Moderate Deviations Asymptotics for Streaming Compression of Correlated Sources

Lin Zhou, Vincent Y. F. Tan, and Mehul Motani

arXiv:1604.07151v1 [cs.IT] 25 Apr 2016

Abstract

In this paper, we consider the problem of blockwise streaming compression of a pair of correlated sources, which we term streaming Slepian-Wolf coding. We study the moderate deviations regime in which the rate pairs of a sequence of codes converges, along a straight line, to various points on the boundary of the Slepian-Wolf region at a speed slower than the inverse square root of the blocklength n, while the error probability decays subexponentially fast in n. Our main result focuses on directions of approaches to corner points of the Slepian-Wolf region. It states that for each correlated source and all corner points, there exists a non-empty subset of directions of approaches such that the moderate deviations constant (the constant of proportionality for the subexponential decay of the error probability) is enhanced (over the non-streaming case) by at least a factor of T, the block delay of decoding symbol pairs. We specialize our main result to the setting of lossless streaming source coding and generalize this result to the setting where we have different delay requirements for each of the two source blocks. The proof of our main result involves the use of various analytical tools and amalgamates several ideas from the recent information-theoretic streaming literature. We adapt the so-called truncated memory idea from Draper and Khisti (2011) and Lee, Tan and Khisti (2015) to ensure that the effect of error accumulation is nullified in the limit of large blocklengths. We also adapt the use of the so-called minimum empirical suffix entropy decoder which was used by Draper, Chang and Sahai (2014) to derive achievable error exponents for streaming Slepian-Wolf coding.

Index Terms

Slepian-Wolf coding, Streaming compression, Moderate deviations, Truncated memory, Weighted empirical suffix-entropy decoder

I. Introduction

Streaming is the dominant mode of consuming videos on smart phones, tablets and laptops. Instead of downloading an entire movie at once, we watch as we download. The importance of streaming video is evidenced by the fact that Netflix had over 65 million subscribers worldwide for its streaming movie service, as of July 2015 [1]. The streaming of video can also be performed in a distributed manner [2], [3], i.e., streaming video from multiple mirror sites to subscribers. This is also the idea in content delivery networks [4].

In this paper, we study a *streaming* version of Slepian-Wolf (SW) source coding [5]. Our setting is shown pictorially in Figure 1. In our setting, the correlated source generates one source block pair (each of length n) and the k-th pair of encoders has access to *all* k source block pairs generated up to and including the current time. The decoders incur a block delay T when decoding each pair of source blocks. We would like to evaluate the minimum of the error probabilities over *all* source blocks under the so-called moderate deviations regime. We note our setting is, in general, *different* from the settings in [6] and [7] where the authors restricted the encoders to have access to accumulated *symbol pairs* and not *blocks* of symbol pairs at each time. Hence, we term the settings in [6] and [7] as *symbolwise* streaming and our setting as *blockwise* streaming.

There are three asymptotic regimes in the study of Shannon-theoretic problems when we want to establish the relationships and tradeoffs between the blocklength, coding rate(s) and the error probability. They are, respectively, the error exponent (large deviations), second-order (central limit) and moderate deviations regimes. Here we use fixed-length lossless source coding as an example to discuss these three regimes. 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 [8], [9]. In second-order or normal approximation analysis [10], [11], [12], 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. In contrast, in the study of moderate deviations [13], [14], [15], 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 roughly $\exp(-\nu n^{1-2t})$ for $t \in (0, 1/2)$. The constant ν is known as the moderate deviations constant and is the object of study in moderate deviations analysis here. The moderate deviations regime can be seen as a bridge between the error exponent and second-order regimes.

The authors are with the Department of Electrical and Computer Engineering, National University of Singapore (emails: lzhou@u.nus.edu, vtan@nus.edu.sg, motani@nus.edu.sg).

The work of the authors is supported in part by a Ministry of Education Tier 2 grant (grant number R-263-000-B61-112).

A. Related Work

The papers that are most related to the present one are those by Lee, Tan and Khisti [16] and Draper, Chang and Sahai [7]. In [16], the authors study the information-theoretic limits of the blockwise streaming version of channel coding in the moderate deviations and central limit regimes. In [7], the authors derived lower bounds for the error exponent (reliability function) of symbolwise streaming SW coding by using random binning, minimum empirical entropy decoding, and maximum likelihood decoding. The authors in [7] considered polynomially many error events and so the one with the smallest exponent dominates. In other works on information-theoretic limits of streaming compression and transmission, Chang and Sahai [17] derived bounds on the error exponent for symbolwise streaming of lossless compression. They demonstrated similar results for the case with both encoder and decoder side information in [18]. They also extended the feedforward decoder idea (which originated from Pinsker [19]) to the case with decoder side information to derive an upper (converse) bound on the error exponent. Other works by Chang on information-theoretic limits in symbolwise streaming are summarized in [6].

Concerning other works on streaming and source coding with delayed decoding, Palaiyanur [20] 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 [21] considered the source coding with feedforward problem and derived lossy source coding theorems for general sources. Matsuta and Uyematsu [22] considered the lossy source coding problem with delayed side information. Ma and Ishwar [23] focused on delayed sequential coding of correlated video sources. Finally, Zhang [24] analyzed the error exponent of lossless streaming compression of a single source using variable-length sequential random binning.

Here we also mention a few works on moderate deviations analysis in information theory. Chen et al. [25] and He et al. [26] initiated the study of moderate deviations for fixed-to-variable length source coding with decoder side information. Moreover, Chen et al. [25] also studied moderate deviations of so-called cyclic symmetric channels. For fixed-to-fixed length analysis, Altuğ and Wagner [13] initiated the study of moderate deviations in Shannon theory in the context of discrete memoryless channels. Polyanksiy and Verdú [14] relaxed some assumptions in the conference version of Altuğ and Wagner's work [27] and they also considered moderate deviations for AWGN channels. Altuğ, Wagner and Kontoyiannis [15] considered moderate deviations for lossless source coding. Other works on moderate deviations in information theory include [28], [29]. 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 this paper, we derive an achievable moderate deviations constant for the blockwise streaming version of SW coding. We show that for each correlated source, there exists a non-empty subset of directions of approaches to the boundary of the SW region for which the moderate deviations constant is enhanced by a factor of T, the block delay of decoding symbol pairs, compared with non-streaming case. Furthermore, we specialize our results to streaming lossless source coding and generalize our results to the setting where we have different delay requirements for each of the two source blocks.

Despite being analogous to [16] in terms of the main result, our proof is significantly different. We adapt two key ideas to our blockwise streaming setting: the *truncated memory* idea in [16] which originated from [30] and the *minimum weighted empirical suffix-entropy* decoder in [7]. The analysis of this paper is different from both [16] and [7]. In [16], the truncated memory idea was used in the central limit regime for streaming channel coding. In this paper, we adapt the idea and apply it to the moderate deviations regime of our streaming SW setting. We argue that truncation is, in general, necessary for us to prove our main result (otherwise the accumulation of error probabilities will result). In [7], the minimum weighted empirical suffix-entropy decoder was used to derive an achievable error exponent for symbolwise streaming SW coding. We adapt the decoding rule to be suited to our moderate deviations setting so as to obtain an analogous term as in the exponent calculation for [7, Theorem 6, Case (iii)]. Subsequently, we Taylor expand the exponent to obtain an achievable moderate deviations constant. We remark that this final Taylor expansion step is not straightforward, requiring some analytical techniques inspired by Polyanskiy [31, Lemma 48]. This is because the exponent involves an additional minimization over a scalar parameter (see [7, Theorem 6, Case (iii)] and Lemma 6).

We emphasize that in [7], a quantification of the improvement of the error exponent in the streaming case vis-à-vis the block coding setting was absent. In this paper, we provide a definitive answer to the question of how much we gain in the moderate deviations regime from streaming setup with a block delay T. We show analytically that there is a multiplicative gain of T in the moderate deviations constant over the non-streaming setting in many scenarios. We analyze several sources in Section III and calculate the various directions of approaches in which we can attain this gain.

C. Organization of the Paper

The rest of the paper is organized as follows. In Section II, we set up the notation, formulate the problem of streaming SW coding and present our achievable moderate deviations constant. In Section III, we provide three numerical examples to

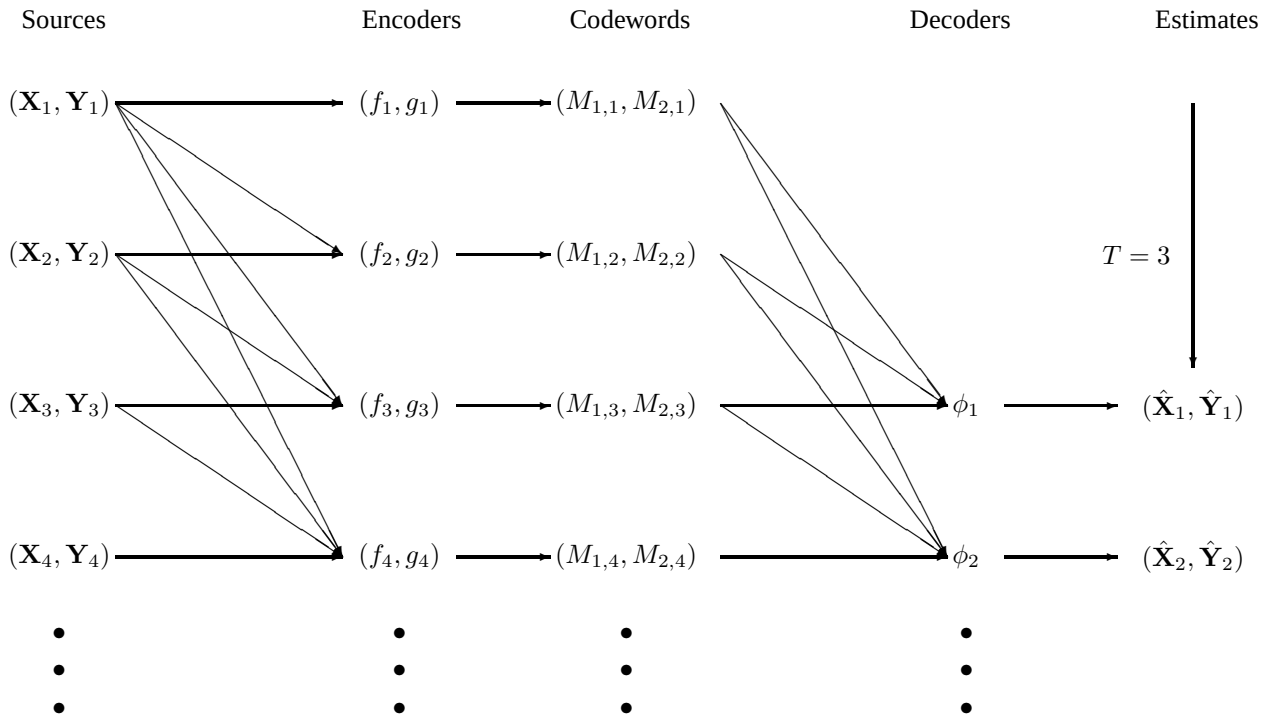


Fig. 1. Lossless streaming compression of correlated sources with $T = 3$. At each time k, a new source block pair $(\mathbf{X}_k, \mathbf{Y}_k)$ is fed into encoders (f_k, g_k). The encoders then produce output as codewords $(M_{1,k}, M_{2,k})$ using the accumulated source blocks $(\mathbf{X}^k, \mathbf{Y}^k)$. For the case of $T = 3$, the decoder is required to estimate (X_k, Y_k) after receiving the codewords with indices from 1 to $k + 2 = k + T - 1$.

illustrate our results. In particular, we delineate the set of directions for which we can achieve a gain of at least the block delay T in the moderate deviations constant. In Section IV, we specialize our results to streaming lossless source coding and generalize our results to the scenario where we impose different delay requirements on different source blocks. In Section V, we present the proof of our achievable moderate deviations constant. In Section VI, we present the proof for the scenario of different delay requirements. Finally, in Section VII, we conclude the paper and propose future research topics. Auxiliary lemmata are proved in the Appendices.

II. Streaming Slepian Wolf Coding

A. 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. We use $\mathbf{X}$ to denote X^n and $\mathbf{X}_a^b$ to denote $X_{n(a-1)+1}^{bn}$. All logarithms are base e (natural logarithm). As usual, for any $k \in \mathbb{N}$, $p \log^k p = 0$ if $p = 0$. Given two integers a and b, we use $[a : b]$ to denote the set $\{a, a + 1, \ldots, b - 1, b\}$. We use standard asymptotic notation such as $O(\cdot)$ and $o(\cdot)$ [32].

The set of all probability distributions on $\mathcal{X}$ is denoted as $\mathcal{P}(\mathcal{X})$ and the set of all conditional probability distributions from $\mathcal{X}$ to $\mathcal{Y}$ is denoted as $\mathcal{P}(\mathcal{Y}|\mathcal{X})$. Given $P \in \mathcal{P}(\mathcal{X})$ and $V \in \mathcal{P}(\mathcal{Y}|\mathcal{X})$, we use $P \times V$ to denote the joint distribution induced by P and V. We use the method of types extensively and we follow the notation in [33]. Given a sequence x^n, the empirical distribution (type) is denoted as $\hat{T}_{x^n}$. The set of types formed from length-n sequences in $\mathcal{X}$ is denoted as $\mathcal{P}_n(\mathcal{X}) = \{\hat{T}_{x^n} \in \mathcal{P}(\mathcal{X}) : x^n \in \mathcal{X}^n\}$. Given $P \in \mathcal{P}_n(\mathcal{X})$, the set of all sequences of length-n with type P (the type class) is denoted as $\mathcal{T}_P$. Given $x^n \in \mathcal{T}_P$, the set of all sequences $y^n \in \mathcal{Y}^n$ such that the joint type of (x^n, y^n) is $P \times V$ is denoted as $\mathcal{T}_V(x^n)$, the V-shell. The set of all stochastic matrices $V \in \mathcal{P}(\mathcal{Y}|\mathcal{X})$ for which the V-shell of a sequence of type P in $\mathcal{X}^n$ is not empty is denoted as $\mathcal{V}_n(\mathcal{Y}; P)$.

For information-theoretic quantities, we interchangeably use $H(P_X)$ and $H(X)$ to denote the entropy of a random variable X with distribution P_X. Similarly, we interchangeably use $H(P_{X|Y}|P_Y)$ or $H(X|Y)$ to denote the conditional entropy. The mutual information and relative entropy are denoted in a similar manner.

B. System Model and Problem Formulation

We consider a streaming version of Slepian-Wolf problem, which is termed *streaming Slepian-Wolf coding*. The standard Slepian-Wolf problem was solved in [5] and [34]. Similarly to [5], we have two correlated sources to be separately compressed and jointly reconstructed. However, our streaming model differs from [5] in three main aspects:

1) We have a *sequence* (a finite number) of source blocks as inputs for each encoder;
2) We have a *sequence* (countably infinite) encoders and decoders;
3) We incur a *delay* at the decoder to decode a specific source block pair.

Our streaming setting is also different from [7] in the sense that we allow encoders to have access to one more source *block* per unit time instead of one more *symbol* as in [7].

Consider a discrete memoryless source (DMS) with joint probability mass function (pmf) P_{XY} and a finite alphabet $\mathcal{X}\times\mathcal{Y}$. We define a streaming SW code formally as follows.

Definition 1 (Streaming Code)**.** *An $(n, N_1, N_2, T, \epsilon_n)$-code for streaming SW coding consists of*

1) *A sequence of correlated source blocks $\{(\mathbf{X}_k, \mathbf{Y}_k)\}_{k\geq 1}$, where $(\mathbf{X}_k, \mathbf{Y}_k) \in \mathcal{X}^n \times \mathcal{Y}^n$ is an i.i.d. sequence with common distribution $P_{XY} \in \mathcal{P}(\mathcal{X}\times\mathcal{Y})$;*
2) *A sequence of encoding function pairs $f_k : \mathcal{X}^{kn} \to \mathcal{M}_1$ that maps the accumulated source blocks $\mathbf{X}^k \in \mathcal{X}^{kn}$ to a codeword $M_{1,k} \in \mathcal{M}_1$ and $g_k : \mathcal{Y}^{kn} \to \mathcal{M}_2$ that maps the accumulated source blocks $\mathbf{y}^k \in \mathcal{Y}^{kn}$ to a codeword $M_{2,k} \in \mathcal{M}_2$. Here $|\mathcal{M}_1| = N_1$ and $|\mathcal{M}_2| = N_2$;*
3) *A sequence of decoding functions $\phi_k : \mathcal{M}_1^{k+T-1} \times \mathcal{M}_2^{k+T-1} \to \mathcal{X}^n \times \mathcal{Y}^n$ that maps the accumulated codewords $\{(M_{1,j}, M_{2,j})\}_{j=1}^{k+T-1} \in \mathcal{M}_1^{k+T-1} \times \in \mathcal{M}_2^{k+T-1}$ to a source block pair $(\hat{\mathbf{X}}_k, \hat{\mathbf{Y}}_k)$,*

which satisfies

$$\sup_{k\in\mathbb{N}} \Pr\left(\left(\hat{\mathbf{X}}_k, \hat{\mathbf{Y}}_k\right) \neq (\mathbf{X}_k, \mathbf{Y}_k)\right) \leq \epsilon_n. \tag{1}$$

We remark that given fixed blocklength n, an $(n, N_1, N_2, T, \epsilon_n)$-code for streaming SW coding consists of a *sequence* of encoding and decoding functions. Specifically, when $T = 1$ and $k = 1$ (only *one* source block is considered), we recover the standard SW coding scenario [5]. To illustrate our streaming setup, we show in Figure 1 the case of $T = 3$.

To define the next concept succinctly, let R_X^* and R_Y^* be two fixed rates. The rate pair (R_X^*, R_Y^*) will be taken to be on the boundary of the optimal rate region for SW coding [5] in Figure 2. There are five cases in total, but Cases (iv) and (v) are symmetric to Cases (ii) and (i) respectively. Hence, in the presentation of definition and main result, we illustrate only the first three cases.

Definition 2 (Moderate Deviations Constant)**.** Consider any correlated source with joint pmf P_{XY} and any positive sequence $\{\xi_n\}_{n\in\mathbb{N}}$ satisfying $\xi_n \to 0$ and $\frac{\log n}{n\xi_n^2} \to 0$ as $n \to \infty$.[1] Let $\boldsymbol{\theta} = (\theta_1, \theta_2) \in \mathbb{R}^2$ be a real vector. A number ν is said to be a $(R_X^*, R_Y^*, \boldsymbol{\theta}, T)$-*achievable moderate deviations constant* if there exists a sequence of $(n, N_1, N_2, T, \epsilon_n)$-codes such that

$$\limsup_{n\to\infty} \frac{\log N_1 - nR_X^*}{n\xi_n} \leq \theta_1 \tag{2}$$

$$\limsup_{n\to\infty} \frac{\log N_2 - nR_Y^*}{n\xi_n} \leq \theta_2. \tag{3}$$

and

$$\liminf_{n\to\infty} -\frac{\log \epsilon_n}{n\xi_n^2} \geq \nu. \tag{4}$$

The supremum of all $(R_X^*, R_Y^*, \boldsymbol{\theta}, T)$-achievable moderate deviations constants is denoted as $\nu^*(R_X^*, R_Y^*, \boldsymbol{\theta}, T)$.

We remark that in the above definition, in order to approach a boundary rate pair from a sequence of non-boundary rate pairs *inside* (in the interior of) the SW coding region, we need to impose different conditions on $\boldsymbol{\theta}$ for the different cases as follows:

- Case (i): The feasible set of $\boldsymbol{\theta}$ is

$$\Theta_{\text{(i)}} := \left\{\boldsymbol{\theta} \in \mathbb{R}^2 : \theta_1 > 0, \ \ -\infty < \theta_2 < \infty\right\}. \tag{5}$$

- Case (ii): The feasible set of $\boldsymbol{\theta}$ is

$$\Theta_{\text{(ii)}} := \left\{\boldsymbol{\theta} \in \mathbb{R}^2 : \theta_1 > 0, \ \ -\theta_1 < \theta_2 < \infty\right\}. \tag{6}$$

- Case (iii): The feasible set of $\boldsymbol{\theta}$ is

$$\Theta_{\text{(iii)}} := \left\{\boldsymbol{\theta} \in \mathbb{R}^2 : -\infty < \theta_1 < \infty, \ \ -\theta_1 < \theta_2 < \infty\right\} \tag{7}$$

$$= \Theta_{\text{(ii)}} \bigcup \left\{\boldsymbol{\theta} \in \mathbb{R}^2 : \theta_1 < 0, \ \ -\theta_1 < \theta_2 < \infty\right\}. \tag{8}$$

The conditions on $\boldsymbol{\theta}$ for Cases (iv) and (v) are omitted since they are similar to conditions for Cases (ii) and (i).

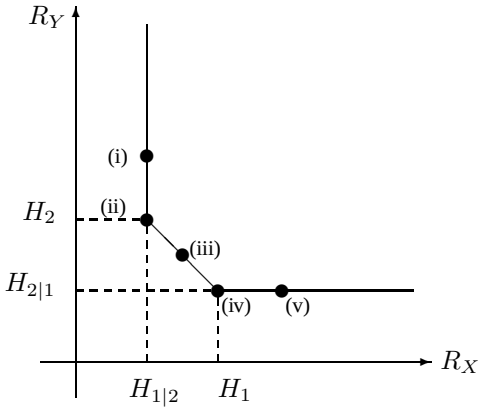


Fig. 2. Illustration of the different cases in Definition 2, Theorems 1 and 2 where $H_1 = H(P_X)$ and $H_{1|2} = H(P_{X|Y}|P_Y)$ etc.

C. Preliminaries and An Assumption

For a given source with pmf P_{XY} on alphabet $\mathcal{X} \times \mathcal{Y}$, the *joint and conditional source dispersions* or *varentropies* [35] are respectively defined as

$$\mathrm{V}(P_{XY}) := \sum_{x,y} P_{XY}(x,y) \left(-\log P_{XY}(x,y) - H(P_{XY})\right)^2, \tag{9}$$

$$\mathrm{V}(P_{X|Y}|P_Y) := \sum_{x,y} P_{XY}(x,y) \left(-\log P_{X|Y}(x|y) - H(P_{X|Y}|P_Y)\right)^2, \tag{10}$$

and $\mathrm{V}(P_{Y|X}|P_X)$ is defined in a similar manner as $\mathrm{V}(P_{X|Y}|P_Y)$ with X and Y interchanged. Note that in (9) and (10), it suffices to only sum over elements $(x,y) \in \mathcal{X} \times \mathcal{Y}$ such that $P_{XY}(x,y) > 0$.

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

D. Moderate Deviations Asymptotics for Standard Slepian-Wolf Coding

In this section, we present the moderate deviations constant for the (standard) Slepian-Wolf problem, i.e., the *non-streaming case*. Let (R_X^*, R_Y^*) be fixed as a boundary rate pair of the Slepian Wolf rate region in Figure 2. Let the optimal moderate deviations constant for non-streaming Slepian-Wolf be denoted as $\nu_{\mathrm{sw}}^*(R_X^*, R_Y^*, \boldsymbol{\theta})$.

Theorem 1 (Non-Streaming Moderate Deviations Constant)**.** *Depending on (R_X^*, R_Y^*), there are five cases, of which we present three here. The moderate deviations constant for non-streaming SW source coding is*

1) *Case (i): $R_X^* = H(P_{X|Y}|P_Y)$ and $R_Y^* > H(P_Y)$*

$$\nu_{\mathrm{sw}}^*(R_X^*, R_Y^*, \boldsymbol{\theta}) = \frac{\theta_1^2}{2\mathrm{V}(P_{X|Y}|P_Y)}. \tag{11}$$

2) *Case (ii): $R_X^* = H(P_{X|Y}|P_Y)$ and $R_Y^* = H(P_Y)$*

$$\nu_{\mathrm{sw}}^*(R_X^*, R_Y^*, \boldsymbol{\theta}) = \min\left\{\frac{\theta_1^2}{2\mathrm{V}(P_{X|Y}|P_Y)}, \frac{(\theta_1+\theta_2)^2}{2\mathrm{V}(P_{XY})}\right\} \tag{12}$$

3) *Case (iii): $R_X^* + R_Y^* = H(P_{XY})$, $H(P_{X|Y}|P_Y) < R_X^* < H(P_X)$ and $H(P_{Y|X}|P_X) < R_Y^* < H(P_Y)$*

$$\nu_{\mathrm{sw}}^*(R_X^*, R_Y^*, \boldsymbol{\theta}) = \frac{(\theta_1+\theta_2)^2}{2\mathrm{V}(P_{XY})}. \tag{13}$$

Hayashi and Matsumoto derived an achievability result for the moderate deviations constants similar to Theorem 1 in [36, Lemma 89]. However, no corresponding converse result was proved.

We remark that there are at least two ways to prove Theorem 1 for both achievability and converse parts. One way is to Taylor expand the error exponents for SW coding in a similar manner as in [13]. In Appendix B, we provide some preliminary lemmas for this calculation. The second way is to leverage information spectrum bounds [37, Lemmas 7.2.1 and 7.2.2] and the moderate deviations theorem [38, Theorem 3.7.1]. We believe that the result in Theorem 1 (sans the achievability part [36]) does not appear explicitly in previous works but it is straightforward and thus we omit its proof.

[1]As an example, $\xi_n = n^{-t}$ for any $t \in (0, 1/2)$ satisfies the two conditions.

E. Moderate Deviations Asymptotics for Streaming Slepian-Wolf Coding

We now state the main result of this paper.

Theorem 2 (Streaming Moderate Deviations Constant)**.** *The moderate deviations constant for streaming SW coding satisfies*

1) *Case (i):* $R_X^* = H(P_{X|Y}|P_Y)$ *and* $R_Y^* > H(P_Y)$

$$\nu^*(R_X^*, R_Y^*, \boldsymbol{\theta}, T) \geq \frac{T\theta_1^2}{2\mathrm{V}(P_{X|Y}|P_Y)}. \tag{14}$$

2) *Case (ii):* $R_X^* = H(P_{X|Y}|P_Y)$ *and* $R_Y^* = H(P_Y)$

$$\nu^*(R_X^*, R_Y^*, \boldsymbol{\theta}, T) \geq T \min\left\{ \inf_{\gamma\in[0,1]} \frac{(\theta_1+(1-\gamma)\theta_2)^2}{2\left(\gamma\mathrm{V}(P_{X|Y}|P_Y)+(1-\gamma)\mathrm{V}(P_{XY})\right)}, \frac{(\theta_1+\theta_2)^2}{2\mathrm{V}(P_{XY})} \right\} \tag{15}$$

3) *Case (iii):* $R_X^* + R_Y^* = H(P_{XY})$, $H(P_{X|Y}|P_Y) < R_X^* < H(P_X)$ *and* $H(P_{Y|X}|P_X) < R_Y^* < H(P_Y)$

$$\nu^*(R_X^*, R_Y^*, \boldsymbol{\theta}, T) \geq \frac{T(\theta_1+\theta_2)^2}{2\mathrm{V}(P_{XY})}. \tag{16}$$

The proof of Theorem 2 is provided in Section V. Several remarks are now in order.

1) At a high level, our proof proceeds by combining the truncated memory idea for second-order analysis from [16] and the so called minimum weighted empirical suffix-entropy decoding idea from [7]. Subsequently, we Taylor expand the resultant exponents at rates near the first-order fundamental limit and invoke some continuity arguments from [31]. We remark that, unlike the moderate deviations analysis in [16], the truncated memory idea appears to be necessary. Otherwise, the decoding error probability of a particular source block pair is upper bounded by the sum of the probabilities of $2(k+T-1)$ "dominant" error events. In general, k (the index of the source block we wish to decode) can be much larger than an exponential function of the blocklength so if truncation is not performed, this upper bound on the error probability would be vacuous as it would exceed one.
2) The first term of the minimization in (15) is somewhat unusual so we comment on its significance here. This term results from the analysis of our streaming setup. Essentially, we perform standard random binning [34] and in the decoding procedure, several dominant error events result from the use of the minimum weighted suffix empirical entropy decoder (see (51)) due to the streaming scenario. More precisely, let l be the first index where the true source sequence $\mathbf{x}^{T_k}$ differs from a competitor source sequence $\tilde{\mathbf{x}}^{T_k}$. Similarly, let m denote the first index where the true source sequence $\mathbf{y}^{T_k}$ differs from a competitor source sequence $\tilde{\mathbf{y}}^{T_k}$. The minimization over $\gamma \in [0,1]$ in (15) results from identifying the dominant error events over all feasible choices of l and m.
3) We also comment on why the slightly stringent condition $\frac{n\xi_n^2}{\log n} \to \infty$ (as opposed to $n\xi_n^2 \to \infty$ which is standard in moderate deviations analysis [13], [14]) is needed in our analysis. This is because when we use the truncated memory coding scheme, it turns out that we have $O(n^{\frac{1}{2}+\delta})$ dominant error events, thus we need the additional logarithm to nullify these error events. This stringent condition on $\{\xi_n\}_{n\in\mathbb{N}}$ *cannot* be relaxed even if we employ an analogue of the (non-universal) maximum likelihood decoder.
4) In the streaming SW setting, for Cases (i) and (iii), we can clearly achieve *at least* T times of the moderate deviations constant compared to the non-streaming case (see Theorem 1). However, for Case (ii), which is most interesting, we cannot guarantee a gain of T in moderate deviations constant for all possible values of $\boldsymbol{\theta} \in \Theta_{(\mathrm{ii})}$ (recall (6)). In Proposition 3, we show that for each correlated source P_{XY}, there exists a non-empty set of values of $\boldsymbol{\theta} \in \Theta_{(\mathrm{ii})}$ such that we have at least a multiplicative gain of T in the moderate deviations constant compared to Case (ii) of the non-streaming setting. Recall that $\boldsymbol{\theta}$ designates the direction of approach of a pair of rates towards the boundary of the SW region.

For simplicity in notation, let the objective function in (15) be denoted as

$$f(P_{XY}, \gamma, \boldsymbol{\theta}) := \frac{(\theta_1+(1-\gamma)\theta_2)^2}{2\left(\gamma\mathrm{V}(P_{X|Y}|P_Y)+(1-\gamma)\mathrm{V}(P_{XY})\right)}. \tag{17}$$

Also define the functions

$$g_1(P_{XY}) := \frac{\mathrm{V}(P_{XY})-\mathrm{V}(P_{X|Y}|P_Y)}{2\mathrm{V}(P_{X|Y}|P_Y)}, \qquad \text{and} \tag{18}$$

$$g_2(P_{XY}) := \min\left\{ \sqrt{\frac{\mathrm{V}(P_{XY})}{\mathrm{V}(P_{X|Y}|P_Y)}} - 1, \frac{\mathrm{V}(P_{XY})-\mathrm{V}(P_{X|Y}|P_Y)}{\mathrm{V}(P_{XY})+\mathrm{V}(P_{X|Y}|P_Y)} \right\}. \tag{19}$$

We remark that $g_1(P_{XY}) \geq g_2(P_{XY})$ for all sources P_{XY} with $\mathrm{V}(P_{X|Y}|P_Y) > 0$. This can be verified by considering the different relationships between $\mathrm{V}(P_{XY})$ and $\mathrm{V}(P_{X|Y}|P_Y)$. For the case in which $\mathrm{V}(P_{XY}) = \mathrm{V}(P_{X|Y}|P_Y)$, we have

$g_1(P_{XY}) = g_2(P_{XY}) = 0$. When $\mathrm{V}(P_{XY}) > \mathrm{V}(P_{X|Y}|P_Y)$, we have

$$g_1(P_{XY}) > \frac{\mathrm{V}(P_{XY}) - \mathrm{V}(P_{X|Y}|P_Y)}{\mathrm{V}(P_{XY}) + \mathrm{V}(P_{X|Y}|P_Y)} \geq g_2(P_{XY}). \tag{20}$$

Lastly, when $\mathrm{V}(P_{XY}) < \mathrm{V}(P_{X|Y}|P_Y)$, we have

$$g_1(P_{XY}) = \frac{1}{2}\left(\frac{\mathrm{V}(P_{XY})}{\mathrm{V}(P_{X|Y}|P_Y)} - 1\right) > \sqrt{\frac{\mathrm{V}(P_{XY})}{\mathrm{V}(P_{X|Y}|P_Y)}} - 1 \geq g_2(P_{XY}) \tag{21}$$

where the strict inequality holds because $\frac{1}{2}(t-1) > \sqrt{t}-1$ for all $t \in [0,1)$ (due to the strict concavity of $t \mapsto \sqrt{t}-1$). From the bounds in (20) and (21), we also infer that for all sources P_{XY}, $g_1(P_{XY}) = g_2(P_{XY})$ if and only if $\mathrm{V}(P_{XY}) = \mathrm{V}(P_{X|Y}|P_Y)$.

We now state necessary and sufficient conditions on $\boldsymbol{\theta} \in \Theta_{\text{(ii)}}$ (recall (6)) for obtaining a multiplicative gain of T.

Proposition 3 (Conditions for Obtaining a Multiplicative Gain of T)**.** *For each source distribution P_{XY}, the moderate deviations constant in the streaming setting is at least a factor of T larger than the non-streaming counterpart, i.e.,*

$$\min\left\{\inf_{\gamma\in[0,1]} f(P_{XY},\gamma,\boldsymbol{\theta}), \frac{(\theta_1+\theta_2)^2}{2\mathrm{V}(P_{X|Y}|P_Y)}\right\} = (12), \tag{22}$$

if and only if $\boldsymbol{\theta} \in \Theta_{\text{(ii)}}$ satisfies

$$\frac{\theta_2}{\theta_1} \geq g_1(P_{XY}), \tag{23}$$

or

$$-1 < \frac{\theta_2}{\theta_1} \leq g_2(P_{XY}). \tag{24}$$

The proof of Proposition 3 is provided in Appendix A.

Note that for any source P_{XY}, $g_1(P_{XY}) \in \mathbb{R}$. Thus, Proposition 3, and in particular (23), allows us to conclude that for each correlated source, there exists a non-empty set of directions of approaching a boundary rate pair (R_X^*, R_Y^*), parametrized by $\boldsymbol{\theta}$, such that we can achieve a multiplicative gain of T in the moderate deviations constant compared with Case (ii) of the non-streaming SW coding. In addition, from Proposition 3 and the observation after (21) (namely that $g_1(P_{XY}) = g_2(P_{XY}) \iff \mathrm{V}(P_{XY}) = \mathrm{V}(P_{X|Y}|P_Y)$), we notice that for sources P_{XY} such that $\mathrm{V}(P_{XY}) \neq \mathrm{V}(P_{X|Y}|P_Y)$, we conclude that there will also be a non-empty set of directions of approaching the boundary in which we *cannot*, in general, achieve a gain of T in the streaming scenario (using the coding scheme and analyses delineated in Section V).

III. Numerical Examples

In this section, we present three sources satisfying the conditions as stated in Section II-C and illustrate conditions on $\boldsymbol{\theta}$ for which we can achieve a multiplicative gain of T in moderate deviations constant compared to non-streaming setting. Throughout, we focus only on Case (ii) of Theorem 2.

A. Doubly Symmetric Binary Source

We consider the example of a doubly symmetric binary source (DSBS) where $\mathcal{X} = \mathcal{Y} = \{0,1\}$, $P_{XY}(0,0) = P_{XY}(1,1) = \frac{1-p}{2}$ and $P_{XY}(0,1) = P_{XY}(1,0) = \frac{p}{2}$ for some $p \in \left[0, \frac{1}{2}\right]$. Note that $P_X(0) = P_Y(0) = \frac{1}{2}$, $P_{X|Y}(x|y) = 1-p$ if $x = y$ and $P_{X|Y}(x|y) = p$ if $x \neq y$. The joint and conditional entropies are

$$H(P_{XY}) = 2\left(\frac{1-p}{2}\log\frac{2}{1-p} + \frac{p}{2}\log\frac{2}{p}\right) = \log 2 + h_{\mathrm{b}}(p), \tag{25}$$

$$H(P_{X|Y}|P_Y) = h_{\mathrm{b}}(p). \tag{26}$$

where $h_{\mathrm{b}}(p) = -p\log p - (1-p)\log(1-p)$ is the binary entropy function.

The joint source dispersion is

$$\mathrm{V}(P_{XY}) = 2\left(\frac{1-p}{2}\left(\log\frac{2}{1-p} - h_{\mathrm{b}}(p) - \log 2\right)^2 + \frac{p}{2}\left(\log\frac{2}{p} - h_{\mathrm{b}}(p) - \log 2\right)^2\right) \tag{27}$$

$$= (1-p)\left(-\log(1-p) - h_{\mathrm{b}}(p)\right)^2 + p\left(-\log p - h_{\mathrm{b}}(p)\right)^2, \tag{28}$$

and similarly we obtain the conditional source dispersion

$$\mathrm{V}(P_{X|Y}|P_Y) = (1-p)\left(-\log(1-p) - h_{\mathrm{b}}(p)\right)^2 + p\left(-\log p - h_{\mathrm{b}}(p)\right)^2. \tag{29}$$

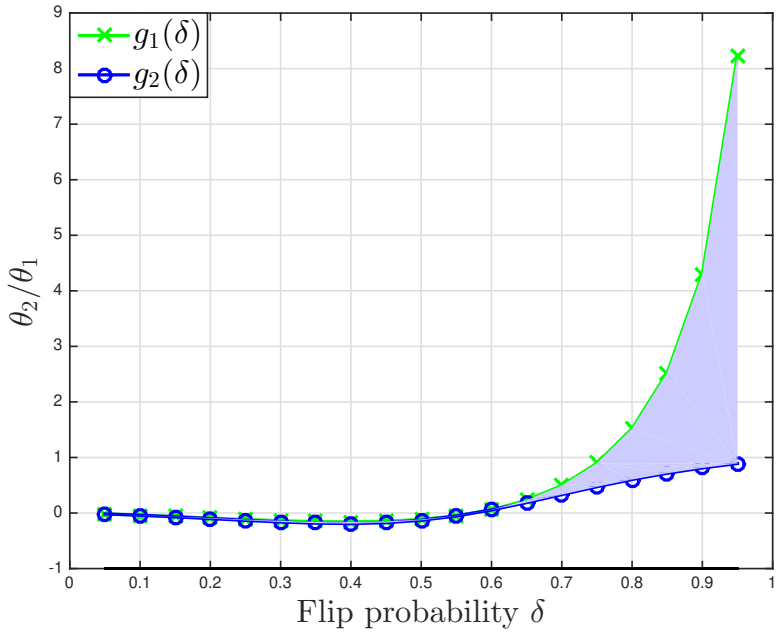


Fig. 3. Bounds on θ_2/θ_1 for the correlated binary source generated by a Z-channel determined by δ. Proposition 3 says that the set of θ_2/θ_1 such that we obtain a multiplicative gain of T in the moderate deviations constant are precisely contained in those areas *excluding the shaded region*, i.e., θ_2/θ_1 above the green line ($\theta_2/\theta_1 \geq g_1(\delta)$) or between the red line and the blue line ($-1 < \theta_2/\theta_1 \leq g_2(\delta)$). As the flip probability δ increases towards one (very noisy Z-channel), the gap between the green and blue lines increases.

Hence, $\mathrm{V}(P_{XY}) = \mathrm{V}(P_{X|Y}|P_Y)$ for this class of sources and we expect to achieve a gain of T.

Indeed, using the definitions of g_1 in (18) and g_2 in (19), we obtain

$$g_1(P_{XY}) = g_2(P_{XY}) = 0. \tag{30}$$

Now invoking Proposition 3 and (6), we conclude that for any DSBS, for all $\boldsymbol{\theta} \in \Theta_{\text{(ii)}}$, we can achieve a multiplicative gain of T in the moderate deviations constant.

B. Correlated Binary Source Generated via a Z-channel

Let $\mathcal{X} = \mathcal{Y} = \{0, 1\}$ and $X \sim \mathrm{Bern}(\frac{1}{2})$. The random variable Y is generated by transmitting X through a Z-channel with flip probability $\delta \in (0, 1)$ [39, Figure 1.7], i.e., $P_{Y|X}(0|0) = 1$ and $P_{Y|X}(0|1) = \delta$. For this source, it can be verified that (using the convention that $0 \log 0 = 0$)

$$H(P_{XY}) = \log 2 + \frac{1}{2} h_{\mathrm{b}}(\delta), \tag{31}$$

$$H(P_{X|Y}|P_Y) = \frac{1+\delta}{2} h_{\mathrm{b}}\left(\frac{1}{1+\delta}\right), \tag{32}$$

and

$$\mathrm{V}(P_{XY}) = \frac{1}{8} h_{\mathrm{b}}(\delta)^2 + \frac{\delta}{2}\left(-\log\delta - \frac{1}{2} h_{\mathrm{b}}(\delta)\right)^2 + \frac{1-\delta}{2}\left(-\log(1-\delta) - \frac{1}{2} h_{\mathrm{b}}(\delta)\right)^2, \tag{33}$$

$$\begin{aligned}\mathrm{V}(P_{X|Y}|P_Y) = {} & \frac{1}{2}\left(\log(1+\delta) - \frac{1+\delta}{2} h_{\mathrm{b}}\left(\frac{1}{1+\delta}\right)\right)^2 + \frac{\delta}{2}\left(\log\frac{1+\delta}{\delta} - \frac{1+\delta}{2} h_{\mathrm{b}}\left(\frac{1}{1+\delta}\right)\right)^2 \\ & + \frac{(1+\delta)(1-\delta^2)}{8}\left(h_{\mathrm{b}}\left(\frac{1}{1+\delta}\right)\right)^2.\end{aligned} \tag{34}$$

Note that $\mathrm{V}(P_{XY})$ and $\mathrm{V}(P_{X|Y}|P_Y)$ are functions of δ. Hence, we use $g_1(\delta)$ and $g_2(\delta)$ in place of (18) and (19) for this correlated source.

Invoking Proposition 3, we know that for any $\boldsymbol{\theta} \in \Theta_{\text{(ii)}}$ (recall (6)) such that $\frac{\theta_2}{\theta_1} \geq g_1(\delta)$ or $-1 < \frac{\theta_2}{\theta_1} \leq g_2(\delta)$, we obtain a gain of T in the moderate deviations constant compared with Case (ii) of the non-streaming setting in Theorem 1. In Figure 3, we plot the two functions $g_1(\delta)$ and $g_2(\delta)$ versus the flip probability δ. From Figure 3, we observe that when $\delta = 0.6$, we obtain a multiplicative gain of T in the moderate deviations constant except when $\boldsymbol{\theta} \in \Theta_{\text{(ii)}}$ is such that $0.056 < \frac{\theta_2}{\theta_1} < 0.058$ (which is a small interval).

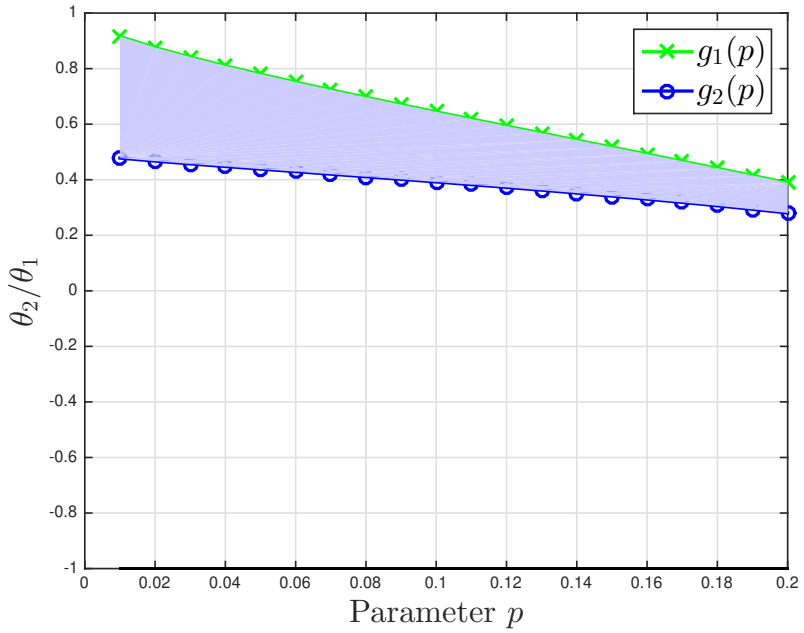


Fig. 4. Bounds on θ_2/θ_1 for the asymmetric correlated binary source determined by p. See the caption of Figure 3 for the set of θ_2/θ_1 such that we obtain a gain of T.

C. Asymmetric Correlated Binary Source

In this subsection, we consider an asymmetric correlated binary sources where $\mathcal{X} = \mathcal{Y} = \{0,1\}$, $P_{XY}(0,0) = 1-3p$ and $P_{XY}(0,1) = P_{XY}(1,0) = P_{XY}(1,1) = p$ for some $p \in \left(0, \frac{1}{4}\right) \bigcup \left(\frac{1}{4}, \frac{1}{3}\right)$. We do not allow p to be 0 (constant/degenerate distribution) or $1/4$ (uniform distribution), otherwise at least one of the joint and conditional source dispersions (varentropies) is zero, which is not permitted.

The joint and conditional entropies are

$$H(P_{XY}) = -(1-3p)\log(1-3p) - 3p\log p, \tag{35}$$

$$H(P_{X|Y}|P_Y) = (1-2p)h_{\mathrm{b}}\left(\frac{1-3p}{1-2p}\right) + 2p\log 2. \tag{36}$$

The joint and conditional dispersions are

$$\mathrm{V}(P_{XY}) = (1-3p)\left(-\log(1-3p) - H(P_{XY})\right)^2 + 3p\left(-\log p - H(P_{XY})\right)^2, \tag{37}$$

$$\begin{aligned}\mathrm{V}(P_{X|Y}|P_Y) = (1-3p)&\left(\log\frac{1-2p}{1-3p} - H(P_{X|Y}|P_Y)\right)^2 + p\left(\log\frac{1-2p}{p} - H(P_{X|Y}|P_Y)\right)^2 \\ &+ 2p\left(\log 2 - H(P_{X|Y}|P_Y)\right)^2\end{aligned} \tag{38}$$

Define $g_1(p), g_2(p)$ similarly as (18) and (19). We plot these two functions $g_1(p)$ and $g_2(p)$ in Figure 4. Note that when $p = 0.1$, for all $\boldsymbol{\theta} \in \Theta_{(\mathrm{ii})}$ (recall (6)) such that $\frac{\theta_2}{\theta_1} \in (-1, 0.40] \cup [0.67, \infty)$, we can achieve a multiplicative gain of T in the moderate deviations constant. When $\frac{\theta_2}{\theta_1} \in (0.40, 0.67)$ we cannot guarantee that we can achieve this gain using the proposed coding scheme.

IV. Some Specializations and Generalizations

In this section, we specialize Theorem 2 to point-to-point lossless source coding and we also generalize our result to the case where the delay requirements on both sources are different.

A. Specialization to Streaming Lossless Source Coding

In this subsection, we specialize the result in Theorem 2 to the one encoder and one decoder case, which we term as *streaming lossless source coding*. Let the optimal moderate deviations constant for this case be defined as $\nu^*(R_X^*, T)$, which is defined similarly as Definition 2 by letting $\theta_1 = 1$ and ignoring the requirement on the rate of the second codeword.

The source dispersion [11], [35] is defined as

$$\mathrm{V}(P_X) := \sum_x P_X(x) \left(-\log P_X(x) - H(P_X)\right)^2. \tag{39}$$

It is also known that the moderate deviations constant for non-streaming lossless source coding is $1/(2\mathrm{V}(P_X))$ [15].

By specializing Theorem 2, we obtain the following result.

Corollary 4. *The moderate deviations constant for streaming lossless source coding satisfies*

$$\nu^*(H(P_X), T) \geq \frac{T}{2\mathrm{V}(P_X)}. \tag{40}$$

Thus when there is only one source, we obtain a multiplicative gain of T in the moderate deviations constant. We remark that Corollary 4 can be proved similarly as [16, Theorem 1] using the information spectrum method [37, Lemma 7.2.1] and the moderate deviations theorem [38, Theorem 3.7.1]. The proof requires that the decoder has knowledge of the source distribution and thus it is *non-universal*. However, the benefit is that we only require the usual condition on $\{\xi_n\}_{n\in\mathbb{N}}$ for the moderate deviations regime [13] (i.e., $n\xi_n^2 \to \infty$ as $n \to \infty$) instead of the more stringent condition that $\frac{\log n}{n\xi_n^2} \to 0$ as $n \to \infty$ in Definition 2.

B. Generalization to Different Delay Requirements

In this subsection, we consider a generalization of Theorem 2. Recall that in Theorem 2, we require that both source blocks are reconstructed with the same delay T. Now assume that we incur delay T_1 for source $\mathbf{X}$ and T_2 for source $\mathbf{Y}$. We are now interested in characterizing the streaming performance in the moderate deviations regime. The formal definition of this setting up is as follows.

Definition 3 ((T_1, T_2)-Streaming Code)**.** *An $(n, N_1, N_2, T_1, T_2, \epsilon_n)$-code for streaming SW coding consists of*

1) *A sequence of correlated source blocks $\{(\mathbf{X}_k, \mathbf{Y}_k)\}_{k\geq 1}$, where $(\mathbf{X}_k, \mathbf{Y}_k) \in \mathcal{X}^n \times \mathcal{Y}^n$ is an i.i.d. sequence;*
2) *A sequence of encoding function pairs $f_k : \mathcal{X}^{kn} \to \mathcal{M}_1$ that maps the accumulated source blocks $\mathbf{X}^k \in \mathcal{X}^{kn}$ to a codeword $M_{1,k} \in \mathcal{M}_1$ and $g_k : \mathcal{Y}^{kn} \to \mathcal{M}_2$ that maps the accumulated source blocks $\mathbf{y}^k \in \mathcal{Y}^{kn}$ to a codeword $M_{2,k} \in \mathcal{M}_2$ where $|\mathcal{M}_i| = N_i,\ i = 1, 2$.*
3) *A sequence of decoding functions $\phi_k : \mathcal{M}_1^{k+T_1-1} \times \mathcal{M}_2^{k+T_1-1} \to \mathcal{X}^n \times \mathcal{Y}^n$ that maps the accumulated codewords $(M_1^{k+T_1-1}, M_2^{k+T_1-1}) \in \mathcal{M}_1^{k+T_1-1} \times \in \mathcal{M}_2^{k+T-1}$ to a source block pair $(\hat{\mathbf{X}}_{k_1}, \hat{\mathbf{Y}}_{k_2})$,*

which satisfies that for each k satisfying $\max\{k_1, k_2\} \geq 1$,

$$\Pr\left(\left(\hat{\mathbf{X}}_{k_1}, \hat{\mathbf{Y}}_{k_2}\right) \neq (\mathbf{X}_{k_1}, \mathbf{Y}_{k_2})\right) \leq \epsilon_n, \tag{41}$$

where $k_1 = k$ and $k_2 = k + T_1 - T_2$.

When $k_1 > 0$, $k_2 \leq 0$, (41) reduces to

$$\Pr\left(\hat{\mathbf{X}}_{k_1} \neq \mathbf{X}_{k_1}\right) \leq \epsilon_n, \tag{42}$$

and similarly when $k_1 \leq 0, k_2 > 0$, (41) reduces to

$$\Pr\left(\hat{\mathbf{Y}}_{k_2} \neq \mathbf{Y}_{k_2}\right) \leq \epsilon_n. \tag{43}$$

The optimal moderate deviations constant for this setting is defined similarly to that in Definition 2 and we denote it as $\nu^*(R_X^*, R_Y^*, \boldsymbol{\theta}, T_1, T_2)$.

Theorem 5. *The moderate deviations constant for streaming compression of correlated sources with two decoding delay requirements satisfies the same lower bounds as in Theorem 2 with T replaced by $\min\{T_1, T_2\}$.*

The proof of Theorem 5 is provided in Section VI.

V. Proof of Theorem 2

In this section, we present the proof of the lower bound on the moderate deviations constant for streaming SW coding. We combine the minimum weighted empirical suffix-entropy decoding in [7] and the truncated memory idea in [16] in the proof. The analysis is, however, rather *different* from both works.

Let the smallest and largest memory at the encoders in terms of blocks being Ω and Ψ respectively. We assume that $\Omega \geq T$ and $\Psi > 2\Omega$ and in the general decoding framework, we will ensure these two conditions hold in the sequel. Note that Ψ is used to control the truncated memory and Ω is set to make sure that every source block is encoded into at least Ω codewords. We will make these points clear later.

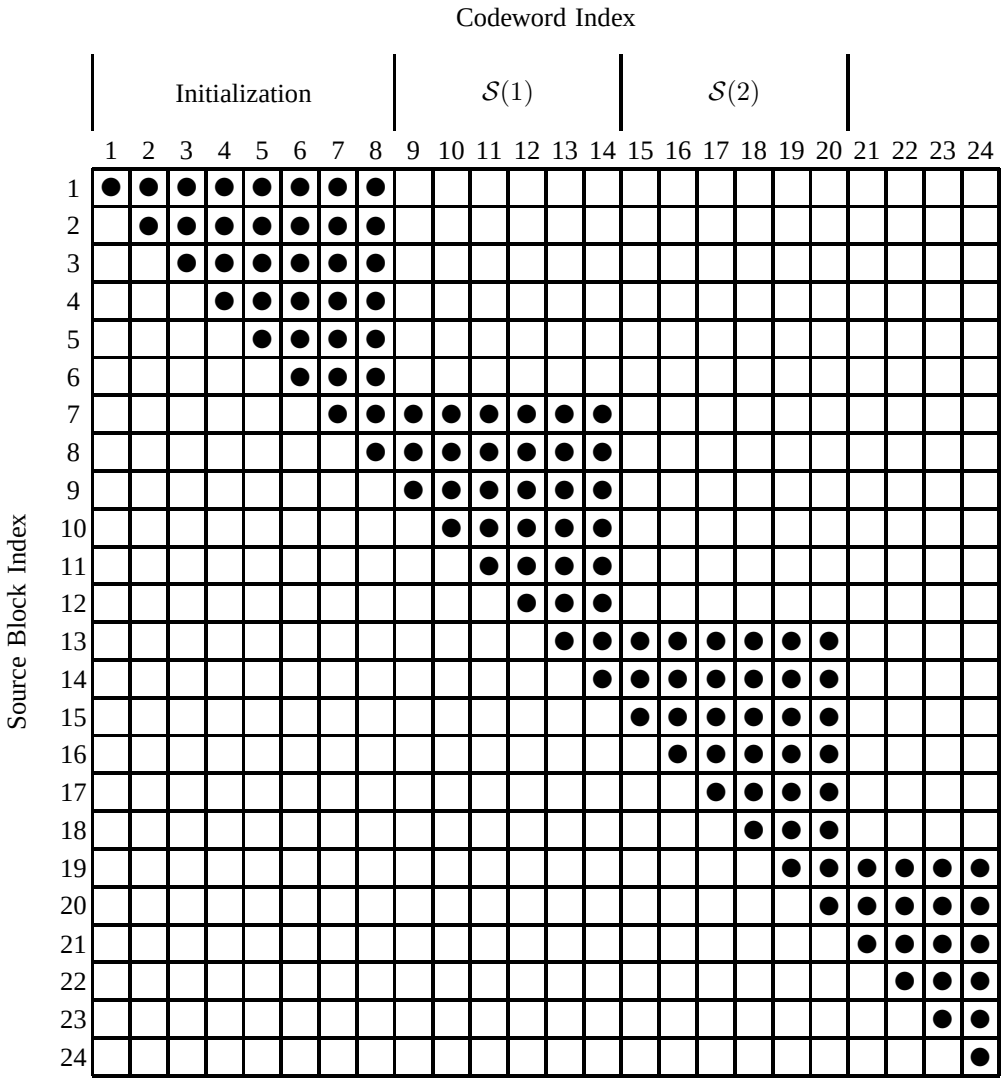


Fig. 5. Encoding strategy for encoder f_k for the case where $\Omega = 3$ and $\Psi = 8$. For the first $\Psi = 8$ timeslots, we encode all accumulated source blocks into codewords $M_{1,k}$. For $k \in \mathcal{S}(1)$, we encode source blocks with indices from $t_1 = 7$ to k into codeword $M_{1,k}$. For $k \in \mathcal{S}(2)$, we encode source blocks with indices from $t_2 = 13$ to k into codeword $M_{1,k}$. This figure is a reproduction of [16, Figure 3] with a slight change in parameters.

A. Codebook Generation and Encoding: An Example

The codebook generation and encoding procedure are similar to [16] for the central limit regime. To illustrate the codebook generation and encoding, we first consider the case of $\Omega = 3$ and $\Psi = 8$ which is shown in Figure 5. Here we describe the encoding procedure for $\mathbf{X}$ sequence only since for $\mathbf{Y}$ sequence, the codebook generation and encoding are similar. Throughout the section, we assume $q \in \mathbb{N}$.

Codebook Generation. In the streaming setting, for each time $k \in \mathbb{N}$, we generate a different codebook $\mathcal{C}_k$. Hence, the codebook generation depends on the source block index. In our scheme here, the codebook generation is divided into two phases. The first phase is called the *Initialization Phase* of the codebook generation in which we randomly and independently generate a codeword $M_{1,k}$ for each $\mathbf{x}^k \in \mathcal{X}^{kn}$ according to a uniform distribution over $\mathcal{M}_1$. The assignments of sequences to codeword $M_{1,k}$ constitute the codebook $\mathcal{C}_k$. This part holds for $k \in [1 : \Psi = 8]$. The second phase is called the *Periodic Phase* of the codebook generation in which we randomly and independently generate a codeword $M_{1,k}$ for a chosen subsequence of each $\mathbf{x}^k \in \mathcal{X}^{kn}$.

Before describing the details of codebook generation for the periodic phase, we define the following important quantities. Let

$$\mathcal{S}(q) = \big[(\Psi - \Omega + 1)q + \Omega : (\Psi - \Omega + 1)(q + 1) + \Omega - 1\big]. \tag{44}$$

Let α_q and β_q be the first and last elements in $\mathcal{S}(q)$, i.e.,

$$\alpha_q := (\Psi - \Omega + 1)q + \Omega, \tag{45}$$
$$\beta_q := (\Psi - \Omega + 1)(q+1) + \Omega - 1. \tag{46}$$

Define

$$t_q := \alpha_q - (\Omega - 1) = (\Psi - \Omega + 1)q + 1. \tag{47}$$

Note that for this example, $\Psi - \Omega + 1 = 6$. Hence,

$$\alpha_q = 6q + 3, \tag{48}$$
$$\beta_q = 6(q+1) + 2, \tag{49}$$
$$t_q = 6q + 1. \tag{50}$$

Now from the columns of Figure 5, we observe that for $k \in [\alpha_1 = 9 : \beta_1 = 14] = \mathcal{S}(1)$, we randomly and independently generate a codeword $M_{1,k}$ for each $\mathbf{x}_{t_1=7}^k \in \mathcal{X}^{n(k-t_1+1)}$ according to a uniform distribution over $\mathcal{M}_1$. In general, for $k \in \mathcal{S}(q)$, we randomly and independently generate a codeword $M_{1,k}$ for each $\mathbf{x}_{t_q}^k \in \mathcal{X}^{n(k-t_q+1)}$ according to a uniform distribution over $\mathcal{M}_1$.

Encoding We are now ready to describe the encoding procedure. For $k \in [1 : \Psi = 8]$, given $\mathbf{x}^k$, we send $f_k(\mathbf{x}^k) = M_{1,k}(\mathbf{x}^k)$. For $k \in [\alpha_1 = 9 : \beta_1 = 14] = \mathcal{S}(1)$, given $\mathbf{x}^k$, we send $f_k(\mathbf{x}^k) = M_{1,k}(\mathbf{x}_{t_1=7}^k)$. In general, for $k \in \mathcal{S}(q)$ where $q \geq 2$, given $\mathbf{x}^k$, we send $f_k(\mathbf{x}^k) = M_{1,k}\left(\mathbf{x}_{t_q}^k\right)$.

After describing the encoding strategy, let us now examine the rows of Figure 5. We observe $\mathbf{X}_j$ where $j \in [1 : t_1 - 1 = 6]$ is mapped to $M_{1,j}, \ldots, M_{1,\Psi=8}$; $\mathbf{X}_j$ where $j \in [t_1 = 7 : t_2 - 1 = 12]$ is mapped to $M_{1,j}, \ldots, M_{1,\beta_1=14}$. Note that this mapping from $j = t_1$ is periodic. In summary, $\mathbf{X}_j$ where $j \in [t_q : t_{q+1} - 1]$ is mapped to $M_{1,j}, \ldots, M_{1,\beta_q}$.

B. Codebook Generation and Encoding: General Case

We now describe the encoding for arbitrary values of Ψ and Ω satisfying the conditions that $\Omega \geq T$, $\Psi > 2\Omega$.

Codebook Generation:

Initialization Phase: for $k \in [1 : \Psi]$, we randomly and independently generate a codeword $M_{1,k}$ for each $\mathbf{X}^k \in \mathcal{X}^{nk}$ according to a uniform distribution over $\mathcal{M}_1$.

Periodic Phase: For each $k \in \mathcal{S}(q)$, we randomly and independently generate a codeword $M_{1,k}$ for each $\mathbf{x}_{t_q}^k \in \mathcal{X}^{n(k-t_q+1)}$ according to a uniform distribution over $\mathcal{M}_1$.

Encoding: For $k \in [1 : \Psi]$, given $\mathbf{x}^k$, we send $f_k(\mathbf{x}^k) = M_{1,k}(\mathbf{x}^k)$. For $k \in \mathcal{S}(q)$, given $\mathbf{x}^k$, we send $f_k(\mathbf{x}^k) = M_{1,k}(\mathbf{x}_{t_q}^k)$. For $k \in \mathcal{S}(q)$, since the output of f_k depends only on $\mathbf{x}_{t_q}^k$, we may define a function $\tilde{f}_k : \mathcal{X}^{n(k-t_q+1)} \to \mathcal{M}_1$ satisfying $\tilde{f}_k(\mathbf{x}_{t_q}^k) = f_k(\mathbf{x}^k)$. Similarly, we define $\tilde{g}_k(\mathbf{y}_{t_q}^k) = M_{2,k}(\mathbf{y}_{t_q}^k)$ for $k \in \mathcal{S}(q)$.

Throughout this section, we assume that a source block pair $\{(\mathbf{X}_k, \mathbf{Y}_k)\}_{k\in\mathbb{N}}$ is generated by the correlated source and the corresponding codewords produced by the encoders are $\{(M_{1,k}, M_{2,k})\}_{k\in\mathbb{N}}$.

Paralleling the observation in the end of Section V-A, using the above encoding rule, we conclude that

1) Source block $\mathbf{X}_j$ where $j \in [1 : t_1 - 1]$ is mapped to codeword $M_{1,j}, \ldots, M_{1,\Psi}$;
2) Source block $\mathbf{X}_j$ where $j \in [t_q : t_{q+1} - 1]$ is mapped to codeword $M_{1,j}, \ldots, M_{1,\beta_q}$.

C. Decoding: An Example

In this subsection, we wish to decode $(\mathbf{X}_k, \mathbf{Y}_k)$ at time $T_k = k + T - 1$. We use $(\hat{\mathbf{X}}_j, \hat{\mathbf{Y}}_j)$ to denote the estimate of $(\mathbf{X}_j, \mathbf{Y}_j)$ at time T_k. Throughout the description of decoding, we suppress the dependency of the estimates on T_k for simplicity.

We remark that our decoding strategy is similar to [16] for the central limit regime of streaming transmission (channel coding). Before going to the details of decoding technique for general case in Section V-D, we explain our decoding for the example where $T = 2$, $\Omega = 3$ and $\Psi = 8$. Note this choice satisfies $\Psi > 2\Omega$ and $\Omega \geq T$. Recall the quantities α_q, β_q and t_q in (48), (49) and (50) respectively. Note that $T_k = k + T - 1 = k + 1$.

Our decoding is done in correspondence with the encoding strategy. Depending on the index of the source block pair we aim to decode, we classify the decoding into three phases: *Initialization Phase I*, *Initialization Phase II* and *Periodic Decoding Phase*. For each phase, given the index k and delay T, our decoding strategy is described in detail as follows.

1) $k \in [1 : \Psi = 8]$ (Initialization Phase I)
   In this case, if we wish to decode $(\mathbf{X}_k, \mathbf{Y}_k)$, we decode all source blocks $(\mathbf{X}_1, \mathbf{Y}_1), \ldots, (\mathbf{X}_k, \mathbf{Y}_k)$ sequentially.
   For example, if we wish to decode $(\mathbf{X}_3, \mathbf{Y}_3)$ after receiving $T_3 = 4$-th codeword pair, we decode $(\mathbf{X}_1, \mathbf{Y}_1,) \ldots, (\mathbf{X}_3, \mathbf{Y}_3)$ sequentially. In the decoding of each source block $(\mathbf{X}_j, \mathbf{Y}_j)$ where $j \in [1 : 3]$, we use codewords with indices from j to $T_3 = 4$. To be specific, given $(\hat{\mathbf{X}}^{j-1}, \hat{\mathbf{Y}}^{j-1})$, we decode $(\mathbf{X}_j, \mathbf{Y}_j) = (\tilde{\mathbf{x}}_j, \tilde{\mathbf{y}}_j)$ if the following conditions are satisfied:

a) Codewords involving $(\tilde{\mathbf{x}}_j, \tilde{\mathbf{y}}_j)$ are same as received, i.e., there exists $(\tilde{\mathbf{x}}^{T_3=4}, \tilde{\mathbf{y}}^{T_3=4})$ such that $f_\tau(\tilde{\mathbf{x}}^\tau) = M_{1,\tau} = f_\tau(\mathbf{X}^\tau)$ and $g_\tau(\tilde{\mathbf{y}}^\tau) = M_{2,\tau} = g_l(\mathbf{Y}^\tau)$ for all $\tau \in [j:4]$;
b) The first j source blocks of $(\tilde{\mathbf{x}}^{T_3=4}, \tilde{\mathbf{y}}^{T_3=4})$ coincide with previous estimates, i.e., $(\tilde{\mathbf{x}}^{j-1}, \tilde{\mathbf{y}}^{j-1}) = (\hat{\mathbf{X}}^{j-1}, \hat{\mathbf{Y}}^{j-1})$;
c) The chosen block $(\tilde{\mathbf{x}}^{T_3=4}, \tilde{\mathbf{y}}^{T_3=4})$ has the highest score using a score function (to be defined later) among all blocks satisfying conditions 1a) and 1b).

The detailed description of decoding for the other cases in this subsection are omitted since they are similar to the above. In Section V-D, we will explain the decoding in detail.

We can also consider another example where the decoding is more complicated. If we wish to decode $(\mathbf{X}_8, \mathbf{Y}_8)$, then we decode $(\mathbf{X}_1, \mathbf{X}_2), \ldots, (\mathbf{X}_8, \mathbf{Y}_8)$ after receiving $T_8 = 9$-th codeword pair. When we decode $(\mathbf{X}_j, \mathbf{Y}_j)$ where $j \in [1 : t_1 - 1 = 6]$, we use codewords with indices from j to $\Psi = 8$. However, when we decode $(\mathbf{X}_j, \mathbf{Y}_j)$ where $j \in [t_1 = 7 : \Psi = 8]$, we use codewords with indices from j to $T_8 = 9$. This is due to the encoding procedure where we encode $(\mathbf{X}_j, \mathbf{Y}_j)$ for $j \in [1 : t_1 - 1 = 6]$ into codewords with indices from j to $\Psi = 8$ only (recall the end of Section V-A). Hence, codewords with indices greater that 8 are redundant if we decode $(\mathbf{X}_j, \mathbf{Y}_j)$ where $j \leq 6$.

2) $k \in [\alpha_1 = 9 : \beta_1 = 14] = \mathcal{S}(1)$ (Initialization Phase II)
In this case, if we wish to decode $(\mathbf{X}_k, \mathbf{Y}_k)$, we decode all source blocks $(\mathbf{X}_1, \mathbf{Y}_1), \ldots, (\mathbf{X}_k, \mathbf{Y}_k)$ sequentially. The difference in decoding from the case where $k \in [1 : \Psi = 8]$ lies in the use of codewords we received.
For example, if we wish to decode $(\mathbf{X}_{10}, \mathbf{Y}_{10})$, we decode all $(\mathbf{X}_1, \mathbf{Y}_1), \ldots, (\mathbf{X}_{10}, \mathbf{Y}_{10})$ sequentially. When we decode $(\mathbf{X}_j, \mathbf{Y}_j)$ where $j \in [1 : t_1 - 1 = 6]$, we use codewords with indices from j to $\Psi = 8$. When we decode $(\mathbf{X}_j, \mathbf{Y}_j)$ where $j \in [t_1 = 7 : 10]$, we use codewords with indices from j to $T_{10} = 11$.
3) $k \in [\alpha_2 = 15 : \beta_2 = 20] = \mathcal{S}(2)$ (Periodic Decoding Phase)
In this case, we decode all source blocks involved into codewords with indices no greater that k and falling in $\mathcal{S}(1)$ or $\mathcal{S}(2)$, i.e, $t_1 = 7$ to k. We consider the example where $T_k = k + T - 1 < 20$. Hence, $k = T_k - T + 1 < 19 = t_3$. To be specific, we first decode $(\mathbf{X}_{t_1=7}, \mathbf{Y}_{t_1=7}), \ldots, (\mathbf{X}_{\alpha_1=9}, \mathbf{Y}_{\alpha_1=9})$ jointly using codewords with indices from $\alpha_1 = 9$ to $\beta_1 = 14$. Next we decode $(\mathbf{X}_j, \mathbf{Y}_j)$ where $j \in [\alpha_1 + 1 = 10 : t_2 - 1 = 12]$ sequentially using codewords with indices from j to $\beta_1 = 14$. Then we decode $(\mathbf{X}_j, \mathbf{Y}_j)$ where $j \in [t_2 = 13 : \beta_1 = 14]$ sequentially using codewords with indices from j to T_k. Finally, we decode $(\mathbf{X}_j, \mathbf{Y}_j)$ where $j \in [\alpha_2 = 15 : k]$ sequentially using codewords with indices from j to T_k. The decoding for last part ($j \in [\alpha_2 = 15 : k]$) is different if $T_k > 20$. Let us consider decoding $(\mathbf{X}_{20}, \mathbf{Y}_{20})$ to illustrate this. For $k = 20$, $T_k = 20 + 1 = 21 > 20$. For this case, we split the decoding of $(\mathbf{X}_j, \mathbf{Y}_j)$ where $j \in [\alpha_2 = 15 : k]$ into two parts. First, we decode $(\mathbf{X}_j, \mathbf{Y}_j)$ where $j \in [\alpha_2 = 15 : t_3 - 1 = 18]$ sequentially using codewords with indices from j to $\beta_2 = 20$ and decode $(\mathbf{X}_j, \mathbf{Y}_j)$ where $j \in [t_3 = 19 : k = 20]$ sequentially using codewords with indices from j to T_k.

We remark that the decoding described for $k \in \mathcal{S}(2)$ is periodic, i.e, it is the same for $k \in \mathcal{S}(q)$ for $q \geq 3$. We will emphasize this point in Section V-D when we describe the decoding for the general case.

D. Decoding: General Case

Before we describe the score function and the decoding procedure, let us introduce some additional notation. For a pair of sequences $(\mathbf{x}, \mathbf{y}) \in \mathcal{X}^r \times \mathcal{Y}^r$ (for some $r \in \mathbb{N}$), we let $\hat{H}(\mathbf{x}, \mathbf{y})$ be the joint empirical entropy, i.e., the entropy of the joint type $\hat{T}_{\mathbf{x},\mathbf{y}}$. Similarly, we let $\hat{H}(\mathbf{x}|\mathbf{y}) = H(\hat{T}_{\mathbf{x}|\mathbf{y}}|\hat{T}_{\mathbf{y}})$ be the conditional empirical entropy. Given two pairs of sequences $(\tilde{\mathbf{x}}_a^b, \tilde{\mathbf{y}}_a^b)$ and $(\bar{\mathbf{x}}_a^b, \bar{\mathbf{y}}_a^b)$, let l be the index of the block where $\tilde{\mathbf{x}}^{T_k}$ and $\bar{\mathbf{x}}^{T_k}$ first differ and similarly let m be the index of the block where $\tilde{\mathbf{y}}^{T_k}$ and $\bar{\mathbf{y}}^{T_k}$ first differ.

We make use of the weighted empirical suffix entropy in [7, Eqn. (54)], i.e., for $l, m \in [a : b]$,

$$\hat{H}_{\mathrm{S}}(l, m, \tilde{\mathbf{x}}_a^b, \tilde{\mathbf{y}}_a^b) = \begin{cases} \hat{H}(\tilde{\mathbf{x}}_l^b, \tilde{\mathbf{y}}_l^b) & l = m \\ \frac{m-l}{b-l+1}\hat{H}(\tilde{\mathbf{x}}_l^{m-1}|\tilde{\mathbf{y}}_l^{m-1}) + \frac{b-m+1}{b-l+1}\hat{H}(\tilde{\mathbf{x}}_m^b, \tilde{\mathbf{y}}_m^b) & l < m \\ \frac{l-m}{b-l+1}\hat{H}(\tilde{\mathbf{x}}_m^{l-1}|\tilde{\mathbf{y}}_m^{l-1}) + \frac{b-l+1}{b-m+1}\hat{H}(\tilde{\mathbf{x}}_l^b, \tilde{\mathbf{y}}_l^b) & l > m \end{cases} \tag{51}$$

The reciprocal of the suffix empirical entropy will serve as the score function.

As explained in the example in Section V-C, our decoding technique differs according to the index k of the source block pair we aim to decode. In summary, if $k \in [1 : \Psi] \bigcup \mathcal{S}(1)$, we decode $(\mathbf{X}_1, \mathbf{Y}_1), \ldots (\mathbf{X}_k, \mathbf{Y}_k)$ sequentially. If $k \in \mathcal{S}(q)$ where $q \geq 2$, we first decode $(\mathbf{X}_{t_{q-1}}^{\alpha_{q-1}}, \mathbf{Y}_{t_{q-1}}^{\alpha_{q-1}})$ jointly, then decode $(\mathbf{X}_{\alpha_{q-1}+1}, \mathbf{Y}_{\alpha_{q-1}+1}), \ldots, (\mathbf{X}_k, \mathbf{Y}_k)$ sequentially. There are three phases of decoding depending on k: Initialization Phase I ($k \in [1 : \Psi]$), Initialization Phase II ($k \in \mathcal{S}(1)$) and Periodic Decoding Phase ($k \in \mathcal{S}(q)$ where $q \geq 2$). We will describe the decoding for each phase in detail.

We make heavy use of the definitions of α_q (see (45)), β_q (see (46)) and t_q (see (47)) to simplify the description of decoding.

1) $k \in [1 : \Psi]$ (Initialization Phase I)
a) $T_k \leq \Psi$

Define the set

$$\mathcal{A}_1(\mathbf{X}^{T_k},\mathbf{Y}^{T_k},j):=\left\{(\tilde{\mathbf{x}}^{T_k},\tilde{\mathbf{y}}^{T_k}):\forall\ \tau\in[j:T_k],f_\tau(\tilde{\mathbf{x}}^\tau)=f_\tau(\mathbf{X}^\tau),\ g_\tau(\tilde{\mathbf{y}}^\tau)=g_\tau(\mathbf{Y}^\tau)\right\}. \tag{52}$$

After receiving T_k-th codeword pair, we decode $(\mathbf{X}_1,\mathbf{Y}_1),\ldots,(\mathbf{X}_k,\mathbf{Y}_k)$ sequentially as follows. Given $(\hat{\mathbf{X}}^{j-1},\hat{\mathbf{Y}}^{j-1})$, we decode $(\hat{\mathbf{X}}_j,\hat{\mathbf{Y}}_j)=(\tilde{\mathbf{x}}_j,\tilde{\mathbf{y}}_j)$ if there exists $(\tilde{\mathbf{x}}^{T_k},\tilde{\mathbf{y}}^{T_k})\in\mathcal{A}_1(\mathbf{X}^{T_k},\mathbf{Y}^{T_k},j)$ satisfying $\tilde{\mathbf{x}}^{j-1}=\hat{\mathbf{X}}^{j-1},\tilde{\mathbf{y}}^{j-1}=\hat{\mathbf{Y}}^{j-1}$ such that for all $(\bar{\mathbf{x}}^{T_k},\bar{\mathbf{y}}^{T_k})\in\mathcal{A}_1(\mathbf{X}^{T_k},\mathbf{Y}^{T_k},j)$ satisfying $\bar{\mathbf{x}}^{j-1}=\hat{\mathbf{X}}^{j-1},\bar{\mathbf{y}}^{j-1}=\hat{\mathbf{Y}}^{j-1}$, we have

$$\hat{H}_{\mathrm{S}}(l,m,\tilde{\mathbf{x}}^{T_k},\tilde{\mathbf{y}}^{T_k})\le\hat{H}_{\mathrm{S}}(l,m,\bar{\mathbf{x}}^{T_k},\bar{\mathbf{y}}^{T_k}), \tag{53}$$

Note that l and m in (53) are uniquely determined by $(\mathbf{x}^{T_k},\mathbf{y}^{T_k})$ and $(\tilde{\mathbf{x}}^{T_k},\tilde{\mathbf{y}}^{T_k})$ according to the rule as described prior to (51). This comment applies verbatim whenever we employ minimum empirical suffix entropy decoding in the sequel.

b) $T_k>\Psi$

The decoding for this scenario is split into two parts depending on the value of j.

i) $1\le j\le\alpha_1-\Omega=t_1-1$

The decoding is exactly the same as the case where $T_k\le\Psi$ except that we replace T_k with Ψ.

ii) $t_1\le j\le k$

Define the set

$$\begin{aligned}\mathcal{A}_2(\mathbf{X}^{T_k},\mathbf{Y}^{T_k},j):=\Big\{(\tilde{\mathbf{x}}^{T_k},\tilde{\mathbf{y}}^{T_k}),&\forall\ \tau\in[j:\Psi]:f_\tau(\tilde{\mathbf{x}}^\tau)=f_\tau(\mathbf{X}^\tau),g_\tau(\tilde{\mathbf{y}}^\tau)=g_l(\mathbf{Y}^\tau);\\&\forall\ \tau\in[\alpha_1:T_k],\tilde{f}_\tau(\tilde{\mathbf{x}}_{t_1}^\tau)=\tilde{f}_\tau(\mathbf{X}_{t_1}^\tau),\tilde{g}_\tau(\tilde{\mathbf{y}}_{t_1}^\tau)=\tilde{g}_\tau(\mathbf{Y}_{t_1}^\tau)\Big\}.\end{aligned} \tag{54}$$

Given $(\hat{\mathbf{X}}^{j-1},\hat{\mathbf{Y}}^{j-1})$, we decode $(\hat{\mathbf{X}}_j,\hat{\mathbf{Y}}_j)=(\tilde{\mathbf{x}}_j,\tilde{\mathbf{y}}_j)$ if there exists $(\tilde{\mathbf{x}}^{T_k},\tilde{\mathbf{y}}^{T_k})\in\mathcal{A}_2(\mathbf{X}^{T_k},\mathbf{Y}^{T_k},j)$ satisfying $\tilde{\mathbf{x}}^{j-1}=\hat{\mathbf{X}}^{j-1},\tilde{\mathbf{y}}^{j-1}=\hat{\mathbf{Y}}^{j-1}$ such that for all $(\bar{\mathbf{x}}^{T_k},\bar{\mathbf{y}}^{T_k})\in\mathcal{A}_2(\mathbf{X}^{T_k},\mathbf{Y}^{T_k},j)$ where $\bar{\mathbf{x}}^{j-1}=\hat{\mathbf{X}}^{j-1},\bar{\mathbf{y}}^{j-1}=\hat{\mathbf{Y}}^{j-1}$, we have

$$\hat{H}_{\mathrm{S}}(l,m,\tilde{\mathbf{x}}^{T_k},\tilde{\mathbf{y}}^{T_k})\le\hat{H}_{\mathrm{S}}(l,m,\bar{\mathbf{x}}^{T_k},\bar{\mathbf{y}}^{T_k}), \tag{55}$$

2) $k\in\mathcal{S}(1)$ (Initialization Phase II)

For $j\in[1:t_1-1]$, the decoding is exactly the same as $T_k\le\Psi$ except that we replace T_k with Ψ. We then proceed to decode $\mathbf{X}_j$ for $t_1\le j\le k$. The decoding is separated into two parts.

a) $T_k\le\beta_1$

The decoding of $(\mathbf{X}_j,\mathbf{Y}_j)$ for $j\in[t_1:\Psi]$ is exactly the same as the case $T_k>\Psi$. We then describe the decoding of $(\mathbf{X}_j,\mathbf{Y}_j)$ for $\alpha_1\le j\le k$.

Define

$$\mathcal{A}_3(\mathbf{X}^{T_k},\mathbf{Y}^{T_k},j):=\left\{(\tilde{\mathbf{x}}^{T_k},\tilde{\mathbf{y}}^{T_k}):\forall\tau\in[j:T_k],\tilde{f}_\tau\left(\tilde{\mathbf{x}}_{t_1}^\tau\right)=\tilde{f}_\tau\left(\mathbf{X}_{t_1}^\tau\right),\tilde{g}_\tau\left(\tilde{\mathbf{y}}_{t_1}^\tau\right)=\tilde{g}_\tau\left(\mathbf{Y}_{t_1}^\tau\right)\right\} \tag{56}$$

Given $(\hat{\mathbf{X}}_{t_1}^{j-1},\hat{\mathbf{Y}}_{t_1}^{j-1})$, we decode $(\hat{\mathbf{X}}_j,\hat{\mathbf{Y}}_j)=(\tilde{\mathbf{x}}_j,\tilde{\mathbf{y}}_j)$ if there exists $(\tilde{\mathbf{x}}^{T_k},\tilde{\mathbf{y}}^{T_k})\in\mathcal{A}_3(\mathbf{X}^{T_k},\mathbf{Y}^{T_k},j)$ satisfying $\tilde{\mathbf{x}}^{j-1}=\hat{\mathbf{X}}^{j-1},\tilde{\mathbf{y}}^{j-1}=\hat{\mathbf{Y}}^{j-1}$ such that for all $(\bar{\mathbf{x}}^{T_k},\bar{\mathbf{y}}^{T_k})\in\mathcal{A}_3(\mathbf{X}^{T_k},\mathbf{Y}^{T_k},j)$ where $\bar{\mathbf{x}}^{j-1}=\hat{\mathbf{X}}^{j-1},\bar{\mathbf{y}}^{j-1}=\hat{\mathbf{Y}}^{j-1}$, we have

$$\hat{H}_{\mathrm{S}}(l,m,\tilde{\mathbf{x}}^{T_k},\tilde{\mathbf{y}}^{T_k})\le\hat{H}_{\mathrm{S}}(l,m,\bar{\mathbf{x}}^{T_k},\bar{\mathbf{y}}^{T_k}), \tag{57}$$

b) $T_k>\beta_1$

The decoding of $(\mathbf{X}_j,\mathbf{Y}_j)$ for $j\in[t_1:\Psi]$ is exactly the same as the case $T_k>\Psi$ except that we need to replace T_k with β_1.

In order to decode $(\mathbf{X}_j,\mathbf{Y}_j)$ for $\alpha_1\le j\le k$, we describe the decoding depending the value of j.

i) $\alpha_1\le j\le t_2-1$

Define

$$\mathcal{A}_4(\mathbf{X}^{\beta_1},\mathbf{Y}^{\beta_1},j):=\left\{(\tilde{\mathbf{x}}^{\beta_1},\tilde{\mathbf{y}}^{\beta_1}):\forall\tau\in[j:\beta_1],\ \tilde{f}_\tau\left(\tilde{\mathbf{x}}_{t_1}^\tau\right)=\tilde{f}_\tau\left(\mathbf{X}_{t_1}^\tau\right),\ \tilde{g}_\tau\left(\tilde{\mathbf{y}}_{t_1}^\tau\right)=\tilde{g}_\tau\left(\mathbf{Y}_{t_1}^\tau\right)\right\}. \tag{58}$$

Given $(\hat{\mathbf{X}}_{t_1}^{j-1},\hat{\mathbf{Y}}_{t_1}^{j-1})$, we decode to $(\hat{\mathbf{X}}_j,\hat{\mathbf{Y}}_j)=(\tilde{\mathbf{x}}_j,\tilde{\mathbf{y}}_j)$ if there exists $(\tilde{\mathbf{x}}^{\beta_1},\tilde{\mathbf{y}}^{\beta_1})\in\mathcal{A}_4(\mathbf{X}^{\beta_1},\mathbf{Y}^{\beta_1},j)$ satisfying $\tilde{\mathbf{x}}^{j-1}=\hat{\mathbf{X}}^{j-1},\tilde{\mathbf{y}}^{j-1}=\hat{\mathbf{Y}}^{j-1}$ such that for all $(\bar{\mathbf{x}}^{\beta_1},\bar{\mathbf{y}}^{\beta_1})\in\mathcal{A}_4(\mathbf{X}^{\beta_1},\mathbf{Y}^{\beta_1},j)$ where $\bar{\mathbf{x}}^{j-1}=\hat{\mathbf{X}}^{j-1},\bar{\mathbf{y}}^{j-1}=\hat{\mathbf{Y}}^{j-1}$, we have

$$\hat{H}_{\mathrm{S}}(l,m,\tilde{\mathbf{x}}^{\beta_1},\tilde{\mathbf{y}}^{\beta_1})\le\hat{H}_{\mathrm{S}}(l,m,\bar{\mathbf{x}}^{\beta_1},\bar{\mathbf{y}}^{\beta_1}). \tag{59}$$

ii) $t_2\le j\le k$.

Define

$$\mathcal{A}_5(\mathbf{X}^{T_k},\mathbf{Y}^{T_k},j):=\Big\{(\tilde{\mathbf{x}}^{T_k},\tilde{\mathbf{y}}^{T_k}):\forall\tau\in[j:\beta_1],\tilde{f}_\tau\left(\tilde{\mathbf{x}}_{t_1}^\tau\right)=\tilde{f}_\tau\left(\mathbf{X}_{t_1}^\tau\right),\tilde{g}_\tau\left(\tilde{\mathbf{y}}_{t_1}^\tau\right)=\tilde{g}_\tau\left(\mathbf{Y}_{t_1}^\tau\right),\\ \forall\tau\in[\alpha_2:T_k],\ \tilde{f}_\tau\left(\tilde{\mathbf{x}}_{t_2}^\tau\right)=\tilde{f}_\tau\left(\mathbf{X}_{t_2}^\tau\right),\ \tilde{g}_\tau\left(\tilde{\mathbf{y}}_{t_2}^\tau\right)=\tilde{g}_\tau\left(\mathbf{Y}_{t_2}^\tau\right)\Big\}. \tag{60}$$

Given $(\hat{\mathbf{X}}_{t_1}^{j-1},\hat{\mathbf{Y}}_{t_1}^{j-1})$, we decode to $(\hat{\mathbf{X}}_j,\hat{\mathbf{Y}}_j)=(\tilde{\mathbf{x}}_j,\tilde{\mathbf{y}}_j)$ if there exists $(\tilde{\mathbf{x}}^{T_k},\tilde{\mathbf{y}}^{T_k})\in\mathcal{A}_5(\mathbf{X}^{T_k},\mathbf{Y}^{T_k},j)$ satisfying $\tilde{\mathbf{x}}^{j-1}=\hat{\mathbf{X}}^{j-1},\tilde{\mathbf{y}}^{j-1}=\hat{\mathbf{Y}}^{j-1}$ such that for all $(\bar{\mathbf{x}}^{T_k},\bar{\mathbf{y}}^{T_k})\in\mathcal{A}_5(\mathbf{X}^{T_k},\mathbf{Y}^{T_k},j)$ where $\bar{\mathbf{x}}^{j-1}=\hat{\mathbf{X}}^{j-1},\bar{\mathbf{y}}^{j-1}=\hat{\mathbf{Y}}^{j-1}$, we have

$$\hat{H}_{\mathrm{S}}(l,m,\tilde{\mathbf{x}}^{T_k},\tilde{\mathbf{y}}^{T_k})\le\hat{H}_{\mathrm{S}}(l,m,\bar{\mathbf{x}}^{T_k},\bar{\mathbf{y}}^{T_k}). \tag{61}$$

3) $k\in\mathcal{S}(q)$ for $q\ge2$ (Periodic Decoding Phase)
We now consider the general case where we decode to $(\mathbf{X}_k,\mathbf{Y}_k)$ where $k\in\mathcal{S}(q)$ and $q\ge2$. In this case, we first jointly decode $(\mathbf{X}_{t_{q-1}}^{\alpha_{q-1}},\mathbf{Y}_{t_{q-1}}^{\alpha_{q-1}})$. Then we sequentially decode $(\mathbf{X}_j,\mathbf{Y}_j)$ with $j\in\mathcal{S}(q-1)$. Finally, we sequentially decode $(\mathbf{X}_j,\mathbf{Y}_j)$ with $j\le k$ and $j\in\mathcal{S}(q)$. To make it clear, we illustrate each part separately.

a) Joint decoding of $\mathbf{X}_{t_{q-1}}^{\alpha_{q-1}}$. Define

$$\mathcal{B}_1(\mathbf{X}_{t_{q-1}}^{\beta_{q-1}},\mathbf{Y}_{t_{q-1}}^{\beta_{q-1}}):=\left\{(\tilde{\mathbf{x}}_{t_{q-1}}^{\beta_{q-1}},\tilde{\mathbf{y}}_{t_{q-1}}^{\beta_{q-1}}):\forall\tau\in[\alpha_{q-1}:\beta_{q-1}],\ \tilde{f}_\tau(\tilde{\mathbf{x}}_{t_{q-1}}^\tau)=\tilde{f}_\tau(\mathbf{X}_{t_{q-1}}^\tau),\ \tilde{g}_\tau(\tilde{\mathbf{y}}_{t_{q-1}}^\tau)=\tilde{g}_\tau(\mathbf{Y}_{t_{q-1}}^\tau)\right\}. \tag{62}$$

We decode to $(\mathbf{X}_{t_{q-1}}^{\alpha_{q-1}},\mathbf{Y}_{t_{q-1}}^{\alpha_{q-1}})=(\tilde{\mathbf{x}}_{t_{q-1}}^{\alpha_{q-1}},\tilde{\mathbf{y}}_{t_{q-1}}^{\alpha_{q-1}})$ if there exists $(\tilde{\mathbf{x}}_{t_{q-1}}^{\beta_{q-1}},\tilde{\mathbf{y}}_{t_{q-1}}^{\beta_{q-1}})\in\mathcal{B}_1(\mathbf{X}_{t_{q-1}}^{\beta_{q-1}},\mathbf{Y}_{t_{q-1}}^{\beta_{q-1}})$ such that for all $(\bar{\mathbf{x}}_{t_{q-1}}^{\beta_{q-1}},\bar{\mathbf{y}}_{t_{q-1}}^{\beta_{q-1}})\in\mathcal{B}_1(\mathbf{X}_{t_{q-1}}^{\beta_{q-1}},\mathbf{Y}_{t_{q-1}}^{\beta_{q-1}})$, we have

$$\hat{H}_{\mathrm{S}}(l,m,\tilde{\mathbf{x}}_{t_{q-1}}^{\beta_{q-1}},\tilde{\mathbf{y}}_{t_{q-1}}^{\beta_{q-1}})\le\hat{H}_{\mathrm{S}}(l,m,\bar{\mathbf{x}}_{t_{q-1}}^{\beta_{q-1}},\bar{\mathbf{y}}_{t_{q-1}}^{\beta_{q-1}}). \tag{63}$$

b) Sequential decoding for $\mathbf{X}_{\alpha_{q-1}+1},\ldots,\mathbf{X}_{\beta_{q-1}}$, i.e., the remaining source blocks with indices in $\mathcal{S}(q-1)$. Let $\lambda_k=\min\{T_k,\beta_q\}$.

i) $\alpha_{q-1}+1\le j\le t_q-1$
Define

$$\mathcal{B}_2(\mathbf{X}_{t_{q-1}}^{\beta_{q-1}},\mathbf{Y}_{t_{q-1}}^{\beta_{q-1}},j):=\left\{(\tilde{\mathbf{x}}_{t_{q-1}}^{\beta_{q-1}},\tilde{\mathbf{y}}_{t_{q-1}}^{\beta_{q-1}}):\forall\tau\in[j:\beta_{q-1}],\ \tilde{f}_\tau(\tilde{\mathbf{x}}_{t_{q-1}}^\tau)=\tilde{f}_\tau(\mathbf{X}_{t_{q-1}}^\tau),\ \tilde{g}_\tau(\tilde{\mathbf{y}}_{t_{q-1}}^\tau)=\tilde{g}_\tau(\mathbf{Y}_{t_{q-1}}^\tau)\right\}. \tag{64}$$

Given $(\hat{\mathbf{X}}_{t_{q-1}}^{j-1},\hat{\mathbf{Y}}_{t_{q-1}}^{j-1})$, we decode to $(\hat{\mathbf{X}}_j,\hat{\mathbf{Y}}_j)=(\tilde{\mathbf{x}}_j,\tilde{\mathbf{y}}_j)$ if there exists $(\tilde{\mathbf{x}}_{t_{q-1}}^{\beta_{q-1}},\tilde{\mathbf{y}}_{t_{q-1}}^{\beta_{q-1}})\in\mathcal{B}_2(\mathbf{X}_{t_{q-1}}^{\beta_{q-1}},\mathbf{Y}_{t_{q-1}}^{\beta_{q-1}},j)$ satisfying $(\tilde{\mathbf{x}}_{t_{q-1}}^{j-1},\tilde{\mathbf{y}}_{t_{q-1}}^{j-1})=(\hat{\mathbf{X}}_{t_{q-1}}^{j-1},\hat{\mathbf{Y}}_{t_{q-1}}^{j-1})$ such that for all $(\bar{\mathbf{x}}_{t_{q-1}}^{\beta_{q-1}},\bar{\mathbf{y}}_{t_{q-1}}^{\beta_{q-1}})\in\mathcal{B}_2(\mathbf{X}_{t_{q-1}}^{\beta_{q-1}},\mathbf{Y}_{t_{q-1}}^{\beta_{q-1}},j)$ where $(\bar{\mathbf{x}}_{t_{q-1}}^{j-1},\bar{\mathbf{y}}_{t_{q-1}}^{j-1})=(\hat{\mathbf{X}}_{t_{q-1}}^{j-1},\hat{\mathbf{Y}}_{t_{q-1}}^{j-1})$, we have

$$\hat{H}_{\mathrm{S}}(l,m,\tilde{\mathbf{x}}_{t_{q-1}}^{\beta_{q-1}},\tilde{\mathbf{y}}_{t_{q-1}}^{\beta_{q-1}})\le\hat{H}_{\mathrm{S}}(l,m,\bar{\mathbf{x}}_{t_{q-1}}^{\beta_{q-1}},\bar{\mathbf{y}}_{t_{q-1}}^{\beta_{q-1}}). \tag{65}$$

ii) $t_q\le j\le\beta_{q-1}$
Define

$$\mathcal{B}_3(\mathbf{X}_{t_{q-1}}^{\lambda_k},\mathbf{Y}_{t_{q-1}}^{\lambda_k},j):=\Big\{(\tilde{\mathbf{x}}_{t_{q-1}}^{\lambda_k},\tilde{\mathbf{y}}_{t_{q-1}}^{\lambda_k}):\forall\tau\in[j:\beta_{q-1}],\ \tilde{f}_\tau(\tilde{\mathbf{x}}_{t_{q-1}}^\tau)=\tilde{f}_\tau(\mathbf{X}_{t_{q-1}}^\tau),\ \tilde{g}_\tau(\tilde{\mathbf{y}}_{t_{q-1}}^\tau)=\tilde{g}_\tau(\mathbf{Y}_{t_{q-1}}^\tau),\\ \forall\tau\in[\alpha_q:\lambda_k],\ \tilde{f}_\tau(\tilde{\mathbf{x}}_{t_q}^\tau)=\tilde{f}_\tau(\mathbf{X}_{t_q}^\tau),\ \tilde{g}_\tau(\tilde{\mathbf{y}}_{t_q}^\tau)=\tilde{g}_\tau(\mathbf{Y}_{t_q}^\tau)\Big\}. \tag{66}$$

Given $(\hat{\mathbf{X}}_{t_{q-1}}^{j-1},\hat{\mathbf{Y}}_{t_{q-1}}^{j-1})$, we decode to $(\hat{\mathbf{X}}_j,\hat{\mathbf{Y}}_j)=(\tilde{\mathbf{x}}_j,\tilde{\mathbf{y}}_j)$ if there exists $(\tilde{\mathbf{x}}_{t_{q-1}}^{\lambda_k},\tilde{\mathbf{y}}_{t_{q-1}}^{\lambda_k})\in\mathcal{B}_3(\mathbf{X}_{t_{q-1}}^{\lambda_k},\mathbf{Y}_{t_{q-1}}^{\lambda_k},j)$ satisfying $(\tilde{\mathbf{x}}_{t_{q-1}}^{j-1},\tilde{\mathbf{y}}_{t_{q-1}}^{j-1})=(\hat{\mathbf{X}}_{t_{q-1}}^{j-1},\hat{\mathbf{Y}}_{t_{q-1}}^{j-1})$ such that for all $(\bar{\mathbf{x}}_{t_{q-1}}^{\lambda_k},\bar{\mathbf{y}}_{t_{q-1}}^{\lambda_k})\in\mathcal{B}_3(\mathbf{X}_{t_{q-1}}^{\lambda_k},\mathbf{Y}_{t_{q-1}}^{\lambda_k},j)$ where $(\bar{\mathbf{x}}_{t_{q-1}}^{j-1},\bar{\mathbf{y}}_{t_{q-1}}^{j-1})=(\hat{\mathbf{X}}_{t_{q-1}}^{j-1},\hat{\mathbf{Y}}_{t_{q-1}}^{j-1})$, we have

$$\hat{H}_{\mathrm{S}}(l,m,\tilde{\mathbf{x}}_{t_{q-1}}^{\lambda_k},\tilde{\mathbf{y}}_{t_{q-1}}^{\lambda_k})\le\hat{H}_{\mathrm{S}}(l,m,\bar{\mathbf{x}}_{t_{q-1}}^{\lambda_k},\bar{\mathbf{y}}_{t_{q-1}}^{\lambda_k}). \tag{67}$$

c) Sequential decoding for $\mathbf{X}_{\alpha_q},\ldots,\mathbf{X}_k$, i.e., the source blocks with indices in $\mathcal{S}(q)$ which is no greater than k. We consider two scenarios which differ in the using of codewords.

i) $T_k\le\beta_q$.
Define

$$\mathcal{B}_4(\mathbf{X}_{t_q}^{T_k},\mathbf{Y}_{t_q}^{T_k},j):=\left\{(\tilde{\mathbf{x}}_{t_q}^{T_k},\tilde{\mathbf{y}}_{t_q}^{T_k}):\forall\tau\in[j:T_k],\ \tilde{f}_\tau(\tilde{\mathbf{x}}_{t_q}^\tau)=\tilde{f}_\tau(\mathbf{X}_{t_q}^\tau),\ \tilde{g}_\tau(\tilde{\mathbf{y}}_{t_q}^\tau)=\tilde{g}_\tau(\mathbf{Y}_{t_q}^\tau)\right\}. \tag{68}$$

Given $(\hat{\mathbf{X}}_{t_{q-1}}^{j-1}, \hat{\mathbf{Y}}_{t_{q-1}}^{j-1})$, we decode to $(\hat{\mathbf{X}}_j, \hat{\mathbf{Y}}_j) = (\tilde{\mathbf{x}}_j, \tilde{\mathbf{y}}_j)$ if there exists $(\tilde{\mathbf{x}}_{t_q}^{T_k}, \tilde{\mathbf{y}}_{t_q}^{T_k}) \in \mathcal{B}_4(\mathbf{X}_{t_q}^{T_k}, \mathbf{Y}_{t_q}^{T_k}, j)$ satisfying $(\tilde{\mathbf{x}}_{t_q}^{j-1}, \tilde{\mathbf{y}}_{t_q}^{j-1}) = (\hat{\mathbf{X}}_{t_q}^{j-1}, \hat{\mathbf{Y}}_{t_q}^{j-1})$ such that for all $(\bar{\mathbf{x}}_{t_q}^{T_k}, \bar{\mathbf{y}}_{t_q}^{T_k}) \in \mathcal{B}_4(\mathbf{X}_{t_q}^{T_k}, \mathbf{Y}_{t_q}^{T_k}, j)$ where $(\bar{\mathbf{x}}_{t_q}^{j-1}, \bar{\mathbf{y}}_{t_q}^{j-1}) = (\hat{\mathbf{X}}_{t_q}^{j-1}, \hat{\mathbf{Y}}_{t_q}^{j-1})$, we have

$$\hat{H}_{\mathrm{S}}(l, m, \tilde{\mathbf{x}}_{t_q}^{T_k}, \tilde{\mathbf{y}}_{t_q}^{T_k}) \leq \hat{H}_{\mathrm{S}}(l, m, \bar{\mathbf{x}}_{t_q}^{T_k}, \bar{\mathbf{y}}_{t_q}^{T_k}). \tag{69}$$

ii) $T_k > \beta_q$

A) $\alpha_q \leq j \leq t_{q+1} - 1$

Define

$$\mathcal{B}_5(\mathbf{X}_{t_q}^{\beta_q}, \mathbf{Y}_{t_q}^{\beta_q}, j) := \left\{(\tilde{\mathbf{x}}_{t_q}^{\beta_q}, \tilde{\mathbf{y}}_{t_q}^{\beta_q}) : \forall \tau \in [j : \beta_q],\ \tilde{f}_\tau(\tilde{\mathbf{x}}_{t_q}^{\tau}) = \tilde{f}_\tau(\mathbf{X}_{t_q}^{\tau}),\ \tilde{g}_\tau(\tilde{\mathbf{y}}_{t_q}^{\tau}) = \tilde{g}_\tau(\mathbf{Y}_{t_q}^{\tau})\right\}. \tag{70}$$

Given $(\hat{\mathbf{X}}_{t_{q-1}}^{j-1}, \hat{\mathbf{Y}}_{t_{q-1}}^{j-1})$, we decode to $(\hat{\mathbf{X}}_j, \hat{\mathbf{Y}}_j) = (\tilde{\mathbf{x}}_j, \tilde{\mathbf{y}}_j)$ if there exists $(\tilde{\mathbf{x}}_{t_q}^{\beta_q}, \tilde{\mathbf{y}}_{t_q}^{\beta_q}) \in \mathcal{B}_5(\mathbf{X}_{t_q}^{\beta_q}, \mathbf{Y}_{t_q}^{\beta_q}, j)$ satisfying $(\tilde{\mathbf{x}}_{t_q}^{j-1}, \tilde{\mathbf{y}}_{t_q}^{j-1}) = (\hat{\mathbf{X}}_{t_q}^{j-1}, \hat{\mathbf{Y}}_{t_q}^{j-1})$ such that for all $(\bar{\mathbf{x}}_{t_q}^{\beta_q}, \bar{\mathbf{y}}_{t_q}^{\beta_q}) \in \mathcal{B}_5(\mathbf{X}_{t_q}^{\beta_q}, \mathbf{Y}_{t_q}^{\beta_q}, j)$ where $(\bar{\mathbf{x}}_{t_q}^{j-1}, \bar{\mathbf{y}}_{t_q}^{j-1}) = (\hat{\mathbf{X}}_{t_q}^{j-1}, \hat{\mathbf{Y}}_{t_q}^{j-1})$, we have

$$\hat{H}_{\mathrm{S}}(l, m, \tilde{\mathbf{x}}_{t_q}^{\beta_q}, \tilde{\mathbf{y}}_{t_q}^{\beta_q}) \leq \hat{H}_{\mathrm{S}}(l, m, \bar{\mathbf{x}}_{t_q}^{\beta_q}, \bar{\mathbf{y}}_{t_q}^{\beta_q}). \tag{71}$$

B) $t_{q+1} \leq j \leq k$

Define

$$\begin{aligned}\mathcal{B}_6(\mathbf{X}_{t_q}^{T_k}, \mathbf{Y}_{t_q}^{T_k}, j) := \Big\{(\tilde{\mathbf{x}}_{t_q}^{T_k}, \tilde{\mathbf{y}}_{t_q}^{T_k}) :& \forall \tau \in [j : \beta_q],\ \tilde{f}_\tau(\tilde{\mathbf{x}}_{t_q}^{\tau}) = \tilde{f}_\tau(\mathbf{X}_{t_q}^{\tau}),\ \tilde{g}_\tau(\tilde{\mathbf{y}}_{t_q}^{\tau}) = \tilde{g}_\tau(\mathbf{Y}_{t_q}^{\tau}) \\ & \forall \tau \in [\alpha_{q+1} : T_k],\ \tilde{f}_\tau(\tilde{\mathbf{x}}_{t_{q+1}}^{\tau}) = \tilde{f}_\tau(\mathbf{X}_{t_{q+1}}^{\tau}),\ \tilde{g}_\tau(\tilde{\mathbf{y}}_{t_{q+1}}^{\tau}) = \tilde{g}_\tau(\mathbf{Y}_{t_{q+1}}^{\tau})\Big\}.\end{aligned} \tag{72}$$

Given $(\hat{\mathbf{X}}_{t_{q-1}}^{j-1}, \hat{\mathbf{Y}}_{t_{q-1}}^{j-1})$, we decode to $(\hat{\mathbf{X}}_j, \hat{\mathbf{Y}}_j) = (\tilde{\mathbf{x}}_j, \tilde{\mathbf{y}}_j)$ if there exists $(\tilde{\mathbf{x}}_{t_q}^{T_k}, \tilde{\mathbf{y}}_{t_q}^{T_k}) \in \mathcal{B}_6(\mathbf{X}_{t_q}^{T_k}, \mathbf{Y}_{t_q}^{T_k}, j)$ satisfying $(\tilde{\mathbf{x}}_{t_q}^{j-1}, \tilde{\mathbf{y}}_{t_q}^{j-1}) = (\hat{\mathbf{X}}_{t_q}^{j-1}, \hat{\mathbf{Y}}_{t_q}^{j-1})$ such that for all $(\bar{\mathbf{x}}_{t_q}^{T_k}, \bar{\mathbf{y}}_{t_q}^{T_k}) \in \mathcal{B}_6(\mathbf{X}_{t_q}^{T_k}, \mathbf{Y}_{t_q}^{T_k}, j)$ where $(\bar{\mathbf{x}}_{t_q}^{j-1}, \bar{\mathbf{y}}_{t_q}^{j-1}) = (\hat{\mathbf{X}}_{t_q}^{j-1}, \hat{\mathbf{Y}}_{t_q}^{j-1})$, we have

$$\hat{H}_{\mathrm{S}}(l, m, \tilde{\mathbf{x}}_{t_q}^{T_k}, \tilde{\mathbf{y}}_{t_q}^{T_k}) \leq \hat{H}_{\mathrm{S}}(l, m, \bar{\mathbf{x}}_{t_q}^{T_k}, \bar{\mathbf{y}}_{t_q}^{T_k}). \tag{73}$$

E. Error Events

In this subsection, we state the error events in our coding scheme. We only consider the case where $q \geq 2$, $k \in \mathcal{S}(q)$ (Periodic Phase of decoding) and $T_k > \beta_q$ since other cases are similar.

Define the set

$$\mathcal{F}(l, m, \mathbf{x}_a^b, \mathbf{y}_a^b) := \left\{(\tilde{\mathbf{x}}_a^b, \tilde{\mathbf{y}}_a^b) : \tilde{\mathbf{x}}_a^{l-1} = \mathbf{x}_a^{l-1}, \tilde{\mathbf{x}}_l \neq \mathbf{x}_l, \tilde{\mathbf{y}}_a^{m-1} = \mathbf{y}_a^{m-1}, \tilde{\mathbf{y}}_m \neq \mathbf{y}_m\right\}. \tag{74}$$

The error in decoding $(\mathbf{X}_k, \mathbf{Y}_k)$ occurs if and only if at least one of the following events occur:

1) $l, m \in [t_{q-1} : \beta_{q-1}]$ and $\min\{l, m\} \leq \alpha_{q-1}$

$$\begin{aligned}\mathcal{E}_{\mathrm{sw}}^1(l, m) := \Bigg\{ & \exists (\tilde{\mathbf{x}}_{t_{q-1}}^{\beta_{q-1}}, \tilde{\mathbf{y}}_{t_{q-1}}^{\beta_{q-1}}) \in \mathcal{B}_1(\mathbf{X}_{t_{q-1}}^{\beta_{q-1}}, \mathbf{Y}_{t_{q-1}}^{\beta_{q-1}}) \bigcap \mathcal{F}(l, m, \mathbf{X}_{t_{q-1}}^{\beta_{q-1}}, \mathbf{Y}_{t_{q-1}}^{\beta_{q-1}}) : \\ & \hat{H}_{\mathrm{S}}(l, m, \tilde{\mathbf{x}}_{t_{q-1}}^{\beta_{q-1}}, \tilde{\mathbf{y}}_{t_{q-1}}^{\beta_{q-1}}) \leq \hat{H}_{\mathrm{S}}(l, m, \mathbf{X}_{t_{q-1}}^{\beta_{q-1}}, \mathbf{Y}_{t_{q-1}}^{\beta_{q-1}})\Bigg\}.\end{aligned} \tag{75}$$

These error events correspond to the joint decoding of $(\mathbf{X}_{t_{q-1}}^{\alpha_{q-1}}, \mathbf{Y}_{t_{q-1}}^{\alpha_{q-1}})$ (refer to (62) and (63)).

2) $j \in [\alpha_{q-1} + 1 : t_q - 1]$, $l, m \in [j : \beta_{q-1}]$ and $\min\{l, m\} = j$.

$$\begin{aligned}\mathcal{E}_{\mathrm{sw},j}^2(l, m) := \Bigg\{ & \exists (\tilde{\mathbf{x}}_{t_{q-1}}^{\beta_{q-1}}, \tilde{\mathbf{y}}_{t_{q-1}}^{\beta_{q-1}}) \in \mathcal{B}_2(\mathbf{X}_{t_{q-1}}^{\beta_{q-1}}, \mathbf{Y}_{t_{q-1}}^{\beta_{q-1}}) \bigcap \mathcal{F}(l, m, \mathbf{X}_{t_{q-1}}^{\beta_{q-1}}, \mathbf{Y}_{t_{q-1}}^{\beta_{q-1}}) : \\ & \hat{H}_{\mathrm{S}}(l, m, \tilde{\mathbf{x}}_{t_{q-1}}^{\beta_{q-1}}, \tilde{\mathbf{y}}_{t_{q-1}}^{\beta_{q-1}}) \leq \hat{H}_{\mathrm{S}}(l, m, \mathbf{X}_{t_{q-1}}^{\beta_{q-1}}, \mathbf{Y}_{t_{q-1}}^{\beta_{q-1}})\Bigg\}.\end{aligned} \tag{76}$$

These error events correspond to the sequential decoding of $\mathbf{X}_{\alpha_{q-1}+1}, \ldots \mathbf{X}_{t_q-1}$ (refer to (64) and (65)).

3) $j \in [t_q : \beta_{q-1}]$, $l, m \in [j : \beta_q]$ and $\min\{l, m\} = j$.

$$\mathcal{E}^3_{\mathrm{sw},j}(l,m) := \Bigg\{ \exists (\tilde{\mathbf{x}}^{\beta_q}_{t_{q-1}}, \tilde{\mathbf{y}}^{\beta_q}_{t_{q-1}}) \in \mathcal{B}_3(\mathbf{X}^{\beta_q}_{t_{q-1}}, \mathbf{Y}^{\beta_q}_{t_{q-1}}) \bigcap \mathcal{F}(l, m, \mathbf{X}^{\beta_q}_{t_{q-1}}, \mathbf{Y}^{\beta_q}_{t_{q-1}}) : \\ \hat{H}_{\mathrm{S}}(l, m, \tilde{\mathbf{x}}^{\beta_q}_{t_{q-1}}, \tilde{\mathbf{y}}^{\beta_q}_{t_{q-1}}) \le \hat{H}_{\mathrm{S}}(l, m, \mathbf{X}^{\beta_q}_{t_{q-1}}, \mathbf{Y}^{\beta_q}_{t_{q-1}}) \Bigg\}. \tag{77}$$

These error events correspond to the sequential decoding of $\mathbf{X}_{t_q}, \dots \mathbf{X}_{\beta_{q-1}}$ (refer to (66) and (67)).

4) $j \in [\alpha_q : t_{q+1} - 1]$, $l, m \in [j : \beta_q]$ and $\min\{l, m\} = j$.

$$\mathcal{E}^5_{\mathrm{sw},j}(l,m) := \Bigg\{ \exists (\tilde{\mathbf{x}}^{\beta_q}_{t_q}, \tilde{\mathbf{y}}^{\beta_q}_{t_q}) \in \mathcal{B}_5(\mathbf{X}^{\beta_q}_{t_q}, \mathbf{Y}^{\beta_q}_{t_q}, j) \bigcap \mathcal{F}(l, m, \mathbf{X}^{\beta_q}_{t_q}, \mathbf{Y}^{\beta_q}_{t_q}) : \\ \hat{H}_{\mathrm{S}}(l, m, \tilde{\mathbf{x}}^{\beta_q}_{t_q}, \tilde{\mathbf{y}}^{\beta_q}_{t_q}) \le \hat{H}_{\mathrm{S}}(l, m, \mathbf{X}^{\beta_q}_{t_q}, \mathbf{Y}^{\beta_q}_{t_q}) \Bigg\}. \tag{78}$$

These error events correspond to the sequential decoding of $\mathbf{X}_{\alpha_q}, \dots \mathbf{X}_{t_{q+1}-1}$ (refer to (70) and (71)).

5) $j \in [t_{q+1} : k]$, $l, m \in [j : T_k]$ and $\min\{l, m\} = j$.

$$\mathcal{E}^6_{\mathrm{sw},j}(l,m) := \Bigg\{ \exists (\tilde{\mathbf{x}}^{T_k}_{t_q}, \tilde{\mathbf{y}}^{T_k}_{t_q}) \in \mathcal{B}_6(\mathbf{X}^{T_k}_{t_q}, \mathbf{Y}^{T_k}_{t_q}, j) \bigcap \mathcal{F}(l, m, \mathbf{X}^{T_k}_{t_q}, \mathbf{Y}^{T_k}_{t_q}) : \\ \hat{H}_{\mathrm{S}}(l, m, \tilde{\mathbf{x}}^{T_k}_{t_q}, \tilde{\mathbf{y}}^{T_k}_{t_q}) \le \hat{H}_{\mathrm{S}}(l, m, \mathbf{X}^{T_k}_{t_q}, \mathbf{Y}^{T_k}_{t_q}) \Bigg\}. \tag{79}$$

These error events correspond to the sequential decoding of $\mathbf{X}_{t_{q+1}}, \dots \mathbf{X}_k$ (refer to (72) and (73)).

For the case where $k \in \mathcal{S}(q)$ and $T_k \le \beta_q$, the error events are almost the same as the case $T_k > \beta_q$ except that we do not have $\mathcal{E}^5_{\mathrm{sw},j}$ and $\mathcal{E}^6_{\mathrm{sw},j}$. Instead, we have for $j \in [\alpha_q : k]$, $l, m \in [j : T_k]$ and $\min\{l, m\} = j$,

$$\mathcal{E}^4_{\mathrm{sw},j}(l,m) : = \Bigg\{ \exists (\tilde{\mathbf{x}}^{T_k}_{t_q}, \tilde{\mathbf{y}}^{T_k}_{t_q}) \in \mathcal{B}_4(\mathbf{X}^{T_k}_{t_q}, \mathbf{Y}^{T_k}_{t_q}, j) \bigcap \mathcal{F}(l, m, \mathbf{X}^{T_k}_{t_q}, \mathbf{Y}^{T_k}_{t_q}) : \tag{80}$$

$$\hat{H}_{\mathrm{S}}(l, m, \tilde{\mathbf{x}}^{T_k}_{t_q}, \tilde{\mathbf{y}}^{T_k}_{t_q}) \le \hat{H}_{\mathrm{S}}(l, m, \mathbf{X}^{T_k}_{t_q}, \mathbf{Y}^{T_k}_{t_q}) \Bigg\} \tag{81}$$

We remark these error events $\mathcal{E}^4_{\mathrm{sw},j}(l,m)$ correspond to the sequential decoding of $\mathbf{X}_{\alpha_q}, \dots \mathbf{X}_k$ (refer to (68) and (69)).

F. Preliminaries

In this subsection, we present some preliminaries which will be used in the analysis of the probabilities of the error events defined in Section V-G.

First we recall some important quantities that are used in the derivation of the error exponent in [7] for distributed streaming compression, i.e., $E_X(R_X, R_Y, \gamma)$ and $E_Y(R_X, R_Y, \gamma)$ (cf. [7, Eqn. (29)]). We recap the expression in both the Gallager [9] and Csiszár-Körner [8] forms. In the proof in Section V-G, we use the fact that these two forms for the error exponent are equivalent (cf. [7, Lemma 5]). Define

$$E_X(R_X, R_Y, \gamma) := \max_{\rho \in [0,1]} \Big[\gamma E_{X|Y}(R_X, \rho) + (1-\gamma) E_{XY}(R_X, R_Y, \rho) \Big] \tag{82}$$

$$:= \inf_{Q_{XY}, \tilde{Q}_{XY}} \gamma D(Q_{XY} \| P_{XY}) + (1-\gamma) D(\tilde{Q}_{XY} \| P_{XY}) \\ + \Big| \gamma \left(R_X - H(Q_{X|Y} | Q_Y) \right) + (1-\gamma) \left(R_X + R_Y - H(\tilde{Q}_{XY}) \right) \Big|^+, \tag{83}$$

where the *Gallager functions* are

$$E_{X|Y}(R_X, \rho) := \rho R_X - \log \sum_y P_Y(y) \left(\sum_x P_{X|Y}(x|y)^{\frac{1}{1+\rho}} \right)^{1+\rho}, \tag{84}$$

$$E_{XY}(R_X, R_Y, \rho) := \rho(R_X + R_Y) - (1+\rho) \log \sum_{x,y} P_{XY}(x,y)^{\frac{1}{1+\rho}}, \tag{85}$$

and $E_Y(R_X, R_Y, \gamma)$ is similar to $E_X(R_X, R_Y, \gamma)$ with X and Y interchanged.

Now, we present the definitions of the coding rates. Let (R_X^*, R_Y^*) be a rate pair on the boundary of the SW region (see Cases (i)-(v) in Figure 2). We choose (n, N_1, N_2) such that

$$R_{X,n} := \frac{1}{n}\log N_1 = R_X^* + \theta_1 \xi_n \tag{86}$$

$$R_{Y,n} := \frac{1}{n}\log N_2 = R_Y^* + \theta_2 \xi_n. \tag{87}$$

We remark that this choice satisfies the conditions in Definition 2.

Additionally, define the rates

$$R_{X,n}^l := \frac{\Psi - \Omega + 1}{n(\beta_{q-1} - l + 1)}\log N_1, \tag{88}$$

$$R_{Y,n}^m := \frac{\Psi - \Omega + 1}{n(\beta_{q-1} - m + 1)}\log N_2. \tag{89}$$

We remark that $R_{X,n}^l$ and $R_{Y,n}^m$ are only used in the analysis of the probability of the error event $\mathcal{E}_{\mathrm{sw}}^1(l,m)$ (see the definition in (75) and the upper bounding in the steps leading to (113)) where $l, m \in [t_{q-1} : \alpha_{q-1}]$.

Finally, we present an alternative form of the weighted empirical suffix entropy in (51) in terms of types and conditional types. Given $P_1, P_2 \in \mathcal{P}(\mathcal{Y})$ and $V_1, V_2 \in \mathcal{P}(\mathcal{X}|\mathcal{Y})$, when $l \le m$, define

$$H_{\mathrm{S}}(P_1, P_2, V_1, V_2) := (m-l)H(V_1|P_1) + (\beta_{q-1} - m + 1)H(P_2 \times V_2). \tag{90}$$

Using the definition of the weighted empirical suffix entropy in (51), we obtain that for $a \le l \le m \le b$,

$$H_{\mathrm{S}}(P_1, P_2, V_1, V_2) = \hat{H}_{\mathrm{S}}(l, m, \tilde{\mathbf{x}}_a^b, \tilde{\mathbf{y}}_a^b). \tag{91}$$

holds if $\tilde{\mathbf{y}}_l^{m-1} \in \mathcal{T}_{P_1}$, $\tilde{\mathbf{y}}_m^b \in \mathcal{T}_{P_2}$, $\tilde{\mathbf{x}}_l^{m-1} \in \mathcal{T}_{V_1}(\tilde{\mathbf{y}}_l^{m-1})$ and $\tilde{\mathbf{x}}_m^b \in \mathcal{T}_{V_2}(\tilde{\mathbf{y}}_m^b)$ for some types $P_1 \in \mathcal{P}_{n(m-l)}(\mathcal{Y})$ and $P_2 \in \mathcal{P}_{n(b-m+1)}(\mathcal{Y})$ and some conditional types $V_1 \in \mathcal{V}_{n(m-l)}(\mathcal{X}; P_1)$ and $V_2 \in \mathcal{V}_{n(b-m+1)}(\mathcal{X}; P_2)$. Recall that given an n-type $P \in \mathcal{P}_n(\mathcal{Y})$, $\mathcal{V}_n(\mathcal{X}; P) := \{V \in \mathcal{P}(\mathcal{X}|\mathcal{Y}) : \mathcal{T}_V(y^n) \neq \emptyset \text{ for some } y^n \in \mathcal{T}_P\}$ is the set of all conditional types "compatible with" P. We use the facts that $|\mathcal{P}_n(\mathcal{Y})| \le (n+1)^{|\mathcal{Y}|}$ and $|\mathcal{V}_n(\mathcal{X}; P)| \le (n+1)^{|\mathcal{X}||\mathcal{Y}|}$ extensively in the following.

G. Evaluation of the Error Probabilities

We are now ready to upper bound the probability of each error event using the definitions in Section V-F.

1) $\mathcal{E}_{\mathrm{sw}}^1(l,m)$ (recall (75)). We then analyze its probability using above definitions.

   a) $l, m \in [t_{q-1} : \alpha_{q-1}]$ and $l \le m$.
   We can upper bound the probability of $\mathcal{E}_{\mathrm{sw}}^1(l,m)$ as follows:

$$\begin{aligned}&\Pr\left(\mathcal{E}_{\mathrm{sw}}^1(l,m)\right)\\ &\le \sum_{(\mathbf{x}_{t_{q-1}}^{\beta_{q-1}}, \mathbf{y}_{t_{q-1}}^{\beta_{q-1}})} P_{XY}^{n\Psi}(\mathbf{x}_{t_{q-1}}^{\beta_{q-1}}, \mathbf{y}_{t_{q-1}}^{\beta_{q-1}}) \min\left\{1, \Pr\left(\mathcal{E}_{\mathrm{sw}}^1(l,m) | (\mathbf{X}_{t_{q-1}}^{\beta_{q-1}}, \mathbf{Y}_{t_{q-1}}^{\beta_{q-1}}) = (\mathbf{x}_{t_{q-1}}^{\beta_{q-1}}, \mathbf{y}_{t_{q-1}}^{\beta_{q-1}})\right)\right\},\end{aligned} \tag{92}$$

   where $\beta_{q-1} - t_{q-1} + 1 = \Psi$ by invoking the definitions of β_q and t_q in (46) and (47) respectively.
   Define

$$q(\mathbf{x}_{t_{q-1}}^{\beta_{q-1}}, \mathbf{y}_{t_{q-1}}^{\beta_{q-1}})) := \Pr\left(\mathcal{E}_{\mathrm{sw}}^1(l,m) | (\mathbf{X}_{t_{q-1}}^{\beta_{q-1}}, \mathbf{Y}_{t_{q-1}}^{\beta_{q-1}}) = (\mathbf{x}_{t_{q-1}}^{\beta_{q-1}}, \mathbf{y}_{t_{q-1}}^{\beta_{q-1}})\right). \tag{93}$$

   Then, given types $P_1 \in \mathcal{P}_{n(m-l)}(\mathcal{Y})$, $P_2 \in \mathcal{P}_{n(\beta_{q-1}-m+1)}(\mathcal{Y})$ and conditional types, $V_1 \in \mathcal{V}_{n(m-l)}(\mathcal{X}; P_1)$, $V_2 \in$

$\mathcal{V}_{n(\beta_{q-1}-m+1)}(\mathcal{X};P_2)$, if $\mathbf{y}_l^{m-1}\in\mathcal{T}_{P_1}$, $\mathbf{y}_m^{\beta_{q-1}}\in\mathcal{T}_{P_2}$, $\mathbf{x}_l^{m-1}\in\mathcal{T}_{V_1}(\mathbf{y}_l^{m-1})$ and $\mathbf{x}_m^{\beta_{q-1}}\in\mathcal{T}_{V_2}(\mathbf{y}_m^{\beta_{q-1}})$, we obtain

$$q(\mathbf{x}_{t_{q-1}}^{\beta_{q-1}},\mathbf{y}_{t_{q-1}}^{\beta_{q-1}}))\leq \sum_{\substack{(\tilde{\mathbf{x}}_{t_{q-1}}^{\beta_{q-1}},\tilde{\mathbf{y}}_{t_{q-1}}^{\beta_{q-1}})\in\mathcal{F}(l,m,\mathbf{x}_{t_{q-1}}^{\beta_{q-1}},\mathbf{y}_{t_{q-1}}^{\beta_{q-1}}):\\ \hat{H}_{\mathrm{S}}(l,m,\tilde{\mathbf{x}}_{t_{q-1}}^{\beta_{q-1}},\tilde{\mathbf{y}}_{t_{q-1}}^{\beta_{q-1}})\leq\hat{H}_{\mathrm{S}}(l,m,\mathbf{x}_{t_{q-1}}^{\beta_{q-1}},\mathbf{y}_{t_{q-1}}^{\beta_{q-1}})}}\Pr\left((\tilde{\mathbf{x}}_{t_{q-1}}^{\beta_{q-1}},\tilde{\mathbf{y}}_{t_{q-1}}^{\beta_{q-1}})\in\mathcal{B}_1(\mathbf{x}_{t_{q-1}}^{\beta_{q-1}},\mathbf{y}_{t_{q-1}}^{\beta_{q-1}})\right) \tag{94}$$

$$\leq \sum_{\substack{(\tilde{\mathbf{x}}_{t_{q-1}}^{\beta_{q-1}},\tilde{\mathbf{y}}_{t_{q-1}}^{\beta_{q-1}})\in\mathcal{F}(l,m,\mathbf{x}_{t_{q-1}}^{\beta_{q-1}},\mathbf{y}_{t_{q-1}}^{\beta_{q-1}}):\\ \hat{H}_{\mathrm{S}}(l,m,\tilde{\mathbf{x}}_{t_{q-1}}^{\beta_{q-1}},\tilde{\mathbf{y}}_{t_{q-1}}^{\beta_{q-1}})\leq\hat{H}_{\mathrm{S}}(l,m,\mathbf{x}_{t_{q-1}}^{\beta_{q-1}},\mathbf{y}_{t_{q-1}}^{\beta_{q-1}})}}\frac{1}{(N_1N_2)^{\Psi-\Omega+1}} \tag{95}$$

$$= \sum_{\substack{\tilde{P}_2\in\mathcal{P}_{n(\beta_{q-1}-m+1)}(\mathcal{Y}),\\ \tilde{V}_1\in\mathcal{V}_{n(m-l)}(\mathcal{X};P_1),\\ \tilde{V}_2\in\mathcal{V}_{n(\beta_{q-1}-m+1)}(\mathcal{X};\tilde{P}_2):\\ H_{\mathrm{S}}(P_1,\tilde{P}_2,\tilde{V}_1,\tilde{V}_2)\leq H_{\mathrm{S}}(P_1,P_2,V_1,V_2)}}\sum_{\tilde{y}_m^{\beta_{q-1}}\in\mathcal{T}_{\tilde{P}_2}}\sum_{\substack{\tilde{\mathbf{x}}_l^{m-1}\in\mathcal{P}_{\tilde{V}_1}(\mathbf{y}_l^{m-1})\\ \tilde{\mathbf{x}}_m^{\beta_{q-1}}\in\mathcal{P}_{\tilde{V}_2}(\tilde{\mathbf{y}}_m^{\beta_{q-1}})}}\frac{1}{(N_1N_2)^{\Psi-\Omega+1}} \tag{96}$$

$$\leq \frac{(n(\beta_{q-1}-l+1)+1)^{3|\mathcal{X}||\mathcal{Y}|}\exp(n(\beta_{q-1}-l+1)H_{\mathrm{S}}(P_1,P_2,V_1,V_2))}{(N_1N_2)^{\Psi-\Omega+1}}, \tag{97}$$

where (95) follows because in the joint decoding of $(\mathbf{X}_{t_{q-1}}^{\alpha_{q-1}},\mathbf{Y}_{t_{q-1}}^{\alpha_{q-1}})$, we make use of $\beta_{q-1}-\alpha_{q-1}+1=\Psi-\Omega+1$ binning codewords (see decoding procedure in (62) and (63), corresponding error events in (75) and the definitions of α_q and β_q in (45) and (46) respectively); (96) follows by invoking the definition of $\mathcal{F}(l,m,\mathbf{x}_{t_{q-1}}^{\beta_{q-1}},\mathbf{y}_{t_{q-1}}^{\beta_{q-1}})$ in (74), expressing the sum of sequences as sum over (conditional) type classes and noting the equivalent expression for the score function in terms of types and conditional types in (91); (97) follows from calculations involving types [8] and the condition $H_{\mathrm{S}}(P_1,\tilde{P}_2,\tilde{V}_1,\tilde{V}_2)\leq H_{\mathrm{S}}(P_1,P_2,V_1,V_2)$ (also refer to [7, Eqns. (82)-(87)]). In particular, noting that $l\leq m\leq\alpha_{q-1}<\beta_{q-1}$,

$$\left|\mathcal{P}_{n(\beta_{q-1}-m+1)}(\mathcal{Y})\right|\leq(n(\beta_{q-1}-m+1)+1)^{|\mathcal{Y}|} \tag{98}$$

$$\leq(n(\beta_{q-1}-l+1)+1)^{|\mathcal{X}||\mathcal{Y}|}, \tag{99}$$

$$\left|\mathcal{V}_{n(m-l)}(\mathcal{X};P_1)\right|\left|\mathcal{V}_{n(\beta_{q-1}-m+1)}(\mathcal{X};\tilde{P}_2)\right|\leq(n(m-l)+1)^{|\mathcal{X}||\mathcal{Y}|}(n(\beta_{q-1}-m+1)+1)^{|\mathcal{X}||\mathcal{Y}|} \tag{100}$$

$$\leq(n(\beta_{q-1}-l+1)+1)^{2|\mathcal{X}||\mathcal{Y}|}, \tag{101}$$

we obtain the polynomial factor in (97).
Combining (92) and (97), we obtain

$$\Pr\left(\mathcal{E}_{\mathrm{sw}}^1(l,m)\right)$$

$$\leq \sum_{(\mathbf{x}_{t_{q-1}}^{\beta_{q-1}},\mathbf{y}_{t_{q-1}}^{\beta_{q-1}})}P_{XY}^{n\Psi}(\mathbf{x}_{t_{q-1}}^{\beta_{q-1}},\mathbf{y}_{t_{q-1}}^{\beta_{q-1}})\min\left\{1,q(\mathbf{x}_{t_{q-1}}^{\beta_{q-1}},\mathbf{y}_{t_{q-1}}^{\beta_{q-1}})\right\} \tag{102}$$

$$= \sum_{(\mathbf{x}_{t_{q-1}}^{l-1},\mathbf{y}_{t_{q-1}}^{l-1})}P_{XY}^{n(l-t_q+1)}(\mathbf{x}_{t_{q-1}}^{l-1},\mathbf{y}_{t_{q-1}}^{l-1})\sum_{(\mathbf{x}_l^{\beta_{q-1}},\mathbf{y}_l^{\beta_{q-1}})}P_{XY}^{n(\beta_q-l+1)}(\mathbf{x}_l^{\beta_{q-1}},\mathbf{y}_l^{\beta_{q-1}})\min\left\{1,q(\mathbf{x}_{t_{q-1}}^{\beta_{q-1}},\mathbf{y}_{t_{q-1}}^{\beta_{q-1}})\right\} \tag{103}$$

$$= \sum_{(\mathbf{x}_{t_{q-1}}^{l-1},\mathbf{y}_{t_{q-1}}^{l-1})}P_{XY}^{n(l-t_q+1)}(\mathbf{x}_{t_{q-1}}^{l-1},\mathbf{y}_{t_{q-1}}^{l-1})\sum_{\substack{P_1\in\mathcal{P}_{n(m-l)}(\mathcal{Y})\\ P_2\in\mathcal{P}_{n(\beta_{q-1}-m+1)}(\mathcal{Y})\\ V_1\in\mathcal{V}_{n(m-l)}(\mathcal{X};P_1)\\ V_2\in\mathcal{V}_{n(\beta_{q-1}-m+1)}(\mathcal{X};P_2)}}\sum_{\substack{\mathbf{y}_l^{m-1}\in\mathcal{T}_{P_1}\\ \mathbf{y}_m^{\beta_{q-1}}\in\mathcal{T}_{P_2}}}\sum_{\substack{\mathbf{x}_l^{m-1}\in\mathcal{T}_{V_1}(\mathbf{y}_l^{m-1})\\ \mathbf{x}_m^{\beta_{q-1}}\in\mathcal{T}_{V_2}(\mathbf{y}_m^{\beta_{q-1}})}}$$

$$P_{XY}^{n(\beta_{q-1}-l+1)}(\mathbf{x}_l^{\beta_{q-1}},\mathbf{y}_l^{\beta_{q-1}})\min\left\{1,q(\mathbf{x}_{t_{q-1}}^{\beta_{q-1}},\mathbf{y}_{t_{q-1}}^{\beta_{q-1}}))\right\} \tag{104}$$

$$\leq\left(\sum_{(\mathbf{x}_{t_{q-1}}^{l-1},\mathbf{y}_{t_{q-1}}^{l-1})}P_{XY}^{n(l-t_q+1)}(\mathbf{x}_{t_{q-1}}^{l-1},\mathbf{y}_{t_{q-1}}^{l-1})\right)\sum_{\substack{P_1\in\mathcal{P}_{n(m-l)}(\mathcal{Y})\\ P_2\in\mathcal{P}_{n(\beta_{q-1}-m+1)}(\mathcal{Y})}}\sum_{\substack{V_1\in\mathcal{V}_{n(m-l)}(\mathcal{X};P_1)\\ V_2\in\mathcal{V}_{n(\beta_{q-1}-m+1)}(\mathcal{X};P_2)}}$$

$$\exp\left\{-n(m-l)D(P_1\times V_1\|P_{XY})-n(\beta_{q-1}-m+1)D(P_2\times V_2\|P_{XY})\right\}$$

$$\times\min\left\{1,\frac{(n(\beta_{q-1}-l+1)+1)^{3|\mathcal{X}||\mathcal{Y}|}\exp(n(\beta_{q-1}-l+1)H_{\mathrm{S}}(P_1,P_2,V_1,V_2))}{(N_1N_2)^{\Psi-\Omega+1}}\right\} \tag{105}$$

$$= \sum_{\substack{P_1 \in \mathcal{P}_{n(m-l)}(\mathcal{Y}) \\ P_2 \in \mathcal{P}_{n(\beta_{q-1}-m+1)}(\mathcal{Y})}} \sum_{\substack{V_1 \in \mathcal{V}_{n(m-l)}(\mathcal{X};P_1) \\ V_2 \in \mathcal{V}_{n(\beta_{q-1}-m+1)}(\mathcal{X};P_2)}}$$
$$\exp\left\{-n(m-l)D(P_1 \times V_1 \| P_{XY}) - n(\beta_{q-1}-m+1)D(P_2 \times V_2 \| P_{XY})\right\}$$
$$\times \left(n(\beta_{q-1}-l+1)+1\right)^{3|\mathcal{X}||\mathcal{Y}|} \min\left\{1, \frac{\exp\left(n(\beta_{q-1}-l+1)H_{\mathrm{S}}(P_1,P_2,V_1,V_2)\right)}{(N_1N_2)^{\Psi-\Omega+1}}\right\} \tag{106}$$
$$\leq \left(n(\beta_{q-1}-l+1)+1\right)^{7|\mathcal{X}||\mathcal{Y}|} \exp\left\{-n(\beta_{q-1}-l+1)E_X\left(R^l_{X,n}, R^m_{Y,n}, \frac{m-l}{\beta_q-l+1}\right)\right\} \tag{107}$$
$$\leq \left(n(\beta_{q-1}-l+1)+1\right)^{7|\mathcal{X}||\mathcal{Y}|} \exp\left\{-n(\beta_{q-1}-l+1) \inf_{\gamma\in[0,1]} E_X(R^l_{X,n}, R^m_{Y,n}, \gamma)\right\}, \tag{108}$$

where in (103) we split the product distribution $P_{XY}^{n\Psi}$ and the sum into two parts (from t_{q-1} to $l-1$ and from l to β_{q-1}); in (104) we split the inner sum into types and conditional types; (105) follows from using standard results from the method of types [8] and the upper bound on $q(\mathbf{x}_{t_{q-1}}^{\beta_{q-1}}, \mathbf{y}_{t_{q-1}}^{\beta_{q-1}})$ in (97); (106) follows by using the fact that the first term in parentheses in (105) is unity; (107) follows from upper bound on the number of (conditional) types similarly as in (97) (also see the calculations in (98)–(101)), the definition of $E_X(R_X, R_Y, \gamma)$ in (82), the codeword sizes N_1 in (86) and N_2 in (87), and the rates $R^l_{X,n}$ in (88) and $R^m_{Y,n}$ in (89).

b) $l \in [t_{q-1} : \alpha_{q-1}]$ and $m \in [\alpha_{q-1}+1 : \beta_{q-1}]$
Similarly, we obtain

$$\Pr\left(\mathcal{E}^1_{\mathrm{sw}}(l,m)\right) \leq \left(n(\beta_{q-1}-l+1)+1\right)^{7|\mathcal{X}||\mathcal{Y}|} \exp\left\{-n(\beta_{q-1}-l+1) \inf_{\gamma\in[0,1]} E_X(R^l_{X,n}, R_{Y,n}, \gamma)\right\}. \tag{109}$$

c) $l, m \in [t_{q-1} : \alpha_{q-1}]$ and $l \geq m$
Symmetrically, we obtain

$$\Pr\left(\mathcal{E}^1_{\mathrm{sw}}(l,m)\right) \leq \left(n(\beta_{q-1}-m+1)+1\right)^{7|\mathcal{X}||\mathcal{Y}|} \exp\left\{-n(\beta_{q-1}-m+1) \inf_{\gamma\in[0,1]} E_Y(R^l_{X,n}, R^m_{Y,n}, \gamma)\right\}. \tag{110}$$

d) $m \in [t_{q-1} : \alpha_{q-1}]$ and $l \in [\alpha_{q-1}+1 : \beta_{q-1}]$
Similarly,

$$\Pr\left(\mathcal{E}^1_{\mathrm{sw}}(l,m)\right) \leq \left(n(\beta_{q-1}-m+1)+1\right)^{7|\mathcal{X}||\mathcal{Y}|} \exp\left\{-n(\beta_{q-1}-m+1) \inf_{\gamma\in[0,1]} E_Y(R_{X,n}, R^m_{Y,n}, \gamma)\right\}. \tag{111}$$

Therefore, for all $n \in \mathbb{N}$,

$$\Pr\left(\bigcup_{\substack{l,m\in[t_{q-1}:\beta_{q-1}]: \\ \min\{l,m\}\leq\alpha_{q-1}}} \mathcal{E}^1_{\mathrm{sw}}(l,m)\right) \leq \sum_{\substack{l,m\in[t_{q-1}:\beta_{q-1}]: \\ \min\{l,m\}\leq\alpha_{q-1}}} \Pr\left(\mathcal{E}^1_{\mathrm{sw}}(l,m)\right)$$
$$\leq \sum_{l=t_{q-1}}^{\alpha_{q-1}} \sum_{m=l}^{\alpha_{q-1}} \left(n(\beta_{q-1}-l+1)+1\right)^{7|\mathcal{X}||\mathcal{Y}|} \exp\left\{-n(\beta_{q-1}-l+1) \inf_{\gamma\in[0,1]} E_X(R^l_{X,n}, R^m_{Y,n}, \gamma)\right\}$$
$$+ \sum_{m=t_{q-1}}^{\alpha_{q-1}} \sum_{l=m}^{\alpha_{q-1}} \left(n(\beta_{q-1}-m+1)+1\right)^{7|\mathcal{X}||\mathcal{Y}|} \exp\left\{-n(\beta_{q-1}-m+1) \inf_{\gamma\in[0,1]} E_Y(R^l_{X,n}, R^m_{Y,n}, \gamma)\right\}$$
$$+ \sum_{l=t_{q-1}}^{\alpha_{q-1}} \sum_{m=\alpha_{q-1}+1}^{\beta_{q-1}} \left(n(\beta_{q-1}-l+1)+1\right)^{7|\mathcal{X}||\mathcal{Y}|} \exp\left\{-n(\beta_{q-1}-l+1) \inf_{\gamma\in[0,1]} E_X(R^l_{X,n}, R_{Y,n}, \gamma)\right\}$$
$$+ \sum_{m=t_{q-1}}^{\alpha_{q-1}} \sum_{l=\alpha_{q-1}+1}^{\beta_{q-1}} \left(n(\beta_{q-1}-m+1)+1\right)^{7|\mathcal{X}||\mathcal{Y}|} \exp\left\{-n(\beta_{q-1}-m+1) \inf_{\gamma\in[0,1]} E_Y(R_{X,n}, R^m_{Y,n}, \gamma)\right\} \tag{112}$$
$$\leq 2\Psi^2 (n\Psi+1)^{7|\mathcal{X}||\mathcal{Y}|} \exp\left\{-n(\Psi-\Omega+1) \min\left\{\inf_{\gamma\in[0,1]} E_X(R^l_{X,n}, R^m_{Y,n}, \gamma), \inf_{\gamma\in[0,1]} E_Y(R^l_{X,n}, R^m_{Y,n}, \gamma)\right\}\right\}$$
$$+ 2\Psi^2 (n\Psi+1)^{7|\mathcal{X}||\mathcal{Y}|} \exp\left\{-n(\Psi-\Omega+1) \min\left\{\inf_{\gamma\in[0,1]} E_X(R^l_{X,n}, R_{Y,n}, \gamma), \inf_{\gamma\in[0,1]} E_Y(R_{X,n}, R^m_{Y,n}, \gamma)\right\}\right\}. \tag{113}$$

where (113) follows by invoking the definitions of α_q, β_q and t_q in (45), (46), and (47) respectively and the conditions that $\Psi > 2\Omega$, $\Omega \geq T$ and concluding that for $l, m \in [t_{q-1} : \alpha_{q-1}]$,

$$\alpha_{q-1} - t_{q-1} + 1 = \Omega < \Psi, \tag{114}$$

$$\beta_{q-1} - (\alpha_{q-1} + 1) + 1 = \Psi - \Omega < \Psi, \tag{115}$$

$$\Psi - \Omega + 1 \leq \beta_{q-1} - l + 1 \leq \Psi, \tag{116}$$

$$\Psi - \Omega + 1 \leq \beta_{q-1} - m + 1 \leq \Psi. \tag{117}$$

Specifically, using the bound in (114), we can upper bound the number of the summands in the first two sums in (112) by $2\Psi^2$. Furthermore, using (116) and (117), each summand in the first two sums in (112) can be upper bounded by the first term in (113) divided by $2\Psi^2$. Hence, the sum of first two sums in (112) is upper bounded by the the first term in (113). Similarly, using (114) to (117), we can upper bound the sum of last two sums in (112) by the second term in (113).

2) $\mathcal{E}^2_{\mathrm{sw},j}(l,m)$ (recall (76)). We consider two different cases.

a) $l \leq m$

$$\Pr\left(\mathcal{E}^2_{\mathrm{sw},j}(l,m)\right) \leq (n(\beta_{q-1} - l + 1) + 1)^{7|\mathcal{X}||\mathcal{Y}|} \exp\left\{-n(\beta_{q-1} - l + 1) \inf_{\gamma \in [0,1]} E_X(R_{X,n}, R_{Y,n}, \gamma)\right\} \tag{118}$$

b) $m \leq l$

$$\Pr\left(\mathcal{E}^2_{\mathrm{sw},j}(l,m)\right) \leq (n(\beta_{q-1} - m + 1) + 1)^{7|\mathcal{X}||\mathcal{Y}|} \exp\left\{-n(\beta_{q-1} - m + 1) \inf_{\gamma \in [0,1]} E_Y(R_{X,n}, R_{Y,n}, \gamma)\right\} \tag{119}$$

Therefore, for all $n \in \mathbb{N}$,

$$\begin{aligned}
&\Pr\left(\bigcup_{\substack{j \in [\alpha_{q-1}+1 : t_q - 1] \\ l,m \in [j : \beta_{q-1}] \\ \min\{l,m\} = j}} \mathcal{E}^2_{\mathrm{sw},j}(l,m)\right) \\
&= \sum_{j=\alpha_{q-1}+1}^{t_q-1} \left(\sum_{m=j}^{\beta_{q-1}} (n(\beta_{q-1} - j + 1) + 1)^{7|\mathcal{X}||\mathcal{Y}|} \exp\left\{-n(\beta_{q-1} - j + 1) \inf_{\gamma \in [0,1]} E_X(R_{X,n}, R_{Y,n}, \gamma)\right\} \right. \\
&\quad \left. + \sum_{l=j}^{\beta_{q-1}} (n(\beta_{q-1} - j + 1) + 1)^{7|\mathcal{X}||\mathcal{Y}|} \exp\left\{-n(\beta_{q-1} - j + 1) \inf_{\gamma \in [0,1]} E_Y(R_{X,n}, R_{Y,n}, \gamma)\right\} \right) \qquad (120) \\
&\leq 2\Psi^2 (n\Psi + 1)^{7|\mathcal{X}||\mathcal{Y}|} \exp\left\{ - n\Omega \min\left\{ \inf_{\gamma \in [0,1]} E_X(R_{X,n}, R_{Y,n}, \gamma), \inf_{\gamma \in [0,1]} E_Y(R_{X,n}, R_{Y,n}, \gamma)\right\}\right\}, \qquad (121)
\end{aligned}$$

where (121) follows by invoking the defintions of α_q, β_q and t_q in (45), (46) and (47) and concluding that for $j \in [\alpha_{q-1} + 1 : t_q - 1]$ and $l, m \in [j : \beta_{q-1}]$,

$$(t_q - 1) - (\alpha_{q-1} + 1) + 1 \leq t_q - 1 - \alpha_{q-1} = \Psi - 2\Omega + 1 < \Psi, \tag{122}$$

$$\beta_{q-1} - l + 1 \leq \beta_{q-1} - \alpha_{q-1} = \Psi - \Omega < \Psi, \tag{123}$$

$$\beta_{q-1} - m + 1 \leq \beta_{q-1} - \alpha_{q-1} = \Psi - \Omega < \Psi, \tag{124}$$

$$\beta_{q-1} - j + 1 \geq \beta_{q-1} - t_q + 2 = \Omega, \tag{125}$$

$$\beta_{q-1} - j + 1 \leq \beta_{q-1} - (\alpha_{q-1} + 1) + 1 = \Psi - \Omega < \Psi. \tag{126}$$

Specifically, using the bounds in (122), (123) and (124), we can upper bound the number of summands in the two sums in (120) by $2\Psi^2$. Furthermore, using the bounds in (125) and (126), each summand in the sums of (120) can be upper bounded by (121) divided by $2\Psi^2$.

We remark the reasoning for last step in (121) holds similarly in the analyses of other error events for the last steps of the bounding of the error probabilities in the sequel. See (130), (134) and (138) where we omit the reasonings for them for the sake of brevity.

3) $\mathcal{E}^3_{\mathrm{sw},j}(l,m)$ (recall (77)). We consider two different cases.

a) $l \leq m$

$$\Pr\left(\mathcal{E}^3_{\mathrm{sw},j}(l,m)\right) \leq (n(\beta_q - l + 1) + 1)^{7|\mathcal{X}||\mathcal{Y}|} \exp\left\{-n(\beta_q - l + 1) \inf_{\gamma \in [0,1]} E_X(R_{X,n}, R_{Y,n}, \gamma)\right\} \tag{127}$$

b) $m \leq l$

$$\Pr\left(\mathcal{E}^3_{\mathrm{sw},j}(l,m)\right) \leq (n(\beta_q - m + 1) + 1)^{7|\mathcal{X}||\mathcal{Y}|} \exp\left\{-n(\beta_q - m + 1) \inf_{\gamma\in[0,1]} E_Y(R_{X,n}, R_{Y,n}, \gamma)\right\} \tag{128}$$

Therefore, for all $n \in \mathbb{N}$,

$$\begin{aligned}
&\Pr\left(\bigcup_{\substack{j\in[t_q:\beta_{q-1}]\\ l,m\in[j,\beta_q]\\ \min\{l,m\}=j}} \mathcal{E}^3_{\mathrm{sw},j}(l,m)\right)\\
&= \sum_{j=t_q}^{\beta_{q-1}} \left(\sum_{m=j}^{\beta_q} (n(\beta_q - j + 1) + 1)^{7|\mathcal{X}||\mathcal{Y}|} \exp\left\{-n(\beta_q - j + 1) \inf_{\gamma\in[0,1]} E_X(R_{X,n}, R_{Y,n}, \gamma)\right\} \right.\\
&\quad \left. + \sum_{l=j}^{\beta_q} (n(\beta_q - j + 1) + 1)^{7|\mathcal{X}||\mathcal{Y}|} \exp\left\{-n(\beta_q - j + 1) \inf_{\gamma\in[0,1]} E_Y(R_{X,n}, R_{Y,n}, \gamma)\right\} \right) \qquad (129)\\
&\leq 2\Psi^2 (n\Psi + 1)^{7|\mathcal{X}||\mathcal{Y}|} \exp\left\{ - n(\Psi - \Omega + 2) \min\left\{ \inf_{\gamma\in[0,1]} E_X(R_{X,n}, R_{Y,n}, \gamma), \inf_{\gamma\in[0,1]} E_Y(R_{X,n}, R_{Y,n}, \gamma)\right\}\right\}. \qquad (130)
\end{aligned}$$

4) $\mathcal{E}^5_{\mathrm{sw},j}(l,m)$ (recall (78)). We consider two different cases.

a) $l \leq m$

$$\Pr\left(\mathcal{E}^5_{\mathrm{sw},j}(l,m)\right) \leq (n(\beta_q - l + 1) + 1)^{7|\mathcal{X}||\mathcal{Y}|} \exp\left\{-n(\beta_q - l + 1) \inf_{\gamma\in[0,1]} E_X(R_{X,n}, R_{Y,n}, \gamma)\right\} \tag{131}$$

b) $m \leq l$

$$\Pr\left(\mathcal{E}^5_{\mathrm{sw},j}(l,m)\right) \leq (n(\beta_q - m + 1) + 1)^{7|\mathcal{X}||\mathcal{Y}|} \exp\left\{-n(\beta_q - m + 1) \inf_{\gamma\in[0,1]} E_Y(R_{X,n}, R_{Y,n}, \gamma)\right\} \tag{132}$$

Therefore, for all $n \in \mathbb{N}$,

$$\begin{aligned}
&\Pr\left(\bigcup_{\substack{j\in[\alpha_q:t_{q+1}-1]\\ l,m\in[j:\beta_q]\\ \min\{l,m\}=j}} \mathcal{E}^5_{\mathrm{sw},j}(l,m)\right)\\
&= \sum_{j=\alpha_q}^{t_{q+1}-1} \left(\sum_{m=j}^{\beta_q} (n(\beta_q - j + 1) + 1)^{7|\mathcal{X}||\mathcal{Y}|} \exp\left\{-n(\beta_q - j + 1) \inf_{\gamma\in[0,1]} E_X(R_{X,n}, R_{Y,n}, \gamma)\right\} \right.\\
&\quad \left. + \sum_{l=j}^{\beta_q} (n(\beta_q - j + 1) + 1)^{7|\mathcal{X}||\mathcal{Y}|} \exp\left\{-n(\beta_q - j + 1) \inf_{\gamma\in[0,1]} E_Y(R_{X,n}, R_{Y,n}, \gamma)\right\} \right) \qquad (133)\\
&\leq 2\Psi^2 (n\Psi + 1)^{7|\mathcal{X}||\mathcal{Y}|} \exp\left\{ - n\Omega \min\left\{ \inf_{\gamma\in[0,1]} E_X(R_{X,n}, R_{Y,n}, \gamma), \inf_{\gamma\in[0,1]} E_Y(R_{X,n}, R_{Y,n}, \gamma)\right\}\right\}. \qquad (134)
\end{aligned}$$

5) $\mathcal{E}^6_{\mathrm{sw},j}(l,m)$ (recall (79)). We consider two different cases.

a) $l \leq m$

$$\Pr\left(\mathcal{E}^6_{\mathrm{sw},j}(l,m)\right) \leq (n(T_k - l + 1) + 1)^{7|\mathcal{X}||\mathcal{Y}|} \exp\left\{-n(T_k - l + 1) \inf_{\gamma\in[0,1]} E_X(R_{X,n}, R_{Y,n}, \gamma)\right\} \tag{135}$$

b) $m \leq l$

$$\Pr\left(\mathcal{E}^6_{\mathrm{sw},j}(l,m)\right) \leq (n(T_k - m + 1) + 1)^{7|\mathcal{X}||\mathcal{Y}|} \exp\left\{-n(T_k - m + 1) \inf_{\gamma\in[0,1]} E_Y(R_{X,n}, R_{Y,n}, \gamma)\right\} \tag{136}$$

Therefore, for all $n \in \mathbb{N}$,

$$\begin{aligned}
&\Pr\left(\bigcup_{\substack{j\in[t_{q+1}:k]\\ l,m\in[j:T_k]\\ \min\{l,m\}=j}} \mathcal{E}^6_{\mathrm{sw},j}(l,m)\right)\\
&= \sum_{j=t_{q+1}}^{k}\left(\sum_{m=j}^{T_k}(n(T_k-j+1)+1)^{7|\mathcal{X}||\mathcal{Y}|}\exp\left\{-n(T_k-j+1)\inf_{\gamma\in[0,1]}E_X(R_{X,n},R_{Y,n},\gamma)\right\}\right.\\
&\quad\left.+\sum_{l=j}^{T_k}(n(T_k-j+1)+1)^{7|\mathcal{X}||\mathcal{Y}|}\exp\left\{-n(T_k-j+1)\inf_{\gamma\in[0,1]}E_Y(R_{X,n},R_{Y,n},\gamma)\right\}\right) \qquad (137)\\
&\le 2\Psi^2(n\Psi+1)^{7|\mathcal{X}||\mathcal{Y}|}\exp\left\{-nT\min\left\{\inf_{\gamma\in[0,1]}E_X(R_{X,n},R_{Y,n},\gamma),\inf_{\gamma\in[0,1]}E_Y(R_{X,n},R_{Y,n},\gamma)\right\}\right\}. \qquad (138)
\end{aligned}$$

We are now ready to bound the error probability in decoding $(\mathbf{X}_k,\mathbf{Y}_k)$ at time T_k for $k\in\mathcal{S}(q)$ with $q\ge 2$. Recalling Section V-E, we conclude that

$$\begin{aligned}
&\Pr\left((\hat{\mathbf{X}}_k,\hat{\mathbf{Y}}_k)\neq(\mathbf{X}_k,\mathbf{Y}_k)\right)\\
&\le \Pr\left(\bigcup_{\substack{l,m\in[t_{q-1}:\beta_{q-1}]:\\ \min\{l,m\}\le\alpha_{q-1}}}\mathcal{E}^1_{\mathrm{sw}}(l,m)\right)+\Pr\left(\bigcup_{\substack{j\in[\alpha_{q+1}:t_q-1]\\ l,m\in[j:\beta_{q-1}]\\ \min\{l,m\}=j}}\mathcal{E}^2_{\mathrm{sw},j}(l,m)\right)+\Pr\left(\bigcup_{\substack{j\in[t_q:\beta_{q-1}]\\ l,m\in[j:\beta_q]\\ \min\{l,m\}=j}}\mathcal{E}^3_{\mathrm{sw},j}(l,m)\right)\\
&\quad+\Pr\left(\bigcup_{\substack{j\in[\alpha_q:t_{q+1}-1]\\ l,m\in[j:\beta_q]\\ \min\{l,m\}=j}}\mathcal{E}^5_{\mathrm{sw},j}(l,m)\right)+\Pr\left(\bigcup_{\substack{j\in[t_{q+1}:k]\\ l,m\in[j:T_k]\\ \min\{l,m\}=j}}\mathcal{E}^6_{\mathrm{sw},j}(l,m)\right). \qquad (139)
\end{aligned}$$

H. Asymptotic Behavior of the Exponents

We choose the largest and smallest memories to be

$$\Psi = n^{\frac{1}{2}+\delta} \qquad (140)$$

$$\Omega = 2T, \qquad (141)$$

where $\delta\in(0,\frac{1}{2})$. Note these choices of Ψ and Ω satisfy the two conditions $\Psi>2\Omega$ and $\Omega\ge T$ for n large enough since T is a constant.

Define the doubly-indexed sequence

$$\kappa_{n,l} := \left(1-\frac{\Psi-\Omega+1}{\beta_{q-1}-l+1}\right)(R_X^*+\theta_1\xi_n). \qquad (142)$$

Recalling (86) and (88), we obtain

$$\begin{aligned}
R^l_{X,n} &= \frac{\Psi-\Omega+1}{n(\beta_{q-1}-l+1)}\log N_1 && (143)\\
&= R_X^*+\theta_1\xi_n-\left(1-\frac{\Psi-\Omega+1}{\beta_{q-1}-l+1}\right)(R_X^*+\theta_1\xi_n) && (144)\\
&= R_X^*+\theta_1\xi_n-\kappa_{n,l} = R_{X,n}-\kappa_{n,l}. && (145)
\end{aligned}$$

Using the definitions in (45), (46) and (47), we obtain

$$\beta_q-\alpha_q = \Psi-\Omega, \qquad (146)$$

$$\beta_q-t_q = \Psi-1. \qquad (147)$$

Hence, for $l \in [t_{q-1} : \alpha_{q-1}]$, we have

$$\Psi - \Omega + 1 \leq \beta_{q-1} - l + 1 \leq \Psi. \tag{148}$$

Starting from (142), we see that for every $l \in [t_{q-1} : \alpha_{q-1}]$, we obtain

$$\kappa_{n,l} \leq \left(1 - \frac{\Psi - \Omega + 1}{\Psi}\right)(R_X^* + \theta_1 \xi_n) \tag{149}$$

$$\leq \frac{\Omega - 1}{\Psi}(R_X^* + \theta_1 \xi_n) \tag{150}$$

$$= o(\xi_n), \tag{151}$$

where (151) holds by invoking (140) and (141) and the asymptotic conditions on the sequence $\{\xi_n\}_{n\in\mathbb{N}}$ (See Definition 2). We remark that (151) holds for all $l \in [t_{q-1} : \alpha_{q-1}]$, i.e., for all such l and for any $\varepsilon > 0$, there exists $N = N_\varepsilon \in \mathbb{N}$ such that $\kappa_{n,l}/\xi_n < \varepsilon$ for all $n > N$.

Define

$$\zeta_{n,m} := \left(1 - \frac{\Psi - \Omega + 1}{\beta_{q-1} - m + 1}\right)(R_Y^* + \theta_2 \xi_n). \tag{152}$$

Similarly as (145) and (151), we can show that

$$R_{Y,n}^m = R_Y^* + \theta \xi_n - \zeta_{n,m} = R_{Y,n} - \zeta_{n,m}, \tag{153}$$

and for $m \in [t_{q-1} : \alpha_{q-1}]$,

$$\zeta_{n,m} = o(\xi_n). \tag{154}$$

We remark that (154) also holds for all $m \in [t_{q-1} : \alpha_{q-1}]$.

The next lemma presents the asymptotic behavior of the exponents of the error probabilities in (113), (121), (130), (134) and (138).

Lemma 6. *For $l, m \in [t_{q-1} : \alpha_{q-1}]$, we have*

$$\liminf_{n\to\infty} \frac{\min\left\{\inf_{\gamma\in[0,1]} E_X(R_{X,n}^l, R_{Y,n}^m, \gamma), \inf_{\gamma\in[0,1]} E_Y(R_{X,n}^l, R_{Y,n}^m, \gamma)\right\}}{\xi_n^2} \geq L(R_X^*, R_Y^*) \tag{155}$$

$$\liminf_{n\to\infty} \frac{\min\left\{\inf_{\gamma\in[0,1]} E_X(R_{X,n}^l, R_{Y,n}, \gamma), \inf_{\gamma\in[0,1]} E_Y(R_{X,n}, R_{Y,n}^m, \gamma)\right\}}{\xi_n^2} \geq L(R_X^*, R_Y^*) \tag{156}$$

$$\liminf_{n\to\infty} \frac{\min\left\{\inf_{\gamma\in[0,1]} E_X(R_{X,n}, R_{Y,n}, \gamma), \inf_{\gamma\in[0,1]} E_Y(R_{X,n}, R_{Y,n}, \gamma)\right\}}{\xi_n^2} \geq L(R_X^*, R_Y^*), \tag{157}$$

where $L(R_X^*, R_Y^*)$ is defined as follows:

1) *Case (i): $R_X^* = H(P_{X|Y}|P_Y)$ and $R_Y^* > H(P_Y)$*

$$L(R_X^*, R_Y^*) = \frac{\theta_1^2}{2\mathrm{V}(P_{X|Y}|P_Y)}, \tag{158}$$

2) *Case (ii): $R_X^* = H(P_{X|Y}|P_Y)$ and $R_Y^* = H(P_Y)$*

$$L(R_X^*, R_Y^*) = \min\left\{\inf_{\gamma\in[0,1]} \frac{(\theta_1 + (1-\gamma)\theta_2)^2}{2\left(\gamma\mathrm{V}(P_{X|Y}|P_Y) + (1-\gamma)\mathrm{V}(P_{XY})\right)}, \frac{(\theta_1+\theta_2)^2}{2\mathrm{V}(P_{XY})}\right\} \tag{159}$$

3) *Case (iii): $R_X^* + R_Y^* = H(P_{XY})$, $H(P_{X|Y}|P_Y) < R_X^* < H(P_X)$ and $H(P_{Y|X}|P_X) < R_Y^* < H(P_Y)$*

$$L(R_X^*, R_Y^*) = \frac{(\theta_1+\theta_2)^2}{2\mathrm{V}(P_{XY})}, \tag{160}$$

and for Cases (iv) and (v), $L(R_X^*, R_Y^*)$ is defined similarly as Cases (ii) and (i) with X and Y, θ_1 and θ_2 interchanged.

The proof of Lemma 6 is presented in Appendix C. We emphasize that the same lower bounds for the limits in (155)–(157) hold regardless of the specific choices of $l, m \in [t_{q-1} : \alpha_{q-1}]$ due to the estimates $\kappa_{n,l} = o(\xi_n)$ (see (151)) and $\zeta_{n,m} = o(\xi_n)$ (see (154)) and the fact that (151) and (154) hold for all l and m in the specified range.

Using Lemma 6, the bounds in (113), (121), (130), (134) and (138), and the facts that $\Psi = n^{\frac{1}{2}+\delta}$ (cf. (140)) and $\Omega = 2T$ (cf. (141)), we conclude that the upper bound of the error probability in (139) is dominated by the final term in (138). Therefore, we obtain

$$\begin{aligned}
&\liminf_{n\to\infty} -\frac{\log \Pr\left((\hat{\mathbf{X}}_k, \hat{\mathbf{Y}}_k) \neq (\mathbf{X}_k, \mathbf{Y}_k)\right)}{n\xi_n^2} \\
&\geq \liminf_{n\to\infty} T\frac{\min\left\{\inf_{\gamma\in[0,1]} E_X(R_{X,n}, R_{Y,n}, \gamma), \inf_{\gamma\in[0,1]} E_Y(R_{X,n}, R_{Y,n}, \gamma)\right\}}{\xi_n^2} \\
&\quad + \frac{\log 2 + 2\log\Psi}{n\xi_n^2} + \frac{7|\mathcal{X}||\mathcal{Y}|\log(n\Psi+1)}{n\xi_n^2} \qquad (161)\\
&\geq TL(R_X^*, R_Y^*), \qquad (162)
\end{aligned}$$

where (162) holds by invoking Lemma 6 and noting that $\frac{\log n}{n\xi_n^2} \to 0$ as $n \to \infty$.

VI. Proof of Theorem 5

Without loss of generality, we assume $T_1 \leq T_2$. Let $T_1 = T$ and $T_k = k + T - 1$. The encoding strategy remains exactly the same as Section V-B. The decoding differs to some extent. In this setup, for $k \leq T_2 - T$, we need only to decode $\mathbf{X}_k$ (see (42)). However, when $k > T_2 - T$, we need to decode $(\mathbf{X}_k, \mathbf{Y}_{k_2})$. Depending (k, k_2), there are 6 cases listed as follows.

1) $1 \leq k \leq T_2 - T_1$ and $k_2 < 0$
2) $T_2 - T + 1 \leq k \leq \Psi$ and $1 \leq k_2 \leq \Psi + T - T_2$
3) $k \in \mathcal{S}(1)$ and $\Psi + T - T_2 + 1 \leq k_2 \leq \Psi$
4) $k \in \mathcal{S}(1)$ and $k_2 \in \mathcal{S}(1)$
5) $k \in \mathcal{S}(q)$ and $k_2 \in \mathcal{S}(q-1)$ where $q \geq 2$
6) $k \in \mathcal{S}(q)$ and $k_2 \in \mathcal{S}(q)$ where $q \geq 2$.

In the proof, we consider only Cases 1) and 6) since other cases can be analyzed similarly. We remark that for all cases, we also need to consider several sub-cases given the value of T_k, similarly as in Section V-D.

We follow the same notation as in Section V. We assume $\Psi > 2\Omega$, $\Omega = 2T_2$ and we will ensure these two conditions are satisfied. As in Section V-C, we let $\hat{\mathbf{X}}_j$ denote our estimate of $\mathbf{X}_j$ at time T_k. Similarly let $\hat{\mathbf{Y}}_j$ denote our estimate of $\mathbf{Y}_j$ at time T_k. We suppress the dependency on T_k for simplicity.

A. Decoding

1) $k \leq T_2 - T$
Given Ψ and Ω, we obtain $k \leq \Omega$ and $T_k = k + T - 1 < \Psi$. For this case, instead of decoding $\mathbf{X}_k$ only, we decode $\mathbf{X}_1, \ldots, \mathbf{X}_k$ sequentially. Our decoding strategy is as follows.
After receiving T_k-th codeword pair, given $\hat{\mathbf{X}}^{j-1}$, we decode $\hat{\mathbf{X}}_j = \tilde{\mathbf{x}}_j$ if there exists $(\tilde{\mathbf{x}}^{T_k}, \tilde{\mathbf{y}}^{T_k}) \in \mathcal{A}(\mathbf{X}^{T_k}, \mathbf{Y}^{T_k}, j)$ (recall (52)) satisfying $\tilde{\mathbf{x}}^{j-1} = \hat{\mathbf{X}}^{j-1}$ such that for all $(\bar{\mathbf{x}}^{T_k}, \bar{\mathbf{y}}^{T_k}) \in \mathcal{A}(\mathbf{X}^{T_k}, \mathbf{Y}^{T_k}, j)$ satisfying $\bar{\mathbf{x}}^{j-1} = \hat{\mathbf{X}}^{j-1}$, we have

$$\hat{H}_{\mathrm{S}}(l, m, \tilde{\mathbf{x}}^{T_k}, \tilde{\mathbf{y}}^{T_k}) \leq \hat{H}_{\mathrm{S}}(l, m, \bar{\mathbf{x}}^{T_k}, \bar{\mathbf{y}}^{T_k}). \qquad (163)$$

2) $k \in \mathcal{S}(q)$ and $k_2 \in \mathcal{S}(q-1)$ for some $q \geq 2$
There are two sub-cases, i.e., $T_k > \beta_q$ or $T_k \leq \beta_q$. We consider $T_k > \beta_q$ here only.
Instead of decoding $(\mathbf{X}_k, \mathbf{Y}_{k_2})$ only, we decode $\mathbf{X}_{t_{q-1}}, \ldots, \mathbf{X}_k$ and $\mathbf{Y}_{t_{q-1}}, \ldots, \mathbf{Y}_{k_2}$. The decoding strategy is as follows. We first jointly decode $(\mathbf{X}_{t_{q-1}}^{\alpha_{q-1}}, \mathbf{Y}_{T_{q-1}}^{\alpha_{q-1}})$, which is done exactly the same as 3a) in Section V-D. Then we decode $(\mathbf{X}_j, \mathbf{Y}_j)$ where $j \in [\alpha_{q-1}+1 : k_2]$ sequentially, which is done exactly the same as 3b) in Section V-D. The last part is to decode $\mathbf{X}_{k_2+1}, \ldots, \mathbf{X}_k$ sequentially.
Note that

$$k_2 = k + T - T_2 > \alpha_q + T - \Omega \geq (\alpha_q - \Omega + 1) = t_q. \qquad (164)$$

We are now ready to describe the decoding strategy for $\mathbf{X}_j$ depending on the index $j \in [k_2 + 1 : k]$.

a) $k_2 + 1 \leq j \leq \beta_{q-1}$
Given $(\hat{\mathbf{X}}_{t_{q-1}}^{j-1}, \hat{\mathbf{Y}}_{t_{q-1}}^{k_2})$, we decode $\hat{\mathbf{X}}_j = \tilde{\mathbf{x}}_j$ if there exists $(\tilde{\mathbf{x}}_{t_{q-1}}^{\beta_q}, \tilde{\mathbf{y}}_{t_{q-1}}^{\beta_q}) \in \mathcal{B}_3(\mathbf{X}_{t_{q-1}}^{\beta_q}, \mathbf{Y}_{t_{q-1}}^{\beta_q})$ (recall (66)) satisfying $(\tilde{\mathbf{x}}_{t_{q-1}}^{j-1}, \tilde{\mathbf{y}}_{t_{q-1}}^{j-1}) = (\hat{\mathbf{X}}_{t_{q-1}}^{j-1}, \hat{\mathbf{Y}}_{t_{q-1}}^{k_2})$ such that for all $(\bar{\mathbf{x}}_{t_{q-1}}^{\beta_q}, \bar{\mathbf{y}}_{t_{q-1}}^{\beta_q}) \in \mathcal{B}_3(\mathbf{X}_{t_{q-1}}^{\beta_q}, \mathbf{Y}_{t_{q-1}}^{\beta_q})$ where $(\bar{\mathbf{x}}_{t_{q-1}}^{j-1}, \bar{\mathbf{y}}_{t_{q-1}}^{k_2}) = (\hat{\mathbf{X}}_{t_{q-1}}^{j-1}, \hat{\mathbf{Y}}_{t_{q-1}}^{k_2})$, we have

$$\hat{H}_{\mathrm{S}}(l, m, \tilde{\mathbf{x}}_{t_{q-1}}^{\lambda_k}, \tilde{\mathbf{y}}_{t_{q-1}}^{\lambda_k}) \leq \hat{H}_{\mathrm{S}}(l, m, \bar{\mathbf{x}}_{t_{q-1}}^{\lambda_k}, \bar{\mathbf{y}}_{t_{q-1}}^{\lambda_k}). \qquad (165)$$

b) $\alpha_q \leq j \leq t_{q+1} - 1$

Given $(\hat{\mathbf{X}}_{t_{q-1}}^{j-1}, \hat{\mathbf{Y}}_{t_{q-1}}^{k_2})$, we decode $\hat{\mathbf{X}}_j = \tilde{\mathbf{x}}_j$ if there exists $(\tilde{\mathbf{x}}_{t_q}^{\beta_q}, \tilde{\mathbf{y}}_{t_q}^{\beta_q}) \in \mathcal{B}_5(\mathbf{X}_{t_q}^{\beta_q}, \mathbf{Y}_{t_q}^{\beta_q}, j)$ (recall (70)) satisfying $(\tilde{\mathbf{x}}_{t_q}^{j-1}, \tilde{\mathbf{y}}_{t_q}^{k_2}) = (\hat{\mathbf{X}}_{t_q}^{j-1}, \hat{\mathbf{Y}}_{t_q}^{k_2})$ such that for all $(\bar{\mathbf{x}}_{t_q}^{\beta_q}, \bar{\mathbf{y}}_{t_q}^{\beta_q}) \in \mathcal{B}_5(\mathbf{X}_{t_q}^{\beta_q}, \mathbf{Y}_{t_q}^{\beta_q}, j)$ where $(\bar{\mathbf{x}}_{t_q}^{j-1}, \bar{\mathbf{y}}_{t_q}^{k_2}) = (\hat{\mathbf{X}}_{t_q}^{j-1}, \hat{\mathbf{Y}}_{t_q}^{k_2})$, we have

$$\hat{H}_{\mathrm{S}}(l, m, \tilde{\mathbf{x}}_{t_q}^{\beta_q}, \tilde{\mathbf{y}}_{t_q}^{\beta_q}) \leq \hat{H}_{\mathrm{S}}(l, m, \bar{\mathbf{x}}_{t_q}^{\beta_q}, \bar{\mathbf{y}}_{t_q}^{\beta_q}). \tag{166}$$

c) $t_q \leq j \leq k$
Given $(\hat{\mathbf{X}}_{t_{q-1}}^{j-1}, \hat{\mathbf{Y}}_{t_{q-1}}^{j-1})$, we decode $\hat{\mathbf{X}}_j = (\tilde{\mathbf{x}}_j$ if there exists $(\tilde{\mathbf{x}}_{t_q}^{T_k}, \tilde{\mathbf{y}}_{t_q}^{T_k}) \in \mathcal{B}_6(\mathbf{X}_{t_q}^{T_k}, \mathbf{Y}_{t_q}^{T_k}, j)$ (recall (72)) satisfying $(\tilde{\mathbf{x}}_{t_q}^{j-1}, \tilde{\mathbf{y}}_{t_q}^{k_2}) = (\hat{\mathbf{X}}_{t_q}^{j-1}, \hat{\mathbf{Y}}_{t_q}^{k_2})$ such that for all $(\bar{\mathbf{x}}_{t_q}^{T_k}, \bar{\mathbf{y}}_{t_q}^{T_k}) \in \mathcal{B}_6(\mathbf{X}_{t_q}^{T_k}, \mathbf{Y}_{t_q}^{T_k}, j)$ where $(\bar{\mathbf{x}}_{t_q}^{j-1}, \bar{\mathbf{y}}_{t_q}^{k_2}) = (\hat{\mathbf{X}}_{t_q}^{j-1}, \hat{\mathbf{Y}}_{t_q}^{k_2})$, we have

$$\hat{H}_{\mathrm{S}}(l, m, \tilde{\mathbf{x}}_{t_q}^{T_k}, \tilde{\mathbf{y}}_{t_q}^{T_k}) \leq \hat{H}_{\mathrm{S}}(l, m, \bar{\mathbf{x}}_{t_q}^{T_k}, \bar{\mathbf{y}}_{t_q}^{T_k}). \tag{167}$$

B. Error Events

In this subsection, we analyze the error probability for cases discussed in Section VI-A.

1) $k \in [1 : T_2 - T]$.
The error in decoding $\mathbf{X}_k$ occurs if and only if one of the following events occur: for $j \in [1 : k]$, $l, m \in [j : T_k]$ and $\min\{l, m\} = j$,

$$\mathcal{E}_{T,T_2,j}(l, m) := \left\{ \exists (\tilde{\mathbf{x}}^{T_k}, \tilde{\mathbf{y}}^{T_k}) \in \mathcal{A}(\mathbf{X}^{T_k}, \mathbf{Y}^{T_k}, j) \bigcap \mathcal{F}(l, m, \mathbf{X}^{T_k}, \mathbf{Y}^{T_k}) : \hat{H}_{\mathrm{S}}(l, m, \tilde{\mathbf{x}}^{T_k}, \tilde{\mathbf{y}}^{T_k}) \leq \hat{H}_{\mathrm{S}}(l, m, \bar{\mathbf{x}}^{T_k}, \bar{\mathbf{y}}^{T_k}) \right\}. \tag{168}$$

2) $k \in \mathcal{S}(q)$, $k_2 \in \mathcal{S}(q-1)$ and $T_k > \beta_q$ for some $q \geq 2$.
The error in decoding of $(\mathbf{X}_K, \mathbf{Y}_{k_2})$ occurs if and only if one of the following events occur:
   a) Error in joint decoding of $(\mathbf{X}_{t_{q-1}}^{\alpha_{q-1}}, \mathbf{Y}_{T_{q-1}}^{\alpha_{q-1}})$.
   For $l, m \in [t_{q-1} : \beta_{q-1}]$ and $\min\{l, m\} \leq \alpha_{q-1}$: $\mathcal{E}_{\mathrm{sw}}^1(l, m)$ (recall (75)).
   b) Error in sequential decoding of $(\mathbf{X}_j, \mathbf{Y}_j)$ where $j \in [\alpha_{q-1} + 1 : t_q - 1]$.
   For each j, for $l, m \in [j : \beta_{q-1}]$ and $\min\{l, m\} = j$: $\mathcal{E}_{\mathrm{sw},j}^2(l, m)$ (recall (76)).
   c) Error in sequential decoding of $(\mathbf{X}_j, \mathbf{Y}_j)$ where $j \in [t_q : k_2]$.
   For each j, for $l, m \in [j : \beta_q]$ and $\min\{l, m\} = j$: $\mathcal{E}_{\mathrm{sw},j}^3(l, m)$ (recall (77)).
   d) Error in sequential decoding of $\mathbf{X}_j$ where $j \in [k_2 + 1 : \beta_{q-1}]$.
   For each j, $l = j$ and $m \in [k_2 + 1 : \beta_q]$: $\mathcal{E}_{\mathrm{sw},j}^3(l, m)$ (recall (77)).
   e) Error in sequential decoding of $\mathbf{X}_j$ where $j \in [\alpha_q : t_{q+1} - 1]$.
   For each j, $l = j$, $m \in [k_2 + 1 : \beta_q]$: $\mathcal{E}_{\mathrm{sw},j}^5(l, m)$ (recall (78)).
   f) Error in sequential decoding of $\mathbf{X}_j$ where $j \in [t_{q+1} : k]$.
   For each j, $l = j$, $m \in [k_2 + 1 : T_k]$: $\mathcal{E}_{\mathrm{sw},j}^6(l, m)$ (recall (79)).

C. Evaluation of the Error Probabilities

The values for Ψ, Ω and the definitions of $(R_{X,n}, R_{Y,n})$ and $(R_{X,n}^l, R_{Y,n}^m)$ are exactly the same as Section V-G.
In this subsection, we analyze the error probability for events considered in Section VI-B.

1) $k \in [1 : T_2 - T]$.
Similarly as (108), we can prove for $l \leq m$

$$\Pr\left(\mathcal{E}_{T,T_2,j}(l, m)\right) \leq \left(n(T_k - l + 1) + 1\right)^{7|\mathcal{X}||\mathcal{Y}|} \exp\left\{ -n(T_k - l + 1) \inf_{\gamma \in [0,1]} E_X(R_{X,n}, R_{Y,n}, \gamma) \right\}, \tag{169}$$

and for $l \geq m$,

$$\Pr\left(\mathcal{E}_{T,T_2,j}(l, m)\right) \leq \left(n(T_k - m + 1) + 1\right)^{7|\mathcal{X}||\mathcal{Y}|} \exp\left\{ -n(T_k - l + 1) \inf_{\gamma \in [0,1]} E_Y(R_{X,n}, R_{Y,n}, \gamma) \right\}. \tag{170}$$

Hence, for all $n \in \mathbb{N}$, we obtain

$$\begin{aligned}
&\Pr\left((\hat{\mathbf{X}}_k, \hat{\mathbf{Y}}_{k_2}) \neq (\mathbf{X}_k, \mathbf{Y}_{k_2})\right) \\
&= \Pr\left(\hat{\mathbf{X}}_k \neq \mathbf{X}_k\right) && (171)\\
&= \Pr\left(\bigcup_{\substack{j\in[1:k]\\ l,m\in[j:T_k]\\ \min\{l,m\}=j}} \mathcal{E}_{T,T_2,j}(l,m)\right) && (172)\\
&\leq \sum_{j=1}^{k}\left(\sum_{m=j}^{T_k}\Pr\left(\mathcal{E}_{T,T_2,j}(j,m)\right)+\sum_{l=j}^{T_k}\Pr\left(\mathcal{E}_{T,T_2,j}(l,j)\right)\right) && (173)\\
&\leq \sum_{j=1}^{k}\left(\sum_{m=j}^{T_k}\left(n(T_k-j+1)+1\right)^{7|\mathcal{X}||\mathcal{Y}|}\exp\left\{-n(T_k-j+1)\inf_{\gamma\in[0,1]}E_X(R_{X,n},R_{Y,n},\gamma)\right\}\right.\\
&\quad\left.+\sum_{l=j}^{T_k}\left(n(T_k-j+1)+1\right)^{7|\mathcal{X}||\mathcal{Y}|}\exp\left\{-n(T_k-j+1)\inf_{\gamma\in[0,1]}E_X(R_{X,n},R_{Y,n},\gamma)\right\}\right) && (174)\\
&\leq 2\Psi^2(n\Psi+1)^{7|\mathcal{X}||\mathcal{Y}|}\exp\left\{-nT\min\left\{\inf_{\gamma\in[0,1]}E_X(R_{X,n},R_{Y,n},\gamma),\inf_{\gamma\in[0,1]}E_Y(R_{X,n},R_{Y,n},\gamma)\right\}\right\}, && (175)
\end{aligned}$$

where (171) follows from (42), (174) follows from the bounds in (169) and (170), and (175) holds by invoking the definition of $T_k = k + T - 1$, the conditions that $k \leq T_2 - T$, $\Omega = 2T_2 < \Psi$ and finally concluding that

$$\begin{aligned}
T_k - j + 1 &\geq T_k - k + 1 = T && (176)\\
T_k - j + 1 &\leq T_k - 1 + 1 \leq k + T - 1 \leq T_2 - 1 < \Psi. && (177)
\end{aligned}$$

Now by choosing $\Psi = n^{\frac{1}{2}+\delta}$ and $\Omega = 2T_2$, we can ensure that the condition $\Psi > 2\Omega$ holds for n large enough. Invoking Lemma 6 similarly as (162), we obtain

$$\liminf_{n\to\infty} -\frac{\log\Pr\left((\hat{\mathbf{X}}_k, \hat{\mathbf{Y}}_{k_2}) \neq (\mathbf{X}_k, \mathbf{Y}_{k_2})\right)}{n\xi_n^2} \geq TL(R_X^*, R_Y^*). \tag{178}$$

2) $k \in \mathcal{S}(q)$, $k_2 \in \mathcal{S}(q-1)$ and $T_k > \beta_q$
Similarly as Section V-G, we obtain

$$\liminf_{n\to\infty} -\frac{\log\Pr\left((\hat{\mathbf{X}}_k, \hat{\mathbf{Y}}_{k_2}) \neq (\mathbf{X}_k, \mathbf{Y}_{k_2})\right)}{n\xi_n^2} \geq TL(R_X^*, R_Y^*). \tag{179}$$

VII. Conclusion

In this paper, we have derived an achievable moderate deviations constant for the blockwise streaming version of SW coding. We showed that the moderate deviations constant is enhanced by a multiplicative factor of T (over the non-streaming setting) in many instances.

In future, one may consider deriving matching converse results using concept of a feedforward decoder as demonstrated in [40] and [18]. Other possible research topics may include streaming lossy source coding [6, Chapter 3] and streaming versions of other multi-terminal coding problems [41, Part II].

Appendix

A. Proof of Proposition 3

Note that for $\boldsymbol{\theta} \in \Theta_{(\text{ii})}$ (cf. (6)), $\theta_1 > 0$ and $\theta_1 + \theta_2 > 0$. The conditions on $\boldsymbol{\theta}$ are essentially only concerning the ratio of θ_1 and θ_2. For ease of notation, we denote $f(P_{XY}, \gamma, \boldsymbol{\theta})$ as $f(P_{XY}, \gamma)$ to suppress the dependency on $\boldsymbol{\theta}$. To further simplify notation, we also use $\mathrm{V_c} := \mathrm{V}(P_{X|Y}|P_Y)$ and $\mathrm{V_j} := \mathrm{V}(P_{XY})$ to denote the conditional and joint varentropies.

The first and second derivatives of $f(P_{XY}, \gamma)$ with respect to γ are

$$\begin{aligned}
f'(P_{XY},\gamma) &= \frac{-(\mathrm{V_c}-\mathrm{V_j})(\theta_1+(1-\gamma)\theta_2)^2}{2\left((\gamma\mathrm{V_c}+(1-\gamma)\mathrm{V_j})\right)^2} - \frac{\theta_2(\theta_1+(1-\gamma)\theta_2)}{(\gamma\mathrm{V_c}+(1-\gamma)\mathrm{V_j})} && (180)\\
f''(P_{XY},\gamma) &= \frac{\left((\theta_1+\theta_2)\mathrm{V_c}-\theta_1\mathrm{V_j}\right)^2}{\left((\gamma\mathrm{V_c}+(1-\gamma)\mathrm{V_j})\right)^3}. && (181)
\end{aligned}$$

Hence, $f(P_{XY},\gamma)$ is a convex function in γ since $f''(P_{XY},\gamma)\geq 0$.

Define

$$\gamma^* = \arg\min_{\gamma\in[0,1]} f(P_{XY},\gamma). \tag{182}$$

We first prove that (23) and (24) are sufficient conditions by considering the scenarios where $\gamma^*=1$ and $\gamma^*=0$. Then we prove that (23) and (24) are also necessary by considering the scenario where $\gamma^*\in(0,1)$.

1) $\gamma^*=1$

In order to achieve the infimum at $\gamma^*=1$, we need $f'(\gamma^*)\leq 0$, i.e.,

$$f'(P_{XY},1) = \frac{-(\mathrm{V_c}-\mathrm{V_j})\theta_1^2}{2\mathrm{V_c^2}} - \frac{\theta_1\theta_2}{\mathrm{V_c}} \leq 0 \tag{183}$$

Hence, we have

$$\frac{\theta_2}{\theta_1} \geq \frac{\mathrm{V_j}-\mathrm{V_c}}{2\mathrm{V_c}}. \tag{184}$$

Note that

$$f(P_{XY},1) = \frac{\theta_1^2}{2\mathrm{V_c}}. \tag{185}$$

Hence, we have proved (22) holds when (23) is satisfied.

2) $\gamma^*=0$

In order to achieve the infimum at $\gamma^*=0$, we need $f'(\gamma^*)\geq 0$, i.e.,

$$f'(P_{XY},0) = \frac{-(\mathrm{V_c}-\mathrm{V_j})(\theta_1+\theta_2)^2}{2\mathrm{V_j^2}} - \frac{\theta_2(\theta_1+\theta_2)}{\mathrm{V_j}} \geq 0. \tag{186}$$

Note that $\theta_1+\theta_2>0$ and $\mathrm{V_j}+\mathrm{V_c}>0$. After performing some algebra, we obtain

$$\frac{\theta_2}{\theta_1} \leq \frac{\mathrm{V_j}-\mathrm{V_c}}{\mathrm{V_j}+\mathrm{V_c}} \tag{187}$$

Further note that

$$f(P_{XY},0) = \frac{(\theta_1+\theta_2)^2}{2\mathrm{V_j}}. \tag{188}$$

Hence, in order to for (22) to hold, invoking (12), we obtain

$$\frac{(\theta_1+\theta_2)^2}{2\mathrm{V_j}} \leq \frac{\theta_1^2}{2\mathrm{V_c}}. \tag{189}$$

Invoking the constraint that $\theta_2>-\theta_1$ and $\theta_1>0$ (refer to the definition of $\Theta_{(\mathrm{ii})}$ in (6)), we obtain

$$-1 < \frac{\theta_2}{\theta_1} \leq \sqrt{\frac{\mathrm{V_j}}{\mathrm{V_c}}} - 1. \tag{190}$$

Combining (187) and (190), we have proved that (22) holds when (24) holds.

3) $\gamma^*\in(0,1)$.

For this case, in order for (22) to hold, we need (189) and

$$f(P_{XY},\gamma^*) \geq \frac{(\theta_1+\theta_2)^2}{2\mathrm{V_j}}. \tag{191}$$

Since γ^* minimizes $f(\gamma)$, we obtain

$$f'(P_{XY},\gamma^*) = \frac{-(\mathrm{V_c}-\mathrm{V_j})(\theta_1+(1-\gamma^*)\theta_2)^2}{2\left((\gamma^*\mathrm{V_c}+(1-\gamma^*)\mathrm{V_j})\right)^2} - \frac{\theta_2(\theta_1+(1-\gamma^*)\theta_2}{(\gamma^*\mathrm{V_c}+(1-\gamma^*)\mathrm{V_j})} = 0. \tag{192}$$

Solving (192), we obtain

$$\gamma^* = \frac{(\mathrm{V_c}+\mathrm{V_j})\theta_2 - (\mathrm{V_j}-\mathrm{V_c})\theta_1}{\theta_2\left(\mathrm{V_j}-\mathrm{V_c}\right)} \tag{193}$$

$$= -\frac{\theta_1}{\theta_2} + \frac{\mathrm{V_c}+\mathrm{V_j}}{\mathrm{V_j}-\mathrm{V_c}} \tag{194}$$

Therefore, we obtain

$$f(P_{XY},\gamma^*)=\frac{2\theta_2\left(\theta_1\mathrm{V_j}-(\theta_1+\theta_2)\mathrm{V_c}\right)}{\left(\mathrm{V_j}-\mathrm{V_c}\right)^2}. \tag{195}$$

Since (184) and (187) are, respectively, the conditions on θ_2/θ_1 that ensure that $\gamma^*=0$ and $\gamma^*=1$, when $\gamma^*\in(0,1)$, θ_2/θ_1 cannot satisfy either of (184) or (187), we conclude that $\gamma^*\in(0,1)$ implies

$$\frac{\mathrm{V_j}-\mathrm{V_c}}{\mathrm{V_j}+\mathrm{V_c}}<\frac{\theta_2}{\theta_1}<\frac{\mathrm{V_j}-\mathrm{V_c}}{2\mathrm{V_c}}. \tag{196}$$

Hence, $f(P_{XY},\gamma^*)>0$. In order to satisfy (191), we have

$$\frac{2\theta_2\left(\theta_1\mathrm{V_j}-(\theta_1+\theta_2)\mathrm{V_c}\right)}{\left(\mathrm{V_j}-\mathrm{V_c}\right)^2}\geq\frac{(\theta_1+\theta_2)^2}{2\mathrm{V_j}}. \tag{197}$$

Solving (197), we obtain

$$-\left(\theta_1\left(\mathrm{V_c}-\mathrm{V_j}\right)+\theta_2\left(\mathrm{V_c}+\mathrm{V_j}\right)\right)^2\geq 0. \tag{198}$$

Thus, the only possible case is

$$\frac{\theta_2}{\theta_1}=\frac{\mathrm{V_j}-\mathrm{V_c}}{\mathrm{V_c}+\mathrm{V_j}} \tag{199}$$

However, using the condition for $\gamma^*\in(0,1)$ given in (196), we conclude that it is impossible to satisfy (191) if $\gamma^*\in(0,1)$.

Therefore, we have proved that (23) and (24) are both sufficient and necessary conditions for (22) to hold.

B. Preliminaries for the Proof of Lemma 6

Define

$$E_{XY}(\rho)=(1+\rho)\log\sum_{x,y}P_{XY}(x,y)^{\frac{1}{1+\rho}}, \tag{200}$$

$$E_{X|Y}(\rho)=\log\sum_{y}P_Y(y)\Big(\sum_{x}P_{X|Y}(x|y)^{\frac{1}{1+\rho}}\Big)^{1+\rho}, \tag{201}$$

$$E_X(\gamma,\rho)=\gamma E_{X|Y}(\rho)+(1-\gamma)E_{XY}(\rho), \tag{202}$$

and $E_{Y|X}(\rho)$ is similar to $E_{X|Y}(\rho)$ with X and Y interchanged; $E_Y(\gamma,\rho)$ is similar to $E_X(\gamma,\rho)$ with X and Y interchanged.

Lemma 7. *For any pmf $P_{XY}\in\mathcal{P}(\mathcal{X}\times\mathcal{Y})$ where $\mathcal{X}$ and $\mathcal{Y}$ are finite sets,*

$$E'_{XY}(\rho)\Big|_{\rho=0}=H(P_{XY}), \tag{203}$$

$$E''_{XY}(\rho)\Big|_{\rho=0}=\mathrm{V}(P_{XY}), \tag{204}$$

$$\inf_{\rho\in[0,1]}E''_{XY}(\rho)\geq 0. \tag{205}$$

Furthermore, there exists a finite positive number M_{XY} such that

$$\sup_{\rho\in[0,1]}|E'''_{XY}(\rho)|\leq M_{XY}. \tag{206}$$

Lemma 8. *For any pmf $P_{XY}\in\mathcal{P}(\mathcal{X}\times\mathcal{Y})$ where $\mathcal{X}$ and $\mathcal{Y}$ are finite sets,*

$$E'_{X|Y}(\rho)\Big|_{\rho=0}=H(P_{X|Y}|P_Y), \tag{207}$$

$$E''_{X|Y}(\rho)\Big|_{\rho=0}=\mathrm{V}(P_{X|Y}|P_Y), \tag{208}$$

$$\inf_{\rho\in[0,1]}E''_{X|Y}(\rho)\geq 0. \tag{209}$$

Furthermore, there exists a finite positive number M_X such that

$$\sup_{\rho\in[0,1]}\left|E'''_{X|Y}(\rho)\right|\leq M_X. \tag{210}$$

The derivatives and properties of $E_{Y|X}(\rho)$ are exactly the same as $E_{X|Y}(\rho)$ with X and Y interchanged.

The proofs of Lemmas 7 and 8 are standard and thus omitted. See for example [7] and [13].

Invoking Lemmas 7 and 8, the Taylor series expansions of (202) is as follows:

$$E_X(\gamma,\rho) = \gamma\rho H(P_{X|Y}|P_Y) + (1-\gamma)\rho H(P_{XY}) + \frac{\gamma\rho^2}{2}\mathrm{V}(P_{X|Y}|P_Y) + \frac{(1-\gamma)\rho^2}{2}\mathrm{V}(P_{XY}) + \frac{\gamma\rho^3}{6}E'''_{X|Y}(\overline{\rho}) + \frac{(1-\gamma)\rho^3}{6}E'''_{XY}(\overline{\rho}), \tag{211}$$

for some $\overline{\rho}\in[0,\rho]$. The Taylor expansion of $E_Y(\gamma,\rho)$ is obtained by interchanging X and Y in (211).

Lemma 9. *Let $f:[0,1]\to\mathbb{R}$, $g:[0,1]\to\mathbb{R}$ and $g_i:[0,1]\to\mathbb{R}$ where $i\in[1:m]$ be continuous functions. Consider arbitrary positive sequences $\{a_n\}_{n=1}^\infty$, $\{b_n\}_{n=1}^\infty$ $\{c_{i,n}\}_{n=1}^\infty$ satisfying $b_n=o(a_n)$ and $c_{i,n}=o(b_n)$. Then we have*

$$\inf_{\gamma\in[0,1]}\left(a_nf(\gamma)+b_ng(\gamma)+\sum_{i=1}^m c_{i,n}g_i(\gamma)\right) = a_nf^*+b_ng^*+o(b_n), \tag{212}$$

where

$$f^* = \inf_{\gamma\in[0,1]} f(\gamma) \tag{213}$$

$$g^* = \inf_{\gamma:f(\gamma)=f^*} g(\gamma). \tag{214}$$

Lemma 9 can be proved similarly as [31, Lemma 48].

C. Proof of Lemma 6

We prove (155) only since others can be done similarly. Let (R_X^*,R_Y^*) be fixed boundary point of the SW region (cf. Cases (i)-(v) in Figure 2). We consider an arbitrary sequence $\{\xi_n\}_{n=1}^\infty$ satisfying $\xi_n\to0$ and $\frac{\log n}{n\xi_n^2}\to0$ as $n\to\infty$.

Recall the definition of $R_{X,n}^l$, $R_{Y,n}^m$, $\kappa_{n,l}$ and $\zeta_{n,m}$ in (88), (89), (142) and (152) respectively.

Define

$$\rho_n = \frac{(\theta_1+(1-\gamma)\theta_2)\xi_n}{\gamma\mathrm{V}(P_{X|Y}|P_Y)+(1-\gamma)\mathrm{V}(P_{XY})}. \tag{215}$$

Note that for n large enough, $\rho_n\in[0,1]$. Hence, for any $\gamma\in[0,1]$,

$$\begin{aligned}
&E_X(R_{X,n}^l,R_{Y,n}^m,\gamma)\\
&= \max_{\rho\in[0,1]}\gamma\rho R_{X,n}^l+(1-\gamma)\rho(R_{X,n}^l+R_{Y,n}^m)-E_X(\gamma,\rho) &(216)\\
&\ge \gamma\rho_nR_{X_n}^l-\gamma\left(\rho_nH(P_{X|Y}|P_Y)+\frac{\rho_n^2}{2}\mathrm{V}(P_{X|Y}|P_Y)+\frac{\rho_n^3}{6}E'''_{X|Y}(\overline{\rho}_n)\right) &(217)\\
&\quad+(1-\gamma)\rho_n(R_{X,n}^l+R_{Y,n}^m)-(1-\gamma)\left(\rho_nH(P_{XY})+\frac{\rho_n^2}{2}\mathrm{V}(P_{XY})+\frac{\rho_n^3}{6}E'''_{XY}(\overline{\rho}_n)\right) &(218)\\
&= \rho_n\left(\gamma R_{X,n}^l+(1-\gamma)(R_{X,n}^l+R_{Y,n}^m)-\gamma H(P_{X|Y}|P_Y)-(1-\gamma)H(P_{XY})\right)\\
&\quad-\gamma\left(\frac{\rho_n^2}{2}\mathrm{V}(P_{X|Y}|P_Y)+\frac{\rho_n^3}{6}M_X\right)-(1-\gamma)\left(\frac{\rho_n^2}{2}\mathrm{V}(P_{XY})+\frac{\rho_n^3}{6}M_{XY}\right) &(219)\\
&= \rho_n\left(\gamma R_X^*+(1-\gamma)(R_X^*+R_Y^*)-\gamma H(P_{X|Y}|P_Y)-(1-\gamma)H(P_{XY})\right)-\rho_n(\theta_1\kappa_{n,l}+(1-\gamma)\theta_2\zeta_{n,m})\\
&\quad+\left(\rho_n(\theta_1+(1-\gamma)\theta_2)\xi_n-\rho_n^2\left(\frac{\gamma\mathrm{V}(P_{X|Y}|P_Y)+(1-\gamma)\mathrm{V}(P_{XY})}{2}\right)\right)-\frac{\rho_n^3(\gamma M_X+(1-\gamma)M_{XY})}{6} &(220)\\
&= \rho_n\left(\gamma R_X^*+(1-\gamma)(R_X^*+R_Y^*)-\gamma H(P_{X|Y}|P_Y)-(1-\gamma)H(P_{XY})\right)-\theta_1\rho_n\kappa_{n,l}-(1-\gamma)\theta_2\rho_n\zeta_{n,m}\\
&\quad+\frac{(\theta_1+(1-\gamma)\theta_2)^2\xi_n^2}{2\left(\gamma\mathrm{V}(P_{X|Y}|P_Y)+(1-\gamma)\mathrm{V}(P_{XY})\right)}-\frac{\rho_n^3(\gamma M_X+(1-\gamma)M_{XY})}{6} &(221)\\
&\ge \rho_n\left(\gamma R_X^*+(1-\gamma)(R_X^*+R_Y^*)-\gamma H(P_{X|Y}|P_Y)-(1-\gamma)H(P_{XY})\right)-\theta_1\rho_n\kappa_{n,l}-(1-\gamma)\theta_2\rho_n\zeta_{n,m}\\
&\quad+\frac{(\theta_1+(1-\gamma)\theta_2)^2\xi_n^2}{2\left(\gamma\mathrm{V}(P_{X|Y}|P_Y)+(1-\gamma)\mathrm{V}(P_{XY})\right)}-\frac{\rho_n^3}{6}\max\{M_{XY},M_X\}, &(222)
\end{aligned}$$

where (218) follows from (211) and $\bar{\rho}_n\in[0,\rho_n]$.

Define

$$\rho'_n = \frac{((1-\gamma)\theta_1+\theta_2)\xi_n}{\gamma\mathrm{V}(P_{Y|X}|P_X)+(1-\gamma)\mathrm{V}(P_{XY})}. \tag{223}$$

Similarly, we obtain

$$E_Y(R^l_{X,n},R^m_{Y,n},\gamma)\geq \rho'_n\left(\gamma R^*_Y+(1-\gamma)(R^*_X+R^*_Y)-\gamma H(P_{Y|X}|P_X)-(1-\gamma)H(P_{XY})\right)-(1-\gamma)\theta_1\rho'_n\kappa_{n,l}-\theta_2\rho'_n\zeta_{n,m}$$
$$+\frac{((1-\gamma)\theta_1+\theta_2)^2\xi_n^2}{2\left(\gamma \mathrm{V}(P_{Y|X}|P_X)+(1-\gamma)\mathrm{V}(P_{XY})\right)}-\frac{\rho'^3_n}{6}\max\left\{M_{XY},M_Y\right\}. \tag{224}$$

We then deal with different cases. Here we only prove the result for Cases (i)-(iii) because Case (iv) is symmetric to Case (ii) and Case (v) is symmetric to Case (i).

1) Case (i): $R^*_X=H(P_{X|Y}|P_Y)$ and $R_Y>H(P_Y)$
For this case, invoking (222) and (224), we obtain

$$E_X(R^l_{X,n},R^m_{Y,n},\gamma)\geq \rho_n(1-\gamma)\left((R^*_X+R^*_Y)-H(P_{XY})\right)-\theta_1\rho_n\kappa_{n,l}-(1-\gamma)\theta_2\rho_n\zeta_{n,m}$$
$$+\frac{(\theta_1+(1-\gamma)\theta_2)^2\xi_n^2}{2\left(\gamma \mathrm{V}(P_{X|Y}|P_Y)+(1-\gamma)\mathrm{V}(P_{XY})\right)}-\frac{\rho^3_n}{6}\max\{M_{XY},M_X\} \tag{225}$$

and

$$E_Y(R^l_{X,n},R^m_{Y,n},\gamma)\geq \rho'_n\left(\gamma R^*_Y+(1-\gamma)(R^*_X+R^*_Y)-\gamma H(P_{Y|X}|P_X)-(1-\gamma)H(P_{XY})\right)$$
$$-(1-\gamma)\theta_1\rho'_n\kappa_{n,l}-\theta_2\rho'_n\zeta_{n,m}+\frac{((1-\gamma)\theta_1+\theta_2)^2\xi_n^2}{2\left(\gamma \mathrm{V}(P_{Y|X}|P_X)+(1-\gamma)\mathrm{V}(P_{XY})\right)}$$
$$-\frac{\rho'^3_n}{6}\max\left\{M_{XY},M_Y\right\}. \tag{226}$$

In order to evaluate the right hand side of (225), we define the following functions:

$$f(\gamma):=\frac{(1-\gamma)\left(R^*_X+R^*_Y-H(P_{XY})\right)(\theta_1+(1-\gamma)\theta_2)}{\gamma \mathrm{V}(P_{X|Y}|P_Y)+(1-\gamma)\mathrm{V}(P_{XY})} \tag{227}$$

$$g(\gamma):=\frac{(\theta_1+(1-\gamma)\theta_2)^2}{2\left(\gamma \mathrm{V}(P_{X|Y}|P_Y)+(1-\gamma)\mathrm{V}(P_{XY})\right)} \tag{228}$$

$$g_1(\gamma):=-\frac{(\theta_1+(1-\gamma)\theta_2)}{\gamma \mathrm{V}(P_{X|Y}|P_Y)+(1-\gamma)\mathrm{V}(P_{XY})} \tag{229}$$

$$g_2(\gamma):=-\frac{(1-\gamma)\theta_2\left(\theta_1+(1-\gamma)\theta_2\right)}{\gamma \mathrm{V}(P_{X|Y}|P_Y)+(1-\gamma)\mathrm{V}(P_{XY})} \tag{230}$$

$$g_3(\gamma):=-\frac{\max\left\{M_{XY},M_Y\right\}}{6}\left(\frac{(\theta_1+(1-\gamma)\theta_2)}{\gamma \mathrm{V}(P_{X|Y}|P_Y)+(1-\gamma)\mathrm{V}(P_{XY})}\right)^3. \tag{231}$$

Invoking the definition of ρ_n in (215), we conclude that minimizing the right hand side of (225) is equivalent to minimizing

$$F(n,\gamma):=f(\gamma)\xi_n+g(\gamma)\xi_n^2+g_1(\gamma)\rho_n\kappa_{n,l}+g_2(\gamma)\rho_n\zeta_{n,m}+g_3(\gamma)\xi_n^3. \tag{232}$$

Note that $\rho_n\kappa_{n,l}=o(\xi_n^2)$ (see (151) and (215)) and similarly $\rho_n\zeta_{n,m}=o(\xi_n^2)$. Invoking Lemma 9, we obtain

$$\inf_{\gamma\in[0,1]}E_X(R_{X,n},R_{Y,n},\gamma)\geq \inf_{\gamma\in[0,1]}F(n,\gamma) \tag{233}$$
$$=\frac{T\theta_1^2\xi_n^2}{2\mathrm{V}(P_{X|Y}|P_Y)}+o(\xi_n^2). \tag{234}$$

However, for any $\gamma\in[0,1]$, the right hand side of (226) is dominated by the first term which is of order $\Theta(\xi_n)$ (see (223)).
Therefore, we obtain

$$\liminf_{n\to\infty}\frac{\min\left\{\inf_{\gamma\in[0,1]}E_X(R^l_{X,n},R^m_{Y,n},\gamma),\inf_{\gamma\in[0,1]}E_Y(R^l_{X,n},R^m_{Y,n},\gamma)\right\}}{\xi_n^2}$$
$$\geq \min\left\{\frac{1}{2\mathrm{V}(P_{X|Y}|P_Y)},\infty\right\}=\frac{1}{2\mathrm{V}(P_{X|Y}|P_Y)}. \tag{235}$$

2) Case (ii): $R^*_X=H(P_{X|Y}|P_Y)$ and $R_Y=H(P_Y)$
Invoking (222) and (224), we obtain

$$E_X(R^l_{X,n},R^m_{Y,n},\gamma)\geq -\theta_1\rho_n\kappa_{n,l}-(1-\gamma)\theta_2\rho_n\zeta_{n,m}+\frac{(\theta_1+(1-\gamma)\theta_2)^2\xi_n^2}{2\left(\gamma \mathrm{V}(P_{X|Y}|P_Y)+(1-\gamma)\mathrm{V}(P_{XY})\right)}$$
$$-\frac{\rho^3_n}{6}\max\{M_{XY},M_X\} \tag{236}$$

and

$$E_Y(R_{X,n}^l, R_{Y,n}^m, \gamma) \geq \rho_n' \gamma I(P_{X|Y}, P_Y) - (1-\gamma)\theta_1 \rho_n' \kappa_{n,l} - \theta_2 \rho_n' \zeta_{n,m} + \frac{((1-\gamma)\theta_1 + \theta_2)^2 \xi_n^2}{2\left(\gamma \mathrm{V}(P_{Y|X}|P_X) + (1-\gamma)\mathrm{V}(P_{XY})\right)} - \frac{\rho_n'^3}{6} \max\left\{M_{XY}, M_Y\right\}. \tag{237}$$

Following the argument leading to (235), we obtain

$$\liminf_{n\to\infty} \frac{\min\left\{\inf_{\gamma\in[0,1]} E_X(R_{X,n}^l, R_{Y,n}^m, \gamma), \inf_{\gamma\in[0,1]} E_Y(R_{X,n}^l, R_{Y,n}^m, \gamma)\right\}}{\xi_n^2} \geq \min\left\{\inf_{\gamma\in[0,1]} \frac{(\theta_1 + (1-\gamma)\theta_2)^2}{2\left(\gamma \mathrm{V}(P_{X|Y}|P_Y) + (1-\gamma)\mathrm{V}(P_{XY})\right)}, \frac{(\theta_1+\theta_2)^2}{2\mathrm{V}(P_{XY})}\right\} \tag{238}$$

3) Case (iii): $H(P_{X|Y}|P_Y) < R_X^* < H(P_X)$, $H(P_{Y|X}|P_X) < R_Y^* < H(P_Y)$ and $R_X^* + R_Y^* = H(P_{XY})$. Invoking (222) and (224), we obtain

$$E_X(R_{X,n}^l, R_{Y,n}^m, \gamma) \geq \rho_n \gamma\left(R_X^* - H(P_{X|Y}|P_Y)\right) - \theta_1 \rho_n \kappa_{n,l} - (1-\gamma)\theta_2 \rho_n \zeta_{n,m} + \frac{(\theta_1 + (1-\gamma)\theta_2)^2 \xi_n^2}{2\left(\gamma \mathrm{V}(P_{X|Y}|P_Y) + (1-\gamma)\mathrm{V}(P_{XY})\right)} - \frac{\rho_n^3}{6} \max\{M_{XY}, M_X\} \tag{239}$$

and

$$E_Y(R_{X,n}^l, R_{Y,n}^m, \gamma) \geq \rho_n' \gamma\left(R_Y^* - H(P_{Y|X}|P_X)\right) - (1-\gamma)\theta_1 \rho_n' \kappa_{n,l} - \theta_2 \rho_n' \zeta_{n,m} + \frac{((1-\gamma)\theta_1 + \theta_2)^2 \xi_n^2}{2\left(\gamma \mathrm{V}(P_{Y|X}|P_X) + (1-\gamma)\mathrm{V}(P_{XY})\right)} - \frac{\rho_n'^3}{6} \max\left\{M_{XY}, M_Y\right\}. \tag{240}$$

Following similar steps leading to (235), we obtain

$$\liminf_{n\to\infty} \frac{\min\left\{\inf_{\gamma\in[0,1]} E_X(R_{X,n}^l, R_{Y,n}^m, \gamma), \inf_{\gamma\in[0,1]} E_Y(R_{X,n}^l, R_{Y,n}^m, \gamma)\right\}}{\xi_n^2} \geq \frac{(\theta_1+\theta_2)^2}{2\mathrm{V}(P_{XY})}. \tag{241}$$

REFERENCES

[1] "Netflix streams its way to another blockbuster quarter, share price soars," http://fortune.com/2015/07/15/netflix-q2-earnings-2015/, accessed: 2015-07-15.
[2] V. N. Padmanabhan, H. J. Wang, P. A. Chou, and K. Sripanidkulchai, "Distributing streaming media content using cooperative networking," in *Proc. ACM NOSSDAV*, 2002, pp. 177–186.
[3] T. P. Nguyen and A. Zakhor, "Distributed video streaming over internet," in *Proc. ACM/SPIE MMCN*, Jan 2002, pp. 186–195.
[4] S. Saroiu, K. P. Gummadi, R. J. Dunn, S. D. Gribble, and H. M. Levy, "An analysis of internet content delivery systems," *SIGOPS Oper. Syst. Rev.*, vol. 36, no. SI, pp. 315–327, Dec. 2002. [Online]. Available: http://doi.acm.org/10.1145/844128.844158
[5] D. Slepian and J. K. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. Inf. Theory*, vol. 19, no. 4, pp. 471–480, 1973.
[6] C. Chang, "Streaming source coding with delay," Ph.D. dissertation, UC Berkeley, 2007.
[7] 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.
[8] I. Csiszár and J. Körner, *Information theory: coding theorems for discrete memoryless systems*. Cambridge University Press, 2011.
[9] R. G. Gallager, *Information Theory and Reliable Communication*. New York: Wiley, 1968.
[10] 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.
[11] M. Hayashi, "Second-order asymptotics in fixed-length source coding and intrinsic randomness," *IEEE Trans. Inf. Theory*, vol. 54, no. 10, pp. 4619–4637, 2008.
[12] V. Strassen, "Asymptotische Abschätzungen in Shannons Informationstheorie," in *Trans. Third Prague Conf. Inf. Theory*, Prague, 1962, pp. 689–723, http://www.math.cornell.edu/∼pmlut/strassen.pdf.
[13] Y. Altug and A. Wagner, "Moderate deviations in channel coding," *IEEE Trans. Inf. Theory*, vol. 60, no. 8, pp. 4417–4426, 2014.
[14] 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.
[15] Y. Altug, A. Wagner, and I. Kontoyiannis, "Lossless compression with moderate error probability," in *Proc. IEEE ISIT*, Jul. 2013, pp. 1744–1748.
[16] S.-H. Lee, V. Y. Tan, and A. Khisti, "Streaming data transmission in the moderate deviations and central limit regimes," *arXiv:1512.06298*, 2015.
[17] C. Chang and A. Sahai, "The error exponent with delay for lossless source coding," in *Proc. IEEE ITW*, Mar. 2006, pp. 252–256.
[18] ——, "The price of ignorance: The impact of side-information on delay for lossless source-coding," *arXiv:0712.0873*, 2007.
[19] 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.
[20] H. Palaiyanur and A. Sahai, "Sequential decoding for lossless streaming source coding with side information," *arXiv cs/0703120*, 2007.
[21] 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.
[22] 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.
[23] N. Ma and P. Ishwar, "On delayed sequential coding of correlated sources," *IEEE Trans. Inf. Theory*, vol. 57, no. 6, pp. 3763–3782, 2011.
[24] N. Zhang, B. Vellambi, and K. Nguyen, "Delay exponent of variable-length random binning for point-to-point transmission," in *Proc. IEEE ITW*, Nov. 2014, pp. 207–211.
[25] J. Chen, D.-K. He, A. Jagmohan, and L. A. Lastras-Montano, "On the redundancy-error tradeoff in Slepian-Wolf coding and channel coding," in *Proc. IEEE ISIT*, 2007.

[26] D.-K. He, L. Lastras-Montaño, E.-H. Yang, A. Jagmohan, and J. Chen, "On the redundancy of Slepian–Wolf coding," *IEEE Trans. Inf. Theory*, vol. 55, no. 12, pp. 5607–5627, 2009.

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

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

[29] 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.

[30] S. C. Draper and A. Khisti, "Truncated tree codes for streaming data: Infinite-memory reliability using finite memory," in *ISWCS*, Nov 2011, pp. 136–140.

[31] Y. Polyanskiy, *Channel coding: non-asymptotic fundamental limits*. Princeton University, 2010.

[32] T. Cormen, C. Leiserson, R. Rivest, and C. Stein, *Introduction to Algorithms*, 2nd ed. McGraw-Hill Science/Engineering/Math, 2003.

[33] V. Y. F. Tan, *Asymptotic estimates in information theory with non-vanishing error probabilities*. Foundations and Trends® in Communications and Information Theory, 2015, vol. 11, no. 1-2.

[34] 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.

[35] 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.

[36] M. Hayashi and R. Matsumoto, "Secure multiplex coding with dependent and non-uniform multiple messages," *IEEE Trans. Inf. Theory*, vol. PP, no. 99, pp. 1–1, 2016.

[37] T. S. Han, *Information-spectrum methods in information theory*. Springer, 2003.

[38] A. Dembo and O. Zeitouni, *Large deviations techniques and applications*. Springer Science & Business Media, 2009, vol. 38.

[39] H. Palaiyanur, "The impact of causality on information-theoretic source and channel coding problems," 2011.

[40] 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.

[41] A. El Gamal and Y.-H. Kim, *Network information theory*. Cambridge University Press, 2011.