INTERMITTENT COMMUNICATION

A Dissertation

Submitted to the Graduate School
of the University of Notre Dame
in Partial Fulfillment of the Requirements
for the Degree of

Doctor of Philosophy

by

Mostafa Khoshnevisan

J. Nicholas Laneman, Director

Graduate Program in Electrical Engineering
Notre Dame, Indiana

January 2014

INTERMITTENT COMMUNICATION

Abstract

by

Mostafa Khoshnevisan

We formulate a model for intermittent communication that can capture bursty transmissions or a sporadically available channel, where in either case the receiver does not know a priori when the transmissions will occur. For the point-to-point case, we develop two decoding structures and their achievable rates for such communication scenarios. One structure determines the transmitted codeword through exhaustive search, and the other structure first detects the locations of codeword symbols and then uses them to decode. We introduce the concept of partial divergence and study some of its properties in order to obtain stronger achievability results. As the system becomes more intermittent, the achievable rates decrease due to the additional uncertainty about the positions of the codeword symbols at the decoder. Additionally, we provide upper bounds on the capacity of binary noiseless intermittent communication with the help of a genie-aided encoder and decoder. The upper bounds imply a tradeoff between the capacity and the intermittency rate of the communication system, even if the receive window scales linearly with the codeword length.

Upon this foundation, we develop two extensions. First, we extend the model to intermittent multi-access communication for two users that captures the bursty transmission of the codeword symbols for each user and the possible asynchronism between the receiver and the transmitters as well as between the transmitters themselves. This model can be viewed as another attempt to combine information-theoretic and

network-oriented multi-access models. We characterize the performance of the system in terms of achievable rate regions. In our achievable schemes, the intermittency of the system comes with a significant cost. Second, we extend the model to packet-level intermittent communication in which codeword and noise symbols are grouped into packets. Depending on the scaling behavior of the packet length relative to the codeword length, we identify some interesting scenarios, and characterize the performance of the system in terms of the achievable rates for each model.

Finally, we apply the insights and tools developed for intermittent communication to several related problems. First, we obtain new results on the capacity of deletion channels and a random access model that drops the collided symbols. Second, we study the problem of lossless source coding in the presence of intermittent side-information.

To my loving parents, Maryam and Mohammad,

and my adorable sister, Farzaneh

CONTENTS

FIGURES

ACKNOWLEDGMENTS

First and foremost, I would like to thank my advisor, Dr. J Nicholas Laneman, for his genius, generosity, and unlimited patience. I am very grateful for all the guidance, advice, and more importantly, the freedom in research he gave me during these five years of my research life. His support and motivation is what made this dissertation possible. I also wish to thank my committee members, Dr. Thomas Fuja, Dr. Martin Haenggi, and Dr. Bertrand Hochwald, who were more than generous with their expertise and precious time.

A special thanks goes to multiple generations of the Laneman research group for all the helpful technical discussions and the friendship we developed during my time at Notre Dame. I would like to thank the senior members of the group: Matthieu, Brian, and Michael; the young members: Sahand and Mingming; and last but not least, the members in the group whom I spent most time with: Ebrahim, Glenn, Utsaw, Zhanwei, and Peyman.

I would like to thank all my friends at Notre Dame, who were the source of support and encouragement during my five-years graduate life. In particular, I would like to thank Elham, Amin, Arman, Ali, Houman, Golnaz, Farzad, and Somayeh to just name a few. My deepest gratitude goes to my parents for their love, support, and encouragement throughout my life. I will not begin to explain how grateful I am to my family here, as that would probably need a dissertation for itself. Finally, I would like to thank three favorite people in my life: my sister, Farzaneh, and my cousins, Nafiseh and Maryam.

CHAPTER 1

INTRODUCTION

Communication systems are traditionally analyzed assuming contiguous transmission of encoded symbols through the channel. However, in many practical applications such an assumption may not be 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 may not explicitly know whether a given output symbol of the channel is the result of sending a symbol of the codeword or is simply a noise symbol containing no information about the message. This dissertation provides one model, called *intermittent communication* for non-contiguous transmission of codewords in such settings. Figure 1.1 compares the contiguous and intermittent transmission of the codeword symbols.

In this chapter, we first provide some motivations and practical examples in Section 1.1. Next, we summarize the contributions of the dissertation in Section 1.2. Finally, we summarize the outline of the dissertation in Section 1.3.

1.1 Motivations and Practical Examples

Intermittent communications captures bursty transmissions or a sporadically available channel, where in either case the receiver does not know a priori when the transmissions will occur. In the system model for intermittent communication, we consider a random process in between the transmitter and the channel, which we call *intermittent process*.

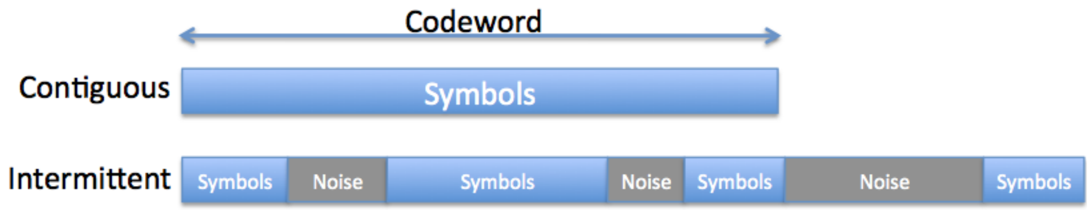

Figure 1.1. Comparing contiguous and intermittent transmission of codeword symbols.

If the intermittent process is considered to be part of the channel behavior, then intermittent communication models a sporadically available channel in which at some times a symbol from the codeword is sent, and at other times the receiver observes only noise. We extend the model beyond individual symbols to consider intermittent transmission of packets that are part of a larger codeword. As another application, we can think of the intermittent process as the random amount of delay that occurs for transmission of each symbol in a communication network, in which the realization of the delay is not known to the receiver, and during this delay the receiver observes only noise. The model can be interpreted as an insertion channel in which some number of silent / noise symbols are inserted between the codeword symbols.

If the intermittent process is considered to be part of the transmitter, then we say that the transmitter is intermittent. Practical examples include energy harvesting systems in which the transmitter harvests energy usually from a natural source and uses it for transmission. Assuming that there is a special input that can be transmitted with zero energy, the transmitter sends the symbols of the codeword if there is enough energy for transmission, and sends the zero-cost symbol otherwise.

The framework for intermittent communication we develop in this dissertation can also be extended to multi-access communication. In this case, practical examples include a cognitive radio in which the primary user is bursty, i.e., sends codeword symbols or packets at some times and remains silent at other times, and a secondary user also wants to communicate with the same receiver and can sense the channel and transmit its codeword symbols whenever the primary user is silent. As another application, consider an ALOHA random access protocol with a collision-avoidance mechanism in which at each time either only one of the users transmits or the channel remains idle.

Packet-level intermittent communication arises in network applications in which the data is packetized, and each packet goes through a network with intermittent connectivity and / or random delay. As a result, packets are received at some random intervals after experiencing some sort of roaming and / or delay. The receiver's task is to identify the packets in order to reconstruct the original data. The intermittency of the system and the size of the packet might allow for individual packet detection at the receiver, in which case the communication problem simplifies to individual packet detection and message decoding. However, if the intermittency of the system increases or the packet length decreases, then we may not reliably detect all the data packets individually, and encoding and decoding across packets could be a better solution. In this dissertation, we extend the symbol-level model for intermittent communication to the packet-level, with the packet lengths having different scaling relative to the codeword length. We refer to the case in which individual packet detection is effective as *large-packet intermittent communication*, and to two other cases as *small-* and *medium-packet intermittent communication.*

1.2 Contributions of the Dissertation

In this section, we summarize the contribution of the dissertation.

- We formulate a model for intermittent communication that can capture bursty transmissions or a sporadically available channel by inserting a random number of silent symbols between each codeword symbol, where the receiver may not know a priori when the transmissions occur. The intermittency rate in this model controls how widely separated, on averages, the transmission bursts are. We consider a linear scaling for the receive window relative to the codeword length.

- We introduce a quantity called partial divergence, which is a generalization of the Kullback-Leibler divergence, and is the normalized 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 characterizing a decoder's ability to distinguish a sequence obtained partially from the codewords and partially from the noise from a codeword sequence or a noise sequence. We study some of the properties of partial divergence that provide insights about the achievable rates for intermittent communication.

- Using the results on partial divergence, we show that as long as the intermittency rate is finite and the capacity of the channel is not zero, rate $R = 0$ with only two messages is achievable for intermittent communication. Therefore, no matter how large the intermittency rate becomes, if it is finite, the receiver can distinguish between two messages with vanishing probability of error.

- We specify two decoding structures in order to develop achievable rates: decoding from exhaustive search and decoding from pattern detection. Decoding from pattern detection, which achieves a larger rate, is based on the partial divergence and its properties. As the system becomes more intermittent, the achievable rates decrease due to the additional uncertainty about the positions of the codeword symbols at the decoder.

- For the case of a binary-input binary-output noiseless channel, we obtain upper bounds on the capacity of intermittent communication 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 and the intermittency rate.

- We derive bounds on the capacity per unit cost of intermittent communication. To obtain the lower bound, we use pulse-position modulation at the encoder, and searched for the position of the pulse at the decoder. The achievable rate per unit cost decreases as the system becomes more intermittent.

- Extending to multi-user communication, we formulate a model for intermittent multi-access for two users that captures the bursty transmission of the codeword symbols for each user and the possible asynchronism between the receiver and the transmitters as well as between the transmitters themselves. This model can be viewed as an attempt to combine information-theoretic and network-oriented multi-access models. By making different assumptions for the intermittent process, we specialize the system to three models: random access with no idle-times and no collisions, random access with idle-times and no collisions, and random access with collisions and no idle-times. For each model, we characterize the performance of the system in terms of an achievable rate region. In our achievable schemes, the intermittency of the system comes with a significant cost, i.e., it reduces the size of the achievable rate regions, which can be interpreted as communication overhead.

- Inspired by network applications, we extend the model to packet-level intermittent communication in which codeword and noise symbols are grouped into packets. Depending on the scaling behavior of the packet length relative to the codeword length, we identify three scenarios: small-packet, medium-packet, and large-packet intermittent communication. For small- and medium-packet intermittent communication, we utilize both decoding from exhaustive search and decoding from pattern detection in order to obtain achievable rates, whereas, for large-packet intermittent communication, we utilize decoding from packet detection in order to obtain achievable rates. In all three cases, the intermittency rate determines the scaling of the receive window relative to the codeword length (or the packet length), even though the scaling behavior itself depends on the scenario. Increasing the intermittency rate generally reduces the achievable rate for each of the three scenarios, because it makes the receive window larger, and therefore, increases the uncertainty about the positions of the codeword packets at the receiver, making the decoder's task more involved.

- We use some of the insights and tools developed in this dissertation to obtain some new results on related problems. Specifically, we first use a similar decoding structure to decoding from exhaustive search in conjunction with a lemma on the longest common subsequence of random sequences to prove a side result on lower bounding the capacity of the deletion channels. Second, we obtain achievability results for a random access model that drops / deletes collided symbols using a similar decoding structure to decoding from pattern detection.

- Inspired by the problem of file synchronization in which we compress a source sequence with the benefit of decoder side-information that is related to the source via insertions, deletions, and substitutions, we study a similar problem in which the side-information at the decoder is related to the source via an intermittent process. Focusing on achievability, we introduce encoding and decoding structures in order to compress the source at the encoder and reconstruct it reliably at the decoder.

1.3 Outline of the Dissertation

In Chapter 2, we briefly review system models and main results of some related work and identify the gaps in the literature. After reviewing some of the results from the method of types, we study the system model and main results on frame synchronization, asynchronous communication, and insertion / deletion channels. Finally, we discuss several networking issues, such as asynchronism, collisions, and random access in the context of multiple-access communication with an information-theoretic approach.

In Chapter 3, we present a lemma, which generalizes some of the results of method of types, and then specialize the lemma to a certain case in which the partial divergence will be defined. Next, we study some of the properties of the partial divergence that provide insights about the achievable rates for intermittent communication in Chapter 4. Finally, we generalize the lemma and the partial divergence to the case of three distributions, which will be used in Chapter 5.

In Chapter 4, we introduce a model for single-user intermittent communication that consists of an intermittent process followed by a discrete memoryless channel (DMC), and develop two coding theorems for achievable rates to lower bound the capacity. Toward this end, we use some of the results on partial divergence and its properties from Chapter 3. We show that, as long as the ratio of the receive window to the codeword length is finite and the capacity of the DMC is not zero, rate $R = 0$ is achievable for intermittent communication. By using decoding from exhaustive search and decoding from pattern detection, we obtain two achievable rates that are also valid for arbitrary intermittent processes. Next, we focus on the binary-input binary-output noiseless channel, and obtain upper bounds on the capacity of intermittent communication. Finally, we develop lower and upper bounds on the capacity per unit cost of intermittent communication.

In Chapter 5, we generalize the single-user intermittent communication model introduced in Chapter 4 to multi-user intermittent communication for two users that captures the bursty transmission of the codeword symbols for each user and the possible asynchronism between the receiver and the transmitters as well as between the transmitters themselves. By making different assumptions for the intermittent process, we specialize the system to three models: random access with no idle-times and no collisions, random access with idle-times and no collisions, and random access with collisions and no idle-times. For each model, we obtain an achievable rate region that depend on the concept of partial divergence and decoding from pattern detection.

In Chapter 6, we consider three extensions. First, we introduce a system model for packet-level intermittent communication in which codeword and noise symbols are grouped into packets. Depending on the scaling behavior of the packet length relative to the codeword length, we identify three scenarios: small-packet, medium-packet, and large-packet intermittent communication. For each model, we obtain some achievability results. Next, we use some of the insights and tools developed in this dissertation to obtain some new results on the capacity of deletion channels and a random access that drops the collided symbols. Finally, we study the problem of lossless source coding in the presence of intermittent side-information, and introduce encoding and decoding structures in order to compress the source at the encoder and reconstruct it reliably at the decoder.

In Chapter 7, we conclude the dissertation and introduce some directions for future research.

CHAPTER 2

BACKGROUND

In this chapter, we briefly summarize the system model and main results of some related work, identify the gaps in the literature, and explain how the models and results in the remaining chapters close the gaps to some extent. Additionally, we would like to provide insights about intermittent communication as well as the techniques used to obtain the results.

Before studying the relevant literature, we briefly review the method of types in Section 2.1 since we make frequent use of this tool, and specifically, the notations in our analysis. In Section 2.2, we review the literature on frame synchronization in single-user communication systems, which studies the problem of locating a sync pattern in a string of data. In Section 2.3, we study the problem of joint frame synchronization and decoding in asynchronous communication, which corresponds to contiguous transmission of codeword symbols in which the receiver observes noise before and after transmission. In Section 2.4, we summarize the results on the lower and upper bounds of the capacity of insertion / deletion channels, which are also called channels with synchronization errors. In Section 2.5, we discuss on several networking issues, such as asynchronism, random access, and collisions in the context of multiple-access communication with an information-theoretic approach.

A key assumption of these communication models is the lack of knowledge of the state of the channel and / or a timing reference at the transmitter and receiver, capturing certain kinds of asynchronism. Generally, this asynchronism makes the task of the receiver more difficult since it must acquire synchronization in the first place. Intermittent communication is just another attempt to model asynchronism at the symbol or packet level.

Before proceeding to the next sections, we summarize several notations that are used throughout the sequel:

- We use $o(\cdot)$ and $\text{poly}(\cdot)$ to denote quantities that grow strictly slower than their arguments and are polynomial in their arguments, respectively.
- By $X \sim P(x)$, we mean that random variable X is distributed according to probability distribution P.
- $h(p) := -p \log p - (1-p)\log(1-p)$ is the binary entropy function, and for $\beta_1 + \beta_2 < 1$, let $h(\beta_1, \beta_2)$ denote the entropy of the ternary probability mass function $(\beta_1, \beta_2, 1-\beta_1-\beta_2)$.
- We use the convention that the binomial coefficient $\binom{n}{k} = 0$ 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.
- We use the conventional definition $x^+ := \max\{x, 0\}$.
- If $0 \leq \rho \leq 1$, then $\bar{\rho} := 1 - \rho$.
- $\doteq$ denotes equality in the exponential sense as $k \to \infty$, i.e., $\lim_{k\to\infty} \frac{1}{k}\log(\cdot)$ of both sides are equal. In the same way, $\dot{\leq}$ and $\dot{\geq}$ denote inequalities in the exponential sense.
- Let $[1:M]$ denote the set of integers $\{1, 2, ..., M\}$.
- We denote sequences / vectors using both boldface and superscript, i.e., $\mathbf{x} = x^k = (x_1, x_2, ..., x_k)$.

2.1 Method of Types

The method of types is a powerful technique in large deviation theory which was developed by Csiszár and Körner [9], [10], and is summarized in [7, Chapter 12.1]. We now briefly review the definitions and results.

Let $\mathcal{P}^{\mathcal{X}}$ denote the set of probability distributions over the finite alphabet $\mathcal{X}$. The *empirical distribution* (or *type*) of a sequence $x^n \in \mathcal{X}^n$ is denoted by $\hat{P}_{x^n} \in \mathcal{P}^{\mathcal{X}}$ and is defined as

$$\hat{P}_{x^n}(x) := \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{[x_i=x]},$$

where $\mathbb{1}_{[\cdot]}$ is the indicator function. A sequence x^n is said to have a *type* $P \in \mathcal{P}^{\mathcal{X}}$ if $\hat{P}_{x^n} = P$. The set of all sequences that have type P is denoted $T_P^n \subseteq \mathcal{X}^n$, or more simply T_P. Joint empirical distributions on $\mathcal{X} \times \mathcal{Y}$ are denoted similarly, i.e.,

$$\hat{P}_{x^n,y^n}(x,y) := \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{[x_i=x,y_i=y]}.$$

Let $\mathcal{P}^{\mathcal{Y}|\mathcal{X}}$ denote the set of probability distributions over the finite alphabet $\mathcal{Y}$ conditioned on the finite alphabet $\mathcal{X}$. A sequence y^n is said to have a conditional empirical distribution $\hat{P}_{y^n|x^n}$ given x^n, if for all $(x,y) \in \mathcal{X} \times \mathcal{Y}$,

$$\hat{P}_{x^n,y^n}(x,y) = \hat{P}_{x^n}(x)\hat{P}_{y^n|x^n}(y|x),$$

and the set of sequences y^n that have a conditional type W given x^n is denoted by $T_W(x^n)$.

For $P, P' \in \mathcal{P}^{\mathcal{X}}$ and $W, W' \in \mathcal{P}^{\mathcal{Y}|\mathcal{X}}$, the *Kullback-Leibler divergence* between P and P' is defined as [7]

$$D(P\|P') := \sum_{x\in\mathcal{X}} P(x)\log\frac{P(x)}{P'(x)},$$

and the conditional information divergence between W and W' conditioned on P is defined as

$$D(W\|W'|P) := \sum_{x\in\mathcal{X}} P(x)\sum_{y\in\mathcal{Y}} W(y|x)\log\frac{W(y|x)}{W'(y|x)}.$$

The *average mutual information* between $X \sim P$ and $Y \sim PW$ and coupled via

$P_{Y|X} = W$ is defined as

$$\mathbb{I}(P, W) := \sum_{x\in\mathcal{X}} P(x) \sum_{y\in\mathcal{Y}} W(y|x) \log \frac{W(y|x)}{(PW)(y)}.$$

We also use $\mathbb{I}(X;Y)$ to denote the average mutual information in this dissertation.

With these definitions, we now state the following lemmas, which are used throughout the dissertation.

Lemma 2.1. *([10, Lemma 1.2.6]): If X^n is an independent and identically distributed (iid) sequence according to P', then the probability that it has a type P is bounded by*

$$\frac{1}{(n+1)^{|\mathcal{X}|}} e^{-nD(P\|P')} \le \mathbb{P}(X^n \in T_P) \le e^{-nD(P\|P')}.$$

Also, if the input $x^n \in \mathcal{X}^n$ to a memoryless channel $W' \in \mathcal{P}^{\mathcal{Y}|\mathcal{X}}$ has type P, then the probability that the observed channel output sequence Y^n has a conditional type W given x^n is bounded by

$$\frac{1}{(n+1)^{|\mathcal{X}||\mathcal{Y}|}} e^{-nD(W\|W'|P)} \le \mathbb{P}(Y^n \in T_W(x^n)) \le e^{-nD(W\|W'|P)}.$$

A sequence $x^n \in \mathcal{X}^n$ is called *P-typical with constant μ*, denoted $x^n \in T_{[P]_\mu}$, if

$$|\hat{P}_{x^n}(x) - P(x)| \le \mu \text{ for every } x \in \mathcal{X},$$

and a sequence $y^n \in \mathcal{Y}^n$ is called *W-typical conditioned on $x^n \in \mathcal{X}^n$ with constant μ*, denoted $y^n \in T_{[W]_\mu}$, if

$$|\hat{P}_{x^n,y^n}(x,y) - \hat{P}_{x^n}(x)W(y|x)| \le \mu \text{ for every } (x,y) \in \mathcal{X} \times \mathcal{Y}.$$

Lemma 2.2. *([10, Lemma 1.2.12]): If X^n is an iid sequence according to P, then*

$$\mathbb{P}(X^n \in T_{[P]_\mu}) \geq 1 - \frac{|\mathcal{X}|}{4n\mu^2}.$$

Also, if the input $x^n \in \mathcal{X}^n$ to a memoryless channel $W \in \mathcal{P}^{\mathcal{Y}|\mathcal{X}}$, and Y^n is the output, then

$$\mathbb{P}(Y^n \in T_{[W]_\mu}(x^n)) \geq 1 - \frac{|\mathcal{X}||\mathcal{Y}|}{4n\mu^2}.$$

Remark 2.1. *([10]): In Lemma 2.2 the terms subtracted from 1 could be replaced even by exponentially small terms $2|\mathcal{X}|e^{-2n\mu^2}$ and $2|\mathcal{X}||\mathcal{Y}|e^{-2n\mu^2}$, respectively.*

Finally, we state a stronger version of the packing lemma [23, Lemma 3.1] that will be useful in typicality decoding, and is proved in [54, Equations (24) and (25)] based on the method of types.

Lemma 2.3. *Assume that X^n and $\tilde{Y}^n$ are independent, X^n is generated iid according to P, and $\tilde{Y}^n$ is generated arbitrarily, i.e., does not need to be iid or even the output of the channel W given input X^n, then*

$$\mathbb{P}(\tilde{Y}^n \in T_{[W]_\mu}(X^n)) \leq \textit{poly}(n)e^{-n(\mathbb{I}(P,W)-\epsilon)}$$

for all n sufficiently large, where $\epsilon > 0$ can be made arbitrarily small by choosing a small enough typicality parameter μ.

2.2 Frame Synchronization

Frame synchronization usually refers to the problem of locating a sync pattern in a string of data, and is treated in various ways in the literature. In this section, we review three different approaches to this problem for single-user communication, namely, the problem of locating a sync word periodically embedded into the data stream [37] and [43], the design of frame markers [49], and the one-shot frame synchronization problem [3].

In this section, the problem of interest is to acquire frame synchronization at the receiver by locating a sync pattern that is inserted into the data stream at the transmitter. For simplicity, we consider binary sequences for now. It has been common engineering practice to detect the sync pattern of length k by passing successive k-digit segments of the received sequence through a "pattern recognizer", which evaluates the similarity between the sync word and the corresponding received segment, e.g., correlation, typicality, or Hamming distance. It has been shown in [37] that this intuitive rule is not optimum in general, because it disregards the effect of the random data surrounding the sync word. However, such sequential pattern recognizer approaches may be optimal for the case of infinite block length in the context of [3].

To see what the optimum frame synchronization algorithm is, we summarize the problem formulation of [37]. It is important to note the underlying assumption in this part, namely, the sync word is periodically embedded into a data stream; the one shot approach is treated in [3] and will be discussed later. Let n denote the frame length and $k < n$ denote the sync word length, such that the binary sync word $\mathbf{s} = (s_0, ..., s_{k-1})$ is followed by $n - k$ random binary data bits $\mathbf{d} = (d_k, ..., d_{n-1})$, where the d_i's are iid random variables with $\mathbb{P}(d_i = +1) = \mathbb{P}(d_i = -1) = 1/2$. Next, let $\mathbf{sd}$ denote the concatenation of the two sequences $\mathbf{s}$ and $\mathbf{d}$, and T be the cyclic shift operator such that $T(\mathbf{sd}) = (d_{n-1}, s_0, ..., s_{k-1}, d_k, ..., d_{n-2})$.

Due to asynchronism, we assume that the sync word is a priori equally likely to begin in any of the n positions of the received signal $\mathbf{y} = (y_0, ..., y_{n-1})$. If the sync word actually begins at position ν, which is unknown to the receiver, then the received sequence is modeled as

$$\mathbf{y} = \sqrt{E}T^{\nu}(\mathbf{sd}) + \mathbf{n}, \tag{2.1}$$

where $\mathbf{n}$ is a sequence of real-valued, iid Gaussian random variables with mean zero and variance $N_0/2$. The problem of frame synchronization is to detect the actual realization of ν at the receiver given the received sequence $\mathbf{y}$ and the sync word $\mathbf{s}$. Let $\hat{\nu}$ denote the detected location. Since the location of the sync word is a priori equally likely, the optimum decision rule in the sense of maximizing the probability of correctly locating the sync word is

$$\begin{aligned}
\hat{\nu} &= \text{argmax}_{\nu \in [1:n]} \mathbb{P}(\mathbf{y}|\nu) \\
&= \text{argmax}_{\nu \in [1:n]} \sum_{\mathbf{d} \in \{-1,+1\}^{n-k}} \mathbb{P}(\mathbf{y}|\mathbf{d},\nu)\mathbb{P}(\mathbf{d}) \\
&= \text{argmax}_{\nu \in [1:n]} \sum_{\mathbf{d} \in \{-1,+1\}^{n-k}} \mathbb{P}(\mathbf{n} = \mathbf{y} - \sqrt{E} T^{\nu}(\mathbf{sd})) \qquad (2.2) \\
&= \text{argmax}_{\nu \in [1:n]} \sum_{i=0}^{k-1} s_i y_{i+\nu} - \sum_{i=0}^{k-1} f(y_{i+\nu}), \qquad (2.3)
\end{aligned}$$

where (2.2) follows from (2.1), and the details of the derivation of (2.3) from (2.2) can be found in [37]. Note that the subscripts in (2.3) are taken modulo n, and the function $f(x) := (N_0/2\sqrt{E}) \ln\cosh(2\sqrt{E}x/N_0)$.

It is important to note that the first summation in (2.3) is the ordinary correlation between the sync word and a k-digit segment of the received sequence. The second summation represents a kind of correction term required to account for the random data surrounding the sync word. In order to obtain some insight into the nature of the optimal decision rule (2.3), we consider two limiting cases of very high and very low signal-to-noise-ratios (SNRs).

If $E/N_0 \gg 1$, then the optimal decision rule can be approximated by

$$\begin{aligned}\hat{\nu} &= \text{argmax}_{\nu\in[1:n]} \sum_{i=0}^{k-1} s_i y_{i+\nu} - \sum_{i=0}^{k-1} |y_{i+\nu}| \\ &= \text{argmax}_{\nu\in[1:n]} \sum_{i=0}^{k-1} |y_{i+\nu}|(\text{sign}(s_i y_{i+\nu}) - 1),\end{aligned} \tag{2.4}$$

which means that whenever s_i and $y_{i+\nu}$ agree in sign, their contribution to the correlation term is exactly canceled out and only negatively correlated terms contribute to the optimal decision rule. Note that the decision rule (2.4) finds the location ν for the sync word that yields the least total negative correlation.

If $E/N_0 \ll 1$, then the optimal decision rule can be approximated by

$$\hat{\nu} = \text{argmax}_{\nu\in[1:n]} \sum_{i=0}^{k-1} s_i y_{i+\nu} - \frac{\sqrt{E}}{N_0} \sum_{i=0}^{k-1} y_{i+\nu}^2,$$

which indicates that the the correction term in the decision rule becomes an energy correction in the low SNR regime.

Performance analysis, i.e., characterization of the probability of a sync error under the optimal decision rule (2.3), depends on the sync word $\mathbf{s}$. Therefore, it is important to find the properties of a "good" sync word, which is called "marker design" in [49].

We now focus on the probability of a sync error, denoted here by $P_s := \mathbb{P}(\hat{\nu} \neq \nu)$, in the high SNR regime in which we approximate the sequence being received without error [43]. Note that even if the channel is error-free, it is still possible to have a sync error, which is the probability that the sync word appears elsewhere than its intended location. In this regime, it is easy to see that it is desirable to choose a sync word with good autocorrelation properties, and in particular, we should ensure that no overlap between a sync word and random data bits may be equal to the sync word itself. Mathematically,

$$(s_0, s_1, ..., s_{l-1}) \neq (s_{k-l}, s_{k-l+1}, ..., s_{k-1}), \quad l = 1, 2, ..., k-1. \tag{2.5}$$

It is shown in [43] that in the high SNR regime if the sync word has the property (2.5), then

$$P_s = \sum_{j=1}^{r} \frac{(-1)^{j+1}}{j+1} \binom{n-k-(k-1)j}{j} 2^{-kj}, \tag{2.6}$$

where $r := \lfloor (n-k)/k \rfloor$.

The property (2.5) of a good sync word is generalized in [49] for a noisy channel, and can be qualitatively expressed as follows: The probability of a random data sequence of length k looking like the sync word should be larger than the probability of a sequence of length k containing both random data and a part of the sync word looking like the sync word.

Next, we turn our attention to the one shot approach for frame synchronization [3]. The receiver's aim is still to locate a sync word $\mathbf{s} = (s_0, ..., s_{k-1})$ of length k. We consider a general discrete memoryless channel (DMC) with probability transition matrix W and input and output alphabets $\mathcal{X}$ and $\mathcal{Y}$, respectively. Also, let $\star \in \mathcal{X}$ denote the data symbol that is transmitted before and after the sync word, and let $W_\star$ denote the distribution of the output of the channel if the input is data, and more generally, let W_x denote the distribution of the output of the channel if the input is $x \in \mathcal{X}$.

The transmission of the sync word starts at a random time ν, uniformly distributed in $[1 : A = e^{\alpha k}]$, where the integer A and the exponent α are called the *asynchronism level* and the *asynchronism exponent*, respectively. The received sequence $\mathbf{y} = (y_1, ..., y_n)$ has length $n = A + k - 1$ and is distributed as

$$\mathbb{P}(\mathbf{y}) = \prod_{i=1}^{\nu-1} W_\star(y_i) \prod_{i=\nu}^{\nu+k-1} W(y_i|s_{i-\nu}) \prod_{i=\nu+k}^{n} W_\star(y_i).$$

The receiver observes the output sequence $\mathbf{y}$ and detects the location where the sync word starts, denoted by $\hat{\nu}$. The probability of sync error is defined as before, i.e., $P_s = \mathbb{P}(\hat{\nu} \neq \nu)$. An asynchronism exponent α is called *achievable* if there exists a sequence of sync words and decoders such that the probability of sync error approaches zero under asynchronism level $A = e^{\alpha k}$ as the length of the sync word, $k \to \infty$. The *asynchronous threshold*, denoted by $\alpha(W)$, is the supremum of the set of achievable asynchronism exponents. Note that in the system model above, the scaling behavior of the received window n with respect to the sync word's length k is assumed to be exponential, and the performance of the system is analyzed in terms of the asynchronous threshold in regime of infinitely large k and n. However, it is worth pointing out that the results in [37, 43, 49] are valid for any value of k and n as long as $k \leq n$.

The main result of [3] is that the asynchronous threshold as defined above is given by

$$\alpha(W) = \max_{x \in \mathcal{X}} D(W_x \| W_\star), \tag{2.7}$$

where $D(\cdot\|\cdot)$ is the Kullback-Leibler divergence.

One scheme that achieves the asynchronism exponent, and is therefore optimal in the problem formulation above, is the sequential detector that looks for the first sequence of length k in the received sequence $\mathbf{y}$ that is jointly typical with the sync word $\mathbf{s}$. Furthermore, it is sufficient to use a sync word $\mathbf{s}$ that is mainly composed of $\bar{x}$'s, where $\bar{x}$ is input symbol that achieves the maximum in (2.7), but with a few $\star$'s mixed in so that shifts of the sync word look sufficiently different from the original sync word. This structure is consistent with the properties of a good sync word described in [43] and [49].

The problem considered so far is about acquiring synchronization in a communication system without emphasizing the encoding of data/messages. The problem of joint or successive synchronization and decoding in a single-user communication system with similar problem settings to those in [3] is studied in Section 2.3. Furthermore, the problem of joint synchronization and decoding in multi-user communication systems with a different model for asynchronism than [3] is studied in Section 2.5.

2.3 Asynchronous Communication

Asynchronous communication arises in practical communication scenarios in which the fundamental performance of the system and the problem of acquiring synchronization might vary depending upon the type of asynchronism and the system model.

One particular information theoretic model for asynchronous communication is developed in [4, 47, 52, 54] with a single block transmission that starts at a random time unknown to the receiver, within an exponentially large window known to the receiver, which is similar to the one shot frame synchronization problem reviewed in Section 2.2. In this model, the transmission is contiguous; once it begins, the whole codeword is transmitted, and the receiver observes only noise both before and after transmission.

A discrete memoryless channel with finite input and output alphabets $\mathcal{X}$ and $\mathcal{Y}$, respectively, and transition probability matrix W is considered. The transmitter wants to communicate a message $m \in [1:M]$ to a receiver through the channel, where $M \geq 2$. For each message m, there is an associated codeword

$$c^k(m) := c_1(m)c_2(m)...c_k(m),$$

which is the output of the encoder and is a sequence of k symbols drawn from $\mathcal{X}$.

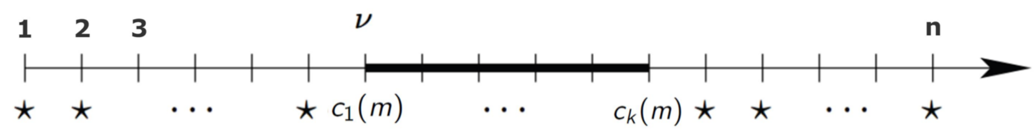

Figure 2.1. Representation of the channel input sequence in asynchronous communication model (obtained from [54] with modification).

The transmitter starts sending the codeword $c^k(m)$ at a random time ν, unknown to the receiver, independent of $c^k(m)$, and uniformly distributed over $[1 : A = e^{\alpha k}]$, and $n = A+k-1$ is the length of the receive window, and α is the asynchronism exponent as before. During the interval that a codeword is transmitted, the distribution of the output Y_t is $W(\cdot|c_{t-\nu+1}(m))$, $t \in \{\nu, \nu+1, ..., \nu+k-1\}$.

Before and after transmitting the codeword, the transmitter is silent, and the receiver observes only noise, containing no information about the message m. In order to characterize the output of the channel for channel uses when the transmitter is silent, we fix a silent / noise symbol denoted by $\star \in \mathcal{X}$, so that $W_\star(\cdot) := W(\cdot|x = \star)$ characterizes the noise distribution of the channel. Figure 2.1, obtained from [54] with modification, illustrates the channel input sequence.

The decoder observes the output sequence y^n sequentially, and decodes the message denoted by $\hat{m}$. Stopping time $k \leq \tau \leq n$ with respect to the output sequence indicates when decoding occurs. For convenience, we define $\hat{\nu} := \tau - k + 1$ as decoders estimate of ν.

There are different definitions for probability of error and communication rate for this problem. In [52], the probability of error is defined as the average (over messages and transmission time ν) of the decoding error probability, i.e., the probability that the decoded message does not correspond to the sent message. In [54], the probability of error is defined as the maximum (over messages), time-averaged decoding error probability. In both [52, 54], the communication rate is defined as the logarithm of the codebook size divided by the average elapsed time between the time the codeword starts being sent and the time the decoder makes a decision, i.e., a delay-compensated definition of rate. However, with this definition for communication rate, the problem of finding the capacity of asynchronous communication becomes more difficult. In fact, only inner and outer bounds for the capacity of asynchronous communication with the delay-compensated definition of rate are given in [52, 54].

In [47], the probability of error is defined as $\mathbb{P}[\hat{m} \neq m \text{ or } \hat{\nu} \geq \nu]$, which is greater or equal to the average probability of error $\mathbb{P}[\hat{m} \neq m]$ and therefore imposes a stronger constraint on the definition of achievable rates. The communication rate in [4, 47] is defined as the logarithm of the codebook size divided by the codeword length, i.e., $R := \log M/k$. With these definitions, the delay constraint is detached from the rate definition by considering the stronger condition for the probability of error. In stating the results on the capacity of asynchronous communication, we focus on the later definitions for probability of error and communication rate.

A pair (R, α) is called *achievable* if for an asynchronous communication system described above with a receive window of size $n = A + k - 1$, where $A = e^{\alpha k}$, there exists a sequence of length k codes of size e^{kR} such that the probability of error vanishes as $k \to \infty$. The *asynchronous capacity with asynchronism exponent* α is defined as

$$C(\alpha) := \sup\{R : (R, \alpha) \text{ is achievable}\}.$$

If $R = 0$, then we assume that $M = 2$, and the problem reduces to a reliable communication with only two possible messages. The *asynchronous threshold* is defined as

$$\alpha(W) := \sup\{\alpha : (0, \alpha) \text{ is achievable}\},$$

which is equivalent to the definition given in Section 2.2 for the asynchronous threshold.

Theorem 2.1 recalls the main result on the synchronization threshold and asynchronous capacity.

Theorem 2.1. *([52], [4], [47]): For any DMC W, we have*

$$\alpha(W) = \max_{x \in \mathcal{X}} D(W_x \| W_\star). \tag{2.8}$$

The asynchronous capacity of the DMC W is

$$C(\alpha) = \max_{P : D(PW \| W_\star) \geq \alpha} \mathbb{I}(P, W), \tag{2.9}$$

where the maximum is defined to be zero if $\alpha \geq \alpha(W)$.

To interpret these results, note that in (2.8), $D(W_x \| W_\star)$ is the negative exponent of the probability that noise outputs are misinterpreted as codeword outputs, and α gives the positive exponent of the number of starting times for the transmission. This result basically states that the divergence has to be larger than α for the probability of synchronization error to go to zero, and maximizing over x gives the largest divergence, and therefore, allows for the largest α. The result of (2.9) essentially allows an exponential number of messages, but constrains the choice of input distributions to ensure that the synchronization error does not dominate.

The performance of training-based schemes, which handle the synchronization separately from the information transmission by using a preamble as an identifier, are studied in [54]. It is shown that such schemes are suboptimal and cannot achieve the asynchronous capacity in general, and the penalty is substantial in the high-rate regime. The authors in [4] focus on the capacity of asynchronous communication per unit cost [58], or equivalently the minimum cost to transmit one bit of information asynchronously.

The finite blocklength regime for asynchronous communication is investigated in [47] and it is shown that the channel dispersion does not increase due to asynchronism. In a recent work [53], the authors study the capacity per unit cost for asynchronous communication if the receiver is constrained to sample only a fraction of the channel outputs, and they show that the capacity remains unchanged under this constraint.

A slotted asynchronous channel model is investigated in [60]. For the entire communication period of length k, the transmitter is either silent, which is modeled by transmitting $\star^k$, or transmits a codeword $\mathbf{x}_m$ of length k from a given codebook of size M. The task of the receiver is to determine whether or not transmission has taken place, and if so, to decode the message. The fundamental limits of asynchronous communication in terms of misdetection, false alarm, and decoding error exponents are characterized and partially solved in [60] and [59] for the case in which at least one of the exponents vanishes; the optimum detection/decoding rule in the sense of the best trade-off among the probabilities of misdetection, false alarm, and decoding error is derived in [39] for the general case of non-vanishing exponents. Denoting the channel output vector of length k by $\mathbf{y}$, then according to this rule, a transmission is detected if and only if

$$e^{k\beta_1}\sum_{m=1}^{M} W(\mathbf{y}|\mathbf{x}_m) + \max_{1\leq m\leq M} W(\mathbf{y}|\mathbf{x}_m) \geq e^{k\beta_2} W(\mathbf{y}|\star^k), \tag{2.10}$$

where β_1 and β_2 are chosen to meet the misdetection and false alarm constraints. If the received $\mathbf{y}$ passes the test (2.10), then decoder outputs the maximum likelihood (ML) estimate. As a final note, [61] studies the same problem in the context of single-message unequal error protection (UEP).

With this context, it is important to compare and contrast asynchronous communication with intermittent communication studied in this dissertation. Although we delay the precise development of the system model for intermittent communica-

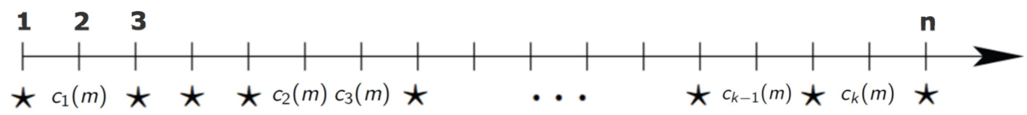

Figure 2.2. Representation of the channel input sequence in intermittent communication model.

tion until Chapter 4, it is adequate at this point to demonstrate the system model graphically. Figure 2.2 characterizes the channel input sequence in intermittent communication. As opposed to asynchronous communication in which the transmission of codeword symbols is contiguous as illustrated in Figure 2.1, the transmission of codeword symbols in intermittent communication can be bursty, i.e., the transmitter becomes silent at random times and the receiver observes only noise, while at the other times codeword symbols are transmitted non-contiguously.

2.4 Insertion / Deletion Channels

Non-contiguous transmission of codeword symbols, as illustrated in Figure 2.2, is reminiscent of insertion channels. In fact, if the intermittent process is considered as part of the channel behavior, then intermittent communication can be described by the following insertion channel: after the i^{th} symbol of the codeword, N_i noise symbols are inserted, where N_i, $i = 1, 2, ..., k$ are random variables, possibly iid. The resulting sequence passes through a discrete memoryless channel, and the receiver should decode the message based on the output of the channel without knowing the positions of the codeword symbols. To the best of our knowledge, this specific insertion channel model has not been studied previously. However, some of our techniques, especially for providing upper bounds for the capacity of intermittent communication, are similar to those of insertion / deletion channels in the literature.

A more general class of channels with synchronization errors is studied in [13], in which every transmitted symbol is independently replaced with a random number of symbols, possibly including the empty string to model a deletion event, and the transmitter and receiver do not know a priori the positions or the pattern of the insertions / deletions. Dobrushin [13] proved the following characterization of the capacity for such iid synchronization error channels.

Theorem 2.2. *([13]): For iid synchronization error channels, let* $X^k := (X_1, X_2, ..., X_k)$ *denote the channel input sequence of length* k*, and* $Y^N := (Y_1, Y_2, ..., Y_N)$ *denote the corresponding output sequence at the decoder, where the output length* N *is a random variable determined by the channel realization. The channel capacity is*

$$C = \lim_{k \to \infty} \max_{P_{X^k}} \frac{1}{k} \mathbb{I}(X^k; Y^N). \tag{2.11}$$

Theorem 2.2 demonstrates that iid synchronization error channels are information stable [44]. However, there are several difficulties related to computing the capacity through this characterization. First, it is challenging to compute the mutual information because of the memory inherent in the joint distribution of the input and output sequences. Second, the optimization over all the input distributions is computationally involved. A single-letter characterization for the capacity of the general class of synchronization error channels is still an open problem, even though there are many papers deriving bounds on the capacity of the insertion / deletion channels [11, 12, 14, 18–20, 26, 31, 41, 42, 55].

We first focus on lower bounds for the capacity of the insertion / deletion channels. Gallager [20] considers a channel model with substitution and insertion / deletion errors and derives a lower bound on the channel capacity as is summarized in the following theorem.

Theorem 2.3. *([20]): Consider a substitution / insertion / deletion binary-input* *binary-output channel in which every bit gets deleted with probability p_d, replaced with two random bits with probability p_i, correctly received with probability $p_c = (1 - p_d - p_i)(1 - p_s)$, and inverted with probability $p_f = (1 - p_d - p_i)p_s$. The capacity of this channel is lower bounded by*

$$C \geq 1 + p_d \log p_d + p_i \log p_i + p_c \log p_c + p_f \log p_f.$$

In [11], codebooks from first-order Markov chains are used to improve the achievability results for deletion channels. The intuition is that it is helpful to put some memory in the codewords if the channel has some inherent memory. Therefore, they consider codewords generated randomly by a first order Markov process, which yields codewords consisting of blocks of alternating 0's and 1's with the lengths geometrically distributed. Most of the subsequent results on lower bounding the capacity of deletion and duplication channels rely on this specific codebook generation. For example [31] directly lower bounds the information capacity given by (2.11) for channels with iid deletions and duplications with an input distribution following a symmetric, first-order Markov chain.

However, the specific case of the insertion channel we described at the beginning of this section is different from the duplication channels, and the same techniques cannot directly be used in order to derive an achievability result.

The first difference is that, in our model, a specific symbol is considered as an insertion symbol, and it is not necessarily a duplication. Intuitively, this makes the decoders task more difficult, because the inserted symbol has no information about the transmitted message and introduces uncertainty about the positions of codeword symbols. This observation suggests that for our insertion channel model, it is a better idea to detect and remove the inserted symbols than treating all the output symbols as information symbols. In Chapter 4, we introduce a decoding structure which is based on this intuitive idea.

The second difference in our insertion channel model is that runs of the input sequence are subject to significant changes once they pass through the channel: In the binary case, if the specific insertion symbol is 0, a block of l consecutive 1's can be split into as many as $2l$ blocks. This will make it difficult to directly bound (2.11) if we would like to generate codewords with iid run lengths (first order Markov chain codeword generation). For comparison, consider sticky channels [41] in which each symbol may be duplicated many times. A convenient feature of a binary sticky channel is that runs at the transmitter correspond to the runs at the receiver, i.e., the length of the runs might change, but no run is deleted and no new run is created. Therefore, the mutual information in (2.11) becomes easier to evaluate and bound.

The last difference is that in our insertion channel model, we consider arbitrary alphabet size for the input and output of the channel, while the insertion channels in the literature mostly focus on the binary alphabets. This difference provides another reason to not focus on runs of the input and output sequences, because considering the corresponding runs in the output sequence becomes computationally involved for non-binary channels. In fact, for our insertion channel model, a simple codebook design combined with detecting the patterns in the output sequence leads to a simpler analysis that gives more general theoretical results.

Next, we consider upper bounds on the capacity of insertion / deletion channels. In fact, only a few upper bounds have been derived on the capacity of iid insertion / deletion channels [12, 18, 19], which numerically evaluate the capacity of an auxiliary channel obtained by a genie-aided decoder (and encoder) with access to side-information about the insertion / deletion process. Although our insertion channel model is different, we can apply some of these ideas and techniques to upper bound the capacity of intermittent communication, as we will see in Chapter 4.

Finally, bounds on the capacity per unit cost for a noisy channel with synchronization errors are given in a recent work [24], in which an encoding / decoding scheme with only loose transmitter-receiver synchronization is proposed. We use the same approach for obtaining a lower bound on the capacity per unit cost of intermittent communication in Chapter 4.

2.5 Networking and Information Theory for Multi-Access Communication

Multi-access communication is treated in various ways in the literature. Gallager [22] reviews both information-theoretic and network-oriented approaches, and emphasizes the need for a perspective that can merge elements from these two approaches. As also pointed out in [16], information-theoretic models focus on accurate analysis of the effect of the noise and interference, whereas network-oriented models focus on bursty transmissions and collision-resolution approaches.

In Chapter 5, we introduce a model for intermittent multi-access communication with two users that can be viewed as an attempt to combine the information-theoretic and network-oriented multi-access models and to characterize the performance of the system in terms of achievable rate regions. The model captures two network-oriented concepts. First, it models bursty transmission of the codeword symbols for each user. Second, it takes into account the possible asynchronism between the receiver and the transmitters as well as between the transmitters themselves. In this section, we summarize the system model and results of some of the works that give an information-theoretic model for multi-access communication with some focus on the network-oriented concepts. Specifically, we review models and results for asynchronous multi-access communication, random access, and the collision channel. As we will see in Chapter 5, some of the assumptions and / or conclusions are similar to those of intermittent multi-access communication.

2.5.1 Asynchronous Multi-Access Communication

Asynchronism in multi-user communication usually refers to the inability of the users to start the transmission at the same time, with the receiver is unaware of this time difference. We can ask, for example, how the lack synchronization affects the capacity region of the channel. In this context, two types of asynchronism have been studied in the literature. One type is frame asynchronism in which the two users are not able to start the transmission of their codewords in unison [8, 17, 25, 45, 57]. The other type is symbol asynchronism in which each codeword symbol modulates a fixed assigned waveform that are not perfectly aligned in time [56].

We briefly review the system models and results concerning frame asynchronous multiple-access channel (MAC). A good review of the results and a unified model for the two-user frame asynchronous memoryless MAC introduced in [8, 25, 45] is given in [23, Section 24.3], while [17] treats frame asynchronism when there are more than two users and [57] treats frame asynchronism if the MAC has memory.

Consider a two-user discrete memoryless multiple-access channel (DM-MAC) with conditional probability mass functions $W(y|x_1, x_2)$ over input alphabets $\mathcal{X}_1$ and $\mathcal{X}_2$ and output alphabet $\mathcal{Y}$. We focus on the frame-asynchronous case in which the system is symbol-synchronous. The system model is depicted in Figure 2.3. Suppose that user 1 and user 2 wish to communicate iid message sequences $\{M_{1l}\}_{l=1}^{\infty}$ and $\{M_{2l}\}_{l=1}^{\infty}$, respectively. These two sequences are independent, and each message pair $(M_{1l}, M_{2l}), \quad l = 1, 2, ...,$ is uniformly distributed over $[1 : 2^{nR_1}] \times [1 : 2^{nR_2}]$. Encoder $j, j = 1, 2$ assigns a sequence of codewords $x_j^n(m_{jl})$ of length n to each message sequence $m_{jl} \in [1 : 2^{nR_j}]$ for $l = 1, 2, ...$

Also, assume that symbols are synchronized, but that the blocks sent by the two encoders incur arbitrary delays $d_1, d_2 \in [0 : d]$, respectively for some $d \leq n - 1$, and that the encoders and the decoder do not know the delays a priori. The received

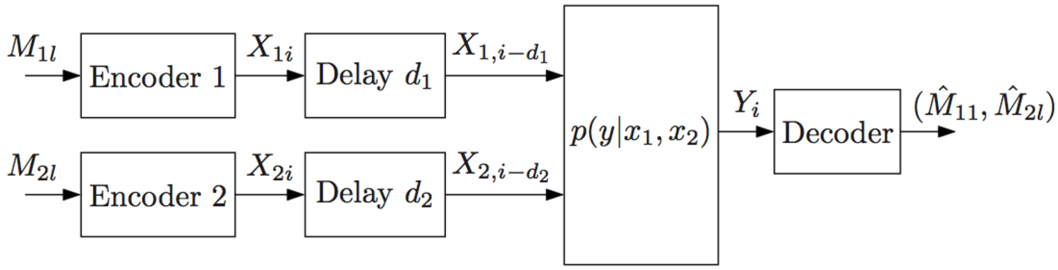

Figure 2.3. Frame asynchronous multiple access communication system [23].

sequence Y^n is distributed according to

$$\mathbb{P}(y^n | x^n_{1,1-d_1}, x^n_{2,1-d_2}) = \prod_{i=1}^{n} p_{Y|X_1,X_2}(y_i | x_{1,i-d_1}, x_{2,i-d_2}).$$

The decoder assigns a sequence of message pairs $(\hat{m}_{1l}, \hat{m}_{2l}) \in [1:2^{nR_1}] \times [1:2^{nR_2}]$ or an error message e to each received sequence $y^{ln+d}_{(l-1)n+1}$ for each $l = 1, 2, ...$ Note that the received sequence $y^{ln+d}_{(l-1)n+1}$ can include parts of the previous and the next blocks. The average probability of error for this model is defined as

$$P_e^{(n)} = \max_{d_1, d_2 \in [0:d]} \sup_l \mathbb{P}((\hat{M}_{1l}, \hat{M}_{2l}) \neq (M_{1l}, M_{2l}) | d_1, d_2).$$

Achievability and the capacity regions are defined as for the synchronous DM-MAC. Two types of asynchrony are considered. *Mild asynchrony* refers to the case in which d/n tends to zero as $n \to \infty$. It is shown in [8] that the capacity region of any DM-MAC under mild asynchrony is the same as for the synchronous case: the closure of the set of rate pairs (R_1, R_2) such that

$$\begin{aligned} R_1 &< \mathbb{I}(X_1; Y|X_2, Q), \\ R_2 &< \mathbb{I}(X_2; Y|X_1, Q), \\ R_1 + R_2 &< \mathbb{I}(X_1, X_2; Y|Q) \end{aligned}$$

for some distribution $p(q)p(x_1|q)p(x_2|q)$, with the cardinality of Q bounded as $|\mathcal{Q}| \leq 2$. In the region above, Q is called the time sharing random variable, which makes the capacity region convex.

Total asynchrony refers to the case in which $d = n - 1$, i.e., d_1 and d_2 can vary from 0 to $n-1$. It is shown in [25] and [45] that the capacity region of any DM-MAC under total asynchrony is the closure of the set of rate pairs (R_1, R_2) such that

$$\begin{aligned} R_1 &< \mathbb{I}(X_1; Y|X_2), \\ R_2 &< \mathbb{I}(X_2; Y|X_1), \\ R_1 + R_2 &< \mathbb{I}(X_1, X_2; Y) \end{aligned}$$

for some input distribution $p(x_1)p(x_2)$. It is important to note that this region is not convex in general, and unlike the synchronous case or mild asynchronous case, the capacity region of DM-MAC with total asynchrony is not necessarily convex. Furthermore, the sum-capacity of the totally asynchronous DM-MAC is the same as that of the synchronous DM-MAC, i.e., if the channel is memoryless, we do not need time sharing in order to achieve the sum-capacity $C_{sum} = \max_{p(x_1)p(x_2)} \mathbb{I}(X_1, X_2; Y)$. However, this does not remain true for channels with memory, as we will discuss shortly.

For asynchrony in between the mild and the total asynchrony, i.e., when $0 < d/n < 1$ as $n \to \infty$, the capacity region of the asynchronous MAC is an open problem [23].

Next, we study the effect of frame asynchronism on the capacity region of MACs with memory. It is shown in [57] that if the channel has memory, the lack of frame synchronism does not allow for non-stationary inputs to achieve any point in the capacity region, and cooperation in the time domain is not possible. This restriction, unlike in the case of DM-MAC which only results in the removal of the convex hull operation from the expression of the capacity region, has more destructive effects on the capacity region of some MACs with memory. In particular, the sum-capacity of some MACs with memory could be drastically reduced as a result of total frame asynchronism, which is not the case in DM-MACs.

As expected, the capacity region of the MAC with memory stated in [57] does not admit single-letter characterizations. Instead, it is given in terms of a limit of regions. One example of the MAC with memory for which the limits are computable, and the frame-synchronous and frame-asynchronous capacity rate regions can be explicitly characterized, is given in [57] and is summarized below.

Consider a MAC with memory with $\mathcal{X}_1 = \mathcal{X}_2 = \{0, 1, 2\}$, $\mathcal{Y} = \{1, 2\}$, and

$$y_i = \begin{cases} x_{1,i}, & \text{if } x_{1,i} \neq 0 \text{ and } x_{2,i} = 0 \text{ and } x_{2,i-1} \neq 0 \\ x_{2,i}, & \text{if } x_{2,i} \neq 0 \text{ and } x_{1,i} = 0 \text{ and } x_{1,i-1} \neq 0 \\ (1/2, 1/2), & \text{otherwise} \end{cases} \tag{2.12}$$

where $(1/2, 1/2)$ indicates that y_i is equally likely to be 1 or 2. Note that simultaneous zeros or non-zeros for both users and consecutive zeros for one user result in equally likely outputs. In this channel, it is necessary for the encoders to use some sort of time sharing to achieve optimum rates. Figure 2.4 shows achievable rate regions with and without frame synchronism of the MAC with memory characterized by (2.12). The region for the synchronous case is an achievable region, whereas the region for

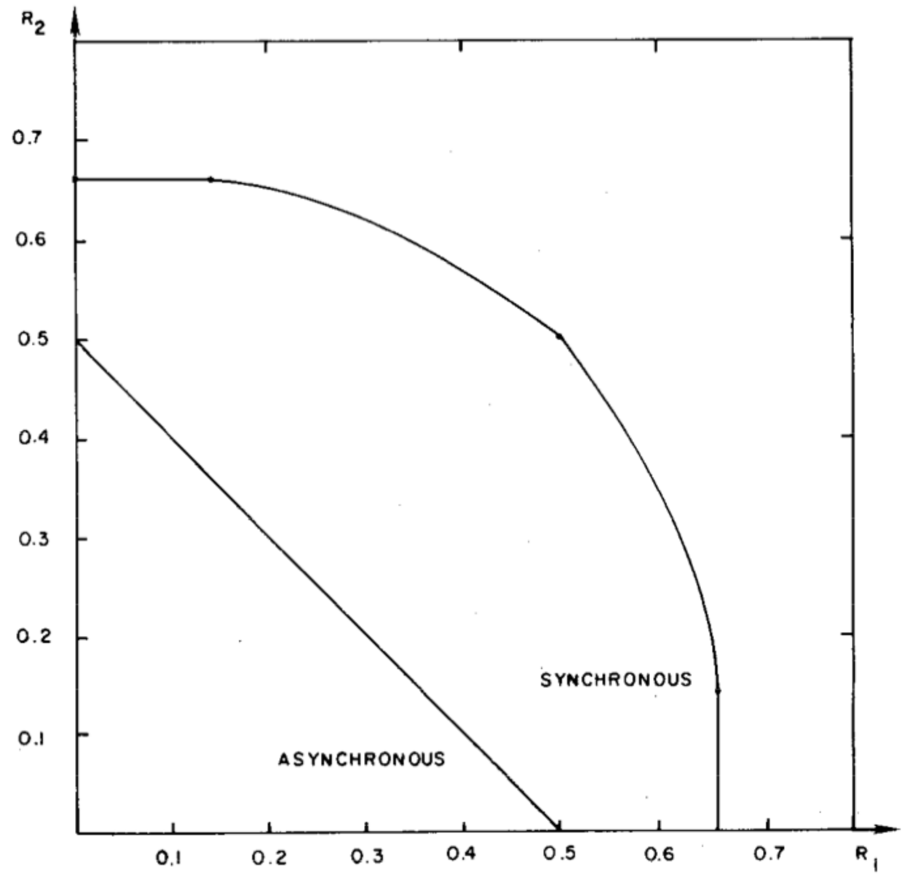

Figure 2.4. Achievable rate regions with and without frame synchronism of an example MAC with memory [57].

the asynchronous case is shown to be the capacity region in [57].

It can be seen that frame asynchronism drastically reduces the capacity region of this example, which is because non-stationary inputs are necessary to achieve the optimum rates, and this is not possible when the system is frame asynchronous. Interestingly, even under mild frame asynchronism, we cannot do better than the inner region of Figure 2.4 for this example, which is in contrast to the memoryless channels discussed earlier. The reason is that mild frame asynchronism allows for cooperation in the large time-scale, but not in the small time-scale, which may be necessary for the encoders if the channels has memory [57].

2.5.2 Random Access

Another model for multi-access communication capturing elements from information-theoretic and network-oriented approaches is introduced in a recent work [40]. Specifically, this work presents an information-theoretic model for a random access communication scenario with two modes of operation for each user, active or inactive. The set of active users is available to the decoder only, and active users encode their messages into two streams: one high priority stream ensures that part of the transmitted information is decoded reliably even in the presence of interference from other users, and one low priority stream opportunistically takes advantage of the channel when other users do not transmit. Therefore, some part of the information can be decoded only in the absence of interference.

A different information-theoretic model for random access communication is introduced in [34] and [33] in which a new channel coding approach is presented for coding within each data packet with built-in support for message underflow and packet collision detection. The key feature of this coding approach is that it does not require joint communication rate determination either among the transmitters or between the transmitters and the receiver. An achievable region is defined such that reliable message recovery is supported for all the rates inside the region and reliable collision detection is supported for all the rates outside the region. The main results of [34] and [33] is that for a DM-MAC, the achievable rate region of the coding scheme equals the Shannon information rate region without a convex hull operation. This removal of the convex hull operation was seen for the capacity region of the totally frame asynchronous DM-MAC we reviewed earlier.

2.5.3 The Collision Channel without Feedback

Finally, we briefly review the system model and the results on the collision channel without feedback [38] in which users share a common communication resource with unknown time offsets among their clocks. We focus on the slot-synchronized case in which a discrete-time communication is considered with the time offsets of the K users being arbitrary integers $\delta_1, \delta_2, ..., \delta_K$. All the time offsets are unknown to all users, and can never be learned as the users receive no feedback from the channel, and are also a priori unknown to the receiver. If two users transmit a packet at the same time slot, it is considered as a "collision" and cannot be decoded; otherwise, the packets are received error-free in the absence of a collision.

It is assumed that each encoder has a periodic protocol sequence denoted by $\mathbf{s}_i$ with period N_i, $i = 1, 2, ..., K$, and the users may jointly choose their protocol sequences and their choice is known by the receiver. Assuming that N is the least common multiple of N_1, N_2,..., and N_K, we would like to design the protocol sequence $\mathbf{s}_i = (s_{i1}, ..., s_{iN})$. Denoting the transmitted signal of the i'th user at time n by $x_i(n)$ for $i = 1, 2, ..., K$ and $n = 1, 2, ..., N$, and an idle signal in a slot by $\star$, we have $x_i(n) = \star$ if $s_{in} = 0$, and otherwise, it contains a packet of information. The output of the channel is

$$y(n) = \begin{cases} \star, & \text{if } x_i(n-\delta_i) = \star \text{ for all } 1 \leq i \leq K \\ x_i(n-\delta_i), & \text{if } x_j(n-\delta_j) = \star \text{ for all } j \neq i \\ \Delta, & \text{otherwise ,} \end{cases}$$

where Δ denotes a collision of two or more packets. The encoding and decoding are defined across all N packets. It is obvious that the randomness of the channel comes from the time offsets $\delta_1, \delta_2, ..., \delta_K$, which are assumed to be iid and uniform on $[1 : N]$. In [38], the capacity region and the zero-error capacity region are obtained and are shown to be equal.

For the case of two users, $K = 2$, the outer boundary of the capacity region is the rates that satisfy the equation $\sqrt{R_1} + \sqrt{R_2} = 1$, and is illustrated in Figure 2.5.

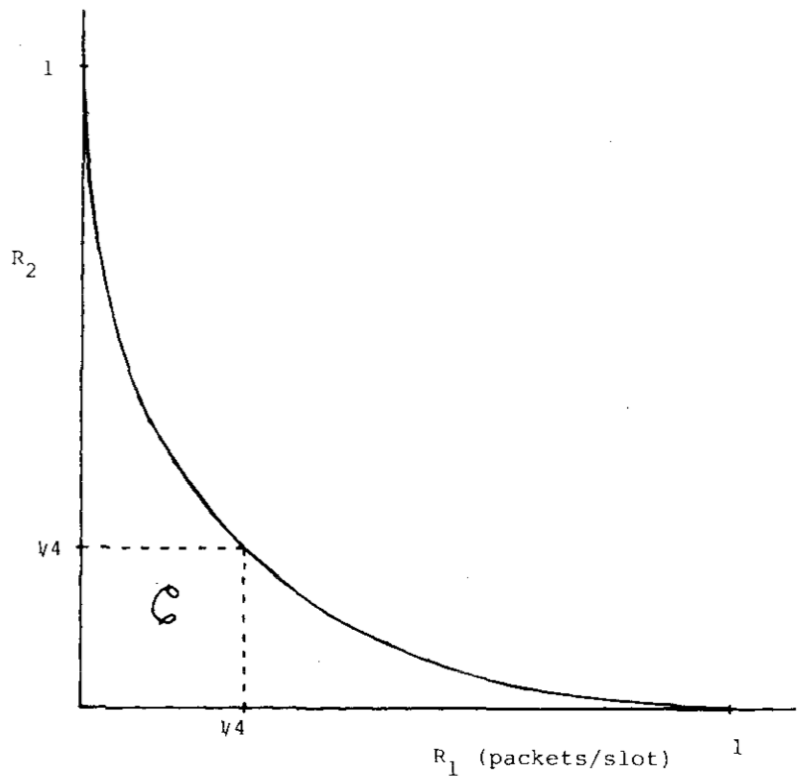

Figure 2.5. Capacity region of two-user collision channel without feedback [38].

Clearly, this capacity rate region is not convex. This is because the users cannot time share due to the unknown time offsets. As we have seen earlier in this section, the capacity regions of neither the total frame asynchronous MAC [25] nor the random access [34] is generally convex.

In Chapter 5, after introducing a model for intermittent multi-access communication and obtaining the results, we compare various assumptions and conclusions with those of the related works reviewed in this section.

2.6 Summary

In this chapter, we summarized the system model and main results of some related work, identified the gaps in the literature, and explained how the models and results in the remaining chapters close the gaps to some extent. First, we briefly reviewed the method of types since we make frequent use of this tool, and specifically, the notations in our analysis. Second, we reviewed the literature on frame synchronization in single-user communication systems, which studies the problem of locating a sync pattern in a string of data. Third, we studied the problem of joint frame synchronization and decoding in asynchronous communication, which corresponds to contiguous transmission of codeword symbols in which the receiver observes noise before and after transmission. Next, we summarized the results on the lower and upper bounds of the capacity of insertion / deletion channels. Finally, we discussed on several networking issues, such as asynchronism, random access, and collisions in the context of multiple-access communication with an information-theoretic approach. A key assumption of these communication models is the lack of knowledge of the state of the channel and / or a timing reference at the transmitter and receiver, capturing certain kinds of asynchronism. Generally, this asynchronism makes the task of the receiver more difficult since it must acquire synchronization in the first place. Intermittent communication is just another attempt to model asynchronism at the symbol or packet level.

CHAPTER 3

PARTIAL DIVERGENCE

We will see in Chapter 4 that that a relevant function is $d_\rho(P\|Q)$, which we call *partial divergence* and view as a generalization of the Kullback-Leibler divergence. Qualitatively speaking, partial divergence is the normalized 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 characterizing a decoder's ability to distinguish a sequence obtained partially from the codewords and partially from noise from a codeword sequence or a noise sequence.

In this chapter, we present a lemma that generalizes some of the results of the method of types reviewed in Section 2.1, and then specialize the lemma to a certain case in which the partial divergence is defined. Next, we study some of the properties of the partial divergence, which are used in Chapter 4 to provide insights about the achievable rates in Section 4.2.3. Finally, we generalize the lemma and the partial divergence to the case of three distributions, which is applied in Chapter 5.

3.1 A Lemma and Specializing to Partial Divergence

In this section, we first present a lemma, which generalizes some of the results of method of types reviewed in Section 2.1, and then specialize the lemma to a certain case in which the partial divergence will be defined.

Lemma 3.1. *Consider the distributions $P, Q, Q' \in \mathcal{P}^{\mathcal{X}}$ on a finite alphabet $\mathcal{X}$. A random sequence X^k is generated as follows: ρk symbols iid according to Q and $\bar{\rho}k$*

symbols iid according to Q', where $0 \leq \rho \leq 1$. The normalized exponent of the probability that X^k has type P is

$$
\begin{aligned}
d(P, Q, Q', \rho) &:= \lim_{k \to \infty} -\frac{1}{k} \log \mathbb{P}(X^k \in T_P) \\
&= \min_{\substack{P_1, P_2 \in \mathcal{P}^{\mathcal{X}}: \\ \rho P_1 + \bar{\rho} P_2 = P}} \rho D(P_1 \| Q) + \bar{\rho} D(P_2 \| Q'). \qquad (3.1)
\end{aligned}
$$

The proof follows from the proof of Lemma 3.2 in Section 3.3. A different characterization and proof of Lemma 3.1 can be found in [27, 28].

Specializing Lemma 3.1 for $Q' = Q$ results in Lemma 2.1, and we have $d(P, Q, Q, \rho) = D(P\|Q)$. However, we will be interested in the special case of Lemma 3.1 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 with a fraction ρ according to Q and the remaining fraction $\bar{\rho}$ according to P. 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 3.1 gives an explicit expression for the partial divergence by solving the optimization problem in (3.1) for the special case of $Q' = P$.

Proposition 3.1. *If $\mathcal{X} = [1 : |\mathcal{X}|]$ and $P, Q \in \mathcal{P}^{\mathcal{X}}$, where $P := (p_1, p_2, ..., p_{|\mathcal{X}|})$ and $Q := (q_1, q_2, ..., q_{|\mathcal{X}|})$ and we assume that all values of the PMF Q are nonzero, then the partial divergence can be written as*

$$
d_\rho(P\|Q) = D(P\|Q) - \sum_{j=1}^{|\mathcal{X}|} p_j \log(c^* + \frac{p_j}{q_j}) + \rho \log c^* + h(\rho), \qquad (3.2)
$$

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

$$c^* \sum_{j=1}^{|\mathcal{X}|} \frac{p_j q_j}{c^* q_j + p_j} = \rho. \tag{3.3}$$

Proof. It can readily be verified that the function $d(P, Q, Q', \rho)$ in (3.1) can be written as [28]

$$d(P, Q, Q', \rho) = H(P) + D(P\|Q) - \bar{\rho} \log \frac{q'_{|\mathcal{X}|}}{q_{|\mathcal{X}|}} - e(P, Q, Q', \rho), \tag{3.4}$$

where

$$e(P, Q, Q', \rho) := \max_{0 \le \theta_j \le 1, j=1,2,...,|\mathcal{X}|-1} \{\rho H(P_1) + \bar{\rho} H(P_2) + \sum_{j=1}^{|\mathcal{X}|-1} \theta_j p_j \log a_j\}, \tag{3.5}$$

$$a_j := \frac{q'_j q_{|\mathcal{X}|}}{q_j q'_{|\mathcal{X}|}}, \quad j = 1, 2, ..., |\mathcal{X}| - 1, \tag{3.6}$$

$$P_1 := (\frac{\bar{\theta}_1 p_1}{\rho}, \frac{\bar{\theta}_2 p_2}{\rho}, ..., \frac{\bar{\theta}_{|\mathcal{X}|-1} p_{|\mathcal{X}|-1}}{\rho}, 1 - \frac{\sum_{j=1}^{|\mathcal{X}|-1} \bar{\theta}_j p_j}{\rho}), \tag{3.7}$$

$$P_2 := (\frac{\theta_1 p_1}{\bar{\rho}}, \frac{\theta_2 p_2}{\bar{\rho}}, ..., \frac{\theta_{|\mathcal{X}|-1} p_{|\mathcal{X}|-1}}{\bar{\rho}}, 1 - \frac{\sum_{j=1}^{|\mathcal{X}|-1} \theta_j p_j}{\bar{\rho}}). \tag{3.8}$$

Now, we prove the proposition by solving the optimization problem in (3.5) and simplifying (3.4) for the special case of $Q' = P$. We would like to find the optimal $\Theta := (\theta_1, \theta_2, ..., \theta_{|\mathcal{X}|-1})$ for which the function

$$\tilde{e}(P, Q, \rho, \Theta) := \rho H(P_1) + \bar{\rho} H(P_2) + \sum_{j=1}^{|\mathcal{X}|-1} \theta_j p_j \log a_j$$

is maximized subject to the constraints

$$0 \leq \theta_j \leq 1, \quad j = 1, 2, ..., |\mathcal{X}| - 1, \tag{3.9}$$

$$\sum_{j=1}^{|\mathcal{X}|-1} \bar{\theta}_j p_j \leq \rho, \tag{3.10}$$

$$\sum_{j=1}^{|\mathcal{X}|-1} \theta_j p_j \leq \bar{\rho}, \tag{3.11}$$

where (3.10) and (3.11) arise from the fact that P_1 and P_2 should be probability mass functions, respectively. Note that $\tilde{e}$ is concave in Θ, because $H(P)$ is a concave function in P, and P_1 and P_2 are linear functions in Θ. Therefore, the optimal Θ can be found by setting the derivative of $\tilde{e}$ with respect to θ_j's, $j = 1, 2, ..., |\mathcal{X}| - 1$ to zero, and verifying that the solution satisfies the constraints [2]. We have

$$\frac{\partial \tilde{e}}{\partial \theta_j} = p_j \log \frac{\theta_j(\rho - \sum_{j=1}^{|\mathcal{X}|-1} \bar{\theta}_j p_j)}{\bar{\theta}_j(\bar{\rho} - \sum_{j=1}^{|\mathcal{X}|-1} \theta_j p_j) a_j} = 0, \quad j = 1, 2, ..., |\mathcal{X}| - 1.$$

Therefore, we have

$$\frac{\bar{\theta}_j}{\theta_j} \frac{p_j}{q_j} = c^* = \frac{p_{|\mathcal{X}|}}{q_{|\mathcal{X}|-1}} \frac{\rho - \sum_{j=1}^{|\mathcal{X}|-1} \bar{\theta}_j p_j}{\bar{\rho} - \sum_{j=1}^{|\mathcal{X}|-1} \theta_j p_j}, \quad j = 1, 2, ..., |\mathcal{X}| - 1, \tag{3.12}$$

where c^* is defined to be equal to the right term in (3.12) since it is fixed for all j's. From the left equality in (3.12), the optimal solution is characterized by

$$\theta_j^* = \frac{p_j}{c^* q_j + p_j}, \quad j = 1, 2, ..., |\mathcal{X}| - 1, \tag{3.13}$$

and from the right equality in (3.12), c^* is found to be the solution to (3.3). Now, we check that this solution satisfies the constraints. Let $g(c) := c \sum_{j=1}^{|\mathcal{X}|} \frac{p_j q_j}{c q_j + p_j} - \rho$ be a function of c for a fixed P, Q, and ρ whose root gives the value of c^*. Note that the equation $g(c) = 0$ cannot have more than one solutions, otherwise, the concave and differentiable function $\tilde{e}$ would have more than one local maximums, which is a contradiction. Also, note that

$$\lim_{c \to 0} g(c) = -\rho \leq 0,$$
$$\lim_{c \to \infty} g(c) = 1 - \rho \geq 0,$$

and because the function $g(c)$ is continuous in c, the root of this function is unique and is non-negative. Therefore, $c^* \geq 0$ is the unique solution to (3.3). From (3.13) and the fact that $c^* \geq 0$, the constraint (3.9) is satisfied. From the right equality in (3.12) and the fact that $c^* \geq 0$, we observe that

$$\frac{\rho - \sum_{j=1}^{|\mathcal{X}|-1} \bar{\theta}_j p_j}{\bar{\rho} - \sum_{j=1}^{|\mathcal{X}|-1} \theta_j p_j} \geq 0.$$

Also, note that numerator and denominator of the above expression cannot simultaneously be negative, and therefore, both constraints (3.10) and (3.11) are satisfied as well.

Using the optimal solution obtained above, the elements of the PMF's P_1 and P_2 in (3.7) and (3.8), respectively, are obtained as

$$p_{1,j} = \frac{p_j q_j}{c^* q_j + p_j} \Big/ \sum_{j=1}^{|\mathcal{X}|} \frac{p_j q_j}{c^* q_j + p_j},$$
$$p_{2,j} = \frac{p_j^2}{c^* q_j + p_j} \Big/ \sum_{j=1}^{|\mathcal{X}|} \frac{p_j^2}{c^* q_j + p_j},$$

for $j = 1, 2, ..., |\mathcal{X}|$. Substituting the optimal solution into (3.5) and then into (3.4), (3.2) can be obtained after some manipulation. $\square$

3.2 Properties of Partial Divergence

In this section, we examine some of the interesting properties of partial divergence that provide insights about the achievable rates in Chapter 4.

Proposition 3.2. *The partial divergence $d_\rho(P\|Q), 0 \leq \rho \leq 1$, where all of the elements of the PMF Q are nonzero, 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. (a) From (3.3), we observe that $c^* \to 0$ as $\rho \to 0$. Therefore, from (3.2), we have

$$d_0(P\|Q) = D(P\|Q) - \sum_{j=1}^{|\mathcal{X}|} p_j \log \frac{p_j}{q_j} + c^* \log c^* + h(0) = 0,$$

for $c^* \to 0$.

(b) From (3.3), we observe that $c^* \to \infty$ as $\rho \to 1$. Therefore, from (3.2), we have

$$d_1(P\|Q) = D(P\|Q) - \sum_{j=1}^{|\mathcal{X}|} p_j \log(1 + \frac{p_j}{c^* q_j}) + h(1) = D(P\|Q),$$

for $c^* \to \infty$.

(c) If $P = Q$, then (3.3) simplifies to $\frac{c^*}{c^*+1} = \rho$, and therefore $c^* = \rho/\bar{\rho}$. By substituting c^* into (3.2) and because $D(P\|P) = 0$, we obtain

$$d_\rho(P\|P) = -\log\left(\frac{\rho}{\bar{\rho}} + 1\right) + \rho \log \frac{\rho}{\bar{\rho}} + h(\rho) = 0.$$

(d) By taking the derivative of (3.2) with respect to ρ, we obtain

$$
\begin{aligned}
d'_\rho(P\|Q) &= -\frac{\partial c^*}{\partial \rho}\sum_{j=1}^{|\mathcal{X}|}\frac{p_j q_j}{c^* q_j + p_j} + \log c^* + \frac{\partial c^*}{\partial \rho}\frac{\rho}{c^*} + \log\frac{\bar{\rho}}{\rho} \\
&= \frac{\partial c^*}{\partial \rho}\Big(\frac{\rho}{c^*} - \sum_{j=1}^{|\mathcal{X}|}\frac{p_j q_j}{c^* q_j + p_j}\Big) + \log\Big(c^*\frac{\bar{\rho}}{\rho}\Big) \\
&= \log\Big(c^*\frac{\bar{\rho}}{\rho}\Big), \qquad (3.14)
\end{aligned}
$$

where (3.14) is obtained by using (3.3). Therefore,

$$
d'_0(P\|Q) = \lim_{\rho\to 0} d'_\rho(P\|Q) = \lim_{\rho\to 0}\log\Big(c^*\frac{\bar{\rho}}{\rho}\Big) = 0,
$$

because we have $\lim_{\rho\to 0}\frac{\rho}{c^*} = \sum_{j=1}^{|\mathcal{X}|}\frac{p_j q_j}{p_j} = 1$ from (3.3).

(e) According to (3.14), in order to prove $d'_\rho(P\|Q) > 0, 0 < \rho \le 1$, it is enough to show that $\frac{\rho}{\bar{\rho}} < c^*, 0 < \rho \le 1$:

$$
\begin{aligned}
1 &= \left(\sum_{j=1}^{|\mathcal{X}|} p_j\right)^2 \\
&< \sum_{j=1}^{|\mathcal{X}|}\big(\sqrt{c^* q_j + p_j}\big)^2 \cdot \sum_{j=1}^{|\mathcal{X}|}\Big(\frac{p_j}{\sqrt{c^* q_j + p_j}}\Big)^2 \qquad (3.15) \\
&= (c^* + 1)\sum_{j=1}^{|\mathcal{X}|}\frac{p_j^2}{c^* q_j + p_j} \\
&= (c^* + 1)\bar{\rho}, \qquad (3.16)
\end{aligned}
$$

where (3.15) follows from the Cauchy-Schwarz inequality, and (3.16) is true because

$$
\begin{aligned}
\bar{\rho} = 1 - \rho &= \sum_{j=1}^{|\mathcal{X}|} p_j - \rho \\
&= \sum_{j=1}^{|\mathcal{X}|} p_j - \sum_{j=1}^{|\mathcal{X}|}\frac{c^* p_j q_j}{c^* q_j + p_j} \qquad (3.17) \\
&= \sum_{j=1}^{|\mathcal{X}|}\frac{p_j^2}{c^* q_j + p_j},
\end{aligned}
$$

where (3.17) follows from (3.3). Note that the Cauchy-Schwarz inequality in (3.15) cannot hold with equality for $0 < \rho$ (and therefore, for $0 < c^*$), because otherwise, $p_j = q_j, j = 1, 2, ..., |\mathcal{X}|$, and $P = Q$. From (3.16), $1 < (c^* + 1)\bar{\rho}$, which results in the desirable inequality $\frac{\rho}{\bar{\rho}} < c^*$.

(f) By taking the derivative of (3.14) with respect to ρ, it can be seen that

$$d''_\rho(P\|Q) = \frac{1}{c^*}\frac{\partial c^*}{\partial \rho} - \frac{1}{\rho\bar{\rho}}. \tag{3.18}$$

Also, by taking the derivative of (3.3) with respect to ρ and after some calculation, we have

$$\sum_{j=1}^{|\mathcal{X}|} \frac{(c^* q_j)^2 p_j}{(c^* q_j + p_j)^2} = \rho - \frac{c^*}{\frac{\partial c^*}{\partial \rho}}. \tag{3.19}$$

Therefore,

$$\begin{aligned}
\rho^2 &= \left(\sum_{j=1}^{|\mathcal{X}|} \frac{c^* p_j q_j}{c^* q_j + p_j} \right)^2 \\
&< \sum_{j=1}^{|\mathcal{X}|} \frac{(c^* q_j)^2 p_j}{(c^* q_j + p_j)^2} \cdot \sum_{j=1}^{|\mathcal{X}|} p_j \qquad (3.20) \\
&= \rho - \frac{c^*}{\frac{\partial c^*}{\partial \rho}}, \qquad (3.21)
\end{aligned}$$

where (3.20) follows from the Cauchy-Schwarz inequality, which cannot hold with equality since otherwise $P = Q$, and where (3.21) follows from (3.19). From (3.21), $\rho^2 < \rho - \frac{c^*}{\frac{\partial c^*}{\partial \rho}}$, which implies that $d''_\rho(P\|Q) > 0$ according to (3.18).

(g) From part (f), $d_\rho(P\|Q)$ is convex in ρ, and therefore, $\frac{d_\rho(P\|Q)}{\rho}$ is increasing in ρ. In addition, from part (d), $\lim_{\rho\to 0} \frac{d_\rho(P\|Q)}{\rho} = d'_0(P\|Q) = 0$, and from part (b), $\lim_{\rho\to 1} \frac{d_\rho(P\|Q)}{\rho} = D(P\|Q)$. Consequently, $0 \le d_\rho(P\|Q) \le \rho D(P\|Q)$.

$\square$

Figure 3.1 shows two examples of the partial divergence for PMF's with alphabets of size 4. Specifically, $d_\rho(P\|Q)$ versus ρ is plotted 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 3.2 are apparent in the figure for these examples.

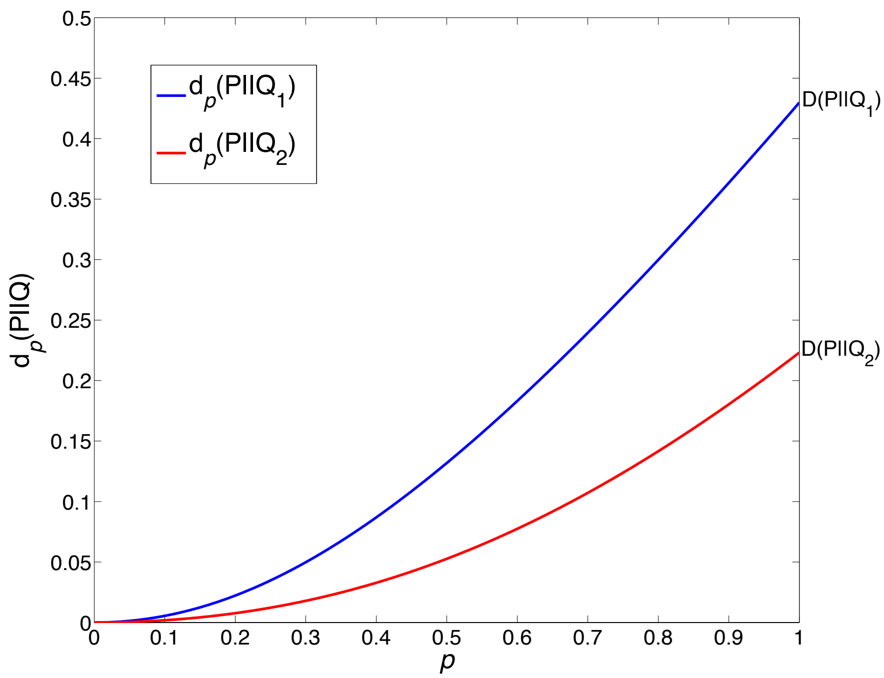

Figure 3.1. Partial divergence $d_\rho(P\|Q)$ versus ρ for $P = (0.25, 0.25, 0.25, 0.25)$, $Q_1 = (0.1, 0.1, 0.1, 0.7)$, and $Q_2 = (0.1, 0.4, 0.1, 0.4)$.

Proposition 3.3. *The partial divergence* $d_\rho(P\|Q), 0 \leq \rho \leq 1$, *satisfies*

$$d_\rho(P\|Q) \geq D(P\|\rho Q + \bar{\rho}P).$$

Proof. From the definition of the partial divergence and (3.1), we have

$$\begin{aligned} d_\rho(P\|Q) &= \min_{\substack{P_1, P_2 \in \mathcal{P}^{\mathcal{X}}: \\ \rho P_1 + \bar{\rho} P_2 = P}} \rho D(P_1\|Q) + \bar{\rho} D(P_2\|P) \\ &\geq \min_{\substack{P_1, P_2 \in \mathcal{P}^{\mathcal{X}}: \\ \rho P_1 + \bar{\rho} P_2 = P}} D(\rho P_1 + \bar{\rho} P_2 \| \rho Q + \bar{\rho} P) \qquad (3.22) \\ &= D(P\|\rho Q + \bar{\rho} P), \qquad (3.23) \end{aligned}$$

where (3.22) follows from the convexity of the Kullback-Leibler divergence [7]; and (3.23) follows from the constraint $\rho P_1 + \bar{\rho} P_2 = P$ in the minimization. □

The interpretation of Proposition 3.3 is that if all the elements of a sequence are generated independently according to a mixture probability $\rho Q + \bar{\rho} P$, then it is more probable that this sequence has type P than in the case that a fraction ρ of its elements are generated independently according to Q and the remaining fraction $\bar{\rho}$ are generated independently according to P. Since the partial divergence $d_\rho(P\|Q)$ is used to obtain achievability results, it can be substituted with the less complicated function $D(P\|\rho Q + \bar{\rho} P)$ in the achievable rates throughout this dissertation with the expense of loosening the bounds according to Proposition 3.3.

3.3 Generalization of the Lemma and Partial Divergence

In this section, we first state a generalization for Lemma 3.1 for which the sequence is generated according to three distributions, and then specialize it to partial divergence with two mismatch factors ρ_1 and ρ_2.

Lemma 3.2. *Consider the distributions $P, Q_1, Q_2, Q_3 \in \mathcal{P}^{\mathcal{X}}$ on a finite alphabet $\mathcal{X}$. A random sequence X^k is generated as follows: $\rho_1 k$ symbols iid according to Q_1, $\rho_2 k$ symbols iid according to Q_2, and $\rho_3 k$ iid according to Q_3, where $\rho_1 + \rho_2 + \rho_3 = 1$. The normalized exponent of the probability that X^k has type P is*

$$\begin{aligned} &\lim_{k\to\infty} -\frac{1}{k} \log \mathbb{P}(X^k \in T_P) \\ &= \min_{\substack{P_1,P_2,P_3 \in \mathcal{P}^{\mathcal{X}}: \\ \rho_1 P_1 + \rho_2 P_2 + \rho_3 P_3 = P}} \rho_1 D(P_1\|Q_1) + \rho_2 D(P_2\|Q_2) + \rho_3 D(P_3\|Q_3) \end{aligned} \tag{3.24}$$

Proof. With some abuse of notation, let $X_1^{\rho_1 k}$, $X_2^{\rho_2 k}$, and $X_3^{\rho_3 k}$ be the sequence of symbols in X^k that are iid according to Q_1, Q_2, and Q_3, respectively. If these sequences have types P_1, P_2, and P_3, respectively, then the whole sequence X^k has type $\rho_1 P_1 + \rho_2 P_2 + \rho_3 P_3$. Therefore, we have

$$
\begin{aligned}
&\mathbb{P}(X^k \in T_P) \\
&= \mathbb{P}\left(\cup_{\substack{P_1,P_2,P_3 \in \mathcal{P}^{\mathcal{X}}: \\ \rho_1 P_1+\rho_2 P_2+\rho_3 P_3=P}} \{X_1^{\rho_1 k} \in T_{P_1}, X_2^{\rho_2 k} \in T_{P_2}, X_3^{\rho_3 k} \in T_{P_3}\} \right) && (3.25) \\
&= \sum_{\substack{P_1,P_2,P_3 \in \mathcal{P}^{\mathcal{X}}: \\ \rho_1 P_1+\rho_2 P_2+\rho_3 P_3=P}} \mathbb{P}(X_1^{\rho_1 k} \in T_{P_1}, X_2^{\rho_2 k} \in T_{P_2}, X_3^{\rho_3 k} \in T_{P_3}) && (3.26) \\
&\doteq \sum_{\substack{P_1,P_2,P_3 \in \mathcal{P}^{\mathcal{X}}: \\ \rho_1 P_1+\rho_2 P_2+\rho_3 P_3=P}} \exp\{-k(\rho_1 D(P_1\|Q_1) + \rho_2 D(P_2\|Q_2) + \rho_3 D(P_3\|Q_3))\}, && (3.27) \\
&\doteq \exp\left\{ -k \min_{\substack{P_1,P_2,P_3 \in \mathcal{P}^{\mathcal{X}}: \\ \rho_1 P_1+\rho_2 P_2+\rho_3 P_3=P}} \rho_1 D(P_1\|Q_1) + \rho_2 D(P_2\|Q_2) + \rho_3 D(P_3\|Q_3) \right\} && (3.28)
\end{aligned}
$$

where: (3.26) follows from the disjointedness of the events in (3.25) since a sequence has a unique type; (3.27) follows from the independence of the three events in (3.26) and obtaining the probability of each of them according to Lemma 2.1 to first order in the exponent; and (3.28) follows from the fact that the number of different types is polynomial in the length of the sequence [10], which makes the total number of terms in the summation (3.27) polynomial in k, and therefore, the exponent equals the largest exponent of the terms in the summation (3.27). □

We will be interested in a special case of Lemma 3.2 for which $Q_3 = P$. In other words, we need to find the exponent of the probability that a sequence has a type P if its elements are generated independently according to Q_1, Q_2, and P. For this case, we denote the right-hand side of (3.24) by $d_{\rho_1,\rho_2}(P\|Q_1, Q_2)$, where $\rho_1 + \rho_2 < 1$. This function will be used in Chapter 5.

3.4 Summary

In this chapter, we introduced a quantity called partial divergence, which is a generalization of the Kullback-Leibler divergence, and is the normalized 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 characterizing a decoder's ability to distinguish a sequence obtained partially from the codewords and partially from the noise from a codeword sequence or a noise sequence. We also studied some of the properties of partial divergence that provide insights about the achievable rates for intermittent communication. Finally, we generalized the partial divergence to the case of three distributions.

CHAPTER 4

SINGLE-USER INTERMITTENT COMMUNICATION

In this chapter, after introducing a model for single-user intermittent communication in Section 4.1, we develop two coding theorems for achievable rates to lower bound the capacity in Section 4.2. Toward this end, we use some of the results on partial divergence and its properties from Chapter 3. We show that, as long as the ratio of the receive window to the codeword length is finite and the capacity of the DMC is not zero, rate $R = 0$ is achievable for intermittent communication; i.e., if there are only two messages, then the probability of decoding error vanishes as the codeword length becomes sufficiently large. By using two decoding structures, we obtain achievable rates that are also valid for arbitrary intermittent processes.

Focusing on the binary-input binary-output noiseless case, we obtain upper bounds on the capacity of intermittent communication in Section 4.3 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. Although the gap between the achievable rates and upper bounds is fairly large, especially for large values of intermittency rate, the results suggest that 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 and the intermittency rate.

Finally, in Section 4.4 we obtain lower and upper bounds on the capacity per unit cost of intermittent communication. To obtain the lower bound, we use a similar approach to the one in [24], in which pulse-position modulation codewords are used at the encoder, and the decoder searches for the position of the pulse in the output

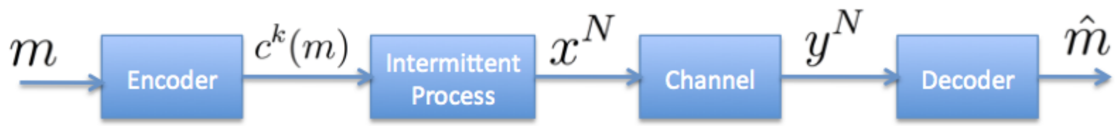

Figure 4.1. System model for intermittent communication.

sequence. The upper bound is the capacity per unit cost of the DMC.

4.1 System Model and Foundations

We consider a communication scenario in which a transmitter communicates a single message $m \in [1 : e^{kR} = M]$ to a receiver over a DMC with probability transition matrix W and input and output alphabets $\mathcal{X}$ and $\mathcal{Y}$, respectively. Let C_W denote the capacity of the DMC. Also, let $\star \in \mathcal{X}$ denote the "silent" or the "noise" symbol, which corresponds to the input of the channel when the transmitter is "silent". For convenience, we define $W_\star(\cdot) := W(\cdot|X = \star)$, and more generally, $W_x(\cdot) := W(\cdot|X = x)$. The transmitter encodes the message as a codeword $c^k(m)$ of length k, which is the input sequence of intermittent process shown in Figure 4.1.

The intermittent process captures the burstiness of the channel or the transmitter and can be described as follows: After the i^{th} symbol from the codeword, N_i noise symbols $\star$ are inserted, where N_i's are iid geometric random variables with mean $\alpha - 1$, where $\alpha \geq 1$ is the *intermittency rate*. As will see later, if $N \geq k$ is the random variable denoting the total number of received symbols, the intermittency rate $N/k \xrightarrow{p} \alpha$ as $k \to \infty$ will be an important parameter of the system. In fact, the larger the value of α, the larger the receive window, and therefore, the more intermittent the system becomes with more uncertainty about the positions of the codeword symbols; if $\alpha = 1$, the system is not intermittent and corresponds to contiguous communication. We call this scenario *intermittent communication* and denote it by the tuple $(\mathcal{X}, \mathcal{Y}, W, \star, \alpha)$.

This model corresponds to an iid insertion channel model in which at each time slot a codeword symbol is sent with probability $p_t := 1/\alpha$ and the noise symbol $\star$ is sent with probability $1 - p_t$ until the whole codeword is transmitted.

The output of the intermittent process then goes through a DMC. At the decoder, there are N symbols, where N is a random variable having a negative binomial distribution with parameters k and p_t:

$$P(N = n) = \binom{n-1}{k-1} p_t^k (1 - p_t)^{n-k}, n \geq k, \tag{4.1}$$

with $\mathbb{E}[N] = \alpha k$, and we have

$$\frac{N}{k} = \frac{k + N_0 + N_1 + N_2 + ... + N_k}{k} \xrightarrow{p} 1 + \mathbb{E}(N_0) = \alpha, \text{ as } k \to \infty, \tag{4.2}$$

Therefore, the receive window N scales linearly with the codeword length k, as opposed to the exponential scaling in asynchronous communication summarized in Section 2.3. The intermittent communication model represents bursty communication in which either the transmitter or the channel is bursty. In a bursty communication scenario, the receiver usually does not know the realization of the bursts. Therefore, we assume that the receiver does not know the positions of the codeword symbols, making the decoder's task more involved. However, we assume that the receiver knows the codeword length k and the realization of the size of the receive window n.

Denoting the decoded message by $\hat{m}$, which is a function of the random sequence Y^N, and defining a code as in [7], we say that rate R is achievable if there exists a sequence of length k codes of size $M = e^{kR}$ with average probability of error $\frac{1}{M}\sum_{m=1}^{M} \mathbb{P}(\hat{m} \neq m) \to 0$ 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. Rate region (R, α) is said to be achievable if the rate R is achievable for the corresponding scenario with the intermittency rate α.

It can be seen that the result of Theorem 2.2 for iid synchronization error channels applies to the intermittent communication model, and therefore, the capacity equals $\lim_{k\to\infty}\max_{P_{C^k}}\frac{1}{k}\mathbb{I}(C^k;Y^N)$. We have the following theorem.

Theorem 4.1. *For intermittent communication* $(\mathcal{X},\mathcal{Y},W,\star,\alpha)$*, rates less than* $R_1 := (C_W - \alpha h(1/\alpha))^+$ *are achievable.*

Proof. We show that R_1 is a lower bound for the capacity of intermittent communication by lower bounding the mutual information. Let vector $T^{k+1} := (N_0, N_1, ..., N_k)$ denote the number of noise insertions in between the codeword symbols, where the N_i's are iid geometric random variables with mean $\alpha - 1$. We have

$$\begin{aligned}
\mathbb{I}(C^k;Y^N) &= \mathbb{I}(C^k;Y^N,T^{k+1}) - \mathbb{I}(C^k;T^{k+1}|Y^N) \\
&= \mathbb{I}(X^k;Y^k) - \mathbb{I}(C^k;T^{k+1}|Y^N) && (4.3)\\
&\geq k\mathbb{I}(X;Y) - H(T^{k+1}) && (4.4)\\
&= k\mathbb{I}(X;Y) - (k+1)H(N_0) && (4.5)\\
&= k\mathbb{I}(X;Y) - (k+1)\alpha h(\frac{1}{\alpha}), && (4.6)
\end{aligned}$$

where (4.3) follows from the fact that C^k is independent of the insertion process T^{k+1}, and conditioned on the positions of noise symbols T^{k+1}, the mutual information between C^k and Y^N equals the mutual information between input and output sequences of the DMC without considering the noise insertions; where (4.4) follows by considering iid codewords and by the fact that conditioning cannot increase the entropy; and (4.5) and (4.6) follow from the fact that N_i's are iid geometric random variables. Finally, the result follows after dividing both sides by k and considering the capacity achieving input distribution of the DMC. □

Although the lower bound on the capacity of intermittent communication in Theorem 4.1 is valid for the specific intermittent process described above, our achievability results in Section 4.2 apply to an arbitrary insertion process as long as $N/k \xrightarrow{p} \alpha$ as $k \to \infty$.

4.2 Achievability

In this section, we obtain achievability results for intermittent communication based upon two decoding structures: *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 codeword symbols; and *decoding from pattern detection*, which attempts to decode the transmitted codeword only if the selected outputs appear to be a pattern of codeword symbols. In order to analyze the probability of error for the second decoding structure, which gives a larger achievable rate, we use some of the results on partial divergence and its properties from Chapter 3 .

In Section 4.2.1, we also show that rate $R = 0$ is achievable for intermittent communication with finite intermittency rate, using the properties of partial divergence. In Section 4.2.2, we introduce decoding from exhaustive search and decoding from pattern detection. Finally, in Section 4.2.3, using these decoding structures, we obtain achievable rates for intermittent communication.

Although the system model in Section 4.1 assumes iid geometric insertions, the results of this section apply to a general intermittent process, as we point out in Remark 4.2.

4.2.1 Achievability of Rate $R = 0$

Using the results on partial divergence, we now state a result about the achievability of rate $R = 0$ in the following theorem. The idea is that no matter how large the intermittency rate becomes, as long as it is finite, the receiver can distinguish between two messages with vanishing probability of error.

Theorem 4.2. *If the intermittency rate is finite and the capacity of the DMC, C_W, is non-zero, then rate $R = 0$ is achievable for intermittent communication $(\mathcal{X}, \mathcal{Y}, W, \star, \alpha)$ for the case of having only two messages.*

Proof. We need to show that the transmission of a message $m \in \{1, 2\}$ is reliable for intermittent communication $(\mathcal{X}, \mathcal{Y}, W, \star, \alpha)$ as $k \to \infty$.

Encoding: If $m = 1$, then transmit symbol $\star$ at all the times, i.e., $c^k(1) = \star^k$. If $m = 2$, then transmit symbol $x \neq \star$ at all the times, i.e., $c^k(2) = x^k$, where we consider the symbol $x^* = \text{argmax}_x D(W_\star \| W_x)$. If the capacity of the DMC is nonzero, then $W_\star \neq W_{x^*}$.

Decoding: For arbitrarily small ϵ, if $|N/k - \alpha| > \epsilon$, then the decoder declares an error; otherwise, if the sequence y^N has type $W_\star$ with a fixed typicality parameter $\mu > 0$, i.e., $y^N \in T_{[W_\star]_\mu}$, then $\hat{m} = 1$; otherwise $\hat{m} = 2$.

Analysis of the probability of error: The average probability of error can be bounded as

$$p_e \leq \mathbb{P}(|N/k - \alpha| > \epsilon) + \sum_{n=k(\alpha-\epsilon)-1}^{k(\alpha+\epsilon)+1} \mathbb{P}(N=n) \left(\mathbb{P}(Y^N \notin T_{[W_\star]_\mu} | m=1, N=n) + \mathbb{P}(Y^N \in T_{[W_\star]_\mu} | m=2, N=n)\right) \tag{4.7}$$

$$\leq o(1) + \max_{n:|n/k-\alpha|<\epsilon} (o(1) + e^{-n d_{k/n}(W_\star \| W_{x^*})}) \tag{4.8}$$

$$\leq o(1) + e^{-k(\alpha-\epsilon) d_{1/(\alpha+\epsilon)}(W_\star \| W_{x^*})} \to 0 \text{ as } k \to \infty, \tag{4.9}$$

where: (4.7) follows from the union bound; (4.8) follows from the fact that $\mathbb{P}(|N/k - \alpha| > \epsilon) \to 0$ as $k \to \infty$, Lemma 2.2 and Lemma 3.1; and (4.9) follows from the fact that $d_{1/(\alpha+\epsilon)}(W_\star \| W_{x^*}) > 0$ according to Proposition 3.2 since $W_\star \neq W_{x^*}$ and $1/(\alpha + \epsilon) > 0$. $\square$

In order to prove achievability results for the case of an exponential number of messages, i.e., $R > 0$, we introduce two decoding structures in the following section.

4.2.2 Decoding Structures

In this section, two decoding structures are introduced. The encoding structure is identical for both: Given an input distribution P, the codebook is randomly and independently generated, i.e., all $C_i(m), i \in [1:k], m \in [1:M]$ are iid according to P. Although we focus on typicality for detection and decoding for ease of analyzing the probability of error, other algorithms such as maximum likelihood decoding could 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 dissertation. Note that the number of received symbols at the decoder, N, is a random variable. However, using the same procedure as in the proof of Theorem 4.2, we can focus on the case that $|N/k - \alpha| < \epsilon$, and essentially assume that the receive window is of length $n = \alpha k$, which makes the analysis of the probability of error for the decoding algorithms more concise.

Decoding from exhaustive search: In this structure, the decoder observes the n symbols of the output sequence y^n, chooses k of them, resulting in a subsequence denoted by $\tilde{y}^k$, and performs joint typicality decoding with a fixed typicality parameter $\mu > 0$, i.e., checks if $\tilde{y}^k \in T_{[W]_\mu}(c^k(m))$ for a unique index m. In words, this condition 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 this condition, it declares m as the transmitted message. Otherwise, it makes another choice for the k symbols from symbols of the sequence y^n and again attempts typicality decoding. If at the end of all $\binom{n}{k}$ choices the typicality decoding procedure does not declare any message as being transmitted, then the decoder declares an error.

Decoding from pattern detection: This structure involves two stages for each choice of the output symbols. As in decoding from exhaustive search, the decoder chooses k of the n symbols from the output sequence y^n. Let $\tilde{y}^k$ denote the subsequence of the chosen symbols, and $\hat{y}^{n-k}$ denote the subsequence of the other symbols. 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 induced 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 with $\tilde{y}^k$ over the codebook as in the decoding from exhaustive search, 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 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 codeword symbols satisfies the conditions in the first stage.

In decoding from pattern detection, the first stage 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 Chapter 3 play a role. This structure always outperforms decoding from exhaustive search, and their difference in performance indicates how much the results on the partial divergence improve the achievable rates.

4.2.3 Achievable Rates

In this section, we obtain two achievable rates for intermittent communication using the decoding algorithms introduced in Section 4.2.2.

Theorem 4.3. *Using decoding from exhaustive search for intermittent communication with $(\mathcal{X}, \mathcal{Y}, W, \star, \alpha)$, rates less than $R_1 = (C_W - \alpha h(1/\alpha))^+$ are achievable.*

Proof. Let P be the capacity achieving input distribution for the DMC with stochastic matrix W, and consider decoding from exhaustive search described in Sections 4.2.2. For any $\epsilon > 0$, we prove that if $R = C_W - \alpha h(1/\alpha) - 2\epsilon$, then the average probability vanishes as $k \to \infty$.

The analysis of the probability of error is similar to that of Theorem 4.4, except that instead of breaking down the first term in (4.14) as in (4.17), we use the union bound over all the $\binom{n}{k}$ choices without trying to distinguish the output symbols based on their empirical distributions.

Specifically, using the union bound, we have (4.14) in which the second term vanishes as $k \to \infty$ according to (4.16). Using the union bound for the first term in (4.14), we have

$$\mathbb{P}(\hat{m} \in \{2, 3, ..., M\} | m = 1) \leq \binom{n}{k} (M - 1) \mathbb{P}(\tilde{Y}^k \in T_{[W]_\mu}(C^k(2)) | m = 1), \quad (4.10)$$

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 Stirling's approximation, we have

$$\binom{n}{k} \leq \frac{e^{\frac{1}{12}}}{\sqrt{2\pi}} \sqrt{\frac{n}{k(n-k)}} e^{nh(k/n)} \doteq e^{k\alpha h(1/\alpha)}, \quad (4.11)$$

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 Lemma 2.3 with the capacity achieving input distribution, we have

$$\mathbb{P}(\tilde{Y}^k \in T_{[W]_\mu}(C^k(2)) | m = 1) \leq \text{poly}(k) e^{-k(C - \epsilon)}. \quad (4.12)$$

Combining (4.10), (4.11), and (4.12), and substituting $R = C_W - \alpha h(1/\alpha) - 2\epsilon$, we have

$$\mathbb{P}(\hat{m} \in \{2, 3, ..., M\}|m = 1) \leq e^{o(k)} e^{-k\epsilon} \quad \to 0 \text{ as } k \to \infty.$$

Therefore, the average probability of error vanishes as $k \to \infty$. □

Note that R_1 is the same as the lower bound in Theorem 4.1, but here we introduced an explicit decoding structure which is also valid for a general intermittent process. The form of the achievable rate is reminiscent of communications overhead as the cost of constraints [32], where the constraint is the system's burstiness or intermittency, and the overhead cost is $\alpha h(1/\alpha)$. Note that the overhead cost is increasing in the intermittency rate α and is equal to zero at $\alpha = 1$. These observations suggest that increasing the receive window makes the decoder's task more difficult.

Theorem 4.4. *Using decoding from pattern detection for intermittent communication with* $(\mathcal{X}, \mathcal{Y}, W, \star, \alpha)$, *rates less than* $\max_P\{(\mathbb{I}(P, W) - f(P, W, \alpha))^+\}$ *are achievable, where*

$$f(P, W, \alpha) := \max_{0 \leq \beta \leq 1} \{(\alpha-1)h(\beta) + h((\alpha-1)\beta) - d_{(\alpha-1)\beta}(PW \| W_\star) - (\alpha-1)d_\beta(W_\star \| PW)\}. \tag{4.13}$$

Proof. Fix the input distribution P, and consider decoding from pattern detection described in Section 4.2.2. For any $\epsilon > 0$, we prove that if $R = \mathbb{I}(P, W) - f(P, W, \alpha) - 2\epsilon$, then the average probability of error vanishes as $k \to \infty$. We have

$$p_e^{avg} \leq \mathbb{P}(\hat{m} \in \{2, 3, ..., M\}|m = 1) + \mathbb{P}(\hat{m} = e|m = 1), \tag{4.14}$$

where (4.14) follows from the union bound in which 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,

$$\mathbb{P}(\hat{m} = e|m = 1) \leq \mathbb{P}(Y^k \notin T_{[PW]_\mu}) + \mathbb{P}(Y_\star^{n-k} \notin T_{[W_\star]_\mu}) + \mathbb{P}(Y^k \notin T_{[W]_\mu}(C^k(1))) \tag{4.15}$$

$$\to 0, \text{ as } k \to \infty, \tag{4.16}$$

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 the union bound to establish (4.15). The limit (4.16) follows because all the three terms in (4.15) vanish as $k \to \infty$ according to Lemma 2.2.

The first term in (4.14) is more challenging. It is the probability that for at least one choice of the output symbols, the decoder passes the first stage and then the typicality decoder declares an incorrect message. We characterize the $\binom{n}{k}$ choices based on the number of incorrectly chosen output symbols, i.e., the number of symbols in $\tilde{y}^k$ that are in fact output symbols corresponding to a noise symbol, which is equal to the number of symbols in $\hat{y}^{n-k}$ that are in fact output symbols corresponding to a codeword symbol. For any $0 \leq k_1 \leq n-k$, there are $\binom{k}{k_1}\binom{n-k}{k_1}$ choices. [1] Using the union bound for all the choices and all the messages $\hat{m} \neq 1$, we have

$$\mathbb{P}(\hat{m} \in \{2, 3, ..., M\}|m = 1) \leq (e^{kR} - 1) \sum_{k_1=0}^{n-k} \binom{k}{k_1} \binom{n-k}{k_1} \mathbb{P}_{k_1}(\hat{m} = 2|m = 1), \tag{4.17}$$

where the index k_1 in (4.17) denotes the condition that the number of incorrectly chosen output symbols is equal to k_1. Note that message $\hat{m} = 2$ is declared at the

[1] According to Vandermonde's identity, we have $\sum_{k_1=0}^{n-k}\binom{k}{k_1}\binom{n-k}{k_1} = \binom{n}{k}$.

decoder only if it passes the first and the second stage. Therefore,

$$\begin{aligned}
&\mathbb{P}_{k_1}(\hat{m}=2|m=1)\\
&=\mathbb{P}_{k_1}\left(\{\tilde{Y}^k\in T_{[PW]_\mu}\}\cap\{\hat{Y}^{n-k}\in T_{[W_\star]_\mu}\}\cap\{\tilde{Y}^k\in T_{[W]_\mu}(C^k(2))\}|m=1\right)\\
&=\mathbb{P}_{k_1}(\tilde{Y}^k\in T_{[PW]_\mu})\cdot\mathbb{P}_{k_1}(\hat{Y}^{n-k}\in T_{[W_\star]_\mu})\\
&\quad\cdot\mathbb{P}\left(\tilde{Y}^k\in T_{[W]_\mu}(C^k(2))|m{=}1,\tilde{Y}^k{\in}T_{[PW]_\mu},\hat{Y}^{n-k}{\in}T_{[W_\star]_\mu}\right) &(4.18)\\
&\le e^{o(k)}e^{-kd_{k_1/k}(PW\|W_\star)}e^{-(n-k)d_{k_1/(n-k)}(W_\star\|PW)}e^{-k(\mathbb{I}(P,W)-\epsilon)}, &(4.19)
\end{aligned}$$

where: (4.18) follows from the independence of the events $\{\tilde{Y}^k\in T_{[PW]_\mu}\}$ and $\{\hat{Y}^{n-k}\in T_{[W_\star]_\mu}\}$ conditioned on k_1 incorrectly chosen output symbols; and (4.19) follows from using Lemma 3.1 for the first two terms in (4.18) with mismatch ratios k_1/k and $k_1/(n-k)$, respectively, and using Lemma 2.3 for the last term in (4.18), 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 (4.19) into the summation in (4.17), we have

$$\begin{aligned}
&\mathbb{P}(\hat{m}\in\{2,3,...,M\}|m=1)\\
&\le e^{o(k)}(e^{kR}-1)e^{-k(\mathbb{I}(P,W)-\epsilon)}\sum_{k_1=0}^{n-k}\binom{k}{k_1}\binom{n-k}{k_1}e^{-kd_{k_1/k}(PW\|W_\star)-(n-k)d_{k_1/(n-k)}(W_\star\|PW)}\\
& &(4.20)\\
&\le e^{o(k)}e^{kR}e^{-k(\mathbb{I}(P,W)-\epsilon)}e^{kf(P,W,\alpha)} &(4.21)\\
&=e^{o(k)}e^{-k\epsilon}\quad\to 0\text{ as }k\to\infty, &(4.22)
\end{aligned}$$

where: (4.22) is obtained by substituting $R=\mathbb{I}(P,W)-f(P,W,\alpha)-2\epsilon$; and (4.21) is obtained by finding the exponent of the sum in (4.20) as follows

$$\lim_{k\to\infty} \frac{1}{k} \log \sum_{k_1=0}^{n-k} \binom{k}{k_1}\binom{n-k}{k_1} e^{-kd_{k_1/k}(PW\|W_\star)-(n-k)d_{k_1/(n-k)}(W_\star\|PW)}$$

$$= \lim_{k\to\infty} \frac{1}{k} \log \sum_{k_1=0}^{n-k} \exp\{kh(\tfrac{k_1}{k}) + (n-k)h(\tfrac{k_1}{n-k}) - kd_{\frac{k_1}{k}}(PW\|W_\star) - (n-k)d_{\frac{k_1}{n-k}}(W_\star\|PW)\} \tag{4.23}$$

$$= \lim_{k\to\infty} \frac{1}{k} \max_{k_1=0,\dots,n-k} \{kh(\tfrac{k_1}{k}) + (n-k)h(\tfrac{k_1}{n-k}) - kd_{\frac{k_1}{k}}(PW\|W_\star) - (n-k)d_{\frac{k_1}{n-k}}(W_\star\|PW)\} \tag{4.24}$$

$$\le \max_{0\le\beta\le1} \{(\alpha-1)h(\beta) + h((\alpha-1)\beta) - d_{(\alpha-1)\beta}(PW\|W_\star) - (\alpha-1)d_\beta(W_\star\|PW)\} \tag{4.25}$$

$$= f(P, W, \alpha), \tag{4.26}$$

where: (4.23) follows by using Stirling's approximation for the binomial terms; (4.24) follows by noticing that the exponent of the summation is equal to the largest exponent of each term in the summation, since the number of terms is polynomial in k; (4.25) is obtained by letting $\beta := k_1/(n-k)$ $(0 \le \beta \le 1)$ and substituting $n = \alpha k$; and (4.26) follows from the definition (4.13).

Now, combining (4.14), (4.16), and (4.22), we have $p_e^{avg} \to 0$ as $k \to \infty$, which proves the Theorem. □

Remark 4.1. *The achievable rate in Theorem 4.4 can be expressed as follows: Rate R is achievable if for any mismatch $0 \le \beta \le 1$, we have*

$$R + (\alpha-1)h(\beta) + h((\alpha-1)\beta) < \mathbb{I}(P, W) + d_{(\alpha-1)\beta}(PW\|W_\star) + (\alpha-1)d_\beta(W_\star\|PW).$$

The interpretation is that the total amount of uncertainty should be smaller than the total amount of information. Specifically, R and $(\alpha-1)h(\beta) + h((\alpha-1)\beta)$ are *the amount of uncertainty in codewords and patterns, respectively, and* $\mathbb{I}(P, W)$ *and* $d_{(\alpha-1)\beta}(PW\|W_\star) + (\alpha - 1)d_\beta(W_\star\|PW)$ *are the amount of information about the codewords and patterns, respectively.*

Remark 4.2. *The results of Theorems 4.2, 4.3, and 4.4 are valid for an arbitrary intermittent process in Figure 4.1 provided* $N/k \xrightarrow{p} \alpha$ *as* $k \to \infty$.

The achievable rate in Theorem 4.4 is larger than the one in Theorem 4.3, because decoding from pattern detection utilizes the fact that the choice of the codeword symbols 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 4.4, the overhead cost for a fixed input distribution is $f(P, W, \alpha)$. Using the properties of partial divergence, we state some properties of this overhead cost in the next proposition.

Proposition 4.1. *The overhead cost* $f(P, W, \alpha)$ *in* (4.13) *has the following properties:*

(a) The maximum of the term in (4.13) *occurs in the interval* $[0, 1/\alpha]$, *i.e., instead of the maximization over* $0 \leq \beta \leq 1$, $f(P, W, \alpha)$ *can be found by the same maximization problem over* $0 \leq \beta \leq 1/\alpha$.

(b) $f(P, W, \alpha)$ *is increasing in* α.

(c) $f(P, W, 1) = 0$.

(d) If $D(PW\|W_\star)$ *is finite, then* $f(P, W, \alpha) \to \infty$ *as* $\alpha \to \infty$.

Proof. (a) The term $(\alpha - 1)h(\beta) + h((\alpha - 1)\beta)$ in (4.13) is maximized at $\beta = 1/\alpha$, because it is concave in β and its derivative with respect to β is zero at $1/\alpha$. Thus, this term is decreasing in β in the interval $[1/\alpha, 1]$. Also, note that the partial divergence terms in (4.13) are increasing with respect to β according to Proposition 3.2 (e). Therefore, the term in the max operator in (4.13) is decreasing in β in the interval $[1/\alpha, 1]$, and the maximum occurs in the interval $[0, 1/\alpha]$.

(b) The term in the max operator in (4.13) is concave in β, because $h(\beta)$ is concave in β and $d_\beta(\cdot\|\cdot)$ is convex in β according to Proposition 3.2 (f). Therefore, the term is maximized at a point β^* where the derivative with respect to β is equal to zero. We then have

$$\log\frac{1-\beta^*}{\beta^*}+\log\frac{1-(\alpha-1)\beta^*}{(\alpha-1)\beta^*}-\log c_1\frac{1-(\alpha-1)\beta^*}{(\alpha-1)\beta^*}-\log c_2\frac{1-\beta^*}{\beta^*}=0, \tag{4.27}$$

where: (3.14) is used to derive (4.27); and c_1 and c_2 are the corresponding c^*'s in (3.3) for the two partial divergence terms in (4.13). Taking the derivative of (4.13) with respect to α assuming that the maximum occurs at β^*, we obtain

$$\begin{aligned}
&\frac{\partial f(P,W,\alpha)}{\partial\alpha}\\
&=\frac{\partial\beta^*}{\partial\alpha}(\alpha-1)(\cdot)+h(\beta^*)-d_{\beta^*}(W_\star\|PW)\\
&\quad+\beta^*(\log\frac{1-(\alpha-1)\beta^*}{(\alpha-1)\beta^*}-\log c_1\frac{1-(\alpha-1)\beta^*}{(\alpha-1)\beta^*})\\
&=h(\beta^*)-d_{\beta^*}(W_\star\|PW)+\beta^*(\log c_2\frac{1-\beta^*}{\beta^*}-\log\frac{1-\beta^*}{\beta^*}) && (4.28)\\
&=-\log(1-\beta^*)+\left(\frac{\partial d_{\beta^*}(W_\star\|PW)}{\partial\beta^*}-d_{\beta^*}(W_\star\|PW)\right) && (4.29)\\
&\geq 0, && (4.30)
\end{aligned}$$

where: $(\cdot)$ in the first line is the left-side of (4.27), which is equal to zero; (4.28) follows from (4.27); (4.29) follows from (3.14); and (4.30) follows from the fact that $-\log(1-\beta^*)$ is always positive for $0\leq\beta^*\leq 1$ and $\partial d_{\beta^*}(W_\star\|PW)/\partial\beta^*-d_{\beta^*}(W_\star\|PW)$ is also always positive, because the partial divergence $d_{\beta^*}(W_\star\|PW)$ is convex in β^* according to Proposition 3.2 (f).

(c) Substituting $\alpha=1$ in (4.13), all the terms would be zero, because $h(0)=0$ and $d_0(P\|Q)=0$ according to Proposition 3.2 (a).

(d) Consider that the maximum in (4.13) occurs at β^*. According to part (a), $0\leq\beta^*\leq 1/\alpha$, and therefore, $\beta^*\to 0$ as $\alpha\to\infty$. Using Proposition 3.2 (d) and (4.27), we have $\alpha\beta^*\to 1$ as $\alpha\to\infty$. Substitiuting $\alpha=1/\beta^*$ in (4.13), we obtain

$$\begin{aligned}
&\lim_{\alpha\to\infty} f(P,W,\alpha)\\
&=\lim_{\beta^*\to 0}\frac{h(\beta^*)}{\beta^*}+h(1)-d_1(PW\|W_\star)-\frac{d_{\beta^*}(W_\star\|PW)}{\beta^*}\\
&=\lim_{\beta^*\to 0}\frac{h(\beta^*)}{\beta^*}-D(PW\|W_\star)-d_0'(P\|Q) && (4.31)\\
&=\lim_{\beta^*\to 0}\frac{h(\beta^*)}{\beta^*}-D(PW\|W_\star) && (4.32)\\
&\to\infty, && (4.33)
\end{aligned}$$

where: (4.31) follows from Proposition 3.2 (b); (4.32) follows from Proposition 3.2 (d);

and (4.33) follows from the definition of the binary entropy function and the assumption that $D(PW\|W_\star)$ is finite.

(e) This part follows directly from the definition (4.13).

$\square$

Note that part (b) in Proposition 4.1 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 $\alpha \to \infty$.

Now consider a binary symmetric channel (BSC) for the DMC in Figure 4.1 with the crossover probability $0 \le p \le 0.5$, and the noise symbol $\star = 0$. Figure 4.2 shows the value of the achievable rates for different p, versus α. $R_{\text{Insertion}}$ denotes the achievable rate obtained from Theorem 4.4 if the channel is noiseless ($p = 0$), and can be proven to be equal to $\max_{0\le p_0\le 1}\{2h(p_0) - \max_{0\le\beta\le 1}\{(\alpha-1)h(\beta) + h((\alpha-1)\beta) + (1-(\alpha-1)\beta)h(\frac{p_0-(\alpha-1)\beta}{1-(\alpha-1)\beta})\}\}$.

As we can see from the plot, the achievable rate in Theorem 4.4 (indicated by "R_2") is always larger than the one in Theorem 4.3 (indicated by "R_1") since decoding from pattern detection 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 3.1 in terms of the partial divergence helps the decoder detect the right symbols, and therefore, achieve a larger rate. The arrows in Figure 4.2 show this difference and suggest that the benefit of using decoding from pattern detection is larger for increasing α. Note that the larger α is, the smaller the achievable rate would be for a fixed p. Not surprisingly, as $\alpha \to 1$, the capacity of the BSC is approached for both of the achievable rates. In this example, we cannot achieve a positive rate if $\alpha \ge 2$, 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 C_W is sufficiently large. The results suggest that, as communication

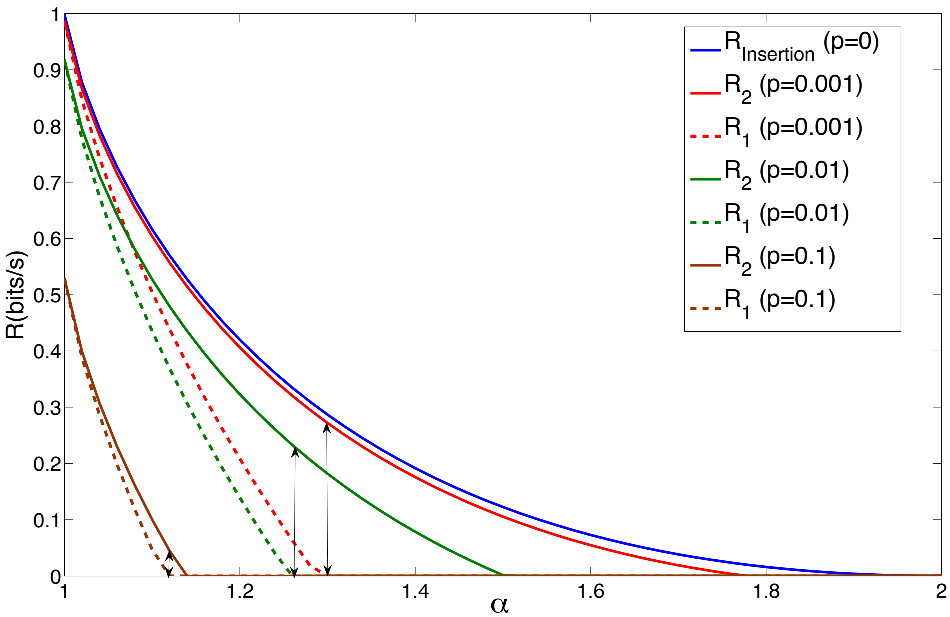

Figure 4.2. Achievable rate region (R, α) for the BSC for different cross over probability p's.

becomes more intermittent and α becomes larger, the achievable rate is decreased due to the additional uncertainty about the positions of the codeword symbols at the decoder.

4.3 Upper Bounds

In this section, we focus on obtaining upper bounds on the capacity of a special case of intermittent communication in which the DMC in Figure 4.1 is binary-input binary-output noiseless with the noise symbol $\star = 0$. The achievable rate for this case is denoted by $R_{\text{Insertion}}$ in Section 4.2.3, and is shown by the blue curve in Figure 4.2. Similar to [18], 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. After introducing a useful function

$g(a,b)$ in Section 4.3.1, we obtain upper bounds in Section 4.3.2. The techniques of in this section can in principle be applied to non-binary and noisy channels as well; however, the computational complexity for numerical evaluation of the genie-aided system grows very rapidly in the size of the alphabets.

The results in this section 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.

4.3.1 Auxiliary Channel: Uniform Insertion

Let a and b be two integer numbers such that $0 \le a \le b$, and consider a discrete memoryless channel for which at each channel use the input consists of a sequence of a bits and the output consists of a sequence of b bits, i.e., the input and output of this channel are $\mathbf{A} \in \{0,1\}^a$ and $\mathbf{B} \in \{0,1\}^b$, respectively. For each channel use, $b-a$ zeroes are inserted randomly and uniformly among the input symbols. The set of positions at which the insertions occur takes on each of the possible $\binom{b}{b-a} = \binom{b}{a}$ realizations with equal probability, and is unknown to the transmitter and the receiver. As an example, the transition probability matrix of this channel for the case of $a = 2$ and $b = 3$ is reported in Figure 4.3.

The capacity of the auxiliary channel is defined as

$$g(a,b) := \max_{P(\mathbf{A})} \mathbb{I}(\mathbf{A};\mathbf{B}), \quad 0 \le a \le b, \tag{4.34}$$

where $P(\mathbf{A})$ is the input distribution. The exact value of the function $g(a,b)$ for finite a and b can be numerically computed by evaluating the transition probabilities $P(\mathbf{B}|\mathbf{A})$ and using the Blahut-Arimoto algorithm [1] to maximize the mutual information. The computational complexity of the Blahut-Arimoto algorithm increases

A \ B	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

Figure 4.3. Transition probabilities $P(\mathbf{B}|\mathbf{A})$ for the auxiliary channel with $a = 2$ and $b = 3$.

exponentially for large values of a and b since the transition probability matrix is of size $2^a \times 2^b$. In order to partially overcome this issue, we recall the following lemma.

Lemma 4.1. *([7, 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 per channel use as*

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

Now, notice that the function $g(a, b)$ can be evaluated by considering the union of $a + 1$ memoryless channels with distinct input and output alphabets, where the input of the i^{th} channel is the set of binary sequences with length a and weight $i - 1$, $i = 1, ..., a + 1$, and the output is binary sequences with length b obtained from the input sequence by inserting $b - a$ zeroes uniformly. The weight of the sequences remains 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 $g_i(a, b)$ and applying Lemma 4.1, we have

$$2^{g(a,b)} = 2^{g_1(a,b)} + ... + 2^{g_{a+1}(a,b)}. \tag{4.35}$$

It is easy to see that $g_1(a,b) = 0$ and $g_{a+1}(a,b) = 0$. For other values of i, the capacity $g_i(a,b), i = 2, ..., a$ can be evaluated numerically using the Blahut-Arimoto algorithm, where input and output alphabets have sizes $\binom{a}{i-1}$ and $\binom{b}{i-1}$, respectively, which are considerably less than those of the original alphabet sizes. This reduction allows us to obtain the function $g(a,b)$ for larger values of a and b, which will be useful in Section 4.3.2. The largest value of b for which we are able to evaluate the function $g(a,b)$ for all values of $a \leq b$ is $b = 17$. Although we cannot obtain a closed-form expression for the function $g(a,b)$, we can find a closed-form upper bound by expanding the mutual information in (4.34) and bounding some of its terms. As a result, we can find upper bounds on the function $g(a,b)$ for larger values of a and b. First, we introduce some notation.

For a binary sequence $x^a \in \{0,1\}^a$, let $w(x^a)$ denote the weight of the sequence, i.e., the number of 1's. Also, let the vector $r^0(x^a) := (r^0_1, ..., r^0_{l_0})$ of length l_0 denote the length of consecutive 0's in the sequence x^a such that $r^0_1 = 0$ if $x_1 = 1$, i.e., the binary sequence starts with 1, and $r^0_{l_0} = 0$ if $x_k = 1$, i.e., the binary sequence ends with 1, and all the other elements of the vector $r^0(x^a)$ are positive integers. In addition, let the vector $r^1(x^a) := (r^1_1, ..., r^1_{l_1})$ of length l_1 denote the length of consecutive 1's in the sequence x^a with length larger than one, i.e., runs of 1's with length one not counted. Finally, let $l(x^a) := l_0 + l_1$. If it is clear from the context, we drop the argument x^a from these functions. For example, if $x^a = 0010111000011$, then $w = 6$, $r^0 = (2,1,4,0)$, $r^1 = (3,2)$, and $l = 4 + 2 = 6$. As another example, if $x^a = 10011101$, 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 upper bound,

$$F(x^a) := \sum_{i_1,...,i_l \in \mathbb{N} \cup \{0\} : \sum_{j=1}^{l} i_j = b-a} p_{i_1,...,i_l}(r^0, r^1) h_{i_1,...,i_l}(r^0), \tag{4.36}$$

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

$$p_{i_1,\ldots,i_l}(r^0,r^1) := \frac{1}{\binom{b}{a}} \prod_{j=1}^{l_0} \binom{r_j^0 + i_j}{i_j} \prod_{j=1}^{l_1} \binom{r_j^1 - 2 + i_{j+l_0}}{i_{j+l_0}}, \tag{4.37}$$

$$h_{i_1,\ldots,i_l}(r^0) := \sum_{j=1}^{l_0} \log \binom{r_j^0 + i_j}{i_j}. \tag{4.38}$$

Proposition 4.2. *The function* $g(a,b)$ *in* (4.34) *satisfies*

$$g(a,b) \le \log \sum_{j=0}^{a} \binom{b}{j} 2^{\max_{x^a : w(x^a)=j} F(x^a)} - \log \binom{b}{a} \tag{4.39}$$

where $F(\cdot)$ *is defined in* (4.36).

Proof. Let P^b denote the random vector describing the positions of the insertions in the output sequence Y^b, such that $P_i = 1$ if and only if Y_i is one of the $b-a$ inserted 0's. We have

$$\begin{aligned} \mathbb{I}(X^a;Y^b) &= H(X^a) - H(X^a|Y^b) \\ &= H(X^a) - H(P^b|Y^b) + H(P^b|X^a,Y^b), \end{aligned} \tag{4.40}$$

where (4.40) follows by the general identity $H(X^a|Y^b) + H(P^b|X^a,Y^b) = H(P^b|Y^b) + H(X^a|P^b,Y^b)$ and noticing that for this choice of P^b, we have $H(X^a|P^b,Y^b) = 0$. For the term $H(P^b|X^a,Y^b)$ in (4.40), we have

$$H(P^b|X^a,Y^b) = \sum_{x^a} P(x^a) \sum_{y^b} P(y^b|x^a) H(P^b|X^a = x^a, Y^b = y^b) \tag{4.41}$$

$$= \sum_{x^a} P(x^a) F(x^a), \tag{4.42}$$

where: $F(\cdot)$ is defined in (4.36); and (4.42) is because instead of the summation over y^b, we can sum over the possible 0 insertions in between the runs of a fixed input sequence x^a such that there are total of $b-a$ insertions. If we denote the number of insertions in between the runs of zeros by $i_1, ..., i_{l_0}$, and the number of insertions in between the runs of ones by $i_{1+l_0}, ..., i_{l_1+l_0}$, then we have $i_1, ..., i_l \in \mathbb{Z}_{\geq 0} : \sum_{j=1}^{l} i_j = b-a$. Given these number of insertions, it is easy to see that $P(y^b|x^a)$ in (4.41) is equal to $p_{i_1,...,i_l}(r^0, r^1)$ in (4.37). Also, $H(P^b|X^a = x^a, Y^b = y^b)$ is equal to $h_{i_1,...,i_l}(r^0)$ in (4.38), 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. Note that from (4.35), we have

$$g(a,b) = \log \sum_{j=0}^{a} 2^{\max_{P(x^a)} \mathbb{I}_j(X^a;Y^b)}, \tag{4.43}$$

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

$$\begin{aligned} H(P^b|Y^b) &= H(Y^b|P^b) + H(P^b) - H(Y^b) \\ &= H(X^a) + \log \binom{b}{a} - H(Y^b), \end{aligned} \tag{4.44}$$

where (4.44) 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 $\log \binom{b}{a}$ due to the uniform insertions. Combining (4.40), (4.42), and (4.44), we have

$$\begin{aligned} \mathbb{I}_j(X^a; Y^b) &= H_j(Y^b) - \log \binom{b}{a} + \sum_{x^a : w(x^a) = j} P(x^a) F(x^a) \\ &\leq \log \binom{b}{j} - \log \binom{b}{a} + \max_{x^a : w(x^a) = j} F(x^a), \end{aligned} \tag{4.45}$$

where: $H_j(Y^b)$ denotes the entropy of the output sequence if it has weight j; and (4.45) follows from the fact that the uniform distribution maximizes the entropy and by maximizing $F(x^a)$ over all input sequences with weight j. Finally, by combining (4.43) and (4.45), we get the upper bound (4.39). $\square$

Similarly, it is possible to obtain a lower bound on the function $g(a, b)$. The lower bound and a numerical comparison between the lower bound, the exact value, and the upper bound on the function $g(a, b)$ can be found in [29]. Although the results of the lower bound are not used throughout the dissertation, we state the result for completeness.

Proposition 4.3. *The function $g(a, b)$ in (4.34) satisfies*

$$g(a, b) \geq \log \sum_{x^a \in \{0,1\}^a} \frac{2^{F(x^a)}}{\binom{b-w(x^a)}{a-w(x^a)}} \tag{4.46}$$

where $F(\cdot)$ is defined in (4.36).

Proof. Consider the term $H(P^b|Y^b)$ in (4.40), we have

$$\begin{aligned} H(P^b|Y^b) &= \sum_{y^b} P(y^b) H(P^b|Y^b = y^b) \\ &\leq \sum_{y^b} P(y^b) \log \binom{b - w(y^b)}{a - w(y^b)} \qquad (4.47) \\ &= \sum_{y^b} \sum_{x^a} P(x^a) P(y^b|x^a) \log \binom{b - w(y^b)}{a - w(y^b)} \\ &= \sum_{x^a} P(x^a) \log \binom{b - w(x^a)}{a - w(x^a)}, \qquad (4.48) \end{aligned}$$

where: (4.47) is because if $y_i = 1$, then it is not an inserted symbol and $p_i = 0$, and therefore, given the output sequence y^b, there are $\binom{b-w(y^b)}{a-w(y^b)}$ possible choices for the position vector P^b, and we can upper bound it by assuming a uniform distribution; and (4.48) is because the weights of the input x^a and output y^b are always the same.

Now, we have

$$g(a,b) = \max_{P(x^a)} \mathbb{I}(X^a;Y^b)$$
$$\geq \max_{P(x^a)} \sum_{x^a} P(x^a) \left[-\log P(x^a) - \log \binom{b-w(x^a)}{a-w(x^a)} + F(x^a) \right] \tag{4.49}$$
$$= \log \sum_{x^a \in \{0,1\}^a} \frac{2^{F(x^a)}}{\binom{b-w(x^a)}{a-w(x^a)}}, \tag{4.50}$$

where: (4.49) follows by combining (4.40), (4.48), and (4.42); and (4.50) is the solution to the optimization problem (4.49). Note that this is a convex optimization problem where the optimal solution can be found to be $P^*(x^a) = D2^{F(x^a)}/\binom{b-w(x^a)}{a-w(x^a)}$ by Karush-Kuhn-Tucker (KKT) conditions [2], where the constant D is obtained such that $\sum_{x^a} P^*(x^a) = 1$, and (4.50) is obtained by substituting $P^*(x^a)$. Therefore, the lower bound (4.46) is proved. □

A numerical comparison between the lower bound, the exact value, and the upper bound on the function $g(a,b)$ can be found in [29].

The following definition will be useful in expressing the upper bounds in Section 4.3.2.

$$\phi(a,b) := a - g(a,b), \tag{4.51}$$

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 4.4. *The functions $g(a,b)$ and $\phi(a,b)$ have the following properties:*

(a) $g(a,b) \leq a$, $\phi(a,b) \geq 0$.

(b) $g(a,a) = a$, $\phi(a,a) = 0$.

(c) $g(1,b) = 1$, $\phi(1,b) = 0$.

(d) $g(a, b+1) \leq g(a, b)$, $\phi(a, b+1) \geq \phi(a, b)$.

(e) $g(a+1, b+1) \leq 1 + g(a, b)$, $\phi(a+1, b+1) \geq \phi(a, b)$.

Proof. We prove the properties for the capacity function $g(a, b)$. The corresponding properties for the function $\phi(a, b)$ easily follows from (4.51).

(a) Since the cardinality of the input alphabet of this channel is 2^a, the capacity of this channel is at most a bits per channel use.

(b) There are no insertions. Therefore, it is a noiseless channel with input and output alphabets of sizes 2^a and capacity a bits per channel use.

(c) The input alphabet is $\{0, 1\}$, and the output consists of binary sequences with length b 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 per channel use.

(d) The capacity $g(a, b+1)$ 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 $g(a, b)$. Therefore, $g(a, b+1) \leq g(a, b)$.

(e) The capacity $g(a+1, b+1)$ 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 $1 + g(a, b)$. Therefore, $g(a+1, b+1) \leq 1 + g(a, b)$.

$\square$

4.3.2 Genie-Aided System and Numerical Upper Bounds

In this section, we focus on upper bounds on the capacity of binary-input binary-output noiseless intermittent communication. The procedure is similar to [18]. 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.

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 iid insertions, the random sequence $\{Z_i\}_{i=1}^{t}$ is iid as well with negative binomial distribution:

$$P(Z_i = b+1) = \binom{b}{s}(1-p_t)^{b-s}p_t^{s+1}, b \geq s, \tag{4.52}$$

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

$$\frac{N}{t} \xrightarrow{p} \mathbb{E}[Z_i] = \frac{s+1}{p_t}. \tag{4.53}$$

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

$$\begin{aligned}
C_1 &= \lim_{k\to\infty} \frac{1}{k} \sum_{i=1}^{t} g(s, Z_i - 1) + 1 \\
&= \lim_{k\to\infty} \frac{N}{k} \frac{t}{N} \frac{1}{t} \sum_{i=1}^{t} g(s, Z_i - 1) + 1 \\
&= \frac{1}{s+1} \lim_{t\to\infty} \frac{1}{t} \sum_{i=1}^{t} g(s, Z_i - 1) + 1 && (4.54) \\
&= \frac{1}{s+1} \mathbb{E}\left[g(s, Z_i - 1) + 1\right] && (4.55) \\
&= \frac{1}{s+1} \left[1 + \sum_{b=s}^{\infty} \binom{b}{s} (1-p_t)^{b-s} p_t^{s+1} g(s,b)\right] && (4.56) \\
&= 1 - \frac{1}{s+1} \sum_{b=s}^{\infty} \binom{b}{s} (1-p_t)^{b-s} p_t^{s+1} \phi(s,b), && (4.57)
\end{aligned}$$

where: (4.54) follows from (4.53); (4.55) follows from the law of large numbers; (4.56) follows from the distribution of Z_i's given in (4.52); and (4.57) follows from the definition (4.51). Note that the capacity C_1 cannot be larger than 1, since the coefficients $\phi(\cdot,\cdot)$ cannot be negative. The negative term in (4.57) can be interpreted as a lower bound on the communication overhead as the cost of intermittency in the context of [32].

The expression in (4.57) 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 (4.57) and using part (d) of Proposition 4.4.

$$C_1 \le 1 - \frac{\phi(s, b_{max})}{s+1} + \frac{1}{s+1} \sum_{b=s}^{b_{max}} \binom{b}{s} p_t^{s+1} (1-p_t)^{b-s} (\phi(s, b_{max}) - \phi(s,b)), \qquad (4.58)$$

The expression (4.58), which we denote by C_1', gives a nontrivial and computable

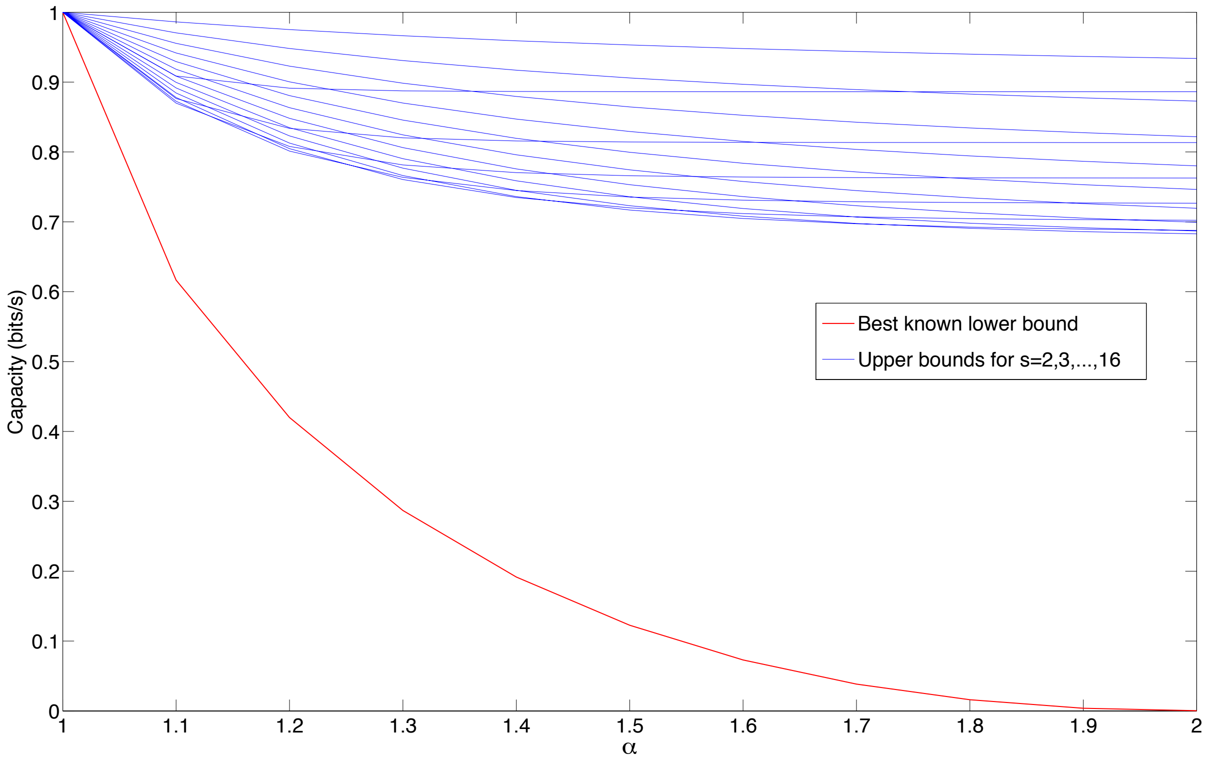

Figure 4.4. Comparison between the best achievability result with different upper bounds obtained from (4.58) for $b_{max} = 17$ and $s = 2, 3, ..., 16$, versus the intermittency rate α.

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 4.4 shows the upper bounds for $b_{max} = 17$ and $s = 2, 3, ..., 16$ versus the intermittency rate α, along with the the achievability result.

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

$$C_2 = 1 - \frac{1}{sp_t} \sum_{a=0}^{s} \binom{s}{a} p_t^a (1-p_t)^{s-a} \phi(a,s). \tag{4.59}$$

Note that the summation in (4.59) 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 4.5 shows the upper bounds for $s = 3, 4, ..., 17$ versus the intermittency rate α, along with the the achievability result. The upper bound corresponding to $s = 17$ is tighter than others for all ranges of α, i.e., (4.59) 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 (4.59) gives better upper bounds for the range of α 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

$$\begin{aligned} \lim_{\alpha\to\infty} C_1' &= 1 - \frac{\phi(s, b_{max})}{s+1}, \\ \lim_{\alpha\to\infty} C_2 &= 1. \end{aligned} \tag{4.60}$$

This is because of the fact that by increasing α, 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 (4.60) is 0.6739 bits per channel use. In principle, we can use the upper bound on $g(a, b)$ in Proposition 4.2 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 (4.39) for larger arguments. It seems that this does not improve the upper bounds significantly for the range of α shown in the figures. However, by

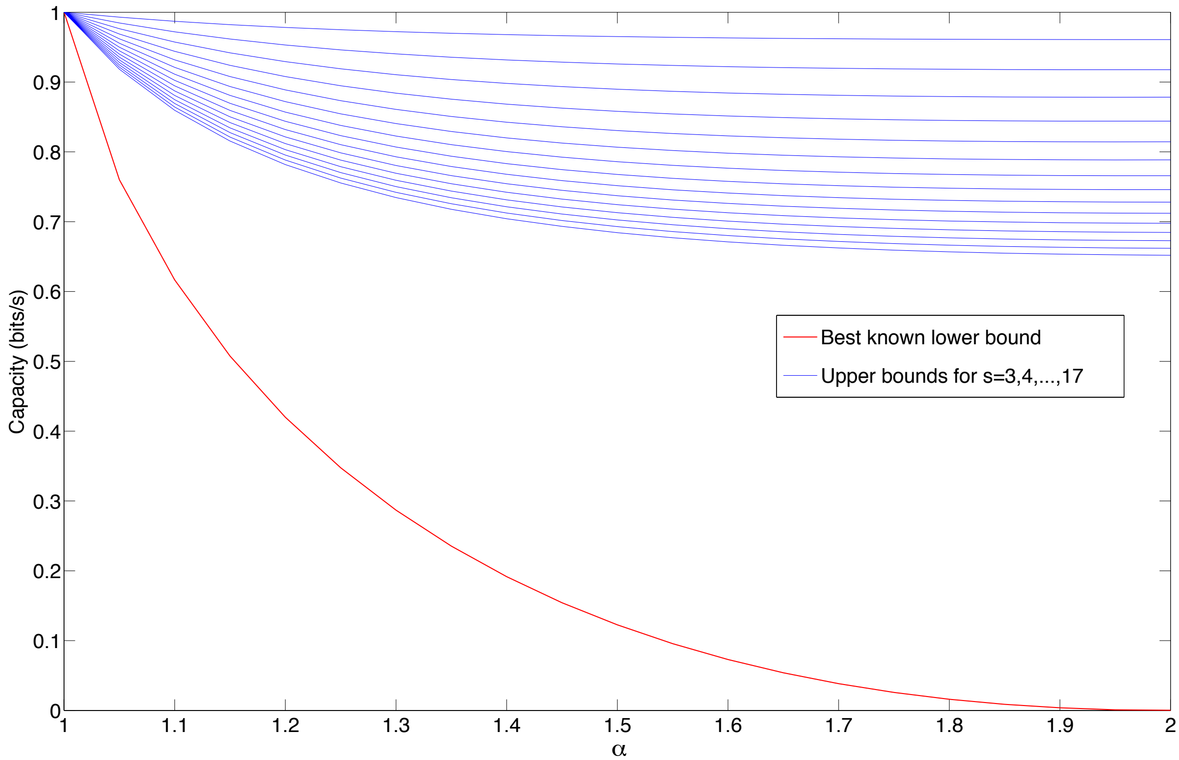

Figure 4.5. Comparison between the best achievability result with different upper bounds obtained from (4.59) for $s = 3, 4, ..., 17$, versus the intermittency rate α.

upper bounding (4.60) via (4.39), we can tighten the upper bound for the limiting case of $\alpha \to \infty$ to 0.6307 bits per channel use.

Although the gap between the achievable rates and upper bounds is not particularly tight, especially for large values of intermittency rate α, 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 [52, 54], where the transmission of the codeword is contiguous, only exponential scaling $n = e^{\alpha k}$ induces a tradeoff between capacity and asynchronism.

4.4 Bounds on Capacity Per Unit Cost

In this section, we obtain bounds on the capacity per unit cost of intermittent communication. Let $\gamma : \mathcal{X} \to [0, \infty]$ be a cost function that assigns a non-negative value to each channel input. We assume that the noise symbol has zero cost, i.e., $\gamma(\star) = 0$. The cost of a codeword is defined as

$$\Gamma(c^k(m)) = \sum_{i=1}^{k} \gamma(c_i(m)).$$

A (k, M, P, ε) code consists of M codewords of length k, $c^k(m), m \in [1 : M]$, each having cost at most P with average probability of decoding error at most ε, where the intermittent process is the same as in Section 4.1. Note that the cost of the input and output sequences of the intermittent process shown in Figure 4.1 is the same since the cost of the noise symbols is zero. We say rate $\hat{R}$ bits per unit cost is achievable if for every $\varepsilon > 0$ and large enough M there exists a (k, M, P, ε) code with $\log(M)/P \geq \hat{R}$. For intermittent communication $(\mathcal{X}, \mathcal{Y}, W, \star, \alpha)$, the capacity per unit cost $\hat{C}_\alpha$ is the supremum of achievable rates per unit cost.

It is shown in [58] that the capacity per unit cost of a general DMC is

$$\max_{x \in \mathcal{X} \setminus \{\star\}} \frac{D(W_x \| W_\star)}{\gamma(x)},$$

where we assume that $\gamma(\star) = 0$, and the optimization is over the input alphabet instead of over the set of all input distributions. The asynchronous capacity per unit cost for asynchronous communication with timing uncertainty per information bit β has been shown to be [4]

$$\frac{1}{1+\beta} \max_{x \in \mathcal{X} \setminus \{\star\}} \frac{D(W_x \| W_\star)}{\gamma(x)}.$$

Therefore, comparing to the capacity per unit cost of a DMC, the rate is penalized by a factor of $1/(1+\beta)$ due to asynchronism. For a channel with iid synchronization errors with average number of duplications equal to μ concatenated with a DMC, bounds on the capacity per unit cost, $\hat{C}_\mu$, have been obtained in [24] as

$$\frac{\mu}{2} \max_{x \in \mathcal{X} \setminus \{\star\}} \frac{D(W_x \| W_\star)}{\gamma(x)} \leq \hat{C}_\mu \leq \mu \max_{x \in \mathcal{X} \setminus \{\star\}} \frac{D(W_x \| W_\star)}{\gamma(x)},$$

where the lower bound is obtained by using a type of pulse position modulation at the encoder and searching for the position of the pulse at the decoder. Using similar encoding and decoding schemes, we obtain a lower bound for the capacity per unit cost of intermittent communication.

Theorem 4.5. *The capacity per unit cost $\hat{C}_\alpha$ for intermittent communication $(\mathcal{X}, \mathcal{Y}, W, \star, \alpha)$ satisfies*

$$\frac{\alpha}{2} \max_{x \in \mathcal{X} \setminus \{\star\}} \frac{D(\frac{1}{\alpha} W_x + (1 - \frac{1}{\alpha}) W_\star \| W_\star)}{\gamma(x)} \leq \hat{C}_\alpha \leq \max_{x \in \mathcal{X} \setminus \{\star\}} \frac{D(W_x \| W_\star)}{\gamma(x)} \tag{4.61}$$

Sketch of the Proof: The upper bound in (4.61) is the capacity per unit cost of the DMC W, and follows by providing the decoder with side-information about the positions of inserted noise symbols $\star$. The derivation of the lower bound is similar to the one in [24, Theorem 3]. Essentially, the encoder uses pulse position modulation, i.e., to transmit message m, it transmits a burst of symbols x of length

$$B := \frac{2 \log(M)}{\alpha D(\frac{1}{\alpha} W_x + (1 - \frac{1}{\alpha}) W_\star \| W_\star)},$$

at a position corresponding to this message and transmits the zero-cost noise symbol $\star$ at the other $k - B$ positions before and after this burst, so that each codeword has cost $P = B\gamma(x)$. In order to decode the message, we search for the location of the pulse using a sliding window with an appropriate length looking for a subsequence that has a type equal to $\frac{1}{\alpha}W_x + (1-\frac{1}{\alpha})W_\star$, because at the receiver, we expect to have approximately $B(\alpha-1)$ inserted noise symbols $\star$ in between the B burst symbols x. Similar to the analysis of [24, Theorem 3], it can be shown that the probability of decoding error vanishes as $M \to \infty$, and the rate per unit cost is

$$\hat{R} = \frac{\log(M)}{P} = \frac{\alpha}{2}\frac{D(\frac{1}{\alpha}W_x + (1-\frac{1}{\alpha})W_\star \| W_\star)}{\gamma(x)}.$$

Finally, by choosing the optimum input symbol x, rate per unit cost equal to the left-hand side of (4.61) can be achieved. $\square$

From the convexity of the Kullback-Leibler divergence, it can be seen that the lower bound is always smaller than half of the upper bound. Consider the BSC example with crossover probability $p = 0.1$ and input costs $\gamma(\star = 0) = 0$ and $\gamma(1) = 1$. The upper bound in (4.61) equals 2.536 bits per unit cost, and the lower bound in (4.61) is plotted in Figure 4.6 versus the intermittency rate α. As we would expect, the lower bound decreases as the intermittency increases.

4.5 Summary

In this chapter, we formulated a model for intermittent communication that can capture bursty transmissions or a sporadically available channel by inserting a random number of silent symbols between each codeword symbol so that the receiver does not know a priori when the transmissions will occur. First, we specified two decoding structures in order to develop achievable rates. Interestingly, decoding from pattern detection, which achieves a larger rate, is based on a generalization of the method of types and properties of partial divergence. As the system becomes more intermittent, the achievable rates decrease due to the additional uncertainty about the positions of the codeword symbols at the decoder. We also showed that as long as the intermittency rate α is finite and the capacity of the DMC is not zero, rate

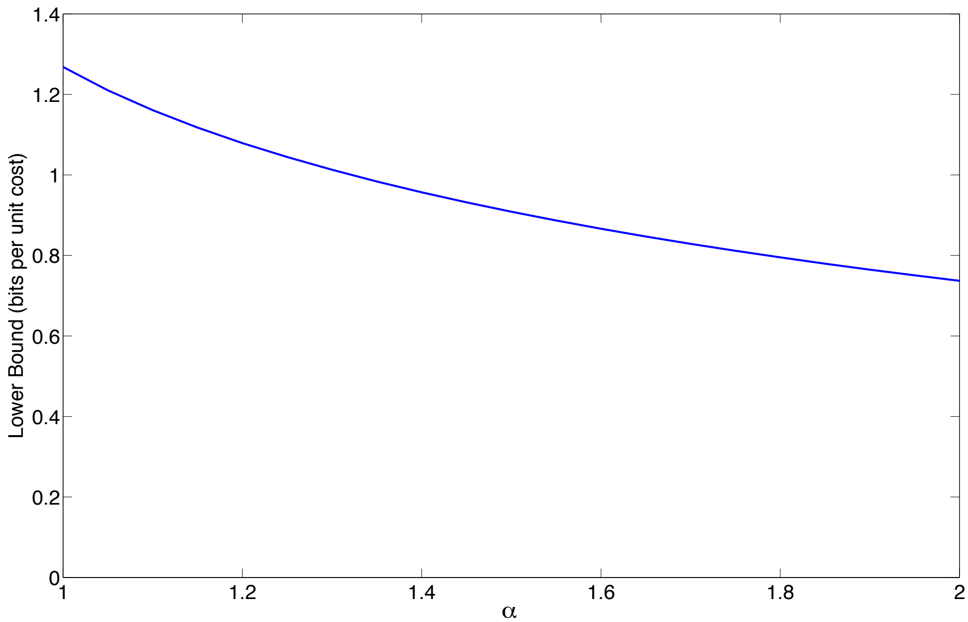

Figure 4.6. The lower bound on the capacity per unit cost of intermittent communication versus the intermittency rate α.

$R = 0$ is achievable for intermittent communication. For the case of binary-input binary-output noiseless channel, we obtained upper bounds on the capacity of intermittent communication 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 and the intermittency rate. Finally, we derived bounds on the capacity per unit cost of intermittent communication. To obtain the lower bound, we used pulse-position modulation at the encoder, and searched for the position of the pulse at the decoder.

CHAPTER 5

MULTI-USER INTERMITTENT COMMUNICATION

This chapter can be viewed as another attempt to combine the information-theoretic and network-oriented multi-access models, as discussed in Section 2.5, and to characterize the performance of the system in terms of the achievable rate regions. We formulate a model for intermittent multi-access communication for two users that captures two network-oriented concepts. First, it models bursty transmission of the codeword symbols for each user. Second, it takes into account the possible asynchronism between the receiver and the transmitters as well as between the transmitters themselves.

A basic system model is introduced in Section 5.1, which generalizes the intermittent communication model introduced in Chapter 4. By making different assumptions for the intermittent process, we specialize the system to three models: random access with no idle-times and no collisions, random access with idle-times and no collisions, and random access with collisions and no idle-times in Sections 5.2, 5.3, and 5.4, respectively. Random access that allows for both idle-times and collisions can be handled similarly, but is removed in order to avoid further complexity. In Section 5.5, we study a simple example, and in Section 5.6, we point out some connections and differences between intermittent multi-access communication and the other models in the literature we reviewed in Section 2.5.

For each model, we obtain achievable rate regions that depend on the concept of partial divergence introduced in Chapter 3. The collisions are treated as interference, and information can be extracted from the collided symbols. Because of the assumption that the receiver does not know a priori that an output symbol corresponds to transmission by a given user, neither user, or both users, the decoder has to both detect the positions and decode the messages.

5.1 System Model

We consider a 2-user discrete memoryless multiple access channel (DM-MAC) with conditional probability mass functions $W(y|x_1, x_2)$ over input alphabets $\mathcal{X}_1$ and $\mathcal{X}_2$ and output alphabet $\mathcal{Y}$. The two senders wish to communicate independent messages $m_1 \in [1 : e^{kR_1} = M_1]$ and $m_2 \in [1 : e^{kR_2} = M_2]$ to a receiver. Let $\star \in \mathcal{X}_1, \mathcal{X}_2$ denote a special symbol, corresponding to the input of the channel when the sender is silent. Let $W_{\cdot\star} := W(y|x_1, x_2 = \star)$ denote the probability transition matrix for the point to point channel for user 1 if user 2 is silent, let $W_{\star\cdot}$ be defined analogously, and let $W_{\star\star} := W(y|x_1 = \star, x_2 = \star)$ denote the output distribution if both users are silent. Each user encodes its message into a codeword of length k: $c_1^k(m_1)$ and $c_2^k(m_2)$ denote the codewords of user 1 and user 2, respectively. Assume that x_1^n and x_2^n are the input sequences and y^n is the output sequence of the channel, where n is the length of the receive window at the decoder.

Figure 5.1 shows a block diagram for the system model in which the intermittent process stores inputs $c_1^k(m_1)$ and $c_2^k(m_2)$ in two separate buffers, and generates outputs x_1^n and x_2^n to capture the burstiness and the asynchronism of the users. The intermittent process, in general, has memory, and can be described as a state-dependent process with four possible states (s_1, s_2), $s_1, s_2 \in \{0, 1\}$ in each time slot. If $s_i = 0$, then user i is silent and transmits the symbol $\star$. If $s_i = 1$, which is only possible if there are codeword symbols remaining in user i's buffer, then user i transmits the next codeword symbol. We assume that neither the encoders nor the decoder know the states of the intermittent process. Note that the intermittent process together with the DM-MAC can be collected into a state-dependent MAC with memory with

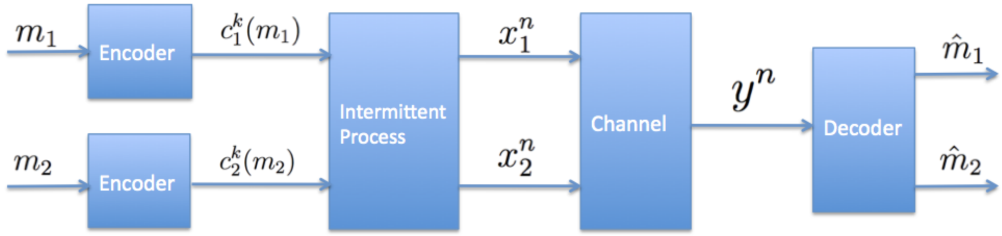

Figure 5.1. System model for intermittent multi-access communication.

the states unknown to the encoders and the decoder. See [23, 51] and the references therein for some treatment of memoryless state-dependent MAC.

Denoting the decoded messages by $\hat{m}_1$ and $\hat{m}_2$, which are functions of the random sequence Y^n, we say that the rate pair (R_1, R_2) is achievable if there exists two sequences of length k codes of sizes $M_1 = e^{kR_1}$ and $M_2 = e^{kR_2}$ for the two encoders with average probability of error $\frac{1}{M_1 M_2} \sum_{m_1=1}^{M_1} \sum_{m_2=1}^{M_2} \mathbb{P}((\hat{m}_1, \hat{m}_2) \neq (m_1, m_2)) \to 0$ as $k \to \infty$. We refer to this general scenario as intermittent multi-access communication, and in Sections 5.2, 5.3, and 5.4, we consider several instances of the intermittent process in Figure 5.1. We have introduced some of these models in [30].

5.2 Random Access: No Idle-Times and No Collisions

In this section, we consider an intermittent process in Figure 5.1 that models a random access channel in which, at each time slot, exactly one of the users sends a codeword symbol and the other remains silent by sending the special symbol $\star$, until both users have finished sending their codewords. In this model, there are only two possible states for the intermittent process $(s_1, s_2) \in \{(1, 0), (0, 1)\}$, and therefore, the output pair (x_1, x_2) of the intermittent process at each time slot takes one of the two following forms: $(c_1, \star)$ or $(\star, c_2)$, where c_1 and c_2 denote the next codeword symbol to be transmitted from the first and the second user, respectively. Note that if both input buffers of the intermittent process are empty, then the transmission terminates, and if exactly one of them is empty, then only the state corresponding to transmission of the codeword symbol from the user with the non-empty buffer is allowed. As a result, the length of the receive window in this model is $n = 2k$. The receiver observes the sequence y^n, wishes to decode both messages, but does not know a priori which output symbol corresponds to which user's codeword.

A potential application of this model include a cognitive radio in which the primary user is bursty, i.e., sends codeword symbols in some time slots and remains silent in the other time slots, and a secondary user also wants to communicate with the same receiver and can sense the channel and transmit its codeword symbols whenever the first user is silent.

In the following theorem, we obtain an achievable rate region for (R_1, R_2).

Theorem 5.1. *For intermittent multi-access communication with no idle-times and no collisions, rates (R_1, R_2) satisfying*

$$R_1 < \mathbb{I}(X_1; Y | X_2 = \star) - f_1(P_1, P_2, W) \tag{5.1}$$

$$R_2 < \mathbb{I}(X_2; Y | X_1 = \star) - f_1(P_1, P_2, W) \tag{5.2}$$

are achievable for any $(X_1, X_2) \sim P_1(x_1)P_2(x_2)$, where

$$f_1(P_1, P_2, W) := \max_{0 \le \beta \le 1} \{2h(\beta) - d_\beta(P_1 W_{\cdot\star} \| P_2 W_{\star\cdot}) - d_\beta(P_2 W_{\star\cdot} \| P_1 W_{\cdot\star})\}, \tag{5.3}$$

and $d_\cdot(\cdot\|\cdot)$ is the partial divergence introduced in Chapter 3.

Proof. Encoding: Fix two input distributions P_1 and P_2 for user 1 and user 2, respectively. Randomly and independently generate e^{kR_1} sequences $c_1^k(m_1)$, $m_1 \in [1 : e^{kR_1}]$ each iid according to P_1 for user 1, and e^{kR_2} sequences $c_2^k(m_2)$, $m_2 \in [1 : e^{kR_2}]$ each iid according to P_2 for user 2. To send message m_1, encoder 1 transmits $c_1^k(m_1)$, and to send message m_2, encoder 2 transmits $c_2^k(m_2)$.

Decoding: Similar to decoding from pattern detection described in Chapter 4, the decoder chooses k of the $2k$ output symbols y^{2k}. Let $\tilde{y}_1^k$ denote the sequence of chosen symbols, and $\tilde{y}_2^k$ denote the other k symbols. For each choice, there are two stages. In the first stage, the decoder checks if $\tilde{y}_1^k$ is induced by user 1, i.e., if $\tilde{y}_1^k \in T_{P_1 W_{\cdot\star}}$, and if $\tilde{y}_2^k$ is induced by user 2, i.e., if $\tilde{y}_2^k \in T_{P_2 W_{\star\cdot}}$. If both of these conditions are satisfied, then we proceed to the second stage; otherwise, we make another choice for the k symbols and restart the two-stage decoding procedure. In the second stage, we perform joint typicality decoding with a fixed typicality parameter $\mu > 0$ for both sequences $\tilde{y}_1^k$ and $\tilde{y}_2^k$, i.e., if $\tilde{y}_1^k \in T_{[W_{\cdot\star}]_\mu}(c_1^k(\hat{m}_1))$ and $\tilde{y}_2^k \in T_{[W_{\star\cdot}]_\mu}(c_2^k(\hat{m}_2))$ for a unique message pair $(\hat{m}_1, \hat{m}_2)$, then we declare them as the transmitted messages; otherwise, we make another choice for the k symbols and repeat the two-stage decoding procedure. If at the end of all $\binom{2k}{k}$ choices the typicality decoding procedure has not declared any message pair as being sent, then the decoder declares an error.

Analysis of the probability of error: For any $\epsilon > 0$, we prove that if $R_1 = \mathbb{I}(X_1; Y | X_2 = \star) - f_1(P_1, P_2, W) - 2\epsilon$, and $R_2 = \mathbb{I}(X_2; Y | X_1 = \star) - f_1(P_1, P_2, W) - 2\epsilon$, then the average probability of error vanishes as $k \to \infty$. Considering independent uniform distributions on the messages and assuming that the message pair $(1,1)$ is transmitted, we have

$$\begin{aligned} p_e^{avg} \leq & \mathbb{P}((\hat{m}_1, \hat{m}_2) = e | (m_1, m_2) = (1,1)) \\ & + \mathbb{P}(\hat{m}_1 \in \{2, 3, ..., e^{kR_1}\} | (m_1, m_2) = (1,1)) \\ & + \mathbb{P}(\hat{m}_2 \in \{2, 3, ..., e^{kR_2}\} | (m_1, m_2) = (1,1)), \end{aligned} \tag{5.4}$$

where (5.4) follows from the union bound in which the first term is the probability that the decoder declares an error (does not find any message pair) at the end of all $\binom{2k}{k}$ choices, which implies that even if we pick the correct output symbols corresponding to user 1 and user 2, the decoder either does not pass the first stage or does not declare $(\hat{m}_1, \hat{m}_2) = (1,1)$ in the second stage. The probability of this event vanishes as $k \to \infty$ according to Lemma 2.2.

The second term in (5.4) is the probability that for at least one choice of the output symbols, the decoder passes the first stage, and then in the second stage, it declares an incorrect message for user 1. We characterize the $\binom{2k}{k}$ choices based on the number of incorrectly chosen output symbols, which is denoted by k_1, i.e., the number of symbols in $\tilde{y}_1^k$ that are in fact output symbols corresponding to the second user, which is equal to the number of symbols in $\tilde{y}_2^k$ that are in fact output symbols corresponding to the first user. For any $0 \le k_1 \le k$, there are $\binom{k}{k_1}\binom{k}{k_1}$ possible choices. Using the union bound for all the choices and all the messages $\hat{m}_1 \neq 1$, we have

$$\begin{aligned}
&\mathbb{P}(\hat{m}_1 \in \{2,3,...,e^{kR_1}\}|(m_1,m_2)=(1,1)) \\
&\le (e^{kR_1}-1)\sum_{k_1=0}^{k}\binom{k}{k_1}\binom{k}{k_1}\mathbb{P}_{k_1}(\hat{m}_1 = 2|(m_1,m_2)=(1,1)),
\end{aligned} \tag{5.5}$$

where the index k_1 in (5.5) denotes the condition that the number of incorrectly chosen output symbols is k_1. Note that message $\hat{m}_1 = 2$ is declared at the decoder only if the choice of the output symbols passes the first stage, and then the condition $\tilde{y}_1^k \in T_{[W\cdot\star]_\mu}(c_1^k(2))$ is satisfied. Therefore,

$$\begin{aligned}
&\mathbb{P}_{k_1}(\hat{m}_1 = 2|(m_1,m_2)=(1,1)) \\
&= \mathbb{P}_{k_1}\Big(\{\tilde{Y}_1^k \in T_{P_1W\cdot\star}\} \cap \{\tilde{Y}_2^k \in T_{P_2W_\star\cdot}\} \cap \{\tilde{Y}_1^k \in T_{[W\cdot\star]_\mu}(c_1^k(2))\}|(m_1,m_2)=(1,1)\Big) \\
&= \mathbb{P}_{k_1}(\tilde{Y}_1^k \in T_{P_1W\cdot\star}) \cdot \mathbb{P}_{k_1}(\tilde{Y}_2^k \in T_{P_2W_\star\cdot}) \cdot \mathbb{P}(\tilde{Y}_1^k \in T_{[W\cdot\star]_\mu}(c_1^k(2))|(m_1,m_2)=(1,1))
\end{aligned} \tag{5.6}$$

$$\le e^{o(k)} e^{-kd_{k_1/k}(P_1W\cdot\star\|P_2W_\star\cdot)} e^{-kd_{k_1/k}(P_2W_\star\cdot\|P_1W\cdot\star)} \cdot e^{-k(\mathbb{I}(X_1;Y|X_2=\star)-\epsilon)}, \tag{5.7}$$

where: (5.6) follows from the independence of the events $\{\tilde{Y}_1^k \in T_{P_1 W_{\cdot\star}}\}$ and $\{\tilde{Y}_2^k \in T_{P_2 W_{\star\cdot}}\}$ conditioned on k_1 (a fixed number of) incorrectly chosen output symbols; and (5.7) follows from the results on the partial divergence in Chapter 3 for the first two terms in (5.6) with mismatch ratios k_1/k, and using Lemma 2.3 for the last term in (5.6), because conditioned on message $m_1 = 1$ being sent, $C_1^k(2)$ and $\tilde{Y}_1^k$ are independent regardless of the number of incorrectly chosen output symbols.

Substituting (5.7) into the summation in (5.5), using Stirling's approximation for the terms $\binom{k}{k_1}$, and finding the largest exponent of the terms in the summation, we have

$$
\begin{aligned}
&\mathbb{P}(\hat{m}_1 \in \{2, 3, ..., e^{kR_1}\} | (m_1, m_2) = (1, 1)) \\
&\leq e^{kR_1} e^{o(k)} e^{k f_1(P_1, P_2, W)} e^{-k(\mathbb{I}(X_1; Y | X_2 = \star) - \epsilon)} \\
&= e^{o(k)} e^{-k\epsilon},
\end{aligned} \tag{5.8}
$$

where (5.8) is obtained by substituting $R_1 = \mathbb{I}(X_1; Y | X_2 = \star) - f_1(P_1, P_2, W) - 2\epsilon$. Therefore, the second term in (5.4) vanishes as $k \to \infty$. Similarly, the third term in (5.4) also vanishes as $k \to \infty$, which proves the theorem. $\square$

Remark 5.1. *The result in Theorem 5.1 is valid for the intermittent process described above with arbitrary probability distribution on the time slots that each user transmits.*

Example 5.1. *As a special case, we might think of an intermittent process in which at each time slot $\mathbb{P}((S_1, S_2) = (1, 0)) = \mathbb{P}((S_1, S_2) = (0, 1)) = 1/2$ if both buffers are non-empty; otherwise only the user with the non-empty buffer transmits. Note that the length of the receive window remains $2k$ in any case, since each codeword has length k and there is neither idle-times nor collisions.*

The function $f_1(P_1, P_2, W)$ can be interpreted as an overhead term due to the system's burstiness or intermittency. Note that the result in Theorem 5.1 implies that there is a tradeoff between the two terms in (5.1) and in (5.2) by choosing the input distributions P_1 and P_2. In order to maximize the first terms we need to choose the capacity achieving input distributions, but at the same time, it is desirable to choose input distributions such that the two distributions $P_1W_{\cdot\star}$ and $P_2W_{\star\cdot}$ have the largest distance to maximize the partial divergences $d_\beta(P_1W_{\cdot\star}\|P_2W_{\star\cdot})$ and $d_\beta(P_2W_{\star\cdot}\|P_1W_{\cdot\star})$ so that we have a smaller overhead term $f_1(P_1, P_2, W)$. Also, note that both rates R_1 and R_2 have the same overhead cost for fixed input distributions P_1 and P_2. This is no longer the case if we consider different codeword lengths for the two users.

5.3 Random Access: With Idle-Times and No Collisions

In this section, we consider an intermittent process in Figure 5.1 that models a random access channel in which at each time slot either one of the users sends a codeword symbol and the other one remains silent or both remain silent until both users send their codewords. In this model, there are three possible states for the intermittent process $(s_1, s_2) \in \{(1,0),(0,1),(0,0)\}$, and therefore, the output pair (x_1, x_2) of the intermittent process at each time slot takes one of the three following forms: $(c_1, \star)$, $(\star, c_2)$, or $(\star, \star)$. The length of the receive window in this model is $n \geq 2k$, and we assume that there are $n - 2k$ idle-times $(\star, \star)$, where $\theta := (n-2k)/k$ shows the ratio of the idle-times to the codeword length. The receiver observes the sequence y^n, but does not know a priori that an output symbol corresponds to which of the three possible input pairs, and wishes to decode both messages.

Practical examples include an ALOHA random access protocol with a collision-avoidance mechanism in which at each time either only one of the users transmits or the channel remains idle.

In the following theorem, we obtain an achievable rate region for (R_1, R_2).

Theorem 5.2. *For intermittent multi-access communication with idle-times and no* *collisions rates* (R_1, R_2) *satisfying*

$$R_1 < \mathbb{I}(X_1; Y | X_2 = \star) - f_2(P_1, P_2, W, \theta)$$

$$R_2 < \mathbb{I}(X_2; Y | X_1 = \star) - f_2(P_1, P_2, W, \theta)$$

are achievable for any $(X_1, X_2) \sim P_1(x_1)P_2(x_2)$, *where*

$$\begin{aligned} & f_2(P_1, P_2, W, \theta) := \\ & \max_{\substack{0 \le \beta_1 + \beta_2 \le 1 \\ 0 \le \beta_1' + \beta_2' \le 1}} \Big\{ h(\beta_1, \beta_2) + h(\beta_1', \beta_2') + \theta h(\frac{\beta_1 + \beta_2 - \beta_1'}{\theta}, \frac{\beta_1' + \beta_2' - \beta_1}{\theta}) \\ & \qquad - d_{\beta_1, \beta_2}(P_1 W_{\cdot\star} \| W_{\star\star}, P_2 W_{\star\cdot}) - d_{\beta_1', \beta_2'}(P_2 W_{\star\cdot} \| W_{\star\star}, P_1 W_{\cdot\star}) \\ & \qquad - \theta d_{(\beta_1 + \beta_2 - \beta_1')/\theta, (\beta_1' + \beta_2' - \beta_1)/\theta}(W_{\star\star} \| P_1 W_{\cdot\star}, P_2 W_{\star\cdot}) \Big\}, \end{aligned} \tag{5.9}$$

where $d_{\cdot,\cdot}(\cdot\|\cdot,\cdot)$ *is the generalized partial divergence function defined in Chapter 3.*

Proof. Encoding is similar to the encoding in the proof of Theorem 5.1. The decoder splits the output sequence y^n into three subsequences of length k, k, and $n-2k$, and denotes them by $\tilde{y}_1^k$, $\tilde{y}_2^k$, and $\hat{y}^{n-2k}$, respectively. For each choice, there are two stages. In the first stage, we check three conditions: $\tilde{y}_1^k \in T_{P_1 W_{\cdot\star}}$, $\tilde{y}_2^k \in T_{P_2 W_{\star\cdot}}$, and $\hat{y}^{n-2k} \in T_{W_{\star\star}}$. If all three conditions are satisfied, then we proceed to the second stage; otherwise, we make another choice for the three output subsequences and restart the two-stage decoding procedure. The second stage is similar to the one in the proof of Theorem 5.1; if a unique message pair $(\hat{m}_1, \hat{m}_2)$ passes the joint typicality test, then we declare them as the transmitted messages; otherwise, we make another choice and repeat the two-stage decoding procedure. If at the end of all $\binom{n}{k,k,n-2k}$ choices the typicality decoding procedure has not declared any message pair as being sent, then the decoder declares an error.

Analysis of the probability of error: Similar to the proof of Theorem 5.1, the first term in (5.4) vanishes as $k \to \infty$. The second term in (5.4) is the probability that for at least one choice of the output subsequences, the decoder passes the first stage,

Decoder's choice	From user 1	From user 2	From idle-times
$\tilde{y}_1^k$	$k - k_1 - k_2$	k_1	k_2
$\tilde{y}_2^k$	k_1'	$k - k_1' - k_2'$	k_2'
$\hat{y}^{n-2k}$	$k_1 + k_2 - k_1'$	$k_1' + k_2' - k_1$	$n - 2k - k_2 - k_2'$

Figure 5.2. Characterizing the number of symbols in $\tilde{y}_1^k$, $\tilde{y}_2^k$, and $\hat{y}^{n-2k}$ from the first user, the second user, and idle times.

and then in the second stage, it declares an incorrect message for user 1.

We characterize the $\binom{n}{k,k,n-2k}$ choices based on the number of incorrectly chosen output symbols, which is denoted by k_1, k_2, k_1', and k_2', where k_1 and k_2 are the number of symbols in $\tilde{y}_1^k$ that are incorrectly chosen from the second user and the idle-times, respectively (note that $\tilde{y}_1^k$ is supposed to contain symbols from the first user only). Similarly, k_1' and k_2' are the number of symbols in $\tilde{y}_2^k$ that are incorrectly chosen from the first user and the idle-times, respectively. Note that $0 \le k_1 + k_2 \le k$ and $0 \le k_1' + k_2' \le k$, and the number of symbols in $\hat{y}^{n-2k}$ that are incorrectly chosen from the first and the second users would be uniquely determined. Figure 5.2 summarizes this division.

Note that for any k_1, k_2, k_1', and k_2', there are

$$\binom{k}{k_1, k_2, k-k_1-k_2}\binom{k}{k_1', k_2', k-k_1'-k_2'}\binom{n-2k}{k_1+k_2-k_1', k_1'+k_2'-k_1, n-2k-k_2-k_2'}$$

possible choices. Using the union bound for all the choices and all the messages $\hat{m}_1 \neq 1$, we can bound $\mathbb{P}(\hat{m}_1 \in \{2,3,...,e^{kR_1}\}|(m_1,m_2) = (1,1))$ in a similar way as we did in the proof of Theorem 5.1. Using Stirling's approximation, generalized partial divergence for three distributions introduced in Chapter 3, Lemma 2.3, and following similar steps as in the proof of Theorem 5.1, we can see that the probability of error vanishes as $k \to \infty$, which proves the theorem. $\square$

Remark 5.2. *The result in Theorem 5.2 is valid for the intermittent process described* *above with arbitrary probability distribution on the time slots that each user transmits as long as the number of idle-times $n - 2k$ is fixed. Furthermore, the result remains valid if the number of idle-times is a random variable denoted by E, such that the ratio of the idle-times to the codeword length converges, i.e., $E/k \xrightarrow{p} \theta$ as $k \to \infty$.*

Example 5.2. *As a special case, we might think of the following intermittent process: If the length of the buffers are equal, then $\mathbb{P}((S_1, S_2) = (0,0)) = \theta/(\theta+1)$ and $\mathbb{P}((S_1, S_2) = (1,0)) = \mathbb{P}((S_1, S_2) = (0,1)) = 1/(2\theta+2)$; otherwise the user with more symbols in its buffer transmits. Note that in this example, the length of the receive window is a random variable $N = 2k + E$, but $E/k \xrightarrow{p} \theta$ as $k \to \infty$.*

5.4 Random Access: With Collisions and No Idle-Times

In this section, we consider an intermittent process in Figure 5.1 that models a random access channel with collisions. In principle, we can consider a random access channel that allows for both idle-times and collisions using a similar approach. However, we assume that there are no idle times in this section in order to avoid overcomplicating the results. In this model, there are three possible states for the intermittent process $(s_1, s_2) \in \{(1,0), (0,1), (1,1)\}$, where the total number of states representing a collision, i.e., $(s_1, s_2) = (1,1)$, is assumed to be $d \leq k$. Therefore, the output pair (x_1, x_2) of the intermittent process with length $n = 2k - d$ consists of $k - d$ of the form $(c_1, \star)$, $k - d$ of the form $(\star, c_2)$, and d of the form (c_1, c_2). In other words, user 1 and user 2 transmit $k - d$ codeword symbols over a point to point channel, $W_{\cdot,\star}$ and $W_{\star,\cdot}$, respectively, and transmit d codeword symbols over the MAC channel W, through which there is interference between the users, but the decoder does not know a priori these positions. Let $\theta := d/k \leq 1$ denote the ratio of the collided symbols of each user to the codeword length.

In the following theorem, we obtain an achievable rate region for (R_1, R_2).

Theorem 5.3. *For intermittent multi-access communication with collisions and no idle-times, rates (R_1, R_2) satisfying*

$$\begin{aligned} R_1 &< \bar{\theta}\mathbb{I}(X_1;Y|X_2=\star)+\theta\mathbb{I}(X_1;Y|X_2)-f_3(P_1,P_2,W,\theta) \\ R_2 &< \bar{\theta}\mathbb{I}(X_2;Y|X_1=\star)+\theta\mathbb{I}(X_2;Y|X_1)-f_3(P_1,P_2,W,\theta) \\ R_1+R_2 &< \bar{\theta}\mathbb{I}(X_1;Y|X_2=\star)+\bar{\theta}\mathbb{I}(X_2;Y|X_1=\star)+\theta\mathbb{I}(X_1,X_2;Y)-f_3(P_1,P_2,W,\theta) \end{aligned}$$

are achievable for any $(X_1, X_2) \sim P_1(x_1)P_2(x_2)$, where

$$\begin{aligned} & f_3(P_1,P_2,W,\theta) := \\ & \max_{\substack{0\le\beta_1+\beta_2\le1 \\ 0\le\beta_1'+\beta_2'\le1}} \Big\{ \bar{\theta}h(\beta_1,\beta_2)+\bar{\theta}h(\beta_1',\beta_2')+\theta h\Big(\frac{\bar{\theta}(\beta_1+\beta_2-\beta_1')}{\theta},\frac{\bar{\theta}(\beta_1'+\beta_2'-\beta_1)}{\theta}\Big) \\ & \qquad -\bar{\theta}d_{\beta_1,\beta_2}(P_1W_{\cdot\star}\|P_1P_2W,P_2W_{\star\cdot})-\bar{\theta}d_{\beta_1',\beta_2'}(P_2W_{\star\cdot}\|P_1P_2W,P_1W_{\cdot\star}) \\ & \qquad -\theta d_{(\beta_1+\beta_2-\beta_1')\bar{\theta}/\theta,(\beta_1'+\beta_2'-\beta_1)\bar{\theta}/\theta}(P_1P_2W\|P_1W_{\cdot\star},P_2W_{\star\cdot})\Big\}, \end{aligned} \tag{5.10}$$

and $d_{\cdot,\cdot}(\cdot\|\cdot,\cdot)$ is the generalized partial divergence function defined in Chapter 3.

Sketch of the Proof: Encoding is the same as in the proof of Theorem 5.1. We briefly explain the decoding procedure. The analysis of the probability of error is lengthy, but similar to the previous ones, and is omitted here.

Decoding: The decoder splits the output sequence y^{2k-d} into three subsequences of length $k-d$, $k-d$, and d, and denotes them by $\tilde{y}_1^{k-d}$, $\tilde{y}_2^{k-d}$, and $\hat{y}^d$, respectively. For each choice, there are two stages. In the first stage, we check three conditions: $\tilde{y}_1^{k-d} \in T_{P_1W_{\cdot\star}}$, $\tilde{y}_2^{k-d} \in T_{P_2W_{\star\cdot}}$, and $\hat{y}^d \in T_{P_1P_2W}$. If all three conditions are satisfied, then we proceed to the second stage; otherwise, we make another choice for the three output subsequences and restart the two-stage decoding procedure.

In the second stage, we perform simultaneous joint typicality decoding. We first split all of the codewords as follows. Let $\tilde{c}_1^{k-d}(m_1)$ and $\hat{c}_1^d(m_1)$ be the subsequences of $c_1^k(m_1)$ corresponding to the positions of the symbols of the chosen subsequences

$\tilde{y}_1^{k-d}$ and $\hat{y}^d$, respectively. Similarly, let $\tilde{c}_2^{k-d}(m_2)$ and $\hat{c}_2^d(m_2)$ be the subsequences of $c_2^k(m_2)$ corresponding to the positions of the symbols of the chosen subsequences $\tilde{y}_2^{k-d}$ and $\hat{y}^d$, respectively. We declare the message pair $(\hat{m}_1, \hat{m}_2)$ as being transmitted if it is the unique message pair such that the following three conditions are satisfied simultaneously: $(\tilde{c}_1^{k-d}(\hat{m}_1), \tilde{y}_1^{k-d})$ is jointly typical, $(\tilde{c}_2^{k-d}(\hat{m}_2), \tilde{y}_2^{k-d})$ is jointly typical, and $(\hat{c}_1^d(\hat{m}_1), \hat{c}_2^d(\hat{m}_2), \hat{y}^d)$ is jointly typical; otherwise, we make another choice for the three output subsequences and repeat the two-stage decoding procedure. If at the end of all $\binom{2k-d}{k-d,k-d,d}$ choices the typicality decoding procedure has not declared any message pair as being sent, then the decoder declares an error. $\square$

Remark 5.3. *The result in Theorem 5.3 is valid for the intermittent process described above with arbitrary probability distribution on the time slots that each user transmits as long as the number of collided symbols d is fixed. Furthermore, the result remains valid if the number of collided symbols is a random variable denoted by D, such that the ratio of the collided symbols to the codeword length converges, i.e., $D/k \xrightarrow{p} \theta$ as $k \to \infty$.*

Example 5.3. *As a special case, we might think of the following intermittent process: If the length of the buffers are equal, then $\mathbb{P}((S_1, S_2) = (1, 1)) = \theta$ and $\mathbb{P}((S_1, S_2) = (1, 0)) = \mathbb{P}((S_1, S_2) = (0, 1)) = (1 - \theta)/2$; otherwise only the user with more symbols in its buffer transmits. Note that in this example, the length of the receive window is a random variable $N = 2k - D$, but $D/k \xrightarrow{p} \theta$ as $k \to \infty$.*

5.5 A Simple Example

Consider a DM-MAC with $\mathcal{X}_1, \mathcal{X}_2 = \{0, 1, 2, 3\}$ and $\mathcal{Y} = \{0, 1, ..., 6\}$ such that $Y = X_1 + X_2$, where $+$ corresponds to real addition. The capacity region of this channel is shown with the blue curve in Figure 5.3. The red dots correspond to achievable rates (R_1, R_2) for intermittent MAC with no idle-times and no collisions obtained from Theorem 5.1 using different input distributions $P_1(x_1)$ and $P_2(x_2)$.

For simplicity, we only focus on the result of Theorem 5.1. Not surprisingly, the plot suggests that the intermittency of the system and lack of knowledge about the position of the symbols at the decoder come with a significant cost. We should mention that achieving the rate pairs shown by points A and B in the figure is surprisingly simple. In order to achieve point A, we use $P_1(x_1) = [0, 1/3, 1/3, 1/3]$ and $P_2(x_2) = [1, 0, 0, 0]$, and to achieve point B, we use $P_1(x_1) = [0, 0, 1/2, 1/2]$ and $P_2(x_2) = [1/2, 1/2, 0, 0]$. In both cases, the overhead function $f_1(P_1, P_2, W)$ in Theorem 5.1 evaluates to zero, since the distributions $P_1W_{\cdot\star}$ and $P_2W_{\star\cdot}$ become disjoint and the partial divergence terms become infinite. It is also worth pointing out that the achievable rate region for the intermittent MAC model does not have to be convex, as can be seen from the figure, because time sharing is not possible due to the intermittency and asynchronism of the system.

5.6 Connections to Related Works

In Section 2.5, we summarized the system models and results of some of the works that give an information-theoretic model for multi-access communication focusing on some of the network-oriented concepts, such as asynchronism, random access, and collision. There are some similarities and differences between the models and results reviewed in Section 2.5 and those of intermittent MAC, which we briefly discuss.

In the models for intermittent MAC, frame asynchronous MAC, and collision channels, there are uncertainties about the positions of the codeword symbols at the receiver, the transmitters do not know this information and can never learn it, and the receiver does not know this information a priori. The transmission of the users' codewords is contiguous in the frame asynchronous MAC model, whereas it can be bursty in both intermittent MAC and the collision channel models. Furthermore, the decoding is performed across all these bursty symbols/packets in both intermittent

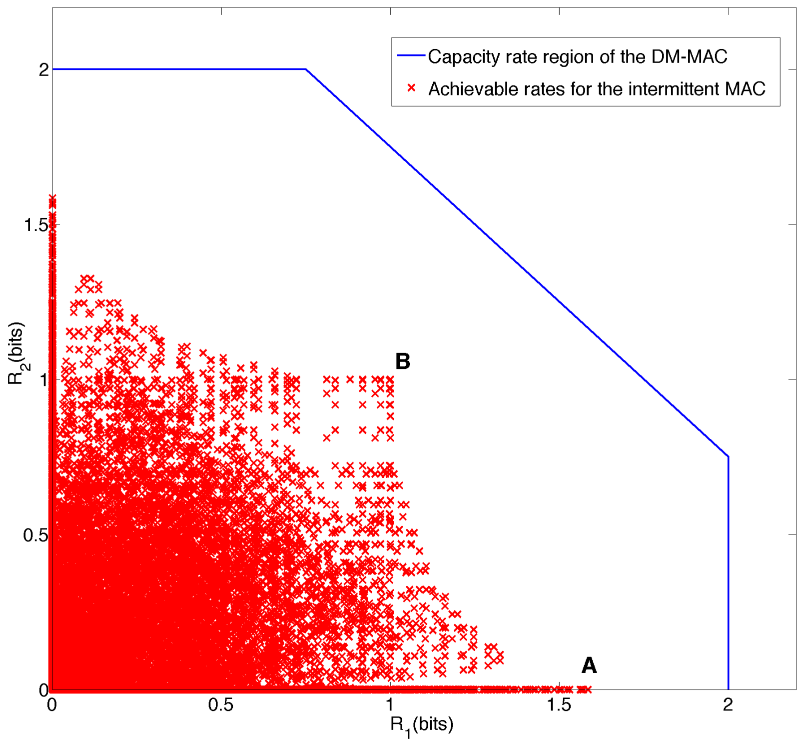

Figure 5.3. Comparing the capacity rate region of the DM-MAC with the achievable rates for intermittent MAC with no idle-times and no collisions obtained from Theorem 5.1.

MAC and the collision channel models. Also, in intermittent MAC, the receiver differentiates the output symbols based on their empirical distributions, whereas in the collision channel model, the receiver differentiates the output symbol based on its knowledge on the protocol sequences and the positions relative to the idle symbol Λ and the collision symbol Δ.

In the model for random access [40], the states of the users (active or inactive) are similar to those of the model for intermittent MAC, but remain constant for the entire communication block and are known to the receiver. As a result, unlike the intermittent MAC, decoding is not performed across different states in this model for random access, and there is no uncertainty about the states of the users at the receiver. In other words, in the intermittent MAC model, random accessing is considered in the symbol/packet level, whereas in the model of [40], random accessing is considered in the block level. However, in both [40] and intermittent MAC with collisions, some part of information is decoded in the absence of interference and some part in the presence of interference.

Finally, similar to the capacity region in the models for frame asynchronous MAC, random access in [34], and collision channels, the achievable rate region for intermittent MAC can be non-convex, as we have seen in Section 5.5. This is because of the impossibility of time sharing and coordination in the time domain in these models. Furthermore, similar to the capacity rate region of the frame asynchronous MAC with memory in [57], the achievable rate region of intermittent MAC is drastically reduced due to the lack of coordination in the time domain, as we have seen in Section 5.5.

5.7 Summary

In this chapter, we formulated a model for intermittent multi-access communication for two users that captures the bursty transmission of the codeword symbols for each user and the possible asynchronism between the receiver and the transmitters as well as between the transmitters themselves. By making different assumptions for the intermittent process, we specialized the system to a random access system with or without idle-times and collisions. For each model, we characterized the performance of the system in terms of achievable rate regions. In our achievable schemes, the intermittency of the system comes with a significant cost, i.e., it reduces the size of the achievable rate regions, which can be interpreted as communication overhead [15]. Note that as opposed to [15], where the constraint is the lack of coordination between the users in multi-access communication, the constraint in our problem is the intermittency of the system.

CHAPTER 6

OTHER EXTENSIONS

In this chapter, we study three relevant extensions. First, inspired by network applications, we extend the model to packet-level intermittent communication in which codeword and noise symbols are grouped into packets. Next, we use some of the insights and tools developed in this dissertation to obtain some new results on the capacity of deletion channels and a random access that drops the collided symbols. Finally, inspired by the problem of file synchronization, we obtain some results on compressing a source sequence with the presence of decoder side-information that is related to the source via an intermittent process.

6.1 Packet-Level Intermittent Communication

In this section, we introduce a system model for packet-level intermittent communication in which codeword and noise symbols are grouped into packets of length l, noise packets are inserted in the input sequence of the channel, and the receiver does not know a priori the positions of the codeword 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. The system model in this chapter recovers intermittent communication introduced in Chapter 4 if the packets correspond to a single symbol, i.e., $l = 1$, and recovers asynchronous communication reviewed in Section 2.3 if the packet corresponds to the codeword, i.e., contiguous transmission of codeword symbols.

Packet-level intermittent communication may arise in network applications in which the data is packetized, and each packet goes through a network with intermittent connectivity and / or random delay. The receiver's task is to identify the packets in order to reconstruct the original data. The intermittency of the system and the size of the packet might allow for individual packet detection at the receiver, in which case the communication problem simplifies to individual packet detection and message decoding. However, if the intermittency of the system increases or the packet length decreases, then we may not reliably detect all the data packets individually, and encoding and decoding across packets could be a better solution. In this section, we refer to the former case as large-packet intermittent communication, and to the latter case as small- or medium-packet intermittent communication.

After introducing a general system model in Section 6.1.1, we will specialize it to small-packet, medium-packet, and large-packet intermittent communication in Sections 6.1.2, 6.1.3, and 6.1.4, respectively. For small-packet and medium-packet intermittent communication, we use similar decoding structures as in Chapter 4, namely, decoding from exhaustive search and decoding from pattern detection, whereas for large-packet intermittent communication, we introduce a new decoding structure, which is called *decoding from packet detection*.

6.1.1 System Model

As before, a transmitter communicates a single message $m \in [1 : e^{kR} = M]$ to a receiver over a DMC with probability transition matrix W and input and output alphabets $\mathcal{X}$ and $\mathcal{Y}$, with the noise symbol denoted $\star \in \mathcal{X}$. Also, let C_W be the capacity of this DMC. The transmitter encodes the message as a codeword $c^k(m)$ of length k. The intermittent process takes $c^k(m)$ as an input sequence and outputs x^n. Assume that x^n and y^n are the input and output sequences of the channel W, respectively, where $n \geq k$ is the length of the receive window at the decoder. Note

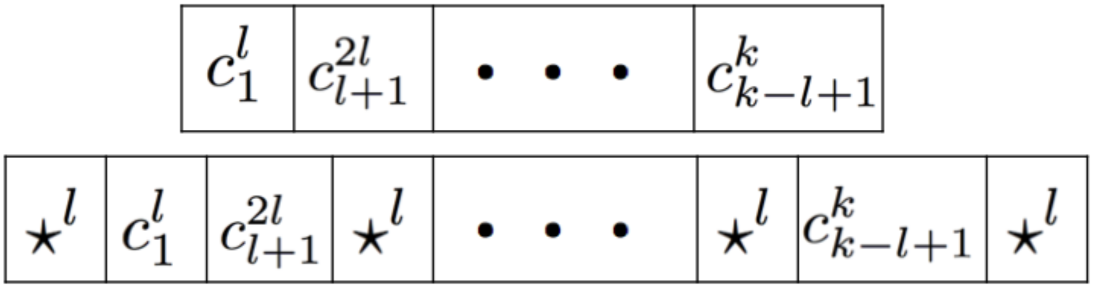

Figure 6.1. Illustration of the input and output of the intermittent process for the packet-level intermittent communication.

that unlike to the model in Chapter 4 with the random length of the receive window, we assume that the length of the receive window is fixed in this section in order to be able to unify the model for different scaling behavior we consider for packet-level intermittent communication.

The intermittent process operates at the packet level, which is illustrated in Figure 6.1 and can be described as follows: First, the codeword symbols $c^k(m)$ are grouped into packets of length l, resulting in a total of k/l codeword packets. Then, $(n-k)/l$ noise packets, i.e., sequences of noise symbols $\star$ of length l, are inserted arbitrarily between the codeword packets, i.e., the output of the intermittent process, x^n, contains a total of n/l packets consisting of the k/l codeword packets and $(n-k)/l$ noise packets, and packets are received in the right order. Analogous to symbol-level intermittent communication, we assume that the transmitter cannot decide on the locations of the codeword packets, so it cannot encode any timing information, and the receiver does not know a priori the positions of the codeword packets.

Generally, increasing the packet length makes the transmission of the codeword symbols more contiguous, and decreases the uncertainty about their locations at the receiver. This observation suggests that the scaling behavior of the receive window relative to the codeword length should be determined based on the scaling behavior of the packet length relative to the codeword length in order to identify the regimes of interest. In this section, we consider three different scaling behaviors for the packet length l relative to the codeword length k, identify the corresponding regimes of interest for the scaling behavior for the receive window n relative to k, and define the associated communication scenarios.

Definition 6.1. (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 6.2. (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 6.3. (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 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 6.1 and 6.2 is the necessary condition that $n \geq k$. Note that in all of the communication scenarios defined above, α determines the scaling of the receive window relative to the codeword length (or the packet length), even though the scaling behavior itself depends on the scenario. In all three cases, the larger the value of α, the larger the receive window, and therefore, the more intermittent the system becomes. Hence, α is called the intermittency rate as in Chapter 4. As we will see, increasing α generally reduces the achievable rate R for each of the three considered scenarios, because it makes the receive window larger, and therefore, increases the uncertainty about the positions of the codeword 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 Chapter 4, and the special case of $\lambda = 1$ (or $l = k$) for large-packet intermittent communication recovers the model and results for slotted asynchronous communication [60]. In the former case, as we saw in Chapter 4, 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 achieved, and the intermittency does not impact the communication rate.

Next, we develop achievable rates for the three scenarios defined above. For small- and medium-packet intermittent communication, we utilize both decoding from exhaustive search and decoding from pattern detection introduced in Chapter 4 in order to obtain achievable rates. However, for large-packet intermittent communication, a new decoding structure, decoding from packet detection, is introduced and an achievable rate is obtained.

The encoding structure is as before: Given an input distribution P, the codebook is randomly and independently generated, i.e., all $C_i(m), i \in [1:k], m \in [1:M]$ are iid according to P. For ease of presentation, let $s := k/l$ denote the number of codeword 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 $b - s$.

6.1.2 Small-Packet Intermittent Communication

In this section, we obtain two achievable rates for small-packet intermittent communication. For obtaining the first achievable rate, we utilize decoding from exhaustive search, which is the same as described in Section 4.2.2, but instead of choosing the k symbols out of the n symbols from the output sequence at each step, we choose s packets out of the b output packets, so that the total number of choices becomes $\binom{b}{s}$. We have the following achievability result.

Theorem 6.1. *For small-packet intermittent communication with parameters l and α, rates less than $(C_W - \alpha h(1/\alpha)/l)^+$ are achievable.*

Proof. The proof is similar to the proof of Theorem 4.3, but instead of (4.10), we have

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

and instead of (4.11), we have

$$\binom{b}{s} \leq \frac{e^{\frac{1}{12}}}{\sqrt{2\pi}}\sqrt{\frac{b}{s(b-s)}}e^{bh(s/b)} \doteq e^{k\alpha h(1/\alpha)/l}, \text{ as } k \to \infty, \tag{6.1}$$

where (6.1) is obtained by substituting $s = k/l$, $b = n/l$, and $n/k = \alpha$. □

As before, we can interpret the term $\alpha h(1/\alpha)/l$ as the overhead cost due to the intermittency. Note that the overhead cost is increasing in the intermittency rate α, is equal to zero 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 codeword packets at the decoder yielding a better achievability result.

Using decoding from pattern detection introduced in Chapter 4, with the modification that instead of choosing symbols we choose packets, for the small-packet intermittent communication model and the results on partial divergence developed in Chapter 3, we obtain the following achievability result.

Theorem 6.2. *For small-packet intermittent communication with parameters l and α, rates less than* $\max_P\{\left(\mathbb{I}(P,W) - f_l^{SP}(P,W,\alpha)\right)^+\}$ *are achievable, where*

$$f_l^{SP}(P,W,\alpha) := \max_{0\leq\beta\leq 1}\left\{\frac{(\alpha-1)h(\beta)+h((\alpha-1)\beta)}{l} - d_{(\alpha-1)\beta}(PW\|W_\star) - (\alpha-1)d_\beta(W_\star\|PW)\right\}. \tag{6.2}$$

Proof. The proof is similar to the proof of Theorem 4.4, but instead of (4.17), we have

$$\mathbb{P}(\hat{m} \in \{2,3,...,M\}|m=1) \leq (e^{kR}-1)\sum_{k_1=0}^{b-s}\binom{s}{k_1}\binom{b-s}{k_1}\mathbb{P}_{k_1}(\hat{m}=2|m=1),$$

where the index k_1 denotes the condition that the number of wrongly chosen output packets is equal to k_1, $0 \leq k_1 \leq b-s$. Using the same steps as in the proof of Theorem 4.4, substituting $s = k/l$, $b = n/l$, and $n/k = \alpha$, applying Stirling's approximation, and invoking Lemma 3.1, the result follows. □

The achievable rate in Theorem 6.2 is larger than the one in Theorem 6.1, because decoding from pattern detection utilizes the fact that the choice of the codeword 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 6.2, 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 6.1. *The overhead cost $f_l^{SP}(P, W, \alpha)$ in (6.2) has the same properties as in Proposition 4.1. In addition, $f_l^{SP}(P, W, \alpha) \leq f_1^{SP}(P, W, \alpha)/l$ for all integers $l \geq 1$, and the overhead cost is decreasing in l.*

The proof is similar to the proof of Proposition 4.1. From Proposition 6.1, we can make the same conclusions for the achievable rate in Theorem 6.2 as we made before for the achievable rate in Theorem 4.4. In addition, Proposition 6.1 implies that increasing the packet length l increases the achievable rate, and if the packet length is sufficiently large, then the achievable rate approaches the capacity of the channel.

As a different example than the one considered in Chapter 4 for the channel W, we consider a DMC with a symmetric transition matrix with input and output alphabets of size 4 as is depicted in Figure 6.2, in which all the crossover 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 Chapter 3 is more apparent for a channel with larger alphabet size.

The boundary of the achievable rate region (R, α) characterizes the tradeoff between the achievable rates and the intermittency rate α. Figure 6.3 illustrates the achievable rate region (R, α) for small-packet intermittent communication over the channel depicted in Figure 6.2 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 6.1 and 6.2, indicated by R_1 and R_2, respectively, for two values of the packet length: $l = 1$ and $l = 2$. The achievable rates are decreasing in the inter-

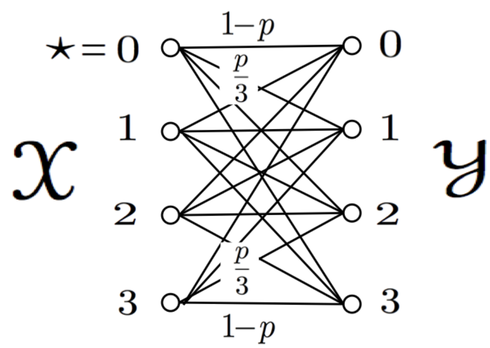

Figure 6.2. Graphical description of the transition matrix for the DMC we consider in this section.

mittency rate α, which is because increasing α increases size of the receive window and therefore the uncertainty about the positions of the codeword 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 longer packets, because increasing l adds more structure to the output sequence and reduces the uncertainty about the positions of the codeword symbols at the receiver.

The arrows in Figure 6.3 highlight the differences between R_2 and R_1, 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.

6.1.3 Medium-Packet Intermittent Communication

Using decoding from exhaustive search introduced in Chapter 4 for medium-packet intermittent communication, we obtain the following achievability result.

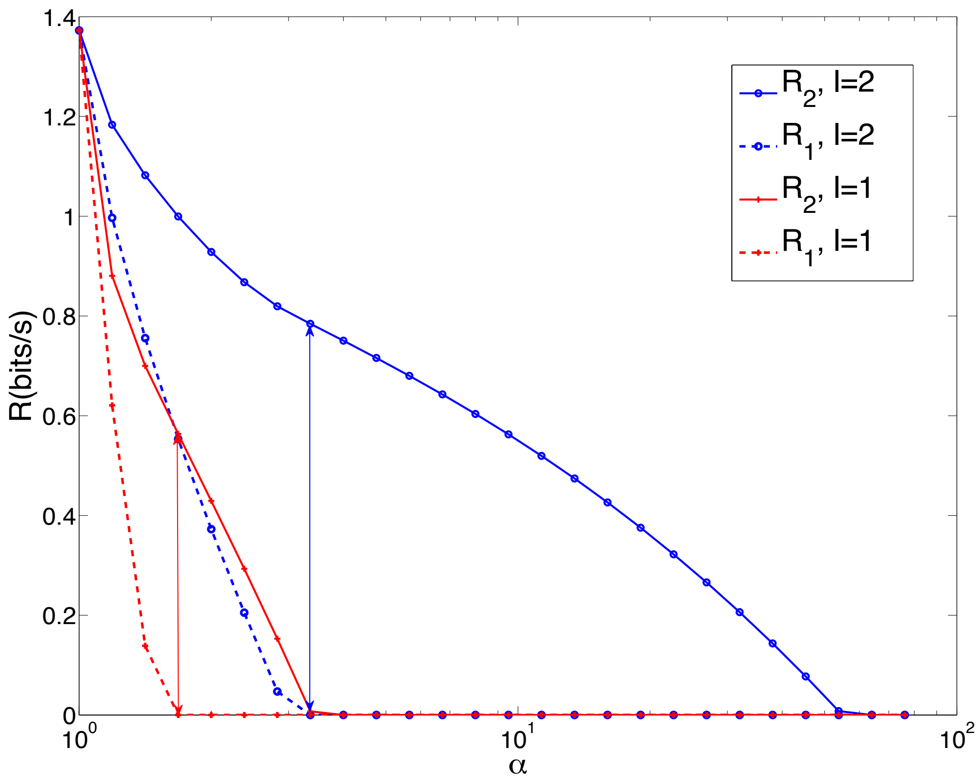

Figure 6.3. Achievable rate region (R, α) for small-packet intermittent communication over the channel depicted in Figure 6.2 with $p = 0.1$.

Theorem 6.3. *For medium-packet intermittent communication with parameters λ and α, rates less than $(C_W - (\alpha - 1)/\lambda)^+$ are achievable.*

Proof. The proof is similar to the proof of Theorem 6.1, except that instead of (6.1), we have

$$\binom{b}{s} \dot{\leq} e^{k(\alpha-1)/\lambda}, \text{ as } k \to \infty,$$

because by substituting $s = k/l = k/(\lambda \log k)$ and $b = n/l = k^{\alpha}$, we have

$$\begin{aligned}
\lim_{k\to\infty} \frac{1}{k} \log \binom{b}{s} &= \lim_{k\to\infty} \frac{1}{k} \log \binom{k^{\alpha}}{k/(\lambda \log k)} \\
&= \lim_{k\to\infty} k^{\alpha-1} h(\frac{1}{\lambda k^{\alpha-1} \log k}) \qquad (6.3) \\
&= \lim_{k\to\infty} \frac{\log(\lambda k^{\alpha-1} \log k)}{\lambda \log k} \qquad (6.4) \\
&= \frac{\alpha-1}{\lambda},
\end{aligned}$$

where (6.3) follows from (4.11), and (6.4) follows from expanding the binary entropy function and using the fact that $\log(1-x) \sim -x$ as $x \to 0$. □

Here, the overhead cost is $(\alpha-1)/\lambda$, which is increasing in the intermittency rate α and decreasing in λ, and similar conclusions as in Section 6.1.2 can be drawn.

Using decoding from pattern detection introduced in Chapter 4 for medium-packet intermittent communication and the results on partial divergence developed in Chapter 3, we obtain the following achievability result.

Theorem 6.4. *For medium-packet intermittent communication with parameters λ and α, rates less than $\max_P\{\left(\mathbb{I}(P,W) - f_{\lambda}^{MP}(P,W,\alpha)\right)^+\}$ are achievable, where*

$$f_{\lambda}^{MP}(P,W,\alpha) := \max_{0\le\beta\le1} \left\{\beta\frac{\alpha-1}{\lambda} - d_{\beta}(PW\|W_{\star})\right\}. \qquad (6.5)$$

Proof. The proof is similar to the proof of Theorem 6.2, except that the summation on k_1 is from 0 to s instead of $b-s$. Note that in general, for the number of incorrectly chosen output packets, k_1, we have $0 \le k_1 \le \min\{s, b-s\}$, but for medium-packet intermittent communication $0 \le k_1 \le s$. Therefore, we have a similar expression to (4.24) except that the maximization is over $k_1 = 0, 1, ..., s$. We proceed by letting $\beta := k_1/s$ $(0 \le \beta \le 1)$, substituting $s = k/(\lambda \log k)$ and $b = k^{\alpha}$, and calculating the four terms in the maximization:

$$\lim_{k\to\infty}\frac{1}{k}sh(\beta)=\lim_{k\to\infty}\frac{1}{\lambda\log k}h(\beta)=0 \tag{6.6}$$

$$\begin{aligned}&\lim_{k\to\infty}\frac{1}{k}(b-s)h(\frac{\beta s}{b-s})\\&=\lim_{k\to\infty}\frac{\lambda k^{\alpha-1}\log k-1}{\lambda\log k}h(\frac{\beta}{\lambda k^{\alpha-1}\log k-1})\\&=\beta\frac{\alpha-1}{\lambda}\end{aligned} \tag{6.7}$$

$$\lim_{k\to\infty}\frac{1}{k}kd_{\frac{\beta sl}{k}}(PW\|W_\star)=d_\beta(PW\|W_\star) \tag{6.8}$$

$$\lim_{k\to\infty}\frac{1}{k}(n-k)d_{\frac{\beta sl}{n-k}}(W_\star\|PW)=\lim_{\rho\to 0}\frac{1}{\rho}d_{\beta\rho}(W_\star\|PW) \tag{6.9}$$

$$=\beta d_0'(W_\star\|PW)=0, \tag{6.10}$$

where : (6.7) follows from expanding the binary entropy function and using the fact that $\log(1-x)\sim -x$ as $x\to 0$; (6.9) follows by substituting $\rho:=\frac{k}{n-k}\to 0$ as $k\to\infty$; and (6.10) follows from Proposition 3.2 (d).

Now, by substituting (6.6), (6.7), (6.8), and (6.10) in the maximization, we have

$$\begin{aligned}&\lim_{k\to\infty}\frac{1}{k}\log\sum_{k_1=0}^{s}\binom{s}{k_1}\binom{b-s}{k_1}e^{-kd_{k_1l/k}(PW\|W_\star)-(n-k)d_{k_1l/(n-k)}(W_\star\|PW)}\\&=\max_{0\le\beta\le 1}\{\beta\frac{\alpha-1}{\lambda}-d_\beta(PW\|W_\star)\}\\&=f_\lambda^{MP}(P,W,\alpha),\end{aligned}$$

and the rest of the proof is the same as the proof of Theorem 4.4. □

For the same reason as in Section 6.1.2, the achievable rate in Theorem 6.4 is larger than the one in Theorem 6.3. The overhead cost in Theorem 6.4 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 (6.5). Similar conclusions as in Section 6.1.2 can be drawn.

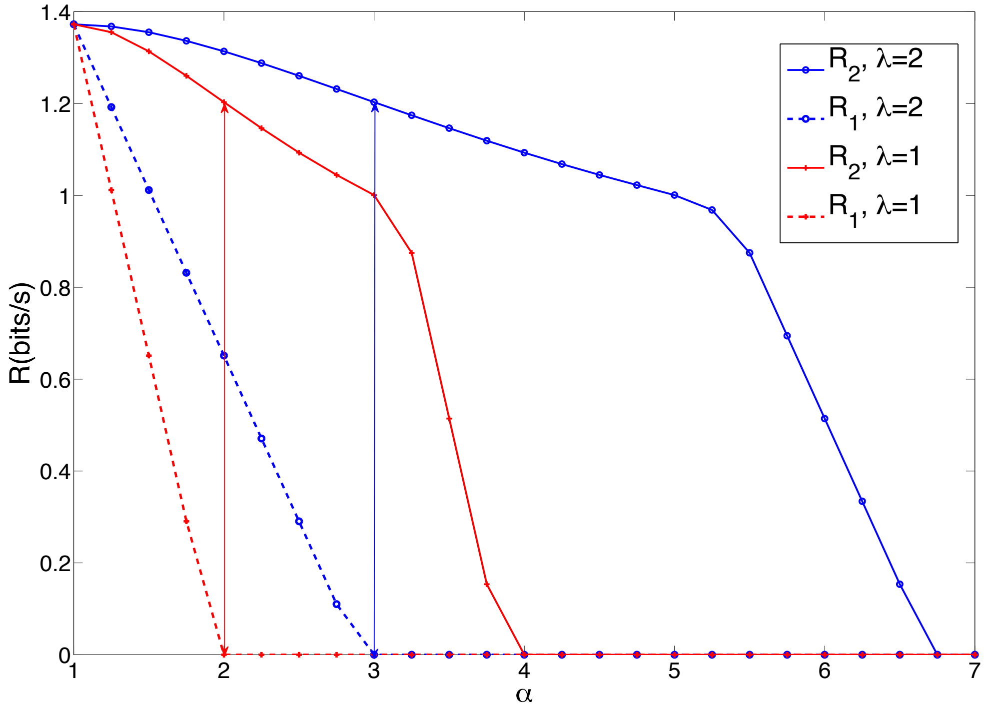

Figure 6.4. Achievable rate region (R, α) for medium-packet intermittent communication over the channel depicted in Figure 6.2 with $p = 0.1$.

Figure 6.4 illustrates the achievable rate region (R, α) for medium-packet intermittent communication model over the channel depicted in Figure 6.2 with $p = 0.1$. The achievable rate regions corresponds to the results in Theorems 6.3 and 6.4, indicated by R_1 and R_2, respectively, for two values of λ: $\lambda = 1$ and $\lambda = 2$. Similar observations and conclusions can be made as those for Figure 6.3.

6.1.4 Large-Packet Intermittent Communication

In order to obtain tight achievable rates for large-packet intermittent communication, we introduce the following decoding structure.

Decoding from packet detection: This structure consists of two separate stages. In the first stage, the decoder completely locates the s codeword packets by checking if the i^{th} packet of the output sequence denoted by $y_{(i-1)l+1}^{il}$ is a codeword 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 codeword packets. In the second stage, the decoder forms a sequence consisting of all the detected codeword 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 codeword packets correctly. Obviously, this structure does not lead to an achievability result if the packet length is finite as in small-packet intermittent communication. As we will see in Remark 6.1, it also turns out that this structure does not work for medium-packet intermittent communication. Therefore, decoding from packet detection is considered only for large-packet intermittent communication.

Using decoding from packet detection for large-packet intermittent communication, we obtain the following achievability result.

Theorem 6.5. *For large-packet intermittent communication with parameters* λ *and* α*, rates not exceeding* $\max_P\{\left(\mathbb{I}(P, W) - f^{LP}(P, W, \alpha)\right)^+\}$ *are achievable, where*

$$f^{LP}(P, W, \alpha) := (\alpha - D(PW\|W_\star))^+ . \tag{6.11}$$

Proof. It is equivalent to prove that rates not exceeding $\max_{P:D(PW\|W_\star)\geq\alpha} \mathbb{I}(X;Y)$

are achievable, because

$$
\begin{aligned}
&\max_P \{\mathbb{I}(X;Y) - f^{LP}(P,W,\alpha)\} \\
&= \max\left\{ \max_{P:D(PW\|W_\star)>\alpha} \mathbb{I}(X;Y), \max_{P:D(PW\|W_\star)\leq\alpha} \{\mathbb{I}(X;Y) + D(PW\|W_\star) - \alpha\} \right\} && (6.12) \\
&= \max\left\{ \max_{P:D(PW\|W_\star)>\alpha} \mathbb{I}(X;Y), \max_{P:D(PW\|W_\star)=\alpha} \mathbb{I}(X;Y) \right\} && (6.13) \\
&= \max_{P:D(PW\|W_\star)\geq\alpha} \mathbb{I}(X;Y),
\end{aligned}
$$

where (6.13) follows from the fact that $\mathbb{I}(X;Y) + D(PW\|W_\star) - \alpha$ is linear in P and $\{P : D(PW\|W_\star) \leq \alpha\}$ is a convex set, and therefore, the second maximization in (6.12) is achieved at the boundary $D(PW\|W_\star) = \alpha$.

Now we prove that for any $\epsilon > 0$ and a fixed input distribution P such that $D(PW\|W_\star) \geq \alpha + \epsilon$, if $R = \mathbb{I}(X;Y) - \epsilon$, then the average probability of error using decoding from packet detection introduced earlier vanishes as $k \to \infty$. To that end, let $\mathcal{E}_1$ denote the union of all events in which at least one noise packet is detected as a codeword packet in the first stage of the decoding structure, and $\mathcal{E}_2$ denote the union of all events in which at least one codeword packet is not detected. Applying the union bound, we have

$$
p_e^{avg} \leq \mathbb{P}(\mathcal{E}_1) + \mathbb{P}(\mathcal{E}_2) + \mathbb{P}(\hat{m} \neq 1 | m = 1, \mathcal{E}_1^C, \mathcal{E}_2^C). \tag{6.14}
$$

The last term in (6.14) is the probability that the decoder does not find the right message in the second stage of the decoding structure after it correctly detects the s codeword packets in the first stage, which vanishes as $k \to \infty$ for all $R < \mathbb{I}(X;Y)$ by the conventional coding theorem. As for the first term in (6.14), we have

$$\mathbb{P}(\mathcal{E}_1) \leq (b-s)\mathbb{P}(Y_\star^l \in T_{[PW]_\mu}) \tag{6.15}$$

$$\leq be^{-lD(PW\|W_\star)} \tag{6.16}$$

$$\leq e^{-l\epsilon} \quad \to 0 \text{ as } k \to \infty, \tag{6.17}$$

where : (6.15) follows from the union bound and the fact that there are $b-s$ noise packets; (6.16) follows from Lemma 2.1; and (6.17) is obtained by $b = e^{\alpha l}$ and $D(PW\|W_\star) \geq \alpha + \epsilon$.

Finally, the second term in (6.14) also vanishes because

$$\mathbb{P}(\mathcal{E}_2) \leq s\mathbb{P}(Y^l \notin T_{[PW]_\mu}) \tag{6.18}$$

$$\leq s2|\mathcal{Y}|e^{-2l\mu^2} \tag{6.19}$$

$$= l^{1/\lambda-1}2|\mathcal{Y}|e^{-2l\mu^2} \tag{6.20}$$

$$\to 0 \text{ as } l \to \infty (\text{or as } k \to \infty), \tag{6.21}$$

where : (6.18) follows from the union bound and the fact that there are s codeword packets, where Y^l denotes the output of the channel if the input is a codeword packet; (6.19) follows from Lemma 2.2 and Remark 2.1; (6.20) is obtained by substituting $s = k/l = l^{1/\lambda-1}$; and (6.21) is obtained by choosing the appropriate typicality parameter μ, e.g., $l\mu^4 \to \infty$ as $l \to \infty$ and $\mu \to 0$.

Now, combining (6.14), (6.17), (6.21), and the fact that the last term in (6.14) is also vanishing, we have $p_e^{avg} \to 0$ as $k \to \infty$, which proves the theorem. □

Remark 6.1. *Decoding from packet detection cannot be used for small-packet intermittent communication in which the packet length l is finite. Decoding from packet detection cannot be used for medium-packet intermittent communication model either, because there are too many smaller packets compared to the large-packet sce-*

nario. Specifically, in medium-packet intermittent communication, instead of (6.20)*, we have* $(1/l) \cdot e^{l/\lambda} \cdot 2|\mathcal{Y}|e^{-2l\mu^2}$*, which is not vanishing as* $l \to \infty$ *and* $\mu \to 0$.

Remark 6.2. *For large-packet intermittent communication, 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 6.5.*

This achievability result is identical to the capacity of asynchronous communication reviewed in Section 2.3. 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 large-packet intermittent communication, 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 for 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.

Figure 6.5 illustrates the achievable rate region (R, α) in Theorem 6.5 for large-packet intermittent communication over the channel depicted in Figure 6.2 with $p = 0.1$. As before, increasing the intermittency rate α reduces the achievable rate since it increases the uncertainty about the codeword packets 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 [47].

6.2 Deletion Channels

In this section, we use some of the insights and tools developed in this dissertation to obtain some new results on the capacity of deletion channels and a random access model that drops the collided symbols. Specifically, we first use a similar decoding structure as in decoding from exhaustive search introduced in Chapter 4 in

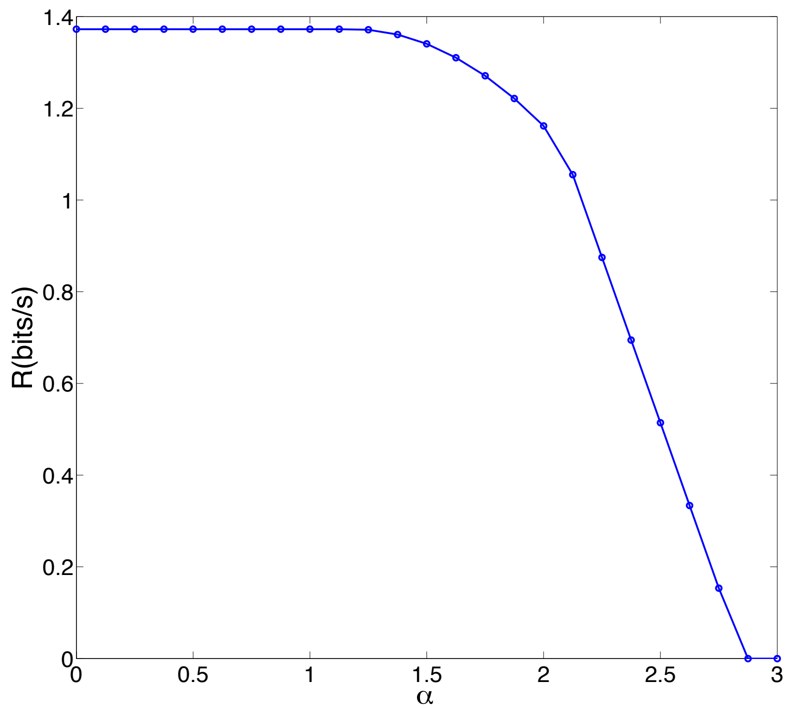

Figure 6.5. Achievable rate region (R, α) for large-packet intermittent communication over the channel depicted in Figure 6.2 with $p = 0.1$.

conjunction with a lemma on the longest common subsequence of random sequences to prove a side result to lower bound the capacity of deletion channels. Then, we obtain achievability results on a model for random access that drops / deletes collided symbols using a similar decoding structure as the decoding from pattern detection introduced in Chapter 4.

6.2.1 A Side Result on Deletion Channels

As we have discussed in Section 2.4, there is considerable work concentrating on achievability results for the deletion channel [11, 14, 31]. In [11], in addition to iid codewords with uniform distribution over the alphabet, codewords from first order Markov chains are used to improve the achievability results for deletion channels.

However, we require iid codewords in order to simplify the analysis of the probability of error for the decoding algorithm used in Section 6.2.2.

We first introduce a different model for the noisy deletion channel for which the number of deleted symbols is assumed to be fixed, and then give an achievability result, which is similar to the one in [11], but allows for arbitrary input distribution rather than the uniform one, and is valid for a general discrete memoryless channel (DMC) rather than the symmetric one. Note that our result is also valid for the iid deletion channel.

Consider the cascade of a deletion channel with a DMC, where the deletion channel deletes d symbols of its input sequence of length $k \geq d$ arbitrarily at random so that the output of the deletion channel and the DMC has length $k - d$, where $\theta := d/k \leq 1$ is the ratio of the deleted symbols to the codeword length. Let $\mathcal{X}$, $\mathcal{Y}$, W, and C_W denote the input alphabet, output alphabet, probability transition matrix, and the capacity of the DMC, respectively. After stating the following lemma from [5], we state the achievability result.

Lemma 6.1. *[5] For a given $|\mathcal{Y}|$-ary sequence y^{k-d} of length $k - d$, the number of $|\mathcal{Y}|$-ary sequences of length k that contain sequence y^{k-d} as a subsequence is given by*

$$\sum_{j=k-d}^{d} \binom{k}{j} (|\mathcal{Y}| - 1)^{d-j} \leq k \binom{k}{k-d} (|\mathcal{Y}| - 1)^{d}$$

Theorem 6.6. *For the noisy deletion channel described above, rates not exceeding $C_s - h(\theta) - \theta \log(|\mathcal{Y}| - 1)$ are achievable.*

Proof. Encoding: Fix an input distribution P. Randomly and independently generate e^{kR} sequences $c^k(m)$, $m \in [1 : e^{kR}]$ each iid according to P. To send message m, the encoder transmits $c^k(m)$.

Decoding: The decoder observes the output sequence y^{k-d} and constructs all possible $\tilde{y}^k$ that contain sequence y^{k-d} as a subsequence. Then it checks if any of these sequences are jointly typical with any of the codewords, i.e., if $\tilde{y}^k \in T_{[W_s]_\mu}(c^k(\hat{m}))$, then we declare that message as being sent. If this condition is not satisfied for any of the sequences $\tilde{y}^k$ and any of the messages, then the decoder declares an error.

Analysis of the probability of error: For any $\epsilon > 0$, we prove that if $R = \mathbb{I}(X;Y) - h(\theta) - \theta \log(|\mathcal{Y}| - 1) - 2\epsilon$, then the average probability of error vanishes as $k \to \infty$. Considering the uniform distribution on the messages and assuming that the message $m = 1$ is transmitted, we have

$$p_e^{avg} \leq \mathbb{P}(\hat{m}{=}e|m{=}1) + \mathbb{P}(\hat{m} \in \{2, 3, ..., e^{kR}\}|m{=}1), \tag{6.22}$$

where (6.22) follows from the union bound in which the first term is the probability that the decoder declares an error, i.e., does not find any codeword being jointly typical with any of the possible sequences $\tilde{y}^k$, that contain sequence y^{k-d} as a subsequence. This implies that even if the correct deletion pattern is considered and all possible choices for the deleted symbols are evaluated, none of them are jointly typical with $c^k(1)$. The probability of this event vanishes as $k \to \infty$ according to Lemma 2.2.

Applying Lemma 6.1 and the union bound for all possible $\tilde{y}^k$'s and all the messages $\hat{m} \neq 1$, we have

$$\begin{aligned}
&\mathbb{P}(\hat{m} \in \{2, 3, ..., e^{kR}\}|m = 1) \\
&\leq k\binom{k}{k-d}(|\mathcal{Y}|-1)^d(e^{kR}-1)\mathbb{P}(\tilde{Y}^k \in T_{[W_s]_\mu}(c^k(2))|m{=}1) \\
&\leq e^{o(k)} e^{k(h(\theta)+\theta\log(|\mathcal{Y}|-1))} e^{kR} e^{-k(\mathbb{I}(X;Y)-\epsilon)} && (6.23) \\
&= e^{o(k)} e^{-k\epsilon}, && (6.24)
\end{aligned}$$

where: (6.23) results from Stirling's approximation and Lemma 2.3 since conditioned on message $m = 1$ being sent, $C^k(2)$ and $\tilde{Y}^k$ are independent; and (6.24) follows by substituting $R = \mathbb{I}(X;Y) - h(\theta) - \theta\log(|\mathcal{Y}|-1) - 2\epsilon$. Therefore, the second term in (6.22) also vanishes as $k \to \infty$, and the lemma is proved by considering the capacity achieving input distribution for the DMC. $\square$

6.2.2 Random Access That Drops the Collided Symbols

In this section, we consider the intermittent MAC introduced in Chapter 5 and allows for collisions, but considers them as deletions. Specifically, we consider the same model for random access with collisions and no idle-times introduced in Section 5.4 with three possible states for the intermittent process $(s_1, s_2) \in \{(1,0),(0,1),(1,1)\}$, where the total number of states representing a collision, i.e., $(s_1, s_2) = (1,1)$, is assumed to be $d \leq k$.

We assume that if a collision occurs, then it will be dropped from the output sequence. Therefore, collisions are considered as deletions in this section. We assume that the output of the intermittent process with length $n = 2(k-d)$ consists of $k-d$ of the pair $(c_1, \star)$, $k-d$ of the pair $(\star, c_2)$, and d collided symbols that are deleted from the output sequence. The encoders and the decoder do not know the positions.

In the following theorem, we obtain an achievable rate region for (R_1, R_2).

Theorem 6.7. *For intermittent multi-access communication with the intermittent process described above, rates* (R_1, R_2) *satisfying*

$$R_1 < \mathbb{I}(X_1;Y|X_2=\star) - f_{deletion}(P_1,P_2,W,\theta)$$
$$R_2 < \mathbb{I}(X_2;Y|X_1=\star) - f_{deletion}(P_1,P_2,W,\theta)$$

are achievable for any $(X_1, X_2) \sim P_1(x_1)P_2(x_2)$, *where*

$$f_{deletion}(P_1,P_2,W,\theta) := (1-\theta)f_1(P_1,P_2,W) + h(\theta) + \theta\log(|\mathcal{Y}|-1), \qquad (6.25)$$

where $f_1(P_1, P_2, W)$ *is given in* (5.3).

The decoding scheme and the techniques for the analysis of the probability of error are a combination of those in the proofs of Theorem 6.6 and Theorem 5.1.

6.3 Lossless Source Coding with Intermittent Side-Information

Inspired by the problem of file synchronization [35, 36] in which we compress a source sequence with the benefit of decoder side-information that is related to the source via insertions, deletions, and substitutions, we study a similar problem in which the side-information at the decoder is related to the source via an intermittent process. Focusing on achievability, we introduce encoding and decoding structures in order to compress the source at the encoder and reconstruct it reliably at the decoder. We now summarize the system model.

Consider a discrete memoryless source (DMS) S^k, which is iid with distribution P, to be described losslessly via an index set $M \in [1 : e^{kR}]$ with small rate R to a decoder with side-information Y^N over a noiseless communication link. The side-information Y^N is related to the source sequence S^k via the cascade of an intermittent process and a DMC W as shown in Figure 6.6.

The intermittent process is the same was described in Chapter 4: After the i^{th} symbol from the source sequence S^k, N_i noise symbols $\star$ are inserted, where the N_i's are iid geometric random variables with mean $\alpha - 1$, with $\alpha \geq 1$ being the intermittency rate. The side-information Y^N of length N, where N is a random variable having a negative binomial distribution, is available to the decoder.

The decoder wishes to reconstruct the source sequence as $\hat{S}^k$ through a function of the index message M and side-information Y^N. The probability of decoding error is defined as $P_e^{(k)} := \mathbb{P}(\hat{S}^k \neq S^k)$. We refer to this compression setup as *source coding with intermittent side-information*. As a practical example for this problem, we can consider the compression of genomic sequences if a genomic sequence of the

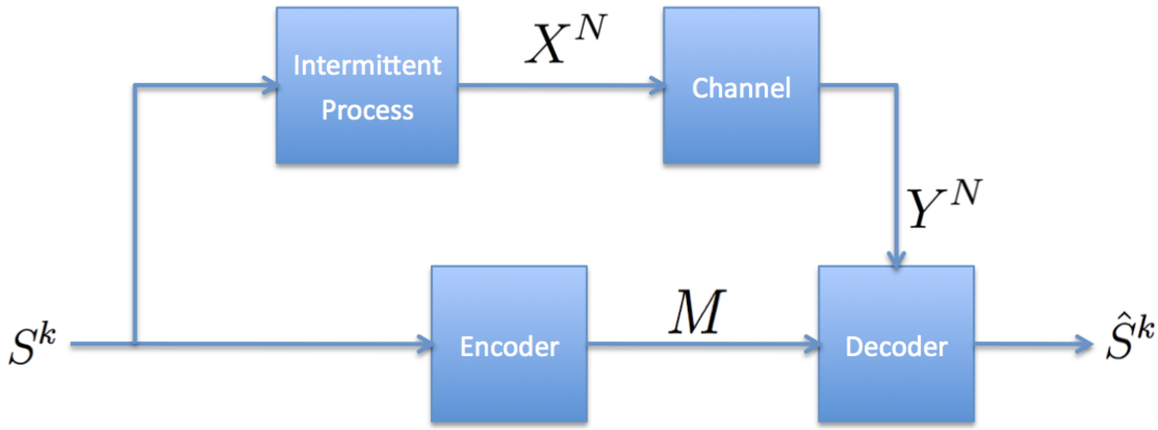

Figure 6.6. System model for source coding with intermittent side-information.

same species, which is different from the original one in terms of insertions and substitutions, is given as side-information [35, 36].

We say that rate R is achievable for the source coding with intermittent side-information if there exists a sequence of encoding and decoding schemes with message size e^{kR} such that $P_e^{(k)} \to 0$ as $k \to \infty$, i.e., the decoding is asymptotically lossless. The optimal lossless compression rate R^* is the infimum of all achievable rates. The following theorem provides an achievability result.

Theorem 6.8. *Assume that $(S, Y) \sim P(s)W(y|s)$. For the source coding problem with intermittent side-information described above with the DMS having distribution P, the DMC having transition probability matrix W, and the intermittency rate being α, we have*

$$R^* \geq H(S|Y) + f(P, W, \alpha), \tag{6.26}$$

where $H(\cdot|\cdot)$ is the conditional entropy and the function $f(\cdot, \cdot, \cdot)$ is the same as in (4.13).

Proof. For the encoding we use the idea of random binning: randomly and independently assign an index $m(s^k) \in [1 : e^{kR}]$ to each sequence $s^k \in \mathcal{S}^k$ uniformly over $[1 : e^{kR}]$. This results in e^{kR} bins denoted by $\mathcal{B}(m), m \in [1 : e^{kR}]$. Assume that the encoder and the decoder both know the chosen bin assignments. Upon observing $s^k \in \mathcal{B}(m)$, the encoder sends the bin index m.

Note that the length of the side-information sequence at the decoder, N, is a random variable. However, using the same procedure as in the proof of Theorem 4.2, we can focus on the case that $|N/k-\alpha| < \epsilon$, and essentially assume that the length of the side-information is of length $n = \alpha k$, which makes the analysis of the probability of error for the decoding algorithms more concise.

The decoding structure is similar to the decoding from pattern detection described in Chapter 4: choose k symbols out of the n symbols of y^n, denote them by $\tilde{y}^k$ and denote the other symbols by $\hat{y}^{n-k}$. In the first stage, check if $\tilde{y}^k \in T_{PW}$ and if $\hat{y}^{n-k} \in T_{W_\star}$. If both of these conditions are satisfied, then continue to the second stage; otherwise make another choice for the k symbols. In the second stage, upon receiving the index m, the decoder declares s^k as the estimated source sequence if it is the unique sequence in $\mathcal{B}(m)$ that satisfies $s^k \in T_{[W^{-1}]_\mu}(\tilde{y}^k)$ for a small enough typicality parameter μ; otherwise the decoder makes another choice for the k symbols and repeats the procedure. If by the end of all $\binom{n}{k}$ possible choices the decoder does not the estimated source sequence, then it declares an error.

The analysis of the probability of error is a combination of the techniques used in Theorem 4.4 and [23, Theorem 10.1]. $\square$

6.4 Summary

In this chapter, we first studied packet-level intermittent communication in which codeword and noise symbols are grouped into packets. Depending on the scaling behavior of the packet length relative to the codeword length, we identified three scenarios: small-packet, medium-packet, and large-packet intermittent communication. For small- and medium-packet intermittent communication, we utilized both decoding from exhaustive search and decoding from pattern detection in order to obtain achievable rates, whereas, for large-packet intermittent communication, we utilized decoding from packet detection in order to obtain achievable rates. Increasing the intermittency rate generally reduces the achievable rate for each of the three scenarios, because it makes the receive window larger, and therefore, increases the uncertainty about the positions of the codeword packets at the receiver, making the decoder's task more involved.

Next, we used a similar decoding structure to decoding from exhaustive search in conjunction with a lemma on the longest common subsequence of random sequences to prove a side result on lower bounding the capacity of the deletion channels. We also obtained achievability results for a random access model that drops / deletes collided symbols using a similar decoding structure to decoding from pattern detection.

Finally, inspired by the problem of file synchronization in which we compress a source sequence with the benefit of decoder side-information that is related to the source via insertions, deletions, and substitutions, we studied a similar problem in which the side-information at the decoder is related to the source via an intermittent process. Focusing on achievability, we introduced encoding and decoding structures in order to compress the source at the encoder and reconstruct it reliably at the decoder.

CHAPTER 7

CONCLUSIONS AND FUTURE WORK

In this final chapter, we conclude the dissertation and introduce some directions for future research.

7.1 Conclusions

We formulated a model for intermittent communication that can capture bursty transmissions or a sporadically available channel by inserting a random number of silent symbols between each codeword symbol so that the receiver does not know a priori when the transmissions will occur. We specified two decoding structures in order to develop achievable rates. Interestingly, decoding from pattern detection, which achieves a larger rate, is based on a generalization of the method of types and properties of partial divergence. As the system becomes more intermittent, the achievable rates decrease due to the additional uncertainty about the positions of the codeword symbols at the decoder. For the case of binary-input binary-output noiseless channel, we obtained upper bounds on the capacity of intermittent communication 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. Despite of the large gap between the lower bounds and the upper bounds, 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 and the intermittency rate, even if the receive window scales linearly with the codeword length.

We extended the model to intermittent multi-access communication for two users that captures the bursty transmission of the codeword symbols for each user and the possible asynchronism between the receiver and the transmitters as well as between the transmitters themselves. This model can be viewed as an attempt to combine information-theoretic and network-oriented multi-access models. We characterized the performance of the system in terms of achievable rate regions. In our achievable schemes, the intermittency of the system comes with a significant cost. Inspired by network applications, we extended the model to packet-level intermittent communication in which codeword and noise symbols are grouped into packets. Depending on the scaling behavior of the packet length relative to the codeword length, we identified some interesting scenarios, and characterize the performance of the system in terms of the achievable rates for each model. Furthermore, we used some of the insights and tools developed in this dissertation to obtain some new results on the capacity of deletion channels and a random access model that drops the collided symbols, and on the problem of source coding with the presence of intermittent side-information.

The results of the dissertation suggest that intermittency / lack of synchronization is quite costly. Therefore, the transmitters should continuously send the symbols / packets whenever possible.

7.2 Future Work

In this section, we introduce some directions for future research. First, we introduce other types of decoding structures that might lead to tighter lower bounds on the capacity of intermittent communication as well as error exponents for intermittent communication. Next, we evaluate possible approaches that can lead to stronger upper bounds for intermittent communication. As another future direction, we introduce the problem of finding the fountain capacity of intermittent communication. Finally, we suggest designing explicit code constructions for intermittent communication.

7.2.1 Error Exponent for Intermittent Communication

As we have seen in Chapter 4, there is a relatively large gap between the lower and upper bounds for the capacity of intermittent communication. Although we have no strong evidence whether the lower bounds, upper bounds, or both need to be improved in order to tighten this gap, one way to improve the lower bounds is to consider other decoding algorithms. Regardless of its tractability in our problem setup, maximum likelihood (ML) decoding is optimal. The ML decoder achieves the error exponents of memoryless channels [21]. Error exponents determine the rate at which the probability of error vanishes exponentially with respect to the block length. In addition, it is an interesting problem to see how the intermittency of the communication system can affect the performance for finite codeword length k. For these reasons, one future direction worth exploring is ML decoding algorithms and analysis for intermittent communication.

Denoting the codeword associated with the message m by $\mathbf{x}(m)$, and the channel output vector by $\mathbf{y}$, ML decoding selects message $\hat{m}$ if

$$\mathbb{P}(\mathbf{y}|\mathbf{x}(\hat{m})) \geq \mathbb{P}(\mathbf{y}|\mathbf{x}(m)), \quad \text{for all } m \neq \hat{m}.$$

Gallager in [21] proves a general coding theorem that for a given number $M \geq 2$ of codewords, where each codeword is chosen independently with probability measure $P_{\mathbf{X}}$, the average probability of error under ML decoding over this ensemble of codes given that message m is sent, denoted by $\bar{P}_{e,m}$, by

$$\bar{P}_{e,m} \leq (M-1)^{\rho} \sum_{\mathbf{y}} \left[\sum_{\mathbf{x}} P_{\mathbf{X}}(\mathbf{x}) \mathbb{P}(\mathbf{y}|\mathbf{x})^{1/(1+\rho)} \right]^{1+\rho}, \text{ for any } 0 \leq \rho \leq 1. \tag{7.1}$$

This bound is surprisingly general and powerful, and applies to both memoryless channels and channels with memory [21]. For memoryless channels, the bound (7.1) leads to single-letter expressions for error exponents. Consider a DMC with with probability transition matrix W and input and output alphabets $\mathcal{X}$ and $\mathcal{Y}$, respectively. The bound (7.1) can be specialize to the following explicit error exponent if using k channel uses of the DMC and each symbol of the codeword is selected independently with the probability assignment $P(x)$:

$$\bar{P}_{e,m} \leq e^{-k(E_0(\rho,W)-\rho R)}, \tag{7.2}$$

where

$$E_0(\rho, W) = -\log \sum_{y\in\mathcal{Y}} \left[\sum_{x\in\mathcal{X}} P(x) W(y|x)^{1/(1+\rho)} \right]^{1+\rho}.$$

However, there is inherent memory in the model for intermittent communication, which makes it difficult to obtain closed form error exponents for this model. Specifically, the input and output vectors $\mathbf{x}$ and $\mathbf{y}$ in (7.1) have length k and N, respectively. Let vector $T^{k+1} := (N_0, N_1, ..., N_k)$ denote the number of noise insertions in between the codeword symbols, where the N_i's are iid geometric random variables with probability mass function P_N as described in Chapter (4), and let $t^{k+1} = (n_0, n_1, ..., n_k)$ denote its realization. Then, the term $\mathbb{P}(\mathbf{y}|\mathbf{x})$ in (7.1) can be expressed as

$$\mathbb{P}(\mathbf{y}|\mathbf{x}) = \sum_{t^{k+1}} P_N(n_0) \prod_{l=1}^{n_0} W_\star(y_l) \prod_{i=1}^{k} \left[P_N(n_i) W(y_{i+\sum_{j=0}^{i-1} n_j}|x_i) \prod_{l=1}^{n_i} W_\star(y_{l+i+\sum_{j=0}^{i-1} n_j}) \right].$$

This is a complicated mixture distribution over all the possible realizations of the intermittent process t^{k+1}. Therefore, using the same techniques as in [21], it is not clear how to obtain an explicit error exponent analogous to (7.2). It is, however, possible in principle to compute the related quantities numerically for finite codeword length k.

We note that problem exists even for asynchronous communication described in Section 2.3, for which the term $\mathbb{P}(\mathbf{y}|\mathbf{x})$ in (7.1) can be expressed as

$$\mathbb{P}(\mathbf{y}|\mathbf{x}) = \sum_{\nu\in[1:n]} \prod_{i=1}^{\nu-1} W_\star(y_i) \prod_{i=\nu}^{\nu+k-1} W(y_i|x_{i-\nu+1}) \prod_{i=\nu+k}^{n} W_\star(y_i),$$

which is again a mixture distribution over all the possible realizations of $\nu \in [1:n]$. The memoryless technique in [21] cannot be used to derive an explicit error exponent as in (7.2). It is worth mentioning that error exponents have been obtained for slotted asynchronous communication in [39, 60, 61]. However, in slotted asynchronous communication, it is assumed that a given sequence is either codeword or noise, and therefore, the problem of handling a mixture distribution, which results from sequences partially from codeword symbols and partially from noise symbols, does not arise.

Another decoding structure that can achieve the error exponents for the probability of error is the maximum mutual information (MMI) decoder. In [10], the same error exponents as in [21], but in a seemingly different form, are obtained using a general packing lemma and MMI decoding. The techniques used in [10] to analyze the performance of the MMI decoder are suited for memoryless channels. Generalizing these techniques to channels with memory may lead to error exponents for intermittent communication using the MMI decoder.

In summary, an important future direction is to consider other types of decoding structures that might help improving the lower bounds on the capacity of intermittent communication as well as characterizing the rate at which the probability of error decays for intermittent communication.

7.2.2 Stronger Upper Bounds for Intermittent Communication

As we have seen in Section 4.3, the gap between the achievable rates and upper bounds is not tight, especially for large values of intermittency rate α. As a future direction, we propose to decrease this gap. We believe we can tighten the upper bounds by providing the encoder and decoder with side-information less frequently, but this would exponentially increase the computational complexity of our calculations, and improvements might be marginal.

One possible solution for improving the upper bounds is to focus on other kinds of side-information than explored in Section 4.3 to get less complex and tighter outer bounds. Another possibility is to explore other approaches to upper bounds on the capacity not only for the binary noiseless insertion channel, but also for the general class of intermittent communication. Ideally, this general approach could lead to outer bounds for the capacity region of different kinds of intermittent MACs introduced in Chapter 5.

Also, it is worth exploring upper bounds for packet-level intermittent communication introduced in Section 6.1. For a special case of small-packet intermittent communication with $l = 1$, namely the noiseless binary-input binary-output channel, we have found upper bounds in Section 4.3, but these are not extendable to more general cases due to their computational complexity. Also, large-packet intermittent communication with $\lambda = 1$ recovers asynchronous communication for which the capacity is known and is given in Theorem 2.1, and therefore, the converse exists for this special case [4, 47]. In [47], the meta-converse principle, initially introduced in [46, 48], is used to prove a strong converse for the capacity of asynchronous communication. We suggest utilizing or extending the concept of meta-converse in order to obtain upper bounds on the capacity of packet-level intermittent communication models.

Specifically, consider large-packet intermittent communication, for which an achievability result is obtained in Theorem 6.5, and consider the following requirement: the decoder in addition to decoding the message m, it should detect the exact positions of all the codeword packets denoted by $n_1, n_2, ..., n_s$, where $s = k/l = l^{1/\lambda - 1}$ is the number of codeword packets. Denoting the decoded message by $\hat{m}$ and detected positions by $\hat{n}_1, \hat{n}_2, ..., \hat{n}_s$, the average probability of error is defined as

$$p_e^{avg} = 1 - \mathbb{P}\left(\hat{m} = m, (n_1, n_2, ..., n_s) = (\hat{n}_1, \hat{n}_2, ..., \hat{n}_s)\right).$$

Under these conditions, it can be easily verified that the rate in Theorem 6.5 can still be achieved. Furthermore, using [47, Lemma 8] and the same meta-converse techniques, a converse with the rate in Theorem 6.5 can be proved. This means that, under this additional requirement, the capacity of large-packet intermittent communication is $\max_P\{\left(\mathbb{I}(P, W) - f^{LP}(P, W, \alpha)\right)^+\}$, where $f^{LP}(P, W, \alpha)$ is given in (6.11).

Although this additional requirement for the decoder to detect all the codeword packets is helpful in order to find the capacity of large-packet intermittent communication, we would face some difficulties in obtaining an achievability result for small- and medium-packet intermittent communication with this additional requirement. It is clear that for small-packet intermittent communication, it is not possible to detect all the codeword packets with vanishing probability of error since the packet length is finite.

7.2.3 Intermittent Receiver: Fountain Capacity of Intermittent Communication

The notion of fountain capacity for arbitrary channels is introduced in [50] in which the definition of rate penalizes the reception of symbols by the receiver rather than their transmission. In this communication model, a message $m \in [1 : M]$ is encoded into a sequence of infinite length. Figure 7.1 shows the basic fountain setup

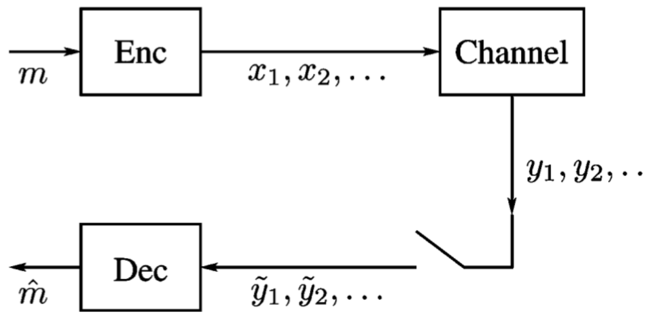

Figure 7.1. Concatenation of noisy and erasure channel [50].

in which the receiver, but not the transmitter, is aware of the schedule of times at which the switch is on [50].

A scheduler determines the times at which the receiver is allowed to observe the channel outputs, which is modeled by the switch in Figure 7.1. The schedule is denoted by $\mathfrak{N}$ with cardinality $|\mathfrak{N}|$, and the receiver is only allowed to see the channel outputs $\{y_i, i \in \mathfrak{N}\}$. The schedule is unknown to the encoder, and is adversely chosen without the knowledge of either the message, codebook, or the channel output. The rate is essentially defined with respect to the cardinality of the schedule $|\mathfrak{N}|$ rather than the number of encoded symbols or channel uses.

We can think of this setup as an intermittent receiver, which is able to observe the channel output at only some time slots with a given maximum number of these time slots, where the definition of the rate is “pay-per-view” rather than “pay-per-use”. The worst case scenario for the schedule in [50] can be interpreted as a worst case intermittency at the receiver. If both the transmitter and receiver are intermittent, where the intermittency of the transmitter is defined in the sense of bursty transmission of the codeword symbols as described in this dissertation, and the intermittency of the receiver is defined with respect to the schedule $\mathfrak{N}$ as described above, then a relevant metric for characterizing the performance of this system is fountain capacity of intermittent communication. Note that in Chapter 4, we focused on the Shannon capacity of intermittent communication and derived some lower and upper bounds.

It is proven in [50] that the fountain capacity equals the Shannon capacity for a stationary memoryless channel, and the fountain capacity is upper bounded by the Shannon capacity for a general channel. Furthermore, it is mentioned in [50] that a general formula for the fountain capacity of channels with memory is an open problem, as the least favorable schedule is heavily dependent on the channel. However, the fountain capacity of some channels with memory has been obtained in [50]. As intermittent communication is an example of channels with memory, its fountain capacity is a non-trivial problem, and could provide some insights on the additional overhead cost caused by the intermittency of the receiver.

7.2.4 Code Construction for Intermittent Communication

As a future direction, we suggest designing explicit (non-random) code constructions for intermittent communication. In Section 4.2, we described some encoding and decoding structures to find achievable rates. However, these structures are based on a random coding argument, can achieve the rates asymptotically as $k \to \infty$, and are computationally complex. Specifically, the decoding structures include as many as $\binom{n}{k}e^{kR}$ typicality tests, which is not feasible in practical systems.

In [6], heavy weight codes are investigated, which are described as good codes for asynchronous communication reviewed in Section 2.3. The authors study $B(n, d, w)$, defined as the maximum number of length n binary sequences with minimum distance d, and such that each sequence has weight at least w, and investigate the exponential growth rate of this function with respect to the sequence length n. In asynchronous communication, it is proved that combined synchronization and information transmission can lead to a significant reduction in error probability compared to a separation architecture, where the problem of synchronization is handled with a

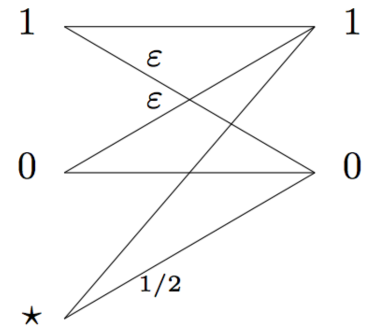

Figure 7.2. Binary symmetric channel with external noise symbol [6].

common preamble at the beginning of each codeword [54]. To understand how heavy weight codes can carry information while also acting as information flags, consider the following asynchronous channel model introduced in [52] and shown in Figure 7.2.

In order to increase the message isolation, we need to increase the minimum distance of a code. Note that with the channel model depicted in Figure 7.2, a typical noise sequence contains equal number of 0's and 1's, because a noise symbol $\star$ produces 0 or 1 with equal probability. In order to increase the message detection, we need to bias the codewords with a common flag, which here is the weight distribution, i.e., we focus on heavy weight codes so that they are easily distinguishable from the noise sequence.

In intermittent communication, the task of code construction is much more difficult since the codewords should be immune to insertions in addition to the noisy channel. Even for the simple case of binary error-free channel (BSC with zero crossover probability) with the noise symbol $\star = 0$, it is not clear how to construct codewords such that they are distinguishable after some number of 0's are inserted in between the codewords' symbols. Note that in this example, the number of 1's remains fixed at the decoder, but there might be additional 0's between them. As a naive code construction, we might consider $k + 1$ codewords with different number of 1's, and the decoder can detect the codeword symbols with probability one for any codeword length k. However, in order to construct a code with positive rate, we need an exponential number of codewords.

BIBLIOGRAPHY

1. R. E. Blahut. Computation of Channel Capacity and Rate-Distortion Functions. *IEEE Transactions on Information Theory*, 18(4):460–473, 1972.

2. S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

3. V. Chandar, A. Tchamkerten, and G. Wornell. Optimal Sequential Frame Synchronization. *IEEE Transactions on Information Theory*, 54(8):3725–3728, 2008.

4. V. Chandar, A. Tchamkerten, and D. N. C. Tse. Asynchronous Capacity per Unit Cost. *IEEE Transactions on Information Theory*, 59(3):1213–1226, 2013.

5. V. Chvatal and D. Sankoff. Longest Common Subsequence of Two Random Sequences. *Journal of Applied Probability*, pages 306–315, 1975.

6. G. Cohen, P. Solé, and A. Tchamkerten. Heavy Weight Codes. In *Proc. IEEE Int. Symp. Information Theory (ISIT)*, pages 1120–1124. Austin, Texas, 2010.

7. T. M. Cover and J. A. Thomas. *Elements of Information Theory*. New York: Wiley, second edition, 2006.

8. T. M. Cover, R. J. McEliece, and E. C. Posner. Asynchronous Multiple-Access Channel Capacity. *IEEE Transactions on Information Theory*, 27(4):409–413, 1981.

9. I. Csiszár. The Method of Types. *IEEE Transactions on Information Theory*, 44 (6):2505–2523, 1998.

10. I. Csiszár and J. Körner. *Information Theory: Coding Theorems for Discrete Memoryless Systems*. New York: Academic, second edition, 2011.

11. S. Diggavi and M. Grossglauser. On Information Transmission over a Finite Buffer Channel. *IEEE Transactions on Information Theory*, 52(3):1226–1237, 2006.

12. S. Diggavi, M. Mitzenmacher, and H. D. Pfister. Capacity Upper Bounds for the Deletion Channels. In *Proc. IEEE Int. Symp. Information Theory (ISIT)*, pages 1716–1720. Nice, France, 2007.

13. R. L. Dobrushin. Shannon's Theorems for Channels with Synchronization Errors. *Problems of Information Transmission*, 3(4):11–26, 1967.

14. E. Drinea and M. Mitzenmacher. Improved Lower Bounds for the Capacity of i.i.d. Deletion and Duplication Channels. *IEEE Transactions on Information Theory*, 53(8):2693–2714, 2007.

15. B. P. Dunn. *Overhead in Communication Systems as the Cost of Constraints*. PhD thesis, University of Notre Dame, Notre Dame, IN, 2010.

16. A. Ephremides and B. Hajek. Information Theory and Communication Networks: An Unconsummated Union. *IEEE Transactions on Information Theory*, 44(6): 2416–2434, 1998.

17. L. Farkas and T. Koi. Capacity Regions of Discrete Asynchronous Multiple Access Channels. In *Proc. IEEE Int. Symp. Information Theory (ISIT)*, pages 2273–2277. St. Petersburg, Russia, 2011.

18. D. Fertonani and T. M. Duman. Novel Bounds on the Capacity of the Binary Deletion Channel. *IEEE Transactions on Information Theory*, 56(6):2753–2765, 2010.

19. D. Fertonani, T. M. Duman, and M. F. Erden. Bounds on the Capacity of Channels with Insertions, Deletions and Substitutions. *IEEE Transactions on Communications*, 59(1):2–6, 2011.

20. R. Gallager. Sequential Decoding for Binary Channels with Noise and Synchronization Errors. *Tech. Rep., MIT Lincoln Lab. Group Report*, 1961.

21. R. Gallager. *Information Theory and Reliable Communication*. Wiley, 1968.

22. R. Gallager. A Perspective on Multiaccess Channels. *IEEE Transactions on Information Theory*, 31(2):124–142, 1985.

23. A. E. Gamal and Y.-H. Kim. *Network Information Theory*. Cambridge University Press, 2012.

24. Y.-C. Huang, U. Niesen, and P. Gupta. Energy-Efficient Communication in the Presence of Synchronization Errors. *arXiv preprint arXiv:1301.6589*, 2013.

25. J. Y. N. Hui and P. A. Humblet. The Capacity Region of the Totally Asynchronous Multiple-Access Channel. *IEEE Transactions on Information Theory*, 31(2):207–216, 1985.

26. Y. Kanoria and A. Montanari. On the Deletion Channel with Small Deletion Probability. In *Proc. IEEE Int. Symp. Information Theory (ISIT)*, pages 1002–1006. Austin, Texas, 2010.

27. M. Khoshnevisan and J. N. Laneman. Intermittent Communication and Partial Divergence. In *Proc. Allerton Conference on Communications, Control, and Computing*, pages 656–662. Monticello, IL, USA, 2012.

28. M. Khoshnevisan and J. N. Laneman. Achievable Rates for Intermittent Communication. In *Proc. IEEE Int. Symp. Information Theory (ISIT)*, pages 1346–1350. Cambridge, MA, USA, 2012.

29. M. Khoshnevisan and J. N. Laneman. Upper Bounds on the Capacity of Binary Intermittent Communication. In *Proc. Workshop on Information Theory and its Applications (ITA)*, pages 1–6. San Diego, CA, 2013.

30. M. Khoshnevisan and J. N. Laneman. Achievable Rates for Intermittent Multi-Access Communication. In *Proc. IEEE Information Theory Workshop (ITW)*. Seville, Spain, 2013.

31. A. Kirsch and E. Drinea. Directly Lower Bounding the Information Capacity for Channels with i.i.d. Deletions and Duplications. *IEEE Transactions on Information Theory*, 56(1):86–102, 2010.

32. J. N. Laneman and B. P. Dunn. Communications Overhead as the Cost of Constraints. In *Proc. IEEE Information Theory Workshop (ITW)*, pages 365–369. Paraty, Brazil, 2011.

33. J. Luo and A. Ephremides. A Channel Coding Approach for Random Access Communication with Bursty Sources. In *Proc. IEEE Int. Symp. Information Theory (ISIT)*, pages 1823–1827. Austin, Texas, 2010.

34. J. Luo and A. Ephremides. A New Approach to Random Access: Reliable Communication and Reliable Collision Detection. *IEEE Transactions on Information Theory*, 58(2):989–1002, 2012.

35. N. Ma, K. Ramchandran, and D. Tse. Efficient File Synchronization: A Distributed Source Coding Approach. In *Proc. IEEE Int. Symp. Information Theory (ISIT)*, pages 583–587. St. Petersburg, Russia, 2011.

36. N. Ma, K. Ramchandran, and D. Tse. A Compression Algorithm Using Mis-Aligned Side-Information. In *Proc. IEEE Int. Symp. Information Theory (ISIT)*, pages 16–20. Cambridge, MA, USA, 2012.

37. J. L. Massey. Optimum Frame Synchronization. *IEEE Transactions on Communications*, 20(2):115–119, 1972.

38. J. L. Massey and P. Mathys. The Collision Channel without Feedback. *IEEE Transactions on Information Theory*, 31(2):192–204, 1985.

39. N. Merhav. Codeword or Noise? Exact Random Coding Exponents for Slotted Asynchronism. *arXiv preprint arXiv:1308.4572*, 2013.

40. P. Minero, M. Franceschetti, and D. N. C. Tse. Random Access: An Information-Theoretic Perspective. *IEEE Transactions on Information Theory*, 58(2):909–930, 2012.

41. M. Mitzenmacher. Capacity Bounds for Sticky Channels. *IEEE Transactions on Information Theory*, 54(1):72–77, 2008.

42. M. Mitzenmacher. A Survey of Results for Deletion Channels and Related Synchronization Channels. *Probability Surveys*, 6:1–33, 2009.

43. P. T. Nielsen. Some Optimum and Suboptimum Frame Synchronizers for Binary Data in Gaussian Noise. *IEEE Transactions on Communications*, 21(6):770–772, 1973.

44. M. S. Pinsker. *Information and Information Stability of Random Variables and Processes*. Holden-Day, 1964.

45. G. S. Poltyrev. Coding in an Asynchronous Multiple-Access Channel. *Problems Information Transmission*, pages 12–21, 1983.

46. Y. Polyanskiy. *Channel Coding: Non-Asymptotic Fundamental Limits*. PhD thesis, Princeton University, Princeton, NJ, 2010.

47. Y. Polyanskiy. Asynchronous Communication: Exact Synchronization, Universality, and Dispersion. *IEEE Transactions on Information Theory*, 59(3):1256–1270, 2013.

48. Y. Polyanskiy, H. V. Poor, and S. Verdú. Channel Coding Rate in the Finite Blocklength Regime. *IEEE Transactions on Information Theory*, 56(5):2307–2359, 2010.

49. R. A. Scholtz. Frame Synchronization Techniques. *IEEE Transactions on Communications*, 28(8):1204–1212, 1980.

50. S. Shamai, E. Telatar, and S. Verdú. Fountain Capacity. *IEEE Transactions on Information Theory*, 53(11):4372–4376, 2007.

51. A. Somekh-Baruch, S. Shamai, and S. Verdú. Cooperative Multiple-Access Encoding with States Available at One Transmitter. *IEEE Transactions on Information Theory*, 54(10):4448–4469, 2008.

52. A. Tchamkerten, V. Chandar, and G. W. Wornell. Communication under Strong Asynchronism. *IEEE Transactions on Information Theory*, 55(10):4508–4528, 2009.

53. A. Tchamkerten, V. Chandar, and G. Caire. Energy and Sampling Constrained Asynchronous Communication. *arXiv preprint arXiv:1302.6574*, 2013.

54. A. Tchamkerten, V. Chandar, and G. W. Wornell. Asynchronous Communication: Capacity Bounds and Suboptimality of Training. *IEEE Transactions on Information Theory*, 59(3):1227–1255, 2013.

55. R. Venkataramanan, S. Tatikonda, and K. Ramchandran. Achievable Rates for Channels with Deletions and Insertions. In *Proc. IEEE Int. Symp. Information Theory (ISIT)*, pages 346–350. St. Petersburg, Russia, 2011.

56. S. Verdú. The Capacity Region of the Symbol-Asynchronous Gaussian Multiple-Access Channel. *IEEE Transactions on Information Theory*, 35(4):733–751, 1989.

57. S. Verdú. Multiple-Access Channels with Memory with and without Frame Synchronism. *IEEE Transactions on Information Theory*, 35(3):605–619, 1989.

58. S. Verdú. On Channel Capacity per Unit Cost. *IEEE Transactions on Information Theory*, 36(5):1019–1030, 1990.

59. D. Wang. Distinguishing Codes From Noise: Fundamental Limits and Applications to Sparse Communication. Master's thesis, Massachusetts Institute of Technology, Cambridge, MA, 2010.

60. D. Wang, V. Chandar, S.-Y. Chung, and G. Wornell. Error Exponents in Asynchronous Communication. In *Proc. IEEE Int. Symp. Information Theory (ISIT)*, pages 1071–1075. St. Petersburg, Russia, 2011.

61. D. Wang, V. Chandar, S.-Y. Chung, and G. Wornell. On Reliability Functions for Single-Message Unequal Error Protection. In *Proc. IEEE Int. Symp. Information Theory (ISIT)*, pages 2934–2938. Cambridge, MA, USA, 2012.

This document was prepared & typeset with pdfLaTeX*, and formatted with* NDdiss2$_\varepsilon$ *classfile (v3.2013[2013/04/16]) provided by Sameer Vijay and updated by Megan Patnott.*