Achievable Rates for 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*—We formulate a model for intermittent communications that can capture bursty transmissions or a sporadically available channel, where in either case the receiver does not know a priori when the transmissions occur. Focusing on the point-to-point case, we develop two decoding schemes and their achievable rates for such communication scenarios. One scheme determines the transmitted codeword, and another scheme first locates the information symbols and then uses them to decode. The two-stage scheme leads to a higher achievable rate because it uses a generalization of the method of types in the first stage, which leads to a notion of *partial divergence*. We illustrate the results in the case of an intermittent binary symmetric channel.**

I. Introduction

Communication systems are traditionally analyzed assuming continuous transmission of encoded symbols through the channel. However, in many practical applications such an assumption is not appropriate, and transmitting a codeword can be intermittent due to lack of synchronization, shortage of transmission energy, or burstiness of the system. 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. In this paper, we model such intermittent communication and determine some achievable rates.

Asynchronous communication is modeled in [1] and [2] by a single block transmission that starts at a random time, unknown to the receiver, in an exponentially large window, known to the receiver. In this model, the transmission is contiguous; once it starts, the whole codeword is transmitted, and the receiver observes only noise before and after transmission. However, in intermittent communication, a codeword is not transmitted contiguously. Instead, some number of noise symbols occur between the symbols of a codeword before going through the channel. We assume that the receive window and the codewords length, and therefore the number of inserted noise symbols, is fixed. In contrast to [1], where the interesting scenario arises if the receive window scales exponentially with the codeword length, the interesting scenario in intermittent communication arises if the receiver window scales linearly with the codeword length. Our system model can be also interpreted as an insertion channel, in which a random number of noise symbols is inserted between consecutive symbols of the codeword. Although different from the insertion channels in the literature [3] and [4], our results may provide some insights about them.

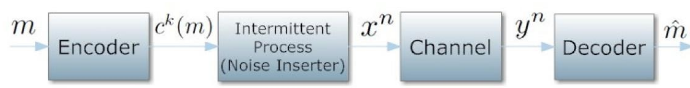

Fig. 1. System model for intermittent communication.

After formulating the problem and specifically mentioning the applications, we describe the first decoding scheme and analyze its achievable rate. Then, we consider a two-stage decoding procedure and characterize its achievable rate using a notion of *partial divergence* that results from a generalization of the method of types. Finally, we illustrate the achievable rates for an intermittent binary symmetric channel (BSC).

II. Model, Applications, and Notations

A transmitter wants to send a message $m \in \{1, 2, ..., e^{kR} = M\}$ to a receiver through a discrete memoryless channel. Let $\mathcal{X}$ and $\mathcal{Y}$ denote the input and output alphabets of the channel, and let W denote its probability transition matrix. Assume that $\{c^k(m), m = 1, 2, ..., M\}$ are the codewords of integer length k available to the transmitter, and let X^n and Y^n denote the integer length n input and output vectors of the channel, respectively, where $n \geq k$, and $\alpha := n/k$, $\alpha \geq 1$. Assume that X^n consists of the transmitted codeword at k arbitrary time slots and is equal to a noise symbol denoted by $\star \in \mathcal{X}$ at the other $n - k$ time slots, where the codeword c^k appears in the sequence X^n in the same order, i.e., c_i cannot appear after c_j in the sequence if $i \leq j$. It is equivalent to say that $n - k$ noise symbols are arbitrarily inserted between the symbols of a codeword. Figure 1 shows a block diagram for the system model, which we call *intermittent communication* and denote by $(\mathcal{X}, \mathcal{Y}, W, \star, \alpha)$. The communication rate is defined as $\log M/k$. Assuming that the decoded message is denoted by $\hat{m}$, we say that rate R 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 $k \to \infty$. The capacity is the supremum of all the achievable rates. A natural question is, what is the capacity of intermittent communication? In this paper, we find some achievable rates.

The intermittent communication model can also represent bursty communication in which either the transmitter or the channel is bursty. We assume that the receiver does not know the positions of the codeword symbols in the channel input

978-1-4673-2579-0/12/$31.00 ©2012 IEEE

sequence. As we will see, this natural assumption seems to make the decoder's task more difficult. We also assume that the transmitter cannot decide on these positions, so it cannot encode any timing information. Note that in a bursty communication scenario, the process of burstiness is usually out of the transmitter's control, and the receiver usually does not know the realization of the bursts, which are consistent with these assumptions. The level of intermittency (or burstiness) is controlled by how large n is compared to k. The larger the value of α, the more intermittent the system is; if $\alpha = 1$, the system is not intermittent and corresponds to contiguous communication. Not surprisingly, the achievable rates obtained in this paper are a function of α: increasing α generally reduces the achievable rates.

The described system model applies to several practical communication scenarios. If the intermittent process component in Figure 1 is considered as a part of the channel behavior, then we say that the channel is intermittent in the sense that it takes a symbol of the codeword as an input at some time slots, and takes a noise symbol as an input at other time slots. As an example, consider an insertion channel in which, after the i^{th} symbol from the codeword, N_i noise symbols are inserted, where N_i's are i.i.d. random variables with mean $\alpha-1$. At the decoder, there are N symbols, where N is a random variable, but we have

$$\frac{N}{k} = \frac{k + N_1 + N_2 + ... + N_k}{k} \xrightarrow{p} 1 + \mathbb{E}(N_1) = \alpha.$$

It turns out that the achievability results in the sequel are valid for such insertion channels, as we discuss in more detail in Section IV. Note that this is a specific class of insertion channels, and we refer the interested reader to See [3], [4] for more general classes of insertion channels and associated results.

On the other hand, if the intermittent process component in Figure 1 is considered as a part of the transmitter, then we say that the transmitter is intermittent. Practical examples are energy harvesting systems, where the transmitter harvests energy usually from a natural source and uses it for transmission. Assuming that the noise symbol can be transmitted with zero energy, the transmitter sends the symbols of the codeword if there is enough energy for transmission, and sends noise symbols otherwise. If the transmitter can store energy in an energy buffer with enough capacity, then it can decide when to send the symbols of the codeword and when to transmit a noise symbol to save some energy based on the amount of the available energy. In that case, the transmitter can also encode some timing information, which is beyond the scope of this paper.

Notation: We use $o(\cdot)$ to denote quantities that grow strictly slower than their arguments. Most of the notation in this paper follows that in [1] and [5]. By $X \sim P(x)$, we mean X is distributed as P. The empirical distribution (or type) of a sequence $x^n \in \mathcal{X}^n$ is denoted by $\hat{P}_{x^n}$. Joint empirical distributions are denoted similarly. We say a sequence x^n has type P if $\hat{P}_{x^n} = P$ and denote it by $x^n \in T_P^n$, where T_P^n or more simply T_P is the set of all sequences that have type P. We use $\mathcal{P}^{\mathcal{X}}$ to denote the set of distributions over the finite alphabet $\mathcal{X}$. For simplicity, we define $W_\star(\cdot) := W(\cdot|x = \star)$. In this paper, we use the convention that $\binom{n}{k} = 0$ if $k < 0$ or $n < k$, and the entropy $H(P) = -\infty$ if P is not a probability mass function, i.e., one of its elements is negative or the sum of its elements is larger than one. $h(\cdot)$ is the binary entropy function. Finally, if $0 \le \rho \le 1$, then $\bar{\rho} := 1 - \rho$.

III. Main Results

We first introduce a simple decoding scheme, obtaining an achievable rate (Proposition 1) in addition to providing some ingredients for obtaining an achievable rate for a two stage decoding scheme (Theorem 1).

A. One-Stage Decoding

For a fixed input distribution P, the codebook is randomly and independently generated, i.e., all $C_i(m), i \in \{1, 2, ..., k\}, m \in \{1, 2, ..., M\}$ are i.i.d. according to P. For a fixed typicality parameter $\mu > 0$, the decoder observes the sequence y^n, chooses k symbols out of those n symbols, denoted by $\tilde{y}^k$, and performs joint typicality decoding, i.e., checks if

$$|\hat{P}_{c^k(m),\tilde{y}^k}(x, y) - P_m(x, y)| \le \mu \tag{1}$$

for all $(x, y) \in \mathcal{X} \times \mathcal{Y}$ and a unique index m, where P_m is the joint probability mass function induced by the type of codeword $c^k(m)$ and the channel W, defined by [1]

$$P_m(x, y) := \hat{P}_{c^k(m)}(x)W(y|x), \ (x, y) \in \mathcal{X} \times \mathcal{Y}.$$

For convenience, we write $\tilde{y}^k \in T_{[W]_\mu}(c^k(m))$, if (1) is satisfied for m. If the decoder finds a unique m satisfying (1), it declares m as the transmitted message. Otherwise, it makes another choice for the k symbols from the sequence y^n and again attempts typicality decoding. If at the end of all $\binom{n}{k}$ choices the typicality decoding procedure did not declare any message, then the decoder declares an error. In order to analyze the probability of error, we state the following fact, which is proved in [1, Equations (24) and (25)] based on the method of types [5, Chapter 1.2].

Fact 1. *Assume that $C^k(m)$ and $\tilde{Y}^k$ are independent, $C^k(m)$ is generated i.i.d. according to P, and $(X, Y) \sim P(x)W(y|x)$, then the probability that $C^k(m)$ together with $\tilde{Y}^k$ satisfy* (1) *for this specific m, is upper bounded as*

$$\mathbb{P}(\tilde{Y}^k \in T_{[W]_\mu}(C^k(m))) \le poly(k)e^{-k(\mathbb{I}(X;Y)-\epsilon)} \tag{2}$$

for all k sufficiently large, where ϵ can be made arbitrarily small by choosing a small enough typicality parameter μ.

Proposition 1. *For intermittent communication with $(\mathcal{X}, \mathcal{Y}, W, \star, \alpha)$, rates not exceeding $C - \alpha h(1/\alpha)$ are achievable, where C is the capacity of the DMC with stochastic matrix W.*

Proof: Let P be the capacity achieving input distribution for the DMC with stochastic matrix W, and consider the encoding and decoding strategies described above. For any

$\epsilon > 0$, we prove that if $R = C - \alpha h(1/\alpha) - 2\epsilon$, then the average probability vanishes as $k \to \infty$.

$$p_e^{avg} = p_e^1 = \mathbb{P}(\hat{m} \neq 1 | m = 1) \\ \leq \mathbb{P}(\hat{m} \in \{2, 3, ..., M\} | m=1) + \mathbb{P}(Y^k \notin T_{[W]_\mu}(C^k(1))), \quad (3)$$

where p_e^1 is the probability of error conditioned on the sending of message $w = 1$, and where (3) results from the union bound in which the second term is the probability that the typicality decoding fails for the right codeword given that the k chosen output symbols are the correct ones, which vanishes as $k \to \infty$ according to [5, Lemma 2.12]. Using the union bound for the first term in (3), we have

$$\mathbb{P}(\hat{m} \in \{2, 3, ..., M\} | m = 1) \\ \leq \binom{n}{k} (M-1) \mathbb{P}(\tilde{Y}^k \in T_{[W]_\mu}(C^k(2)) | m = 1), \quad (4)$$

because there are $\binom{n}{k}$ choices for the k output symbols, and for each choice, we use the union bound for all $M - 1 = e^{kR} - 1$ messages other than $m = 1$. Using the Stirling's approximation, we have

$$\binom{n}{k} \leq \frac{e^{\frac{1}{12}}}{\sqrt{2\pi}} \sqrt{\frac{n}{k(n-k)}} e^{nh(k/n)} = e^{k(o(1) + \alpha h(1/\alpha))}, \quad (5)$$

as $k \to \infty$. Note that, conditioned on message $m = 1$ being sent, $C^k(2)$ and $\tilde{Y}^k$ are independent for any choice of output symbols. Therefore, using Fact 1 with the capacity achieving input distribution, we have

$$\mathbb{P}(\tilde{Y}^k \in T_{[W]_\mu}(C^k(2)) | m = 1) \leq \text{poly}(k) e^{-k(C - \epsilon)} \quad (6)$$

Combining (3), (4), (5), and (6), we obtain

$$p_e^{avg} \leq e^{k(o(1) + \alpha h(1/\alpha))} e^{k(C - \alpha h(1/\alpha) - 2\epsilon)} e^{-k(C - \epsilon)} + o(1) \\ \leq e^{-k(\epsilon - o(1))} \text{poly}(k) + o(1) \to 0,$$

as $k \to \infty$, which proves the proposition. ∎

Note that the maximum probability of error vanishes as well using standard expurgation arguments. The form of the achievable rate is reminiscent of communications overhead as the cost of constraints [6], where the constraint is the system's burstiness, and the overhead cost is $\alpha h(1/\alpha)$.

B. Two-Stage Decoding

In this section, we introduce a two-stage decoding procedure that strictly improves upon the achievable rate in Proposition 1. As before, the decoder chooses k of the n symbols from the output vector y^n. Let $\tilde{y}^k$ denote the chosen output symbols, and $\hat{y}^{n-k}$ denote the other output symbols. The first stage consists of checking if $\tilde{y}^k$ is the transmitted sequence (if $\tilde{y}^k \in T_{PW}$), and if $\hat{y}^{n-k}$ is generated by noise (if $\hat{y}^{n-k} \in T_{W_\star}$). If both of these conditions are satisfied, then we perform the typicality decoding as described in Section III-A, which is called the second stage here. Otherwise, we make another choice for the k symbols and repeat the two-stage decoding procedure. At any step that we run the second stage, if the typicality decoding declares a message as being sent, then decoding ends. If the decoder does not declare any message as being sent by the end of all $\binom{n}{k}$ choices, then the declares an error.

In order to analyze the probability of error, we need to generalize the method of types in [5] to be able to bound the probability of certain events in the first stage of the two-stage decoding procedure. To that end, we establish Lemma 1, which is a generalization of [5, Lemma 2.6].

Lemma 1. *Consider an alphabet with t symbols, i.e., $\mathcal{X} = \{0, 1, ..., t-1\}$. Consider three distributions $P, Q, Q' \in \mathcal{P}^{\mathcal{X}}$, where $P := (p_0, p_1, ..., p_{t-1})$, $Q := (q_0, q_1, ..., q_{t-1})$, and $Q' := (q'_0, q'_1, ..., q'_{t-1})$. We assume that all of the elements of these three PMF's are nonzero. A random sequence X^k is generated as follows: k_1 symbols are i.i.d. according to Q and k_2 symbols are i.i.d. according to Q', where $k_1 + k_2 = k$ and $\rho := k_1/k$. The probability that X^k has type P is upper bounded as*

$$\mathbb{P}(X^k \in T_P) \leq e^{o(k)} e^{-kd(P, Q, Q', \rho)}, \quad (7)$$

where

$$d(P, Q, Q', \rho) := H(P) + D(P||Q) \\ - \bar{\rho} \log \frac{q'_{t-1}}{q_{t-1}} - e(P, Q, Q', \rho), \quad (8)$$

$$e(P, Q, Q', \rho) := \max_{0 \leq \theta_j \leq 1, j=0,1,...,t-2} \{\rho H(P_1) + \bar{\rho} H(P_2) \\ + \sum_{j=0}^{t-2} \theta_j p_j \log a_j\}, \quad (9)$$

$$a_j := \frac{q'_j q_{t-1}}{q_j q'_{t-1}}, \; j = 0, 1, ..., t-2,$$

$$P_1 := (\frac{\bar{\theta}_0 p_0}{\rho}, \frac{\bar{\theta}_1 p_1}{\rho}, ..., \frac{\bar{\theta}_{t-2} p_{t-2}}{\rho}, 1 - \frac{\sum_{j=0}^{t-2} \bar{\theta}_j p_j}{\rho}),$$

$$P_2 := (\frac{\theta_0 p_0}{\bar{\rho}}, \frac{\theta_1 p_1}{\bar{\rho}}, ..., \frac{\theta_{t-2} p_{t-2}}{\bar{\rho}}, 1 - \frac{\sum_{j=0}^{t-2} \theta_j p_j}{\bar{\rho}}).$$

Because of space limitations, we prove the lemma for the case of binary alphabets, i.e, $t = 2$, in Appendix A. The generalization to $t > 2$ is straightforward, but lengthy. Specializing Lemma 1 for $Q' = Q$ results in [5, Lemma 2.6], and we have $d(P, Q, Q, \rho) = D(P||Q)$. Lemma 1 enables us to evaluate the probability that a sequence consisting of symbols from two different distributions has a specific type P, which will be useful in proving the result in Theorem 1. Basically, we are interested in a special case of Lemma 1 for which $Q' = P$. In other words, we need to upper bound the probability that a sequence has type P when partially generated according to Q and partially according to P, where the ratio of the mismatched symbols (generated from Q) to all the symbols is $\rho = k_1/k$. For this special case, we define $d_\rho(P||Q) := d(P, Q, P, \rho)$. The function $d_\rho(P||Q)$, which we call *partial divergence* between P and Q, has some interesting properties (such as $0 \leq d_\rho(P||Q) \leq \rho D(P||Q)$), on which we will elaborate in another paper.

Theorem 1. *For intermittent communication with* $(\mathcal{X}, \mathcal{Y}, W, \star, \alpha)$, *rates not exceeding* $\max_P\{\mathbb{I}(X;Y) - f(P,W,\alpha)\}$ *are achievable, where*

$$f(P,W,\alpha){:=}\max_{0\le\beta\le 1}\{(\alpha-1)h(\beta)+h((\alpha-1)\beta) -d_{(\alpha-1)\beta}(PW||W_\star)-(\alpha-1)d_\beta(W_\star||PW)\}. \tag{10}$$

Proof: Fix the input distribution P, and consider the encoding and decoding strategies described in Sections III-A and III-B, respectively. For any $\epsilon > 0$, we prove that if $R = \mathbb{I}(X;Y) - f(P,W,\alpha) - 2\epsilon$, then the average probability vanishes as $k \to \infty$.

$$p_e^{avg} = p_e^1 \le \mathbb{P}(\hat{m} \in \{2,3,...,M\}|m=1) + \mathbb{P}(\hat{m} = e|m=1), \tag{11}$$

where the second term is the probability that the decoder declares an error (does not find any message) at the end of all $\binom{n}{k}$ choices, which implies that even if we pick the correct output symbols, the decoder either does not pass the first stage or does not declare $m = 1$ in the second stage. Therefore,

$$\mathbb{P}(\hat{m} = e|m=1) \le \mathbb{P}(Y^k \notin T_{[PW]_\mu}) + \mathbb{P}(Y_\star^{n-k} \notin T_{[W_\star]_\mu}) + \mathbb{P}(Y^k \notin T_{[W]_\mu}(C^k(1))) \tag{12}$$

$$\to 0, \text{ as } k \to \infty \tag{13}$$

where Y^k is the output of the channel if the input is $C^k(1)$, and $Y_\star$ is the output of the channel if the input is the noise symbol, and where we use union bound to establish (12). The limit (13) is because all the three terms in (12) vanish as $k \to \infty$ according to [5, Lemma 2.12].

The first term in (11) is more challenging. It is the probability that for at least one choices of the output symbols, the decoder passes the first stage and then the typicality decoder declares an incorrect message. Let the index $\underline{k}_i$ denote the condition on the i^{th} choice of the output symbols out of all $\binom{n}{k}$ choices. We characterize the choices based on the number of incorrectly chosen output symbols, i.e., the number of symbols in $\tilde{y}^k$ that are in fact outputs corresponding to a noise input, which is equal to the number of symbols in $\hat{y}^{n-k}$ that are in fact outputs corresponding to an input symbol of the codeword. Let the index k_1 denote the condition that the number of wrongly chosen output symbols is equal to k_1. For a specific k_1 there are $\binom{k}{k_1}\binom{n-k}{k_1}$ choices (According to Vandermonde's identity, we have $\sum_{k_1=0}^{n-k}\binom{k}{k_1}\binom{n-k}{k_1}=\binom{n}{k}$). Using the union bound for all the choices and all the messages $\hat{w} \ne 1$, we have

$$\mathbb{P}(\hat{m}{\in}\{2,3,...,M\}|m{=}1) \le \sum_{i=1}^{\binom{n}{k}} (M-1)\mathbb{P}_{\underline{k}_i}(\hat{m}=2|m=1)$$

$$= (e^{kR}-1)\sum_{k_1=0}^{n-k}\binom{k}{k_1}\binom{n-k}{k_1}\mathbb{P}_{k_1}(\hat{m}=2|m=1). \tag{14}$$

Note that message $\hat{m} = 2$ is declared at the decoder only if it passes the first stage, and then it is the unique message that satisfies (1). Therefore,

$$\mathbb{P}_{k_1}(\hat{m}=2|m=1) = \mathbb{P}_{k_1}\Big(\{\tilde{Y}^k \in T_{PW}\} \cap \{\hat{Y}^{n-k} \in T_{W_\star}\} \cap \{\tilde{Y}^k \in T_{[W]_\mu}(C^k(2))\}|m=1\Big)$$

$$= \mathbb{P}_{k_1}(\tilde{Y}^k \in T_{PW}) \cdot \mathbb{P}_{k_1}(\hat{Y}^{n-k} \in T_{W_\star}) \cdot \mathbb{P}(\tilde{Y}^k \in T_{[W]_\mu}(C^k(2))|m=1, \tilde{Y}^k \in T_{PW}, \hat{Y}^{n-k} \in T_{W_\star}) \tag{15}$$

$$\le e^{o(k)} e^{-kd_{k_1/k}(PW||W_\star)} e^{-(n-k)d_{k_1/(n-k)}(W_\star||PW)} \cdot e^{-k(\mathbb{I}(X;Y)-\epsilon)}, \tag{16}$$

where (15) follows from the independence of the events $\{\tilde{Y}^k \in T_{[W]_\mu}\}$ and $\{\hat{Y}^{n-k} \in T_{W_\star}\}$ conditioned on k_1 mismatched symbols, and (16) follows from using Lemma 1 for the first two terms in (15) and using Fact 1 for the last term in (15), because conditioned on message $m = 1$ being sent, $C^k(2)$ and $\tilde{Y}^k$ are independent regardless of the other conditions in the last term. Substituting (16) into the summation in (14), we have

$$\mathbb{P}(\hat{m} \in \{2,3,...,M\}|m=1) \le e^{o(k)}(e^{kR}-1)e^{-k(\mathbb{I}(X;Y)-\epsilon)} \cdot \sum_{k_1=0}^{n-k}\binom{k}{k_1}\binom{n-k}{k_1} e^{-kd_{k_1/k}(PW||W_\star)-(n-k)d_{k_1/(n-k)}(W_\star||PW)} \tag{17}$$

$$\le e^{o(k)} e^{kR} e^{-k(\mathbb{I}(X;Y)-\epsilon)} e^{kf(P,W,\alpha)} \tag{18}$$

$$= e^{o(k)} e^{-k\epsilon} \quad \to 0 \text{ as } k \to \infty, \tag{19}$$

where (18) is obtained by finding the exponent of the sum in (17), which is equal to the largest exponent of each term in the summation, since the number of terms is polynomial in k. To that end, let $\beta := k_1/(n-k)$ $(0 \le \beta \le 1)$, using Stirling's approximation as in (5) for each of the combinatorial terms in (17), we find that the largest exponent is $f(P,W,\alpha)$, which is defined in (10); and where (19) is obtained by substituting $R = \mathbb{I}(X;Y) - f(P,W,\alpha) - 2\epsilon$. Now, combining (11), (13), and (19), we have $p_e^{avg} \to 0$ as $k \to \infty$, which proves the Theorem. ■

IV. Example: BSC

Consider a BSC with crossover probability $0 \le p \le 0.5$. Figure 2 shows the two achievable rates obtained in Proposition 1 and Theorem 1 versus p for different values of $\alpha \ge 1$.

Not surprisingly, the achievable rate in Theorem 1 (indicated by "Thm 1") is always larger than the one in Proposition 1 (indicated by "Prop 1") since the two-stage decoding procedure takes advantage of the fact that the choice of the k output symbols might not be a good one. Specifically, the exponent obtained in Lemma 7 in terms of the partial divergence helps the decoder detect the right symbols, and therefore, achieve a larger rate. The arrows in Figure 2 show this difference and suggests that the benefit of using the two-stage decoding is larger for increasing α. Note that the larger α is, the smaller the achievable rate would be for a fixed p. Also, note that as $p \to 0$ the achievable rates approach a limit, which is equal

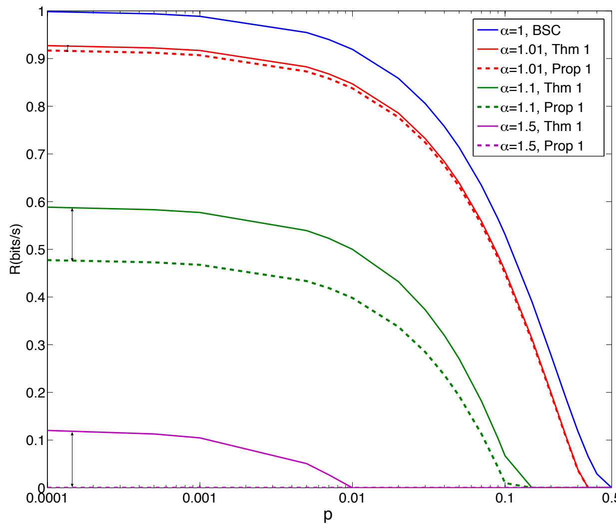

Fig. 2. Achievable rate for the BSC verses cross over probability p for different intermittency levels α's.

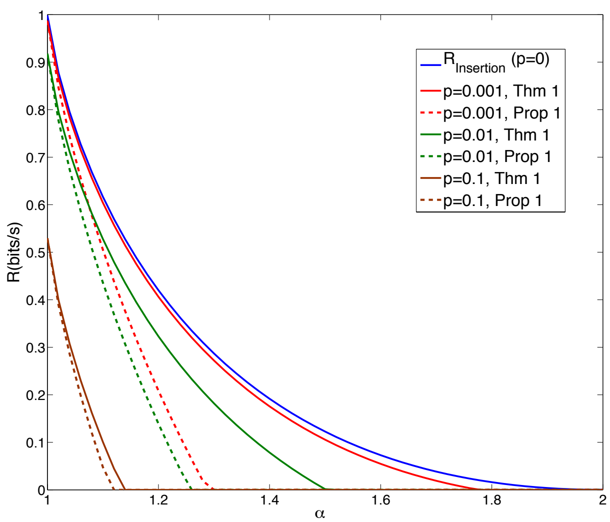

Fig. 3. Achievable rate for the BSC verses the intermittency level α for different cross over probabilities p's.

to $1-\alpha h(\frac{1}{\alpha})$ (bits/s) for the first achievable rate. However, it is more involved to obtain it for the second achievable rate, which is denoted by $R_{\text{Insertion}}$ and can be proven to be equal to $\max_{0\leq p_0\leq 1}\{2h(p_0)-\max_{0\leq\beta\leq 1}\{(\alpha-1)h(\beta)+h((\alpha-1)\beta)+(1-(\alpha-1)\beta)h(\frac{p_0-(\alpha-1)\beta}{1-(\alpha-1)\beta})\}\}$. Note that this is the achievable rate if the channel is noiseless ($p=0$), which models the insertion channel we discussed in Section II.

Figure 3, shows the value of $R_{\text{Insertion}}$ and the value of achievable rates for different p, versus α. Not surprisingly, as $\alpha\to 1$, the capacity of the BSC is approach for both of the achievable rates. Again, the difference between the achievable rates is more obvious for larger α. In this example, we cannot achieve a positive rate if $\alpha\geq 2$, even for the case of a noiseless channel ($p=0$). However, this is not true in general, because even the first achievable rate can be positive for a large α, if the capacity of the channel is sufficiently large. Figure 3 can be interpreted as the achievable region for (R,α). The results suggest that, as communication becomes more intermittent and α becomes larger, the achievable rate is decreased due to the additional uncertainty in the positions of the information symbols at the decoder.

APPENDIX A
PROOF OF LEMMA 1 FOR THE BINARY CASE

We have $t=2$, $\mathcal{X}=\{0,1\}$, $P=(p_0,p_1)$, $Q=(q_0,q_1)$, and $Q'=(q'_0,q'_1)$. Let $N(0)$ denote the number of 0's in $X_1^{k_1}$, and $N'(0)$ denote the number of 0's in $X_{k_1+1}^{k}$. We have

$$\mathbb{P}(X^k\in T_P)=\mathbb{P}(N(0)+N'(0)=kp_0)$$

$$=\sum_{l=0}^{kp_0}\mathbb{P}(N'(0)=l)\mathbb{P}(N(0)=kp_0-l) \tag{20}$$

$$=\sum_{l=0}^{kp_0}\binom{k_2}{l}(q'_0)^l(q'_1)^{k_2-l}\binom{k_1}{kp_0-l}q_0^{kp_0-l}q_1^{k_1-(kp_0-l)} \tag{21}$$

$$=(q'_1)^{k_2}q_0^{kp_0}q_1^{k_1-kp_0}\sum_{l=0}^{kp_0}\binom{k_2}{l}\binom{k_1}{kp_0-l}a_0^l;\quad a_0=\frac{q'_0q_1}{q_0q'_1}$$

$$=e^{-k(H(P)+D(P||Q)-\bar{\rho}\log\frac{q'_1}{q_1})}\sum_{l=0}^{kp_0}\binom{k_2}{l}\binom{k_1}{kp_0-l}a_0^l \tag{22}$$

$$\leq e^{-k(H(P)+D(P||Q)-\bar{\rho}\log\frac{q'_1}{q_1})}e^{o(k)}e^{ke(P,Q,Q',\rho)} \tag{23}$$

$$=e^{o(k)}e^{-kd(P,Q,Q',\rho)}, \tag{24}$$

where (20) and (21) follow from the fact that $N(0)$ and $N'(0)$ are independent and have binomial distributions, (22) follows from the definition of the entropy and divergence functions, (23) follows by the same procedure used to obtain (18) from (17), i.e., finding the largest exponent of the sum by defining $\theta_0 := l/(kp_0)$, where with simple algebraic manipulation the exponent becomes $e(P,Q,Q',\rho)$ defined in (9); and where (24) follows from definition (8).

REFERENCES

[1] A. Tchamkerten, V. Chandar, and G. W. Wornell, "Asynchronous communication: Capacity bounds and suboptimality of training," 2011. [Online]. Available: http://arxiv.org/abs/1105.5639v1

[2] D. Wang, V. Chandar, S.-Y. Chung, and G. Wornell, "Error exponents in asynchronous communication," in *Proc. IEEE Int. Symp. Information Theory (ISIT)*, St. Petersburg, Russia, Aug. 2011.

[3] R. Venkataramanan, S. Tatikonda, and K. Ramchandran, "Achievable Rates for Channels with Deletions and Insertions," in *Proc. IEEE Int. Symp. Information Theory (ISIT)*, St. Petersburg, Russia, Aug. 2011.

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

[5] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems;* 2nd edition, Cambridge University Press, 2011.

[6] 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.