N^{2/3})$, we obtain

$$\mathcal{B}_{t,k} \leq O(N^{2/3}) + O(\log N \cdot N^{1/3}) = O(N^{2/3})$$

with probability at least $1 - N^{-\Omega(\log N)}$.

For $\mathbf{Pr}[T=t, K=k] \leq \frac{\log^2 N}{N}$ and $\gamma_{t,k} = \frac{O(\log^2 N)}{Pr[T=t,K=k]\cdot N}$, standard Chernoff bounds (e.g., [10, eq. (4.10)]) for large $\gamma_{t,k}$ yield that the probability that such a pair exceeds its expected number of appearances by more than $O(\log^2 N)$ is at most $2^{-\Omega(\log^2 N)} = N^{-\Omega(\log N)}$. An entirely similar reasoning to the previous case yields that $\mathcal{B}_{t,k} < O(N^{2/3})$ with probability at least $1 - N^{-\Omega(\log N)}$. Hence, for $|\mathbb{T} \times \mathbb{K}|$ upper-bounded by (9), the probability that any of the first two events in the proposition statement fails to happen is at most

$$|\mathbb{T} \times \mathbb{K}| \cdot \left(e^{-\Omega(N^{-1/3})} + N^{-\Omega(\log N)}\right) < 2^{O(\log\log N) + 2(\log N)^{2/3} - \Omega(\log^2 N)} = N^{-\Omega(\log N)}. \quad (17)$$

The total number of appearances of pairs (t,k) such that $\log^{2/3} N < |t| \leq \gamma_3 \log^4 N$ and $k \geq 1$ is upper-bounded in Lemma 2, which we prove immediately below; this bound holds with probability at least $1 - N^{-\Omega(\log N)}$ for sufficiently large N. The proposition follows from the discussion in Section III-A showing that the probability of a type longer than $\gamma_3 \log^4 N$ is at most $N^{-\Omega(\log N)}$. □

Remark 1: Proposition 1 can be used directly to provide bounds for the deviations γ_k of the block lengths $k \in \mathbb{K}$, if in the first part of the proposition (i.e., the part referring to pairs $(t,k) \in \mathbb{T} \times \mathbb{K}$), pairs $(t,k) \in \mathbb{T} \times \mathbb{K}$ are replaced with lengths $k \in \mathbb{K}$. This further implies that $\gamma_{t,k}$ is replaced with γ_k, $\mathbf{Pr}[T=t, K=k]$ with $\mathcal{P}_k = \mathbf{Pr}[K=k]$, and $\mathbb{S}$ with $\{k \in \mathbb{K}, \mathcal{P}_k \geq N^{-1/3}\}$. The proposition may be used to provide bounds for the deviations from the conditional probabilities $\mathbf{Pr}[T=t \,|\, K=k]$ of the pairs $(t,k) \in \mathbb{T} \times \mathbb{K}$ in an entirely similar fashion: $\mathbf{Pr}[T=t, K=k]$ is replaced with $\mathbf{Pr}[T=t \,|\, K=k]$ and $\mathbb{S}$ with $\{(t,k) \in \mathbb{T} \times \mathbb{K}, \mathbf{Pr}[T=t \,|\, K=k] \geq N^{-1/3}\}$. This latter observation will actually be used in Lemma 5.

Lemma 2: Let $0 < \kappa < 1$ be a constant given by

$$\kappa = \begin{cases} \alpha + (1-\alpha)\sqrt{d}, & \text{if } P_j = (1-\alpha)\alpha^{j-1},\ j \geq 1,\ 0 < \alpha < 1 \\ D^{\frac{1}{2U}}, & \text{if } P_j = 0 \text{ for } j > U \end{cases}$$

for U as in the definition of $\mathbb{P}$ and $D = \sum_z P_z d^z$. For any positive constant M, the total number of appearances of pairs (t,k) such that $\log^{2/3} N < |t| \leq \gamma_3 \log^4 N$ and $k \geq 1$, denoted by V_*, is upper-bounded as

$$V_* \leq \Theta\left(\kappa^{(\log N)^{2/3}} \cdot N\right) < \Theta\left(\frac{N}{\log^M N}\right)$$

with probability at least $1 - N^{-\Omega(\log N)}$.

Proof: Note that V_* is essentially the number of appearances of types t with $\log^{2/3} N < |t| \leq \gamma_3 \log^4 N$ (since there is no constraint on k). Given this interpretation for V_*, we prove the lemma in two steps. We start by computing the expected value for V_*. To this end, we need compute the probability that a type t with $2i+1$ blocks, $i \geq 0$, has length $|t| = \ell$ for distributions $P \in \mathbb{P}$. Then we apply Chernoff bounds (e.g., [10, Theorems 4.2 and 4.3]) to show that V_* does not deviate significantly from its expectation.

First, suppose that the block lengths in the codeword follow a geometric distribution. In this case, the only restriction for the blocks is that they consist of at least one bit. Therefore, the number of ways to assign ℓ bits into $2i+1$ blocks is $\binom{\ell-1}{2i}$: the last block of the type ends at the ℓth bit, so we only need choose endpoints for the remaining $2i$ blocks. It follows that when blocks are geometrically distributed with parameter α, each arrangement of ℓ bits into $2i+1$ blocks is equiprobable, and occurs with probability $(1-\alpha)^{2i+1} \cdot \alpha^{\ell-2i-1}$. Also, since at least i blocks of the type are deleted, the deleted bits contribute at least d^i to the probability of the type. The latter is now upper-bounded as

$$\begin{aligned} \mathbf{Pr}[|t| = \ell] &\leq \frac{1-\alpha}{\alpha}\alpha^\ell \sum_{i=0}^{\ell} \binom{\ell-1}{2i} \left(\frac{(1-\alpha)\sqrt{d}}{\alpha}\right)^{2i} \\ &\leq \frac{1-\alpha}{\alpha} \cdot \left(\frac{[\alpha + (1-\alpha)\sqrt{d}]^{\ell-1}}{2} + \frac{[\alpha - (1-\alpha)\sqrt{d}]^{\ell-1}}{2}\right) \\ &\leq \frac{1-\alpha}{\alpha}[\alpha + (1-\alpha)\sqrt{d}]^{\ell-1}. \end{aligned}$$

Next, consider distributions P with finite support. Recall that the probability of a type with $2i+1$ blocks is upper-bounded by D^i, where D is the probability that a block is deleted. Then the probability of a type with total length ℓ is at most $D^{-1/2} \cdot D^{\frac{\ell}{2U}}$, since at least $\lceil \frac{\ell}{U} \rceil$ blocks are required to obtain a total length ℓ.

Let $\kappa = \alpha + (1-\alpha)\sqrt{d}$ for geometric P, and $\kappa = D^{\frac{1}{2U}}$ for distributions P with finite support; in both cases, $0 < \kappa < 1$. There are $\Theta(\mathcal{B})$ types in X. The expected number of types with lengths between $(\log N)^{2/3}$ and $\gamma_3(\log N)^4$ (in other words, the expected value of V_*), is at most

$$\Theta(\kappa^{(\log N)^{2/3}} \cdot \mathcal{B}) = \Theta(\kappa^{(\log N)^{2/3}} \cdot N).$$

Now the Chernoff bounds in Theorems 4.2 and 4.3 in [10] guarantee that V_* does not exceed its expectation by more than $\kappa^{(\log N)^{2/3}} \cdot N^{2/3}$ with probability at least

$$1 - e^{-\Theta(\kappa^{(\log N)^{2/3}} \cdot N^{1/3})} \geq 1 - 2^{-\Omega(\log^2 N)}.$$

Therefore, $V_* \leq \Theta(\kappa^{(\log N)^{2/3}} \cdot N)$ with the same probability.

Since

$$\kappa^{(\log N)^{2/3}} N \leq \frac{N}{\log^M N}$$

for sufficiently large N for any positive constant M, we can conclude the derivation. □

We now proceed to describe our decoding algorithm in more detail. Our algorithm will attempt to solve multiple jigsaw puzzles. A jigsaw puzzle is determined by the numbers of various pieces. Every pair $(t,k) \in \mathbb{T} \times \mathbb{K}$ introduces a piece; the number of such pieces is allowed to be any number in the range satisfying the constraint that the output is typical. Similarly, $\mathcal{B}_k$ must take on an appropriate value under this constraint. Of course for each k, $\mathcal{B}_k - \sum_{t \in \mathbb{T}} \mathcal{B}_{t,k} \geq 0$. For each $k \in \mathbb{K}$ such that $\mathcal{B}_k - \sum_{t \in \mathbb{T}} \mathcal{B}_{t,k} > 0$, the decoder considers all remaining blocks of this length as arising from the same special type $*$, which could correspond to *any* binary string of length between $(\log N)^{2/3}$ and $\gamma_3(\log N)^4$. The jigsaw puzzle will have a corresponding number of pieces for the pair $(*, k)$. (Of course, allowing *any* string of a suitable length for puzzle pieces with a special type $*$ will simply overestimate the number of possible resulting codewords. We are ignoring, for example, that this binary string must be such that the output block could be derived from this string. Since we are seeking a lower bound on the capacity, such overestimates are valid, and because pieces with type $*$ are so rare, they do not affect our asymptotic bound.) Note that each $k \in \mathbb{K}$ may introduce some $(*, k)$ pieces to a puzzle, while every $k \notin \mathbb{K}$ introduces only such pieces. Given the received Y (i.e., given the value of $\mathcal{B}_k$ for every k), and given a valid assignment of values to $\mathcal{B}_{t,k}$ for $t \in \mathbb{T}$, $\mathcal{B}_{*,k}$ is fully determined for every k (and, therefore, no new puzzles need to be introduced because of these pieces). The number J of jigsaw puzzles formed by the decoding algorithm is upper-bounded by the following lemma.

Lemma 3: The number J of jigsaw puzzles introduced by deviations in the $\mathcal{B}_{t,k}$ within the ranges permitted by Proposition 1 is at most $2^{o(N)}$ with probability at least $1 - N^{-\Omega(\log N)}$.

Proof: For every pair $(t, k) \in \mathbb{S}$, Proposition 1 specifies at most $\mathbf{Pr}[T = t, K = k] \cdot \mathcal{B} \cdot O(N^{-1/6}) \leq O(N^{5/6})$ deviations from expectation. Therefore, there are at most $O(N^{5/6})$ possible values for each such $\mathcal{B}_{t,k}$. Similarly, for $(t, k) \in (\mathbb{T} \times \mathbb{K}) - \mathbb{S}$, $\mathcal{B}_{t,k}$ may assume any value up to $O(N^{2/3})$. As these values determine each puzzle to be solved, for $|\mathbb{T} \times \mathbb{K}|$ upper-bounded by (9), J is upper-bounded by

$$\begin{aligned} J &< \prod_{(t,k)\in\mathbb{S}} O(N^{5/6}) \cdot \prod_{(t,k)\in(\mathbb{T}\times\mathbb{K})-\mathbb{S}} O(N^{2/3}) \\ &< (O(N^{5/6}))^{|\mathbb{T}\times\mathbb{K}|} < 2^{O(\log^{7/3} N \cdot 2^{2(\log N)^{2/3}})} = 2^{o(N)}. \end{aligned}$$

This bound fails with probability at most $N^{-\Omega(\log N)}$ which is the probability that there exists a pair (t, k) in $\mathbb{T} \times \mathbb{K}$ that exceeds its expectation by an amount larger than the amount specified by Proposition 1 for this pair. □

Each jigsaw puzzle is input to a slightly modified version of the decoding algorithm of Section III-B that handles the $*$ types: when such a type is encountered, the decoder lists all binary strings with lengths between $(\log N)^{2/3}$ and $\gamma_3(\log N)^4$ as possible substrings of the codeword at the position of the $*$. Clearly, the total number of appearances of pairs $(*, k)$, given by $\sum_{k\geq 1} \mathcal{B}_{*,k}$, is given by V_*, for V_* as in Proposition 1. Since each $*$ type may be replaced by at most $2 \cdot 2^{\gamma_3(\log N)^4}$ strings, the number S of sequences which the new decoder considers as potentially transmitted codewords is at most

$$\begin{aligned} S &\leq J \cdot \prod_{k\in\mathbb{K}} \frac{\mathcal{B}_k!}{\left(\prod_{t\in\mathbb{T}} \mathcal{B}_{t,k}!\right) \cdot \mathcal{B}_{*,k}!} \cdot \prod_{k\geq 1} \left(2^{\gamma_3(\log N)^4+1}\right)^{\mathcal{B}_{*,k}} \\ &\leq 2^{o(N)} \cdot 2^{(\gamma_3(\log N)^4+1)\cdot V_*} \cdot \prod_{k\in\mathbb{K}} \frac{\mathcal{B}_k!}{\left(\prod_{t\in\mathbb{T}} \mathcal{B}_{t,k}!\right) \cdot \mathcal{B}_{*,k}!} \\ &\leq 2^{o(N)} \cdot \prod_{k\in\mathbb{K}} \frac{\mathcal{B}_k!}{\left(\prod_{t\in\mathbb{T}} \mathcal{B}_{t,k}!\right) \cdot \mathcal{B}_{*,k}!} \end{aligned} \tag{18}$$

where the last equation follows from the bound on V_* in Proposition 1 (e.g., for $M \geq 5$). The upper bound on S holds with probability at least $1 - N^{-\Omega(\log N)}$, since each of the upper bounds on J and V_* fails with probability at most $N^{-\Omega(\log N)}$ (see Lemma 3 and Proposition 1). The new decoder is successful if only one of these sequences (the actual sequence sent) corresponds to a codeword in our codebook.

We are now ready to show the main result of this section in Lemma 4 below: the expression in (18) is upper-bounded by the right-hand side of (10), with the possible addition of a multiplicative term of $2^{o(N)}$, with probability at least $1 - N^{-\Omega(\log N)}$. We can then infer that the analysis of our new decoding algorithm results in the same capacity bounds proven in Theorems 2 and 3 under the simplified analysis.

Lemma 4: With probability at least $1 - N^{-\Omega(\log N)}$, expression (19) at the bottom of the page holds.

Proof: We first observe that ignoring the effect of the $*$ types yields an upper bound for the left-hand side of (19). We write $\mathcal{B}_k$ as $\mathcal{P}_k\mathcal{B}(1 \pm \Theta(N^{-1/3}))(1 \pm \gamma_k)$ (from the discussion preceding Proposition 1). Recall also that $\mathcal{B}_{t,k} = \mathbf{Pr}[T = t, K = k]\mathcal{B}(1 \pm \Theta(N^{-1/3}))(1 \pm \gamma_{t,k})$ (from Proposition 1). We introduce a further expression $\gamma_{t|k}$, and rewrite $\mathcal{B}_{t,k} = \mathbf{Pr}[T = t \,|\, K = k]\mathcal{P}_k\mathcal{B}(1 \pm \Theta(N^{-1/3}))(1 \pm \gamma_{t|k})(1 \pm \gamma_k)$, where $(1 \pm \gamma_{t|k})(1 \pm \gamma_k) = 1 \pm \gamma_{t,k}$. The role of $\gamma_{t|k}$ will be clarified below. Using the standard entropy-based bound for combinations we now obtain the following upper bound for the left-hand side of (19):

$$\begin{aligned} &\mathrm{pow}\Bigg(2, -\sum_{k\in\mathbb{K}} \mathcal{B}_k(1 \pm \Theta(N^{-1/3}))(1 \pm \gamma_k) \\ &\qquad \cdot \sum_{t\in\mathbb{T}} \mathbf{Pr}[T = t \,|\, K = k] \cdot (1 \pm \gamma_{t|k}) \\ &\qquad \cdot \log(\mathbf{Pr}[T = t \,|\, K = k] \cdot (1 \pm \gamma_{t|k}))\Bigg) \\ &= \mathrm{pow}\Bigg(2, -\mathcal{B} \cdot \sum_{t\in\mathbb{T}} \sum_{k\in\mathbb{K}} \mathbf{Pr}[T = t, K = k] \\ &\qquad \cdot (1 \pm \Theta(N^{-1/3} + \gamma_{t,k})) \\ &\qquad \cdot \log\left(\mathbf{Pr}[T = t \,|\, K = k] \cdot (1 \pm \gamma_{t|k})\right)\Bigg). \end{aligned} \tag{20}$$

In order to prove the lemma, if suffices to show that, with probability at least $1 - N^{-\Omega(\log N)}$, the total effect of terms in (20) which include the $\Theta(N^{-1/3} + \gamma_{t,k})$ and/or the $\log(1 \pm \gamma_{t|k})$ factor is $2^{o(N)}$. This would imply that the expression in (20) is equivalent, up to $2^{o(N)}$ terms, to

$$2^{-\mathcal{B}\sum_{t\in\mathbb{T}}\sum_{k\in\mathbb{K}} \mathbf{Pr}[T=t,K=k] \cdot \log \mathbf{Pr}[T=t \,|\, K=k]}$$

which is clearly upper bounded by the expression in (10).

$$\prod_{k\in\mathbb{K}} \frac{\mathcal{B}_k!}{\left(\prod_{t\in\mathbb{T}} \mathcal{B}_{t,k}!\right) \mathcal{B}_{*,k}!} \leq 2^{-\mathcal{B}\left[\sum_k \sum_F \sum_{t\in F} \mathbf{Pr}[T=t,K=k] \log(\mathbf{Pr}[T=t|K=k])\right]+o(N)}. \tag{19}$$

Let A denote a set of pairs (t,k). We define

$$R_1^A = \left| \mathcal{B} \cdot \sum_{(t,k)\in A} \mathbf{Pr}[T=t, K=k] \cdot \log(1 \pm \gamma_{t|k}) \right| \tag{21}$$

$$R_2^A = \left| \mathcal{B} \cdot \sum_{(t,k)\in A} \mathbf{Pr}[T=t, K=k] \cdot \Theta(N^{-1/3} + \gamma_{t,k}) \cdot \log \mathbf{Pr}[T=t \,|\, K=k] \right| \tag{22}$$

$$R_3^A = \left| \mathcal{B} \cdot \sum_{(t,k)\in A} \mathbf{Pr}[T=t, K=k] \cdot \Theta(N^{-1/3} + \gamma_{t,k}) \cdot \log(1 \pm \gamma_{t|k}) \right|.$$

Every term in the exponent of (20) where a $\Theta(N^{-1/3} + \gamma_{t,k})$ and/or $\log(1 \pm \gamma_{t|k})$ factor appears is included in exactly one of the $R_1^{\mathbb{T}\times\mathbb{K}}$, $R_2^{\mathbb{T}\times\mathbb{K}}$, and $R_3^{\mathbb{T}\times\mathbb{K}}$. Lemma 5 below shows that $R_1^{\mathbb{T}\times\mathbb{K}} = R_2^{\mathbb{T}\times\mathbb{K}} = R_3^{\mathbb{T}\times\mathbb{K}} = o(N)$, with probability at least $1 - N^{-\Omega(\log N)}$. By our discussion above, the proof of the lemma is complete. □

Lemma 5: With probability at least $1 - N^{-\Omega(\log N)}$

$$R_1^{\mathbb{T}\times\mathbb{K}} = R_2^{\mathbb{T}\times\mathbb{K}} = R_3^{\mathbb{T}\times\mathbb{K}} = o(N).$$

Proof: We will prove this statement by considering pairs in $\mathbb{S}$ and $(\mathbb{T}\times\mathbb{K}) - \mathbb{S}$ separately.

First, consider pairs $(t,k) \in \mathbb{S}$. For each such pair, $\mathbf{Pr}[T = t, K = k]$ is upper-bounded by some constant $0 < c < 1$. Since $N^{-1/3} < \mathbf{Pr}[T=t, K=k] \leq \mathbf{Pr}[T=t \,|\, K=k] \leq 1$, we have

$$\begin{aligned} &|\log \mathbf{Pr}[T=t \,|\, K=k]| \\ &\qquad \leq |\log \mathbf{Pr}[T=t, K=k]| \leq |\log N^{-1/3}| = \frac{1}{3}\log N. \end{aligned}$$

From Proposition 1, $\gamma_{t,k} = N^{-1/6}$ for each pair in $\mathbb{S}$, hence $\Theta(N^{-1/3} + \gamma_{t,k}) = \Theta(N^{-1/6})$. Since $\mathbf{Pr}[T=t \,|\, K=k] \geq N^{-1/3}$, Remark 1 yields $\gamma_{t|k} = N^{-1/6}$ for all $(t,k) \in \mathbb{S}$, and we therefore have that $\log(1 \pm \gamma_{t|k}) = \Theta(N^{-1/6})$. Finally, by (9)

$$S \leq |\mathbb{T}\times\mathbb{K}| < N^{\frac{4}{3}\frac{\log\log N}{\log N} + \frac{2}{\log^{1/3} N}} = N^{o(1)}.$$

Substituting the above into (21), (22), and (23), we obtain for $R_1^{\mathbb{S}}$, $R_2^{\mathbb{S}}$, and $R_3^{\mathbb{S}}$, respectively

$$\begin{aligned} R_1^{\mathbb{S}} &\leq \mathcal{B} \cdot |\mathbb{T}\times\mathbb{K}| \cdot c \cdot \Theta(N^{-1/6}) \leq N^{\frac{5}{6}+o(1)} = o(N), \\ R_2^{\mathbb{S}} &\leq \mathcal{B} \cdot |\mathbb{T}\times\mathbb{K}| \cdot c \cdot \Theta(N^{-1/6}) \cdot \frac{1}{3}\log N \\ &\leq N^{\frac{5}{6}+o(1)} = o(N), \\ R_3^{\mathbb{S}} &\leq \mathcal{B} \cdot |\mathbb{T}\times\mathbb{K}| \cdot c \cdot \Theta(N^{-1/6}) \cdot \Theta(N^{-1/6}) \\ &\leq N^{\frac{2}{3}+o(1)} = o(N). \end{aligned}$$

The above bounds hold with probability at least $1 - 2|\mathbb{T}\times\mathbb{K}| \cdot e^{-\Omega(N^{1/3})} = 1 - e^{-\Omega(N^{1/3})}$.

Next, consider pairs (t,k) in $(\mathbb{T}\times\mathbb{K}) - \mathbb{S}$. In our analysis for these pairs, we will assume that all blocks in X and all pairs (t,k) arise with probability at least $N^{-\Omega(\log N)}$. By a union bound over the $\Theta(N)$ blocks in X and the $|\mathbb{T}\times\mathbb{K}| = N^{o(1)}$ pairs, the subsequent analysis will hold with probability at least $1 - N^{-\Omega(\log N)}$.

In the process of bounding the effect of the deviations of pairs (t,k) in $(\mathbb{T}\times\mathbb{K}) - \mathbb{S}$, we first show that the absolute value of the effect of all these pairs in (10) is at most $2^{o(N)}$. Then it will be easy to deduce that their deviations also contribute at most $2^{o(N)}$ to (20).

More specifically, let $\mathcal{A}$ denote the logarithm of the absolute value of the effect of all $(t,k) \in (\mathbb{T}\times\mathbb{K}) - \mathbb{S}$ in (10), i.e.,

$$\mathcal{A} = \mathcal{B} \sum_{(t,k)\in(\mathbb{T}\times\mathbb{K})-\mathbb{S}} |\mathbf{Pr}[T=t, K=k] \cdot \log(\mathbf{Pr}[T=t \,|\, K=k])|. \tag{24}$$

We will show that $\mathcal{A}$ is $o(N)$. As argued previously, $|\log \mathbf{Pr}[T=t, K=k]|$ provides an upper bound for $|\log \mathbf{Pr}[T=t \,|\, K=k]|$. Therefore, to upper-bound the right-hand side of (24), a lower bound for $\mathbf{Pr}[T=t, K=k]$ is required. A rather crude yet sufficient for our purposes such bound arises from (6)

$$\begin{aligned} \mathbf{Pr}[T=t, K=k] &\geq (1-d)^{\gamma_3 \log^4 N} \cdot d^{\gamma_3 \log^4 N} \\ &\quad \cdot \left(N^{-\Omega(\log N)}\right)^{2\gamma_2 \log^2 N + 1} \\ &= 2^{-\Omega(\log^4 N)}. \end{aligned} \tag{25}$$

Here, we used our assumption that all blocks in X arise with probability at least $N^{-\Omega(\log N)}$. Since for each (t,k) in $(\mathbb{T}\times\mathbb{K}) - \mathbb{S}$, $\mathbf{Pr}[T=t, K=k] < N^{-1/3}$, substituting the result of (25) into (24), we obtain

$$\begin{aligned} \mathcal{A} &\leq \Theta(N) \cdot |\mathbb{T}\times\mathbb{K}| \cdot N^{-1/3} \cdot O(\log^4 N) \\ &\leq N^{\frac{2}{3}+o(1)} = o(N). \end{aligned} \tag{26}$$

We now move to computing each of $R_1^{(\mathbb{T}\times\mathbb{K})-\mathbb{S}}$, $R_2^{(\mathbb{T}\times\mathbb{K})-\mathbb{S}}$, and $R_3^{(\mathbb{T}\times\mathbb{K})-\mathbb{S}}$. We start by providing upper bounds for $\gamma_{t,k}$ and $\gamma_{t|k}$. For the former, Proposition 1 guarantees that

$$\begin{aligned} &\mathcal{B} \cdot \mathbf{Pr}[T=t, K=k] \cdot \gamma_{t,k} \\ &\quad \leq O(\log N) \cdot (\mathbf{Pr}[T=t, K=k] \cdot N)^{1/2} \\ &\quad \leq O(N^{1/3} \log N), \end{aligned}$$

since $\mathbf{Pr}[T=t, K=k] < N^{-1/3}$.

Next, we bound the deviations $\gamma_{t|k}$ of the pairs in $(\mathbb{T}\times\mathbb{K}) - \mathbb{S}$. To this end, we observe that, from Proposition 1 and our assumption that no pair occurring with probability less than $N^{-\Omega(\log N)}$ arises, $\gamma_{t,k}$ is at most

$$\frac{O(\log^2 N)}{N^{-O(\log\{N\})} \cdot N} = N^{O(\log N)}.$$

Again, since $\mathbf{Pr}[T=t \,|\, K=k] \geq \mathbf{Pr}[T=t, K=k]$, Remark 1 yields $\gamma_{t|k} \leq N^{O(\log N)}$ for *all* $(t,k) \in (\mathbb{T}\times\mathbb{K}) - \mathbb{S}$, with probability at least $1 - |\mathbb{T}\times\mathbb{K}| \cdot N^{-\Omega(\log N)} = 1 - N^{-\Omega(\log N)}$.

Hence, $\log(1 \pm \gamma_{t|k}) \le \log(N^{O(\log N)}) = O(\log^2 N)$. We immediately obtain from (21)

$$\begin{aligned} R_1^{(\mathbb{T}\times\mathbb{K})-\mathbb{S}} &\le \Theta(N)\cdot|\mathbb{T}\times\mathbb{K}|\cdot N^{-1/3}\cdot O(\log^2 N)\\ &= N^{\frac{2}{3}+o(1)} = o(N). \end{aligned}$$

Using (25) and (26), we now obtain from (22)

$$\begin{aligned} R_2^{(\mathbb{T}\times\mathbb{K})-\mathbb{S}} &\le \mathcal{A}\cdot\Theta(N^{-1/3})\\ &\quad + |\mathbb{T}\times\mathbb{K}|\cdot O(N^{1/3}\log N)\cdot O(\log^4 N)\\ &\le N^{\frac{1}{3}+o(1)} = o(N). \end{aligned}$$

Similarly, from (23) we obtain

$$\begin{aligned} R_3^{(\mathbb{T}\times\mathbb{K})-\mathbb{S}} &\le \Theta(N)\cdot|\mathbb{T}\times\mathbb{K}|\cdot N^{-1/3}\Theta(N^{-1/3})\cdot O(\log^2 N)\\ &\quad + |\mathbb{T}\times\mathbb{K}|\cdot O(N^{1/3}\log N)\cdot O(\log^2 N)\\ &\le N^{\frac{1}{3}+o(1)} = o(N). \end{aligned}$$

The lemma follows. □

We should mention here that Lemma 5 (and in particular, (26)) further implies that with probability at least $1 - N^{-\Omega(\log N)}$, (10) is an asymptotic upper bound for

$$2^{\mathcal{B}\left(\sum_{(t,k)\in\mathbb{S}} \mathbf{Pr}[T=t,K=k]\cdot\log\frac{1}{\mathbf{Pr}[T=t\,|\,K=k]}\right)}.$$

It is clear that $\mathbb{S}$ above may be substituted by any subset of $\mathbb{T}\times\mathbb{K}$ in which pairs occur with probability at least $N^{-\zeta}$, for small positive constant $0 < \zeta < 1$.

IV. Geometric Distributions

In this section, we use Theorem 3 to derive a lower bound for the capacity of the i.i.d. deletion channel in the special case where the block lengths in X are geometrically distributed, i.e., $P_j = (1-p)p^{j-1}$.

In this case, the probability that m blocks from X have length n given by (4) is simply $Q_{n,m} = \binom{n-1}{m-1}p^{n-m}(1-p)^m$, 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 nth bit). Hence, a family $F(i,z,r,s)$ consists of $\binom{s-1}{i-1}\cdot\binom{r-1}{i-1}$ equiprobable members and the probability of a type becomes

$$\mathbf{Pr}[T=t] = p^{z+s+r}\cdot\left(\frac{1-p}{p}\right)^{2i+1}\cdot(1-d^z)\cdot d^s. \quad (27)$$

(Again, the case $i = 0$ is special but the final (27) is correct.) Then the joint probability of length k and type t becomes

$$\begin{aligned} \mathbf{Pr}[T=t,K=k] = \left(\frac{1-d}{d}\right)^k &\cdot\left[\binom{z+r}{k}-\binom{r}{k}\right]\\ &\cdot(pd)^{z+r+s}\cdot\left(\frac{1-p}{p}\right)^{2i+1}. \quad (28) \end{aligned}$$

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 [2], [4]). It is easy to show that $H(\mathcal{P}) = \frac{H(q)}{1-q}$. Since $\sum_k k\mathcal{P}_k = \frac{1}{1-q}$, we immediately obtain the following corollary to Theorem 3.

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 in bits is lower-bounded by

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

where $q = 1 - \frac{1-p}{1+d-2pd}$.

Note that $\mathbf{Pr}[T\in F(i,z,r,s),K=k]$ is given by the right-hand side of (28) multiplied by the size of $F(i,z,r,s)$, which is $\binom{r-1}{i-1}\binom{s-1}{i-1}$ for $i > 0$.

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. Hence, the graph in Fig. 2 presents an underestimate of the actual rates. Numerical results are given in Table I.

V. A Lower Bound for (m, M, x) Distributions

In this section, we use Theorem 3 to derive a lower bound for the capacity of the i.i.d. deletion channel in the special case where codewords are generated by (m, M, x) distributions. Under such distributions, a block in X is either assigned a short integer length m with probability x or a larger integer length $M > m$ with probability $1-x$. Under the less effective decoding methods of [4], such distributions outperformed the geometric distribution for $d \ge 0.35$.

Consider the concatenation of $b \ge 0$ blocks in X, a of which are short; here $0 \le a \le b$. Since $M > m$, it follows easily that for each choice of b, there are exactly $b+1$ distinct lengths that the concatenation of b blocks may undertake with nonzero probability; each length corresponds to a value of a ranging from 0 to b. Moreover, if n is the length of the concatenation, for fixed n and b there is a unique $0 \le a \le b$ such that $am+(b-a)M = n$. Therefore, using our standard notation $Q_{n,b}$ for the probability that b blocks have length n, we have $Q_{n,b} = \binom{b}{a}x^a(1-x)^{b-a}$, if $n = am+(b-a)M$, for some $0 \le a \le b$; otherwise, $Q_{n,b} = 0$.

Given our preceding discussion, for $i \ge 0$, z in $\{m, M\}$, $0 \le i_r \le i$, $0 \le i_s \le i$, a type t is in the family

$$F(i, z, i_r m + (i-i_r)M, i_s m + (i-i_s)M)$$

if the following happens: it consists of $2i+1$ blocks; the first block has length z; exactly i_r of the blocks which are the same as the first block have length m; and, exactly i_s of the blocks

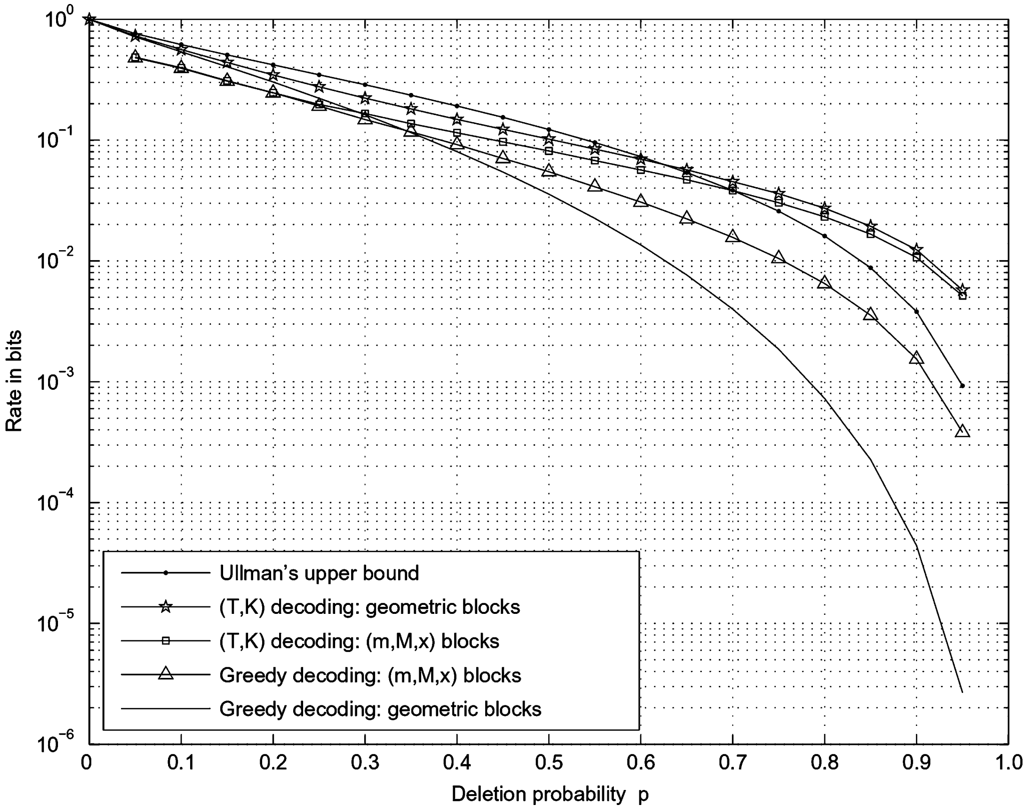

Fig. 2. Improvement in rate with our framework for codewords with geometrically distributed block lengths (Section IV). Comparison with lower bounds for (m, M, x) distributions (Section V), lower bounds for geometric and (m, M, x) distributions from [4], and Ullman's upper bound.

TABLE I
LOWER BOUNDS BASED ON CODEBOOKS DERIVED FROM GEOMETRIC DISTRIBUTIONS

d	p	Rate R in bits
0.05	0.53	0.72829
0.10	0.57	0.56196
0.15	0.62	0.43918
0.20	0.67	0.34669
0.25	0.72	0.27588
0.30	0.77	0.22243
0.35	0.81	0.18101
0.40	0.84	0.14841
0.45	0.87	0.12286
0.50	0.89	0.10186
0.55	0.91	0.084323
0.60	0.92	0.069564
0.65	0.93	0.056858
0.70	0.94	0.045324
0.75	0.96	0.035984
0.80	0.97	0.027266
0.85	0.98	0.019380
0.90	0.985	0.012378
0.95	0.993	0.005741

which differ from the first block have length m. This family is well defined since it is unique for each quadruple (i, z, i_r, i_s). Further, each type in $F(i, z, i_r m + (i - i_r)M, i_s m + (i - i_s)M)$ occurs with the same probability given below

$$\mathbf{Pr}[T = t] = P_z \cdot (1 - d^z) \cdot x^{i_r + i_s}(1 - x)^{2i - i_r - i_s} \cdot d^{i_s \cdot m + (i - i_s) \cdot M}$$

and the joint probability of length k and type t from (6) becomes

$$\begin{aligned} \mathbf{Pr}[T = t, K = k] &= \left(\frac{1 - d}{d}\right)^k \left[\binom{z + i \cdot M - i_r \cdot (M - m)}{k} - \binom{i \cdot M - i_r \cdot (M - m)}{k}\right] \cdot P_z \cdot d^z \\ &\quad \cdot \left(\frac{x \cdot d^{-(M-m)}}{1 - x}\right)^{i_r} \cdot \left(\frac{x \cdot d^{-(M-m)}}{1 - x}\right)^{i_s} \\ &\quad \cdot [(1 - x) \cdot d^M]^{2i}. \end{aligned} \tag{29}$$

For $D = x \cdot d^m + (1 - x) \cdot d^M$, (8) yields

$$\begin{aligned} \mathcal{P}_k = \left(\frac{1 - d}{d}\right)^k \cdot \sum_i [(1 - x) \cdot d^M \cdot D]^i \cdot \sum_{z, i_r} P_z \cdot d^z \\ \cdot \left(\frac{x \cdot d^{-(M-m)}}{1 - x}\right)^{i_r} \cdot \left[\binom{z + i \cdot M - i_r \cdot (M - m)}{k} - \binom{i \cdot M - i_r \cdot (M - m)}{k}\right]. \end{aligned} \tag{30}$$

We obtain the following corollary to Theorem 3.

Corollary 2: Consider a channel that deletes every transmitted bit independently and with probability $0 < d < 1$, a binary input alphabet and (m, M, x) block-length distribution P. The capacity of this channel in bits is lower-bounded by

$$C_{\text{del}} \geq \sup_{0<x<1} \left[\frac{1-d}{\sum_k k\mathcal{P}_k} \cdot \left(\sum_k \sum_{(i,z,i_r,i_s)} \mathbf{Pr}[T \in F(i, z, i_r m + (i-i_r)M, i_s m + (i-i_s)M), K = k] \cdot \log \left[\binom{z + iM - i_r(M-m)}{k} - \binom{iM - i_r(M-m)}{k} \right] \right) + \frac{1-d}{\sum_k k\mathcal{P}_k} H(\mathcal{P}) - H(d) \right]$$

where $\mathcal{P}$ is given by (30).

Although deriving a closed form for $\mathcal{P}$ seems difficult, one can easily compute it numerically for fixed m, M, x, and d. Then it is a matter of optimizing over all integers of $m \geq 1$ and $M > m$, and real $0 < x < 1$. The rates in Fig. 2 are underestimates of the best achievable rates under (m, M, x) distributions, 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 Corollary 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 graph in Fig. 2 shows $\hat{R}^d$; clearly, $R^d \geq \hat{R}^d$. Numerical results are given in Table II.

VI. Discussion of Our Results

As discussed in the Introduction, upper bounds for the capacity of channels with synchronization errors are provided by Ullman [11] and Dolgopolov [6].

Ullman considers binary channels for which the limiting value of the fraction of synchronization errors over the block length of the code as the latter goes to infinity is d. Specifically, he considers a channel that introduces $d \cdot N$ insertions in the first $(1-d) \cdot N$ bits of the codeword. His upper bound on the zero-error capacity of such channels is given by

$$1 - (1+d)\log_2(1+d) + d\log_2(2d). \tag{31}$$

All previous theoretical lower bounds for the capacity of the i.i.d. deletion channel ([2], [4], [6]) were strictly below this bound, and it was used in [2] as a comparison point for their lower bounds, but of course the channel studied is quite different. Our current bounds exceed this upper bound for $d \geq 0.65$. This is not a contradiction, as the channels studied are different, but it is the first time this bound has been provably exceeded for the i.i.d. deletion channel.

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	0.484061
0.10	1	3	0.61	0.397447
0.15	1	3	0.57	0.309695
0.20	1	4	0.57	0.246006
0.25	2	6	0.62	0.197568
0.30	2	7	0.61	0.165961
0.35	2	8	0.59	0.136975
0.40	3	11	0.63	0.114967
0.45	3	12	0.61	0.096723
0.50	4	15	0.63	0.081228
0.55	5	19	0.64	0.067797
0.60	5	20	0.61	0.056580
0.65	7	27	0.64	0.046903
0.70	8	32	0.63	0.038220
0.75	10	41	0.63	0.030344
0.80	13	53	0.63	0.023206
0.85	18	75	0.64	0.016682
0.90	28	117	0.64	0.010686
0.95	59	245	0.64	0.005145

Dolgopolov [6] relies on a theorem by Dobrushin [5] 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 the following upper bounds for the binary i.i.d. deletion channel:

$$\left(1 - \frac{d}{2}\right)\log(2-d) + \frac{d}{2}\log d. \tag{32}$$

These bounds hold for codebooks with nonzero probability of error and therefore are closer to the nature of our bounds. However, they rely on an unproven assumption, and arise from considering codebooks where each codeword is chosen uniformly at random. Therefore, it is not surprising that we exceed (32) for $d > 0.8$: random codes seem to perform well only for very small deletion probabilities (e.g., see [2], [4]).

VII. A General Framework for Analyzing Channels With i.i.d. Insertions (Duplications) and Deletions

As mentioned in the Introduction, a critical advantage of our new approach is that it can be applied to other types of channels with i.i.d. synchronization errors besides channels where deletions only occur. In this section, we show how to use our technique to analyze the following class of channels. Let G be a probability distribution over the nonnegative integers, and let G_j be the probability of integer j. We only consider G with geometrically decreasing tails, i.e., there exist real constants $0 < c_G \leq 1$, $0 \leq \alpha_G < 1$ and integer constant $U_G \geq 1$ such that for $0 \leq j \leq U_G$, $G_j \leq c_G$ and for $j > U_G$, $G_j \leq (1-\alpha_G)\alpha_G^{j-1}$. Consider a channel which, independently for each transmitted bit, either deletes the bit with probability

$0 \leq G_0 < 1$ or transmits the bit together with $j-1$ copies of it with probability G_j, $j \geq 1$. We call this class of channels *i.i.d. channels with effective probability* G. Since G_0 is the deletion probability, we shall henceforth denote it by d to emphasize the correspondence with the deletions-only channel. For example, for the i.i.d. deletion channel, we have $G_0 = d$, $G_1 = 1-d$, and $G_j = 0$ for $j \geq 2$. Also, the average number of bits that a single bit from X contributes to the received sequence Y is given by $\sum_j jG_j$, which in the case of the i.i.d. deletion channel is $1-d$.

We derive a generalization of Theorem 2 for i.i.d. channels with effective distribution G. We first consider distributions G such that $0 < d = G_0 < 1$. For codebooks generated as in Section II, the distribution of the types is still given by

$$\mathbf{Pr}[T=t] = P_z \cdot (1-d^z) \cdot P_{s_1} P_{r_1} \cdots P_{s_i} P_{r_i} \cdot d^s.$$

Let $\rho_{a,b}$ be the probability that $a \geq 1$ bits transmitted over an i.i.d. channel with effective distribution G generate $b \geq 0$ bits. Let i_j be the number of bits that the jth of the a bits generates; clearly, $0 \leq i_j \leq b$ and $\sum_{j=1}^{a} i_j = b$. Then

$$\rho_{a,b} = \sum_{\substack{0 \leq i_1 \leq b \\ \vdots \\ 0 \leq i_a \leq b \\ \sum_{j=1}^{a} i_j = b}} G_{i_1} \cdot G_{i_2} \cdots G_{i_a}. \tag{33}$$

Consider a type $t \in F(i,z,r,s)$, giving rise to a block of length k in Y. Since at least one of the k bits comes from the first z bits, the joint probability of type t and block length k becomes

$$\mathbf{Pr}[T=t, K=k] = P_z P_{r_1} P_{s_1} \cdots P_{r_i} P_{s_i} d^s \cdot (\rho_{z+r,k} - \rho_{r,k} d^z). \tag{34}$$

Then for $D = \sum_z P_z d^z$, the block-length distribution in the received sequence is given by

$$\mathcal{P}_k = \sum_{i=0}^{\infty} D^i \sum_{z=1}^{\infty} \sum_{r=i}^{\infty} P_z Q_{r,i} (\rho_{z+r,k} - \rho_{r,k} \cdot d^z). \tag{35}$$

In order to extend our analysis for the i.i.d. deletion channel, we will need large deviation bounds on the number of blocks in the received sequence. To apply such bounds, we need to show that the moment-generating function of $\mathcal{P}$, denoted by $M_{\mathcal{P}}(t)$, is finite in a small interval around 0 for distributions $P \in \mathbb{P}$. The following lemma computes $M_{\mathcal{P}}(t)$.

Lemma 6: Let $M_G(t) = \sum_{j=0}^{\infty} G_j e^{tj}$ be the moment-generating function of the effective distribution G with $0 < G_0 < 1$. The moment-generating function $M_{\mathcal{P}}(t)$ of the block-length distribution $\mathcal{P}$ in the received sequence is given by

$$M_{\mathcal{P}}(t) = \frac{L(t) - D}{1 - L(t)D}$$

where $d = G_0$, $D = \sum_z P_z d^z$, and $L(t) = \sum_{z=1}^{\infty} P_z \cdot [M_G(t)]^z$. Moreover, the average block length in the received sequence is given by

$$\sum_k k\mathcal{P}_k = \sum_j jG_j \cdot \frac{1+D}{1-D} \cdot \sum_z zP_z \tag{36}$$

Proof: Using (35), we obtain for the moment-generating function $M_{\mathcal{P}}(t)$

$$\sum_k \mathcal{P}_k \cdot e^{tk} = \sum_{i=0}^{\infty} D^i \sum_{z=1}^{\infty} \sum_{r=i}^{\infty} P_z Q_{r,i} \sum_{k=1}^{\infty} e^{tk} \left(\rho_{r+z,k} - \rho_{r,k} \cdot d^z\right) \tag{37}$$

Let $\mathcal{R}_x(t) = \sum_{k=0}^{\infty} \rho_{x,k} e^{tk}$. By (33), we obtain

$$\begin{aligned}
\mathcal{R}_x(t) &= \sum_{k=0}^{\infty} \sum_{i_1=0}^{k} G_{i_1} e^{ti_1} \sum_{\substack{0 \leq i_2 \leq k-i_1 \\ \vdots \\ 0 \leq i_a \leq k-i_1 \\ \sum_{j=2}^{a} i_j = k-i_1}} G_{i_2} \cdots G_{i_a} \cdot e^{t(k-i_1)} \\
&= \sum_{i_1=0}^{\infty} G_{i_1} e^{ti_1} \sum_{k=i_1}^{\infty} \sum_{\substack{0 \leq i_2 \leq k-i_1 \\ \vdots \\ 0 \leq i_a \leq k-i_1 \\ \sum_{j=2}^{a} i_j = k-i_1}} G_{i_2} \cdots G_{i_a} \cdot e^{t(k-i_1)} \\
&= \sum_{i_1=0}^{\infty} G_{i_1} e^{ti_1} \cdot \sum_{k=0}^{\infty} \sum_{\substack{0 \leq i_2 \leq k \\ \vdots \\ 0 \leq i_a \leq k \\ \sum_{j=2}^{a} i_j = k}} G_{i_2} \cdots G_{i_a} \cdot e^{tk} \\
&= \sum_{i_1=0}^{\infty} G_{i_1} e^{ti_1} \cdot \sum_{k=0}^{\infty} \rho_{x-1,k} e^{tk} \\
&= M_G(t) \cdot \mathcal{R}_{x-1}(t).
\end{aligned}$$

It follows that $\mathcal{R}_x(t) = [M_G(t)]^x$. Observing that

$$\sum_{k=1}^{\infty} \rho_{x,k} e^{tk} = \mathcal{R}_x(t) - \rho_{x,0} = \mathcal{R}_x(t) - d^x$$

(37) becomes

$$\begin{aligned}
\sum_k \mathcal{P}_k \cdot e^{t \cdot k} &= \sum_{i=0}^{\infty} D^i \sum_{z=1}^{\infty} \sum_{r=i}^{\infty} P_z Q_{r,i} \\
&\quad \cdot \left([M_G(t)]^{r+z} - d^{r+z} - ([M_G(t)]^r - d^r) d^z\right) \\
&= \sum_{i=0}^{\infty} D^i \left(\sum_{r=i+1}^{\infty} Q_{r,i+1} \cdot [M_G(t)]^r \right. \\
&\quad \left. - D \cdot \sum_{r=i}^{\infty} Q_{r,i} \cdot [M_G(t)]^r \right) \\
&= \sum_{i=0}^{\infty} D^i \cdot \left(L(t)^{i+1} - D \cdot L(t)^i\right) \\
&= \frac{L(t) - D}{1 - L(t) \cdot D},
\end{aligned} \tag{38}$$

where (38) is obtained by an argument entirely similar to the proof of Lemma 1 in [4], which shows that

$$\sum_{r=i}^{\infty} Q_{r,i} \cdot [M_G(t)]^r = L(t)^i, \quad \text{for } i \geq 0$$

(the argument is a simple application of the recursive definition of $Q_{n,m}$).

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

$$\begin{aligned} L'(t) &= \left(\sum_{z=1}^{\infty} P_z \cdot \left(\sum_{j=0}^{\infty} G_j e^{tj}\right)^z\right)' \\ &= \sum_{z=1}^{\infty} P_z \left(\left(\sum_{j=0}^{\infty} G_j e^{tj}\right)^z\right)' \\ &= \sum_{z=1}^{\infty} zP_z \cdot \left(\sum_{j=0}^{\infty} G_j e^{tj}\right)^{z-1} \cdot \left(\sum_{j=0}^{\infty} G_j e^{tj}\right)' \\ &= \sum_{z=1}^{\infty} zP_z \cdot \left(\sum_{j=0}^{\infty} G_j e^{tj}\right)^{z-1} \cdot \sum_{j=0}^{\infty} jG_j e^{tj}. \end{aligned}$$

Hence $L'(0) = \sum_j jG_j \cdot \sum_z zP_z$ and we obtain for $\sum_k k\mathcal{P}_k = M'_{\mathcal{P}}(0)$

$$\begin{aligned} M'_{\mathcal{P}}(0) &= \frac{L'(0)(1 - D\cdot L(0)) + D\cdot L'(0)\cdot (L(0) - D)}{(1 - D\cdot L(0))^2} \\ &= \frac{L'(0)(1-D^2)}{(1 - D\cdot L(0))^2} = \sum_j jG_j \cdot \frac{1+D}{1-D}\cdot \sum_z zP_z. \quad \Box \end{aligned}$$

Since we only consider distributions G with geometrically decreasing tails, $M_G(t)$ is upper-bounded by

$$\begin{aligned} M_G(t) &\leq \sum_{j=0}^{U_G} c_G e^{tj} + \sum_{j>U_G} (1-\alpha_G)\alpha_G^{j-1} e^{tj} \\ &= c_G \frac{e^{t(U_G+1)} - 1}{e^t - 1} + \frac{(1-\alpha_G)e^t(\alpha_G e^t)^{U_G}}{1 - \alpha_G e^t}. \qquad (39) \end{aligned}$$

Therefore, the moment-generating function of G is finite in an interval around 0. It follows that $L(t)$ and hence $M_{\mathcal{P}}(t)$ are also finite in an interval around 0. Since $\sum_j jG_j = M_G(0)$, $\sum_j jG_j$ is finite for all permitted G. A standard application of the Chernoff bounds in Theorems 4.2 and 4.3 of [10] shows that for $\epsilon = N^{-1/3}$, the received sequence consists of $N\cdot \sum_j jG_j \cdot (1\pm\epsilon)$ bits, with probability at least $1 - e^{-\Theta(N^{1/3})}$. Hence, for $\delta = N^{-1/3}$ and

$$\mathcal{B} = \frac{N\sum_j jG_j}{\sum_k k\mathcal{P}_k}$$

Proposition 1 in [4] guarantees that the number of blocks in the received sequence is $\mathcal{B}(1\pm\epsilon)(1\pm\delta)$ with probability at least $1 - e^{-\Theta(N^{1/3})}$.

As mentioned in Section III-A, for distributions $P \in \mathbb{P}$, the longest type in X consists of at most $\gamma_3(\log N)^4$ bits with probability at least $1 - N^{-\Omega(\log N)}$. Unlike the i.i.d. deletion channel, where a received block is always shorter than the type it arises from, in an i.i.d. channel with effective distribution G a received block may be longer than the type from which it arises. Therefore, for $\mathbb{T}$ as in Section III-A (i.e., $t \in \mathbb{T}$ if and only if $|t| \leq \log^{2/3} N$), the set $\mathbb{K}$ of possible block lengths arising from types in $\mathbb{T}$ is now a superset of $\{1,\ldots,\log^{2/3} N\}$. We argue that this increase does not affect in any critical way the analysis in Section III. Consequently, the analysis in Section III holds for i.i.d. channels with effective distribution G, with the sole change of replacing $1-d$ (which is the average of G for the deletion channel) with $\sum_j jG_j$. We conclude the section with the analog of Theorem 3 for channels with i.i.d. deletions and duplications.

We first upper-bound the size of the set $\mathbb{K}$ in this context. To upper-bound the length of the longest block in Y, we reason as follows. Consider the longest type in X. At any individual bit of the type, the effective distribution introduces more than $\log^2 N$ copies with probability at most $N^{-\Omega(\log N)}$. By a union bound, the probability that there is at least one bit of the type where more than $\log^2 N$ bits are introduced is at most $\gamma_3(\log^4 N)(N^{-\Omega(\log N)}) = N^{-\Omega(\log N)}$. Therefore, with probability at least $1 - N^{-\Omega(\log N)}$, there is no insertion of more than $\log^2 N$ copies at any individual bit of the type. This implies that, conditioned on the longest type having length at most $\gamma_3 \log^4 N$, every block in Y has length at most $\gamma_3 \log^6 N$ with probability at least $1 - N^{-\Omega(\log N)}$. It follows that the latter is also the unconditional probability of a block in Y having length at most $\gamma_3 \log^6 N$, hence $|\mathbb{K}| \leq \gamma_3 \log^6 N$.

The increase in $|\mathbb{K}|$ only affects arguments where the quantity $|\mathbb{T}\times\mathbb{K}|$ appears, as well as the first line of (18) and (25). It is easy to check that these changes affect only $o(1)$ terms in our arguments, and the remaining analysis holds unchanged.

We therefore obtain the following theorem, following the same argument as Theorem 2.

Theorem 4: Consider an i.i.d. channel with effective distribution G such that $0 < G_0 < 1$, and a binary input alphabet. The capacity of this channel in bits is lower bounded by

$$C_f \geq \sup_{P\in\mathbb{P}} \frac{1}{\frac{1+D}{1-D}\cdot\sum_z zP_z} \cdot \left[\frac{1+D}{1-D}\cdot H(P) - (H(T,K) - H(\mathcal{P}))\right]$$

for $d = G_0$ and $D = \sum_z P_z d^z$.

Following the proof of Lemma 1, we also obtain

$$H(T,K) = \frac{1+D}{1-D}\cdot H(P) - \sum_k \sum_F \sum_{t\in F} \mathbf{Pr}[T = t, K = k] \cdot \log\left[d^s \cdot (\rho_{z+r,k} - \rho_{r,k}\cdot d^z)\right].$$

Here F is shorthand for $F(i,z,r,s)$ as usual. Therefore, we obtain the following corollary.

Corollary 3: Consider an i.i.d. channel with effective distribution G such that $0 < G_0 < 1$, and a binary input alphabet. The capacity of this channel in bits is lower-bounded by

$$C_f \geq \sup_{P\in\mathbb{P}} \frac{1}{\frac{1+D}{1-D}\cdot\sum_z zP_z} \cdot \left[H(\mathcal{P}) + \sum_k\sum_F\sum_{t\in F} \mathbf{Pr}[T = t, K = k] \cdot \log\left[d^s\cdot(\rho_{z+r,k} - \rho_{r,k}\cdot d^z)\right]\right]$$

for $d = G_0$ and $D = \sum_z P_z d^z$, and F standing for $F(i,z,r,s)$.

VIII. I.I.D. Channels With Duplications

In the special case of channels that only duplicate bits ($G_0 = d = 0$), the type of a block in the received sequence Y is simply the length of the block in the original codeword from which it arose. Then the equivalent of (3) for such channels is $\mathbf{Pr}[T=t] = \mathbf{Pr}[Z=z]$. This implies that $H(T) = H(P)$ and for $b \geq 0$

$$\rho_{a,a+b} = \sum_{\substack{1\leq i_1 \leq b+1\\ \vdots \\ 1\leq i_a \leq b+1 \\ \sum_{j=1}^{a} i_j = a+b}} G_{i_1}\cdot G_{i_2}\cdots G_{i_a}. \tag{40}$$

Hence

$$\mathbf{Pr}[T=t, K=k] = \mathbf{Pr}[Z=z, K=k] = P_z \cdot \rho_{z,k} \tag{41}$$

and

$$\mathcal{P}_k = \sum_{z=1}^{\infty} P_z \cdot \rho_{z,k}. \tag{42}$$

Note that for i.i.d. insertion channels which introduce copies only, the number of blocks in the received sequence exactly equals the number of blocks in the codeword (no blocks disappear and no new blocks are generated). From (41) and the proof of Lemma 1, we easily conclude that

$$H(Z,K) = H(P) + \sum_z \sum_k \mathbf{Pr}[Z=z, K=k] \cdot \log \rho_{z,k}.$$

Therefore, we obtain the following corollary to Theorem 4.

Corollary 4: Consider an i.i.d. channel with effective distribution G such that $G_0 = 0$, and a binary input alphabet. The capacity of this channel in bits is lower-bounded by

$$C_{\text{ins}} \geq \sup_{P\in\mathbb{P}} \frac{1}{\sum_z zP_z} \cdot \left[H(\mathcal{P}) + \sum_z \sum_k \mathbf{Pr}[Z=z,K=k]\cdot \log \rho_{z,k} \right]$$

for $\rho_{a,b}$ given by (40), $\mathcal{P}$ given by (42), and $\mathbf{Pr}[Z=z, K=k]$ given by (41).

Finally, it is easy to follow Lemma 6 and show that $M_{\mathcal{P}}(t) = L(t)$. Therefore, the average block length in the received sequence given by $M'_{\mathcal{P}}(0) = L'(0)$ is

$$\sum_k k\mathcal{P}_k = \sum_j jG_j \cdot \sum_z zP_z. \tag{43}$$

A. Elementary i.i.d. Duplication Channel

In this section, we apply Corollary 4 to lower-bound the capacity C_{dupl} of the elementary i.i.d. duplication channel when $P \in \mathbb{P}$.

The effective distribution of the elementary i.i.d. duplication channel is $G_0 = d = 0$, $G_1 = 1-v$, $G_2 = v$, $G_j = 0$ for $j > 2$; hence, $\sum_j jG_j = 1+v$. Then the probability that a block of length a generates a block of length $a+b \leq 2a$ from (40) becomes

$$\begin{aligned}\rho_{a,a+b} &= \sum_{\substack{1\leq i_1 \leq 2\\ \vdots \\ 1\leq i_a \leq 2 \\ \sum_{j=1}^{a} i_j = a+b}} G_{i_1}\cdot G_{i_2}\cdots G_{i_a} \\ &= \binom{a}{b} \cdot v^b \cdot (1-v)^{a-b}.\end{aligned}$$

Therefore, the joint probability of a block in Y having length k and arising from type z from (41) becomes

$$\begin{aligned}\mathbf{Pr}[Z=z,K=k] &= P_z \cdot \binom{z}{k-z} \cdot v^{k-z} \cdot (1-v)^{2z-k} \\ &= \left(\frac{v}{1-v}\right)^k \cdot P_z \cdot \binom{z}{k-z} \\ &\quad \cdot \left(\frac{(1-v)^2}{v}\right)^z. \end{aligned} \tag{44}$$

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

$$\mathcal{P}_k = \left(\frac{v}{1-v}\right)^k \cdot \sum_{z=1}^{\lfloor \frac{k}{2} \rfloor} P_z \cdot \binom{z}{k-z} \cdot \left(\frac{(1-v)^2}{v}\right)^z. \tag{45}$$

By (43), for $D = 0$ and $\sum_j jG_j = 1+v$, the average block length in the received sequence is given by

$$\sum_{k\geq 1} k\mathcal{P}_k = (1+v)\cdot \sum_{z\geq 1} zP_z. \tag{46}$$

The following lemma simplifies the quantity inside the summation of Corollary 4 (the proof appears in the Appendix).

Lemma 7: Consider the elementary duplication channel with duplication probability v. Then

$$\begin{aligned}\sum_z \sum_k & \mathbf{Pr}[Z=z,K=k]\cdot \log \rho_{z,k} \\ &= -\sum_z \sum_k \mathbf{Pr}[Z=z,K=k] \cdot \log \binom{z}{k-z} \\ &\quad + H(v) \cdot \sum_z zP_z. \end{aligned} \tag{47}$$

Substituting into Corollary 4, we obtain the following theorem for the capacity of the elementary i.i.d. duplication channel.

Theorem 5: Consider a channel that duplicates every transmitted bit independently and with probability v and a binary input alphabet. The capacity of this channel in bits is lower bounded by

$$\begin{aligned}C_{\text{dupl}} \geq \sup_{P\in\mathbb{P}} \Bigg[\frac{1}{\sum_j jP_j} \Bigg(& H(\mathcal{P}) + \\ & + \sum_k \sum_z \mathbf{Pr}[Z=z,K=k] \cdot \log \binom{z}{k-z} \Bigg) \\ & - H(v) \Bigg] \end{aligned} \tag{48}$$

for $\mathbf{Pr}[Z=z,K=k]$ given by (44) and $\mathcal{P}$ given by (45).

Although it might be difficult to derive a closed formula for the summation in (47) for arbitrary distributions P, one can easily compute it numerically for fixed P and v over a limited range of z and k. Similarly to the i.i.d. deletion channel, summing over a finite number of these terms yields strict underestimates of the actual lower bounds derivable by our approach.

We use (48) to derive a lower bound for the capacity C_{dupl} of the elementary i.i.d. duplication channel in the special case where the block lengths in X are geometrically distributed, i.e., $P_j = (1-p)p^{j-1}$. Since the probability that the type of a block in Y is z is given by $\mathbf{Pr}[Z = z] = (1-p)\cdot p^{z-1}$, the joint probability of type z and length k from (44) becomes

$$\begin{aligned}\mathbf{Pr}[Z=z, K=k] &= \mathbf{Pr}[Z=z]\cdot\mathbf{Pr}[K=k \mid Z=z]\\ &= (1-p)\cdot p^{z-1}\cdot\binom{z}{k-z}\\ &\quad\cdot v^{k-z}\cdot(1-v)^{2z-k}\\ &= \frac{1-p}{p}\left(\frac{v}{1-v}\right)^{k}\binom{z}{k-z}\\ &\quad\cdot\left(\frac{p(1-v)^2}{v}\right)^{z}. \qquad (49)\end{aligned}$$

The following combinatorial lemma provides a closed form for $\mathcal{P}$ (the proof appears in the Appendix).

Lemma 8: Let $A = \frac{p\cdot(1-v)^2}{v}$. Then for all $k \geq 1$

$$\begin{aligned}\mathcal{P}_k = &\frac{1-p}{p\sqrt{1+4/A}}\cdot(p(1-v))^{k}\\ &\cdot\left[\left(\frac{1+\sqrt{1+4/A}}{2}\right)^{k+1} - \left(\frac{1-\sqrt{1+4/A}}{2}\right)^{k+1}\right].\end{aligned}$$

By (46), the average block length in Y equals $\frac{1+v}{1-p}$. Also, $H(\mathcal{P})$ is given by

$$\begin{aligned}H(\mathcal{P}) &= -\sum_{k=1}^{\infty}\mathcal{P}_k\cdot\log\mathcal{P}_k\\ &= -\log\frac{1-p}{p\sqrt{1+4/A}}\\ &\quad-\frac{1+v}{1-p}\left(\log p(1-v) + \log\frac{1+\sqrt{1+4/A}}{2}\right)\\ &\quad-\log\frac{1+\sqrt{1+4/A}}{2}\\ &\quad-\sum_{k=1}^{\infty}\mathcal{P}_k\log\left(1-\left(\frac{1-\sqrt{1+4/A}}{1+\sqrt{1+4/A}}\right)^{k+1}\right). \qquad (50)\end{aligned}$$

Each term inside the infinite summation above is negative, therefore, its contribution to (50) is positive. Since $H(\mathcal{P})$ appears unnegated in (48), summing over a finite range of k strictly underestimates our final derived bounds. Finally, $\sum_j jP_j = \frac{1}{1-p}$ and we obtain the following corollary to Theorem 5.

Corollary 5: Consider a channel that duplicates every transmitted bit independently and with probability $0 < v < 1$, a binary input alphabet and geometric block length distribution P. The capacity of this channel in bits is lower bounded by

$$\begin{aligned}C_{\text{dupl}} \geq \sup_{0<p<1}&\Bigg[(1-p)\Bigg(H(\mathcal{P})\\ &+\sum_{z}\sum_{k}\mathbf{Pr}[Z=z,K=k]\cdot\log\binom{z}{k-z}\Bigg)\\ &-H(v)\Bigg],\end{aligned}$$

where $H(\mathcal{P})$ is given in (50) and $\mathbf{Pr}[Z=z, K=k]$ is given in (49).

B. I.I.D. Geometric Duplication Channel, With Geometric P

In this subsection, we use Corollary 4 to derive a lower bound for the capacity $C_{\text{g.d.}}$ of the i.i.d. geometric duplication channel in the special case where the block lengths in X are geometrically distributed, i.e., $P_j = (1-p)p^{j-1}$.

Consider a binary i.i.d. channel with the following effective distribution: $G_0 = d = 0$ and $G_j = (1-v)\cdot v^{j-1}$ for $j \geq 1$ (we remind that for $j \geq 1$, G_j is the probability that the channel transmits the original bit together with $j-1$ new copies of the bit). Then the probability that a block of length a generates a block of length $a+b \geq a$ from (40) becomes

$$\begin{aligned}\rho_{a,a+b} &= \sum_{\substack{1\leq i_1\leq b+1\\ \vdots\\ 1\leq i_a\leq b+1\\ \sum_{j=1}^{a} i_j = a+b}} G_{i_1}G_{i_2}\cdots G_{i_a}\\ &= \sum_{\substack{0\leq i_1\leq b\\ \vdots\\ 0\leq i_a\leq b\\ \sum_{j=1}^{a} i_j = b}} (1-v)v^{i_1}(1-v)v^{i_2}\cdots(1-v)v^{i_a}\\ &= (1-v)^a\cdot v^b\cdot\binom{a+b-1}{b}.\end{aligned}$$

Therefore, the joint probability of a block of length z in X giving rise to a block of length $k \geq z$ in Y is

$$\begin{aligned}\mathbf{Pr}[T=t, K=k] &= P_z\cdot\rho_{z,k}\\ &= P_z\cdot(1-v)^z v^{k-z}\binom{k-1}{k-z}. \qquad (51)\end{aligned}$$

If P is geometrically distributed, i.e., $P_j = (1-p)p^{j-1}$ for all $j \geq 1$, then for $q = p + v - pv < 1$, we obtain

$$\begin{aligned}\mathcal{P}_k &= \frac{1-p}{p}\cdot v^k\cdot\sum_{z=1}^{\infty}\binom{k-1}{k-z}\left(\frac{1-v}{v}\right)^{z}\\ &= (1-p)(1-v)\cdot(v+p-pv)^{k-1}\\ &= (1-q)\cdot q^{k-1}.\end{aligned}$$

This immediately yields that the entropy $H(\mathcal{P})$ is $\frac{H(q)}{1-q}$ and that the average block length in Y in $\frac{1}{1-q}$. Also, since

$$\sum_z zP_z = \frac{1}{1-p}$$

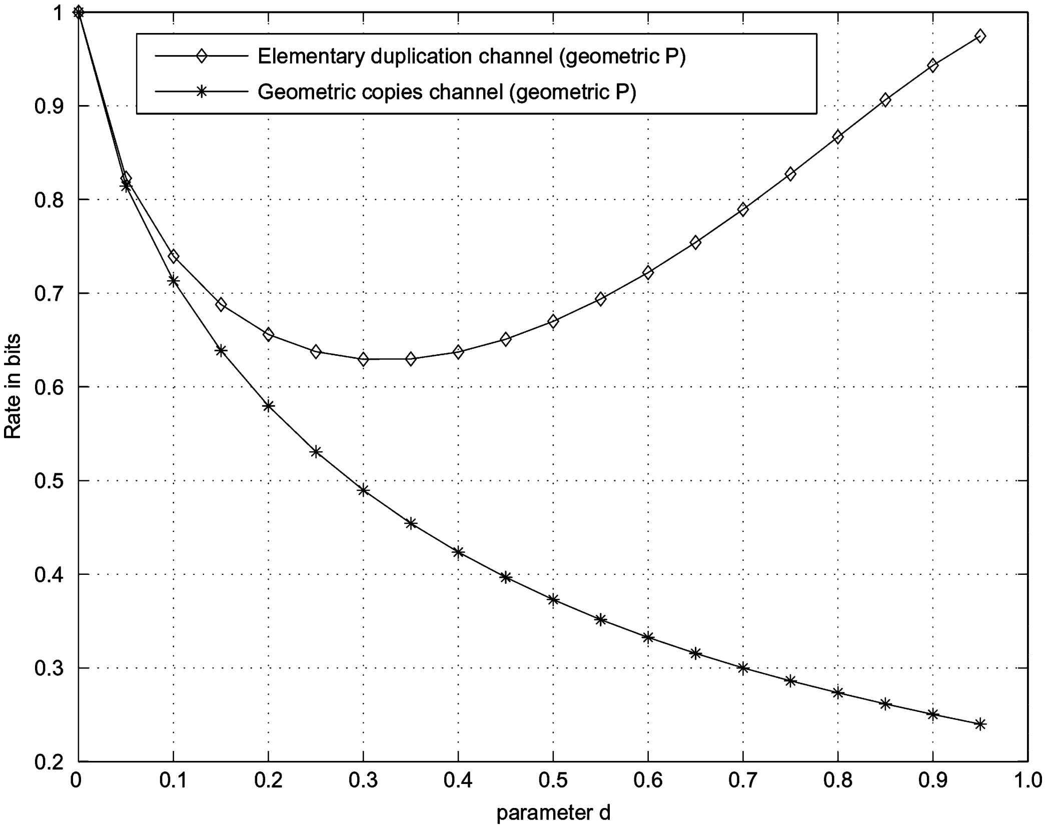

Fig. 3. Rates for the elementary i.i.d. duplication channel and the i.i.d. geometric duplication channel (Sections VIII-A and VIII-B) for codewords with geometrically distributed block lengths.

it is easy to show that

$$\begin{aligned}\sum_z \sum_k &\mathbf{Pr}[Z=z, K=k]\cdot \log \rho_{z,k}\\ &= \frac{\log(1-v)}{1-p} - \frac{\log v}{1-p} + \frac{\log v}{(1-p)(1-v)}\\ &\quad + \sum_{k,z} \mathbf{Pr}[Z=z,K=k]\cdot \log\binom{k-1}{k-z}\\ &= -\frac{H(v)}{(1-p)(1-v)} + \sum_{k,z}\mathbf{Pr}[Z=z,K=k]\\ &\quad \cdot \log\binom{k-1}{k-z}.\end{aligned}$$

Hence, by Corollary 4, the capacity of this channel is lower-bounded by the following theorem.

Theorem 6: Consider a channel that independently at every transmitted bit inserts a nonnegative integer number of copies distributed according to a geometric distribution with parameter v, and a binary input alphabet. The capacity of this channel in bits is lower-bounded by

$$C_{\text{g.d.}} \geq \sup_{P\in\mathbb{P}} \left[(1-p)\sum_k \sum_z \mathbf{Pr}[Z=z,K=k] \cdot \log\binom{k-1}{k-z} + \frac{H(q)-H(v)}{1-v}\right]$$

for $\mathbf{Pr}[Z=z,K=k]$ given in (51) and $1-q=(1-p)(1-v)$.

C. Discussion of Our Results

Fig. 3 presents an underestimate of the rates for the elementary i.i.d. duplication channel and the i.i.d. geometric duplication channel derived in Sections VIII-A and VIII-B, respectively. The graphs are underestimates because we optimized over only two decimal digits for the parameter p, and our numerical calculations were over a limited range of z and k. Again, we performed extensive simulations for codewords with geometrically distributed block lengths and $N = 2.5\cdot 10^{10}$. The simulations verified the convergence of $H(\mathcal{P})$ and $H(Z,K)$ to the values predicted by the theory, giving us confidence in the results of Fig. 3.

Notice that as one would expect, for the i.i.d. duplication channel the capacity approaches 1 as the probability that every bit is duplicated approaches 1.

IX. Conclusion

We have presented new lower bounds for the capacity of the i.i.d. binary deletion channel, improving on previous analysis by using a stronger definition of a typical output. We focused on channels with binary alphabets although our results generalize in a natural way. We also presented a general lower bound for binary channels with i.i.d. deletions and duplications, as well as lower bounds for binary channels with i.i.d. copies only.

In an earlier version of this paper we suggested exploring stronger notions of typical outputs to achieve a better bound. In particular, considering types for several consecutive blocks in the received sequence at a time, instead of just one, would attempt "block decoding" in the standard information-theoretic sense. Recently, Drinea and Kirsch [3] characterized the improvement induced by block decoding when considering *any* number of consecutive blocks in the received sequence at a time by analyzing the *information* capacity of the deletion channel (instead of the *operational* capacity addressed in this paper). Their work basically expands on the ideas from renewal theory presented here, and gives clean theoretical results by using elementary facts from renewal theory instead of combinatorial arguments. Open directions to advance this work include the following.

- There may be better ways of selecting codewords, such as using different distributions. For example, the experimental bounds in [8] and [12] suggest that higher order Markov chains yield better bounds.
- Our approach could possibly be applied to channels with more general insertions and deletions, or other errors such as transpositions. The most natural next step would be to consider the insertion channel where a random bit can be inserted after each bit with some probability. While we expect our approach can be used for this channel, there are difficulties to consider. In particular, our work thus far to handle duplication makes strong use of the block structure of the input and output sequences. With insertions of random bits, this block structure would be lost at the output, and thus it appears there is additional work necessary to analyze this kind of channel.

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

Appendix

Lemma 1: Consider a binary i.i.d. deletion channel. The joint entropy $H(T,K)$ of the distribution of the types in X and the block lengths in Y is given by

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

Proof: For $\mathbf{Pr}[T=t,K=k]$ given by (6), we have

$$\begin{aligned} \log\mathbf{Pr}[T=t,K=k] = k\log\frac{1-d}{d} &+ (z+r+s)\log d \\ &+ \log\left(P_z P_{s_1} P_{r_1}\cdots P_{s_i}P_{r_i}\right) \\ &+ \log\left(\binom{z+r}{k} - \binom{r}{k}\right). \end{aligned}$$

The first two terms as well as the last one depend only on the family $F(i,z,r,s)$ where the type t belongs, and not on which specific member of the family is represented by t. On the other hand, the third term explicitly depends on the specific member of $F(i,z,r,s)$ represented by t. Therefore, we obtain for $H(T,K)$

$$\begin{aligned} H(T,K) &= -\sum_k\sum_t \mathbf{Pr}[T=t,K=k]\log\mathbf{Pr}[T=t,K=k] \\ &= -\log\frac{1-d}{d}\cdot\left(\sum_{k=1}^{\infty}\sum_{i=0}^{\infty}\sum_{z=1}^{\infty}\sum_{r=i}^{\infty}\sum_{s=i}^{\infty}\right. \\ &\qquad \left.\mathbf{Pr}[T\in F(i,z,r,s), K=k]\cdot k\right) \qquad (52) \\ &\quad -\log d\cdot\left(\sum_{k=1}^{\infty}\sum_{i=0}^{\infty}\sum_{z=1}^{\infty}\sum_{r=i}^{\infty}\sum_{s=i}^{\infty}(r+z+s)\right. \\ &\qquad \left.\cdot\,\mathbf{Pr}[T\in F(i,z,r,s), K=k]\right) \qquad (53) \\ &\quad -\left(\sum_{k=1}^{\infty}\sum_{i=0}^{\infty}\sum_{z=1}^{\infty}\sum_{\substack{r_1=1\\ r_2=1\\ \cdots\\ r_i=1}}^{\infty}\sum_{\substack{s_1=1\\ s_2=1\\ \cdots\\ s_i=1}}^{\infty}\mathbf{Pr}[T=t,K=k]\right. \\ &\qquad \left.\cdot\,\log(P_zP_{r_1}P_{s_1}\cdots P_{r_i}P_{s_i})\right) \qquad (54) \\ &\quad -\left(\sum_{k=1}^{\infty}\sum_{i=0}^{\infty}\sum_{z=1}^{\infty}\sum_{r=i}^{\infty}\sum_{s=i}^{\infty}\right. \\ &\quad \mathbf{Pr}[T\in F(i,z,r,s), K=k] \\ &\qquad \left.\cdot\log\left[\binom{r+z}{k}-\binom{r}{k}\right]\right). \qquad (55) \end{aligned}$$

We will show that for all distributions P, the sum of (52) and (53) equals $H(d)\cdot\frac{\sum_k k\mathcal{P}_k}{1-d}$, while the term (54) equals

$$\frac{H(P)}{\sum_j jP_j}\cdot\frac{\sum_k k\mathcal{P}_k}{1-d}.$$

First, we derive closed forms for two quantities that will provide useful in the following computations. For $m\geq 0$, let $X_m = \sum_{n=m}^{\infty} nQ_{n,m}$ and $Y_m = \sum_{n=m}^{\infty} nQ_{n,m}d^n$. Let $C = \sum_z zP_zd^z$ (hence, C stands for the expected length of a deleted block). Then

$$\begin{aligned} X_m = \sum_{n=m}^{\infty} nQ_{n,m} &= \sum_{n=m-1}^{\infty}\sum_{z=1}^{\infty}(n+z)\cdot Q_{n,m-1}\cdot P_z \\ &= \sum_{n=m-1}^{\infty} nQ_{n,m-1}\sum_{z=1}^{\infty}P_z + \sum_{n=m-1}^{\infty} Q_{n,m-1}\sum_{z=1}^{\infty} zP_z \\ &= X_{m-1} + \sum_{z=1}^{\infty} zP_z. \qquad (56) \end{aligned}$$

The solution to the above recurrence is $X_m = m \cdot \sum_{z=1}^{\infty} zP_z$. Similarly

$$\begin{aligned} Y_m &= \sum_{n=m-1}^{\infty} \sum_{z=1}^{\infty} (n+z) \cdot Q_{n,m-1} \cdot P_z \cdot d^{z+n} \\ &= \sum_{n=m-1}^{\infty} nQ_{n,m-1} d^n \sum_{z=1}^{\infty} P_z d^z \\ &\quad + \sum_{n=m-1}^{\infty} Q_{n,m-1} d^n \sum_{z=1}^{\infty} zP_z d^z \\ &= Y_{m-1} \cdot D + D^{m-1} \cdot C. \end{aligned} \tag{57}$$

The solution to the above recurrence is $Y_m = m \cdot D^{m-1} \cdot C$. Now let

$$A = \sum_k \sum_{i,z,r,s} (r+z+s) \cdot \mathbf{Pr}[T \in F(i,z,r,s), K = k]$$

where $\mathbf{Pr}[T \in F(i,z,r,s), K = k]$ is given by (7). Then A is simplified as follows:

$$\begin{aligned} A &= \sum_{i=0}^{\infty} \sum_{z=1}^{\infty} \sum_{r=i}^{\infty} \sum_{s=i}^{\infty} (r+z+s) \cdot P_z Q_{r,i} Q_{s,i} \\ &\quad \cdot d^{z+r+s} \cdot \sum_{k=1}^{\infty} \left(\frac{1-d}{d} \right)^k \left[\binom{z+r}{k} - \binom{r}{k} \right] \\ &= \sum_{i=0}^{\infty} \sum_{z=1}^{\infty} \sum_{r=i}^{\infty} \sum_{s=i}^{\infty} (r+z+s) \cdot P_z Q_{r,i} Q_{s,i} \\ &\quad \cdot d^{z+r+s} \cdot [d^{-r-z} - d^{-r}] \\ &= \sum_{i=0}^{\infty} \sum_{z=1}^{\infty} \sum_{r=i}^{\infty} \sum_{s=i}^{\infty} P_z Q_{r,i} \cdot Q_{s,i} d^s \cdot [(r+z) - rd^z] \\ &\quad + \sum_{i=0}^{\infty} \sum_{z=1}^{\infty} \sum_{r=i}^{\infty} \sum_{s=i}^{\infty} P_z Q_{r,i} \cdot sQ_{s,i} d^s \cdot (1-d^z) \\ &\quad - \sum_{i=0}^{\infty} \sum_{z=1}^{\infty} \sum_{r=i}^{\infty} \sum_{s=i}^{\infty} Q_{r,i} \cdot Q_{s,i} d^s \cdot zP_z d^z \\ &= \sum_{i=0}^{\infty} \sum_{z=1}^{\infty} \sum_{r=i}^{\infty} \sum_{s=i}^{\infty} P_z Q_{r,i} Q_{s,i} \cdot d^s \cdot [(r+z) - rd^z] \\ &\quad + \sum_{i=0}^{\infty} i \cdot D^{i-1} \cdot C \cdot (1-D) - \sum_{i=0}^{\infty} D^i \cdot C \\ &= \sum_{i=0}^{\infty} \sum_{z=1}^{\infty} \sum_{r=i}^{\infty} \sum_{s=i}^{\infty} P_z Q_{r,i} Q_{s,i} \cdot d^s \cdot [(r+z) - rd^z] \\ &\quad + \frac{C}{(1-D)^2} \cdot (1-D) - \frac{1}{1-D} \cdot C \\ &= \sum_{i=0}^{\infty} D^i \left(\sum_{r=i+1}^{\infty} rQ_{r,i+1} - D \cdot \sum_{r=i}^{\infty} rQ_{r,i} \right) \\ &= \sum_{i=0}^{\infty} D^i (X_{i+1} - D \cdot X_i) \\ &= \frac{1+D}{1-D} \cdot \sum_z zP_z = \frac{\sum_k k\mathcal{P}_k}{1-d}. \end{aligned}$$

The last equality follows from Lemma 1 in [4] giving the average block length in the received sequence after the i.i.d. deletion channel as $\sum_k k\mathcal{P}_k = (1-d)\frac{1+D}{1-D} \sum_z zP_z$.

Since the term (52) equals $\log \frac{1-d}{d} \cdot \sum_k k\mathcal{P}_k$, we conclude that the sum of the terms (52) and (53) equals

$$-\log \frac{1-d}{d} \cdot \sum_k k\mathcal{P}_k - \log d \frac{\sum_k k\mathcal{P}_k}{1-d} = \frac{\sum_k k\mathcal{P}_k}{1-d} \cdot H(d). \tag{58}$$

Now consider (54) and let

$$B = \sum_{k,i,z,r_1,s_1,\dots,r_i,s_i} \mathbf{Pr}[T=t, K=k] \cdot \log\left(P_z P_{s_1} P_{r_1} \cdots P_{s_i} P_{r_i}\right)$$

with $\mathbf{Pr}[T = t, K = k]$ given by (6). Summing first over k, we obtain that

$$d^{r+z} \sum_k \left[\binom{r+z}{k} - \binom{r}{k} \right] \left(\frac{1-d}{d} \right)^k = 1 - d^z.$$

Then after some rearranging of the terms, B becomes

$$\begin{aligned} B &= \sum_{i=0}^{\infty} \sum_{s=i}^{\infty} Q_{s,i} d^s \cdot \sum_{r=i}^{\infty} Q_{r,i} \cdot \sum_{z=1}^{\infty} P_z (1-d^z) \cdot \log P_z \\ &\quad + \sum_{i=0}^{\infty} \sum_{s=i}^{\infty} Q_{s,i} d^s \cdot \sum_{z=1}^{\infty} P_z (1-d^z) \cdot i \\ &\quad \cdot \sum_{r=i-1}^{\infty} Q_{r,i-1} \sum_{r_1=1}^{\infty} P_{r_1} \log P_{r_1} \\ &\quad + \sum_{i=0}^{\infty} \sum_{r=i}^{\infty} Q_{r,i} \cdot \sum_{z=1}^{\infty} P_z (1-d^z) \cdot i \\ &\quad \cdot \sum_{s=i-1}^{\infty} Q_{s,i-1} d^s \sum_{s_1=1}^{\infty} P_{s_1} d^{s_1} \log P_{s_1} \\ &= \frac{1}{1-D} \cdot (H(P) - \sum_{z=1}^{\infty} P_z d^z \log P_z) \\ &\quad + \sum_i D^i \cdot (1-D) \cdot i \cdot H(P) \\ &\quad + \sum_{i=0}^{\infty} (1-D) i \cdot D^{i-1} \cdot \sum_{z=1}^{\infty} P_z d^z \log P_z \\ &= \frac{1+D}{1-D} \cdot H(P) = \frac{\sum_k k\mathcal{P}_k}{(1-d)\sum_j jP_j} \cdot H(P) \end{aligned}$$

where the last equality again follows from the formula for the average block length in the received sequence. The proof is complete. □

We now derive a formula for the entropy $H(T)$ of the distribution of types.

Lemma 9: The entropy $H(T)$ of the distribution of the types in X is given by

$$\begin{aligned} H(T) = \frac{1+D}{1-D} \cdot H(P) - \frac{1}{1-D} \cdot \sum_z P_z \\ \cdot \left(d^z \cdot \log d^z + (1-d^z) \cdot \log(1-d^z)\right). \end{aligned} \tag{59}$$

Proof: By (8) and the proof of Lemma 1, we have

$$\begin{aligned} H(T) &= -\sum_{i,z,r,s}\sum_{t\in F(i,z,r,s)} \mathbf{Pr}[T=t]\cdot\log\mathbf{Pr}[T=t] \\ &= -\log d\cdot\sum_{i=0}^{\infty}\sum_{z=1}^{\infty}P_z(1-d^z)\sum_{r=i}^{\infty}Q_{r,i}\sum_{s=i}^{\infty}sQ_{s,i}d^s \qquad (60) \\ &\quad -\sum_{i=0}^{\infty}\sum_{z=1}^{\infty}\cdot\sum_{\substack{r_1=1\\ r_2=1\\ \cdots\\ r_i=1}}^{\infty}\sum_{\substack{s_1=1\\ s_2=1\\ \cdots\\ s_i=1}}^{\infty} P_z(1-d^z)P_{r_1}\cdots P_{s_i}d^s \\ &\quad \cdot\log(P_zP_{r_1}P_{s_1}\cdots P_{r_1}P_{s_i}) \qquad (61) \\ &\quad -\sum_{i=0}^{\infty}\sum_{z=1}^{\infty}P_z(1-d^z)\log(1-d^z) \\ &\quad \cdot\sum_{r=i}^{\infty}Q_{r,i}\sum_{s=i}^{\infty}Q_{s,i}\cdot d^s \qquad (62) \end{aligned}$$

We compute (60), (61), and (62) separately. First, for $C = \sum_z zP_zd^z$ and $\sum_{s=i}^{\infty} sQ_{s,i}d^s$ given by (57), (60) becomes

$$-\log d\cdot\sum_{i=0}^{\infty}(1-D)\cdot i\cdot D^{i-1}\cdot C = -\log d\cdot\frac{C}{1-D}.$$

Also, (61) coincides with (54), and therefore equals $\frac{1+D}{1-D}\cdot H(P)$. Finally, (62) is simplified to

$$-\frac{1}{1-D}\cdot\sum_{z=1}^{\infty}P_z(1-d^z)\log(1-d^z). \qquad \square$$

We immediately obtain the following corollary for geometrically distributed codewords.

Corollary 6: When the blocks in X are geometrically distributed with parameter p, the entropy $H(T)$ of the distribution of the types is given by

$$\begin{aligned} H(T) = \frac{H(p)}{(1-d)(1-q)} &- \frac{d(1-p)\log d}{(1-d)(1-pd)} \\ &+\frac{(1-p)(1-pd)}{p(1-d)}\cdot\sum_{z=1}^{\infty}p^z(1-d^z)\log(1-d^z). \end{aligned}$$

Lemma 7: Consider the elementary duplication channel with duplication probability v. Then

$$\begin{aligned} &\sum_z\sum_k \mathbf{Pr}[Z=z,K=k]\cdot\log\rho_{z,k} \\ &= -\sum_z\sum_k \mathbf{Pr}[Z=z,K=k]\cdot\log\binom{z}{k-z} \\ &\quad + H(v)\cdot\sum_z zP_z. \end{aligned}$$

Proof: By (42), we have

$$\begin{aligned} &\sum_z\sum_k \mathbf{Pr}[Z=z,K=k]\cdot\log\rho_{z,k} \\ &= -\sum_{k=1}^{\infty}\sum_{z=1}^{\infty}\mathbf{Pr}[z,k]\cdot\log\binom{z}{k-z} \\ &\quad -\sum_{k=1}^{\infty}\sum_{z=1}^{\infty}\mathbf{Pr}[z,k]\cdot k\log\frac{v}{1-v} \qquad (63) \\ &\quad -\sum_{k=1}^{\infty}\sum_{z=1}^{\infty}\mathbf{Pr}[z,k]\cdot z\log\frac{(1-v)^2}{v} \qquad (64) \end{aligned}$$

Let

$$\begin{aligned} A = \sum_k\sum_z z\cdot\mathbf{Pr}[z,k] &= \sum_z z\cdot\sum_k\mathbf{Pr}[z,k] \\ &= \sum_{z=1}^{\infty}zP_z. \end{aligned}$$

Since the term (63) equals $-\log\frac{v}{1-v}\cdot\sum_k k\mathcal{P}_k$, by (46), we conclude that the sum with the term (64) simplifies to

$$\begin{aligned} -\log\frac{v}{1-v}\cdot(1+v)\cdot\sum_z zP_z - \log\frac{(1-v)^2}{v}\cdot\sum_z zP_z \\ = H(v)\cdot\sum_z zP_z \end{aligned}$$

The proof is complete. $\square$

Lemma 8: Let $A = \frac{p\cdot(1-v)^2}{v}$. Then for all $k\geq 1$, the block length distribution in the sequence received after the i.i.d. elementary duplication channel with duplication probability v is given by

$$\begin{aligned} \mathcal{P}_k = \frac{1-p}{p\sqrt{1+4/A}}\cdot(p(1-v))^k\cdot&\left[\left(\frac{1+\sqrt{1+4/A}}{2}\right)^{k+1}\right. \\ &\left. -\left(\frac{1-\sqrt{1+4/A}}{2}\right)^{k+1}\right]. \end{aligned}$$

Proof: Let $T_k = \sum_{z=0}^{k}\binom{z}{k-z}\cdot A^z$, for $k\geq 0$. Then

$$\begin{aligned} T_k &= \sum_{z=0}^{k}\binom{z}{k-z}A^z = \sum_{z=0}^{k}\binom{k-z}{z}A^{k-z} \\ &= A^k\cdot\sum_{z=0}^{k}\binom{k-z}{z}A^{-z} \\ &= \frac{A^k}{\sqrt{1+4/A}}\cdot\left[\left(\frac{1+\sqrt{1+4/A}}{2}\right)^{k+1}\right. \\ &\quad\left. -\left(\frac{1-\sqrt{1+4/A}}{2}\right)^{k+1}\right] \qquad (65) \end{aligned}$$

where the last equality follows from standard texts, e.g., [7, p. 204] (otherwise, it is easy to observe that $T_k = A\cdot(T_{k-1}+T_{k-2})$, with $T_0 = 1$ and $T_1 = A$; the solution to this recurrence is given by (65)). The lemma follows. $\square$

References

[1] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley , 1991.

[2] S. Diggavi and M. Grossglauser, *IEEE Trans. Inf. Theory*, vol. 52, no. 3, pp. 1226–1237, Mar. 2006.

[3] E. Drinea and A. Kirsch, “Directly lower bounding the information capacity for channels with i.i.d. deletions and duplications,” in *Proc. IEEE Int. Symp. Information Theory*, Nice, France, Jun. 2007, pp. 1731–1735.

[4] E. Drinea and M. Mitzenmacher, "On lower bounds for the capacity of deletion channels," *IEEE Trans. Inf. Theory*, vol. 52, no. 10, pp. 4648–4657, Oct. 2006.

[5] 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. Inform.*, vol. 3, no. 4, pp 18-36, 1967.

[6] 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, April–June, 1990.

[7] R. L. Graham, D. E. Knuth, and O. Patashnik*, Concrete Mathematics*. Reading, MA: Addison-Wesley , 1988.

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

[9] M. Mitzenmacher and E. Drinea, "A simple lower bound for the capacity of the deletion channel," *IEEE Trans. Inf. Theory*, vol. 52, no. 10, pp. 4657–4660, Oct. 2007.

[10] R. Motwani and P. Raghavan*, Randomized Algorithms*. Cambridge, U.K.: Cambridge Univ. Press, 1995.

[11] 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.

[12] 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, no. 3, pp 92-95, 1968.