On Error Exponents and Moderate Deviations for Lossless Streaming Compression of Correlated Sources

Lin Zhou, Vincent Y. F. Tan, and Mehul Motani
Department of Electrical and Computer Engineering,
National University of Singapore
Emails:{lzhou, vtan, motani}@nus.edu.sg

Abstract

We derive upper and lower bounds for the error exponents of lossless streaming compression of two correlated sources under the blockwise and symbolwise settings. We consider the linear scaling regime in which the delay is a scalar multiple of the number of symbol pairs of interest. We show that for rate pairs satisfying certain constraints, the upper and lower bounds for the error exponent of blockwise codes coincide. For symbolwise codes, the bounds coincide for rate pairs satisfying the aforementioned constraints and a certain condition on the symbol pairs we wish to decode—namely, that their indices are asymptotically comparable to the blocklength. We also derive moderate deviations constants for blockwise and symbolwise codes, leveraging the error exponent results, and using appropriate Taylor series expansions. In particular, for blockwise codes, we derive an information spectrum-type strong converse, giving the complete characterization of the moderate deviations constants. For symbolwise codes, under an additional limiting requirement, we can show that the moderate deviations constants are the same as the blockwise setting.

Index Terms

Lossless source coding, Streaming sources, Error exponents, Moderate deviations, Feedforward decoder, Information spectrum method, Random binning

I. INTRODUCTION

Streaming is the dominant mode of consuming videos on smart phones, tablets and laptops. Instead of downloading an entire whole movie at once, we watch as we download. This illustrates the decoding nature of streaming communication: The decoder incurs a delay. In wireless body sensor networks, sensors collect and monitor vital data. These sensors can potentially generate a huge amount of data, which need to be compressed and transmitted in real time. This highlights the encoding nature of streaming communication: The data enters the encoder in real time.

In this paper, we consider lossless compression of streaming data. We assume there are two correlated sources to be compressed. Instead of knowing all the source symbol pairs at the time of compression, as in [1], [2], the encoder pair has access to only one symbol pair per unit time in our streaming scenario. The decoder in the streaming scenario takes some extra time, called a fixed delay $\Delta \in \mathbb{N}$, to either produce an estimate of a symbol pair or an estimate of a sequence of symbol pairs. The former and latter are called symbolwise and blockwise decoders respectively. We assume that the delay scales linearly with the number of symbol pairs we would like to reconstruct. This is motivated by real-world applications in communications with delays, in which receivers are typically more tolerant of longer delays when receiving large files (e.g., movies, high-quality audio files), but are less tolerant of delays when they receive smaller files (e.g., emails, text messages). Also, this setting allows us to prove that the derived information-theoretic limits are tight in some regimes. The system illustrating the blockwise decoder is shown in Fig. 1. At time $j + \Delta$, each encoder has access to $j + \Delta$ symbols and the decoder produces an estimate of the set of symbol pairs generated until time $j$. The model can be simplified to lossless streaming compression of single source with (resp. without) side information by removing the second encoder (resp. second source). Our interest in this paper is to explore the error exponents (i.e., the speed of exponential decay of the error probability) and moderate deviations constants for lossless streaming compression of multiple correlated sources, i.e., streaming Slepian-Wolf (SW) source coding.

In the traditional study of error exponents for source coding, the rate of the code $R$ is fixed at a value strictly above entropy $H(P_X)$ (which is the first-order fundamental limit) and the exponential rate of the decay of the error probability is sought [3], [4]. In contrast, in the study of moderate deviations, the rate of the code $R_n$ depends on the blocklength $n$ and converges to the entropy $H(P_X)$ (the first-order fundamental limit) at a speed slower than $1/\sqrt{n}$ while the error probability decays to zero at a sub-exponential speed of $\exp(-bn^{1/2})$ for $t \in (0, 1/2)$. The constant $b$ is known as the moderate deviations constant and is the object of study in moderate deviations analysis here. We note that there is yet another asymptotic regime, known as second-order or normal approximation analysis [5], where the rate $R_n$ converges to the first-order fundamental limit at a rate of the order $1/\sqrt{n}$ and the error probability converges to a constant between 0 and 1. We do not explore this regime here.
rates and, additionally, we consider the empirical entropy decoding. For symbolwise codes, we rely on the random binning idea and we use both maximum likelihood decoding and minimum empirical entropy decoding. It turns out that, under both conditions, the error exponents for blockwise and symbolwise codes coincide. We show that the upper and lower bounds for blockwise codes coincide for a non-empty set of rate pairs. We utilize the information spectrum method to derive a strong converse for blockwise codes; we resort to the change of measure technique, we derive upper bounds on the optimal error exponents. To derive lower bounds on the optimal error exponents, for both blockwise and symbolwise codes, we use both maximum likelihood decoding and minimum empirical entropy decoding.

A. Related Work

The key reference that is related to the present work is the paper by Draper, Chang and Sahai [6]. The authors derived lower bounds for the error exponent of streaming SW coding by using random binning, minimum empirical entropy decoding, and maximum likelihood decoding. The authors considered polynomially many error events and so the one with the smallest exponent dominates. Chang and Sahai [7] derived the error exponent for lossless streaming compression. They demonstrated similar results for lossless streaming compression of single sources with both encoder and decoder side information in [8]. They also extended the feedforward decoder idea (which originated from Pinsker [9]) to streaming compression of sources with only decoder side information to derive an upper bound on the error exponent. Other works by Chang are summarized in [10].

In other (non-exhaustive) works on streaming and source coding with delayed decoding, Palaiyanur [11] studied lossless streaming compression of a source with side information with and without a discrete memoryless channel between the encoder and the decoder. Venkataramanan and Pradhan [12] considered the source coding with feedforward problem and derived lossy source coding theorems for general sources. Matsuta and Uyematsu [13] considered the lossy source coding problem with delayed side information. Ma and Ishwar [14] focused on delayed sequential coding of correlated video sources. Finally, Zhang [15] analyzed the error exponent of lossless streaming compression of a single source using variable-length sequential random binning.

Altuğ and Wagner [16] initiated the study of moderate deviations in information theory in the context of discrete memoryless channels. Polyanskiy and Verdú [17] relaxed some assumptions in the conference version of Altuğ and Wagner's work [18] and they also considered moderate deviations for AWGN channels. Altuğ, Wagner and Kontoyiannis [19] considered moderate deviations for lossless source coding. Other works on moderate deviations in information theory include [20], [21]. Note, however, that all these cited works on moderate deviations analysis in information theory pertain to point-to-point systems with a single rate parameter. In this paper, we perform moderate deviations analysis on a multi-terminal problem involving two rates and, additionally, we consider the streaming scenario.

B. Main Contributions

In the sequel, we define blockwise and symbolwise codes precisely. Blockwise codes correspond to the scenario where we decode a block of symbol pairs with a delay. Symbolwise codes correspond to the scenario where we decode certain symbol pairs with the same delay. Given these definitions, the contribution of this paper is twofold.

For error exponents, we derive an upper bound and a lower bound for the optimal error exponent of streaming SW coding for both blockwise and symbolwise codes. We show that the upper and lower bounds for blockwise codes coincide for a non-empty set of rate pairs. For symbolwise codes, for a limited set of symbol sequences, the upper and lower bounds coincide for the same non-empty set of rate pairs. It turns out that, under both conditions, the error exponents for blockwise and symbolwise codes coincide. We utilize the information spectrum method to derive a strong converse for blockwise codes; we resort to the feedforward decoder idea to derive a weak converse for symbolwise codes. Then, with a change of measure technique, we derive upper bounds on the optimal error exponents. To derive lower bounds on the optimal error exponents, for both blockwise and symbolwise codes, we rely on the random binning idea and we use both maximum likelihood decoding and minimum empirical entropy decoding.

To obtain the moderate deviations constants for blockwise and symbolwise codes, we leverage our error exponent results and use Taylor series expansions appropriately. One distinguishing feature relative to the existing works on moderate deviations in information theory is that we need to consider multiple first-order fundamental limits on the boundary of the optimum rate region (similar to the study of second-order coding rates in network information theory [5], [22], [23]). Thus, the moderate deviation constants depend on the boundary point we are operating in the vicinity of. Furthermore, the strong converse for blockwise codes, established using the information spectrum method [24], [25], allows us to prove our result for rates that tend towards the boundary at a speed slower than \( \Delta^{-1/4} \), where \( \Delta \in \mathbb{N} \) is the delay in decoding. This speed is usual in studies on moderate derivation analysis in information theory. For symbolwise codes, because we are unable to show a strong converse, we are only able to derive moderate deviations constants for rates that tend to the boundary at a speed slower than \( \Delta^{-1/4} \).
C. Organization Of The Paper

This paper is organized as follows. In Section II, we set up the notation and define the various problems we aim to solve precisely. In particular, we define the notation of streaming error exponents in the blockwise and symbolwise settings and their moderate deviations counterparts. We also review existing error exponent results for the traditional (non-streaming) SW problem. In Section III, we present our error exponent results and state a condition on the rate pairs for which our result is tight. We also present a simple numerical example to illustrate the derived exponents and to gain intuition for the large delay and small delay scenarios. In Section IV, we present our moderate deviation results. We conclude our discussion and state avenues for further research in Section V. To ensure a smooth flow in the presentation of the results, we defer all proofs to the appendices.

II. Preliminaries

Notation: Random variables and their realizations are in capital (e.g., \(X\)) and lower case (e.g., \(x\)) respectively. All sets are denoted in calligraphic font (e.g., \(\mathcal{X}\)). Let \(X^n := (X_1, \ldots, X_n)\) be a random vector of length \(n\). Information-theoretic quantities are denoted as in the book by Csiszár and Körner [3]. We use \(\exp(x)\) to denote \(2^x\). We use \(\lceil x \rceil\) to denote the largest integer not greater than \(x\). All logarithms are base 2.

A. Problem Formulation

We consider a discrete memoryless correlated source \((X,Y)\) with joint distribution \(P_{XY}\) on \(\mathcal{X} \times \mathcal{Y}\). Both sources produce one symbol per unit time. The decoder decodes a sequence of symbols or a symbol pair with a fixed delay.

The encoders and the decoders are defined in a similar manner as in [6] and [8]. We recapitulate the definitions here.

**Definition 1.** A sequential rate-\((R_X, R_Y)\) encoder pair \(\{(f_j, g_j)\}_{j \in \mathbb{N}}\) is a sequence of mappings such that
\[
\begin{align*}
f_j &: \mathcal{X}^j \rightarrow \{0,1\}^{\lfloor (j+\Delta)R_X \rfloor - \lfloor (j-1)R_X \rfloor}, \\
g_j &: \mathcal{Y}^j \rightarrow \{0,1\}^{\lfloor (j+\Delta)R_Y \rfloor - \lfloor (j-1)R_Y \rfloor}.
\end{align*}
\]

In this paper, we consider the scenario where \(n + \Delta\) independent copies of \((X_j, Y_j)\) are generated according to \(P_{XY}\). Here, \(\Delta \in \mathbb{N}\) denotes the delay in decoding and \(n\) denotes the number of source symbols we are interested to decode. We remark that in the above definition, \(f_j\) (resp. \(g_j\)) maps \(x^j\) (resp. \(y^j\)) to approximately \(R_X\) (resp. \(R_Y\)) bits.

The encoded messages for the source \(\{X_j\}_{j=1}^{n+\Delta}\) is denoted as \(\prod_{j=1}^{n+\Delta} f_j(X^j)\) and similarly the encoded messages for the source \(\{Y_j\}_{j=1}^{n+\Delta}\) is denoted as \(\prod_{j=1}^{n+\Delta} g_j(Y^j)\).

In this paper, the delay \(\Delta\) scales linearly with the number of symbol pairs of interest \(n\) as follows:
\[
\Delta = \lfloor \alpha n \rfloor.
\]
The finite positive parameter \(\alpha\) controls the tradeoff between the delay and the number of symbols of interest. This scaling law was motivated in the Introduction.

**Definition 2.** A sequential rate-\((R_X, R_Y)\) blockwise decoder with fixed delay \(\Delta\) is a sequence of mappings,
\[
\phi_{j+\Delta}^B : \{0,1\}^{\lfloor (j+\Delta)R_X \rfloor} \times \{0,1\}^{\lfloor (j+\Delta)R_Y \rfloor} \rightarrow \mathcal{X}^j \times \mathcal{Y}^j
\]
where \(j \in \mathbb{N}\). A sequential blockwise rate-\((R_X, R_Y)\) code with fixed delay \(\Delta\) is a sequential rate-\((R_X, R_Y)\) encoder pair and a sequential rate-\((R_X, R_Y)\) blockwise decoder with fixed delay \(\Delta\).

Note that at time \(j + \Delta\), only the first \(j\) symbol pairs \(\{(X_i, Y_i)\}_{i=1}^{j}\) are to be decoded. The blockwise error event at blocklength \(n\) is
\[
\left\{ \phi_{n+\Delta}^B \left( \prod_{j=1}^{n+\Delta} f_j(X^j), \prod_{j=1}^{n+\Delta} g_j(Y^j) \right) \neq (X^n, Y^n) \right\},
\]
or simply \(\{(\hat{X}^n, \hat{Y}^n) \neq (X^n, Y^n)\}\).

**Definition 3.** A sequential rate-\((R_X, R_Y)\) symbolwise decoder with fixed delay \(\Delta\) is a sequence of mappings,
\[
\phi_{j+\Delta}^S : \{0,1\}^{\lfloor (j+\Delta)R_X \rfloor} \times \{0,1\}^{\lfloor (j+\Delta)R_Y \rfloor} \rightarrow \mathcal{X} \times \mathcal{Y}
\]
where \(j \in \mathbb{N}\). A sequential symbolwise rate-\((R_X, R_Y)\) code with fixed delay \(\Delta\) is a sequential rate-\((R_X, R_Y)\) encoder pair and a sequential rate-\((R_X, R_Y)\) symbolwise decoder with fixed delay \(\Delta\).

At time \(j + \Delta\), only symbol pair \((X_j, Y_j)\) is decoded. The \(j\)-th error event is
\[
\left\{ \phi_{j+\Delta}^S \left( \prod_{i=1}^{j+\Delta} f_i(X^j), \prod_{i=1}^{j+\Delta} g_i(Y^j) \right) \neq (X_j, Y_j) \right\},
\]
or more simply \( \{(\hat{X}_j, \hat{Y}_j) \neq (X_j, Y_j)\} \).

In traditional SW coding \([1]\), there is only one pair of encoders. At the time of compression \( n \), encoder \( X \) has knowledge of all \( n \) symbols \( x^n \) and compresses these symbols into a binary sequence with length \( nR_X \) (and similarly for the \( Y \) encoder). However, in the streaming context, as (1) and (2) indicate, there are numerous encoder pairs at blocklength \( n \).

The decoder in (4) incurs a delay of \( \Delta \) to decode the first \( j \) symbol pairs because its domain is a pair of length-\((j + \Delta)R_X, (j + \Delta)R_Y \) bits. The decoder in (6) is similar to the one in (4) except that it estimates only the \( j \)-th symbol pair. See Fig. 2 for an illustration.

**Definition 4.** A number \( E \geq 0 \) is said to be a rate-\((R_X, R_Y)\) achievable blockwise error exponent if for any \( \delta > 0 \), there exists a sequential blockwise rate-\((R_X, R_Y)\) code with fixed delay \( \Delta \) such that

\[
\Pr \left( (\hat{X}^n, \hat{Y}^n) \neq (X^n, Y^n) \right) \leq \exp \left( -\Delta (E - \delta) \right)
\]

for all \( \Delta \) large enough. The supremum of all rate-\((R_X, R_Y)\) achievable blockwise error exponents is denoted as \( E_{bi}^R(R_X, R_Y) \).

**Definition 5.** Consider an arbitrary sequence \( \{j_n\}_{n \in \mathbb{N}} \subset \mathbb{N} \) where \( j_n \leq n \). A number \( \nu \geq 0 \) is said to be a rate-\((R_X, R_Y)\) achievable symbolwise error exponent with respect to \( \{j_n\}_{n \in \mathbb{N}} \) if for any \( \delta > 0 \), there exists a sequential symbolwise rate-\((R_X, R_Y)\) code with fixed delay \( \Delta \) such that

\[
\Pr \left( (\hat{X}_{j_n}, \hat{Y}_{j_n}) \neq (X_{j_n}, Y_{j_n}) \right) \leq \exp \left( -\Delta (E - \delta) \right)
\]

for all \( \Delta \) large enough. The supremum of all rate-\((R_X, R_Y)\) achievable symbolwise error exponents with respect to \( \{j_n\}_{n \in \mathbb{N}} \) is denoted as \( E_{sy}^R(R_X, R_Y, \{j_n\}) \).

We emphasize that the symbolwise decoder is only interested in the \( j_n \)-th symbol pair \( (X_{j_n}, Y_{j_n}) \) at time \( j_n + \Delta \).

To define the next concept succinctly, let \( R_X^{(\Delta)} \) and \( R_Y^{(\Delta)} \) be two fixed rates. The rate pair \( \{R_X^{(\Delta)}, R_Y^{(\Delta)}\} \) will be taken to be on the boundary of the optimal rate region subsequently in Section IV. To motivate the next two definitions, observe that in Definitions 4 and 5, the rate pair \( \{R_X, R_Y\} \) is fixed and it is usually in the interior of the (first-order) optimal rate region, while the error probability decays exponentially fast in the delay. We would like our rates to be arbitrarily close to the boundary of the optimal rate region. The price we have to pay is in terms of the error probability which now decays subexponentially fast. We formalize these concepts in the next two definitions.

**Definition 6** (Blockwise Moderate Deviations Constant). Consider any correlated source with joint probability mass function (pmf) \( P_{XY} \) and any sequence \( \{\epsilon_{\Delta}\}_{\Delta=1}^{\infty} \) satisfying

\[
\epsilon_{\Delta} \to 0 \quad \text{as} \quad \Delta \to \infty, \\
\Delta\epsilon_{\Delta}^2 \to 0 \quad \text{as} \quad \Delta \to \infty.
\]

Let \( \theta_1 \) and \( \theta_2 \) be two real numbers. Define

\[
R_X^{(\Delta)} = R_X^* + \theta_1 \epsilon_{\Delta}, \quad R_Y^{(\Delta)} = R_Y^* + \theta_2 \epsilon_{\Delta}.
\]

A number \( \nu \) is said to be a rate-\((R_X^{(\Delta)}, R_Y^{(\Delta)})\) achievable blockwise moderate deviations constant if for any \( \delta > 0 \), there exists a sequential blockwise rate-\((R_X^{(\Delta)}, R_Y^{(\Delta)})\) code with fixed delay \( \Delta \) such that

\[
\Pr \left( (\hat{X}^n, \hat{Y}^n) \neq (X^n, Y^n) \right) \leq \exp \left( -\Delta \epsilon_{\Delta}^2 (\nu - \delta) \right)
\]

for all \( \Delta \) large enough. The supremum of all achievable blockwise rate-\((R_X^{(\Delta)}, R_Y^{(\Delta)})\) moderate deviations constants is denoted as \( \nu_{bi}^* \).}

**Definition 7** (Symbolwise Moderate Deviations Constant). Consider any correlated source with joint pmf \( P_{XY} \) and any sequence \( \{\epsilon_{\Delta}\}_{\Delta=1}^{\infty} \) satisfying (10). Consider an arbitrary sequence \( \{j_n\}_{n \in \mathbb{N}} \subset \mathbb{N} \) where \( j_n \leq n \). Let \( \theta_1 \) and \( \theta_2 \) be two real numbers. A number \( \nu \) is said to be a rate-\((R_X^*, R_Y^*)\) achievable symbolwise moderate deviations constant with respect to \( \{j_n\}_{n \in \mathbb{N}} \) if for any \( \delta > 0 \), there exists a sequential symbolwise rate-\((R_X^{(\Delta)}, R_Y^{(\Delta)})\) (defined in (11)) code with fixed delay \( \Delta \) such that

\[
\Pr \left( (\hat{X}_{j_n}, \hat{Y}_{j_n}) \neq (X_{j_n}, Y_{j_n}) \right) \leq \exp \left( -\Delta \epsilon_{\Delta}^2 (\nu - \delta) \right)
\]
for all $\Delta$ large enough. The supremum of all achievable symbolwise rate-$(R_X^*, R_Y^*)$ moderate deviations constants with respect to $\{j_n\}_{n \in \mathbb{N}}$ is denoted as $\nu_{sw}^*(R_X^*, R_Y^*, \{j_n\})$.

### B. Error Exponents for Traditional SW Problem

In this section, we review known results pertaining to the error exponents for the traditional (non-streaming) SW problem. In SW coding [1], the encoders have access to all $n$ symbols at the time of compression and the decoder decodes all $n$ symbol pairs. Rate-$(R_X, R_Y)$ encoders can be defined as $f : \mathcal{X}^n \rightarrow \{0, 1\}^{nR_X}$ and $g : \mathcal{Y}^n \rightarrow \{0, 1\}^{nR_Y}$. The decoder is defined as $\phi : \{0, 1\}^{nR_X} \times \{0, 1\}^{nR_Y} \rightarrow \mathcal{X}^n \times \mathcal{Y}^n$.

For correlated sources with joint pmf $P_{XY}$, the optimal rate region of SW coding is given by

$$\mathcal{R}_{sw}(P_{XY}) := \left\{ (R_X, R_Y) : R_X \geq H(P_X|P_{X|Y}), \ R_Y \geq H(P_Y|P_{X|Y}), \ R_X + R_Y \geq H(P_{XY}) \right\}. \quad (14)$$

Denote the optimal error exponent as $E_{sw}^*(R_X, R_Y)$, which is the exponential speed at which the error probability decays with respect to $n$.

**Definition 8.** The Gallager exponents [26] are defined as

$$E_{X|Y}(R_X, \rho) := \rho R_X - \log \sum_y P_Y(y) \left( \sum_x P_{X|Y}(x|y)^{\frac{1}{1+\rho}} \right)^{1+\rho}, \quad (15)$$

$$E_{XY}(R_X, R_Y, \rho) := \rho (R_X + R_Y) - (1 + \rho) \log \sum_{x,y} P_{XY}(x,y)^{\frac{1}{1+\rho}}, \quad (16)$$

and $E_{X|Y}(R_X, \rho)$ is similar to $E_{X|Y}(R_X, \rho)$ with $X$ and $Y$ interchanged.

**Definition 9.** The random coding error exponent is

$$E_{sw}(R_X, R_Y) := \min \left\{ E_{sw}^{X|Y}(R_X), \ E_{sw}^{Y|X}(R_Y), \ E_{sw}^{XY}(R_X, R_Y) \right\}. \quad (17)$$

where

$$E_{sw}^{X|Y}(R_X) := \max_{\rho \in [0,1]} E_{X|Y}(R_X, \rho),$$

$$E_{sw}^{Y|X}(R_Y) := \min_{\rho \in [0,1]} D(Q_{XY} \| P_{XY}) + |H(Q_{X|Y}) - R_X|^+, \quad (18)$$

$$E_{sw}^{XY}(R_X, R_Y) := \max_{\rho \in [0,1]} E_{XY}(R_X, R_Y, \rho),$$

and $E_{sw}^{Y|X}(R_Y)$ is similar to $E_{sw}^{X|Y}(R_X)$ with $X$ and $Y$ interchanged.

**Definition 10.** The sphere packing error exponent is

$$E_{sw}(R_X, R_Y) := \min \left\{ E_{sw}^{X|Y}(R_X), \ E_{sw}^{Y|X}(R_Y), \ E_{sw}^{XY}(R_X, R_Y) \right\}. \quad (20)$$

where

$$E_{sw}^{X|Y}(R_X) := \sup_{\rho \geq 0} E_{X|Y}(R_X, \rho),$$

$$E_{sw}^{X|Y}(R_X) := \min_{Q_{XY} : H(Q_{X|Y}) \geq R_X} D(Q_{XY} \| P_{XY}), \quad (21)$$

$$E_{sw}^{XY}(R_X, R_Y) := \sup_{\rho \geq 0} E_{XY}(R_X, R_Y, \rho),$$

and $E_{sw}^{Y|X}(R_Y)$ is similar to $E_{sw}^{X|Y}(R_X)$ with $X$ and $Y$ interchanged.

**Theorem 1.** The optimal error exponent for traditional SW coding satisfies

$$E_{sw}(R_X, R_Y) \leq E_{sw}^*(R_X, R_Y) \leq \mathcal{E}_{sw}(R_X, R_Y). \quad (23)$$
This classical result is from [3] and [26]. Further from Exercise 2.15 of [3], we can conclude that the lower and upper bounds coincide in the low rate setting in Proposition 2. Define the set

\[ R(P_{XY}) := \{(R_X, R_Y) \in \mathbb{R}^2_+ : H(P_{X|Y}|P_Y) \leq R_X \leq H(P_{X|Y}|P_Y)_{\rho=1} \]

\[ H(P_{Y|X}|P_X) \leq R_Y \leq H(P_{Y|X}|P_X)_{\rho=1} \]

\[ H(P_{XY}) \leq R_X + R_Y \leq H(P_{XY})_{\rho=1} \],

(24)

where \( P_{X|Y}, P_{Y|X} \) are tilted conditional distributions, \( P_{X,Y} \) is a tilted joint distribution and \( P_{X}, P_{Y} \) are tilted marginal distributions, i.e.,

\[ P_{X|Y}(x|y) := \frac{P_{X|Y}(x|y)^{1+\rho}}{\sum_x P_{X|Y}(x|y)^{1+\rho}}, \]

(25)

\[ P_{Y}(y) := \frac{P_{Y}(y) \left( \sum_x P_{X|Y}(x|y)^{1+\rho} \right)^{1+\rho}}{\sum_y P_{Y}(y) \left( \sum_x P_{X|Y}(x|y)^{1+\rho} \right)^{1+\rho}}, \]

(26)

\[ P_{X,Y}(x,y) := \frac{P_{X,Y}(x,y)^{1+\rho}}{\sum_{x,y} P_{X,Y}(x,y)^{1+\rho}}. \]

(27)

**Proposition 2.** For any rate pair \((R_X, R_Y) \in R(P_{XY})\),

\[ E_{sw}(R_X, R_Y) = E_{sw}(R_X, R_Y). \]

(28)

Thus, for all rate pairs in \( R(P_{XY}) \), the error exponent of SW coding is known and equal to the common value of \( E_{sw}(R_X, R_Y) \) and \( E_{sw}(R_X, R_Y) \).

### III. Error Exponents

In this section, we present our error exponent results and in addition, we provide a condition for which we have a conclusive (tight) characterization of the error exponents. We also provide a numerical example to illustrate the various exponents. This example also serves as a tool to understand the effects of a small delay (\( \alpha \downarrow 0 \)) and a large delay (\( \alpha \uparrow \infty \)).

#### A. Basic Definitions

Before doing so, we define some notation. Define the set

\[ R'(P_{XY}) := \{(R_X, R_Y) \in \mathbb{R}^2_+ : H(P_{X|Y}|P_Y) \leq R_X < R_{X|Y}(\infty) \]

\[ H(P_{Y|X}|P_X) \leq R_Y < R_{Y|X}(\infty) \]

\[ H(P_{XY}) \leq R_X + R_Y < R_{XY}(\infty) \}, \]

(29)

where

\[ R_{X|Y}(\infty) = \arg \min_{R_X} \left\{ \inf_{Q_{XY}: H(Q_{X|Y}|Q_Y) \geq R} D(Q_{XY}\|P_{XY}) = \infty \right\} = \log |X|, \]

(30)

\[ R_{XY}(\infty) = \arg \min_{R} \left\{ \inf_{Q_{XY}: H(Q_{X|Y}) \geq R} D(Q_{XY}\|P_{XY}) = \infty \right\} = \log (|X||Y|), \]

(31)

and \( R_{Y|X}(\infty) \) is similar to \( R_{X|Y}(\infty) \) with \( X \) and \( Y \) interchanged. Note that \( R'(P_{XY}) \supset R(P_{XY}). \)

#### B. Results and Interpretations

We now present bounds on the blockwise error exponent.

**Theorem 3** (Blockwise Error Exponent). For any rate pair \((R_X, R_Y)\),

\[ \frac{1}{\alpha}E_{sw}((1+\alpha) R_X, (1+\alpha) R_Y) \leq E_{bl}^*(R_X, R_Y) \leq \frac{1}{\alpha}E_{sw}((1+\alpha) R_X, (1+\alpha) R_Y). \]

(32)

The proof of Theorem 3 is provided in Appendices A and B.

Observe that \( E_{bl}^*(R_X, R_Y) \) is positive if and only if \((1+\alpha) R_X, (1+\alpha) R_Y) \in \text{int}(R_{sw})\). This implies the optimal rate region is enlarged vis-à-vis regular SW coding. However, the system suffers from a delay of \( \Delta = \lfloor \alpha n \rfloor \).
We now compare the result in Theorem 3 to a relevant existing work in the literature. Specifically, Draper, Chang and Sahai (DCS) [6] derived a lower bound for the error exponent for blockwise streaming SW codes. In their setting, they decode \( n' := n - \Delta \) symbol pairs using the encoded messages of \( n = n' + \Delta \) symbol pairs. Define
\[
E_X(R_X, R_Y, \gamma) := \max_{\rho \in [0,1]} \left[ \gamma E_{XY}(R_X, \rho)(1 - \gamma)E_{XY}(R_X, R_Y, \rho) \right],
\]
\[
E_Y(R_X, R_Y, \gamma) := \max_{\rho \in [0,1]} \left[ \gamma E_{XY}(R_X, \rho)(1 - \gamma)E_{XY}(R_X, R_Y, \rho) \right].
\]

In [6, Theorem 6], DCS showed that a lower bound for their setting, in which the delay \( \Delta \) does not necessarily scale linearly with the number of symbol pairs of interest \( n' \), is given by
\[
E_{\text{DCS}}^{\text{B}}(R_X, R_Y) = \min \left\{ \inf_{\gamma \in [0,1]} E_X(R_X, R_Y, \gamma), \inf_{\gamma \in [0,1]} E_Y(R_X, R_Y, \gamma) \right\}.
\]

For \( \alpha > 0 \), we can verify that (cf. Appendix E)
\[
E_{\text{DCS}}^{\text{B}}(R_X, R_Y) \leq E_{\text{sw}}(R_X, R_Y) \leq \frac{1}{\alpha} E_{\text{sw}}((1 + \alpha)R_X, (1 + \alpha)R_Y).
\]

Note that the setting in [6] is more general because the authors do not require the delay to scale linearly with the number of symbol pairs of interest as we do in (3). Because the additional restriction on \( \Delta \) and \( n \) imposed by (3), we can show that our lower bound for the error exponent matches the upper bound for blockwise streaming SW codes in the low rate regime. See Corollary 5 in the sequel.

We now present bounds on the symbolwise error exponent.

**Theorem 4 (Symbolwise Error Exponent).** For any rate pair \((R_X, R_Y)\) satisfying \((1 + \alpha)R_X, (1 + \alpha)R_Y) \in \mathcal{R}'(P_{XY})\) and any sequence \(\{j_n\}\),
\[
E_{\text{sw}}^*(R_X, R_Y, \{j_n\}) \leq \frac{1}{\alpha} E_{\text{sw}}((1 + \alpha)R_X, (1 + \alpha)R_Y).
\]

For any sequence \(\{j_n\}\) satisfying
\[
\lim_{n \to \infty} \frac{j_n}{n} = 1,
\]
we have
\[
E_{\text{sw}}^*(R_X, R_Y, \{j_n\}) \geq \frac{1}{\alpha} E_{\text{sw}}((1 + \alpha)R_X, (1 + \alpha)R_Y).
\]

The proof of Theorem 4 is given in Appendices B and C. The upper bound for the symbolwise error exponent only holds in the low rate setting (i.e., \((1 + \alpha)R_X, (1 + \alpha)R_Y) \in \mathcal{R}'(P_{XY})\)). This arises due to our adaptation of the feedforward proof technique by Chang and Sahai [8] in our proofs. For symbolwise decoders with delay \(\Delta\), the lower bound in (39) holds only if we wish to decode symbols with indices \(j_n\) where \(j_n\) is comparable to \(n\); cf. (38).

The results in Theorems 3 and 4 are similar to the focusing bound [7, 27], with the only difference that there is no minimization over \(\alpha > 0\) in the derived exponents. The authors of [7] and [27] minimize over all \(\alpha\) since they consider all possible relationships between the delay \(\Delta\) and the number of symbols of interest \(n\). However, we fix the delay \(\Delta\) to scale linearly with \(n\) in (3).

It is natural to ask whether the bounds in Theorems 3 and 4 are tight. This is answered by the following corollary which is a consequence of Corollary 2.

**Corollary 5.** Consider the set \(\mathcal{R}(P_{XY})\) defined in (24). For any rate pair \((R_X, R_Y)\) such that
\[
((1 + \alpha)R_X, (1 + \alpha)R_Y) \in \mathcal{R}(P_{XY}),
\]
we have
\[
E_{\text{sw}}((1 + \alpha)R_X, (1 + \alpha)R_Y) = E_{\text{sw}}((1 + \alpha)R_X, (1 + \alpha)R_Y) = E_{\text{sw}}^*((1 + \alpha)R_X, (1 + \alpha)R_Y).
\]

Furthermore, for any rate pair \((R_X, R_Y)\) satisfying (40) and for all \(\{j_n\}\) satisfying (38), we have
\[
E_{\text{bl}}^*(R_X, R_Y, \{j_n\}) = E_{\text{sw}}^*((1 + \alpha)R_X, (1 + \alpha)R_Y).
\]

From (42), we see that the symbolwise and blockwise error exponents coincide under conditions (38) and (40). The intuition behind this is as follows: For symbolwise codes, if the condition in (38) holds, asymptotically, we decode all symbol pairs with indices 1 to \(j_n \approx n\) (with delay \(\Delta\)). This is similar to blockwise codes where we also decode symbol pairs with indices 1 to \(j_n\). However, we note that for symbolwise codes at time \(j_n + \Delta\), we are only interested in the symbol with index \(j_n\) but for blockwise codes we are interested in all symbols with indices 1 to \(j_n\). Somewhat surprisingly, the optimal exponents are the same in the regime indicated by (40) in Corollary 5.
C. A Numerical Example

In this subsection, we present a numerical example depicting the exponents in Theorems 3 and 4 and Corollary 5. For ease of presentation, we explore the symmetric rate setting, i.e., $R_X = R_Y = r \geq 0$. In Fig. 3, we illustrate the functions $\overline{E}(\alpha, r) := \frac{1}{\alpha} E_{\text{sw}}((1 + \alpha) r, (1 + \alpha) r)$ and $\overline{E}(\alpha, r) := \frac{1}{\alpha} E_{\text{sw}}((1 + \alpha) r, (1 + \alpha) r)$ for various values of the scaling parameter $\alpha$ and the fixed rate $r$.

First, we notice that Fig. 3 corroborates Corollary 5 since $\overline{E}(\alpha, r) = \overline{E}(\alpha, r)$ for (small) values of $\alpha$ which satisfy (40). In fact for all $\alpha \in [0, 0.2]$ and for all chosen values of $r$, the two sets of curves in the left and right plots coincide.

Second, we observe that when $\alpha \downarrow 0$, the error exponents $\overline{E}(\alpha, r)$ and $\overline{E}(\alpha, r)$ diverge to infinity. For blockwise codes, the intuitive reason for this is the delay $\Delta$ is now an infinitesimal fraction of the length of the block $n$ and we defined the error exponents in terms of the delay $\Delta$ (cf. Definitions 4 and 5). However, when $\alpha \downarrow 0$, we revert to the original SW setting in which it is known that the error probability decays exponentially fast with the blocklength $n$ when the rate pair $(R_X, R_Y)$ is in the interior of the SW region (14), i.e., $\Pr\{(\hat{X}^n, \hat{Y}^n) \neq (X^n, Y^n)\} \approx \exp(-nE)$ for some constant $E > 0$. Since $\Delta = o(n)$, the corresponding error exponent defined with respect to $\Delta$ is infinite.

Finally, we also observe that when $\alpha \uparrow \infty$, $\overline{E}(\alpha, r)$ converges to $r$ while $\overline{E}(\alpha, r)$ diverge to infinity. The first fact can be shown analytically as follows: Recalling the definition of $E_{\text{sw}}(R_X, R_Y)$ in (17), as $\alpha \uparrow \infty$, we obtain

$$\frac{1}{\alpha} E_{\text{sw}}^{XY}((1 + \alpha) R_X, (1 + \alpha) R_Y) = \max_{\rho \in [0, 1]} \left\{ \rho (R_X + R_Y) - \frac{1}{\alpha} E_{\text{sw}}^{XY}(R_X, R_Y) \right\} \rightarrow R_X + R_Y = 2r, \quad (43)$$

$$\frac{1}{\alpha} E_{\text{sw}}^{X|Y}((1 + \alpha) R_X) = \max_{\rho \in [0, 1]} \left\{ \rho R_X - \frac{1}{\alpha} E_{\text{sw}}^{X|Y}(R_X) \right\} \rightarrow R_X = r, \quad (44)$$

$$\frac{1}{\alpha} E_{\text{sw}}^{Y|X}((1 + \alpha) R_Y) = \max_{\rho \in [0, 1]} \left\{ \rho R_Y - \frac{1}{\alpha} E_{\text{sw}}^{Y|X}(R_Y) \right\} \rightarrow R_Y = r. \quad (45)$$

Since $r$ is non-negative, as $\alpha \uparrow \infty$,

$$\overline{E}(\alpha, r) = \frac{1}{\alpha} \min \left\{ E_{\text{sw}}^{XY}((1 + \alpha) R_X, (1 + \alpha) R_Y), E_{\text{sw}}^{X|Y}((1 + \alpha) R_X), E_{\text{sw}}^{Y|X}((1 + \alpha) R_Y) \right\} \rightarrow r. \quad (46)$$

Similarly, we can conclude $\overline{E}(\alpha, r) \uparrow \infty$ as $\alpha \uparrow \infty$ because the maximizations in (43)–(45) over $\rho$ are now constrained to be the interval $[0, 1]$. In this large-$\alpha$ regime, we do not have a tight characterization of the optimal error exponent (cf. Corollary 5). All we can infer from the results is that the optimal error exponent is at least $r$ and potentially much larger. This can be understood intuitively by considering blockwise codes. Assume, for the moment, we decode $\Delta$ (and not the much smaller $n$) symbol pairs. Then we revert to the original SW setting in which it is known that the error probability decays exponentially fast with delay $\Delta$ when the rate pair $(R_X, R_Y)$ is in the interior of the SW region (14), i.e., $\Pr\{(\hat{X}^\Delta, \hat{Y}^\Delta) \neq (X^\Delta, Y^\Delta)\} \approx \exp(-\Delta E)$ for
some constant $E > 0$. Since we in fact only decode $n \ll \Delta$ symbol pairs, the exponent is potentially much larger than the constant $E$.

IV. MODERATE DEVIATIONS

In this section, we present corresponding results for moderate deviations asymptotics in which the rate pairs converge to a first-order fundamental limit while the error probability decays subexponentially fast. To present our results succinctly, we define the following fundamental quantities for a joint source $P_{XY}$.

For a given source with pmf $P_{XY}$ on alphabet $\mathcal{X} \times \mathcal{Y}$, the joint and conditional source dispersions or varentropy [28] are defined respectively as

$$V(P_{XY}) := \sum_{x,y} P_{XY}(x,y) \left( -\log P_{XY}(x,y) - H(P_{XY}) \right)^2,$$

and $V(P_{X|Y}|P_Y)$ is defined similar to $V(P_{X|Y}|P_Y)$ with $X$ and $Y$ interchanged. We assume that for the correlated source with joint pmf $P_{XY}$ on the alphabet $\mathcal{X} \times \mathcal{Y}$, the three source dispersions $V(P_{XY})$, $V(P_{X|Y}|P_Y)$, and $V(P_{Y|X}|P_X)$ are positive.

**Theorem 6** (Blockwise Moderate Deviations Constant). For any sequence $\{\epsilon_\Delta\}$ satisfying (10), depending on $(R^*_X, R^*_Y)$, there are five cases, as shown in Fig. 4, of which we present three.

- **Case (i):** $(1 + \alpha)R^*_X = H(P_{X|Y}|P_Y)$ and $(1 + \alpha)R^*_Y > H(P_Y)$ (vertical boundary)

$$\nu^*_b(R^*_X, R^*_Y) = \frac{(1 + \alpha)^2\theta_1^2 \log e}{2\alpha V(P_{XY})}. \quad (49)$$

- **Case (ii):** $(1 + \alpha)R^*_X = H(P_{X|Y}|P_Y)$ and $(1 + \alpha)R^*_Y = H(P_Y)$ (left corner point)

$$\nu^*_b(R^*_X, R^*_Y) = \min \left\{ \frac{(1 + \alpha)^2(\theta_1 + \theta_2)^2 \log e}{2\alpha V(P_{XY})}, \frac{(1 + \alpha)^2(\theta_1 + \theta_2)^2 \log e}{2\alpha V(P_{X|Y}|P_Y)} \right\}. \quad (50)$$

- **Case (iii):** $(1 + \alpha)(R^*_X + R^*_Y) = H(P_{XY})$, $(1 + \alpha)R^*_X > H(P_{X|Y}|P_Y)$ and $(1 + \alpha)R^*_Y > H(P_{Y|X}|P_X)$ (diagonal face)

$$\nu^*_b(R^*_X, R^*_Y) = \frac{(1 + \alpha)^2(\theta_1 + \theta_2)^2 \log e}{2\alpha V(P_{XY})}. \quad (51)$$

**Theorem 7** (Symbolwise Moderate Deviations Constant). For any sequence $\{\epsilon_\Delta\}$ satisfying (10) and

$$\Delta \epsilon^2_\Delta \to \infty, \quad \Delta \to \infty,$$  

and any sequence $\{j_\Delta\}$ satisfying (38), for all the cases in Theorem 6,

$$\nu^*_s(R^*_X, R^*_Y, \{j_\Delta\}) = \nu^*_b(R^*_X, R^*_Y). \quad (53)$$

The proofs of Theorems 6 and 7 are provided in Appendix D. As can be seen from Fig. 4, cases (iv) and (v) are symmetrical to cases (ii) and (i) respectively (with $X$ and $Y$ interchanged) and we omit the results. We now provide four remarks concerning the moderate deviations results above.

First, case (ii) in (50) (corner point) is the most interesting one. Here, we observe that the moderate deviations constant is the minimum of two terms. Each of these two terms corresponds to a certain error event. The first term in the minimization in
(50) corresponds to the joint error event while the second corresponds to the marginal event concerning the reconstruction of the source $X^n$. Similar to the large deviations principle, in the moderate deviations regime, the smallest moderate deviations constant dominates and so we observe that the moderate deviations constant in case (ii) is as in (50). This behavior is different from that in second-order analysis [22], [23] in which the second-order rate region depends on both rate constraints.

Second, for blockwise codes, we use the information spectrum method to first prove the strong converse for blockwise streaming SW codes. Next, we utilize the strong converse and a measure changing step to establish an appropriate upper bound on the blockwise error exponent. Using the same change of measure technique to establish the upper bound on the blockwise error exponent.

Finally, Hayashi and Matsumoto derived an achievability result for the moderate deviations constants for traditional SW coding in Lemma 89 of [29]. Our result differs from theirs primarily in the additional term $(1+\alpha)^2$, which again results from the fact that we fix the delay to scale linearly with the number of symbol pairs of interest in (3). Furthermore, we observe the presence of constants $\theta_1$ and $\theta_2$ in Theorem 6 is due to the definition of $R_X^\Delta$ and $R_Y^\Delta$ in (11).

V. CONCLUSION

In this paper, the error exponent functions and moderate deviations constants are derived for streaming SW coding under both blockwise and symbolwise settings. We show that the bounds for the error exponent function are tight for blockwise codes in a certain low rate regime. For symbolwise codes, under the same rate pair requirement and with a certain constraint on the indices of the symbol pairs, the bounds are also tight. Somewhat surprisingly, the bounds under the above conditions are the same for blockwise and symbolwise codes. The derived error exponent results hold when the delay $\Delta$ scales linearly with the number of symbol pairs of interest $n$. It may be interesting to explore if tight error exponents can be derived when $\Delta$ varies arbitrary with $n$.

We also derived moderate deviations constants for streaming SW coding under both blockwise and symbolwise settings. To obtain the full moderate deviations regime for the blockwise setting, we appealed to the information spectrum method to first obtain a strong converse. For the symbolwise setting, under an additional limiting requirement, we showed that the moderate deviations constants are the same as the blockwise setting. The natural question is whether this additional requirement can be relaxed to obtain the full moderate deviations regime for the symbolwise setting.

There are several other avenues in which the current research can be extended. One is to provide bounds on the excess distortion exponent when the reconstruction of the sources may be done in a lossy manner. In addition, one may also study the streaming counterpart of the source coding with compressed side-information problem [30], [31].

APPENDIX

A. Proof of Upper Bound in Theorem 3

For blockwise codes, we use the information spectrum method [24] to obtain the strong converse. Subsequently, we use a change of measure technique to establish the upper bound on the blockwise error exponent.

1) Strong Converse: For blockwise streaming codes, when the inflated rate pair $((1+\alpha)R_X,(1+\alpha)R_Y)$ is in the exterior of the usual SW region, the error probability tends to one. To prove this result, we first define an $(n,N_X,N_Y,\Delta,\varepsilon)$-blockwise code. This is a slight refinement of the definition of a sequential blockwise rate-$(R_X,R_Y)$ code with delay $\Delta$ per Definition ?? except that we make the dependence of the code on the error probability $\epsilon$ explicit.

Definition 11. An $(n,N_X,N_Y,\Delta,\varepsilon)$-blockwise code consists of $(n+\Delta)$ sequential rate-$(R_X,R_Y)$ encoders $\{(f_j,g_j)\}_{j=1}^{n+\Delta}$ and one sequential rate-$(R_X,R_Y)$ decoder $\phi_{n+\Delta}$ operating at time $n+\Delta$ where

$$N_X = \exp((n+\Delta)R_X), \quad N_Y = \exp((n+\Delta)R_Y),$$

and

$$\Pr\left( (\hat{X}^n,\hat{Y}^n) \neq (X^n,Y^n) \right) \leq \varepsilon.$$
Lemma 8. Fix $\gamma > 0$. For any $(n, N_X, N_Y, \Delta, \epsilon)$-blockwise code for the joint source $Q_{XY}^n$,

$$
\Pr \left\{ (\hat{X}^n, \hat{Y}^n) \neq (X^n, Y^n) \right\} \geq \Pr \left\{ \frac{1}{n} \log \frac{1}{Q^n_{X,Y}(X^n|Y^n)} \geq \frac{1}{n} \log N_X + \gamma \text{ or } \frac{1}{n} \log \frac{1}{Q^n_{Y,X}(Y^n|X^n)} \geq \frac{1}{n} \log N_Y + \gamma \text{ or } \frac{1}{n} \log \frac{1}{Q^n_{XY}(X^n,Y^n)} \geq \frac{1}{n} \log N_X N_Y + \gamma \right\} - 3\exp(-n\gamma).
$$

(56)

The proof of Lemma 8 is analogous to that in Miyake and Kanaya [25] (see also Han [24]) with some differences. See Remark 2. For the sake of completeness, its proof is provided in Appendix F.

Remark 1. We remark that the bound in Lemma 8 does not depend on $\Delta$ because for blockwise codes, we are only interested in the first $n$ (and not all $n+\Delta$) symbols. Also, we emphasize that the message $N_X$ and $N_X$ consist of approximately $(n+\Delta)R_X$ and $(n+\Delta)R_Y$ bits respectively.

Remark 2. The proof is similar to Lemma 7.2.2 in [24] except in two important aspects:

- $N_X$ and $N_X$ denote the number of encoded messages for sequences $X^{n+\Delta}$ and $Y^{n+\Delta}$ respectively;
- By considering any one-to-one mapping from $X^{n+\Delta}$ to $\{1, 2, \ldots, N_X\}$, we observe that at most $N_X$ length-$n$ sequences $x^n$ can be decoded correctly. An analogous statement holds for the other source.

When the joint source $(X^n, Y^n)$ is independent and identically distributed,

$$
\frac{1}{n} \log \frac{1}{Q^n_{X,Y}(X^n|Y^n)} = \frac{1}{n} \sum_{i=1}^{n} \log \frac{1}{Q_{X,Y}(X_i|Y_i)}.
$$

(57)

Invoking Chebyshev’s inequality, for all $n \in \mathbb{N}$ and any $\xi > 0$,

$$
\Pr \left\{ \frac{1}{n} \log \frac{1}{Q^n_{X,Y}(X^n|Y^n)} - H(Q_{X|Y}|Q_Y) > \xi \right\} \leq \frac{\text{Var} \left( \frac{1}{n} \log \frac{1}{Q^n_{X,Y}(X^n|Y^n)} \right)}{n\xi^2}.
$$

(58)

Similarly,

$$
\Pr \left\{ \frac{1}{n} \log \frac{1}{Q^n_{Y,X}(Y^n|X^n)} - H(Q_{Y|X}|Q_X) > \xi \right\} \leq \frac{\text{Var} \left( \frac{1}{n} \log \frac{1}{Q^n_{Y,X}(Y^n|X^n)} \right)}{n\xi^2},
$$

(59)

and

$$
\Pr \left\{ \frac{1}{n} \log \frac{1}{Q^n_{XY}(X^n,Y^n)} - H(Q_{XY}) > \xi \right\} \leq \frac{\text{Var} \left( \frac{1}{n} \log \frac{1}{Q^n_{XY}(X^n,Y^n)} \right)}{n\xi^2}.
$$

(60)

Hence, if $((1 + \frac{\Delta}{n})R_X, (1 + \frac{\Delta}{n})R_Y) \notin R_{sw}(Q_{XY})$, there exists a constant $\gamma > 0$ such that

$$
\frac{1}{n} \log N_X = H(Q_{X|Y}|Q_Y) - 2\gamma \quad \text{ or, }
$$

(61)

$$
\frac{1}{n} \log N_Y = H(Q_{Y|X}|Q_X) - 2\gamma \quad \text{ or, }
$$

(62)

$$
\frac{1}{n} \log (N_X N_Y) = H(Q_{XY}) - 2\gamma.
$$

(63)

Define $V := \max \{ \text{Var} \left( \frac{1}{n} \log \frac{1}{Q^n_{X,Y}(X^n|Y^n)} \right), \text{Var} \left( \frac{1}{n} \log \frac{1}{Q^n_{Y,X}(Y^n|X^n)} \right), \text{Var} \left( \frac{1}{n} \log \frac{1}{Q^n_{XY}(X^n,Y^n)} \right) \}$. It is easy to see that $V$ is finite. Consequently, we have

$$
Q^n_{XY} \left\{ (\hat{X}^n, \hat{Y}^n) \neq (X^n, Y^n) \right\} \geq 1 - \frac{V}{n\xi^2}.
$$

(64)

Thus in this blockwise streaming scenario, the strong converse holds, i.e.,

$$
Q^n_{XY} \left\{ (\hat{X}^n, \hat{Y}^n) \neq (X^n, Y^n) \right\} \to 1.
$$

(65)
2) Change of Measure: We derive a lower bound for the error probability for blockwise codes under the true source distribution $P_{XY}$. Define

$$\delta_n = \frac{V}{n\xi^2}. \quad (66)$$

Clearly, as $n \rightarrow \infty$, $\frac{\Delta}{n} \rightarrow \alpha$, and $\delta_n \rightarrow 0$. With this definition, we are ready to state an important lower bound on the error probability.

**Lemma 9** (Change of Measure). For any rate-$(R_X, R_Y)$ blockwise code, suppose the correlated source has distribution $P_{XY}$, we have

$$\Pr_{P_{XY}} \left( (\hat{X}^n, \hat{Y}^n) \neq (X^n, Y^n) \right) \geq \exp \left( -\frac{nD(Q_{XY} || P_{XY}) + H_0(1 - \delta_n)}{1 - \delta_n} \right). \quad (67)$$

The argument leading to the proof of Lemma 9 is contained in [3, Theorem 10.3] and so we omit its proof.

3) Error Exponent: Take logarithm of (67), divide both sides by $-\Delta$, and let $\Delta \rightarrow \infty$ on both sides of (67) to obtain

$$\limsup_{\Delta \rightarrow \infty} -\frac{1}{\Delta} \log P^n_{XY}(\hat{X}^n, \hat{Y}^n) (X^n, Y^n) \leq \frac{1}{\alpha} D(Q_{XY} || P_{XY}). \quad (68)$$

We complete the proof by minimizing the right-hand-side of (68) over all the auxiliary distributions $Q_{XY}$ such that $(1 + \alpha)R_X, (1 + \alpha)R_Y \notin R_{aw}(Q_{XY})$.

**Remark 3.** We remark that the strong converse technique here will also turn out useful in establishing the converser part of moderate deviations results in Theorem 6. Without the strong converse technique, we can establish an upper bound on the error exponent for blockwise codes only for rate pairs $(R_X, R_Y)$ satisfying $((1 + \alpha)R_X, (1 + \alpha)R_Y) \in R^*(P_{XY})$.

B. Proof of Lower Bound in Theorem 3 and 4

We use random binning to encode both sources by assigning random bit strings with lengths $\lfloor jR \rfloor - \lfloor (j - 1)R \rfloor$ at time $j$ where $R$ is $R_X$ or $R_Y$. Define $B_X(x^n) := \{\bar{x}^n : f^n_X(\bar{x}^j) = f^n_X(x^j), j = 1, 2, \ldots, n\}$ and $B_Y(y^n) := \{\bar{y}^n : f^n_Y(\bar{y}^j) = f^n_Y(y^j), j = 1, 2, \ldots, n\}$ as bins that contain sequences with same encoded messages.

Suppose the correlated sources generate a random sequence pair $(X^{n+\Delta}, Y^{n+\Delta})$. We use two techniques to derive an achievable error exponent, namely Gallager’s form [4], [26] and Csiszár-Körner [3] (CK) form.

1) Gallager’s Form: The decoding rule under maximum likelihood decoding is,

$$(\hat{x}^n, \hat{y}^n) = \arg \max_{(\bar{x}^n, \bar{y}^n) : (\bar{x}^{n+\Delta}, \bar{y}^{n+\Delta}) \in B_X(X^{n+\Delta}) \times B_Y(Y^{n+\Delta})} P^n_{XY}(\bar{x}^n, \bar{y}^n). \quad (69)$$

Hence, an error occurs in the following three scenarios:

$$\mathcal{E}_1 : \{\exists \bar{z}^{n+\Delta} \in B_X(X^{n+\Delta}) : P^n_{XY}(\bar{x}^n, Y^n) \geq P^n_{XY}(X^n, Y^n)\}, \quad (70)$$

$$\mathcal{E}_2 : \{\exists \bar{y}^{n+\Delta} \in B_Y(Y^{n+\Delta}) : P^n_{XY}(X^n, \bar{y}^n) \geq P^n_{XY}(X^n, Y^n)\}, \quad (71)$$

$$\mathcal{E}_3 : \{\exists (\bar{x}^{n+\Delta}, \bar{y}^{n+\Delta}) \in B_X(X^{n+\Delta}) \times B_Y(Y^{n+\Delta}) : P^n_{XY}(\bar{x}^n, \bar{y}^n) \geq P^n_{XY}(X^n, Y^n)\}. \quad (72)$$

Note that the above sets contain sequences (or pairs of sequences) of length $n+\Delta$. However, the condition on the probabilities is over a subsequence of length $n$. Equipped with these definitions, we can calculate the total error probability as follows:

$$\Pr(\mathcal{E}) = \Pr(\mathcal{E}_1 \cup \mathcal{E}_2 \cup \mathcal{E}_3) \leq \Pr(\mathcal{E}_1) + \Pr(\mathcal{E}_2) + \Pr(\mathcal{E}_3). \quad (73)$$
We now bound each of the three terms.

\[
\Pr(\mathcal{E}_1) = \Pr(\exists \hat{x}^{n+\Delta} \in B_X(X^{n+\Delta}) : P^n_{X,Y}(\hat{x}^n, Y^n) \geq P^n_{XY}(X^n, Y^n))
\]

\[= \sum_{(x^{n+\Delta}, y^{n+\Delta})} P^n_{XY}(x^{n+\Delta}, y^{n+\Delta}) \Pr\left(\exists \hat{x}^{n+\Delta} \in B_X(X^{n+\Delta}) : P^n_{X,Y}(\hat{x}^n, y^n) \geq P^n_{XY}(x^n, y^n)\right) \tag{74}\]

\[= \sum_{(x^{n+\Delta}, y^{n+\Delta})} P^n_{XY}(x^{n+\Delta}, y^{n+\Delta}) \left(\sum_{\hat{x}^{n+\Delta} : P^n_{X,Y}(\hat{x}^n, y^n) \geq P^n_{XY}(x^n, y^n)} \Pr\left(\exists \hat{x}^{n+\Delta} \in B_X(X^{n+\Delta})\right)\right)^\rho \tag{75}\]

\[\leq \sum_{(x^{n+\Delta}, y^{n+\Delta})} P^n_{XY}(x^{n+\Delta}, y^{n+\Delta}) \left(\sum_{\hat{x}^{n+\Delta} : P^n_{X,Y}(\hat{x}^n, y^n) \geq P^n_{XY}(x^n, y^n)} \exp\left(-(n + \Delta) R_X\right)\right)^\rho \tag{76}\]

\[\leq \sum_{(x^{n+\Delta}, y^{n+\Delta})} P^n_{XY}(x^{n+\Delta}, y^{n+\Delta}) \left(\sum_{\hat{x}^{n+\Delta} : P^n_{X,Y}(\hat{x}^n, y^n) \geq P^n_{XY}(x^n, y^n)} \exp\left(-(n + \Delta) R_X\right)\right)^\rho \tag{77}\]

where \(\rho \in [0, 1]\). Observe that the difference of this derivation vis-à-vis the usual derivation of the SW error exponent [26, 32] is that we are only decoding \(n\) symbol pairs using \((n + \Delta) R_X\) and \((n + \Delta) R_Y\) encoded bits.

Similarly, we can obtain

\[
\Pr(\mathcal{E}_2) \leq \exp\left(-n E_{X|Y}(1 + \Delta / n) R_X, \rho\right) \tag{78}\]

\[
\Pr(\mathcal{E}_3) \leq \exp\left(-n E_{X|Y}(1 + \Delta / n) R_X, (1 + \Delta / n) R_Y, \rho\right) \tag{79}\]

Therefore, we conclude

\[
\lim inf_{\Delta \to \infty} -\frac{1}{\Delta} \log \Pr(\mathcal{E}) \geq \frac{1}{\alpha} E_{sw}(1 + \Delta / n) R_X, (1 + \alpha) R_Y. \tag{80}\]

2) **CK Form:** We use the minimum empirical entropy decoding rule as follows:

\[
(\hat{x}^n, \hat{y}^n) = \arg \min_{(\hat{x}^n, \hat{y}^n) : (\hat{x}^{n+\Delta}, \hat{y}^{n+\Delta}) \in B_X(X^{n+\Delta}) \times B_Y(Y^{n+\Delta})} \hat{H}(\hat{x}^n, \hat{y}^n), \tag{81}\]

where \(\hat{H}(\hat{x}^n, \hat{y}^n) = H_{P^n_{XY}}(XY)\) is the empirical entropy. Error occurs in the following three scenarios:

\[
\mathcal{E}_1 : \{\exists \hat{x}^{n+\Delta} \in B_X(X^{n+\Delta}) : \hat{H}(\hat{x}^n, Y^n) \leq \hat{H}(X^n, Y^n)\}, \tag{82}\]

\[
\mathcal{E}_2 : \{\exists \hat{y}^{n+\Delta} \in B_Y(Y^{n+\Delta}) : \hat{H}(X^n, \hat{y}^n) \leq \hat{H}(X^n, Y^n)\}, \tag{83}\]

\[
\mathcal{E}_3 : \{\exists (\hat{x}^{n+\Delta}, \hat{y}^{n+\Delta}) \in B_X(X^{n+\Delta}) \times B_Y(Y^{n+\Delta}) : \hat{H}(\hat{x}^n, \hat{y}^n) \leq \hat{H}(X^n, Y^n)\}. \tag{84}\]
Hence the total error probability is as in (73). We can upper bound each term via a standard method of types [3] calculation, resulting in

\[
\Pr(\mathcal{E}_1) \leq (n + 1)^{2|Y|} \exp \left( -n \min_{Q_{XY}} \left( D(Q_{XY} || P_{XY}) + \left| H(Q_{XY} | Q_Y) - (1 + \frac{\Delta}{n}) R_X \right| \right) \right),
\]

\[
\Pr(\mathcal{E}_2) \leq (n + 1)^{2|Y|} \exp \left( -n \min_{Q_{XY}} \left( D(Q_{XY} || P_{XY}) + \left| H(Q_{Y|X} | Q_X) - (1 + \frac{\Delta}{n}) R_Y \right| \right) \right),
\]

\[
\Pr(\mathcal{E}_3) \leq (n + 1)^{2|Y|} \exp \left( -n \min_{Q_{XY}} \left( D(Q_{XY} || P_{XY}) + \left| H(Q_{XY}) - (1 + \frac{\Delta}{n}) (R_X + R_Y) \right| \right) \right).
\]

Finally, take logarithm on the total error probability, divide by \(-\frac{1}{\Delta}\) and let \(\Delta \to \infty\) to obtain

\[
\lim \inf_{\Delta \to \infty} -\frac{1}{\Delta} \log \Pr(\mathcal{E}) \geq \frac{1}{\alpha} E_{sw}(1 + \alpha) R_X, (1 + \alpha) R_Y).
\]

**Remark 4.** To prove (37), replace \(n\) with \(j_n\) and we use the fact that \(\lim_{n \to \infty} \frac{j_n}{n} = 1\).

### C. Proof of Upper Bound in Theorem 4

We generalize the symbolwise feedforward decoder with fixed delay \(\Delta\) for single streaming source with decoder side information in [8] to the two-streaming-sources scenario. We lower bound the symbolwise error probability \(\Pr(\hat{X}_j \neq X_j, \hat{Y}_j \neq Y_j)\) under some auxiliary distribution. Then by using properties of an appropriately defined typical set, we derive a lower bound for the symbolwise error probability under the true distribution. The whole procedure can be split into four parts following [7].

1) **Feedforward decoders:**

**Definition 12** (Type-I Feedforward Decoder). A symbolwise decoder with fixed delay \(\Delta\) is said to be a Type-I feedforward decoder if it has access to the encoded messages of symbols \((X_j^\Delta, Y_j^\Delta)\) as well past source symbols \((X_{j-1}, Y_{j-1})\) at time \(j\), where \(j \geq 1\).

**Definition 13** (Type-II Feedforward Decoder). Define the error sequence as \((\hat{X}_j, \hat{Y}_j) = (X_{j-1}, Y_{j-1} - (X_j, Y_j)).\) A symbolwise decoder with fixed delay is said to be a Type-II feedforward decoder if it has access to the encoded messages of symbols \((X_j^\Delta, Y_j^\Delta)\) as well error sequence of past source symbols, (i.e. \((X_{j-1}^\Delta, Y_{j-1}^\Delta)\)), at time \(j\), where \(j \geq 1\).

**Definition 14** (Type-III Feedforward Decoder). A symbolwise decoder with fixed delay is said to be a Type-III feedforward decoder for source \(X\) if it has access to the encoded messages of symbols \((X_j^\Delta, Y_j^\Delta)\) as well symbols of source \(Y\) (i.e. \(Y_{j+\Delta}\)), at time \(j\), where \(j \geq 1\). Similarly, we can define Type-III feedforward decoder for source \(Y\).

**Definition 15** (Type-IV Feedforward Decoder). Define the marginal error sequence of source \(X\) as \(\hat{X}_j = X_j - \hat{X}_j\). A symbolwise decoder with fixed delay is said to be a Type-IV feedforward decoder for source \(X\) if it has access to the encoded messages of symbols \((X_j^\Delta, Y_j^\Delta)\) as well as error sequence of past symbols of source \(X\), (i.e. \(X_{j-1}\)), and all symbols of source \(Y\), (i.e. \(Y_{j+\Delta}\)), at time \(j\), where \(j \geq 1\). Similarly, we can define Type-IV feedforward decoder for source \(Y\).

**Lemma 10.** Type-I and Type-II feedforward decoders are equivalent in the sense that their estimates of the same symbol pair are the same. Type-III and Type-IV feedforward decoders for the same source is also equivalent in the sense that their estimate of the same source symbol is the same.

The proof of Lemma 10 is similar to that of Lemma 4.2 in [27] and Lemma 3 in [8], so we omit the proof.

2) **Error free systems:** We can construct joint error-free system in Fig. 5 by the concatenation of Type-I and Type-II feedforward decoders. Similarly, we can construct marginal error free system in Fig. 6 of source \(X\) (resp. \(Y\)) by constructing Type-III and Type-IV feedforward decoders for source \(X\) (resp. \(Y\)). With the three error free systems, we have the following lemma regarding the existence of corresponding Markov chains. Let \(M_{X}^{n+\Delta | R_X} = \prod_{j=1}^{n+\Delta} f_j(X_j)\) and \(M_{Y}^{n+\Delta | R_Y} = \prod_{j=1}^{n+\Delta} g_j(Y_j)\) be the accumulated encoded messages of source \(X\) and \(Y\) from time \(1\) to \(n+\Delta\).

**Lemma 11.** With the joint and marginal error free system for source \(X\) and \(Y\), we have the following Markov chains.

\[
(X^n, Y^n) \to (M_{X}^{n+\Delta | R_X}, M_{Y}^{n+\Delta | R_Y}, \hat{X}_n, \hat{Y}_n) \to (X^n, Y^n),
\]

\[
X^n \to (M_{X}^{n+\Delta | R_X}, \hat{X}_n, X^{n+\Delta}) \to X^n,
\]

\[
Y^n \to (M_{Y}^{n+\Delta | R_Y}, \hat{Y}_n, Y^{n+\Delta}) \to Y^n.
\]

The proof of Lemma 11 is a modification of Lemmas 4 in [8] and hence we omit it.
3) Lower bound error probability under an auxiliary distribution: According to the three Markov chains, we can derive a lower bound for the entropy of the joint error sequence and entropies of marginal error sequences. Based on chain rule and the fact that conditioning does not increase entropy, we can lower bound the average entropy of joint error pairs and average entropies of marginal error symbols. Finally, we can lower bound the joint and marginal symbolwise error probability under an auxiliary distribution.

Before stating the lemma, we define three quantities:

\[ n_{XY} = \frac{n}{2} \log |\mathcal{X}||\mathcal{Y}| - \frac{1}{2} (H(Q_{XY}) - (1 + \alpha)(R_X + R_Y)) \]

\[ n_X = \frac{n}{2} \log |\mathcal{X}| - \frac{1}{2} (H(Q_{X|Y}) - (1 + \alpha)R_X) \]

\[ n_Y = \frac{n}{2} \log |\mathcal{Y}| - \frac{1}{2} (H(Q_{Y|X}) - (1 + \alpha)R_Y) \]

**Lemma 12.** If sources have joint pmf \(Q_{XY}\) from time 1 to \(n\) and \((R_X, R_Y)\) satisfies \((1 + \alpha)R_X, (1 + \alpha)R_Y \notin \mathcal{R}_{sw}(Q_{XY})\), then for any symbolwise feedforward decoder, when \(n\) is sufficiently large, there exists at least \(n_t = \max\{n_{XY}, n_X, n_Y\}\) symbol pairs \(j_1, \ldots, j_{n_t}\) with symbolwise error probability satisfying

\[ \Pr \left( (\hat{X}_{j_k}, \hat{Y}_{j_k}) \neq (X_{j_k}, Y_{j_k}) \right) \geq \beta, \]

for some \(\beta \in (0, 1)\) and for \(k = 1, \ldots, n_t\).

The proof of Lemma 12 is a modification of Lemma 5 in [8] and hence we omit it.

4) Change of measure: The final step is to lower bound the symbolwise joint and marginal error probabilities under the true distribution \(P_{XY}\). In Lemma 12, the sources are required to have joint distribution \(Q_{XY}\) from time 1 to \(n\). In time slots \(n + 1\) to \(n + \Delta\), the distribution of the sources can be arbitrary; we let the distribution be \(P_{XY}\) for these time slots.
Define the set of incorrectly decoded $j$-th source pair as $E_j := \{(x^{n+\Delta}, y^{n+\Delta}) : (\tilde{x}_j, \tilde{y}_j) \neq (x_j, y_j)\}$. Noting that the sources have joint pmf $P_{XY}$ from time $n + 1$ to $n + \Delta$, We know from Lemma 12 that

$$Q^n_{XY} P^n_{XY}(E_j) = \sum_{(x^{n+\Delta}, y^{n+\Delta}) \in E_j} Q^n_{XY}(x^n, y^n) P^n_{XY}(x^{n+\Delta}_{n+1}, y^{n+\Delta}_{n+1}) \geq \beta = 1 - \beta'. \quad (105)$$

**Lemma 13 (Change of Measure).** Fix $\tau_n > 0$. If the true source has pmf $P_{XY}$ and $Q_{XY}$ is any auxiliary distribution for time slots 1 to $n$ (as in Lemma 12), then

$$P^n_{XY}(E_j) \geq (\beta - \eta_n) \exp \left(-n \left(D(Q_{XY} \| P_{XY}) + \tau_n\right)\right). \quad (106)$$

Proof: We prove the lemma by modifying the proof for Theorem 1.2 and Corollary 1.2 in [3]. Define the relative entropy typical set $B(n, \tau_n)$ as

$$B(n, \tau_n) := \{(x^n, y^n) : -\tau_n \leq \frac{1}{n} \log \frac{Q^n_{XY}(x^n, y^n)}{P^n_{XY}(x^n, y^n)} - D(Q_{XY} \| P_{XY}) \leq \tau_n\}. \quad (107)$$

Invoking Chebyshev’s inequality,

$$\Pr((X^n, Y^n) \notin B(n, \tau_n)) = \Pr \left(\frac{1}{n} \sum_{i=1}^{n} \log \frac{Q_{XY}(X_i, Y_i)}{P_{XY}(X_i, Y_i)} - D(Q_{XY} \| P_{XY}) > \tau_n\right) \leq \frac{V(Q_{XY} \| P_{XY})}{n \tau_n^2} := \eta_n, \quad (108)$$

where $V(Q_{XY} \| P_{XY}) := \text{Var}_{P_{XY}}(\log Q_{XY}(X, Y) - \log P_{XY}(X, Y))$ is the relative entropy variance. Define the set $C(n + \Delta, \tau) := \{(x^{n+\Delta}, y^{n+\Delta}) : (x^n, y^n) \in B(n, \tau_n)\}$. Then we have

$$Q^n_{XY}(C(n + \Delta, \tau_n)) = Q^n_{XY} P^n_{XY}(C(n + \Delta, \tau_n)) = Q^n_{XY}(B(n, \tau_n)) \geq 1 - \eta_n. \quad (109)$$

Consequently,

$$P_{XY}^{n+\Delta}(E_j) \geq P_{XY}^{n+\Delta}(E_j \cap C(n + \Delta, \tau_n)) = \sum_{(x^{n+\Delta}, y^{n+\Delta}) \in E_j \cap C(n + \Delta, \tau_n)} P_{XY}^{n+\Delta}(x^{n+\Delta}, y^{n+\Delta}) \quad (110)$$

$$= \sum_{(x^{n+\Delta}, y^{n+\Delta}) \in E_j \cap C(n + \Delta, \tau_n)} Q^n_{XY}(x^n, y^n) P_{XY}^{n+\Delta}(x^{n+\Delta}_{n+1}, y^{n+\Delta}_{n+1}) \exp \{-n \left(D(Q_{XY} \| P_{XY}) + \tau_n\right)\} \quad (111)$$

$$= \sum_{(x^{n+\Delta}, y^{n+\Delta}) \in E_j \cap C(n + \Delta, \tau_n)} \exp \{-n \left(D(Q_{XY} \| P_{XY}) + \tau_n\right)\} \quad (112)$$

$$= \exp \{-n \left(D(Q_{XY} \| P_{XY}) + \tau_n\right)\} \quad (113)$$

$$= \exp \{-n \left(D(Q_{XY} \| P_{XY}) + \tau_n\right)\} \quad (114)$$

$$= \exp \{-n \left(D(Q_{XY} \| P_{XY}) + \tau_n\right)\} \quad (115)$$

Take the logarithm on both sides of (106), divide by $\Delta$ and let $\Delta \to \infty$. Hence we observe that for all $Q_{XY}$ such that $((1 + \alpha)R_X, (1 + \alpha)R_Y) \notin \mathcal{R}_{sw}(Q_{XY})$, we have

$$\limsup_{\Delta \to \infty} -\frac{1}{\Delta} \log P_{XY}^{n+\Delta}(E_j) \leq \limsup_{\Delta \to \infty} \frac{n}{\Delta} D(Q_{XY} \| P_{XY}) + \frac{n \tau_n}{\Delta} - \frac{1}{\Delta} \log(\beta - \eta_n). \quad (116)$$

Choosing $\tau_n$ such that $\tau_n \to 0$ and $n \tau_n^2 \to \infty$ as $n \to \infty$, then $\eta_n \to 0$ as $n \to \infty$. Hence,

$$\limsup_{\Delta \to \infty} -\frac{1}{\Delta} \log P_{XY}^{n+\Delta}(E_j) \leq \min_{Q_{XY} : ((1 + \alpha)R_X, (1 + \alpha)R_Y) \notin \mathcal{R}_{sw}(Q_{XY})} \frac{1}{\alpha} D(Q_{XY} \| P_{XY}) = \frac{1}{\alpha} F_{sw}((1 + \alpha)R_X, (1 + \alpha)R_Y). \quad (117)$$

Note that for $((1 + \alpha)R_X, (1 + \alpha)R_Y) \in \mathcal{R}'(P_{XY})$, we conclude at least one of the following three equalities hold:

$$H(Q_{XY} \| Q_Y) = (1 + \alpha)R_X, \quad (118)$$

$$H(Q_Y \| Q_X) = (1 + \alpha)R_Y, \quad (119)$$

$$H(Q_{XY}) = (1 + \alpha)(R_X + R_Y). \quad (120)$$

Thus $n_t = n$. This means that the upper bound of error exponents holds for all symbol pairs, i.e. $j = 1, 2, \ldots, n$. 
D. Proofs of Theorems 6 and 7

1) Preliminaries: Define

\[ E_{XY}(\rho) = (1 + \rho) \log \sum_{x,y} P_{XY}(x,y)^{1+\rho}, \]  

\[ E_{X|Y}(\rho) = \log \sum_{y} P_Y(y) \left( \sum_{x} P_{X|Y}(x|y)^{1+\rho} \right)^{1+\rho}, \]  

and \( E_{Y|X}(\rho) \) is similar to \( E_{X|Y}(\rho) \) with \( X \) and \( Y \) interchanged.

Next, given a joint pmf \( P_{XY} \) on the alphabet \( \mathcal{X} \times \mathcal{Y} \), define the third central moment of the source as

\[ T(P_{XY}) := \sum_{x,y} P_{XY}(x,y) \left( -\log P_{XY}(x,y) - H(P_{XY}) \right)^3. \]

**Lemma 14.** For any pmf \( P_{XY} \) on finite alphabet \( \mathcal{X} \times \mathcal{Y} \) such that \( P_{XY}(x,y) > 0 \) for all \( (x,y) \in \mathcal{X} \times \mathcal{Y} \),

\[ E'_{XY}(\rho) |_{\rho=0} = H(P_{XY}), \]  

\[ E''_{XY}(\rho) |_{\rho=0} = \frac{1}{\log e} V(P_{XY}), \]  

\[ E'''_{XY}(\rho) \geq 0, \]  

\[ E'''_{XY}(\rho) = \frac{-1}{(1 + \rho)^2} \left( \frac{3}{\log e} V(P_{X,Y}^\rho) + \frac{1}{\log e} T(P_{X,Y}^\rho) \right). \]

There exists finite positive number \( M_{XY} \) such that

\[ \left| E'''_{XY}(\rho) |_{\rho \in [0,1]} \right| \leq M_{XY}. \]

For a conditional distribution \( P_{X|Y} \), we define the following quantities in the usual manner:

\[ H(P_{X|Y}) := -\sum_x P_{X|Y}(x|y) \log P_{X|Y}(x|y), \]  

\[ V(P_{X|Y}) := \sum_x P_{X|Y}(x|y) \left( -\log P_{X|Y}(x|y) - H(P_{X|Y}) \right)^2, \]  

\[ T(P_{X|Y}) := \sum_x P_{X|Y}(x|y) \left( -\log P_{X|Y}(x|y) - H(P_{X|Y}) \right)^3. \]

**Lemma 15.** For any pmf \( P_{XY} \) on finite alphabet \( \mathcal{X} \times \mathcal{Y} \) such that \( P_{XY}(x,y) > 0 \), \( P_{Y|X}(y|x) > 0 \) for any \( (x,y) \in \mathcal{X} \times \mathcal{Y} \),

\[ E'_{X|Y}(\rho) |_{\rho=0} = H(P_{X|Y}|P_Y), \]  

\[ E''_{X|Y}(\rho) |_{\rho=0} = \frac{1}{\log e} V(P_{X|Y}|P_Y), \]  

\[ E'''_{X|Y}(\rho) = \frac{1}{\log^2 e} \sum_y P_{Y}(y) \left( H(P_{X,Y}^\rho|Y=y) - H(P_{X,Y}^\rho|P_Y) \right)^3 \]  

\[ + \frac{3}{(1 + \rho)^2 \log^2 e} \sum_y P_{Y}(y) \left( H(P_{X,Y}^\rho|Y=y) - H(P_{X,Y}^\rho|P_Y) \right) V(P_{X,Y}^\rho|Y=y) \]  

\[ - \frac{1}{(1 + \rho)^2 \log^2 e} \sum_y P_{Y}(y) \left( 3V(P_{X,Y}^\rho|Y=y) \log e + T(P_{X,Y}^\rho|P_Y) \right). \]

There exists a finite number \( M_X \) such that

\[ \left| E'''_{X|Y}(\rho) |_{\rho \in [0,1]} \right| \leq M_X. \]

The derivatives and properties of \( E_{Y|X}(\rho) \) is similar to \( E_{X|Y}(\rho) \) with \( X \) and \( Y \) interchanged.

The proofs of Lemmas 14 and 15 are similar to the proofs of Lemma 1 in [16] and are thus omitted. The Taylor series expansions of (121) and (122) are as follows:

\[ E_{XY}(\rho) = \rho H(P_{XY}) + \frac{\rho^2}{2 \log e} V(P_{XY}) + \frac{\rho^3}{6} E'''_{XY}(\overline{T}_{XY}), \]  

\[ E_{X|Y}(\rho) = \rho H(P_{X|Y}|P_Y) + \frac{\rho^2}{2 \log e} V(P_{X|Y}|P_Y) + \frac{\rho^3}{6} E'''_{X|Y}(\overline{T}_X), \]

for some \( \overline{T} \in [0,\rho] \). The Taylor expansion of \( E_{Y|X}(\rho) \) is obtained by interchanging \( X \) and \( Y \) in (137).
2) Converse Proofs of Theorems 6 and 7: Recall Lemma 9. For blockwise codes with rate pair \((R_X^{(\Delta)}, R_Y^{(\Delta)})\), it holds that for all pmfs \(Q_{XY}\) such that \(((1 + \Delta)\alpha R_X^{(\Delta)}, (1 + \Delta)\alpha R_Y^{(\Delta)}) \notin R_{sw}(Q_{XY})\), a lower bound for the error probability is given by

\[
\Pr \left( (\hat{X}^n, \hat{Y}^n) \neq (X^n, Y^n) \right) \geq \exp \left( - \frac{nD(Q_{XY}||P_{XY}) + H_b(1 - \delta_n)}{1 - \delta_n} \right).
\]  

(138)

Take the logarithm on both sides of (138) to obtain

\[
- \log \Pr \left( (\hat{X}^n, \hat{Y}^n) \neq (X^n, Y^n) \right) \leq \frac{nD(Q_{XY}||P_{XY})}{1 - \delta_n} - \frac{H_b(\delta_n)}{1 - \delta_n}.
\]  

(139)

Now we invoke Lemma 13. For symbolwise codes, we have for all \(Q_{XY}\) such that \(((1 + \Delta)\alpha R_X, (1 + \Delta)\alpha R_Y) \notin R_{sw}(Q_{XY})\), a lower bound for the error probability is given by

\[
P_{XY}^{n+\Delta}(\mathcal{E}_j) \geq (\beta - \eta_n) \times \exp \left( - n \left( D(Q_{XY}||P_{XY}) + \tau_n \right) \right),
\]  

(140)

where \(\mathcal{E}_j := \{(x^n+\Delta, y^n+\Delta) : (\tilde{x}_j, \tilde{y}_j) \neq (x_j, y_j)\}\). For \(((1 + \Delta)\alpha R_X, (1 + \Delta)\alpha R_Y) \in \mathcal{R}'(P_{XY})\), (140) holds for all \(j = 1, 2, \ldots, n\).

Take the logarithm on both sides of (140) to obtain

\[
- \log P_{XY}^{n+\Delta}(\mathcal{E}_j) \leq - \log(\beta - \eta_n) + n\tau_n + nD(Q_{XY}||P_{XY}).
\]  

(141)

Because the relative entropy \(D(Q_{XY}||P_{XY})\) is the common term of blockwise codes and symbolwise codes, we then derive an upper bound on this divergence term for the three different cases of Theorems 6 and 7.

First consider the top left corner point. According to Corollary 2,

\[
\min_{Q_{XY} : ((1 + \Delta)\alpha R_X^{(\Delta)}, (1 + \Delta)\alpha R_Y^{(\Delta)}) \notin R_{sw}(Q_{XY})} D(Q_{XY}||P_{XY}) = \min \left\{ \max_{\rho \in [0, 1]} \rho \left( 1 + \frac{\Delta}{n} \right) \left( R_X^{(\Delta)} + R_Y^{(\Delta)} \right) - E_XY(\rho), \max_{\rho \in [0, 1]} \rho \left( 1 + \frac{\Delta}{n} \right) R_X^{(\Delta)} - E_X|Y(\rho), \sup_{\rho \geq 0} \rho \left( 1 + \frac{\Delta}{n} \right) R_Y^{(\Delta)} - E_Y|X(\rho) \right\}.
\]  

(142)

Dropping the third term in the above equation, we upper bound (142) to obtain

\[
\min_{Q_{XY} : ((1 + \Delta)\alpha R_X^{(\Delta)}, (1 + \Delta)\alpha R_Y^{(\Delta)}) \notin R_{sw}(Q_{XY})} D(Q_{XY}||P_{XY}) \leq \min \left\{ \max_{\rho \in [0, 1]} \rho \left( 1 + \frac{\Delta}{n} \right) \left( R_X^{(\Delta)} + R_Y^{(\Delta)} \right) - E_XY(\rho), \max_{\rho \in [0, 1]} \rho \left( 1 + \frac{\Delta}{n} \right) R_X^{(\Delta)} - E_X|Y(\rho) \right\}.
\]  

(143)

Suppose \(\rho_{X,\Delta}^*\) and \(\rho_{XY,\Delta}^*\) achieve the maximum of the two functions of \(\rho\) in (143). Recalling Taylor series expansions in (136), we obtain

\[
\max_{\rho \in [0, 1]} \rho \left( 1 + \frac{\Delta}{n} \right) \left( R_X^{(\Delta)} + R_Y^{(\Delta)} \right) - E_XY(\rho)
\]

\[
= \rho_{XY,\Delta}^* \left[ \frac{1 + \Delta}{1 + \alpha} \left( H(P_{XY}) + (1 + \alpha)(\theta_1 + \theta_2)\epsilon_\Delta \right) - \rho_{XY,\Delta}^* H(P_{XY}) - \frac{(\rho_{XY,\Delta}^*)^2}{2 \log e} V(P_{XY}) \right]
\]

\[- \frac{(\rho_{XY,\Delta}^*)^3}{6} E''_{XY}(\rho_{XY,\Delta}^*)\]  

\[
\leq \rho_{XY,\Delta}^* H(P_{XY}) \left( \frac{1 + \Delta}{1 + \alpha} - 1 \right) + \left( 1 + \frac{\Delta}{n} \right) (\theta_1 + \theta_2)\epsilon_\Delta - \frac{(\rho_{XY,\Delta}^*)^2}{2 \log e} V(P_{XY}) \left( \rho - \frac{(\rho_{XY,\Delta}^*)^2}{2 \log e} (\theta_1 + \theta_2)\epsilon_\Delta \log e \right)^2
\]

\[+ \frac{(1 + \Delta)^2 (\theta_1 + \theta_2)^2 (\epsilon_\Delta)^2}{2V(P_{XY})} \left( \frac{(\rho_{XY,\Delta}^*)^3}{6} (\epsilon_\Delta)^3 \right) M_{XY} \]

\[= \rho_{XY,\Delta}^* H(P_{XY}) \left( \frac{1 + \Delta}{1 + \alpha} - 1 \right) + \frac{(1 + \Delta)^2 (\theta_1 + \theta_2)^2 (\epsilon_\Delta)^2}{2V(P_{XY})} \left( \frac{(\rho_{XY,\Delta}^*)^3}{6} (\epsilon_\Delta)^3 \right) M_{XY}.
\]  

(144)
Similarly,
\[
\max_{\rho \in [0, 1]} \rho \left(1 + \frac{\Delta}{n}\right) R_X(\Delta) - E_{X\mid Y}(\rho) \\
= \rho_X^* \Delta \left(1 + \frac{\Delta}{1 + \alpha} (H(P_X) + (1 + \alpha)\theta_1\epsilon_\Delta) - \rho_X^* \Delta H(P_X) - \frac{(\rho_X^* \Delta)^2}{2 \log e} V(P_X) + \frac{(\rho_X^* \Delta)^3}{6} M_X \right) \\
\leq \rho_X^* \Delta \left(1 + \frac{\Delta}{1 + \alpha} - 1\right) + \left(1 + \frac{\Delta}{n}\right) \theta_1 \rho_X^* \Delta \epsilon_\Delta - \frac{(\rho_X^* \Delta)^2}{2 \log e} V(P_X) + \frac{(\rho_X^* \Delta)^3}{6} M_X \\
\leq \rho_X^* \Delta \left(1 + \frac{\Delta}{1 + \alpha} - 1\right) + \sup_{\rho} \left(\frac{V(P_X) + \epsilon_\Delta}{2 \log e} \left(\rho - \frac{(1 + \Delta/\alpha)\theta_1 \epsilon_\Delta \log e}{V(P_X)}\right)^2 \right) \\
+ \left(1 + \frac{\Delta}{\alpha}\right)^2 \frac{\rho_X^* \Delta \epsilon_\Delta}{2 \log e} + \frac{(\rho_X^* \Delta)^3}{6} M_X \\
= \rho_X^* \Delta \left(1 + \frac{\Delta}{1 + \alpha} - 1\right) + \left(1 + \frac{\Delta}{\alpha}\right)^2 \frac{\rho_X^* \Delta \epsilon_\Delta}{2 \log e} + \frac{(\rho_X^* \Delta)^3}{6} M_X. 
\]
(145)

**Proposition 16.**
\[
\rho_X^* \Delta \rightarrow 0 \text{ as } \Delta \rightarrow \infty, \\
\rho_X^* \Delta \rightarrow 0 \text{ as } \Delta \rightarrow \infty.
\]
(146)

(147)

Refer to Lemma 7 in [16] for the proof of Proposition 16.

We are now ready to conclude the proof of the moderate deviations constants for top left corner point. For blockwise codes, note that \(\Delta \epsilon_\Delta \rightarrow \infty, \epsilon_\Delta \rightarrow 0, n \rightarrow \infty, \frac{\Delta}{x} \rightarrow \alpha \text{ as } \Delta \rightarrow \infty,
\]
\[
\limsup_{\Delta \rightarrow \infty} \frac{-\log \Pr\left((\hat{X}_n, \hat{Y}_n) \neq (X^n, Y^n)\right)}{\Delta \epsilon_\Delta} \\
\leq \limsup_{\Delta \rightarrow \infty} \left\{-\frac{H_\Delta(\epsilon_\Delta)}{1 - \epsilon_\Delta} \frac{1}{\Delta \epsilon_\Delta} + \frac{n}{\Delta(1 - \epsilon_\Delta)} \min_{Q_X: (1 + \Delta \epsilon_\Delta)R_X, (1 + \Delta \epsilon_\Delta)R_Y \notin R_{sw}(Q_{XY})} D(Q_{XY} \parallel P_{XY}) \right\} \\
= \limsup_{\Delta \rightarrow \infty} \frac{1}{\alpha} \frac{\min_{Q_X: (1 + \alpha)R_X, (1 + \alpha)R_Y \notin R_{sw}(Q_{XY})} D(Q_{XY} \parallel P_{XY})}{\epsilon_\Delta^2} \\
\leq \frac{1}{\alpha} \min\left\{\frac{(1 + \alpha)^2(\theta_1 + \theta_2)^2 \log e}{2V(P_{XY})}, \frac{(1 + \alpha)^2 \theta_1^2 \log e}{2V(P_{X\mid Y})}\right\}. 
\]
(148)
(149)
(150)
(151)

For symbolwise codes, note that \(\Delta \epsilon_\Delta \rightarrow \infty, \Delta \epsilon_\Delta^2 \rightarrow \infty, n \rightarrow \infty, \frac{\Delta}{x} \rightarrow \alpha, \tau_n \rightarrow 0, n \epsilon_\Delta \rightarrow \eta_n \rightarrow 0 \text{ as } \Delta \rightarrow \infty,
\]
\[
\limsup_{\Delta \rightarrow \infty} \frac{-\log \Pr\left((\hat{X}_n, \hat{Y}_n) \neq (X^n, Y^n)\right)}{\Delta \epsilon_\Delta^2} \\
\leq \limsup_{\Delta \rightarrow \infty} \left\{-\frac{\log (\beta - \eta_n)}{\Delta \epsilon_\Delta^2} + \frac{\Delta \tau_n}{n \epsilon_\Delta} + \frac{\Delta \min_{Q_X: (1 + \Delta \epsilon_\Delta)R_X, (1 + \Delta \epsilon_\Delta)R_Y \notin R_{sw}(Q_{XY})} D(Q_{XY} \parallel P_{XY})}{\epsilon_\Delta^2} \right\} \\
= \limsup_{\Delta \rightarrow \infty} \frac{\Delta \min_{Q_X: (1 + \alpha)R_X, (1 + \alpha)R_Y \notin R_{sw}(Q_{XY})} D(Q_{XY} \parallel P_{XY})}{n \epsilon_\Delta^2} \\
\leq \frac{1}{\alpha} \min\left\{\frac{(1 + \alpha)^2(\theta_1 + \theta_2)^2 \log e}{2V(P_{XY})}, \frac{(1 + \alpha)^2 \theta_1^2 \log e}{2V(P_{X\mid Y})}\right\}, 
\]
(152)
(153)
(154)

where from (152) to (153) we choose \(\tau_n \) such that \(\tau_n / \epsilon_\Delta \rightarrow 0 \) as \(\Delta \rightarrow \infty \) and use the fact that \(\beta \in (0, 1)\) is non-vanishing. From (153) to (154) we use (144) and (145) as well as Proposition 16.

**Remark 5.** It is possible to choose \(\tau_n \) such that \(\tau_n \rightarrow 0, n \epsilon_\Delta \rightarrow \infty \) and \(\frac{\Delta}{\epsilon_\Delta} \rightarrow 0 \) due to the assumption that \(n \epsilon_\Delta \rightarrow \infty \) as \(\Delta \rightarrow \infty \). For example, we can choose \(\tau_n = \epsilon_\Delta^{1+\psi} \) for any \(\psi > 0 \). Then \(\tau_n = \epsilon_\Delta^{1+\psi} \rightarrow 0 \) and \(\frac{\Delta}{\epsilon_\Delta} = \epsilon_\Delta^\psi \rightarrow 0 \) as \(\Delta \rightarrow \infty \). Finally, \(\Delta \epsilon_\Delta \rightarrow \infty \) implies that \(n \epsilon_\Delta^2 = \frac{\epsilon_\Delta^{1+\psi}}{\Delta^{1+\psi}} \rightarrow \infty \) as \(\Delta \rightarrow \infty \) by letting \(\psi \) tend to 0 from above.

**Remark 6.** The strong converse in (64) and Lemma 9 lead to (149). Without the strong converse, we will have the additional term \(\frac{\Delta}{\epsilon_\Delta} \) which requires additional requirement \(\Delta \epsilon_\Delta \rightarrow \infty \) as mentioned in Remark 5.

For the vertical boundary, the proof can be done by upper bounding the minimization by \(E_{sw}^{X\mid Y}(1 + \alpha)R_X \). For any point in diagonal face, we can upper bound the minimization of three terms by \(E_{sw}^{X\mid Y}(1 + \alpha)R_X, (1 + \alpha)R_Y \).
3) Achievability Proofs of Theorems 6 and 7: Recall Appendix B, especially (86), (87), and (88). For blockwise codes with rate pair \((R_X^{(\Delta)}, R_Y^{(\Delta)})\), an upper bound on the error probability is given by

\[
\Pr \left(\tilde{X}^n, \tilde{Y}^n \neq (X^n, Y^n)\right) \leq \exp \left( -nE_{\mathrm{w}}\left((1 + \frac{\Delta}{n})R_X^{(\Delta)}, (1 + \frac{\Delta}{n})R_Y^{(\Delta)}\right)\right).
\]  

(155)

Recall the definition of \(E_{\mathrm{w}}(R_X, R_Y)\) in (17). We first consider the top left corner point.

Choosing \(\rho_\Delta = \frac{(1 + \frac{\Delta}{n})\theta_1 + \theta_2}{V^{(\rho)}}\) and for some \(\overline{\rho}_\Delta \in [0, \rho_\Delta]\), we have

\[
\sup_{\rho \in [0,1]} \left( \rho \left(1 + \frac{\Delta}{n}\right)(R_X^{(\Delta)} + R_Y^{(\Delta)}) - E_{XY}(\rho) \right) 
\]

(156)

\[
\geq \rho_\Delta \frac{1 + \frac{\Delta}{n}}{1 + \alpha} \left( H(P_{XY}) + (1 + \alpha)(\theta_1 + \theta_2)\epsilon_\Delta \right) - \rho_\Delta H(P_{XY}) - \frac{\rho_\Delta^2}{2\log e} V(P_{XY}) - \frac{\rho_\Delta^3}{6} E''_{XY}(\overline{\rho}_\Delta) 
\]

(157)

\[
\geq \left( \frac{1 + \frac{\Delta}{n}}{1 + \alpha} - 1 \right) \rho_\Delta H(P_{XY}) + \left(1 + \frac{\Delta}{n}\right)(\theta_1 + \theta_2)\rho_\Delta \epsilon_\Delta - \frac{\rho_\Delta^2}{2\log e} V(P_{XY}) - \frac{\rho_\Delta^3}{6} \Delta Y 
\]

(158)

\[
= \left( \frac{1 + \frac{\Delta}{n}}{1 + \alpha} - 1 \right) \rho_\Delta H(P_{XY}) - \frac{V(P_{XY})}{2\log e} \left( \rho_\Delta - \frac{(1 + \frac{\Delta}{n})(\theta_1 + \theta_2)\epsilon_\Delta \log e}{V(P_{XY})} \right)^2 
\]

\[
+ \frac{(1 + \frac{\Delta}{n})^2(\theta_1 + \theta_2)^2\epsilon_\Delta^2\log e}{2V(P_{XY})} - \frac{\rho_\Delta^3}{6} \Delta Y 
\]

(159)

\[
= \left( \frac{1 + \frac{\Delta}{n}}{1 + \alpha} - 1 \right) \frac{(1 + \frac{\Delta}{n})(\theta_1 + \theta_2)\epsilon_\Delta \log e}{V(P_{XY})} H(P_{XY}) + \frac{(1 + \frac{\Delta}{n})^2(\theta_1 + \theta_2)^2\epsilon_\Delta^2\log e}{2V(P_{XY})} 
\]

\[
- \frac{(1 + \frac{\Delta}{n})^3(\theta_1 + \theta_2)^3\epsilon_\Delta^3\log^3 e}{6V(P_{XY})^3} \Delta Y. 
\]

(160)

Choosing \(\rho_\Delta = \frac{(1 + \frac{\Delta}{n})\theta_1 + \theta_2}{V^{(\rho)}}\) and for some \(\overline{\rho}_\Delta \in [0, \rho_\Delta]\), we obtain

\[
\sup_{\rho \in [0,1]} \left( \rho \left(1 + \frac{\Delta}{n}\right)R_Y^{(\Delta)} - E_{Y|X}(\rho) \right) 
\]

(161)

\[
\geq \rho_\Delta \frac{1 + \frac{\Delta}{n}}{1 + \alpha} \left( H(P_{XY}|P_{X}) + (1 + \alpha)(\theta_1 \epsilon_\Delta) \right) - \rho_\Delta H(P_{XY}|P_{X}) - \frac{\rho_\Delta^2}{2\log e} V(P_{XY}|P_{X}) - \frac{\rho_\Delta^3}{6} E''_{Y|X}(\overline{\rho}_\Delta) 
\]

(162)

\[
\geq \left( \frac{1 + \frac{\Delta}{n}}{1 + \alpha} - 1 \right) \rho_\Delta H(P_{XY}|P_{X}) + \left(1 + \frac{\Delta}{n}\right)\theta_1 \rho_\Delta \epsilon_\Delta - \frac{\rho_\Delta^2}{2\log e} V(P_{XY}|P_{X}) - \frac{\rho_\Delta^3}{6} \Delta X 
\]

(163)

\[
= \left( \frac{1 + \frac{\Delta}{n}}{1 + \alpha} - 1 \right) \rho_\Delta H(P_{XY}|P_{X}) - \frac{V(P_{XY}|P_{X})}{2\log e} \left( \rho_\Delta - \frac{(1 + \frac{\Delta}{n})\theta_1 \epsilon_\Delta \log e}{V(P_{XY}|P_{X})} \right)^2 
\]

\[
+ \frac{(1 + \frac{\Delta}{n})^2\theta_1^2\epsilon_\Delta^2\log e}{2V(P_{XY}|P_{X})} - \frac{\rho_\Delta^3}{6} \Delta X 
\]

(164)

\[
= \left( \frac{1 + \frac{\Delta}{n}}{1 + \alpha} - 1 \right) \frac{(1 + \frac{\Delta}{n})\theta_1 \epsilon_\Delta \log e}{V(P_{XY}|P_{X})} H(P_{XY}|P_{X}) + \frac{(1 + \frac{\Delta}{n})^2\theta_1^2\epsilon_\Delta^2\log e}{2V(P_{XY}|P_{X})} 
\]

\[
- \frac{(1 + \frac{\Delta}{n})^3\theta_1^3\epsilon_\Delta^3\log^3 e}{6V(P_{XY}|P_{X})^3} \Delta X. 
\]

(165)

Finally, choosing \(\rho_\Delta = \frac{(1 + \frac{\Delta}{n})\theta_2 + \theta_3}{V^{(\rho)}}\) and for some \(\overline{\rho}_\Delta \in [0, \rho_\Delta]\), we obtain

\[
\sup_{\rho \in [0,1]} \left( \rho \left(1 + \frac{\Delta}{n}\right)R_Y^{(\Delta)} - E_{X|Y}(\rho) \right) 
\]

(166)

\[
\geq \rho_\Delta \frac{1 + \frac{\Delta}{n}}{1 + \alpha} \left( H(P_{Y}) + (1 + \alpha)(\theta_2 \epsilon_\Delta) \right) - \rho_\Delta H(P_{Y}|P_{X}) - \frac{\rho_\Delta^2}{2\log e} V(P_{Y}|P_{X}) - \frac{\rho_\Delta^3}{6} E''_{X|Y}(\overline{\rho}_\Delta) 
\]

(167)

\[
\geq \rho_\Delta I(P_{X}, P_{Y}|X) + \left( \frac{1 + \frac{\Delta}{n}}{1 + \alpha} - 1 \right) \rho_\Delta H(P_{Y}|P_{X}) + \left(1 + \frac{\Delta}{n}\right)\theta_2 \rho_\Delta \epsilon_\Delta - \frac{\rho_\Delta^2}{2\log e} V(P_{Y}|P_{X}) - \frac{\rho_\Delta^3}{6} \Delta Y 
\]

(168)

\[
= \rho_\Delta I(P_{X}, P_{Y}|X) + \left( \frac{1 + \frac{\Delta}{n}}{1 + \alpha} - 1 \right) \rho_\Delta H(P_{Y}|P_{X}) - \frac{V(P_{Y}|P_{X})}{2\log e} \left( \rho_\Delta - \frac{(1 + \frac{\Delta}{n})\theta_2 \epsilon_\Delta \log e}{V(P_{Y}|P_{X})} \right)^2 
\]

\[
+ \frac{(1 + \frac{\Delta}{n})^2\theta_2^2\epsilon_\Delta^2\log e}{2V(P_{Y}|P_{X})} - \frac{\rho_\Delta^3}{6} \Delta Y 
\]

(169)

\[
= \left( \frac{1 + \frac{\Delta}{n}}{1 + \alpha} - 1 \right) \frac{(1 + \frac{\Delta}{n})\theta_2 \epsilon_\Delta \log e}{V(P_{Y}|P_{X})} I(P_{X}, P_{Y}|X) + \frac{(1 + \frac{\Delta}{n})^2\theta_2^2\epsilon_\Delta^2\log e}{2V(P_{Y}|P_{X})} - \frac{(1 + \alpha)^3\theta_2^3\epsilon_\Delta^3\log^3 e}{6V(P_{Y}|P_{X})^3} \Delta Y. 
\]

(170)
Noting that $\epsilon_\Delta \to 0$, $\frac{\Delta}{n} \to \alpha$ as $\Delta \to \infty$, we acquire

$$\liminf_{\Delta \to \infty} \frac{E^{XY}((1 + \frac{\Delta}{n})R_X^{(\Delta)},(1 + \frac{\Delta}{n})R_Y^{(\Delta)})}{\epsilon_\Delta^2} \geq \frac{1}{\alpha} \frac{(1 + \alpha)^2(\theta_1 + \theta_2)^2 \log e}{2V(P_{XY})},$$  \hspace{1cm} (171)

$$\liminf_{\Delta \to \infty} \frac{E^{X|Y}((1 + \frac{\Delta}{n})R_X^{(\Delta)},(1 + \frac{\Delta}{n})R_Y^{(\Delta)})}{\epsilon_\Delta^2} \geq \frac{1}{\alpha} \frac{(1 + \alpha)^2\theta_2^2 \log e}{2V(P_{X|Y|P_Y})},$$  \hspace{1cm} (172)

$$\liminf_{\Delta \to \infty} \frac{E^{Y|X}((1 + \frac{\Delta}{n})R_X^{(\Delta)},(1 + \frac{\Delta}{n})R_Y^{(\Delta)})}{\epsilon_\Delta^2} \geq \infty,$$  \hspace{1cm} (173)

where (173) results from the fact that for dependent correlated sources, $I(P_X, P_{Y|X}) > 0$ and hence

$$\liminf_{\Delta \to \infty} \frac{(1 + \frac{\Delta}{n})\epsilon_\Delta^2 log e}{V(P_{Y|X|P_X})} I(P_X, P_{Y|X}) \liminf_{\Delta \to \infty} \frac{1}{\epsilon_\Delta} = \infty.$$  \hspace{1cm} (174)

Recalling (155), we obtain

$$\liminf_{\Delta \to \infty} \frac{-\log Pr(\mathcal{F})}{\epsilon_\Delta^2} \geq \frac{1}{\alpha} \frac{(1 + \alpha)^2(\theta_1 + \theta_2)^2 \log e}{2V(P_{XY})}.$$  \hspace{1cm} (176)

For the vertical boundary, only (161) can result in a finite value after being divided by $\epsilon_\Delta^2$ and taking the limit. For diagonal face, only (156) can result in a finite value after being divided by $\epsilon_\Delta^2$ and taking the limit. These observations complete the proofs of the other cases.

**Remark 7.** To prove the achievability part of Theorem 7, replace $n$ with $j_n$, then use the fact that $\frac{\Delta}{n} \to 1$ and $n \to \infty$ as $\Delta \to \infty$.

**E. Verification of (36)**

The verification of (36) is done as follows:

$$E^{DCP}_{\Phi}(R_X, R_Y) \leq \min \left\{ \min \{ E_X(R_X, R_Y, 0), E_X(R_X, R_Y, 1) \}, \min \{ E_Y(R_X, R_Y, 0), E_Y(R_X, R_Y, 1) \} \right\} \hspace{1cm} (177)

\leq \min \left\{ \max_{\rho \in [0,1]} E_{X|Y}(R_X, \rho), \max_{\rho \in [0,1]} E_{Y|X}(R_Y, \rho), \max_{\rho \in [0,1]} E_{XY}(R_X, R_Y, \rho) \right\} \hspace{1cm} (178)

= E^{sw}_{\Phi}(R_X, R_Y). \hspace{1cm} (179)

Furthermore,

$$\frac{1}{\alpha} E^{sw}_{\Phi}((1 + \alpha)R_X, (1 + \alpha)R_Y) := \frac{1}{\alpha} \min \left\{ \begin{array}{c}
\max_{\rho \in [0,1]} E_{X|Y}(R_X, (1 + \alpha)R_Y, \rho), \\
\max_{\rho \in [0,1]} E_{X|Y}(R_X, (1 + \alpha)R_Y, \rho), \\
\max_{\rho \in [0,1]} E_{XY}(R_X, (1 + \alpha)R_Y, \rho)
\end{array} \right\} \hspace{1cm} (180)

\geq \min \left\{ \begin{array}{c}
\max_{\rho \in [0,1]} \rho R_X, \\
\max_{\rho \in [0,1]} \rho R_Y, \\
\max_{\rho \in [0,1]} \rho (R_X + R_Y)
\end{array} \right\} \hspace{1cm} (182)

\geq \min \left\{ \begin{array}{c}
\max_{\rho \in [0,1]} \rho R_X - \log \sum_y P_Y(y) \left( \sum_x P_{X|Y}(x|y)^{1+\rho} \right)^{1+\rho}, \\
\max_{\rho \in [0,1]} \rho R_Y - \log \sum_x P_X(x) \left( \sum_y P_{Y|X}(y|x)^{1+\rho} \right)^{1+\rho}, \\
\max_{\rho \in [0,1]} \rho (R_X + R_Y) - (1 + \rho) \log \sum_{x,y} P_{XY}(x,y)^{1+\rho}
\end{array} \right\} \hspace{1cm} (183)

:= E^{sw}_{\Phi}(R_X, R_Y), \hspace{1cm} (184)
Define the set $T$ and it follows that we decode at most $\rho (1 + \rho)$ from $(\hat{X}, \hat{Y})$. Then our objective is to prove $\Pr\{X^n, Y^n \neq (\hat{X}, \hat{Y})\} \geq \Pr\{(X^n, Y^n) \in T_n\} - 3\exp(-n\gamma)$. (189)

It follows that

$$\Pr\{T_n\} = \Pr\{T_n \cap S_n^c\} + \Pr\{T_n \cap S_n\}$$

$$\leq \Pr\{S_n^c\} + \Pr\{T_n \cap S_n\}$$

$$= \Pr\{X^n, \hat{Y}^n \neq (X^n, Y^n)\} + \Pr\{T_n^{(1)} \cap S_n\} + \Pr\{T_n^{(2)} \cap S_n\} + \Pr\{T_n^{(12)} \cap S_n\}. \quad (192)$$

Define the set $S_n(y^n) := \{x^n : (x^n, y^n) \in S_n\}$. By considering any one-to-one mapping from $\lambda'_{n+\Delta}$ to $N_X$ messages, we observe that we decode at most $N_X$ sequences $x^n \in \lambda'_{n+\Delta}$ correctly. Thus $|S_n(y^n)| \leq N_X$.

By using the fact that for $(x^n, y^n) \in T_n^{(1)}$, $Q^n_{X,Y}(x^n|y^n) \leq \frac{\exp(-n\gamma)}{N_X}$, we obtain

$$\Pr\{T_n^{(1)} \cap S_n\} = \sum_{(x^n, y^n) \in T_n^{(1)} \cap S_n} Q^n_{X,Y}(y^n)Q^n_{X,Y}(x^n|y^n)$$

$$\leq \sum_{(x^n, y^n) \in T_n^{(1)} \cap S_n} Q^n_{Y}(y^n)\frac{\exp(-n\gamma)}{N_X} \quad (194)$$

$$\leq \sum_{(x^n+\Delta, y^n+\Delta) \in S_n} Q^n_{Y}(y^n)\frac{\exp(-n\gamma)}{N_X}$$

$$= \sum_{y^n \in Y^n} Q^n_{Y}(y^n)|S_n(y^n)|\frac{\exp(-n\gamma)}{N_X} \quad (195)$$

$$\leq \exp(-n\gamma) \quad (196)$$

Similarly, we can prove prove

$$\Pr\{T_{n+\Delta}^{(2)} \cap S_{n+\Delta}\} \leq \exp(-n\gamma)$$

$$\Pr\{T_{n+\Delta}^{(12)} \cap S_{n+\Delta}\} \leq \exp(-n\gamma) \quad (199)$$

This completes the proof of Lemma 8.

**REFERENCES**

[1] D. Slepian and J. K. Wolf, “Noiseless coding of correlated information sources,” IEEE Trans. Inf. Theory, vol. 19, no. 4, pp. 471–480, 1973.

[2] T. M. Cover, “A proof of the data compression theorem of Slepian and Wolf for ergodic sources (corresp.),” IEEE Trans. Inf. Theory, vol. 21, no. 2, pp. 226–228, 1975.

[3] I. Csiszar and J. Körner, Information theory: coding theorems for discrete memoryless systems. Cambridge University Press, 2011.

[4] R. G. Gallager, Information Theory and Reliable Communication. New York: Wiley, 1968.

[5] V. Y. F. Tan, “Asymptotic estimates in information theory with non-vanishing error probabilities,” Foundations and Trends® in Communications and Information Theory, vol. 11, no. 1–2, pp. 1–184, 2014.

[6] S. C. Draper, C. Chang, and A. Sahai, “Lossless coding for distributed streaming sources,” IEEE Trans. Inf. Theory, vol. 60, no. 3, pp. 1447–1474, 2014.

[7] C. Chang and A. Sahai, “The error exponent with delay for lossless source coding,” in Proc. Inf. Theory Workshop, Mar. 2006, pp. 252–256.

[8] ———, “The price of ignorance: The impact of side-information on delay for lossless source-coding,” arXiv:0712.0873, 2007.
[9] M. S. Pinsker, “Bounds of the probability and of the number of correctable errors for nonblock codes,” *Probl. Per. Inform.*, vol. 3, no. 4, pp. 58–71, 1967.

[10] C. Chang, “Streaming source coding with delay,” Ph.D. dissertation, UC Berkeley, 2007.

[11] H. Palaiyanur and A. Sahai, “Sequential decoding for lossless streaming source coding with side information,” *arXiv cs/0703120*, 2007.

[12] R. Venkataramanan and S. S. Pradhan, “Source coding with feed-forward: Rate-distortion theorems and error exponents for a general source,” *IEEE Trans. Inf. Theory*, vol. 53, no. 6, pp. 2154–2179, 2007.

[13] T. Matsuta and T. Uyematsu, “On the Wyner-Ziv source coding problem with unknown delay,” *IEICE Trans. Fundamentals*, vol. 97, no. 12, pp. 2288–2299, 2014.

[14] N. Ma and P. Ishwar, “On delayed sequential coding of correlated sources,” *IEEE Trans. Inf. Theory*, vol. 57, no. 6, pp. 3763–3782, 2011.

[15] N. Zhang, B. Vellambi, and K. Nguyen, “Delay exponent of variable-length random binning for point-to-point transmission,” in *Proc. Inf. Theory Workshop*, Nov. 2014, pp. 207–211.

[16] Y. Altug and A. Wagner, “Moderate deviations in channel coding,” *IEEE Trans. Info. Theory*, vol. 60, no. 8, pp. 4417–4426, 2014.

[17] Y. Polyanskiy and S. Verdu, “Channel dispersion and moderate deviations limits for memoryless channels,” in *Proc. 48th Annu. Allerton Conf.*, Sept. 2010, pp. 1334–1339.

[18] Y. Altug and A. Wagner, “Moderate deviation analysis of channel coding: Discrete memoryless case,” in *Proc. IEEE ISIT*, Jun. 2010, pp. 265–269.

[19] Y. Altug, A. Wagner, and I. Kontoyiannis, “Lossless compression with moderate error probability,” in *Proc. IEEE ISIT*, Jul. 2013, pp. 1744–1748.

[20] V. Y. F. Tan, “Moderate-deviations of lossy source coding for discrete and Gaussian sources,” in *Proc. IEEE ISIT*, Jul. 2012, pp. 920–924.

[21] V. Y. F. Tan, S. Watanabe, and M. Hayashi, “Moderate deviations for joint source-channel coding of systems with Markovian memory,” in *Proc. IEEE ISIT*, Jun. 2014, pp. 1687–1691.

[22] V. Y. F. Tan and O. Kosut, “On the dispersions of three network information theory problems,” *IEEE Trans. Info. Theory*, vol. 60, no. 2, pp. 881–903, 2014.

[23] R. Nomura and T. S. Han, “Second-order Slepian-Wolf coding theorems for non-mixed and mixed sources,” *IEEE Trans. Info. Theory*, vol. 60, no. 9, pp. 5553–5572, 2014.

[24] T. S. Han and H. Koga, *Information-spectrum methods in information theory*. Springer, 2003.

[25] S. Miyake and F. Kanaya, “Coding theorems on correlated general sources,” *IEICE Trans. Fundamentals*, vol. 78, no. 9, pp. 1063–1070, 1995.

[26] R. G. Gallager, “Source coding with side information and universal coding,” LIDS, MIT, Tech. Rep., 1976.

[27] A. Sahai, “Why do block length and delay behave differently if feedback is present?” *IEEE Trans. Inf. Theory*, vol. 54, no. 5, pp. 1860–1886, 2008.

[28] S. Verdú and I. Kontoyiannis, “Optimal lossless data compression: Non-asymptotics and asymptotics,” *IEEE Trans. Inf. Theory*, vol. 60, no. 2, pp. 777–795, 2014.

[29] M. Hayashi and R. Matsumoto, “Secure multiplex coding with dependent and non-uniform multiple messages,” *arxiv:1202.1332*, 2012.

[30] A. D. Wyner, “On source coding with side information at the decoder,” *IEEE Trans. Inf. Theory*, vol. 21, no. 3, pp. 294–300, 1975.

[31] R. Ahlswede and J. Körner, “Source coding with side information and a converse for the degraded broadcast channel,” *IEEE Trans. Inf. Theory*, vol. 21, no. 6, pp. 629–637, 1975.

[32] V. N. Koshelev, “On a problem of separate coding of two dependent sources,” *Probl. Per. Inform.*, vol. 13, no. 1, pp. 26–32, 1977.