Upper Bounds on the Capacity of Deletion Channels Using Channel Fragmentation

Mojtaba Rahmati and Tolga M. Duman, *Fellow, IEEE*

***Abstract*— We study memoryless channels with synchronization errors as defined by a stochastic channel matrix allowing for symbol drop-outs or symbol insertions with particular emphasis on the binary and non-binary deletion channels. We offer a different look at these channels by considering equivalent models 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. We consider both random and deterministic types of fragmentation processes applied to binary and nonbinary deletion channels. With two specific applications of this idea, a random fragmentation applied to a binary deletion channel and a deterministic fragmentation process applied to a nonbinary deletion channel, we prove certain inequality relations among the capacities of the original channels and those of the introduced subchannels. The resulting inequalities prove useful in deriving tighter capacity upper bounds for: 1) independent identically distributed (i.i.d.) deletion channels when the deletion probability exceeds 0.65 and 2) nonbinary deletion channels. Some extensions of these results, for instance, to the case of deletion/substitution channels are also explored.**

***Index Terms*— Binary deletion channel, non-binary deletion channel, deletion/substitution channel, channel capacity, capacity upper bounds.**

I. Introduction

CHANNELS with synchronization errors can be modeled using symbol drop-outs and/or symbol insertions as well as random errors. There are many different models adopted in the literature to describe the resulting channels in different applications. Among them, a relatively general model is employed by Dobrushin [1] where 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. 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. In this paper, we consider the i.i.d. deletion channel model for both binary and non-binary input cases. In an i.i.d. deletion channel, the transmitted symbols are either received correctly and in the right order or 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 [2]–[6].

Another special case of the general model by Dobrushin is the Gallager model allowing for insertions, deletions and substitution errors in a binary input channel in which every transmitted bit is either deleted with probability d, replaced by two random bits with probability i, flipped with probability f or received correctly with probability $1-d-i-f$. With $i = 0$, the Gallager model boils down to the deletion/substitution channel model which is also considered in this paper. 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 (BSC) with cross error probability of $s = \frac{f}{1-d}$. There are some capacity upper and lower bounds for the Gallager's insertion/deletion channel model in the literature, see [7], [8].

In this paper, for both binary and non-binary input deletion channels, it is shown 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. Furthermore, this new formulation provides a means of relating the capacities of the original channels and those of the subchannels considered through certain inequalities, thereby allowing us to obtain tighter capacity upper bounds for certain synchronization error channels.

Manuscript received September 19, 2013; revised June 6, 2014; accepted September 25, 2014. Date of publication November 7, 2014; date of current version December 22, 2014. This work was supported in part by the National Science Foundation under Contract NSF-TF 0830611 and in part by the European Commission Marie Curie Career Integration under Grant PCIG12-GA-2012-334213. This paper was presented at the 2013 IEEE International Symposium on Information Theory [17] and 2013 Allerton Conference on Communication, Control and Computing [18].

M. Rahmati was with the School of Electrical, Computer and Energy Engineering, Arizona State University, Phoenix, AZ 85004 USA. He is now with Samsung Modem Research and Development, San Diego, CA 92121 USA (e-mail: m.rahmati@samsung.com).

T. M. Duman is with the Department of Electrical and Electronics Engineering, Bilkent University, Bilkent, Ankara, 06800, Turkey and he is on leave from the School of ECEE of Arizona State University. (e-mail: duman@ee.bilkent.edu.tr).

Communicated by H. D. Pfister, Associate Editor for Coding Theory.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIT.2014.2368553

For the binary input deletion channel, we prove that the capacity of an i.i.d. deletion channel with deletion probability d can be upper bounded in terms of the capacities of i.i.d. 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 $\lambda \in [0, 1]$. The proof relies on a simple observation that the deletion channel with deletion probability d can be considered as a "parallel concatenation" of two independent deletion channels with deletion probabilities d_1 and d_2 where each bit is either transmitted over the first channel with probability λ or the second channel with probability $1 - \lambda$ independently of each other. We formalize the equivalence in Section III. 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 $d \geq 0.65$ [5]. The improvement comes from the fact that the currently known best upper bounds are not convex for some range of deletion probabilities. More precisely, our result allows us to convexify the existing deletion channel capacity upper bound for $d \geq 0.65$, leading to a significant improvement for a wide range of deletion probabilities. More precisely, we are able to prove that for $0 \leq \lambda \leq 1$, $C_2(\lambda d + 1 - \lambda) \leq \lambda C_2(d)$ (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 broad generalization of the one obtained in [9] which only holds asymptotically as $d \to 1$. We also demonstrate that a similar improvement is possible for the case of deletion/substitution channels. As an example, we can prove that for substitution probability of $s = 0.03$, an improved capacity upper bound is obtained for $d \geq 0.6$ over the best existing result given in [7].

For the non-binary case, we derive the first non-trivial capacity upper bound for the i.i.d. deletion channel, and reduce the gap with the existing achievable rates. To derive the results we first prove an inequality between the capacity of a $2K$-ary deletion channel with deletion probability d, denoted by $C_{2K}(d)$, and the capacity of the binary deletion channel with the same deletion probability, $C_2(d)$, that is, $C_{2K}(d) \leq C_2(d) + (1 - d)\log(K)$. As a result, any upper bound on the binary deletion channel capacity can be used to derive an upper bound on the $2K$-ary deletion channel capacity. Therefore by employing existing upper bounds on the capacity of the binary deletion channel, we obtain upper bounds on the capacity of the $2K$-ary deletion channel. For example, using the result on the binary deletion channel stated in the previous paragraph, we obtain $C_{2K}(d) \leq (\log(K) + 0.4143)(1 - d)$ for $d \geq 0.65$. Furthermore, we illustrate via examples the use of the new bounds and discuss their asymptotic behavior as $d \to 0$.

The paper is organized as follows. In Section II, we first provide the model for non-binary deletion channels, and then review the previous work on the capacity of both binary and non-binary input deletion channels. In Section III, 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. We provide our new upper bound on the capacity of the non-binary deletion channels in Section IV. In Section V, we present tighter upper bounds on the capacity of the deletion and deletion/substitution channels based on previously known best upper bounds (for binary channels), and comment on the limit of the capacity as the deletion probability approaches unity. Furthermore, we provide several implications of the result for the non-binary case where we compare the resulting capacity upper bounds with the existing capacity upper and lower bounds, and provide a discussion of the non-binary input channel capacity behavior as the deletion probability approaches zero. We conclude the paper in Section VI.

II. Preliminaries

In this section, we first introduce the general model for i.i.d. deletion channels, and then review the existing work on the deletion channel capacity in the literature.

A. Channel Model

An i.i.d. Q-ary input deletion channel with input alphabet $\mathcal{X} = \{1, \ldots, Q\}$ is considered in which every transmitted symbol is either randomly deleted with probability d or received correctly with probability $1 - d$ while there is no information about the values or the positions of the lost symbols at the transmitter or at the receiver. In transmission of N symbols through the channel, the input sequence is denoted by $\boldsymbol{X} = (x_1, \ldots, x_N)$ in which $x_n \in \mathcal{X}$ and $\boldsymbol{X} \in \mathcal{X}^N$, and the output sequence is denoted by $\boldsymbol{Y} = (y_1, \ldots, y_M)$ in which M is a binomial random variable with parameters N and d (due to the characteristics of the i.i.d. deletion channel). With $Q = 2$, we obtain the usual binary input i.i.d. deletion channel.

B. Brief Literature Review

Capacity of binary deletion channels has received significant attention in the existing literature, see [10] and references therein. Examples of the deletion channel capacity lower bounds include [4], [11], [12]. Gallager [11] provided the first lower bound on the transmission capacity of the channels with random insertion, deletion and substitution errors which provides a lower bound on the binary deletion channel capacity as well. The tightest lower bound on the binary deletion channel capacity is provided in [4] where the information capacity of the binary deletion channel is directly lower bounded by considering input sequences as alternating blocks of zeros and ones (runs) and the length of the runs L as i.i.d. random variables following a particular distribution over positive integers with a finite expectation and finite entropy.

There are also several upper bounds on the binary deletion channel capacity, see [5], [13]. In [13] a genie-aided channel is considered in which the receiver is provided with the side information about the completely deleted runs. For example, if "110001" is transmitted but "111" is received (i.e., the entire run of "000" is lost), the genie aided channel considers the received signal as "11 − 1", i.e., the position of the complete lost run is marked by a different symbol. An upper bound on the capacity per unit cost of the genie-aided channel is computed by running the Blahut-Arimoto algorithm (BAA). Fertonani and Duman [5] take a similar approach, but consider

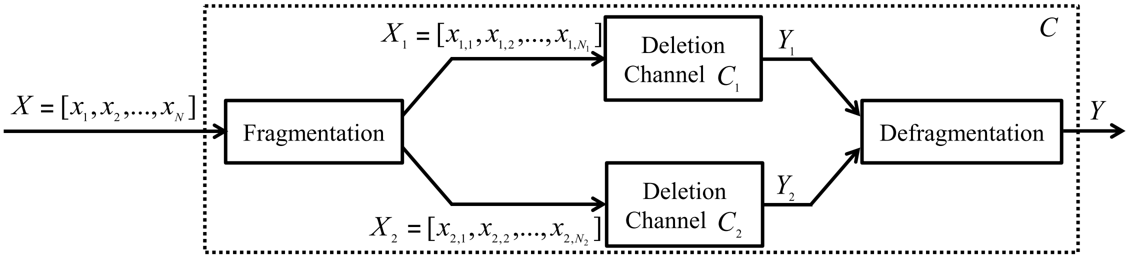

Fig. 1. Illustration of the new channel $\mathcal{C}'$.

different genie-aided channels, along with the BAA, and obtain tighter upper bounds on the binary deletion channel capacity.

Despite the extensive work on binary deletion channels, the case of non-binary deletion channels has not received significant attention so far. To the best of our knowledge, the only non-trivial lower bounds on the capacity of the non-binary deletion channels are provided in [6] where two different bounds are derived. More precisely, the achievable rates for Q-ary input deletion channels are computed for i.i.d. and Markovian codebooks by considering a simple decoder which decides in favor of a sequence if the received sequence is a subsequence of only one transmitted sequence. The derived achievable rates are given by

$$C_Q \geq \log\left(\frac{Q}{Q-1}\right) + (1-d)\log(Q-1) - H_b(d), \quad (1)$$

by considering i.i.d. codebooks, where $H_b(d) = -d\log(d) - (1-d)\log(1-d)$, and

$$C_Q \geq \sup_{\gamma>0,\ 0<p<1} \left[-(1-d)\log\left((1-q)A + qB\right) - \gamma\log(e)\right] \quad (2)$$

by considering Markovian codebooks, with d being the deletion probability, $q = \frac{1}{Q}\left(1 + \frac{(1-d)(Q-1)(Qp-1)}{Q-1-d(Qp-1)}\right)$, $A = \frac{e^{-\gamma}(1-p)}{(Q-1)(1-e^{-\gamma}(1-\frac{1-p}{Q-1}))}$ and $B = e^{-\gamma}\left((1-p)A + p\right)$. Non-binary input alphabet channels with synchronization errors are also considered in [14] where the capacity of memoryless synchronization error channels in the presence of noise and the capacity of channels with weak synchronization errors (i.e., the transmitter and receiver are partly synchronized) have been studied. The main focus of the work in [14] is on the asymptotic behavior of the channel capacity for large values of Q.

III. An Improved Upper Bound on the Capacity of Binary Deletion Channels

As stated in the introduction section, the main idea explored in this paper is the fragmentation of the input and output sequences of a deletion channel in an effort to come up with alternate representations of the channel input and output processes. We do so in such a way that the alternate representations are helpful in the channel capacity study. As a first fragmentation approach, 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 present a simple proof for the special case with $d_2 = 0$, i.e., $C_2(\lambda d_1 + \bar{\lambda}) \leq \lambda C_2(d_1)$, and generalize the result to the case of binary input deletion/substitution channel and randomized parallel concatenation of more than two deletion channels.

A. A Novel Upper Bound on $C_2(d)$

The following theorem states our result on the binary deletion channel capacity whose proof hinges on a simple observation.

Theorem 1: Let $C_2(d)$ denote the capacity of the i.i.d. binary 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

$$C_2(d) \leq \lambda C_2(d_1) + \bar{\lambda} C_2(d_2) + \bar{d}\log(\bar{d}) - \lambda\bar{d}_1\log(\lambda\bar{d}_1) - \bar{\lambda}\bar{d}_2\log(\bar{\lambda}\bar{d}_2). \quad (3)$$

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 $\mathcal{C}'$ (shown 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 $\bar{\lambda}$, 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, and which specific subchannel each output symbol is received from. Lemmas 1 and 2 (given below) demonstrate that 1) the new channel is a new i.i.d. deletion channel with deletion probability $d = \lambda d_1 + \bar{\lambda} d_2$, 2) if appropriate side information be provided for the transmitter and the receiver then the capacity of the genie-aided channel 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 easily by noting that the capacity of the new channel $\mathcal{C}'$ cannot decrease with side information. ■

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

Lemma 1: $\mathcal{C}'$ as defined in the proof of Theorem 1 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 Theorem 1 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: We first define a new genie-aided channel which is obtained by providing the transmitter and the receiver of the channel $\mathcal{C}'$ with appropriate side information, then derive an upper bound on the capacity of the genie-aided channel which is also an upper bound on the capacity of the channel $\mathcal{C}'$. More precisely, we provide the transmitter with side information on which channel is being used for each transmitted symbol ($\boldsymbol{X} = \boldsymbol{X}_1\boldsymbol{X}_2$), and the receiver with side information on which channel the received symbol comes from ($\boldsymbol{Y} = \boldsymbol{Y}_1\boldsymbol{Y}_2$), and reveal the side information on the fragmentation information, i.e., random process $\boldsymbol{F}_y$, to the receiver such that by knowing $\boldsymbol{F}_y$, $\boldsymbol{Y}_1$ and $\boldsymbol{Y}_2$, one can retrieve Y. $\boldsymbol{F}_y$ is defined as an M-tuple $\boldsymbol{F}_y = (f_y[1], \ldots, f_y[M])$, 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. We also define $\boldsymbol{F}_x$ which determines the fragmentation process from the random process $\boldsymbol{X}$ to $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ as an N-tuple $\boldsymbol{F}_x = (f_x[1], \ldots, f_x[N])$, where $f_x[i] \in \{1, 2\}$ denotes the index of the channel the i-th bits is going through.

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

$$\begin{aligned} I(\boldsymbol{X}; \boldsymbol{Y}) &\le I(\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{F}_x; \boldsymbol{Y}_1, \boldsymbol{Y}_2, \boldsymbol{F}_y) \\ &= I_1 + I_2 + I_3, \end{aligned} \tag{4}$$

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

$$\begin{aligned} I_1 &= I(\boldsymbol{X}_1; \boldsymbol{Y}_1) + I(\boldsymbol{X}_2, \boldsymbol{F}_x; \boldsymbol{Y}_1|\boldsymbol{X}_1) \\ &= I(\boldsymbol{X}_1; \boldsymbol{Y}_1), \end{aligned} \tag{5}$$

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

$$\begin{aligned} I_2 &= I(\boldsymbol{X}_2; \boldsymbol{Y}_2|\boldsymbol{Y}_1) + I(\boldsymbol{X}_1, \boldsymbol{F}_x; \boldsymbol{Y}_2|\boldsymbol{Y}_1, \boldsymbol{X}_2) \\ &= H(\boldsymbol{Y}_2|\boldsymbol{Y}_1) - H(\boldsymbol{Y}_2|\boldsymbol{X}_2) \\ &\le I(\boldsymbol{X}_2; \boldsymbol{Y}_2). \end{aligned} \tag{6}$$

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 $I(\boldsymbol{X}, \boldsymbol{Y})$. For I_3, if we define $N_i = |X_i|$ and $M_i = |Y_i|$ as the length of the transmitted and received sequences form the i-th channel, respectively, then we can write

$$\begin{aligned} I_3 &= H(\boldsymbol{F}_y|\boldsymbol{Y}_1, \boldsymbol{Y}_2) - H(\boldsymbol{F}_y|\boldsymbol{Y}_1, \boldsymbol{Y}_2, \boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{F}_x) \\ &\le H(\boldsymbol{F}_y|\boldsymbol{Y}_1, \boldsymbol{Y}_2) \\ &= H(\boldsymbol{F}_y|\boldsymbol{M}_1, \boldsymbol{M}_2). \end{aligned} \tag{7}$$

For fixed M_1 and M_2, there are $\binom{M_1+M_2}{M_2}$ possibilities for $\boldsymbol{F}_y = (f_y[1], \ldots, f_y[M_1])$. Therefore, we obtain (see Appendix A)

$$\begin{aligned} H(\boldsymbol{F}_y|\boldsymbol{M}_1 = M_1, \boldsymbol{M}_2 = M_2) &\le \log\left(\binom{M_1 + M_2}{M_2}\right) \\ &\le (M_1 + M_2)\log(M_1 + M_2) \\ &\quad - M_1 \log(M_1) - M_2 \log(M_2). \end{aligned} \tag{8}$$

Furthermore, since $g([M_1, M_2]) = (M_1 + M_2)\log(M_1 + M_2) - M_1 \log(M_1) - M_2 \log(M_2)$ is a concave function of $[M_1, M_2]$ (see Appendix B), by applying Jensen's inequality, we obtain

$$\begin{aligned} I_3 &\le E_{\boldsymbol{M}_1, \boldsymbol{M}_2}\left\{H(\boldsymbol{F}_y|\boldsymbol{M}_1, \boldsymbol{M}_2)\right\} \\ &\le E\{\boldsymbol{M}_1 + \boldsymbol{M}_2\}\log(E\{\boldsymbol{M}_1 + \boldsymbol{M}_2\}) \\ &\quad - E\{\boldsymbol{M}_1\}\log(E\{\boldsymbol{M}_1\}) - E\{\boldsymbol{M}_2\}\log(E\{\boldsymbol{M}_2\}) \end{aligned}$$

Due to the structure of the channel $\mathcal{C}'$, $\boldsymbol{M}_i$ is binomially distributed, i.e., $P(\boldsymbol{M}_i = M_i) = \binom{N}{M_i}(\lambda_i \bar{d}_i)^{M_i}(\bar{\lambda}_i + \lambda_i d_i)^{N-M_i}$, and as a result $E\{\boldsymbol{M}_i\} = N\lambda_i(1-d_i)$. Therefore, we arrive at

$$\begin{aligned} I_3 &\le N(\lambda\bar{d}_1 + \bar{\lambda}\bar{d}_2)\log(\lambda\bar{d}_1 + \bar{\lambda}\bar{d}_2) \\ &\quad - N\lambda\bar{d}_1 \log(\lambda\bar{d}_1) - N\bar{\lambda}\bar{d}_2 \log(\bar{\lambda}\bar{d}_2). \end{aligned} \tag{9}$$

On the other hand, for $I(\boldsymbol{X}_i; \boldsymbol{Y}_i)$ ($i \in \{1, 2\}$), we have (see Appendix C)

$$\begin{aligned} I(\boldsymbol{X}_i; \boldsymbol{Y}_i) &\le E\{\boldsymbol{N}_i\}C_2(d_i) + 2\log(N+1) \\ &= \lambda_i N C_2(d_i) + 2\log(N+1). \end{aligned} \tag{10}$$

Finally, by substituting (10), (9), (6) and (5) into (4), we obtain

$$\begin{aligned} I(\boldsymbol{X}; \boldsymbol{Y}) &\le N\lambda C_2(d_1) + N\bar{\lambda} C_2(d_2) + 4\log(N+1) \\ &\quad + N\bar{d}\log(\bar{d}) - N\lambda\bar{d}_1 \log(\lambda\bar{d}_1) - N\bar{\lambda}\bar{d}_2 \log(\bar{\lambda}\bar{d}_2). \end{aligned}$$

By 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. ■

B. Special Case With $d_2 = 1$

In this subsection, we note that for the special case of $d_2 = 1$, the derived upper bound (3) results in $C_2(\lambda d_1 + \bar{\lambda}) \le \lambda C_2(d_1)$. This expression provides a tighter upper bound on the capacity of the binary deletion channel for deletion probabilities larger than 0.65 compared to the existing ones in the literature if $d_1 = 0.65$ is used and the capacity upper bound for this value of the deletion probability is substituted from [5].

We digress here to point out that to prove the special case $C_2(\lambda d_1 + \bar{\lambda}) \le \lambda C_2(d_1)$, there is no need for the entire proof given in Lemma 2. More precisely, when $\mathcal{C}_2$ is a pure deletion channel, $\boldsymbol{X} \to \boldsymbol{X}_1 \to \boldsymbol{Y}_1 \to \boldsymbol{Y}$ form a Markov chain ($\boldsymbol{Y} = \boldsymbol{Y}_1$), therefore we can write

$$\begin{aligned} I(\boldsymbol{X};\boldsymbol{Y}) &\le I(\boldsymbol{X}_1;\boldsymbol{Y}_1) \\ &\le \lambda N C_2(d_1) + \log(\lambda_1 N + 1) + \log(N+1), \end{aligned} \tag{11}$$

where the last inequality holds due to the result in Appendix C. Furthermore, by dividing both sides of the above inequality by N, letting N go to infinity, and the fact that the inequality is valid for any input distribution $P(\boldsymbol{X})$, we arrive at $C_2(\lambda d_1 + \bar{\lambda}) \le \lambda C_2(d_1)$.

Another observation from the result $C_2(\lambda d_1 + \bar{\lambda}) \le \lambda C_2(d_1)$ is that by series concatenation of two independent deletion channels with deletion probabilities d_1 and $\bar{\lambda}$, we also arrive at a deletion channel with deletion probability of $d = \lambda d_1 + \bar{\lambda}$. Therefore we can say that the capacity of the series concatenation of two independent deletion channels can be upper bounded in terms of the capacity of one of them and the parameters of the other.

C. Generalization to the Case of Deletion/Substitution Channels

Deletion/substitution channel is a special case of the Gallager channel model without any insertions. In a deletion/substitution channel 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 $1-d-f$. The position of the deleted and flipped bits are not known either at the transmitter or the receiver. It is easy to show that the result of Theorem 1 can also be generalized 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

$$\begin{aligned} C_2(d,f) \le \lambda C_2(d_1,f_1) + \bar{\lambda} C_2(d_2,f_2) + \bar{d}\log(\bar{d}) \\ -\lambda\bar{d}_1\log(\lambda\bar{d}_1) - \bar{\lambda}\bar{d}_2\log(\bar{\lambda}\bar{d}_2). \end{aligned} \tag{12}$$

Proof: The proof of Lemma 1 simply holds if we consider $\mathcal{C}_1$ in Fig. 1 as a deletion/substitution channel with parameters (d_1, f_1) and $\mathcal{C}_2$ as another deletion/substitution channel with parameters (d_2, f_2). Then $\mathcal{C}$ also becomes a deletion/substitution channel with parameters $(\lambda d_1 + \bar{\lambda} d_2, \lambda f_1 + \bar{\lambda} f_2)$. Furthermore, replacing the deletion channel $\mathcal{C}_i$ with deletion probability d_i with a deletion/substitution channel with parameters (d_i, f_i) does not change the distribution of $\boldsymbol{N}_i$ and $\boldsymbol{M}_i$. Therefore, the proof of Lemma 2 holds for the deletion/substitution channel as well. ∎

One can also consider a deletion/substitution channel with parameters (d,f) as a series concatenation of two independent channels. The first channel is a deletion only channel with deletion probability d and the second one is a binary symmetric channel (BSC) with crossover probability $s = \frac{f}{1-d}$ ($1-d-f \le 1$ and, if $d=1$ then $s=0$). Therefore, we can represent any deletion/substitution channel with another set of parameters d and s as well, and denote the capacity by $C_s(d,s)$ with the understanding that $C_s(d,s) = C_2(d,f) = C_2(d,(1-d)s)$. With this new representation, for $d_2 = 1$ and $f_2 = 0$, we can write

$$C_s(\lambda d_1 + \bar{\lambda}, s) \le \lambda C_s(d_1, s). \tag{13}$$

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 Section V).

D. Parallel Concatenation of More Than Two Channels

So far, we considered the parallel concatenation of two independent deletion channels which is useful for improving upon the existing upper bounds. However, we can also consider the parallel concatenation of more than two deletion channels by considering a different "random" fragmentation process. If we define the deletion channel $\mathcal{C}$ as a parallel concatenation of P independent deletion channels $\mathcal{C}_p$ with deletion probability d_p ($p = \{1,\ldots,P\}$) where each input bit is transmitted with probability λ_p over $\mathcal{C}_p$, and modify the definition of $\boldsymbol{F}_y$ such that $f_y[i] \in \{1,\ldots,P\}$ denotes the index of the channel the i-th bit is coming from, then for $d = \sum_{p=1}^{P} \lambda_p d_p$, we have

$$C_2(d) \le \sum_{p=1}^{P} \lambda_p C_2(d_p) + \bar{d}\log(\bar{d}) - \sum_{p=1}^{P} \lambda_p \bar{d}_p \log(\lambda_p \bar{d}_p), \tag{14}$$

where $\sum_{p=1}^{P} \lambda_p = 1$. Note, however, that the above inequality does not result in any tighter upper bounds on the deletion channel capacity than the one obtained by considering the parallel concatenation of only two independent deletion channels.

IV. Non-Binary Deletion Channels

We now switch gears and consider non-binary deletion channels. The main objective is to explore a different (deterministic) input-output fragmentation process and use it to obtain the first non-trivial capacity upper bounds for non-binary deletion channels. Before proceeding with the main result, we provide a discussion on BAA based upper bounds for non-binary input deletion channels in the following subsection. Then, the subsequent subsections describe the fragmentation process under consideration for $2K$-ary deletion channel and study the case of general Q-ary input deletion channels.

A. Discussion on BAA Based Upper Bounds

One approach to derive upper bounds on the Q-ary deletion channel capacity is to modify the numerical approaches in [5] and [13] in which the decoder (and possibly the encoder) of the deletion channel is provided with some side information about the deletion process and the capacity (or an upper bound on the capacity) of the resulting genie-aided channel is computed by the Blahut-Arimoto algorithm. Although this approach is useful for binary input channels (even when other impairments such as insertions and substitutions are considered [7]), for the non-binary case, running the BAA for large values of Q is not computationally feasible. For instance,

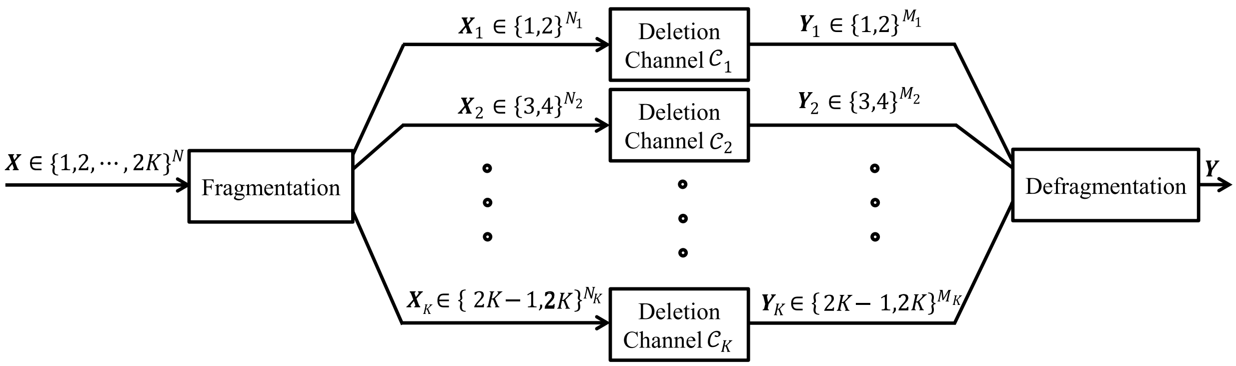

Fig. 2. Fragmentation of the $2K$-ary deletion channel into K independent binary input deletion channels.

one of the upper bounds in [5] is obtained by computing the capacity of the binary deletion channel with finite transmission length $L = 17$. Obviously, by increasing the alphabet size Q, the maximum possible value of L in running the BAA algorithm decreases. Therefore, to achieve meaningful upper bounds, L needs to be increased which makes the numerical computations infeasible.

The main contribution of the present section is that we are able to relate the capacity of the Q-ary deletion channel to the capacity of the lower order deletion channels through an inequality which enables us to upper bound the Q-ary deletion channel capacity and avoid the computationally prohibitive BAA directly for the Q-ary deletion channel.

B. A Different Look at the $2K$-Ary Deletion Channel

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, as shown in Fig. 2, 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. $\boldsymbol{X}_k$ and $\boldsymbol{Y}_k$ denote the input and output sequences of the k-th channel, respectively, and N_k and M_k denote the length of $\boldsymbol{X}_k$ and $\boldsymbol{Y}_k$, respectively.

To be able to relate the mutual information between the input and output sequences of the $2K$-ary deletion channel, $I(\boldsymbol{X};\boldsymbol{Y})$, with the mutual information between the input and output sequences of the considered binary deletion channels, $I(\boldsymbol{X}_k;\boldsymbol{Y}_k)$, we define two new random vectors $\boldsymbol{F}_x = (f_x[1],\ldots,f_x[N])$ and $\boldsymbol{F}_y = (f_y[1],\ldots,f_y[M])$ where $f_x[n] \in \{1,\ldots,K\}$ and $f_y[m] \in \{1,\ldots,K\}$ denote the label of the subchannel the n-th input symbol and m-th output symbol belong to, respectively. Clearly, by knowing $\boldsymbol{X}$, one can determine $(\boldsymbol{X}_1,\ldots,\boldsymbol{X}_K,\boldsymbol{F}_x)$ and by knowing $(\boldsymbol{X}_1,\ldots,\boldsymbol{X}_K,\boldsymbol{F}_x)$ can determine $\boldsymbol{X}$. The same situation holds for $\boldsymbol{Y}$ and $(\boldsymbol{Y}_1,\ldots,\boldsymbol{Y}_K,\boldsymbol{F}_y)$. Therefore, we have

$$\begin{aligned} I(\boldsymbol{X};\boldsymbol{Y}) &= I(\boldsymbol{X}_1,\ldots,\boldsymbol{X}_K,\boldsymbol{F}_x;\boldsymbol{Y}_1,\ldots,\boldsymbol{Y}_K,\boldsymbol{F}_y) \\ &= \sum_{k=1}^{K} I_k + I_F, \end{aligned} \tag{15}$$

where $I_k = I(\boldsymbol{X}_1,\ldots,\boldsymbol{X}_K,\boldsymbol{F}_x;\boldsymbol{Y}_k|\boldsymbol{Y}_1,\ldots,\boldsymbol{Y}_{k-1})$ and

$$I_F = I(\boldsymbol{X}_1,\ldots,\boldsymbol{X}_K,\boldsymbol{F}_x;\boldsymbol{F}_y|\boldsymbol{Y}_1,\ldots,\boldsymbol{Y}_K). \tag{16}$$

In Section IV-C, we will derive upper bounds on I_k and I_F which will enable us to relate the non-binary and binary deletion channels capacities, and will lead to the main result of the paper.

C. A Novel Upper Bound on $C_{2K}(d)$

As discussed in Section IV-B, a $2K$-ary deletion channel can be considered as a parallel concatenation of K independent binary deletion channels. This new look at a $2K$-ary deletion channel enables us to relate the $2K$-ary deletion channel capacity to the binary deletion channel capacity with the same deletion error probability as given in the following theorem.

Theorem 2: Let $C_{2K}(d)$ denote the capacity of a $2K$-ary i.i.d. deletion channel with deletion probability d, then

$$C_{2K}(d) \leq C_2(d) + (1-d)\log(K). \tag{17}$$

As given in (15), the mutual information $I(\boldsymbol{X};\boldsymbol{Y})$ can be expanded in terms of several other mutual information terms, I_k for $k \in \{1,\ldots,K\}$ and I_F. To prove the theorem, we first derive upper bounds for I_k and I_F in the following two lemmas.

Lemma 3: For any input distribution $P(\boldsymbol{X}_1,\ldots,\boldsymbol{X}_K,\boldsymbol{F}_x)$, the mutual information I_k given in (15) can be upper bounded by

$$I_k \leq E\{\boldsymbol{N}_k\}C_2(d) + 2\log(N+1).$$

Proof: For I_k, since $P(\boldsymbol{Y}_k|\boldsymbol{Y}_1,\ldots,\boldsymbol{Y}_{k-1},\boldsymbol{X}_k) = P(\boldsymbol{Y}_k|\boldsymbol{X}_k)$ and $P(\boldsymbol{Y}_k|\boldsymbol{X}_1,\ldots,\boldsymbol{X}_K,\boldsymbol{F}_x,\boldsymbol{Y}_1,\ldots,\boldsymbol{Y}_{k-1}) = P(\boldsymbol{Y}_k|\boldsymbol{X}_k)$, we can write

$$\begin{aligned} I_k &= I(\boldsymbol{X}_k;\boldsymbol{Y}_k|\boldsymbol{Y}_1,\ldots,\boldsymbol{Y}_{k-1}) \\ &\quad + I(\boldsymbol{X}_1,\ldots,\boldsymbol{X}_{k-1},\boldsymbol{X}_{k+1},\ldots,\boldsymbol{X}_K, \\ &\qquad \boldsymbol{F}_x;\boldsymbol{Y}_k|\boldsymbol{Y}_1,\ldots,\boldsymbol{Y}_{k-1},\boldsymbol{X}_k) \\ &= I(\boldsymbol{X}_k;\boldsymbol{Y}_k|\boldsymbol{Y}_1,\ldots,\boldsymbol{Y}_{k-1}) \\ &= H(\boldsymbol{Y}_k|\boldsymbol{Y}_1,\ldots,\boldsymbol{Y}_{k-1}) - H(\boldsymbol{Y}_k|\boldsymbol{Y}_1,\ldots,\boldsymbol{Y}_{k-1},\boldsymbol{X}_k) \\ &= H(\boldsymbol{Y}_k) - I(\boldsymbol{Y}_1,\ldots,\boldsymbol{Y}_{k-1};\boldsymbol{Y}_k) - H(\boldsymbol{Y}_k|\boldsymbol{X}_k) \\ &\leq I(\boldsymbol{X}_k;\boldsymbol{Y}_k). \end{aligned} \tag{18}$$

Furthermore, for $I(\boldsymbol{X}_k;\boldsymbol{Y}_k)$, we have (see Appendix C)

$$I(\boldsymbol{X}_k;\boldsymbol{Y}_k) \leq E\{\boldsymbol{N}_k\}C_2(d) + 2\log(N+1).$$

Finally, by substituting the above inequality into (18), the proof follows. ■

Lemma 4: For any input distribution, the mutual information I_F given in (16) can be upper bounded by

$$I_F \leq N(1-d)\log(K).$$

Proof: Using the definition of the mutual information, we can write

$$\begin{aligned} I_F &= H(\boldsymbol{F}_y|\boldsymbol{Y}_1,\ldots,\boldsymbol{Y}_K) \\ &\quad -H(\boldsymbol{F}_y|\boldsymbol{Y}_1,\ldots,\boldsymbol{Y}_K,\boldsymbol{X}_1,\ldots,\boldsymbol{X}_K,\boldsymbol{F}_x) \\ &\leq H(\boldsymbol{F}_y|\boldsymbol{Y}_1,\ldots,\boldsymbol{Y}_K) \\ &\leq H(\boldsymbol{F}_y|\boldsymbol{M}_1,\ldots,\boldsymbol{M}_K), \end{aligned} \tag{19}$$

where the last inequality follows since $(\boldsymbol{M}_1,\ldots,\boldsymbol{M}_K)$ is a function of $(\boldsymbol{Y}_1,\ldots,\boldsymbol{Y}_K)$, i.e., $H(\boldsymbol{M}_1,\ldots,\boldsymbol{M}_K|\boldsymbol{Y}_1,\ldots,\boldsymbol{Y}_K) = 0$. For fixed m_k with $\sum_{k=1}^{K} m_k = m$, there are $\binom{m}{m_1,\ldots,m_K}$ possibilities for $\boldsymbol{F}_y$ leading to $H(\boldsymbol{F}_y|\boldsymbol{M}_1 = m_1,\ldots,\boldsymbol{M}_K = m_K) \leq \log\binom{m}{m_1,\ldots,m_K}$. It follows from the inequality (see Appendix A)

$$\log\binom{m}{m_1,\ldots,m_K} \leq m\log(m) - \sum_{k=1}^{K} m_k\log(m_k), \tag{20}$$

that $H(\boldsymbol{F}_y|\boldsymbol{M}_1 = m_1,\ldots,\boldsymbol{M}_K = m_K) \leq m\log(m) - \sum_{k=1}^{K} m_k\log(m_k)$. Since, for $\sum_{k=1}^{K} m_k = m$,

$$g([m_1,\ldots,m_k]) = m\log(m) - \sum_{k=1}^{K} m_k\log(m_k)$$

is a concave function of $[m_1,\ldots,m_K]$ (see Appendix B), employing the Jensen's inequality yields

$$\begin{aligned} I_F &\leq \left(\sum_{k=1}^{K} E\{\boldsymbol{M}_k\}\right)\log\left(\sum_{k=1}^{K} E\{\boldsymbol{M}_k\}\right) \\ &\quad -\sum_{k=1}^{K} E\{\boldsymbol{M}_k\}\log(E\{\boldsymbol{M}_k\}). \end{aligned}$$

On the other hand, due to the fact that $\mathcal{C}_k$ are i.i.d. binary input deletion channels, we have $E\{\boldsymbol{M}_k\} = N\bar{d}\alpha_k$ where α_k's depend on the input distribution $P(\boldsymbol{X})$ and $\sum_{k=1}^{K}\alpha_k = 1$. Hence, we obtain

$$\begin{aligned} I_F &\leq N\bar{d}\left(\log\left(N\bar{d}\right) - \sum_{k=1}^{K}\alpha_k\log\left(N\bar{d}\alpha_k\right)\right) \\ &= -N\bar{d}\sum_{k=1}^{K}\alpha_k\log\alpha_k = N\bar{d}H(\alpha_1,\ldots,\alpha_K) \\ &\leq N\bar{d}\log(K), \end{aligned} \tag{21}$$

which concludes the proof. ■

Proof of Theorem 2: Substituting the results of Lemmas 3 and 4 into (15), we obtain

$$\begin{aligned} I(\boldsymbol{X};\boldsymbol{Y}) &\leq E_{\boldsymbol{N}_1,\ldots,\boldsymbol{N}_K}\left\{\sum_{k=1}^{K}\boldsymbol{N}_k\right\}C_2(d) \\ &\quad + 2K\log(N+1) + N\bar{d}\log(K) \\ &= NC_2(d) + 2K\log(N+1) + N\bar{d}\log(K), \end{aligned}$$

where we have used the fact that $\sum_{k=1}^{K}\boldsymbol{N}_k = N$ independent of the input distribution $P(\boldsymbol{X})$. Since the above inequality holds for any input distribution $P(\boldsymbol{X})$ and any value of N, we can write

$$\begin{aligned} C_{2K}(d) &= \lim_{N\to\infty}\max_{P(\boldsymbol{X})}\frac{1}{N}I(\boldsymbol{X};\boldsymbol{Y}) \\ &\leq C_2(d) + \bar{d}\log(K), \end{aligned}$$

which concludes the proof.

D. Generalization of the Result to Q-Ary Input Deletion Channels

So far, we have focused on $2K$-ary input deletion channels and provided a new upper bound on their capacity in terms of the binary deletion channel capacity. In this section, we generalize the results to the Q-ary input deletion channels for arbitrary values of Q by providing upper bounds on the Q-ary input deletion channel capacity in terms of the capacity of the lower order deletion channels, e.g., the capacity of the 5-ary input deletion channel can be upper bounded in terms of the capacity of the ternary and binary input deletion channels.

Let us define $\mathcal{S} = \{1,\ldots,Q\}$ as the set of input symbols to the Q-ary input deletion channel. There are different possibilities to fragment the given set $\mathcal{S}$ into subsets with smaller length. For example, there are $\binom{Q}{q_1,\ldots,q_P}$ possibilities to fragment $\mathcal{S}$ into P subsets $\mathcal{S}_p$ ($p \in \{1,\ldots,P\}$) where $|\mathcal{S}_p| = q_p$ and $\sum_{p=1}^{P} q_p = Q$. Similar to the case of $2K$-ary deletion channels, for each possible fragmentation of the input symbols into P subsets, we can define a Q-ary input channel as a parallel concatenation of P parallel independent deletion channels where the symbols of the subset $\mathcal{S}_p$ travel through the p-th deletion channel. Therefore, there are $J = \binom{Q}{q_1,\ldots,q_P}$ possibilities for considering the Q-ary input deletion channel as a parallel concatenation of P independent deletion channels.

Let us first focus on a specific input distribution $P(\boldsymbol{X})$ and a specific fragmentation process, the jth one, for which the subset $\mathcal{S}_{p,j}$ of length q_p represents the set of input symbols to the deletion channel $\mathcal{C}_{p,j}$ with input sequence of $\boldsymbol{X}_{p,j}$ and output sequence of $\boldsymbol{Y}_{p,j}$. When transmitting N symbols over the new channel, as defined in Section IV-B, the fragmentation and defragmentation processes can be represented by $\boldsymbol{F}_x = (f_x[1],\ldots,f_x[N])$ and $\boldsymbol{F}_y = (f_y[1],\ldots,f_y[N])$, respectively. Similar to (15), by defining $I_{p,j} = I(\boldsymbol{X}_{1,j},\ldots,\boldsymbol{X}_{p,j},\boldsymbol{F}_x;\boldsymbol{Y}_{p,j}|\boldsymbol{Y}_{1,j},\ldots,\boldsymbol{Y}_{p-1,j})$ and I_F as in (16), we can write

$$I(\boldsymbol{X};\boldsymbol{Y}) = \sum_{p=1}^{P} I_{p,j} + I_F. \tag{22}$$

For $I_{p,j}$, by generalizing the result of Lemma 3 to the new defined fragmented channel, we obtain

$$I_{p,j} \leq E\{\boldsymbol{N}_{p,j}\}C_{q_p}(d) + 2\log(N+1), \tag{23}$$

where $N_{p,j}$ denotes the length of the input sequence to the channel $\mathcal{C}_{p,j}$ and $E\{\boldsymbol{N}_{p,j}\}$ depends on the input distribution $P(\boldsymbol{X})$. Obviously, the result of Lemma 4 also holds for the newly defined channel, i.e.,

$$I_F \leq N\bar{d}\log(P). \tag{24}$$

Substituting (23) and (24) into (22), for a given input distribution $P(\boldsymbol{X})$ and a specific j, we obtain

$$I(\boldsymbol{X};\boldsymbol{Y}) \leq \sum_{p=1}^{P} E\{\boldsymbol{N}_{p,j}\}C_{q_p}(d) + 2P\log(N+1) + N\bar{d}\log(P). \tag{25}$$

Since the above inequality is valid for any possible fragmentation process, i.e., all j, by averaging the right hand side of the inequality for all possible fragmentation processes, we arrive at

$$I(\boldsymbol{X};\boldsymbol{Y}) \leq \frac{1}{J}\sum_{p=1}^{P}\left(\sum_{j=1}^{J} E\{\boldsymbol{N}_{p,j}\}\right)C_{q_p}(d) + N\bar{d}\log(P) + 2P\log(N+1). \tag{26}$$

Let us consider a specific input symbol $s \in \{1,\ldots,Q\}$ and define $N_s(X)$ as the number of times s occurs in the input sequence X ($\sum_{s\in\mathcal{S}} N_s(X) = N$ for any possible input sequence X). Furthermore, among all possible $J = \binom{Q}{q_1,\ldots,q_P}$ fragmentation processes, there are $\binom{Q-1}{q_1,\ldots,q_p-1,\ldots,q_P}$ possible fragmentations for which s belongs to the p-th subset of length q_p. Therefore, we can write

$$\begin{aligned}\frac{1}{J}\sum_{j=1}^{J} E\{\boldsymbol{N}_{p,j}\} &= \frac{1}{J}E\left\{\sum_{j=1}^{J}\sum_{s\in\mathcal{S}_{p,j}}\boldsymbol{N}_s(\boldsymbol{X})\right\}\\ &= \frac{1}{J}\binom{Q-1}{q_1,\ldots,q_p-1,\ldots,q_P}E\left\{\sum_{s\in\mathcal{S}}N_s(\boldsymbol{X})\right\}\\ &= N\frac{\binom{Q-1}{q_1,\ldots,q_p-1,\ldots,q_P}}{J}\\ &= N\frac{q_p}{Q}\end{aligned} \tag{27}$$

which is independent of the specific channel input (hence it is independent of the input distribution used). Substituting this result into (26), we obtain

$$I(\boldsymbol{X};\boldsymbol{Y}) \leq N\sum_{p=1}^{P}\frac{q_p}{Q}C_{q_p}(d) + 2P\log(N+1) + N\bar{d}\log(P).$$

Since the above inequality holds for any input distribution $P(\boldsymbol{X})$, it holds for the capacity achieving input distribution as well. Therefore, for the capacity achieving input distribution, by dividing both sides of the above expression by N and letting N go to infinity, we arrive at

$$C_Q(d) \leq \sum_{p=1}^{P}\frac{q_p}{Q}C_{q_p}(d) + \bar{d}\log(P). \tag{28}$$

As an example, for the $2K+1-$ary input deletion channel, by considering parallel concatenation of $K-1$ binary

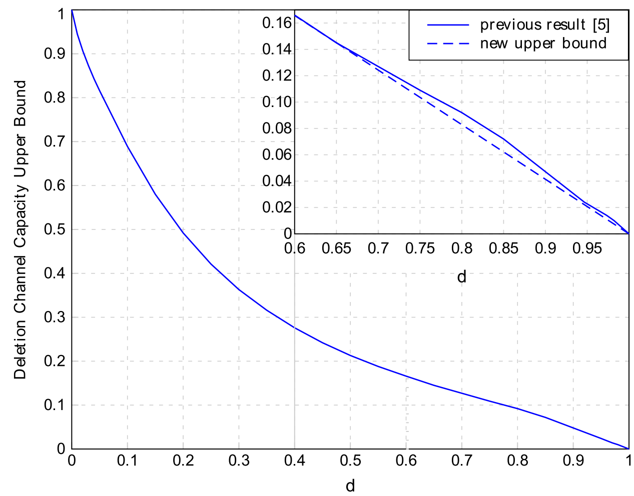

Fig. 3. Upper bounds on the i.i.d. deletion channel capacity.

deletion channels and a ternary input deletion channel, we obtain

$$C_{2K+1}(d) \leq \frac{2(K-1)}{2K+1}C_2(d) + \frac{3}{2K+1}C_3(d) + \bar{d}\log(K). \tag{29}$$

The newly derived capacity upper bound (28) provides some upper bounds on $C_Q(d)$ for arbitrary values of Q in terms of the capacity of the lower order input deletion channels, $C_{q_p}(d)$ for $p \in \{1,\ldots,P\}$, as long as $\sum_{p=1}^{P} q_p = Q$. As stated in Section IV-A the tightest available upper bounds on $C_2(d)$ are provided in [5] in which BAA is employed to compute the capacity of some genie-aided channels which are upper bounds on $C_2(d)$, however running BAA to obtain upper bounds on $C_Q(d)$ is infeasible for large values of Q. Running BAA for ternary and even quaternary input deletion channels is still feasible. Specifically, one can obtain upper bounds on $C_3(d)$ by employing the approach used in [5] in a straightforward manner. This bound can then be employed in (29) to come up with improved capacity upper bounds for non-binary deletion channels with odd number of inputs.

V. Examples of the Newly Derived Capacity Upper Bounds

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

A. Binary Deletion Channel

An interesting application of the result (3) 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 not convex for $d \geq 0.65$ as shown in Fig. 3 (with values taken from the boldfaced values in [5, Table IV]). As clarified in the

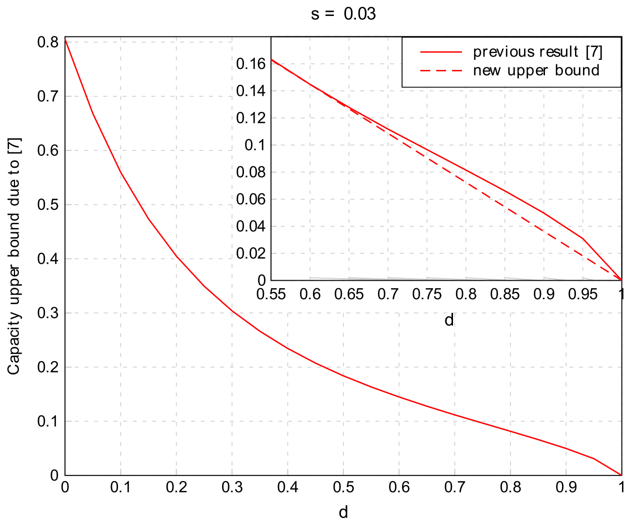

Fig. 4. Upper bounds on the deletion/substitution channel capacity for $s = 0.03$.

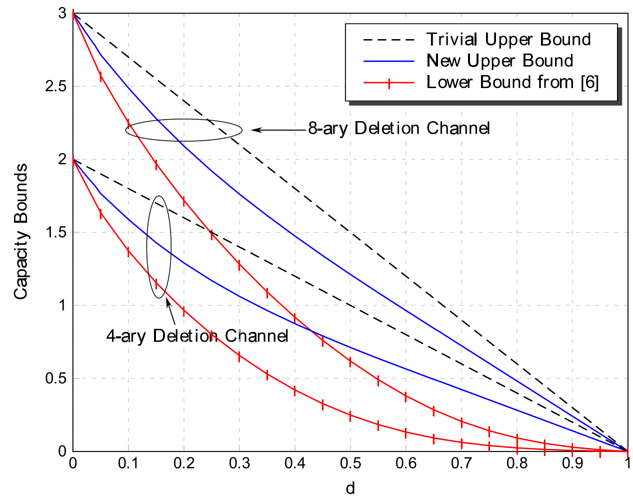

Fig. 5. Comparison among the new upper bound (17), the lower bound (2) and the trivial erasure channel upper bound for the 4-ary and 8-ary deletion channels.

table, the best known values for small d are due to [13], for a wide range (up to $d \sim 0.8$) are due to the "fourth version" of the upper bound (named C_4 in [5] which we refer to it as $C_4^{[5]}$), 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 $C_2(1 - 0.35\lambda) \leq \lambda C_2(0.65) \leq \lambda C_4^{[5]}(0.65)$ with $0 \leq \lambda \leq 1$. That is, we have $C_2(d) \leq 0.4143(1 - d)$ for $d \in (0.65, 1)$. This is illustrated in Fig. 3.

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 paper, we can consider the capacity of the deletion/substitution channel. The best known capacity upper bound for this case is given in [7]. For example [7, Fig. 1] presents several upper bounds for fixed $s = 0.03$ which show that the 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 as illustrated in Fig. 4.

B. Non-Binary Deletion Channels

As stated earlier, a trivial upper bound on the capacity of the Q-ary deletion channel is given by $(1 - d)\log(Q)$ which is the capacity of the Q-ary erasure channel. We have shown in the previous section that substituting any upper bound on $C_{q_p}(d)$ into (28) results in an upper bound on the Q-ary deletion channel capacity. For $Q = 2K$, by employing $C_2(d) \leq (1 - d)$, which is the trivial upper bound on the binary deletion channel capacity, the erasure channel upper bound on the $2K$-ary deletion channel capacity is obtained. Therefore, any upper bound tighter than $(1 - d)$ on the binary deletion channel capacity gives an upper bound tighter than $\log(2K)(1 - d)$ on the $2K$-ary deletion channel capacity. The amount of improvement is $1-d-C_2^{UB}(d)$, where C_2^{UB} denotes the upper bound on the binary deletion channel capacity.

As it is shown in [14], $(1 - d)\log(Q) - 1 \leq C_Q(d) \leq (1 - d)\log(Q)$, where the lower bound is implied from (1), therefore the existing trivial upper and lower bounds are tight enough for asymptotically large values of Q, and i.i.d. distributed input sequences are sufficient to achieve the capacity. However, the importance of the result in (28) (Theorem 2 for the special case of $2K$-ary deletion channel) is for moderate values of Q, where the amount of improvement in closing the gap between the existing upper and lower bounds is significant.

To demonstrate the improvement over the trivial erasure channel upper bound, we compare the upper bound $C_{2K}(d) \leq C_2^{UB}(d) + (1 - d)\log(K)$ with the erasure channel upper bound $\log(2K)(1 - d)$ and the tightest existing lower bound (2) (from [6]) in Fig. 5 for 4-ary and 8-ary deletion channels. Here we utilize the binary deletion channel capacity upper bounds $C_2^{UB}(d)$ from [5, Table III] for $d \leq 0.65$ and the upper bound $C_2(d) \leq 0.4143(1 - d)$ provided in Section III for $d \geq 0.65$.

Another implication of the result in Theorem 2 is in studying the asymptotic behavior of the $2K$-ary deletion channel capacity for $d \to 0$. It is shown in [15] that

$$C_2(d) = 1 + d\log(d) - A_1 d + A_2 d^2 + O(d^{3-\epsilon}), \quad (30)$$

for small d and any $\epsilon > 0$ with $A_1 \approx 1.15416377$, $A_2 \approx 1.78628364$ and $O(.)$ denoting the standard Landau (big-O) notation. Employing this result in (17), leads to an upper bound expansion for small values of d as

$$\begin{aligned} C_{2K}(d) \leq\ & 1 + d\log(d) - (A_1 + \log(K))d + A_2 d^2 \\ & + \log(K) + O(d^{3-\epsilon}). \end{aligned} \quad (31)$$

In Fig. 6, we compare the above upper bound (by ignoring the $O(d^{3-\epsilon})$ term) which serves as an estimate, with the lower bound (2) for $d \leq 0.1$. We observe that by employing the

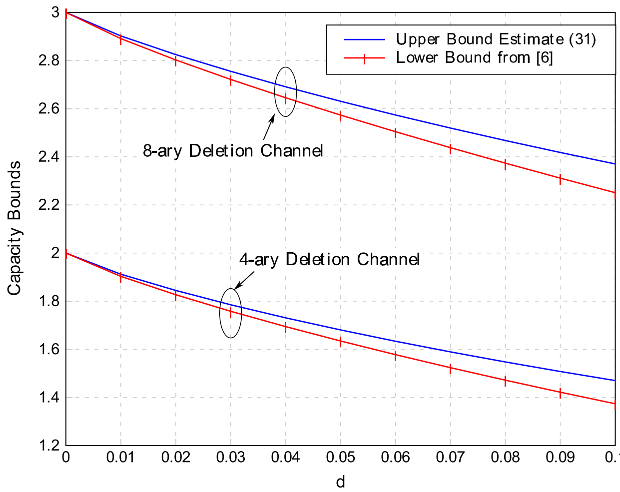

Fig. 6. Comparison between the upper bound (31) (ignoring the $O(d^{3-\epsilon})$ term) and the lower bound (2).

capacity expansion (30) in (17), a good characterization for the asymptotic behavior of the $2K$-ary deletion channel capacity is obtained as $d \to 0$.

VI. Conclusions

In this paper, we present a new upper bound on the capacity of the binary-input deletion channel and show that it improves on all previous results for $d \geq 0.65$. We also introduce the first non-trivial upper bound on the non-binary input deletion channel capacity. For both binary and non-binary input cases, the approach is based on fragmentation of the input symbol sequences into smaller subsequences which travel through independent deletion channels and the surviving symbols are combined without changing the order in the original sequence. For the binary case, 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 deletion channels with non-binary inputs, a deterministic fragmentation of the input sequence is considered which results in capacity upper bounds in terms of lower order input deletion channel capacities. For instance, the capacity of the non-binary deletion channel is upper bounded in terms of the binary deletion channel capacity. 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 an i.i.d. deletion channel, we prove 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 $d \to 1$. Furthermore, for non-binary deletion channels, the provided upper bound results in tighter characterizations than the trivial erasure channel upper bound for the entire range of deletion probabilities. We also describe a generalization of the result to the case of deletion/substitution channels and provide a tighter capacity upper bound for this case as well.

Appendix A
Proof of Inequalities (8) and (20)

It follows from the inequality $\log\binom{m}{m_1} \leq mH_b(\frac{m_1}{m}) = m\log(m) - m_1\log(m_1) - (m-m_1)\log(m-m_1)$ given in [16, p. 353] that

$$\begin{aligned}
\log\binom{m}{m_1,\ldots,m_K} &= \sum_{j=1}^{K-1}\log\binom{m-\sum_{k=1}^{j-1}m_k}{m_j}\\
&\leq \sum_{j=1}^{K-1}\left(m-\sum_{k=1}^{j-1}m_k\right)\log\left(m-\sum_{k=1}^{j-1}m_k\right)\\
&\quad -m_j\log m_j - \sum_{j=1}^{K-1}\left(m-\sum_{k=1}^{j}m_k\right)\\
&\quad \times\log\left(m-\sum_{k=1}^{j}m_k\right)\\
&= m\log(m) - \sum_{k=1}^{K}m_k\log(m_k).
\end{aligned}$$

Appendix B
Concavity of $g([m_1,\ldots,m_k])$

For the Hessian of $g([m_1,\ldots,m_k])$, we have

$$\begin{aligned}
\nabla^2 g([m_1,\ldots,m_k]) &= \frac{1}{\sum_{k=1}^{K}m_k}\mathbf{1}\mathbf{1}^T\\
&\quad -diag\left\{\frac{1}{m_1},\ldots,\frac{1}{m_K}\right\},
\end{aligned}$$

where $\mathbf{1}$ is an all one vector of length K, i.e., $\mathbf{1}=[1,\ldots,1]^T$, and $diag\left\{\frac{1}{m_1},\ldots,\frac{1}{m_K}\right\}$ denotes a diagonal matrix whose k-th diagonal element is $\frac{1}{m_k}$. Furthermore, by defining $\boldsymbol{a}=[a_1,\ldots,a_K]$, we can write

$$\begin{aligned}
\boldsymbol{a}\nabla^2 g\boldsymbol{a}^T &= \frac{(\sum_{k=1}^{K}a_k)^2}{\sum_{k=1}^{K}m_k} - \sum_{k=1}^{K}\frac{a_k^2}{m_k}\\
&= \frac{1}{\sum_{k=1}^{K}m_k}\Bigg(\sum_{k=1}^{K}a_k^2 + 2\sum_{k=1}^{K-1}\sum_{j=k+1}^{K}a_ka_j\\
&\quad -\sum_{k=1}^{K}a_k^2 - \sum_{k=1}^{K}\frac{\sum_{j\neq k}m_j}{m_k}a_k^2\Bigg)\\
&= \frac{1}{\sum_{k=1}^{K}m_k}\sum_{k=1}^{K-1}\sum_{j=k+1}^{K}\left(2a_ka_j - \frac{m_j}{m_k}a_k^2 - \frac{m_k}{m_j}a_j^2\right)\\
&= \frac{-1}{\sum_{k=1}^{K}m_k}\sum_{k=1}^{K-1}\sum_{j=k+1}^{K}\frac{m_j}{m_k}(a_k - \frac{m_k}{m_j}a_j)^2,
\end{aligned}$$

which is negative for all $m_k,\ m_j > 0$. Therefore, $\nabla^2 g([m_1,\ldots,m_k])$ is a negative semi-definite matrix and as a result $g([m_1,\ldots,m_k])$ is a concave function of $[m_1,\ldots,m_k]$.

Appendix C
Upper Bounding $I(\boldsymbol{X}_i;\boldsymbol{Y}_i)$

For $I(\boldsymbol{X}_i;\boldsymbol{Y}_i)$, we can write

$$\begin{aligned}
I(\boldsymbol{X}_i;\boldsymbol{Y}_i) &= I(\boldsymbol{X}_i;\boldsymbol{Y}_i,\boldsymbol{N}_i) - I(\boldsymbol{X}_i;\boldsymbol{N}_i|\boldsymbol{Y}_i)\\
&= I(\boldsymbol{X}_i;\boldsymbol{Y}_i|\boldsymbol{N}_i) + I(\boldsymbol{X}_i;\boldsymbol{N}_i) - I(\boldsymbol{X}_i;\boldsymbol{N}_i|\boldsymbol{Y}_i)
\end{aligned}$$

$$
\begin{aligned}
&\le I(\boldsymbol{X}_i; \boldsymbol{Y}_i|\boldsymbol{N}_i) + H(\boldsymbol{N}_i) \\
&\le I(\boldsymbol{X}_i; \boldsymbol{Y}_i|\boldsymbol{N}_i) + \log(N+1) \\
&= \sum_{N_i=0}^{N} P(\boldsymbol{N}_i = N_i) I(\boldsymbol{X}_i; \boldsymbol{Y}_i|\boldsymbol{N}_i = N_i) \\
&\quad + \log(N+1), \qquad (32)
\end{aligned}
$$

where in deriving the first inequality we have used the facts that $H(\boldsymbol{N}_i|\boldsymbol{X}_i) = 0$ and $I(\boldsymbol{X}_i; \boldsymbol{N}_i|\boldsymbol{Y}_i) \ge 0$, and in deriving the second equality the fact that

$$
\begin{aligned}
H(\boldsymbol{N}_i) &= -\sum_{n=0}^{N} \binom{N}{n} \lambda^n \bar{\lambda}^{N-n} \log\left(\binom{N}{n} \lambda^n \bar{\lambda}^{N-n}\right) \\
&\le \log(N+1). \qquad (33)
\end{aligned}
$$

Furthermore, as it is shown in [5], for a finite length transmission over the deletion channel, the mutual information rate between the transmitted and received sequences can be upper bounded in terms of the capacity of the channel after adding some appropriate term, which can be spelled out as [5, eq. (39)]

$$
I(\boldsymbol{X}_i; \boldsymbol{Y}_i|\boldsymbol{N}_i = N_i) \le N_i C_2(d_i) + H(\boldsymbol{D}_i|\boldsymbol{N}_i = N_i), \quad (34)
$$

where D_i denotes the number of deletions through the transmission of N_i bits over the i-th channel and

$$
\begin{aligned}
H(\boldsymbol{D}_i|\boldsymbol{N}_i = N_i) \\
&= -\sum_{n=0}^{N_i} \binom{N_i}{n} d_i^n \bar{d}_i^{N_i-n} \log\left(\binom{N_i}{n} d_i^n \bar{d}_i^{N_i-n}\right) \\
&\le \log(N_i+1).
\end{aligned}
$$

Substituting (34) into (32), we have

$$
\begin{aligned}
I(\boldsymbol{X}_i; \boldsymbol{Y}_i) &\le \sum_{N_i=0}^{N} P(\boldsymbol{N}_i = N_i) \left(N_i C_2(d_i) + \log(N_i+1)\right) \\
&\quad + \log(N+1) \\
&\le E\{\boldsymbol{N}_i\} C_2(d_i) + 2\log(N+1),
\end{aligned}
$$

where the last inequality results since $\log(N_i+1) \le \log(N+1)$.

Acknowledgments

The authors would like to thank Marco Dalai for his insightful comments on the paper.

References

[1] R. L. Dobrushin, "Shannon's theorems for channels with synchronization errors," *Problems Inf. Transmiss.*, 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, Oct. 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, Jun. 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, Mar. 2006.

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

[8] M. Rahmati and T. M. Duman, "Bounds on the capacity of 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)*, Jul./Aug. 2011, pp. 499–502.

[10] M. Mitzenmacher, "A survey of results for deletion channels and related synchronization channels," *Probab. Surv.*, vol. 6, pp. 1–33, Jun. 2009.

[11] R. G. Gallager, "Sequential decoding for binary channels with noise and synchronization errors," Lincoln Lab., Massachusetts Inst. Technol., Cambridge, MA, USA, Tech. Rep., Oct. 1961.

[12] 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.

[13] S. Diggavi, M. Mitzenmacher, and H. Pfister, "Capacity upper bounds for deletion channels," in *Proc. Int. Symp. Inf. Theory (ISIT)*, Jun. 2007, pp. 1716–1720.

[14] H. Mercier, V. Tarokh, and F. Labeau, "Bounds on the capacity of discrete memoryless channels corrupted by synchronization and substitution errors," *IEEE Trans. Inf. Theory*, vol. 58, no. 7, pp. 4306–4330, Jul. 2012.

[15] Y. Kanoria and A. Montanari, "Optimal coding for the binary deletion channel with small deletion probability," *IEEE Trans. Inf. Theory*, vol. 59, no. 10, pp. 6192–6219, Oct. 2013.

[16] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York, NY, USA: Wiley, 2006.

[17] M. Rahmati and T. M. Duman, "An upper bound on the capacity of non-binary deletion channels," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2013, pp. 2940–2944.

[18] M. Rahmati and T. M. Duman, "An improvement of the deletion channel capacity upper bound," in *Proc. 51st Annu. Allerton Conf. Commun., Control Comput.*, Oct. 2013, pp. 1221–1225.

Mojtaba Rahmati received his B. S. degree in electrical engineering in 2007 from University of Tehran, Iran, his M. S. degree in telecommunication systems in 2009 from Sharif University of Technology, Tehran, Iran and his PhD in electrical engineering in 2013 from Arizona State University, Tempe, AZ. He is currently a research engineer at Samsung Modem R&D. His research interests include wireless and mobile communication, digital communications, digital signal processing and information theory.

Tolga M. Duman (S'95–M'98–SM'03–F'11) is a Professor of Electrical and Electronics Engineering Department of Bilkent University in Turkey, and is on leave from the School of ECEE at Arizona State University. He received the B.S. degree from Bilkent University in Turkey in 1993, M.S. and Ph.D. degrees from Northeastern University, Boston, in 1995 and 1998, respectively, all in electrical engineering. Prior to joining Bilkent University in September 2012, he has been with the Electrical Engineering Department of Arizona State University first as an Assistant Professor (1998-2004), then as an Associate Professor (2004-2008), and starting August 2008 as a Professor. Dr. Duman's current research interests are in systems, with particular focus on communication and signal processing, including wireless and mobile communications, coding/modulation, coding for wireless communications, data storage systems and underwater acoustic communications.

Dr. Duman is a Fellow of IEEE, a recipient of the National Science Foundation CAREER Award and IEEE Third Millennium medal. He served as an editor for IEEE Trans. on Wireless Communications (2003-08), IEEE Trans. on Communications (2007-2012) and IEEE Online Journal of Surveys and Tutorials (2002-07). He is currently the coding and communication theory area editor for IEEE Trans. on Communications (2011-present) and an editor for Elsevier Physical Communications Journal (2010-present).