# Upper Bounds on the Capacity of Binary Intermittent Communication

Mostafa Khoshnevisan and J. Nicholas Laneman
Department of Electrical Engineering
University of Notre Dame
Notre Dame, Indiana 46556
Email:{mkhoshne, jnl}@nd.edu

***Abstract*—This paper focuses on obtaining upper bounds on the capacity of a special case of intermittent communication, introduced in [1], in which the channel is binary-input binary-output noiseless with i.i.d. number of zeros inserted in between the binary codeword symbols. Upper bounds are obtained by providing the encoder and the decoder with various amounts of side-information, and calculating or upper bounding the capacity of this genie-aided system. The results suggest that the linear scaling of the receive window with respect to the codeword length considered in the system model is relevant since the upper bounds imply a tradeoff between the capacity of the channel and the intermittency rate.**

## I. Introduction

Intermittent communication is introduced in [1], which models non-continuous transmission of encoded symbols through the channel. For some practical applications transmitting a codeword can be intermittent due to lack of synchronization, shortage of transmission energy, or burstiness of the system. The system model can also be interpreted as an insertion channel in which some number of noise symbols are inserted between the codeword symbols, where the ratio of the receive window $n$ to the codeword length $k$ is assumed to be fixed and is called the intermittency rate $\alpha$. The challenge is that the receiver does not explicitly know whether a given output symbol of the channel is a result of sending a symbol of the codeword or is simply a noise symbol containing no information about the message.

The focus of [1] and [2] is on obtaining achievable rates for intermittent communication by introducing the notion of partial divergence. In this paper, we focus on obtaining upper bounds for the capacity of intermittent communication. We consider a special case of intermittent communication in which the channel is binary-input binary-output noiseless, and the uncertainty is only due to insertions. Specifically, we assume that some number of 0's are inserted in between the binary codeword symbols. After introducing a useful function $g(k, n)$ through uniform insertion model in Section II, we obtain upper bounds on the capacity of the intermittent communication with i.i.d. insertions in Section III by giving some kind of side-information to the encoder and decoder, and calculating or upper bounding the capacity of this genie-aided channel, which is similar to the method used in [3] and [4]. Also, by obtaining an upper bound for the function $g(k, n)$, we are able to tighten the upper bounds for the i.i.d. insertion model in certain regimes.

Although the gap between the achievable rates and upper bounds is not particularly tight, especially for large values of intermittency rate $\alpha$, the upper bounds suggest that the linear scaling of the receive window with respect to the codeword length considered in the system model is relevant since there is a tradeoff between the capacity of the channel and the intermittency rate. By contrast, in asynchronous communication [5], [6], where the transmission of the codeword is contiguous, an exponential scaling $n = e^{\alpha k}$ is most relevant in terms of capacity.

## II. Uniform Insertion Model

### A. Channel Model

Consider binary-input binary-output intermittent communication in which the intermittency is considered as uniform insertions over the whole codeword sequence with the insertions being a fixed input symbol, e.g., inserted symbols are all 0's. The input and output sequences of the channel are $X^k \in \{0,1\}^k$ and $Y^n \in \{0,1\}^n$, respectively, where $k$ and $n$ are two positive integers with $k \leq n$. The output sequence $Y^n$ is constructed as follows: $n-k$ 0's are inserted randomly and uniformly in between the input symbols. The positions at which the insertions occur takes on each of the possible $\binom{n}{n-k} = \binom{n}{k}$ realizations with equal probability, and is unknown to the transmitter and the receiver.

The transmitter communicates a single message $m \in \{1, 2, ..., e^{kR} = M\}$ to the receiver over this channel by encoding the message as a codeword $X^k(m)$, which is called the sequence of codeword symbols. The receiver observes the channel output sequence $Y^n$ and forms an estimate of the message denoted by $\hat{m}$, which is a function of the random sequence $Y^n$. We denote the intermittency rate by $\alpha := n/k$ and assume that it is fixed. This determines the rate that the receive window $n$ scales with the codeword length $k$, i.e., the larger the value of $\alpha$, the larger the receive window, and therefore, the more intermittent the system becomes; if $\alpha = 1$, the system is not intermittent and corresponds to contiguous communication.

We say that the pair $(R, \alpha)$ is achievable if there exists a sequence of length $k$ codes of size $e^{kR}$ with $\frac{1}{M}\sum_{m=1}^{M} \mathbb{P}(\hat{m} \neq m) \to 0$ as $k \to \infty$ for the uniform insertion model with

| $X^k$ \ $Y^n$ | 000 | 001 | 010 | 100 | 011 | 101 | 110 | 111 |
|---|---|---|---|---|---|---|---|---|
| 00 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 01 | 0 | 2/3 | 1/3 | 0 | 0 | 0 | 0 | 0 |
| 10 | 0 | 0 | 1/3 | 2/3 | 0 | 0 | 0 | 0 |
| 11 | 0 | 0 | 0 | 0 | 1/3 | 1/3 | 1/3 | 0 |

TABLE I
TRANSITION PROBABILITIES $P(Y^n|X^k)$ FOR $k=2$ AND $n=3$.

intermittency rate $\alpha$. Note that the communication rate is defined as $\log M/k$. The capacity $C(\alpha)$ is the supremum of all $R$ such that $(R,\alpha)$ is achievable. As we will see, increasing $\alpha$ generally reduces the lower and upper bounds on the capacity. This is because increasing $\alpha$ makes the receive window larger, and therefore, increases the uncertainty about the positions of the information symbols at the receiver making the decoder's task more challenging.

Let $g(k,n)$ be the maximum value of the mutual information between the input and output sequence of the channel over the input distribution, i.e.,

$$g(k,n) := \max_{P(x^k)} \mathbb{I}(X^k;Y^n), \quad 0 \le k \le n \tag{1}$$

Although we have not been able to analytically prove that this channel is information stable, we know by Fano's inequality that $\lim_{k\to\infty} g(k,\alpha k)/k$ is an upper bound on the capacity of the channel with intermittency rate $\alpha$. Note that if the channel is information stable, then this value is equal to the capacity of the channel. Using the achievability result in [1] for this channel model, $(R,\alpha)$ is achievable if

$$R < \max_{0\le p_0\le 1}\{2h(p_0) - \max_{0\le\beta\le 1}\{(\alpha-1)h(\beta) + h((\alpha-1)\beta) + (1-(\alpha-1)\beta)h(\frac{p_0-(\alpha-1)\beta}{1-(\alpha-1)\beta})\}\}. \tag{2}$$

## B. Lower and Upper Bounds on $g(k,n)$

The exact value of the function $g(k,n)$ for finite $k$ and $n$ can be numerically computed by evaluating the transition probabilities $P(Y^n|X^k)$ and using the Blahut algorithm [7] to maximize the mutual information between $X^k$ and $Y^n$. As an example, the transition probability matrix between the input and output sequences for the case of $k=2$ and $n=3$ is given in Table I.

The computational complexity of the Blahut algorithm increases exponentially for large values of $k$ and $n$ since the transition probability matrix is of size $2^k \times 2^n$. In order to partially overcome this issue, we recall the following lemma.

**Lemma 1.** *( [8, Problem 7.28]): Consider a channel that is the union of $i$ memoryless channels $(\mathcal{X}_1, P_1(y_1|x_1), \mathcal{Y}_1), ..., (\mathcal{X}_i, P_i(y_i|x_i), \mathcal{Y}_i)$ with capacities $C_1, ..., C_i$, where at each time one can send a symbol over exactly one of the channels. If the output alphabets are distinct and do not intersect, then the capacity $C$ of this channel can be characterized in terms of $C_1, ..., C_i$ in bits/s as*

$$2^C = 2^{C_1} + ... + 2^{C_i}.$$

Now, notice that the function $g(k,n)$ can be evaluated by considering the union of $k+1$ memoryless channels with distinct input and output alphabets, where the input of the $i^{th}$ channel is binary sequences with length $k$ and weight $i-1$, $i=1,...,k+1$, and the output is binary sequences with length $n$ obtained from the input sequence by inserting $n-k$ zeroes uniformly. Note that the weight of the sequences remain fixed after zero insertions, and therefore, the output alphabets are also distinct and do not intersect. Assuming that the capacity of the $i^{th}$ channel is $g_i(k,n)$ and applying Lemma 1, we have

$$2^{g(k,n)} = 2^{g_1(k,n)} + ... + 2^{g_{k+1}(k,n)}. \tag{3}$$

It is easy to see that $g_1(k,n)=0$ and $g_{k+1}(k,n)=0$. For other values of $i$, the capacity $g_i(k,n), i=2,...,k$ can be evaluated numerically using the Blahut algorithm, where input and output alphabets have sizes $\binom{k}{i-1}$ and $\binom{n}{i-1}$, respectively, which is considerably less than those of the original alphabet sizes. This allows us to obtain the function $g(k,n)$ for larger values of $k$ and $n$. The largest value of $n$ for which we are able to evaluate the function $g(k,n)$ for all values of $k \le n$ is $n=17$. In Section III, we make frequent use of the function $g(k,n)$, and introduce some of its properties.

Although we cannot obtain a closed-form expression for the function $g(k,n)$, we can find closed-form lower and upper bounds by expanding the mutual information in (1) and bounding some of its terms. As a result, we can find lower and upper bounds on the function $g(k,n)$ for larger values of $k$ and $n$. Before stating the results on lower and upper bounds, we introduce some notations.

For a binary sequence $x^k \in \{0,1\}^k$, let $w(x^k)$ denote the weight of the sequence, i.e., number of 1's. Also, let the vector $r^0(x^k) := (r_1^0, ..., r_{l_0}^0)$ of length $l_0$ denote the length of consecutive 0's in the sequence $x^k$ such that $r_1^0 = 0$ if $x_1 = 1$, i.e., the binary sequence starts with 1, and $r_{l_0}^0 = 0$ if $x_k = 1$, i.e., the binary sequence ends with 1, and all the other elements of the vector $r^0(x^k)$ are positive integers. In addition, let the vector $r^1(x^k) := (r_1^1, ..., r_{l_1}^1)$ of length $l_1$ denote the length of consecutive 1's in the sequence $x^k$ with length larger than one, i.e., runs of 1's with length one are not counted. Finally, let $l(x^k) := l_0 + l_1$. When it is clear from the context, we drop the argument $x^k$ from these functions. For example, if $x^k = 0010111000011$, then $w=6$, $r^0=(2,1,4,0)$, $r^1=(3,2)$, and $l=4+2=6$. As another example, if $x^k = 10011101$, then $w=5$, $r^0=(0,2,1,0)$, $r^1=(3)$, and $l=4+1=5$. Now, we define the following function, which will be used for expressing the lower and upper bounds.

$$F(x^k) := \sum_{i_1,...,i_l \in \mathbb{Z}_{\ge 0}: \sum_{j=1}^{l} i_j = n-k} p_{i_1,...,i_l}(r^0, r^1) h_{i_1,...,i_l}(r^0), \tag{4}$$

where $l$, $r^0$, and $r^1$ are a function of $x^k$ as defined before, and we have

$$p_{i_1,\dots,i_l}(r^0,r^1) := \frac{1}{\binom{n}{k}} \prod_{j=1}^{l_0} \binom{r_j^0 + i_j}{i_j} \prod_{j=1}^{l_1} \binom{r_j^1 - 2 + i_{j+l_0}}{i_{j+l_0}}, \tag{5}$$

$$h_{i_1,\dots,i_l}(r^0) := \sum_{j=1}^{l_0} \log \binom{r_j^0 + i_j}{i_j}. \tag{6}$$

**Proposition 1.** *We have the following lower and upper bounds on the function $g(k,n)$:*

$$\log \sum_{x^k \in \{0,1\}^k} \frac{2^{F(x^k)}}{\binom{n-w(x^k)}{k-w(x^k)}} \leq g(k,n) \tag{7}$$

$$g(k,n) \leq \log \sum_{j=0}^{k} \binom{n}{j} 2^{\max_{x^k: w(x^k)=j} F(x^k)} - \log \binom{n}{k} \tag{8}$$

*where $F(\cdot)$ is defined in* (4).

*Proof:* Let $P^n$ denote the random vector describing the positions of the insertions in the output sequence $Y^n$, such that $P_i = 1$ if and only if $Y_i$ is one of the $n-k$ inserted 0's. We first prove the lower bound. We have

$$\begin{aligned}\mathbb{I}(X^k;Y^n) &= H(X^k) - H(X^k|Y^n)\\ &= H(X^k) - H(P^n|Y^n) + H(P^n|X^k,Y^n),\end{aligned} \tag{9}$$

where (9) follows by the general identity $H(X^k|Y^n) + H(P^n|X^k,Y^n) = H(P^n|Y^n) + H(X^k|P^n,Y^n)$ and noticing that for this choice of $P^n$, we have $H(X^k|P^n,Y^n) = 0$. For the term $H(P^n|Y^n)$ in (9), we have

$$\begin{aligned}H(P^n|Y^n) &= \sum_{y^n} P(y^n) H(P^n|Y^n = y^n)\\ &\leq \sum_{y^n} P(y^n) \log \binom{n-w(y^n)}{k-w(y^n)} \qquad (10)\\ &= \sum_{y^n} \sum_{x^k} P(x^k) P(y^n|x^k) \log \binom{n-w(y^n)}{k-w(y^n)}\\ &= \sum_{x^k} P(x^k) \log \binom{n-w(x^k)}{k-w(x^k)}, \qquad (11)\end{aligned}$$

where (10) is because if $y_i = 1$, then it is not an inserted symbol and $p_i = 0$, and therefore, given the output sequence $y^n$, there are $\binom{n-w(y^n)}{k-w(y^n)}$ possible choices for the position vector $P^n$, and we can upper bound it by assuming a uniform distribution; and where (11) is because the weights of the input $x^k$ and output $y^n$ are always the same. For the term $H(P^n|X^k,Y^n)$ in (9), we have

$$H(P^n|X^k,Y^n) = \sum_{x^k} P(x^k) \sum_{y^n} P(y^n|x^k) H(P^n|X^k = x^k, Y^n = y^n) \tag{12}$$

$$= \sum_{x^k} P(x^k) F(x^k), \tag{13}$$

where $F(\cdot)$ is defined in (4); and where (13) is because instead of the summation over $y^n$, we can sum over the possible 0 insertions in between the runs of a fixed input sequence $x^k$ such that there are total of $n-k$ insertions. If we denote the number of insertions in between the runs of zeros by $i_1, \dots, i_{l_0}$, and the number of insertions in between the runs of ones by $i_{1+l_0}, \dots, i_{l_1+l_0}$, then we have $i_1, \dots, i_l \in \mathbb{Z}_{\geq 0} : \sum_{j=1}^{l} i_j = n-k$. Given these number of insertions, it is easy to see that $P(y^n|x^k)$ in (12) is equal to $p_{i_1,\dots,i_l}(r^0,r^1)$ in (5). Also, $H(P^n|X^k = x^k, Y^n = y^n)$ is equal to $h_{i_1,\dots,i_l}(r^0)$ in (6), because given the input and output sequences, the only uncertainty about the position sequence is where there is a run of zeros in the input sequence, i.e., for a run of ones, we know that all the zeros in between them are insertions. Also, the uncertainty is uniformly distributed over all the possible choices. Now, we have

$$\begin{aligned}g(k,n) &= \max_{P(x^k)} \mathbb{I}(X^k;Y^n)\\ &\geq \max_{P(x^k)} \sum_{x^k} P(x^k) \Big[ -\log P(x^k) - \log \binom{n-w(x^k)}{k-w(x^k)} + F(x^k) \Big] \qquad (14)\\ &= \log \sum_{x^k \in \{0,1\}^k} \frac{2^{F(x^k)}}{\binom{n-w(x^k)}{k-w(x^k)}}, \qquad (15)\end{aligned}$$

where (14) follows by combining (9), (11), and (13); and where (15) is the solution to the optimization problem (14). Note that this is a convex optimization problem where the optimal solution can be found to be $P^*(x^k) = D 2^{F(x^k)} / \binom{n-w(x^k)}{k-w(x^k)}$ by Karush-Kuhn-Tucker (KKT) conditions [9], where the constant $D$ is obtained such that $\sum_{x^k} P^*(x^k) = 1$, and (15) is obtained by substituting $P^*(x^k)$. Therefore, the lower bound (7) is proved.

Now, we focus on the upper bound. Note that from (3), we have

$$g(k,n) = \log \sum_{j=0}^{k} 2^{\max_{P(x^k)} \mathbb{I}_j(X^k;Y^n)}, \tag{16}$$

where $\mathbb{I}_j(X^k;Y^n)$ denotes the mutual information if the input sequence, and therefore, the output sequence have weight $j$, and the maximization is over the distribution of all such input sequences. Using the chain rule, we have

$$\begin{aligned}H(P^n|Y^n) &= H(Y^n|P^n) + H(P^n) - H(Y^n)\\ &= H(X^k) + \log \binom{n}{k} - H(Y^n),\end{aligned} \tag{17}$$

where (17) is because the entropy of the output sequence given the insertion positions equals the entropy of the input sequence, and because the entropy of the position sequence equals $\log \binom{n}{k}$ due to the uniform insertions. Combining (9), (13), and (17), we have

$$\begin{aligned}\mathbb{I}_j(X^k;Y^n) &= H_j(Y^n) - \log \binom{n}{k} + \sum_{x^k: w(x^k)=j} P(x^k) F(x^k)\\ &\leq \log \binom{n}{j} - \log \binom{n}{k} + \max_{x^k: w(x^k)=j} F(x^k),\end{aligned} \tag{18}$$

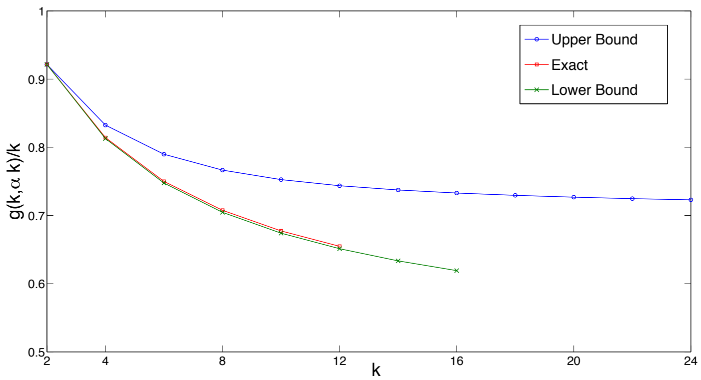


Fig. 1. Comparison between the exact value, the lower bound, and the upper bound on $g(k, \alpha k)/k$ for $\alpha = 1.5$ versus $k$.

where $H_j(Y^n)$ denotes the entropy of the output sequence if it has weight $j$; and where (18) follows from the fact that the uniform distribution maximizes the entropy and by maximizing $F(x^k)$ over all input sequences with weight $j$. Finally, by combining (16) and (18), we get the upper bound (8). ∎

Figure 1 compares the exact value, the lower bound, and the upper bound on $g(k, \alpha k)/k$ for $\alpha = 1.5$ versus $k$. The computational complexity of the exact value is more than the one for the lower and upper bounds, and therefore, cannot be computed numerically for large values of $k$. Note that the upper bound can be computed for relatively large values of $k$, although it seems to be loose. The limit of the exact value (and therefore, the upper bound) as $k \to \infty$ is an upper bound on the capacity of the channel with the intermittency rate $\alpha = 1.5$, as discussed before.

## III. IID Insertion Model

### A. Channel Model

Consider the same binary-input binary-output intermittent communication model introduced in Section II-A, but instead of uniform insertions, after the $i^{th}$ symbol from the codeword, $N_i$ 0's are inserted, where $N_i$'s are i.i.d. geometric random variables with mean $\alpha - 1$. This is equivalent to an i.i.d. insertion channel model in which at each time slot a codeword symbol is sent with probability $p_t := 1/\alpha$ and the insertion symbol 0 is sent with probability $1 - p_t$ until the whole codeword is transmitted. At the decoder, there are $N$ symbols, where $N$ is a random variable with negative binomial distribution with parameters $k$ and $p_t$:

$$P(N = n) = \binom{n-1}{k-1} p_t^k (1 - p_t)^{n-k}, n \geq k, \qquad (19)$$

with $\mathbb{E}[N] = \alpha k$ and $N/k \xrightarrow{p} \alpha$ as $k \to \infty$. As before, $\alpha$ denotes the intermittency rate, and the achievability result (2) is also valid for this channel model.

In contrast to the channel model in Section II, the channel model with the i.i.d. insertions is information stable and falls into the category of the channels with synchronization errors [10] for which the information and transmission capacities are equal and can be obtained by $\lim_{k\to\infty} g(k, N)/k$. However, the single-letter characterization of the capacity is an open problem.

### B. Upper Bounds on the Capacity

In this section, we focus on upper bounds on the capacity of the channel model introduced in Section III-A. The procedure is similar to [3]. Specifically, we obtain upper bounds by giving some kind of side-information to the encoder and decoder, and calculating or upper bounding the capacity of this genie-aided channel. The following definition will be useful in expressing the upper bounds:

$$\phi(k, n) := k - g(k, n), \qquad (20)$$

where the function $g(k, n)$ is defined in (1). Note that the function $\phi(a, b)$ quantifies the loss in capacity due to the uncertainty about the positions of the insertions, and cannot be negative. The following proposition characterizes some of the properties of the functions $g(a, b)$ and $\phi(a, b)$, which will be used later.

**Proposition 2.** *The functions $g(k, n)$ and $\phi(k, n)$ have the following properties:*

*(a)* $g(k, n) \leq k$, $\phi(k, n) \geq 0$.
*(b)* $g(k, k) = a$, $\phi(k, k) = 0$.
*(c)* $g(1, n) = 1$, $\phi(1, n) = 0$.
*(d)* $g(k, n+1) \leq g(k, n)$, $\phi(k, n+1) \geq \phi(k, n)$.
*(e)* $g(k+1, n+1) \leq 1 + g(k, n)$, $\phi(k+1, n+1) \geq \phi(k, n)$.

*Proof:* We prove the properties for the function $g(k, n)$. The corresponding properties for the function $\phi(a, b)$ easily follows from (20).

(a) Since the cardinality of the input sequence is $2^k$, the mutual information is at most $k$ bits/s.
(b) There are no insertions. Therefore, it is a noiseless channel with input and output alphabets of sizes $2^k$ and capacity $k$ bits/s.
(c) The input alphabet is $\{0, 1\}$, and the output consists of binary sequences with length $n$ and weight 0 or 1, because only 0's can be inserted in the sequence. Considering all the output sequences with weight 1 as a super-symbol, the channel becomes binary noiseless with capacity 1 bits/s.
(d) The capacity $g(k, n+1)$ cannot decrease if, at each channel use, the decoder knows exactly one of the positions at which an insertion occurs, and the capacity of the channel with this genie-aided encoder and decoder becomes $g(k, n)$. Therefore, $g(k, n+1) \leq g(k, n)$.
(e) The capacity $g(k+1, n+1)$ cannot decrease if, at each channel use, the encoder and decoder know exactly one of the positions at which an input bit remains unchanged, so that it can be transmitted uncoded and the capacity of the channel with this genie-aided encoder and decoder becomes $1 + g(k, n)$. Therefore, $g(k+1, n+1) \leq 1 + g(k, n)$.

∎

Now we introduce one form of side-information. Assume that the position of the $[(s+1)i]^{th}$ codeword symbol in the output sequence is given to the encoder and decoder for all $i = 1, 2, ...$ and a fixed integer number $s \geq 1$. We assume that the codeword length is a multiple of $s+1$, so that $t = k/(s+1)$ is an integer, and is equal to the total number of positions that are provided as side-information. This assumption does not impact the asymptotic behavior of the channel as $k \to \infty$. We define the random sequence $\{Z_i\}_{i=1}^t$ as follows: $Z_1$ is equal to the position of the $[s+1]^{th}$ codeword symbol in the output sequence, and for $i \in \{2, 3, ..., t\}$, $Z_i$ is equal to the difference between the positions of the $[(s+1)i]^{th}$ codeword symbol and $[(s+1)(i-1)]^{th}$ codeword symbol in the output sequence.

Since we assumed i.i.d. insertions, the random sequence $\{Z_i\}_{i=1}^t$ is i.i.d. too with negative binomial distribution:

$$P(Z_i = b+1) = \binom{b}{s}(1-p_t)^{b-s}p_t^{s+1}, b \geq s, \tag{21}$$

with mean $\mathbb{E}[Z_i] = (s+1)/p_t$. Also, note that as $k \to \infty$, by the law of large numbers, we have

$$\frac{N}{t} \xrightarrow{p} \mathbb{E}[Z_i] = \frac{s+1}{p_t}. \tag{22}$$

Let $C_1$ denote the capacity of the channel if we provide the encoder and decoder with side-information on the random sequence $\{Z_i\}_{i=1}^t$, which is clearly an upper bound on the capacity of the original channel. With this side-information, we essentially partition the transmitted and received sequences into $t$ contiguous blocks that are independent from each other. In the $i^{th}$ block the place of the $[s+1]^{th}$ codeword symbol is given, which can convey one bit of information. Other than that, the $i^{th}$ block has $s$ input bits and $Z_i - 1$ output bits with uniform 0 insertions. Therefore, the information that can be conveyed through the $i^{th}$ block equals $g(s, Z_i - 1) + 1$. Thus, we have

$$\begin{aligned}
C_1 &= \lim_{k\to\infty} \frac{1}{k}\sum_{i=1}^{t} g(s, Z_i - 1) + 1 \\
&= \lim_{k\to\infty} \frac{N}{k}\frac{t}{N}\frac{1}{t}\sum_{i=1}^{t} g(s, Z_i - 1) + 1 \\
&= \frac{1}{s+1}\lim_{t\to\infty}\frac{1}{t}\sum_{i=1}^{t} g(s, Z_i - 1) + 1 \qquad (23)\\
&= \frac{1}{s+1}\mathbb{E}\left[g(s, Z_i - 1) + 1\right] \qquad (24)\\
&= \frac{1}{s+1}\left[1 + \sum_{b=s}^{\infty}\binom{b}{s}(1-p_t)^{b-s}p_t^{s+1}g(s,b)\right] \qquad (25)\\
&= 1 - \frac{1}{s+1}\sum_{b=s}^{\infty}\binom{b}{s}(1-p_t)^{b-s}p_t^{s+1}\phi(s,b), \qquad (26)
\end{aligned}$$

where (23) follows from (22); where (24) follows from the law of large numbers; where (25) follows from the distribution of $Z_i$'s given in (21); and where (26) follows from the definition (20). Note that the capacity $C_1$ cannot be larger than 1, since the coefficients $\phi(\cdot,\cdot)$ cannot be negative. The negative term in (26) can be interpreted as the communication overhead as the cost of intermittency in the context of [11]. The expression in (26) gives an upper bound on the capacity of the original channel with $p_t = 1/\alpha$. However, it is infeasible to numerically evaluate the coefficients $\phi(s, b)$ for large values of $b$. As we discussed before, the largest value of $b$ for which we are able to evaluate the function $\phi(s, b)$ is $b_{max} = 17$. The following upper bound on $C_1$ results by truncating the summation in (26) and using part (d) of Proposition 2.

$$\begin{aligned}
C_1 \leq & 1 - \frac{\phi(s, b_{max})}{s+1} \\
&+ \frac{1}{s+1}\sum_{b=s}^{b_{max}}\binom{b}{s}p_t^{s+1}(1-p_t)^{b-s}(\phi(s,b_{max}) - \phi(s,b)),
\end{aligned} \tag{27}$$

The expression (27), which we denote by $C_1'$, gives a non-trivial and computable upper bound for each value of $s = 2, 3, ..., b_{max} - 1$ on $C_1$, and therefore, an upper bound on the capacity of the original channel with $p_t = 1/\alpha$. Figure 2 shows the upper bounds for $b_{max} = 17$ and $s = 2, 3, ..., 16$ versus the intermittency rate $\alpha$, along with the the achievability result.

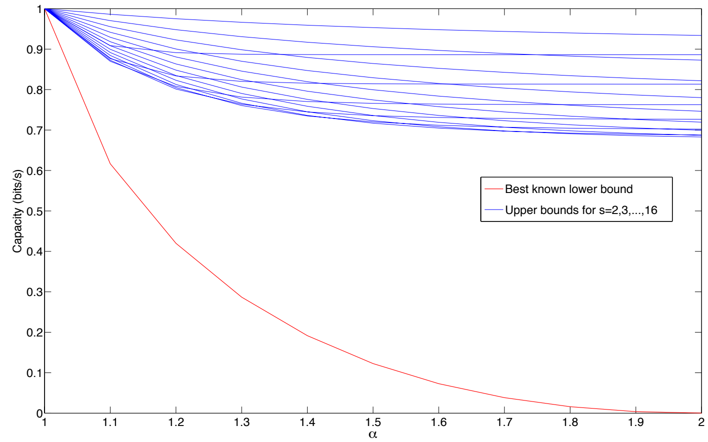

Fig. 2. Comparison between the best achievability result with different upper bounds obtained from (27) for $b_{max} = 17$ and $s = 2, 3, ..., 16$, versus the intermittency rate $\alpha$.

Next, we introduce a second form of side-information. Assume that for consecutive blocks of length $s$ of the output sequence, the number of codeword symbols within that block is given to the encoder and decoder as side-information, i.e., the number of codeword symbols in the sequence $(y_{(i-1)s+1}, y_{(i-1)s+2}, ..., y_{is})$, $i = 1, 2, ...$ for a fixed integer number $s \geq 2$. Let $C_2$ denote the capacity of the channel if we provide the encoder and decoder with this side-information. Using a similar procedure, we obtain

$$C_2 = 1 - \frac{1}{sp_t}\sum_{a=0}^{s}\binom{s}{a}p_t^a(1-p_t)^{s-a}\phi(a,s). \tag{28}$$

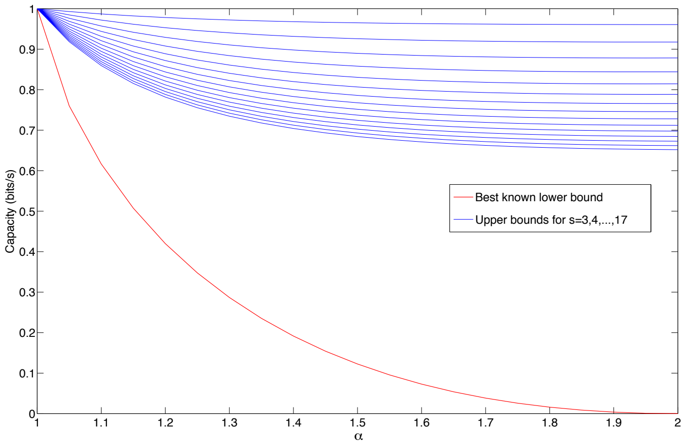


Fig. 3. Comparison between the best achievability result with different upper bounds obtained from (28) for $s = 3, 4, ..., 17$, versus the intermittency rate $\alpha$.

Note that the summation in (28) is finite, and we do not need to upper bound $C_2$ as we did for $C_1$. The value of $C_2$ gives nontrivial and computable upper bounds on the capacity of the original channel. Figure 3 shows the upper bounds for $s = 3, 4, ..., 17$ versus the intermittency rate $\alpha$, along with the the achievability result. The upper bound corresponding to $s = 17$ is tighter than others for all ranges of $\alpha$, i.e., (28) is decreasing in $s$. Intuitively, this is because by decreasing $s$, we provide the side-information more frequently, and therefore, the capacity of the resulting genie-aided system becomes larger.

It seems that (28) gives better upper bounds for the range of $\alpha$ shown in the figures ($1 < \alpha \leq 2$). However, the other upper bound $C_1'$ can give better results for the limiting values of $\alpha \to \infty$ or $p_t \to 0$. We have

$$\lim_{\alpha\to\infty} C_1' = 1 - \frac{\phi(s, b_{max})}{s+1}, \qquad (29)$$
$$\lim_{\alpha\to\infty} C_2 = 1.$$

The upper bound $C_2$ give the trivial upper bound 1. Therefore, the upper bound $C_1'$ is tighter for large values of $\alpha$. This is because of the fact that by increasing $\alpha$, and thus decreasing $p_t$, we have more zero insertions and the first kind of genie-aided system provides side-information less frequently leading to tighter upper bounds. The best upper bound for the limiting case of $\alpha \to \infty$ found by (29) is $0.6739$ bits/s. In principle, we can use the upper bound on $g(k, n)$ in Proposition 1 to upper bound $C_1$ and $C_2$. By doing so, we can find the bounds for larger values of $s$ and $b_{max}$, because we can calculate the upper bound (8) for larger arguments. It seems that this does not improve the upper bounds significantly for the range of $\alpha$ shown in the figures. However, by upper bounding (29) via (8), we can tighten the upper bound for the limiting case of $\alpha \to \infty$ to $0.6307$ bits/s.

Although the gap between the achievable rates and upper bounds is not particularly tight, especially for large values of intermittency rate $\alpha$, the upper bounds suggest that the linear scaling of the receive window with respect to the codeword length considered in the system model is natural since there is a tradeoff between the capacity of the channel and the intermittency rate. By contrast, in asynchronous communication [5], [6], where the transmission of the codeword is contiguous, an exponential scaling $n = e^{\alpha k}$ is most relevant in terms of capacity.

We would like to mention that the techniques used in this section can in principle be utilized for non-binary and noisy i.i.d. insertion channels as well. However, the computational complexity for numerical evaluation of the genie-aided system becomes cumbersome.

## IV. Conclusion

We obtained upper bounds on the capacity of a special case of intermittent communication in which the channel is binary-input binary-output noiseless with i.i.d. number of $0$ insertions in between the binary codeword symbols. Upper bounds are obtained by providing the encoder and the decoder with two forms of side-information, and calculating or upper bounding the capacity of this genie-aided system. Also, by obtaining an upper bound for the function $g(k, n)$, we tightened the upper bounds for certain regimes of the intermittency rate $\alpha$. The results suggest that the linear scaling of the receive window with respect to the codeword length considered in the system model is natural since there is a tradeoff between the capacity of the channel and the intermittency rate.

## References


[1] M. Khoshnevisan and J. N. Laneman, "Achievable Rates for Intermittent Communication," in *Proc. IEEE Int. Symp. Information Theory (ISIT),* pp. 1346-1350, Cambridge, MA, USA, July 2012.

[2] M. Khoshnevisan and J. N. Laneman, "Intermittent Communication and Partial Divergence," in *Proc. Allerton Conf. Communications, Control, and Computing,* Monticello, IL, Oct. 2012.

[3] D. Fertonani and T. M. Duman, "Novel bounds on the capacity of the binary deletion channel," *IEEE Trans. on Inf. Theory,* vol. 56, no. 6, pp. 2753-2765, 2010.

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

[5] A. Tchamkerten, V. Chandar, and G. W. Wornell, "Communication under strong asynchronism," *IEEE Trans. on Inf. Theory,* vol. 55, no. 10, pp. 4508-4528, Oct. 2009.

[6] ——, "Asynchronous communication: Capacity bounds and suboptimality of training," *IEEE Trans. on Inf. Theory*, submitted 2012. [Online]. Available: http://arxiv.org/pdf/1105.5639v2.pdf

[7] R. E. Blahut, "Computation of Channel Capacity and Rate Distortion Functions," *IEEE Trans. on Inf. Theory,* vol. 18, no. 1, pp. 460-473, Jan. 1972.

[8] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, Wiley, Second Edition, 2006.

[9] S. Boyd and L. Vandenberghe, *Convex Optimization.* Cambridge, U.K.: Cambridge Univ. Press, 2004.

[10] R. L. Dobrushin, "Shannon's theorems for channels with synchronization errors," *Problems of Information Transmission,* vol. 3, no. 4, pp. 11-26, 1967.

[11] J. N. Laneman and B. P. Dunn, "Communications Overhead as the Cost of Constraints," in *Proc. IEEE Information Theory Workshop (ITW)*, Paraty, Brazil, Oct. 2011.