An Improvement of the Deletion Channel Capacity Upper Bound

Mojtaba Rahmati* and Tolga M. Duman*†

*Arizona State University, School of Electrical, Computer and Energy Engineering, Tempe, AZ 85287-5706
†Bilkent University, Dept. of Electrical and Electronics Engineering, TR-06800, Bilkent, Ankara, Turkey
Emails: mrahmati@asu.edu, duman@ee.bilkent.edu.tr

Abstract— In this paper, we offer an alternative look at channels with deletion errors by considering equivalent models for deletion channels by "fragmenting" the input sequence where different subsequences travel through different channels. The resulting output symbols are combined appropriately to come up with an equivalent input-output representation of the original channel which allows for derivation of new upper bounds on the channel capacity. Considering a random fragmentation processes applied to binary deletion channels, we prove an inequality relation among the capacities of the original binary deletion channel and the introduced binary deletion subchannels. This inequality enables us to provide an improved upper bound on the capacity of the i.i.d. deletion channels, i.e., $C(d) \leq 0.4143(1 - d)$ for $d \geq 0.65$. We also consider a deterministic fragmentation process suitable for the study of non-binary deletion channels which results in improved capacity upper bounds.

I. INTRODUCTION

Channels with synchronization errors can be well modeled using symbol drop-outs and/or symbol insertions as well as random errors. There are many models adopted in the literature to describe the resulting channels in different applications. In [1], memoryless channels with synchronization errors are described by a channel matrix allowing for the channel outputs to be of different lengths for different uses of the channel. As proved in the same paper, for such channels, information stability holds and Shannon capacity exists. However, the determination of the capacity remains elusive as the mutual information term to be maximized does not admit a single letter or finite letter form.

In the existing literature, several specific instances of this model are more widely studied, e.g., [2]–[8]. For instance, by a proper selection of the stochastic channel transition matrix, one obtains the i.i.d. deletion channel which represents one of the simplest models allowing for symbol drop-outs which is also the main focus of this paper. In an i.i.d. deletion channel, the transmitted symbols are either received correctly and in the right order, or they are deleted from the transmitted

sequence altogether with a certain probability d independent of each other. Neither the receiver nor the transmitter knows the positions of the deleted symbols. Despite the simplicity of the model, the capacity for this channel is still unknown and only a few upper and lower bounds are available in the literature [2]–[8].

In this paper, for both binary and non-binary input i.i.d. deletion channels, it is observed that if we define a new channel in which the input sequence is fragmented into subsequences of smaller lengths where the resulting subsequences travel through independent i.i.d. deletion channels and the surviving symbols of the deletion channels are combined without changing their order in the original input sequence, then the resulting channel is an i.i.d. deletion channel with parameters which depend on the parameters of the considered subchannels.

This new formulation enables us to prove that the capacity of an i.i.d. binary deletion channel with deletion probability d can be upper bounded in terms of the capacities of i.i.d. binary deletion channels with deletion probabilities d_1 and d_2 where d is a weighted average of d_1 and d_2, i.e., $d = \lambda d_1 + (1 - \lambda) d_2$ for any $\lambda \in [0, 1]$. Thanks to the derived inequality relation among the deletion channel capacities, we are able to improve upon the existing upper bounds on the capacity of the binary deletion channel for $d \geq 0.65$. The improvement is the result of the fact that the currently known best upper bounds [5] are not convex for some range of deletion probabilities. More precisely, we are able to prove that for any $0 \leq \lambda \leq 1$, $C_2(\lambda d + 1 - \lambda) \leq \lambda C_2(d)$ (where $C_2(d)$ stands for the binary deletion channel capacity), resulting in $C_2(d) \leq 0.4143(1 - d)$ for $d \geq 0.65$. This result is also a generalization of the one obtained in [9] which only holds asymptotically as $d \to 1$.

A different fragmentation idea is employed by the authors in [10] to derive the first non-trivial capacity upper bound for the i.i.d. $2K$-ary input deletion channel, and reduce the gap with the existing achievable rates in [10]. In this case, the fragmentation process used is a deterministic one where different subsets of channel inputs travel through different channels. Specifically, it is proved that $C_{2K}(d) \leq C_2(d) + (1 - d) \log(K)$ where $C_{2K}(d)$ denotes the capacity of a $2K$-ary deletion channel with deletion probability d.

The paper is organized as follows. In Section II, we prove a result on the binary deletion channel capacity which relates the capacity of the three different binary deletion channels through an inequality, and generalize it to the case of deletion/substitution channels. In Section III, we provide a discussion on applying a similar fragmentation idea to obtain upper bounds on the capacity of non-binary input deletion channels. In Section IV, we present tighter upper bounds on the capacity of the deletion channel based on previously known best upper bounds, and comment on the limit of the capacity bounds as the deletion probability approaches unity. We conclude the paper in Section V.

II. A NOVEL UPPER BOUND ON $C_2(d)$

In this section, we consider a "random" fragmentation for the binary input i.i.d. deletion channel. That is, we show a simple result that the parallel concatenation of two different independent deletion channels with deletion probabilities d_1 and d_2, in which every input bit is either transmitted over the first channel with probability of λ or over the second one with probability of $\bar{\lambda} = 1 - \lambda$, independently of each other, and the surviving output bits are combined without changing the order, is nothing but another deletion channel with deletion probability of $d = \lambda d_1 + \bar{\lambda} d_2$. This formulation allows us to provide an upper bound on the concatenated deletion channel capacity $C_2(d)$ in terms of a weighted average of $C_2(d_1)$, $C_2(d_2)$ and the parameters of the three channels. Furthermore, we argue that for the special case with $d_2 = 1$, i.e., $C_2(\lambda d_1 + \bar{\lambda}) \leq \lambda C_2(d_1)$, and generalize the result to the case of a binary input deletion/substitution channel.

This new look at the deletion channel and the resulting formulation allows us to prove the main result of this paper which is stated in the following theorem.

Theorem 1. *Let $C_2(d)$ denote the capacity of the i.i.d. deletion channel with deletion probability d, $\lambda \in [0,1]$ and $d = \lambda d_1 + \bar{\lambda} d_2$, then by defining $\bar{d} = 1 - d$, we have*

$$
\begin{aligned}
C_2(d) \leq {} & \lambda C_2(d_1) + \bar{\lambda} C_2(d_2) + \bar{d} \log(\bar{d}) - \lambda \bar{d_1} \log(\lambda \bar{d_1}) \\
& - \bar{\lambda} \bar{d_2} \log(\bar{\lambda} \bar{d_2}).
\end{aligned} \tag{1}
$$

Proof: Let us consider two different deletion channels, $\mathcal{C}_1$ and $\mathcal{C}_2$, with deletion probabilities d_1 and d_2, input sequences of bits $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$, and output sequences of bits $\boldsymbol{Y}_1$ and $\boldsymbol{Y}_2$, respectively. Denote their Shannon capacities by $C_2(d_1)$ and $C_2(d_2)$, respectively. Given a specific $\lambda \in [0,1]$, define a new binary input channel $\mathcal{C}'$ (as illustrated in Fig. 1) with input sequence of bits $\boldsymbol{X}$ and output sequence of bits $\boldsymbol{Y}$ as follows: each channel input symbol is transmitted through $\mathcal{C}_1$ with probability λ, and through $\mathcal{C}_2$ with probability $1 - \lambda$, independently of each other. Neither the transmitter nor the receiver knows the specific realization of the "individual channel selection events," i.e., they do not know which specific subchannel a symbol is transmitted through, or which specific subchannel each output symbol is received from. Lemmas 1 and 2 (given below) demonstrate that 1) the new channel is an i.i.d. deletion channel with deletion probability

$d = \lambda d_1 + \bar{\lambda} d_2$, 2) the capacity of the i.i.d. deletion channel with deletion probability d is upper bounded by

$$
\lambda C_2(d_1) + \bar{\lambda} C_2(d_2) + \bar{d} \log \bar{d} - \lambda \bar{d_1} \log(\lambda \bar{d_1}) - \bar{\lambda} \bar{d_2} \log(\bar{\lambda} \bar{d_2}).
$$

Combining these two results, the proof of the theorem follows. ∎

The following two lemmas are employed in the proof of the above theorem.

Lemma 1. *$\mathcal{C}'$ as defined in the proof of the theorem above is nothing but a deletion channel with deletion probability $d = \lambda d_1 + \bar{\lambda} d_2$.*

Proof: For each use of the channel $\mathcal{C}'$, for any input symbol $x \in \mathcal{X}$ and channel output $y \in \mathcal{Y}$, the transition probability is given by $P\{\mathcal{C}_1 \text{ is used}\} d_1 + P\{\mathcal{C}_2 \text{ is used}\} d_2 = \lambda d_1 + \bar{\lambda} d_2$. Noting that the subchannels are memoryless and the channel selection events are independent of each other, this transition matrix precisely defines a deletion channel with deletion probability $d = \lambda d_1 + \bar{\lambda} d_2$. ∎

Lemma 2. *The capacity of the channel $\mathcal{C}'$ as defined in the proof of the theorem above is upper bounded by*

$$
\lambda C_2(d_1) + \bar{\lambda} C_2(d_2) + \bar{d} \log \bar{d} - \lambda \bar{d_1} \log(\lambda \bar{d_1}) - \bar{\lambda} \bar{d_2} \log(\bar{\lambda} \bar{d_2}).
$$

Proof: Define the fragmentation process $\boldsymbol{F}_x$ as an N-tuple $\boldsymbol{F}_x = (f_x[1], \cdots, f_x[N])$, with $f_x[i] \in \{1, 2\}$ where the elements of the vector denotes the index of the channel that ith transmitted bit is going through, and similarly the fragmentation process $\boldsymbol{F}_x$ as an M-tuple $\boldsymbol{F}_y = (f_y[1], \cdots, f_y[M])$, where M denotes the length of the received sequence $\boldsymbol{Y}$, i.e., $M = |Y|$, and $f_y[i] \in \{1, 2\}$ denotes the index of the channel the i-th received bit is coming from. With this definition, clearly by knowing $\boldsymbol{F}_x$, $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$, one can retrieve $\boldsymbol{X}$, and by knowing $\boldsymbol{F}_y$, $\boldsymbol{Y}_1$ and $\boldsymbol{Y}_2$, one can retrieve $\boldsymbol{Y}$.

Since $\boldsymbol{X} \to (\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{F}_x) \to (\boldsymbol{Y}_1, \boldsymbol{Y}_2, \boldsymbol{F}_y) \to \boldsymbol{Y}$ form a Markov chain, we can write

$$
I(\boldsymbol{X}; \boldsymbol{Y}) \leq I(\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{F}_x; \boldsymbol{Y}_1, \boldsymbol{Y}_2, \boldsymbol{F}_y) = I_1 + I_2 + I_3, \tag{2}
$$

where

$$
I_1 = I(\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{F}_x; \boldsymbol{Y}_1),
$$

$$
I_2 = I(\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{F}_x; \boldsymbol{Y}_2 | \boldsymbol{Y}_1),
$$

and $I_3 = I(\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{F}_x; \boldsymbol{F}_y | \boldsymbol{Y}_1, \boldsymbol{Y}_2)$.

For I_1, we have

$$
I_1 = I(\boldsymbol{X}_1; \boldsymbol{Y}_1) + I(\boldsymbol{X}_2, \boldsymbol{F}_x; \boldsymbol{Y}_1 | \boldsymbol{X}_1) = I(\boldsymbol{X}_1; \boldsymbol{Y}_1), \tag{3}
$$

where we used the fact that $P(\boldsymbol{Y}_1 | \boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{F}_x) = P(\boldsymbol{Y}_1 | \boldsymbol{X}_1)$, i.e., $\boldsymbol{Y}_1$ is independent of $\boldsymbol{X}_2$ and $\boldsymbol{F}_x$ conditioned on $\boldsymbol{X}_1$.

Furthermore, by using the facts that $P(\boldsymbol{Y}_2 | \boldsymbol{X}_2, \boldsymbol{Y}_1) = P(\boldsymbol{Y}_2 | \boldsymbol{X}_2)$ and $P(\boldsymbol{Y}_2 | \boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{F}_x, \boldsymbol{Y}_1) = P(\boldsymbol{Y}_2 | \boldsymbol{X}_2)$, we obtain

$$
\begin{aligned}
I_2 = {} & I(\boldsymbol{X}_2; \boldsymbol{Y}_2 | \boldsymbol{Y}_1) + I(\boldsymbol{X}_1, \boldsymbol{F}_x; \boldsymbol{Y}_2 | \boldsymbol{Y}_1, \boldsymbol{X}_2) \\
= {} & H(\boldsymbol{Y}_2 | \boldsymbol{Y}_1) - H(\boldsymbol{Y}_2 | \boldsymbol{X}_2) \leq I(\boldsymbol{X}_2; \boldsymbol{Y}_2). \tag{4}
\end{aligned}
$$
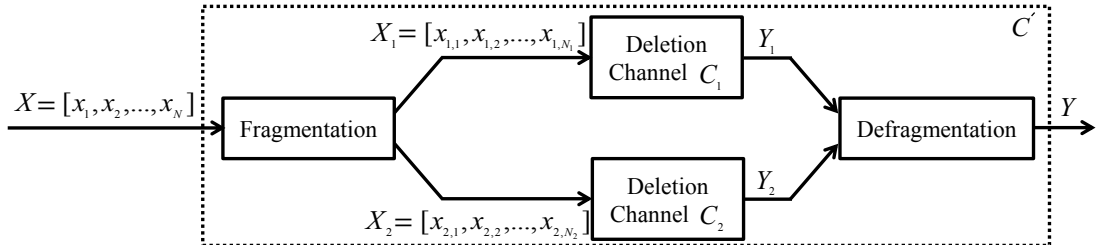
Fig. 1. Illustration of the new channel $\mathcal{C}'$.

On the other hand, for $I(\boldsymbol{X}_i; \boldsymbol{Y}_i)$ ($i \in \{1, 2\}$), we obtain (see [11] for details)

$$I(\boldsymbol{X}_i; \boldsymbol{Y}_i) \leq \lambda_i N C_2(d_i) + \log(\lambda_i N + 1) + \log(N + 1). \quad (5)$$

We are not able to derive the exact value of I_3, therefore we resort to an upper bound which results in an upper bound on $I(\boldsymbol{X}, \boldsymbol{Y})$. Namely, we can prove that (see [11] for details)

$$\begin{aligned} I_3 &\leq N(\lambda \bar{d}_1 + \bar{\lambda} \bar{d}_2) \log(\lambda \bar{d}_1 + \bar{\lambda} \bar{d}_2) - N \lambda \bar{d}_1 \log(\lambda \bar{d}_1) \\ &\quad - N \bar{\lambda} \bar{d}_2 \log(\bar{\lambda} \bar{d}_2). \end{aligned} \quad (6)$$

Finally, by substituting (5), (6), (4) and (3) into (2), we obtain

$$\begin{aligned} I(\boldsymbol{X}; \boldsymbol{Y}) &\leq N \lambda C_2(d_1) + N \bar{\lambda} C_2(d_2) - N \lambda \bar{d}_1 \log(\lambda \bar{d}_1) \\ &\quad - N \bar{\lambda} \bar{d}_2 \log(\bar{\lambda} \bar{d}_2) + N \bar{d} \log \bar{d} + \log(\bar{\lambda} N + 1) \\ &\quad + 2 \log(N + 1) + \log(\lambda N + 1). \end{aligned}$$

Dividing both sides of the above inequality by N, letting N go to infinity, and noting that the inequality is valid for any input distribution $P(\boldsymbol{X})$, the proof follows. ∎

A. Generalization to the Case of Deletion/Substitution Channels

In a deletion/substitution channel (special case of the Gallager channel model [12] without any insertions) with parameters (d, f), any transmitted bit is either deleted with probability of d or flipped with probability of f or received correctly with probability of $1 - d - f$, where neither the transmitter nor the receiver have any information about the position of the deleted and flipped bits. Another look at the deletion/substitution channel can be as a series concatenation of two independent channels such that the first one is a deletion-only channel with deletion probability of d and the second one is binary symmetric channel with cross error probability of $s = \frac{f}{1-d}$. It is easy to show that the result of Theorem 1 can also be extended to the deletion/substitution channel as given in the following corollary.

Corollary 1. *Let $C_2(d, f)$ denote the capacity of the deletion/substitution channel with deletion probability d and flipping probability f, $\lambda \in [0, 1]$, $d = \lambda d_1 + \bar{\lambda} d_2$ and $f = \lambda f_1 + \bar{\lambda} f_2$, then we have*

$$\begin{aligned} C_2(d, f) &\leq \lambda C_2(d_1, f_1) + \bar{\lambda} C_2(d_2, f_2) + \bar{d} \log \bar{d} \\ &\quad - \lambda \bar{d}_1 \log(\lambda \bar{d}_1) - \bar{\lambda} \bar{d}_2 \log(\bar{\lambda} \bar{d}_2). \end{aligned} \quad (7)$$

Proof: See [11] for details. ∎

Similar to the case of deletion-only channels, this expression provides a tighter upper bound on the deletion/substitution channel capacity compared to the existing bounds in the literature for a wide range of channel parameters (which is discussed further in the numerical examples section).

III. NON-BINARY DELETION CHANNELS

In this section, we review a recent result reported in [10] where we have considered non-binary input deletion channels using a different fragmentation approach. Loosely speaking, instead of using a random fragmentation, a deterministic fragmentation is used and an arbitrary Q-ary channel is decomposed into parallel deletion channels with smaller alphabets. For the Q-ary input deletion channel, by choosing an appropriate deterministic fragmentation of the input symbol set it is possible to derive a new upper bound on the Q-ary input deletion channel capacity in terms of the lower order deletion channels.

As a special case, when Q is an even number, i.e., for a $2K$-ary input deletion channel, it is proved in [10] that $C_{2K}(d) \leq C_2(d) + (1 - d) \log(K)$. The main idea is that any $2K$-ary input deletion channel with deletion probability d can be considered as a parallel concatenation of K independent binary deletion channels $\mathcal{C}_k$ ($k \in \{1, \ldots, K\}$) all with the same deletion probability d, in which the input symbols $2k-1$ and $2k$ travel through $\mathcal{C}_k$ and the surviving output symbols of the subchannels are combined based on the order in which they go through the subchannels. This is a useful result to come up with improved upper bounds on the capacity of non-binary deletion channels by exploiting already existing results for the binary case. Examples of improved bounds for this setting are provided in [10].

IV. EXAMPLES OF THE NEWLY DERIVED CAPACITY UPPER BOUNDS

In this section, we provide several implications of the results presented in the paper on the binary deletion channel capacity. Namely, we explicitly demonstrate the tightest upper bound on the binary input deletion channel capacity for $d \geq 0.65$.

An interesting application of the result (1) on the capacity of the binary deletion and deletion/substitution channels is in obtaining improved capacity upper bounds. For instance, the best known upper bound on the deletion channel capacity is
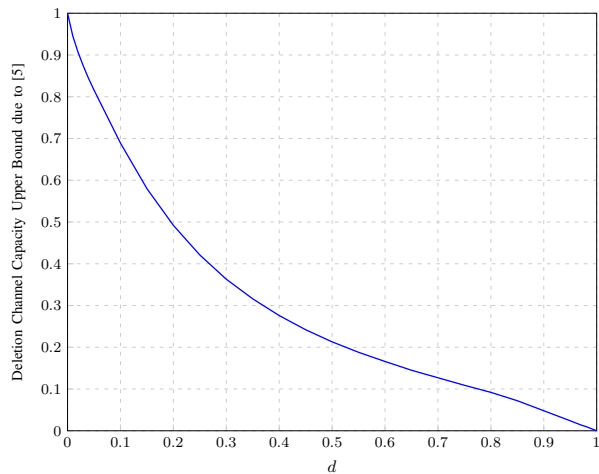
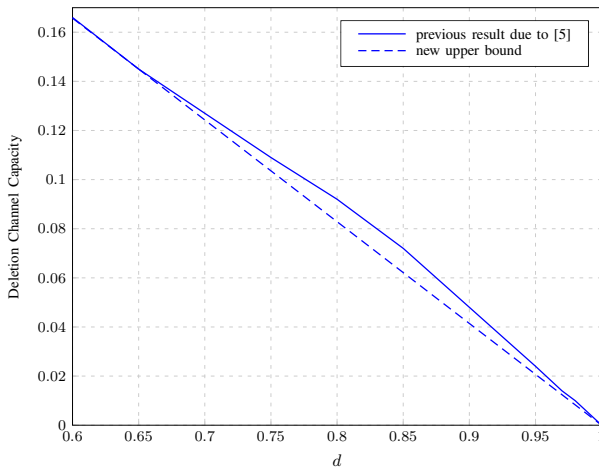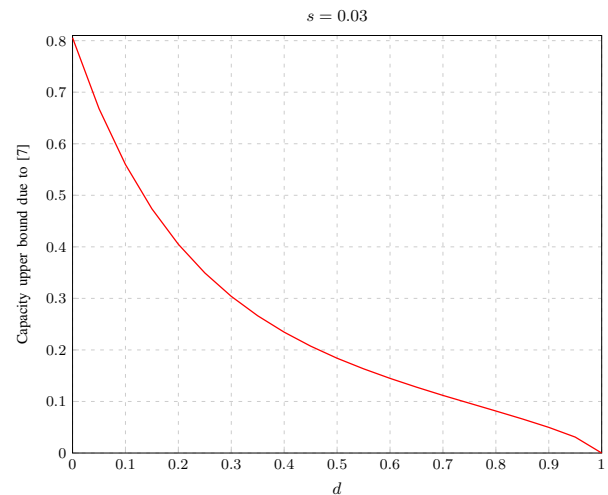Fig. 2. Previously best known upper bound on the i.i.d. deletion channel capacity.

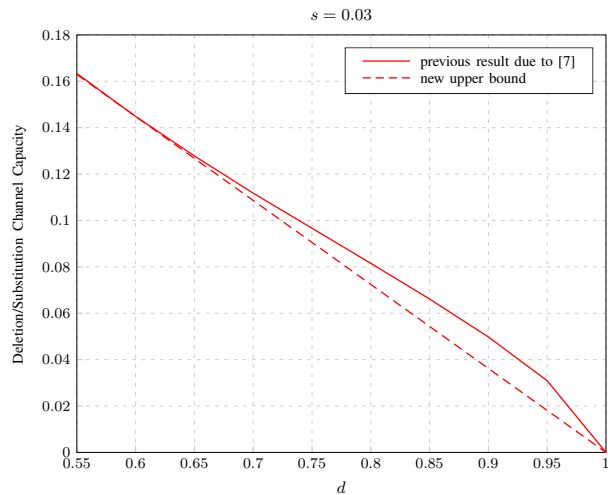Fig. 4. Previously best known upper bound on the deletion/substitution channel capacity for $s = 0.03$.

Fig. 3. Improved upper bound on the deletion channel capacity employing $C(\lambda d + \bar{\lambda}) \leq \lambda C(d)$.

Fig. 5. Improved upper bound on the deletion/substitution channel capacity for $s = 0.03$.

not convex for $d \geq 0.65$ as shown in Fig. 2 (with values taken from the boldfaced values in Table IV of [5]). As clarified in the table, the best known values for small d are due to [13], for a wide range (up to $d \sim .8$) are due to the "fourth version" of the upper bound (named C_4 in [5]), and for large values of d are due to the "second version" named C_2^* in the same paper. Therefore, the deletion channel capacity upper bound can be improved for $d \in (0.65, 1)$ as $C_2(1 - 0.35\lambda) \leq \lambda C_2(0.65)$ with $0 \leq \lambda \leq 1$. That is, we have $C_2(d) \leq 0.4143(1-d)$ for $d \in (0.65, 1)$ as illustrated in Fig. 3. Clearly, this new version of the i.i.d. deletion channel upper bound is tighter than the previous version, as we are able to "convexify" the existing upper bound.

We note that our result is a generalization of the one in [9] where it was shown that $C_2(d) \leq 0.4143(1 - d)$ as $d \to 1$. We also note an earlier asymptotic result on a lower bound derived in [2] which states that $C_2(d)$ as $d \to 1$ is larger than $0.1185(1-d)$.

As another application of the approach proposed in this pa-per, we can consider the capacity of the deletion/substitution channel. The best known capacity upper bound for this case is given in [7], e.g., Fig. 1 of [7] presents several upper bounds for fixed $s = 0.03$ (see Fig. 4). It is clear that this bound is not a convex function of the deletion probability for $d \geq 0.6$, hence it can be improved. That is, applying the result in our paper, we obtain, for instance for $s = 0.03$, $C_s(d, 0.03) \leq 0.3621(1 - d)$ for $d \geq 0.6$ which is a tighter bound than the existing one as illustrated in Fig. 5.

We conclude this section by noting that examples of improved channel capacity bounds can be given for the case of non-binary deletion channels as well. In particular, it can be shown that for moderate channel alphabet sizes, the bound presented in the previous section exploits the available bounds on the binary deletion channels effectively, and provides significant improvements over the only other alternative bound, namely, the one based on the erasure channel assumption. We omit details of these results here and refer the reader to [10].

V. Conclusions

In this paper, we considered an alternative look at the i.i.d. deletion channels. Specifically, we assumed that the input of the channel is fragmented into smaller sequences traveling through different (independent) deletion channels, and the outputs are combined without changing their order in the transmitted sequence. We showed that this new look provides a different way to study the capacity of deletion channels. For the binary deletion channel by considering a random fragmentation process, an inequality relating the capacity of a binary deletion channel to two other binary deletion channels is found. For non-binary input deletion channels, considering a deterministic fragmentation of the input and output sequences provides us with a way to upper bound the channel capacity with deletion channel capacities with smaller alphabets. An immediate application of the result for the binary input case is in obtaining improved upper bounds on the capacity of the deletion channel. For instance, for an i.i.d. deletion channel, we proved that $C_2(d) \leq 0.4143(1-d)$ for all $d \geq 0.65$. This is a stronger result than the earlier characterization in [9] which is valid only asymptotically as $d \to 1$.

References

[1] R. L. Dobrushin, "Shannon's theorems for channels with synchronization errors," *Probs. Inf. Transm.*, vol. 3, no. 4, pp. 11–26, 1967.

[2] 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, 2006.

[3] E. Drinea and M. Mitzenmacher, "Improved lower bounds for the capacity of i.i.d. deletion and duplication channels," *IEEE Trans. Inf. Theory*, vol. 53, no. 8, pp. 2693–2714, Aug. 2007.

[4] A. Kirsch and E. Drinea, "Directly lower bounding the information capacity for channels with i.i.d. deletions and duplications," *IEEE Trans. Inf. Theory*, vol. 56, no. 1, pp. 86 –102, Jan. 2010.

[5] D. Fertonani and T. M. Duman, "Novel bounds on the capacity of the binary deletion channel," *IEEE Trans. Inf. Theory*, vol. 56, no. 6, pp. 2753–2765, June 2010.

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

[7] D. Fertonani, T. M. Duman, and M. F. Erden, "Bounds on the capacity of channels with insertions, deletions and substitutions," *IEEE Trans. on Communications*, vol. 59, no. 1, pp. 2–6, Jan. 2011.

[8] M. Rahmati and T. M. Duman, "Bounds on the capacity of the random insertion and deletion-additive noise channels," *IEEE Trans. Inf. Theory*, vol. 59, no. 9, pp. 5534–5546, Sep. 2013.

[9] M. Dalai, "A new bound on the capacity of the binary deletion channel with high deletion probabilities," *in Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, pp. 499–502, Aug. 2011.

[10] M. Rahmati and T. M. Duman, "An upper bound on the capacity of non-binary deletion channels," *presented in ISIT 2013, ArXiv e-prints:1301.6599[cs.IT]*, July 2013.

[11] ——, "Some results on the capacity of deletion channels," *submitted to IEEE Trans. Inf. Theory*, Sep. 2013.

[12] R. Gallager, "Sequential decoding for binary channels with noise and synchronization errors," *Tech. Rep., MIT Lincoln Lab. Group Report*, Oct. 1961.

[13] S. Diggavi, M. Mitzenmacher, and H. Pfister, "Capacity upper bounds for deletion channels," in *Proceedings of IEEE International Symposium on Information Theory (ISIT)*, 2007, pp. 1716–1720.