

Modeling and Information Rates for Synchronization Error Channels

Aravind R. Iyengar, Paul H. Siegel, and Jack K. Wolf
 University of California San Diego, La Jolla, CA 92093 – 0401, USA
 Email: {aravind, psiegel, jwolf}@ucsd.edu

Abstract—We propose a new channel model for channels with synchronization errors. Using this model, we give simple, non-trivial and, in some cases, tight lower bounds on the capacity for certain synchronization error channels.

I. INTRODUCTION

Channels with synchronization errors have been of interest from the very beginnings of information theory. However, little is known of their capacities or of good coding schemes. In the last decade, a flurry of activity has led to significant progress in estimating achievable information rates for certain synchronization error channels (SECs). A “good” coding scheme continues to be elusive.

In this paper, we model an SEC as a channel with states and use this model to arrive at some simple lower bounds on the capacity. Although the idea behind the alternative model is straightforward, the model itself has been absent in literature. While the present paper deals only with a few asymptotic results on information rates of the SEC, we think that the model presented here can be utilized to design codes for SECs in general.

The remainder of this paper is organized as follows. In Section II, we recall a few of the main results on capacity of SECs. We consider a special case of the generic SEC—the deletion, duplication channel (DDC)—and construct an equivalent channel by viewing the SEC as a channel with states in Section III. We use the model to obtain bounds on the capacity in Section IV. We conclude by highlighting the possible advantages of the model discussed in Section V.

II. SYNCHRONIZATION ERROR CHANNELS

Remark 1 (Notation): Non-random variables are written as lowercase letters, e.g. n . We denote sets by double-stroke uppercase letters, \mathbb{X} , and define $[n] = \{1, 2, \dots, n\}$, $[0] = \emptyset$ and $[m : n] = \{m, m + 1, \dots, n\}$, $m \leq n$. We assume an underlying probability space $(\Omega, \mathcal{F}, \mathbb{P})$ over which all random variables, denoted by uppercase letters X , are defined. Random vectors are denoted by uppercase letters with the set of indices as subscripts, e.g. $X_{[n]} = (X_1, X_2, \dots, X_n)$ or $X_{Y_{[n]}}$ when the set of indices is itself a random vector $Y_{[n]}$. Random processes are denoted by script letters \mathcal{X} , or as $X_{\mathbb{N}}$.

Let \mathbb{X} be a finite set. A *memoryless* synchronization error channel is specified by a stochastic matrix $\{q(\bar{y}|x), \bar{y} \in \bar{\mathbb{Y}}, x \in \mathbb{X}\}$ where \mathbb{Y} is the output alphabet and $\bar{\mathbb{Y}}$ is the set of all strings (including the empty string λ) over \mathbb{Y} . We assume that the expected length of the output string arising

from one input symbol is strictly positive and finite, i.e. $0 < \sum_{\bar{y} \in \bar{\mathbb{Y}}} |\bar{y}| q(\bar{y}|x) < \infty$, where $|\bar{y}|$ denotes the length of the string \bar{y} (the number of symbols in \bar{y}). For $x_{[n]} = (x_1, x_2, \dots, x_n) \in \mathbb{X}^n$ and $\bar{y}_{[n]} = (\bar{y}_1, \bar{y}_2, \dots, \bar{y}_n) \in \bar{\mathbb{Y}}^n$, we write $q_n(\bar{y}_{[n]}|x_{[n]}) = \prod_{i=1}^n q(\bar{y}_i|x_i)$. Let $\overline{\bar{y}_{[n]}}$ denote the concatenation of strings $\bar{y}_i, i \in [n]$. Then a memoryless SEC Q_n is specified by the input alphabet \mathbb{X} , output alphabet \mathbb{Y} and transition probabilities

$$Q_n(\bar{y}|x_{[n]}) = \sum_{\overline{\bar{y}_{[n]}} = \bar{y}} q_n(\bar{y}_{[n]}|x_{[n]}) \quad (1)$$

for $\bar{y} \in \bar{\mathbb{Y}}$ and $x_{[n]} \in \mathbb{X}^n$. Consider the sequence of channels $\{Q_n\}_{n=1}^{\infty}$ where Q_n is as defined above. Then, the following was shown by Dobrushin in 1967.

Theorem 1 (Coding Theorem [1]): Let $X_{[n]}$ and \bar{Y} denote the input and the output of the SEC Q_n . Let

$$C_n = \sup_{\mathbb{P}(X_{[n]})} \frac{1}{n} I(X_{[n]}; \bar{Y}).$$

Then $C = \lim_{n \rightarrow \infty} C_n = \inf_{n \geq 1} C_n$ exists and is equal to the transmission capacity of the SEC. Furthermore,

$$C = \sup_{\mathcal{X}_{\mathcal{M}}} \lim_{n \rightarrow \infty} \frac{1}{n} I(X_{[n]}; \bar{Y})$$

where $\mathcal{X}_{\mathcal{M}}$ is a stationary, ergodic, Markov process over \mathbb{X} .

We will consider an example of an SEC and confine our attention to this channel throughout this paper.

Example 2 (Deletion-Duplication Channel (DDC)): Consider the binary SEC with $\mathbb{X} = \mathbb{Y} = \{0, 1\}$ and the following stochastic matrix

$$q(\bar{y}|x) = \begin{cases} p_d, & \bar{y} = \lambda \\ p_t p_i^{r-1}, & \bar{y} = x^r, \forall r \geq 1 \end{cases}$$

with $p_d + \sum_{r=1}^{\infty} p_t p_i^{r-1} = 1 \Rightarrow p_t = (1 - p_d)(1 - p_i)$ for $p_i < 1$. This model implies that deletions and duplications occur i.i.d. (and mutually exclusively) with probabilities p_d and p_i respectively. The expected output string length is

$$0 < \sum_{r=1}^{\infty} r p_t p_i^{r-1} = \frac{p_t}{(1 - p_i)^2} = \frac{1 - p_d}{1 - p_i} < \infty.$$

Hence $(p_i, p_d) \in [0, 1]^2$. Since the capacity is zero when either p_i or p_d is 1 (they cannot simultaneously be 1), this model does indeed represent the entire class of deletion-duplication channels. Note that when $p_i = 0$, the DDC is the same as the

$$\begin{aligned}
P_n(\bar{y}|x_{[n]}, Z_0 = 0) &= \sum_{\{\bar{z}:|\bar{z}|=|\bar{y}|\}} \mathbb{P}(\bar{Z} = \bar{z}, \bar{Y} = \bar{y} | X_{[n]} = x_{[n]}, Z_0 = 0) \\
&= \sum_{\{\bar{z}:|\bar{z}|=|\bar{y}|\}} \mathbb{P}(\bar{Z} = \bar{z} | Z_0 = 0) \mathbb{P}(\bar{Y} = \bar{y} | X_{[n]} = x_{[n]}, Z_0 = 0, \bar{Z} = \bar{z}) \\
&= \sum_{\{\bar{z}:|\bar{z}|=|\bar{y}|\}} \prod_{i=1}^{|\bar{y}|} \left(\mathbb{P}(Z_i = z_i | Z_{i-1} = z_{i-1}, Z_0 = 0) \mathbb{P}(Y_i = y_i | X_{[n]} = x_{[n]}, Z_i = z_i) \right) \tag{2}
\end{aligned}$$

binary deletion channel (BDC), and when $p_d = 0$, it is the so-called *binary sticky channel* [2].

The BDC has been the most well-studied SEC. In [3], the author surveys the results that were known prior to 2009. To summarize, the best known lower bounds were obtained, chronologically, through bounds on the cutoff rate for sequential decoding [4], bounding the rate with a first-order Markov input [5], reduction to a Poisson-repeat channel [6], analyzing a ‘‘jigsaw-puzzle’’ coding scheme [7], or by directly bounding the information rate by analyzing the channel as a joint renewal process [8]. Recently, [9] and [10] independently gave the capacity of a BDC with small deletion probabilities, and showed that it is achieved by independent and uniformly distributed (i.u.d.) inputs. The known upper bounds for the BDC have been obtained by genie-aided decoder arguments [11], [12]. An idea from [12] was extended to obtain some analytical lower bounds on the capacity of channels that involve substitution errors as well as insertions or deletions [13]. In contrast to these existing results, our approach explicitly characterizes the achievable information rates in terms of ‘‘subsequence-weights’’, which is a measure relevant in ML decoding for the BDC [3]. Additionally, the method proposed here gives the tight bound on capacity for small deletion probabilities obtained in [9] more directly.

For the sticky channel, [2] obtained lower bounds on the capacity by numerically estimating the capacity per unit cost of the equivalent channel of runs through optimization of 8 and 16 bit codes. Here, we obtain direct analytical lower bounds on the capacity. These, to the best of our knowledge, represent the only analytical bounds for the capacity of the sticky channel.

III. SEC AS A CHANNEL WITH STATES

For the DDC Q_n , we write

$$Y_i = X_{\Gamma_i} = X_{i-Z_i}, i \in [N_n] \tag{3}$$

where $Z_i \in \mathbb{Z}$ is the ‘‘state’’ of the channel and we define the length of the output to be $N_n \triangleq \sup\{i \geq 0 : \Gamma_i \leq n | \Gamma_0 = 0\}$. The state process \mathcal{Z} is independent of the channel input process \mathcal{X} , and is a *time-homogeneous* Markov chain over the set of integers \mathbb{Z} with transition probabilities

$$\mathbb{P}(Z_i = z_i | Z_{i-1} = z_{i-1}) = \begin{cases} p_i, & z_i = z_{i-1} + 1 \\ p_t p_d^r, & z_i = z_{i-1} - r, r \geq 0 \end{cases} \tag{4}$$

where we define $p_t = (1 - p_d)(1 - p_i)$ for normalization. We also assume the boundary condition that $Z_0 = 0$, i.e., that

there was perfect synchronization initially. It is easy to see that $N_n < \infty \forall n \in \mathbb{N}$ a.s. since we impose $p_i < 1$, and that $N_n \rightarrow \infty$ a.s. as $n \rightarrow \infty$ since $p_d < 1$. We refer to the Γ -process as the *index process*. The index process and the channel state process have a one-to-one correspondence, and consequently, we will use them interchangeably. From the state transition probabilities in (4), it is also clear that the \mathcal{Z} process is *shift-invariant*, i.e., $\mathbb{P}(Z_i = z_i | Z_{i-1} = z_{i-1}) = \mathbb{P}(Z_i = z_i - z_{i-1} | Z_{i-1} = 0)$. The index process inherits this property from the \mathcal{Z} process.

For $\bar{y} \in \bar{\mathbb{Y}}$ and $x_{[n]} \in \mathbb{X}^n$, the channel transition probabilities are given as in Equation (2). Note that in the terms within the parenthesis on the right hand side of Equation (2), the first term is completely specified by the transition probabilities (4) of the channel state process \mathcal{Z} , and the second term is 0 or 1 accordingly as $y_i = x_{i-z_i}$ or $y_i \neq x_{i-z_i}$ respectively. The input and output alphabets of P_n are $\mathbb{X} = \mathbb{Y} = \{0, 1\}$. The equivalence between the DDC Q_n and P_n , for any n , is evident by noting that for every parsing of $\bar{y} \in \bar{\mathbb{Y}}$ as $\bar{y}_{[n]}$ in Equation (1), there is a corresponding state path $\bar{z} \in \bar{\mathbb{Z}}$ in Equation (2) (and vice versa) and that the terms within the parenthesis in (2), when grouped according to the output symbols arising from the same input symbol, spell out exactly the same probability as the terms $q(\bar{y}_i | x_i)$.

As a consequence of the above equivalence, Theorem 1 applies to the sequence of channels $\{P_n\}_{n=1}^\infty$ specified by Equations (3) and (4). We hence have for input $X_{[n]}$ and output $Y_{[N_n]}$ of P_n ,

$$\begin{aligned}
C &= \limsup_{n \rightarrow \infty} \sup_{\mathbb{P}(X_{[n]})} \frac{1}{n} I(X_{[n]}; Y_{[N_n]}) \\
&= \sup_{\mathcal{X}_{\mathcal{M}}} \lim_{n \rightarrow \infty} \frac{1}{n} I(X_{[n]}; Y_{[N_n]}).
\end{aligned}$$

We will restrict our attention to stationary, ergodic, Markov sources $\mathcal{X}_{\mathcal{M}}$. Under this assumption, the output process \mathcal{Y} is also stationary, and the *entropy rate* $\mathcal{H}(\mathcal{Y})$ is well-defined.

IV. BOUNDS ON CAPACITY

With the setting of Section III, it is possible to immediately obtain some non-trivial bounds on the capacity. We start with some simple bounds and bounding techniques for the DDC and consider the BDC and the sticky channel in separate subsections.

Proposition 3: For the deletion-duplication channel,

$$\left((1 - p_d) \left(1 - \frac{h_2(p_i)}{1 - p_i} \right) - h_2(p_d) \right)^+ \leq C \leq 1 - p_d,$$

where $(x)^+ = \max\{0, x\}$ and $h_2(\cdot)$ is the *binary entropy function*.

Proof: We can write

$$\begin{aligned} I(X_{[n]}; Y_{[N_n]}) &= I(X_{[n]}; Y_{[N_n]}, Z_{[N_n]}) - I(X_{[n]}; Z_{[N_n]} | Y_{[N_n]}) \\ &\stackrel{(a)}{=} I(X_{[n]}; Y_{[N_n]} | Z_{[N_n]}) - I(X_{[n]}; Z_{[N_n]} | Y_{[N_n]}) \\ &\stackrel{(b)}{=} (1 - p_d)H(X_{[n]}) - I(X_{[n]}; Z_{[N_n]} | Y_{[N_n]}), \end{aligned} \quad (5)$$

where (a) is true because $\mathcal{X} \perp \mathcal{Z}$ and (b) from the fact that the DDC, given the \mathcal{Z} process realization, is equivalent to a BEC with erasure rate p_d . Then,

$$n(1 - p_d) \geq I(X_{[n]}; Y_{[N_n]}) \geq (1 - p_d)H(X_{[n]}) - H(Z_{[N_n]}).$$

Since the \mathcal{Z} process is a Markov chain, we can easily show $H(Z_{[N_n]}) \leq \mathbb{E}(N_n) \left(h_2(p_i) + \frac{1-p_d}{1-p_d} h_2(p_d) \right)$, where the inequality follows because, for any finite n , we have the extra knowledge that $Z_i \geq i - n$ by definition. This extra knowledge becomes tautological when $n \rightarrow \infty$ so that

$$\lim_{n \rightarrow \infty} \frac{H(Z_{[N_n]})}{n} = \left(\lim_{n \rightarrow \infty} \frac{\mathbb{E}(N_n)}{n} \right) \left(h_2(p_i) + \frac{1 - p_d}{1 - p_d} h_2(p_d) \right).$$

Writing the Γ -process as a renewal process, from the strong law of large numbers, $\frac{N_n}{n} \rightarrow \frac{1-p_d}{1-p_i}$ a.s. as $n \rightarrow \infty$. ■

Note that the above result implies the following for the BDC ($p_i = 0, p_d = p$), the *symmetric* DDC (with $p_i = p_d = p$) and the sticky channel ($p_i = p, p_d = 0$) respectively.

$$\begin{aligned} (1 - p - h_2(p))^+ &\leq C_{\text{BDC}} \leq 1 - p, \\ (1 - p - 2h_2(p))^+ &\leq C_{\text{SDDC}} \leq 1 - p, \\ \left(1 - \frac{h_2(p)}{1 - p} \right)^+ &\leq C_{\text{Sticky}} \leq 1. \end{aligned}$$

Although these bounds have simple closed-form expressions, they are far from the best known bounds for the capacity of these channels. We can, however, improve these bounds. We have from Equation (5),

$$\begin{aligned} I(X_{[n]}; Y_{[N_n]}) &= (1 - p_d)H(X_{[n]}) + I(Y_{[N_n]}; Z_{[N_n]}) \\ &\quad - H(Z_{[N_n]}) + H(Z_{[N_n]} | X_{[n]}, Y_{[N_n]}). \end{aligned} \quad (6)$$

We can easily show that $\mathcal{H}(\mathcal{Z} | \mathcal{X}, \mathcal{Y}) = \frac{1-p_d}{1-p_i} H(Z_1 | \mathcal{X}, \mathcal{Y})$, and hence

$$\begin{aligned} C &\geq \sup_{\mathcal{X}} \left(\mathcal{H}(\mathcal{X}) + \frac{H(Z_1 | \mathcal{X}, \mathcal{Y})}{1 - p_i} \right) (1 - p_d) \\ &\quad - \frac{1 - p_d}{1 - p_i} h_2(p_i) - h_2(p_d). \end{aligned} \quad (7)$$

It is not easy to evaluate the right hand side of the above inequality. However, we can lower bound it further by introducing some conditioning.

$$\begin{aligned} C &\geq \sup_{\mathcal{X}} \left(\mathcal{H}(\mathcal{X}) + \frac{H(Z_1 | Z_i, \mathcal{X}, \mathcal{Y})}{1 - p_i} \right) (1 - p_d) \\ &\quad - \frac{1 - p_d}{1 - p_i} h_2(p_i) - h_2(p_d) \triangleq \sup_{\mathcal{X}} L_i^{\mathcal{X}} \quad \forall i \in \mathbb{N}. \end{aligned}$$

Lemma 4: The sequence $\{L_i^{\mathcal{X}}\}_{i=1}^{\infty}$ is non-decreasing.

Proof: We have

$$\begin{aligned} H(Z_1 | Z_{i+1}) &= H(Z_1, Z_i | Z_{i+1}) - H(Z_i | Z_1, Z_{i+1}) \\ &\stackrel{(a)}{=} H(Z_i | Z_{i+1}) + H(Z_1 | Z_i) - H(Z_i | Z_1, Z_{i+1}) \\ &= H(Z_1 | Z_i) + I(Z_1; Z_i | Z_{i+1}) \geq H(Z_1 | Z_i), \end{aligned}$$

where (a) follows from the Markovity of the \mathcal{Z} process. Since conditioning on \mathcal{X} and \mathcal{Y} retains the above chain of inequalities, we have $H(Z_1 | Z_{i+1}, \mathcal{X}, \mathcal{Y}) \geq H(Z_1 | Z_i, \mathcal{X}, \mathcal{Y}) \quad \forall i \geq 1$. Hence $\{L_i^{\mathcal{X}}\}_{i=1}^{\infty}$ is non-decreasing, and maximizing over stationary, ergodic, Markov input processes \mathcal{X} gives the bound in Proposition 3 for $i = 1$. Therefore, for increasing i , we have bounds better than the one in Proposition 3, and in the limit as $i \rightarrow \infty$, we approach the bound in (7). ■

For the case of the BDC and the sticky channel, evaluating some of these bounds is easier, owing to the fact that the \mathcal{Z} process is monotonic, i.e., in these cases, the output is just a subsequence of the input sequence and vice versa respectively.

A. Information Rates for the BDC

For the BDC with i.u.d. inputs, we can easily show that \mathcal{Y} is also an i.u.d. sequence. Consequently, $\mathcal{I}(\mathcal{Y}; \mathcal{Z}) = 0$ because the only information obtained from $Y_{[N_n]}$ about $Z_{[N_n]}$ is the length of the vector, and this information vanishes in the limit as $n \rightarrow \infty$ as a result of the concentration. Therefore, we have from Equation (6) that the lower bound in Equation (7) is actually the *symmetric information rate* (SIR) in this case. Let us denote by $w_{y_{[i]}}(x_{[j]})$ the number of subsequences of $x_{[j]}$ that are the same as $y_{[i]}$, and define $w_{\lambda}(x_{[j]}) = 1 \quad \forall x_{[j]}$. We will refer to $w_{y_{[i]}}(x_{[j]})$ as the *$y_{[i]}$ -subsequence weight* of the vector $x_{[j]}$.

We will focus on the term $H(Z_1 | Z_i, \mathcal{X}, \mathcal{Y})$ which is the only term to be evaluated to get an estimate of L_i^{iud} . First note that $H(Z_1 | Z_i, \mathcal{X}, \mathcal{Y}) = H(Z_1 | Z_i, X_{[i-1-Z_i]}, Y_{[i-1]})$. Given $Z_i = -m$, $X_{[m+i-1]} = x_{[m+i-1]}$ and $Y_{[i-1]} = y_{[i-1]}$, we have $Z_1 \in \{j \in \{0, -1, \dots, -m\} : x_{1-j} = y_1, w_{y_{[2:i-1]}}(x_{[2-j:m+i-1]}) > 0\}$. Further, it is easy to see that

$$\begin{aligned} \mathbb{P}(Z_1 = z | Z_i = -m, X_{[m+i-1]} = x_{[m+i-1]}, Y_{[i-1]} = y_{[i-1]}) \\ = \mathbb{1}_{\{x_{1-z} = y_1\}} \frac{w_{y_{[2:i-1]}}(x_{[2-z:m+i-1]})}{w_{y_{[i-1]}}(x_{[m+i-1]})}, \quad -m \leq z \leq 0. \end{aligned}$$

Since $\mathbb{P}(X_{[m+i-1]} = x_{[m+i-1]} | Z_i = -m) = 2^{-(m+i-1)}$,

$$\begin{aligned} \mathbb{P}(Y_{[i-1]} = y_{[i-1]} | X_{[m+i-1]} = x_{[m+i-1]}, Z_i = -m) \\ = \frac{w_{y_{[i-1]}}(x_{[m+i-1]})}{\binom{m+i-1}{m}}, \end{aligned}$$

and $\mathbb{P}(Z_i = -m | Z_0 = 0) = \mathbf{p}^{\otimes i}(-m)$, where we write $\mathbf{p}(-m) = \mathbb{P}(Z_1 = -m | Z_0 = 0)$, $\mathbf{p}^{\otimes i}(-m) = (\mathbf{p} \otimes \mathbf{p}^{\otimes i-1})(-m)$, $\mathbf{p}^{\otimes 1}(-m) = \mathbf{p}(-m)$ with \otimes denoting convolution, we have for any $i \in \mathbb{N}$

$$C \geq L_i^{\text{iud}} = \left(1 + \sum_{m \geq 0} \mathbf{p}^{\otimes i}(-m) \mathfrak{H}_m^{(i)} \right) (1 - p) - h_2(p)$$

$$\mathfrak{H}_m^{(i)} = \sum_{x_{[m+i-1]}} \frac{1}{2^{m+i-1}} \left(\sum_{y_{[i-1]}} \frac{w_{y_{[i-1]}}(x_{[m+i-1]})}{\binom{m+i-1}{m}} \mathfrak{h}(x_{[m+i-1]}, y_{[i-1]}) \right), \quad (8)$$

$$\mathfrak{h}(x_{[m+i-1]}, y_{[i-1]}) = - \sum_{z=-m}^0 \mathbf{1}_{\{x_{1-z}=y_1\}} \frac{w_{y_{[2:i-1]}}(x_{[2-z:m+i-1]})}{w_{y_{[i-1]}}(x_{[m+i-1]})} \log_2 \left(\mathbf{1}_{\{x_{1-z}=y_1\}} \frac{w_{y_{[2:i-1]}}(x_{[2-z:m+i-1]})}{w_{y_{[i-1]}}(x_{[m+i-1]})} \right).$$

where $\mathfrak{H}_m^{(i)}$ is as given in Equation (8). Unfortunately, evaluating $\mathfrak{H}_m^{(i)}$ for $i > 2$ is hard since counting subsequences is not easy. For the case of $i = 2$, we can easily evaluate

$$L_2^{\text{iud}} = \left(1 + (1-p)^2 \sum_{m \geq 0} (m+1)p^m \mathfrak{H}_m^{(2)} \right) (1-p) - h_2(p) \quad (9)$$

with

$$\mathfrak{H}_m^{(2)} = \log_2(m+1) - \frac{1}{2^{m+1}} \sum_{i=0}^{m+1} \binom{m+1}{i} h_2\left(\frac{i}{m+1}\right).$$

Although evaluating L_i^{iud} for $i > 2$ is hard, we can further lower bound it as follows.

$$\begin{aligned} L_i^{\text{iud}} &= (1 + H(Z_1|Z_i, \mathcal{X}, \mathcal{Y}))(1-p) - h_2(p) \\ &\geq \left(1 + \sum_{m=0}^j \mathbb{P}(Z_i = -m) H(Z_1|Z_i = -m, \mathcal{X}, \mathcal{Y}) \right) \\ &\quad \times (1-p) - h_2(p) \\ &\triangleq 1 - p - h_2(p) + (1-p)a_j^{(i)} \\ &\triangleq \mathfrak{L}_j^{(i)} \quad \forall j \geq 0, i \geq 1. \end{aligned}$$

We can then use $C \geq \sup_{i \geq 1} \mathfrak{L}_j^{(i)} \triangleq \mathfrak{L}_j^{\text{iud}}$ for some $j \geq 0$ as a lower bound for the capacity. We proceed as follows

$$\begin{aligned} a_j^{(i)} &= a_{j-1}^{(i)} + p^{\otimes i}(-j) \cdot H(Z_1|Z_i = -j, \mathcal{X}, \mathcal{Y}) \\ &\triangleq a_{j-1}^{(i)} + p^{\otimes i}(-j)b_j^{(i)}. \end{aligned} \quad (10)$$

Since $a_0^{(i)} = b_0^{(i)} = 0$, we have $a_1^{(i)} = ip(1-p)^i b_1^{(i)}$. Let us denote by $r_1(x_{[n]})$ the length of the first run in the vector $x_{[n]}$. Then, we can show that, for $y_{[i-1]}$ received from $x_{[i]}$ with a single deletion,

$$H(Z_1|Z_i = -1, X_{[i]} = x_{[i]}, Y_{[i-1]} = y_{[i-1]}) = h_2\left(\frac{1}{r_1(x_{[i]})}\right),$$

and hence

$$b_1^{(i)} = \sum_{x_{[i]}} \frac{1}{2^i} \sum_{y_{[i-1]}} \frac{1}{i} h_2\left(\frac{1}{r_1(x_{[i]})}\right). \quad (11)$$

From (10) and (11),

$$a_1^{(i)} = p(1-p)^i \sum_{j=1}^i \frac{j}{2^j} h_2\left(\frac{1}{j}\right) + p(1-p)^i \frac{i}{2^i} h_2\left(\frac{1}{i}\right)$$

and thus

$$C \geq 1 + p \log_2 p - cp + O(p^2) \quad (12)$$

where $c = \log_2(2e) - \frac{1}{2} \sum_{j \geq 1} \frac{j}{2^j} \log_2 j \approx 1.154163$. Note that this is exactly the expression obtained for capacity for small p in [9]. In the evaluation of the above bound, we were helped by the fact that when we restrict to the case of a single deletion, the ambiguity in the first channel state Z_1 arises only when $r_1(x_{[i]}) > 1$, in which case the uncertainty is exactly $h_2\left(\frac{1}{r_1(x_{[i]})}\right)$. This, however, is not true when there are 2 or more deletions, wherein we will have to count subsequence weights of sequences.

We can obtain similar bounds for symmetric first-order Markov input processes. But these calculations will have to keep track of *ascents* and *descents* in sequences, and are therefore more tedious. We can write for $\mathbb{P}(X_i = x \oplus 1|X_{i-1} = x) = \alpha$,

$$\begin{aligned} L_2^{\mathcal{M}1} &= \left[\sup_{\alpha} \left(h_2(\alpha) + (1-p)^2 \sum_{m \geq 0} (m+1)p^m \ell_m(\alpha) \right) \right] \\ &\quad \times (1-p) - h_2(p), \text{ where} \\ \ell_m(\alpha) &= \log_2(m+1) - \sum_{i=0}^{m+1} h_2\left(\frac{i}{m+1}\right) \pi(\alpha, i, m+1), \end{aligned}$$

and $\pi(\dots)$ is defined recursively as

$$\begin{aligned} \pi(\alpha, i, m) &= \frac{1}{2} \pi_0(\alpha, i, m) + \frac{1}{2} \pi_1(\alpha, i, m) \\ \pi_0(\alpha, i, m) &= (1-\alpha) \pi_0(\alpha, i, m-1) + \alpha \pi_1(\alpha, i-1, m-1) \\ \pi_1(\alpha, i, m) &= (1-\alpha) \pi_1(\alpha, i-1, m-1) + \alpha \pi_0(\alpha, i, m-1) \end{aligned}$$

with $\pi_j(\alpha, i, m) = \frac{1}{2}(1-\alpha)^{m-1}$, $i \in \{0, m\}$, $j \in \{0, 1\}$ and $\pi_j(\alpha, i, m) = 0$ for $i \notin [m]$, $j \in \{0, 1\}$. We can also evaluate

$$\begin{aligned} \mathfrak{L}_1^{\mathcal{M}1} &= -h_2(p) + (1-p) \times \\ &\sup_{\alpha} \left[h_2(\alpha) + p \cdot \sup_{i \geq 1} (1-p)^i \left(\alpha \sum_{j=1}^i j(1-\alpha)^{j-1} h_2\left(\frac{1}{j}\right) \right. \right. \\ &\quad \left. \left. + i(1-\alpha)^i h_2\left(\frac{1}{i}\right) \right) \right]. \end{aligned}$$

However, both $L_2^{\mathcal{M}1}$ and $\mathfrak{L}_1^{\mathcal{M}1}$ turn out to be better than their SIR counterparts by less than 1%.

B. Information Rates for the Sticky Channel

The analysis for the sticky channel is very similar to that for the BDC in the previous subsection. Since $\lim_{n \rightarrow \infty} \frac{1}{n} I(Y_{[N_n]}; Z_{[N_n]}) \neq 0$ when the input is i.u.d., we bound the $\mathcal{H}(Z|Y)$ term (see Equation (6)) differently. In this case, we obtain

$$C \geq \sup_{\mathcal{X}, \mathcal{M}} \left[\sup_{i \geq 1} \left(\mathcal{H}(\mathcal{X}) + \frac{H(Z_1|Z_i, \mathcal{X}, \mathcal{Y}) - H(Z_1|Y_1)}{1-p} \right) \right]$$

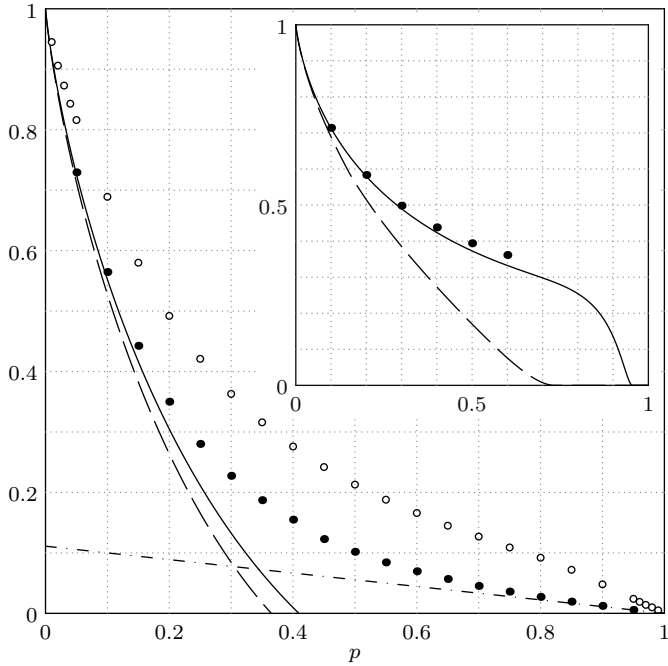


Fig. 1. Bounds on the capacity for the binary deletion channel. L_2^{iid} in (9) is shown as the long-dashed line and L_1^{iid} in (12) (with the $O(p^2)$ term dropped) as the solid line. The best known numerical lower [8] and upper bounds [12] are shown as black and white circles respectively. The best known lower bound as p approaches 1 [6] is shown as the dash-dotted line. The inset plots the bound (13) as the long-dashed line for the sticky channel, and the Markov-1 rate (14) as the solid line. The lower bounds from [2] are shown as black circles.

$$\geq \sup_{\alpha} \left[h_2(\alpha) + \sup_{i \geq 1} \left(ip(1-p)^{i-2} H(Z_1 | Z_i = 1, \mathcal{X}, \mathcal{Y}) \right) - \frac{p + (1-\alpha)(1-p)}{1-p} h_2\left(\frac{p}{p + (1-\alpha)(1-p)}\right) \right], \quad (13)$$

where

$$H(Z_1 | Z_i = 1, \mathcal{X}, \mathcal{Y}) = \frac{1}{i} \sum_{j=1}^{i-1} (j+1) h_2\left(\frac{1}{j+1}\right) (1-\alpha)^j \alpha + h_2\left(\frac{1}{i}\right) (1-\alpha)^i.$$

For $\alpha = \frac{1}{2}$, we get $C \geq 1 + p \log_2 p + dp - O(p^2)$ where $d = \log_2\left(\frac{2}{e}\right) + \frac{1}{2} \sum_{j \geq 1} \frac{j}{2^j} \log_2 j \approx 0.845836$. As was the case for the BDC, we expect this to be a tight bound for the capacity for small p . In fact, for the sticky channel, we can exactly characterize the maximum rate achievable by a first-order Markov process $C^{\mathcal{M}1}$ as

$$C^{\mathcal{M}1} = \sup_{\alpha} \left[h_2(\alpha) + \alpha \sum_{r \geq 1} \left((1-\alpha) \frac{1-p}{p} \right)^r \left(\sum_{s \geq r} \binom{s}{r} p^s h_2\left(\frac{r}{s}\right) \right) - \frac{p + (1-\alpha)(1-p)}{1-p} h_2\left(\frac{p}{p + (1-\alpha)(1-p)}\right) \right]. \quad (14)$$

Figure 1 plots all the bounds obtained for BDC and sticky channel.

We note that the Markov-1 rate (14) is larger than $1-p$ for a range of p values. This disproves the conjecture that the capacity is convex in p for the sticky channel, unlike what is expected for the BDC [14].

V. CONCLUSIONS

The model presented here provides a unified framework to handle a broad class of channels with synchronization errors over any finite alphabet. For channels with only deletions or only duplications, we obtain analytical lower bounds on the capacity, including some bounds that are expected to be tight for small deletion or duplication probabilities. More generally, the model has an immediate factor-graph interpretation, and this could potentially be used to explore reliable coding schemes. Moreover, it could facilitate the exploration of some fundamental theoretical questions, e.g., establishing a coding theorem for synchronization error channels with memory. A more detailed treatment of some of these questions, results for the BDC and sticky channel, and generalizations to other channels of interest is in preparation [15].

ACKNOWLEDGMENT

The work of A. R. Iyengar is supported by the Center for Magnetic Recording Research and the National Science Foundation under the Grant CCF-0829865.

REFERENCES

- [1] R. L. Dobrushin, "Shannon's theorems for channels with synchronization errors," *Problems Inform. Transmission*, vol. 3, no. 4, pp. 11–26, 1967.
- [2] M. Mitzenmacher, "Capacity bounds for sticky channels," *IEEE Trans. Inf. Theory*, vol. 54, no. 1, pp. 72–77, Jan. 2008.
- [3] —, "A survey of results for deletion channels and related synchronization channels," *Probability Surveys*, no. 6, pp. 1–33, 2009.
- [4] R. G. Gallager, "Sequential decoding for binary channels with noise and synchronization errors," *Lincoln Lab. Group Report*, 1961.
- [5] 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.
- [6] 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.
- [7] 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.
- [8] 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.
- [9] Y. Kanoria and A. Montanari, "On the deletion channel with small deletion probability," in *Proc. IEEE Int. Symp. Inf. Theory*, Austin, TX, USA, Jun. 13–18, 2010, pp. 1002–1006.
- [10] A. Kalai, M. Mitzenmacher, and M. Sudan, "Tight asymptotic bounds for the deletion channel with small deletion probability," in *Proc. IEEE Int. Symp. Inf. Theory*, Austin, TX, USA, Jun. 13–18, 2010, pp. 997–1001.
- [11] S. Diggavi, M. Mitzenmacher, and H. D. Pfister, "Capacity upper bounds for the deletion channel," in *Proc. IEEE Int. Symp. Inf. Theory*, Nice, France, Jun. 24–29, 2007, pp. 1716–1720.
- [12] D. Fertonani and T. Duman, "Novel bounds on the capacity of the binary deletion channel," *IEEE Trans. Inf. Theory*, vol. 56, no. 6, pp. 2753–2765, Jun. 2010.
- [13] M. Rahmati and T. M. Duman, "Analytical lower bounds on the capacity of insertion and deletion channels," *CoRR*, vol. abs/1101.1310, 2011.
- [14] M. Dalai, "A new bound for the capacity of the deletion channel with high deletion probabilities," *CoRR*, vol. abs/1004.0400, 2010.
- [15] A. R. Iyengar, P. H. Siegel, and J. K. Wolf, "On the capacity of channels with synchronization errors," *In Preparation*, 2011.