Improved Capacity Lower Bounds for Channels with Deletions and Insertions

Ramji Venkataramanan
Dept. of Engineering
University of Cambridge
ramji.v@eng.cam.ac.uk

Sekhar Tatikonda
Dept. of Electrical Engineering
Yale University
sekhar.tatikonda@yale.edu

***Abstract*—New lower bounds are obtained for the capacity of a binary channel with deletions and insertions. Each input bit to the channel is deleted with probability d, or an extra bit is inserted after it with probability i, or it is transmitted unmodified with probability $1-d-i$. This paper builds on the idea introduced in [1] of using a sub-optimal decoder that decodes the positions of the deleted and inserted runs, in addition to the transmitted codeword. The mutual information between the channel input and output sequences is expressed as the sum of the rate achieved by this decoder and the rate loss due to its sub-optimality. The main contribution is an analytical lower bound for the rate loss term which leads to an improvement in the capacity lower bound of [1]. For the special case of the deletion channel, the new bound is larger than the previous best lower bound for deletion probabilities up to 0.3.**

I. Introduction

We consider the problem of obtaining computable capacity bounds for the following binary channel with deletions and insertions. Each input bit is independently transformed by the channel in one of three ways: it is deleted with probability d, or an extra bit is inserted after it with probability i, or the bit is transmitted unmodified with probability $1-d-i$. When an insertion occurs, the inserted bit is equal to the input bit (*duplication*) with probability α, and is the complement of the input bit (*complementary insertion*) with probability $1-\alpha$. We refer to this channel as the 'InDel' channel with parameters (d, i, α). When $i = 0$, we obtain the well-studied i.i.d deletion channel [2]–[9]. Setting $d = 0$ and $\alpha = 1$ gives us the elementary sticky channel [10].

Capacity lower bounds for the InDel channel were obtained in [1] by characterizing the rate achievable by using a first-order Markov codebook with a run-syncing decoder. In addition to the transmitted codeword, this decoder decodes auxiliary sequences which indicate the positions of deleted and inserted runs in the received sequence. The run-syncing decoder is sub-optimal because of the extra information decoded, but its analysis is tractable. The rate achievable via such a decoder yields a lower bound on the InDel capacity [1]. The mutual information between the input and the output sequences can be expressed as the sum of two terms: 1) the rate achieved by the run-syncing decoder, and 2) the rate loss due to the sub-optimality of this decoder (compared to a max-likelihood decoder). The main contribution of this paper is an analytical lower bound on the rate loss term, which combined with the first term, results in an improved capacity lower bound.

In Section III, we consider the special case of the deletion channel. We lower bound the rate penalty incurred by decoding the positions of the deleted runs, and use it to obtain a new capacity lower bound which is higher than the best previous bound [3] for deletion probabilities up to 0.3. We also compare the mutual information decomposition used here with the one obtained from jigsaw decoding scheme of [3], [4].

In Section IV, we bound the rate penalty incurred by decoding the complementary insertions (inserted runs). This bound is combined with the results of Sections III to obtain an improved lower bound on the InDel capacity. Due to space constraints, the discussion in Section IV is brief. The reader is referred to [11] for additional details, including proofs.

II. Preliminaries

Notation: For $\alpha \in [0,1]$, $\bar{\alpha} \triangleq 1-\alpha$. Logarithms are with base 2 and entropy is measured in bits. $h(.)$ denotes the binary entropy function. We use uppercase letters to denote random variables, bold-face letters for random processes, and superscript notation to denote random vectors.

The communication over the channel is characterized by three random processes defined over the same probability space: the input process $\mathbf{X} = \{X_n\}_{n\geq 1}$, the output process $\mathbf{Y} = \{Y_n\}_{n\geq 1}$, and $\mathbf{M} = \{M_n\}_{n\geq 1}$, where M_n is the number of output symbols corresponding to the first n input symbols. The length n channel input sequence is $X^n = (X_1, \dots X_n)$ and the output sequence is Y^{M_n}. Note that M_n is a random variable determined by the channel realization. For brevity, we sometimes use underlined notation for random vectors when we do not need to be explicit about their length. Thus $\underline{X} = (X_1, X_2, \dots, X_n)$ and $\underline{Y} = (Y_1, \dots, Y_{M_n})$.

Dobrushin [12] provided a capacity characterization for a general class of synchronization channels which includes the InDel channel.

Fact 1. *Let* $C_n = \sup_{P_{X^n}} \frac{1}{n} I(X^n; Y^{M_n})$. *Then* $C = \lim_{n\to\infty} C_n$ *exists, and is equal to the InDel capacity.*

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 work, we choose the input distribution to be the class of 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 the deletion channel [2], [3] and the elementary sticky channel [10], both special cases of the InDel channel. First-order Markov sequences have runs that are independent, and the average run length can be controlled via the Markov parameter. This fits well with the techniques used in this paper, which are based on the relationship between input and output runs in an InDel channel.

The input process $\mathbf{X} = \{X_n\}_{n\geq 1}$ is characterized by the following distribution for all n:

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

where for $x \in \{0,1\}$, $P(X_1 = x) = \frac{1}{2}$ and

$$P(X_j = x|X_{j-1} = x) = \gamma, \quad j > 1. \tag{1}$$

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). The value of the first bit of $\mathbf{X}$ can be communicated to the decoder with vanishing rate; 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 the binary first-order Markov process in (1), the run-lengths are independent and geometrically distributed, i.e., $\Pr(L_j^X = r) = \gamma^{r-1}(1-\gamma)$ for $r = 1,2,\ldots$ 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.

We use subscript P to denote the mutual information and entropy quantities computed with the input distribution (1). For all n, we have

$$C_n = \sup_{P_{X^n}} \frac{1}{n} I(X^n; Y^{M_n}) > \frac{1}{n} I_P(X^n; Y^{M_n}). \tag{2}$$

Therefore

$$\begin{aligned} C &> \liminf_{n\to\infty} \frac{1}{n} I_P(X^n; Y^{M_n}) \\ &= h(\gamma) - \limsup_{n\to\infty} \frac{1}{n} H_P(X^n|Y^{M_n}) \end{aligned} \tag{3}$$

where $h(\gamma)$ is the entropy rate of the Markov process $\mathbf{X}$. We derive upper bounds on $\limsup_{n\to\infty} \frac{1}{n} H_P(X^n|Y^{M_n})$ and use them in (3) to obtain lower bounds on the capacity.

III. Deletion Channel

Consider the following pair of input and output sequences for the deletion channel: $\underline{X} = 000111000$, $\underline{Y} = 0010$. For this pair, we can associate each run of $\underline{Y}$ uniquely with a run in $\underline{X}$. Therefore the conditional probability $P(\underline{Y} = 0010|\underline{X} = 000111000)$ equals

$$P(L_1^Y = 2|L_1^X = 3)P(L_2^Y = 1|L_2^X = 3)P(L_3^Y = 1|L_3^X = 3)$$

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. In general, there *are* 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 Y_1, and S_{M_n+1} is the number of runs deleted after 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 $\underline{S} = (0,0,1,0)$. On the other hand, if the last six bits were all deleted, i.e., $\underline{X} = 000\underbrace{\mathit{111000}}$, then $\underline{S} = (0,0,0,2)$. Thus $\underline{S}$ is not uniquely determined given $(\underline{X}, \underline{Y})$.

The auxiliary sequence $\underline{S}$ lets us augment $\underline{Y}$ with the positions of missing runs. If the decoder is 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 $\underline{X}$. Denote by $L_1^{Y'}, L_2^{Y'}, \ldots$ the run-lengths of the augmented sequence $\underline{Y}'$, where $L_j^{Y'} = 0$ if the jth run is a $-$. Then we have

$$P(\underline{X}, \underline{Y}') = P(L_1^X)P(L_1^{Y'}|L_1^X) \cdot P(L_2^X)P(L_2^{Y'}|L_2^X)\ldots \tag{4}$$

where $\forall j$:

$$\begin{aligned} P(L_j^X = r) &= \gamma^{r-1}(1-\gamma), \quad r = 1,2,\ldots \\ P(L_j^{Y'} = s|L_j^X = r) &= \binom{r}{s} d^{r-s}(1-d)^s, \quad 0 \leq s \leq r. \end{aligned} \tag{5}$$

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

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

We therefore have

$$\begin{aligned} &\liminf_{n\to\infty} \tfrac{1}{n} I_P(X^n; Y^{M_n}) = h(\gamma) - \limsup_{n\to\infty} \tfrac{1}{n} H_P(X^n|Y^{M_n}) \\ &\geq \underbrace{h(\gamma) - \limsup_{n\to\infty} \tfrac{1}{n} H_P(X^n, S^{M_n+1}|Y^{M_n})}_{\text{rate of sub-optimal decoder}} \\ &\quad + \underbrace{\liminf_{n\to\infty} \tfrac{1}{n} H_P(S^{M_n+1}|X^n, Y^{M_n})}_{\text{penalty term}}. \end{aligned} \tag{6}$$

The first part of (6) is the rate achieved by using a decoder that decodes the positions of the deleted runs in addition to the transmitted codeword. An analytical expression for this term was obtained in [1]. The second term measures the rate penalty incurred by decoding the auxiliary sequence $\underline{S}$.

A. Bounding the penalty term

To get some intuition about the penalty term $H(S^{M_n+1}|Y^{M_n}, X^n)$, let us consider the following example.

$$\underline{X} = \overbrace{00000}^{z \text{ bits}}111\overbrace{00000}^{r \text{ bits}} \quad \longrightarrow \quad \underline{Y} = \overbrace{000}^{s \text{ bits}} \tag{7}$$

Given $(\underline{X}, \underline{Y})$ the uncertainty in $\underline{S}$ corresponds to how many of the s output bits came from the first run of zeros in $\underline{X}$, and how many came from the second. In (7), $\underline{S}$ can be

$$\mathcal{F}_{j,z,r,s} = \left\{ (\underline{X}, \underline{Y}, \Theta(\underline{Y} \backslash \underline{Y}(j))) : (\underbrace{c, c, , \ldots, c}_{z \text{ bits}}, \underbrace{\bar{c}, \bar{c}, \ldots, \bar{c}}_{k \text{ bits}}, \underbrace{c, c, \ldots, c}_{r \text{ bits}}) \longrightarrow \underline{Y}(j) = (\underbrace{c, c, \ldots, c}_{s \text{ bits}}) \text{ for some } k \geq 1, \ c \in \{0, 1\} \right\} \tag{10}$$

$$\begin{aligned}
&\sum_{\underline{x},\underline{y}} \sum_{j=1}^{R(\underline{y})} \sum_{\substack{\underline{\theta}:(\underline{x},\underline{y},\underline{\theta}) \in \\ \mathcal{F}_{j,z,r,s}}} P(\underline{X}, \underline{Y}, \Theta(\underline{Y} \backslash \underline{Y}(j)) = \underline{x}, \underline{y}, \underline{\theta}) \\
&= \sum_{\underline{y}} P(\underline{Y} = \underline{y}) \sum_{j=1}^{R(\underline{y})} \mathbf{1}(j\text{th } \underline{y}\text{-run has length } s) \ P(\text{runs of } \underline{X} \text{ with lengths } (z, k, r) \ \to \ j\text{th } \underline{y}\text{-run, for some } k \geq 1) \\
&= \mathbb{E}\Big[\mathbb{E}\Big[\sum_{j=1}^{R(\underline{Y})} \mathbf{1}(\text{runs of } \underline{X} \text{ with lengths } (z, k, r) \ \to \ j\text{th } \underline{Y}\text{-run, for some } k \geq 1) \ \mathbf{1}(j\text{th } \underline{Y}\text{-run has length } s) \mid R(\underline{Y}) \Big]\Big] \\
&\overset{(a)}{=} \mathbb{E}\Big[\sum_{j=1}^{R(\underline{Y})} \gamma^{z-1}\bar{\gamma} \ \Big(\sum_{k=1}^{\infty} \gamma^{k-1}\bar{\gamma} d^k\Big) \ \gamma^{r-1}\bar{\gamma} \cdot \binom{z+r}{s} (1-d)^s d^{z+r-s} \Big] \ \overset{(b)}{=} n\bar{d}\bar{q} \cdot \frac{\bar{\gamma}^3 d}{\gamma^2(1-\gamma d)} (\gamma d)^{z+r} \binom{z+r}{s} \left(\frac{1-d}{d}\right)^s .
\end{aligned} \tag{11}$$

---

one of four sequences: $(2, 0, 0, 0), (0, 0, 0, 2)$, $(0, 1, 0, 0)$ and $(0, 0, 1, 0)$. The first case corresponds to all the output bits coming from the second run of zeros; in the second case all the output bits come from the first run. The third and fourth cases correspond to the output bits coming from both input runs of zeros. The probability of the deletion patterns that result in each of these possibilities can be calculated. We can thus compute $H(\underline{S}|\underline{X}, \underline{Y})$ precisely for this example. For general $(\underline{X}, \underline{Y})$, we lower bound $H(\underline{S}|\underline{X}, \underline{Y})$ by considering patterns in $(\underline{X}, \underline{Y})$ of the form shown in (7).

Lemma 1. *$\liminf_{n\to\infty} \frac{1}{n} H_P(S^{M_n+1}|Y^{M_n}, X^n) \geq \Phi(d, \gamma)$ where*

$$\Phi(d, \gamma) = \frac{\bar{d} \ \bar{q} \ \bar{\gamma}^3 \ d}{\gamma^2 \ (1-\gamma d)} \sum_{z,r=1}^{\infty} (\gamma d)^{z+r} \sum_{s=1}^{z+r} \left(\frac{\bar{d}}{d}\right)^s \binom{z+r}{s} \cdot H\left(\left\{ \frac{\binom{z}{l}\binom{r}{s-l}}{\binom{z+r}{s}} \right\}_{l=0,\ldots,s} \right) \tag{8}$$

where $q = \frac{\gamma+d-2\gamma d}{1+d-2\gamma d}$ and $H(\{p_i\})$ is the entropy of the pmf $\{p_i\}$. (In (8) is assumed that $\binom{n}{k} = 0$ for $k > n$.)

Proof: We expand $H(\underline{S}|\underline{Y}, \underline{X})$ in terms of the runs of $\underline{Y}$. We denote the number of runs in $\underline{Y}$ by $R(\underline{Y})$, the jth run of $\underline{Y}$ by $\underline{Y}(j)$ and the corresponding part of $\underline{S}$ by $\underline{S}(j)$. We have

$$\begin{aligned}
H(S^{M_n+1} \mid Y^{M_n}, X^n) &= \sum_{\underline{x},\underline{y}} P(\underline{X} = \underline{x}, \underline{Y} = \underline{y}) \\
&\quad \cdot \sum_{j=1}^{R(\underline{y})} H(\underline{S}(j) \mid \underline{S}(1), \ldots, \underline{S}(j-1), \underline{X} = \underline{x}, \underline{Y} = \underline{y}) \\
&\geq \sum_{\underline{x},\underline{y}} P(\underline{X} = \underline{x}, \underline{Y} = \underline{y}) \sum_{j=1}^{R(\underline{y})} H(\underline{S}(j)|\underline{X} = \underline{x}, \underline{Y} = \underline{y}, \Theta(\underline{Y} \backslash \underline{Y}(j)))
\end{aligned} \tag{9}$$

where $\underline{Y} \backslash \underline{Y}(j)$ is the sequence obtained by removing $\underline{Y}(j)$ from $\underline{Y}$. $\Theta(\underline{Y} \backslash \underline{Y}(j))$ denotes the exact deletion pattern corresponding to the output bits $\underline{Y} \backslash \underline{Y}(j)$, i.e., it tells us which bit in $\underline{X}$ corresponds to each bit in $\underline{Y} \backslash \underline{Y}(j)$. The inequality in (9) holds since $\underline{S}(1), \ldots, \underline{S}(j-1)$ is determined by $\Theta(\underline{Y} \backslash \underline{Y}(j))$.

We obtain an analytical lower bound for the right side of (9) by considering only those terms for which the run $\underline{Y}(j)$ is generated from either one or three adjacent runs in $\underline{X}$, as in (7). For $z, r \geq 1$ and $1 \leq s \leq z + r$, define the set $\mathcal{F}_{j,z,r,s}$ as in (10) at the top of this page. We allow the possibility that all of $\underline{Y}(j)$ is generated from just one of the three runs; further note that the $\bar{c}$-run in the middle is always deleted. The right side of (9) is lower bounded by considering only triples $(\underline{X}, \underline{Y}, \Theta(\underline{Y} \backslash \underline{Y}(j)) \in \mathcal{F}_{j,z,r,s}$, as follows.

$$\begin{aligned}
&H(S^{M_n+1} \mid Y^{M_n}, X^n) \\
&\geq \sum_{z,r\geq 1} \sum_{s=1}^{z+r} \sum_{\underline{x},\underline{y}} \sum_{j=1}^{R(\underline{y})} \sum_{\substack{\underline{\theta}:(\underline{x},\underline{y},\underline{\theta}) \in \\ \mathcal{F}_{j,z,r,s}}} P(\underline{X}, \underline{Y}, \Theta(\underline{Y} \backslash \underline{Y}(j)) = \underline{x}, \underline{y}, \underline{\theta}) \\
&\quad \cdot H(\underline{S}(j) \mid \underline{X}, \underline{Y}, \Theta(\underline{Y} \backslash \underline{Y}(j)) = \underline{x}, \underline{y}, \underline{\theta}).
\end{aligned} \tag{12}$$

$H(\underline{S}(j) \mid \underline{Y}, \underline{X}, \Theta(\underline{Y} \backslash \underline{Y}(j)) = \underline{x}, \underline{y}, \underline{\theta})$ can be computed for $(\underline{x}, \underline{y}, \underline{\theta}) \in \mathcal{F}_{j,z,r,s}$ as follows. The length-s vector $\underline{S}(j)$ has at most one non-zero element: For $l = 0, \ldots, s-1$, if $\underline{Y}(j)$ was formed with l bits from the first length-z run and $s - l$ bits from the third length-r run, $\underline{S}(j)$ will have a non-zero in position $l + 1$. If $\underline{Y}(j)$ was formed with all s bits from the first length-z run, then all the s elements of $\underline{S}(j)$ are zero and the symbol in $\underline{S}$ immediately after $\underline{S}(j)$ is non-zero. We thus have

$$H(\underline{S}(j) | \ \underline{X}, \underline{Y}, \Theta(\underline{Y} \backslash \underline{Y}(j)) = \underline{x}, \underline{y}, \underline{\theta}) = H\Big(\Big\{ \frac{\binom{z}{l}\binom{r}{s-l}}{\binom{z+r}{s}} \Big\}_{l=0}^{s} \Big) \tag{13}$$

Next, for a fixed (z, r, s), we compute the three innermost sums of $P(\underline{X}, \underline{Y}, \Theta(\underline{Y} \backslash \underline{Y}(j)) = \underline{x}, \underline{y}, \underline{\theta})$ in (12). This is done in (11) at the top of the previous page. In the third line of (11), each term of the inner expectation is the probability of three

TABLE I. CAPACITY LOWER BOUND FOR THE DELETION CHANNEL

d	0.05	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5	0.55	0.6
LB Thm.1	**0.7291**	**0.5638**	**0.4414**	**0.3482**	**0.2770**	**0.2225**	0.1805	0.1478	0.1217	0.1005	0.0830	0.0682
Optimal γ	0.535	0.575	0.62	0.67	0.72	0.77	0.81	0.84	0.87	0.89	0.91	0.925
LB of [3]	0.7283	0.5620	0.4392	0.3467	0.2759	0.2224	0.1810	0.1484	0.1229	0.1019	0.0843	0.0696

successive $\underline{X}$-runs having the specified lengths and giving rise to a $\underline{Y}$-run of length s. For (a), we use the fact that $\underline{X}$ is first-order Markov and thus has independent runs. (b) holds because $\underline{Y}$ is first-order Markov with parameter q and has expected length $n\bar{d}$. The expected number of runs in $\underline{Y}$ equals $\bar{q}$ times the expected length of $\underline{Y}$. Substituting (13) and (11) in (12), and dividing both sides by n yields the lemma. ■

Theorem 1. *The deletion capacity $C(d)$ satisfies*

$$C(d) \geq \max_{0<\gamma<1} h(\gamma) - \bar{d}\, H(S_2|Y_1Y_2) - \bar{\gamma}\, H(L_X|L_{Y'}) + \Phi(d,\gamma)$$

where

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

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

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

Proof: The first part of (6) (rate achieved by the sub-optimal decoder) was shown in [1, Theorem 3] to be equal to $h(\gamma) - \bar{d}\, H(S_2|Y_1Y_2) - \bar{\gamma}\, H(L_X|L_{Y'})$. The result is obtained by using Lemma 1 to lower bound the penalty term in (6). ■

Table I shows the lower bound of Theorem 1 for various values of d together with $\gamma \in (0,1)$ optimized with a resolution of 0.005. The values shown in bold are those for which the bound improves on the previous best lower bound [3].

A sharper lower bound on the penalty term will further improve the capacity bound of Theorem 1. The lower bound for $\frac{1}{n}H(\underline{S}|\underline{X},\underline{Y})$ in Lemma 1 can be refined by separately considering the cases where $\underline{Y}$-run arose from $3/5/7/\ldots$ $\underline{X}$-runs. However, this will imply a more complicated expression than the one in (8).

It is interesting to compare the penalty term $\frac{1}{n}H(\underline{S}|\underline{X},\underline{Y})$ with the rate penalty incurred by the jigsaw decoding scheme of [3]. The jigsaw decoder decodes the type of each run in the output sequence $\underline{Y}$. The type of a $\underline{Y}$-run is the set of input runs that gave rise to it, with the first input run in the set contributing at least one bit. The penalty for decoding the codeword by first decoding the sequence of types is $\frac{1}{n}H(\text{types of } \underline{Y}\ |\underline{X},\underline{Y})$. A characterization of this conditional entropy in terms of the joint distribution of $(\underline{X},\underline{Y})$ is derived in [4].

Given a pair $(\underline{X},\underline{Y})$, knowledge of $\underline{S}$ uniquely determines the sequence of types of $\underline{Y}$, but not vice versa. For example, consider the pair $\underline{X} = 1010101$, $\underline{Y} = 1101$. Suppose we know that $\underline{S} = (0,3,0,0,0)$, i.e., $\underline{Y}$ can be augmented with deleted runs as $1---101$. Then the types of the three $\underline{Y}$ runs are

$$\{10101\} \to 11, \quad \{0\} \to 0, \quad \{1\} \to 1. \tag{14}$$

In contrast, suppose we know that the set of types for the $(\underline{X},\underline{Y})$ pair above is given by (14). Then $\underline{S} = (0,1,2,0,0)$ and $\underline{S} = (0,3,0,0,0)$ (corresponding to deletion patterns $1-1--01$ and $1---101$, respectively) are both consistent $\underline{S}$-sequences with the given set of types. In summary, since the set of types is a function of $(\underline{X},\underline{Y},\underline{S})$ we have

$$\tfrac{1}{n}H(\text{types of } \underline{Y}\ |\underline{X},\underline{Y}) \leq \tfrac{1}{n}H(\underline{S}|\underline{X},\underline{Y}).$$

In other words, the rate penalty incurred by the jigsaw decoder is *less* than the penalty of the sub-optimal decoder considered here. However, the penalty term for our decoder can be lower bounded analytically, which leads to improved lower bounds on the deletion capacity for $d \leq 0.3$. The jigsaw penalty term is harder to lower bound and is estimated via simulation for a few values of d in [4].

IV. InDel Channel

This channel is defined by three parameters (d,i,α) with $d+i<1$. Each input bit undergoes a deletion with probability d, a duplication with probability $i\alpha$, or a complementary insertion with probability $i\bar{\alpha}$. Given that a bit is *not* deleted, the probability that it undergoes an insertion is $\frac{i}{1-d}$. Hence the InDel channel is equivalent to a cascade of two channels: the first is a deletion channel with parameter d; the second is an insertion channel with parameters (i',α), where $i' = \frac{i}{1-d}$. The input sequence to the second channel is denoted Z^{L_n} and its output is Y^{M_n}. Z^{L_n} is Markov with parameter $q = \frac{\gamma+d-2\gamma d}{1+d-2\gamma d}$ as it is obtained by passing the Markov(γ) sequence X^n through a deletion channel [3].

We work with the cascade channel and use two auxiliary sequences, $S^{M_n+1} = (S_1,\ldots,S_{M_n+1})$ and $T^{M_n} = (T_1,\ldots T_{M_n})$. As in Section III, S^{M_n+1} indicates the positions of the missing runs: $S_j = k$, if k runs were completely deleted between Y_{j-1} and Y_j. 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. We write $H_P(X^n|Y^{M_n})$ as

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

We therefore have

$$\begin{aligned}&\liminf_{n\to\infty}\tfrac{1}{n}I_P(X^n;Y^{M_n}) \geq \\ &\underbrace{h(\gamma) - \limsup_{n\to\infty}\tfrac{1}{n}H_P(X^n,T^{M_n},S^{M_n+1}|Y^{M_n})}_{\text{rate of sub-optimal decoder}} \\ &+ \underbrace{\liminf_{n\to\infty}\tfrac{1}{n}H_P(T^{M_n},S^{M_n+1}|X^n,Y^{M_n})}_{\text{penalty term}}.\end{aligned} \tag{15}$$

The first part of (15) is the rate achieved by using a decoder that decodes the sequences $\underline{S}, \underline{T}$ (positions of deleted and inserted runs) in addition to the transmitted codeword. An analytical lower bound for this term was obtained in [1]. We now briefly discuss how the second term, the penalty incurred due to decoding $(\underline{S}, \underline{T})$, can be lower bounded.

The term $H_P(T^{M_n}, S^{M_n+1}|X^n, Y^{M_n})$ can be written as

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

The second term above can be lower bounded using the technique of Section III-A. (See [11, Lemma 18] for details.) The first term can be bounded as follows.

$$\begin{aligned} H_P(T^{M_n}|X^n, Y^{M_n}) \geq & H_P(T^{M_n}|X^n, Y^{M_n}, Z^{L_n}) \\ = & H_P(T^{M_n}|Y^{M_n}, Z^{L_n}). \end{aligned} \tag{16}$$

The equality holds due to the Markov chain $(\mathbf{X}, \mathbf{\Lambda}_{del}) - \mathbf{Z} - (\mathbf{\Lambda}_{ins}, \mathbf{Y})$, where $\mathbf{\Lambda}_{del}$ and $\mathbf{\Lambda}_{ins}$ denote the deletion and insertion patterns of the first and second channels in the cascade, respectively.

The right side of (16) is the uncertainty in the positions of complementary insertions given both the input and output sequences of the insertion channel. Consider the example

$$\underline{Z} = \; 0 \overbrace{1111}^{k_1 \text{ bits}} 0, \qquad \underline{Y} = \; 0 \; \overbrace{111111}^{k_1+k_2 \text{ bits}} \; 0. \tag{17}$$

Given this $(\underline{Z}, \underline{Y})$ pair, the only uncertainty in $\underline{T}$ is in the value of T_2 (the first bit in the run of ones). The remaining bits of $\underline{T}$ are all 0. Indeed,

- $T_2 = 1$ if the 0 undergoes a complementary insertion leading to the first 1. Then $(k_2 - 1)$ out of the k_1 1's in the $\underline{Z}$-run undergo duplications, the remaining 1's are transmitted without any insertions.
- $T_2 = 0$ if the 0 is transmitted without any insertions. k_2 out of the k_1 1's in the $\underline{Z}$-run undergo duplications, the remaining are transmitted without insertions.

Calculating the binary entropy associated with the two cases yields a lower bound for $H(\underline{T}|\underline{Y}, \underline{Z})$ in this example. For each $k_1 \geq 1$ and $1 \leq k_2 \leq k_1$, we can estimate the number of times the pattern in (17) appears in a typical $(\underline{Z}, \underline{Y})$ pair. This leads to a lower bound for the insertion penalty term.

Lemma 2. *a)* $\liminf_{n\to\infty} \frac{1}{n} H_P(T^{M_n}|Y^{M_n}, Z^{L_n}) \geq (1 - d)\Gamma(i', \alpha, q)$*, where*

$$\Gamma(i', \alpha, q) = \bar{q}^2 \sum_{k_1=1}^{\infty} \sum_{k_2=1}^{k_1} \binom{k_1}{k_2} q^{k_1-1} (i'\alpha)^{k_2} (1-i')^{k_1-k_2+1} \cdot \left(1 + \frac{\bar{\alpha}k_2}{\alpha(k_1-k_2+1)}\right) h\left(\frac{\bar{\alpha}k_2}{\bar{\alpha}k_2 + \alpha(k_1-k_2+1)}\right).$$

b) $\liminf_{n\to\infty} \frac{1}{n} H_P(S^{M_n+1}|X^n, Y^{M_n}, T^{M_n}) \geq \Phi(d, \gamma)$.

Proof: The proof of the first part is found in [11, Lemma 11]. For the second part, see [11, Lemma 18].

Using Lemma 2 in (15) together with the bound obtained in [1] for the rate of the sub-optimal decoder yields the following improved lower bound on the InDel capacity.

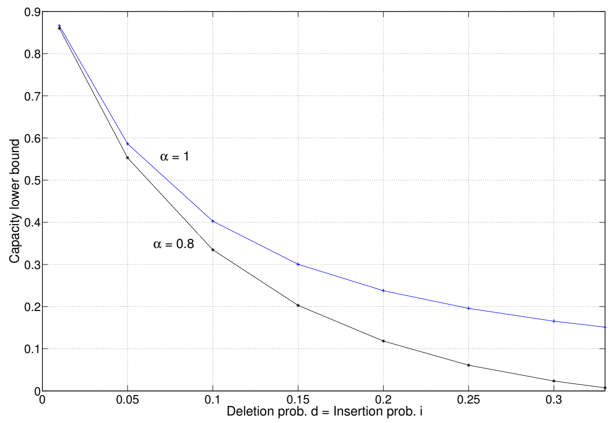

Fig. 1. Lower bound on the InDel capacity $C(d, i, \alpha)$ for $d = i$.

Theorem 2. *The InDel capacity can be lower bounded as*

$$C(d, i, \alpha) \geq \max_{0<\gamma<1} R_{sub}(\gamma) + (1-d)\Gamma(i', \alpha, q) + \Phi(d, \gamma).$$

where $R_{sub}(\gamma)$ denotes the lower bound given in [1, Theorem 4] for the rate achieved by the sub-optimal run-syncing decoder with a Markov(γ) codebook.

The bound is plotted in Figure 1 for various values of $d = i$. For $\alpha = 0.8$, the bound is positive up to $d = i = 0.33$. In comparison, the earlier lower bound [1, Theorem 4] was zero beyond $d = i = 0.3$.

REFERENCES

[1] R. Venkataramanan, S. Tatikonda, and K. Ramchandran, "Achievable rates for channels with deletions and insertions," in *Proc. IEEE Int. Symp. Inf. Theory*, 2011.

[2] 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.

[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, 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, pp. 86–102, January 2010.

[5] 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.

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

[7] Y. Kanoria and A. Montanari, "On the deletion channel with small deletion probability," in *Proc. IEEE Int. Symp. on Inf. Theory*, 2010. Expanded version at http://arxiv.org/abs/1104.5546.

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

[9] A. R. Iyengar, P. H. Siegel, and J. K. Wolf, "Modeling and information rates for synchronization error channels," in *Proc. IEEE Int. Symp. Inf. Theory*, 2011. Expanded version at http://arxiv.org/abs/1302.2702.

[10] M. Mitzenmacher, "Capacity bounds for sticky channels," *IEEE Trans. on Inf. Theory*, vol. 54, pp. 72–77, Jan. 2008.

[11] R. Venkataramanan, S. Tatikonda, and K. Ramchandran, "Achievable rates for channels with deletions and insertions," *To appear in IEEE Trans. Inf. Theory*. Available at http://arxiv.org/abs/1102.5112.

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