Abstract

Streaming of Markov Sources over Burst Erasure Channels

Farrokh Etezadi
Doctor of Philosophy
Graduate Department of Electrical and Computer Engineering (ECE)
University of Toronto
2015

Real-time streaming communication systems require both the sequential encoding of information sources and playback under strict latency constraints. The central focus of this thesis is on the fundamental limits of such communication systems in the presence of packet losses. In practice packet losses are unavoidable due to fading in wireless channels or congestion in wired networks. While several ad hoc approaches are used to deal with packet losses in streaming systems, in this thesis we examine these approaches using an information theoretic framework.

In our setup, the source process is a sequence of vectors sampled from a spatially i.i.d. and temporally a first-order stationary Markov distribution. The encoder sequentially compresses these source vectors into channel packets. The channel may introduce a burst erasure of length up to $B$ in an unknown location during the transmission period, and perfectly reveals the rest of the packets to the destination. The decoder is interested in reconstructing the source vectors with zero delay, except those at the time of erasure and a window of length $W$ following it. The minimum attainable compression rate for this setup $R(B,W)$, termed the rate-recovery function, is investigated for discrete source with lossless recovery, and Gauss-Markov sources with a quadratic distortion measure.

The above setup introduces a new problem in network information theory. Our key contributions include: (1) Upper and lower bounds on the rate-recovery function for discrete memoryless sources and lossless recovery, which coincide in some special cases. (2) A new coding scheme for the Gauss-Markov sources and a quadratic distortion measure. This scheme can be interpreted as a hybrid between predictive coding and memoryless quantization-and-binning. (3) Extensions of our zero-delay setup to incorporate non-zero decoding delays. We further show that our proposed hybrid coding scheme yields significant performance gains over baseline schemes such as predictive coding, memoryless quantization-and-binning and interleaving, over statistical channels such as the i.i.d. erasure channel and the Gilbert Elliott channel, and performs close to optimally, over a wide range of channel parameters. While our information theoretic framework involves coding theorems for burst-erasure channels our resulting schemes are applicable for much broader class of erasure channels and can yield significant performance gains in practice.
To my parents and sister
**Acknowledgements**

First and foremost, I would like to thank my advisor Professor Ashish Khisti for his excellent academic supervision during the course of my doctoral work. Ashish is an outstanding teacher and a brilliant researcher from whom I learned enormously. I really admire his dedication and passion for research, clarity of thought and his intuitive and fundamental attitude toward problems. I am honored to be his student, and would like to express my hearty gratitude for all I learned from him.

I also wish to express my gratitude to the members of my Ph.D. committee, namely, Professors Stark Draper, Frank R. Kschischang and Wei Yu. I wish to extend my deep appreciation to my external examiner, Professor Jun Chen, from McMaster University, for his constructive comments.

I would like also to sincerely thank my family, specially my parents Mohammad-Bagher and Shafigheh, and my sister Shabnam for their endless love and support toward me. This thesis is dedicated to them.

Finally, I thank my friends and colleagues for the fun times, which will leave many fond memories.
# Contents

1 Introduction ................................................................. 1

1.1 Related Works ............................................................. 2

1.1.1 Structural Results on Sequential Compression ................. 2

1.1.2 Information Theoretic Models for Sequential Compression .... 2

1.1.3 Sequential Joint Source-Channel Coding ....................... 3

1.1.4 Practical Distributed Video Coding ............................ 3

1.2 Thesis Outline and Contributions ..................................... 4

1.3 Notation ........................................................................ 5

2 Zero-Delay Lossless Streaming ........................................... 6

2.1 Introduction ................................................................. 6

2.2 Problem Statement ........................................................ 6

2.2.1 Source Model and Encoder ......................................... 7

2.2.2 Channel Model and Decoder ......................................... 7

2.2.3 Rate-Recovery Function ............................................. 8

2.2.4 Practical Motivation .................................................. 8

2.3 Main Results ................................................................. 9

2.3.1 Upper and Lower Bounds on Rate-Recovery .................... 9

2.3.2 Symmetric Sources and Memoryless Encoders ................ 10

2.3.3 Linear Semi-Deterministic Markov Sources .................... 10

2.4 Upper and Lower Bounds on Lossless Rate-Recovery Function . 12

2.4.1 Connection to Multi-terminal Source Coding Problem ........ 12

2.4.2 Lower Bound on Lossless Rate-Recovery Function ............ 14

2.4.3 Upper Bound on Lossless Rate-Recovery Function ............ 16

2.5 Symmetric Sources and Memoryless Encoders ...................... 18

2.5.1 Multi-Terminal Source Coding ...................................... 19

2.5.2 Lower Bound for Multi-terminal Source Coding Problem ..... 20

2.5.3 Extension to Arbitrary $W > 1$ .................................... 21

2.6 Semi-Deterministic Markov Sources and Prospicient Coding .... 22

2.6.1 Diagonally Correlated Semi-Deterministic Source Model .... 22

2.6.2 Sufficiency of $K = B + W$ ......................................... 24

2.6.3 Prospicient Coding .................................................... 24

2.6.4 Rate-Optimality of the Coding Scheme ........................ 26
3 Zero-Delay Streaming of Gauss-Markov Sources: Immediate Recovery 34

3.1 Introduction ............................................................. 34
3.2 Problem Setup .......................................................... 35
  3.2.1 Source Model and Encoder ........................................ 35
  3.2.2 Channel Model and Decoder ....................................... 35
  3.2.3 Rate-Recovery Function .......................................... 36
3.3 Main Results ............................................................ 36
  3.3.1 Channels with Single erasure burst ............................... 36
  3.3.2 Channels with Multiple erasure bursts ............................ 38
  3.3.3 High Resolution Regime ......................................... 40
  3.3.4 Gaussian Sources with Sliding Window Recovery Constraints . . 41
3.4 Lower Bound on Rate-Recovery Function ............................ 42
3.5 Coding Scheme: Single erasure burst ................................ 44
  3.5.1 Analysis of Achievable Rate ..................................... 45
  3.5.2 Numerical Evaluation ............................................. 50
3.6 Coding Scheme: Multiple erasure bursts with Guard Intervals .......... 51
  3.6.1 Analysis of Achievable Rate ..................................... 52
  3.6.2 Numerical Evaluation ............................................. 53
3.7 Upper Bound in the High Resolution Regime .......................... 54
3.8 Independent Gaussian Sources with Sliding Window Recovery .......... 56
  3.8.1 Sufficiency of $K = B + W$ ...................................... 56
  3.8.2 Coding Scheme ..................................................... 56
  3.8.3 Converse ............................................................ 61
  3.8.4 Illustrative Suboptimal Schemes ................................ 62
3.9 Conclusion .............................................................. 64

4 Hybrid Coding 65

4.1 Introduction ............................................................. 65
4.2 Problem Statement ...................................................... 66
4.3 Zero-Delay Sequential Coding Scheme for Erasure-free Channels ....... 67
  4.3.1 General Coding Scheme for DMS .................................. 68
  4.3.2 General Coding Scheme for Gauss-Markov Sources ............. 69
  4.3.3 Special Cases ...................................................... 72
  4.3.4 Rate Analysis of General Coding Scheme for Erasure-free channels . 73
4.4 Rate Analysis for erasure burst Channels ................................ 76
  4.4.1 General Coding Scheme for DMS .................................. 76
  4.4.2 Conventional Coding Schemes for Gauss-Markov Sources ........ 77
  4.4.3 Hybrid Coding Scheme ............................................. 80
  4.4.4 Lower Bound on Rate-Recovery Function ......................... 91
  4.4.5 Performance Over Statistical Channel Models ...................... 91
4.5 Upper Bounds on Rate-Recovery .................................................. 96
  4.5.1 Predictive Coding .............................................................. 96
  4.5.2 Memoryless Q-and-B ......................................................... 97
  4.5.3 Memoryless Q-and-B in the High Resolution Regime ................. 101
  4.5.4 GOP-Based Coding Scheme .................................................. 102
  4.5.5 Hybrid Coding Scheme ...................................................... 103
  4.5.6 Hybrid Coding Scheme In the High Resolution Regime .............. 105
4.6 Lower Bound on Rate-Recovery Function ....................................... 113
  4.6.1 Connection to the Multi-terminal Source Coding Problem ............ 113
  4.6.2 General Lower Bound ....................................................... 115
  4.6.3 Lower Bound in the High Resolution Regime ............................ 118
4.7 Conclusion ............................................................................ 119

5 Delay-Constrained Streaming ......................................................... 120
  5.1 Introduction ........................................................................... 120
  5.2 Problem Statement ............................................................... 121
    5.2.1 Streaming with Controlled-Interruption ............................... 122
    5.2.2 Streaming with Ideal-Playback .......................................... 123
  5.3 Main Results .......................................................................... 123
    5.3.1 Streaming with Controlled-Interruption ............................... 123
    5.3.2 Streaming with Ideal-Playback .......................................... 128
  5.4 Lossless Streaming with Controlled-Interruption ......................... 131
    5.4.1 Achievability ................................................................. 131
    5.4.2 Converse ..................................................................... 131
  5.5 Lossless Streaming with Controlled-Interruption over
    Sliding-Window erasure burst Channel ........................................ 133
    5.5.1 Achievability ................................................................. 133
    5.5.2 Converse .................................................................. 135
  5.6 Lossy Streaming of Gauss-Markov Sources with Controlled-Interruption
  ....................................................................................................... 136
    5.6.1 Converse ................................................................. 136
    5.6.2 Coding Scheme ............................................................ 137
    5.6.3 High Resolution Regime .................................................. 146
  5.7 Lossless Streaming with Ideal-Playback ....................................... 147
    5.7.1 Converse ................................................................. 147
    5.7.2 Achievability ............................................................... 149
  5.8 Lossy Streaming of Gauss-Markov Sources
    with Ideal-Playback ................................................................... 149
    5.8.1 Converse ................................................................. 149
    5.8.2 Achievability ............................................................... 152
  5.9 Conclusion ............................................................................ 156

6 Conclusion ............................................................................ 157

Appendices ............................................................................. 159
List of Tables

4.1 Numerical values of $\rho^*$ for $W \in \{1, 2, 3, 4\}$ .................................................. 88
4.2 High resolution excess rates for different coding schemes ............................................. 89
List of Figures

1.1 The different decoding delay and source/recovery models treated in different chapters of the thesis .................................................. 4

2.1 Problem Setup: The encoder output \( f_i \) is a function of all the past source sequences. The channel introduces an erasure burst of length up to \( B \). The decoder produces \( \hat{s}_{ni} \) upon observing the channel outputs up to time \( i \). As indicated, the decoder is not required to produce those source sequences that are observed either during the erasure burst, or a period of \( W \) following it. The first sequence, \( s_{n1} \) is a synchronization frame available to both the source and destination. .................................................. 7

2.2 Multi-terminal problem setup associated with our proposed streaming setup when \( W = B = 1 \). The erasure at time \( t = j - 1 \) leads to two virtual decoders with different side information as shown. .................................................. 12

2.3 Connection between the streaming problem and the multi-terminal source coding problem. The setup on the right is identical to the setup on the left, except with the side information sequence \( s^{n}_{j-1} \) replaced with \( s^{n}_{j+1} \). However the rate region for both problems are identical for symmetric Markov sources. .................................................. 18

2.4 Schematic of Diagonally Correlated Deterministic Markov Source. The first row of sub-symbols are innovation symbols. They are generated independently of all past symbols. On each diagonal the sub-symbol is a deterministic function of the sub-symbols above it. .................................................. 23

2.5 Schematic of Coding Scheme: Codeword structure. We set \( p = B + W + 1 \). .................................................. 24

2.6 Schematic of Coding Scheme: Rate reduction. .................................................. 25

2.7 Block diagram of the system described in Case 1. .................................................. 28

2.8 Two sub-blocks of \( T \). .................................................. 29

3.1 Lower and upper bounds of lossy rate-recovery function. .................................................. 37

3.2 Achievable rates for multiple erasure bursts model for different values of guard length \( L \) separating erasure bursts comparing to single erasure burst. As \( L \) grows, the rate approaches the single erasure case. The lower bound for single erasure case is also plotted for comparison (\( B = 1 \)). .................................................. 39

3.3 A comparison of achievable rates for the Gauss-Markov source (\( B = 1 \)). .................................................. 40
3.4 Schematic of the Gaussian sources with sliding window recovery constraints for $K = 2$. The source $s_t$, drawn as white circles, are independent sources and $t_i$ is defined as a collection of $K + 1 = 3$ most recent sources. The source symbols along the diagonal lines are the same. The decoder at time $i$ recovers $s_i$, $s_{i-1}$ and $s_{i-2}$ within distortions $d_0$, $d_1$ and $d_2$, respectively where $d_0 \leq d_1 \leq d_2$. In figure the colour density of the circle represents the amount of reconstruction distortion.

3.5 Flowchart summarizing the proof steps of Prop. 3.

3.6 Schematic of single erasure burst channel model. The channel inputs in the interval $\{t - B' - k, \ldots, t - k - 1\}$ is erased for some $0 \leq B' \leq B$ and $k \in \{0, 2, \ldots, t - B'\}$. The rest are available at the decoder, as shown by check mark in the figure.

3.7 Replacing $u_{t-B'-k}$ by $u_{t-k}$ improves the estimate of $s_t$ and $u_t$.

3.8 Schematic of the erasure patterns in Lemma 11 for $t = 18$, $L = 3$ and $B = 2$. Fig. 3.8a illustrates, $\Omega^*_t(\theta)$ in part 1 of Lemma 11. The non-erased symbols are denoted by checkmarks. Fig. 3.8b illustrates $\Omega^*_t$ as stated in part 2 of Lemma 11.

3.9 Schematic of encoder and decoder for i.i.d. Gaussian source with sliding window recovery constraint. SR and LR indicate successive refinement and layer rearrangement (Sections 3.8.2 and 3.8.2), respectively.

3.10 $(B + 1)$-layer coding scheme based on successive refinement (SR). Note that for each $k \in \{0, \ldots, B\}$, $m_{i,k}$ is of rate $R_k$ and $M_{i,k}$ is of rate $R_k$. The dashed box represents the SR code.

3.11 Schematic of the auxiliary codewords defined in (3.117). The codewords are temporally correlated in a diagonal form depicted using ellipses. In particular, as shown in Fig. 3.10, $M_{i-j,j} \subseteq M_{i-j,j-1}$. Based on this diagonal correlation structure, the codewords depicted in the boxes are sufficient to know all the codewords.

3.12 Comparison of rate-recovery of suboptimal systems to minimum possible rate-recovery function for different recovery window length $W$. We assume $K = 5$, $B = 2$ and a distortion vector $d = (0.1, 0.25, 0.4, 0.55, 0.7, 0.85)$.

4.1 Problem Setup: Consider the example of $B = 2$ and $W = 3$. The encoder output $f_j$ is a function of the source sequences up to time $j$, i.e., $s^0_j, s^1_j, \ldots, s^n_j$. The channel introduces an erasure burst of length $B$. The decoder produces $\hat{s}^n_j$ upon observing the sequence $\{g_0, g_1, \ldots, g_j\}$. The decoder is not required to produce those source sequences that fall in a window of length $B + W$ following the start of an erasure burst. However, the decoder recovers the rest of the source sequences within zero-delay and average distortion constraint.

4.2 Block diagram of the proposed coding scheme: a) Encoder, b) Decoder.

4.3 Structure of encoder for special cases.

4.4 The trade-off between the rate-reduction gains of successive quantization and binning techniques. In predictive coding, the binning gain is zero and successive quantization gain is the maximum. In memoryless Q-and-B the successive quantization gain is zero and the binning gain is the maximum. The overall rate-reduction is independent of the choice of $A$.

4.5 For jointly Gaussian sources the two system (a) and (b) attain the same rate-distortion pair with different coding schemes.
4.6 Excess Rates based on Hybrid Coding in comparison with Predictive Coding, Memoryless
Q-and-B and the lower bound for \( D = 0.3 \). The excess rate is the difference between the
achievable rate by the that coding scheme and the achievable rate of the predictive coding
in case of erasure-free channel. ................................................................. 81

4.7 Schematic of encoder and decoder for \( B = W = 1 \) and high resolution. The \( f_{t-2} \) is erased
and \( \hat{s}_t \) is required to be reconstructed at time \( t \). (a) The predictive coding which fails in
reconstructing \( s^n_t \) (b) The memoryless Q-and-B scheme, (c) The simplified scheme, and
(d) The generalization of the idea. ............................................................... 82

4.8 High resolution Excess Rates based on Hybrid Coding in comparison with Memoryless
Q-and-B and the lower bound. ................................................................. 89

4.9 Comparison excess rate of different schemes in high resolution when \( B \to \infty \).
.................................................................................................................. 90

4.10 Gilbert Channel model. In the “Good” state, the channel perfectly reveals the encoder
output to the decoder, while in the “Bad” state, the channel erases the decoder output. . 91

4.11 Comparison of different schemes for 2% rate-overhead over erasure-free channel rate, i.e.,
\( R = 1.02 R_I = 1.0629 \text{ Bits/Symbol}, \rho = 0.8 \) and \( D = 0.1 \). ................................. 93

4.12 Comparison of different schemes \( \rho = 0.8, D = 0.1, \alpha_G = 5 \times 10^{-3} \) and \( R = 1.0629 \)
Bits/Symbol. .................................................................................................. 95

4.13 Flowchart summarizing the proof steps of Theorem 8. .................................................. 97

4.14 erasure burst channel model parametrized by starting time of erasure, \( \tau \). ...................... 98

4.15 The sum-rate and marginal rates as a function of coefficient \( w_1 \), for \( B = W = 1 \) and
\( \rho = 0.7 \). The marginal rate is upper bounded for approximation. ............................... 112

4.16 Multi-terminal source coding problem as an enhanced version of original streaming problem.114

5.1 Problem Setup: Consider the example of \( B = 3 \) and \( W = T = 2 \). The encoder output
\( f_j \) is a function of the source sequences up to time \( j \), i.e., \( s^n_0, s^n_1, \ldots, s^n_j \). The channel
introduces an erasure burst of length \( B \). The decoder produces \( \hat{s}^n_j \) upon observing the
sequence \( \{g_0, g_1, \ldots, g_{j+T}\} \). a) In Controlled-Interruption, the decoder is not required to
produce those source sequences that fall in a window of length \( B + W \) following the start
of an erasure burst. However, the decoder recovers the rest of the source sequences within
a delay of \( T \). b) In Ideal-Playback the decoder recovers all the source sequences within a
delay of \( T \). ........................................................................................................ 121

5.2 Sliding-window erasure burst channel model. The channel introduces multiple erasure
bursts each of length up to \( B \). The consecutive erasure bursts are separated by a guard
interval of length at least \( G \). .............................................................................. 125

5.3 Upper and lower bounds on the delay-constrained streaming rate function for Gauss-
Markov with correlation \( \rho \) sources over erasure burst channel of maximum burst length
\( B \) and delay \( T \). Both ideal-playback and controlled-interruption streaming rates are shown.127

5.4 The encoders for the two coding schemes for ideal-playback streaming with delay-constraint
\( T \). The coding scheme I consists of zero-delay predictive coding source code and delay-
optimal channel code with delay \( T \). The coding scheme II for any \( T > 1 \), consists of
unit-delay optimal source code and delay-optimal channel code with delay \( T - 1 \). The
overall delay of both schemes is \( T \). ........................................................................ 129

5.5 Periodic erasure burst channel considered in proof of converse. ........................................ 131
5.6 An example of streaming with controlled-interruption over sliding-window erasure burst channel model. \( B = 3 \), \( W = 1 \) and \( G = 3 \). (a) \( T = 2 \), and (b) \( T_{\text{opt}} = G - W - 1 = 1 \). Note that the two systems are equivalent because the decoder in both cases recovers the source sequences \( \hat{s}_4^n \) and \( \hat{s}_5^n \) from \( \{f_3, f_4, f_5\} \) which is available at time 5.

5.7 A schematic of the erasure burst channel model and four different regions for \( B' = 2 \), \( T = 3 \), \( j = 6 \).

B.1 Relationship of the Variables for Lemma 31.

C.1 Schematic of the variables in Lemma 35.

C.2 Example of Lemma 36 with \( B = 2 \), \( W = 4 \), \( \tau = 7 \). In this case \( k = 3 \), \( L_\tau = \{9, 10, 11, 12, 13\} \) and \( K^* = \{9, 10, 11\} \). According to Lemma 36, among any subset of \( L_\tau \) with size \( k = 3 \), differential entropy of \( [u|K^*] \) give \( [u]_{1}^{\tau-1} \) is the minimum.
Chapter 1

Introduction

“As you yourself have said, what other explanation can there be?” Poirot stared straight ahead of him. “That is what I ask myself,” he said. “That is what I never cease to ask myself.”

Agatha Christie, Murder on the Orient Express

The recent proliferation of mobile devices such as Smart-phones, Tablet PCs and Netbooks has truly revolutionized our day-to-day activities and has opened up new possibilities for collaboration, communication and social networking that we could never have imagined just a few years ago. Naturally such a phenomenal growth has significantly increased the stress on wireless communication infrastructure and created an unprecedented demand for high quality multimedia streaming over both wireless and wired networks.

A short-term solution for service providers to satisfy this increasing demand is to acquire more wireless spectrum. A longer term solution however is to develop fundamentally new techniques for efficient multimedia streaming over the Internet as well as wireless networks. An average consumer today can routinely watch high-definition programming over television sets. Enabling a similar high quality experience over the Internet today is highly expensive, if not impossible. Wireless adds its own set of challenges. Thus fundamentally new techniques for compression and communication are essential to support high end streaming applications over such communication networks.

Any real-time multimedia streaming application requires both the sequential compression, and playback of multimedia frames under strict latency constraints. Linear predictive techniques such as differential pulse-code modulation (DPCM) have long been used to exploit the source memory in such systems to remove as much redundancy as possible and only send information essential for source reconstruction to the destination [1]. While predictive coding is a very powerful technique to remove the source redundancy and increase the compression efficiency it also exhibits a significant level of error propagation in the presence of packet losses [2]. In practice one must develop transmission schemes that satisfy both the real-time constraints and are robust to channel errors.

Commonly used video compression formats such as H.264/MPEG and HEVC use a combination of intra-coded and predictively-coded frames to limit the amount of error propagation. The predictively-
Chapter 1. Introduction

coded frames are used to improve the compression efficiency whereas the intra-coded frames limit the amount of error propagation. Other techniques including forward error correction codes [3, 4], leaky DPCM [5–7] and distributed video coding [8, 9] can also be used to trade off the transmission rate with error propagation. Despite these efforts, such a tradeoff is not well understood even in the case of a single isolated packet loss [10]. A central objective of this thesis is to understand this trade off between the compression efficiency and error propagation from an information-theoretic standpoint.

We first present related literature in Section 1.1. In Section 1.2 we summarize the outline of the thesis and the main contributions of each chapter. Section 1.3 summarizes the notation used throughout the thesis.

1.1 Related Works

Problems involving real-time coding and compression have been studied from many different perspectives in the related literature. In this section we categorize the related works into the following classes.

1.1.1 Structural Results on Sequential Compression

The compression of a Markov source, with zero encoding and decoding delays, was studied in an early work by Witsenhausen [11]. In this setup, the encoder must sequentially compress a (scalar) Markov source and transmit it over an ideal channel. The channel can be viewed as an ideal bit-pipe between the transmitter and destination. The decoder must reconstruct the source symbols with zero-delay and under an average distortion constraint. It was shown in [11] that for a $k$-th order Markov source model, an encoding rule that only depends on the $k$ most recent source symbols, and the decoder’s memory, is sufficient to achieve the optimal rate. Similar structural results have been obtained in a number of followup works, see e.g., [12] and references therein. The authors in [13] considered real-time communication of a memoryless source over memoryless channels, with or without the presence of unit-delay feedback. The encoding and decoding is sequential with a fixed finite lookahead at the encoder. The authors propose conditions under which symbol-by-symbol encoding and decoding, without lookahead, is optimal and more generally characterize the optimal encoder as a solution to a dynamic programming problem.

1.1.2 Information Theoretic Models for Sequential Compression

The problem of sequential coding of correlated vector sources in a multi-terminal source coding framework was introduced by Viswanathan and Berger [14]. In this setup, a set of correlated sources must be sequentially compressed by the encoder, whereas the decoder at each stage is required to reconstruct the corresponding source sequence, given all the encoder outputs up to that time. It is noted in [14] that the correlated source sequences can model consecutive video frames and each stage at the decoder maps to sequential reconstruction of a particular source frame. This setup is an extension of the well-known successive refinement problem in source coding [15]. In followup works, in reference [16] the authors consider the case where the encoders at each time have access to previous encoder outputs rather than previous source frames. Reference [17] considers an extension where the encoders and decoders can introduce non-zero delays. All these works assume ideal channel conditions. Reference [18] considers an extension of [14] where at any given stage the decoder has either all the previous outputs, or only the
present output. A robust extension of the predictive coding scheme is proposed and shown to achieve the minimum sum-rate. However this setup does not capture the effect of packet losses over a channel, where the destination has access to all the non erased symbols. To our knowledge, only reference [5] considers the setting of sequential coding over a random packet erasure channel. The source is assumed to be Gaussian, spatially i.i.d. and temporally autoregressive. A class of linear predictive coding schemes is studied and an optimal scheme within this class, with respect to the excess distortion ratio metric is proposed.

1.1.3 Sequential Joint Source-Channel Coding

In other related works, the joint source-channel coding of a vector Gaussian source over a vector Gaussian channel with zero reconstruction delay has also been extensively studied. While optimal analog mappings are not known in general, a number of interesting approaches have been proposed in, e.g., [19, 20] and related references. Reference [21] studies the problem of sequential coding of the scalar Gaussian source over a channel with random erasures. In [10], the authors consider a joint source-channel coding setup and propose the use of distributed source coding to compensate the effect of channel losses. However no optimality results are presented for the proposed scheme. Sequential random binning techniques for streaming scenarios have been proposed in, e.g., [22], [23] and the references therein.

1.1.4 Practical Distributed Video Coding

There has been a recent line of research on distributed video coding which attracted a lot of attention (see [9] and references therein). The idea is to apply distributed source coding techniques, specifically Slepian-Wolf coding [24] and Wyner-Ziv coding [25], to develop a new paradigm of video compression schemes with a low complexity encoder and a high complexity decoder. Unlike conventional video compression, the encoder independently encodes the video frames by only taking into account the inter-frame correlation, whereas the decoder must reconstruct each source frame, taking into account all the past observations [26]. The authors in [10] have proposed a two-layer compression scheme for error resilient video communication. The base-layer is the standard distributed source coding while the refinement layer consists of additional parity-check bits for reconstruction in the presence of packet losses. A similar approach is also presented in [27] where the proposed scheme was called Systematic Lossy Error Protection (SLEP). SLEP uses a two-layer error resilient video compression scheme. The primary layer is used during non-erasures. The additional refinement layer uses the syndrome bits that can recover the frames with a higher distortion during erasures using the available outputs of the first layer as side-information.

In this thesis, we build upon the information theoretic framework of sequential compression of vector sources [14] by considering the effect of packet losses. Following this line of work, we too consider source vectors that are drawn i.i.d. along the spatial dimension and form a first-order Markov process over the time. By assuming very large spatial dimension, we establish single-letter characterization of fundamental information-theoretic rate-distortion functions. We refer the reader to Section 2.2.4 for some discussion on the practical justification of the problem formulation. To the best of our knowledge, this thesis is the only work which studies an information theoretic tradeoff between error-propagation and compression efficiency in real-time streaming systems.
Chapter 1. Introduction

<table>
<thead>
<tr>
<th>Decoding Delay</th>
<th>Chapter 2</th>
<th>Chapter 3</th>
<th>Chapter 4</th>
<th>Chapter 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zero-Delay</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>-</td>
</tr>
<tr>
<td>Delay-Constrained</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>✓</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Source and Recovery Model</th>
<th>Chapter 2</th>
<th>Chapter 3</th>
<th>Chapter 4</th>
<th>Chapter 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Discrete Source/Lossless Recovery</td>
<td>✓</td>
<td>-</td>
<td>-</td>
<td>✓</td>
</tr>
<tr>
<td>Gauss-Markov Source/Lossy Recovery</td>
<td>-</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

Figure 1.1: The different decoding delay and source/recovery models treated in different chapters of the thesis

1.2 Thesis Outline and Contributions

The primary objective of this work is to develop coding schemes for real-time streaming that minimize the residual loss probability over statistical channels such as the i.i.d. erasure channel and Gilbert-Elliott erasure channel. However directly addressing such channels appears to be analytically intractable. Therefore for most of this thesis we will focus on a simpler class of channel models and study optimal coding schemes for such channels. In Chapters 4 and 5 we will explain how parameters in our simplified model can be selected to find near optimal schemes for the statistical models.

We consider a setup where the encoder must sequentially compress the stream of source vectors in a causal fashion and the decoder must reconstruct each source vector with zero delay. The channel introduces an erasure burst of maximum length $B$ and the decoder is not required to reconstruct those source vectors that correspond to the period of the erasure burst as well as those within a window of length $W$ that follows it. The minimum attainable rate is called the rate-recovery function. In this thesis, we consider two classes of sources and recovery constraints. For the class of discrete sources, where the source symbols are drawn from a finite alphabet, we consider the lossless recovery at the decoder. For the class of continuous sources, we consider the Gauss-Markov sources and lossy reconstruction with quadratic distortion measure. Fig 1.1 schematically explains the main focus of different chapters of the thesis and the relations among them.

Our achievability technique throughout this thesis is based on the quantization-and-binning technique which has been proposed in multi-terminal source coding [14]. However the choice of the test channel used provides considerable insight in the different scenarios as discussed below. First, in Chapter 2, we consider discrete sources and lossless recovery. In this setup we develop an achievable rate expression (i.e., an upper bound on the rate-recovery function) using a memoryless binning scheme. A corresponding lower bound is also proposed by establishing a connection to a multi-terminal source coding problem that captures the recovery constraints similar to our streaming setup. Both the upper and lower bound expressions are structurally similar — they involve a predictive coding rate, i.e., the rate required for the streaming of the source vectors through an ideal channel without erasure, plus a mutual information term that decays at least as fast as $1/(W+1)$. The upper and lower bounds do not coincide in general except for the case of $W = 0$ and $W \to \infty$, where we establish the actual rate-recovery function. We also show that for symmetric sources when restricted to memoryless encoders, the proposed memoryless binning scheme is optimal. In contrast, for a class of linear semi-deterministic Markov sources, we show that a suitable quantization and binning scheme attains a rate that matches our lower bound, thus
Chapter 1. Introduction

establishing the optimal scheme for this special class of sources.

In Chapter 3, we consider the important class of Gauss-Markov sources with quadratic distortion measure for the special case of $W = 0$, where the decoder is interested in reconstructing the source vectors immediately after the burst ends. While we again consider the memoryless quantize and binning scheme, the analysis of the achievable rate is more difficult. The main technical challenge in this case is locating the worst-case erasure burst from the rate and distortion constraint perspective. A lower bound is also developed based on the approach similar to that in discrete sources and exploiting connections to certain multi-terminal source coding problems. We show that the upper and lower bounds are very close for a wide range of problem parameters and indeed coincide in high resolution regime. In Chapter 3 we also consider an extension when the channel introduces multiple erasure bursts with a certain guard intervals separating these. We show that for relatively small values of guard period the achievable rate is very close to the single burst case.

In Chapter 4 we consider the case of Gauss-Markov source when $W > 0$, where the decoder is not interested in reconstructing the source vectors in a window of length $W$ after the erasure ends. We observe that for certain choices of parameters the predictive coding scheme can outperform the memoryless quantize and binning scheme. Motivated by this observation, we introduce a new hybrid coding scheme that involves successive quantization of sources followed by random binning. Through a suitable choice of parameters, the proposed scheme can outperform both predictive coding and memoryless quantize and binning and is provably optimal in certain special cases. We further provide simulation results over statistical channels such as the i.i.d. erasure channel and the Gilbert-Elliott channel, and demonstrate that the new hybrid scheme can provide significant gains over all baseline schemes for a wide range of parameters of practical interest.

Finally in Chapter 5 we consider the streaming setup where the decoder is allowed a delay of at most $T$. We study both the case of discrete sources and lossless reconstruction and Gauss-Markov sources with a quadratic distortion measure in this chapter. While some of our results are natural extensions of the zero-delay case others require new insights arising from the non-zero decoding delay requirements.

1.3 Notation

Throughout this thesis we represent the Euclidean norm operator by $|| \cdot ||$ and the expectation operator by $E[\cdot]$. The notation “log” is used for the binary logarithm, and rates are expressed in bits. The operations $H(\cdot)$ and $h(\cdot)$ denote the entropy and the differential entropy, respectively. The “slanted sans serif” font $a$ and the normal font $a$ represent random variables and their realizations respectively. The notation $a_i^n = \{a_{i,1}, \ldots, a_{i,n}\}$ represents a length-$n$ sequence of symbols at time $i$. The notation $[f]_i^j$ for $i < j$ represents $f_i, f_{i+1}, \ldots, f_j$. Capital bold letter, such as $A$, is used for matrices. The calligraphic font $\mathcal{A}$ represents sets. The notation $[x]^+$ is used for $\max\{0, x\}$.
Chapter 2

Zero-Delay Lossless Streaming

2.1 Introduction

In this chapter, as our first step to formulate the source streaming problem over erasure burst channels, we consider zero-delay streaming of discrete sources with a lossless recovery constraint at the destination. This is the simplest streaming setup considered in the thesis. The decoding analysis is much simpler than the case of lossy recovery as the reconstructed sequences also inherit the Markov property of the source process. We first define the source and channel model as well as the notion of lossless rate-recovery function that captures the inherent trade off between the compression-rate and the error-propagation in the presence of packet losses. General upper and lower bounds on the rate-recovery function are derived. The general upper bound (achievability) is based on memoryless binning at the encoder. For the class of symmetric sources and memoryless encoders this upper bound is shown to be tight. For another class of semi-deterministic sources we show that simple memoryless binning is sub-optimal and show that the rate-recovery function can be attained by another coding scheme.

In the rest of this chapter, we present the problem setup of lossless streaming of Markov sources in Section 2.2. The main results of this chapter are presented in Section 2.3. Section 2.4 presents the proof for the general upper and lower bounds on the lossless rate-recovery function. The case of symmetric sources and memoryless encoders is treated in Section 2.5, and the case of semi-deterministic Markov source and the proposed coding technique is treated in Section 2.6. Section 2.7 concludes the chapter.

2.2 Problem Statement

In this section we introduce our source and channel models and the associated definition of the rate-recovery function.
### 2.2.1 Source Model and Encoder

We assume that the communication spans the interval \( i \in \{-1, 0, 1, \ldots, \Upsilon\} \). At each time \( i \), a source vector \( \{s^n_i\} \) is sampled, whose symbols are drawn independently across the spatial dimension, and from a first-order Markov chain across the temporal dimension, i.e.,

\[
\Pr(s^n_i = s^n_{i-1} \mid s^n_{i-2}, s^n_{i-3}, \ldots, s^n_{-1}) = \prod_{k=1}^{n} p_1(s_{i,k} \mid s_{i-1,k}), \quad 0 \leq i \leq \Upsilon. \tag{2.1}
\]

The underlying random variables \( \{s_i\} \) constitute a time-invariant, stationary and a first-order Markov chain with a common marginal distribution denoted by \( p_{S}(\cdot) \) over an alphabet \( \mathcal{S} \). Throughout this thesis, we will treat the source process as many correlated source vectors drawn i.i.d. in spatial domain. The sequence \( s^n_{-1} \) is sampled i.i.d. from \( p_{S}(\cdot) \) and revealed to both the encoder and decoder before the start of the communication. It plays the role of a synchronization frame.

A rate-\( R \) encoder computes an index \( f_i \in \{1, 2, \ldots, 2^{nR}\} \) at time \( i \), according to an encoding function

\[
f_i = F_i(s^n_{-1}, s^n_0, \ldots, s^n_i), \quad 0 \leq i \leq \Upsilon. \tag{2.2}
\]

Note that the encoder in (2.2) is a causal function of the source sequences. A memoryless encoder satisfies \( F_i(\cdot) = F_i(s^n_i) \), i.e., the encoder does not use the knowledge of the past sequences. Naturally a memoryless encoder is very restrictive, and we will only use it to establish some special results.

### 2.2.2 Channel Model and Decoder

The channel takes each \( f_i \) as input and either outputs \( g_i = f_i \) or an erasure symbol, i.e., \( g_i = \ast \). We consider the class of erasure burst channels. For some particular \( j \geq 0 \), the channel introduces a erasure burst such that

\[
g_i = \begin{cases} 
\ast, & i \in \{j, j+1, \ldots, j + B' - 1\} \\
 f_i, & \text{otherwise},
\end{cases} \tag{2.3}
\]

![Error Propagation Window](image)

Figure 2.1: Problem Setup: The encoder output \( f_i \) is a function of all the past source sequences. The channel introduces a erasure burst of length up to \( B \). The decoder produces \( \hat{s}^n_i \) upon observing the channel outputs up to time \( i \). As indicated, the decoder is not required to produce those source sequences that are observed either during the erasure burst, or a period of \( W \) following it. The first sequence, \( s^n_{-1} \) is a synchronization frame available to both the source and destination.
where the burst length $B'$ is upper bounded by $B$.

Upon observing the sequence $\{g_i\}_{i \geq 0}$, the decoder is required to reconstruct each source sequence with zero delay, i.e.,

$$\hat{s}_i^n = G_i(g_0, g_1, \ldots, g_i, s_{i-1}^n), \quad i \notin \{j, \ldots, j + B' + W - 1\}$$

(2.4)

where $\hat{s}_i^n$ denotes the reconstruction sequence and $j$ denotes the time at which erasure burst starts in (2.3). The destination is not required to produce the source vectors that appear either during the erasure burst or in the period of length $W$ following it. We call this period the error propagation window. Fig. 2.1 provides a schematic of the causal encoder (2.2), the channel model (2.3), and the decoder (2.4).

### 2.2.3 Rate-Recovery Function

We define the lossless rate-recovery function under lossless and lossy reconstruction constraints. We assume that the source alphabet is discrete and the entropy $H(s)$ is finite. A rate $R_\Upsilon(B, W)$ is feasible if there exists a sequence of encoding and decoding functions and a sequence $\epsilon_n$ that approaches zero as $n \to \infty$ such that $\Pr(s^n_i \neq \hat{s}_i^n) \leq \epsilon_n$ for all source sequences reconstructed as in (2.4). We seek the minimum feasible rate $R_\Upsilon(B, W)$, which is the lossless rate-recovery function. In this chapter, we will focus on the infinite-horizon case, $R(B, W) = \lim_{\Upsilon \to \infty} R_\Upsilon(B, W)$, which will be called the rate-recovery function for simplicity.

**Remark 1.** Note that our proposed setup only considers a single erasure burst during the entire duration of communication. When we consider lossless recovery at the destination our results immediately extend to channels involving multiple erasure bursts with a certain guard interval separating consecutive bursts.

### 2.2.4 Practical Motivation

Note that our setup assumes that the size of both the source frames and channel packets is sufficiently large. A relevant application for the proposed setup is video streaming. Video frames are generated at a rate of approximately 60 Hz and each frame typically contains several hundred thousand pixels. The inter-frame interval is thus $\Delta_s \approx 17$ ms. Suppose that the underlying broadband communication channel has a bandwidth of $W_s = 2.5$ MHz. Then in the interval of $\Delta_s$ the number of symbols transmitted using ideal synchronous modulation is $N = 2\Delta_s W_s \approx 83,000$. Thus the block length between successive frames is sufficiently long that capacity approaching codes could be used and the erasure model and large packet sizes is justified. Our source model throughout the thesis implies that the sources are spatially independent but temporally dependent. While this is rarely an accurate statistical model for unprocessed frames of a video it is a reasonable approximation for the evolution of the video innovation process along optical-flow motion trajectories for groups of adjacent pixels (see [28] and reference therein). The temporal statistical dependence among the frames is assumed to be known here. In practice, this may be learned using offline training using video database by video-codec standardization groups such as H.26x and MPEG-x. Such source models have been widely used in earlier works e.g., [5, 14, 16–18].

Possible applications of the burst loss model considered in our setup include fading wireless channels and congested wired networks. We note that the present chapter does not consider a statistical channel model but instead considers a adversarial channel model. As mentioned before even the effect of such a single burst loss has not been well understood in the video streaming setup and therefore our proposed
setup is a natural starting point. Furthermore when we study Gaussian sources in subsequent chapters we will see that the coding schemes that result of such models also provide significant gains in the simulations involving statistical models. For a related approach in channel coding, see e.g., [29–32].

2.3 Main Results

In this section, we summarize the main results of the chapter.

2.3.1 Upper and Lower Bounds on Rate-Recovery

Theorem 1. (Lossless Rate-Recovery Function) For the stationary, first-order Markov, discrete source process, the lossless rate-recovery function satisfies the following upper and lower bounds:

\[ R^- (B, W) \leq R(B, W) \leq R^+ (B, W), \]

where

\[ R^+ (B, W) = H(s_1 | s_0) + \frac{1}{W+1} I(s_B; s_{B+1} | s_0), \]

\[ R^- (B, W) = H(s_1 | s_0) + \frac{1}{W+1} I(s_B; s_{B+W+1} | s_0). \] (2.5)

Notice that the upper and lower bounds (2.5) and (2.6) coincide for \( W = 0 \) and \( W \to \infty \), yielding the rate-recovery function in these cases. We can interpret the term \( H(s_1 | s_0) \) as the amount of uncertainty in \( s_1 \) when the past sources are perfectly known. This term is equivalent to the rate associated with ideal predictive coding in absence of any erasures. The second term in both (2.5) and (2.6) is the additional penalty that arises due to the recovery constraint following a erasure burst. Notice that this term decreases at-least as \( H(s) / (W + 1) \), thus the penalty decreases as we increase the recovery period \( W \). Note that the mutual information term associated with the lower bound is \( I(s_B; s_{B+W+1} | s_0) \) while that in the upper bound is \( I(s_B; s_{B+1} | s_0) \). Intuitively this difference arises because in the lower bound we only consider the reconstruction of \( s_{B+W+1} \) following an erasure bust in \( \{1, 2, \ldots, B\} \) while, as explained below in Corollary 1 the upper bound involves a binning based scheme that reconstructs all sequences \( (s_{B+1}, \ldots, s_{B+W+1}) \), though not required, at time \( t = B + W + 1 \).

A proof of Theorem 1 is provided in Section 2.4. The lower bound involves a connection to a multi-terminal source coding problem. This model captures the different requirements imposed on the encoder output following a erasure burst and in the steady state. The following Corollary provides an alternate expression for the achievable rate and makes the connection to the binning technique explicit.

Corollary 1. The upper bound in (2.5) is equivalent to the following expression

\[ R^+ (B, W) = \frac{1}{W+1} H(s_{B+1}, s_{B+2}, \ldots, s_{B+W+1} | s_0). \] (2.7)

The proof of Corollary 1 is provided in Appendix A.1. We make several remarks. First, the entropy term in (2.7) is equivalent to the sum-rate constraint associated with the Slepian-Wolf coding scheme in...
simultaneously recovering \( \{s^n_{B+1}, s^n_{B+2}, \ldots, s^n_{B+W+1}\} \) when \( s^n_0 \) is known. Note that due to the stationarity of the source process, the rate expression in (2.7) suffices for recovering from any erasure burst of length up to \( B \), spanning an arbitrary interval. Second, note that in (2.7) we amortize over a window of length \( W+1 \) as \( \{s^n_{B+1}, s^n_{B+2}, \ldots, s^n_{B+W+1}\} \) are recovered simultaneously at time \( t = B + W + 1 \). Note that this is the maximum window length over which we can amortize due to the decoding constraint. Third, the results in Theorem 1 immediately apply when the channel introduces multiple bursts with a guard spacing of at least \( W+1 \). This property arises due to the Markov nature of the source. Given a source sequence at time \( i \), all the future source sequences \( \{s^n_t\}_{t>i} \) are independent of the past \( \{s^n_t\}_{t<i} \) when conditioned on \( s^n_i \). Thus when a particular source sequence is reconstructed at the destination, the decoder becomes oblivious to past erasures. Finally, while the results in Theorem 1 are stated for the rate-recovery function over an infinite horizon, upon examining the proof of Theorem 1, it can be verified that both the upper and lower bounds hold for the finite horizon case, i.e., \( R_T(B,W) \), when \( T \geq B+W \).

### 2.3.2 Symmetric Sources and Memoryless Encoders

A **symmetric source** is defined as a Markov source such that the underlying Markov chain is also reversible, i.e., the random variables satisfy \( (s_0,...,s_l) \equiv (s_l,...,s_0) \), where the equality is in the sense of distribution [33]. Of particular interest to us is the following property satisfied for each \( t \):

\[
p_{s_{t+1},s_t}(s_a,s_b) = p_{s_{t-1},s_t}(s_a,s_b), \quad \forall s_a, s_b \in S,
\]

i.e., we can “exchange” the source pair \( (s^n_{t+1},s^n_t) \) with \( (s^n_{t-1},s^n_t) \) without affecting the joint distribution.

An example of a symmetric source is the binary symmetric source:

\[ s^n_t = s^n_{t-1} \oplus z^n_t, \]

where \( \{z^n_t\}_{t \geq 0} \) is an i.i.d. binary source process (in both temporal and spatial dimensions) with the marginal distribution \( \Pr(z_{t,i} = 0) = p \), the marginal distribution \( \Pr(s_{t,i} = 0) = \Pr(s_{t,i} = 1) = \frac{1}{2} \) and \( \oplus \) denotes modulo-2 addition.

**Corollary 2.** For the class of symmetric Markov sources that satisfy (2.8), the lossless rate-recovery function when restricted to the class of memoryless encoders, i.e., \( f_i = F_i(s^n_i) \), is given by

\[
R(B,W) = \frac{1}{W+1}H(s_{B+1}, s_{B+2}, \ldots, s_{B+W+1}|s_0).
\]

The proof of Corollary 2 is presented in Section 2.5. The converse is obtained by using a multi-terminal source coding problem, but obtaining a tighter bound by exploiting the memoryless property of the encoders and the symmetric structure (2.8).

### 2.3.3 Linear Semi-Deterministic Markov Sources

We propose a special class of source models — linear semi-deterministic Markov sources — for which the lower bound in (2.6) is tight. Our proposed coding scheme is most natural for a subclass of deterministic sources defined below.
Definition 1. (Linear Diagonally Correlated Deterministic Sources) The alphabet of a linear diagonally correlated deterministic source consists of $K$ sub-symbols, i.e.,
\[
s_i = (s_{i,0}, \ldots, s_{i,K}) \in S_0 \times S_1 \times \ldots \times S_K,
\] (2.10)
where each $S_j = \{0, 1\}^{N_j}$ is a binary sequence. The sub-sequence $\{s_{i,0}\}_{i \geq 0}$ is an i.i.d. sequence sampled uniformly over $S_0$ and for $1 \leq j \leq K$, the sub-symbol $s_{i,j}$ is a linear deterministic function of $s_{i-1,j-1}$, i.e.,
\[
s_{i,j} = R_{j,j-1} \cdot s_{i-1,j-1}, \quad 1 \leq j \leq K.
\] (2.11)
for fixed matrices $R_{1,0}, R_{2,1}, \ldots, R_{K,K-1}$ each of full row-rank, i.e., $\text{rank}(R_{j,j-1}) = N_j$ and $N_j \leq N_{j-1}$.

For such a class of sources we establish that the lower bound in Theorem 1 is tight and the binning based scheme is sub-optimal.

Proposition 1. For the class of Linear\(^2\) Diagonally Correlated Deterministic Sources in Def. 1 the rate-recovery function is also given by:
\[
R(B,W) = R^{-}(B,W) = H(s_1|s_0) + \frac{1}{W+1} I(s_B; s_{B+W+1}|s_0)
\] (2.12)
\[
= N_0 + \frac{1}{W+1} \sum_{k=1}^{\min(K-W,B)} N_{W+k}.
\] (2.13)

Sec. 2.6 provides the proof of Prop. 1. Our coding scheme exploits the special structure of such sources and achieves a rate that is strictly lower than the binning based scheme. We call this technique \textit{prospicient coding} because it exploits non-causal knowledge of some future symbols.

The proposed coding scheme can also be generalized to a broader class of semi-deterministic sources.

Definition 2. (Linear Semi-Deterministic Sources) The alphabet of a linear semi-deterministic source\(^3\) consists of two sub-symbols, i.e.,
\[
s_i = (s_{i,0}, s_{i,1}) \in S_0 \times S_1,
\] (2.14)
where each $S_j = \{0, 1\}^{N_j}$ for $j = 0, 1$. The sequence $\{s_{i,0}\}$ is an i.i.d. sequence sampled uniformly over $S_0$ whereas
\[
s_{i,1} = \begin{bmatrix} A & B \end{bmatrix} \cdot \begin{bmatrix} s_{i-1,0} \\ s_{i-1,1} \end{bmatrix}
\] (2.15)
for some fixed matrices $A$ and $B$.

We show that through a suitable invertible memoryless linear transform, this apparently more general source model can be transformed into a diagonally correlated deterministic Markov source. The prospicient coding can be applied to this class.

\(^1\)All multiplication is over the binary field.
\(^2\)The assumption of linearity in Def. 1 is not required to achieve the lower bound. However we use linearity to generalize to the class of semi-deterministic sources in Thm. 2.
\(^3\)Since each sub-symbol is a (fixed length) binary sequence we use the bold-face font $s_{i,j}$ to represent it. Similarly since each source symbol is a collection of sub-symbols we use a bold-face font to represent it. This should not be confused with a length $n$ source sequence at time $i$, which will be represented as $s_i^n$. 
Chapter 2. Zero-Delay Lossless Streaming

12

Theorem 2. For the class of Linear Semi-Deterministic Sources in Def. 2 the rate-recovery function is given by:

\[ R(B, W) = R^{-}(B, W) = H(s_1 | s_0) + \frac{1}{W+1} I(s_B, s_{B+W+1} | s_0). \]  \hspace{1cm} (2.16)

The proof of Theorem 2 is provided in Sec. 2.6.5.

2.4 Upper and Lower Bounds on Lossless Rate-Recovery Function

In this section we present the proof of Theorem 1. In particular, we show that the rate-recovery function satisfies the following lower bound.

\[ R \geq R^{-}(B, W) = H(s_1 | s_0) + \frac{1}{W+1} I(s_B, s_{B+W+1} | s_0), \]  \hspace{1cm} (2.17)

which is inspired by a connection to a multi-terminal source coding problem introduced in Section 2.4.1. Based on this connection, the proof of the lower bound in general form in (2.17) is presented in Section 2.4.2. Then by proposing a coding scheme based on random binning, we show in Section 2.4.3 that the following rate is achievable.

\[ R \geq R^{+}(B, W) = H(s_1 | s_0) + \frac{1}{W+1} I(s_B, s_{B+1} | s_0). \]  \hspace{1cm} (2.18)

2.4.1 Connection to Multi-terminal Source Coding Problem

We first present a multi-terminal source coding setup which captures the tension inherent in the streaming setup. We focus on the special case when \( B = 1 \) and \( W = 1 \). At any given time \( j \) the encoder output \( f_j \) must satisfy two objectives simultaneously: 1) if \( j \) is outside the error propagation period then the decoder should use \( f_j \) and the past sequences to reconstruct \( s^n_j \); 2) if \( j \) is within the recovery period then \( f_j \) must only help in the recovery of a future source sequence.

Fig. 2.2 illustrates the multi-terminal source coding problem with one encoder and two decoders that captures these constraints. The sequences \((s^n_j, s^n_{j+1})\) are revealed to the encoder which produces outputs...
are then generalized rather naturally in the formal proof of the lower bound in the next sub-section. Nevertheless the main steps of the lower bound developed in the multi-terminal setup are then generalized rather naturally in the formal proof of the lower bound in the next sub-section.

For the above multi-terminal problem, we establish a lower bound on the sum rate as follows:

\[
n(R_1 + R_2) \geq H(f_j, f_{j+1}) \\
\geq H(f_j, s_{j+1}^n | s_{j-2}^n) \\
= H(f_j, s_{j+1}^n | s_{j-2}^n) - H(s_{j+1}^n | f_j, f_{j+1}, s_{j-2}^n) \\
= H(f_j, s_{j+1}^n | s_{j-2}^n) + H(f_j | f_{j+1}, s_{j-2}^n, s_{j+1}^n) - H(s_{j+1}^n | f_j, f_{j+1}, s_{j-2}^n) \\
\geq H(f_j, s_{j+1}^n | s_{j-2}^n) - n\varepsilon_n \\
= H(s_{j+1}^n | s_{j-2}^n) + H(f_j | s_{j+1}^n, s_{j-2}^n) - n\varepsilon_n \\
\geq H(s_{j+1}^n | s_{j-2}^n) + H(f_j | s_{j+1}^n, s_{j-1}^n, s_{j-2}^n) - n\varepsilon_n \\
\geq H(s_{j+1}^n | s_{j-2}^n) + H(s_{j+1}^n | s_{j-1}^n, s_{j-2}^n) - 2n\varepsilon_n \\
= H(s_{j+1}^n | s_{j-2}^n) - 2n\varepsilon_n \\
= nH(s_1 | s_0) + nH(s_1 | s_2, s_0) - 2n\varepsilon_n \\
\]  

where (2.19) follows from the chain rule of entropy, (2.20) follows from the fact that \( s_{j+1}^n \) must be recovered from \( \{f_j, f_{j+1}, s_{j-2}^n\} \) at decoder 2 hence Fano’s inequality applies and (2.21) follows from the fact that conditioning never increases the entropy. Eq. (2.22) follows from Fano’s inequality applied to decoder 1 and (2.23) follows from the Markov chain associated with the source process. Finally (2.24) follows from the fact that the source process is memoryless. Dividing throughout by \( n \) in (2.24) and taking \( n \to \infty \) yields

\[
R_1 + R_2 \geq H(s_1 | s_0, s_2) + H(s_1 | s_0). \tag{2.25}
\]

**Tightness of Lower Bound:** As a side remark, we note that the sum-rate lower bound in (2.25) can be achieved if Decoder 1 is further given \( s_{j+1}^n \). Note that the lower bound (2.25) also applies in this case since the Fano’s Inequality applied to decoder 1 in (2.22) has \( s_{j+1}^n \) in the conditioning. We claim that \( R_1 = H(s_j | s_{j+1}, s_{j-1}) \) and \( R_2 = H(s_{j+1} | s_{j-2}) \) are achievable. The encoder can achieve \( R_1 \) by random binning of source \( s_j^n \) with \( \{s_{j-1}^n, s_{j+1}^n\} \) as decoder 1’s side information and achieve \( R_2 \) by random binning of source \( s_{j+1}^n \) with \( s_{j-2}^n \) as decoder 2’s side information. Thus revealing the additional side information of \( s_{j+1}^n \) to decoder 1, makes the link connecting \( f_j \) to decoder 2 unnecessary.

Also note that the setup in Fig. 2.2 reduces to the source coding problem in [34] if we set \( s_{j-2}^n = \phi \) where \( \phi \) is an empty set. It is also a successive refinement source coding problem with different side information at the decoders and special distortion constraints at each of the decoders. However to the best of our knowledge the multi-terminal problem in Fig. 2.2 has not been addressed in the literature nor has the connection to our proposed streaming setup been considered in earlier works.

In the streaming setup, the symmetric rate, i.e., \( R_1 = R_2 = R \), is of interest. Setting this in (2.25)
we obtain:

\[ R \geq \frac{1}{2} H(s_1|s_0, s_2) + \frac{1}{2} H(s_1|s_0). \]  

(2.26)

It can be easily shown that the expression in (2.26) and the right hand side of the general lower bound in (2.6) for \( B = W = 1 \) are the equivalent using a simple calculation.

\[ R - (B = 1, W = 1) = H(s_1|s_0) + \frac{1}{2} I(s_1; s_3|s_0) - \frac{1}{2} H(s_1|s_0, s_2) + \frac{1}{2} H(s_1|s_0) \]  

(2.27)

\[ = \frac{1}{2} H(s_2|s_0) + \frac{1}{2} H(s_1|s_0, s_2) - \frac{1}{2} H(s_1|s_0) \]  

(2.28)

\[ = \frac{1}{2} H(s_1|s_0, s_2) + \frac{1}{2} H(s_3|s_0) - \frac{1}{2} H(s_3|s_1) \]  

(2.29)

where the first term in (2.27) follows from the Markov Chain property \( s_0 \rightarrow s_1 \rightarrow s_2 \), the last term in (2.27) follows from the Markov Chain property \( s_1 \rightarrow s_2 \rightarrow s_3 \) and (2.29) follows from the fact that the source model is stationary, thus the first and last term in (2.28) are the same.

As noted before the above proof does not directly apply to the streaming setup as it does not take into account that the decoders have access to all the past encoder outputs, and that the encoder has access to all the past source sequences. We next provide a formal proof of the lower bound that shows that this additional information does not help.

### 2.4.2 Lower Bound on Lossless Rate-Recovery Function

For any sequence of \( (n, 2^{nR}) \) codes we show that there is a sequence \( \varepsilon_n \) that vanishes as \( n \to \infty \) such that

\[ R \geq H(s_1|s_0) + \frac{1}{W+1} I(s_{B+W+1}; s_B|s_0) - \varepsilon_n. \]  

(2.30)

We consider that an erasure burst of length \( B \) spans the interval \( \{t-B-W, \ldots, t-W-1\} \) for some \( t \geq B+W \). It suffices to lower bound the rate for this erasure pattern. By considering the interval \( \{t-W, \ldots, t\} \), following the erasure burst we have

\[ (W + 1)nR \geq H([f]_{t-W}^t) \]  

\[ \geq H([f]_{t-W}^t|f]_{0}^{t-B-W-1}, s_{n-1}^n), \]  

(2.31)

where (2.31) follows from the fact that conditioning never increases the entropy. By definition, the source sequence \( s^n_t \) must be recovered from \( \{[f]_{0}^{t-B-W-1}, [f]_{t-W}^t, s_{n-1}^n\} \). Applying Fano’s inequality we have that

\[ H(s^n_t|[f]_{0}^{t-B-W-1}, [f]_{t-W}^t, s_{n-1}^n) \leq n \varepsilon_n. \]  

(2.32)
Therefore we have
\[
H([f]_t | f_0^{t-B-W-1}, s^n) = H(s^n, [f]_t | f_0^{t-B-W-1}, s^n) - H(s^n | f_0^{t-B-W-1}, [f]_t, s^n) 
\geq H(s^n | f_0^{t-B-W-1}, s^n) + H([f]_t | s^n, f_0^{t-B-W-1}, s^n) - n \varepsilon_n,
\]
(2.33)

where (2.33) and the first two terms of (2.34) follow from the application of chain rule and the last term in (2.34) follows form (2.32). Now we bound each of the two terms in (2.34). First we note that:
\[
H(s^n | f_0^{t-B-W-1}, s^n) = H(s^n | f_0^{t-B-W-1}, s^n_{-B-W-1}, s^n) 
= H(s^n | s^n_{-B-W-1}) 
= H(s^n_{B+W+1}|s^n) 
= nH(s^n_{B+W+1}|s^n),
\]
(2.35)
(2.36)
(2.37)
(2.38)

where (2.35) follows from the fact that conditioning never increases entropy and (2.36) follows from the Markov relation
\[ (s^n_{-B-W-1}, f_0^{t-B-W-1}) \rightarrow s^n_{-B-W-1} \rightarrow s^n. \]

Eq. (2.37) and (2.38) follow from the stationary and memoryless source model.

Furthermore the second term in (2.34) can be lower bounded using the following series of inequalities.
\[
H([f]_t | s^n_t, f_0^{t-B-W-1}, s^n) \geq H([f]_t^{t-1} | s^n_t, f_0^{t-1}, s^n) 
= H([f]_t^{t-1} | s^n_t, f_0^{t-1}, s^n_{-1}) 
- H(s^n_t, s^n_{-1} | f_0^{t-1}, s^n) 
\geq H(s^n_t, s^n_{-1} | f_0^{t-1}, s^n_{-1}) - Wn\varepsilon_n 
\geq H(s^n_t, s^n_{-1} | f_0^{t-1}, s^n_{-1}) - Wn\varepsilon_n 
\geq H(s^n_t, s^n_{-1} | f_0^{t-1}, s^n_{-1}) - Wn\varepsilon_n 
\geq nH(s^n_{B+W+1}|s^n) - Wn\varepsilon_n 
\geq nH(s^n_{B+W+1}|s^n) - Wn\varepsilon_n 
\geq n(W+1)H(s_1|s_0) - nH(s_{B+W+1}|s_B) - Wn\varepsilon_n
\]
(2.39)
(2.40)
(2.41)
(2.42)
(2.43)
(2.44)

Note that in (2.39), in order to lower bound the entropy term, we reveal the codewords \([f]_t^{t-1}W-1\) which are not originally available at the decoder and exploit the fact that conditioning reduces the entropy. This step in deriving the lower bound may not necessarily be tight, however it is the best lower bound we have for the general problem. Also (2.40) follows from the fact that according to the problem setup \(\{s^n_{-W}, \ldots, s^n_{-1}\}\) must be decoded when \(s^n_{-1}\) and all the channel codewords before time \(t\), i.e., \([f]_t^{t-1}\), are available at the decoder, hence Fano’s inequality again applies. The expression (2.41) also follows from
the fact that conditioning never increases the entropy. Eq. (2.42) follows from the fact that
\[(s^n_1, [f]_0^{t-W-1}) \rightarrow s^n_{t-W} \rightarrow (s^n_{t-W}, \ldots, s^n_{t-1}). \tag{2.45}\]

Eq. (2.43) and (2.44) follow from memoryless and stationarity of the source sequences. Combining (2.34), (2.38) and (2.44) we have that
\[H ([f]_{t-W}^t \mid [f]_0^{t-B-W-1}, s^n_{t-1}) \geq nH(s_{B+W+1}^t | s^n_0) + n(W + 1)H(s^n_0) - nH(s_{B+W+1}^t | s_B) - (W + 1)n\varepsilon_n \tag{2.46}\]

Finally from (2.46) and (2.31) we have that,
\[nR \geq nH(s^n_0) + \frac{n}{W + 1} H(s_{B+W+1}^t | s^n_0) - H(s_{B+W+1}^t | s_B) - (W + 1)n\varepsilon_n \]
\[= nH(s^n_0) + \frac{n}{W + 1} [H(s_{B+W+1}^t | s^n_0) - (W + 1)n\varepsilon_n] \]
\[= nH(s^n_0) + \frac{n}{W + 1} I(s_{B+W+1}; s_B | s^n_0) - n\varepsilon_n \tag{2.47}\]

where the second step above follows from the Markov condition \(s_0 \rightarrow s_B \rightarrow s_{B+W+1}\). As we take \(n \rightarrow \infty\) we recover (2.30). This completes the proof of the lower bound in Theorem 1.

We remark that the derived lower bound holds for any \(t \geq B + W\). Therefore, the lower bound (2.30) on lossless rate-recovery function also holds for finite-horizon rate-recovery function whenever \(T \geq B + W\).

Finally we note that in our setup we are assuming a peak rate constraint on \(f_t\). If we assume the average rate constraint across \(f_t\) the lower bound still applies with minor modifications in the proof.

### 2.4.3 Upper Bound on Lossless Rate-Recovery Function

In this section we establish the achievability of \(R^+(B, W)\) in Theorem 1 using a binning based scheme. At each time the encoding function \(f_t\) in (2.2) is the bin-index of a Slepian-Wolf codebook [24, 35]. Following a burst erasure in \(\{j, \ldots, j + B - 1\}\), the decoder collects \(f_{j+B}, \ldots, f_{j+W+B}\) and attempts to jointly recover all the underlying sources at \(t = j + W + B\). Using Corollary 1 it suffices to show that
\[R^+ = \frac{1}{W + 1} H(s_{B+1}, \ldots, s_{B+W+1} | s^n_0) + \varepsilon \tag{2.48}\]
is achievable for any arbitrary \(\varepsilon > 0\).

We use a codebook \(\mathcal{C}\) which is generated by randomly partitioning the set of all typical sequences \(T^n(s)\) into \(2^nR^+\) bins. The partitions are revealed to the decoder ahead of time.

Upon observing \(s^n_t\) the encoder declares an error if \(s^n_t \notin T^n(s)\). Otherwise it finds the bin to which \(s^n_t\) belongs to and sends the corresponding bin index \(f_t\). We separately consider two possible scenarios at the decoder.

First suppose that the sequence \(s^n_{t-1}\) has already been recovered. Then the destination attempts to recover \(s^n_t\) from \((f_t, s^n_{t-1})\). This succeeds with high probability if \(R^+ > H(s^n_1 | s^n_0)\), which is guaranteed via (2.48). If we define probability of the error event \(E_i \triangleq \{s^n_t \neq s^n_i\}\) conditioned on the correct recovery of \(s^n_{t-1}\), i.e., \(E_{t-1}\), as follows
\[P_{E_{t-1}}^{(n)} \triangleq P(E_i | E_{t-1}) \tag{2.49}\]
then for the rates satisfying $R^+ > H(s_1|s_0)$ and in particular for $R^+$ in (2.48), it is guaranteed that

$$\lim_{n \to \infty} P_{e,1}^{(n)} = 0. \quad (2.50)$$

Next consider the case where $s^n_i$ is the first sequence to be recovered after the erasure burst. In particular the erasure burst spans the interval $\{i - B', \ldots, i - W - 1\}$ for some $B' \leq B$. The decoder thus has access to $s^n_{i - B' - W - 1}$, before the start of the erasure burst. Upon receiving $f_{i-W}, \ldots, f_i$ the destination simultaneously attempts to recover $(s^n_{i-W}, \ldots, s^n_i)$ given $(s^n_{i-B' - W - 1}, f_{i-W}, \ldots, f_i)$. This succeeds with high probability if

$$(W + 1)nR = \sum_{j = i-W}^i H(f_j) \quad (2.51)$$

$$> nH(s_{i-W}, \ldots, s_i|s_{i-B' - W - 1}) \quad (2.52)$$

$$= nH(s_{B' + 1}, \ldots, s_{B' + W + 1}|s_i). \quad (2.53)$$

where (2.53) follows from the fact that the sequence of variables $s_i$ is a stationary process. Whenever $B' \leq B$ it immediately follows that (2.53) is also guaranteed by (2.48). Define $P_{e,2}^{(n)}$ as the probability of error in $s^n_i$ given $(s^n_{i-B' - W - 1}, f_{i-W}, \ldots, f_i)$, i.e.,

$$P_{e,2}^{(n)} \triangleq P(\bar{E}_i|\bar{E}_{i-B'-W-1}). \quad (2.54)$$

For a rate $R$ satisfying (2.53), which is satisfied through (2.48), it is guaranteed that

$$\lim_{n \to \infty} P_{e,2}^{(n)} = 0. \quad (2.55)$$

**Analysis of the Streaming Decoder:** As described in problem setup, the decoder is interested in recovering all the source sequences outside the error propagation window with vanishing probability of error. Assume a communication duration of $T$ and a single erasure burst of length $0 < B' \leq B$ spanning the interval $\{j, \ldots, j + B' - 1\}$, for $0 \leq j \leq T$. The decoder fails if at least one source sequence outside the error propagation window is erroneously recovered, i.e., $\hat{s}^n_i \neq s^n_i$ for some $i \in \{0, \ldots, j - 1\} \cup \{j + B' + W + 1, \ldots, T\}$. For this particular channel erasure pattern, the probability of decoder’s failure, denoted by $P_{F}^{(n)}$, can be bounded as follows.

$$P_{F}^{(n)} \leq \sum_{k=0}^{j-1} P(\bar{E}_k|\bar{E}_0, \bar{E}_1, \ldots, \bar{E}_{k-1}) + P(\bar{E}_{j+B'+W+1}|\bar{E}_0, \ldots, \bar{E}_{j-1}) +$$

$$\sum_{k=j+B'+W+2}^{T} P(\bar{E}_k|\bar{E}_0, \ldots, \bar{E}_{j-1}, \bar{E}_{j+B'+W+1}, \ldots, \bar{E}_{k-1}) \quad (2.56)$$

$$= (\Upsilon - B' - W)P_{e,1}^{(n)} + P_{e,2}^{(n)} \leq \Upsilon P_{e,1}^{(n)} + P_{e,2}^{(n)} \quad (2.57)$$

where $P_{e,1}^{(n)}$ and $P_{e,2}^{(n)}$ are defined in (2.49) and (2.54). Eq. (2.57) follows from the fact that, because of the Markov property of the source model, all the terms in the first and the last summation in (2.56) are the same and equal to $P_{F}^{(n)}$.

According to (2.50) and (2.55), for any rate satisfying (2.48) and for any $\Upsilon$, $n$ can be chosen large enough such that the upper bound on $P_{F}^{(n)}$ in (2.57) approaches zero. Thus the decoder fails with
vanishing probability for any fixed $\Upsilon$. This in turn establishes the upper bound on $R(B, W)$, when $\Upsilon \to \infty$. This completes the justification of the upper bound.

2.5 Symmetric Sources and Memoryless Encoders

In this section we establish that the lossless rate-recovery function for symmetric Markov sources restricted to class of memoryless encoders is given by

$$R(B, W) = \frac{1}{W + 1} H(s_{B+1}, \ldots, s_{B+W+1}|s_0).$$  \hspace{1cm} (2.58)

The achievability follows from Theorem 1 and Corollary 1. We thus only need to prove the converse to improve upon the general lower bound in (2.6). The lower bound for the special case when $W = 0$ follows directly from (2.6) and thus we only need to show the lower bound for $W \geq 1$. For simplicity in exposition we illustrate the case when $W = 1$. Then we need to show that

$$R(B, W = 1) \geq \frac{1}{2} H(s_{B+1}, s_{B+2}|s_0)$$  \hspace{1cm} (2.59)

The proof for general $W > 1$ will follow along similar lines and will be sketched later.

Assume that an erasure burst spans time indices $j-B, \ldots, j-1$. The decoder must recover

$$\hat{s}_{j+1}^n = G_{j+1}\left([f]_{0}^{j-1-B}, f_j, f_{j+1}, s_{j-1}^n\right).$$  \hspace{1cm} (2.60)

Furthermore if there is no erasure until time $j$ then

$$\hat{s}_j^n = G_j\left([f]_0^n, s_{j-1}^n\right)$$  \hspace{1cm} (2.61)
must hold. Our aim is to establish the following lower bound on the sum-rate:

\[ 2R \geq H(s_{j+1}|s_j) + H(s_j|s_{j-B-1}). \]  

(2.62)

The lower bound (2.59) then follows since

\[ R \geq \frac{1}{2}(H(s_{j+1}|s_j) + H(s_j|s_{j-B-1})) \]
\[ = \frac{1}{2}(H(s_{j+1}|s_j, s_{j-B-1}) + H(s_j|s_{j-B-1})) \]  
\[ = \frac{1}{2} H(s_{j+1}, s_j|s_{j-B-1}) = \frac{1}{2} H(s_{B+1}, s_{B+2}|s_0), \]  

(2.63)

(2.64)

where (2.63) follows from the Markov chain property \( s_{j-B-1} \rightarrow s_j \rightarrow s_{j+1} \), and the last step in (2.64) follows from stationarity of the source model.

To establish (2.62) we make a connection to a multi-terminal source coding problem in Fig. 2.3(a). We accomplish this in several steps as outlined below.

### 2.5.1 Multi-Terminal Source Coding

Consider the multi-terminal source coding problem with side information illustrated in Fig. 2.3(a). In this setup there are four source sequences drawn i.i.d. from a joint distribution \( p(s_{j+1}, s_j, s_{j-1}, s_{j-B-1}) \).

The two source sequences \( s^n_j \) and \( s^n_{j+1} \) are revealed to the encoders \( j \) and \( j+1 \) respectively and the two sources \( s^n_{j-1} \) and \( s^n_{j-B-1} \) are revealed to the decoders \( j \) and \( j+1 \) respectively. The encoders operate independently and compress the source sequences to \( f_j \) and \( f_{j+1} \) at rates \( R_j \) and \( R_{j+1} \) respectively. Decoder \( j \) has access to \( (f_j, s^n_{j-1}) \) while decoder \( j+1 \) has access to \( (f_{j+1}, s^n_{j-B-1}) \). The two decoders are required to reconstruct

\[ \hat{s}^n_j = \hat{G}_j(f_j, s^n_{j-1}) \]
\[ \hat{s}^n_{j+1} = \hat{G}_{j+1}(f_{j+1}, s^n_{j-B-1}) \]  

(2.65)

(2.66)

respectively such that \( \Pr(\hat{s}^n_i \neq s^n_i) \leq \varepsilon_n \) for \( i = \{j, j+1\} \).

Note that the multi-terminal source coding setup in Fig. 2.3(a) is similar to the setup in Fig. 2.2, except that the encoders do not cooperate and \( \hat{f}_i = \mathcal{F}_i(s^n_i) \), due to the memoryless property. We exploit this property to directly show that a lower bound on the multi-terminal source coding setup in Fig. 2.3(a) also constitutes a lower bound on the rate of the original streaming problem.

**Lemma 1.** For the class of memoryless encoding functions, i.e., \( f_j = \mathcal{F}_j(s^n_i) \), the decoding functions \( \hat{s}^n_j = \hat{G}_j([f]_{0}^{j-2}, f_j, s^n_{j-2}) \) and \( \hat{s}^n_{j+1} = \hat{G}_{j+1}(f_{j+1}, s^n_{j-B-1}) \) can be replaced by the following decoding functions

\[ \tilde{s}^n_j = \tilde{G}_j(f_j, s^n_{j-1}) \]
\[ \tilde{s}^n_{j+1} = \tilde{G}_{j+1}(f_{j+1}, s^n_{j-B-1}) \]  

(2.67)

(2.68)

such that

\[ \Pr(\tilde{s}^n_i \neq s^n_i) \leq \Pr(\hat{s}^n_i \neq s^n_i) \]
\[ \Pr(\tilde{s}^n_{j+1} \neq s^n_{j+1}) \leq \Pr(\hat{s}^n_{j+1} \neq s^n_{j+1}). \]  

(2.69)

(2.70)
Proof. Assume that the extra side-informations $s_{j-1}^n$ is revealed to the decoder $j$. Now define the maximum a posteriori probability (MAP) decoder as follow.

$$\tilde{s}_j^n = \overline{G}_j([f]_0^j, s_{j-1}^n) \triangleq \arg\max_{s_j^n} p(s_j^n|[f]_0^j, s_{j-1}^n, s_{j-1}^n)$$

(2.71)

where we dropped the subscript in the conditional probability density for the sake of simplicity. It is known that the MAP decoder is optimal and minimizes the decoding error probability, therefore

$$\Pr(\tilde{s}_j^n \neq s_j^n) \leq \Pr(\hat{s}_j^n \neq s_j^n).$$

(2.72)

Also note that

$$\tilde{s}_j^n = \overline{G}_j([f]_0^j, s_{j-1}^n) = \arg\max_{s_j^n} p(s_j^n|[f]_0^j, s_{j-1}^n)$$

(2.73)

$$= \arg\max_{s_j^n} p(s_j^n|f_j, s_{j-1}^n)$$

(2.74)

$$\triangleq \tilde{G}_j(f_j, s_{j-1}^n),$$

(2.75)

where (2.74) follows from the following Markov property:

$$([f]_0^{j-1}, s_{j-1}^n) \rightarrow (f_j, s_{j-1}^n) \rightarrow s_j^n.$$  

(2.76)

It can be shown through similar steps that the decoder defined in (2.68) exists with the error probability satisfying (2.70). This completes the proof.

The conditions in (2.67) and (2.68) show that any rate that is achievable in the streaming problem in Fig. 2.1 is also achieved in the multi-terminal source coding setup in Fig. 2.3(a). Hence a lower bound to this source network also constitutes a lower bound to the original problem. In the next section we find a lower bound on the rate for the setup in Fig. 2.3(a).

### 2.5.2 Lower Bound for Multi-terminal Source Coding Problem

In this section, we establish a lower bound on the sum-rate of the multi-terminal source coding setup in Fig. 2.3(a), i.e., $R \geq \frac{1}{2} H(s_{B+1}, s_{B+2}|s_0)$. To this end, we observe the equivalence between the setup in Fig. 2.3(a) and Fig. 2.3(b) as stated below.

**Lemma 2.** The set of all achievable rate-pairs $(R_j, R_{j+1})$ for the problem in Fig. 2.3(a) is identical to the set of all achievable rate-pairs for the problem in Fig. 2.3(b) where the side information sequence $s_{j-1}^n$ at decoder 1 is replaced by the side information sequence $s_{j+1}^n$.

The proof of Lemma 2 follows by observing that the capacity region for the problem in Fig. 2.3(a) depends on the joint distribution $p(s_j, s_{j+1}, s_{j-1}, s_{j-B-1})$ only via the marginal distributions $p(s_j, s_{j-1})$ and $p(s_{j+1}, s_{j-B-1})$. Indeed the decoding error at decoder $j$ depends on the former whereas the decoding error at decoder $j + 1$ depends on the latter. When the source is symmetric, the joint distributions $p(s_j, s_{j-1})$ and $p(s_j, s_{j+1})$ are identical and thus exchanging $s_{j-1}^n$ with $s_{j+1}^n$ does not change the
error probability at decoder $j$ and leaves the functions at all other terminals unchanged. The formal proof is straightforward and will be omitted.

Thus it suffices to lower bound the achievable sum-rate for the problem in Fig. 2.3(b). First note that

\begin{align*}
nR_{j+1} &= H(f_{j+1}) \\
&\geq I(f_{j+1}; s^n_{j+1}| s^n_{j-B-1}, f_j) \\
&= H(s^n_{j+1}| s^n_{j-B-1}, f_j) - H(s^n_{j+1}| s^n_{j-B-1}, f_j, f_{j+1}) \\
&\geq H(s^n_{j+1}| s^n_{j-B-1}, f_j) - n\varepsilon_n
\end{align*}

(2.77)

where (2.77) follows by applying Fano’s inequality for decoder $j+1$ in Fig. 2.3(b) since $s^n_{j+1}$ can be recovered from $(s^n_{j-B-1}, f_j, f_{j+1})$. To bound $R_j$

\begin{align*}
nR_j &= H(f_j) \\
&\geq I(f_j; s^n_j| s^n_{j-B-1}) \\
&= H(s^n_j| s^n_{j-B-1}) - H(s^n_j| s^n_{j-B-1}, f_j) \\
&\geq nH(s_j| s_{j-B-1}) - H(s^n_j| s^n_{j-B-1}, f_j) + H(s^n_j| s^n_{j-B-1}, s^n_{j+1}, f_j) - n\varepsilon_n
\end{align*}

(2.78)

(2.79)

where (2.78) follows by applying Fano’s inequality for decoder $j$ in Fig. 2.3(b) since $s^n_j$ can be recovered from $(s^n_{j+1}, f_j)$ and hence $H(s^n_j| s^n_{j-B-1}, s^n_{j+1}, f_j) \leq n\varepsilon_n$ holds and (2.79) follows from the Markov relation $s^n_{j+1} \rightarrow s^n_j \rightarrow (f_j, s^n_{j-B-1})$. By summing (2.77) and (2.79) and using $R_j = R_{j+1} = R$, we have

\begin{align*}
R_j + R_{j+1} &\geq H(s_{j+1}| s_j) + H(s_j| s_{j-B-1}) \\
&= H(s_j, s_{j+1}| s_{j-B-1}).
\end{align*}

(2.80)

(2.81)

which is equivalent to (2.62).

**Remark 2.** One way to interpret the lower bound in (2.81) is by observing that the decoder $j+1$ in Fig. 2.3(b) is able to recover not only $s^n_{j+1}$ but also $s^n_j$. In particular, the decoder $j+1$ first recovers $s^n_{j+1}$. Then, similar to decoder $j$, it also recovers $s^n_j$ from $f_j$ and $s^n_{j+1}$ as side information. Hence, by only considering decoder $j+1$ and following standard source coding argument, the lower bound on the sum-rate satisfies (2.81).

### 2.5.3 Extension to Arbitrary $W > 1$

To extend the result for arbitrary $W$, we use the following result which is a natural generalization of Lemma 1.

**Lemma 3.** Consider memoryless encoding functions $f_k = F_k(s^n_k)$ for $k \in \{j, \ldots, j+W\}$. Any set of
decoding functions
\[
\hat{s}_k^n = G_k([f]_0^k, s_{n-1}^n) \quad k \in \{j, \ldots, j+W-1\} \\
\hat{s}_{j+W}^n = G_{j+W}([f]_{j-B-1}^j, [f]_{j+W}^j, s_{n-1}^n)
\]

(2.82)

(2.83)

can be replaced by a new set of decoding functions as
\[
\tilde{s}_k^n = \tilde{G}_k(f_k, s_{n-1}^k) \quad k \in \{j, \ldots, j+W-1\} \\
\tilde{s}_{j+W}^n = \tilde{G}_{j+W}(s_{j-B-1}^n, [f]_{j}^{j+W})
\]

(2.84)

(2.85)

where
\[
Pr(\tilde{s}_n^l \neq s_n^l) \leq Pr(\hat{s}_n^l \neq s_n^l) \quad j \leq l \leq j+W.
\]

(2.86)

The proof is an immediate extension of Lemma 1 and is excluded here. The lemma suggests a natural multi-terminal problem for establishing the lower bound with \(W+1\) encoders and decoders. For concreteness we discuss the case when \(W = 2\). Consider three encoders \(t \in \{j, j+1, j+2\}\). Encoder \(t\) observes \(s_n^t\) and compresses it into an index \(f_t \in \{1, \ldots, 2^nR_t\}\). The sequence \(s_{n-1}^t\) for \(t \in \{j, j+1\}\) are revealed to the corresponding decoders and \(s_{j-B-1}^n\) is revealed to the decoder \(j+2\). Using an argument analogous to Lemma 2, the rate region is equivalent to the case when \(s_{j+1}^n\) and \(s_{j+2}^n\) are instead revealed to decoders \(j\) and \(j+1\) respectively. For this new setup we can argue that decoder \(j+2\) can always reconstruct \((s_n^j, s_{n+1}^j, s_{n+2}^j)\) given \((s_{j-B-1}^n, f_j, f_{j+1}, f_{j+2})\). In particular, following the same argument in Remark 2, the decoder \(j+2\) first recovers \(s_{j+2}^n\), then using \(f_{j+1}, s_{j+2}^n\) recovers \(s_{j+1}^n\) and finally using \(f_j, s_{j+1}^n\) recovers \(s_j^n\). And hence if we only consider decoder \(j+2\) with side information \(s_{j-B-1}^n\) the sum-rate must satisfy:
\[
3R = R_j + R_{j+1} + R_{j+2} \geq H(s_j, s_{j+1}, s_{j+2}|s_{j-B-1}).
\]

(2.87)

Using Lemma 3 for \(W = 2\) it follows that the proposed lower bound also continues to hold for the original streaming problem. This completes the proof. The extension to an arbitrary \(W\) is completely analogous.

2.6 Semi-Deterministic Markov Sources and Prospicient Coding

We establish Prop. 1 in this section.

2.6.1 Diagonally Correlated Semi-Deterministic Source Model

We consider the semi-deterministic source model with a special diagonal correlation structure as described in Def. 1. The diagonal correlation structure appears to be the most natural structure to consider in developing insights into our proposed coding scheme. As we will see later in Theorem 2, the underlying coding scheme can also be generalized to a broader class of linear semi-deterministic
sources. Furthermore this class of semi-deterministic sources also provides a solution to the Gaussian source model as discussed in Theorem 3 in Chapter 3.

We first provide an alternate characterization of the sources defined in Def. 1. Let us define

\[ R_{k,l} = R_{k,k-1}R_{k-1,k-2} \cdots R_{l+2,l+1}R_{l+1,l}, \quad (2.88) \]

where \( k > l \). Note that since each \( R_{j,j-1} \) is assumed to have a full row-rank (c.f. Def. 1) the matrix \( R_{k,l} \) is a \( N_k \times N_l \) full-rank matrix of rank \( N_k \). From Def. 1

\[ s_i = \begin{pmatrix} s_{i,0} \\ R_{1,0} s_{i-1,0} \\ R_{2,0} s_{i-2,0} \\ \vdots \\ R_{K,0} s_{i-K,0} \end{pmatrix}, \quad (2.89) \]

where \( \{s_{i-K,0}, s_{i-K+1,0}, \ldots, s_{i,0}\} \) are innovation sub-symbols of each source. This is expressed in Fig. 2.4. Any diagonal in Fig. 2.4 consists of the same set of innovation bits. In particular the innovation bits are introduced on the upper-left most entry of the diagonal. As we traverse down, each sub-symbol consists of some fixed linear combinations of these innovation bits. Furthermore the sub-symbol \( s_{i,j} \) is completely determined given the sub-symbol \( s_{i-1,j-1} \).

In this section, we first argue that analyzing the coding scheme for the case \( K = B + W \) is sufficient.
Then we explain the prospicient coding scheme which achieves the rate specified in (2.13). Finally, the proof of the rate-optimality of the prospicient coding scheme is provided by establishing the equality of the rate expression (2.13) and the general lower bound in (2.12).

2.6.2 Sufficiency of $K = B + W$

We first argue that for our coding scheme, it suffices to assume that each source symbol $s_i$ consists of one innovation sub-symbol and a total of $K = B + W$ deterministic symbols. In particular when $K < B + W$, by simply adding $K - B - W$ zeros, the source can be turned into a source with $B + W$ deterministic sub-symbols.

For the case $K > B + W$ we argue that it suffices to construct a coding scheme with $K = B + W$. The remainder of the sub-symbols can be trivially computed by the receiver. In particular, at any time $i$, either $s_{i-1}$ or $s_{i-B-W-1}$ is guaranteed to be available to the destination. In the former case, except the innovation bits of $s_i$, all other bits are known. Thus all the deterministic sub-symbols, including those corresponding to $K > B + W$ can be computed. In the latter case, because of the diagonal structure of the source, the sub-symbols $s_{i,j}$, for $j \geq B + W + 1$, are deterministic functions of $s_{i-B-W-1}$ (c.f. (2.89)), and therefore, are known and can be ignored. Thus without loss of generality we assume that $K = B + W$ is sufficient.

2.6.3 Prospicient Coding

Our coding scheme is based on the following observation, illustrated in Fig. 2.5. Suppose that an erasure happens between $t \in \{i-W-B-1, \ldots, i-W-1\}$ and after the “don’t care” period of $\{i-W, \ldots, i-1\}$ we need to recover $s^n_i$. Based on the structure of the source, illustrated in Fig. 2.5 we make the following
observations:

- Sub-symbols \( \{ \mathbf{s}_{i,1}, \ldots, \mathbf{s}_{i,W} \} \) can be directly computed from the innovation sub-symbols \( \{ \mathbf{s}_{i-1,0}, \ldots, \mathbf{s}_{i-W,0} \} \) respectively.

- Sub-symbols \( \{ \mathbf{s}_{i,W+1}, \ldots, \mathbf{s}_{i,W+B} \} \) can be computed from sub-symbols \( \{ \mathbf{s}_{i-W,1}, \ldots, \mathbf{s}_{i-W,B} \} \) respectively.

Thus if we send the first \( B + 1 \) sub-symbols at each time, i.e., \( \mathbf{x}_i = (\mathbf{s}_{i,0}, \ldots, \mathbf{s}_{i,B}) \) then we are guaranteed that the destination will be able to decode \( \mathbf{s}_{i,W} \) when an erasure happens between \( \{i - B - W, \ldots, i - W - 1\} \). To achieve the optimal rate, we further compress \( \mathbf{x}_i \) as discussed below. Our coding scheme consists of two steps.

**Source Re-arrangement**

The source symbols \( \mathbf{s}_i \) consisting of innovation and deterministic sub-symbols as in Def. (1) are first rearranged to produce an auxiliary set of codewords

\[
\mathbf{c}_i = \begin{pmatrix} c_{i,0} \\ c_{i,1} \\ c_{i,2} \\ \vdots \\ c_{i,B} \end{pmatrix} \begin{pmatrix} \mathbf{s}_{i,0} \\ \mathbf{s}_{i+W,W+1} \\ \mathbf{s}_{i+W,W+2} \\ \vdots \\ \mathbf{s}_{i+W,W+B} \end{pmatrix} \begin{pmatrix} \mathbf{s}_{i,0} \\ \mathbf{R}_{W+1,1} \mathbf{s}_{i,1} \\ \mathbf{R}_{W+2,2} \mathbf{s}_{i,2} \\ \vdots \\ \mathbf{R}_{W+B,B} \mathbf{s}_{i,B} \end{pmatrix},
\]

(2.90)

where the last relation follows from (2.89).

Note that the codeword \( \mathbf{c}_i \) consists of the innovation symbol \( \mathbf{s}_{i,0} \), as well as symbols

\[ \{ \mathbf{s}_{i+W,W+1}, \ldots, \mathbf{s}_{i+W,W+B} \} \]

that enable the recovery of symbols in \( \mathbf{s}_{i+W} \).

It can be verified from (2.90) that the rate associated with the codewords \( \mathbf{c}_i \) is given by

\[
R_0 = N_0 + \sum_{k=W+1}^{W+B} N_k,
\]

(2.91)
which is larger than the rate-expression in (2.13). In particular it is missing the \( \frac{1}{W+1} \) factor in the second term. This factor can be recovered by binning the sequences \( \mathbf{c}_i^n \) as described next.

**Slepian-Wolf Coding**

There is a strong temporal correlation between the sequences \( \mathbf{c}_i^n \) in (2.90). As shown in Fig 2.6 as we proceed along any diagonal the sub-symbols \( \mathbf{c}_{i,j} \) and \( \mathbf{c}_{i+1,j+1} \) contain the same underlying set of innovation bits, i.e., from sub-symbol \( \mathbf{s}_{i-j,0} \).

To exploit the correlation, we independently bin the codeword sequences \( \mathbf{c}_i^n \) into \( 2^{nR} \) bins at each time. We let \( R = R(B,W) + \varepsilon \) is as given in (2.13), and only transmits the bin index of the associated codeword, i.e., \( f_i = \mathcal{F}(\mathbf{c}_i^n) \in \{1,2,\ldots,2^{nR}\} \).

It remains to show that given the bin index \( f_i \), the decoder is able to recover the underlying codeword symbols \( \mathbf{c}_i^n \).

**Analysis of Slepian-Wolf Coding**

Recall that we only transmit the bin index \( f_i \) of \( \mathbf{c}_i^n \). The receiver first recovers the underlying sequence \( \mathbf{c}_i^n \) as follows:

1) If the receiver has access to \( \mathbf{s}_{i-1} \) in addition to \( f_i \) it can recover \( \mathbf{c}_i^n \) if

\[
R \geq H(\mathbf{c}_i|\mathbf{s}_{i-1}) = H(\mathbf{c}_i,0) = N_0. \tag{2.92}
\]

where the second quality follows since \( \mathbf{c}_{i,1}, \ldots, \mathbf{c}_{i,W} \) are all deterministic functions of \( \mathbf{s}_{i,1}, \ldots, \mathbf{s}_{i,W} \), which in turn are deterministic functions of \( \mathbf{s}_{i-1} \). Clearly (2.92) is satisfied by our choice of \( R \) in (2.13).

2) The decoder has access to \( \mathbf{s}_{i-B-W-1} \) and \( \{f_{i-W}, f_{i-W+1}, \ldots, f_i\} \). The decoder is able to recover \( \{\mathbf{c}_{i-W}, \ldots, \mathbf{c}_i\} \) if

\[
(W + 1)R \geq H(\mathbf{c}_i, \mathbf{c}_{i-1}, \ldots, \mathbf{c}_{i-W} | \mathbf{s}_{i-B-W-1})
\]

\[
= \sum_{k=0}^{W+1} H(\mathbf{c}_{i-k,0}) + \sum_{k=1}^{B} H(\mathbf{c}_{i-W,k}) \tag{2.93}
\]

\[
= (W + 1)N_0 + \sum_{k=1}^{B} N_{W+k}, \tag{2.94}
\]

where (2.93) comes from the diagonal correlation property illustrated in Fig. 2.6. Our choice of \( R \) (2.13) guarantees that (2.94) is satisfied.

**2.6.4 Rate-Optimality of the Coding Scheme**

We specialize the general lower bound established in Theorem 1 to the case of diagonally correlated deterministic sources. Using (2.89) and \( p = B + W + 1 \) we have

\[
R \geq H(\mathbf{s}_i|\mathbf{s}_0) + \frac{1}{W+1}I(\mathbf{s}_p; \mathbf{s}_B|\mathbf{s}_0)
\]

\[
= H(\mathbf{s}_i|\mathbf{s}_{i-1}) + \frac{1}{W+1} \{H(\mathbf{s}_i|\mathbf{s}_{i-p}) - H(\mathbf{s}_i|\mathbf{s}_{i-W-1})\}
\]
Chapter 2. Zero-Delay Lossless Streaming

\[ H(s_i|s_{i-1}) + \frac{1}{W+1} H(s_{i,0}, R_{1,0}s_{i-1,0}, \ldots, R_{p-1,0}s_{i-p+1,0}) \]

\[ - \frac{1}{W+1} H(s_{i,0}, R_{1,0}s_{i-1,0}, \ldots, R_{W,0}s_{i-W,0}) \]  

(2.95)

According to the fact that innovation bits of each source are drawn i.i.d. (2.95) reduces to

\[ R \geq H(s_i, 0) + \frac{1}{W+1} \left( H(s_i, 0) + \sum_{k=1}^{p-1} H(R_{k,0}s_{i-k,0}) \right) - \frac{1}{W+1} \left( H(s_i, 0) + \sum_{k=1}^{W} H(R_{k,0}s_{i-k,0}) \right) \]

(2.96)

\[ = N_0 + \frac{1}{W+1} \left( \sum_{k=1}^{p-1} N_k - \sum_{k=1}^{W} N_k \right) \]

(2.97)

\[ = N_0 + \frac{1}{W+1} \sum_{k=W+1}^{p-1} N_k, \]  

(2.98)

where (2.97) follows from the fact that \( R_{k,0} \) are \( N_k \times N_0 \) full-rank matrices of rank \( N_k \). Since (2.98) equals (2.13) the optimality of the proposed scheme is established.

### 2.6.5 General Linear Semi-Deterministic Sources

We consider the class of linear deterministic sources as defined in Def. 2 in this section. Recall that for such a source the deterministic component \( s_{i,d} \) is obtained from the previous sub-symbol \( s_{i-1} \) through a linear transformation, i.e.,

\[ s_{i,d} = \begin{bmatrix} A & B \end{bmatrix} \begin{bmatrix} s_{i-1,0} \\ s_{i-1,d} \end{bmatrix}. \]

As discussed below, the transfer matrix \( \begin{bmatrix} A & B \end{bmatrix} \) can be converted into a block-diagonal form through suitable invertible linear transformations, thus resulting in a diagonally correlated deterministic source. The prospliant coding scheme discussed earlier can then be applied to such a transformed source.

**Case 1: Full-Rank A**

Our transformation is most natural for the case when \( A \) is a full row-rank matrix. So we treat this case first. Let

\[ N_1 \triangleq \text{Rank}(A) \leq \min\{N_0, N_d\}. \]  

(2.99)

In this section we restrict to the special case where \( N_1 = N_d \), i.e., \( A \) is a full-row-rank matrix with \( N_d \) independent non-zero rows. For this case, we explain the coding scheme by describing the encoder and decoder shown in Fig 2.7.

**Encoder:** As in Fig 2.7, the encoder applies a memoryless transformation block \( T(.) \) onto each symbol \( s_i \) to yield \( \tilde{s}_i = L(s_i) \).

Suppose that \( X \) is a matrix of dimensions \( N_0 \times N_d \). Define

\[ M \triangleq \begin{bmatrix} I & X \\ 0 & I \end{bmatrix} \]  

(2.100)
and observe that

\[
M^{-1} = \begin{bmatrix}
I & -X \\
0 & I
\end{bmatrix}.
\] (2.101)

For a certain \(X\) to be specified later, let

\[
s_{i,d} = \begin{bmatrix}
A \\
B
\end{bmatrix} M^{-1} M \begin{bmatrix}
s_{i-1,0} \\
s_{i-1,d}
\end{bmatrix} = \begin{bmatrix}
A \\
B - AX
\end{bmatrix} \begin{bmatrix}
s_{i-1,0} + Xs_{i-1,d} \\
s_{i-1,d}
\end{bmatrix}.
\] (2.102)

(2.103)

Since \(A\) is a full-rank matrix, we may select \(X\) such that

\[
B - AX = 0
\] (2.104)

With this choice of \(X\), (2.103) reduces to

\[
s_{i,d} = \begin{bmatrix}
A \\
0
\end{bmatrix} \begin{bmatrix}
s_{i-1,0} + Xs_{i-1,d} \\
s_{i-1,d}
\end{bmatrix}
\] (2.105)

Now, define the linear transformation \(T(.)\) as follows.

\[
\tilde{s}_i = \begin{pmatrix}
\tilde{s}_{i,0} \\
\tilde{s}_{i,1}
\end{pmatrix} = T(s_i) \triangleq \begin{pmatrix}
 s_{i,0} + Xs_{i,d} \\
 s_{i,d}
\end{pmatrix} = Ms_i
\] (2.106)

Note that 1) The transformation \(T(.)\) is memoryless and requires no knowledge of the past source sequences, 2) The innovation bits \(s_{i,0}\) are independently drawn and independent of \(s_{i,d}\). Hence \(\tilde{s}_{i,0}\) are drawn i.i.d. according to Bernoulli-(1/2), and are independent of \(s_{i,d}\), 3) The map between the two sources \(s_i\) and \(\tilde{s}_i\) is one-to-one.

Observe that \(\tilde{s}_i\) is diagonally correlated Markov source with \(N_0\) innovation bits \(\tilde{s}_{i,0}\) and \(N_d\) deterministic bits \(\tilde{s}_{i,1}\) that satisfy

\[
\tilde{s}_{i,1} = A \tilde{s}_{i-1,0}.
\] (2.107)
We transmit the source sequence \( \{ \tilde{s}_i \} \) using the Prospicient Coding scheme.

**Decoder:** At the receiver, first the Prospicient decoder recovers the diagonally correlated source \( \tilde{s}_i \) at any time except error propagation window. Then whenever \( \tilde{s}_i \) is available, the decoder directly constructs \( s_i \) as

\[
s_i = T^{-1}(\tilde{s}_i) = M^{-1}\tilde{s}_i. \tag{2.108}
\]

**Rate-optimality:** Suppose that our two step approach in Fig. 2.7 is sub-optimal. Then, in order to transmit the \( \tilde{s}_i \) through the channel, one can first transform it into \( s_i \) via \( T^{-1} \) and achieve lower rate than the prospicient coding scheme. However this is impossible because prospicient scheme is optimal. This shows the optimality of the coding scheme.

**Case 2: General A**

Now we consider the general case of semi-deterministic Markov sources defined in Def. 2. As illustrated in Fig. 2.8 the reduction to the diagonally correlated source is done in two steps using two linear transforms: \( L_f(\cdot) \) and \( L_b(\cdot) \).

**Lemma 4.** Any semi-deterministic Markov source specified in Def. 2, or equivalently by (2.15), can be transformed into an equivalent source \( \hat{s}_i \) consisting of innovation component \( s_{i,0} \in \{0,1\}^{N_0} \) and \( K \) deterministic components that satisfy the following.

\[
\hat{s}_{i,d} = \begin{pmatrix}
  s_{i,1} \\
  s_{i,2} \\
  \vdots \\
  s_{i,K-1} \\
  s_{i,K}
\end{pmatrix} =
\begin{pmatrix}
  R_{1,0} & R_{1,1} & \cdots & R_{1,K-2} & R_{1,K-1} & R_{1,K} \\
  0 & R_{2,1} & \cdots & R_{2,K-2} & R_{2,K-1} & R_{2,K} \\
  \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\
  0 & 0 & \cdots & R_{K-1,K-2} & R_{K-1,K-1} & R_{K-1,K} \\
  0 & 0 & \cdots & 0 & R_{K,K-1} & R_{K,K}
\end{pmatrix}
\begin{pmatrix}
  s_{i-1,0} \\
  s_{i-1,1} \\
  \vdots \\
  s_{i-1,K-2} \\
  s_{i-1,K-1} \\
  s_{i-1,K}
\end{pmatrix}. \tag{2.109}
\]
Lemma 5. Consider the source $\hat{s}_i = T_f(s_i)$, where $s_i$ is a semi-deterministic Markov source and $\hat{s}_i$ is defined in (2.109). There exists a one-to-one linear transformation $T_h$ which maps $\hat{s}_i$ to a diagonally correlated deterministic Markov source $\hat{s}_i$ that satisfies (2.110).

Using a one-to-one linear transformation $L_f$ where

1. $s_{i,j} \in \{0,1\}^{N_j}$ for $j \in \{0,\ldots,K\}$ where

$$N_0 \geq N_1 \geq \ldots \geq N_K,$$

and $\sum_{k=1}^{K} N_k = N_d$.

2. $R_{j,j-1}$ is $N_j \times N_{j-1}$ full-rank matrix of rank $N_j$ for $j \in \{1,\ldots,K-1\}$.

3. The matrix $R_{K,K-1}$ is either full-rank of rank $N_K$ or zero matrix.

The transformation to $\hat{s}_i$ involves repeated application the technique in case 1. The proof is provided in Appendix A.2. The proof provides an explicit construction of $L_f$.

To illustrate the idea, here we study a simple example. The complete proof is available in Appendix A.3. Assume $K = 2$ and consider the source $\hat{s}_i$ consisting of $N_0$ innovation bits $s_{i,0}$ and $N_1 + N_2$ deterministic bits as

$$\hat{s}_{i,d} = \begin{pmatrix} \hat{s}_{i,1} \\ \hat{s}_{i,2} \\ \vdots \\ \hat{s}_{i,K-1} \\ \hat{s}_{i,K} \end{pmatrix} = \begin{pmatrix} R_{1,0} & 0 & \cdots & 0 & 0 \\ 0 & R_{2,1} & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & R_{K-1,K-2} & 0 \\ 0 & 0 & \cdots & 0 & R_{K,K-1} \end{pmatrix} \begin{pmatrix} s_{i-1,0} \\ s_{i-1,1} \\ \vdots \\ s_{i-1,K-2} \\ s_{i-1,K-1} \end{pmatrix},$$

where $R_{1,0}$ and $R_{2,1}$ are full-rank (non-zero) matrices of rank $N_1$ and $N_2$, respectively.

The following steps transforms the source $\hat{s}_i$ into diagonally correlated Markov source.

**Step 1:** Define

$$\begin{pmatrix} \hat{s}_{i,1} \\ \hat{s}_{i,2} \end{pmatrix} \triangleq \begin{pmatrix} I_{N_1} & X_1 \\ 0 & I_{N_2} \end{pmatrix} \begin{pmatrix} s_{i,1} \\ s_{i,2} \end{pmatrix}$$

and

$$D_1 \triangleq \begin{pmatrix} I_{N_0} & 0 & 0 \\ 0 & I_{N_1} & X_1 \\ 0 & 0 & I_{N_2} \end{pmatrix}$$
and note that
\[
D_1^{-1} = \begin{pmatrix}
I & 0 & 0 \\
0 & I & -X_1 \\
0 & 0 & I
\end{pmatrix}.
\] (2.115)

By these definitions it is not hard to check that
\[
\left( \tilde{s}_{i,1}, \tilde{s}_{i,2} \right) = 
\begin{pmatrix}
I & X_1 \\
0 & I
\end{pmatrix}
\begin{pmatrix}
R_{1,0} & R_{1,1} & R_{1,2} \\
0 & R_{2,1} & R_{2,2}
\end{pmatrix}
D_1^{-1}
\begin{pmatrix}
s_{i-1,0} \\
\tilde{s}_{i-1,1} \\
\tilde{s}_{i-1,2}
\end{pmatrix} = 
\begin{pmatrix}
R_{1,0} & R_{1,1} & \tilde{R}_{1,2} \\
0 & R_{2,1} & R_{2,2} - R_{2,1} X_1
\end{pmatrix}
\begin{pmatrix}
s_{i-1,0} \\
\tilde{s}_{i-1,1} \\
\tilde{s}_{i-1,2}
\end{pmatrix}
\] (2.116)

where
\[
\tilde{R}_{1,1} = R_{1,1} + X_1 R_{2,1} 
\] (2.117)

and
\[
\tilde{R}_{1,2} = R_{1,2} + X_1 R_{2,2} - X_1 R_{2,1} X_1 - R_{1,1} X_1.
\] (2.118)

Note that \( R_{2,1} \) is full-row-rank of rank \( N_2 \) and \( R_{2,2} \) is \( N_2 \times N_2 \) matrix, thus \( X_1 \) can be selected such that
\[
R_{2,2} - R_{2,1} X_1 = 0
\] (2.119)

and (2.116) reduces to
\[
\left( \tilde{s}_{i,1}, \tilde{s}_{i,2} \right) = 
\begin{pmatrix}
R_{1,0} & \tilde{R}_{1,1} & \tilde{R}_{1,2} \\
0 & R_{2,1} & 0
\end{pmatrix}
\begin{pmatrix}
s_{i-1,0} \\
\tilde{s}_{i-1,1} \\
\tilde{s}_{i-1,2}
\end{pmatrix}.
\] (2.120)

**Step 2:** Define
\[
\tilde{s}_{i-1,0} \triangleq \begin{pmatrix}
I & X_{1,2} \\
0 & X_{2,2}
\end{pmatrix}
\begin{pmatrix}
s_{i-1,0} \\
\tilde{s}_{i-1,1} \\
\tilde{s}_{i-1,2}
\end{pmatrix} 
\] (2.121)

and
\[
D_2 \triangleq 
\begin{pmatrix}
I & X_{1,2} & X_{2,2} \\
0 & I & 0 \\
0 & 0 & I
\end{pmatrix}
\] (2.122)
and note that

\[ D_2^{-1} = \begin{pmatrix} I & -X_{1,2} & -X_{2,2} \\ 0 & I & 0 \\ 0 & 0 & I \end{pmatrix}. \]  

(2.123)

It can be observed that

\[ \begin{pmatrix} \tilde{s}_{i,1} \\ \tilde{s}_{i,2} \end{pmatrix} = \begin{pmatrix} R_{1,0} & \tilde{R}_{1,1} & \tilde{R}_{1,2} \\ 0 & R_{2,1} & 0 \end{pmatrix} D_2^{-1} \begin{pmatrix} \tilde{s}_{i-1,0} \\ \tilde{s}_{i-1,1} \\ \tilde{s}_{i-1,2} \end{pmatrix} \]  

(2.124)

\[ = \begin{pmatrix} R_{1,0} & \tilde{R}_{1,1} - R_{1,0}X_{1,2} & \tilde{R}_{1,2} - R_{1,0}X_{2,2} \\ 0 & R_{2,1} & 0 \end{pmatrix} \begin{pmatrix} \tilde{s}_{i-1,0} \\ \tilde{s}_{i-1,1} \\ \tilde{s}_{i-1,2} \end{pmatrix}. \]  

(2.125)

Similarly, \( X_{1,2} \) and \( X_{2,2} \) are selected such that

\[ \tilde{R}_{1,1} - R_{1,0}X_{1,2} = 0 \]  

(2.126)

\[ \tilde{R}_{1,2} - R_{1,0}X_{2,2} = 0. \]  

(2.127)

Therefore, the source \( \tilde{s}_i \) consists of \( N_0 \) innovation bits and \( N_1 + N_2 \) deterministic bits as

\[ \begin{pmatrix} \tilde{s}_{i,1} \\ \tilde{s}_{i,2} \end{pmatrix} = \begin{pmatrix} R_{1,0} & 0 & 0 \\ 0 & R_{2,1} & 0 \end{pmatrix} \begin{pmatrix} \tilde{s}_{i-1,0} \\ \tilde{s}_{i-1,1} \\ \tilde{s}_{i-1,2} \end{pmatrix} \]  

(2.128)

\[ = \begin{pmatrix} R_{1,0} & 0 \\ 0 & R_{2,1} \end{pmatrix} \begin{pmatrix} \tilde{s}_{i-1,0} \\ \tilde{s}_{i-1,1} \end{pmatrix}. \]  

(2.129)

Clearly, \( \tilde{s}_i = T_b(\hat{s}_i) \) is a diagonally correlated deterministic Markov source and the mapping is invertible.

Exploiting Lemmas 4 and 5, any linear semi-deterministic source \( s_i \) is first transformed into a diagonally correlated deterministic Markov source \( \tilde{s}_i = T_b(T_f(s_i)) \) and then is transmitted through the channel using prospecient coding scheme. The block diagram of encoder and decoder is shown in Fig 2.8. The optimality of the scheme can be shown using a similar argument in Sec. 2.6.5.

### 2.7 Conclusion

In this chapter we introduced the zero-delay streaming problem setup and the notion of rate-recovery function for discrete sources and lossless recovery at the decoder. The main objective is to capture the fundamental trade-off between the compression efficiency and the error propagation in any multimedia streaming application from an information-theoretic viewpoint. General upper and lower bounds on the lossless rate-recovery function were provided. The upper and lower bounds do not coincide in general. We studied a class of memoryless encoders and symmetric sources where the general lower bound was improved and established the optimality of the upper bounds. In addition we provided the example of semi-deterministic Markov sources where the general upper bound was improved to establish the optimality of the lower bound. This completes the study of discrete sources and lossless recovery
constraint for zero-delay streaming. In Chapter 5 we extend the results to the case of delay-constrained rather than zero-delay streaming.
Chapter 3

Zero-Delay Streaming of Gauss-Markov Sources: Immediate Recovery

3.1 Introduction

While the extension of the lossless streaming setup to the lossy case is rather natural, it turns out that the analysis of the achievable rate is highly nontrivial. In this thesis, we focus on the important special case of Gauss-Markov sources with quadratic distortion measure. In particular we consider the scenario where the source vectors, drawn from a first-order Gauss-Markov process, are sequentially compressed and sent through the channel. The channel may introduce an erasure burst of length up to $B$ in an unknown location during the communication period, but perfectly reveals the rest of the packets to the destination. This chapter investigates the lossy rate-recovery function for Gauss-Markov sources for the special case of $W = 0$, i.e., the decoder is required to reconstruct the source sequences within an average quadratic distortion $D$ at any time except when the channel introduces erasures. Upper and lower bounds are established, which asymptotically coincide at high resolutions. The upper bound is based on the quantization-and-binning (Q-and-B) technique and involves a non-trivial step of locating the worst-case erasure burst pattern. The lower bound is a generalization of the lower bound for the lossless case using standard Gaussian source coding techniques. Motivated by the prosipient coding and semi-deterministic Markov sources introduced in Chapter 2, we study a special class of i.i.d. Gaussian sources with a sliding window recovery constraint where a coding scheme similar to prosipient coding attains the optimal lossy rate-recovery function.

The rest of the chapter is organized as follows. The problem setup is introduced in Section 3.2. The main results of the chapter are summarized in Section 3.3. The lower bound on the rate-recovery function

If I had an hour to solve a problem I'd spend 55 minutes thinking about the problem and 5 minutes thinking about solutions.

Albert Einstein
is studied in Section 3.4. The upper bounds for channel models having single and multiple erasure bursts are treated in Section 3.5 and Section 3.6, respectively. The achievable rate in high resolution regime is treated in Section 3.7. The example of independent Gaussian sources with sliding window recovery constraint is treated in Section 3.8. Section 3.9 concludes the chapter.

3.2 Problem Setup

3.2.1 Source Model and Encoder

We assume that the communication spans the interval \( t \in \{0, 1, \ldots, \Upsilon\} \). At each time \( t \), a Gaussian source vector \( \{s^n_t\} \) is sampled i.i.d. from a zero-mean Gaussian distribution \( \mathcal{N}(0, \sigma_z^2) \) along the spatial dimension, and forms a first-order Gauss-Markov chain across the temporal dimension, i.e.,

\[
s_{t} = \rho s_{t-1} + n_t,
\]

where \( \rho \in (0, 1) \) and \( n_t \sim \mathcal{N}(0, (1-\rho^2)\sigma_z^2) \). Without loss of generality we assume \( \sigma_z^2 = 1 \). The sequence \( s^n_0 \) is sampled i.i.d. from \( \mathcal{N}(0, \sigma_z^2) \) and revealed to both the encoder and decoder before the start of the communication. It plays the role of a synchronization frame.

An encoder computes an index \( f_t \in \{1, 2, \ldots, 2^n R_t\} \) at time \( t \), according to an encoding function

\[
f_t = F_i(s^n_0, \ldots, s^n_t), \quad 1 \leq t \leq \Upsilon.
\]

(3.2)

Note that the encoder in (3.2) is a causal function of the source sequences.

3.2.2 Channel Model and Decoder

The channel takes each \( f_t \) as input and either outputs \( g_t = f_t \) or an erasure symbol i.e., \( g_t = \star \). We consider the class of erasure burst channels. For some particular \( j \geq 1 \), the channel introduces an erasure burst such that

\[
g_t = \begin{cases} 
\star, & t \in \{j, j+1, \ldots, j+B-1\} \\
 f_t, & \text{otherwise}
\end{cases}
\]

(3.3)

Upon observing the sequence \( \{g_t\}_{t \geq 1} \), the decoder is required to reconstruct each source sequence with zero delay i.e.,

\[
\hat{s}_t^n = \tilde{G}_t(g_1^n, g_2^n, \ldots, g_t^n), \quad t \notin \{j, \ldots, j + B - 1\},
\]

(3.4)

where \( \hat{s}_t^n \) denotes the reconstruction sequence and \( j \) denotes the time at which erasure burst starts in (2.3). The destination is not required to produce the source vectors that appear during the erasure burst. We call this period the error propagation window. We consider the case where the reconstruction in (2.4) satisfies the following average distortion constraint.

\[
\limsup_{n \to \infty} E \left[ \frac{1}{n} \sum_{k=1}^{n} d(s_{i,k}, \hat{s}_{i,k}) \right] \leq D
\]

(3.5)
for the quadratic distortion measure, \(d(s_i, \hat{s}_i) = (s_i - \hat{s}_i)^2\).

### 3.2.3 Rate-Recovery Function

The rate \(R\) is feasible if a sequence of encoding and decoding functions exists that satisfies the average distortion constraint. The minimum feasible rate \(R_\gamma(B, D)\), is the rate-recovery function.

### 3.3 Main Results

#### 3.3.1 Channels with Single erasure burst

In this channel model, as stated in (2.3), we assume that the channel can introduce a single erasure burst of length up to \(B\) during the transmission period. Define \(R_{\text{GM-SE}}(B, D) \triangleq R(B, W = 0, D)\) as the lossy rate-recovery function of Gauss-Markov sources with single erasure burst channel model.

**Proposition 2** (Lower Bound–Single Burst). The lossy rate-recovery function of the Gauss-Markov source for single erasure burst channel model when \(W = 0\) satisfies

\[
R_{\text{GM-SE}}(B, D) \geq R_{\text{GM-SE}}^-(B, D) \triangleq \frac{1}{2} \log \left( \frac{D\rho^2 + 1 - \rho^2(B+1) + \sqrt{\Delta}}{2D} \right)
\]

where \(\Delta \triangleq (D\rho^2 + 1 - \rho^2(B+1))^2 - 4D\rho^2(1 - \rho^2B)\). \(\blacksquare\)

The proof of Prop. 2 is presented in Section 3.4. The proof considers the recovery of a source sequence \(s^n\), given an erasure burst in the interval \(\{t - B, \ldots, t - 1\}\) and extends the lower bounding technique in Theorem 1 to incorporate the distortion constraint.

**Proposition 3** (Upper Bound–Single Burst). The lossy rate-recovery function of the Gauss-Markov source for single erasure burst channel model when \(W = 0\) satisfies

\[
R_{\text{GM-SE}}(B, D) \leq R_{\text{GM-SE}}^+(B, D) \triangleq \frac{1}{2} \log \left( \frac{D\rho^2 + 1 - \rho^2(B+1) + \sqrt{\Sigma}}{2D} \right)
\]

where \(\Sigma \triangleq \frac{1}{2} \sqrt{1 - \sigma_z^2/(1 - \Sigma^2)}\) and \(\sigma_z^2 > 0\) is chosen to satisfy

\[
\left[ \frac{1}{\sigma_z^2} + \frac{1}{1 - \rho^2(1 - \Sigma^2)} \right]^{-1} \leq D.
\]

This is equivalent to \(\sigma_z^2\) satisfying

\[
E[(s_i - \hat{s}_i)^2] \leq D,
\]
(a) $R_{\text{GM-SE}(B, D)}$ versus $\rho$ for $D = 0.2$, $D = 0.3$ and $B = 1, B = 2$.

(b) $R_{\text{GM-SE}(B, D)}$ versus $D$ for $\rho = 0.9$, $\rho = 0.7$ and $B = 1, B = 2$.

Figure 3.1: Lower and upper bounds of lossy rate-recovery function.
where \( \hat{s}_i \) denotes the minimum mean square estimate (MMSE) of \( s_i \) from \( \{ \hat{s}_{i-B}, u_t \} \).

The following alternative rate expression for the achievable rate in Prop. 3, provides a more explicit interpretation of the coding scheme.

\[
R_{GM-SE}^+(B, D) = \lim_{t \to \infty} I(s_t; u_t | [u_{t-1}^{t-B-1}])
\]  

(3.11)

where the random variables \( u_t \) are obtained using the same test channel in Prop. 3. Notice that the test channel noise \( \sigma_s^2 > 0 \) is chosen to satisfy \( E \left[ (s_t - \hat{s}_t)^2 \right] \leq D \) where \( \hat{s}_t \) denotes the MMSE of \( s_t \) from \( \{ [u_{t-1}^{t-B-1}, u_t] \} \) in steady state, i.e., \( t \to \infty \). Notice that (3.11) is based on a Q-and-B scheme when the receiver has side information sequences \( \{ u_0^n, \ldots, u_{t-B-1}^n \} \). The proof of Prop. 3 which is presented in Section 3.5 also involves establishing that the worst case erasure pattern during the recovery of \( \hat{s}_t^n \) spans the interval \( \{ t-B, \ldots, t-1 \} \). The proof is considerably more involved as the reconstruction sequences \( \{ u_t^n \} \) do not form a Markov chain.

As we will show subsequently, the upper and lower bounds in Prop. 2 and Prop. 3 coincide in the high resolution limit. Numerical evaluations suggest that the bounds are close for a wide range of parameters. Fig. 3.1a and Fig. 3.1b illustrate some sample comparison plots.

### 3.3.2 Channels with Multiple erasure bursts

We also consider the case where the channel can introduce multiple erasure bursts, each of length no greater than \( B \) and with a guard interval of length at-least \( L \) separating consecutive bursts. The encoder is defined as in (2.2). We again only consider the case when \( W = 0 \). Upon observing the sequence \( \{ g_i \}_{i \geq 0} \), the decoder is required to reconstruct each source sequence with zero delay, i.e.,

\[
\hat{s}_t^n = G_t(g_0, g_1, \ldots, g_{t-1}), \quad \text{whenever } g_i \neq \star
\]  

(3.12)

such that the reconstructed source sequence \( \hat{s}_t^n \) satisfies an average mean square distortion of \( D \). The destination is not required to produce the source vectors that appear during any of the erasure bursts. The rate \( R(L, B, D) \) is feasible if a sequence of encoding and decoding functions exists that satisfies the average distortion constraint. The minimum feasible rate \( R_{GM-ME}(L, B, D) \), is the lossy rate-recovery function.

**Proposition 4 (Upper Bound–Multiple Bursts).** The lossy rate-recovery function \( R_{GM-ME}(L, B, D) \) for Gauss-Markov sources over the multiple erasure bursts channel satisfies the following upper bound:

\[
R_{GM-ME}(L, B, D) \leq R_{GM-ME}^+(L, B, D) \triangleq I(u_t; \hat{s}_{t-L-B}, [u_{t-L-B+1}^{t-B-1}])
\]  

(3.13)

where \( \hat{s}_{t-L-B} = s_{t-L-B} + \epsilon \), where \( \epsilon \sim \mathcal{N}(0, D/(1-D)) \). Also for any \( i, u_i \triangleq s_i + z_i \) and \( z_i \) is sampled i.i.d. from \( \mathcal{N}(0, \sigma_z^2) \) and the noise in the test channel, \( \sigma_z^2 > 0 \) satisfies

\[
E \left[ (s_t - \hat{s}_t)^2 \right] \leq D
\]  

(3.14)

and \( \hat{s}_t \) denotes the MMSE estimate of \( s_t \) from \( \{ \hat{s}_{t-L-B}, [u_{t-L-B+1}^{t-B-1}], u_t \} \).

The proof of Prop. 4 presented in Section 3.6 is again based on Q-and-B technique and involves characterizing the worst-case erasure pattern by the channel. Note also that the rate expression in
Figure 3.2: Achievable rates for multiple erasure bursts model for different values of guard length $L$ separating erasure bursts comparing to single erasure burst. As $L$ grows, the rate approaches the single erasure case. The lower bound for single erasure case is also plotted for comparison ($B = 1$).
(3.13) depends on the minimum guard spacing $L$, the maximum erasure burst length $B$ and distortion $D$, but is not a function of time index $t$, as the test channel is time invariant and the source process is stationary. An expression for computing $\sigma^2_z$ is provided in Section 3.6. While we do not provide a lower bound for $R_{GM,ME}(L,B,D)$ we remark that the lower bound in Prop. 2 also applies to the multiple erasure bursts setup.

Fig. 3.2 provides numerical evaluation of the achievable rate for different values of $L$. We note that even for $L$ as small as 4, the achievable rate in Prop. 4 is virtually identical to the rate for single erasure burst in Prop. 3. This strikingly fast convergence to the single erasure burst rate appears due to the exponential decay in the correlation coefficient between source samples as the time-lag increases.

### 3.3.3 High Resolution Regime

For both the single and multiple erasure bursts models, the upper and lower bounds on lossy rate-recovery function for $W = 0$ denoted by $R(L,B,D)$ coincide in the high resolution limit as stated below.

**Corollary 3.** In the high resolution limit, the Gauss-Markov lossy rate-recovery function satisfies the following:

$$R(L,B,D) = \frac{1}{2} \log \left( \frac{1 - \rho^{2(B+1)}}{D} \right) + o(D).$$  \hspace{1cm} (3.15)

where $\lim_{D \to 0} o(D) = 0$.

The proof of Corollary 3 is presented in Section 3.7. It is based on evaluating the asymptotic behaviour of the lower bound in (3.6) and the upper bound in Prop. 4, in the high resolution regime. Notice that the rate expression in (3.15) does not depend on the guard separation $L$. The intuition behind this is as follows. In the high resolution regime, the output of the test channel, i.e., $u_k$, becomes very close to
the original source \( s_t \). Therefore the Markov property of the original source is approximately satisfied by these auxiliary random variables and hence the past sequences are not required. The rate in (3.15) can also be approached by a *Naive Wyner-Ziv* coding scheme that only makes use of the most recently available sequence at the decoder. The rate of this scheme is given by [36]:

\[
R_{NWZ}(B, D) \triangleq I(s_t; u_i | u_{-B-1})
\]  

(3.16)

where for each \( i \), \( u_i = s_i + z_i \) and \( z_i \sim \mathcal{N}(0, \sigma_z^2) \) and \( \sigma_z^2 \) satisfies the following distortion constraint

\[
E[(s_i - \hat{s}_i)^2] \leq D
\]  

(3.17)

where \( \hat{s}_i \) is the MMSE estimate of \( s_i \) from \( \{u_{-B-1}, u_i\} \).

Fig. 3.3 reveals that while the rate in (3.16) is near optimal in the high resolution limit, it is in general sub-optimal when compared to the rates in (3.13) when \( \rho = 0.9 \). As we decrease \( \rho \), the performance loss associated with this scheme appears to reduce.

### 3.3.4 Gaussian Sources with Sliding Window Recovery Constraints

In this section we consider a specialized source model and distortion constraint, where it is possible to improve upon the binning-based upper bound. Our proposed scheme attains the rate-recovery function for this special case and is thus optimal. This example illustrates that the binning-based scheme can be sub-optimal in general.

**Source Model:** We consider a sequence of i.i.d. Gaussian source sequences i.e., at time \( i \), \( s^n_i \) is sampled i.i.d. according to a zero mean unit variance Gaussian distribution \( \mathcal{N}(0,1) \), independent of the past sources. At each time we associate an auxiliary source

\[
t^n_i = \left( s^n_i, s^n_{i-1}, \ldots, s^n_{i-K} \right)
\]  

(3.18)

which is a collection of the past \( K + 1 \) source sequences. Note that \( t^n_i \) constitutes a first-order Markov chain. We will define a reconstruction constraint with the sequence \( t^n_i \).

**Encoder:** The (causal) encoder at time \( i \) generates an output given by \( f_i = F_i(s^n_{i+1}, \ldots, s^n_i) \) where \( f_i \in \{1, 2, \ldots, 2^{nR}\} \).

**Channel Model:** The channel can introduce a burst erasure of length up to \( B \) in an arbitrary interval \( \{j, \ldots, j + B - 1\} \).

**Decoder:** At time \( i \) the decoder is interested in reproducing a collection of past \( K + 1 \) sources within a distortion vector \( d = (d_0, d_1, \ldots, d_K) \) i.e., at time \( i \) the decoder is interested in reconstructing \( (\hat{s}^n_i, \ldots, \hat{s}^n_{i-K}) \) where \( E[||s^n_{i-l} - \hat{s}^n_{i-l}||^2] \leq nd_{l} \) must be satisfied for \( l \in \{0, \ldots, K\} \). We assume throughout that \( d_0 \leq d_1 \leq \ldots \leq d_K \) which corresponds to the requirement that the more recent source sequences must be reconstructed with a smaller average distortion.

In Fig. 3.4, the source symbols \( s_t \) are shown as white circles. The symbols \( t_i \) and \( \hat{t}_i \) are also illustrated for \( K = 2 \). The different shading for the sub-symbols in \( \hat{t}_i \) corresponds to different distortion constraints.

If a erasure burst spans the interval \( \{j, \ldots, j + B - 1\} \), the decoder is not required to output a reproduction of the sequences \( t^n_i \) for \( i \in \{j, \ldots, j + B + W - 1\} \).

\(^\dagger\)In this section it is sufficient to assume that any source sequence with a time index \( j \leq -1 \) is a constant sequence.
Chapter 3. Zero-Delay Streaming of Gauss-Markov Sources: Immediate Recovery

\[ \hat{s}_i - 2 \]
\[ \hat{s}_i - 1 \]
\[ \hat{s}_i \]
\[ \hat{t}_i \]
\[ \hat{t}_i - 1 \]
\[ \hat{t}_i - 2 \]
\[ \{\hat{s}_i\}_{d_0} \]
\[ \{\hat{s}_i - 1\}_{d_1} \]
\[ \{\hat{s}_i - 2\}_{d_2} \]

Figure 3.4: Schematic of the Gaussian sources with sliding window recovery constraints for \( K = 2 \). The source \( s_i \), drawn as white circles, are independent sources and \( t_i \) is defined as a collection of \( K + 1 = 3 \) most recent sources. The source symbols along the diagonal lines are the same. The decoder at time \( i \) recovers \( s_i, s_i - 1 \) and \( s_i - 2 \) within distortions \( d_0, d_1 \) and \( d_2 \), respectively where \( d_0 \leq d_1 \leq d_2 \). In figure the colour density of the circle represents the amount of reconstruction distortion.

The lossy rate-recovery function denoted by \( R(B, W, d) \) is the minimum rate required to satisfy these constraints.

**Remark 3.** One motivation for considering the above setup is that the decoder might be interested in computing a function of the last \( K + 1 \) source sequences at each time e.g., \( v_i = \sum_{j=0}^{K} \alpha^j s_i - j \). A robust coding scheme, when the coefficient \( \alpha \) is not known to the encoder is to communicate \( s_i^n, s_i - 1, s_i - 2 \) with distortion \( d_j \) at time \( i \) to the decoder.

**Theorem 3.** For the proposed Gaussian source model with a non-decreasing distortion vector \( d = (d_0, \ldots, d_K) \) with \( 0 < d_i \leq 1 \), the lossy rate-recovery function is given by

\[
R(B, W, d) = \frac{1}{2} \log \left( \frac{1}{d_0} \right) + \frac{1}{W + 1} \min\{K - W, B\} \sum_{k=1}^{\min\{K - W, B\}} \frac{1}{2} \log \left( \frac{1}{d_{W+k}} \right). \tag{3.19}
\]

The proof of Theorem 3 is provided in Section 3.8. The coding scheme for the proposed model involves using a successive refinement codebook for each sequence \( s_i^n \) to produce \( B + 1 \) layers and carefully assigning the sequence of layered codewords to each channel packet. A simple quantize and binning scheme in general does not achieve the rate-recovery function in Theorem 3. A numerical comparison of the lossy rate-recovery function with other schemes is presented in Section 3.8.

This completes the statement of the main results in this chapter.

### 3.4 Lower Bound on Rate-Recovery Function

Consider any rate \( R \) code that satisfies an average distortion of \( D \) as stated in (3.5). For each \( i \geq 0 \) we have

\[
nR \geq H(f_i) \\
\geq H(f_i | [f]_{0}^{i-B-1}, s_{i-1}^n) \tag{3.20}
\]
Thus we have

\[ I(s^n_i; f_i | [f]_0^{i-B-1}, s_{i-1}^n) + H(f_i | s^n_i, [f]_0^{i-B-1}, s_{i-1}^n) \]

for each \( i \geq B \). Since \( I(s^n_i; f_i | [f]_0^{i-B-1}, s_{i-1}^n) \) is no greater than \( D \), and applies standard arguments [37, Ch. 13]. We first establish an upper bound for the second term in (3.21). Suppose that the erasure burst occurs in the interval \( \{i-B, \ldots, i-1\} \). The reconstruction sequence \( \hat{s}_i^n \) must be a function of \( (f_i, [f]_0^{i-B-1}, s_{i-1}^n) \). Thus we have

\[
\begin{align*}
\text{h}(s^n_i | [f]_0^{i-B-1}, f_i, s_{i-1}^n) &= \text{h}(s^n_i - \hat{s}_i^n | [f]_0^{i-B-1}, f_i, s_{i-1}^n) \\
&\leq \text{h}(s^n_i - \hat{s}_i^n) \\
&\leq \frac{n}{2} \log(2\pi eD),
\end{align*}
\]

where the last step uses the fact that the expected average distortion between \( s^n_i \) and \( \hat{s}_i^n \) is no greater than \( D \), and applies standard arguments [37, Ch. 13].

To lower bound the first term in (3.21), we successively use the Gauss-Markov relation (3.1) to express:

\[
s_i = \rho^{(B+1)} s_{i-B-1} + \hat{n}
\]

for each \( i \geq B \) and \( \hat{n} \sim N(0, 1 - \rho^{2(B+1)}) \) is independent of \( s_{i-B-1} \). Using the Entropy Power Inequality [37] we have

\[
2^{\frac{n}{2}} h(s^n_i | [f]_0^{i-B-1}, s_{i+1}^n) \geq 2^{\frac{n}{2}} h(\rho^{(B+1)} s_{i-B-1} | [f]_0^{i-B-1}, s_{i+1}^n) + 2^{\frac{n}{2}} h(\hat{n})
\]

This further reduces to

\[
\text{h}(s^n_i | [f]_0^{i-B-1}, s_{i+1}^n) \geq \frac{n}{2} \log \left( \rho^{2(B+1)} 2^{\frac{n}{2}} h(s_{i-B-1} | [f]_0^{i-B-1}, s_{i+1}^n) + 2\pi e(1 - \rho^{2(B+1)}) \right).
\]

It remains to lower bound the entropy term in the right hand side of (3.25). We show the following in Appendix B.1.

**Lemma 6.** For any \( k \geq 0 \)

\[
2^{\frac{n}{2}} h(s^n_i | [f]_0^{i-B-1}, s_{i+1}^n) \geq 2^{\frac{n}{2}} e(1 - \rho^2) \left( 1 - \left( \frac{\rho^2}{2\pi R} \right)^k \right)
\]

Upon substituting, (3.26), (3.25), and (3.22) into (3.21) we obtain that for each \( i \geq B + 1 \)

\[
R \geq \frac{1}{2} \log \left( \frac{\rho^{2(B+1)} (1 - \rho^2)}{D(2^{2R} - \rho^2)} \left( 1 - \left( \frac{\rho^2}{2\pi R} \right)^{i-B-1} \right) + \frac{1 - \rho^{2(B+1)}}{D} \right).
\]

Selecting \( i = T \), yields the tightest lower bound. As mentioned earlier, we are interested in infinite
horizon when $T \to \infty$, which yields the tightest lower bound, we have

$$R \geq \frac{1}{2} \log \left( \frac{\rho^{2(B+1)}(1-\rho^2)}{D(2^R-\rho^2)} + \frac{1-\rho^{2(B+1)}}{D} \right)$$

(3.28)

Rearranging (3.28) we have that

$$D2^{4R} - (D\rho^2 + 1 - \rho^{2(B+1)})2^R + \rho^2(1 - \rho^{2B}) \geq 0$$

(3.29)

Since the left hand side is quadratic in $2^{2R}$, (3.29) results in a lower bound and an upper bound on $2^{2R}$. Exploiting the condition $R > 0$ results in the lower bound in (3.6) in Prop. 2. This completes the proof.

\textbf{Remark 4.} Upon examining the proof of the lower bound of Prop. 2, we note that it applies to any source process that satisfies (3.1) and where the additive noise is i.i.d. $\mathcal{N}(0,1-\rho^2)$. We do not use the fact that the source process is itself a Gaussian process.

### 3.5 Coding Scheme: Single erasure burst

The achievable rate is based on quantization and binning. For each $i \geq 0$, we consider the test channel

$$u_i = s_i + z_i,$$

(3.30)

where $z_i \sim \mathcal{N}(0, \sigma_z^2)$ is independent Gaussian noise. At time $i$ we sample a total of $2^{n(I(u_i; s_i) + \epsilon)}$ codeword sequences i.i.d. from $\mathcal{N}(0, 1 + \sigma_z^2)$. The codebook at each time is partitioned into $2^{nR}$ bins. The encoder finds the codeword sequence $u^n_i$ jointly typical with the source sequence $s^n_i$ and transmits the bin index $f_i$ assigned to $u^n_i$.

The decoder, upon receiving $f_i$ attempts to decode $u^n_i$ at time $i$, using all the previously recovered codewords $\{u^n_j : 0 \leq j \leq i-1, g_j \neq \star\}$ and the source sequence $s^n_{i-1}$ as side information. The reconstruction sequence $\hat{s}^n_i$ is the minimum mean square error (MMSE) estimate of $s^n_i$ given $u^n_i$ and the past sequences. The coding scheme presented here is based on binning, similar to lossless case discussed in Section 2.4.3. The main difference in the analysis is that, unlike the lossless case, neither the recovered sequences $u^n_i$ nor reconstructed source sequences $\hat{s}^n_i$ inherit the Markov property of the original source sequences $s^n_i$. Therefore, unlike the lossless case, the decoder does not reset following a erasure burst, once the error propagation is completed. Since the effect of a erasure burst persists throughout, the analysis of achievable rate is significantly more involved.

Fig. 3.5 summarizes the main steps in proving Prop. 3. In particular, in Lemma 7, we first derive necessary parametric rate constraints associated with every possible erasure pattern. Second, through the Lemma 8, we characterize the worst-case erasure pattern that dominates the rate and distortion constraints. Finally in Lemma 9 and Section 3.5.2, we evaluate the achievable rate to complete the proof of Prop. 3.
Lemma 7: Connection to Gaussian Many-help-one Source Coding Problem.

Lemma 8: Worst-case Characterization of Burst Erasure and Steady State Analysis.

Lemma 9 and Section 3.5.2: Rate Evaluation

Figure 3.5: Flowchart summarizing the proof steps of Prop. 3.

Figure 3.6: Schematic of single erasure burst channel model. The channel inputs in the interval \( \{t - B' - k, \ldots, t - k - 1\} \) is erased for some \( 0 \leq B' \leq B \) and \( k \in \{0, 2, \ldots, t - B'\} \). The rest are available at the decoder, as shown by check mark in the figure.

3.5.1 Analysis of Achievable Rate

Given a collection of random variables \( \mathcal{V} \), we let the MMSE estimate of \( s_i \) be denoted by \( \hat{s}_i(\mathcal{V}) \), and its associated estimation error is denoted by \( \sigma^2_t(\mathcal{V}) \), i.e.,

\[
\hat{s}_i(\mathcal{V}) = E[s_i | \mathcal{V}] \tag{3.31}
\]

\[
\sigma^2_t(\mathcal{V}) = E[(s_i - \hat{s}_i(\mathcal{V}))^2]. \tag{3.32}
\]

We begin with a parametric characterization of the achievable rate.

**Lemma 7.** A rate-distortion pair \((R,D)\) is achievable if, for every \( t \geq 0, B' \in \{0, \ldots, B\} \) and \( k \in \{0, \ldots, t - B'\} \), we have

\[
R \geq \lambda_t(k, B') \triangleq I(s_i; u_i | [u]_{t-B'-k-1}^{t-1}, [u]_{t-k-1}^{t-1}, s_{-1}), \tag{3.33}
\]

and the test-channel (3.30) satisfies

\[
\gamma_t(k, B') \triangleq E \left[ \left( \hat{s}_i - \hat{s}_i([u]_{t-B'-k-1}^{t-1}, [u]_{t-k-1}^{t-1}) \right)^2 \right] = \sigma^2_t([u]_{t-B'-k-1}^{t-1}, [u]_{t-k-1}^{t-1}, s_{-1}) \leq D, \tag{3.34}
\]

where \( \sigma^2_t(\cdot) \) and \( \hat{s}_i(\cdot) \) are defined in (3.32) and (3.31) respectively. \qed
Proof. Consider the decoder at any time $t \geq 0$ outside the error propagation window. Assume that a single erasure burst of length $B' \in \{0, 1, \ldots, B\}$ spans the interval $\{t - B' - k, \ldots, t - k - 1\}$ for some $k \in \{0, \ldots, t - B'\}$, i.e.,

$$g_j = \begin{cases} * \quad & j \in \{t - B' - k, \ldots, t - k - 1\} \\ f_j \quad & \text{else.} \end{cases} \quad (3.35)$$

The schematic of the erasure channel is illustrated in Fig. 3.6. Notice that $k = 0$ represents the case of the most recent erasure burst spanning the interval $\{t - B' - 1, \ldots, t - 1\}$. The decoder is interested in first successfully recovering $u^n_t$ and then reconstructing $z^n_i$ within distortion $D$ by performing MMSE estimation of $z^n_i$ from all the previously recovered sequences $u^n_i$ where $i \leq t$ and $g_i \neq *$. The decoder succeeds with high probability if the rate constraint satisfies (3.33) (see e.g., [38]) and the distortion constraint satisfies (3.34). If these constraints hold for all the possible triplets $(t, B', k)$, the decoder is guaranteed to succeed in reproducing any source sequence within desired distortion $D$.

Finally in the streaming setup, we can follow the argument similar to that in Section 2.4.3 to argue that the decoder succeeds in the entire horizon of $L$ provided we select the source length $n$ to be sufficiently large. The formal proof is omitted here. \qed

As a result of Lemma 7, in order to compute the achievable rate, we need to characterize the worst case values of $(t, k, B')$ that simultaneously maximize $\lambda_t(k, B)$ and $\gamma_t(k, B)$. We present such a characterization next.

Lemma 8. The functions $\lambda_t(k, B)$ and $\gamma_t(k, B)$ satisfy the following properties:

1. For all $t \geq B'$ and $k \in \{0, \ldots, t - B'\}$, $\lambda_t(k, B') \leq \lambda_t(0, B')$ and $\gamma_t(k, B') \leq \gamma_t(0, B')$, i.e., the worst-case erasure pattern contains the erasure burst in the interval $\{t - B, \ldots, t - 1\}$.

2. For all $t \geq B$ and $0 \leq B' \leq B$, $\lambda_t(0, B') \leq \lambda_t(0, B)$ and $\gamma_t(0, B') \leq \gamma_t(0, B)$, i.e., the worst-case erasure pattern includes maximum burst length.

3. For a fixed $B$, the functions $\lambda_t(0, B)$ and $\gamma_t(0, B)$ are both increasing with respect to $t$, for $t \geq B$, i.e., the worst-case erasure pattern happens in steady state (i.e., $t \to \infty$) of the system.

4. For all $t < B$, $0 \leq B' \leq t$ and $k \in \{0, \ldots, t - B'\}$, $\lambda_t(k, B') \leq \lambda_B(0, B)$ and $\gamma_t(k, B') \leq \gamma_B(0, B)$, i.e., the erasure burst spanning $\{0, \ldots, B - 1\}$ dominates all erasure bursts that terminate before time $B - 1$.

\qed

Proof. Before establishing the proof, we state two inequalities which are established in Appendix B.2. For each $k \in \{1, \ldots, t - B'\}$ we have that:

$$h(u_i || u_{i-k}^{t-B'-k-1}, [u]_{t-k}^{t-1}, s_{-1}) \leq h(u_i || [u]_{i-k}^{t-B'-k-1}, [u]_{i-k+1}^{t-1}, s_{-1}), \quad (3.36)$$

$$h(s_i || u_{i-k}^{t-B'-k-1}, [u]_{t-k}^{t}, s_{-1}) \leq h(s_i || [u]_{i-k}^{t-B'-k-1}, [u]_{t-k+1}^{t}, s_{-1}). \quad (3.37)$$

The above inequalities state that the conditional differential entropy of $u_i$ and $s_i$ is reduced if the variable $u_{i-B'-k}$ is replaced by $u_{i-k}$ in the conditioning and the remaining variables remain unchanged. Fig. 3.7
provides a schematic interpretation of the above inequalities. The proof in Appendix B.2 exploits the specific structure of the Gaussian test channel (3.30) and Gaussian sources to establish these inequalities.

In the remainder of the proof, we establish each of the four properties separately.

1) We show that both $\lambda_t(k, B')$ and $\gamma_t(k, B')$ are decreasing functions of $k$ for $k \in \{1, \ldots, t - B'\}$.

$$
\lambda_t(k, B') = I(s_t; u_t| [u]_0^{t-B'-k-1}, [u]_{k-1}^{t-1}, s_{-1}) \\
= h(u_t||[u]_0^{t-B'-k-1}, [u]_{k-1}^{t-1}, s_{-1}) - h(u_t|s_t) \\
\leq h(u_t||[u]_0^{t-B'-k}, [u]_{k}^{t-1}, s_{-1}) - h(u_t|s_t) \\
= I(s_t; u_t| [u]_0^{t-B'-k}, [u]_{k}^{t-1}, s_{-1}) \\
= \lambda_t(k-1, B'),
$$

(3.38)

where (3.38) follows from using (3.36). In a similar fashion since

$$
\gamma_t(k, B') = \sigma_t^2 \left( [u]_0^{t-B'-k}, [u]_{k}^{t-1}, s_{-1} \right)
$$

is the MMSE estimation error of $s_t$ given $([u]_0^{t-B'-k}, [u]_{k}^{t-1}, s_{-1})$, we have

$$
\frac{1}{2} \log (2\pi e \cdot \gamma_t(k, B')) = h(s_t||[u]_0^{t-B'-k-1}, [u]_{k-1}^{t-1}, s_{-1}) \\
\leq h(s_t||[u]_0^{t-B'-k}, [u]_{k}^{t-1}, s_{-1}) \\
= \frac{1}{2} \log (2\pi e \cdot \gamma_t(k-1, B'))
$$

(3.40)

(3.41)

where (3.40) follows from using (3.37). Since $f(x) = \frac{1}{2} \log(2\pi e x)$ is a monotonically increasing function it follows that $\gamma_t(k, B') \leq \gamma_t(k-1, B')$. By recursively applying (3.39) and (3.41) until $k = 1$, the proof of property (1) is complete.

2) We next show that the worst case erasure pattern also has the longest burst. This follows intuitively since the decoder can just ignore some of the symbols received over the channel. Thus any rate achieved with the longest burst is also achieved for the shorter burst. The formal justification is as follows. For any $B' \leq B$ we have

$$
\lambda_t(0, B') = I(s_t; u_t| [u]_0^{t-B'-1}, s_{-1}) \\
= h(u_t||[u]_0^{t-B'-1}, s_{-1}) - h(u_t|s_t) \\
\leq h(u_t||[u]_0^{t-B-1}, [u]_{B-B'}^{t-1}, s_{-1}) - h(u_t|s_t) \\
= \frac{1}{2} \log (2\pi e \cdot \gamma_t(k-1, B'))
$$

(3.42)

(3.43)
\begin{align*}
\gamma_t &= I(s_t; u_t | u_0^{t-B-1}, s_{-1}) \\
&= \lambda_t(0, B), \tag{3.44}
\end{align*}

where (3.42) and (3.44) follows from the Markov chain property

$$u_t \rightarrow s_t \rightarrow \{[u]_0^{t-j-1}, s_{-1}\}, \quad j \in \{B, B'\} \tag{3.46}$$

and (3.43) follows from the fact that conditioning reduces differential entropy. In a similar fashion the

inequality $\gamma_t(0, B') \leq \gamma_t(0, B)$ follows from the fact that the estimation error can only be reduced by

having more observations.

3) We show that both $\lambda_t(0, B)$ and $\gamma_t(0, B)$ are increasing functions with respect to $t$. Intuitively

as $t$ increases the effect of having $s_{-1}$ at the decoder vanishes and hence the required rate increases. Consider

\begin{align*}
\lambda_{t+1}(0, B) &= I(s_{t+1}; u_{t+1} | u_0^{t-B}, s_{-1}) \\
&= h(u_{t+1} | u_0^{t-B}, s_{-1}) - h(u_{t+1} | s_{t+1}) \\
&= h(u_{t+1} | u_0^{t-B}, s_{-1}) - h(u_t | s_t) \\
&\geq h(u_{t+1} | u_0^{t-B}, s_{-1}, s_0) - h(u_t | s_t) \\
&= h(u_{t+1} | u_0^{t-B}, s_0) - h(u_t | s_t) \\
&= h(u_t | u_0^{t-B-1}, s_{-1}) - h(u_t | s_t) \\
&= I(s_t; u_t | u_0^{t-B-1}, s_{-1}) \\
&= \lambda_t(0, B), \tag{3.51}
\end{align*}

where (3.47) and (3.50) follow from time-invariant property of the source model and the test channel, (3.48) follows from the fact that conditioning reduces differential entropy and (3.49) uses the following

Markov chain property

$$\{u_0, s_{-1}\} \rightarrow \{[u]_0^{t-B}, s_0\} \rightarrow u_{t+1}. \tag{3.52}$$

Similarly,

\begin{align*}
\frac{1}{2} \log (2\pi e \cdot \gamma_{t+1}(0, B)) &= h(s_{t+1} | u_0^{t-B}, u_{t+1}, s_{-1}) \\
&\geq h(s_{t+1} | u_0^{t-B}, u_{t+1}, s_0, s_{-1}) \\
&= h(s_{t+1} | u_0^{t-B}, u_{t+1}, s_0) \\
&= h(s_{t} | u_0^{t-B-1}, u_{t}, s_{-1}) \\
&= \frac{1}{2} \log (2\pi e \cdot \gamma_t(0, B)), \tag{3.54}
\end{align*}

where (3.53) follows from the following Markov chain property

$$\{u_0, s_{-1}\} \rightarrow \{[u]_0^{t-B}, u_{t+1}, s_0\} \rightarrow s_{t+1}. \tag{3.55}$$

Since (3.51) and (3.54) hold for every $t \geq B$ the proof of property (3) is complete.
4) Note that for $t < B$ we have $0 \leq B' \leq t$ and thus we can write

$$\lambda_t(k, B') \leq \lambda_t(0, B') \quad \text{(3.56)}$$
$$\leq \lambda_t(0, t) \quad \text{(3.57)}$$
$$= h(u_t | s_{-1}) - h(u_t | s_i)$$
$$= h(u_t | s_{-1}) - h(u_B | s_B)$$
$$= h(u_B | s_{B-1}) - h(u_B | s_B)$$
$$\leq h(u_B | s_{-1}) - h(u_B | s_B) \quad \text{(3.58)}$$
$$= \lambda_B(0, B), \quad \text{(3.59)}$$

where (3.56) follows from part 1 of the lemma, (3.57) is based on the fact that the worse-case erasure pattern contains most possible erasures and follows from the similar steps used in deriving (3.45) and using the fact that if $t < B$, the erasure burst length is at most $t$. Eq. (3.58) follows from the fact that whenever $t < B$ the relation $s_{-1} \rightarrow s_{B-1} \rightarrow u_B$ holds since $t < B$ is assumed. In a similar fashion we can show that $\gamma_t(k, B') \leq \gamma_B(0, B)$.

This completes the proof of Lemma 8.

Following the four parts of Lemma 8, it follows that the worst-case erasure pattern happens at steady state, i.e., $t \to \infty$ when there is a burst of length $B$ which spans $\{t-B, \ldots, t-1\}$. According to this and Lemma 7, any pair $(R, D)$ is achievable if

$$R \geq \lim_{t \to \infty} \lambda_t(0, B) \quad \text{(3.60)}$$
$$D \geq \lim_{t \to \infty} \gamma_t(0, B). \quad \text{(3.61)}$$

**Lemma 9.** Consider $u_t = s_t + z_t$ and suppose the noise variance $\sigma_z^2$ satisfies

$$\Gamma(B, \sigma_z^2) \triangleq \lim_{t \to \infty} E \left[ (s_t - \hat{s}_t([u_{t-1}^{t-B}], u_t))^2 \right] \quad \text{(3.62)}$$
$$= \lim_{t \to \infty} \sigma_z^2 \left( [u_{t-1}^{t-B}], u_t \right) \leq D. \quad \text{(3.63)}$$

The following rate is achievable:

$$R = \Lambda(B, \sigma_z^2) \triangleq \lim_{t \to \infty} I(s_t; u_t | [u_{t-1}^{t-B}]). \quad \text{(3.64)}$$

**Proof.** It suffices to show that any test channel satisfying (3.63) also implies (3.61) and any rate satisfying (3.64) implies (3.60). These relations can be established in a straightforward manner as shown below:

$$R = \Lambda(B, \sigma_z^2) = \lim_{t \to \infty} I(s_t; u_t | [u_{t-1}^{t-B}])$$
$$= \lim_{t \to \infty} \left( h(u_t | [u_{t-1}^{t-B}]) - h(u_t | s_t) \right) \quad \text{(3.65)}$$
$$\geq \lim_{t \to \infty} \left( h(u_t | [u_{t-1}^{t-B}], s_{-1}) - h(u_t | s_t) \right) \quad \text{(3.66)}$$
$$= \lim_{t \to \infty} \lambda_t(0, B) \quad \text{(3.67)}$$
and

\[ D \geq \Gamma(B, \sigma_z^2) = \lim_{t \to \infty} E \left[ (s_t - \hat{s}_t([u]_{0}^{t-B-1}, u_t))^2 \right] \]  

(3.68)

\[ \geq \lim_{t \to \infty} E \left[ (s_t - \hat{s}_t([u]_{0}^{t-B-1}, u_t, s_{t-1}))^2 \right] \]  

(3.69)

\[ = \lim_{t \to \infty} \gamma_i(0, B). \]  

(3.70)

We conclude that \( \Gamma(B, \sigma_z^2) = D \), the rate \( \Lambda^+_{\text{GM-SE}}(B, D) = \Lambda(B, \sigma_z^2) \) is achievable.

### 3.5.2 Numerical Evaluation

We derive an expression for numerically evaluating the noise variance \( \sigma_z^2 \) in (3.30) and also establish (3.7) and (3.10).

To this end it is helpful to consider the following single-variable discrete-time Kalman filter for \( i \in \{0, \ldots, t-B-1\} \):

\[ s_i = \rho s_{i-1} + n_i, \quad n_i \sim N(0, 1 - \rho^2) \]  

(3.71)

\[ u_i = s_i + z_i, \quad z_i \sim N(0, \sigma_z^2). \]  

(3.72)

Note that \( s_i \) can be viewed as the state of the system updated according a Gauss-Markov model and \( u_i \) as the output of the system at each time \( i \), which is a noisy version of the state \( s_i \). Consider the system in steady state, i.e., \( t \to \infty \). The MMSE estimation error of \( s_{t-B} \) given all the previous outputs up to time \( t-B-1 \), i.e., \([u]_{0}^{t-B-1}\) is expressed as (see, e.g., [39, Example V.B.2]):

\[ \Sigma(\sigma_z^2) \triangleq \lim_{t \to \infty} \sigma_{t-B}^2([u]_{0}^{t-B-1}) \]  

(3.73)

\[ = \frac{1}{2} \sqrt{(1 - \sigma_z^2)^2(1 - \rho^2)^2 + 4\sigma_z^2(1 - \rho^2)} + \frac{1 - \rho^2}{2} \]  

(3.74)

Also using the orthogonality principle for MMSE estimation we have

\[ [u]_{0}^{t-B-1} \to \hat{s}_{t-B}([u]_{0}^{t-B-1}) \to s_{t-B} \to s_t. \]  

(3.75)

Thus we can express

\[ s_{t-B} = \hat{s}_{t-B}([u]_{0}^{t-B-1}) + \hat{e}, \]  

(3.76)

where the noise \( \hat{e} \sim N(0, \Sigma(\sigma_z^2)) \) is independent of the observation set \([u]_{0}^{t-B-1}\). Equivalently we can express (see e.g. [40])

\[ \hat{s}_{t-B}([u]_{0}^{t-B-1}) = \tilde{\alpha} s_{t-B} + \hat{e}, \]  

(3.77)

where

\[ \tilde{\alpha} \triangleq 1 - \Sigma(\sigma_z^2). \]  

(3.78)
and \( \tilde{e} \sim \mathcal{N}(0, \Sigma(\sigma_z^2)/(1 - \Sigma(\sigma_z^2))) \) is independent of \( s_{t-B} \). Thus we have

\[
\Lambda(B, \sigma_z^2) = \lim_{t \to \infty} I(s_t; u_t | [u]_{0}^{t-B-1}) \\
= \lim_{t \to \infty} I(s_t; u_t | \tilde{s}_{t-B}([u]_{0}^{t-B-1})) \\
= \lim_{t \to \infty} I(s_t; u_t | \tilde{s}_{t-B} + \tilde{e}) \\
= \lim_{t \to \infty} I(s_t; u_t | \tilde{s}_{t-B} + e) \\
= I(s_t; u_t | \tilde{s}_{t-B}), \tag{3.79}
\]

where we have used (3.77) and \( e \sim \mathcal{N}(0, \Sigma(\sigma_z^2)/(1 - \Sigma(\sigma_z^2))) \). This establishes (3.7) in Prop. 3. In a similar manner,

\[
\Gamma(B, \sigma_z^2) = \lim_{t \to \infty} \sigma_z^2([u]_{0}^{t-B-1}, u_t) \\
= \lim_{t \to \infty} \sigma_z^2(\tilde{s}_{t-B}([u]_{0}^{t-B-1}), u_t) \\
= \lim_{t \to \infty} \sigma_z^2(\tilde{s}_{t-B} + \tilde{e}, u_t) \\
= \lim_{t \to \infty} \sigma_z^2(\tilde{s}_{t-B} + e, u_t) \\
= \sigma_z^2(\tilde{s}_{t-B}, u_t), \tag{3.80}
\]

which establishes (3.10). Furthermore since

\[
s_t = \rho^B s_{t-B} + \tilde{n}, \tag{3.81}
\]

where \( \tilde{n} \sim \mathcal{N}(0, 1 - \rho^{2B}) \),

\[
\Gamma(B, \sigma_z^2) = \sigma_z^2(\tilde{s}_{t-B}, u_t) \tag{3.82}
\]

where (3.83) follows from the application of MMSE estimator and using (3.81), (3.76) and the definition of the test channel in (3.30). Thus the noise \( \sigma_z^2 \) in the test channel (3.30) is obtained by setting

\[
\Gamma(B, \sigma_z^2) = D. \tag{3.84}
\]

This completes the proof of Prop. 3.

### 3.6 Coding Scheme: Multiple erasure bursts with Guard Intervals

We study the achievable rate using the quantize and binning scheme with test channel (3.30) when the channel introduces multiple erasure bursts each of length no greater than \( B \) and with a guard interval of at-least \( L \) symbols separating consecutive erasure bursts. While the coding scheme is the same as the single erasure burst channel model and is based on quantize and binning and MMSE estimation at the
3.6.1 Analysis of Achievable Rate

We introduce the following notation in our analysis. Let $\Omega_t$ denote the set of time indices up to time $t-1$ when the channel packets are not erased, i.e.,

$$\Omega_t = \{i : 0 \leq i \leq t-1, g_i \neq *\}, \tag{3.85}$$

and let us define

$$s_t = \{s_i : i \in \Omega\}, \tag{3.86}$$

$$u_t = \{u_i : i \in \Omega\}. \tag{3.87}$$

Given the erasure sequence $\Omega_t$, and given $g_t = f_t$, the decoder can reconstruct $u^n_t$ provided that the test channel is selected such that the rate satisfies (see e.g., [38])

$$R \geq \lambda_t(\Omega_t) \triangleq I(s_t; u_t|u_{\Omega_t}, s_{t-1}) \tag{3.88}$$

and the distortion constraint satisfies

$$\gamma_t(\Omega_t) \triangleq E[(s_t - \hat{s}_t(u_{\Omega_t}, u_t, s_{t-1}))^2] = \sigma_t^2(u_{\Omega_t}, u_t, s_{t-1}) \leq D \tag{3.89}$$

for each $t \geq 0$ and each feasible set $\Omega_t$. Thus we are again required to characterize the $\Omega_t$ for each value of $t$ corresponding to the worst-case erasure pattern. The following two lemmas are useful towards this end.

**Lemma 10.** Consider two sets $A, B \subseteq \mathbb{N}$ each of size $r$, as $A = \{a_1, a_2, \cdots, a_r\}$, $B = \{b_1, b_2, \cdots, b_r\}$ such that $1 \leq a_1 < a_2 < \cdots < a_r$ and, $1 \leq b_1 < b_2 < \cdots < b_r$ and for any $i \in \{1, \ldots, r\}$, $a_i \leq b_i$. Then the test channel (3.30) satisfies the following:

$$h(s_t|u_A, s_{-1}) \geq h(s_t|u_B, s_{-1}), \quad \forall t \geq b_r \tag{3.90}$$

$$h(u_t|u_A, s_{-1}) \geq h(u_t|u_B, s_{-1}), \quad \forall t > b_r \tag{3.91}$$

The proof of Lemma 10 is available in Appendix B.3.

**Lemma 11.** Assume that at time $t$, $g_t = f_t$ and let $\Omega_t$ be as defined in (3.85).

1. Among all feasible sets $\Omega_t$ of size $|\Omega_t| = \theta$, $\lambda_t(\Omega_t)$ and $\gamma_t(\Omega_t)$ are maximized by a set $\Omega^*_t(\theta)$ where all the erasures happen in the closest possible locations to time $t$.

2. For each fixed $t$, both $\lambda_t(\Omega^*_t(\theta))$ and $\gamma_t(\Omega^*_t(\theta))$ are maximized by the minimum possible value of $\theta$. Equivalently, the maximizing set, denoted by $\Omega^*_t$, corresponds to the erasure pattern with maximum number of erasures.
Chapter 3. Zero-Delay Streaming of Gauss-Markov Sources: Immediate Recovery

3. Zero-Delay Streaming of Gauss-Markov Sources: Immediate Recovery

3. Both \( \lambda_t(\Omega^*_t) \) and \( \gamma_t(\Omega^*_t) \) are increasing functions with respect to \( t \).

The proof of Lemma 11 is presented in Appendix B.4. We present an example in Fig. 3.8 to illustrate Lemma 11. We assume \( t = 18 \). The total number of possible erasures up to time \( t = 18 \) is restricted to be 5, or equivalently the number of non-erased packets is \( \theta = 13 \) in Fig 3.8a. The set \( \Omega^*_{18}(13) \) indicates the set of non-erased indices associated with the worst case erasure pattern. Based on part 2 of Lemma 11, Fig. 3.8b shows the worst case erasure pattern for time \( t = 18 \), which includes the maximum possible erasures.

Following the three steps in Lemma 11 a rate-distortion pair \( (R, D) \) is achievable if

\[
R \geq \lim_{t \to \infty} \lambda_t(\Omega^*_t) \\
D \geq \lim_{t \to \infty} \gamma_t(\Omega^*_t)
\]

(3.92)  
(3.93)

**Lemma 12.** Any test channel noise \( \sigma^2_z \) satisfying (3.13) and (3.14) in Prop. 4, i.e.,

\[
R \geq I(s_t; u_t | \hat{s}_{t-L-B+1}, [u]_{t-L-B+1}^{t-B-1}) \\
D \geq \sigma^2_z(\hat{s}_{t-L-B+1}, [u]_{t-L-B+1}^{t-B-1}, u_t)
\]

(3.94)  
(3.95)

where \( \hat{s}_{t-L-B} = s_{t-L-B} + \epsilon \), where \( \epsilon \sim \mathcal{N}(0, D/(1-D)) \), also satisfies (3.92) and (3.93).

**Proof.** See Appendix B.5.

This completes the proof of Prop 4.

### 3.6.2 Numerical Evaluation

We derive the expression for numerically evaluating \( \sigma^2_z \). To this end, first note that the estimation error of estimating \( s_{t-B-1} \) from \( \{\hat{s}_{t-L-B}, [u]_{t-L-B+1}^{t-B-1}\} \) can be computed as follows.

\[
\eta(\sigma^2_z) = \sigma^2_z(\hat{s}_{t-L-B+1}, [u]_{t-L-B+1}^{t-B-1})
\]

where \( \hat{s}_{t-L-B} = s_{t-L-B} + \epsilon \), where \( \epsilon \sim \mathcal{N}(0, D/(1-D)) \), also satisfies (3.92) and (3.93).
\[ R = E [\hat{s}_{i-B-1}^2] - E [s_{i-B-1} U] (E [U^T U])^{-1} E [s_{i-B-1} U^T] \]  
\[ = 1 - A_1(A_2)^{-1}A_1^T \]  
(3.96)  
(3.97)

where we define

\[ U \triangleq \left[ u_{i-B-1} \ u_{i-B-2} \ \ldots \ u_{i-L-B+1} \ \tilde{s}_{i-L-B} \right] \]

and \((.)^T\) denotes the transpose operation. Also note that \(A_1\) and \(A_2\) can be computed as follows.

\[
A_1 = (1, \rho, \rho^2, \ldots, \rho^{L-1})
\]

\[
A_2 = \begin{pmatrix}
1 + \sigma_z^2 & \rho & \ldots & \rho^{L-2} & \rho^{L-1} \\
\rho & 1 + \sigma_z^2 & \ldots & \rho^{L-3} & \rho^{L-2} \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
\rho^{L-2} & \rho^{L-3} & \ldots & 1 + \sigma_z^2 & \rho \\
\rho^{L-1} & \rho^{L-2} & \ldots & \rho & 1 + \frac{D}{2}\eta
\end{pmatrix}
\]

According to (3.95) we can write

\[
D = \sigma_i^2 (\tilde{s}_{i-L-B}, [u]_{i-L-B+1}^{i-B-1}, u_i) = \sigma_i^2 (\tilde{s}_{i-L-B}, [u]_{i-L-B+1}^{i-B-1}, u_i)
\]

\[
= \left[ \frac{1}{\sigma_z^2} + \frac{1}{1 - \rho^2(B+1)(1 - \eta(\sigma_z^2))} \right]^{-1}
\]

(3.100)  
(3.101)

Therefore by solving (3.101) the expression for \(\sigma_z^2\) can be obtained. Finally the achievable rate is computed as:

\[
R_{\text{GM-ME}}^+(L, B, D) = I(\tilde{s}_i; u_i | \tilde{s}_{i-L-B}, [u]_{i-L-B+1}^{i-B-1})
\]

\[
= h(\tilde{s}_i | \tilde{s}_{i-L-B}, [u]_{i-L-B+1}^{i-B-1}) - h(\tilde{s}_i | \tilde{s}_{i-L-B}, [u]_{i-L-B+1}^{i-B-1}, u_i)
\]

\[
= h(\tilde{s}_i | \tilde{s}_{i-L-B}, [u]_{i-L-B+1}^{i-B-1}) - \frac{1}{2} \log(2\pi e D)
\]

\[
= \frac{1}{2} \log \left( \frac{1 - \rho^2(B+1)(1 - \eta(\sigma_z^2))}{D} \right) - \frac{1}{2} \log(2\pi e D)
\]

\[
= \frac{1}{2} \log \left( \frac{1 - \rho^2(B+1)(1 - \eta(\sigma_z^2))}{D} \right)
\]

(3.102)

### 3.7 Upper Bound in the High Resolution Regime

We investigate the behavior of the lossy rate-recovery functions for Gauss-Markov sources for single and multiple erasure burst channel models, i.e., \(R_{\text{GM-SE}}(B, D)\) and \(R_{\text{GM-ME}}(L, B, D)\), in the high resolution regime and establish Corollary 3. The following inequalities can be readily verified.

\[
R_{\text{GM-SE}}^-(B, D) \leq R_{\text{GM-SE}}(B, D) \leq R_{\text{GM-ME}}(L, B, D) \leq R_{\text{GM-ME}}^+(L, B, D)
\]

(3.103)

The first and the last inequalities in (3.103) are by definition and the second inequality follows from the fact that the rate achievable for multiple erasure model is also achievable for single erasure burst as the decoder can simply ignore the available codewords in reconstructing the source sequences. According to
(3.103), it suffices to characterize the high resolution limit of $R^{-}_{\text{GM-SE}}(B, D)$ and $R^{+}_{\text{GM-ME}}(L, B, D)$ in Prop. 2 and Prop. 4 respectively.

For the lower bound, note that as $D \to 0$ the expression for $\Delta$ in (3.6) satisfies

$$\Delta \triangleq (D\rho^2 + 1 - \rho^2(B+1))^2 - 4D\rho^2(1 - \rho^2B) \to (1 - \rho^2(B+1))^2.$$ 

Upon direct substitution in (3.6) we have that

$$\lim_{D \to 0} \left\{ R^{-}_{\text{GM-SE}}(B, D) - \frac{1}{2} \log \left( \frac{1 - \rho^2(B+1)}{D} \right) \right\} = 0, \quad (3.104)$$ 

as required.

To establish the upper bound note that according to Prop. 4 we can write

$$R^{+}_{\text{GM-ME}}(L, B, D) = I(\hat{s}_t; u_t|\hat{s}_{t-L-B}, [u]_{t-L-B+1}^{t-B})$$

$$= h(\hat{s}_t|\hat{s}_{t-L-B}, [u]_{t-L-B+1}^{t-B}) - h(\hat{s}_t|\hat{s}_{t-L-B}, [u]_{t-L-B+1}^{t-B}, u_t)$$

$$= h(\hat{s}_t|\hat{s}_{t-L-B}, [u]_{t-L-B+1}^{t-B}) - \frac{1}{2} \log(2\pi eD) \quad (3.105)$$

where the last term follows from the definition of $\hat{s}_{t-L-B}$ in Prop. 4. Also we have

$$h(\hat{s}_t|\hat{s}_{t-B-1}) \leq h(\hat{s}_t|\hat{s}_{t-L-B}, [u]_{t-L-B+1}^{t-B}) \leq h(\hat{s}_t|u_{t-B-1}) \quad (3.106)$$

where the left hand side inequality in (3.106) follows from the following Markov property,

$$\{\hat{s}_{t-L-B}, [u]_{t-L-B+1}^{t-B} \} \to s_{t-B-1} \to s_t \quad (3.107)$$

and the fact that conditioning reduces the differential entropy. Also, the right hand side inequality in (3.106) follows from the latter fact. By computing the upper and lower bounds in (3.106) we have

$$\frac{1}{2} \log \left( 2\pi e(1 - \rho^2(B+1)) \right) \leq h(\hat{s}_t|\hat{s}_{t-L-B}, [u]_{t-L-B+1}^{t-B}) \leq \frac{1}{2} \log \left( 2\pi e \left( 1 - \frac{\rho^2(B+1)}{1 + \sigma_z^2} \right) \right) \quad (3.108)$$

Now note that

$$D \geq \sigma_t^2(\hat{s}_{t-L-B}, [u]_{t-L-B+1}^{t-B}, u_t) \geq \sigma_t^2(u_t, s_{t-1})$$

$$= \left( \frac{1}{\sigma_z^2} + \frac{1}{1 - \rho^2} \right)^{-1} \quad (3.109)$$

which equivalently shows that if $D \to 0$ we have that $\sigma_z^2 \to 0$. By computing the limit of the upper and lower bounds in (3.108) as $D \to 0$, we can see that

$$\lim_{D \to 0} \left\{ h(\hat{s}_t|\hat{s}_{t-L-B}, [u]_{t-L-B+1}^{t-B}) - \frac{1}{2} \log \left( 2\pi e(1 - \rho^2(B+1)) \right) \right\} = 0 \quad (3.111)$$

Finally (3.111) and (3.105) results in

$$\lim_{D \to 0} \left\{ R^{+}_{\text{GM-ME}}(L, B, D) - \frac{1}{2} \log \left( \frac{1 - \rho^2(B+1)}{D} \right) \right\} = 0 \quad (3.112)$$
as required. Equations (3.104), (3.112) and (3.103) establish the results of Corollary 3.

3.8 Independent Gaussian Sources with Sliding Window Recovery

In this section we study the memoryless Gaussian source model discussed in Section 3.3.4. The source sequences are drawn i.i.d. both in spatial and temporal dimension according to a unit-variance, zero-mean, Gaussian distribution \( \mathcal{N}(0,1) \). The rate-\( R \) causal encoder sequentially compresses the source sequences and sends the codewords through the erasure burst channel. The channel erases a single burst of maximum length \( B \) and perfectly reveals the rest of the packets to the decoder. The decoder at each time \( i \) reconstructs \( K + 1 \) past source sequences, i.e., \( (s_i^n, s_{i-1}^n, \ldots, s_{i-K}^n) \) within a vector distortion measure \( d = (d_0, \ldots, d_K) \). More recent source sequences are required to be reconstructed with less distortion, i.e., \( d_0 \leq d_1 \leq \ldots \leq d_K \). The decoder however is not interested in reconstructing the source sequences during the error propagation window, i.e., during the erasure burst and a window of length \( W \) after the erasure burst ends.

For this setup, we establish the rate-recovery function stated in Theorem 3. We do this by presenting the coding scheme in Section 3.8.2 and the converse in Section 3.8.3. We also study some baseline schemes and compare their performance with the rate-recovery function at the end of this section.

3.8.1 Sufficiency of \( K = B + W \)

In our analysis we only consider the case \( K = B + W \). The coding scheme can be easily extended to a general \( K \) as follows. If \( K < B + W \), we can assume that the decoder, instead of recovering the source \( t_i = (s_i, s_{i-1}, \ldots, s_{i-K})^T \) at time \( i \) within distortion \( d \), aims to recover the source \( t'_i = (s_i, \ldots, s_{i-K'})^T \) within distortion \( d' \) where \( K' = B + W \) and

\[
d'_j = \begin{cases} d_j & \text{for } j \in \{0,1,\ldots,K\} \\ 1 & \text{for } j \in \{K+1,\ldots,K'\}, \end{cases}
\]  

and thus this case is a special case of \( K = B + W \). Note that layers \( K+1,\ldots,K' \) require zero rate as the source sequences have unit variance.

If \( K > B + W \), for each \( j \in \{B + W + 1, \ldots, K\} \) the decoder is required to reconstruct \( s^n_{i-j} \) within distortion \( d_j \). However we note the rate associated with these layers is again zero. In particular there are two possibilities during the recovery at time \( i \). Either, \( t^n_{i-1} \) or, \( t^n_{i-B-W-1} \) are guaranteed to have been reconstructed. In the former case \( \{s^n_{i-j}\}_{d_{j-1}} \) is available from time \( i-1 \) and \( d_{j-1} \leq d_j \). In the latter case \( \{s^n_{i-j}\}_{d_{j-W-B-1}} \) is available from time \( i-B-W-1 \) and again \( d_{j-W-B-1} \leq d_j \). Thus the reconstruction of any layer \( j \geq B + W \) does not require any additional rate and it again suffices to assume \( K = B + W \).

3.8.2 Coding Scheme

Throughout our analysis, we assume the source sequences are of length \( n \cdot r \) where both \( n \) and \( r \) will be assumed to be arbitrarily large. The block diagram of the scheme is shown in Fig. 3.9. We

\footnote{The notation \( \{s^n_{i-j}\}_d \) indicates the reconstruction of \( s^n_i \) within average distortion \( d \).}
partition \( s_{1}^{nr} \) into \( r \) blocks each consisting of \( n \) symbols \((s_{i}^{n})_{i}\). We then apply a successive refinement quantization codebook to each such block to generate \( B + 1 \) indices \((\{m_{i,j}\}_{j=0}^{B})_{i}\). Thereafter these indices are carefully rearranged in time to generate \((c_{i})_{i}\). At each time we thus have a length \( r \) sequence \( c_{r} \triangleq \{(c_{i})_{1}, \ldots, (c_{i})_{r}\} \). We transmit the bin index of each sequence over the channel as in Section 2.4.3. At the receiver the sequence \( \hat{c}_{r} \) is first reconstructed by the inner decoder. Thereafter upon rearranging the refinement layers in each packet, the required reconstruction sequences are produced. We provide the details of the encoding and decoding below.

![Figure 3.9: Schematic of encoder and decoder for i.i.d. Gaussian source with sliding window recovery constraint. SR and LR indicate successive refinement and layer rearrangement (Sections 3.8.2 and 3.8.2), respectively.](image)

\( s_{1}^{nr} \)

Source Partitioning Block

\( (s_{1}^{n})_{1} \)

SR Code 1

\( (m_{1,j})_{j=0}^{B} \)

LR Block 1

\( (c_{1})_{1} \)

Bin Index Assignment

\( f_{i} \in [0, 2^{nrR}] \)

Burst Erasure Channel

\( s_{1}^{nr} \)

Source Reproduction Block

\( \hat{c}_{r} \)

Joint Typicality Decoder

\( i-W \leq j \leq i \)


\( \{\hat{s}_{i}^{n}\}_{d_{B+w}} \)

Encoder 0

\( m_{0} \)

Decoder 0

\( \{\hat{s}_{i}^{n}\}_{d_{0}} \)

Encoder 1

\( m_{1} \)

Decoder 1

\( \{\hat{s}_{i}^{n}\}_{d_{1}} \)

Encoder \( B-1 \)

\( m_{B-1} \)

Decoder \( B-1 \)

\( \{\hat{s}_{i}^{n}\}_{d_{B+w-1}} \)

Encoder \( B \)

\( m_{B} \)

Decoder \( B \)

\( \{\hat{s}_{i}^{n}\}_{d_{B+w}} \)


\( s_{1}^{n} \)


Figure 3.10: \((B + 1)\)-layer coding scheme based on successive refinement (SR). Note that for each \( k \in \{0, \ldots, B\}, m_{i,k} \) is of rate \( \tilde{R}_{k} \) and \( M_{i,k} \) is of rate \( R_{k} \). The dashed box represents the SR code.

![Figure 3.10: (B + 1)-layer coding scheme based on successive refinement (SR). Note that for each k ∈ {0, ..., B}, mi,k is of rate R̃k and Mi,k is of rate Rk. The dashed box represents the SR code.](image)
Successive Refinement (SR) Encoder

The encoder at time $i$, first partitions the source sequence $s_i^n$ into $r$ source sequences $(s_i^l)_l$, $l \in \{1, \ldots, r\}$. As shown in Fig. 3.10, we encode each source signal $(s_i^l)_l$ using a $(B + 1)$-layer successive refinement codebook [15,41] to generate $(B + 1)$ codewords whose indices are given by $\{(m_{i,0})_l, (m_{i,1})_l, \ldots, (m_{i,B})_l\}$ where $(m_{i,j})_l \in \{1, 2, \ldots, 2^{nR_i}\}$ for $j \in \{0, 1, \ldots, B\}$ and

$$
\tilde{R}_j = \begin{cases} 
\frac{1}{2} \log\left( \frac{d_{W+1}}{d_0} \right) & \text{for } j = 0 \\
\frac{1}{2} \log\left( \frac{d_{W+j+1}}{d_{W+j}} \right) & \text{for } j \in \{1, 2, \ldots, B - 1\} \\
\frac{1}{2} \log\left( \frac{1}{d_{W+B}} \right) & \text{for } j = B,
\end{cases}
$$

(3.114)

The $j$-th layer uses indices

$$(M_{i,j})_l \triangleq \{(m_{i,j})_l, \ldots, (m_{i,B})_l\}$$

(3.115)

for reproduction and the associated rate with layer $j$ is given by:

$$
R_j = \begin{cases} 
\sum_{k=0}^{B} \tilde{R}_k = \frac{1}{2} \log\left( \frac{1}{d_0} \right) & \text{for } j = 0 \\
\sum_{k=j}^{B} \tilde{R}_k = \frac{1}{2} \log\left( \frac{1}{d_{W+j}} \right) & \text{for } j \in \{1, 2, \ldots, B\},
\end{cases}
$$

(3.116)

and the corresponding distortion associated with layer $j$ equals $d_0$ for $j = 0$ and $d_{W+j}$ for $j \in \{1, 2, \ldots, B\}$.

From Fig. 3.10 it is clear that for any $i$ and $j \in \{0, \ldots, B\}$, the $j$-th layer $M_{i,j}$ is a subset of $j - 1$-th layer $M_{i,j-1}$, i.e., $M_{i,j} \subseteq M_{i,j-1}$.

Layer Rearrangement (LR) and Binning

In this stage the encoder rearranges the outputs of the SR blocks associated with different layers to produce an auxiliary set of sequences as follows:\footnote{We suppress the index $l$ in (3.115) for compactness.}

$$
c_i \triangleq \begin{pmatrix} M_{i,0} \\
M_{i-1,1} \\
M_{i-2,2} \\
\vdots \\
M_{i-B,B} \end{pmatrix}
$$

(3.117)

In the definition of (3.117) we note that $M_{i,0}$ consists of all the refinement layers associated with the source sequence at time $i$. It can be viewed as the “innovation symbol” since it is independent of all past symbols. It results in a distortion of $d_0$. The symbol $M_{i-1,1}$ consists of all refinement layers of the source sequence at time $i - 1$, except the last layer and results in a distortion of $d_1$. Recall that $M_{i-1,1} \subseteq M_{i-0,0}$. In a similar fashion $M_{i-B,B}$ is associated with the source sequence at time $i - B$ and results in a distortion of $d_B$. Fig. 3.11 illustrates a schematic of these auxiliary codewords.

Note that as shown in Fig. 3.10 the encoder at each time generates $r$ independent auxiliary codewords $(c_i)_l, \ldots, (c_i)_r$. Let $c_i^r$ be the set of all $r$ codewords. In the final step, the encoder generates $f_i$, the
bin index associated with the codewords $e_i^r$ and transmit this through the channel. The bin indices are randomly and independently assigned to all the codewords beforehand and are revealed to both encoder and decoder.

**Decoding and Rate Analysis**

To analyze the decoding process, first consider the simple case where the actual codewords $e_i^r$ defined in (3.117), and not the assigned bin indices, are transmitted through the channel. In this case, whenever the channel packet is not erased by the channel, the decoder has access to the codewords $c_i$. According to (3.117) the decoder also knows $(M_i^r, M_i^r, \ldots, M_i^r)$. Thus we have shown that if actual codewords $c_i^r$ defined in (3.117) are transmitted the required distortion constraints are satisfied. It can be verified from (3.117) and (3.116) that the rate associated is known. According to (3.117), $(M_i^r, M_i^r, \ldots, M_i^r)$ are each known within distortion $d_0$. In addition, since $c_i-W$ is known, according to (3.117), $(M_i-W-1, M_i-W-2, \ldots, M_i-B-W-B)$ is known and according to SR structure depicted in Fig. 3.10 the source sequences $(s_{i-W-1}^{nr}, s_{i-W-2}^{nr}, \ldots, s_{i-B-W-B}^{nr})$ are known within distortion $(d_{i-W-1}, d_{i-W-2}, \ldots, d_{i-B-W-B})$ which satisfies the distortion constraint. Now consider the case where $\hat{t}_{i-1}^{nr}$ and $\hat{e}_i^r$ are available, i.e., $\hat{t}_{i-1}^{nr}$ is already reconstructed within the required distortion vector, the decoder is able to reconstruct $\hat{t}_{i}^{nr}$ from $\hat{t}_{i-1}^{nr}$ and $e_i^r$. In particular, from $M_i^r$ the source sequence $s_i^{nr}$ is reconstructed within distortion $d_0$. Also reconstruction of $s_i^{nr}$ within distortion $d_{k-1}$ is already available from $\hat{t}_i$ for $k \in \{1, \ldots, B + W\}$ which satisfies the distortion constraint as $d_{k-1} \leq d_k$.

Thus we have shown that if actual codewords $c_i^r$ defined in (3.117) are transmitted the required distortion constraints are satisfied. It can be verified from (3.117) and (3.116) that the rate associated

![Figure 3.11: Schematic of the auxiliary codewords defined in (3.117). The codewords are temporally correlated in a diagonal form depicted using ellipses. In particular, as shown in Fig. 3.10, $M_{i-j,j} \subseteq M_{i-j,j-1}$. Based on this diagonal correlation structure, the codewords depicted in the boxes are sufficient to know all the codewords.](image-url)
with the $c_i^r$ is given by

$$R_c = \sum_{k=0}^{B} R_k = \frac{1}{2} \log \left( \frac{1}{d_0} \right) + \sum_{j=1}^{B} \frac{1}{2} \log \left( \frac{1}{d_{W+j}} \right).$$

(3.118)

Thus compared to the achievable rate (3.19) in Theorem 3 we are missing the factor of $\frac{1}{W+1}$ in the second term. To reduce the rate, note that, as shown in Fig. 3.11 and based on definition of the auxiliary codewords in (3.117), there is a strong temporal correlation among the consecutive codewords. We therefore bin the set of all sequences $c_i^r$ into $2^{nR}$ bins as in Section 2.4.3. The encoder, upon observing $c_i^r$, only transmits its bin index $f_i$ through the channel. We next describe the decoder and compute the minimum rate required to reconstruct $c_i^r$.

Outside the error propagation window, one of the following cases can happen as discussed below. We claim that in either case the decoder is able to reconstruct $c_i^r$ as follows.

- In the first case, the decoder has already recovered $c_1^{r-1}$ and attempts to recover $c_i^r$ given $(f_i, c_i^{r-1})$. This succeeds with high probability if

$$nR \geq H(c_i|c_{i-1})$$

(3.119)

$$= H(M_{i,0}, M_{i-1,1}, \ldots, M_{i-1,B}, B|c_{i-1})$$

(3.120)

$$= H(M_{i,0}, M_{i-1,1}, \ldots, M_{i-B,B}|M_{i-1,0}, M_{i-2,1}, \ldots, M_{i-B-B-1,1})$$

(3.121)

$$= H(M_{i,0})$$

(3.122)

$$= nR_0$$

(3.123)

where we use (3.117) in (3.120) and (3.121), and the fact that layer $j$ is a subset of layer $j-1$, i.e., $M_{i-j,j} \subseteq M_{i-j,j-1}$ in (3.122). Thus the reconstruction of $c_i^r$ follows since the choice of (3.19) satisfies (3.123). Thus according to the second part of Claim 1, the decoder is able to reconstruct $\hat{t}_i^{w-r}$.

- In the second case we assume that the decoder has not yet successfully reconstructed $c_i^{r-1}$ but is required to reconstruct $c_i^r$. In this case $c_i^r$ is the first sequence to be recovered following the end of the error propagation window. Our proposed decoder uses $(f_i, f_i-1, \ldots, f_i-W)$ to simultaneously reconstruct $(c_i^r, \ldots, c_i^{r-W})$. This succeeds with high probability provided:

$$n(W+1)R \geq H(c_i-W, c_i-W+1, \ldots, c_i)$$

(3.124)

$$= H(c_i-W, M_{i-W+1,0}, M_{i-W+2,0}, \ldots, M_{i,0})$$

(3.125)

$$= H(c_i-W) + \sum_{k=1}^{W} H(M_{i-W+k,0})$$

$$= H(M_{i-W,0}, M_{i-W-1,1}, \ldots, M_{i-B-W,B}) + \sum_{k=1}^{W} H(M_{i-W+k,0})$$

$$= n \sum_{k=1}^{B} R_k + n(W+1)R_0$$

(3.126)

where in (3.124) we use the fact that the sub-symbols satisfy $M_{i,j+1} \subseteq M_{i,j}$ as illustrated in Fig. 3.11. In particular, in computing the rate in (3.124) all the sub-symbols in $c_i-W$ and the
sub-symbols $M_{j,0}$ for $j \in \{i-W+1, \ldots, i\}$ need to be considered. From (3.123), (3.126) and (3.116), the rate $R$ is achievable if

$$R \geq R_0 + \frac{1}{W+1} \sum_{k=1}^{B} R_k$$

(3.127)

$$= \frac{1}{2} \log \left( \frac{1}{d_0} \right) + \frac{1}{2(W+1)} \sum_{k=1}^{B} \log \left( \frac{1}{d_{W+k}} \right).$$

(3.128)

as required. Thus, the rate constraint in (3.128) is sufficient for the decoder to recover the codewords $(c_{i}^r, \ldots, c_{f-W}^r)$ right after the error propagation window and to reconstruct $\hat{t}_i^n$ according to Claim 1.

Thus, the rate constraint in (3.128) is sufficient for the decoder to succeed in reconstructing the source sequences within required distortion constraints at the any time $i$ outside the error propagation window. This completes the justification of the upper bound in Theorem 3.

### 3.8.3 Converse

Here we study the converse proof for Theorem 3. We need to show that for any sequence of codes that achieve a distortion tuple $(d_0, \ldots, d_{W+B})$ the rate is lower bounded by (3.128). As in the proof of Theorem 1, we consider a erasure burst of length $B$ spanning the time interval $\{t-B-W, \ldots, t-W-1\}$. Consider,

$$(W+1)nR \geq H([f]_{l-W})$$

$$\geq H([f]_{l-W}|[f]_{0}^{t-B-W-1}, s_{n-1}^n)$$

(3.129)

where the last step follows from the fact that conditioning reduces entropy. We need to lower bound the entropy term in (3.129). Consider

$$H([f]_{l-W}|[f]_{0}^{t-B-W-1}, s_{n-1}^n) = I([f]_{l-W}; t_n^n|[f]_{0}^{t-B-W-1}, s_{n-1}^n) + H([f]_{l-W}|[f]_{0}^{t-B-W-1}, t_n^n, s_{n-1}^n)$$

(3.130)

$$= h(t_n^n|[f]_{0}^{t-B-W-1}, s_{n-1}^n) - h(t_n^n|[f]_{0}^{t-B-W-1}, [f]_{l-W}, s_{n-1}^n) + H([f]_{l-W}|[f]_{0}^{t-B-W-1}, t_n^n, s_{n-1}^n)$$

(3.131)

where (3.131) follows since $t_n^n = (s_{n-B-W}, \ldots, s_{n})$ is independent of $([f]_{0}^{t-B-W-1}, s_{n-1}^n)$ as the source sequences $s_i^n$ are generated i.i.d. By expanding $t_n^n$ we have that

$$h(t_n^n) = h(s_{n-B-W}, \ldots, s_{n-W-1}) + h(s_{n-W}, \ldots, s_{n})$$

(3.132)

and

$$h(t_n^n|[f]_{0}^{t-B-W-1}, [f]_{l-W}, s_{n-1}^n) = h(s_{n-B-W}, \ldots, s_{n-W-1}|[f]_{0}^{t-B-W-1}, [f]_{l-W}, s_{n-1}^n)$$

$$+ h(s_{n-W}, \ldots, s_{n}|[f]_{0}^{t-B-W-1}, [f]_{l-W}, s_{n-B-W}, \ldots, s_{n-W-1}, s_{n-1}^n).$$

(3.133)
We next establish the following claim, whose proof is given in Appendix B.6.

**Lemma 13.** The two inequalities

\[ h(s^n_{t-B-W}, \ldots, s^n_{t-W-1}) - h(s^n_{t-B-W}, \ldots, s^n_{t-W-1}) \geq \sum_{i=1}^{B} \frac{n}{2} \log \left( \frac{1}{d_{W+i}} \right) \]  

and

\[ h(s^n_{t-W}, s^n_{t}) - h(s^n_{t-W}, s^n_{t}) \geq \frac{n(W + 1)}{2} \log \left( \frac{1}{d_0} \right) + \frac{n(W + 1)}{2} \log \left( \frac{1}{d_0} \right) \]  

hold.

**Proof.** See Appendix B.6.

From (3.131), (3.132), (3.133), (3.134) and (3.135), we can write

\[ H([f]_{t-W}^t| [f]_{0}^{t-B-W-1}, s^n_{t-W}, s^n_{t-1}) \geq \frac{n(W + 1)}{2} \log \left( \frac{1}{d_0} \right) \]  

as required.

### 3.8.4 Illustrative Suboptimal Schemes

We compare the optimal lossy rate-recovery function with the following suboptimal schemes.

**Still-Image Compression**

In this scheme, the encoder ignores the decoder’s memory and at time \( i \geq 0 \) encodes the source \( t_i \) in a memoryless manner and sends the codewords through the channel. The rate associated with this scheme is

\[ R_{SI}(d) = I(t_i; t_i) = \sum_{k=0}^{K} \frac{1}{2} \log \left( \frac{1}{d_k} \right) \]  

In this scheme, the decoder is able to recover the source whenever its codeword is available, i.e., at all the times except when erasures happen.
Wyner-Ziv Compression with Delayed Side Information

At time $i$ the encoders assumes that $t_{i-B-1}$ is already reconstructed at the receiver within distortion $d$. With this assumption, it compresses the source $t_i$ according to Wyner-Ziv scheme and transmits the codewords through the channel. The rate of this scheme is

$$R_{WZ}(B, d) = I(t_i; \hat{t}_i|t_{i-B-1}) = \sum_{k=0}^{B} \frac{1}{2} \log \left( \frac{1}{d_k} \right)$$ (3.139)

Note that, if at time $i$, $\hat{t}_{i-B-1}$ is not available, $\hat{t}_{i-1}$ is available and the decoder can still use it as side-information to construct $\hat{t}_i$ since $I(t_i; t_i|\hat{t}_{i-B-1}) \geq I(t_i; t_i|\hat{t}_{i-1})$.

As in the case of Still-Image Compression, the Wyner-Ziv scheme also enables the recovery of each source sequence except those with erased codewords.

Predictive Coding plus FEC

This scheme consists of predictive coding followed by a Forward Error Correction (FEC) code to compensate the effect of packet losses of the channel. As the contribution of $B$ erased codewords need to be recovered using $W + 1$ available codewords, the rate of this scheme can be computed as follows.

$$R_{FEC}(B, W, d) = \frac{B + W + 1}{W + 1} I(t_i; \hat{t}_i|\hat{t}_{i-1})$$

$$= \frac{B + W + 1}{2(W + 1)} \log \left( \frac{1}{d_0} \right)$$ (3.141)

Group-of-Picture (GOP)-Based Compression

This scheme consists of predictive coding where synchronization frames (I-frames) are inserted periodically to prevent error propagation. The synchronization frames are transmitted with the rate $R_1 = I(t_i; t_i)$ and the rest of the frames are transmitted at the rate $R_2 = I(t_i; t_i|\hat{t}_{i-1})$ using predictive coding. Whenever an erasure happens the decoder fails to recover the source sequences until the next synchronization frame and then the decoder becomes synced to the encoder. In order to guarantee the recovery of the source sequences, the synchronization frames have to be inserted with a period of at most $W + 1$. This results in the following average rate expression.

$$R = \frac{1}{(W+1)} I(t_i; \hat{t}_i) + \frac{W}{(W+1)} I(t_i; \hat{t}_i|\hat{t}_{i-1})$$

$$= \frac{1}{2(W+1)} \sum_{k=0}^{K} \log \left( \frac{1}{d_k} \right) + \frac{W}{2(W+1)} \log \left( \frac{1}{d_0} \right)$$ (3.143)

In Fig. 3.12, we compare the result in Theorem 3 with the described schemes. It can be observed from Fig. 3.12 that except when $W = 0$ none of the other schemes are optimal. The Predictive Coding plus FEC scheme, which is a natural separation based scheme and the GOP-based compression scheme are suboptimal even for relatively large values of $W$. Also note that the GOP-based compression scheme reduces to Still-Image compression for $W = 0$. 
Figure 3.12: Comparison of rate-recovery of suboptimal systems to minimum possible rate-recovery function for different recovery window length $W$. We assume $K = 5$, $B = 2$ and a distortion vector $d = (0.1, 0.25, 0.4, 0.55, 0.7, 0.85)$.

### 3.9 Conclusion

In this chapter we studied the lossy rate-recovery function for Gauss-Markov sources with a quadratic distortion measure. We considered the case where the decoder is interested in reconstructing the source vectors with zero-delay whenever the channel packet is available and not erased by the channel. In other words, the decoder is required to start reconstructing the source vectors immediately after the erasure burst ends. The upper and lower bounds are provided for the rate-recovery function which coincide at high resolutions. The upper bound is based on Q-and-B technique. We separately treated the two channel models: channels with single erasure burst and channels with multiple erasure bursts with guard interval. In addition, motivated by the prospliant coding for discrete sources and lossless recovery presented in Chapter 2, we studied a class of i.i.d. Gaussian sources and sliding window recovery constraints where we established the rate-recovery function.

In Chapter 5 we extend the problem studied in this chapter to the delay-constrained streaming setup rather than zero-delay streaming. In the next chapter, i.e., Chapter 4, we investigate the case where the decoder is allowed to not reconstruct the source sequences for a window of length $W$ after the erasure burst ends. We show that the simple coding scheme based on Q-and-B is in fact sub-optimal for this case. We suggest hybrid coding scheme which outperforms conventional coding schemes.
Chapter 4

Hybrid Coding

4.1 Introduction

In Chapter 3, we studied the lossy rate-recovery function for Gauss-Markov sources and quadratic distortion when $W = 0$, i.e., the destination is required to reconstruct the source vectors within distortion $D$ right after the erasure burst ends. It was shown that, the coding scheme based on the quantization-and-binning technique performs close to optimal and indeed attains the optimal performance at high resolutions. In this chapter we investigate the lossy rate-recovery function for general $W$, where the decoder is not required to reconstruct the source vectors in a window of length $W$ after the erasure burst. We show that the simple memoryless Q-and-B technique is suboptimal. For instance we show that the conventional predictive coding scheme outperforms the Q-and-B scheme for some ranges of source parameters.

In this chapter, we propose a new sequential coding scheme by serially concatenating the successive quantization and random binning schemes. We first show that the conventional coding schemes of predictive coding and memoryless Q-and-B in Chapter 3 are in fact special cases of the proposed coding scheme and, in addition, all attain the optimal performance in the case of an erasure-free channel without erasures.

In the case of a erasure burst channel model, we specialize the general sequential coding scheme to propose a novel hybrid coding scheme. The hybrid coding scheme outperforms the conventional schemes in all range of parameters and attains close-to-optimal performance. Indeed, for some specific range of parameters we are able to prove the optimality of hybrid coding scheme at high resolutions. Several properties of hybrid coding are studied and a lower bound on the lossy rate-recovery function is derived.

The rest of the chapter is organized as follows. Section 4.2 presents the problem setup. In Section 4.3, the general coding scheme is introduced and its connection to special cases is studied. The rate analysis in case of an erasure-free channel is presented in Section 4.3.4. The hybrid coding scheme as well as the
performance analysis of different coding schemes are presented in Section 4.4. The involved treatment of different coding schemes are provided in Section 4.5 and the proof of lower bound on rate-recovery is presented in Section 4.6. Section 4.7 concludes the chapter.

4.2 Problem Statement

At each time $t \geq 1$, a Gaussian source vector $\{s^n_t\}$ is sampled i.i.d. form a zero-mean Gaussian distribution $\mathcal{N}(0, \sigma^2_z)$ along the spatial dimension, and forms a first-order Gauss-Markov chain across the temporal dimension, i.e.,

$$s_t = \rho s_{t-1} + n_t$$

where $\rho \in (0, 1)$ and $n_t \sim \mathcal{N}(0, (1 - \rho^2)\sigma^2_z)$. Without loss of generality we assume $\sigma^2_z = 1$. The sequence $s^n_0$ is sampled i.i.d. from $\mathcal{N}(0, \sigma^2_z)$ and revealed to both the encoder and decoder before the start of the communication. It plays the role of a synchronization frame. We assume that the communication spans the interval $t \in \{0, 1, \ldots, \Upsilon\}$. Without loss of generality, one can consider the source sequences $\{x^n_t\}$, instead of the original source sequences $\{s^n_t\}$, i.e.,

$$x_t = \rho x_{t-1} + n_t.$$  

Note that $x_t \sim (0, 1 - \rho^{2t})$ and the source sequences $\{x^n_t\}$ inherits the first order Markov property from the original source sequences $\{s^n_t\}$, i.e.,

$$x_t = \rho x_{t-1} + n_t.$$  

Throughout the chapter, based on the equivalence of the two models and in order to present the results in their simplest form, we will interchangeably use the two source models.

An encoder computes an index $f_t \in \{1, 2, \ldots, 2^{nR_t}\}$ at time $t$, according to an encoding function

$$f_t = F_i(x^n_t_0, \ldots, x^n_t), \quad 1 \leq t \leq \Upsilon.$$  

An encoder computes an index $f_t \in \{1, 2, \ldots, 2^{nR_t}\}$ at time $t$, according to an encoding function

$$f_t = F_i(x^n_t_0, \ldots, x^n_t), \quad 1 \leq t \leq \Upsilon.$$  

Note that the encoder in (4.4) is a causal function of the source sequences.

The channel takes each $f_t$ as input and either outputs $g_t = f_t$ or an erasure symbol, i.e., $g_t = \star$. We consider the class of erasure burst channels. For some particular $j \geq 1$, it introduces a erasure burst such that

$$g_t = \begin{cases} \star, & t \in \{j, j + 1, \ldots, j + B - 1\} \\ f_t, & \text{otherwise} \end{cases}.$$ 

As illustrated in Fig 4.1, upon observing the sequence $\{g_t\}_{t \geq 1}$, the decoder is required to reconstruct each source sequence with zero delay, i.e.,

$$\hat{x}^n_t = G_i(g_1, g_2, \ldots, g_t), \quad t \notin \{j, \ldots, j + B + W - 1\}$$

where $\hat{x}^n_t$ denotes the reconstruction sequence and $j$ denotes the time at which erasure burst starts.
Figure 4.1: Problem Setup: Consider the example of $B = 2$ and $W = 3$. The encoder output $f_j$ is a function of the source sequences up to time $j$, i.e., $s^n_0, s^n_1, \ldots, s^n_j$. The channel introduces an erasure burst of length $B$. The decoder produces $\hat{s}^n_j$ upon observing the sequence $\{g_0, g_1, \ldots, g_j\}$. The decoder is not required to produce those source sequences that fall in a window of length $B + W$ following the start of an erasure burst. However, the decoder recovers the rest of the source sequences within zero-delay and average distortion constraint.

in (4.5). The destination is not required to produce the source vectors that appear either during the erasure burst or in the period of length $W$ following it. We call this period the error propagation window.

We consider the case where the reconstruction in (4.6) satisfies the average distortion constraint

$$\lim_{n \to \infty} \sup_{t \notin \{j, \ldots, j + B + W - 1\}} \frac{1}{n} \sum_{k=1}^{n}(x_{t,k} - \hat{x}_{t,k})^2 \leq D.$$  \hspace{1cm} (4.7)

For any $t \in \{1, \ldots, \Upsilon\}$ define $P_t^{(n)}$ as the probability of the event that the decoder fails in reproducing $\hat{x}^n_t$ within average distortion $D$ as in (4.7). A tuple $(R_1, R_2, \ldots, R_\Upsilon, D)$ is achievable if there exists a sequence of encoding and decoding functions and a sequence $\epsilon_n$ that approaches zero as $n \to \infty$ such that $P_t^{(n)} \leq \epsilon_n$ for any $t$ outside the error propagation window. Define $R_\Upsilon(D)$ be the closure of the achievable tuples $(R_1, R_2, \ldots, R_\Upsilon, D)$. We define the rate-recovery function as follows.

$$R_\Upsilon(B, W, D) \triangleq \inf_{(R_1, R_2, \ldots, R_\Upsilon, D) \in R_\Upsilon(D)} \left\{ \sup_{k \in \{1, \ldots, \Upsilon\}} R_k \right\}.$$  \hspace{1cm} (4.8)

In particular we are interested in the rate-recovery function in the large $\Upsilon$ asymptotic, i.e.,

$$R(B, W, D) \triangleq \lim_{\Upsilon \to \infty} R_\Upsilon(B, W, D),$$  \hspace{1cm} (4.9)

which we will simply call the rate-recovery function.

### 4.3 Zero-Delay Sequential Coding Scheme for Erasure-free Channels

In this section we consider the zero-delay streaming problem for the case of erasure-free channels ($B = 0$) where the channel perfectly reveals all the encoder outputs to the destination.
4.3.1 General Coding Scheme for DMS

Here we present a general zero-delay sequential coding scheme. We first present the coding scheme for the discrete memoryless sources (DMS) and then discuss the generalization to the Gauss-Markov sources.

**Theorem 4.** Let \( \{x_1, x_2, \ldots, x_T\} \in X_1 \times X_2 \times \cdots \times X_T \) be a \( \Upsilon \)-DMS and \( d(x_i, \hat{x}_i) \) be a distortion measure \( d : X_1 \times \hat{X}_1 \rightarrow [0, \infty) \).

The tuple \((R_1, R_2, \ldots, R_T, D)\) is achievable if for any \( t \in \{1, \ldots, \Upsilon\} \),

\[
R_t \geq I(u_t; [x_1^t, [u_1^t]_1^{-1})
\]

for some pmf

\[
p([x_1^t, [u_1^t]) = p([x_1^t]_1) \prod_{i=1}^{\Upsilon} p(u_i|[x_1^t]_1,[u_1^t]_1^{-1})
\]

and mappings

\[
\psi_t : U_1 \times U_2 \times \cdots \times U_t \rightarrow \hat{X}_t
\]

such that for any \( t \in \{1, \ldots, \Upsilon\} \), \( E \{d(x_t, \psi_t([u_1^t]))\} \leq D \).

**Proof.** The proposed coding scheme is based on quantizing the source sequences via a specialized test channel and randomly binning the quantization codewords.

**Codebook Generation:** Fix a conditional pmf \( \prod_{t=1}^{\Upsilon} p(u_t|[x_1^t]_1,[u_1^t]_1^{-1}) \) and functions \( \psi_t([u_1^t]_1) \) for all \( t \in \{1, \ldots, \Upsilon\} \) such that for any \( t \in \{1, \ldots, \Upsilon\} \), \( E \{d(x_t, \psi_t([u_1^t]))\} \leq D/(1+\epsilon) \). For each time \( t \in \{1, \ldots, \Upsilon\} \), randomly and independently generate \( 2^{nR_t} \) sequences \( u^n_t(l_t) \), \( l_t \in \{1, 2, \ldots, 2^{nR_t}\} \), each according to

\[
\prod_{k=1}^{n} p(u_{t,k}|x_{t,k})_{\tau \in \{1, \ldots, t\}}, [u_{t,k}]_{\tau \in \{1, \ldots, t-1\}}
\]

where we used the notation \([x_{t,k}]_{\tau \in \{1, \ldots, t\}} \triangleq \{x_{1,k}, x_{2,k}, \ldots, x_{t,k}\}\). Partition the set of indices \( l_t \in \{1, \ldots, 2^{nR_t}\} \) into equal-size bins

\[
B_i(m_t) = [(m_t - 1)2^{n(R_t - R)} + 1 : m_t 2^{n(R_t - R)}],
\]

where \( m_t \in \{1, \ldots, 2^{nR_t}\} \). The codebook is revealed to the encoder and decoder.

**Encoding:** At time \( t \), upon observing \([x_1^t, [u_1^t]_1^{-1})\), the encoder finds an index \( l_t \in \{1, \ldots, 2^{nR_t}\} \) such that \((x_1^t, u^n_t(l_t)) \in T_i^{(n)}\). If there is more than one such index \( l_t \), the encoder selects one of them uniformly at random. If there is no such index \( l_t \), the encoder selects an index from \( \{1, \ldots, 2^{nR_t}\} \) uniformly at random. Encoder sends the index \( m_t \) such that \( l_t \in B_i(m_t)\).

**Decoding:** At time \( t \), the decoder finds the unique index \( \hat{l}_t \in B_i(m_t) \) such that

\[
(u^n_t(\hat{l}_1), u^n_t(\hat{l}_2), \ldots, u^n_t(\hat{l}_t)) \in T_i^{(n)}
\]

If there is such a unique index \( \hat{l}_t \), the reconstructions are computed as \( x_{k,t} = \psi_t([u_{t,k}]_{\tau \in \{1, \ldots, t\}}) \) for \( k \in \{1, \ldots, n\} \); otherwise \( \hat{x}_t^n = \psi_{t-1}([u_{t-1,k}]_{\tau \in \{1, \ldots, t-1\}}) \) for \( k \in \{1, \ldots, n\} \).
bin index, and $\hat{L}_t$ be the decoded index. Define the error event as follows.

$$\mathcal{E} \triangleq \left\{ (u^n_t(\hat{L}_1), u^n_2(\hat{L}_2), \ldots, u^n_t(\hat{L}_t), [x^n]_1^t) \notin \mathcal{T}^{(n)}_t \text{ for some } t \in \{1, \ldots, T\} \right\}$$  \hspace{1cm} (4.11)

Also define the error event conditioned on error free recovery up to time $t - 1$.

$$\mathcal{E}_t \triangleq \left\{ (u^n_t(L_1), u^n_2(L_2), \ldots, u^n_t(L_{t-1}), u^n_t(\hat{L}_t), [x^n]_1^t) \notin \mathcal{T}^{(n)}_t \right\}$$  \hspace{1cm} (4.12)

First note that $Pr(\mathcal{E}) \leq \sum_{t=1}^{T} Pr(\mathcal{E}_t)$. Consider the following events.

$$\mathcal{E}_{t,1} = \left\{ (u^n_t(L_1), u^n_2(L_2), \ldots, u^n_t(L_{t-1}), u^n_t(l_t), [x^n]_1^t) \notin \mathcal{T}^{(n)}_t \text{ for all } l_t \in \{1, \ldots, 2^{n\tilde{R}_t}\} \right\}$$  \hspace{1cm} (4.13)

$$\mathcal{E}_{t,2} = \left\{ (u^n_t(L_1), u^n_2(L_2), \ldots, u^n_t(L_{t-1}), u^n_t(\hat{l}_t), [x^n]_1^t) \in \mathcal{T}^{n}_t \text{ for some } \hat{l}_t \in \mathcal{B}_t(M_t), \hat{l}_t \neq L_t \right\}$$  \hspace{1cm} (4.14)

By the union of events bound,

$$P(\mathcal{E}_t) \leq P(\mathcal{E}_{t,1}) + P(\mathcal{E}_{t,1}^c \cap \mathcal{E}_{t,2}).$$  \hspace{1cm} (4.15)

By the covering lemma \[42\] $P(\mathcal{E}_{t,1})$ tends to zero as $n \to \infty$ if

$$\tilde{R}_t \geq I(u_t; [x^n]_1^t, [u^n]_1^{t-1}) + \delta(\epsilon').$$  \hspace{1cm} (4.16)

It is also straightforward by packing lemma to show that $P(\mathcal{E}_{t,1}^c \cap \mathcal{E}_{t,2})$ tends to zero as $n \to \infty$, if

$$\tilde{R}_t - R_t < I(u_t; [u^n]_1^{t-1}) - \delta(\epsilon).$$  \hspace{1cm} (4.17)

Combining (4.16) and (4.17) and eliminating $\tilde{R}_t$ we have shown that $P(\mathcal{E}_t)$ tends to zero as $n \to \infty$ if

$$R_t \geq I(u_t; [x^n]_1^t | [u^n]_1^{t-1}) + \delta(\epsilon) + \delta(\epsilon'),$$  \hspace{1cm} (4.18)

where $\delta(\epsilon) + \delta(\epsilon')$ approaches to zero when $n \to \infty$. Following standard lossy source coding achievability proofs, it can be shown that the asymptotic distortion averaged over the random codebook and encoding is bounded as

$$\limsup_{n \to \infty} E \{ d(x^n_t, \hat{x}^n_t) \} \leq D,$$  \hspace{1cm} (4.19)

if the inequality in (4.18) is satisfied. \hfill $\square$

### 4.3.2 General Coding Scheme for Gauss-Markov Sources

The achievability results of the DMS can be generalized to the Gauss-Markov sources and quadratic distortion which is our primary interest. Here we do not provide separate rigorous proofs for Gaussian sources. Interested readers are encouraged to see \[36\]. In fact deriving achievability results for the Gauss-Markov sources reduces to selecting a suitable test channel between the auxiliary random variables $u_t$ and the sources $x_t$ for $t \in \{1, \ldots, T\}$. The test channel has to satisfy the following constraints:
1. The joint pdf of input and output of the test channel satisfies:

\[ p([x_T^T], [u_T^T]) = p([x_1^T]) \prod_{t=1}^{\Upsilon} p(u_t|x_t^t, [u_t^{t-1}]) \]

2. There exist functions \( \psi_t([u_1^t]) \) for any \( t \in \{1, \ldots, \Upsilon\} \) such that \( E \left\{ (x_t - \psi_t([u_1^t]))^2 \right\} \leq D \).

For the general coding scheme we assume that the source sequences \( \{x_1, x_2, \ldots, x_\Upsilon\} \) are successively quantized into the quantization sequences \( \{u_1, u_2, \ldots, u_\Upsilon\} \) according to the following test channel.

\[
\begin{pmatrix}
1 & 0 & 0 & \ldots & 0 \\
0 & 1 & 0 & \ldots & 0 \\
0 & 0 & 1 & \ldots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & 0 & \ldots & 1
\end{pmatrix}
\begin{pmatrix}
u_1 \\
u_2 \\
u_3 \\
\vdots \\
u_\Upsilon
\end{pmatrix}
= \begin{pmatrix}
x_1 \\
x_2 \\
x_3 \\
\vdots \\
x_\Upsilon
\end{pmatrix}
+ \begin{pmatrix}
z_1 \\
z_2 \\
z_3 \\
\vdots \\
z_\Upsilon
\end{pmatrix},
\]

(4.20)

or equivalently,

\[
A[u_1^\Upsilon] = [x_1^\Upsilon] + [z_1^\Upsilon],
\]

(4.21)

where \( A \) is the \( \Upsilon \times \Upsilon \) lower triangular matrix in (4.20) with diagonal elements equal to 1, and \( [z_1^\Upsilon] \) is an i.i.d. zero-mean vector Gaussian random variable, i.e., \( \mathcal{N}(0, \sigma_e^2 I_\Upsilon) \).

Note that,

- By the definition in (4.20), \( x_t + z_t \) can be written as a linear combination of quantization sequences up to time \( t \), i.e.,

\[
x_t + z_t = u_t + \sum_{k=1}^{t-1} a_{t,k} u_k.
\]

(4.22)

Thus, we have

\[
u_t = x_t - \sum_{k=1}^{t-1} a_{t,k} u_k + z_t
\]

(4.23)

\[
= e_t + z_t.
\]

(4.24)

Note from (4.23) that the test channel in (4.20) satisfies the required constraint on the joint pdf, i.e., \( u_t \) only depends on \( x_t \) and \( [u_1^{t-1}] \). According to (4.24) the encoder quantizes the estimation error \( e_t \), which is some linear estimation of \( x_t \) from \( [u_1^{t-1}] \), into \( u_t \).

- We assume that for any \( t \in \{1, \ldots, \Upsilon\} \) the decoding functions \( \psi_t([u_1^t]) \) to be the linear MMSE estimate of source \( x_t \) from the observations \( [u_1^t] \).

- The lower triangular square matrix \( A \) is invertible and has a lower triangular inverse \( Q = A^{-1} \).
Figure 4.2: Block diagram of the proposed coding scheme: a) Encoder, b) Decoder.

with identity diagonals.

\[
\begin{bmatrix}
    u_1 \\
    u_2 \\
    u_3 \\
    \vdots \\
    u_T \\
\end{bmatrix} =
\begin{bmatrix}
    1 & 0 & 0 & \ldots & 0 \\
    q_{2,1} & 1 & 0 & \ldots & 0 \\
    q_{3,1} & q_{3,2} & 1 & \ldots & 0 \\
    \vdots & \vdots & \vdots & \ddots & \vdots \\
    q_{T,1} & q_{T,2} & q_{T,3} & \ldots & 1 \\
\end{bmatrix}
\begin{bmatrix}
    x_1 \\
    x_2 \\
    x_3 \\
    \vdots \\
    x_T \\
\end{bmatrix} + 
\begin{bmatrix}
    z_1 \\
    z_2 \\
    z_3 \\
    \vdots \\
    z_T \\
\end{bmatrix},
\]

or equivalently,

\[
[u_t]_T = Q([x_t]_T^{\top} + [z_t]_T^{\top}),
\]

where \( Q \) is the lower triangular matrix in (4.25). Note that the test channel (4.25) satisfies the zero-delay constraint at the encoder. In particular,

\[
u_t = \sum_{k=1}^{t} q_{t,k} (x_k + z_k),
\]

Note from (4.27) that \( u_t \) is a linear combination of the sources and the test channel noises up to time \( t \) and is indeed independent of the future source sequences and test channel noises, i.e \( \{x_{t+1}, \ldots, x_T\} \) and \( \{z_{t+1}, \ldots, z_T\} \).

- In this chapter, we only consider the special case where the noise vector is i.i.d. However we can extend the analysis to correlated noise vectors with unequal noise powers over time.

To summarize we separately describe the encoding and decoding schemes. The block diagram of such a test channel is shown in Fig. 4.2.

**Encoding**

The block diagram of the encoder is shown in Fig. 4.2a. The encoding consists of two steps; successive quantization and random binning.

**Step 1 (Successive Quantization):** The encoder first successively quantizes the source sequences \( \{x_1, x_2, \ldots, x_T\} \).

**Step 2 (Random Binning):** The encoder randomly and independently places all the quantization sequence \( u_t \) into \( 2^{nR_t} \) bins and, while observing a particular quantization sequence, sends its bin index.
Decoding

The block diagram of the decoder is shown in Fig. 4.2b. The decoder consists of two steps.

Step 1 (Decoding with Side-Information): The decoder, while receiving the channel outputs, applies the joint typicality decoding to recover the quantization codewords. Note that in this stage, all the previously recovered quantization sequences are used by the decoder as the side information.

Step 2 (MMSE Estimation): Furthermore, in order to reconstruct the source sequences, the decoder applies the minimum mean square estimation (MMSE) over all the recovered quantization sequences, in a sequential manner.

4.3.3 Special Cases

In this section we look at two coding schemes as special cases of the proposed general scheme.

Predictive Coding

In this coding scheme the encoder at each time $t$ first computes the MMSE estimation error of the source sequence $x_t$ from all the previous quantization sequences, i.e.,

$$e_t = x_t - E\{x_t|\{u_1^{t-1}\}\}$$

$$= x_t - \sum_{k=1}^{t-1} a_{t,k} u_k$$  \hspace{1cm} (4.28)

where $a_{t,k}$ are the optimal MMSE coefficients of estimation $x_t$ from $\{u_1, \ldots, u_{t-1}\}$. The decoder then quantizes the estimation error through the following test channel.

$$u_t = e_t + z_t$$

$$= x_t - \sum_{k=1}^{t-1} a_{t,k} u_k + z_t$$  \hspace{1cm} (4.29)

as required in (4.23). In the predictive coding, however, the random binning part can be excluded. The schematic of the encoder of the predictive coding is shown in Fig. 4.3a.

Remark 5. The predictive coding described here is in fact the idealized differential pulse code modulation (DPCM) for the vector-valued sources and large spatial asymptotic, similar to [14, 17]. It should be distinguished from the conventional scalar DPCM systems based on scalar quantization (see [43] and references therein).

Memoryless Q-and-B

The schematic of the encoder of the memoryless Q-and-B coding is shown in Fig. 4.3b. In this scheme, the encoder at each time $t$ quantizes the source sequence $x^n_t$ according to the test channel $x_t + z_t = u_t$ and sends the bin index to which the quantization sequence $u^n_t$ belongs. Note that the encoder is in fact memoryless as the quantization sequence and encoder’s output only depends on the source sequence of
time $t$. It can be readily verified that the memoryless Q-and-B scheme is a special case of the proposed general coding scheme with the choice of $A_{QB} = Q_{QB} = I_T$. In particular there is no feedback loop in the encoder structure of the memoryless Q-and-B scheme.

### 4.3.4 Rate Analysis of General Coding Scheme for Erasure-free channels

The following theorem characterizes the achievable rate-distortion by the proposed general coding scheme.

**Theorem 5.** For the erasure-free channel model, any rate-distortion pair $(R, D)$ is achievable by the general coding scheme if

$$R \geq R_I(D) \triangleq \frac{1}{2} \log \left( 1 - \rho^2 D + \rho^2 \right)$$  

(4.30)

Note that the rate-distortion pair is independent of the choice of $A$ in the test channel (4.20).

**Remark 6.** In the proposed encoding scheme, we serially combine two techniques with originally different rate-reduction gains and take advantage of both. (See Fig. 4.2(a) for explanation.)

- “Successive quantization gain”: Recall that the successive quantization technique, quantizes the error sequences, $e_t$ in (4.24) rather than the original sources $x_t$. The gain comes from the fact that
the error process in (4.24) has smaller variance than the original source process and therefore is easier to compress.

- “Binning gain”: This rate-reduction gain comes from exploitation of the remaining temporal correlation among the outputs of the successive quantizer.

There exists a trade-off between the two rate-reduction gains, i.e., the higher successive quantization gain in the first step, the lower binning gain is attainable by the second step. Note that the successive quantization gain is maximized if the prediction coefficient are selected using the MMSE filter. This causes \( \{e_t\} \) to be independent of past and hence the binning gain vanishes. In contrast selecting \( A = I \) will result in the maximum gain from binning. In the case of erasure-free channel, the successive quantization gain and the binning gain perfectly compensate for any choice of \( A \). This results in the overall rate independent of the choice of \( A \). We illustrate this in Fig. 4.4. It is known that the predictive coding attains the optimal rate-distortion over the erasure-free channels \([14, 17]\). Therefore, the optimality results of the predictive coding for erasure-free channels, also holds for the general coding scheme with any choice of \( A \).

**Remark 7.** The explained results of the erasure-free channel case does not hold for the case of erasure burst channels. In this situation, the performance of the general coding scheme depends on the choice of \( A \). In the sequel we will show how a judicious choice of \( A \) results in a hybrid scheme that is more efficient in the rate-recovery function.

**Remark 8.** The results of Theorem 5 is closely related to the classical Wyner-Ziv problem of Gaussian source coding with decoder side-information \([25]\). It is shown in \([25]\) that for jointly Gaussian sources, the same rate-distortion is achievable for the two problems in Fig. 4.5, regardless of the fact that the side-information is or is not available at the encoder. The two systems of Fig. 4.5(a) and Fig. 4.5(b) resembles the extreme cases of predictive coding and memoryless Q-and-B in our streaming setup.

**Proof.** According to Theorem 4 specialized for the test channel in (4.20), the rate-distortion pair \((R, D)\) is achievable if it satisfies the following for all \( t \in \{1, \ldots, T\} \).

\[
R \geq R_t = I(u_t; [x]_1^t | [u]_1^{t-1}) \tag{4.31}
\]

\[
\text{Var}(x_t | [u]_1^t) \leq D \tag{4.32}
\]

Note that

\[
R_t = h(u_t | [u]_1^{t-1}) - \frac{1}{2} \log(2\pi e\sigma_z^2) \tag{4.33}
\]
where (4.35) follows from the fact that \( Q_t \) is invertible, and (4.36) follows from the fact that \( q_{t,t} = 1 \). Note that (4.36) is independent of the choice of \( Q \).

Furthermore, the decoder at each time \( t \) computes \( \hat{x}_t \), i.e., the MMSE estimation of \( x_t \) from all the available codewords \( \{u_1, \ldots, u_k\} \), as the reproduction of the source \( x_t \). The distortion at time \( t \), denoted as \( D_t \), is the MMSE estimation error, given by

\[
D_t = \text{Var}(x_t | u_1^t) = \text{Var}(x_t | Q_t (|x_t^1| + |z_t^1|)) = \text{Var}(x_t | |x_t^1| + |z_t^1|) \triangleq D_t(t, \sigma_z^2),
\]

where (4.37) follow from the fact that \( Q_t \), the square matrix consisting of first \( t \) rows and columns of \( Q \), is invertible. Again, it can be observed that (4.38) is independent of the choice of \( Q \).

First note that \( R_t(t, \sigma_z^2) \) and \( D_t(t, \sigma_z^2) \) are increasing functions with respect to \( t \), i.e.,

\[
R_t(t, \sigma_z^2) \triangleq h \left( x_t + z_t \big| |x_t^1| + |z_t^1| \right) - \frac{1}{2} \log (2\pi e \sigma_z^2) \geq h \left( x_t + z_t, |x_t| + |z_t| \right) - \frac{1}{2} \log (2\pi e \sigma_z^2) \triangleq R_t(t, \sigma_z^2),
\]

where (4.39) follows from the fact that conditioning reduces the differential entropy, and (4.40) follows from the stationarity of the sources. The monotonicity of \( D_t(t, \sigma_z^2) \) can be similarly shown. Thus, for any the test channel noise \( \sigma_z^2 \) such that

\[
\lim_{t \to \infty} D_t(t, \sigma_z^2) = D,
\]

the following rate is achievable:

\[
R_t(D) \triangleq \lim_{t \to \infty} R_t(t, \sigma_z^2) = \frac{1}{2} \log \left( \rho^2 + \frac{1 - \rho^2}{D} \right),
\]

as required in (4.30). Note that (4.42) follows from the fact that \( x_t = \rho x_{t-1} + n_t \). This completes the proof.
4.4 Rate Analysis for erasure burst Channels

As stated earlier, the rate-reduction gain associated with the successive quantization and the binning techniques, results in the same overall gain in the ideal channel model. This makes the overall performance of the streaming system in the lossless channel model to be independent of the test channel design, i.e., the choice of \( \mathbf{A} \). This is not true for lossy channels. In this section we consider the erasure burst channel model where the channel introduces an erasure burst in a unknown location during the transmission period. We first analyze the performance of the special schemes of predictive coding and memoryless Q-and-B. We also study the schemes based on source-channel separation and group-of-picture (GOP) idea of practical video coding. Then we propose the hybrid coding scheme, which outperforms all the schemes.

4.4.1 General Coding Scheme for DMS

Let \( \{x_1, x_2, \ldots, x_T\} \in X_1 \times X_2 \times \ldots \times X_T \) be a T-DMS and \( d(x_t, \hat{x}_t) \) be a distortion measure as the following mapping.

\[
d : X_t \times \hat{X}_t \to [0, \infty).
\]

The erasure burst channel which introduces a single erasure burst of length \( B \) in an unknown location during the transmission period, can be characterized by a single variable \( \tau \) which indicates the time that the erasure starts. Fix the joint distribution \( p([x]_T, [u]_T) \) such that

1. It satisfies the following marginal constraints.

\[
p([x]_T, [u]_T) = p([x]_T) \prod_{t=1}^{T} p(u_t|[x]_1^t, [u]_1^{t-1}),
\]

i.e., the random variable \( u_t \) only depends on \( \{[x]_1^t, [u]_1^{t-1}\} \). This is in fact the zero-delay encoding constraint.

2. The decoder is not interested in reconstructing the sources during the erasure period and a window of length \( W \) after the erasure ends. Thus, for any parameter \( \tau \in \{1, \ldots, T - B + 1\} \), there exists \( T - B - W \) functions as follows.

\[
\hat{x}_{\tau,t} \triangleq \begin{cases} 
\psi_{\tau,t}([u]_1^t) & \text{for } t < \tau \\
\psi_{\tau,t}([u]_{1-1}^t[u]_{1+1}^t) & \text{for } t \geq \tau + B + W,
\end{cases}
\]

such that for any \( \tau \in \{1, \ldots, T - B + 1\} \) and \( t \notin \{\tau, \ldots, \tau + B + W - 1\} \),

\[
E \{d(x_t, \hat{x}_{\tau,t})\} \leq D.
\]

For any \( \tau \in \{1, \ldots, T - B + 1\} \) and \( t \notin \{\tau, \ldots, \tau + B + W - 1\} \), define the following functions.

\[
R_{\tau,t}(B, W, D) \triangleq \]

\[
\]
\[
\begin{cases}
H(u_t[x]^{-1}) - H(u_t[x], [u]_{t-1}) \\
\max_{M \subseteq \{t-W, \ldots, t\}} \left\{ \frac{1}{|M|} \sum_{k \in M} H(u_k[x], [u]_{k-1}) - \sum_{k \in M} H(u_k[x], [u]_{k-1}) \right\} \\
H(u_t[x]^{-1}u_{t+B}) - H(u_t[x], [u]_{t-1})
\end{cases}
\]

for \( t < \tau \)
\[
H(u_t[x]^{-1}) - H(u_t[x], [u]_{t-1})
\]

for \( t = \tau + B + W \)
\[
H(u_t[x]^{-1}u_{t+B}) - H(u_t[x], [u]_{t-1})
\]

for \( t > \tau + B + W \)

(4.47)

**Theorem 6.** Any rate \( R \geq R^+(B, W, D) \) is achievable where

\[
R^+(B, W, D) \triangleq \sup_{\tau \in \{1, \ldots, \Upsilon - B + 1\}} R_{\tau, t}(B, W, D).
\]

The proof of Theorem 6 follows similar to Theorem 4. In particular, the first and third case of (4.47) indicate the rate requirement for recovering the unique \( u_t^n \) before and after the erasure burst. The second case of (4.47) indicates the rate requirement for simultaneous recovery of the unique sequences \( \{u_{\tau+B}, u_{\tau+B+1}, \ldots, u_{\tau+B+W}\} \) at time \( \tau + B + W \) right after the error propagation window, which follows from Burger-Tung achievability results [44]. Similar to the erasure-free channel case, in the following we apply the results of the DMS in Theorem 6 to the Gaussian sources by designing the proper test channel.

### 4.4.2 Conventional Coding Schemes for Gauss-Markov Sources

In this section we study the behaviour of some conventional coding scheme over erasure burst channels.

**Predictive Coding**

We start with the predictive coding scheme. The following theorem characterizes the achievable rate of the predictive coding.

**Theorem 7.** The predictive coding scheme achieves any rate \( R \geq R_{PC}^+(B, W, D) \) for \( D \geq \rho^{2(W+1)}(1 - \rho^{2B}) \) where,

\[
R_{PC}^+(B, W, D) \triangleq \frac{1}{2} \log \left( \frac{1 - \rho^{2(W+1)}(1 - \rho^{2B}) - (1 - D)\rho^2}{D - \rho^{2(W+1)}(1 - \rho^{2B})} \right).
\]

The proof of Theorem 7 is presented in Sec. 4.5.1.

**Remark 9.** The predictive coding requires the distortion to be greater than \( \rho^{2(W+1)}(1 - \rho^{2B}) \). In particular, this amount of distortion is caused by the erasure times where the decoder is not able to recover the quantization sequences. According to this, the predictive coding fails in high resolution when \( D \to 0 \).

**Memoryless Quatize-and-Binning**

The following theorem characterizes the achievable rate by the memoryless Q-and-B scheme.

**Theorem 8.** The memoryless Q-and-B coding scheme achieves any rate \( R \geq R_{QB}^+(B, W, D) \) where

\[
R_{QB}^+(B, W, D) \triangleq \frac{1}{W+1} h([u]^{t+W}|\tilde{x}_{t-B}) - \frac{1}{2} \log (2\pi e\sigma^2) \cdot
\]

(4.50)
where for any $u_i \triangleq s_i + z_i$ and $z_i$ is sampled i.i.d. from $\mathcal{N}(0, \sigma_z^2)$. Also $\tilde{s}_{t-B} \triangleq s_{t-B} + e$ and $e \sim \mathcal{N}(0, \Sigma(\sigma_z^2)/(1 - \Sigma(\sigma_z^2)))$, with

$$\Sigma(\sigma_z^2) \triangleq \frac{1}{2} \sqrt{(1 - \sigma_z^2)^2(1 - \rho^2)^2 + 4\sigma_z^2(1 - \rho^2) + \frac{1 - \rho^2}{2}(1 - \sigma_z^2)}. \quad (4.51)$$

The test channel noise $\sigma_z^2 > 0$ is chosen to satisfy

$$\text{Var}(s_{t+W} | u_t \| u_t + W, \tilde{s}_{t-B}) \leq D. \quad (4.52)$$

The proof of Theorem 8 is presented in Sec. 4.5.2. To better compare the performance of different schemes, we define the excess rate as follows.

**Definition 3.** The excess rate $R_E$ is the rate $R$ subtracted by the erasure-free channel rate in (4.30), i.e.,

$$R_E \triangleq R - R_{I}(D). \quad (4.53)$$

The following corollary characterizes the high resolution performance of the memoryless Q-and-B.

**Corollary 4.** In high resolution regime when $D \to 0$, the achievable excess rate by the memoryless Q-and-B scheme satisfies.

$$\lim_{D \to 0} R_{E, Q_B}(B,W,D) = \lim_{D \to 0} \left( R_{Q_B}(B,W,D) - \frac{1}{2} \log \left( \frac{1 - \rho^2}{D} \right) \right) = \frac{1}{2(W+1)} \log \left( \frac{1 - \rho^2(B+1)}{1 - \rho^2} \right). \quad (4.54)$$

The proof of Corollary 4 is presented in Sec. 4.5.3. The proof is based on the observation that, in high resolution regime, the quantization source sequence $u_t$ is very close to the original source sequence $x_t$. Therefore, the Markov chain property among the original source sequences also approximately hold for quantization sequences.

**Separation Based Scheme**

Many practical scheme are design based on separation of source and channel coding. Motivated by this we consider a coding scheme consisting of predictive coding followed by forward-error-correcting (FEC) codes to protect the channel packets from channel erasures. The following theorem characterizes the achievable rate.

**Theorem 9.** The separation based coding scheme achieves any rate $R \geq R_{PC-FEC}^+(B,W,D)$ where

$$R_{PC-FEC}^+(B,W,D) \triangleq \frac{B + W + 1}{2(W+1)} \log \left( \frac{1 - \rho^2}{D} + \rho^2 \right). \quad (4.55)$$

The proof is very simple and is omitted here. In particular, the source vectors are first encoded by the predictive coding scheme which results in generation of $nR_I$ bits per source vector. A rate-$(B + W + 1)/(W+1)$ FEC code is then applied over the predictive encoder’s outputs which enables the recovery of all $B + W + 1$ codewords (including $B$ erased codewords and $W + 1$ after the erasure
The following theorem characterizes the achievable rate. It is not hard to observe that the excess rate of separation based coding scheme in high resolution scheme is

\[
R_{E,\text{PC-FEC}}^+(B, W, D) = \lim_{D \to 0} \left( R_{\text{PC-FEC}}^+(B, W, D) - \frac{1}{2} \log \left( \frac{1 - \rho^2}{D} \right) \right) = \lim_{D \to 0} \frac{B}{2(W + 1)} \log \left( \frac{1 - \rho^2}{D} \right),
\]

which grows to infinity.

**GOP-Based Coding**

In video compression applications, in order to limit the unavoidable error propagation effect of the predictive coding scheme, the group of picture (GOP) structure is considered. A zero-delay GOP, in its simplest form, contains the following picture types:

- **I-frame** (intra coded frame)- a picture that is coded independently of all other pictures. Each GOP begins (in decoding order) with this type of picture.
- **P-frame** (predictive coded frame) contains motion-compensated difference information relative to previously decoded pictures.

In order to control the error-propagation, the I-frames are transmitted periodically and the P-frames are transmitted in between. When the decoder fails in recovery of any frame during a GOP, the rest of the frames of that GOP are not recovered. However the decoder gets back to recovery of the frames from the time of next I-frame. The following theorem characterizes the achievable rate of the GOP-based scheme for our problem of interest.

**Theorem 10.** The GOP-based coding scheme achieves the average rate \( \bar{R}_{\text{GOP}}(W, D) \) for any \( B \geq 0 \) where

\[
\bar{R}_{\text{GOP}}(W, D) = \frac{1}{2} \log \left( \frac{1 - \rho^2}{D} \right) + \frac{1}{2(W + 1)} \log \left( \frac{1}{1 - (1 - D)\rho^2} \right)
\]

(4.57)

**Remark 10.** The GOP-based coding scheme is a time-variant scheme and the rate associated with the I-frames and P-frames are not the same. In this chapter we compare the average rate of the GOP-based scheme rather than its peak rate with other schemes.

The proof of Theorem 10 is presented in Sec. 4.5.4. It can be readily observed that in high resolution regime when \( D \to 0 \), the excess rate for the GOP-based scheme scales as follows.

\[
\lim_{D \to 0} \bar{R}_{E,\text{GOP}}(W, D) = \frac{1}{2(W + 1)} \log \left( \frac{1}{1 - \rho^2} \right).
\]

(4.58)

**Numerical Comparison**

Fig. 4.6 shows the excess rates as a function of correlation between the source sequences \( \rho \), based on predictive coding, memoryless Q-and-B scheme, separation based scheme and GOP-based scheme. The hybrid coding scheme will be explained in the sequel. It can be observed from the figure that the
memoryless Q-and-B outperforms the GOP based scheme. We conjecture that this is true for any parameter set. It can be also observed from the figure that predictive coding outperforms the Q-and-B based scheme for small values of $\rho$. In the next section, we take a closer look at the two schemes of predictive coding and memoryless Q-and-B coding.

### 4.4.3 Hybrid Coding Scheme

In this section we introduce a class of coding schemes based on the discussed notations which are more robust in case of erasure burst channel model.

**Illustrative Example**

Before introducing the hybrid coding scheme in its general form, it is more insightful to consider the coding scheme for the special case of $B = W = 1$ in high resolution regime. The channel may erase a single channel packet in an unknown location, but losslessly reveals the rest of the packets to the decoder. The decoder is expected to generate the high resolution reproduction of the source sequences with zero delay, except the source sequences associated with the erased packet and one source sequence after the erasure. We focus on the performance of the general coding scheme presented in Sec. 4.3.2. First note that because of the high resolution assumption, the test channel noise $\sigma^2_z$ in (4.20) approaches to zero. In particular we choose $\sigma^2_z = D$, where $D \to 0$.

Consider the case where the erasure happens at time $t - 2$. Up to the time $t - 3$, the decoder reconstructs the source sequences within high resolution and zero delay. At time $t - 1$ the decoder keeps the channel packet and waits for time $t$, when it attempts to simultaneously recover the source sequences $\{u_{t-1}, u_t\}$ from the channel packets of time $t - 1$ and $t$. This succeeds with high probability if the rates satisfy.

$$R_t \geq h(u_t \mid u(t-3)_1, u_{t-1}) - \frac{1}{2} \log (2\pi eD) \quad (4.59)$$

$$R_{t-1} \geq h(u_{t-1} \mid u(t-3)_1, u_t) - \frac{1}{2} \log (2\pi eD) \quad (4.60)$$

$$R_{t-1} + R_t \geq R_{sum} \triangleq h(u_{t-1}, u_t \mid u(t-3)_1) - \log (2\pi eD) \quad (4.61)$$

For sake of illustration, we only focus on the sum-rate constraint in (4.61). It is not hard to show that

$$\lim_{D \to 0} h(u_{t-1}, u_t \mid u(t-3)_1) = \lim_{D \to 0} h(u_{t-1}, u_t \mid x(t-3)_1) = \lim_{D \to 0} h(u_{t-1}, u_t \mid n(t-3)_1). \quad (4.62)$$

Thus,

$$\lim_{D \to 0} R_{sum} = \lim_{D \to 0} \left\{ h(u_{t-1}, u_t \mid n(t-3)_1) - \log (2\pi eD) \right\}. \quad (4.64)$$

Fig 4.7(a) schematically shows the encoder and decoder of the predictive coding scheme. Recall that the predictive coding only sends the innovation part of the source at any time $t$, i.e., $n_t$. A $n_{t-2}$ associated with the erasure time is not available at the decoder, the source $s_t = n_t + \rho n_{t-1} + \rho^2 n_{t-2}$ is not recoverable at time $t$. Thus the predictive coding fails in this scenario.
Figure 4.6: Excess Rates based on Hybrid Coding in comparison with Predictive Coding, Memoryless Q-and-B and the lower bound for $D = 0.3$. The excess rate is the difference between the achievable rate by the that coding scheme and the achievable rate of the predictive coding in case of erasure-free channel.
Chapter 4. Hybrid Coding

82

\[ u_{t-1}, u_t \]

\[ u_{t-1}, u_t \]

\[ u_{t-1}, u_t \]

\[ u_{t-1}, u_t \]

\[ \hat{s}_t = u_t \]

\[ \hat{s}_t = u_t + wu_{t-1} \]

Figure 4.7: Schematic of encoder and decoder for \( B = W = 1 \) and high resolution. The \( f_{t-2} \) is erased and \( \hat{s}_t \) is required to be reconstructed at time \( t \). (a) The predictive coding which fails in reconstructing \( s_t^0 \) (b) The memoryless Q-and-B scheme, (c) The simplified scheme, and (d) The generalization of the idea.
contains all the random variables $n_i$ where

$$u_t = x_t + z_t = n_t + \rho n_{t-1} + \rho^2 n_{t-2} + \cdots + \rho^{t-1} n_1 + z_t,$$  

(4.65)

where $n_t \sim \mathcal{N}(0, 1 - \rho^2)$. Notice that according to (4.65), in this coding scheme random variable $u_t$ contains all the random variables $n_i$ for $i \leq t$. According to Corollary 4 for $B = W = 1$, this scheme as $D \to 0$ achieves the following sum-rate.

$$\lim_{D \to 0} \left\{ R_{\text{sum}} - \log \left( \frac{1 - \rho^2}{D} \right) \right\} = \frac{1}{2} \log (1 + \rho^2)$$  

(4.66)

Note that in this coding scheme, $u_{t-1}$ and $u_t$ are the high resolution version of the source sequences $x_{t-1}$ and $x_t$, respectively. Thus, the decoder at time $t$, having access to $\{u_{t-1}, u_t\}$, is able to generate the high resolution reproduction of not only the required source sequence $x_t$, but also the extra sequence $x_{t-1}$ whose recovery deadline has already passed. A natural question arises: either the decoder is always restricted to additionally reconstruct $x_{t-1}$ or it may relax the rate requirement by not recovering $x_{t-1}$. Interestingly, the latter is correct. Before introducing the hybrid coding which is motivated by this observation, let us consider a simplified version of the scheme that uses the test channel at time $t$. Fig 4.7(c) illustrates this test channel.

$$u_t = n_t + \rho^2 n_{t-2} + \rho^4 n_{t-4} + \cdots + \rho^{t-1} n_1 + z_t - \rho z_{t-1} + \rho^2 z_{t-2} + \cdots + \rho^{t-1} z_1, \text{ for odd } t.$$  

(4.67)

$$u_t = n_t + \rho^2 n_{t-2} + \rho^4 n_{t-4} + \cdots + \rho^{t-2} n_2 + z_t - \rho z_{t-1} + \rho^2 z_{t-2} + \cdots - \rho^{t-1} z_1, \text{ for even } t.$$  

(4.68)

Unlike the memoryless Q-and-B in (4.65), the random variable $u_t$ in this coding scheme consists of the every other term of the past $n_i$, i.e., $i \in \{t, t-2, t-4, \ldots\}$. First note that this particular test channel example, belongs to the class of general coding scheme described in Section 4.3.2. In particular $u_t$ can be written, for any $t$, as

$$u_t = x_t - \rho x_{t-1} + \rho^2 x_{t-2} + \cdots + \rho^{t-1} x_1 + z_t - \rho z_{t-1} + \rho^2 z_{t-2} + \cdots + \rho^{t-1} z_1.$$  

(4.69)

Also note from (4.67) and (4.68) that the high resolution reproduction of the source sequence at any time can be computed using the quantization sequence of that time and the previous time, i.e.,

$$\hat{x}_t = x_t + z_t = u_t + \rho u_{t-1}.$$  

(4.70)

Therefore, if the channel packet $t-2$ is erased, the decoder recovers $\{u_{t-1}, u_t\}$ at time $t$ and reconstructs the source $x_t$ within high resolution from $\{u_{t-1}, u_t\}$ by applying (4.70). The sum-rate in (4.64) for this
coding scheme when \( D \to 0 \), can be computed as

\[
\lim_{D \to 0} R_{\text{sum}} = \lim_{D \to 0} \{ h(u_{t-1}, u_t | [n]_1^{t-3}) - \log (2\pi eD) \}
\]

\[
= \lim_{D \to 0} \left\{ h \left( \left( n_{t-1} + \rho^2 n_{t-3} + \cdots + z_{t-1} - \rho z_{t-2} + \rho^2 z_{t-3} - \cdots \right) | [n]_1^{t-3} \right) - \log (2\pi eD) \right\}
\]

\[
= \lim_{D \to 0} \left\{ h(n_{t-1}) + h(n_t + \rho^2 n_{t-2} | n_{t-1}) - \log (2\pi eD) \right\}
\]

\[
= \lim_{D \to 0} \left\{ \frac{1}{2} \log (2\pi e(1 - \rho^2)) + \frac{1}{2} \log (2\pi e(1 + \rho^4)(1 - \rho^2)) - \log (2\pi eD) \right\}. \quad (4.71)
\]

It can be easily shown from (4.71) that for this scheme we have

\[
\lim_{D \to 0} \left\{ R_{\text{sum}} - \log \left( \frac{1 - \rho^2}{D} \right) \right\} = \frac{1}{2} \log (1 + \rho^4). \quad (4.72)
\]

By comparing (4.66) and (4.72), it can be observed that the sum-rate constraint of the second scheme is lower in comparison with the memoryless Q-and-B.

Note that in the mentioned simplified scheme, unlike the memoryless Q-and-B, the decoder does not reconstruct the source sequence \( x_{t-1} \) even at time \( t \), as \( u_{t-1} \) is not the high resolution version of \( x_{t-1} \) anymore. However the random variable \( u_{t-1} \) contains a specific part of the source \( x_{t-1} \) with the following properties:

- When \( u_{t-2} \) is available, \( u_{t-1} \) is used to reconstruct \( x_{t-1} \) at time \( t - 1 \),
- When \( u_{t-2} \) is not available, \( u_{t-1} \) helps \( u_t \) to reconstruct the source \( x_t \) at time \( t \) according to (4.70).

As illustrated in Fig 4.7(d), by exploiting these observations, we can further generalize the simplified scheme with the test channel in (4.67) and (4.68) to a class of test channels parameterized by \( w \), such that

\[
\hat{x}_t = x_t + z_t = u_t + w u_{t-1}. \quad (4.73)
\]

This is equivalent to defining the test channel at time \( t \), as follows.

\[
u_t = x_t - w x_{t-1} + w^2 x_{t-2} + \cdots + w^{t-1} x_1 + z_t - w z_{t-1} + w^2 z_{t-2} + \cdots + w^{t-1} z_1. \quad (4.74)
\]

Note that \( w = 0 \) reduces to memoryless Q-and-B and \( w = \rho \) reduces to the simplified scheme. In suffices in general to restrict \( w \in [0, \rho] \).

From (4.74) we can write,

\[
u_t = n_t + \rho n_{t-1} + \rho^2 n_{t-2} + \cdots + z_t - w n_{t-1} - w \rho n_{t-2} - w^2 n_{t-3} - \cdots - w z_{t-1} + w^2 n_{t-2} + w^2 \rho n_{t-3} + \cdots + w^2 z_{t-2} = n_t + (\rho - w) n_{t-1} + (\rho^2 - w \rho + w^2) n_{t-2} + (\rho^3 - w \rho^2 + w^2 \rho - w^3) n_{t-3} + \cdots \quad (4.75)
\]
\[ z_t - w z_{t-1} + w^2 z_{t-2} - w^3 z_{t-3} + \cdots. \] (4.76)

It can be shown that the choice of \( w^* = \rho/(1+\rho^2) \), minimizes the sum-rate constraint in (4.64) as follows.

\[
\lim_{D \to 0} R \sum = \lim_{D \to 0} \left\{ h \left( \begin{pmatrix} 1 & \rho - w^* & \rho^2 - w^* \rho + w^* \rho^2 \vspace{0.5em} 0 \end{pmatrix} \right) \left( \begin{pmatrix} n_t \vspace{0.5em} n_{t-1} \end{pmatrix} \right) - \log (2\pi e D) \right\}
= \lim_{D \to 0} \left\{ \log \left( \frac{1-\rho^2}{D} \right) \right\}
+ \frac{1}{2} \log \left( \det \left( \begin{pmatrix} 1 & \rho - w^* & \rho^2 - w^* \rho + w^* \rho^2 \vspace{0.5em} 0 \end{pmatrix} \right) \right) \vspace{0.5em} (4.77)
\]

And from (4.77) we have,

\[
\lim_{D \to 0} \left\{ R \sum - \log \left( \frac{1-\rho^2}{D} \right) \right\} = \frac{1}{2} \log \left( 1 + \frac{\rho^4}{1+\rho^2} \right) \vspace{0.5em} (4.78)
\]

Note that for any \( \rho \in [0,1] \)

\[
\frac{1}{2} \log \left( 1 + \frac{\rho^4}{1+\rho^2} \right) \leq \frac{1}{2} \log (1+\rho^4) \leq \frac{1}{2} \log (1+\rho^2) \vspace{0.5em} (4.79)
\]

In particular, as we will see Corollary 5 in the sequel, the sum-rate in (4.78) is the best achievable sum-rate for this setup.

While the above analysis only considers the sum-rate, a similar computation for the marginal rate constraints in (4.59) and (4.60) also applies. In particular, we show in Theorem 12 that for \( \rho \geq \rho^* = 0.922 \), the sum-rate constraint dominates the marginal constraints of (4.59) and (4.60), and thus the proposed hybrid scheme is optimal. Fig. 4.8(a) also illustrates that the hybrid coding performs very close to optimal for all ranges of \( \rho \) at least for the case \( B = 1 \).

In what follows we show that the hybrid coding scheme can be generalized to any parameters \( B \) and \( W \) and yields rate-gains over conventional test channels, such as memoryless Q-and-B.

**General Case**

Now we define the hybrid coding scheme in general form. Consider the vector \( \mathbf{w} \triangleq \{w_1, w_2, \ldots, w_W\} \) of size \( W < T - 1 \), such that \( 0 \leq w_k \leq \rho^k \). Define the \( T \times T \) Toeplitz lower triangular matrix \( \mathbf{A}_H(\mathbf{w}) \) parametrized by the vector \( \mathbf{w} \) with the following \((i,j)\)-th elements.

\[
\{\mathbf{A}_H(\mathbf{w})\}_{i,j} \triangleq \begin{cases} w_{i-j} & \text{if } 1 \leq i - j \leq W \\ 1 & \text{if } i = j \\ 0 & \text{otherwise.} \end{cases} \vspace{0.5em} (4.80)
\]

We define the class of coding scheme where the test channel in (4.20) is defined by \( \mathbf{A} = \mathbf{A}_H(\mathbf{w}) \), as the *hybrid coding scheme*. 
• From the definition of this test channel, for this family of coding schemes, the noisy version of the source \( x_t \), i.e., \( x_t + z_t \) become a linear combination of the last \( W + 1 \) quantization sequences, i.e.,

\[
x_t + z_t = u_t + \sum_{k=1}^{W} w_k u_{t-k}.
\]  

(4.81)

• The matrix \( Q_H(w) \triangleq A_H^{-1}(w) \) is a \( T \times T \) Toeplitz lower triangular matrix. Let \( v_k \) denotes the elements of \( k \)-th diagonal of the inverse matrix \( Q_H(w) \), i.e.,

\[
Q_H(w) = \begin{pmatrix}
1 & 0 & 0 & \cdots \\
v_1 & 1 & 0 & \cdots \\
v_2 & v_1 & 1 & \cdots \\
v_3 & v_2 & v_1 & \cdots \\
\vdots & \vdots & \vdots & \ddots \\
0 & 0 & 0 & \cdots & 1
\end{pmatrix}.
\]  

(4.82)

The element \( v_k \) can be recursively computed as follows.

\[
v_k = - \sum_{j=0}^{k-1} w_{k-j} v_j,
\]  

(4.83)

where \( w_k = 0 \) for \( k > W \) and \( w_0 = 1 \).

• Some examples of \( A_H(w) \) and \( Q_H(w) \) are shown in (4.84), (4.85) and (4.86).

\[
A_H(\phi) = A_{QB} = I_T \Rightarrow Q_H(\phi) = Q_{QB} = I_T,
\]  

(4.84)
i.e., the hybrid coding scheme with \( W = 0 \) is the memoryless Q-and-B scheme.
Define the set \( \mathcal{L} \triangleq \{B + 1, \ldots, B + W + 1\} \). For any choice of \( \mathbf{w} \triangleq \{w_1, w_2, \ldots, w_W\} \), the rate

\[
R \geq R_H(\sigma_z^2, \mathbf{w}) \triangleq \max_{\mathcal{M} \subseteq \mathcal{L}} \mathbb{E} \left\{ \frac{1}{|\mathcal{M}|} h(\hat{\mathbf{u}}_{|\mathcal{M}|} | \hat{\mathbf{u}}_{\mathcal{M}}, \hat{s}_1) \right\} - \frac{1}{2} \log(2\pi e \sigma_z^2) \tag{4.87}
\]

is achievable by the hybrid coding scheme for the test channel

\[
\begin{pmatrix}
\tilde{u}_{B+1} \\
\tilde{u}_{B+2} \\
\vdots \\
\tilde{u}_{B+W+1}
\end{pmatrix} = \mathbf{Q}_{\text{eff}} \begin{pmatrix}
\mathbf{s}_1 \\
\mathbf{s}_2 \\
\vdots \\
\mathbf{s}_{B+W+1}
\end{pmatrix} + \begin{pmatrix}
\mathbf{z}_1 \\
\mathbf{z}_2 \\
\vdots \\
\mathbf{z}_{B+W+1}
\end{pmatrix} \tag{4.88}
\]

where \( \mathbf{z}_1, \ldots, \mathbf{z}_{B+W+1} \) are drawn i.i.d. according to \( \mathcal{N}(0, \sigma_z^2) \), \( \mathbf{Q}_{\text{eff}} \) is the \((W + 1) \times (B + W + 1)\) matrix consisting of the rows \( \{B + 1, \ldots, B + W + 1\} \) and the columns \( \{1, \ldots, B + W + 1\} \) of the matrix \( \mathbf{Q}_H(\mathbf{w}) \) in (4.82), i.e.,

\[
\mathbf{Q}_{\text{eff}} \triangleq \begin{pmatrix}
v_B & v_{B+1} & \cdots & 1 & 0 & \cdots & 0 & 0 \\
v_{B+1} & v_{B+2} & \cdots & v_1 & 1 & \cdots & 0 & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots \\
v_{B+W} & v_{B+W-1} & \cdots & v_W & v_{W-1} & \cdots & v_1 & 1
\end{pmatrix}, \tag{4.89}
\]

Also \( \hat{s}_1 \triangleq s_1 + \mathbf{e} \) and \( \mathbf{e} \sim \mathcal{N}(0, \Sigma(\sigma_z^2)/(1 - \Sigma(\sigma_z^2))) \), with \( \Sigma(\sigma_z^2) \) defined in (4.51). The test channel noise \( \sigma_z^2 \) has to satisfy

\[
D \geq \Sigma_H(\sigma_z^2, \mathbf{w}) \triangleq \text{Var}(\tilde{s}_{B+W+1} | \tilde{u}_{B+1}, \hat{s}_1). \tag{4.90}
\]

Therefore, the rate

\[
R_H^+(B, W, D) = \min_{\mathbf{w}, \sigma_z^2 \leq D} R_H(\sigma_z^2, \mathbf{w}) \tag{4.91}
\]

is achievable by the hybrid coding scheme.

**Remark 11.** The rate expression in (4.87) is in fact the rate constraint of recovery after the error propagation window and is equivalent to the second rate expression in (4.47). Unlike the memoryless Q-and-B where the sum-rate constraint is always dominant, the maximizing constraint depends on the specific test channel.
The proof of Theorem 11 is provided in Section 4.5.5. Fig. 4.6 shows the achievable excess rates as a function of correlation between the source sequences $\rho$. It can be observed that the hybrid coding scheme always outperforms the other schemes.

**Hybrid Coding in the High Resolution Regime**

In this section we provide some results on the high resolution performance of the hybrid coding scheme.

**Theorem 12.** In the high resolution regime, for $B = 1$ and any $W$, the sum-rate constraint of the hybrid coding scheme is minimized by the following choice of the vector $w$:

$$w_k^* \triangleq \rho^k \frac{1 - \rho^{2(w-k+1)}}{1 - \rho^{2W+1}} \quad \text{for } k \in \{1, \ldots, W\}. \quad (4.92)$$

Furthermore, there exists a $\rho^* \in (0,1)$ such that for $\rho \geq \rho^*$ the hybrid coding scheme is the optimal scheme. Some examples of $\rho^*$ is shown in Table 4.1.

<table>
<thead>
<tr>
<th>$W$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\rho^*$</td>
<td>0.9220</td>
<td>0.9604</td>
<td>0.9764</td>
<td>0.9845</td>
</tr>
</tbody>
</table>

**Table 4.1**: Numerical values of $\rho^*$ for $W \in \{1, 2, 3, 4\}$.

The proof of Theorem 12 is presented in Section 4.5.6. Although Theorem 12 establishes the high resolution optimality of the hybrid coding scheme for a specific range of $\rho \geq \rho^*$, it is important to emphasize that, as Fig. 4.8 suggests, the hybrid coding scheme performs very close to optimal for a wider range of $\rho$ at least when $B = 1$.

**Theorem 13.** In high resolution regime, when $W = 1$, the excess rate of the hybrid coding scheme (See (4.53)), denoted by $R_{E,HR}(\rho, B)$, satisfies the following.

- For $B = 1$, $R_{E,HR}(\rho, B = 1)$ is upper bounded as follows.

$$R_{E,HR}(\rho, B = 1) \leq \frac{1}{4} \log \left( 1 + \frac{2\rho^4}{(1+\rho)^2} \right). \quad (4.93)$$

- For $B \to \infty$, we have

$$R_{E,HR}(\rho, B \to \infty) = \frac{1}{4} \min_w \{ \log (f(w)^2 - g(w)^2) \} \quad (4.94)$$

where

$$f(w) \triangleq \left( \frac{\rho^2}{1-\rho^2} + \frac{1}{1-w^2} \right) \frac{1}{(1+w\rho)^2} \quad (4.95)$$

$$g(w) \triangleq \rho f(w) - \frac{w}{w^2(1-w^2)}. \quad (4.96)$$

Table 4.2 summarizes the high resolution results of the above theorems. The proofs are provided in Section 4.5.6 and Section 4.5.6 respectively.

Fig. 4.8 shows the high resolution performance of the hybrid coding scheme in comparison with the memoryless Q-and-B, GOP-based coding and the lower bound on rate recovery. Recall that the
Figure 4.8: High resolution Excess Rates based on Hybrid Coding in comparison with Memoryless Q-and-B and the lower bound.

<table>
<thead>
<tr>
<th>Coding Scheme</th>
<th>Memoryless Q-and-B</th>
<th>Hybrid Coding</th>
<th>Predictive Coding</th>
<th>Lower Bound</th>
</tr>
</thead>
<tbody>
<tr>
<td>$B = W = 1$</td>
<td>$\frac{1}{4} \log (1 + \rho^2)$</td>
<td>$\leq \frac{1}{4} \log \left( 1 + \frac{2\rho^2}{(1+\rho)^2} \right)$</td>
<td>$\infty$</td>
<td>$\frac{1}{4} \log \left( 1 + \frac{\rho^4}{(1+\rho^2)} \right)$</td>
</tr>
<tr>
<td>$B \to \infty$</td>
<td>$\frac{1}{4} \log \left( \frac{1}{1-\rho^2} \right)$</td>
<td>$\min_w \frac{1}{4} \log (f(w)^2 - g(w)^2)$</td>
<td>$\infty$</td>
<td>$\frac{1}{4} \log \left( \frac{1}{1-\rho^2} \right)$</td>
</tr>
</tbody>
</table>

Table 4.2: High resolution excess rates for different coding schemes.
Figure 4.9: Comparison excess rate of different schemes in high resolution when $B \to \infty$, $W = 1$

predictive coding scheme cannot be applied in this regime. It can be observed from Fig. 4.8 that the hybrid coding scheme performs close to optimal at least for the examples of $B = W = 1$ and $B = 1$, $W = 2$. In fact for some ranges of $\rho$, $B$ and $W$, the hybrid coding scheme is indeed optimal in high resolution.

Fig. 4.9 shows the high resolution excess rate for $W = 1$ and $B \to \infty$ for the following schemes.

- Hybrid Coding Scheme: According to Theorem 13, the excess rate is computed by minimizing (4.94) for any $\rho$.

- Memoryless Q-and-B: According to Corollary 4, the excess rate is as follows.

\[
R_{E, HR, QB} = \frac{1}{4} \log \left( \frac{1}{1 - \rho^2} \right),
\]

(4.97)

- GOP-based Coding: According to (4.58) we have

\[
R_{E, HR, GOP} = R_{E, HR, QB} = \frac{1}{4} \log \left( \frac{1}{1 - \rho^2} \right),
\]

(4.98)

i.e., the GOP-based coding and memoryless Q-and-B scheme are equivalent in the case of High resolution and $B \to \infty$ for any $W$.

- Still-Image Coding: In this scheme the encoder ignores the decoder’s memory and at each time quantizes the source sequence $x_i$ within distortion $D$ and sends the quantization codewords through the channel. The rate of this coding scheme is simply computed as

\[
R_{SI} \triangleq \frac{1}{2} \log \left( \frac{1}{D} \right).
\]

(4.99)
Thus the excess rate in high resolution regime is computed as follows.

\[ R_{E,HR,SI} = \frac{1}{2} \log \left( \frac{1}{1 - \rho^2} \right). \]  

(4.100)

- Lower Bound: As will be shown in Corollary 5, the excess rate has to satisfy.

\[ R_{E,HR} = \frac{1}{4} \log \left( \frac{1}{1 - \rho^4} \right). \]  

(4.101)

### 4.4.4 Lower Bound on Rate-Recovery Function

In order to study the optimality of the achievable rates by the proposed coding scheme over erasure burst channels, it is useful to develop lower bounds for the achievable rates. The following theorem characterizes lower bound on the rate-recovery function.

**Theorem 14.** The rate-recovery function satisfies

\[ R(B, W, D) \geq R(B, W, D) \overset{\Delta}{=} \frac{1}{2} \log (x^*) \],

where \( x^* \) is the unique solution of the following polynomial for \( x \geq 1 \).

\[ p(x) \overset{\Delta}{=} x^W (x - \rho^2)^2 - \left( 1 - \rho^2(B+W+1) \right) \alpha(\rho, D)x + \left( \rho^2 - \rho^2(B+W+1) \right) \alpha(\rho, D) = 0, \]  

(4.102)

where

\[ \alpha(\rho, D) = \frac{1 - \rho^2}{D \left( 1 - (1 - D) \rho^{2(W+1)} \right)} \left( \frac{1 - \rho^2}{D} + \rho^2 \right)^W. \]  

(4.103)

The proof of Theorem 14 is presented in Sec. 4.6.

**Corollary 5.** In the high resolution regime when \( D \to 0 \), the lower bound on the rate-recovery function satisfies.

\[ \lim_{D \to 0} \left( R(B, W, D) - \frac{1}{2} \log \left( \frac{1 - \rho^2}{D} \right) \right) = \frac{1}{2(W+1)} \log \left( \frac{1 - \rho^{2(B+W+1)}}{1 - \rho^{2(W+1)}} \right). \]  

(4.104)

The proof of Corollary 5 is presented in Sec. 4.6.3. The proof is based on studying the behavior of the lower bound of Theorem 14 in the regime \( D \to 0 \).

### 4.4.5 Performance Over Statistical Channel Models

In this section we study the performance of the proposed hybrid coding scheme over statistical channels.
Chapter 4. Hybrid Coding

Channel with Independent Erasures

We first consider the zero-delay streaming of a unit-variance Gauss-Markov source process with correlation \( \rho \) over a simple statistical channel. The channel at each time may introduce an erasure with probability of \( \gamma \) and with a probability of \( 1 - \gamma \) perfectly reveals the channel packet to the destination. The decoder declares a loss at time \( t \) if it can not reproduce the source sequence of time \( t \) within required distortion \( D \). The probability of excess loss is defined as the probability of losses whenever the channel packets are not erased.

In Fig. 4.11 we consider the example of \( \rho = 0.8 \) and \( D = 0.1 \). The communication rate is assumed to be 2% additional over the erasure-free channel rate \( R_I \), i.e., \( R = 1.02R_I = 1.0629 \). Fig. 4.11(a) illustrates the required waiting time after the erasure burst ends until the decoder is able to reconstruct the source sequence within the desired distortion \( D = 0.2 \), as a function of burst length \( B \), for different coding schemes. The parameters for different coding schemes are designed as follows.

- Predictive Coding: For a fixed operational rate \( R \), the smallest feasible test channel noise is specified by the following (See Sec. 4.5.1 for details.)
  \[
  \tilde{\sigma}_z^2 = \frac{1 - \rho^2}{2(1 - \rho^2)} = 0.0967. 
  \]
  (4.105)

  Note that \( \tilde{\sigma}_z^2 \) is the distortion of the source at time \( t \) if all the channel packets up to time \( t \) are available at the decoder. The decoder at each time \( t \) reproduces the source sequence \( \hat{s}^n_t \) as follows.
  \[
  \hat{s}^n_t = \sum_{k \in K(t)} \rho^k u^n_{t-k} 
  \]
  (4.106)

  where \( K(t) \triangleq \{ \tau : \tau \leq t, g_t \neq * \} \) is the set of non-erasure times up to time \( t \). The decoder declares loss whenever the reconstructed source violates the distortion constraint \( D = 0.1 \).

- GOP-based Coding: By fixing the operational rate, the period of transmission of I-frames, i.e., \( W + 1 \) is specified as follows (See Theorem 10 for details.)
  \[
  W + 1 = \left\lfloor \frac{1}{2} \log \left( \frac{1 - (1 - D)10^2}{1 - (1 - D)10^2} \right) \right\rfloor = 29. 
  \]
  (4.107)

  In this scheme whenever the erasure happens at the channel the decoder declares loss up to the time of the next non-erased I-frame when the decoder gets back to recovery. Note that for the GOP-based scheme, unlike the other schemes, the average rate is considered and not the peak rate.

- Memoryless Q-and-B: The test channel noise for this scheme is the only design parameter. Small values of the test channel noise makes the recovery of the auxiliary random variables, i.e., the communication of the compressed sources, easier. However this will cause the final reproduction to have higher distortion. Thus we first choose the test channel noise which satisfies the distortion constraints. The rate determines the required waiting time after any erasure burst. In Fig. 4.11 the numerical value for the test channel noise is \( \sigma_z^2 = 0.1307 \).

- Hybrid Coding: The hybrid coding scheme has more design parameters compared with the other schemes. First, the hybrid coding scheme in general has \( W \) coefficients and the test channel noise
Figure 4.11: Comparison of different schemes for 2% rate-overhead over erasure-free channel rate, i.e., $R = 1.02R_1 = 1.0629$ Bits/Symbol, $\rho = 0.8$ and $D = 0.1$. 

(a) Waiting time ($W$) versus burst length ($B$).

(b) Probability of excess loss versus probability of erasure ($\gamma$).
as design parameters. In addition, the choice of $W$ itself is a parameter to choose: increasing $W$, decreases the required rate and increases the number of sources that the decoder gives up in recovery after the erasure burst. For any communication rate all these parameters for hybrid scheme are selected to minimize the overall loss probability. In particular for smaller rates close to $R_I$, higher values of $W$ are desired which reduces the compression rates. As the rate increases to $R_{SI}$, the smaller values are chosen for $W$. Here we considered the hybrid coding scheme with $W$ of at most 2. The coefficient $[w_1, w_2, w_3]$ and the test channel noise in Fig. 4.11 are designed for the case of $B = 3$, $W = 3$, and $D = 0.1$. In particular we have $[w_1, w_2, w_3] = [0.4832, 0.2858, 0.1121]$ and $\sigma_z^2 = 0.1302$.

• Lower Bound: In order to derive the lower bound on the waiting time for any burst length, we invoke the lower bound on the rate-recovery function to find the minimum required $W$ for any $B$ and fixed operational rate $R = 1.02R_I$. Note that in order to derive lower bound on probability of excess loss for the statistical channel, we assume there exists a genie that, whenever the erasure happens, reveals all the previous erased packets to the decoder. This is similar to the case where the decoder is treating an isolated erasure burst. By invoking the lower bound on the rate-recovery function, the minimum waiting time for such a case can be computed, which results in computing the minimum number of losses.

Fig. 4.11(b) illustrates the performance of different schemes over independent erasure channel as a function of erasure probability. It can be observed that the hybrid coding scheme outperforms the other schemes. The performance of the different coding scheme follows similar patterns to those in Fig. 4.11(a) for $B = 1$.

**Gilbert Channel Model**

We further consider the two-state Gilbert channel model [45, 46] (Fig. 4.10) in which no packet is lost in “good state” and all the packets are lost in “bad state”. Let $\alpha_g$ and $\beta_g$ denote the probability of transition from “good” to “bad” state and vice versa. In steady state, the probability of being in “bad state” and thus the erasure probability is $\alpha_g/(\alpha_g + \beta_g)$. It is not hard to verify that the mean burst length is equal to $1/\beta_g$.

Fig. 4.12(a) illustrates the performance of different schemes over Gilbert channel with $\alpha_G = 5 \times 10^{-3}$ as a function of mean burst length. It can be observed that the performance of the different coding scheme follows similar patterns to those in Fig. 4.11(a). Note that for all the mean burst lengths, the same hybrid scheme is considered.

**Gilbert-Elliott Channel Model**

Now consider the Gilbert-Elliott channel model [47]. The model is similar to the Gilbert model except that in the “Good” state the channel may introduce an erasure with a probability of $\epsilon$. In Fig. 4.12(b) we studied the performance of different schemes for transition of a Gauss-Markov source with $\rho = 0.8$ and $D = 0.1$ over the Gilbert-Elliott channel. In Fig. 4.12(a) we fixed the channel parameters $(\alpha_G, \beta_G) = (5 \times 10^{-3}, 1/3)$. The performance of different schemes are illustrated as a function of the parameter $\epsilon$. The operational rate is again $R = 1.02R_I = 1.0629$. It can be observed that the GOP-based coding is more sensitive to the increase of random erasures by increasing $\epsilon$. This is because the GOP-based scheme is insensitive to the burst length and even a single erasure forces this scheme to wait up to the time of
(a) Probability of excess loss versus mean burst length \(1/\beta_G\) of a Gilbert Channel.

(b) Probability of excess loss versus probability of erasure in good states \(\epsilon\) for Gilbert-Elliott Channel with \(\beta_G = \frac{1}{3}\).

Figure 4.12: Comparison of different schemes \(\rho = 0.8\), \(D = 0.1\), \(ao_G = 5 \times 10^{-3}\) and \(R = 1.0629\) Bits/Symbol.
next I-frame. It can be also observed that the memoryless Q-and-B performs slightly better than the GOP-based and predictive coding schemes as $\epsilon$ increases. The reason can be observed from Fig. 4.11(a) that the memoryless Q-and-B introduces smaller waiting time for erasure of length $B = 1$. Note that the frequency of such isolated erasures increases with $\epsilon$. Hybrid coding outperforms the other schemes and introduces lower increase in probability of excess loss as $\epsilon$ increases. This is mainly because of the low waiting time of the hybrid scheme for isolated erasures (Fig. 4.11(a)).

### 4.5 Upper Bounds on Rate-Recovery

In this section we study the performance of the predictive coding, the memoryless Q-and-B and the hybrid coding schemes in the case of erasure burst channel. Although not explicitly stated, throughout our discussion, our rate analysis at each time $t$ will permit a small error probability $\epsilon_n$ of decoding failure. By selecting the block length $n$ sufficiently large, the union bound argument as explained in Section 2.4.3 in Chapter 2 can be invoked to show that the streaming block error probability can be made vanishingly small, for any arbitrary duration $\Upsilon$.

#### 4.5.1 Predictive Coding

Here we present the Proof of Theorem 7. In predictive coding, as described in Sec. 4.3.3, the encoder at each time $t$, computes the MMSE estimation error of the source $x_t$ from all the previous codewords $u_i, i \leq t - 1$. Based on the optimality of the MMSE estimator for jointly Gaussian sources, the estimation error $e_t$, and thus $u_t$, is independent of the random variables $u_i, i \leq t - 1$. In the analysis, it is more convenient to use a backward test channel:

$$e_t = u_t + \tilde{z}_t,$$

where $\tilde{z}_t \sim \mathcal{N}(0, \tilde{\sigma}_z^2)$ is independent of $u_t, \forall i \leq t$. Using the orthogonality principle, one can show that

$$e_t = \rho \tilde{z}_{t-1} + n_t,$$

and furthermore we can show that

$$x_t = \rho^t u_0 + \rho^{t-1} u_1 + \ldots + \rho u_{t-1} + u_t + \tilde{z}_t$$

Furthermore, the encoder at each time $t$ quantizes $e_t$ where the quantization rate satisfies,

$$R \geq R_{PC}(\sigma_z^2) \triangleq I(e_t; u_t)$$

$$= \frac{1}{2} \log \left( \frac{\sigma_z^2}{\tilde{\sigma}_z^2} \right)$$

$$= \frac{1}{2} \log \left( \frac{1 - (1 - \tilde{\sigma}_z^2)\rho^2}{\tilde{\sigma}_z^2} \right).$$

The value of $\tilde{\sigma}_z^2$ will be specified in the sequel.

For the analysis of the erasure burst channel model observe that the decoder at anytime $t$, when the channel output $f_t$ is not erased, recovers $u_t$. Thus the reconstruction at time $t = \tau + B + W$, following
an erasure burst in \( \{\tau, \ldots, \tau + B - 1\} \) is
\[
\hat{x}_t = \rho^t u_0 + \rho^{t-1} u_1 + \ldots + \rho^B u_{\tau-1} + \rho^W u_{\tau+B} + \rho^{W-1} u_{\tau+B+1} + \ldots + u_{\tau+B+W}.
\] (4.114)

One can show that this corresponds to the worst case distortion which is
\[
E \left[ (x_t - \hat{x}_t)^2 \right] | t = \tau + B + W \triangleq \Sigma_{PC} (\sigma_z^2) \]
\[
= \tilde{\sigma}_z^2 + \sigma_u^2 \sum_{k=W+1}^{W+B} \rho^{2k} \]
\[
= \tilde{\sigma}_z^2 + \sigma_u^2 \rho^{2(W+1)} \frac{1 - \rho^{2B}}{1 - \rho^2} \]
\[
= \tilde{\sigma}_z^2 + (1 - \tilde{\sigma}_z^2) \rho^{2(W+1)} (1 - \rho^{2B}).
\] (4.117)

By setting \( \Sigma_{PC} (\sigma_z^2) = D \), we have
\[
\tilde{\sigma}_z^2 = \frac{D - \rho^{2(W+1)} (1 - \rho^{2B})}{1 - \rho^{2(W+1)} (1 - \rho^{2B})} \] (4.118)

for \( D \geq \rho^{2(W+1)} (1 - \rho^{2B}) \). By replacing \( \tilde{\sigma}_z^2 \) into the rate expression (4.113), we can observe that, for \( D \geq \rho^{2(W+1)} (1 - \rho^{2B}) \), any rate \( R \) satisfying
\[
R \geq R_{PC}^+ (B, W, D) \triangleq \frac{1}{2} \log \left( \frac{1 - \rho^{2(W+1)} (1 - \rho^{2B}) - (1 - D) \rho^2}{D - \rho^{2(W+1)} (1 - \rho^{2B})} \right)
\] (4.119)
is achievable.

### 4.5.2 Memoryless Q-and-B

Fig. 4.13 summarizes the main steps in proving Theorem 8. In particular, in Sec. 4.5.2, we first derive necessary parametric rate constraints associated with every possible erasure pattern. Second, through the Lemma 14, Lemma 15 and Lemma 16, we characterize the the worst-case erasure pattern that dominates the rate and distortion constraints. Finally in Section 4.5.2, we evaluate the achievable rate
to complete the proof of Theorem 8.

**Connection to DMS problem**

In order to study the rate of the memoryless Q-and-B scheme, consider the channel with an erasure burst spanning \( \{\tau, \ldots, \tau + B - 1\} \). Fig. 4.14 illustrates a erasure burst channel model parametrized by \( \tau \), i.e., the time where the erasure burst of length \( B \) starts. We identify three different time regions. Note that the achievable rate expression for any rate region follows from the generalization of Theorem 6 to Gauss-Markov sources.

- **Region 1:** \( t < \tau \), where there is no previous erasure by the channel. The decoder recovers \( u_t \) given \( \{u_1, \ldots, u_{t-1}\} \). This succeeds with high probability if
  \[
  R \geq R_{1,\tau}(t, \sigma_z^2) \triangleq h(u_t | u_{1}^{t-1}) - \frac{1}{2} \log (2\pi e \sigma_z^2). \tag{4.120}
  \]
  Furthermore, the decoder reconstructs the source sequence \( x_t \) within distortion
  \[
  D_{1,\tau}(t, \sigma_z^2) \triangleq \text{Var}(x_t | u_{1}^{t-1}). \tag{4.121}
  \]

- **Region 2:** \( t = \tau + B + W \), right after the erasure burst of length \( B \) spanning \( \{\tau, \ldots, \tau + B - 1\} \) and a window of length \( W \) after that. The decoder simultaneously recovers all the codewords \( [u]_{\tau+B+W}^{t} \) given \( \{u_1, \ldots, u_{\tau-1}\} \). This succeeds with high probability if
  \[
  R \geq R_{2,\tau}(\sigma_z^2) \triangleq \max_{M \subseteq \mathcal{L}_\tau, M \neq \phi} \frac{1}{|M|} h([u]_{\mathcal{M}} [u]_{1}^{\tau-1}, [u]_{\bar{M}}) - \frac{1}{2} \log (2\pi e \sigma_z^2), \tag{4.122}
  \]
  where
  \[
  \mathcal{L}_\tau \triangleq \{\tau + B, \ldots, \tau + B + W\}. \tag{4.123}
  \]
  Furthermore, the decoder reconstructs the source sequence \( x_{\tau+B+W} \) within distortion
  \[
  D_{2,\tau}(\sigma_z^2) \triangleq \text{Var}(x_{\tau+B+W} | [u]_{1}^{\tau-1}, [u]_{\tau+B}). \tag{4.124}
  \]

- **Region 3:** \( t > \tau + B + W \), the time after Region 2. The decoder recovers \( u_t \) given
  \[
  \{u_1, \ldots, u_{\tau-1}, u_{\tau+B+W}, \ldots, u_{t-1}\}.
  \]
This succeeds with high probability if
\[ R \geq R_{3,\tau}(t, \sigma_z^2) \triangleq h(u_t || u_{1:t-1}, [u]_{\tau + B}) - \frac{1}{2} \log(2\pi e \sigma_z^2). \] (4.125)

Furthermore, the decoder reconstructs the source sequence \( x_t \) within the following distortion.
\[ D_{3,\tau}(t, \sigma_z^2) \triangleq \text{Var}(x_t || u_{1:t-1}, [u]_{\tau + B}). \] (4.126)

For any parameter \( \tau \), define
\[ R_{\tau}(t, \sigma_z^2) \triangleq \begin{cases} 
R_{1,\tau}(t, \sigma_z^2), & t < \tau \\
R_{2,\tau}(\sigma_z^2), & t = \tau + B + W \\
R_{3,\tau}(t, \sigma_z^2), & t > \tau + B + W 
\end{cases} \] (4.127)

\[ D_{\tau}(t, \sigma_z^2) \triangleq \begin{cases} 
D_{1,\tau}(t, \sigma_z^2), & t < \tau \\
D_{2,\tau}(\sigma_z^2), & t = \tau + B + W \\
D_{3,\tau}(t, \sigma_z^2), & t > \tau + B + W 
\end{cases} \] (4.128)

The rate and distortion constraints have to be satisfied for all possible parameters \( \tau \). In particular, the following rate is achievable.
\[ R \geq \max_{\tau} \max_t R_{\tau}(t, \sigma_z^2), \] (4.129)

for any test channel noise satisfying,
\[ \max_{\tau} \max_t D_{\tau}(t, \sigma_z^2) \leq D. \] (4.130)

**Remark 12.** Although in this section we consider the special case of memoryless Q-and-B scheme, the rate and distortion requirements in (4.129) and (4.130) are indeed valid for any test channel of the general coding scheme described in 4.3.2.

**Worst-Case Characterization of the erasure burst**

We prove Theorem 8 by simplifying (4.129) and (4.130) through the following steps. These will be stated rigorously in the sequel.

- (Lemma 14) The worst case erasure burst, with respect to both rate and distortion constraints in (4.129) and (4.130), happens at \( \tau \to \infty \).

- (Lemma 15) The worst case rate and distortion constraints at \( \tau \to \infty \) are indeed those of region 2, i.e., \( R_{2,\infty}(\sigma_z^2) \) and \( D_{2,\infty}(\sigma_z^2) \).

- (Lemma 16) Among the constraints of \( R_{2,\infty}(\sigma_z^2) \), the sum-rate constraint is always dominant.

- (Sec. 4.5.2) The sum-rate constraint is equivalent to the expression in Theorem 8.
Lemma 14. The functions $R_\tau(\sigma_z^2)$ and $D_\tau(\sigma_z^2)$, defined as,

\[
R_\tau(\sigma_z^2) \triangleq \max_t R_\tau(t, \sigma_z^2) \\
D_\tau(\sigma_z^2) \triangleq \max_t D_\tau(t, \sigma_z^2).
\] (4.131)

are increasing functions with respect to $\tau$.

Proof. See Appendix C.1.

We can rewrite (4.129), as follows.

\[
R \geq \max_\tau \max_t R_\tau(t, \sigma_z^2) = \max_\tau R_\tau(\sigma_z^2) = \lim_{\tau \to \infty} R_\tau(\sigma_z^2) \triangleq R_\infty(\sigma_z^2). \tag{4.133}
\]

By using the similar notation for distortion expressions, we can show that the rate constraint in (4.130) is equivalent to the following.

\[
D_\infty(\sigma_z^2) \leq D. \tag{4.135}
\]

Furthermore, the recovery immediately following the erasure burst dominates.

Lemma 15. In the limit that $\tau \to \infty$, the rate and distortion constraints of region 2 is the worst, i.e.,

\[
R_\infty(\sigma_z^2) = R_{2,\infty}(\sigma_z^2) \tag{4.136}
\]

\[
D_\infty(\sigma_z^2) = D_{2,\infty}(\sigma_z^2). \tag{4.137}
\]

Proof. See Appendix C.2.

Lemma 16. The sum-rate constraint in (4.122) is indeed the dominant constraint, i.e., as $\tau \to \infty$,

\[
\arg\max_{\mathcal{M} \in \mathcal{C}, \mathcal{M} \neq \emptyset} \frac{1}{|\mathcal{M}|} h([u]_\mathcal{M}||u|_1^{-1}, |u|_{\mathcal{M}}) = \mathcal{L}_\tau, \tag{4.138}
\]

where $\mathcal{L}_\tau$ is defined in (4.123). In particular,

\[
\sup_{\tau \in \{1,2,...\}} \lim_{\tau \to \infty} \frac{1}{|\mathcal{L}_\tau|} h([u]_{\tau+B+W}||u|_1^{-1}) - \frac{1}{2} \log (2\pi e \sigma_z^2). \tag{4.139}
\]

Proof. See Appendix C.3.

According to Lemma 14, Lemma 15 and Lemma 16, the following rate and distortion constraints are active.

\[
R \geq \lim_{\tau \to \infty} \frac{1}{W+1} h([u]_{\tau+B+W}||u|_1^{-1}) - \frac{1}{2} \log (2\pi e \sigma_z^2) \tag{4.140}
\]

\[
\lim_{\tau \to \infty} \text{Var}(x_{\tau+B+W}||u|_1^{-1}, [u]_{\tau+B+W}^{-1}) \leq D \tag{4.141}
\]
Rate Evaluation

We need to show that the following rate and distortion constraints
\[ R \geq \frac{1}{W + 1} h([u]_{i+B}^{t+B+W} | \tilde{s}_t) - \frac{1}{2} \log(2\pi e \sigma_z^2) \]  
\[ \text{Var}(s_{t+B+W} | \tilde{s}_t, [u]_{i+B}^{t+B+W}) \leq D \]

satisfies the constraints in (4.140) and (4.141) and thus are achievable. To this end it is helpful to consider the following Kalman filter for \( i < \tau \).

\[ s_i = \rho s_{i-1} + n_i, \quad n_i \sim \mathcal{N}(0, 1 - \rho^2) \]  
\[ u_i = s_i + z_i, \quad z_i \sim \mathcal{N}(0, \sigma_z^2) \]

At time \( \tau - 1 \), the Kalman filter is in steady state. According to the orthogonality principle of MMSE for Gaussian sources, observation of all \([u]_{1}^{\tau-1}\), is equivalent to having access to the following noisy version of the \( s_\tau \).

\[ \tilde{s}_\tau = s_\tau + e \]

where \( e \sim \mathcal{N}(0, \Sigma(\sigma_z^2)/(1 - \Sigma(\sigma_z^2))) \), and \( \Sigma(\sigma_z^2) \) is defined in (4.51). Thus the expressions in (4.140) and (4.141) at \( \tau \to \infty \) can be written as (4.142) and (4.143). Thus the rate in (4.142) is achievable for any test channel noise \( \sigma_z^2 \) satisfying (4.143). This completes the proof of Theorem 8.

4.5.3 Memoryless Q-and-B in the High Resolution Regime

In order to analyze the high resolution behavior of the memoryless quantization-and-binning scheme, it suffices to study the rate expression in (4.50), or equivalently (4.140), in the limit \( D \to 0 \). In particular we need to show that,

\[ R_{QB}^+(B, W, D) = \frac{1}{2} \log \left( \frac{1 - \rho^2}{D} \right) + \frac{1}{2(W + 1)} \log \left( \frac{1 - \rho^{2(B+1)}}{1 - \rho^2} \right) + o(D). \]  

where \( \lim_{D \to 0} o(D) = 0 \). First we set \( \sigma_z^2 = D \), which satisfies the distortion constraint in (4.52), i.e.,

\[ \text{Var}(s_{t+W} | [u]_{t+W}^{t+W}, \tilde{s}_{t-B}) \leq \text{Var}(s_{t+W} | u_{t+W}) \]

\[ = \frac{D}{1+D} \leq D. \]  

Note that when \( \sigma_z^2 = D \to 0 \), random variable \( u_t \) becomes asymptotically close to \( s_t \). Thus the Markov property among the sources \( s_t \) approximately holds among \( u_t \). Based on this observation, the high resolution limit of the first differential entropy term in (4.140) can be calculated as

\[ \lim_{D \to 0} \lim_{\tau \to \infty} h([u]_{\tau+B}^{\tau+B+W} | [u]_1^{\tau-1}) = \lim_{\tau \to \infty} h([s]_{\tau+B}^{\tau+B+W} | [s]_1^{\tau-1}) \]

\[ = h([s]_{B+2}^{B+W} | s_1) \]

\[ = \frac{1}{2} \log \left( (2\pi e)^{W+1} (1 - \rho^{2(B+1)})(1 - \rho^2)^W \right) \]  

(4.149)
Finally by replacing (4.149) into (4.140) with \( \sigma_z^2 = D \), the expression in (4.147) is derived. This completes the proof.

### 4.5.4 GOP-Based Coding Scheme

The GOP-Based coding scheme for the zero-delay streaming setup periodically transmits the \textit{I-frames} as the \textit{intra-coded} pictures that can be decoded at the decoder without the use of any other frame. Between the two consecutive I-frames, the \textit{P-frames} as the \textit{predicted} pictures are transmitted which require the use of previous frames to be decoded at the decoder.

According to the problem setup, in case of erasure burst of the channel, the decoder is required to start decoding the source vectors at most \( W + 1 \) times after the erasure ends. It is not hard to observe that in the GOP-based scheme the worst erasure pattern erases the I-frame and reveals the packets right after the I-frame. This suggest that in order to guarantee the recovery after \( W + 1 \) times, the I-frames have to be sent at least with a period of \( W + 1 \).

Let us define \( v_t \) as the quantization of the source vector \( s_t \) as the I-frame. Using the Gaussian test channel, the quantization can be modeled as follows.

\[
s_t = v_t + z_t, \tag{4.150}
\]

Note that \( z_t \sim \mathcal{N}(0, D) \) which guarantees the average distortion constraint. The decoder succeeds in reconstructing the source by only using the encoder output at time \( t \) if the rate satisfy

\[
R_t \geq \frac{1}{2} \log \left( \frac{1}{D} \right). \tag{4.151}
\]

For the time interval \( \{t+1, \ldots, t+W\} \) the encoder sends \( u_t \) as the output of the predictive encoder, i.e., the P-frame. Using the similar notation for the predictive coding it is not hard to observe that the source \( s_i \) for any \( i \in \{t+1, t+W\} \) can be represented as follows.

\[
s_i = \rho^{i-t} v_t + \rho^{i-t-1} u_{t+1} + \ldots + \rho u_{t-1} + u_i + z_i. \tag{4.152}
\]

At each time \( i \in \{t+1, \ldots, t+W\} \) the decoder succeeds in recovering \( u_i \) if the rate satisfy

\[
R_i \geq \frac{1}{2} \left( \frac{1 - \rho^2}{D} + \rho^2 \right). \tag{4.153}
\]

From (4.151) and (4.153) it can be observed that the scheme requires the following average rate.

\[
\bar{R}_{\text{GOP}}(W, D) = \frac{1}{W+1} \sum_{i=t}^{t+W} R_i = \frac{1}{2} \log \left( \frac{1 - \rho^2}{D} + \rho^2 \right) + \frac{1}{2(W+1)} \log \left( \frac{1}{1 - (1-D)\rho^2} \right). \tag{4.154}
\]

Note that the rate expression in (4.154) is independent of the burst length \( B \).
4.5.5 Hybrid Coding Scheme

As we noted in Remark 12, the rate and distortion constraints of (4.127) and (4.128) are valid not only for the memoryless Q-and-B but also for the general coding scheme described in Section 4.3.2. This includes the hybrid coding scheme as a special case. Consider the test channel for the hybrid coding scheme with parameters $B$ and $W$ described in Section 4.4.3. In this section, we prove Theorem 11 for the hybrid coding scheme by simplifying (4.127) and (4.128) through the following steps.

**Worst Case Reconstruction:** First we show that for any test channel noise $\sigma_z^2$, the worst-case rate constraint (4.127), is the following.

$$\sup_{\tau,t} R_\tau(t,\sigma_z^2) = \lim_{\tau \to \infty} R_2,\tau(\sigma_z^2)$$

$$= \lim_{\tau \to \infty} \max_{M \subseteq L_\tau, M \neq \emptyset} \frac{1}{|M|} h([u]_{M}||[u]_{M}^{\tau-1}, [u]_{M}), - \frac{1}{2} \log(2\pi e \sigma_z^2)$$ (4.155)

where $L_\tau$ is defined in (4.123). In addition the test channel noise $\sigma_z^2$ has to satisfy the following worst-case distortion constraint.

$$\sup_{\tau,t} D_\tau(t,\sigma_z^2) = \lim_{\tau \to \infty} D_2,\tau(\sigma_z^2)$$

$$= \lim_{\tau \to \infty} \text{Var}(x_{\tau+B+W}||[u]_{1}^{\tau-1}, [u]_{\tau+B+W}) \leq D.$$ (4.156)

These are proved by the following lemmas whose proofs are presented in associated appendices.

**Lemma 17.** For any fixed $\tau$, and any $t \leq \tau$ we have

$$R_2,\tau(\sigma_z^2) \geq R_1,\tau(t,\sigma_z^2)$$ (4.157)

$$D_2,\tau(\sigma_z^2) \geq D_1,\tau(t,\sigma_z^2).$$ (4.158)

The proof of Lemma 17 in presented in Appendix C.4. The following lemma is the generalization of Lemma 14 to the hybrid coding scheme.

**Lemma 18.** The two functions $R_\tau(\sigma_z^2)$ and $D_\tau(\sigma_z^2)$ defined similarly to (4.131) and (4.132), respectively, are increasing functions with respect to $\tau$ for the case of hybrid coding scheme.

The proof of Lemma 18 in presented in Appendix C.5.

**Lemma 19.** For any $t > \tau + B + W$, as $\tau \to \infty$, we have:

$$R_2,\tau(\sigma_z^2) \geq R_3,\tau(t,\sigma_z^2)$$ (4.159)

$$D_2,\tau(\sigma_z^2) \geq D_3,\tau(t,\sigma_z^2).$$ (4.160)

The proof of Lemma 19 is provided in Appendix C.6.

According to Lemma 17 for any $t$, the rate and distortion constraints of region 2, i.e., $R_2,\tau(\sigma_z^2)$ and $D_2,\tau(\sigma_z^2)$, always dominate the constraints of region 1. According to lemma 18, we only need to focus on the case where the erasure burst happens at $\tau \to \infty$. Finally according to Lemma 19, as $\tau \to \infty$, the rate and distortion constraints of region 2 also dominate the constraints of region 3. By combining

...
these results, it can be concluded that \( \lim_{\tau \to \infty} R_{2,\tau}(\sigma_2^2) \) and \( \lim_{\tau \to \infty} D_{2,\tau}(\sigma_2^2) \) are the dominating rate and distortion constraints as required in (4.155) and (4.156).

**Rate Computation:** In this step we show that the rate and distortion constraints in (4.155) and (4.156) are equivalent to the equations in Theorem 11. Define the \((\tau - 1) \times (\tau - 1)\) matrix \(Q_{H,\tau}\) which consists of the first \(\tau - 1\) rows and columns of \(Q_H\). This matrix is lower triangular and thus invertible. Therefore, all the observations \([u]_1^{\tau-1}\), can be replaced by the following.

\[
Q_{H,\tau}^{-1} \begin{pmatrix} u_1 \\ \vdots \\ u_{\tau-1} \end{pmatrix} = \begin{pmatrix} x_1 \\ \vdots \\ x_{\tau-1} \end{pmatrix} + \begin{pmatrix} z_1 \\ \vdots \\ z_{\tau-1} \end{pmatrix} \triangleq \begin{pmatrix} v_1 \\ \vdots \\ v_{\tau-1} \end{pmatrix} \quad (4.161)
\]

For instance, the first differential entropy term in (4.155) can be written as follows,

\[
h([u]_1^{\tau-1}, [u]_{\tau-1}) = h([u]_1^{\tau-1}, [v]_{\tau-1}) = h([u]_1^{\tau-1}, [u]_{\tau-1}),
\]

where \(\mathcal{M} \subseteq \mathcal{L}_\tau\). Similarly, the distortion constraint in (4.156) can be replaced by:

\[
\text{Var}(x_{\tau+B+W} [u]_1^{\tau-1}, [u]_{\tau+B+W}) = \text{Var}(x_{\tau+B+W} [Q_{H,\tau}]^{-1}[u]_1^{\tau-1}, [u]_{\tau+B+W})
\]

\[
= \text{Var}(x_{\tau+B+W} [v]_1^{\tau-1}, [u]_{\tau}),
\]

(4.163)

Furthermore, note that all the random variables \([u]_{\mathcal{L}_\tau}\) can be written as follows.

\[
[u]_{\mathcal{L}_\tau} = Q_{1} ([x]_1^{\tau+B+W} + [z]_1^{\tau+B+W})
\]

\[
= [Q_{1}, Q_{\text{eff}}] ([x]_1^{\tau+B+W} + [z]_1^{\tau+B+W})
\]

\[
= Q_{1} ([x]_1^{\tau-1} + [z]_1^{\tau-1}) + Q_{\text{eff}} ([x]_1^{\tau+B+W} + [z]_1^{\tau+B+W})
\]

\[
= Q_{1}[v]_1^{\tau-1} + Q_{\text{eff}}[v]_1^{\tau+B+W}
\]

(4.164)

where we defined the matrix \(Q_{1}\) of size \((W + 1) \times (\tau + B + W)\) consisting of rows with index \(\mathcal{L}_\tau\) and columns with index \(\{1, \ldots, \tau + B + W\}\) of matrix \(Q_H\). In addition we defined the matrices \(Q_{1}\) and \(Q_{\text{eff}}\) of sizes \((W + 1) \times (\tau - 1)\) and \((W + 1) \times (B + W + 1)\), respectively such that

\[
Q_{1} = [Q_{1}, Q_{\text{eff}}].
\]

(4.165)

When \([v]_1^{\tau-1}\) is known, which is the case in (4.162) and (4.163), the part related to \([v]_1^{\tau-1}\) in (4.164) can be subtracted out from \([u]_{\mathcal{L}_\tau}\). For instance the term in (4.163) can be written as follows.

\[
\text{Var}(x_{\tau+B+W} [v]_1^{\tau-1}, [u]_{\mathcal{L}_\tau}) = \text{Var}(x_{\tau+B+W} [v]_1^{\tau-1}, Q_{\text{eff}}[v]_1^{\tau+B+W}).
\]

(4.166)

Note that at \(\tau \to \infty\), \(x_i\) can be replaced by \(s_i\). By invoking the Kalman filter argument described in Section 4.5.2, one can replace \([v]_1^{\tau-1}\) by \(\bar{s}_\tau \triangleq s_\tau + e\), where \(e\) is defined in Theorem 11. Finally according to the Toeplitz property of the matrix \(Q_H\), and therefore \(Q_{\text{eff}}[v]_1^{\tau+B+W}\), we can write (4.166) as follows.

\[
\max_{\tau,t} D_{\tau}(t, \sigma_2^2) = \lim_{\tau \to \infty} \text{Var}(\bar{s}_{\tau+B+W}|\bar{s}_\tau, Q_{\text{eff}}[v]_1^{\tau+B+W})
\]
\[= \text{Var}(s_{B+W+1}|\tilde{s}_1, Q_{\text{eff}}([s]^{B+W+1}_1 + [z]^{B+W+1}_1))
\]
\[= \text{Var}(s_{B+W+1}|\tilde{s}_1, [\tilde{u}]^{B+W+1}_{B+1}), \quad (4.167)\]

with the test channel defined in (4.88). Also using similar argument the rate constraint in (4.155) reduces to the following.

\[
\max_{\tau, t} R_t(t, \sigma_z^2) = \lim_{\tau \to \infty} \max_{M \subseteq L, M \neq \emptyset} \frac{1}{|M|} h([u]_M|\tilde{s}_\tau, [u]_{\tilde{M}}) - \frac{1}{2} \log(2\pi e \sigma_z^2) 
\]
\[
\geq \max_{M \subseteq L, M \neq \emptyset} \frac{1}{|M|} h([\tilde{u}]_M|\tilde{s}_1, [\tilde{u}]_{\tilde{M}}) - \frac{1}{2} \log(2\pi e \sigma_z^2). \quad (4.168)\]

This completes the proof.

### 4.5.6 Hybrid Coding Scheme In the High Resolution Regime

**Proof of Theorem 12**

In order to prove Theorem 12, which states the behavior of the hybrid coding scheme in high resolution asymptotic, we note that it suffices to set the test channel noise \(\sigma_z^2\) to be equal to \(D\). This test channel noise satisfies the distortion constraint. In particular consider the case where the channel packets of time span \(\{t - B + 1, \ldots, t\}\) are erased and the decoder is interested in reconstruct the source at time \(t + W + 1\) while having access to \(\{u_{t+1}, \ldots, u_{t+W+1}\}\). According to the hybrid test channel, we have

\[x_{t+W+1} + z_{t+W+1} = \sum_{j=0}^{W} w_j u_{t+W-j+1}. \quad (4.169)\]

Therefore at least \(x_{t+W+1} + z_{t+W+1}\) is available at the decoder while reconstructing \(x_{t+W+1}\). Now note that

\[\text{Var}(x_{t+W+1}|[u]_t^{t-B}, [u]_{t+1}^{t+W+1}) \leq \text{Var}(x_{t+W+1}|x_{t+W+1} + z_{t+W+1}) = \frac{D}{1+D} \leq D. \quad (4.170)\]

The sum-rate constraint of the hybrid coding scheme is

\[R_{\text{sum}} = \lim_{t \to \infty} \frac{1}{W+1} h([u]_t^{t+W+1}|[u]_1^{t-B}) - \frac{1}{2} \log (2\pi e D). \quad (4.171)\]

Thus the choice of hybrid coding weights \(w_k\) that minimizes the sum-rate constraint, minimizes the first term in (4.171).

The following two lemmas show that the choice of weights in (4.92) in fact minimizes the sum-rate constraint. First consider the following lemma which is valid for any arbitrary distortion.

**Lemma 20.** For any \(B, W\) and a fixed test channel noise, the choice of the hybrid coding scheme parameters \([w]_1^W\) which minimizes the sum-rate constraint, i.e.,

\[\frac{1}{W+1} \lim_{t \to \infty} h([u]_t^{t+W+1}|[u]_1^{t-B})
\]
also minimizes

$$
\lim_{t \to \infty} I([u]_{t-B+1}^t; [u]_{t+1}^{t+W+1}|[u]_1^{t-B}, s_{t+W+1}),
$$

(4.172)

which is the steady state mutual information between the test channel outputs in the interval of erasure burst and the those of a window of length $W + 1$ after the erasure given all the test channel outputs up to the erasure time and the source after the error propagation window.

The proof of Lemma 20 is presented in Appendix C.7.

Consider the case $B = 1$ and the hybrid coding in high resolution regime when the test channel noise variance $\sigma^2$ approaches arbitrarily close to zero. Before the erasure burst starts, the decoder is able to reconstruct all the source sequences $x$ with high resolution. Equivalently, the innovation process $n_t$ are recovered at the decoder within high resolution. Approximately we assume that the original process $n_t$ are available at the decoder for those times, thus the mutual information term in (4.172) at high resolution can be computed as follows.

$$
\lim_{D \to 0} \lim_{t \to \infty} I(u_t; [u]_{t+1}^{t+W+1}|[u]_1^{t-1}, s_{t+W+1}) = \lim_{D \to 0} \lim_{t \to \infty} I(u_t; [u]_{t+1}^{t+W+1}|[n]_1^{t-1}, s_{t+W+1})
$$

$$
= \lim_{D \to 0} I(\hat{u}_1; [\hat{u}]_2^{W+2}|x_{W+2}),
$$

(4.173)

where in (4.173) we defined

$$
[\hat{u}]_1^{W+2} = Q_{\text{eff}} ([x]_1^{W+2} + [z]_1^{W+2}),
$$

(4.174)

where $Q_{\text{eff}}$ is the square matrix of size $W + 2$ consists of first $W + 2$ rows and columns of $Q_H$. In particular, (4.174) can be written as follows.

$$
\begin{pmatrix}
\hat{u}_1 \\
v_1 \\
\vdots \\
v_{W+1}
\end{pmatrix}
\begin{pmatrix}
1 & 0 & 0 & \cdots & 0 \\
0 & 1 & 0 & \cdots & 0 \\
0 & 0 & 1 & \cdots & 0 \\
0 & 0 & 0 & \cdots & 1
\end{pmatrix}
\begin{pmatrix}
x_1 \\
x_2 \\
\vdots \\
x_{W+1}
\end{pmatrix}
+ \begin{pmatrix}
z_1 \\
z_2 \\
\vdots \\
z_{W+1}
\end{pmatrix}
$$

(4.175)

**Lemma 21.** Consider the high resolution test channel of the hybrid coding scheme with the vector $w^*$ of size $W$, with the following elements.

$$
w^*_k = \rho^k \frac{1 - \rho^{2(W-k+1)}}{1 - \rho^{2W+1}} \quad \text{for } k \in \{1, \ldots, W\},
$$

(4.176)
Then we have the following.

\[
\lim_{D \to 0} I(\hat{u}_1; [\hat{u}_2]_W^{W+2}|x_{W+2}) = 0. \quad (4.177)
\]

**Proof.** First note that \(D \to 0\) requires \(\sigma^2 \to 0\). In fact it is not hard to show that in order to have (4.177), we can ignore the noise in the test channel of (4.175) and show (4.177) for the following test channel.

\[
\begin{pmatrix}
\hat{u}_1 \\
\vdots \\
\hat{u}_{W+2}
\end{pmatrix} = 
\begin{pmatrix}
1 & 0 & 0 & \cdots & 0 \\
v_1 & 1 & 0 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
v_{W+1} & v_W & v_{W-1} & \cdots & 1
\end{pmatrix} 
\begin{pmatrix}
1 & 0 & 0 & \cdots & 0 \\
\rho & 1 & 0 & \cdots & 0 \\
\rho^2 & \rho & 1 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
\rho^{W+1} & \rho^{W} & \rho^{W-1} & \cdots & 1
\end{pmatrix} 
\begin{pmatrix}
m_1 \\
\vdots \\
m_{W+2}
\end{pmatrix} \quad (4.178)
\]

Also remember from (4.83) that the elements of the matrix \(Q_{\text{eff}}\), i.e. \(v_m\), are related to hybrid coding parameters \(w_j\), through the following equation.

\[
v_m = -\sum_{j=0}^{m-1} w_{m-j}v_j, \quad \forall m \geq 1 \quad (4.179)
\]

where \(v_0 = 1\). By defining \(w_0 = 1\), (4.179) can be written as

\[
\sum_{j=0}^{m} w_{m-j}v_j = 0. \quad (4.180)
\]

According to the chain rule of mutual information, we have

\[
I(\hat{u}_1; [\hat{u}_2]_W^{W+2}|x_{W+2}) = I(\hat{u}_1; [\hat{u}_2]_W^{W+1}|x_{W+2}) + I(\hat{u}_1; \hat{u}_{W+2}|[\hat{u}_2]_W^{W+1}, x_{W+2}). \quad (4.181)
\]

- We first show that the second term in (4.181) approaches to zero. According to the definition of the test channel of the hybrid scheme, as \(D \to 0\), \(x_{W+2}\) can be expressed as linear combination of \(\hat{u}_j\) for \(j \in \{2, \ldots, W+2\}\). Equivalently, \(\hat{u}_{W+2}\) is a linear combination of \(\hat{u}_j\) for \(j \in \{2, \ldots, W+1\}\) and \(x_{W+2}\). Thus

\[
\lim_{D \to 0} I(\hat{u}_1; \hat{u}_{W+2}|[\hat{u}_2]_W^{W+1}, x_{W+2}) = 0 \quad (4.182)
\]

- It remains to show that the first term in (4.181) approaches to zero as \(D \to 0\). This is equivalent to show that there exists a vector \(h\) of size \(W\) such that

\[
[\hat{u}]_W^{W+1} = hx_{W+2} + e, \quad (4.183)
\]

where the noise vector \(e\) is independent of \(x_{W+2}\) and \(\hat{u}_1\). According to the fact that all the random variables are jointly Gaussian, \(e\) will be jointly independent of independent of \(\{x_{W+2}, \hat{u}_1\}\) and thus we will have

\[
\lim_{D \to 0} I(\hat{u}_1; [\hat{u}_2]_W^{W+1}|x_{W+2}) = \lim_{D \to 0} I(\hat{u}_1; hx_{W+2} + e|x_{W+2})
\]
According to the test channel in (4.178), \( \hat{u}_1 = \nu_1 \). We show (4.183) through the following two steps.

**Step 1:** We first show that the choice of \( h \) where

\[
h_k = \rho^{k-W-2} \sum_{j=0}^{k-1} v_j \rho^{-j}, \quad \forall k \in \{2, \ldots, W + 2\}
\]  

(4.185)

guarantees that \( e \) is independent of \( \nu_1 \). To see this, note from (4.178) that for any \( k \in \{2, \ldots, W + 2\} \),

\[
\hat{u}_k = \rho^{k-1} \sum_{j=0}^{k-1} v_j \rho^{-j} n_1 + F(n_2, \ldots, n_{W+2}),
\]  

(4.186)

where \( F(n_2, \ldots, n_{W+2}) \) is a linear combination of \( \{n_2, \ldots, n_{W+2}\} \) and not \( \nu_1 \). Thus with the choice of \( w_k \) in (4.185), we have

\[
e_k = \hat{u}_k - h_k x_{W+2}
\]

\[
= \rho^{k-1} \sum_{j=0}^{k-1} v_j \rho^{-j} n_1 + F(n_2, \ldots, n_{W+2}) - \left( \rho^{k-W-2} \sum_{j=0}^{k-1} v_j \rho^{-j} \right) x_{W+2}
\]

\[
= F(n_2, \ldots, n_{W+2}) - \left( \rho^{k-W-2} \sum_{j=0}^{k-1} v_j \rho^{-j} \right) \sum_{i=0}^{W} \rho^i n_{W+2-i}.
\]  

(4.187)

Note that (4.187) is linear combination of \( n_i \) for \( i \in \{2, \ldots, W + 2\} \) and thus is independent of \( \nu_1 \).

**Step 2:** We now need to show that \( e \) is also independent of \( x_{W+2} \). We show that the choice of \( h \) in (4.185) is the MMSE estimation coefficients of [\( \hat{u}_2 \)] from \( x_{W+2} \). Thus from the orthogonality principle of jointly Gaussian sources, \( e_k \), the estimation error is independent of \( x_{W+2} \), the observation.

The MMSE estimation coefficient can be computed as follows for any \( k \in \{2, \ldots, W + 1\} \).

\[
\tilde{h}_k = \frac{E\{x_{W+2} \hat{u}_k\}}{E\{x_{W+2}^2\}}
\]  

(4.188)

Note from (4.178) that for any \( k \in \{2, \ldots, W + 1\} \) we have

\[
E\{x_{W+2} \hat{u}_k\} = \left( \sum_{l=0}^{k-1} v_l x_{W+l} \right)
\]

\[
= \sum_{l=0}^{k-1} v_l E\{x_{W+2} x_{k-l}\}
\]

\[
= \sum_{l=0}^{k-1} v_l \left( \sum_{z=1}^{W+2} n_z \sum_{z'=1}^{k-l} \rho^{k-l-z'} n_{z'} \right)
\]  

\[
= \sum_{l=0}^{k-1} v_l \left( \sum_{z=1}^{W+2} \rho^{W+2-z} n_z \sum_{z'=1}^{k-l} \rho^{k-l-z'} n_{z'} \right)
\]  

\[
= \sum_{l=0}^{k-1} v_l \left( \sum_{z=1}^{W+2} \rho^{W+2-z} \sum_{z'=1}^{k-l} \rho^{k-l-z'} n_{z'} \right)
\]  

\[
= \sum_{l=0}^{k-1} v_l \left( \sum_{z=1}^{W+2} \rho^{W+2-z} \sum_{z'=1}^{k-l} \rho^{k-l-z'} n_{z'} \right)
\]  

(4.189)
\[ k \sum_{l=0}^{k-1} v_l \sum_{z=1}^{k-l} (1 - \rho^2) \rho^{W+2-z} \rho^{k-l-z} \]
\[ = \rho^{W-k+2} \sum_{l=0}^{k-1} v_l \rho^l (1 - \rho^{2(k-l)}). \]

(4.189)

By replacing (4.189) into (4.188), the MMSE estimation coefficients are

\[ \tilde{h}_k = \frac{\rho^{W-k+2}}{1 - \rho^{2(W+2)}} \sum_{l=0}^{k-1} v_l \rho^l (1 - \rho^{2(k-l)}). \]

(4.190)

It remains to show from (4.190) and (4.185) that \( \tilde{h}_k = h_k \) for any \( k \in \{2, \ldots, W+1\}, \) i.e.,

\[ h_k - \tilde{h}_k = \left( \rho^{k-W-2} + \frac{\rho^{W-k+2}}{1 - \rho^{2(W+2)}} \right) \sum_{l=0}^{k-1} v_l \rho^{-l} - \frac{\rho^{W-k+2}}{1 - \rho^{2(W+2)}} \sum_{l=0}^{k-1} v_l \rho^l \]
\[ = \rho^{-W-1} \left( \frac{1 - \rho^{2(W+1)}}{1 - \rho^{2(W+2)}} \right) \sum_{l=0}^{k-1} \rho^{-k-l-1} - \frac{1 - \rho^{2(W-k+2)}}{1 - \rho^{2(W+1)}} v_l \]
\[ = \rho^{-W-1} \frac{1 - \rho^{2(W+1)}}{1 - \rho^{2(W+2)}} \sum_{l=0}^{k-1} w^*_k v_l = 0, \]

(4.191)

where in the last step \( w^* \) is defined in (4.92) and it follows from (4.180) for \( m = k - 1 \). This completes the proof.

\[ \square \]

Finally we show that the sum-rate constraint in (4.171) coincides with the high resolution lower bound in Corollary 5, and thus is optimal. In particular we want to show that, with the choice of hybrid coding weights in (4.92), we have

\[ \lim_{D \to 0} \left\{ R_{\text{sum}} - \frac{1}{2} \log \left( \frac{1 - \rho^2}{D} \right) \right\} = \lim_{D \to 0} \frac{1}{2(W+1)} \log \left( \frac{1 - \rho^{2(W+2)}}{1 - \rho^{2(W+1)}} \right). \]

(4.192)

We have

\[ \lim_{D \to 0} R_{\text{sum}} = \lim_{D \to 0} \lim_{t \to \infty} \frac{1}{W+1} h([u]_t^{t+1} | [u]_1^{t-B}) - \frac{1}{2} \log (2\pi e D) \]

(4.193)

First note that by similar argument used in (4.173), (4.193) can be written as follows.

\[ \lim_{D \to 0} R_{\text{sum}} = \lim_{D \to 0} \frac{1}{W+1} h([\tilde{u}]_2^{W+2}) - \frac{1}{2} \log (2\pi e D) \]

(4.194)

where \( \tilde{u} \) are defined in (4.174). Now note that

\[ h([\tilde{u}]_2^{W+2}) = I(x_{W+2}; [\tilde{u}]_2^{W+2}) + h([\tilde{u}]_2^{W+2} | x_{W+2}) \]
\[ = h(x_{W+2}) - h(x_{W+2} | [\tilde{u}]_2^{W+2}) + h([\tilde{u}]_2^{W+2} | x_{W+2}) \]

(4.195)
When \( D \to 0 \), the mutual information term in (4.195) approaches to zero according to Lemma 21. Now consider the last term in (4.195), we have

\[
\lim_{D \to 0} h([\hat{u}]_{W+2}^W | x_{W+2}) = \lim_{D \to 0} h([\hat{u}]_{W+1}^W | x_{W+1})
\]

\[
= h([x]_{W+1}^W) - h(x_{W+1}) + \lim_{D \to 0} h(x_{W+1} | [u]_{W+1}^W)
\]

\[
= \frac{1}{2} \log \left( \frac{2\pi e^{W+1}}{1 - \rho^2(W+1)} \right) + \lim_{D \to 0} h(x_{W+1} | [u]_{W+1}^W).
\]  

(4.196)

Note that the second term in (4.195) and the last term in (4.196) cancel each other. Thus (4.195) can be written as

\[
h([\hat{u}]_{W+2}^W) = \frac{1}{2} \log \left( \frac{1 - \rho^2(W+2)}{1 - \rho^2(W+1)} \right) + \frac{W + 1}{2} \log (2\pi e(1 - \rho^2)).
\]  

(4.197)

Finally by replacing (4.197) into (4.194), (4.192) is verified. This completes the proof.

**Proof of Theorem 13**

We separately provide the proof for different parts of Theorem 13.

**Part 1:** We first show that the high resolution excess rate for \( B = W = 1 \) is upper bounded by

\[
R_{E,HR}(\rho, B = 1) \leq \frac{1}{4} \log \left( 1 + \frac{2\rho^4}{(1 + \rho)^2} \right).
\]  

(4.198)

We prove this part through the following steps.

1. Consider the test channel noise \( \sigma^2 = D \) when \( D \to 0 \). From Theorem 11 for any \( w_1 \) in the high resolution regime, the rate \( R \) as

\[
R \geq R_H(D, w_1)
\]

\[
\triangleq \max_{t \to \infty} \lim_{t \to \infty} \left\{ \frac{1}{2} h(u_t, u_{t-1} | u_{t-3}^t), h(u_{t-1} | u_t, [u]_{t-3}^t), h(u_t | u_{t-1}, [u]_{t-3}^t) \right\} - \frac{1}{2} \log(2\pi e D)
\]  

(4.199)

is achievable. Note that the choice of test channel noise satisfies the distortion constraint, i.e.,

\[
\Sigma_H(D, w_1) \triangleq \text{Var} \left( s_t | [u]_{t-3}^t, u_t, u_{t-1} \right) \leq \text{Var} \left( s_t | u_t, u_{t-1} \right) = \text{Var} \left( s_t | u_t + w_1 u_{t-1} \right) = \text{Var} \left( s_t | s_t + z_t \right) = \frac{D}{1+D} \leq D.
\]  

(4.200)
2. Note that in the high resolution regime, the reconstruction of the source sequence is very close to the original source sequences. In particular, while the codewords up to time $t-3$ is observed, the source sequences $[s]_{t-3}$ are known with high resolution. Thus, the rate expression in (4.199) can be rewritten as follows.

$$R_{H}(D, w_1) = \max \left\{ \frac{1}{2} h(\tilde{u}_2, \tilde{u}_3), h(\tilde{u}_2|\tilde{u}_3), h(\tilde{u}_3|\tilde{u}_2) \right\} - \frac{1}{2} \log(2\pi eD)$$

where

$$\begin{pmatrix} \tilde{u}_3 \\ \tilde{u}_2 \end{pmatrix} \triangleq \begin{pmatrix} 1 & -w_1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & \rho^2 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} n_3 \\ n_2 \end{pmatrix}$$

$$= \begin{pmatrix} 1 & \rho - w_1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \rho^2 - w_1 \rho + w_1^2 \\ \rho - w_1 \end{pmatrix}$$

3. First from (4.203) note that for any choice of $w_1 \in [0, \rho]$, we have

$$\text{Var}(\tilde{u}_3) = (1 - \rho^2)(1 + (\rho - w_1)^2 + (\rho^2 - \rho w_1 + w_1^2)^2) \geq (1 - \rho^2)(1 + (\rho - w_1)^2) = \text{Var}(\tilde{u}_2)$$

and therefore,

$$h(\tilde{u}_3|\tilde{u}_2) = h(\tilde{u}_2, \tilde{u}_3) - h(\tilde{u}_2) \geq h(\tilde{u}_2, \tilde{u}_3) - h(\tilde{u}_3) = h(\tilde{u}_2|\tilde{u}_3).$$

Thus (4.201) reduces to the following.

$$R_{H}(D, w_1) = \max \left\{ \frac{1}{2} h(\tilde{u}_2, \tilde{u}_3), h(\tilde{u}_3|\tilde{u}_2) \right\} - \frac{1}{2} \log(2\pi eD)$$

$$= \frac{1}{2} h(\tilde{u}_3|\tilde{u}_2) + \frac{1}{2} \max \left\{ h(\tilde{u}_2), h(\tilde{u}_3|\tilde{u}_2) \right\} - \frac{1}{2} \log(2\pi eD).$$

4. Fig. 4.15 shows an example of terms $h(\tilde{u}_2)$ and $h(\tilde{u}_3|\tilde{u}_2)$ for $\rho = 0.7$ and $w_1 \in [0, \rho]$. Finding the close form expression of the value of $w_1, opt$ at the intersection is not straightforward. We apply the following approximation.

$$h(\tilde{u}_3|\tilde{u}_2) \leq h(\tilde{u}_3) - (\rho - w_1)\tilde{u}_2)$$

$$= 1 + \rho^2 w_1^2.$$
\[ \hat{w}_1 = \rho/(1 + \rho). \] By replacing this value to compute

\[ \frac{1}{2} h(\tilde{u}_2, \tilde{u}_3) |_{w_1=\hat{w}_1} = \frac{1}{2} \begin{pmatrix} 1 & \rho - \hat{w}_1 & \rho^2 - \hat{w}_1 \rho + \hat{w}_1^2 \\ 0 & 1 & \rho - \hat{w}_1 \end{pmatrix} \begin{pmatrix} n_3 \\ n_2 \\ n_1 \end{pmatrix} \]

\[ = \frac{1}{4} \log \left( 2\pi e(1 - \rho^2)^2 \left( 1 + 2 \frac{\rho^4}{(1 + \rho)^2} \right) \right). \]

Thus the rate expression in (4.209) is upper bounded as follows.

\[ R_{H}(D, w_1) \leq \frac{1}{2} \log \left( \frac{1 - \rho^2}{D} \right) + \frac{1}{4} \log \left( 1 + 2 \frac{\rho^4}{(1 + \rho)^2} \right). \]

This completes the proof of part 1.

**Part 2:** Now we assume the case where \( W = 1 \) and \( B \to \infty \). Consider the system at time \( t \) where \( t \to \infty \) and the erasure burst spans the interval \( \{1, \ldots, t - 2\} \). We have,

\[ \begin{pmatrix} u_t \\ u_{t-1} \end{pmatrix} = \begin{pmatrix} 1 & -w_1 & \cdots \\ 0 & 1 & \cdots \end{pmatrix} \begin{pmatrix} s_t \\ s_{t-1} \\ \vdots \end{pmatrix}. \]

Now consider the following lemma.
Lemma 22. For the random variables defined in (4.215), we have

\[ E\{|u_t|^2\} = E\{|u_{t-1}|^2\} = (1 - \rho^2) \left( \frac{\rho^2}{1 - \rho^2} + \frac{1}{1 - w_1^2} \right) \frac{1}{(1 + w_1 \rho)^2} \]

\[ = (1 - \rho^2)f(w_1) \]

\[ E\{u_t u_{t-1}\} = (1 - \rho^2) \left( \rho f(w_1) - \frac{w_1}{(1 + w_1 \rho)(1 - w_1^2)} \right) \]

\[ = (1 - \rho^2)g(w_1) \]

where \( f(.) \) and \( g(.) \) are defined in (4.95) and (4.96), respectively.

The proof of Lemma 22 is provided in Appendix C.8. By application of Lemma 22, the sum-rate constraint is

\[ 2R \geq \frac{1}{2} \log \left( \frac{(2\pi e)^2 \det \begin{pmatrix} E\{|u_t|^2\} & E\{u_t u_{t-1}\} \\ E\{u_t u_{t-1}\} & E\{|u_{t-1}|^2\} \end{pmatrix}}{\log(2\pi e D)} \right) - \log(2\pi e D) \]

\[ = \log \left( \frac{1 - \rho^2}{D} \right) + \frac{1}{2} \log \left( f(w_1)^2 - g(w_1)^2 \right). \]

Now it suffices to show that the sum-rate is indeed the dominant constraint. In particular, note that

\[ h(u_{t-1}|u_t) = h(u_{t-1}, u_t) - h(u_t) \]

\[ = h(u_{t-1}, u_t) - h(u_{t-1}) = h(u_t|u_{t-1}), \]

i.e., the two marginal constraints are the same, and

\[ \frac{1}{2} h(u_{t-1}, u_t) = \frac{h(u_t) + h(u_{t-1}|u_t)}{2} \]

\[ \geq \frac{h(u_{t-1}|u_t) + h(u_{t-1}|u_t)}{2} = h(u_{t-1}|u_t), \]

i.e., the sum-rate constraint dominates the marginal rate constraints. This completes the proof.

4.6 Lower Bound on Rate-Recovery Function

4.6.1 Connection to the Multi-terminal Source Coding Problem

Before stating the general lower bound on \( R(B,W,D) \), we consider a special case of \( B = W = 1 \). For this case, we propose a lower bound by exploiting a connection between the streaming setup and the multi-terminal source coding problem illustrated in Fig. 4.16. The encoder observes two sources \( s^n_j \) and \( s^n_{j-1} \). Decoder \( j \) is required to reconstruct \( s^n_j \) within distortion \( D \) while knowing \( s^n_{j-1} \) whereas decoder \( j + 1 \) requires to reconstruct \( s^n_{j+1} \) within distortion \( D \) while knowing \( s^n_{j-2} \) and having access to the codewords \( \{f_j, f_{j+1}\} \). Decoder \( j \) resembles a steady state decoder when the previous source sequence has been reconstructed whereas decoder \( j + 1 \) resembles the decoder following an erasure and the associated recovery period. The proposed multi-terminal setup is different from the original one in that the decoders are given actual source sequences rather than the encoder output. Nevertheless the study of this model captures one source of tension inherent in the streaming setup. When encoding \( s^n_j \)
we need to simultaneously satisfy two requirements: the sequence \( s_j^n \) must be reconstructed within a distortion of \( D \) at encoder \( j \) and it can also be used as a helper by decoder \( j+1 \). In general these requirements can be conflicting. If we set \( s_{j-2}^n = \phi \) then the setup is reminiscent of zig-zag source coding problem [34].

Of particular interest to us in this section is a lower bound on the sum-rate. In particular we show that for any \( D \in (0, 1 - \rho^2) \),

\[
2R \geq \frac{1}{2} \log \left( \frac{1 - \rho^2}{D} \right) + \frac{1}{2} \log \left( \frac{1 - \rho^6}{D} \right) - \frac{1}{2} \log \left( \frac{1 - \rho^4}{1 - (1 - D)\rho^2} \right) \tag{4.226}
\]

To show (4.226), note that

\[
2nR \geq H(f_j, f_{j+1}) \geq H(f_j, f_{j+1} | s_{j-2}^n) \geq I(f_j, f_{j+1} | s_{j+1}^n, s_{j-2}^n) + H(f_j, f_{j+1} | s_{j-2}^n, s_{j+1}^n) \tag{4.227}
\]

\[
\geq h(s_{j+1}^n | s_{j-2}^n) - h(s_{j+1}^n | f_j, f_{j+1}, s_{j-2}^n) + H(f_j | s_{j-2}^n, s_{j+1}^n) \tag{4.228}
\]

\[
\geq n \frac{1}{2} \log \left( \frac{1 - \rho^6}{D} \right) + H(f_j | s_{j-2}^n, s_{j+1}^n) \tag{4.229}
\]

where (4.229) follows from the fact that \( s_{j+1}^n \) must be reconstructed from \((f_j, f_{j+1}, s_{j-2}^n)\) within distortion \( D \) at decoder \( j+1 \). The first term is the minimum rate associated with decoder \( j+1 \). We next lower bound the second term by using the fact that \( f_j \) must also be used by decoder \( j \).

\[
H(f_j | s_{j-2}^n, s_{j+1}^n) \geq H(f_j | s_{j-2}^n, s_{j-1}^n, s_{j+1}^n) \tag{4.230}
\]

\[
\geq I(f_j; s_j^n | s_{j-2}^n, s_{j-1}^n, s_{j+1}^n) \tag{4.231}
\]

\[
= h(s_j^n | s_{j-1}^n, s_{j+1}^n) - h(s_j^n | s_{j-2}^n, s_{j-1}^n, s_{j+1}^n, f_j) \tag{4.232}
\]

\[
= nh(s_j^n | s_{j-2}^n, s_{j-1}^n, s_{j+1}^n, f_j) \tag{4.233}
\]

\[
\geq n \frac{1}{2} \log \left( \frac{2\pi e (1 - \rho^2)^2}{(1 - \rho^2)} \right) - h(s_j^n | s_{j-2}^n, s_{j-1}^n, s_{j+1}^n, f_j) \tag{4.234}
\]

One direct way to upper bound the last term in (4.234) is to use the fact that \( s_j \) can be reconstructed within distortion \( D \) using \((f_j, s_{j-1})\). Thus by ignoring the fact that \( s_{j+1} \) is also available, one can find
Proof. See Appendix C.9.

However knowing \( s_{j+1} \) can provide an extra observation to improve the estimation of \( s_j \) as well as the upper bound in (4.236). In particular, we can show that

\[
h(s_j^n | s_{j-2}^n, s_{j-1}^n, s_{j+1}^n, f_j) \leq h(s_j^n | s_{j-1}^n, f_j) \leq \frac{n}{2} \log (2\pi e D).
\]

(4.235)

Note that the upper bound in (4.237) is strictly tighter than (4.236), as

\[
h(s_j^n | s_{j-2}^n, s_{j-1}^n, s_{j+1}^n, f_j) \leq \frac{n}{2} \log \left( \frac{D(1-\rho^2)}{1-(1-D)\rho^2} \right).
\]

(4.237)

where the first term in (4.239) follows from the fact that at decoder \( j \), \( s_j^n \) is reconstructed within distortion \( D \) knowing \( \{s_{j-1}^n, f_j\} \) and hence

\[
h(s_j^n | s_{j-2}^n, s_{j-1}^n, f_j) \leq h(s_j^n | s_{j-1}^n, f_j) \leq \frac{n}{2} \log(2\pi e D),
\]

(4.240)

and using the Lemma 23 stated below. Eq. (4.226) follows from (4.229), (4.234) and (4.239).

Lemma 23. Assume \( s_0 \sim \mathcal{N}(0, 1) \) and \( s_0 = \rho^m s_0 + n \) for \( n \sim \mathcal{N}(0, 1-\rho^{2m}) \). Also assume the Markov chain property \( f_a \rightarrow s_a \rightarrow s_b \). If \( h(s_a | f_a) \leq \frac{1}{2} \log(2\pi e r) \), then

\[
h(s_a | f_a) - h(s_b | f_a) \leq \frac{1}{2} \log \left( \frac{r}{1-(1-r)\rho^{2m}} \right)
\]

(4.241)

Proof. See Appendix C.9.

In our original streaming setup, as will become apparent in the following, this bound can be tightened by noting that the side information to the decoders in Fig. 4.16 is actually encoder outputs rather than the true source sequences.

### 4.6.2 General Lower Bound

In order to derive a lower bound on the rate-recovery function in general case, consider the case where the erasure burst of length \( B \) spans the interval \( \{t - B - W, \ldots, t - W - 1\} \) and the decoder is interested
We now separately derive the lower bound for first and second term in (4.248). First consider the first term according to the source model where (4.244) follows from the application of Shannon’s entropy power inequality (EPI) and the fact in reconstructing the source sequence

\[ n(W + 1)R \geq H(\mathcal{f}_t^n|\mathcal{f}_{t-W}^0) \]

\[ \geq H(\mathcal{f}_t^n|\mathcal{f}_{t-W}^0|\mathcal{f}_{t-W}^{t-B-W-1}, s_{n-1}) \]

\[ = I(s^n_1; f_{t-W}^t|f_{t-W}^{t-B-W-1}, s_{n-1}) + H(f_{t-W}^t|s^n_1, f_{t-W}^{t-B-W-1}, s_{n-1}) \]

\[ \geq h(s^n_1|f_{t-W}^{t-B-W-1}, s_{n-1}) - h(s^n_1|f_{t-W}^{t-B-W-1}|f_{t-W}^t, s_{n-1}) + H(f_{t-W}^t|s^n_1, f_{t-W}^{t-B-W-1}, s_{n-1}) \] (4.243)

where (4.242) follows from the fact that conditioning reduces the differential entropy.

The first term in (4.243) can be lower bounded as

\[ h(s^n_1|f_{t-W}^{t-B-W-1}, s_{n-1}) \geq \frac{n}{2} \log \left( \frac{\rho^2(B+W+1)2^{\frac{1}{2}D}h(s^n_{-\rho-W-1}|f_{t-W}^{t-B-W-1}, s_{n-1}) + 2\pi e(1 - \rho^2(B+W+1))}{2^{2R} - \rho^2} \frac{1 - \frac{\rho^2}{2^{2R}}}{2^{2R}} + 2\pi e(1 - \rho^2(B+W+1)) \right) \] (4.244)

where (4.244) follows from the application of Shannon’s entropy power inequality (EPI) and the fact that according to the source model

\[ s^n_t = \rho^B+W+1 s^n_{t-B-W-1} + \tilde{n}^n_t \]

where \( \tilde{n}^n_t \) is i.i.d. drawn from \( \mathcal{N}(0, 1 - \rho^2(B+W+1)) \). Inequality in (4.245) also follows from the application of Lemma 6 in Chapter 3.

The second term in (4.243) is lower bounded based on the fact that the decoder is able to reconstruct the source sequence \( s^n_t \) within distortion \( D \) knowing \( \{f_{t-W}^{t-B-W-1}|f_{t-W}^t, s_{n-1}\} \) and following standard source coding arguments. In fact,

\[ h(s^n_1|f_{t-W}^{t-B-W-1}|f_{t-W}^t, s_{n-1}) \leq \frac{n}{2} \log(2\pi eD) \] (4.246)

Deriving lower bound for the third term in (4.243) is more challenging. First note that revealing the erased codewords can only reduce the entropy term as follows.

\[ H(f_{t-W}^t|s^n_1, f_{t-W}^{t-B-W-1}, s_{n-1}) \geq H(f_{t-W}^{t-1}|s^n_1, f_{t-W}^{t-B-W-1}, s_{n-1}) \]

\[ \geq H(f_{t-W}^{t-1}|s^n_1, f_{t-W}^{t-B-W-1}, s_{n-1}) \]

\[ = I(f_{t-W}^{t-1}; s^n_1|f_{t-W}^{t-B-W-1}, s_{n-1}) + H(f_{t-W}^{t-1}|s^n_1|f_{t-W}^{t-B-W-1}, s_{n-1}) \]

\[ \geq h(s^n_1|f_{t-W}^{t-B-W-1}, s_{n-1}) - h(s^n_1|f_{t-W}^{t-B-W-1}|f_{t-W}^t, s_{n-1}) \] (4.247)

\[ \geq h(s^n_1|f_{t-W}^{t-B-W-1}, s_{n-1}) - h(s^n_1|f_{t-W}^{t-B-W-1}|f_{t-W}^t, s_{n-1}) + nW h(s_t|s_0) - h(s^n_1|f_{t-W}^{t-B-W-1}, s_{n-1}) \] (4.248)

We now separately derive the lower bound for first and second term in (4.248). First consider the first term:

\[ h([s^n_{t-W}]_{t-W}^t|s^n_t, f_{t-W}^{t-W-1}, s_{n-1}^t) = h([s^n_{t-W}]_{t-W}^t|f_{t-W}^{t-W-1}, s_{n-1}^t) - h(s^n_1|f_{t-W}^{t-W-1}, s_{n-1}^t) \]

\[ = h(s^n_{t-W}||f_{t-W}^{t-W-1}, s_{n-1}^t) + nW h(s_t|s_0) - h(s^n_1|f_{t-W}^{t-W-1}, s_{n-1}^t) \] (4.249)
where the second and the third terms in (4.257) follows from the Markov Chain properties.

Now consider the second term in (4.248). We can write

\[ h(s_t^n | f_0^n) = h(s_t^n | f_0^n, s_{-1}^n) \]

\[ = h(s_t^n | f_0^n, s_{-1}^n) - h(s_t^n | f_0^n) \]

\[ \geq \frac{n}{2} \log \left( \frac{2\pi e(1 - \rho^2) \rho^2}{2^R - \rho^2} \right) \]

\[ + nW \left( h(s_0^n) - \frac{n}{2} \log \left( 2\pi e(1 - D) \rho^{2(W + 1)} \right) \right) \]

(4.251)

Note that

- The first term in (4.251) follows from the following inequalities.

\[ h(s_t^n | f_0^n, s_{-1}^n) \geq \frac{n}{2} \log \left( \frac{2\pi e(1 - \rho^2) \rho^2}{2^R - \rho^2} \right) \]

\[ + nW \left( h(s_0^n) - \frac{n}{2} \log \left( 2\pi e(1 - D) \rho^{2(W + 1)} \right) \right) \]

(4.252)

where (4.252) follows from Shannon’s entropy power inequality (EPI) and (4.253) follows from the application of Lemma 6 in Chapter 3 for \( k = t - W - 1 \).

- The third term in (4.251) is based on the following.

\[ h(s_t^n | f_0^n, s_{-1}^n) \leq h(s_t^n - h(s_t^n | f_0^n, s_{-1}^n)) \]

\[ \leq \frac{n}{2} \log \left( 2\pi e(1 - D) \rho^{2(W + 1)} \right) + \frac{n}{2} \log \left( 2\pi e(1 - D) \rho^{2(W + 1)} \right) \]

(4.255)

(4.256)

where (4.254) follows from the fact that knowing \( \{f_0^n, s_{-1}^n\} \) the decoder is able to reproduce an estimate of \( s_t \) as

\[ \hat{s}_t^n (f_0^n, s_{-1}^n) = \rho^{W + 1} s_{t-W-1} (f_0^n, s_{-1}^n) + \tilde{n} \]

where \( \tilde{n} \sim N(0, 1 - \rho^{2(W + 1)}) \). (4.255) also follows from the fact that the Gaussian distribution has the largest differential entropy.

Now consider the second term in (4.248). We can write

\[ q(W) \triangleq h(s_t^n | f_0^n) + h(s_{-1}^n | f_0^n) - h(s_t^n | f_0^n) \]

\[ = h(s_t^n | f_0^n) + h(s_{-1}^n | f_0^n) - h(s_t^n | f_0^n) \]

\[ \leq \frac{n}{2} \log \left( \frac{D}{1 - (1 - D) \rho^2} \right) + nW(s_0^n) + q(W - 1) \]

(4.258)

where the second and the third terms in (4.257) follows from the Markov Chain properties

\[ \{f_0^n, s_{-1}^n\} \Rightarrow s_t^n \Rightarrow s_{t-W}^n \]

(4.259)

and

\[ s_t^n \Rightarrow \{s_{t-W}^n, f_0^n\} \Rightarrow \{s_{t-W}^n, f_0^n\} \]

(4.260)
Inequality of (4.258) also follows from the application of the Lemma 23.

By repeating the same steps in (4.258) for \( W \) times we have

\[
q(W) \leq \frac{nW}{2} \log \left( \frac{D}{1 - (1 - D)\rho^2} \right) + nWh(s_1|s_0) + q(0)
\]

(4.261)

where \( q(0) = 0 \).

Now note that based on (4.245) and (4.251) our tightest lower bound happens when \( t \to \infty \). Based on this fact, by replacing (4.251) and (4.261) into (4.248) and then replacing the resulting term as well as (4.245) and (4.246) into (4.243) the following lower bound is derived.

\[
(W + 1)R \geq \frac{1}{2} \log \left( \frac{(1 - \rho^2)^2(2R - 2\rho^2)}{D} + 1 - \rho^2 \right) + \frac{1}{2} \log \left( \frac{1 - \rho^2}{1 - \rho^2} \right) \left( \frac{1 - (1 - D)\rho^2}{D} \right)^{W}
\]

(4.262)

By defining \( x \triangleq 2^R \), (4.262) is equivalent to the polynomial in (4.102).

It remains to show that (4.102) has a unique solution for \( x \geq 1 \). We need this in two steps as follows.

• First we note that the second derivative of the polynomial \( p(x) \) defined in (4.102) is strictly positive for \( x \geq 1 \). In particular,

\[
p''(x) = x^{W-2} \left((W + 2)(W + 1)x^2 - 2\rho^2W(W + 1)x + \rho^4W(W - 1)\right) > 0 \quad \text{for } x \geq 1.
\]

(4.263)

• Second we note that \( \lim_{x \to \infty} p(x) > 0 \) and \( p(1) \leq 0 \). In particular,

\[
p(1) = (1 - \rho^2) (1 - \rho^2 - \alpha_1) \\
\leq (1 - \rho^2) (1 - \rho^2 - \alpha_1) \leq 0
\]

(4.264)

where \( \alpha_1 \) is \( \alpha \) evaluated at \( D = 1 \) and (4.264) follows from the fact that \( \alpha \) is a decreasing function of \( D \) and \( \alpha_1 = 1 - \rho^2 \).

Thus, the polynomial \( p(x) \) has a unique root \( x^* \geq 1 \) which is of particular interest to us. This completes the proof.

### 4.6.3 Lower Bound in the High Resolution Regime

We consider the lower bound in of Theorem 14 in the limit \( D \to 0 \). In particular, in order to prove Corollary 5, we need to show that as \( D \to 0 \), the root of the polynomial \( p(x) \) approaches to

\[
x = \frac{1 - \rho^2}{D} \sqrt{\frac{1 - \rho^2(2W + 1)}{1 - \rho^2(W + 1)}}.
\]

(4.265)

First note that the parameter \( \alpha(\rho, D) \) at \( D \to 0 \), behaves as follows.

\[
\alpha(\rho, D) \to \frac{1}{1 - \rho^2(W + 1)} \left( \frac{1 - \rho^2}{D} \right)^{W+1},
\]

(4.266)
which becomes very large as $D \to 0$. From the definition of $p(x)$ it is not hard to observe that $\alpha(\rho, D) \to \infty$ requires $x \to \infty$. Thus, in order to have $p(x) = 0$ in this asymptotic regime, we require

$$x W^2 (x - \rho^2)^2 = \frac{1 - \rho^2(B + W + 1)}{1 - \rho^2(W + 1)} \left( \frac{1 - \rho^2}{D} \right)^{W+1} x + \frac{\rho^2 - \rho^2(B + W + 1)}{1 - \rho^2(W + 1)} \left( \frac{1 - \rho^2}{D} \right)^{W+1},$$  \hspace{0.5cm} (4.267)

which as $x \to \infty$, reduces to

$$x^{W+2} = \frac{1 - \rho^2(B + W + 1)}{1 - \rho^2(W + 1)} \left( \frac{1 - \rho^2}{D} \right)^{W+1} x,$$  \hspace{0.5cm} (4.268)

which results in (4.265). This completes the proof.

### 4.7 Conclusion

In this chapter we considered the zero-delay streaming of Gauss-Markov sources over erasure burst channels. We introduced a general zero-delay streaming coding scheme which combines the two conventional techniques of successive quantization and random binning. We first showed that the classical predictive coding and the Q-and-B scheme presented in Chapter 3 are in fact special cases of this proposed general coding scheme. Furthermore, we showed that in case of erasure-free channels all the coding schemes attain the optimal rate-distortion performance. In the case of a erasure burst channel, however, we studied the attainable rate by predictive coding and Q-and-B schemes. We observed that each of the two schemes outperforms the other for some range of source parameters. Based on this observation, we proposed hybrid coding scheme which outperforms the two coding scheme by exploiting the benefits of each. Several properties of this hybrid coding scheme were studied. By providing a lower bound on the rate-recovery function, the optimality of this scheme was established for some range of source parameters.

This completes the study of the zero-delay streaming setup. In the next chapter, i.e., Chapter 5 we treat the streaming problem with delay-constrained decoders for both discrete sources and lossless recovery constraint, and Gauss-Markov sources and quadratic distortion measure.
Chapter 5

Delay-Constrained Streaming

5.1 Introduction

In this Chapter we study the sequential transmission of stationary first-order Markov source sequences over burst erasure channel with the delay-constrained decoder. The spatially i.i.d. and temporally first-order Markov source process is causally observed by the rate-$R$ encoder whose outputs are transmitted through the burst erasure channel. The channel introduces a single erasure burst of length up to $B$ spanning an interval unknown to the encoder and perfectly delivers the rest of the codewords to the destination. Two delay-constrained streaming setups are considered to recover the source sequences upon observing the channel outputs with a delay of $T$. In the controlled-interruption setup, the decoder is not required to recover the source sequences for the time when the channel introduces erasures and a waiting window of length $W$ after the erasure burst ends. In the ideal-playback setup however the decoder recovers all the source sequences within the delay of $T$. For each setup, we provide the lower and upper bounds on the minimum required compression rate.

In case of lossless streaming of discrete sources the upper and lower bounds coincide for ideal-playback setup and for controlled-interruption setup in two cases: i) $W = 0$ and ii) either $W$ or $T$ becomes very large.

For the lossy streaming of Gauss-Markov sources with quadratic distortion constraint when $W = 0$, the upper and lower bounds coincide in high resolution and large delay asymptotic for ideal-playback setup and in high resolution asymptotic for controlled-interruption setup.

The remainder of the chapter is organized as follows. The problem setup is described in Section 5.2 and a summary of main results is provided in Section 5.3. We treat the case of discrete sources with lossless recovery and Gaussian sources with lossy reconstruction for the case of controlled-interruption in Section 5.4 and Section 5.7, respectively where we establish upper and lower bounds on the minimum rate. We consider the case of streaming with ideal-playback for lossless recovery of discrete sources and
Figure 5.1: Problem Setup: Consider the example of \( B = 3 \) and \( W = T = 2 \). The encoder output \( f_j \) is a function of the source sequences up to time \( j \), i.e., \( s^n_{-1}, s^n_0, \ldots, s^n_j \). The channel introduces an erasure burst of length \( B \). The decoder produces \( \hat{s}^n_j \) upon observing the sequence \( \{g_0, g_1, \ldots, g_{j+T}\} \). a) In Controlled-Interruption, the decoder is not required to produce those source sequences that fall in a window of length \( B + W \) following the start of an erasure burst. However, the decoder recovers the rest of the source sequences within a delay of \( T \). b) In Ideal-Playback the decoder recovers all the source sequences within a delay of \( T \).

lossy reconstruction of the Gaussian sources in Section 5.7 and Section 5.7, respectively. Section 5.9 concludes the chapter.

5.2 Problem Statement

In this section we describe the source and channel models as well as our notion of an error-propagation window, delay-constrained decoder and the associated rate functions.

We consider a semi-infinite stationary vector source process \( \{s^n_i\}_{i \geq -1} \) whose symbols (defined over alphabet \( S \)) are drawn independently across the spatial dimension and from a first-order Markov chain across the temporal dimension:

\[
\Pr( s^n_i = s^n_i | s^n_{i-1}, s^n_{i-2}, \ldots ) = \prod_{j=1}^{n} p_{s_i|s_{i-j}}(s_{i-j} | s_{i-1,j}), \quad \forall i \geq 0.
\] (5.1)

We assume that the underlying random variables \( \{s_i\} \) constitute a time-invariant, stationary and a first-
order Markov chain with a common marginal distribution denoted by $p_s(\cdot)$. Such models are used in earlier works on sequential source coding. See e.g., [17]. We assume that the source sequence $s_{n-1}$ is revealed to both the encoder and decoder before the communication starts. This plays the role of a synchronization frame.

A rate-$R$ encoder maps the sequence $\{s^n_i\}_{i \geq -1}$ to an index $f_i \in \{1, 2, \ldots, 2^{nR}\}$ according to some function

$$f_i = F_i(s_{n-1}, s^n_i, \ldots, s^n_i)$$

for each $i \geq 0$.

The channel introduces an erasure burst of size $B$, i.e., for some particular $j \geq 0$, it introduces an erasure burst such that

$$g_i = \begin{cases} \star, & i \in \{j, j+1, \ldots, j+B-1\} \\ f_i, & \text{else.} \end{cases}$$

We consider a communication duration of $\Upsilon$ and two notions of delay-constraint decoder.

### 5.2.1 Streaming with Controlled-Interruption

As shown in Fig. 5.1(a), the delay-constrained decoder in this model, upon observing the sequence $\{g_i\}_{i \geq 0}$, is required to recover all the source sequences using decoding functions

$$\hat{s}^n_i = G_i(s^n_{i-1}, g_0, g_1, \ldots, g_{i+T}), \quad i \notin \{j, \ldots, j+B+W-1\}.$$ 

where $j$ denotes the time at which the erasure burst starts in (5.3). It is however not required to produce the source sequences in the window of length $B+W$ following the start of an erasure burst. In fact the decoder in case of erasure burst, freezes the last recovered frame and gives up on streaming of $B+W$ consecutive frames. After this error propagation window, it starts streaming the rest of the source sequences. We define the rate function under lossless and lossy reconstruction constraints.

#### Lossless Case

We first consider the case where the source alphabet $S$ is finite and the reconstruction at the decoder in (5.4), is lossless. A rate $R$ is feasible if there exists a sequence of encoding and decoding functions and a sequence $\epsilon_n$ that approaches zero as $n \to \infty$ such that, $\Pr(s^n_i \neq \hat{s}^n_i) \leq \epsilon_n$ for all $i \notin \{j, \ldots, j+B+W-1\}$. We seek the minimum feasible rate denoted by $R_{CI}(B,W,T)$.

#### Lossy Case

We also consider the case where the source alphabet is real numbers, i.e., $S = \mathbb{R}$, and reconstruction in (5.4) is required to satisfy an average distortion constraint:

$$\limsup_{n \to \infty} E \left[ \frac{1}{n} \sum_{k=1}^{n} d(s_{i,k}, \hat{s}_{i,k}) \right] \leq D$$

(5.5)
for some distortion measure $d : \mathbb{R}^2 \to [0, \infty)$. The rate $R$ is feasible if a sequence of encoding and decoding functions exists that satisfies the average distortion constraint. The minimum feasible rate is denoted by $R_{CI}(B, W, T, D)$. In this chapter we will focus on the class of Gaussian-Markov sources, with quadratic distortion measure, i.e., $d(s, \hat{s}) = (s - \hat{s})^2$, where the analysis simplifies.

5.2.2 Streaming with Ideal-Playback

As shown in Fig. 5.1(b), the delay-constrained decoder in this model is required to recover all the source sequences using decoding functions

$$s^n_i = G_i(s^n_{i-1}, g_0, g_1, \ldots, g_i + T), \quad i \geq 0,$$

where $j$ denotes the time at which the erasure burst starts in (5.3). In fact the decoder is required to reconstruct all the source sequences within a delay of $T$ even in the case of channel erasure bursts. We study the minimum rate required for the two cases of lossless and lossy reconstruction constraints. It seems reasonable to only consider the ideal-playback streaming in case of $T \geq B$, i.e., the minimum delay is not less than the burst length.

**Lossless Case**

In this case the decoder is interested in lossless recovery of the source sequences. A rate $R$ is feasible if there exists a sequence of encoding and decoding functions and a sequence $\epsilon_n$ that approaches zero as $n \to \infty$ such that, $\Pr(s^n_i \neq \hat{s}^n_i) \leq \epsilon_n$ for all $i \geq 0$. We seek the minimum feasible rate $R_{IP}(B, T)$.

**Lossy Case**

Similar to lossy streaming with controlled-interruption, we consider the case where reconstruction in (5.6) is required to satisfy an average distortion constraint in (5.5). The rate $R$ is feasible if a sequence of encoding and decoding functions exists that satisfies the average distortion constraint. Define the minimum feasible rate as $R_{IP}(B, T, D)$. We will focus on the class of Gaussian-Markov sources, with quadratic distortion measure.

5.3 Main Results

In this section we present the main results of the chapter. We note in advance that throughout the chapter, the upper bound on rate function indicates the rate achievable by a proposed coding scheme and the lower bound corresponds to a necessary condition that the rate of any feasible coding scheme has to satisfy.

5.3.1 Streaming with Controlled-Interruption

In this section we present the results of streaming with controlled-interruption for two cases of lossless and lossy recovery at the decoder.
Lossless Case

**Single erasure burst Channel:** We consider the case where the channel introduces an isolated erasure burst of length up to $B$ during the transmission duration as described in (5.3). The following theorem characterizes the upper and lower bounds on lossless rate function.

**Theorem 15.** The rate of the lossless streaming of discrete Markov sources with controlled-interruption satisfies

$$R_{CI}^-(B, W, T) \leq R_{CI}(B, W, T) \leq R_{CI}^+(B, W, T)$$

where

$$R_{CI}^-(B, W, T) = H(s_1|s_0) + \frac{1}{W + T + 1} I(s_{B:B+W+1}|s_0)$$

(5.7)

$$R_{CI}^+(B, W, T) = H(s_1|s_0) + \frac{1}{W + T + 1} I(s_{B:B+1}|s_0)$$

(5.8)

□

It can be observed from Theorem 15 that both the upper and lower bounds consist of a term equal to the predictive coding rate plus another mutual information term inversely scaled by $(W + T + 1)$. The proof of Theorem 15 is provided in Sec. 5.4. The lower bound is based on the idea of considering a periodic erasure burst channel rather than single erasure burst channel. The upper bound is achieved by random-binning coding scheme and Slepian-Wolf decoding. The following proposition makes the rate expression more explicit.

**Corollary 6.** For any first order Markov source process defined, the upper bound in (5.8) can also be expressed as

$$R_{CI}^+(B, W, T) = \frac{H(s_{B+1}, s_{B+2}, \ldots, s_{B+W+T+1}|s_0)}{W + T + 1}$$

(5.9)

□

The proof of Corollary 6 is provided in Appendix D.1. Note that the upper and lower bounds of Theorem 15 coincide for some special cases discussed below.

- When $W = 0$, i.e., when the decoder is interested in recovering all the source sequences corresponding to non-erased codewords, the lossless rate function is

$$R_{CI}(B, W = 0, T) = H(s_1|s_0) + \frac{1}{T + 1} I(s_{B+1}|s_0)$$

(5.10)

$$= \frac{1}{T + 1} H(s_{B+1}, s_{B+2}, \ldots, s_{B+T+1}|s_0).$$

(5.11)

- When each or both of the variables $W$ and $T$ become very large, i.e., $W$ or $T \to \infty$, the lossless rate function reduces to the rate required for predictive coding.

Note also that Theorem 15 can be viewed as a generalization of the zero-delay results of Theorem 1 in Chapter 2 as the upper and lower bounds when $T = 0$ reduce to

$$R_{CI}^-(B, W, T = 0) = R^-(B, W)$$

$$\triangleq H(s_1|s_0) + \frac{1}{W + 1} I(s_{B:B+W+1}|s_0)$$

(5.12)
Figure 5.2: Sliding-window erasure burst channel model. The channel introduces multiple erasure bursts each of length up to $B$. The consecutive erasure bursts are separated by a guard interval of length at least $G$.

$$R^+(B, W, T = 0) = R^+(B, W)$$

$$\triangleq H(s_1|s_0) + \frac{1}{W + 1} I(s_B; s_{B+1} | s_0).$$  (5.13)

**Remark 13.** Even though we consider a single isolated erasure burst in (5.3), the results of the discrete sources and lossless recovery immediately apply when the channel introduces multiple bursts with a guard spacing of at least $W + T + 1$. The upper and lower bound expressions also hold for such a channel model.

**Sliding-Window erasure burst Channel:** In order to investigate the effect of channels with multiple erasures, we consider the sliding-window erasure burst channel model. In this model, which is illustrated in Fig. 5.2, the channel can introduce multiple erasure bursts each of length up to $B$ during the transmission period, however there is a guaranteed guard interval of length at least $G$ between each consecutive erasure bursts. The rest of the setup is similar to single erasure case. Note that in our setting $G > W$, i.e., the guard between the erasures has to be larger than the waiting non-recovery period. The following corollary characterizes the upper and lower bounds on minimum rate function for sliding-window erasure burst channel model denoted as $R_{CI,ME}(B, W, G, T)$.

**Corollary 7.** The rate of the lossless streaming of discrete Markov sources with controlled-interruption delay-constrained decoders over sliding-Window erasure burst channel satisfies

$$R^-_{CI,ME}(B, W, G, T) \leq R_{CI,ME}(B, W, G, T) \leq R^+_{CI,ME}(B, W, G, T)$$  (5.14)

where

$$R^-_{CI,ME}(B, W, G, T) \triangleq H(s_1|s_0) + \frac{1}{\min\{G, T + W + 1\}} I(s_B; s_{B+W+1} | s_0)$$  (5.15)

$$R^+_{CI,ME}(B, W, G, T) \triangleq H(s_1|s_0) + \frac{1}{\min\{G, T + W + 1\}} I(s_B; s_{B+1} | s_0)$$  (5.16)

The proof of Corollary 7 is provided in Section 5.5. It can be observed from Theorem 7 that for $T \leq G - W - 1$, the results of Theorem 15 for minimum rate function of single erasure burst channel model also hold for the sliding-window erasure burst model. The main intuition behind this fact is that as soon as the decoder recovers the source sequences at a specific time, because of the Markov property of the source model, it becomes oblivious to the erasure bursts that happened in the past. Thus it treats the new erasure burst as a single erasure burst as if there has been no previous erasures. On the other hand when $T \geq G - W - 1$ our lower and upper bounds in Theorem 7 surprisingly does not depend on the delay parameter $T$. The upper bound is based on random binning scheme and interestingly reveals
that if $T > G - W - 1$ there is no benefit of delay more than $G - W - 1$. In other words, the best rate-performance is achieved by restricting the decoder to perform within the delay of $G - W - T$ which is strictly lower than $T$.

**Lossy Case (Gauss-Markov Sources)**

We study the lossy rate function when $\{s^n\}$ is sampled i.i.d. from a zero-mean Gaussian distribution, $\mathcal{N}(0, \sigma^2_s)$, along the spatial dimension and forms a first-order Markov chain across the temporal dimension, i.e.,

$$s_i = \rho s_{i-1} + n_i$$

(5.17)

where $\rho \in (0, 1)$ and $n_i \sim \mathcal{N}(0, \sigma^2_n(1-\rho^2))$. Without loss of generality we assume $\sigma^2_s = 1$. We consider the quadratic distortion measure $d(s_i, \hat{s}_i) = (s_i - \hat{s}_i)^2$ between the source symbol $s_i$ and its reconstruction $\hat{s}_i$.

In this chapter we focus on the special case of $W = 0$, where the reconstruction must begin immediately after the erasure burst. Define $R_{\text{CI,GM}}(B, T, D)$ as the lossy rate function with delay-constrained decoder for Gauss-Markov sources.

**Remark 14.** Unlike the lossless case, the results of Gauss-Markov sources for single erasure burst channels do not readily extend to the multiple erasure bursts case. In Chapter 2, in addition to the single erasure burst channel model, we presented the results for the channels with multiple erasure bursts with a guaranteed guard interval of $G$ between the consecutive bursts. We showed that even for guard values as small as $G = 4$, because of the exponentially decaying factor of the autocorrelation among the sources $\rho$, the achievable rate-distortion approaches very close to the single burst case. Thus, in this chapter, we do not consider the extension of the results to the case of channels with multiple erasures.

**Proposition 5.** The rate of the lossy streaming of Gauss-Markov sources with quadratic distortion and delay-constrained decoders with controlled-interruption, satisfies

$$R_{\text{CI,GM}}(B, T, D) \geq R_{\text{CI,GM}}^+(B, T, D) \triangleq \frac{1}{2(T+1)} \log \left( \frac{(1-\rho^2(1+2B))(1-\rho^2)^T}{D^{T+1}} \right)^+$$

(5.18)

The proof of Prop. 5 is provided in Section 5.6.1. The proof is based on considering a periodic erasure burst channel similar to the lossless case. The following proposition, characterizes the upper bound on the rate.

**Proposition 6.** The rate of the lossy streaming of Gauss-Markov sources with quadratic distortion with controlled-interruption, satisfies

$$R_{\text{CI,GM}}(B, T, D) \leq R_{\text{CI,GM}}^-(B, T, D) \triangleq \frac{1}{T+1} I([s]_{0+T}^B; [u]_{0+T}^B|\tilde{s}_0)$$

(5.19)

where for each $i$, $u_i = s_i + z_i$ and $z_i \sim \mathcal{N}(0, \sigma^2_z)$. Also

$$\tilde{s}_i = s_i + e$$
Figure 5.3: Upper and lower bounds on the delay-constrained streaming rate function for Gauss-Markov with correlation $\rho$ sources over erasure burst channel of maximum burst length $B$ and delay $T$. Both ideal-playback and controlled-interruption streaming rates are shown.
and \( e \sim \mathcal{N}(0, \Sigma(\sigma_z^2)/(1 - \Sigma(\sigma_z^2))) \) with

\[
\Sigma(\sigma_z^2) \triangleq 1 \frac{1}{2} \sqrt{(1 - \sigma_z^2)^2(1 - \rho^2)^2 + 4\sigma_z^2(1 - \rho^2) + \frac{1}{2}(1 - \sigma_z^2)}, \quad (5.20)
\]

is independent of all other random variables. The test channel noise \( \sigma_z^2 > 0 \) is chosen to satisfy the distortion constraint

\[
\max \left\{ \mathbb{E}[(s_t - \hat{s}_1)^2], \mathbb{E}[(s_t - \hat{s}_2)^2] \right\} \leq D, \quad (5.21)
\]

where \( \hat{s}_1 \) and \( \hat{s}_2 \) denote the MMSE estimate of \( s_t \) from \( \{\tilde{s}_t, u_t\} \) and \( \{\tilde{s}_{t-B}, [u_{t-B}]^T\} \), respectively, where \( \tilde{s}_{t-B} = s_{t-B} + e \).

The proof of Prop. 6 is provided in Section 5.6.2. The coding scheme is in fact based on quantization of the source sequence of each time through the Gaussian test channel and binning the generated quantization codewords at the encoder, and recovering the quantization codewords and performing minimum mean square error (MMSE) estimation at the decoder. The following corollary whose proof is provided in Section 5.6.3, characterizes the high resolution behavior of the rate function.

**Corollary 8.** In the high resolution regime when \( D \to 0 \), the rate of the lossy streaming of Gauss-Markov sources with controlled-interruption, satisfies

\[
R_{CI,GM}(B, T, D \to 0) = \frac{1}{2(T + 1)} \log \left( \frac{(1 - \rho^2(B+1))(1 - \rho^2)^T}{DT+1} \right) + o(D) \quad (5.22)
\]

where \( \lim_{D \to 0} o(D) = 0 \).

**5.3.2 Streaming with Ideal-Playback**

In this section we present the results of streaming with ideal-playback for two cases of lossless and lossy recovery at the decoder.

**Lossless Case**

The following theorem establishes the optimal rate function in case of lossless recovery.

**Theorem 16.** The rate of the lossless streaming of discrete Markov sources with ideal-playback, when \( T \geq B \), satisfies

\[
R_{IP}(B, T) = \frac{B + T}{T} H(s_1|s_0). \quad (5.23)
\]

The proof of Theorem 16 is provided in Section 5.7. The converse proof is derived by using the technique of periodic erasure channel, similar to the controlled-interruption case. The achievability is derived by the separation of source-channel coding. The source code is the optimal predictive coding and the channel code is delay-optimal code for erasure burst channels.
Lossy Case (Gauss-Markov Source)

The following propositions establish the lower and upper bounds on the rate function in case of lossy recovery.

**Proposition 7.** The rate of the lossy streaming of Gauss-Markov sources with quadratic distortion and ideal-playback, satisfies

\[ R_{IP,GM}(B, T, D) \geq \tilde{R}_{IP,GM}(B, T, D) \]

\[ \tilde{R}_{IP,GM}(B, T, D) \triangleq \max \left( \tilde{R}_{IP,GM}^{-}(B, T, D), R_{CI,GM}^{-}(B, T, D) \right) , \quad (5.24) \]

where

\[ \tilde{R}_{IP,GM}^{-}(B, T, D) \triangleq \begin{cases} \frac{B+T}{2T} \log \left( \frac{1-\rho^2}{D} \right) , & T > B \\ \frac{1}{2} \log(x^*) , & T = B \end{cases} , \quad (5.25) \]

where \( x^* \) is the unique root of the following polynomial equation belonging to the interval \([1, \infty)\).

\[ x^T - \rho^2 x^{T-1} = \left( \frac{1-\rho^2}{D} \right)^2 B . \quad (5.26) \]

The proof of Prop. 7 is provided in Section 5.8.1. We use the fact that the lower bound derived for the rate of streaming with controlled-interruption is also a lower bound for the rate of streaming with perfect-playback. We also establish a separate lower bound in Section 5.8.1, based on a similar approach used in the proof of Theorem 16 and by assuming a periodic erasure channel. As presented in Section 5.8.1, when \( T = B \), we are able to further improve the lower bound. By combining these two lower bounds, the lower bound in Prop. 7 is derived. It can be shown that for a fixed \( B, T \) and \( \rho \), the first constraint in (5.25) is active for \( D \leq d \) for some \( d \geq 0 \) and the second term, i.e., \( R_{CI,GM}^{-}(B, T, D) \) is active for \( D \geq d \).

**Proposition 8.** For the lossy streaming of Gauss-Markov sources with ideal-playback, when \( T \geq B \), any
rate $R$ satisfying

$$R \geq \frac{B + T}{2T} \log \left( \frac{1 - \rho^2}{D} + \rho^2 \right)$$

is achievable. In addition when $T > 1$, for any $D \leq (1 - \rho)/(1 + \rho)$, the rate

$$R \geq \frac{B + T - 1}{2(T - 1)} \log \left( \frac{1 - \rho^2}{D} \right)$$

is achievable.

The proof of Prop. 8 is provided in Section 5.8.2. The proof consists of proposing two different coding schemes as illustrated in Fig. 5.4. Coding scheme I, first applies a zero-delay predictive coding as the source code and then the delay-optimal channel code [48] over the source codes. Coding scheme II, however, when $T > 1$ and $D \leq (1 - \rho)/(1 + \rho)$, first applies the optimal source code over the source sequences. Interestingly, as previously shown in [17], the optimal source code in this case only requires a delay of $T_1 = 1$. This allows a delay budget of $T_2 = T - 1$ for the channel code which is again based on the delay-optimal channel coding. In Section 5.8.2, we provide an alternative proof for sufficiency of one time delay for the optimal source code and rigorously derive the distortion requirement, i.e., $D \leq (1 - \rho)/(1 + \rho)$, for the optimal source code.

The following corollaries specialize the results of Prop. 8 for the two asymptotic cases of high resolution and large decoding delay. In particular, they establish the optimality of the coding scheme I in high resolution and the coding scheme II in the limit of large delay. The proofs are straight-forward and omitted here.

**Corollary 9.** In high resolution regime when $D \to 0$, the optimal streaming rate of a Gauss-Markov source over erasure burst channel with ideal-playback, i.e.,

$$R_{IP,GM}(B, T, D \to 0) = \frac{B + T}{2T} \log \left( \frac{1 - \rho^2}{D} \right) + o(D).$$

is achieved by the coding scheme I.

**Corollary 10.** In high delay regime when $T \to \infty$, the minimum streaming rate of a Gauss-Markov source over erasure burst channel with ideal-playback, when $D \leq (1 - \rho)/(1 + \rho)$, i.e.,

$$\lim_{T \to \infty} R_{IP,GM}(B, T, D) = \frac{1}{2} \log \left( \frac{1 - \rho^2}{D} \right).$$

is achieved by the coding scheme II.

Fig. 5.3 shows the upper and lower bounds on the rate function of streaming of Gauss-Markov sources with delay-constrained decoders for both controlled-interruption and ideal-playback cases.

This completes the statement of the main results in this chapter.
The basic idea behind the converse is illustrated in Fig. 5.5. We consider a periodic erasure channel with period \( P = B + W + T + 1 \). The \( k \)-th period, for \( k \geq 1 \), spans the interval \( \{(k - 1)(B + W + T + 1) + 1, \ldots, k(B + W + T + 1)\} \). In each period the first \( B \) packets are erased, whereas the remaining \( T + W + 1 \) packets \( \{(k - 1)P + B + 1, \ldots, kP\} \) are not erased. For sake of compactness we denote the \( n \)-letter sequence \( s^n \) by \( s \), i.e., using the bold-face font.

The idea behind considering a periodic erasure channel is that when the decoder recovers the source sequence because of the Markov property of the source process, it becomes oblivious to the past erasures. Thus, we expect that it can accept a new erasure bursts as if it is the only erasure burst during the whole transmission period.

Based on this observation, we now derive the lower bound on the rate function in (5.7) as follows. To this end, we consider \( N \) periods of the periodic erasure channel explained before. Rate \( R \) should satisfy the following constraint.

\[
N(P-B)nR \geq H([f]_B^P;[f]_{P+B+1}^{2P};[f]_{2P+B+1}^{3P};\ldots,[f]_{(N-1)P+B+1}^{NP})
\]
where (5.31) follows from the fact that conditioning reduces the entropy. We provide the proof of the lower bound in four steps.

**Step 1:** First consider the first period in Fig. 5.5. According to Fano’s inequality and based on the fact that $s_{p-T}$ can be recovered from $\{f_0, [f]_{B+1}^P\}$, we can write

$$H(s_{p-T}|f_0, [f]_{B+1}^P) \leq n\epsilon_n. \quad (5.32)$$

Using this, the entropy term in (5.31) can be lower bounded as follows.

$$H([f]_{B+1}^P, [f]_{2B+2}^{2P}, \ldots, [f]_{(N-1)P+B+1}^{NP}|s_0)
\geq H(s_{p-T}, [f]_{B+1}^P, [f]_{2B+2}^{2P}, \ldots, [f]_{(N-1)P+B+1}^{NP}|s_0) - n\epsilon_n \quad (5.33)$$

$$= nH(s_{p-T}|s_0) + H([f]_{B+1}^P, [f]_{2B+2}^{2P}, \ldots, [f]_{(N-1)P+B+1}^{NP}|s_0, s_{p-T}, [f]_B^P) - n\epsilon_n. \quad (5.34)$$

where (5.33) follows from (5.32) and the first term in (5.34) follows from the properties of the source sequences.

**Step 2:** In this step, based on the fact that conditioning never increases the entropy, we further lower bound the second term in (5.34) by revealing the erased codewords as follows.

$$H([f]_{B+1}^P, [f]_{2B+2}^{2P}, \ldots, [f]_{(N-1)P+B+1}^{NP}|s_0, s_{p-T})
\geq H([f]_{B+1}^P, [f]_{2B+2}^{2P}, \ldots, [f]_{(N-1)P+B+1}^{NP}|s_0^B, s_{p-T}, [f]_B^P). \quad (5.35)$$

After revealing the erased codewords of the first period, the source sequences in the interval $\{B+1, \ldots, P-T-1\}$ can be recovered. Thus the following inequality holds.

$$H([s]_{B+1}^{P-T-1}|[f]_0^{P-1}) \leq nW\epsilon_n. \quad (5.36)$$

Now the entropy term in (5.35) can be written as.

$$H([f]_{B+1}^P, \ldots, [f]_{(N-1)P+B+1}^{NP}|s_0^B, s_{p-T}, [f]_B^P)
\geq H([s]_{B+1}^{P-T-1}, [f]_{B+1}^P, [f]_{2B+2}^{2P}, \ldots, [f]_{(N-1)P+B+1}^{NP}|s_0^B, s_{p-T}, [f]_B^P) - nW\epsilon_n \quad (5.37)$$

$$\geq H([s]_{B+1}^{P-T-1}|s_B, s_{p-T}) - nW\epsilon_n + H([f]_{P-T+1}^P, \ldots, [f]_{(N-1)P+B+1}^{NP}|s_0^P, [f]_0^{P-T}) \quad (5.38)$$

$$\geq n(W + 1)H(s_1|s_0) - nH(s_{B+1}^P|s_B)
+ H([f]_{P-T+1}^P, [f]_{2B+2}^{2P}, \ldots, [f]_{(N-1)P+B+1}^{NP}|s_0^P, [f]_0^{P-T}) - nW\epsilon_n. \quad (5.39)$$

Note that (5.37) follows from (5.36), and (5.38) follows from following Markov chain property:

$$\{[f]_0^B, [s]_{B+1}^{B-1}\} \rightarrow s_B \rightarrow [s]_{B+1}^{P-T-1}. \quad (5.40)$$

**Step 3:** In this step we exploit the fact that the source sequences in the interval $\{P-T+1, \ldots, P\}$
can also be recovered according to the following inequality.

\[ H([s]_0^{P-T+1}||[f]_0^P, [f]_{P+B+1}^{P+T}) \leq nT\epsilon_n. \]  

(5.41)

Inequality in (5.41) can be used to lower bound the last entropy term in (5.39) as follows.

\[
\begin{align*}
H([f]_0^P, [f]_{P+B+1}^{P+T}) & \geq H([s]_0^{P-T+1}, [f]_0^P, [f]_{P+B+1}^{P+T}) \\
& \geq H([s]_0^{P-T+1}, [f]_0^P, [f]_{P+B+1}^{P+T}) - nT\epsilon_n \\
& = nT(H([f]_0^P, [f]_{P+B+1}^{P+T}) - nT\epsilon_n) \\
& = nTnR - nT\epsilon_n.
\end{align*}
\]

(5.42)

where (5.42) follows from (5.41).

**Step 4:** The last step is considering all the \( N \) periods simultaneously and repeatedly exploiting the same methods in steps 1 to 3. In particular by combining (5.34), (5.39) and (5.43) we have

\[
N(P-B)nR \geq H([f]_0^P, [f]_{P+B+1}^{P+T}) - nT\epsilon_n.
\]

(5.44)

We now repeat the same methods used in steps 1–3 for \( (N-1) \) periods and lower bound the entropy term in (5.44) as follows.

\[
N(P-B)nR \geq H([f]_0^P, [f]_{P+B+1}^{P+T}) - nT\epsilon_n.
\]

(5.45)

Finally by dividing (5.45) by \( N(T+W+1)n \) and taking \( n \to \infty \) and thereafter \( N \to \infty \) we recover (5.7). This completes the proof of the lower bound.

### 5.5 Lossless Streaming with Controlled-Interruption over Sliding-Window erasure burst Channel

#### 5.5.1 Achievability

The coding scheme is the random binning scheme similar to the single erasure burst case. In particular, all the \( 2^{nH(s)} \) typical sequences are randomly and independently placed into \( 2^R \) bins and the partitions are revealed to both the encoder and the decoder beforehand. The encoder at each time observes the source sequence \( s^n \) and sends its bin index through the channel. The decoder keeps collecting the received packets and performs jointly typicality decoding to recover the source sequences by their required time. Consider two cases as follows.

- **\( T \leq G - W - 1 \):** This case is similar to the single erasure burst setup. The decoder at time \( i \)
Chapter 5. Delay-Constrained Streaming

134

Figure 5.6: An example of streaming with controlled-interruption over sliding-window erasure burst channel model. $B = 3$, $W = 1$ and $G = 3$. (a) $T = 2$, and (b) $T_{\text{opt}} = G - W - 1 = 1$. Note that the two systems are equivalent because the decoder in both cases recovers the source sequences $\hat{s}_4^n$ and $\hat{s}_5^n$ from $\{f_3, f_4, f_5\}$ which is available at time 5.

reverses $\hat{s}_{i-T}^n$ in one of the following ways. It either has already recovered $i - T - 1$ and tries to recover $\hat{s}_{i-T}^n$ from $\{s_{i-T-1}^n, f_{i-T}\}$. This succeeds with high probability if $R \geq H(s_1|s_0)$. Or it has recovered $s_{i-B-W-T-1}^n$ and has access to $[f]_{i-W-T}$ and thus succeeds in recovering $\hat{s}_{i-T}^n$ if

$$R \geq H(s_1|s_0) + \frac{1}{T + W + 1} I(s_B; s_{B+1}|s_0),$$

which establishes (5.16) for this case.

- $T > G - W - 1$: Again assume that the decoder at time $i$ is interested in recovering $\hat{s}_{i-T}^n$. Fig. 5.6 illustrates an example of this case. If $s_{i-T-1}^n$ has been recovered, the decoder succeeds in recovering $s_{i-T}^n$ from $\{s_{i-T-1}^n, f_{i-T}\}$ if $R \geq H(s_1|s_0)$. Now consider the case where $s_{i-B-W-T-1}^n$ is already recovered and the decoder keep collecting all the non-erased codewords in the interval $[i - W - T, i]$. Note that for this case, because of large value of the delay $T$ comparing to the guard length $G$, not all the codewords in the specified interval are necessarily available at the decoder. However, according to the sliding-window erasure burst model the codewords in the interval $[i - W - T, i + G - T - W]$ is guaranteed to be available to the decoder. Thus the following rate is achievable.

$$R \geq R_d(B, W; G - W - 1) = H(s_1|s_0) + \frac{1}{G} I(s_B; s_{B+1}|s_0).$$

(5.47)

Note that as the rate expression in (5.47) suggests, when $T > G - W - 1$ the coding scheme is
designed for delay of $G - W - 1$ which is strictly less than the required delay $T$. As the following lemma indicates, for the proposed binning-based coding scheme there is no gain in exploiting larger delays.

**Corollary 11.** Consider the sliding-window erasure burst channel with parameters $B$ and $G$. The rate associated to the binning-based coding scheme described in Sec. 5.5.1 is minimized for the delay $T_{opt} = G - W - 1$ at the decoder. In other words any other delay $T \neq T_{opt}$, requires higher rate.

The proof of Corollary 11 is provided in Appendix D.2. By combining (5.46) and (5.47), the rate expression in (5.16) is derived.

### 5.5.2 Converse

The proof of converse is also very similar to the case of single erasure burst. Consider two cases as follows.

- $T \leq G - W - 1$: The proof of the converse for this case is exactly equivalent to the single erasure burst case presented in Section 5.4.2. In particular we consider a periodic erasure pattern with the period $P = B + W + T + 1$ and all the steps used in Section 5.4.2 holds in this case. This establishes the lower bound on rate as

$$R \geq H(s_1|s_0) + \frac{1}{W + T + 1}I(s_B:s_{B+1}|s_0). \quad (5.48)$$

- $T > G - W - 1$: The proof of the converse for this case is slightly different from what we have in single erasure case. In fact we consider a periodic erasure pattern with period $P = G + B$ such that $k$-th period spans the interval $[kP + 1, (k + 1)p]$ and the channel erases the first $B$ codewords of each period while revealing the rest of the codewords. Now consider

$$(NL + T)nR \geq H([f]_{B+1}^P, [f]_{P+B+1}^{2P}, \ldots, [f]_{(N-1)p+B+1}^{NP}, [f]_{NP+T}^{NP+T}) \quad (5.49)$$

$$\geq H([f]_{B+1}^P, [f]_{P+B+1}^{2P}, \ldots, [f]_{(N-1)p+B+1}^{NP}, [f]_{NP+T}^{NP+T}|s_0) \quad (5.50)$$

$$\geq nH(s_{B+W+1}|s_0) + H([f]_{B+1}^P, [f]_{P+B+1}^{2P}, \ldots, [f]_{(N-1)p+B+1}^{NP}, [f]_{NP+T}^{NP+T}|s_0, s_{B+W+1}) - n\epsilon_n. \quad (5.51)$$

This is very similar to the first step in Section 5.4.2. Using slightly modified methods used in step 2 and step 3 in Section 5.4.2, we can lower bound the entropy term in (5.51) as

$$H([f]_{B+1}^P, [f]_{(N-1)p+B+1}^{NP}, [f]_{NP+T}^{NP+T}|s_0, s_{B+W+1}) \geq H([f]_{B+1}^P, [f]_{(N-1)p+B+1}^{NP}, [f]_{NP+T}^{NP+T}|s_0^B, s_{B+W+1})$$

$$\geq H([s]_{B+W+1}^{B}, [s]_{B+W+1}^{P}, [f]_{P+B+1}^{2P}, \ldots, [f]_{(N-1)p+B+1}^{NP}, [f]_{NP+T}^{NP+T}|s_0^B, s_{B+W+1}) - n(L - 1)\epsilon_n$$

$$\geq H([s]_{B+W+1}^{B}, [s]_{B+W+1}^{P}, [f]_{P+B+1}^{2P}, \ldots, [f]_{(N-1)p+B+1}^{NP}, [f]_{NP+T}^{NP+T}|s_0^P) - n(L - 1)\epsilon_n$$

$$\geq nLH(s_1|s_0) - nH(s_{W+1}|s_0) + H([f]_{P+B+1}^{2P}, \ldots, [f]_{(N-1)p+B+1}^{NP}, [f]_{NP+T}^{NP+T}|s_0^P) - n(L - 1)\epsilon_n. \quad (5.52)$$

Similar to the step 4 in Section 5.4.2, we can combine (5.51) and (5.52) to write

$$(NL + T)nR \geq nNH(s_{B+W+1}|s_0) + nLH(s_1|s_0) - nNH(s_{W+1}|s_0) + H([f]_{NP+T}^{NP+T}|s_0^P) - nL\epsilon_n. \quad (5.53)$$
In what follows, we derive a lower bound on \( (5.56) \).

We consider \( p + 1 \) periods of a periodic erasure channel with period \( p = B + T + 1 \), such that the first \( B \) channel outputs of each period are erased by the channel and the rest of the packets are revealed to the decoder. Now consider the following entropy inequality.

\[
(T + 1)(K + 1)nR \geq H([f^B_B f^{p+B+T}_{p+B} \ldots f^{Kp+B+T}_{Kp+B} | s_{-1}])
\]

(5.56)

In what follows, we derive a lower bound on (5.56).

**Step 1**: We first exploit the fact that the decoder reconstruct the source sequence \( s^n_B \) within distortion \( D \), from \( \{s^n_B, [f^B_B]\} \), i.e.,

\[
h(s_B | f^B_B, s^n_{-1}) \leq \frac{n}{2} \log (2\pi eD)
\]

(5.57)

We have,

\[
H([f^B_B f^{p+B+T}_{p+B} \ldots f^{Kp+B+T}_{Kp+B} | s_{-1}]) \\
= I(s_B; [f^B_B f^{p+B+T}_{p+B} \ldots f^{Kp+B+T}_{Kp+B} | s_{-1}]) + H([f^B_B f^{p+B+T}_{p+B} \ldots f^{Kp+B+T}_{Kp+B} | s_{-1} | s_B]) \\
\geq h(s_B | s_{-1}) - h(s_B | f^B_B, s_{-1}) + H([f^B_B f^{p+B+T}_{p+B} \ldots f^{Kp+B+T}_{Kp+B} | s_{-1}, s_B]) \\
\geq \frac{n}{2} \log \left( \frac{1 - \rho^{2(B+1)}}{D} \right) + H([f^B_B f^{p+B+T}_{p+B} \ldots f^{Kp+B+T}_{Kp+B} | s_{-1}, s_B]),
\]

(5.58)

where (5.58) follows from application of (5.57) and the fact that

\[
h(s_B | s_{-1}) = nh(s_B | s_{-1}) \\
= \frac{n}{2} \log \left( 2\pi e(1 - \rho^{2(B+1)}) \right).
\]

(5.59)

**Step 2**: In this step, we first lower bound the last entropy term in (5.58) by conditioning the term with the source sequences associated with the erasure times of the first period, i.e., \( \{s^n_0, s^n_1, \ldots, s^n_{B-1}\} \).
We have

\[ H([f]_{B+1}^{B+T} | [f]_{p+B}^{p+B+T} \ldots [f]_{Kp+B}^{Kp+B+T} | [s]_{-1}, s_B) \geq H([f]_{B+1}^{B+T} [f]_{p+B}^{p+B+T} \ldots [f]_{Kp+B}^{Kp+B+T} | [s]_{-1}). \]  

(5.60)

Then we exploit the following inequality

\[ h([s]_{B+1}^{B+T} | s_{-1}, [f]_{0}^{B+T} [f]_{p+B}^{p+B+T} \ldots [f]_{Kp+B}^{Kp+B+T}) \leq \frac{nT}{2} \log (2\pi e D), \]

(5.61)

to lower bound the rate in (5.60) as follows.

\[
\begin{align*}
H([f]_{B+1}^{B+T} | [f]_{p+B}^{p+B+T} \ldots [f]_{Kp+B}^{Kp+B+T} | [s]_{B+1}) \\
= I([s]_{B+1}^{B+T} : [f]_{B+1}^{B+T} \ldots [f]_{Kp+B}^{Kp+B+T} | [s]_{-1}) + H([f]_{B+1}^{B+T} [f]_{p+B}^{p+B+T} \ldots [f]_{Kp+B}^{Kp+B+T} | [s]_{-1}) \\
\geq h([s]_{B+1}^{B+T} | s_B) - \frac{nT}{2} \log (2\pi e D) + H([f]_{p+B}^{p+B+T} \ldots [f]_{Kp+B}^{Kp+B+T} | [s]_{-1}^{B+T}) \\
= \frac{nT}{2} \log \left(1 - \frac{\rho^2}{D}\right) + H([f]_{p+B}^{p+B+T} \ldots [f]_{Kp+B}^{Kp+B} | [s]_{-1}^{B+T}),
\end{align*}
\]

(5.62)

where (5.62) follows from the application of (5.61) and (5.63) follows from the fact that

\[ h([s]_{B+1}^{B+T} | s_B) = nh([s]_{B+1}^{B+T} | s_B) \\
= \frac{nT}{2} \log (2\pi e (1 - \rho^2)). \]

(5.64)

**Step 3:** In this step, using the same methods in steps 2 and 3, we can lower bound the last term in (5.63). By repeating the iteration for \( K \) times, the following lower bound on (5.56) is derived.

\[ (T + 1)(K + 1)nR \geq H([f]_{B}^{B+T} [f]_{p+B}^{p+B+T} \ldots [f]_{Kp+B}^{Kp+B+T} | [s]_{-1}) \]

\[ \geq \frac{nK}{2} \log \left(1 - \frac{\rho^2(B+1)(1-\rho^2)^T}{D^{T+1}}\right) + H([f]_{Kp+B}^{Kp+B+T} | [s]_{-1}^{(K-1)p+B+T}). \]

(5.65)

Finally, by dividing the two sides of (5.65) by \((T + 1)(K + 1)n\) and then letting \( K \to \infty \) and then \( n \to \infty \), the lower bound in Proposition 5 is derived.

### 5.6.2 Coding Scheme

In this section we present the proof of Proposition 6. The coding scheme is based on Q-and-B scheme.

**Codebook Generation:** For any time \( t \), the source sequence \( s_t^n \) is quantized through the following Gaussian test channel.

\[ u_t = s_t + z_t \]

(5.66)

where \( z_t \sim \mathcal{N}(0, \sigma^2_t) \) is independent noise. All the typical codewords \( u_t^n \) are randomly and independently placed into \( 2^{nR} \) bins and codebook consisting of the bin indices are revealed to both the encoder and decoder beforehand.

**Encoder:** At each time \( t \), the encoder first finds the quantization sequence \( u_t^n \) typical with \( s_t^n \). Then it sends the bin index associated with \( u_t^n \) through the channel.
Figure 5.7: A schematic of the erasure burst channel model and four different regions for $B' = 2$, $T = 3$, $j = 6$.

**Decoder:** The decoder at any time $t$, first attempts to recover the quantization sequence $u^n_t^\mathcal{c}$ based on jointly typicality decoding. Then, whenever required, it produces the MMSE estimate of the source sequence $s^n_t$ from the available quantization codewords.

Given a collection of random variables $\mathcal{V}$, we let the MMSE estimate of $s_t$ be denoted by $\hat{s}_t(\mathcal{V})$, and its associated estimation error is denoted by $\sigma_t^2(\mathcal{V})$, i.e.,

$$\hat{s}_t(\mathcal{V}) = E[s_t|\mathcal{V}]$$

$$\sigma_t^2(\mathcal{V}) = E[(s_t - \hat{s}_t(\mathcal{V}))^2].$$

**Lemma 24.** A rate-distortion $(R,D)$ is achievable if for any $t \geq 0$, $B' \leq B$ and $j \geq B' - 1$ we have

$$R \geq \lambda_t(j, B')$$

and the test channel in (5.66) satisfies

$$\gamma_t(j, B') \leq D,$$

where $\lambda_t(j, B')$ and $\gamma_t(j, B')$ are defined as follows.

$$\lambda_t(j, B') \triangleq \begin{cases} I(s_t; u_t|[u]_{0}^{t-1}, s_{-1}) & \text{if } t \leq j - B' \\ \max_{\mathcal{M} \subseteq \{j+1, \ldots, j+T+1\}} \frac{1}{|\mathcal{M}|} I(s_{\mathcal{M}}; u_{\mathcal{M}}|[u]_{0}^{j-B'}, u_{\mathcal{M}^c}, s_{-1}) & \text{if } t = j + T + 1 \\ I(s_t; u_t|[u]_{0}^{j-B'}, [u]_{j+1}^{t-1}, s_{-1}) & \text{if } t > j + T + 1 \end{cases}$$

$$\gamma_t(j, B') \triangleq \begin{cases} \sigma_t^2([u]_0^j, s_{-1}) & \text{if } t < j - B' \\ \max_{\mathcal{M}} \sigma_t^2([u]_{0}^{j-B'}, s_{-1}) & \text{if } t = j - B', \mathcal{M} \triangleq \{j - B' - T, \ldots, j - B'\} \\ \sigma_{j+1}^2([u]_{0}^{j-B'}, [u]_{j+1}^{j+T+1}, s_{-1}) & \text{if } t = j + T + 1 \\ \sigma_{j-T}^2([u]_{0}^{j-B'}, [u]_{j+1}^{t}, s_{-1}) & \text{if } t > j + T + 1 \end{cases}$$

**Proof.** Assume there is an erasure burst of length $B' \leq B$ spanning the time interval $\{j - B' + 1, \ldots, j\}$ for some $j \geq B' - 1$. Fig. 5.7 illustrates an example of such a channel model for $B' = 2$, $T = 3$ and $j = 6$. We identify four operational regions for the decoder as follows. The different regions are shown by letter R in the figure.
Region 1: $t < j - B'$ when there is no erasure up to time $t$. The decoder at time $t$, has access to all the sequences $u^n_i$, $i < t$ and attempts the recovery $u^n_t$. The decoder succeeds with high probability if ( [38])

$$R \geq I(s_t; u^n_t | [u^n]_0^{t-1}, s_{-1}).$$

(5.73)

The decoder then computes the MMSE estimate of $s^n_{t-T}$ from $[[u^n]_0^t, s_{-1}]$. The test channel has to satisfy

$$\sigma^2_{t-T}(s^n_t, s_{-1}) \leq D.$$  

(5.74)

Region 2: $t = j - B'$ when the decoder recovers $u^n_t$ if the rate satisfies the same rate constraints in (5.73). It then computes the MMSE estimates of the source sequences $s^n_k$ for $k \in \{j - B' - T, \ldots, j - B'\}$ with the following constraints.

$$\sigma^2_k(s^n_t | [u^n]_0^{t-1}, s_{-1}) \leq D.$$  

(5.75)

Therefore we need to have

$$\max_{k \in \mathcal{M}} \sigma^2_k(s^n_t | [u^n]_0^{t-1}, s_{-1}) \leq D,$$

(5.76)

where $\mathcal{M} \triangleq \{j - B' - T, \ldots, j - B'\}$.

Region 3: $t = j + T + 1$ when the decoder after collecting all the channel outputs of interval $\{j + 1, \ldots, j + T + 1\}$, simultaneously recovers the sequences $\{u^n_{j+1}, \ldots, u^n_{j+T+1}\}$. It succeeds if for any $\mathcal{M} \subseteq \{j + 1, \ldots, j + T + 1\}$, the rate satisfies the following ( [44]).

$$R \geq \frac{1}{|\mathcal{M}|} I(s_\mathcal{M}; u_\mathcal{M}| [u^n]_{j-B'}^{j+B'}, s_{-1}).$$

(5.77)

The decoder in addition computes the MMSE estimate of $s^n_{j+1}$ with the following distortion constraint.

$$\sigma^2_{j+1}(s^n_{j+1} | [u^n]_{j-B'}^{j+B'}, s_{-1}) \leq D.$$  

(5.78)

Region 4: $t > j + T + 1$ when the decoder recovers $u^n_t$ with the rate constraint

$$R \geq I(s_t; u^n_t | [u^n]_{0}^{j-B'}, [u^n]_{j+1}^{t-1}, s_{-1}).$$  

(5.79)

and reconstruct the MMSE estimate of $s^n_{t-T}$ with the distortion constraint

$$\sigma^2_{t-T}(s^n_{t-T} | [u^n]_{0}^{j-B'}, [u^n]_{j+1}^{t-1}, s_{-1}) \leq D.$$  

(5.80)

Note that the rate of any achievable scheme has to simultaneously satisfy the rate constraints in (5.73), (5.77) and (5.79) and the distortion constraints in (5.74), (5.76), (5.78) and (5.80) for all possible values of $t \geq 0$, $B' \leq B$ and $j \geq B' - 1$. This completes the proof of Lemma 24. \qed
Lemma 25. Define

\[ \Delta_1^*(\sigma^2) \triangleq \lim_{t \to \infty} \sigma^2_t ([u]_t^0, s_{-1}) \]  
(5.81)

\[ \Delta_2^*(\sigma^2) \triangleq \lim_{t \to \infty} \sigma^2_t ([u]_t^{t-B-1}, [u]_t^{t+T}, s_{-1}) . \]  
(5.82)

For any test channel noise \( \sigma^2 \) satisfying

\[ d^*(B, T, \sigma^2) \triangleq \max \{ \Delta_1^*(\sigma^2), \Delta_2^*(\sigma^2) \} \leq D, \]  
(5.83)

the following rate is achievable.

\[ R \geq \lim_{t \to \infty} \frac{1}{T+1} I(s_{t+T}; [u]_{t+T}^t [u]_{t-B}^{t-1}, s_{-1}). \]  
(5.84)

\[ \square \]

Proof. We prove the lemma by locating the worst-case erasure pattern of the channel with respect to rate and distortion constraint separately.

Rate Constraint: We first prove the rate constraint in (5.84) by identifying the dominating rate constraint.

Consider the following steps.

Step 1: We first show that for any \( t, B' \) and \( j \), among the rate constraints in region 1 and 2 and region 4 defined through the proof of Lemma 24, the following constraint is dominant.

\[ R \geq \lim_{t \to \infty} I(s_{t}; u_t [u]_{0}^{t-T-B-1}, [u]_{t-1}^{t-1}, s_{-1}). \]  
(5.85)

To show this, first note that for any fixed \( t \) and \( B' \), over all \( B' - 1 \leq j < t - T - 1 \), i.e., region 4, we have

\[ \lambda_t(j, B') = I(s_t; u_t [u]_{0}^{j-B'}, [u]_{j+1}^{t-1}, s_{-1}) \]

\[ = h(u_t | [u]_{0}^{j-B'}, [u]_{j+1}^{t-1}, s_{-1}) - h(u_t | s_t) \]

\[ \leq h(u_t | [u]_{0}^{j-T-B'}, [u]_{j-1}^{t-1}, s_{-1}) - h(u_t | s_t) \]  
(5.86)

\[ = I(s_t; u_t [u]_{0}^{j-T-B'}, [u]_{j-1}^{t-1}, s_{-1}) \]

\[ = \lambda_t(t - T - 2, B'), \]  
(5.87)

where (5.86) follows from the application of Lemma 10.

Second, note that \( \lambda_t(t - T - 2, B') \) is increasing function with respect to \( t \), because

\[ \lambda_{t+1}(t - T - 1, B') \].

\[ \triangleq I(s_{t+1}; u_{t+1} [u]_{0}^{t-T-B'-1}, [u]_{t}^{t-1}, s_{-1}) \]

\[ = h(u_{t+1} | [u]_{0}^{t-T-B'-1}, [u]_{t}^{t-1}, s_{-1}) - h(u_{t+1} | s_{t+1}) \]

\[ \geq h(u_{t+1} | [u]_{0}^{t-T-B'-1}, [u]_{t}^{t-1}, s_{0}, s_{-1}) - h(u_{t+1} | s_{t+1}) \]

\[ = h(u_{t+1} | [u]_{0}^{t-T-B'-2}, [u]_{t-1}^{t-1}, s_{0}) - h(u_t | s_t) \]
Thus for any $t$

$$\lambda_t(t - T - 2, B') \leq \lim_{\tau \to \infty} \lambda_\tau(\tau - T - 2, B').$$

(5.89)

Third, it is not hard to show that for any $B' \leq B$,

$$\lim_{\tau \to \infty} \lambda_\tau(\tau - T - 2, B') \leq \lim_{\tau \to \infty} \lambda_\tau(\tau - T - 2, B).$$

(5.90)

From (5.87), (5.89) and (5.90), the step 1 is shown.

Step 2: We now show that, among all the rate constraints in region 3, the following constraint is dominant.

$$R \geq \lim_{t \to \infty} \frac{1}{T + 1} I([s]_t^T; [u]_t^T || [u]_{t-B}^{t-2}, s_{-1}).$$

(5.91)

First, for any set of index $\mathcal{M} \subseteq \{0, \ldots, T\}$ and $t \geq T$, define the following notations.

$$\mathcal{M}(+t) \triangleq \{t + i | i \in \mathcal{M}\}$$

(5.92)

$$\eta(\mathcal{M}, t, B') \triangleq \frac{1}{|\mathcal{M}|} I([s]_{\mathcal{M}(+t)}; [u]_{\mathcal{M}(+t)} || [u]_{t-B}^{t-1}, [u]_{\mathcal{M}(+t)}, s_{-1})$$

(5.93)

First consider the following lemma.

**Lemma 26.** For any set $\mathcal{M} \subseteq \{0, \ldots, T\}$, we have

$$\eta(\mathcal{M}, t, B') \leq \eta(\mathcal{M}, t + 1, B'),$$

i.e., the rate associated with the same subset $\mathcal{M}$ is an increasing function with respect to $t$. \hfill \Box

**Proof.** Note that

$$|\mathcal{M}| \cdot \eta(\mathcal{M}, t + 1, B') \triangleq I([s]_{\mathcal{M}(+(t+1))}; [u]_{\mathcal{M}(+(t+1))} || [u]_{t-B}^{t-1}, [u]_{\mathcal{M}(+(t+1))}, s_{-1})$$

$$\geq I([s]_{\mathcal{M}(+(t+1))}; [u]_{\mathcal{M}(+(t+1))} || [u]_{t-B}^{t-1}, [u]_{\mathcal{M}(+(t+1))}, [s]_{-1})$$

(5.95)

$$= I([s]_{\mathcal{M}(+(t+1))}; [u]_{\mathcal{M}(+(t+1))} || [u]_{t-B}^{t-1}, [u]_{\mathcal{M}(+(t+1))}, s_0)$$

(5.96)

$$= I([s]_{\mathcal{M}(+(t))}; [u]_{\mathcal{M}(+(t))} || [u]_{t-B}^{t-1}, [u]_{\mathcal{M}(+(t))}, s_{-1})$$

(5.97)

$$\triangleq |\mathcal{M}| \cdot \eta(\mathcal{M}, t, B'),$$

(5.98)

where (5.95) follows from the fact that conditioning reduces the differential entropy, (5.96) follows from the Markov chain property among the sources and (5.97) follows form the stationarity property of the source model. \hfill \Box
According to Lemma 26, in order to locate the dominant rate constraint it suffices to consider the steady state regime when \( t \to \infty \). Define

\[
\bar{\eta}(\mathcal{M}, B') \triangleq \lim_{t \to \infty} \eta(\mathcal{M}, t, B').
\]  

(5.99)

Now consider the following lemma.

**Lemma 27.** For any fixed \( B' \) in region 3, we have

\[
\arg \max_{\mathcal{M} \subseteq \{0, \ldots, T\}} \bar{\eta}(\mathcal{M}, B') = \{0, \ldots, T\}. \tag{5.100}
\]

\[\square\]

**Proof.** See Appendix D.3.

\[\square\]

By exploiting Lemma 27, we can conclude that for a fixed \( B' \), the following rate constraint is dominant.

\[
R \geq \lim_{t \to \infty} \frac{1}{T+1} I([s]_{t-T}; [u]_{t-T}^t | [u]_0^{t-B'-1}, s_{-1}).
\]  

(5.101)

Finally, it can be observed that for any \( B' \leq B \)

\[
\lim_{t \to \infty} \frac{1}{T+1} I([s]_{t-T}; [u]_{t-T}^t | [u]_0^{t-B'-1}, s_{-1}) \leq \lim_{t \to \infty} \frac{1}{T+1} I([s]_{t-T}; [u]_{t-T}^t | [u]_0^{t-B-1}, s_{-1}). \tag{5.102}
\]

This verifies Step 2.

Step 3: Finally we show that the rate constraint (5.85) derived in step 1 is included in the constraint (5.91) of step 2. Note that,

\[
\lim_{t \to \infty} \frac{1}{T+1} I([s]_{t-T}; [u]_{t-T}^t | [u]_0^{t-B-1}, s_{-1}) = \lim_{t \to \infty} \frac{1}{T+1} I([s]_{t-T}; [u]_{t-T}^t | [u]_0^{t-B-1})
\]

\[
= \lim_{t \to \infty} \frac{1}{T+1} h([u]_{t-T}^t | [u]_0^{t-B-1}) - h(u_1 | s_1)
\]

\[
= \lim_{t \to \infty} \frac{1}{T+1} \sum_{k=0}^{T} h(u_{t-T+k} | [u]_0^{t-B-1} | [u]_{t-T}^{t-k-1}) - h(u_1 | s_1)
\]

\[
= \lim_{t \to \infty} \frac{1}{T+1} \sum_{k=0}^{T} h(u_{t-k} | [u]_0^{t-B-1} | [u]_{t-k}^{t-1}) - h(u_1 | s_1)
\]

\[
\geq \lim_{t \to \infty} \frac{1}{T+1} \sum_{k=0}^{T} h(u_{t-k} | [u]_0^{t-B-2} | [u]_{t-k-1}^{t-1}) - h(u_1 | s_1) \tag{5.103}
\]

\[
= \lim_{t \to \infty} h(u_0 | [u]_0^{t-B-2} | [u]_{t-T}^{t-1}) - h(u_1 | s_1)
\]

\[
= \lim_{t \to \infty} I(s_t; u_t | [u]_0^{t-B-2} | [u]_{t-T}^{t-1})
\]

\[
= \lim_{t \to \infty} I(s_t; u_t | [u]_0^{t-B-2} | [u]_{t-T-1}^{t-1}, s_{-1}), \tag{5.104}
\]

where (5.103) follows from the application of Lemma 10, which results the following for any \( k \in \{0, \ldots, T\} \)
and \( t \to \infty \):
\[
h(u_t|u_0^{t-B-k-1}[u]_{t-k}^{t-1}) \geq h(u_t|u_0^{t-T-B-2}[u]_{t-T-1}^{t-1}).
\]
This completes Step 3.

Application of the results of Steps 1–3, concludes that for any test channel if the rate satisfies (5.84), the decoder succeeds in recovering the quantization sequences whenever required.

**Distortion Constraint:** We now prove the distortion constraint in (5.83) by identifying the dominating distortion constraint. Consider the four regions used in proof of Lemma 24.

**Step 1:** We first show that, for any \( B' \leq B \), and \( j > t \), i.e., regions 1 and 2, the following distortion constraint is dominant.
\[
\Delta_1(\sigma^2_s) \triangleq \lim_{t \to \infty} \sigma^2_t ([u]_0^t, s_{-1}). \tag{5.105}
\]
Note that the term in (5.105) refers to the MMSE estimation error of the source sequence \( s^n \) in steady state when the erasure burst spans the interval \( \{t + 1, \ldots, t + B'\} \). First we show that for a fixed \( t \) and \( B' \leq B \), and any \( j > t \) and \( k \in \{0, \ldots, T\} \),
\[
\sigma^2_t ([u]_{t+k}^t, s_{-1}) \leq \sigma^2_t ([u]_0^t, s_{-1}), \tag{5.106}
\]
which is obvious because removing the term \( [u]_{t+1}^t \) from the observation set can only reduces the estimation error. Then we show that the distortion expression in right hand side of (5.106) is an increasing function of \( t \). Note that,
\[
\frac{1}{2} \log (2\pi e \cdot \sigma^2_t ([u]_{t+1}^t, s_{-1})) = h(s_{t+1}|[u]_{t+1}^t, s_{-1}) \tag{5.107}
\]
\[
\leq h(s_{t+1}|[u]_{t+1}^t, s_0, s_{-1})
\]
\[
= h(s_{t+1}|[u]_1^t, s_0)
\]
\[
= h(s_{t}|[u]_{t-1}^t, s_{-1})
\]
\[
= \frac{1}{2} \log (2\pi e \cdot \sigma^2_t ([u]_0^t, s_{-1}) \tag{5.108}
\]
where (5.107), follows from the fact that for jointly Gaussian distribution, MMSE is the optimal estimator. From (5.108) and the fact that \( f(x) = \log(2\pi e x)/2 \) is a monotonically increasing function with respect to \( x \), we conclude that, for any \( t \),
\[
\sigma^2_t ([u]_0^t, s_{-1}) \leq \Delta_1(\sigma^2_s). \tag{5.109}
\]
Finally from (5.109) and (5.106), Step 1 is shown.

**Step 2:** In this step, we show that for any \( t \), \( B' \leq B \), and \( j < t - T - B \), i.e., regions 3 and 4, the distortion constraint
\[
\Delta_2(\sigma^2_s) \triangleq \lim_{t \to \infty} \sigma^2_t ([u]_{0}^{t-B-1}, [u]_t^{t+T}, s_{-1}) \tag{5.110}
\]
is dominant. Note that the term in (5.110) refers to the MMSE estimation error of the source sequence \( s^n \) in steady state right after the erasure burst of length \( B \) spanning the interval \( \{t - B, \ldots, t - 1\} \).
First we show that for a fixed \( t \) and \( B' \leq B \) and any \( j < t \),
\[
\sigma_t^2 ([u]_{0}^{t-B'-1}, [u]_{j+1}^{t+T}, s_{-1}) \leq \sigma_t^2 ([u]_{0}^{t-B-1}, [u]_{t}^{t+T}, s_{-1}).
\]

This is equivalent to showing
\[
h \left( s_{t}[[u]_{0}^{t-B'-1}, [u]_{j+1}^{t+T}, s_{-1}] \right) \leq h \left( s_{t}[[u]_{0}^{t-B-1}, [u]_{t}^{t+T}, s_{-1}] \right), \tag{5.111}
\]
which immediately follows from the application of Lemma 10.

Then, using similar methods used before, we can show that the term in right hand side of (5.111) is increasing function of \( t \). For sake of compactness, we omit the detailed proof here. This proves Step 2.

Note that depending on the test channel noise \( \sigma_z^2 \) and the values of \( B \) and \( T \), each distortion constraint in (5.105) and (5.110) may be active. Thus the test channel noise in achievable scheme has to simultaneously satisfy the two constraints, as required in the distortion constraint of (5.83).

The following lemma completes the proof of Proposition 6.

**Lemma 28.** Any rate-distortion pair \((R,D)\) satisfying the constraints in Proposition 6, also satisfies the rate and distortion constraints in (5.84) and (5.83), in Lemma 25, and thus is achievable. \(\square\)

**Proof.** Define
\[
\hat{R}(B,T,\sigma_z^2) \triangleq \lim_{t \to \infty} \frac{1}{T+1} I([s]_t^{t+T};[u]_t^{t+T}|[u]_{0}^{t-B}) \tag{5.112}
\]
and
\[
\hat{d}(B,T,\sigma_z^2) \triangleq \max \{ \hat{\Delta}_1(\sigma_z^2), \hat{\Delta}_2(\sigma_z^2) \}, \tag{5.113}
\]
where
\[
\hat{\Delta}_1(\sigma_z^2) \triangleq \lim_{t \to \infty} \sigma_t^2 ([u]_{0}^{t-B}) \tag{5.114}
\]
\[
\hat{\Delta}_2(\sigma_z^2) \triangleq \lim_{t \to \infty} \sigma_t^2 ([u]_{0}^{t-B-1}, [u]_t^{t+T}). \tag{5.115}
\]

**Step 1:** First we show that for any test channel noise \( \sigma_z^2 \),
\[
\hat{R}(B,T,\sigma_z^2) \geq R^*(B,T,\sigma_z^2) \tag{5.116}
\]
\[
\hat{d}(B,T,\sigma_z^2) \geq d^*(B,T,\sigma_z^2). \tag{5.117}
\]

In particular,
\[
\hat{R}(B,T,\sigma_z^2) \triangleq \lim_{t \to \infty} \frac{1}{T+1} I([s]_t^{t+T};[u]_t^{t+T}|[u]_{0}^{t-B})
\]
\[
= \lim_{t \to \infty} \frac{1}{T+1} h([u]_t^{t+T}|[u]_{0}^{t-B}) - h(u_{1}|s_{1})
\]
\[
\geq \lim_{t \to \infty} \frac{1}{T+1} h([u]_t^{t+T}|[u]_{0}^{t-B}, s_{-1}) - h(u_{1}|s_{1})
\]
\[
= \lim_{t \to \infty} \frac{1}{T+1} I([s]_t^{t+T};[u]_t^{t+T}|[u]_{0}^{t-B}, s_{-1}) \tag{5.118}
\]
\[ R \triangleq R^*(B, T, \sigma_z^2). \]  

(5.119)

Also

\[ \tilde{\Delta}_1(\sigma_z^2) \triangleq \lim_{t \to \infty} \sigma_t^2([u]_t^0) \]

\[ \geq \lim_{t \to \infty} \sigma_t^2([u]_t^0, s_{-1}) \triangleq \Delta_t^*(\sigma_z^2) \]  

(5.120)

and

\[ \tilde{\Delta}_2(\sigma_z^2) \triangleq \lim_{t \to \infty} \sigma_t^2([u]_{t-B}^{t-1}, [u]_{t}^{t+T}) \]

\[ \geq \lim_{t \to \infty} \sigma_t^2([u]_{t-B}^{t-1}, [u]_{t}^{t+T}, s_{-1}) \triangleq \Delta_t^*(\sigma_z^2) \]  

(5.121)

conclude (5.117).

Thus the rate pair \((R, D)\) satisfying \(R \geq \tilde{R}(B, T, \sigma_z^2)\) and \(\tilde{d}(B, T, \sigma_z^2) \leq D\) is achievable.

**Step 2:** Similar to the method used in Chapter 3, we consider the following single-variable discrete-time Kalman filter in steady state, i.e.,

\[ s_i = \rho s_{i-1} + n_i, \quad n_i \sim \mathcal{N}(0, 1 - \rho^2) \]

(5.122)

\[ u_i = s_i + z_i, \quad z_i \sim \mathcal{N}(0, \sigma_z^2). \]  

(5.123)

Thus, the MMSE estimate of \(s_i\) at steady state when \(i \to \infty\) can be replaced by its equivalent representation as follows.

\[ \hat{s}_i([u]_0^{i-1}) = s_i + e \]

\[ \triangleq \hat{s}_i, \]  

(5.124)

(5.125)

where \(e \sim \mathcal{N}(0, \Sigma(\sigma_z^2)/(1 - \Sigma(\sigma_z^2)))\) is independent noise. Thus

\[ \tilde{R}(B, T, \sigma_z^2) \triangleq \lim_{\tau \to \infty} \frac{1}{T + 1} I([s]_{\tau}^{T+T}; [u]_{\tau}^{T+T}|[u]_0^{T-B}) \]

\[ = \lim_{\tau \to \infty} \frac{1}{T + 1} I([s]_{\tau}^{T+T}; [u]_{\tau}^{T+T}|\hat{s}_{\tau-B+1}) \]  

(5.126)

\[ = \frac{1}{T + 1} I([s]_{\tau}^{T+T}; [u]_{\tau}^{T+T}|\hat{s}_{\tau-B+1}) \]

\[ \triangleq R_{\text{GM-d}}^+(B, T, d) \]  

(5.127)

and

\[ \tilde{\Delta}_1(\sigma_z^2) \triangleq \lim_{\tau \to \infty} \sigma_t^2([u]_0^\tau) \]

\[ = \lim_{\tau \to \infty} \sigma_t^2(\hat{s}_\tau, u_\tau) \]  

(5.128)

\[ = E[(s_\tau - \hat{s}_\tau)^2] \]

(5.129)

\[ (5.130) \]
and similarly
\[ \Delta_2(\sigma_2^2) \triangleq \lim_{\tau \to \infty} \sigma_\tau^2 ([u]_{0}^{\tau-B-1}, [u]_{\tau}^{\tau+T}) \]
\[ = \lim_{\tau \to \infty} \sigma_\tau^2 (\hat{s}_{\tau-B}, [u]_{\tau}^{\tau+T}) \]
\[ = E[(s_t - \hat{s}_2)^2], \tag{5.131} \]
where \( \hat{s}_1 \) and \( \hat{s}_2 \) are defined in Proposition 6. This completes the proof.

\section*{5.6.3 High Resolution Regime}

In order to prove Corollary 8, it suffices to show that the following rate is achievable in high resolution regime when \( D \to 0 \).
\[ R \geq \frac{1}{2(T + 1)} \log \left( \frac{(1 - \rho^{2(B+1)})(1 - \rho^T)}{D^{T+1}} \right) + o(D). \tag{5.132} \]

We choose \( \sigma_2^2 = D \). This satisfies the distortion constraint because we have
\[ \max \{ \tilde{\Delta}_1(\sigma_2^2), \Delta_2(\sigma_2^2) \} = \max \left\{ \lim_{\tau \to \infty} \sigma_\tau^2 ([u]_{0}^{\tau}), \lim_{\tau \to \infty} \sigma_\tau^2 ([u]_{0}^{\tau-B-1}, [u]_{\tau}^{\tau+T}) \right\} \leq \sigma_1^2 (u_\tau) = \frac{D}{1 + D} \leq D. \tag{5.133} \]

Also note that according to (5.128), the rate \( \tilde{R}(B, T, \sigma_2^2 = D) \) is achievable and we have
\[ \tilde{R}(B, T, \sigma_2^2 = D) \triangleq \lim_{\tau \to \infty} \frac{1}{T + 1} I([s]_{\tau}^{\tau+T}; [u]_{\tau}^{\tau+T}|[u]_{0}^{\tau-B-1}) \]
\[ = \lim_{\tau \to \infty} \frac{1}{T + 1} \left( h([s]_{\tau}^{\tau+T}|[u]_{0}^{\tau-B-1}) - h([s]_{\tau}^{\tau+T}|[u]_{0}^{\tau-B-1}[u]_{\tau}^{\tau+T}) \right). \tag{5.134} \]

We need to show that the limit of the term in (5.134) when \( D \to 0 \), satisfies (5.132).

\textit{Step 1:} We first compute the limit of the first term in (5.134) in high resolution regime. The term can be rewritten as
\[ h([s]_{\tau}^{\tau+T}|[u]_{0}^{\tau-B-1}) = h(s_\tau|[u]_{0}^{\tau-B-1}) + \sum_{k=\tau+1}^{\tau+T} h(s_k|s_{k-1}) \]
\[ = h(s_\tau|[u]_{0}^{\tau-B-1}) + Th(s_1|s_0). \tag{5.135} \]

The first term in (5.135) can be bounded as
\[ h(s_\tau|s_{\tau-B-1}) \leq h(s_\tau|[u]_{0}^{\tau-B-1}) \leq h(s_\tau|u_{\tau-B-1}) \tag{5.136} \]
and thus
\[ \frac{1}{2} \log \left( 2\pi e (1 - \rho^{2(B+1)}) \right) \leq h(s_\tau|[u]_{0}^{\tau-B-1}) \leq \frac{1}{2} \log \left( 2\pi e \left( 1 - \frac{\rho^{2(B+1)}}{1+D} \right) \right) \tag{5.137} \]
This conclude that
\[
\lim_{D \to 0} h(s_\tau | [u]_0^{\tau-B-1}) = \frac{1}{2} \log \left( 2\pi e (1 - \rho^2 (B+1)) \right).
\] (5.138)

Thus from (5.135) we can write
\[
\lim_{D \to 0} h([s]_\tau^{\tau+T} | [u]_0^{\tau-B-1}) = \frac{1}{2} \log \left( (2\pi e)^{T+1} (1 - \rho^2 (B+1)) (1 - \rho^2)^T \right).
\] (5.139)

**Step 2:** In this step we compute the limit of the second term in (5.134). The term can be rewritten as follows.
\[
h([s]_\tau^{\tau+T} | [u]_0^{\tau-B-1} [u]_\tau^{\tau+T}) = h(s_\tau | [u]_0^{\tau-B-1} [u]_\tau^{\tau+T}) + \sum_{k=\tau+1}^{\tau+T} h(s_k | s_{k-1}, [u]_\tau^{\tau+T}).
\] (5.140)

First note that (5.140) can be lower bounded as follows.
\[
h(s_\tau | [u]_0^{\tau-B-1} [u]_\tau^{\tau+T}) + \sum_{k=\tau+1}^{\tau+T} h(s_k | s_{k-1}, [u]_\tau^{\tau+T}) \geq h(s_\tau | s_{\tau-1}, u_\tau, s_{\tau+1}) + \sum_{k=\tau+1}^{\tau+T} h(s_k | s_{k-1}, u_k, s_{k+1})
\]
\[
= (T + 1) h(s_1 | s_0, u_1, s_2)
\]
\[
= \frac{T + 1}{2} \log \left( 2\pi e \left( \frac{1}{D} + \frac{1 + \rho^2}{1 - \rho^2} \right)^{-1} \right).
\] (5.141)

Second, (5.140) can be upper bounded as follows.
\[
h(s_\tau | [u]_0^{\tau-B-1} [u]_\tau^{\tau+T}) + \sum_{k=\tau+1}^{\tau+T} h(s_k | s_{k-1}, [u]_\tau^{\tau+T}) \leq h(s_\tau | u_\tau) + \sum_{k=\tau+1}^{\tau+T} h(s_k | u_k)
\]
\[
= (T + 1) h(s_1 | u_1)
\]
\[
= \frac{T + 1}{2} \log \left( 2\pi e \left( \frac{1}{D} + 1 \right)^{-1} \right).
\] (5.142)

As \( D \to 0 \), it can be easily observed that the lower and upper bound of (5.140) in (5.141) and (5.142) and thus (5.140) itself can be written as.
\[
\frac{T + 1}{2} \log (2\pi e D) + o(D).
\] (5.143)

By replacing (5.139) and (5.143) in (5.134) when \( D \to 0 \), (5.132) is derived which completes the proof.

### 5.7 Lossless Streaming with Ideal-Playback

#### 5.7.1 Converse

Consider \( K + 1 \) periods of a periodic erasure channel with period \( p = B + T \), such that at each period the first \( B \) channel outputs are erased and the rest \( T \) channel outputs are perfectly revealed to the decoder.
Now consider the following entropy term.

\[ n(K + 1)TR \geq H([f]_{B}^{B+T-1}, [f]_{p+B}^{2p-1}, [f]_{2p+B}^{3p-1}, \ldots, [f]_{(K-1)p+B}^{Kp-1}) \]

(5.144)

\[ \geq H([f]_{B}^{B+T-1}, [f]_{p+B}^{2p-1}, \ldots, [f]_{(K-1)p+B}^{Kp-1} | s_{-1}). \]

(5.145)

**Step 1:** We apply the following Fano’s inequality.

\[ H([s]_{0}^{B-1} | [f]_{B}^{B+T-1}, s_{-1}) \leq nB\epsilon_n. \]

(5.146)

Then we can derive the following lower bound on (5.145).

\[
H([f]_{B}^{B+T-1}, [f]_{p+B}^{2p-1}, \ldots, [f]_{(K-1)p+B}^{Kp-1} | s_{-1}) \\
= H([s]_{0}^{B-1} | s_{-1}) + H([f]_{B}^{B+T-1}, \ldots, [f]_{(K-1)p+B}^{Kp-1} | s_{-1}) - nB\epsilon_n \\
= nB H(s_{1} | s_{0}) + H([f]_{B}^{B+T-1}, \ldots, [f]_{(K-1)p+B}^{Kp-1} | s_{-1}) - nB\epsilon_n. \]

(5.147)

**Step 2:** We first derive a lower bound on the second term in (5.147) by conditioning the entropy term as follows.

\[
H([f]_{B}^{B+T-1}, \ldots, [f]_{(K-1)p+B}^{Kp-1} | s_{-1}) \geq H([f]_{B}^{B+T-1}, \ldots, [f]_{(K-1)p+B}^{Kp-1} | s_{-1}^{B-1}, [f]_{0}^{B-1}) \]

(5.148)

Then we apply the following Fano’s inequality,

\[
H([s]_{B}^{B+T-1} | s_{0}^{B+T-1}, [f]_{2B+T}^{B+2T-1}, s_{-1}) \leq nT\epsilon_n, \]

(5.149)

to lower bound (5.159) as follows.

\[
H([f]_{B}^{B+T-1}, \ldots, [f]_{(K-1)p+B}^{Kp-1} | s_{-1}, [f]_{0}^{B-1}) \\
\geq H([s]_{B}^{B+T-1} | s_{-1}, [f]_{0}^{B-1}) + H([f]_{B}^{B+T-1}, \ldots, [f]_{(K-1)p+B}^{Kp-1} | s_{-1}^{B+T-1}, [f]_{0}^{B-1}) - nT\epsilon \\
\geq nTH(s_{1} | s_{0}) + H([f]_{p+B}^{2p-1}, \ldots, [f]_{(K-1)p+B}^{Kp-1} | s_{-1}^{p-1}, [f]_{0}^{p-1}) - nT\epsilon. \]

(5.150)

From (5.147) and (5.150), we have

\[
H([f]_{B}^{B+T-1}, [f]_{p+B}^{2p-1}, \ldots, [f]_{(K-1)p+B}^{Kp-1} | s_{-1}) \\
= n(B + T) H(s_{1} | s_{0}) + H([f]_{p+B}^{2p-1}, \ldots, [f]_{(K-1)p+B}^{Kp-1} | s_{-1}^{p-1}, [f]_{0}^{p-1}) - n(B + T)\epsilon_n. \]

(5.151)

**Step 3:** Now we apply the same method in step 1 and 2, to lower bound the second term in (5.151).

If we repeat this up to the \(K\)-th period, we can derive the following bound on rate.

\[
n(K + 1)TR \geq nK (B + T) H(s_{1} | s_{0}) + H([f]_{(K-1)p+B}^{(K-1)p-1} | s_{-1}^{p-1}, [f]_{0}^{(K-1)p-1}) - nK (B + T)\epsilon_n. \]

(5.152)

Finally by taking the limit \(K \rightarrow \infty\) and then \(n \rightarrow \infty\), the lower bound on rate as

\[
R \geq \frac{B + T}{T} H(s_{1} | s_{0}) - \frac{B + T}{T} \epsilon_n \]

(5.153)
is derived. This completes the converse proof.

### 5.7.2 Achievability

In this section we need to show that any rate $R$ satisfying

$$R > \frac{B + T}{T} H(s_t|s_0)$$

(5.154)

is achievable. We show that the coding scheme based on separation of the source and channel coding is optimal. In particular, the source encoder performs the optimal source coding, i.e., predictive coding, assuming the channel is ideal bit pipe. The encoder at each time $t$, while observing the source sequence $s^n_t$, produces the source code $m_t \in \{1 : 2^{nR_s}\}$ with the conditional entropy rate

$$R_s = H(s_t|s_{t-1}) = H(s_1|s_0).$$

Now we apply delay-optimal erasure burst code, based on the following results.

**Theorem 17.** ([29]) The delay-optimal erasure burst code with the rate $R$ can correct all erasure bursts of length $B$ with decoding delay $T$ if

$$\frac{T}{B} \geq \max \left[1, \frac{R}{1-R} \right].$$

(5.155)

This is the smallest attainable rate and thus, the code is delay-optimal.

**Remark 15.** According to Theorem 17, if $T \geq B$, there exists channel codes with the optimal rate $R = (B + T)/T$ which guarantees the correct recovery of the channel inputs with a delay of $T$, for the erasure burst channel model we considered in this chapter. Note that such an optimal rate is not achievable by random channel codes and the delay-optimal erasure burst code introduced in [29,48] exploits the structure of the erasure burst channel.

The delay-optimal erasure burst code of Theorem 17 is applied on the $nR_s$ bits of the source code output at each time $t$ to achieve the capacity of $T/(T + B)$. This requires $R = nR_sT/(T + B)$ channel use per source sequence which achieves the rate in (5.154). Note that the delay-optimal erasure burst code guarantees the recovery of the source codes within delay $T$.

### 5.8 Lossy Streaming of Gauss-Markov Sources with Ideal-Playback

#### 5.8.1 Converse

In this section we establish lower bound on the streaming rate with ideal-playback for Gauss-Markov sources. First note that $R_{GLM}(B,T, D)$, i.e., the lower bound on the rate for the streaming with controlled-interruption, also establishes a lower bound on the rate for the streaming with ideal-playback. In fact, instead of deriving lower bound for streaming with ideal-playback scenario, we assume the case where the decoder is not required to reproduce the source sequences for which the channel packets are erased. In what follows we derive a separate lower bound on the rate. The rate has to simultaneously
satisfy both lower bounds and the general lower bound in Theorem 7 is the maximum of the two lower bounds.

**Case \( T > B \)**

Similar to the converse proof in Section 5.7.1, we consider \( K + 1 \) periods of a periodic erasure channel with period \( p = B + T \). We have

\[
n(K + 1)TR \geq \mathcal{H}([f]_{B+T}^{B+T-1}, [f]_{p+B}^{2p-1}, [f]_{(K-1)p+B}^{3p-1}, \ldots, [f]_{(K-1)p+B}^{Kp-1})
\]

\[
\geq \mathcal{H}([f]_{B+T}^{B+T-1}, [f]_{p+B}^{2p-1}, \ldots, [f]_{(K-1)p+B}^{Kp-1}, [s]_{-1}).
\] (5.156)

**Step 1:** According to the problem setup, if \( B \) consecutive channel outputs of times \( \{0, \ldots, B-1\} \) are erased by the channel, we have

\[
h([s]_{0}^{B-1}[[f]_{B}^{B+T-1}, [s]_{-1}) \leq \frac{nB}{2} \log (2\pi eD).
\] (5.157)

Then we can derive the following lower bound on (5.156).

\[
\mathcal{H}([f]_{B+T}^{B+T-1}, [f]_{p+B}^{2p-1}, \ldots, [f]_{(K-1)p+B}^{Kp-1}, [s]_{-1})
\]

\[
= I([s]_{0}^{B-1}, [f]_{B+T}^{B+T-1}, \ldots, [f]_{(K-1)p+B}^{Kp-1}, [s]_{-1}) + \mathcal{H}([f]_{B+T}^{B+T-1}, \ldots, [f]_{(K-1)p+B}^{Kp-1}, [s]_{-1})
\]

\[
\geq h([s]_{0}^{B-1}([s]_{-1} - \frac{nB}{2} \log (2\pi eD) + \mathcal{H}([f]_{B+T}^{B+T-1}, \ldots, [f]_{(K-1)p+B}^{Kp-1}, [s]_{-1})
\]

\[
= \frac{nB}{2} \log \left( \frac{1 - \rho^2}{D} \right) + \mathcal{H}([f]_{B+T}^{B+T-1}, \ldots, [f]_{(K-1)p+B}^{Kp-1}, [s]_{-1}).
\] (5.158)

**Step 2:** We first derive a lower bound on the second term in (5.158) by conditioning the entropy term as follows.

\[
\mathcal{H}([f]_{B+T}^{B+T-1}, \ldots, [f]_{(K-1)p+B}^{Kp-1}, [s]_{-1}) \geq \mathcal{H}([f]_{B+T}^{B+T-1}, \ldots, [f]_{(K-1)p+B}^{Kp-1}, [s]_{-1}, [f]_{0}^{B-1}).
\] (5.159)

Then we use the fact that,

\[
h([s]_{B+T}^{B+T-1}([f]_{0}^{B+T-1}, [f]_{2B+T-1}^{B+T-1}, [s]_{-1}) \leq \frac{nT}{2} \log (2\pi eD).
\] (5.160)

to lower bound (5.159) as follows.

\[
\mathcal{H}([f]_{B+T}^{B+T-1}, \ldots, [f]_{(K-1)p+B}^{Kp-1}, [s]_{-1}, [f]_{0}^{B-1})
\]

\[
\geq h([s]_{B+T}^{B+T-1}, [s]_{-1}, [f]_{0}^{B-1}) + \mathcal{H}([f]_{B+T}^{B+T-1}, \ldots, [f]_{(K-1)p+B}^{Kp-1}, [s]_{-1}, [f]_{0}^{B-1}) - \frac{nT}{2} \log (2\pi eD)
\]

\[
= \frac{nT}{2} \log \left( \frac{1 - \rho^2}{D} \right) + \mathcal{H}([f]_{p+B}^{2p-1}, \ldots, [f]_{(K-1)p+B}^{Kp-1}, [s]_{-1}, [f]_{0}^{B-1})).
\] (5.161)

From (5.158) and (5.161), we have

\[
\mathcal{H}([f]_{B+T}^{B+T-1}, [f]_{p+B}^{2p-1}, \ldots, [f]_{(K-1)p+B}^{Kp-1}, [s]_{-1})
\]
Finally by taking the limit $K \to \infty \text{ and then } n \to \infty$, the following lower bound on rate is derived.

$$R \geq \frac{B + T}{2T} \log \left( \frac{1 - \rho^2}{D} \right). \quad (5.164)$$

**Case $T = B$**

Now consider the case that $T = B$. In this case, unlike before, we are able to improve the lower bound derived in the previous section, by considering a single erasure burst channel rather than a periodic one. In particular, we assume that an erasure burst of length $B$ spans the interval $\{j - B + 1, \ldots, j\}$. We have

$$TnR \geq H([f]_{j+1}^{j+B} | [f]_0^{j-B}, s_{-1})$$

$$= I([s]_{j-B+1}^{j-B} | [f]_{j+1}^{j+B}, [f]_0^{j-B}, s_{-1}) + H([f]_{j+1}^{j+B} | [f]_0^{j-B}, s_{-1}, [s]_{j-B+1}^{j-B})$$

$$= h([s]_{j-B+1}^{j-B} | [f]_0^{j-B}, s_{-1}) - h([s]_{j-B+1}^{j-B} | [f]_0^{j-B}, [f]_{j+1}^{j+B}, s_{-1}) + H([f]_{j+1}^{j+B} | [f]_0^{j-B}, s_{-1}, [s]_{j-B+1}^{j-B}). \quad (5.165)$$

The first term in (5.165), can be written as follows.

$$h([s]_{j-B+1}^{j-B} | [f]_0^{j-B}, s_{-1}) = h(s_{j-B+1} | [f]_0^{j-B}, s_{-1}) + n(B - 1)h(s_1 | s_0)$$

$$\geq \frac{n}{2} \log \left( \rho^2 e^{2h(s_{j-B+1}^{j-B} | [f]_0^{j-B}, s_{-1}) - 2\pi e(1 - \rho^2)} \right) + \frac{n(B - 1)}{2} \log \left( 2\pi e(1 - \rho^2) \right)$$

$$\geq \frac{n}{2} \log \left( \rho^2 e^{2h(s_{j-B+1}^{j-B} | [f]_0^{j-B}, s_{-1}) - 2\pi e(1 - \rho^2)} \right) + \frac{n(B - 1)}{2} \log \left( 2\pi e(1 - \rho^2) \right)$$

$$= \frac{n}{2} \log \left( \frac{2^2R}{2nR - \rho^2} \right) + \frac{nB}{2} \log \left( 2\pi e(1 - \rho^2) \right), \quad (5.166)$$

where (5.166) follows from the application of Shannon’s EPI and (5.167) follows from the application of Lemma 6 for large $j$.

The second term in (5.165), is upper bounded as follows.

$$h([s]_{j-B+1}^{j-B} | [f]_0^{j-B}, [f]_{j+1}^{j+B}, s_{-1}) \leq \frac{nB}{2} \log \left( 2\pi eD \right). \quad (5.169)$$

The third term in (5.165), is lower bounded as follows.

$$H([f]_{j+1}^{j+B} | [f]_0^{j-B}, s_{-1}, [s]_{j-B+1}^{j-B}) \geq H([f]_{j+1}^{j+B} | [f]_0^{j-B}, s_{-1}, [s]_{j-B+1}^{j-B})$$

$$\geq I([s]_{j+1}^{j+B} | [f]_0^{j-B}, s_{-1}, [s]_{j-B+1}^{j-B})$$
\section{Delay-Constrained Streaming}

\begin{equation}
\begin{aligned}
&= h([s]_{j+1}^{B}|[f]_0^j, s_{-1}, [s]_{j-B+1}^j) - h([s]_{j+1}^{B+1}|[f]_0^{j+1}, s_{-1}, [s]_{j-B+1}^j) \\
&\geq \frac{nB}{2} \log \left( \frac{1-\rho^2}{D} \right). 
\end{aligned}
\tag{5.170}
\end{equation}

By replacing (5.168), (5.169) and (5.170) into (5.165), we have that
\begin{equation}
2^{2(T-1)R} \geq \frac{1}{2^{2R}} \left( \frac{1-\rho^2}{D} \right)^{2B}.
\tag{5.171}
\end{equation}

By solving (5.171) for $R$, the following lower bound is derived.
\begin{equation}
R_{\text{IP,GM}}(B, T, D) \geq \bar{R}_{\text{IP,GM}}(B, T, D) \triangleq \begin{cases} 
\frac{B+T}{2T} \log \left( \frac{1-\rho^2}{D} \right), & T > B \\
\frac{1}{2} \log (x^*) , & T = B,
\end{cases}
\tag{5.172}
\end{equation}

As mentioned the general lower bound of Proposition 7 is derived by combining the two lower bounds. This completes the proof.

\section{Achievability}

In this section we propose two coding scheme for lossy streaming of Gauss-Markov sources over erasure burst channel with ideal-playback delay-constrained decoders. The overall achievable rate in Proposition 8 is the least achievable rate of the two coding scheme. Note that the second coding scheme requires $T > 1$ and a upper bound on $D$ as sill be clear in sequel. The block diagram of the two coding schemes are shown in Fig. 5.4.

**Coding Scheme I**

The proposed coding scheme is based on separation of the source-channel coding. The encoder at time $t$, first applies the predictive coding on the source sequence $s_t$. In the predictive coding scheme for ideal channel without erasures, the encoder at each time first estimates the source sequence, using the information from the past. Then it quantizes the estimation error, known as innovation process, and sends the quantization codewords through the channel. In particular for the problem setup considered in this chapter, the decoder at each time $t \geq 0$ computes the estimation error $q^n_t$ as
\begin{equation}
q^n_t = s^n_t - E\{s^n_t|s_{-1}, [u^n]|_0^{t-1}\},
\tag{5.173}
\end{equation}
which is further quantized via the Gaussian test channel
\begin{equation}
q_t = u_t + z_t,
\tag{5.174}
\end{equation}
where $z_t \sim \mathcal{N}(0, \sigma^2_z)$ is independent noise. The quantization rate $R$ satisfies.
\begin{equation}
R \geq I(q_t; u_t) = \frac{1}{2} \log \left( \frac{\sigma_q^2}{\sigma_z^2} \right) \\
= \frac{1}{2} \log \left( 1 + \frac{\sigma_u^2}{\sigma_z^2} \right),
\tag{5.175}
\end{equation}
where $\sigma_q^2$ and $\sigma_u^2$ are the estimation error (the variance of the innovation process) and the variance of the quantization codewords, respectively. It is not hard to observe that at any time $t$ the source sequence $s_t^n$ can be written as follows.

$$s_t^n = \rho^{t+1}s_{t-1}^n + \rho^tu_0^n + \rho^{t-1}u_1^n + \ldots + \rho u_{t-1}^n + \rho z_{t-1}^n + n_t^n$$

(5.176)

Thus the encoder computes the estimation error $q_t^n$ and quantization codewords $u_t^n$ as

$$q_t^n = \rho z_{t-1}^n + n_t^n = u_t^n + z_t^n.$$  

(5.177)

From (5.177), we have that $\sigma_u^2 = (1 - \rho^2)(1 - \sigma_z^2)$. By replacing this in (5.175), the predictive coding rate $R_s$ can be computed as follows.

$$R_s = \frac{1}{2} \log \left( \frac{\rho^2 + 1 - \rho^2}{\sigma_z^2} \right).$$

(5.178)

The decoder at each time provides an estimate of the source sequence $s_t^n$ as

$$\hat{s}_t^n = \rho^{t+1}s_{t-1}^n + \sum_{k=0}^t \rho^{t-k}u_k^n$$

(5.179)

with the estimation error $\sigma_z^2$. Thus the rate-distortion pair $(R,D)$ as

$$R \geq R_s(D) \triangleq \frac{1}{2} \log \left( \frac{1 - \rho^2}{D} + \rho^2 \right)$$

(5.180)

is achievable by the predictive coding.

The encoder then applies a delay-optimal erasure burst code of Theorem 17 on the $nR_s(D)$ bits of the source code output at each time $t$ to achieve the erasure burst channel capacity of $T/(T + B)$. This requires $R = nR_s(D)/T/(T + B)$ channel use per source sequence which achieves the rate

$$R_c^+(B,T,D) \triangleq \frac{T + B}{2T} \log \left( \frac{1 - \rho^2}{D} + \rho^2 \right).$$

(5.181)

This completes the achievability of the first scheme.

**Coding Scheme II**

The second coding scheme is again a combination of a source and a delay-optimal erasure burst channel code [48]. We first focus on the source code. Define the source sequence

$$x_t^n \triangleq s_t^n - \rho^{t+1}s_{t-1}^n$$

(5.182)

and assume that there are $M + 1$ source sequences with time indices $\{0, 1, \ldots, M\}$. Also define the following notation.

$$x \triangleq [x_0, x_1, \ldots, x_M]^\dagger.$$  

(5.183)
Because the source sequences $s^n$ is known at the decoder, the encoder is only interested in transmitting $x^n$ to the decoder. We have

$$x = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ \rho & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \rho^M & \rho^{M-1} & \cdots & 1 \end{pmatrix} \begin{pmatrix} n_0 \\ n_1 \\ \vdots \\ n_M \end{pmatrix} \triangleq A_n. \quad (5.184)$$

Note that by this definition the covariance matrix of the source vector $x$ is $\Sigma_X = (1 - \rho^2)AA^\dagger$. For the joint-coding problem when the encoder has access to all the source sequences beforehand, the forward Gaussian test channel

$$\hat{x} = x + z \quad (5.185)$$

with $z \sim \mathcal{N}(0, DI)$ and $\hat{x} \sim \mathcal{N}(0, \Sigma_X - DI)$, achieves the optimal sum-rate of the joint rate-distortion problem, when $(\Sigma_X - DI) \succeq 0$, i.e., the test channel achieves the sum-rate,

$$\sum_{k=0}^{M-1} R_k = \frac{1}{2} \log \left( \frac{(1 - \rho^2)^M \det(AA^T)}{D^M} \right) \quad (5.186)$$

$$= \frac{1}{2} \log \left( \frac{(1 - \rho^2)^M (\det(A))^2}{D^M} \right)$$

$$= \frac{M}{2} \log \left( \frac{1 - \rho^2}{D} \right). \quad (5.187)$$

By considering the time-independent rate $R_t = R$, the optimal achievable rate is as follows.

$$R \geq \frac{1}{2} \log \left( \frac{1 - \rho^2}{D} \right). \quad (5.188)$$

First consider the following lemma that characterizes the range of distortion $D$ for which the non-negativity constraints in $(\Sigma_X - DI) \succeq 0$ holds and the rate in (5.188) is achievable.

**Lemma 29.** For sufficiently large $M$ and $D \leq (1 - \rho)/(1 + \rho)$, $(\Sigma_X - DI) \succeq 0$. \hfill $\square$

**Proof.** See Appendix D.4 \hfill $\square$

The general forward test channel framework, similar to the one in (5.185), does not guarantee any delay constraint at the encoder. In particular, the codeword $\hat{x}_t$ can be a function of all the sources $\{x_0, x_1, \ldots, x_M\}$. However, interestingly, the test channel in (5.185), requires a lookahead of length one. See the following lemma.

**Lemma 30.** The test channel in (5.185) can operate with delay of length one over the source sequences, i.e., the codeword $\hat{x}_0$ can be computed from the sources $\{x_0, x_1, \ldots, x_{t+1}\}$. \hfill $\square$

**Remark 16.** Results similar to Lemma 30 are reported in [17, Corollary 3.2]. In particular, the paper establishes the sum-rate optimality of one-stage delayed systems for Gauss-Markov sources and quadratic
distortion. Here we provide an alternative proof to derive the achievable rate (and not the sum-rate). In addition for the case of equal distortion constraint, i.e., $D_i = D$ and sufficiently large communication duration $M$, we derive an explicit distortion constraint in Lemma 29.

Proof. First note that from the application of the standard MMSE operation, the backward test channel in (5.185) can be equivalently expressed by the following forward expression.

\[ \hat{x} = \hat{A}x + \hat{z}, \]  
\[ \text{(5.189)} \]

where

\[ \hat{A} = \Sigma_{XX} \Sigma_{X}^{-1} \]
\[ = (\Sigma_X - DI) \Sigma_{X}^{-1} = I - D \Sigma_{X}^{-1} \]  
\[ \text{(5.190)} \]

and the covariance of the noise vector $\hat{z}$ is expressed as follows.

\[ \Sigma_{\hat{z}} = \Sigma_X - \Sigma_{XX} \Sigma_{X}^{-1} \Sigma_X \hat{x} \]
\[ = (\Sigma_X - DI) - (\Sigma_X - DI) \Sigma_{X}^{-1} (\Sigma_X - DI) \]
\[ = DI - D^2 \Sigma_{X}^{-1} = D\hat{A}. \]  
\[ \text{(5.191)} \]

From (D.32) in Appendix D.4, it can be observed that the matrix $\hat{A}$ is in the following form.

\[ \hat{A} = \begin{pmatrix} a_1 & a_2 & 0 & 0 & \cdots \\ a_2 & a_1 & a_2 & 0 & \cdots \\ 0 & a_2 & a_1 & a_2 & \cdots \\ 0 & 0 & a_2 & a_1 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}, \]  
\[ \text{(5.192)} \]

where

\[ a_1 \triangleq 1 - D \frac{1 + \rho^2}{1 - \rho^2} \]  
\[ a_2 \triangleq \frac{\rho D}{1 - \rho^2}. \]  
\[ \text{(5.193, 5.194)} \]

In addition it is not hard to observe that $\hat{A} = \hat{B}\hat{B}^\dagger$, where

\[ \hat{B} = \begin{pmatrix} b_1 & b_2 & 0 & 0 & \cdots \\ 0 & b_1 & b_2 & 0 & \cdots \\ 0 & 0 & b_1 & b_2 & \cdots \\ 0 & 0 & 0 & b_1 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}, \]  
\[ \text{(5.195)} \]

where

\[ b_1 \triangleq \frac{1}{2} (\sqrt{a_1 + 2a_2} + \sqrt{a_1 - 2a_2}) \]  
\[ \text{(5.196)} \]
\[ b_2 \doteq \frac{1}{2} \left( \sqrt{a_1 + 2a_2} - \sqrt{a_1 - 2a_2} \right). \]  

(5.197)

Thus the test channel in (5.189) can be written as follows.

\[ \hat{x} = \hat{A}x + \hat{B}z_1, \]  

(5.198)

where \( z_1 \sim \mathcal{N}(0, DI) \). From (5.192), (5.195) and (5.198), it can be verified that the codeword \( \hat{x}_t \) can be computed with unit source delay as function of the sources \( \{x_0, x_1, \ldots, x_{t+1}\} \). This completes the proof.

According to Lemma 30, the rate in (5.188) is achieved by source code with an encoder with lookahead of length one and ideal channel without erasure. Note that this lookahead causes a delay of one at the decoder. If we apply a delay-optimal erasure burst code of Theorem 17 with delay \( T - 1 \) over the source code, the ideal-playback streaming of the source sequences within an overall delay of \( T \) is guaranteed. This completes the achievability of the second scheme.

### 5.9 Conclusion

In this chapter we considered the streaming of Markov sources over erasure burst channels with delay-constrained decoder. The rate-\( R \) causal encoder generates the channel inputs to be sent through the channel. The channel may introduce a single erasure burst of length \( B \) in an unknown location during the transmission period. We studied two classes of decoders with delay constraint.

In the first setup, i.e., streaming with controlled interruption, the decoder is required to reconstruct the source sequences with a delay of \( T \), except the source vectors associated with the erasure times and a window of length \( W \) after the erasure burst ends. For the case of discrete sources and lossless recovery, we derived upper and lower bounds for the minimum rate which coincide when \( W = 0 \) and when either \( W \) or \( T \) becomes very large. This can be viewed as the generalization of the upper and lower bounds of zero-delay case in Chapter 2. We also extended the results to the channels with multiple erasure bursts and guard interval between the consecutive erasure bursts. For the Gauss-Markov sources and quadratic distortion measure, we derived upper and lower bounds which coincide in high resolution asymptotic.

In the second setup, i.e., streaming with ideal-playback, the decoder is required to reconstruct all the source sequences with a delay of \( T \). For the case of discrete sources and lossless recovery, we established the optimal rate which is achieved by source-channel separation. The source code is the zero-delay predictive code and the channel code is the delay-optimal channel code. For the Gauss-Markov sources and quadratic distortion measure, we derived upper and lower bounds which coincide in high resolution and large delay regimes. The two coding scheme was provided again based on source-channel separation. The first coding scheme consists of the zero-delay predictive code as the source code and delay-optimal channel code with delay \( T \). The second coding scheme however, consists of unit-delay optimal source code and delay-optimal channel code with delay \( T - 1 \). We established the exact distortion constraint for the optimality of unit-delay source code and thus, the achievability of the second coding scheme. This completes the scenario of streaming with delay-constraint.
Chapter 6

Conclusion

A hair divides what is false and true.

Omar Khayyam

Motivated by real-time multimedia streaming applications, we studied the streaming of Markov sources over burst erasure channels. When the underlying channel is an ideal bit-pipe, the predictive coding scheme (or conditional source coding for discrete sources) is known to attain the optimal rate. In practice, however, packet losses are unavoidable and the predictive schemes exhibit a significant amount of error propagation under such imperfect channel constraints. Several heuristic techniques have been developed to overcome the effect of packet losses. However even the effect of single erasure was not well understood. In this thesis we proposed to study the fundamental trade off between the compression efficiency and error propagation in real-time streaming over channels with packet losses. While the coding theorems were established for a somewhat specialized setup of burst erasure channels and spatially i.i.d. source vector we believe that the insights developed from our study can be applied to a much broader class of source and channel models.

The common approach for achievability in this thesis is the quantize and binning technique. For the case of discrete sources and lossless recovery in Chapter 2, we proposed an achievable rate based on memoryless binning where the quantization step was not used. However for a special class of semi-deterministic sources we showed that a judicious quantization step followed by binning is optimal. For the case of Gauss-Markov sources and quadratic distortion measure we also studied a quantize and binning strategy in Chapter 4. Our proposed scheme includes an imperfect prediction of the source sequence from past sequences, quantization of the resulting error sequence, and binning. It includes predictive coding and memoryless quantize and binning as special cases and demonstrates considerable improvements over these. By examining the structure of the test channel in our proposed scheme, we obtained insights into the performance gains over baseline schemes over statistical channels such as the Gilbert-Elliott channel and i.i.d. erasure channels. As such the coding theorems for the lossy case are considerably more difficult than the lossless case as the reconstruction sequences do not inherit the Markov property. We also developed lower bounds on the minimum required compression rate by exploiting connection to multi-terminal source coding problems that capture similar constraints as in the streaming problem. The lower bound was used to establish the optimality of our coding schemes for some special cases. Finally in Chapter 5 we extended our results to the case when the decoder is permitted to reconstruct
each source sequence with a fixed delay of $T$.

We believe that the present work can be extended in a number of directions.

• **New Open Problems in Network Information Theory:** The present thesis introduces open problems in network information theory which, to the best of our knowledge, are not addressed in the literature. The upper and lower bounds on the lossless rate-recovery function, studied in Chapter 2, do not generally coincide except for some special cases. Except for specific range of parameters, the optimality of the hybrid coding scheme, introduced in Chapter 4, is not known even in the high resolution regime. Establishing the rate-recovery function will lead to either better streaming strategies, by improving the upper bounds, or highlighting the importance of proposed streaming strategies, by proving their optimality while improving lower bounds.

• **Extensions in Problem Setup:** The problem setup can be extended in several directions.
  
  – In the proposed setup for the rate-recovery function, we considered a time interval of length $W$ during which the decoder is allowed to declare a complete outage, and no reconstruction is necessary. This model may be further generalized by considering partial recovery with a higher distortion during such a recovery period.
  
  – This thesis only addressed the case of lossless recovery for discrete sources. Extensions to lossy reconstruction, analogous to the case of Gaussian sources in Chapters 2 and 3, may be interesting. However, this will require characterization of the worst-case erasure sequence for a general source model, which appears challenging.
  
  – In this thesis we only focused on the case of causal encoders, i.e. the source sequence of time $t$ is revealed to the encoder at time $t$. One interesting extension is to consider more general encoders with finite-lookahead.
  
  – In this thesis, we only considered the rate-recovery function for erasure burst channels. Motivated by burst-erasure channel models such as Gilbert-Elliott model, one extension will be considering channels that introduce both erasure bursts and isolated erasures as considered recently in the channel coding context [30].

Studying the behaviour of suggested generalized rate-recovery function will undoubtedly lead to a finer understanding of tradeoffs between compression rate and error propagation in video transmission systems.
Appendices
Appendix A

Zero-Delay Lossless Streaming

A.1 Proof of Corollary 1: Alternative Expression for Upper Bound

We want to show the following equality.

\[ R^+(B, W) \triangleq H(s_1|s_0) + \frac{1}{W+1}I(s_B; s_{B+1}|s_0) \]
\[ = \frac{1}{W+1}H(s_{B+1}, s_{B+2}, \ldots, s_{B+W+1}|s_0). \]  \hspace{1cm} (A.1)

According to the chain rule of entropies, the term in (A.1) can be written as

\[ H(s_{B+1}, s_{B+2}, \ldots, s_{B+W+1}|s_0) = H(s_{B+1}|s_0) + \sum_{k=1}^{W} H(s_{B+k+1}|s_0, s_{B+1}, \ldots, s_{B+k}) \]
\[ = H(s_{B+1}|s_0) + WH(s_1|s_0) \] \hspace{1cm} (A.2)
\[ = H(s_{B+1}|s_0) - H(s_{B+1}|s_B, s_0) + H(s_{B+1}|s_B, s_0) + WH(s_1|s_0) \] \hspace{1cm} (A.3)
\[ = H(s_{B+1}|s_0) - H(s_{B+1}|s_B, s_0) + H(s_{B+1}|s_B) + WH(s_1|s_0) \] \hspace{1cm} (A.4)
\[ = I(s_{B+1}; s_B|s_0) + (W + 1)H(s_1|s_0) \] \hspace{1cm} (A.5)
\[ = (W + 1)R^+(B, W), \] \hspace{1cm} (A.6)

where (A.2) follows from the Markov property

\[ (s_0, s_{B+1}, \ldots, s_{B+k-1}) \rightarrow s_{B+k} \rightarrow s_{B+k+1} \] \hspace{1cm} (A.7)

for any \( k \) and from the stationarity of the sources which for each \( k \) implies that

\[ H(s_{B+k+1}|s_{B+k}) = H(s_1|s_0). \] \hspace{1cm} (A.8)

Note that in (A.3) we add and subtract the same term and (A.4) also follows from the Markov property of (A.7) for \( k = 0 \).
A.2 Proof of Lemma 4: Transformation $\mathcal{L}_f$

First let us define the following notations.

- For a vector $\mathbf{x}$ of size $x$, define $\mathbf{x}^{(u,a)}$ and $\mathbf{x}^{(d,a)}$ such that

$$
\mathbf{x} = \begin{bmatrix}
    (x-a) & \mathbf{x}^{(u,a)} \\
    \mathbf{x}^{(d,a)}
\end{bmatrix},
$$

(A.9)

- For a matrix $\mathbf{X}$ of size $x \times y$, define $\mathbf{X}^{(l,a)}$, $\mathbf{X}^{(r,a)}$, $\mathbf{X}^{(u,b)}$ and $\mathbf{X}^{(d,b)}$ as

$$
\mathbf{X} = \begin{bmatrix}
    (y-a) & \mathbf{X}^{(l,a)} \\
    (x-a) & \mathbf{X}^{(r,a)}
\end{bmatrix},
$$

(A.10)

and

$$
\mathbf{X} = \begin{bmatrix}
    (x-b) & \mathbf{X}^{(u,b)} \\
    (x-b) & \mathbf{X}^{(d,b)}
\end{bmatrix},
$$

(A.11)

- For a square matrix $\mathbf{X}$ of size $x$, define matrices $\mathbf{X}^{(ul,a)}$, $\mathbf{X}^{(ur,a)}$, $\mathbf{X}^{(dl,a)}$ and $\mathbf{X}^{(dr,a)}$ such that

$$
\mathbf{X} = \begin{bmatrix}
    (x-a) & \mathbf{X}^{(ul,a)} \\
    (x-a) & \mathbf{X}^{(ur,a)} \\
    \mathbf{X}^{(dl,a)} & \mathbf{X}^{(dr,a)}
\end{bmatrix}.
$$

(A.12)

We introduce an iterative method to define the transformation $\mathcal{L}_f$.

**Step 0:** If $\mathbf{A} = \mathbf{0}$ or $N_1 = N_d$, the source is in the form of (2.109). Thus $\mathcal{L}_f(s_i) = s_i$. Otherwise, continue to next step.

**Step 1:** Without loss of generality we assume that the first $N_1$ rows of matrix $\mathbf{A}$ are independent.\(^1\)

Let $\mathbf{R}_{1,0}$ denotes the first $N_1$ rows of $\mathbf{A}$ and

$$
\mathbf{A}^{(d,N_1)} = \mathbf{V}_1 \mathbf{R}_{1,0},
$$

(A.13)

where $\mathbf{V}_1$ is an $(N_d - N_1) \times N_1$ matrix relating dependent rows of $\mathbf{A}$ to $\mathbf{R}_{1,0}$. Also define invertible square matrix $\mathbf{M}_1$ as

$$
\mathbf{M}_1 \triangleq \begin{bmatrix}
    N_1 & (N_d - N_1) \\
    (N_d - N_1) & \mathbf{I} \\
    \mathbf{0}_1 & \mathbf{M}_1
\end{bmatrix},
$$

(A.14)

---

\(^1\)By rearranging the rows of matrices $\mathbf{A}$ and $\mathbf{B}$, this assumption can always be satisfied.
Appendix A. Zero-Delay Lossless Streaming

Note that
\[ M_{1}^{-1} = \begin{bmatrix} I & 0 \\ V_{1} & I \end{bmatrix}. \]  
(A.15)

Define
\[ \begin{pmatrix} s_{i,1} \\ \bar{s}_{i,1} \end{pmatrix} \triangleq \begin{pmatrix} (M_{1} s_{i,d})^{(u,N_{1})} \\ (M_{1} s_{i,d})^{(d,N_{1})} \end{pmatrix} = M_{1} s_{i,d}. \]  
(A.16)

We have
\[ \begin{pmatrix} s_{i,1} \\ \bar{s}_{i,1} \end{pmatrix} = (\begin{pmatrix} M_{1} A M_{1}^{-1} \\ M_{1} s_{i-1,d} \end{pmatrix}) \begin{pmatrix} s_{i-1,0} \\ M_{1} s_{i-1,d} \end{pmatrix} \]  
(A.17)

\[ = \begin{pmatrix} R_{1,0} & (M_{1} B M_{1}^{-1})^{(u,N_{1})} \\ 0 & (M_{1} B M_{1}^{-1})^{(d,N_{1})} \end{pmatrix} \begin{pmatrix} s_{i-1,0} \\ (M_{1} s_{i-1,d})^{(u,N_{1})} \\ (M_{1} s_{i-1,d})^{(d,N_{1})} \end{pmatrix} \]  
(A.18)

\[ = N_{1} \begin{bmatrix} N_{0} & N_{1} & N_{d-N_{1}} \\ R_{1,0} & R_{1,1} & R_{1,2} \\ 0 & A^{(1)} & B^{(1)} \end{bmatrix} \begin{pmatrix} s_{i-1,0} \\ s_{i-1,1} \\ s_{i-1,1} \end{pmatrix}, \]  
(A.19)

where \( A^{(1)} = (M_{1} B M_{1}^{-1})^{(d,N_{1})} \) and \( B^{(1)} = (M_{1} B M_{1}^{-1})^{(d,N_{1})} \) and the other matrices are defined similarly. Up to now \( s_{i,1} \) is defined.

**Step 2:** Define \( N_{2} \triangleq \text{Rank}(A^{(1)}). \) Generally
\[ N_{2} \leq \min\{N_{1}, N_{d} - N_{1}\}. \]  
(A.20)

If \( N_{2} = N_{d} - N_{1} \) or if \( A^{(1)} \) is zero matrix, set \( s_{i,2} = s_{i,1} \) and
\[ L_{f}(s_{i}) = \begin{pmatrix} s_{i,0} \\ s_{i,1} \\ s_{i,1} \end{pmatrix}. \]  
(A.21)

If \( A^{(1)} \neq 0 \) and \( N_{2} < N_{d} - N_{1}, \) again we assume that the first \( N_{2} \) rows of \( A^{(1)} \) denoted by \( R_{2,1} \) contains independent rows and
\[ A^{(1)(d,N_{2})} = V_{2} R_{2,1}. \]  
(A.22)

Also define invertible matrix \( M_{2} \) as
\[ M_{2} \triangleq \begin{bmatrix} N_{2} & (N_{d} - N_{1} - N_{2}) \\ I & 0 \\ -V_{2} & I \end{bmatrix}. \]  
(A.23)
and

\[
\begin{pmatrix}
    s_{i,2} \\
    \bar{s}_{i,2}
\end{pmatrix} \triangleq \begin{pmatrix}
    (M_2 \bar{s}_{i,1})^{u,N_2} \\
    (M_2 \bar{s}_{i,1})^{d,N_2}
\end{pmatrix} = M_2 \bar{s}_{i,1}.
\]  

\[\text{(A.24)}\]

We have

\[
\begin{pmatrix}
    s_{i,1} \\
    s_{i,2} \\
    \bar{s}_{i,2}
\end{pmatrix} = \begin{pmatrix}
    R_{1,0} & R_{1,1} & (R_{1,2} M_2^{-1})^{l,N_2} & (R_{1,2} M_2^{-1})^{r,N_2} \\
    0 & M_2 A^{(1)} & M_2 B^{(1)} M_2^{-1} & (M_2 B^{(1)} M_2^{-1})^{d,N_2} \\
    0 & (M_2 A^{(1)})^{d,N_2} & (M_2 B^{(1)} M_2^{-1})^{d,N_2} & (M_2 B^{(1)} M_2^{-1})^{d,N_2}
\end{pmatrix} \begin{pmatrix}
    s_{i-1,0} \\
    s_{i-1,1} \\
    \bar{s}_{i-1,1}
\end{pmatrix}
\]

\[\text{and (A.25) is equivalent to (A.26)}\]

\[\text{which can be written as}\]

\[
\begin{pmatrix}
    s_{i,1} \\
    s_{i,2} \\
    \bar{s}_{i,2}
\end{pmatrix} = \begin{pmatrix}
    R_{1,0} & R_{1,1} & (R_{1,2} M_2^{-1})^{l,N_2} & (R_{1,2} M_2^{-1})^{r,N_2} \\
    0 & R_{2,1} & R_{2,2} & R_{2,3} \\
    0 & 0 & A^{(2)} & B^{(3)}
\end{pmatrix} \begin{pmatrix}
    s_{i-1,0} \\
    s_{i-1,1} \\
    \bar{s}_{i-1,2}
\end{pmatrix}
\]

\[\text{Note that } s_{i,2} \text{ is defined in this step.}\]

This procedure can be repeated through next steps until \((K-1)\text{th step}\) where \(A^{(K-1)}\) is either full-rank of rank \(N_K\) or zero matrix. In this step define \(R_{K,K-1} = A^{(K-1)}\) and \(s_{i,K} = \bar{s}_{i,K-1}\). The result is

\[
\hat{s}_i = L_{\hat{f}}(s_i) = \begin{pmatrix}
    s_{i,0} \\
    \vdots \\
    s_{i,K}
\end{pmatrix}.
\]

\[\text{Similar to (2.99) and (A.20), (2.111) can be verified for all the steps. Note that all the steps are invertible.}\]

\[\text{This completes the proof of lemma 4.}\]

### A.3 Proof of Lemma 5: Transformation \(L_b\)

Consider a source \(\hat{s}_i\) consisting of \(N_0\) innovation bits and \(K\) deterministic sub-symbols \(\hat{s}_{i,d}\) defined in \((2.109)\). The following iterative method characterizes transformation \(L_b\).

**Step 0:** If \(R_{K,K-1} = 0\), we have

\[
\begin{align*}
    s_{i,K} &= R_{K,K} s_{i-1,K} \\
    &= R_{K,K}^{i+1} s_{-1,K}.
\end{align*}
\]

\[\text{Similarly, we can express } \hat{s}_i \text{ as a function of } s_{i,K}.\]
Note that \( s_{-1} \), and thus \( s_{-1,K} \), is known at the decoder. Therefore, we can eliminate sub-symbol \( s_{i,K} \) and consider the source \( \hat{s}_i \) with \( N_0 \) innovation bits and deterministic bits characterized by the following

\[
\hat{s}_{i,d} = \begin{pmatrix}
\hat{s}_{i,1} \\
\hat{s}_{i,2} \\
\vdots \\
\hat{s}_{i,K-1} \\
0 & 0 & \cdots & R_{K-1,K-2} & R_{K-1,K-1} & 0
\end{pmatrix} = \begin{pmatrix}
R_{1,0} & R_{1,1} & \cdots & R_{1,K-2} & R_{1,K-1} & R_{1,K} \\
0 & R_{2,1} & \cdots & R_{2,K-2} & R_{2,K-1} & R_{2,K} \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\
0 & 0 & \cdots & R_{K-1,K-2} & R_{K-1,K-1} & R_{K-1,K}
\end{pmatrix} \begin{pmatrix}
\hat{s}_{i-1,0} \\
\hat{s}_{i-1,1} \\
\vdots \\
\hat{s}_{i-1,K-2} \\
\hat{s}_{i-1,K-1} \\
\hat{s}_{i-1,K}
\end{pmatrix}, \tag{A.31}
\]

and continue to next step with \( K - 1 \). Note that knowing \( \hat{s}_i, \tilde{s}_i \) can be constructed.

If \( R_{K,K-1} \) is full-rank of rank \( K \), continue to next step.

**Step 1:** Define

\[
\begin{pmatrix}
\hat{s}_{i,K-1} \\
\tilde{s}_i
\end{pmatrix} = \begin{pmatrix}
I_{N_{K-1}} & X_1 \\
0 & I_{N_k}
\end{pmatrix} \begin{pmatrix}
\hat{s}_{i,K-1} \\
\tilde{s}_i
\end{pmatrix}, \tag{A.32}
\]

and

\[
D_1 = \begin{bmatrix}
\sum_{j=0}^{N_{K-1}} N_j & N_{K-1} & N_K \\
0 & I & X_1 \\
0 & 0 & I
\end{bmatrix} \tag{A.33}
\]

and note that

\[
D_1^{-1} = \begin{pmatrix}
I & 0 & 0 \\
0 & I & -X_1 \\
0 & 0 & I
\end{pmatrix} \tag{A.34}
\]

Also \( X_1 \) can be defined such that

\[
R_{K,K} - R_{K,K-1} X_1 = 0. \tag{A.35}
\]

By these definitions, (2.109) can be reformulated to get (A.36).

\[
\begin{pmatrix}
s_{i,1} \\
s_{i,2} \\
\vdots \\
\hat{s}_{i,K-1} \\
\hat{s}_{i,K}
\end{pmatrix} = D_{i}^{\text{rd},N_0} \begin{pmatrix}
R_{1,0} & R_{1,1} & \cdots & R_{1,K-2} & R_{1,K-1} & R_{1,K} \\
0 & R_{2,1} & \cdots & R_{2,K-2} & R_{2,K-1} & R_{2,K} \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\
0 & 0 & \cdots & R_{K-1,K-2} & R_{K-1,K-1} & R_{K-1,K} \\
0 & 0 & \cdots & 0 & R_{K,K-1} & R_{K,K}
\end{pmatrix} D_1^{-1} \begin{pmatrix}
s_{i-1,0} \\
s_{i-1,1} \\
\vdots \\
\hat{s}_{i-1,K-2} \\
\hat{s}_{i-1,K-1} \\
\hat{s}_{i-1,K}
\end{pmatrix},
\]
Appendix A. Zero-Delay Lossless Streaming

\[
\begin{pmatrix}
R_{1,0} & R_{1,1} & \cdots & R_{1,K-2} & R_{1,K-1}^{(1)} & R_{1,K}^{(1)} \\
0 & R_{2,1} & \cdots & R_{2,K-2} & R_{2,K-1}^{(1)} & R_{2,K}^{(1)} \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\
0 & 0 & \cdots & R_{K-1,K-2} & R_{K-1,K-1}^{(1)} & R_{K-1,K}^{(1)} \\
0 & 0 & \cdots & 0 & R_{K,K-1} & 0
\end{pmatrix}
\begin{pmatrix}
s_{i-1,0} \\
s_{i-1,1} \\
\vdots \\
s_{i-1,K-2} \\
s_{i-1,K-1} \\
s_{i-1,K}
\end{pmatrix}
= 
\begin{pmatrix}
\tilde{s}_{i-1,0} \\
\tilde{s}_{i-1,1} \\
\vdots \\
\tilde{s}_{i-1,K-2} \\
\tilde{s}_{i-1,K-1} \\
\tilde{s}_{i-1,K}
\end{pmatrix}
\] (A.36)

Matrices \(\hat{R}_{(j)}^{(1)}\) can be defined accordingly.

**Step** \(j \in [2 : K]\): Define \(l = K - j\). At step \(j\), the source is transformed into the form of (A.37).

\[
\begin{pmatrix}
s_{i,1} \\
s_{i,l} \\
\tilde{s}_{i,l+1} \\
\tilde{s}_{i,l+2} \\
\vdots \\
\tilde{s}_{i,K}
\end{pmatrix}
= 
\begin{pmatrix}
R_{1,0} & \cdots & R_{1,l-1} & R_{1,l} & \tilde{R}_{1,l+1}^{(j-1)} & \tilde{R}_{1,l+2}^{(j-1)} & \cdots & \tilde{R}_{1,K-1}^{(j-1)} & \tilde{R}_{1,K}^{(j-1)} \\
\vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
0 & \cdots & R_{l,l-1} & R_{l,l} & \tilde{R}_{l,l+1}^{(j-1)} & \tilde{R}_{l,l+2}^{(j-1)} & \cdots & \tilde{R}_{l,K-1}^{(j-1)} & \tilde{R}_{l,K}^{(j-1)} \\
0 & \cdots & 0 & R_{l+1,l} & \tilde{R}_{l+1,l+1}^{(j-1)} & \tilde{R}_{l+1,l+2}^{(j-1)} & \cdots & \tilde{R}_{l+1,K-1}^{(j-1)} & \tilde{R}_{l+1,K}^{(j-1)} \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
0 & \cdots & 0 & 0 & 0 & 0 & \cdots & R_{K,K-1} & 0
\end{pmatrix}
\begin{pmatrix}
s_{i-1,0} \\
\vdots \\
s_{i-1,l-1} \\
\tilde{s}_{i-1,l+1} \\
\tilde{s}_{i-1,l+2} \\
\vdots \\
\tilde{s}_{i-1,K}
\end{pmatrix}
\] (A.37)

\[
D_j \triangleq 
\begin{bmatrix}
\sum_{j=0}^{l-1} N_j & N_j & N_{j+1} & \cdots & N_K \\
\sum_{j=0}^{l-1} N_j & N_j & X_{1,j} & \cdots & X_{j,j} \\
N_l & 0 & I & \cdots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
N_K & 0 & 0 & 0 & I
\end{bmatrix}
\] (A.38)

and note that

\[
D_j^{-1} = 
\begin{bmatrix}
I & 0 & 0 & \cdots & 0 \\
0 & I & -X_{1,j} & \cdots & -X_{j,j} \\
0 & 0 & I & \cdots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & 0 & \cdots & I
\end{bmatrix}
\] (A.39)
Also define

\[ \hat{s}_{i,t} \triangleq \left( I \; X_{1,j} \; X_{2,j} \; \cdots \; X_{j,j} \right) \begin{pmatrix} s_{i,l} \\ \vdots \\ s_{i,l-1} \\ \tilde{s}_{i,l} \\ \tilde{s}_{i,l+1} \\ \tilde{s}_{i,l+2} \\ \vdots \\ \tilde{s}_{i,K} \end{pmatrix}. \]

By these definitions, (A.37) reduces to

\[ \begin{pmatrix} s_{i,1} \\ \vdots \\ s_{i,l-1} \\ \tilde{s}_{i,l} \\ \tilde{s}_{i,l+1} \\ \tilde{s}_{i,l+2} \\ \vdots \\ \tilde{s}_{i,K} \end{pmatrix} = D_j^{(dr,N_0)} \Psi^{(j-1)} D_j^{-1}, \]

By defining \( X_{k,j} \)'s such that for each \( k \in \{1, 2, \ldots, j\} \)

\[ \tilde{R}_{l+1,l+k}^{(j-1)} - R_{l+1,l} X_{k,j} = 0, \]

it is not hard to see that (A.41) can be rewritten as (A.43) whose \((l+1)\)th row is block-diagonalized.

\[ \begin{pmatrix} s_{i,1} \\ \vdots \\ \tilde{s}_{i,l} \\ \tilde{s}_{i,l+1} \\ \tilde{s}_{i,l+2} \\ \vdots \\ \tilde{s}_{i,K} \end{pmatrix} = \begin{pmatrix} R_{1,0} & \cdots & R_{1,l-1} & \tilde{R}_{1,l}^{(j)} & \tilde{R}_{1,l+1}^{(j)} & \tilde{R}_{1,l+2}^{(j)} & \cdots & \tilde{R}_{1,K-1}^{(j)} & \tilde{R}_{1,K}^{(j)} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\ 0 & \cdots & 0 & R_{l,l} & R_{l,l+1}^{(j)} & R_{l,l+2}^{(j)} & \cdots & R_{l,K-1}^{(j)} & R_{l,K}^{(j)} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\ 0 & \cdots & 0 & 0 & R_{l+1,l} & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\ 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & R_{K,K-1} & 0 \end{pmatrix} \tilde{s}_{i,0} \]

\[ \tilde{s}_{i,0} \]

After these steps, the source \( s_i \) is changed into the diagonally correlated Markov source \( \tilde{s}_i \), with \( N_0 \) innovation bits \( \tilde{s}_{i,0} \) and deterministic bits as (A.44).

\[ \tilde{s}_{i,d} = \begin{pmatrix} \tilde{s}_{i,1} \\ \tilde{s}_{i,2} \\ \vdots \\ \tilde{s}_{i,K-1} \\ \tilde{s}_{i,K} \end{pmatrix} = \begin{pmatrix} R_{1,0} & \cdots & 0 & 0 \\ 0 & R_{2,1} & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & R_{K-1,K-2} \\ 0 & 0 & \cdots & R_{K,K-1} \end{pmatrix} \begin{pmatrix} \tilde{s}_{i,-1,0} \\ \tilde{s}_{i,-1,1} \\ \vdots \\ \tilde{s}_{i,-1,K-2} \\ \tilde{s}_{i,-1,K-1} \end{pmatrix}. \]
All the steps are invertible and this completes the proof.
Appendix B

Zero-Delay Streaming of Gauss-Markov Sources: Immediate Recovery

B.1 Proof of Lemma 6

Define $q_k \triangleq 2^\frac{2}{R}E(h(s^n_k|y^k,x^n_{k-1}))$. We need to show that

$$q_k \geq 2\pi e(1 - \rho^2) \left( 1 - \left( \frac{\rho^2}{2R} \right)^k \right).$$

(B.1)

Consider the following entropy term.

\begin{align*}
h(s^n_k|y^k, x^n_{k-1}) &= h(s^n_k|y^{k-1}, x^n_{k-1}) - I(f_k; s^n_k|y^{k-1}, x^n_{k-1}) \\
&= h(s^n_k|y^{k-1}, x^n_{k-1}) - H(f_k|y^{k-1}, x^n_{k-1}) \\
&\geq h(s^n_k|y^{k-1}, x^n_{k-1}) - H(f_k) \\
&\geq \frac{n}{2} \log \left( \rho^2 2^\frac{2}{R}E(h(s^n_{k-1}|y^{k-1}, x^n_{k-1})) + 2\pi e(1 - \rho^2) \right) - nR,
\end{align*}

where (B.2) follows from the fact that conditioning reduces entropy and (B.3) follows from the Entropy Power Inequality similar to (3.25). Thus

$$q_k \geq \frac{\rho^2}{2R}q_{k-1} + \frac{2\pi e(1 - \rho^2)}{2R}. \quad (B.4)$$

By repeating the iteration in (B.4), we have

$$q_k \geq \left( \frac{\rho^2}{2R} \right)^k q_0 + \frac{2\pi e(1 - \rho^2)}{2R} \sum_{l=0}^{k-1} \left( \frac{\rho^2}{2R} \right)^l. \quad (B.5)$$
Appendix B. Zero-Delay Streaming of Gauss-Markov Sources: Immediate Recovery

Figure B.1: Relationship of the Variables for Lemma 31.

\[
\begin{align*}
\rho_1 & \quad N_1 \\
X_0 & \quad X_1 \\
Z_1 & \quad Y_1 \\
\rho_2 & \quad N_2 \\
X_2 & \quad X_3 \\
Z_2 & \quad Y_2
\end{align*}
\]

\[\geq \frac{2\pi e (1 - \rho^2)}{2^2 R - \rho^2} \left(1 - \left(\frac{\rho^2}{2^2 R}\right)^k\right), \quad (B.6)\]

where (B.6) follows from the fact \(0 < \rho^2 R < 1\) for any \(\rho \in (0, 1)\) and \(R > 0\). This completes the proof.

B.2 Proof of Equations (3.36) and (3.37)

We need to show (3.36) and (3.37), i.e., we need to establish the following two inequities for each \(k \in \{1, \ldots, t - B'\}\)

\[
\begin{align*}
h(u_i||u|^{t-B'-k-1},|u|^{t-1}_{i-k},s_{-1}) & \leq h(u_i||u|^{t-B'-k},|u|^{t-1}_{i-k+1},s_{-1}) \quad (B.7) \\
h(s_i||u|^{t-B'-k-1},|u|^{t-1}_{i-k},s_{-1}) & \leq h(s_i||u|^{t-B'-k},|u|^{t}_{i-k+1},s_{-1}). \quad (B.8)
\end{align*}
\]

We first establish the following Lemmas.

**Lemma 31.** Consider random variables \(\{X_0, X_1, X_2, Y_1, Y_2\}\) that are jointly Gaussian, \(X_k \sim N(0, 1)\), \(k \in \{0, 1, 2\}\), \(X_0 \rightarrow X_1 \rightarrow X_2\) and that for \(j \in \{1, 2\}\) we have:

\[
\begin{align*}
X_j & = \rho_j X_{j-1} + N_j, \quad (B.9) \\
Y_j & = X_j + Z_j. \quad (B.10)
\end{align*}
\]

Assume that \(Z_j \sim N(0, \sigma_j^2)\) are independent of all random variables and likewise \(N_j \sim N(0, 1 - \rho_j^2)\) for \(j \in \{1, 2\}\) are also independent of all random variables. The structure of correlation is sketched in Fig. B.1. Then we have that:

\[
\sigma_{X_2}^2(X_0, Y_2) \leq \sigma_{X_2}^2(X_0, Y_1), \quad (B.11)
\]

where \(\sigma_{X_2}^2(X_0, Y_j)\) denotes the minimum mean square error of estimating \(X_2\) from \(\{X_0, Y_j\}\).

**Proof.** By applying the standard relation for the MMSE estimation error we have (see e.g. [40])

\[
\begin{align*}
\sigma_{X_2}^2(X_0, Y_1) & = E[X_2^2] - \left(E[X_2 Y_1] E[X_2 X_0]\right) \begin{pmatrix} E[Y_1^2] & E[X_0 Y_1] \ E[X_0^2] & E[X_2 Y_1] \ E[X_2 X_0] \end{pmatrix}^{-1} \begin{pmatrix} E[X_2 Y_1] \\
E[X_2 X_0] \end{pmatrix} \\
& = 1 - \rho_2^2 \begin{pmatrix} 1 & \rho_1 \end{pmatrix} \begin{pmatrix} 1 + \sigma_z^2 & \rho_1 \ \rho_1 \ 1 \end{pmatrix}^{-1} \begin{pmatrix} 1 \\
\rho_1 \end{pmatrix} \quad (B.12)
\end{align*}
\]

\[
\begin{align*}
& = 1 - \rho_2^2 \begin{pmatrix} 1 & \rho_1 \end{pmatrix} \begin{pmatrix} 1 + \sigma_z^2 & \rho_1 \ \rho_1 \ 1 \end{pmatrix}^{-1} \begin{pmatrix} 1 \\
\rho_1 \end{pmatrix} \quad (B.13)
\end{align*}
\]
This completes the proof.

To establish (B.11) we only need to show that,

\[ \sigma^2_{X_k}(X_0, Y_2) = 1 - \left( 1 - \rho_1 \rho_2 \right) \left( 1 + \sigma^2_z \rho_1 \rho_2 \right)^{-1} \left( 1 - \rho_1 \rho_2 \right) \]

which is equivalent to showing

\[ 1 - \frac{\rho_1^2 \rho_2^2 - \rho_1^2 \rho_2^2 + 1}{1 + \sigma^2_z - \rho_1^2 \rho_2^2} \geq 1 - \frac{\rho_1^2 \rho_2^2 - \rho_1^2 \rho_2^2 + 1}{1 + \sigma^2_z - \rho_1^2 \rho_2^2} \]

or equivalently

\[ 1 - \frac{\rho_1^2 (1 - \rho_2^2)}{1 + \sigma^2_z - \rho_1^2 \rho_2^2} \geq 1 - \frac{1 - \rho_2^2}{1 + \sigma^2_z - \rho_1^2 \rho_2^2} \]

which is equivalent to showing

\[ \frac{\rho_1^2}{1 + \sigma^2_z - \rho_1^2 \rho_2^2} \leq \frac{1}{1 + \sigma^2_z - \rho_1^2 \rho_2^2} \]

However note that (B.20) can be immediately verified since the left hand side has the numerator smaller than the right hand side and the denominator greater than the right hand side whenever \( \rho_1^2 \in (0, 1) \). This completes the proof.

Lemma 32. Consider the Gauss-Markov source model (3.1) and the test channel in Prop. 4. For a fixed \( t, k \in \{1, \ldots, t\} \) and a set \( \Omega \subseteq \{ t - k, \ldots, t \} \), consider two sets of random variables \( W_1 \) and \( W_2 \) each jointly Gaussian with \( s_{-k} \) such that the following Markov property holds:

\[ W_1 \rightarrow s_{-k} \rightarrow \{ s_t, u_0 \} \]  
(B.21)
\[ W_2 \rightarrow s_{-k} \rightarrow \{ s_t, u_0 \}. \]  
(B.22)

If the MMSE error in \( s_{-k} \) satisfies , \( \sigma^2_{-k}(W_1) \leq \sigma^2_{-k}(W_2) \) then we have

\[ h(s_t|W_1, u_0) \leq h(s_t|W_2, u_0), \quad \forall \Omega \subseteq [t - k, t] \]  
(B.23)
\[ h(u_t|W_1, u_0) \leq h(u_t|W_2, u_0), \quad \forall \Omega \subseteq [t - k, t - 1]. \]  
(B.24)

Proof. Since the underlying random variables are jointly Gaussian, we can express the MMSE estimates
Appendix B. Zero-Delay Streaming of Gauss-Markov Sources: Immediate Recovery

of $s_{t-k}$ from $W_j$, $j \in \{1, 2\}$ as follows (see e.g. [40])

\begin{align}
\hat{s}_{t-k}(W_1) &= \alpha_1 s_{t-k} + e_1 \\
\hat{s}_{t-k}(W_2) &= \alpha_2 s_{t-k} + e_2,
\end{align}

where $e_1 \sim \mathcal{N}(0, E_1)$ and $e_2 \sim \mathcal{N}(0, E_2)$ are Gaussian random variables both independent of $s_{t-k}$. Furthermore the constants in (B.25) and (B.26) are given by

\begin{align}
\alpha_j &= 1 - \sigma^2_{t-k}(W_j) \\
E_j &= \sigma^2_{t-k}(W_j) (1 - \sigma^2_{t-k}(W_j))
\end{align}

for $j = 1, 2$. To establish (B.23), we have

\begin{align}
\mathbb{H}(s_t|W_1, u_{\Omega}) &= \mathbb{H}(s_t|\hat{s}_{t-k}(W_1), u_{\Omega}) \\
&= \mathbb{H}(s_t|\alpha_1 s_{t-k} + e_1, u_{\Omega}) \\
&\leq \mathbb{H}(s_t|\alpha_2 s_{t-k} + e_2, u_{\Omega}) \\
&= \mathbb{H}(s_t|\hat{s}_{t-k}(W_2), u_{\Omega}) \\
&= \mathbb{H}(s_t|W_2, u_{\Omega}),
\end{align}

where (B.29) and (B.33) follows from the Markov properties

\begin{align}
W_1 \rightarrow \hat{s}_{t-k}(W_1) \rightarrow \{s_t, u_{\Omega}\} \\
W_2 \rightarrow \hat{s}_{t-k}(W_2) \rightarrow \{s_t, u_{\Omega}\}
\end{align}

Equations (B.30) and (B.32) follows from (B.25) and (B.26) and (B.31) follows from the fact that $\sigma^2_{t-k}(W_1) \leq \sigma^2_{t-k}(W_2)$ implies that

$$\frac{E_1}{\alpha^2_1} \leq \frac{E_2}{\alpha^2_2}.$$  

Thus the only difference between (B.30) and (B.31) is that the variance of the independent noise component in the first term is smaller in the former. Clearly we obtain a better estimate of $s_t$ in (B.30), which justifies the inequality in (B.31).

Eq. (B.24) can be established as an immediate consequence of (B.23). Since the noise $z_t$ in the test channel is Gaussian and independent of all other random variables, we have

$$\text{Var}(u_t|W_j, u_{\Omega}) = \text{Var}(s_t|W_j, u_{\Omega}) + \sigma^2_z,$$

where the notation $\text{Var}(a|W)$ indicates the noise variance of estimating $a$ from $W$. As a result,

$$h(u_t|W_j, u_{\Omega}) = \frac{1}{2} \log \left(2^{2h(s_t|W_j, u_{\Omega})} + 2\pi e \sigma^2_z\right).$$

Thus (B.23) immediately implies (B.24).

We now establish (B.7) and subsequently establish (B.8) in a similar fashion. Consider the following
two steps.

1) First by applying Lemma 31 we show.

\[ \sigma^2_{t-k}(\{u\}_{0}^{t-B'-k-1}, u_{t-k}, s_{-1}) \leq \sigma^2_{t-k}(\{u\}_{0}^{t-B'-k-1}, u_{t-B'-k}, s_{-1}), \]  

(B.39)
i.e., knowing \{ \{u\}_{0}^{t-B'-k-1}, u_{t-k}, s_{-1} \} rather than \{ \{u\}_{0}^{t-B'-k-1}, u_{t-B'-k}, s_{-1} \}, improves the estimate of the source \( s_{t-k} \). Let \( \tilde{s}_{t-B'-k}(\{u\}_{0}^{t-B'-k-1}, s_{-1}) \) be the MMSE estimator of \( s_{t-B'-k} \) given \{ \{u\}_{0}^{t-B'-k-1}, s_{-1} \}.

Note that \( \tilde{s}_{t-B'-k}(\{u\}_{0}^{t-B'-k-1}, s_{-1}) \) is a sufficient statistic of \( s_{t-B'-k} \) given \{ \{u\}_{0}^{t-B'-k-1}, s_{-1} \} and thus we have that:

\[ \{ \{u\}_{0}^{t-B'-k-1}, s_{-1} \} \rightarrow \tilde{s}_{t-B'-k}(\{u\}_{0}^{t-B'-k-1}, s_{-1}) \rightarrow s_{t-B'-k} \rightarrow s_{t-k}. \]  

(B.40)

Therefore, by application of Lemma 31 for \( X_0 = \tilde{s}_{t-B'-k}(\{u\}_{0}^{t-B'-k-1}, s_{-1}) \), \( X_1 = s_{t-B'-k} \), \( Y_1 = u_{t-B'-k} \), \( X_2 = s_{t-k} \) and \( Y_2 = u_{t-k} \), we have

\[ \sigma^2_{t-k}(\{u\}_{0}^{t-B'-k-1}, u_{t-k}, s_{-1}) = \sigma^2_{t-k}(\tilde{s}_{t-B'-k}(\{u\}_{0}^{t-B'-k-1}, s_{-1}), u_{t-k}) \]
\[ \leq \sigma^2_{t-k}(\tilde{s}_{t-B'-k}(\{u\}_{0}^{t-B'-k-1}, s_{-1}), u_{t-B'-k}) \]
\[ = \sigma^2_{t-k}(\{u\}_{0}^{t-B'-k-1}, u_{t-B'-k}, s_{-1}). \]  

(B.41)

(B.42)

(B.43)

where (B.41) and (B.43) both follow from (B.40). This completes the claim in (B.39).

2) In the second step, we apply Lemma 32 for

\[ \mathcal{W}_1 = \{ \{u\}_{0}^{t-B'-k-1}, u_{t-k}, s_{-1} \} \]  

(B.44)
\[ \mathcal{W}_2 = \{ \{u\}_{0}^{t-B'-k-1}, u_{t-B'-k}, s_{-1} \} \]  

(B.45)
\[ \Omega = \{ t-k+1, \ldots, t-1 \}. \]  

(B.46)

We have

\[ h\left( u_{t} \mid \{u\}_{0}^{t-B'-k-1}, \{u\}_{t-k}^{t-1}, s_{-1} \right) \leq h\left( u_{t} \mid \{u\}_{0}^{t-B'-k-1}, \{u\}_{t-k+1}^{t-1}, s_{-1} \right), \]  

(B.47)

and again by applying Lemma 32 for \( \mathcal{W}_1 \) and \( \mathcal{W}_2 \) in (B.44) and (B.45) and \( \Omega = [t-k+1, t] \), we have

\[ h\left( s_{t} \mid \{u\}_{0}^{t-B'-k-1}, \{u\}_{t-k}^{t-1}, s_{-1} \right) \leq h\left( s_{t} \mid \{u\}_{0}^{t-B'-k-1}, \{u\}_{t-k+1}^{t}, s_{-1} \right). \]  

(B.48)

This establishes (B.7) and (B.8) and equivalently (3.36) and (3.37).

### B.3 Proof of Lemma 10

For the reader’s convenience, we first repeat the statement of the Lemma. Consider the two sets \( A, B \subseteq \mathbb{N} \) each of size \( r \) as \( A = \{a_1, a_2, \ldots, a_r\} \), \( B = \{b_1, b_2, \ldots, b_r\} \) such that \( 1 \leq a_1 < a_2 < \cdots < a_r \) and \( 1 \leq b_1 < b_2 < \cdots < b_r \) and for any \( i \in \{1, \ldots, r\} \), \( a_i \leq b_i \). Then the test channel (3.30) satisfies the
Then we have

\[ h(s_t | u_A, s_{-1}) \geq h(s_t | u_B, s_{-1}), \quad \forall t \geq b_r \]  
(B.49)

\[ h(u_t | u_A, s_{-1}) \geq h(u_t | u_B, s_{-1}), \quad \forall t > b_r. \]  
(B.50)

We first prove (B.49) by induction as follows. The proof of (B.50) follows directly from (B.49) as discussed at the end of this section.

- First we show that (B.49) is true for \( r = 1 \), i.e., given \( 0 \leq a_1 \leq b_1 \) and for all \( t \geq b_1 \) we need to show

\[ h(s_t \mid u_{a_1}, s_{-1}) \geq h(s_t \mid u_{b_1}, s_{-1}). \]  
(B.51)

We apply Lemma 31 in Appendix B.2 for \( \{X_0, X_1, X_2, Y_1, Y_2\} = \{s_{-1}, s_{a_1}, s_{b_1}, u_{a_1}, u_{b_1}\} \) which results in

\[ h(s_{a_1} \mid u_{a_1}, s_{-1}) \geq h(s_{b_1} \mid u_{b_1}, s_{-1}). \]  
(B.52)

Thus (B.51) holds for \( t = b_1 \). For any \( t > b_1 \) we can always express \( s_t = \rho^{t-b_1} s_{b_1} + \tilde{n} \) where \( \tilde{n} \sim \mathcal{N}(0, 1 - \rho^{2(t-b_1)}) \) and also we can express \( s_{a_1} = \tilde{s}_{a_1}(u_j, s_{-1}) + w_j \) for \( j \in \{a_1, b_1\} \) where \( w_j \sim \mathcal{N}(0, \sigma^2_j(u_j, s_{-1})) \) is the MMSE estimation error. For \( j \in \{a_1, b_1\} \), we have

\[ s_t = \rho^{t-b_1} \tilde{s}_{a_1}(u_j, s_{-1}) + \rho^{t-b_1} w_j + \tilde{n}. \]  
(B.53)

Then we have

\[ \sigma^2_t(u_{a_1}, s_{-1}) = \rho^{2(t-b_1)} \sigma^2_{b_1}(u_{a_1}, s_{-1}) + 1 - \rho^{2(t-b_1)} \]  
(B.54)

\[ \geq \rho^{2(t-b_1)} \sigma^2_{b_1}(u_{b_1}, s_{-1}) + 1 - \rho^{2(t-b_1)} \]  
(B.55)

\[ = \sigma^2_t(u_{b_1}, s_{-1}), \]  
(B.56)

where (B.55) immediately follows from (B.52). Thus (B.56) establishes (B.51) and the proof of the base case is now complete.

- Now assume that (B.49) is true for \( r \), i.e., for the sets \( A_r, B_r \) of size \( r \) satisfying \( a_i \leq b_i \) for \( i \in \{1, \ldots, r\} \) and any \( t \geq b_r \),

\[ h(s_t \mid u_{A_r}, s_{-1}) \geq h(s_t \mid u_{B_r}, s_{-1}) \]  
(B.57)

We show that the lemma is also true for the sets \( A_{r+1} = \{A_r, a_{r+1}\} \) and \( B_{r+1} = \{B_r, b_{r+1}\} \) where \( a_r \leq a_{r+1}, b_r \leq b_{r+1} \) and \( a_{r+1} \leq b_{r+1} \). We establish this in two steps.

1) We show that

\[ h(s_t \mid u_{A_{r+1}}, s_{-1}) \geq h(s_t \mid u_{B_{r+1}}, s_{-1}). \]  
(B.58)

By application of Lemma 31 for

\[ \{X_0, X_1, X_2, Y_1, Y_2\} = \{\tilde{s}_{a_r}(u_A, s_{-1}), s_{a_{r+1}}, s_{b_{r+1}}, u_{a_{r+1}}, u_{b_{r+1}}\}, \]  
(B.59)
Appendix B. Zero-Delay Streaming of Gauss-Markov Sources: Immediate Recovery

we have

\[ h(s_{b+1} | \hat{s}_r (u_{A_r}, s_{r-1}), u_{b+1}) \geq h(s_{b+1} | \hat{s}_r (u_{A_r}, s_{r-1}), u_{b+1}). \]  

(B.60)

Thus (B.58) holds for \( t = b_{r+1} \). For \( t \geq b_{r+1} \) we can use the argument analogous to that leading to (B.56). We omit the details as they are completely analogous. This establishes (B.58).

2) Next we show that

\[ h(s|u_{A_r}, u_{b+1}, s_{r-1}) \geq h(s|u_{B_r}, s_{r-1}). \]  

(B.61)

First note that based on the induction hypothesis in (B.57) for \( t = b_{r+1} \) we have

\[ h(s_{b+1} | u_{A_r}, s_{r-1}) \geq h(s_{b+1} | u_{B_r}, s_{r-1}) \]  

(B.62)

and equivalently

\[ \sigma^2_{b+1} (u_{A_r}, s_{r-1}) \geq \sigma^2_{b+1} (u_{B_r}, s_{r-1}). \]  

(B.63)

Now by application of Lemma 32 for \( k = t - b_r \) and

\[ W_1 = \{ u_{B_r}, s_{r-1} \} \]  

(B.64)

\[ W_2 = \{ u_{A_r}, s_{r-1} \} \]  

(B.65)

\[ \Omega = \{ b_{r+1} \} \]  

(B.66)

and noting that \( W_j \to s_{b_r} \to (s_{b_r+1}, u_{\Omega}) \) for \( j = 1, 2 \) we have

\[ h(s|u_{A_r}, u_{b+1}, s_{r-1}) \geq h(s|u_{B_r}, u_{b+1}, s_{r-1}). \]  

(B.67)

which is equivalent to (B.61).

Combining (B.58) and (B.61) we have \( h(s|u_{A_r+1}, s_{r-1}) \geq h(s|u_{B_r+1}, s_{r-1}) \) which shows that (B.49) is also true for \( r + 1 \).

This completes the induction and the proof of (B.49) for general \( r \).

Finally note that (B.49) implies (B.50) as follows.

\[ h(u|u_{A_r}, s_{r-1}) = \frac{1}{2} \log \left( 2^{2h(s|u_{A_r}, s_{r-1})} + 2\pi e \sigma^2_z \right) \]  

(B.68)

\[ \geq \frac{1}{2} \log \left( 2^{2h(s|u_{B_r}, s_{r-1})} + 2\pi e \sigma^2_z \right) \]  

(B.69)

\[ = h(u|u_{B_r}, s_{r-1}), \]  

(B.70)

where (B.68) follows from the fact that the noise in the test channel is independent. Also (B.69) follows from (B.49). This completes the proof.

### B.4 Proof of Lemma 11

We prove each part separately as follows.
1) For any feasible set $\Omega_t$ with size $\theta$ we have

$$
\lambda_t(\Omega_t) = I(s_t; u_t | u_{\Omega_t}, s_{-1})
= h(u_t | u_{\Omega_t}, s_{-1}) - h(u_t | s_t)
\leq h(u_t | u_{\Omega_t^*}(\theta), s_{-1}) - h(u_t | s_t)
= I(s_t; u_t | u_{\Omega_t^*}(\theta), s_{-1})
= \lambda_t(\Omega_t^*(\theta)),
$$

where (B.71) follows from the application of Lemma 10 with $A = \Omega_t^*(\theta)$ and $B = \Omega_t$, which by construction of $\Omega_t^*(\theta)$ clearly satisfy the required condition. Also note that

$$
\frac{1}{2} \log (2\pi e \gamma_t(\Omega_t)) = h(s_t | u_t, u_{\Omega_t}, s_{-1})
\leq h(s_t | u_t, u_{\Omega_t^*}(\theta), s_{-1})
= \frac{1}{2} \log (2\pi e \gamma_t(\Omega_t^*(\theta)))
$$

where (B.73) follows from Lemma 10 for the sets $A = \{\Omega_t^*(\theta), t\}$ and $B = \{\Omega_t, t\}$. Thus we have $\gamma_t(\Omega_t) \leq \gamma_t(\Omega_t^*(\theta))$.

2) We next argue that both $\lambda_t(\Omega_t^*(\theta))$ and $\gamma_t(\Omega_t^*(\theta))$ attain their maximum values with the minimum possible $\theta$. Recall from Part 1 that when the number of erasures $n_e = t - \theta$ is fixed, the worst case sequence must have all erasure positions as close to $t$ as possible. Thus if $n_e \leq B$ the worst case sequence consists of a single burst spanning $\{t - n_e, \ldots, t - 1\}$. If $B < n_e \leq 2B$, the worst case sequence must have two burst erasures spanning $\{t - n_e - L, \ldots, t - B - L - 1\} \cup \{t - B, \ldots, t - 1\}$. More generally the worst case sequence will consist of a sequence of burst erasures each (except possibly the first one) of length $B$ separated by a guard interval of length $L$. Thus the non-erased indices associated with decreasing values of $\theta$ are nested, i.e., $\theta_1 \leq \theta_2$ implies that $\Omega_t^*(\theta_1) \subseteq \Omega_t^*(\theta_2)$. Further note that adding more elements in the non-erased indices $\Omega_t^*(\cdot)$ can only decrease both $\lambda_t(\cdot)$ and $\gamma_t(\cdot)$, i.e., $\Omega_t^*(\theta_1) \subseteq \Omega_t^*(\theta_2)$ implies that $\lambda_t(\Omega_t^*(\theta_1)) \geq \lambda_t(\Omega_t^*(\theta_2))$ and $\gamma_t(\Omega_t^*(\theta_1)) \geq \gamma_t(\Omega_t^*(\theta_2))$. Thus the worst case $\Omega_t^*(\theta)$ must constitute the non-erased indices associated with minimum possible value of $\theta$. The formal proof, which is analogous to the second part of Lemma 8 will be skipped.

3) This property follows from the fact that in steady state the effect of knowing $s_{-1}$ vanishes. In particular we show below that $\lambda_{t+1}(\Omega_{t+1}^*) \geq \lambda_t(\Omega_t^*)$ and $\gamma_{t+1}(\Omega_{t+1}^*) \geq \gamma_t(\Omega_t^*)$.

$$
\lambda_{t+1}(\Omega_{t+1}^*) = I(s_{t+1}; u_{t+1} | u_{\Omega_{t+1}^*}, s_{-1})
= h(u_{t+1} | u_{\Omega_{t+1}^*}, s_{-1}) - h(u_{t+1} | s_{t+1})
\geq h(u_{t+1} | u_{\Omega_{t+1}^*, s_{-1}, s_0}) - h(u_{t+1} | s_{t+1})
= h(u_{t+1} | u_{\Omega_{t+1}^*}, s_0) - h(u_{t+1} | s_{t+1})
= h(u_t | u_{\Omega_t^*}, s_{-1}) - h(u_t | s_t)
= I(s_t; u_t | u_{\Omega_t^*}, s_{-1})
= \lambda_t(\Omega_t^*). \quad \text{(B.78)}
$$
where (B.75) follows from the fact that conditioning reduces the differential entropy. Also in (B.76) the notation $\Omega_{t+1}^* \setminus \{0\}$ indicates the set $\Omega_{t+1}^*$ when the index 0 is excluded if $0 \in \Omega_{t+1}^*$. It can be easily verified that the set $\Omega_{t}^*$ is equivalent to the set obtained by left shifting the elements of the set $\Omega_{t+1}^* \setminus \{0\}$ by one. Then (B.76) follows from this fact and the following Markov property.

$$\{u_0, s_{-1}\} \rightarrow \{u_{\Omega_{t+1}^* \setminus \{0\}}, s_0\} \rightarrow u_{t+1}$$  \hspace{1cm} (B.79)$$

Eq. (B.77) follows from the time-invariant property of source model and the test channel. Also note that

$$\frac{1}{2} \log \left(2\pi e \gamma_{t+1}(\Omega_{t+1}^*)\right) = h(s_{t+1}\mid u_{t+1}, u_{\Omega_{t+1}^*}, s_{-1})$$

$$\geq h(s_{t+1}\mid u_{t+1}, u_{\Omega_{t+1}^*}, s_{-1}, s_0)$$

$$= h(s_{t+1}\mid u_{t+1}, u_{\Omega_{t+1}^* \setminus \{0\}}, s_0)$$

$$= h(s_t\mid u_t, u_{\Omega_t^*}, s_{-1})$$

$$= \frac{1}{2} \log \left(2\pi e \gamma_t(\Omega_t^*)\right)$$

where (B.80) follows from the fact that conditioning reduces the differential entropy, (B.81) follows from the following Markov property

$$\{u_0, s_{-1}\} \rightarrow \{u_{\Omega_{t+1}^* \setminus \{0\}}, u_{t+1}, s_0\} \rightarrow s_{t+1}$$  \hspace{1cm} (B.84)$$

and (B.82) again follows from the time-invariant property of source model and the test channel.

### B.5 Proof of Lemma 12

We need to show

$$I(s_t; u_t\mid \hat{s}_{t-L-B}, [u]_{t-B+1}^{t-1}) \geq \lim_{t \to \infty} \lambda_t(\Omega_t^*)$$

$$= \lim_{t \to \infty} I(s_t; u_t\mid u_{\Omega_t^*}, s_{-1})$$  \hspace{1cm} (B.85)$$

and

$$\sigma_t^2(\hat{s}_{t-L-B}, [u]_{t-B+1}^{t-1}, u_t) \geq \lim_{t \to \infty} \gamma_t(\Omega_t^*)$$

$$= \lim_{t \to \infty} \sigma_t^2(u_{\Omega_t^*}, u_t, s_{-1}).$$  \hspace{1cm} (B.86)$$

For any $t > L + B$, we can write

$$\lambda_t(\Omega_t^*) = I(s_t; u_t\mid u_{\Omega_t^*}, s_{-1})$$

$$= I(s_t; u_t\mid u_{\Omega_t^* \setminus \{0\}}, [u]_{t-L-B+1}^{t-B-1}, s_{-1})$$

$$= I(s_t; u_t\mid \hat{s}_{t-L-B}(u_{\Omega_t^* \setminus \{0\}}, s_{-1}), [u]_{t-B+1}^{t-1})$$

$$\leq I(s_t; u_t\mid \tilde{a}\hat{s}_{t-L-B} + \hat{e}, [u]_{t-L-B+1}^{t-1})$$

$$= I(s_t; u_t\mid \hat{s}_{t-L-B} + \hat{e}, [u]_{t-B+1}^{t-1})$$

(B.87)
where (B.88) follows from the structure of \( \Omega^*_i \) in Lemma 11, \( (B.89) \) follows from the Markov relation

\[
\{ \mathbf{u}_{\Omega^*_i}^{t-L-B} \colon \mathbf{s}_{-1} \} \rightarrow \{ \mathbf{s}_{t-L-B} (\mathbf{u}_{\Omega^*_i}^{t-L-B} \colon \mathbf{s}_{-1}), [\mathbf{u}]^{t-B-1}_{t-L-B+1} \} \rightarrow \mathbf{s}_t
\]

and in \( (B.90) \) we introduce \( \hat{\alpha} = 1 - D \) and \( \hat{\mathbf{e}} \sim \mathcal{N}(0, D(1 - D)) \). This follows from the fact that the estimate \( \mathbf{s}_{t-L-B} (\mathbf{u}_{\Omega^*_i}^{t-L-B} \colon \mathbf{s}_{-1}) \) satisfies the average distortion constraint of \( D \). In (B.91) we re-normalize the test channel so that \( \mathbf{e} \sim \mathcal{N}(0, D/(1 - D)) \). Taking the limit of (B.92) when \( t \rightarrow \infty \), results in (B.85). Also note that

\[
\gamma_t(\Omega^*_i) = \sigma_i^2 (\mathbf{u}_{\Omega^*_i}, \mathbf{u}_t, \mathbf{s}_{-1})
\]

\[
= \sigma_i^2 (\mathbf{u}_{\Omega^*_i}^{t-L-B}, [\mathbf{u}]^{t-B-1}_{t-L-B+1}, \mathbf{u}_t, \mathbf{s}_{-1})
\]

\[
= \sigma_i^2 (\mathbf{s}_{t-L-B} (\mathbf{u}_{\Omega^*_i}^{t-L-B} \colon \mathbf{s}_{-1}), [\mathbf{u}]^{t-B-1}_{t-L-B+1}, \mathbf{u}_t)
\]

\[
\leq \sigma_i^2 (\hat{\alpha} \mathbf{s}_{t-L-B} + \hat{\mathbf{e}}, [\mathbf{u}]^{t-B-1}_{t-L-B+1}, \mathbf{u}_t)
\]

\[
= \sigma_i^2 (\hat{\mathbf{s}}_{t-L-B} + \mathbf{e}, [\mathbf{u}]^{t-B-1}_{t-L-B+1}, \mathbf{u}_t)
\]

\[
= \sigma_i^2 (\hat{\mathbf{s}}_{t-L-B}, [\mathbf{u}]^{t-B-1}_{t-L-B+1}, \mathbf{u}_t),
\]

where (B.94) follows from the Markov property in (B.93) and (B.95) again follows from the fact that the estimate \( \hat{\mathbf{s}}_{t-L-B} (\mathbf{u}_{\Omega^*_i}^{t-L-B} \colon \mathbf{s}_{-1}) \) satisfies the distortion constraint. All the constants and variables in (B.95) and (B.96) are as defined before. Again, taking the limit of (B.97) when \( t \rightarrow \infty \) results in (B.86).

According to (B.92) and (B.97) if we choose the noise in the test channel \( \sigma_i^2 \) to satisfy

\[
\sigma_i^2 (\hat{\mathbf{s}}_{t-L-B}, [\mathbf{u}]^{t-B-1}_{t-L-B+1}, \mathbf{u}_t) = D,
\]

the test channel and the rate \( R^{+}_{\text{GM-ME}}(L, B, D) \) defined in (B.92) both satisfy rate and distortion constraints in (3.92) and (3.93) and therefore \( R^{+}_{\text{GM-ME}}(L, B, D) \) is achievable.

### B.6 Proof of Lemma 13

We first show that (3.134) which is repeated in (B.99),

\[
h(s^n_{t-L-2W} \ldots , s^n_{t-W-1} - h(s^n_{t-L-2W} \ldots , s^n_{t-W-1} || f^n_{1}, f^n_{t-W} \ldots , s^n_{t-W-I}) \geq \frac{B}{2} \log \left( \frac{1}{d_{W+1}} \right).
\]

(B.99)

From the fact that conditioning reduces the differential entropy, we can lower bound the left hand side in (B.99) by

\[
\sum_{i=0}^{B-1} \left( h(s^n_{t-L-W+i} - h(s^n_{t-L-W+i} || f^n_{1}, f^n_{t-W} \ldots , s^n_{t-W-I}) \right).
\]

(B.100)
We show that for each $i \in \{0, 1, \ldots, B - 1\}$

$$\begin{align*}
    h(s^n_{i-B-W+i}) - h(s^n_{i-B-W+i} | f^t_{0}^{t-B-W-1}, f^t_{l-W}, s^n_{-1}) &\geq \frac{n}{2} \log \left( \frac{1}{d_{B+W-i}} \right),
\end{align*}$$

(B.101)

which then establishes (B.99). Recall that since there is a burst erasure between time $t \in \{t - B - W, \ldots, t - W - 1\}$ the receiver is required to reconstruct

$$\hat{t}_i^n = [\hat{s}_t^n, \ldots, \hat{s}_{i-B-W}]$$

with a distortion vector $(d_0, \ldots, d_{B+W})$, i.e., a reconstruction of $\hat{s}^{n}_{i-B-W+i}$ is desired with a distortion of $d_{B+W-i}$ for $i = 0, 1, \ldots, B + W$ when the decoder is revealed $(|f^t_{0}^{t-B-W-1}, f^t_{l-W})$. Hence

$$\begin{align*}
    &h(s^n_{i-B-W+i}) - h(s^n_{i-B-W+i} | f^t_{0}^{t-B-W-1}, f^t_{l-W}, s^n_{-1}) \\
    &= h(s^n_{i-B-W+i}) - h(s^n_{i-B-W+i} | f^t_{0}^{t-B-W-1}, f^t_{l-W} \{ \hat{s}_{i-B-W+i} \}_{d_{B+W-i}}) \\
    &\geq h(s^n_{i-B-W+i}) - h(s^n_{i-B-W+i} | \{ \hat{s}_{i-B-W+i} \}_{d_{B+W-i}}) \\
    &\geq h(s^n_{i-B-W+i}) - h(s^n_{i-B-W+i} - \{ \hat{s}^n_{i-B-W+i} \}_{d_{B+W-i}})
\end{align*}$$

(B.103)

(B.104)

(B.105)

Since we have

$$E \left[ \frac{1}{n} \sum_{j=1}^{n} (s_{i-B-W+i,j} - \hat{s}_{i-B-W+i,j})^2 \right] \leq d_{B+W-i},$$

(B.106)

it follows from standard arguments that [37, Chapter 13] that

$$h(s^n_{i-B-W+i} - \{ \hat{s}^n_{i-B-W+i} \}_{d_{B+W-i}}) \leq \frac{n}{2} \log 2\pi e (d_{B+W-i}).$$

(B.107)

Substituting (B.107) into (B.105) and the fact that $h(s^n_{i-B-W+i}) = \frac{n}{2} \log 2\pi e$ establishes (B.101).

Now we establish (3.135) which is repeated in (B.108) as follows.

$$\begin{align*}
    &h(s^n_{i-W}, \ldots, s^n_{t}, f^t_{l-W}) - h(s^n_{i-W}, \ldots, s^n_{t} | f^t_{0}^{t-B-W-1}, f^t_{l-W}, s^n_{i-W-1}, s^n_{-1}) \\
    &+ H([f^t_{l-W}], [f^t_{0}^{t-B-W-1}], t^n_i, s^n_{-1}) \geq \frac{n(W+1)}{2} \log \left( \frac{1}{d_0} \right).
\end{align*}$$

(B.108)

Since $(s^n_{i-W}, \ldots, s^n_{t})$ are independent we can express the left-hand side in (B.108) as:

$$I(s^n_{i-W}, \ldots, s^n_{t}, f^t_{l-W} | f^t_{0}^{t-B-W-1}, s^n_{i-W-1}, s^n_{-1}) + H([f^t_{l-W}], [f^t_{0}^{t-B-W-1}], t^n_i, s^n_{-1})$$

(B.109)

(B.110)

(B.111)

The above mutual information term can be bounded as follows:

$$h(s^n_{i-W}, \ldots, s^n_{t} | f^t_{0}^{t-W-1}, s^n_{i-B-W}, \ldots, s^n_{l-W-1}, s^n_{-1}) - h(s^n_{i-W}, \ldots, s^n_{t} | f^t_{0}, s^n_{i-B-W}, \ldots, s^n_{l-W-1}, s^n_{-1})$$
\begin{align*}
&= h(s^n_{t-W}, \ldots, s^n_t) - h(s^n_{t-W}, \ldots, s^n_t | f^T_0, s^n_{t-W-B}, \ldots, s^n_{t-W-1}, \hat{s}^n_{t-1}) \\ &\geq h(s^n_{t-W}, \ldots, s^n_t) - h(s^n_{t-W}, \ldots, s^n_t | \{\hat{s}^n_{t-W}\}_{d_0}, \ldots, \{\hat{s}^n_t\}_{d_0}) \\ &\geq \sum_{i=0}^{W} \left( h(s^n_{t-W+i}) - h(s^n_{t-W+i} - \{\hat{s}^n_{t-W+i}\}_{d_0}) \right) \\ &\geq \sum_{i=0}^{W} \frac{n}{2} \log \left( \frac{1}{d_0} \right) = \frac{n(W + 1)}{2} \log \left( \frac{1}{d_0} \right),
\end{align*}

where (B.112) follows from the independence of \((s^n_{t-W}, \ldots, s^n_t)\) from the past sequences, and (B.113) follows from the fact that given the entire past \([f]_0^T\) each source sub-sequence needs to be reconstructed with a distortion of \(d_0\) and the last step follows from the standard approach in the proof of the rate-distortion theorem. This establishes (B.108).

This completes the proof.
Appendix C

Hybrid Coding

C.1 Proof of Lemma 14

First consider the following lemma.

**Lemma 33.** *(Time-Shifting Lemma)* For the memoryless Q-and-B scheme and for any \( k < t \),

\[
\begin{align*}
h(x_t|[u]_M, x_k) &= h(x_{t-k}|[u]_{M-k}) \quad \text{for} \quad M \subseteq \{1, 2, \ldots, t\} \quad \text{(C.1)} \\
h(u_t|[u]_M, x_k) &= h(u_{t-k}|[u]_{M-k}) \quad \text{for} \quad M \subseteq \{1, 2, \ldots, t-1\}, \quad \text{(C.2)}
\end{align*}
\]

where \( M - k \triangleq \{m - k| m \in M, m > k\} \).

**Remark 17.** Similar equality holds for estimation error function rather than differential entropy. In particular,

\[
\begin{align*}
\text{Var}(x_t|[u]_M, x_k) &= \text{Var}(x_{t-k}|[u]_{M-k}) \quad \text{for} \quad M \subseteq \{1, 2, \ldots, t-1\}. \quad \text{(C.3)}
\end{align*}
\]

This follows from the fact that for jointly Gaussian sources the estimation error satisfies

\[
h(\cdot) = \frac{1}{2} \log (2\pi e \text{Var}(\cdot)).\]

**Proof.** First consider (C.1) and note that for any \( k < j \leq t \), we have

\[
x_j = \rho^{j-k} x_k + \sum_{l=k+1}^{j} \rho^{j-l} n_l. \quad \text{(C.4)}
\]

Now for any \( M \subseteq \{1, 2, \ldots, t\} \) we have

\[
\begin{align*}
h(x_t|[u]_M, x_k) &= h\left( \rho^{t-k} x_k + \sum_{l=k+1}^{t} \rho^{t-l} n_l \left| \left\{ \rho^{j-k} x_k + \sum_{l=k+1}^{j} \rho^{j-l} n_l + z_j \right\}_{j \in M, j > k} \right. \right) \\
&= h\left( \sum_{l=k+1}^{t} \rho^{t-l} n_l \left| \left\{ \sum_{l=k+1}^{j} \rho^{j-l} n_l + z_j \right\}_{j \in M, j > k} \right. \right) \quad \text{(C.5)}
\end{align*}
\]
= h \left( \sum_{l=1}^{t-k} \rho^{t-l} n_l \left( \sum_{l=1}^{j} \rho^{j-l} n_l + z_j \right) \right)_{j \in M-k} \tag{C.6}
= h(x_{t-k}||u_{M-k}), \tag{C.7}

where (C.5) follows from the following Markov chain,
\{u_j\}_{j \in M, j > k} \rightarrow x_k \rightarrow x_t. \tag{C.8}

and subtracting $x_k$ from the argument of differential entropy. Also (C.6) follows from the stationarity of the source sequences and time-invariant property of the test channel. Using similar steps (C.2) can be proved.

**Lemma 34.** For any test channel noise $\sigma_z^2$ and any $\tau$ and $t$, we have

\[ R_{\tau+1}(t+1, \sigma_z^2) \geq R_{\tau}(t, \sigma_z^2) \tag{C.9} \]
\[ D_{\tau+1}(t+1, \sigma_z^2) \geq D_{\tau}(t, \sigma_z^2). \tag{C.10} \]

**Proof.** First consider the rate inequality in (C.9). It suffices to show that the inequality holds for all the rate expressions in (4.127). For instance using (4.120) for $R_{1,\tau}(t, \sigma_z^2)$, we have

\[ R_{1,\tau+1}(t+1, \sigma_z^2) \triangleq h(u_{t+1}||u_1) - \frac{1}{2} \log (2\pi e \sigma_z^2) \]
\[ \geq h(u_{t+1}||u_1^t, x_1) - \frac{1}{2} \log (2\pi e \sigma_z^2) \]
\[ = h(u_{t+1}||u_1^{t-1}) - \frac{1}{2} \log (2\pi e \sigma_z^2) \triangleq R_{1,\tau}(t, \sigma_z^2), \tag{C.11} \]

where (C.11) follows from the fact that conditioning reduces the differential entropy, and (C.12) follows from the application of second equality in Lemma 33 at time $t+1$ for $M = \{1, \ldots, t\}$ and $k = 1$. The similar inequalities can be derived for $R_{2,\tau}(\sigma_z^2)$ and $R_{3,\tau}(t, \sigma_z^2)$. This verifies (C.9).

The same method can be applied for the distortion constraints to show (C.10). For example for $D_{1,\tau+1}(t+1, \sigma_z^2)$ from (4.121) we have

\[ D_{1,\tau+1}(t+1, \sigma_z^2) \triangleq \text{Var}(x_{t+1}||u_1^{t+1}) \]
\[ \geq \text{Var}(x_{t+1}||u_1^{t+1}, x_1) \]
\[ = \text{Var}(x_t||u_1^t) \triangleq D_{1,\tau}(t, \sigma_z^2), \tag{C.13} \]

where (C.13) follows from the fact that revealing the additional information $x_1$ can only reduces the distortion and (C.14) follows from application of first equality Lemma 33. The similar inequalities can be derived for $D_{2,\tau}(\sigma_z^2)$ and $D_{3,\tau}(t, \sigma_z^2)$. This verifies (C.10).

Note that

\[ R_{\tau}(\sigma_z^2) = \max_t R_{\tau}(t, \sigma_z^2) \]
\[ \leq \max_t R_{\tau+1}(t+1, \sigma_z^2) \tag{C.15} \]
\[ R_{\tau+1}(\sigma_z^2), \]  

\[ (C.16) \]

where (C.15) follows from (C.9). This completes the proof for \( R_{\tau}(\sigma_z^2) \). The proof for \( D_{\tau}(\sigma_z^2) \) follows similarly.

### C.2 Proof of Lemma 15

First we show that,

\[ R_{2,\tau}(\sigma_z^2) \geq R_{1,\tau}(t, \sigma_z^2), \quad \forall t \leq \tau. \]  

\[ (C.17) \]

For any \( t < \tau \), from (4.122) we have

\[ R_{2,\tau}(\sigma_z^2) \triangleq \max_{M \subseteq L, M \neq \phi} \frac{1}{|M|} \left[ h([u]_M | [u]_1^{t-1}, [u]_{M^c}) - \frac{1}{2} \log(2\pi e \sigma_z^2) \right] \]  

\[ \geq h(u_{t+B+W} | [u]_1^{t-1}, [u]_{t+B}^{t+B+W-1}) - \frac{1}{2} \log(2\pi e \sigma_z^2) \]  

\[ \geq h(u_{t+B+W} | [u]_{t+B}^{t+B+W-1}, x_{t+B+W-t}) - \frac{1}{2} \log(2\pi e \sigma_z^2) \]  

\[ = h(u_t | [u]_{t-1}^{t-1}) - \frac{1}{2} \log(2\pi e \sigma_z^2) \triangleq R_{1,\tau}(t, \sigma_z^2), \]  

\[ (C.18) \]  

\[ (C.19) \]  

\[ (C.20) \]  

\[ (C.21) \]

where \( L_\tau \) in (C.18) is defined in (4.123). Also (C.19) follows from the choice of \( M = \{ \tau + B + W \} \) in (C.18), (C.20) follows from the fact that conditioning reduces the differential entropy, and (C.21) follows from the application of Lemma 33 at time \( \tau + B + W \), with \( M = \{ 1, \ldots, \tau + B + W - 1 \} \) and \( k = \tau + B + W - t \). This proves (C.17).

We also need to show

\[ R_{2,\tau}(\sigma_z^2) \geq R_{3,\tau}(t, \sigma_z^2), \quad \forall t > \tau + B + W \]  

\[ (C.22) \]

as \( \tau \to \infty \). From (C.19) and definition of \( R_{3,\tau}(t, \sigma_z^2) \) in (4.125), in order to show (C.22), it suffices to show

\[ h(u_t | [u]_{1}^{t-1}, [u]_{t+B}^{t-1}) \geq h(u_{t+1} | [u]_{1}^{t-1}, [u]_{t+B}^{t}), \quad \forall t > \tau + B + W \]  

\[ (C.23) \]

as \( \tau \to \infty \). Note that by the definition of the test channel for the memoryless Q-and-B scheme, \( u_t = x_t + z_t \), where \( z_t \) is independent noise. Thus showing (C.23) is equivalent to show that

\[ \text{Var}(x_t | [u]_1^{t-1}, [u]_{t+B}^{t-1}) - \text{Var}(x_{t+1} | [u]_1^{t-1}, [u]_{t+B}^{t}) \geq 0. \]  

\[ (C.24) \]

Also note that from the orthogonality principle all the observations \( [u]_1^{t-1}, [u]_{t+B}^{t-1} \) can be replaced by a noisy version of the source \( x_t \), i.e., \( x_t + \bar{n}_1 \), where \( \bar{n}_1 \sim \mathcal{N}(0, \hat{\sigma}_1^2) \). Therefore, showing (C.24) is equivalent to show that

\[ \text{Var}(x_t | x_t + \bar{n}_1) - \text{Var}(x_{t+1} | x_t + \bar{n}_1, x_{t+1} + z_{t+1}) \geq 0, \]  

\[ (C.25) \]

as \( \tau \to \infty \). Consider the following lemma.
Lemma 35. Consider jointly Gaussian random variables \( \{X_1, X_2, Y_1, Y_2\} \) as shown in Fig. C.1, such that for \( k \in \{1, 2\} \)

\[
X_k \sim \mathcal{N}(0, 1) \quad (C.26)
\]
\[
Z_k \sim \mathcal{N}(0, \epsilon_k) \quad (C.27)
\]
\[
Y_k = X_k + Z_k. \quad (C.28)
\]

Also \( X_2 = \rho X_1 + N \). Define

\[
\delta(e_1, e_2, \rho) \triangleq \text{Var}(X_1|Y_1) - \text{Var}(X_2|Y_1, Y_2) \quad (C.29)
\]

then for any \( e_2, \rho \geq 0 \),

\[
\frac{d\delta(e_1, e_2, \rho)}{de_1} \geq 0. \quad (C.30)
\]

Proof.

\[
\delta(e_1, e_2, \rho) \triangleq \text{Var}(X_1|Y_1) - \text{Var}(X_2|Y_1, Y_2)
\]

\[
= 1 - \frac{1}{1 + e_1} - 1 + \left( \rho \begin{pmatrix} 1 \\ \rho \end{pmatrix} \begin{pmatrix} 1 + e_1 & \rho \\ \rho & 1 + e_2 \end{pmatrix} \begin{pmatrix} \rho \\ 1 \end{pmatrix} \right)
\]

\[
= 1 + e_1 - \rho^2(1 - e_2) \frac{1}{(1 + e_1)(1 + e_2) - \rho^2} - \frac{1}{1 + e_1}. \quad (C.33)
\]

We have

\[
\frac{d\delta(e_1, e_2, \rho)}{de_1} = \frac{1}{(1 + e_1)^2} - \frac{\rho^2 e_2^2}{(1 + e_1)(1 + e_2) - \rho^2)^2}. \quad (C.34)
\]

It can be readily seen that (C.34) is non-negative, by simple manipulation of the following inequality.

\[
(1 + e_1)(1 + (1 - \rho)e_2) \geq 1 \geq \rho^2. \quad (C.35)
\]

This completes the proof. \( \square \)

According to Lemma 35, in order to show the positivity of \( \delta(e_1, e_2, \rho) \) for any range of \( e_1 \), it suffices to show its positivity for the smallest value of \( e_1 \).

Now recall from (C.25) that all the observations \( \{|u|^{-1}, |u|^{-1}_{r+\hat{B}}\} \) are replaced with the noisy version of the source \( x_t \), i.e., \( x_t + \hat{n}_1 \) where \( \hat{n}_1 \sim \mathcal{N}(0, \sigma_1^2) \). Now define \( x_t + \hat{n}_2, \hat{n}_2 \sim \mathcal{N}(0, \sigma_2^2) \) which is the noisy
version of the source $x_t$ equivalent to the observations $\{[u]_1^{t-1}, [u]^\tau_{t+B}\}$ plus the additional observations $\{[u]^{\tau+B-1}_t\}$. Note that as more observations is provided to produce $x_t + \tilde{n}_2$ comparing to $x_t + \tilde{n}_1$, we have $\tilde{\sigma}_2^2 \leq \tilde{\sigma}_1^2$. Based on this inequality and from the application of Lemma 35 with the following parameters

$$\{X_1, X_2, Z_1, e_1, Z_2, e_2\} = \{x_t, x_{t+1}, \tilde{n}_1, \tilde{\sigma}_1^2, z_{t+1}, \sigma_z^2\},$$

in order to show (C.25), it suffices to to show the following

$$\text{Var}(x_t|x_t + \tilde{n}_2) - \text{Var}(x_{t+1}|x_t + \tilde{n}_2, x_{t+1} + z_{t+1}) \geq 0. \quad (C.36)$$

for $t \geq \tau + B + W$, as $\tau \to \infty$. This is equivalent to show

$$\text{Var}(x_t|[u]_1^{t-1}) - \text{Var}(x_{t+1}|[u]_1^{t+1}) \geq 0. \quad (C.37)$$

As $\tau \to \infty, t \geq \tau + B + W \to \infty$. Finally note that the inequality in (C.37) holds, because according to the steady state behavior of the system, we have

$$\lim_{t \to \infty} \text{Var}(x_t|[u]_1^{t-1}) = \lim_{t \to \infty} \text{Var}(x_{t+1}|[u]_1^{t+1}). \quad (C.38)$$

This completes the proof of (C.22).

For the distortion constraint, the following two constraints can be shown using similar method.

$$D_{2,2} \geq D_{1,2}(t), \quad \forall t \leq \tau \quad (C.39)$$

$$D_{2,2} \geq D_{3,2}(t), \quad \forall t > \tau + B + W, \text{ as } \tau \to \infty. \quad (C.40)$$

This completes the proof.

### C.3 Proof of Lemma 16

We want to show that for a fixed $\tau$ such that $\tau \to \infty$, we have

$$\arg\max_{\mathcal{M} \subseteq \mathcal{L}_{\tau}, \mathcal{M} \neq \emptyset} \frac{1}{|\mathcal{M}|} h([u]_{\mathcal{M}}|[u]_1^{\tau-1}, [u]_{\mathcal{M}}) = \mathcal{L}_{\tau}, \quad (C.41)$$

where $\mathcal{L}_{\tau}$ is defined in (4.123). We prove the lemma through the following steps.

**Step 1:** We first show that, for any fixed $m \in \{1, \ldots, W+1\}$, among all $\mathcal{M} \subseteq \mathcal{L}_{\tau}$ such that $|\mathcal{M}| = m$, the maximum rate is attained by the subset $\{\tau + B + W - m + 1, \ldots, \tau + B + W\}$, i.e.,

$$\arg\max_{\mathcal{M} \subseteq \mathcal{L}_{\tau}} \frac{1}{|\mathcal{M}|} h([u]_{\mathcal{M}}|[u]_1^{\tau-1}, [u]_{\mathcal{M}}) = \{\tau + B + W - m + 1, \ldots, \tau + B + W\}. \quad (C.42)$$

To show (C.42) note that, for any $\mathcal{M} \subseteq \mathcal{L}_{\tau}$ such that $|\mathcal{M}| = m$, we have

$$h([u]_{\mathcal{M}}|[u]_0^{\tau-1}, [u]_{\mathcal{M}}) = h([u]_{\mathcal{M}}, [u]_{\mathcal{M}}|([u]_0^{\tau-1}) - h([u]_{\mathcal{M}}|[u]_0^{\tau-1}),$$

$$= h([u]_{\tau+B+W}^{\tau+B+W}|[u]_0^{\tau-1}) - h([u]_{\mathcal{M}}|[u]_0^{\tau-1}) \quad (C.43)$$
of Lemma 10 in Chapter 3. According to Lemma 36, we have
\( K \subseteq L \) where \( u \). Consider the following lemma.

**Lemma 36.** For any set \( K \subseteq L \) such that \( |K| = k \), we have:

\[
    h([u]_K \| [u]_1^{\tau-1}) \geq h([u]_{K^*} \| [u]_1^{\tau-1}),
\]

where \( K^* \triangleq \{ \tau + B, \ldots, \tau + B + k - 1 \}. \)

Fig C.2 schematically illustrates an example of Lemma 36. The proof follows from the application of Lemma 10 in Chapter 3. According to Lemma 36, we have

\[
    \arg\min_{\mathcal{M} \subseteq L \atop \mathcal{M} \neq \emptyset} h([u]_M \| [u]_0^{\tau-1}) = \{ \tau + B + W - m + 1, \ldots, \tau + B + W \},
\]

as required in (C.42).

According to Step 1,

\[
    \max_{\mathcal{M} \subseteq L \atop \mathcal{M} \neq \emptyset} \frac{1}{|\mathcal{M}|} h([u]_M \| [u]_1^{\tau-1}, [u]_M) = h([u]_{\tau+B+W}^{\tau+B+W-m+1} \| [u]_{\tau+B}^{\tau+B+W-m}).
\]

It remains to show that the term in (C.46) is an increasing function of \( m \).

**Step 2:** For \( \tau \to \infty \) and any \( m \leq W \),

\[
    \frac{1}{m+1} h([u]_{\tau+B+W}^{\tau+B+W-m} \| [u]_0^{\tau-1}, [u]_{\tau+B}^{\tau+B+W-m-1}) \geq \frac{1}{m} h([u]_{\tau+B+W}^{\tau+B+W-m+1} \| [u]_{\tau+B}^{\tau+B+W-m}).
\]

To show (C.47), it suffices to show that

\[
    mh([u]_{\tau+B+W}^{\tau+B+W-m} \| [u]_0^{\tau-1}, [u]_{\tau+B}^{\tau+B+W-m-1}) \geq (m+1)h([u]_{\tau+B+W}^{\tau+B+W-m+1} \| [u]_{\tau+B}^{\tau+B+W-m}).
\]

or equivalently,

\[
    mh([u]_{\tau+B+W-m} \| [u]_0^{\tau-1}, [u]_{\tau+B}^{\tau+B+W-m-1}) \geq h([u]_{\tau+B+W-m+1} \| [u]_{\tau+B}^{\tau+B+W-m}).
\]
Appendix C. Hybrid Coding

To show (C.49), note that
\[
\begin{align*}
h([u]_{\tau+B+W}^{\tau+B+W-m+1} | [u]_{\tau+B}^{\tau+B-W-m-1}) &= \sum_{k=0}^{m-1} h([u]_{\tau+B+W-m+k+1}^{\tau+B+W-m+k} | [u]_{\tau+B}^{\tau+B+W-m-1}) \\
&= \sum_{k=0}^{m-1} h([u]_{\tau+B+W-m}^{\tau+B+W-m-k} | [u]_{\tau+B-W-k-1}^{\tau+B}) \\
&\leq m h([u]_{\tau+B+W-m}^{\tau+B+W-m-1}) 
\end{align*}
\]

where (C.50) follows from the time invariant property among the random variables at steady state when \( \tau \to \infty \), (C.51) again follows from the application of Lemma 10 in Chapter 3. According to Lemma 10 in Chapter 3 if the random variables \( u \) with indices closer to a particular time are erased, the conditional entropy is the largest.

According to Step 2, (C.46) is an increasing function of \( m \), thus is maximized with \( m = W \). This proves (C.41) as required.

### C.4 Proof of Lemma 17

First consider the following lemma which is the generalization of Lemma 33 for the hybrid coding scheme.

**Lemma 37.** For the hybrid coding scheme and for any \( k < t \),
\[
\begin{align*}
h(x_i | [u]_{\mathcal{M}}, [x]_1^t, [z]_1^t) &= h(x_{i-k} | [u]_{\mathcal{M}-k}) \quad \text{for} \quad \mathcal{M} \subseteq \{1, 2, \ldots, t\} \\
h(u_i | [u]_{\mathcal{M}}, [x]_1^t, [z]_1^t) &= h(u_{i-k} | [u]_{\mathcal{M}-k}) \quad \text{for} \quad \mathcal{M} \subseteq \{1, 2, \ldots, t-1\},
\end{align*}
\]
where \( \mathcal{M} - k \triangleq \{m-k | m \in \mathcal{M}, m > k \} \).

**Proof.** First consider (C.52) and note that for any \( k < j \leq t \), we have
\[
x_j = \rho^{j-k} x_k + \sum_{l=k+1}^{j} \rho^{j-l} n_l. \tag{C.54}
\]

Now for any \( \mathcal{M} \subseteq \{1, 2, \ldots, t\} \) we have
\[
\begin{align*}
h(x_i | [u]_{\mathcal{M}}, [x]_1^t, [z]_1^t) &= h \left( \rho^{t-k} x_k + \sum_{l=k+1}^{t} \rho^{t-l} n_l \left\{ \sum_{l=k+1}^{\mathcal{M}, j > k} q_{j,l} (x_l + z_l) \right\} \right) \quad \text{for} \quad \mathcal{M} \subseteq \{1, 2, \ldots, t\}
\end{align*}
\]

where (C.55) follows from the fact that \( u_j \) for \( j \leq k \) are function of \( \{[x]_1^k, [z]_1^k\} \). Also (C.53) can be verified using similar methods.
To show (4.157), note that from the definition of $R_2^\tau(\tau^2_{z})$, we have

$$R_2^\tau(\tau^2_{z}) \geq h(u_{\tau+B+W}||[u]_1^{\tau-1},[u]_{\tau+B}^{\tau+B+W-1})$$

$$\geq h(u_{\tau+B+W}||[u]_1^{\tau+B+W-1},[x]_1^{\tau+B+W-t},[z]_1^{\tau+B+W-t})$$

$$= h(u||[u]_1^{\tau-1}),$$

(C.57)

where (C.57) follows from the fact that conditioning reduces the differential entropy and (C.58) follows from the application of Lemma 37 at time $\tau+B+W$ for $M = \{1, \ldots, \tau + B + W\}$ and $k = \tau + B + W - t$.

The distortion constraint in (4.158) can be verified through the similar steps.

### C.5 Proof of Lemma 18

By the application of Lemma 37 instead of Lemma 33, the proof of Lemma 18 for the hybrid coding scheme follows very similarly to the proof of Lemma 14 for the memoryless Q-and-B.

### C.6 Proof of Lemma 19

We want to show that for any $t > \tau + B + W$,

$$R_2^\tau(\tau^2_{z}) \geq R_3^\tau(t, \tau^2_{z})$$

$$D_2^\tau(\tau^2_{z}) \geq D_3^\tau(t, \tau^2_{z}),$$

(C.59)

(C.60)

as $\tau \to \infty$.

We first focus on the rate constraint (C.59). Note that by the definition of $R_2^\tau(\tau^2_{z})$ in (4.122), we have

$$R_2^\tau(\tau^2_{z}) \geq h(u_{\tau+B+W}||[u]_1^{\tau-1},[u]_{\tau+B}^{\tau+B+W-1}) - \frac{1}{2} \log (2\pi e\sigma^2_z).$$

(C.61)

According to (C.61) and the definition of $R_3^\tau(t, \tau^2_{z})$ in (4.125), in order to show (C.59), it suffices to show the following for any $t \geq \tau + B + W$,

$$h(u_t||[u]_1^{t-1},[u]_{\tau+B}^{t-1}) \geq h(u_{t+1}||[u]_1^{t-1},[u]_{\tau+B}^{t-1}).$$

(C.62)

Note that according to the definition of hybrid test channel, we have

$$x_t + z_t = u_t + \sum_{k=1}^{W} w_k u_{t-k},$$

(C.63)

i.e., $x_t + z_t$ is only the function of the current and the past $W$ test channel outputs. Thus for any $t \geq \tau + B + W$,

$$h(u_t||[u]_1^{t-1},[u]_{\tau+B}^{t-1}) = h(x_t + z_t||[u]_1^{t-1},[u]_{\tau+B}^{t-1}).$$

(C.64)
and hence (C.62) can be written as
\[ h(x_t + z_t | [u]_{t-1}^1, [u]_{t+B}^{t-1}) \geq h(x_{t+1} + z_{t+1} | [u]_{t-1}^1, [u]_{t+B}^{t-1}), \]  \( \text{(C.65)} \)
and based on the fact that \( z_t \) and \( z_{t+1} \) are i.i.d. noises, to show (C.65) it suffices to show that
\[ h(x_t | [u]_{t-1}^1, [u]_{t+B}^{t-1}) \geq h(x_{t+1} | [u]_{t-1}^1, [u]_{t+B}^{t-1}), \]  \( \text{(C.66)} \)
which is equivalent to show,
\[ \text{Var}(x_t | [u]_{t-1}^1, [u]_{t+B}^{t-1}) \geq \text{Var}(x_{t+1} | [u]_{t-1}^1, [u]_{t+B}^{t-1}). \]  \( \text{(C.67)} \)

In addition note that for any \( t \geq \tau + B + W \),
\[ \text{Var}(x_t | [u]_{t-1}^1, [u]_{t+B}^{t-1}) = \text{Var}(\rho x_{t-1} + n_t | [u]_{t-1}^1, [u]_{t+B}^{t-1}) \]
\[ = \rho^2 \text{Var}(x_{t-1} | [u]_{t-1}^1, [u]_{t+B}^{t-1}) + (1 - \rho^2). \]  \( \text{(C.68)} \)
Thus to show (C.67), it suffice to show
\[ \text{Var}(x_{t-1} | [u]_{t-1}^1, [u]_{t+B}^{t-1}) \geq \text{Var}(x_t | [u]_{t-1}^1, [u]_{t+B}^{t-1}). \]  \( \text{(C.69)} \)
and we only need to show (C.69) when \( \tau \to \infty \), which is always true according to the following lemma.

\textbf{Lemma 38.} In the hybrid coding scheme, for any test channel noise \( \sigma_z^2 \) and any \( t \geq \tau + B + W \) we have
\[ \text{Var}(x_{t-1} | [u]_{t-1}^1, [u]_{t+B}^{t-1}) \geq \text{Var}(x_t | [u]_{t-1}^1, [u]_{t+B}^{t-1}). \]  \( \text{(C.70)} \)
as \( \tau \to \infty \).

\textit{Proof.} We need to show that, in the case of hybrid coding scheme with any test channel noise \( \sigma_z^2 \), for any \( t \geq \tau + B + W \),
\[ \text{Var}(x_{t-1} | [u]_{t-1}^1, [u]_{t+B}^{t-1}) - \text{Var}(x_t | [u]_{t-1}^1, [u]_{t+B}^{t-1}) \geq 0. \]  \( \text{(C.71)} \)
as \( \tau \to \infty \).

- There exists \( n_t \sim \mathcal{N}(0, \sigma_n^2) \), such that
\[ \text{Var}(x_{t-1} | [u]_{t-1}^1, [u]_{t+B}^{t-1}) = \text{Var}(x_{t-1} x_{t-1} + n_t). \]  \( \text{(C.72)} \)
Using (C.72), for \( t \geq \tau + B + W \) one can write
\[ \text{Var}(x_t | [u]_{t-1}^1, [u]_{t+B}^{t-1}, u_t) = \text{Var}(x_t | [u]_{t-1}^1, [u]_{t+B}^{t-1}, u_t + \sum_{k=1}^{W} w_k u_{t-k}) \]
\[ = \text{Var}(x_t | [u]_{t-1}^1, [u]_{t+B}^{t-1}, x_t + z_t), \]  \( \text{(C.73)} \)
where (C.73) follows from (C.63). Therefore,

$$\text{Var}(x_{t-1} | [u]^t_{1-1}, [u]^{t-1}_{t+B}) - \text{Var}(x_t | [u]^t_{1-1}, [u]^{t}_{t+B}) = \text{Var}(x_{t-1} | x_{t-1} + \tilde{n}_1) - \text{Var}(x_t | x_{t-1} + \tilde{n}_1, x_t + z_t),$$

i.e., in order to show (C.71), it suffices to show the non-negativity of (C.74).

- There also exists \( \tilde{n}_2 \sim \mathcal{N}(0, \tilde{\sigma}_2^2) \), such that.

$$\text{Var}(x_{t-1} | [u]^t_{1-1}) = \text{Var}(x_{t-1} | x_{t-1} + \tilde{n}_2).$$

Similarly to (C.74), it can be shown that

$$\text{Var}(x_{t-1} | [u]^t_{1-1}) - \text{Var}(x_t | [u]^t_{1}) = \text{Var}(x_{t-1} | x_{t-1} + \tilde{n}_2) - \text{Var}(x_t | x_{t-1} + \tilde{n}_2, x_t + z_t).$$

- The fact that

$$\text{Var}(x_{t-1} | [u]^t_{1-1}, [u]^{t-1}_{t+B}) \geq \text{Var}(x_{t-1} | [u]^t_{1-1}),$$

results in \( \tilde{\sigma}_2^2 \leq \tilde{\sigma}_1^2 \). Thus, according to Lemma 35, in order to prove the non-negativity of (C.74), it suffices to prove it for (C.76).

- Based on the fact that in steady state as \( \tau \to \infty, t \to \infty \), and we have

$$\lim_{t \to \infty} \text{Var}(x_{t-1} | [u]^t_{1-1}) = \lim_{t \to \infty} \text{Var}(x_t | [u]^t_{1}).$$

This verifies the non-negativity of (C.76) and completes the proof.

Now consider the distortion constraint in (C.60). By definition, it suffices to show that for any \( t \geq \tau + B + W \),

$$\text{Var}(x_t | [u]^t_{1-1}, [u]^{t}_{t+B}) \geq \text{Var}(x_{t+1} | [u]^t_{1-1}, [u]^{t+1}_{t+B}),$$

which is readily justified according to Lemma 38. This proves the distortion constraint in (4.156).

### C.7 Proof of Lemma 20

Note that,

$$I([u]^t_{t-B+1}; [u]^{t+B}_{t+1} | [u]^{t-1}_1, s_t + w_{t+1}) = h([u]^{t+W+1}_{t+1} | [u]^{t-1}_1, s_t + w_{t+1}) - h([u]^{t+W+1}_{t+1} | [u]^{t+1}_{t+1}, s_t + w_{t+1})
\begin{align*}
= h([u]^{t+W+1}_{t+1} | [u]^{t-1}_1) &+ h(s_{t+w+1} | [u]^{t+B}_{t+1}) \\
&- h([u]^{t+W+1}_{t+1} | [u]^{t+1}_{t+1}) &- h(s_{t+w+1} | [u]^{t+W+1}_{t+1}).
\end{align*}$$

(C.80)
Note that the second term in (C.80) is equal to \( \frac{1}{2} \log (2\pi eD) \) as \( t \to \infty \) and is independent of \( \mathbf{w} \). The third, fifth and sixth terms are also independent of \( \mathbf{w} \), because of the invertibility of matrix \( \mathbf{A} \) in hybrid coding scheme. For instance, for the third term we have

\[
h(s_t + W + 1 | [u]_1^{t-B}) = h(s_t + W + 1 | [s]_1^{t-B} + [z]_{1}^{t-B}).
\]  

(C.81)

Also the fourth term is independent of \( \mathbf{w} \), because

\[
h([u]_{t+1}^{t+W+1} | [u]_1^{t-1}) = \sum_{j=t+1}^{t+W+1} h(u_j | [u]_1^{j-1}),
\]  

(C.82)

which is independent of the choice of \( \mathbf{w} \), because, for any \( j \), \( h(u_j | [u]_1^{j-1}) \) is independent of the choice of \( \mathbf{w} \), i.e.,

\[
h(u_j | [u]_1^{j-1}) = h(u_j | [s]_1^{j-1} + [z]^{j-1})
\]  

(C.83)

\[= h\left((s_j + z_j) + \sum_{k=1}^{j-1} q_{j,k} (s_k + z_k) | [s]_1^{j-1} + [z]^{j-1}\right)
\]  

(C.84)

\[= h(s_j + z_j | [s]_1^{j-1} + [z]^{j-1}).
\]  

(C.85)

Thus the choice of \( \mathbf{w} \) which minimizes the first term in (C.80), minimizes the mutual information. This completes the proof.

**C.8 Proof of Lemma 22**

From (4.215), consider the following definition.

\[
\begin{pmatrix}
  u_t \\
  u_{t-1}
\end{pmatrix}
= \begin{pmatrix}
  1 & -w_1 & w_1^2 & -w_1^3 & \cdots \\
  0 & 1 & -w_1 & w_1^2 & \cdots
\end{pmatrix}
\begin{pmatrix}
  s_t \\
  s_{t-1} \\
  s_{t-2} \\
  \vdots
\end{pmatrix}
\]  

(C.86)

\[
\Delta \triangleq \begin{pmatrix}
  a_0 & a_1 & a_2 & \cdots \\
  0 & a_0 & a_1 & \cdots
\end{pmatrix}
\begin{pmatrix}
  n_t \\
  n_{t-1} \\
  \vdots
\end{pmatrix}.
\]  

(C.87)

It is not hard to observe that, for any \( m \geq 0 \),

\[
a_m = \sum_{k=0}^{m} (-w_1)^k \rho^{m-k}.
\]  

(C.88)

We have

\[
E\{|u_t|^2\} = E\{|u_{t-1}|^2\} = (1 - \rho^2) \sum_{m=0}^{\infty} a_m^2.
\]  

(C.89)
Note that

\[
a_m^2 = \sum_{l=0}^{m} \sum_{k=0}^{m} (-w_1)^{k+l} \rho^{2m-k-l}
\]

\[
= \sum_{l=0}^{m} \sum_{k=0}^{m} \rho^{2m} \left( \frac{-w_1}{\rho} \right)^{k+l}
\]

\[
= \rho^{2m} \sum_{j=0}^{m-1} (j+1) \left( \frac{-w_1}{\rho} \right)^j + \left( \frac{-w_1}{\rho} \right)^{2m-j} + \rho^{2m} (m+1) \left( \frac{-w_1}{\rho} \right)^m,
\]

(C.91)

and therefore,

\[
\sum_{m=0}^{\infty} a_m^2 = \sum_{m=0}^{\infty} (m+1)(-w_1 \rho)^m + \sum_{m=0}^{\infty} \rho^{2m} (j+1) \left( \frac{-w_1}{\rho} \right)^j + \left( \frac{-w_1}{\rho} \right)^{2m-j}
\]

\[
= \frac{1}{(1+w_1 \rho)^2} + \sum_{j=0}^{\infty} \rho^{2m} (j+1) \left( \frac{-w_1}{\rho} \right)^j + \left( \frac{-w_1}{\rho} \right)^{2m-j}
\]

\[
= \frac{1}{(1+w_1 \rho)^2} + \sum_{j=0}^{\infty} \left( (j+1) \left( \frac{-w_1}{\rho} \right)^j \sum_{m=j+1}^{\infty} \rho^{2m} + (j+1) \left( \frac{-w_1}{\rho} \right)^{-j} \sum_{m=j+1}^{\infty} \rho^{2m} \left( \frac{-w_1}{\rho} \right)^{2m} \right)
\]

\[
= \frac{1}{(1+w_1 \rho)^2} + \sum_{j=0}^{\infty} \left( (j+1) \left( \frac{-w_1}{\rho} \right)^j \rho^{2(j+1)} + (j+1) \left( \frac{-w_1}{\rho} \right)^{-j} \frac{w_1^{2(j+1)}}{1-w_1^2} \right)
\]

\[
= \frac{1}{(1+w_1 \rho)^2} \left( \frac{\rho^2}{1-\rho^2} + \frac{w_1^2}{1-w_1^2} \right) \frac{1}{1-w_1^2} 
\]

\[
= \left( \frac{\rho^2}{1-\rho^2} + \frac{1}{1-w_1^2} \right) \frac{1}{1+w_1 \rho} \rho f(w_1).
\]

(C.92)

Similarly,

\[
\sum_{m=0}^{\infty} a_m a_{m+1} = \sum_{m=0}^{\infty} \sum_{l=0}^{m} \sum_{k=0}^{l} \rho^{2m+1} \left( \frac{-w_1}{\rho} \right)^{k+l}
\]

\[
= \rho f(w_1) + \sum_{m=0}^{\infty} \sum_{l=0}^{m} \rho^{2m+1} \left( \frac{-w_1}{\rho} \right)^{m+l+1}
\]

\[
= \rho f(w_1) + \sum_{l=0}^{\infty} \sum_{m=l}^{\infty} \rho^{2m+1} \left( \frac{-w_1}{\rho} \right)^{m+l+1}
\]

\[
= \rho f(w_1) - \frac{w_1}{(1+w_1 \rho)(1-w_1^2)} g(w_1).
\]

(C.93)
C.9 Proof of Lemma 23

First note that for any $\rho \in (0, 1)$ and $x \in \mathbb{R}$ the function
\[
f(x) = x - \frac{1}{2} \log (\rho^{2m} 2^{2x} + 2\pi e(1 - \rho^{2m}))
\]
(C.94)
is an monotonically increasing function with respect to $x$, because
\[
f'(x) = \frac{2\pi e(1 - \rho^{2m})}{\rho^{2m} 2^{2x} + 2\pi e(1 - \rho^{2m})} > 0.
\]
(C.95)
By applying Shannon’s EPI we have,
\[
h(s_b|f_a) \geq \frac{1}{2} \log \left( \rho^{2m} 2^{h(s_a|f_a)} + 2\pi e(1 - \rho^{2m}) \right)
\]
(C.96)
and thus,
\[
h(s_a|f_a) - h(s_b|f_a) \\
\leq h(s_a|f_a) - \frac{1}{2} \log \left( \rho^{2m} 2^{h(s_a|f_a)} + 2\pi e(1 - \rho^{2m}) \right)
\]
(C.97)
\[
\leq \frac{1}{2} \log (2\pi e r) - \frac{1}{2} \log \left( \rho^{2m} 2\pi e r + 2\pi e(1 - \rho^{2m}) \right)
\]
(C.98)
\[
= \frac{1}{2} \log \left( \frac{r}{1 - (1-r)\rho^{2m}} \right).
\]
(C.99)
where (C.98) follows from the assumption that $h(s_a|f_a) \leq \frac{1}{2} \log (2\pi e r)$ and the monotonicity property of $f(x)$. This completes the proof.
Appendix D

Delay-Constrained Streaming

D.1 Proof of Corollary 6

According to the chain rule of entropies, the term in (2.7) can be written as

\[ H(s_{B+1}, s_{B+2}, \ldots, s_{B+W+T+1}|s_0) = H(s_{B+1}|s_0) + \sum_{k=1}^{W+T} H(s_{B+k+1}|s_0, s_{B+1}, \ldots, s_{B+k}) \]

\[ = H(s_{B+1}|s_0) + (W + T)H(s_1|s_0) \quad (D.1) \]
\[ = H(s_{B+1}|s_0) - H(s_{B+1}|s_B, s_0) + H(s_{B+1}|s_B) + (W + T)H(s_1|s_0) \quad (D.2) \]
\[ = H(s_{B+1}|s_0) - H(s_{B+1}|s_B, s_0) + H(s_{B+1}|s_B) + (W + T)H(s_1|s_0) \quad (D.3) \]
\[ = I(s_{B+1}; s_B|s_0) + (W + T + 1)H(s_1|s_0) \quad (D.4) \]
\[ = (W + T + 1)R^+(B, W), \quad (D.5) \]

where (D.1) follows from the Markov property

\[ s_0, s_{B+1}, \ldots, s_{B+k-1} \rightarrow s_{B+k} \rightarrow s_{B+k+1} \quad (D.6) \]

for any \( k \) and from the temporally independency and stationarity of the sources which for each \( k \) implies that

\[ H(s_{B+k+1}|s_{B+k}) = H(s_1|s_0). \quad (D.7) \]

Note that in (D.2) we add and subtract the same term and (D.3) also follows from the Markov property of (D.6) for \( k = 0 \).
D.2 Proof of Corollary 11

Consider a specific setup with periodic erasure pattern where the decoder at time $i$ has access to $s_{i-B-W-T-1}$ and multiple erasure bursts each of length $B$ spans the interval

$$\{kG + i - B - W - T, \ldots, kG + i - W - T - 1\}$$

for $k \geq 0$. Consider the following two cases:

- For any $T' \leq G - W - 1$ we have
  $$R_{CI}^+(B, W, T', G) \triangleq H(s_1|s_0) + \frac{1}{T' + W + 1} I(s_B; s_{B+1}|s_0) \geq H(s_1|s_0) + \frac{1}{G} I(s_B; s_{B+1}|s_0) \triangleq R_{CI}^+(B, W, G - W - 1, G).$$

This suggests that the coding scheme for delay less than $G - W - 1$ requires higher rate comparing to $T_{opt} = G - W - 1$.

- For any $G - W - 1 \leq T'$, define integers $q$ and $r \in \{0, \ldots, B + G - 1\}$ such that
  $$T' + W - G + 1 = q(B + G) + r.$$

1) If $r \in \{0, \ldots, B\}$, the rate should satisfy

$$R \geq \frac{1}{(q + 1)G} H \left( [s]_1^{W-T+G-1}, [s]_{G+B+i-W-T+G-1}, \ldots, [s]_{q(G+B)+i-W-T} | s_{i-B-W-T-1} \right)$$

$$= \frac{q + 1}{(q + 1)G} \left( H(s_{B+1}|s_0) + (G - 1)H(s_1|s_0) \right)$$

$$= H(s_1|s_0) + \frac{1}{G} I(s_B; s_{B+1}|s_0) \triangleq R_{CI}^+(B, W, G - W - 1, G).$$

2) If $r \in \{B + 1, \ldots, B + G - 1\}$, the rate should satisfy

$$R \geq \frac{H(\{[s]_{k(G+B)+i-W-T+G-1}|0 \leq k \leq q\}, [s]_{(q+1)(G+B)+i-W-T+r-B} | s_{i-B-W-T-1})}{(q + 1)G + r - B}.$$  

The right hand side of (D.13) is equivalent to the following.

$$\frac{q + 1}{(q + 1)G + r - B} (H(s_{B+1}|s_0) + (G - 1)H(s_1|s_0))$$

$$+ \frac{1}{(q + 1)G + r - B} (H(s_{B+1}|s_0) + (r - B - 1)H(s_1|s_0))$$

$$= \frac{1}{(q + 1)G + r - B} \left( ((q + 1)G + r - B)H(s_1|s_0) + (q + 2)I(s_B; s_{B+1}|s_0) \right)$$

$$= H(s_1|s_0) + \frac{q + 2}{(q + 1)G + r - B} I(s_B; s_{B+1}|s_0).$$
Also note that $G \geq 1$ and the fact that $r - B \in [1, G - 1]$ imply that
\[
\frac{q + 2}{(q + 1)G + r - B} \geq \frac{1}{G}. \tag{D.17}
\]
According to (D.17) it can be observed that the rate requirement in (D.16) is higher than $R_{Cl}^+(B, W, G - W - 1, G)$.

Based on the above reasoning and (D.12), for the random binning over sliding-window channel model with parameter $G$, delays more than $T_{\text{opt}} = G - W - 1$ does not reduce the rate.

### D.3 Proof of Lemma 27

We want to show that for any fixed $B'$ in region 3 and for $t \to \infty$, we have
\[
\max_{\mathcal{M} \subseteq \{t - T, \ldots, t\}} \frac{1}{|\mathcal{M}|} I(s_{\mathcal{M}}; u_{\mathcal{M}}|u_{0}^{t-T-B'-1}, u_{\mathcal{M}'}_{t}, s_{-1}) = \frac{1}{T + 1} I([s]_{t-T}^{t}; [u]_{t}^{t-T-B'-1}, s_{-1}). \tag{D.18}
\]

**Claim 2.** For any fixed $m \in \{0, 1, \ldots, T + 1\}$, among all $\mathcal{M} \subseteq \{t - T, \ldots, t\}$ such that $|\mathcal{M}| = m$, the maximum is attained by the subset $\{t - m + 1, \ldots, t\}$, i.e.,
\[
\max_{\mathcal{M} \subseteq \{t - T, \ldots, t\}} \frac{1}{|\mathcal{M}|} I(s_{\mathcal{M}}; u_{\mathcal{M}}|u_{0}^{t-T-B'-1}, u_{\mathcal{M}'}_{t}, s_{-1}) = \frac{1}{m} I([s]_{t-m+1}^{t}; [u]_{t-m+1}^{t-T-B'-1}, [u]_{t-T}, s_{-1}). \tag{D.19}
\]

**Proof.** For any $\mathcal{M} \subseteq \{t - T, \ldots, t\}$ such that $|\mathcal{M}| = m$, we have
\[
I([s]_{\mathcal{M}}; [u]_{\mathcal{M}}|u_{0}^{t-T-B'-1}, u_{\mathcal{M}'}_{t}, s_{-1}) = h([u]_{\mathcal{M}}|u_{0}^{t-T-B'-1}, u_{\mathcal{M}'}_{t}, s_{-1})
\]
\[
- h([u]_{\mathcal{M}}|s_{\mathcal{M}}); [u]_{0}^{t-T-B'-1}, [u]_{\mathcal{M}'}_{t}, s_{-1})
\]
\[
= h([u]_{\mathcal{M}}|s_{\mathcal{M}}); [u]_{0}^{t-T-B'-1}, s_{-1})
\]
\[
- h([u]_{\mathcal{M}}|s_{\mathcal{M}}); [u]_{0}^{t-T-B'-1}, s_{-1})
\]
\[
= h([u]_{0}^{t-T-B'-1}|s_{-1}) - m h(u_{1}|s_{1}) \tag{D.20}
\]
\[
= h([u]_{0}^{t-T-B'-1}|s_{-1}) - m h(u_{1}|s_{1}) \tag{D.21}
\]
\[
= h([u]_{0}^{t-T-B'-1}|s_{-1}) - m h(u_{1}|s_{1}) \tag{D.22}
\]
\[
= I([s]_{t-m+1}^{t}; [u]_{t-m+1}^{t-T-B'-1}, [u]_{t-T}, s_{-1}) \tag{D.23}
\]

**Claim 3.** As $t \to \infty$, the term in (D.19) is an increasing function with respect to $m$. \qed
Proof. Consider the term in the regime when $t \to \infty$. For any $m \leq T$ we want to show

$$\frac{1}{m+1} I([s]_t^t; [u]_t^t, [u]_0^{t-T-B-1}, [u]_t^t, s) \geq \frac{1}{m} I([s]_t^t; [u]_t^t, [u]_0^{t-B-1}, [u]_t^t, s)$$

(D.24)

This is equivalent to show that

$$mh([u]_t^t, [u]_0^{t-B-1}, [u]_t^t, s) \geq (m+1)h([u]_t^t, [u]_0^{t-B-1}, [u]_t^t, s)$$

(D.25)

or equivalently,

$$mh([u]_t^t, [u]_0^{t-B-1}, [u]_t^t, s) \geq h([u]_t^t, [u]_0^{t-B-1}, [u]_t^t, s).$$

(D.26)

Note that because we are considering the regime when $t \to \infty$, (D.26) is equivalent to showing

$$mh([u]_t^t, [u]_0^{t-B-1}, [u]_t^t, s) \geq h([u]_t^t, [u]_0^{t-B-1}, [u]_t^t)$$

(D.27)

To show (D.27), note that

$$h([u]_t^t, [u]_0^{t-B-1}, [u]_t^t) = \sum_{k=0}^{m-1} h([u]_{t-m+k}, [u]_0^{t-B-1}, [u]_{t-B-k})$$

(D.28)

$$= \sum_{k=0}^{m-1} h([u]_{t-m}, [u]_0^{t-B-k}, [u]_{t-B-k-1})$$

(D.29)

where (D.29) follows from the application of Lemma 10. This completes the proof.

D.4 Proof of Lemma 29

Note that

$$\Sigma_X^{-1} = ((1 - \rho^2)AA^T)^{-1}$$

$$= \frac{1}{1 - \rho^2} AA^T$$

(D.30)

$$= \frac{1}{1 - \rho^2} \begin{pmatrix} 1 & -\rho & 0 & \cdots & 0 \\ 0 & 1 & -\rho & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & \cdots & 0 \\ -\rho & 1 & \cdots & 0 \\ 0 & -\rho & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}$$
\[
\begin{pmatrix}
1 + \rho^2 & -\rho & 0 & \cdots & 0 & 0 \\
-\rho & 1 + \rho^2 & -\rho & \cdots & 0 & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & 0 & \cdots & 1 + \rho^2 & -\rho \\
0 & 0 & 0 & \cdots & -\rho & 1
\end{pmatrix}
\]

Also note that \((\Sigma_X - DI) \succeq 0\) if and only if \(\Theta \triangleq (D^{-1}I - \Sigma_X^{-1}) \succeq 0\). Thus we only need to show that \(\Theta\), i.e.,

\[
\Theta = \begin{pmatrix}
\alpha & b \\
\frac{b^\dagger}{\alpha} & C
\end{pmatrix}
\]

is non-negative definite, where

\[
\begin{align*}
\alpha & \triangleq \frac{1}{D} - \frac{1 + \rho^2}{1 - \rho^2}, \\
\beta & \triangleq \frac{\rho}{1 - \rho^2}, \\
\frac{b}{\alpha} & \triangleq (\beta, 0, \cdots, 0), \\
C & \triangleq \begin{pmatrix}
\alpha & \beta & \cdots & 0 \\
\beta & \alpha & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & \beta \\
0 & 0 & \cdots & \frac{1}{D} - \frac{1}{1 - \rho^2}
\end{pmatrix},
\end{align*}
\]

In addition, note that \(\Theta\) is non-negative definite if and only if

- \(\alpha \geq 0\), i.e., \(D \leq \frac{1 - \rho^2}{1 + \rho^2}\), and
- Its Schur complement \(S_1 \triangleq C - \frac{b^\dagger b}{\alpha}\) is non-negative definite, i.e.,

\[
S_1 = \begin{pmatrix}
\alpha - \frac{\beta^2}{\alpha} & \beta & \cdots & 0 \\
\beta & \alpha & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & \beta \\
0 & 0 & \cdots & \frac{1}{D} - \frac{1}{1 - \rho^2}
\end{pmatrix}
\]

The matrix \(S_1\) is itself non-negative definite if and only if

- \(\alpha - \frac{\beta^2}{\alpha} \geq 0\), i.e., \(D \leq \frac{1 - \rho^2}{1 + \rho + \rho^2}\), and
Its similarly defined Schur complement $S_2$ is non-negative definite, i.e.,

$$S_2 = \begin{pmatrix}
\alpha - \frac{\beta^2}{\alpha - \frac{\beta^2}{\alpha}} & \beta & \cdots & 0 \\
\beta & \alpha & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & \beta \\
0 & 0 & \cdots & \frac{1}{D} - \frac{1}{1 - \rho^2}
\end{pmatrix}. \tag{D.39}$$

We can define the following recursive formula for any $k \geq 1$,

$$\alpha_{k+1} = \alpha_k - \frac{\beta^2}{\alpha_k}, \tag{D.40}$$

where $\alpha_1 = \alpha$. Using this notation, in order to show that $\Theta$ is non-negative definite, it suffice to show $\alpha_k \geq 0$ for any $k \geq 1$. First from (D.40) note that $\alpha_k \geq 0$ results in $\alpha_{k+1} \leq \alpha_k$. Thus if

$$\alpha_\infty \triangleq \lim_{l \to \infty} \alpha_l \geq 0 \tag{D.41}$$

then $\alpha_k \geq 0$ for any $k$. We can write

$$\alpha_\infty = \alpha - \frac{\beta^2}{\alpha - \frac{\beta^2}{\alpha - \frac{\beta^2}{\alpha}}}, \tag{D.42}$$

and therefore,

$$\alpha_\infty = \alpha - \frac{\beta^2}{\alpha_\infty}, \tag{D.43}$$

and accordingly to have $\alpha_\infty \geq 0$, we need to have $\alpha \geq 2\beta$ which results in

$$D \leq \frac{1 - \rho}{1 + \rho}, \tag{D.44}$$

which completes the proof.
Bibliography


