Intermittent Communication and Partial Divergence

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 communication in which the transmissions of information symbols are bursty or the channel is sporadically available. We consider a general scenario in which information and noise symbols are grouped into packets of length l, and noise packets are inserted in the input sequence of the channel, and the receiver does not know a priori the positions of the information packets. Depending on the scaling behavior of the packet length relative to the codeword length, we identify some interesting scenarios for the scaling behavior of the receive window relative to the codeword length, and find achievable rates using different decoding structures. Interestingly, some of the decoding structures are based on a generalization of the method of types and properties of partial divergence. The achievable rates and the numerical results confirm the intuitive idea that increasing the intermittency decreases the achievable rates by increasing the uncertainty about the positions of the information packets at the receiver.**

I. Introduction

Intermittent communication models non-contiguous transmission of information symbols, i.e., communication scenarios in which the receiver observes some number of noise symbols between information symbols, but does not know a priori the locations of the noise symbols. In many practical applications transmitting a codeword can be intermittent due to lack of synchronization, shortage of transmission energy, or burstiness of the system.

Asynchronous communication is modeled in [1]–[5] by a single block transmission that starts at a random time and the receiver observes only noise before and after transmission, which is a special case of the intermittent communication model developed in this paper. Our system model can also be interpreted as an insertion channel in which some number of noise symbols are inserted between the codeword symbols. Although different from the insertion channels in the literature [7]–[9], our results may provide some insights about them. As another application, if the intermittent process is considered as a part of the transmitter, then we say that the transmitter is intermittent. Practical examples include 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.

Depending on the structure of the intermittency, the difficulty of locating the information symbols at the receiver can vary, leading to different intermittent communication scenarios and corresponding decoding structures. In this paper, we model the intermittency at the packet level, i.e., some number of noise packets are inserted between the information packets. Generally, increasing the packet length makes the transmission of the information symbols more contiguous, which decreases the uncertainty about their locations at the receiver. This observation suggests that the scaling behavior of the receive window should be determined based on the scaling behavior of the packet length in order to identify the regimes of interest. One extreme case arises if the packets correspond to a single symbol, so that the intermittent process operates at the symbol level, and the receive window scales linearly with the codeword length [6]. Another extreme arises if the codeword is represented by a single packet, so that the transmission of the information symbols is contiguous, and the receive window scales exponentially with the codeword length [3]. In this paper, we explore cases that lie between these two extremes. Specifically, we consider three different scenarios for the packet length relative to the codeword length, identify the corresponding relevant scaling behavior for the receive window, and find some achievability results for these communication scenarios using different decoding structures. Interestingly, some of the decoding structures are based on a generalization of the method of types and properties of partial divergence, which we explore in Section IV.

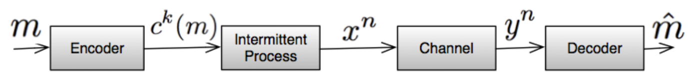

Fig. 1. System model for intermittent communication.

II. System Model

We consider a communication scenario in which a transmitter communicates a single message $m \in \{1, 2, ..., e^{kR} = M\}$ to a receiver over a discrete memoryless channel (DMC) with probability transition matrix W and input and output alphabets $\mathcal{X}$ and $\mathcal{Y}$. Let $\star \in \mathcal{X}$ denote the noise symbol, which is the input of the channel when the transmitter is silent. The transmitter encodes the message as a codeword $c^k(m)$ of length k, which is called the sequence of information

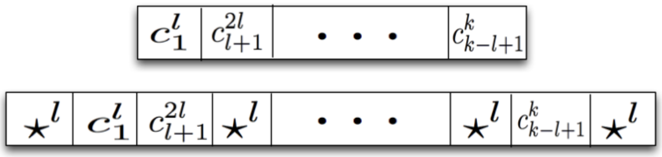

Fig. 2. Illustration of the intermittent process.

symbols. Assume that x^n and y^n are the input and output sequences of the channel, respectively, where $n \geq k$ is the length of the receive window at the decoder. Figure 1 shows a block diagram for the system model in which the intermittent process delivers the information symbols in some time slots and outputs the noise symbol $\star$ in the other time slots. We refer to this general scenario as intermittent communication.

The intermittent process considered in this paper captures the burstiness of the channel or the transmitter, is illustrated in Figure 2, and can be described as follows: First, the information symbols are grouped into packets of length l, resulting in a total of k/l information packets. Then, $(n-k)/l$ noise packets, i.e., sequences of noise symbols $\star$ of length l, are inserted arbitrarily between the information packets, i.e., the sequence x^n contains a total of n/l packets consisting of the k/l information packets and $(n-k)/l$ noise packets. 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. Therefore, we assume that the transmitter cannot decide on the positions of the information packets, so it cannot encode any timing information, and the receiver does not know the positions of the information packets, making the decoder's task more involved.

Assuming that the decoded message is denoted by $\hat{m}$, which is a function of the random sequence Y^n, 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 $k \to \infty$. Note that the communication rate is defined as $\log M/k$. The capacity is the supremum of all the achievable rates. In this paper, we focus on achievability results for different intermittent communication scenarios. Specifically, we consider three different scaling behaviors for the packet length l relative to the codeword length k, identify the corresponding regimes of interest for scaling behavior for the receive window n, and define the associated communication scenarios.

Definition 1: (Small packet intermittent communication) If the packet length l is finite and the receive window scales linearly relative to the codeword length with factor $\alpha \geq 1$, i.e., $n = \alpha k$, then the scenario is called *small packet intermittent communication*.

Definition 2: (Medium packet intermittent communication) If the packet length scales logarithmically relative to the codeword length, i.e., $l = \lambda \log k, \lambda > 0$, and the receive window relative to the codeword length follows a power law with power $\alpha \geq 1$, i.e., $n = lk^{\alpha}$, then the scenario is called *medium packet intermittent communication*.

Definition 3: (Large packet intermittent communication) If the packet length relative to the codeword length follows a power law, i.e., $l = k^{\lambda}$,$0 < \lambda \leq 1$, and the receive window scales exponentially relative to the packet length l with exponent $\alpha > 0$, i.e., $n = le^{\alpha l}$, then the scenario is called *large packet intermittent communication*.

For all the intermittent communication scenarios defined above, rate region (R, α) is said to be achievable if the rate R is achievable for the corresponding scenario with a given α. The reason that α is assumed to be larger than or equal to one in Definitions 1 and 2 is the necessary condition that $n \geq k$. Note that in all of the communication scenarios defined above, α determines the rate that the receive window scales with the codeword length (or the packet length), even though the scaling behavior itself depends on the scenario. In any case, the larger the value of α, the larger the receive window, and therefore, the more intermittent the system becomes. Hence, α is called *intermittency rate* throughout the paper. As we will see, increasing α generally reduces the achievable rate R for each of the above scenarios, because it makes the receive window larger, and therefore, increases the uncertainty about the positions of the information packets at the receiver making the decoder's task more involved.

The special case of $l = 1$ for small packet intermittent communication recovers the model and results in [6], and the special case of $\lambda = 1$ (or $l = k$) for large packet intermittent communication recovers the slotted asynchronous communication [3]. In the former case, the interesting scenario arises if the receive window scales linearly with the codeword length, whereas in the later case, the exponential scaling of the receive window with the codeword length is desirable. To motivate the different scaling behaviors for the receive window n in the definitions, note that increasing the packet length l adds more structure to the output sequence and decreases the uncertainty about the positions of the the codeword symbols at the receiver. As a result, some scalings of the receive window length do not lead to interesting tradeoffs. For example, if the receive window n scales linearly relative to the codeword length k, and the packet length l scales logarithmically with k, then the capacity of the channel can be always achieved, and the intermittency does not impact the communication rate.

Notation: We use $o(\cdot)$ and poly$(\cdot)$ to denote quantities that grow strictly slower than their arguments and are polynomial in their arguments, respectively. Most of the notation in this paper follows that in [2] and [10]. By $X \sim P(x)$, we mean X is distributed according to 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. We use the conventional definition $x^+ := \max\{x, 0\}$. Finally, if $0 \le \rho \le 1$, then $\bar{\rho} := 1 - \rho$.

III. Decoding Structures

In this section, three decoding structures are introduced of which one or more is used in Section V to obtain achievable rates for the communication scenarios defined in Section II. The encoding structure is identical for all the schemes: Given an 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. The three decoding structures include: *decoding from exhaustive search*, which attempts to decode the transmitted codeword from a selected set of output symbols without any attempt to first locate or detect the information symbols; *decoding from pattern detection*, which attempts to decode the transmitted codeword only if the selected outputs appear to be a pattern of information packets; and *decoding from packet detection*, which first detects the individual information packets, and then uses them to decode based on a conventional channel decoding procedure.

Although we focus on typicality for detection and decoding for ease of analyzing the probability of error, other algorithms such as maximum likelihood decoding can in principle be used in the context of these decoding structures. However, detailed specification and analysis of such structures and algorithms are beyond the scope of this paper.

For ease of presentation, let $s := k/l$ denote the number of information packets, and $b := n/l$ denote the total number of packets at the receiver, so that the number of inserted noise packets is equal to $b - s$.

A. Decoding From Exhaustive Search

In decoding from exhaustive search, the decoder observes the b packets of the output sequence y^n, chooses s of them, resulting in a sequence of $sl = k$ symbols denoted by $\tilde{y}^k$, and performs joint typicality decoding with a fixed typicality parameter $\mu > 0$, i.e., checks if

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

for all $(x, y) \in \mathcal{X} \times \mathcal{Y}$ and a unique index m, where P_m denotes the joint probability mass function induced by the type of codeword $c^k(m)$ and the channel W, defined by [2]

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

For convenience, we write $\tilde{y}^k \in T_{[W]_\mu}(c^k(m))$, if (1) is satisfied for m. In words, the condition (1) corresponds to the joint type for codeword $c^k(m)$ and selected outputs $\tilde{y}^k$ being close to the joint distribution induced by $c^k(m)$ and the channel $W(y|x)$. If the decoder finds a unique m satisfying (1), it declares m as the transmitted message. Otherwise, it makes another choice for the s packets from the b packets of the sequence y^n and again attempts typicality decoding. If at the end of all $\binom{b}{s}$ choices the typicality decoding procedure did not declare any message as being transmitted, then the decoder declares an error.

B. Decoding From Pattern Detection

Decoding from pattern detection involves two stages for each choice of the output symbols. As in decoding from exhaustive search, the decoder chooses s of the b packets from the output vector y^n. Let $\tilde{y}^k$ denote the sequence of the chosen output packets, and $\hat{y}^{n-k}$ denote the sequence of the other output packets. For each choice, the first stage checks if this choice of the output symbols is a good one, which consists of checking if $\tilde{y}^k$ is induced by a codeword, i.e., if $\tilde{y}^k \in T_{PW}$, and if $\hat{y}^{n-k}$ is generated by noise, i.e., if $\hat{y}^{n-k} \in T_{W_\star}$. If both of these conditions are satisfied, then we perform typicality decoding to $\tilde{y}^k$ over the codebook as described in Section III-A, which is called the second stage here. Otherwise, we make another choice for the s packets 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 $\binom{b}{s}$ choices, then the decoder declares an error. In this structure, we constrain the search domain for the typicality decoding (the second stage) only to typical patterns by checking that our choice of information packets satisfies the conditions in the first stage.

The first stage in this structure essentially distinguishes a sequence obtained partially from the codewords and partially from the noise from a codeword sequence or a noise sequence. As a result, in the analysis of the probability of error, partial divergence and its properties described in Section IV play a role. This structure always outperforms the decoding from exhaustive search structure, and their difference in performance indicates how much the results on the partial divergence improve the achievable rates.

C. Decoding From Packet Detection

Decoding from packet detection consists of two separate stages. In the first stage, the decoder completely locates the s information packets by checking if the i^{th} packet of the output sequence denoted by $y_{(i-1)l+1}^{il}$ is an information packet, i.e., if $y_{(i-1)l+1}^{il} \in T_{PW}, i = 1, 2, ..., b$. The decoder declares an error if after the first stage there are not exactly s detected information packets. In the second stage, the decoder forms a sequence consisting of all the detected information symbols in the first stage, and decodes the message with a conventional channel decoding procedure.

The complexity of this structure is significantly less than decoding from exhaustive search and decoding from pattern detection, because decoding from packet detection requires b typicality tests for locations, whereas the two other structures require $\binom{b}{s}$ typicality tests. However, this structure requires the packet length l to be sufficiently large in order to locate the individual information packets correctly. Obviously, this structure does not lead to an achievability result if the packet length is finite as in small packet intermittent communication. It also turns out that this structure does not work for medium packet intermittent communication. Therefore, decoding from packet detection is considered only for the large packet intermittent communication.

IV. Partial Divergence

We will see in Section V that the functional $d_\rho(P||Q)$, which we call partial divergence is relevant. Partial divergence is a generalization of the Kullback-Leibler divergence. In this section, we examine some of the interesting properties of partial divergence, which help develop insights about some of the achievable rates in Section V. Loosely speaking, partial divergence is the exponent of the probability that a sequence with independent elements generated partially according to one distribution and partially according to another distribution has a specific type. This exponent is useful in order to distinguish a sequence obtained partially from the codewords and partially from the noise from a codeword sequence or a noise sequence. The following Lemma from [6] is a generalization of the method of types in [10].

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 $Q' := (q'_0, q'_1, ..., q'_{t-1})$, and where P is a type with denominator k. 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

$$\mathbb{P}(X^k \in T_P) \leq e^{o(k)} e^{-kd(P,Q,Q',\rho)}, \tag{2}$$

where

$$d(P, Q, Q', \rho) := H(P) + D(P||Q) - \bar{\rho} \log \frac{q'_{t-1}}{q_{t-1}} - e(P, Q, Q', \rho), \tag{3}$$

$$e(P, Q, Q', \rho) := \max_{0 \leq \theta_j \leq 1, j=0,1,...,t-2} \{\rho H(P_1) + \bar{\rho} H(P_2) + \sum_{j=0}^{t-2} \theta_j p_j \log a_j\}, \tag{4}$$

$$a_j := \frac{q'_j q_{t-1}}{q_j q'_{t-1}},\ j = 0, 1, ..., t-2, \tag{5}$$

$$P_1 := (\frac{\bar{\theta}_0 p_0}{\rho}, \frac{\bar{\theta}_1 p_1}{\rho}, ..., \frac{\bar{\theta}_{t-2} p_{t-2}}{\rho}, 1 - \frac{\sum_{j=0}^{t-2} \bar{\theta}_j p_j}{\rho}), \tag{6}$$

$$P_2 := (\frac{\theta_0 p_0}{\bar{\rho}}, \frac{\theta_1 p_1}{\bar{\rho}}, ..., \frac{\theta_{t-2} p_{t-2}}{\bar{\rho}}, 1 - \frac{\sum_{j=0}^{t-2} \theta_j p_j}{\bar{\rho}}). \tag{7}$$

Proof: See [12]. ■

Specializing Lemma 1 for $Q' = Q$ results in [10, Lemma 2.6], and we have $d(P, Q, Q, \rho) = D(P||Q)$. However, we will be 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 a type P if its elements are generated independently partially according to Q and partially according to P, where the ratio of the mismatched symbols, i.e., generated from Q, to all the symbols is $\rho = k_1/k$. For this case, we call $d_\rho(P||Q) := d(P, Q, P, \rho)$ the partial divergence between P and Q with mismatch ratio $0 \leq \rho \leq 1$. Proposition 1 gives an explicit expression for the partial divergence by solving the optimization problem in (4) and simplifying (3) for the special case of $Q' = P$.

Proposition 1: Partial divergence can be written as

$$d_\rho(P||Q) = D(P||Q) - \sum_{j=0}^{t-1} p_j \log(c^* + \frac{p_j}{q_j}) + \rho \log c^* + h(\rho), \tag{8}$$

where c^* is a function of ρ, P, and Q, and can be uniquely determined from

$$c^* \sum_{j=0}^{t-1} \frac{p_j q_j}{c^* q_j + p_j} = \rho. \tag{9}$$

Proof: See [12]. ■

The next proposition states some of the properties of the partial divergence, which will be used in Section V to prove some of the properties of the achievable rates.

Proposition 2: The partial divergence $d_\rho(P||Q), 0 \leq \rho \leq 1$ has the following properties:

(a) $d_0(P||Q) = 0$.
(b) $d_1(P||Q) = D(P||Q)$.
(c) Partial divergence is zero if $P = Q$, i.e., $d_\rho(P||P) = 0$.
(d) Let $d'_\rho(P||Q) := \frac{\partial d_\rho(P||Q)}{\partial \rho}$ denote the derivative of the partial divergence with respect to ρ, then $d'_0(P||Q) = 0$.
(e) If $P \neq Q$, then $d'_\rho(P||Q) > 0$, for all $0 < \rho \leq 1$, i.e., partial divergence is increasing in ρ.
(f) If $P \neq Q$, then $d''_\rho(P||Q) > 0$, for all $0 \leq \rho \leq 1$, i.e., partial divergence is convex in ρ.
(g) $0 \leq d_\rho(P||Q) \leq \rho D(P||Q)$.

Proof: See [12]. ■

Figure 3 shows examples of the partial divergence for PMF's with alphabets of size 4. Specifically, $d_\rho(P||Q)$ versus ρ is sketched for $P = (0.25, 0.25, 0.25, 0.25)$, and two different Q's, $Q_1 = (0.1, 0.1, 0.1, 0.7)$ and $Q_2 = (0.1, 0.4, 0.1, 0.4)$. The properties in Proposition 2 are apparent in the figure for these examples.

V. Achievable Rates

In this section, we develop achievable rates for each of the intermittent communication scenarios defined in Section II based on the decoding structures introduced in Section III and the results on partial divergence developed in Section IV.

A. Small Packet Intermittent Communication

Using decoding from exhaustive search introduced in Section III-A for small packet intermittent communication model, we obtain the following achievability result.

Theorem 1: For small packet intermittent communication with parameters l and α, rates not exceeding $(C - \alpha h(1/\alpha)/l)^+$ are achievable, where C is the capacity of the DMC with stochastic matrix W.

Proof: See [12]. ■

The form of the achievable rate is reminiscent of communications overhead as the cost of constraints [11], where the constraint is the system's burstiness or intermittency, and the overhead cost is $\alpha h(1/\alpha)/l$. Note that the overhead cost is increasing in the intermittency rate α, is equal to zero

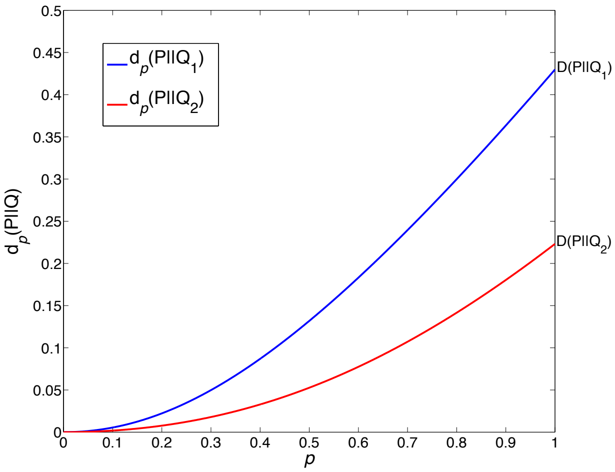

Fig. 3. Partial divergence $d_\rho(P||Q)$ versus ρ for $P = (0.25, 0.25, 0.25, 0.25)$, $Q_1 = (0.1, 0.1, 0.1, 0.7)$, and $Q_2 = (0.1, 0.4, 0.1, 0.4)$.

at $\alpha = 1$, and approaches infinity as $\alpha \to \infty$. These observations suggest that increasing the receive window makes the decoder's task more difficult. Also, note that the overhead cost and the packet length l are inversely proportional, which indicates that if the packet length is sufficiently large, then the achievable rate approaches the capacity of the channel. This is because increasing the packet length decreases the uncertainty about the positions of the information symbols at the decoder yielding a better achievability result.

Using decoding from pattern detection introduced in Section III-B for small packet intermittent communication model and the results on partial divergence developed in Section IV, we obtain the following achievability result.

Theorem 2: For small packet intermittent communication with parameters l and α, rates not exceeding $\max_P\{\left(\mathbb{I}(X;Y) - f_l^{SP}(P,W,\alpha)\right)^+\}$ are achievable, where

$$f_l^{SP}(P,W,\alpha) := \max_{0\le\beta\le1}\{\frac{(\alpha-1)h(\beta)+h((\alpha-1)\beta)}{l} - d_{(\alpha-1)\beta}(PW||W_\star) - (\alpha-1)d_\beta(W_\star||PW)\}. \quad (10)$$

Proof: See [12]. ∎

The achievable rate in Theorem 2 is larger than the one in Theorem 1, because decoding from pattern detection utilizes the fact that the choice of the information packets at the receiver might not be a good one, and therefore, restricts the typicality decoding only to the typical patterns and decreases the search domain. In Theorem 2, the overhead cost for a fixed input distribution is $f_l^{SP}(P,W,\alpha)$, and the next proposition states some of its properties.

Proposition 3: The overhead cost $f_l^{SP}(P,W,\alpha)$ in (10) has the following properties:

(a) The maximum of the term in (10) occurs in the interval $[0, 1/\alpha]$, i.e., instead of the maximization over $0 \le \beta \le 1$, $f_l^{SP}(P,W,\alpha)$ can be found by the same maximization problem over $0 \le \beta \le 1/\alpha$.
(b) $f_l^{SP}(P,W,\alpha)$ is increasing in α.
(c) $f_l^{SP}(P,W,1) = 0$.
(d) If l and $D(PW||W_\star)$ are finite, then $f_l^{SP}(P,W,\alpha) \to \infty$ as $\alpha \to \infty$.
(e) $f_l^{SP}(P,W,\alpha) \le f_1^{SP}(P,W,\alpha)/l$ for all integers $l \ge 1$, and the overhead cost is decreasing in l.

Proof: See [12]. ∎

Note that part (b) in Proposition 3 indicates that increasing the intermittency rate or the receive window increases the overhead cost, resulting in a smaller achievable rate. Parts (c) and (d) show that the achievable rate is equal to the capacity of the channel for $\alpha = 1$ and approaches zero as $\alpha \to \infty$. Also, part (e) implies that if the packet length is sufficiently large, then the achievable rate approaches the capacity of the channel.

B. Medium Packet Intermittent Communication

Using decoding from exhaustive search introduced in Section III-A for the medium packet intermittent communication model, we obtain the following achievability result.

Theorem 3: For medium packet intermittent communication with parameters λ and α, rates not exceeding $(C-(\alpha-1)/\lambda)^+$ are achievable, where C is the capacity of the DMC with stochastic matrix W.

Proof: The proof is similar to the proof of Theorem 1. See [12] for details. ∎

Here, the overhead cost is $(\alpha-1)/\lambda$, which is increasing in the intermittency rate α and decreasing in λ, and the same conclusions as in Section V-A can be drawn.

Using decoding from pattern detection introduced in Section III-B for medium packet intermittent communication model and the results on partial divergence developed in Section IV, we obtain the following achievability result.

Theorem 4: For medium packet intermittent communication with parameters λ and α, rates not exceeding $\max_P\{\left(\mathbb{I}(X;Y) - f_\lambda^{MP}(P,W,\alpha)\right)^+\}$ are achievable, where

$$f_\lambda^{MP}(P,W,\alpha) := \max_{0\le\beta\le1}\{\beta\frac{\alpha-1}{\lambda} - d_\beta(PW||W_\star)\}. \quad (11)$$

Proof: The proof is similar to the proof of Theorem 2. See [12] for details. ∎

For the same reason as in Section V-A, the achievable rate in Theorem 4 is larger than the one in Theorem 3. The overhead cost in Theorem 4 is equal to $f_\lambda^{MP}(P,W,\alpha)$, which is increasing in the intermittency rate α, equals zero at $\alpha = 1$, approaches infinity as $\alpha \to \infty$, and is decreasing in λ as can be seen from (11). The same conclusions as in Section (V-A) can be drawn.

C. Large Packet Intermittent Communication

Using decoding from packet detection introduced in Section III-C for large packet intermittent communication model, we obtain the following achievability result.

Theorem 5: For large packet intermittent communication with parameters λ and α, rates not exceeding $\max_P\{\left(\mathbb{I}(X;Y) - f^{LP}(P,W,\alpha)\right)^+\}$ are achievable, where

$$f^{LP}(P,W,\alpha) := (\alpha - D(PW||W_\star))^+ . \quad (12)$$

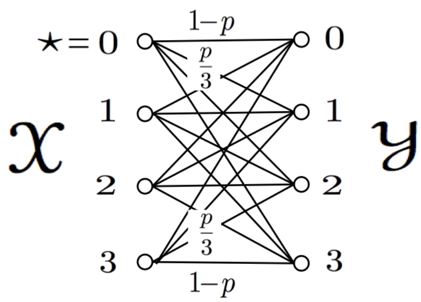

Fig. 4. Graphical description of the transition matrix for the DMC w consider in this section.

Proof: See [12].

Remark 1. As we mentioned before, decoding fror packet detection cannot be used for the small packet inter mittent communication model in which the packet length l is finite. Decoding from packet detection cannot be used for the medium packet intermittent communication model either, because there are too many smaller packets compared to the large packet scenario. See [12] for more details.

Remark 2. For the large packet intermittent communication model, we can use decoding from exhaustive search and decoding from pattern detection. However, it turns out that the resulting achievable rates are strictly smaller than the one in Theorem 5. Therefore, we only focus on decoding from packet detection for the large packet intermittent communication model.

This achievability result is identical to the capacity of the asynchronous communication obtained in [5]. Note that the overhead cost $f^{LP}(P, W, \alpha)$ is independent of the value of λ. Intuitively, this happens because, as λ increases in the large packet intermittent communication model, the scaling behavior of the receive window relative to the codeword length changes (the receive window is exponentially scaled with the packet length l) in a way that compensates this increase in the packet length, and therefore, the achievable rate does not change. As before, the overhead cost $f^{LP}(P, W, \alpha)$ is increasing in the intermittency rate α, which indicates that increasing the receive window results in a smaller achievable rate.

VI. Numerical Results

We consider a DMC with a symmetric transition matrix with input and output alphabets of size 4 as is depicted in Figure 4 in which all the cross over probabilities are equal to $p/3$ and the direct probabilities are equal to $1-p$. The reason that we consider a channel with 4-ary input and output is that the benefit of using the concept of partial divergence and the results in Section IV is more apparent for a channel with non-binary alphabets. Numerical results for the case of binary symmetric channel are provided in [6] for the special case of $l = 1$ for the small packet intermittent communication model.

The boundary of the achievable rate region (R, α) characterizes the tradeoff between the achievable rates and the intermittency rate α. Figure 5 illustrates the achievable rate region (R, α) for the small packet intermittent communication model over the channel depicted in Figure 4 with $p = 0.1$, for which the capacity is approximately 1.37 bits per channel use. The achievable rate regions correspond to the results in Theorems 1 and 2 for two values of the packet length: $l = 1$ and $l = 2$. The achievable rates are decreasing in the intermittency rate α as we have discussed in Section V, which is because increasing α increases size of the receive window and therefore the uncertainty about the positions of the information packets at the receiver. All the achievable rates approach the capacity of the channel as $\alpha \to 1$. Also, note that the achievable rate region is larger for a larger packet length, because increasing l adds more structure to the input and output of the channel reducing the uncertainty about the positions of the output symbols at the receiver.

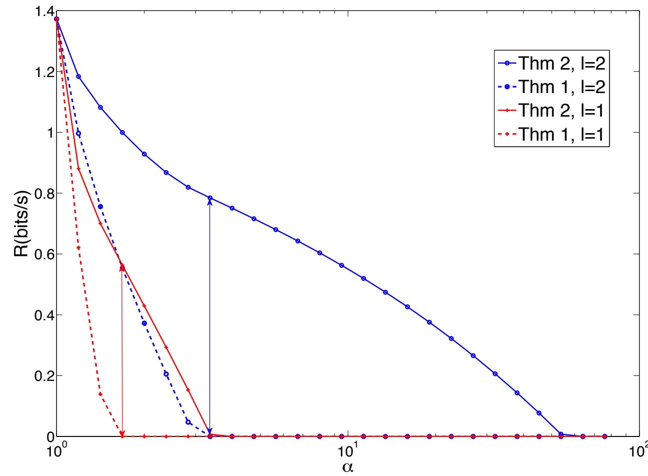

Fig. 5. Achievable rate region (R, α) for the small packet intermittent communication model over the channel depicted in Figure 4 with $p = 0.1$.

The arrows in Figure 5 show the differences between the rates obtained from Theorems 1 and 2, i.e., how much decoding from pattern detection outperforms decoding from exhaustive search. As can be seen from the figure, decoding from pattern detection increases the achievable rate as well as substantially increases the range of intermittency rates for which the achievable rate is non-zero. The reason is that as the receive window becomes larger, the search domain increases exponentially, and the need for restricting the search domain by decoding the codeword only from typical patterns becomes more critical.

Figure 6 illustrates the achievable rate region (R, α) for the medium packet intermittent communication model over the channel depicted in Figure 4 with $p = 0.1$. The achievable rate regions corresponds to the results in Theorems 3 and 4 for two values of λ: $\lambda = 1$ and $\lambda = 2$. Similar observations and conclusions can be made as those for Figure 5.

Finally, Figure 7 illustrates the achievable rate region (R, α) in Theorem 5 for the large packet intermittent communication model over the channel depicted in Figure 4

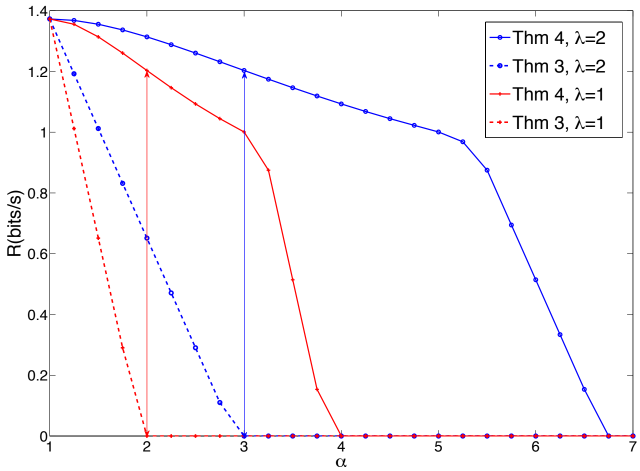

Fig. 6. Achievable rate region (R, α) for the medium packet intermittent communication model over the channel depicted in Figure 4 with $p = 0.1$.

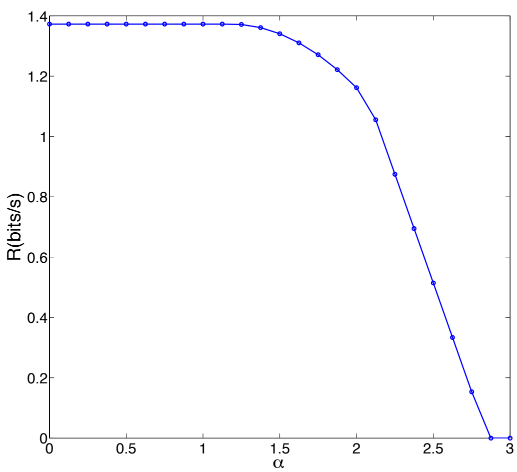

Fig. 7. Achievable rate region (R, α) for the large packet intermittent communication model over the channel depicted in Figure 4 with $p = 0.1$.

with $p = 0.1$. As before, increasing the intermittency rate α reduces the achievable rate since it increases the uncertainty about the information symbols at the receiver, and if α is small enough, then the capacity of the DMC can be achieved, which is similar to the observation in [5].

VII. Conclusion

In this paper, we focused on obtaining achievable rates for intermittent communication, which models a bursty transmitter or a sporadically available channel. Inspired by network applications, we considered packetized transmission of data and introduced three intermittent communication scenarios depending on the scaling behavior of the packet length, which models different levels of asynchronism. We introduced three receiver structures. We can conclude that if the packet length is sufficiently large, then decoding from packet detection gives the largest achievable rate. The achievable rates and the numerical results confirm the intuitive idea that increasing the intermittency and therefore the receive window decreases the achievable rates by increasing the uncertainty about the positions of the information packets at the receiver, and increasing the packet length increases the achievable rates by adding more structure to the output sequence or reducing the level of asynchronism at the receiver. We introduced the concept of partial divergence and studied some of its properties in order to obtain stronger achievability results. The results on the partial divergence may be of independent interest, such as in asynchronous random access communication models in which the decoder does not know a priori that a received symbol or packet corresponds to which user.

References

[1] 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.

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

[3] 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.

[4] V. Chandar, A. Tchamkerten, and D. N. C. Tse, "Asynchronous capacity per unit cost," *IEEE Trans. on Inf. Theory*, submitted 2011. [Online]. Available: `http://arxiv.org/abs/1007.4872v1`

[5] Y. Polyanskiy, "Asynchronous communication: universality, strong converse and dispersion," *IEEE Trans. on Inf. Theory*, submitted 2012.

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

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

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

[9] 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.

[10] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems,* New York: Academic, 1981.

[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.

[12] M. Khoshnevisan and J. N. Laneman, "Intermittent Communication and Partial Divergence," in preparation for submission to *IEEE Trans. on Inf. Theory*, 2012.