

An Upper Bound on the Capacity of non-Binary Deletion Channels

Mojtaba Rahmati

School of Electrical, Computer and Energy Engineering
Arizona State University, Tempe, AZ 85287-5706, USA
Email: mojtaba@asu.edu

Tolga M. Duman

Department of Electrical and Electronics Engineering
Bilkent University, Bilkent, Ankara, 06800, Turkey
Email: duman@ee.bilkent.edu.tr

Abstract— We derive an upper bound on the capacity of non-binary deletion channels. Although binary deletion channels have received significant attention over the years, and many upper and lower bounds on their capacity have been derived, such studies for the non-binary case are largely missing. The state of the art is the following: as a trivial upper bound, capacity of an erasure channel with the same input alphabet as the deletion channel can be used, and as a lower bound the results by Diggavi and Grossglauser in [1] are available. In this paper, we derive the first non-trivial non-binary deletion channel capacity upper bound 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, $C_{2K}(d) \leq C_2(d) + (1-d)\log(K)$. Then by employing some existing upper bounds on the capacity of the binary deletion channel, we obtain upper bounds on the capacity of the $2K$ -ary deletion channel. We illustrate via examples the use of the new bounds and discuss their asymptotic behavior as $d \rightarrow 0$.

I. INTRODUCTION

Non-binary deletion channels can be used to model information transmission over a finite buffer channel [1], where a packet (non-binary symbol) loss occurs whenever a packet arrives at a full buffer. When the channel drop-outs are independent and identically distributed (i.i.d.), the channel is referred as a non-binary i.i.d. deletion channel. Dobrushin [2] proved the existence of Shannon's theorem for discrete memoryless channels with synchronization errors. As a result, Shannon's theorem holds in non-binary deletion channels and information and transmission capacities are equal.

In this paper, we focus on a $2K$ -ary deletion channel in which every transmitted symbol is either lost through the transmission with probability of d or received correctly with probability of $1-d$. There is no information about the position of the lost symbols at either the transmitter or the receiver. Clearly the capacity of a $2K$ -ary erasure channel with erasure probability d is an upper bound on the capacity of the $2K$ -ary deletion channel since by revealing information about the position of the lost symbols to the receiver, the corresponding genie-aided deletion channel is nothing but an erasure channel. Therefore, for the capacity of the $2K$ -ary input deletion channel $C_{2K}(d)$, the relation $C_{2K}(d) \leq (1-d)\log(2K)$ holds. Besides this trivial upper bound, to the best of our knowledge,

there are no other (tighter) upper bounds on the capacity of non-binary deletion channels.

Our main result is to relate the capacity of a $2K$ -ary deletion channel with deletion probability d to the capacity of the binary deletion channel with deletion probability d by the inequality $C_{2K} \leq C_2(d) + (1-d)\log(K)$. 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. For example, by using the result from [3], we obtain $C_{2K}(d) \leq (\log(K) + 0.4143)(1-d)$ for $d \geq 0.65$.

The paper is organized as follows. In Section II, we briefly review the existing work on the capacity of binary and non-binary deletion channels. In Section III, we first give the general $2K$ -ary deletion channel model and then we observe that it can be considered as a parallel concatenation of K independent deletion channels (where each input is binary). Also in the same section, we discuss the possible generalization of the existing Blahut-Arimoto algorithm (BAA) based upper bounding approaches (useful for the binary deletion channels) to the case of $2K$ -ary deletion channels. In Section IV, we prove the main result of the paper providing an upper bound on $C_{2K}(d)$ in terms of $C_2(d)$. In Section V, several implications of the result are given where we compare the resulting capacity upper bounds with the existing capacity upper and lower bounds, and we provide a discussion of the channel capacity behavior as the deletion probability approaches zero. Finally, we conclude the paper in Section VI.

II. PREVIOUS WORKS

Capacity of binary deletion channels has received significant attention in the existing literature, e.g., see [4] and references therein. There are several results on capacity lower bounds [5]–[7]. Gallager [5] 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 [7] 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.

T. M. Duman is currently with Bilkent University in Turkey, and on leave from Arizona State University, Tempe, AZ.

There are also several upper bounds on the binary deletion channel capacity, e.g., [3], [8], [9]. In [8] a genie-aided channel is considered in which the receiver is provided by side information about the completely deleted runs, e.g., in transmitting “110001” over the original channel by deleting the entire run of zeros, the sequence “111” is received while in the considered genie-aided channel “11 – 1” represents the received sequence. Then an upper bound on the capacity per unit cost of the genie-aided channel is computed by running the BAA algorithm. Fertoni and Duman [9], by considering several different genie-aided channels, are able to derive tighter upper bounds on the binary deletion channel capacity compared to the results in [8] for $d > 0.05$. In [3], authors improve upon the upper bounds provided in [9] for $d > 0.65$ where they first derive an inequality relation among the capacity of three different binary deletion channels and as a special case they obtain $C_2(\lambda d + 1 - \lambda) \leq \lambda C_2(d)$ which shows that $C_2(d) \leq 0.4143(1 - d)$ for $d \geq 0.65$.

To the best of our knowledge, the only non-trivial lower bounds on the capacity of the non-binary deletion channels are provided in [1] where two different bounds are derived. More precisely, the achievable rates of the $2K$ -ary input deletion channel 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

$$C_{2K} \geq \log \left(\frac{2K}{2K-1} \right) + (1-d) \log(2K-1) - H_b(d), \quad (1)$$

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

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

by considering Markovian codebooks, with $q = \frac{1}{2K} \left(1 + \frac{(1-d)(2K-1)(2Kp-1)}{2K-1-d(2Kp-1)} \right)$, $A = \frac{e^{-\gamma}(1-p)}{(2K-1)(1-e^{-\gamma}(1-\frac{1-p}{2K-1}))}$ and $B = e^{-\gamma}((1-p)A + p)$. Non-binary input alphabet channels with synchronization errors are also considered in [10] 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 [10] is on the asymptotic behavior of the channel capacity for large values of K .

III. PRELIMINARIES

A. Channel Model

An i.i.d. $2K$ -ary input deletion channel with input alphabet $\mathcal{X} = \{1, \dots, 2K\}$ 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 position of the lost symbols at the transmitter and the receiver. In transmission of N symbols through the channel, the input sequence is denoted by

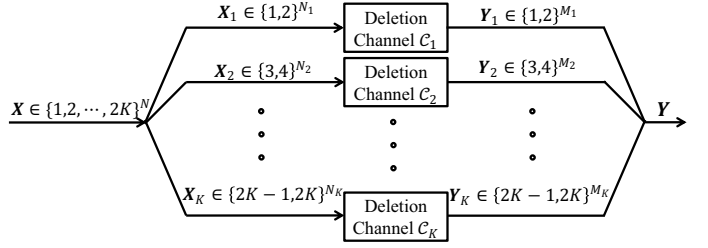


Fig. 1. $2K$ -ary deletion channel as a parallel concatenation of K independent binary input deletion channels.

$\mathbf{X} = (x_1, \dots, x_N)$ in which $x_n \in \mathcal{X}$ and $\mathbf{X} \in \mathcal{X}^N$, and the output sequence is denoted by $\mathbf{Y} = (y_1, \dots, y_M)$ in which M is a binomial random variable with parameters N and d (due to the characteristics of the i.i.d. deletion channel).

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 ($k \in \{1, \dots, K\}$) all with the same deletion probability d , as shown in Fig. 1, 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. \mathbf{X}_k and \mathbf{Y}_k denote the input and output sequences of the k -th channel, respectively, and N_k and M_k denote the length of \mathbf{X}_k and \mathbf{Y}_k , respectively.

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

$$I(\mathbf{X}; \mathbf{Y}) = I(\mathbf{X}_1, \dots, \mathbf{X}_K, \mathbf{F}_x; \mathbf{Y}_1, \dots, \mathbf{Y}_K, \mathbf{F}_y) \\ = \sum_{k=1}^K I_k + I_F, \quad (3)$$

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

$$I_F = I(\mathbf{X}_1, \dots, \mathbf{X}_K, \mathbf{F}_x; \mathbf{F}_y | \mathbf{Y}_1, \dots, \mathbf{Y}_K). \quad (4)$$

In Section IV, 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.

B. Discussion on BAA Based Upper Bounds

One approach to derive upper bounds on the $2K$ -ary deletion channel capacity is to modify the numerical approaches in [8], [9] 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 [11]), for the non-binary case, running the BAA for large values of K is not computationally feasible. E.g., one of the upper bounds in [9] is obtained by computing the capacity of the binary deletion channel with finite length of transmission $L = 17$. Obviously, by increasing the alphabet size, $2K$, 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 paper is that we are able to relate the capacity of the $2K$ -ary deletion channel to the binary deletion channel capacity through an inequality which enables us to upper bound the $2K$ -ary deletion channel capacity avoiding computationally formidable BAA directly for the $2K$ -ary deletion channel.

IV. A NOVEL UPPER BOUND ON $C_{2K}(d)$

As introduced in Section III-A, 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 1. *Let $C_{2K}(d)$ denote the capacity of a $2K$ -ary i.i.d. deletion channel with deletion probability d , then*

$$C_{2K}(d) \leq C_2(d) + (1-d) \log(K). \quad (5)$$

As given in (3), the mutual information $I(\mathbf{X}; \mathbf{Y})$ can be expanded in terms of several other mutual information terms, I_k for $k \in \{1, \dots, K\}$ and I_F . To prove the theorem, we first derive upper bounds on I_k and I_F in the following two lemmas.

Lemma 1. *For any input distribution $P(\mathbf{X}_1, \dots, \mathbf{X}_K, \mathbf{F}_x)$, the mutual information I_k given in (3) can be upper bounded by*

$$I_k \leq E\{N_k\}C_2(d) + 2\log(N+1),$$

where $E\{\cdot\}$ denotes the expected value.

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

$$\begin{aligned} I_k &= I(\mathbf{X}_k; \mathbf{Y}_k|\mathbf{Y}_1, \dots, \mathbf{Y}_{k-1}) \\ &+ I(\mathbf{X}_1, \dots, \mathbf{X}_{k-1}, \mathbf{X}_{k+1}, \dots, \mathbf{X}_K, \mathbf{F}_x; \mathbf{Y}_k|\mathbf{Y}_1, \dots, \mathbf{Y}_{k-1}, \mathbf{X}_k) \\ &= I(\mathbf{X}_k; \mathbf{Y}_k|\mathbf{Y}_1, \dots, \mathbf{Y}_{k-1}) \\ &= H(\mathbf{Y}_k|\mathbf{Y}_1, \dots, \mathbf{Y}_{k-1}) - H(\mathbf{Y}_k|\mathbf{Y}_1, \dots, \mathbf{Y}_{k-1}, \mathbf{X}_k) \\ &= H(\mathbf{Y}_k) - I(\mathbf{Y}_1, \dots, \mathbf{Y}_{k-1}; \mathbf{Y}_k) - H(\mathbf{Y}_k|\mathbf{X}_k) \\ &\leq I(\mathbf{X}_k; \mathbf{Y}_k). \end{aligned} \quad (6)$$

Furthermore, $I(\mathbf{X}_k; \mathbf{Y}_k)$ can be written as

$$\begin{aligned} I(\mathbf{X}_k; \mathbf{Y}_k) &= I(\mathbf{X}_k; \mathbf{Y}_k, N_k) - I(\mathbf{X}_k; N_k|\mathbf{Y}_k) \\ &= I(\mathbf{X}_k; \mathbf{Y}_k|N_k) + I(\mathbf{X}_k; N_k) - I(\mathbf{X}_k; N_k|\mathbf{Y}_k). \end{aligned}$$

Since $H(N_k|\mathbf{X}_k) = 0$ and $I(\mathbf{X}_k; N_k|\mathbf{Y}_k) \geq 0$, we arrive at

$$\begin{aligned} I(\mathbf{X}_k; \mathbf{Y}_k) &\leq I(\mathbf{X}_k; \mathbf{Y}_k|N_k) + H(N_k) \\ &\leq I(\mathbf{X}_k; \mathbf{Y}_k|N_k) + \log(N+1) \\ &= \sum_{n_k=0}^N P(N_k=n_k) I(\mathbf{X}_k; \mathbf{Y}_k|n_k) + \log(N+1), \end{aligned} \quad (7)$$

where the second inequality results since there are $N+1$ possibilities for N_k and as a result $H(N_k) \leq \log(N+1)$. Furthermore, as shown in [9], 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 [9, Eqn. (39)]

$$I(\mathbf{X}_k; \mathbf{Y}_k|N_k = n_k) \leq n_k C_2(d) + H(\mathbf{D}_k|N_k = n_k), \quad (8)$$

where \mathbf{D}_k denotes the number of deletions through the transmission of N_k bits over the k -th channel. We have

$$\begin{aligned} H(\mathbf{D}_k|N_k = n_k) &= - \sum_{n=0}^{n_k} P(n_k, n, d) \log(P(n_k, n, d)) \\ &\leq \log(n_k + 1) \leq \log(N+1), \end{aligned} \quad (9)$$

with $P(n_k, n, d) = \binom{n_k}{n} d^n (1-d)^{n_k-n}$. Substituting (9) and (8) into (7), we obtain

$$\begin{aligned} I(\mathbf{X}_k; \mathbf{Y}_k) &\leq \sum_{n_k=0}^N P(N_k = n_k) (n_k C_2(d) + 2\log(N+1)) \\ &= E\{N_k\}C_2(d) + 2\log(N+1). \end{aligned}$$

Finally, by substituting the above inequality into (6), the proof follows. \blacksquare

Lemma 2. *For any input distribution, the mutual information I_F given in (4) can be upper bounded by*

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

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

$$\begin{aligned} I_F &= H(\mathbf{F}_y|\mathbf{Y}_1, \dots, \mathbf{Y}_K) - H(\mathbf{F}_y|\mathbf{Y}_1, \dots, \mathbf{Y}_K, \mathbf{X}_1, \dots, \mathbf{X}_K, \mathbf{F}_x) \\ &\leq H(\mathbf{F}_y|\mathbf{Y}_1, \dots, \mathbf{Y}_K) \\ &\leq H(\mathbf{F}_y|M_1, \dots, M_K), \end{aligned} \quad (10)$$

where the last inequality follows since (M_1, \dots, M_K) is a function of $(\mathbf{Y}_1, \dots, \mathbf{Y}_K)$, i.e., $H(M_1, \dots, M_K|\mathbf{Y}_1, \dots, \mathbf{Y}_K) = 0$.

For fixed m_k with $\sum_{k=1}^K m_k = m$, there are $\binom{m}{m_1, \dots, m_K}$ possibilities for \mathbf{F}_y leading to $H(\mathbf{F}_y|m_1, \dots, m_K) \leq \log \binom{m}{m_1, \dots, m_K}$. It follows from the inequality (see Appendix A)

$$\log \binom{m}{m_1, \dots, m_K} \leq m \log(m) - \sum_{k=1}^K m_k \log(m_k), \quad (11)$$

that $H(\mathbf{F}_y|m_1, \dots, m_K) \leq m \log(m) - \sum_{k=1}^K m_k \log(m_k)$. Since $g([m_1, \dots, m_K]) = \left(\sum_{k=1}^K m_k\right) \log\left(\sum_{k=1}^K m_k\right) - \sum_{k=1}^K m_k \log(m_k)$ is a concave function of $[m_1, \dots, m_K]$ (see Appendix B), employing the Jensen's inequality yields

$$I_F \leq \left(\sum_{k=1}^K E\{M_k\}\right) \log\left(\sum_{k=1}^K E\{M_k\}\right) - \sum_{k=1}^K E\{M_k\} \log(E\{M_k\}).$$

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

$$\begin{aligned} I_F &\leq N(1-d) \left(\log(N(1-d)) - \sum_{k=1}^K \alpha_k \log(N(1-d)\alpha_k) \right) \\ &= -N(1-d) \sum_{k=1}^K \alpha_k \log \alpha_k = N(1-d) H(\alpha_1, \dots, \alpha_K) \\ &\leq N(1-d) \log(K), \end{aligned} \quad (12)$$

which concludes the proof. \blacksquare

A. Proof of Theorem 1

Substituting the results of Lemmas 1 and 2 into (3), we obtain

$$\begin{aligned} I(\mathbf{X}; \mathbf{Y}) &\leq E_{N_1, \dots, N_K} \left\{ \sum_{k=1}^K N_k \right\} C_2(d) + 2K \log(N+1) \\ &\quad + N(1-d) \log(K) \\ &= NC_2(d) + 2K \log(N+1) + N(1-d) \log(K), \end{aligned}$$

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

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

which concludes the proof of Theorem 1. \square

V. SOME IMPLICATIONS

As stated earlier, a trivial upper bound on the capacity of the $2K$ -ary deletion channel is given by $(1-d) \log(2K)$ which is the capacity of the $2K$ -ary erasure channel. We have shown in the previous section that by substituting any upper bound on the capacity of the binary deletion channel into (5), an upper bound on the $2K$ -ary deletion channel capacity results. Obviously, 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 [10], $(1-d) \log(2K) - 1 \leq C_{2K}(d) \leq (1-d) \log(2K)$, where the lower bound is implied from (1), therefore the existing trivial upper and lower bounds are tight enough for asymptotically large values of K , and i.i.d. distributed input sequences are sufficient to achieve the capacity. However, the importance of the result in Theorem 1 is for moderate values of K , 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 [1]) in Fig. 2 for 4-ary and 8-ary deletion channels. Here we utilize the binary deletion channel capacity upper bounds $C_2^{UB}(d)$ in [3], [9], where for $d \leq 0.65$ we use the results in [9, Table III] and for $d \geq 0.65$ we use the upper bound $C_2(d) \leq 0.4143(1-d)$ given in [3].

Another implication of the result in Theorem 1 is in studying the asymptotic behavior of the $2K$ -ary deletion channel capacity for $d \rightarrow 0$. It is shown in [12] that

$$C_2(d) = 1 + d \log(d) - A_1 d + A_2 d^2 + O(d^{3-\epsilon}), \quad (13)$$

for small d and any $\epsilon > 0$ with $A_1 \approx 1.15416377$, $A_2 \approx 1.78628364$ and $O(\cdot)$ denoting the standard Landau (big-O) notation. Employing this result in (5), leads to an upper bound expansion for small values of d as

$$\begin{aligned} C_{2K}(d) &\leq 1 + d \log(d) - (A_1 + \log(K))d + A_2 d^2 + \log(K) \\ &\quad + O(d^{3-\epsilon}). \end{aligned} \quad (14)$$

In Fig. 3, 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 capacity expansion (13) in (5), a good characterization for the asymptotic behavior of the $2K$ -ary deletion channel capacity is obtained as $d \rightarrow 0$.

VI. CONCLUSIONS

We have derived the first non-trivial upper bound on the $2K$ -ary deletion channel capacity. We first considered the $2K$ -ary deletion channel as a parallel concatenation of K independent binary deletion channels, all with the same deletion probability. We then related the capacity of the original channel to that of the binary deletion channel. By doing so we obtained an upper bound on the capacity of the $2K$ -ary deletion channel in terms of the capacity of the binary deletion channel and as a result any upper bound on the capacity of the binary deletion channel. The provided upper bound results in tighter upper bounds on the K -ary deletion channel capacity than the trivial erasure channel upper bound for the entire range of deletion probabilities.

APPENDIX A

PROOF OF INEQUALITY (11)

It follows from the inequality $\log\left(\frac{m}{m_1}\right) \leq m H_b\left(\frac{m_1}{m}\right) = m \log(m) - m_1 \log(m_1) - (m - m_1) \log(m - m_1)$ given

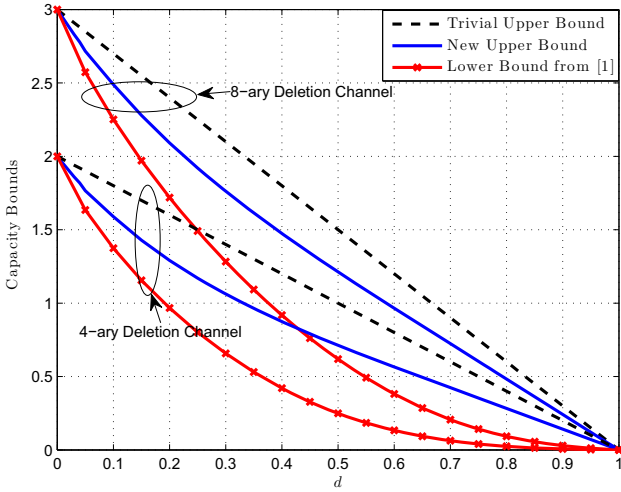


Fig. 2. Comparison among the new upper bound (5), the lower bound (2) and the trivial erasure channel upper bound for the 4-ary and 8-ary deletion channels.

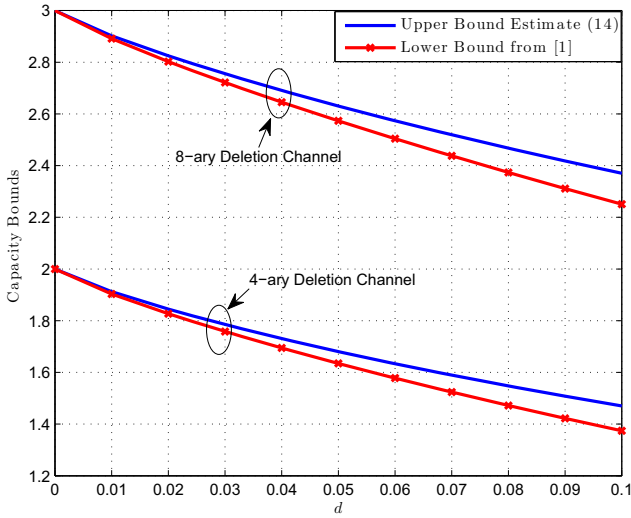


Fig. 3. Comparison between the upper bound (14) (ignoring the $O(d^{3-\epsilon})$ term) and the lower bound (2).

in [13, p. 353] that

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

APPENDIX B

CONCAVITY OF $g([m_1, \dots, m_k])$

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

$$\nabla^2 g([m_1, \dots, m_k]) = \frac{1}{\sum_{k=1}^K m_k} \mathbf{1}\mathbf{1}^T - \text{diag} \left\{ \frac{1}{m_1}, \dots, \frac{1}{m_K} \right\},$$

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

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

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

ACKNOWLEDGMENT

The authors are supported by the National Science Foundation under the contract NSF-TF 0830611. Tolga M. Duman is also supported by the EC Marie Curie Career Integration Grant PCIG12-GA-2012-334213.

REFERENCES

- [1] 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.
- [2] R. L. Dobrushin, "Shannon's theorems for channels with synchronization errors," *Probs. Inf. Transm.*, vol. 3, no. 4, pp. 11–26, 1967.
- [3] M. Rahmati and T. M. Duman, "A note on the deletion channel capacity," *submitted to IEEE Trans. Inf. Theory*, ArXiv e-prints:1211.2497, Nov. 2012.
- [4] M. Mitzenmacher, "A survey of results for deletion channels and related synchronization channels," *Probability Surveys*, vol. 6, pp. 1–33, 2009.
- [5] R. Gallager, "Sequential decoding for binary channels with noise and synchronization errors," *Tech. Rep., MIT Lincoln Lab. Group Report*, Oct. 1961.
- [6] 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.
- [7] 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.
- [8] S. Diggavi, M. Mitzenmacher, and H. Pfister, "Capacity upper bounds for deletion channels," in *Proceedings of the International Symposium on Information Theory (ISIT)*, 2007, pp. 1716–1720.
- [9] 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.
- [10] 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, July 2012.
- [11] 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.
- [12] Y. Kanoria and A. Montanari, "Optimal coding for the deletion channel with small deletion probability," *ArXiv e-prints:1104.5546*, Apr. 2011.
- [13] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Wiley, 2006.