Achievable Rates for Channels with Deletions and Insertions

Ramji Venkataramanan
Dept. of Electrical Engineering
Yale University, USA
Email: ramji.venkataramanan@yale.edu

Sekhar Tatikonda
Dept. of Electrical Engineering
Yale University, USA
Email: sekhar.tatikonda@yale.edu

Kannan Ramchandran
Dept. of EECS
University of California, Berkeley, USA
Email: kannanr@eecs.berkeley.edu

***Abstract*—Consider a binary channel with deletions and insertions, where each input bit is transformed in one of the following ways: it is deleted with probability d, or an extra bit added after it with probability i, or it is transmitted unmodified with probability $1-d-i$. We obtain a lower bound on the capacity of this channel. The transformation of the input sequence by the channel may be viewed in terms of runs as follows: some runs of the input sequence get shorter/longer, some runs get deleted, and some new runs are added. The capacity is difficult to compute mainly due to the last two phenomena: deleted runs, and new inserted runs. We consider a decoder which first decodes the positions of the deleted and inserted runs, and then the transmitted codeword. Analyzing the performance of such a decoder leads to a computable lower bound on the capacity.**

I. Introduction

Consider a binary input channel where for each bit (denoted x), the output is generated in one of the following ways:

- The bit is deleted with probability d,
- An extra bit is inserted after x with probability i. The extra bit is equal to x with probability α, and equal to $1-x$ with probability $1-\alpha$,
- No deletions or insertions occur, and the output is x with probability $1-d-i$.

The channel acts independently on each bit. We refer to this channel as the deletion+insertion channel. If the channel input is a sequence of n bits, the length of the output will be close to $n(1+i-d)$ for large n due to the law of large numbers.

When $i=0$, the above model is the deletion channel which has been studied in several recent papers, e.g., [1]–[8]. When $d=0$, we obtain the insertion channel. The insertion channel with $\alpha=1$ is the sticky channel [9], where all insertions are duplications. We also note that a different channel model with bit flips and synchronization errors was studied in [10], [11].

In this work, we obtain lower bounds on the capacity of the deletion+insertion channel. Our starting point is the result of Dobrushin [12] for general synchronization channels which states that the capacity is given by the maximum of the mutual information per bit between the input and output sequences. There are two challenges to computing the capacity through this characterization. The first is evaluating the mutual information, which is a difficult task because of the memory inherent in the joint distribution of the input and output sequences. The second challenge is to optimize the mutual information over all input distributions.

In this paper, we choose the input distribution to be the class of binary first-order Markov processes and focus on the problem of evaluating the mutual information. It is known that first-order Markov input distributions yield good capacity lower bounds for deletion channels [1], [2] and sticky channels [9], which are special cases of the deletion+insertion channel. This suggests they are likely to perform well on the deletion+insertion channel as well.

For a synchronization channel, it is useful to think of the input and output sequences in terms of *runs* of symbols rather than individual symbols. (The runs of a binary sequence are its alternating blocks of contiguous zeros and ones.) If there was a one-to-one correspondence between the runs of the input sequence $\underline{X}$ and those of the output sequence $\underline{Y}$, we could write the conditional distribution $P(\underline{Y}|\underline{X})$ as a product distribution of run-length transformations; computing the mutual information would then be straightforward. Unfortunately, such a correspondence is not possible since deletions can lead to some runs being lost, and insertions to new runs being inserted. The main idea of the paper is to use auxiliary sequences which indicate the positions (in the output sequence) where runs were deleted and inserted. We will consider a decoder that first decodes the auxiliary sequences, and then the input sequence $\underline{X}$. We derive a computable expression for the maximum rate achievable with such a decoder, and thus obtain a lower bound on the capacity.

A related idea was used in [4], where a genie-aided decoder with access to the locations of deleted runs was used to obtain an upper bound on the deletion capacity. We note that Dobrushin's capacity characterization was also used in [3] to obtain bounds on the deletion capacity. The auxiliary sequences in this paper are quite different from the one used in [3], resulting in a different mutual information decomposition. The decomposition in [3] results in the best known achievable rates for channels with deletions and duplications, but does not apply to channels with general insertions. Our approach yields the first characterization of achievable rates for channels with deletions as well as insertions.

To develop our ideas, we start with the insertion channel ($d=0$) in Section III, and derive a lower bound on its capacity. For this channel, previous bounds exist only for the special case of sticky channels ($\alpha=1$) [9]. In Section IV, we derive a lower bound on the capacity of the deletion channel ($i=0$), and compare it with the best known lower bounds. In Section V, the ideas of Sections III and IV are combined to obtain a lower bound for the deletion+insertion capacity. Due to space constraints, we only give a brief sketch of the proofs of two results. Detailed proofs will be given in an extended version.

II. Preliminaries

Notation: $\mathbb{N}_0$ denotes the set of non-negative integers, and $\mathbb{N}$ the set of natural numbers. $h(.)$ is the binary entropy function, and for any $0 < \alpha \leq 1$, $\bar{\alpha} \triangleq 1 - \alpha$. We use uppercase letters to denote random variables, bold-face letters for random processes, and superscript notation to denote random vectors. Thus the channel input sequence of length n is denoted $X^n \triangleq (X_1, \ldots, X_n)$. The corresponding output sequence at the decoder has length M_n (a random variable determined by the channel realization), and is denoted Y^{M_n}. For brevity, we sometimes use underlined notation for random vectors when we do not need to be explicit about their length. Thus $\underline{X} \triangleq X^n = (X_1, X_2, \ldots, X_n)$, and $\underline{Y} \triangleq Y^{M_n} = (Y_1, \ldots, Y_{M_n})$.

Definition 1: An $(n, 2^{nR})$ code with block length n and rate R consists of

1) An encoder mapping $e : \{1, \ldots, 2^{nR}\} \to \{0,1\}^n$, and
2) A decoder mapping $g : \{0,1\}^* \to \{1, \ldots, 2^{nR}\}$ where $\{0,1\}^*$ is $\cup_{k=0}^{n}\{0,1\}^k$ for the deletion channel, $\cup_{k=n}^{2n}\{0,1\}^k$ for the insertion channel, and $\cup_{k=0}^{2n}\{0,1\}^k$ for the deletion+insertion channel.

Assuming the message W is drawn uniformly on the set $\{1, \ldots, 2^{nR}\}$, the probability of error of a $(n, 2^{nR})$ code is

$$P_{e,n} = \frac{1}{2^{nR}} \sum_{l=1}^{2^{nR}} \Pr(g(Y^{M_n}) \neq l | W = l)$$

A rate R is achievable if there exists a sequence of $(n, 2^{nR})$ codes such that $P_{e,n} \to 0$ as $n \to \infty$. The supremum of all achievable rates is the capacity C. The following characterization of capacity follows from a result proved for a general class of synchronization channels by Dobrushin.

Fact [12]: Let $C_n = \max_{P_{X^n}} \frac{1}{n} I(X^n; Y^{M_n})$. Then $C \triangleq \lim_{n\to\infty} C_n$ exists, and is equal to the capacity of the deletion+insertion channel.

In this paper, we fix the input process to be the class of binary symmetric first-order Markov processes and focus on evaluating the mutual information. This will give us a lower bound on the capacity. The input process $\mathbf{X} = \{X_n\}_{n\geq 1}$ is characterized by the following distribution for all n:

$$P(X_1, \ldots, X_n) = P(X_1) \prod_{j=2}^{n} P(X_j | X_{j-1}), \text{ with}$$

$$P(X_1 = 0) = P(X_1 = 1) = 0.5,$$
$$P(X_j = 1 | X_{j-1} = 1) = P(X_j = 0 | X_{j-1} = 0) = \gamma, \; j \geq 1. \tag{1}$$

A binary sequence may be represented by a sequence of positive integers representing the lengths of its runs, and the value of the first bit (to indicate whether the first run has zeros or ones). For example, the sequence 0001100000 can be represented as $(3, 2, 5)$ if we know that the first bit is 0. The value of the first bit of $\mathbf{X}$ can be communicated to the decoder with vanishing rate, and we will assume this has been done at the outset. Hence, denoting the length of the jth run of $\mathbf{X}$ by L_j^X we have the following equivalence: $\mathbf{X} \leftrightarrow (L_1^X, L_2^X, \ldots)$. For a first-order Markov binary source of (1), the run-lengths are independent and geometrically distributed, i.e.,

$$\Pr(L_j^X = r) = \gamma^{r-1}(1-\gamma), \qquad r = 1, 2, \ldots \tag{2}$$

The average length of a run in $\mathbf{X}$ is $\frac{1}{1-\gamma}$, so the number of runs in a sequence of length n is close to $n(1-\gamma)$ for large n. Our bounding techniques aim to establish a one-to-one correspondence between input runs and output runs. The independence of run-lengths of $\mathbf{X}$ enables us to obtain analytical bounds on the capacity. We denote by $I_P(X^n; Y^{M_n}), H_P(X^n), H_P(X^n|Y^{M_n})$ the mutual information and entropies computed with the channel input sequence X^n distributed as in (1). For all n, we have

$$C_n = \max_{P_{X^n}} \frac{1}{n} I(X^n; Y^{M_n}) > \frac{1}{n} I_P(X^n; Y^{M_n}). \tag{3}$$

Therefore

$$C > \liminf_{n\to\infty} \frac{1}{n} I_P(X^n; Y^{M_n}) = h(\gamma) - \limsup_{n\to\infty} \frac{H_P(X^n|Y^{M_n})}{n}. \tag{4}$$

We will derive upper bounds on $\limsup_{n\to\infty} \frac{H_P(X^n|Y^{M_n})}{n}$ and use it in (4) to obtain a lower bound on the capacity.

III. Insertion Channel

In this channel, an insertion occurs after each bit of $\underline{X}$ with probability $i \in (0,1)$. When a bit is inserted after X_j, the inserted bit is equal to X_j (a *duplication*) with probability α, and equal to $\bar{X}_j$ (a *complementary insertion*) with probability $\bar{\alpha}$. We note that this insertion model is different from the one considered in [10], [11], where an insertion is defined as an input bit replaced by two random bits.

When $\alpha = 1$, we have only duplications - this is the sticky channel studied in [9]. Here, we can associate each run of $\underline{Y}$ with a unique run in $\underline{X}$, which leads to a computable single-letter characterization of the best achievable rates with a first-order distribution. For $0 < \alpha < 1$, the inserted bits may create new runs, and so we cannot associate each run of $\underline{Y}$ with a run in $\underline{X}$, as shown by the following example. Let

$$\underline{X} = 000111000, \quad \underline{Y} = 00\mathit{1}0111\mathit{0}000\mathit{0}, \tag{5}$$

where the inserted bits are indicated in large italics. There is one duplication - in the third run, and two complementary insertions - in the first and second runs. While a duplication never introduces a new run, a complementary insertion introduces a new run, except when it occurs at the end of a run of $\underline{X}$ (e.g., the 0 inserted at the end of the second run in (5)). We derive two lower bounds on the the insertion channel capacity.

A. *Lower Bound* 1

For any input-pair (X^n, Y^{M_n}), define an auxiliary sequence $I^{M_n} = (I_1, \dots, I_{M_n})$ where $I_j = 1$ if Y_j is an inserted bit, and $I_j = 0$ otherwise. The sequence I^{M_n} indicates the positions of the inserted bits in Y^{M_n}. Using I^{M_n}, we can decompose $H_P(X^n|Y^{M_n})$ as

$$\begin{aligned} H_P(X^n|Y^{M_n}) &= H_P(X^n, I^{M_n}|Y^{M_n}) - H_P(I^{M_n}|X^nY^{M_n}) \\ &= H_P(I^{M_n}|Y^{M_n}) - H_P(I^{M_n}|X^nY^{M_n}) \end{aligned} \tag{6}$$

since $H(X^n|Y^{M_n}, I^{M_n}) = 0$. Therefore,

$$\limsup_{n\to\infty} \frac{H_P(X^n|Y^{M_n})}{n} \leq \limsup_{n\to\infty} \frac{H_P(I^{M_n}|Y^{M_n})}{n}. \tag{7}$$

We use this inequality in (4) and upper bound $H_P(I^{M_n}|Y^{M_n})$ to obtain the following lower bound on the insertion capacity.

Theorem 1: (LB 1) The capacity of the insertion channel with parameters (i, α) can be lower bounded as

$$\begin{aligned} C(i,\alpha) \geq \max_{0<\gamma<1}\ & h(\gamma) - (i\alpha + (1-i)\gamma) h\left(\frac{i\alpha}{i\alpha + (1-i)\gamma}\right) \\ & - (i\bar{\alpha} + (1-i)\bar{\gamma}) h\left(\frac{i\bar{\alpha}}{i\bar{\alpha} + (1-i)\bar{\gamma}}\right). \end{aligned}$$

One can interpret the lower bound as the rate achieved by the following coding scheme. Choose a codebook of 2^{nR} codewords of length n, each chosen independently according to the distribution in (1). The decoder receives Y^{M_n}, decodes the inserted bits, and then obtains the codeword by removing them from Y^{M_n}.

B. *Lower Bound* 2

For any input-pair (X^n, Y^{M_n}), define an auxiliary sequence $T^{M_n} = (T_1, \dots, T_{M_n})$ where $T_j = 1$ if Y_j is a *complementary* insertion, and $T_j = 0$ otherwise. Note that T^{M_n}, which indicates the positions of the complementary insertions, is different from the sequence I^{M_n}, which indicates *all* the insertions. Using T^{M_n}, we can decompose $H_P(X^n|Y^{M_n})$ as

$$\begin{aligned} H_P(X^n|Y^{M_n}) &= H_P(X^n, T^{M_n}|Y^{M_n}) - H_P(T^{M_n}|X^nY^{M_n}) \\ &\leq H_P(T^{M_n}|Y^{M_n}) + H_P(X^n|T^{M_n}Y^{M_n}) \\ &\overset{(a)}{\leq} H_P(T^{M_n}|Y^{M_n}) + H_P(X^n|\tilde{Y}^{M_n}) \end{aligned} \tag{8}$$

where $\tilde{Y}^{M_n}$ is the sequence obtained from (T^{M_n}, Y^{M_n}) by flipping Y_j whenever $T_j = 1$, $1 \leq j \leq M_n$. (a) holds in (8) because $\tilde{Y}^{M_n}$ is a function of (T^{M_n}, Y^{M_n}). In words, $\tilde{Y}^{M_n}$ is formed by flipping the complementary insertions in Y^{M_n}. Therefore,

$$\limsup_{n\to\infty} \frac{H_P(X^n|Y^{M_n})}{n} \leq \limsup_{n\to\infty} \frac{H_P(T^{M_n}|Y^{M_n}) + H_P(X^n|\tilde{Y}^{M_n})}{n} \tag{9}$$

We use (9) in (4) to obtain a lower bound on the insertion capacity.

Theorem 2: (LB 2) The capacity of the insertion channel with parameters (i, α) can be lower bounded as

$$C(i,\alpha) \geq \max_{0<\gamma<1} h(\gamma) - (\bar{\gamma} + \gamma i\bar{\alpha})\, h\left(\frac{i\bar{\alpha}}{\bar{\gamma} + \gamma i\bar{\alpha}}\right) - \bar{\gamma} H(L_X|L_{\tilde{Y}})$$

where $H(L_X|L_{\tilde{Y}})$ is computed using the following joint distribution:

$$\begin{aligned} P(L^X = r) &= \gamma^{r-1}(1-\gamma), \quad r = 1, 2, \dots, \\ P(L^{\tilde{Y}} = s|L^X = r) &= \binom{r}{s-r} i^{s-r}(1-i)^{2r-s}, \quad r \leq s \leq 2r. \end{aligned} \tag{10}$$

Proof Sketch: Recall that $\tilde{Y}^{M_n}$ has insertions in the same locations as Y^{M_n}, but the insertions are all duplications. Hence $\tilde{Y}^{M_n}$ has the same number of runs (say, R_n) as X^n. We can represent the sequences in terms of their run-lengths as

$$X^n \leftrightarrow (L_1^X, \dots, L_{R_n}^X), \quad \tilde{Y}^{M_n} \leftrightarrow (L_1^{\tilde{Y}}, \dots, L_{R_n}^{\tilde{Y}}).$$

Thus $H_P(X^n|\tilde{Y}^{M_n}) = H_P(L_1^X, \dots, L_{R_n}^X|L_1^{\tilde{Y}}, \dots, L_{R_n}^{\tilde{Y}})$. Since there is a one-to-one correspondence between the runs of $\mathbf{X}$ and the runs of $\tilde{\mathbf{Y}}$, the process $\{(L_1^X, L_1^{\tilde{Y}}), (L_2^X, L_2^{\tilde{Y}}), \dots\}$ is an i.i.d process characterized by the joint distribution in (10). Combining this with the fact that $\frac{R_n}{n} \to (1-\gamma)$ almost surely, one can prove that $\lim_{n\to\infty} \frac{1}{n} H_P(X^n|\tilde{Y}^{M_n}) = \bar{\gamma} H(L_X|L_{\tilde{Y}})$.

To use the bound (9), we also need to compute $\limsup_{n\to\infty} \frac{1}{n} H_P(T^{M_n}|Y^{M_n})$. Noting that $\frac{M_n}{n} \to (1+i)$ almost surely, one can show that

$$\limsup_{n\to\infty} \frac{1}{n} H_P(T^{M_n}|Y^{M_n}) = (1+i) \limsup_{m\to\infty} \frac{1}{m} H_P(T^m|Y^m).$$

Since conditioning cannot increase entropy, we have

$$\frac{1}{m} H_P(T^m|Y^m) \leq \frac{1}{m} \sum_j H(T_j|T_{j-1}, Y_{j-1}, Y_j). \tag{11}$$

We can formally show that the limit as $m \to \infty$ of the term on the right hand side of (11) exists, and is equal to $\frac{1-\gamma+\gamma i\bar{\alpha}}{1+i} h\left(\frac{i\bar{\alpha}}{1-\gamma+\gamma i\bar{\alpha}}\right)$. We have thus obtained an analytical upper bound on $\limsup_{n\to\infty} \frac{1}{n} H_P(T^{M_n}|Y^{M_n})$. Using this in (9), and substituting the resulting bound in (4) completes the proof.

Combining the bounds of Theorems 1 and 2, we observe that $\max\{LB\ 1, LB\ 2\}$ is a lower bound to the insertion capacity. This is plotted in Figure 1 for various values of i for $\alpha = 1, 0.8, 0.5$. For $\alpha = 1$, the bound is very close to the near-optimal lower bound in [9]. We found that $LB\ 2$ is generally a better bound that $LB\ 1$, except when i is large with $\alpha < 1$. The intuition for the latter case is that it is more efficient to decode the positions of all the insertions (since i is large) rather than just the complementary insertions.

IV. Deletion Channel

In this channel, each input bit is deleted with probability d, or retained with probability $1-d$. To motivate our bounding technique, consider the pair $\underline{X} = 000111000$, $\underline{Y} = 0010$. For

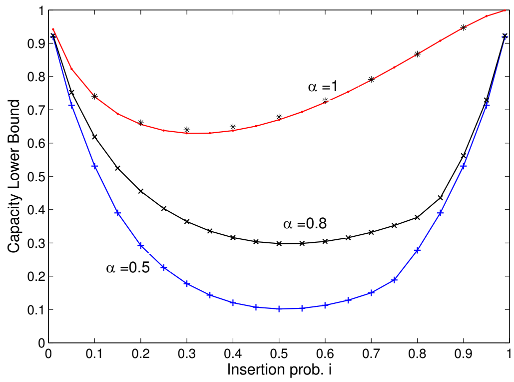

Fig. 1: Lower bound $\max\{LB1, LB2\}$ on the insertion capacity $C(i,\alpha)$. For $\alpha = 1$, the lower bound of [9] is also shown using '*'.

this pair, we can associate each *run* of $\underline{Y}$ uniquely with a run in $\underline{X}$, and hence we can write

$$\begin{aligned} &P(\underline{Y} = 0010|\underline{X} = 000111000) \\ &= P(L_1^Y = 2|L_1^X = 3)P(L_2^Y = 1|L_2^X = 3)P(L_3^Y = 1|L_3^X = 3) \end{aligned}$$

where L_j^X, L_j^Y denote the lengths of the jth runs of X and Y, respectively. We observe that if no runs in $\underline{X}$ are completely deleted, then the conditional distribution of $\underline{Y}$ given $\underline{X}$ may be written as a product distribution of run-length transformations:

$$P(\underline{Y}|\underline{X}) = P(L_1^Y|L_1^X)P(L_2^Y|L_2^X)P(L_3^Y|L_3^X)\ldots \tag{12}$$

where for all runs j,

$$P(L_j^Y = s|L_j^X = r) = \binom{r}{s} d^{r-s}(1-d)^s,\ 1 \le s \le r. \tag{13}$$

Therefore, if the deletion process acting on $\underline{X}$ to generate $\underline{Y}$ did not completely delete any runs of $\underline{X}$, the joint distribution of $(\underline{X}, \underline{Y})$ can be characterized in terms of a single-letter distribution of run-lengths determined by (2) and (13). However, we do have runs of $\underline{X}$ that are completely deleted. For example, if $\underline{X} = 000111000$ and $\underline{Y} = 000$, we cannot associate the single run in $\underline{Y}$ uniquely with a run in $\underline{X}$.

For any input-output pair (X^n, Y^{M_n}), define an auxiliary sequence $S^{M_n+1} = (S_1, S_2, \ldots, S_{M_n+1})$, where $S_j \in \mathbb{N}_0$ is the number of runs *completely* deleted in X^n between the bits corresponding to Y_{j-1} and Y_j. S_1 is the number of runs deleted before the output symbol Y_1, and S_{M_n+1} is the number of runs deleted after the last output symbol Y_{M_n}. For example, if $\underline{X} = 00\underbrace{\mathit{011100}}0$ and the bits shown in italics were deleted to give $\underline{Y} = 000$, then $\underline{S} = (0, 0, 1, 0)$.

The auxiliary sequence $\underline{S}$ enables us to augment $\underline{Y}$ with the positions of missing runs. Consider $\underline{X} = 000111000$, as before. If the decoder were given $\underline{Y} = 000$ and $\underline{S} = (0, 0, 0, 2)$, it can form the augmented sequence $\underline{Y}' = 000 - -$, where a $-$ denotes a missing run, or equivalently a 'run of length 0' in $\underline{Y}$. With the "$-$" markers indicating deleted runs, we can associate each run of the augmented sequence $\underline{Y}'$ uniquely with a run in $\underline{X}$. Thus we have

$$P(\underline{X}, \underline{Y}') = P(L_1^X)P(L_1^{Y'}|L_1^X) \cdot P(L_2^X)P(L_2^{Y'}|L_2^X)\ldots$$

where $\forall j$,

$$P(L_j^{Y'} = s|L_j^X = r) = \binom{r}{s} d^{r-s}(1-d)^s,\ 0 \le s \le r.$$

Using the auxiliary sequence S^{M_n+1}, we can decompose $H_P(X^n|Y^{M_n})$ as

$$H_P(X^n|Y^{M_n}) = H_P(X^n, S^{M_n+1}|Y^{M_n}) - H_P(S^{M_n+1}|X^n Y^{M_n}).$$

We therefore have

$$\limsup_{n\to\infty} \frac{H_P(X^n|Y^{M_n})}{n} \le \limsup_{n\to\infty} \frac{H_P(X^n, S^{M_n+1}|Y^{M_n})}{n}. \tag{14}$$

Using this in (4), we obtain a lower bound on the deletion capacity.

Theorem 3: The deletion channel capacity $C(d)$ can be lower bounded as

$$C(d) \ge \max_{0<\gamma<1} h(\gamma) - (1-d)H(S_2|Y_1Y_2) - (1-\gamma)H(L_X|L_{Y'})$$

where

$$\begin{aligned} H(S_2|Y_1Y_2) =& \gamma\bar{\theta}\log_2\frac{q}{\bar{\theta}} + \frac{\beta\theta}{\bar{\theta}^2}\log_2\frac{1}{\theta} + \frac{\beta\theta}{1-\theta^2}\log_2\frac{q}{\beta} \\ &+ \frac{\beta}{1-\theta^2}\log_2\frac{\bar{q}}{\beta}, \end{aligned}$$

$$q = \frac{\gamma + d - 2\gamma d}{1 + d - 2\gamma d},\quad \theta = \frac{\bar{\gamma} d}{1-\gamma d},\quad \beta = \frac{\bar{\gamma}\bar{d}}{(1-\gamma d)^2},\ \text{and}$$

$$\begin{aligned} H(L_X|L_{Y'}) =& \left(\frac{d}{\bar{\gamma}} - \frac{d\bar{\gamma}}{(1-\gamma d)^2}\right)\log_2\frac{1}{\gamma d} + \frac{d\bar{\gamma}h(d\gamma)}{(1-d\gamma)^2} \\ &- \frac{\bar{d}(2-\gamma-\gamma d)\log_2(1-\gamma d)}{\bar{\gamma}(1-\gamma d)} \\ &- \frac{\bar{\gamma}}{\gamma}\sum_{k=1}^{\infty}\sum_{j=1}^{\infty}(\bar{d}\gamma)^k (d\gamma)^j \binom{j+k}{k}\log_2\binom{j+k}{k}. \end{aligned}$$

Proof Sketch: In (14), we first show that $\lim_{n\to\infty}\frac{1}{n}H(S^{M_n+1}|Y^{M_n})$ exists and is equal to $(1-d)H(S_2|Y_1Y_2)$. This is because the process $\{(S_n, Y_n)\}_{n\ge1}$ is first-order Markov with the following joint distribution for all $m \in \mathbb{N}$:

$$P(S^m, Y^m) = P(Y_1, S_1)\prod_{j=2}^{m} P(Y_j, S_j|Y_{j-1}).$$

The other term in (14) is $\frac{1}{n}H_P(X^n|S^{M_n+1}Y^{M_n})$, which can be shown to converge to $(1-\gamma)H(L_X|L_{Y'})$ since there is a one-to-one correspondence between the runs of X^n and the runs of $Y'^{M_n} \leftrightarrow (S^{M_n+1}, Y^{M_n})$.

Figure 2 shows the lower bound of Theorem 3 for various values of d. We observe that our bound is close to, but smaller than the bound of [2]. ([2] contains the best lower-bounds on the deletion capacity, except for some values of d for which a slight improvement is reported in [3].) Our approach incurs a small loss in performance for the deletion channel, but is

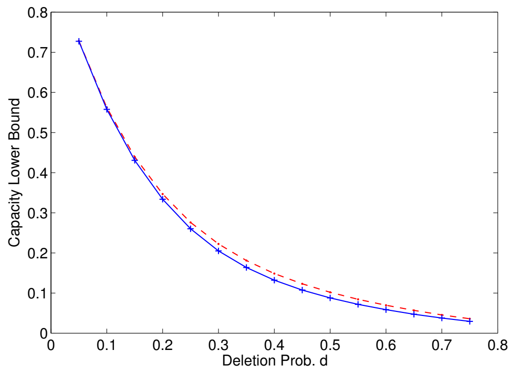

Fig. 2: Lower bound of Theorem 3 on the deletion capacity. The lower bound from [2] is shown in dashed lines.

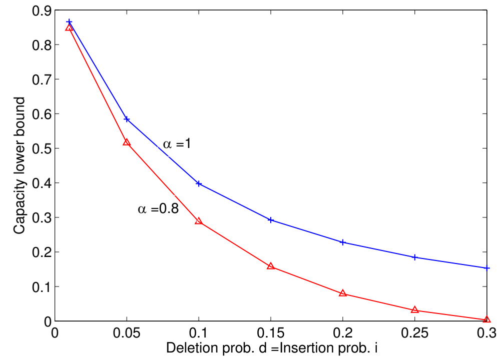

Fig. 3: Lower bound on deletion+insertion channel capacity

general enough to yield computable bounds for channels with both insertions and deletions. In contrast, the techniques of [3] are apply only to channels with deletions and duplications.

V. The Deletion+Insertion Channel

This channel is defined by three parameters (d, i, α). Each input bit undergoes a deletion with probability d, a duplication with probability $i\alpha$, a complementary insertion with probability $i\bar{\alpha}$. We obtain a lower bound on the capacity by combining the ideas of Sections III-B and IV. Introduce two auxiliary sequences $T^{M_n} = (T_1, \dots, T_{M_n})$, and $S^{M_n+1} = (S_1, \dots, S_{M_n+1})$. T^{M_n} indicates the complementary insertions in Y^{M_n}: $T_j = 1$ if Y_j is a complementary insertion, and $T_j = 0$ otherwise. S^{M_n+1} indicates the positions of the missing runs: $S_j = k$, if k runs were deleted between Y_{j-1} and Y_j. Using these auxiliary sequences, we have

$$
\begin{aligned}
&H_P(X^n|Y^{M_n}) \leq H(X^n, T^{M_n}, S^{M_n+1}|Y^{M_n}) \\
&\leq H(T^{M_n}|Y^{M_n}) + H(S^{M_n+1}|T^{M_n}Y^{M_n}) + H(X^n|S^{M_n+1}\tilde{Y}^{M_n})
\end{aligned}
$$

where $\tilde{Y}^{M_n}$ is the sequence formed from (T^{M_n}, Y^{M_n}) by flipping bit Y_j whenever $T_j = 1$. Using the above inequality to upper bound for $\limsup_{n\to\infty} \frac{1}{n} H_P(X^n|Y^{M_n})$ in (4), we obtain a lower bound on the capacity of the deletion+insertion channel.

Definition 2: Define a joint distribution $P(L^X, L^{Y'})$ with $P(L^X)$ given by (2), and for $r \in \mathbb{N}$, $0 \leq s \leq 2r$, $p_{s|r} \triangleq P(L^{Y'} = s|L^X = r)$ is given by

$$
p_{s|r} = \sum_{n_i \in \mathcal{I}} \binom{r}{n_i, r+n_i-s} i^{n_i} d^{r+n_i-s} (1-d-i)^{s-2n_i}
$$

where $\mathcal{I}$, the set of possible values for the number of insertions n_i, is given by

$$
\mathcal{I} = \{0, 1, \dots, \lfloor \frac{s}{2} \rfloor\} \text{ for } s \leq r, \text{ and } \{s-r, \dots, \lfloor \frac{s}{2} \rfloor\} \text{ for } s > r.
$$

Theorem 4: The capacity of the deletion+insertion channel can be lower bounded as

$$
\begin{aligned}
C(d, i, \alpha) \geq \max_{0<\gamma<1} &\, h(\gamma) - (\bar{q}\bar{d} + qi\bar{\alpha}) h\left(\frac{i\bar{\alpha}}{\bar{q}\bar{d} + qi\bar{\alpha}}\right) \\
&- \bar{d}(A_1 + A_2 - \frac{\theta\beta}{(1-\theta)^2}\log_2\theta) - \bar{\gamma} H_P(L_1^X|L_1^{Y'})
\end{aligned}
$$

where q, β, θ are defined in Theorem 3, $H_P(L^X|L^{Y'})$ is computed using the joint distribution in Definition 2, and

$$
\begin{aligned}
A_1 = &\frac{\theta\beta(1-i'\bar{\alpha})}{1-\theta^2}\log_2\frac{i'\alpha + (1-i'\bar{\alpha})q + i'\bar{\alpha}\bar{q}}{\beta(1-i'\bar{\alpha})} \\
&+ \frac{\theta^2\beta i'\bar{\alpha}}{1-\theta^2}\log_2\frac{i'\alpha + (1-i'\bar{\alpha})q + i'\bar{\alpha}\bar{q}}{\beta i'\bar{\alpha}} \\
&+ D_1\log_2\frac{i'\alpha + (1-i'\bar{\alpha})q + i'\bar{\alpha}\bar{q}}{D_1}, \\
A_2 = &\frac{\theta^2\beta(1-i'\bar{\alpha})}{1-\theta^2}\log_2\frac{(1-i'\bar{\alpha})\bar{q} + i'\bar{\alpha}q}{\beta(1-i'\bar{\alpha})} + \\
&\frac{\theta\beta i'\bar{\alpha}}{1-\theta^2}\log_2\frac{(1-i'\bar{\alpha})\bar{q} + i'\bar{\alpha}q}{\beta i'\bar{\alpha}} + D_2\log_2\frac{(1-i'\bar{\alpha})\bar{q} + i'\bar{\alpha}q}{D_2}
\end{aligned}
$$

with $i' \triangleq i/\bar{d}$, $D_1 \triangleq \gamma(1-i'\bar{\alpha})\bar{\theta} + i'\bar{\alpha}\beta + i'\alpha$, and $D_2 \triangleq \gamma i'\bar{\alpha}\bar{\theta} + (1-i'\bar{\alpha})\beta$.

The lower bound is plotted in Figure 3 for various values of $d = i$, for $\alpha = 0.8$ and for $\alpha = 1$.

References

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

[2] 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, 2007.

[3] E. Drinea and A. Kirsch, "Directly lower bounding the information capacity for channels with i.i.d.deletions and duplications," *IEEE Trans. Inf. Theory*, vol. 56, pp. 86–102, January 2010.

[4] S. Diggavi, M. Mitzenmacher, and H. Pfister, "Capacity upper bounds for the deletion channel," in *Proc. Int. Symp. on Inf. Theory*, 2007.

[5] M. Mitzenmacher, "A survey of results for deletion channels and related synchronization channels," *Probability Surveys*, vol. 6, pp. 1–33, 2009.

[6] D. Fertonani and T. M. Duman, "Novel bounds on the capacity of the binary deletion channel," *IEEE Trans. Inf. Theory*, vol. 56, pp. 2753–2765, June 2010.

[7] A. Kalai, M. Mitzenmacher, and M. Sudan, "Tight asymptotic bounds for the deletion channel with small deletion probabilities," in *Proc. Int. Symp. on Inf. Theory*, 2010.

[8] Y. Kanoria and A. Montanari, "On the deletion channel with small deletion probability," in *Proc. Int. Symp. on Inf. Theory*, 2010.

[9] M. Mitzenmacher, "Capacity bounds for sticky channels," *IEEE Trans. on Inf. Theory*, pp. 72–77, January 2008.

[10] R. G. Gallager, "Sequential decoding for binary channels with noise and synchronization errors," October 1961. Lincoln Lab Group Report.

[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, 2011.

[12] R. L. Dobrushin, "Shannon's theorems for channels with synchronization errors," *Problemy Peredachi Informatsii*, vol. 3, no. 4, pp. 18–36, 1967.