Zero-Delay Sequential Transmission of Markov Sources over Burst Erasure Channels

Farrokh Etezadi, Ashish Khisti and Mitchell Trott

Abstract

A setup involving sequential transmission Markov sources over a burst-erasure channel is studied. A sequence of source vectors is compressed in a causal fashion and the resulting output is transmitted over a burst erasure channel. The channel erases up to \( B \) packets in a single burst, but reveals all the other packets to the destination. The destination is required to reconstruct each source vector with zero delay, but those source vectors that are either observed during the erasure burst, or in the interval of length \( W \) following the erasure burst need not be reconstructed. The minimum achievable compression rate is called the rate-recovery function. Each source vector is sampled i.i.d. across the spatial dimension from a stationary, first-order Markov process across the temporal dimension.

For discrete sources, the case of lossless recovery is considered and upper and lower bounds on the rate-recovery function are established. The proposed coding scheme is based on random binning, whereas the lower bound is obtained based on a connection to a multi-terminal source coding problem. The upper and lower bounds coincide when \( W = 0 \) and as \( W \) approaches \( \infty \). More generally, both the upper and lower bounds equal the rate for predictive coding, plus a term that decreases at least inversely with the recovery window length \( W \). In the special case where the encoder is restricted to be memoryless, and the class of symmetric Markov sources is considered, the optimality of a binning scheme is established.

For Gauss-Markov sources and a quadratic distortion measure, upper and lower bounds on the minimum rate are established when \( W = 0 \). The upper bound is based on a natural quantize and binning scheme, but the analysis of the achievable rate involves an explicit characterization of the worst-case erasure burst. The upper and lower bounds coincide in the high resolution limit. Another setup involving i.i.d. Gaussian sources and a sliding window recovery constraint is also studied, and the rate-recovery function completely characterized for this case. The proposed coding scheme involves a successive refinement codebook for quantization, a judicious assignment of the refinement layers to each channel packet, and random binning.

Finally another extension where the zero-delay constraint is relaxed to permit finite decoding delays is considered. For the case of discrete sources, \( W = 0 \), and lossless recovery, the optimality of a binning scheme is again established.

Index Terms

Multi-terminal Information Theory, Rate-distortion Theory, Sequential and Predictive Coding, Joint Source-Channel coding.

I. INTRODUCTION

Compression rate and error resilience are two conflicting requirements in video compression systems. At one extreme, predictive coding achieves the maximum possible compression but is sensitive to packet losses. At the other extreme, still-image coding does not incur any error propagation but incurs a significant increase in the compression rate. A variety of techniques are used in practice to strike a balance between these extremes. Many video codecs use a combination of still image coding and predictive coding to limit error propagation. Techniques such as forward error correction [1], [2], leaky DPCM [3], [4] and distributed video coding [5], [6] are also used for dealing with error propagation. However a fundamental tradeoff between error resilience and compression rate, using an information theoretic framework, has not been addressed before. As such even the effect of a single isolated packet loss has not been fully understood [7].

In this paper we introduce an information theoretic framework to study the tradeoff between the compression rate and error propagation at the destination, over a class of burst-erasure channels. The encoder observes a sequence of vector sources and compresses them in a causal fashion. The encoder output is transmitted over a channel that can introduce an erasure burst of a given maximum length. The receiver is required to reconstruct all the source vectors with zero delay, except those that occur during the error propagation window. This period spans the duration of the erasure burst and an interval of length \( W \) immediately following the burst. We study the minimum rate required in this setup and develop information theoretic upper and lower bounds, that coincide in some special cases. For the case of discrete sources and lossless recovery, both our upper and lower bounds equal the rate of predictive coding plus a term that decreases at least as \( H(s) / (W + 1) \) where \( H(s) \) denotes the entropy of the source symbol. For the case of Gauss-Markov sources and a quadratic distortion measure we derive upper and lower bounds on minimum rate when \( W = 0 \) that coincide in the high resolution limit. We also present another setup involving i.i.d. Gaussian sources, and a sliding window recovery constraint, where an exact characterization of the minimum rate is obtained.

Problems involving real-time coding and compression have been studied from many different perspectives in related literature. The sequential compression of a Markov source with zero encoding and decoding delay was studied in an early work by...
\[ s_0^n \rightarrow s_1^n \rightarrow \ldots \rightarrow s_{n-1}^n \rightarrow s_n^n \rightarrow s_{n+1}^n \rightarrow \ldots \rightarrow s_{j_0+B+W-1}^n \rightarrow s_{j_0+B+W}^n \]

<table>
<thead>
<tr>
<th>( f_0 )</th>
<th>( f_1 )</th>
<th>( f_2 )</th>
<th>( f_{j-1} )</th>
<th>( f_j )</th>
<th>( f_{j+1} )</th>
<th>( f_{j+B-B-1} )</th>
<th>( f_{j+B} )</th>
<th>( f_{j+B+W-1} )</th>
<th>( f_{j+B+W} )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Error Propagation Window

\[
\hat{s}_0^n \quad \hat{s}_1^n \quad \hat{s}_2^n \quad \hat{s}_{j-1}^n
\]

\[
\text{Erased} \quad \not\text{to be recovered}
\]

Fig. 1: Problem Setup: The encoder output \( f_i \) is a function of the source sequences up to time \( i \) i.e., \( s_0^n, s_1^n, \ldots, s_i^n \). The channel introduces an erasure burst of length \( B \). The decoder produces \( \hat{s}_i^n \) upon observing the sequence \( \{g_0, g_1, \ldots, g_l\} \). 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.

Witsenhausen [8]. It was shown that, without loss of optimality, it suffices to restrict the encoding function to depend on the last \( k \) source symbols and the memory of the decoder, where \( k \) denotes the order of the Markov source. Similar structural results involving noisy channels, side information and limited lookahead have also been obtained recently, see, e.g., [9], [10] and references therein. Lossless sequential compression of memoryless sources, with and without side information at the decoder and fixed decoding delay has been studied in [11], [12]. 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. [13], [14] and related references.

Viswanathan and Berger introduce a multi-terminal source coding problem [15] motivated by the application to video coding. A set of correlated sources is required to 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 until that time. The correlated source sequences model consecutive video frames and each stage at the decoder maps to sequential reconstruction of a particular source sequence. More recently in reference [16], the authors study robust extension of this setup where at any given stage the decoder has either all the previous outputs, or only the present output. However this too does not capture the effect of packet losses over a channel, where the destination has access to all the non erased packets. To the best of our knowledge our proposed framework has not been studied in the literature.

The remainder of the paper is organized as follows. The problem setup is described in section II and a summary of the main results is provided in section III. We treat the case of discrete sources and lossless recovery in section IV and establish upper and lower bounds on the minimum rate. The optimality of binning for the special case of symmetric sources and memoryless encoders is established in section V. In section VI we consider the case of Gauss-Markov source with a quadratic distortion constraint. Section VII studies another setup involving independent Gaussian sources and a sliding window recovery constraint, where an exact characterization of the minimum rate is obtained.

**Notations:** Throughout this paper 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 “slanted sans serif” font \( \mathbf{a} \) 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 \( i \). For compactness we will use bold-face font \( \mathbf{a} \) to denote \( n \)-letter sequence \( a^n \). The notation \( [f_i]^j \) for \( i < j \) represents \( f_i, f_{i+1}, \ldots, f_j \).

**II. PROBLEM STATEMENT**

In this section we introduce our source and channel models and the associated definition of the rate-recovery function. We consider a sequence of source vectors \( \{s_i^n\}_{i \geq 1} \) whose symbols are drawn independently across the spatial dimension and from a first-order Markov chain across the temporal dimension, i.e.,

\[
\Pr(s_i^n | s_{i-1}^n, s_{i-2}^n, \ldots) = \prod_{j=1}^{n} p(s_i | s_{i-1,j}), \quad \forall i \geq 0.
\]

The underlying random variables \( \{s_i\} \) constitute a time-invariant, stationary and a first-order Markov chain with a common marginal distribution denoted by \( p_s(\cdot) \). The sequence \( s_i^n \) is sampled i.i.d. from \( p_s(\cdot) \) and revealed to both the encoder and decoder before the start of the communication. It is a synchronization frame at the start of the communication.
A rate-$R$ encoder computes an index $f_i \in [1, 2^{nR}]$ at time $i$, according to an encoding function

$$f_i = F_i(s^n_{i-1}, s^n_0, \ldots, s^n_i). \tag{2}$$

Note that the encoder in (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.

The channel takes each $f_i$ as input and either outputs $f_i$ or an erasure symbol. We consider the class of burst erasure channels. For some particular $j \geq 0$, it introduces an erasure burst such that $g_i = *$ for $i \in \{j, j+1, \ldots, j+B'-1\}$ and $g_i = f_i$ otherwise i.e.,

$$g_i = \begin{cases} * & i \in \{j, j+1, \ldots, j+B'-1\} \\ f_i, & \text{else} \end{cases} \tag{3}$$

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.,

$$\hat{s}^n_i = G_i(g_0, g_1, \ldots, g_i, s^n_{i+1}), \quad i \notin \{j, \ldots, j+B'+W-1\} \tag{4}$$

where $\hat{s}^n_i$ denotes the reconstruction sequence and $j$ denotes the time at which erasure burst starts (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. 1 provides a schematic of the causal encoder (2), the channel model (3), and the decoder (4). We next consider the definition of the rate-recovery function.

A. Rate-Recovery Function

We first consider the case when the reconstruction in (4) is required to be lossless. In this case we assume that the source alphabet is discrete and the entropy $H(s)$ is finite. A rate $R(B, W)$ is feasible if there exists a sequence of encoding and decoding functions and a sequence $e_n$ that approaches zero as $n \to \infty$ such that, $\Pr(s^n_i \neq \hat{s}^n_i) \leq e_n$, for all source sequences reconstructed as in (4). We seek the minimum feasible rate $R(B, W)$, which is the lossless rate-recovery function.

We also consider the case where reconstruction in (4) is required to satisfy an average distortion constraint:

$$\limsup_{n \to \infty} E \left[ \frac{1}{n} \sum_{k=1}^{n} d(s_{i,k}, \hat{s}_{i,k}) \right] \leq D \tag{5}$$

for some distortion measure $d : \mathbb{R}^2 \to [0, \infty)$. The rate $R(B, W, D)$ is feasible if a sequence of encoding and decoding functions exists that satisfy the average distortion constraint. The minimum feasible rate $R(B, W, D)$, is the lossy rate-recovery function. In this paper we will focus on the two classes of Gaussian sources and associated distortion measures that will be explained in the sequel.

B. Practical Motivation

Note that our setup assumes that both the size of the source frames and channel packets is sufficiently large. In practice such a scenario may be justified in video streaming applications. 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 $N = 2\Delta_s W_s \approx 84,000$. Thus the block length between successive frames is sufficiently long that capacity achieving codes could be used and the erasure model and large packet sizes is justified. The assumption of spatially i.i.d. frames could reasonably approximate the video innovation process generated by applying suitable transform on original video frames. Such source models have been used in earlier works on the information theoretic treatment of video coding, e.g., [15].

The burst loss model considered in our setup is often used to capture effects such as fading in wireless channels and congestion in wired networks. We note that the present paper does not consider a statistical channel model but instead considers a worst case 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 while the statistical models are used to capture the typical behaviour of channel errors, the atypical behaviour is often modelled (see e.g., [17, Sec. 6.10]) using a worst-case approach. Therefore in low-latency applications where the local channel dynamics are relevant such models are often used (see e.g., [18], [19]).

In practice the communication duration is finite however we will not focus on the edge effects due to the truncation. Instead, we study the steady state behaviour by assuming that the duration is as large as possible. Finally patterns beyond burst losses can also be of interest, but will not be treated in this paper.
In this section we discuss the main results of this paper.

A. Lossless 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

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

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

Notice that the upper and lower bounds (6) and (7) coincide for \( W = 0 \) and \( W \to \infty \), yielding the rate-recovery function in these cases. More generally we can interpret the term \( H(s_1|s_0) \) as the rate associated with ideal predictive coding in absence of any erasures. Theorem 1 suggests that the rate-recovery function equals \( H(s_1|s_0) \) plus a term that decreases at least as \( H(s)/W + 1 \). A proof of Theorem 1 is provided in section IV. The lower bound in (7) is established by drawing connection to a multi-terminal source coding problem. The upper bound is obtained using a binning based scheme. The following corollary also provides an alternate expression for the achievable rate and makes the connection to the binning technique more explicit.

**Corollary 1.** The upper bound in (6) is equivalent to the following expression

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

The proof of Corollary 1 is provided in Appendix A.

A symmetric source is defined as a Markov source such that the underlying Markov chain is also reversible i.e., the random variables satisfy \( (s_0, \ldots, s_t) \overset{d}{=} (s_t, \ldots, s_0) \), where the equality is in the sense of distribution [20]. Of particular interest to us is the following property satisfied for each \( t \):

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

i.e., we can “exchange” the source pair \((s_{t+1}^a, s_t^a)\) with \((s_{t+1}^b, s_t^b)\) without affecting the joint distribution. An example of a symmetric source is the binary symmetric source: \( s_t^a = s_{t-1}^a \oplus z_t^a \), where \( \{z_t^a\}_{t \geq 0} \) is an i.i.d. binary source process (in both temporal and spatial dimensions) with the marginal distribution \( \Pr(z_{t,j} = 0) = p \), the marginal distribution \( \Pr(s_{t,j} = 0) = \Pr(s_{t,j} = 1) = \frac{1}{2} \) and \( \oplus \) denotes modulo-2 addition. We show that for such sources the binning based scheme is optimal.

**Theorem 2.** For the class of symmetric Markov sources that satisfy (9), the lossless rate-recovery function when restricted to the class of memoryless encoders, 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 Theorem 2 is presented in section V. The converse is obtained by again using a multi-terminal source coding problem, but obtaining a tighter bound by exploiting the memoryless encoders and (9).

**Remark 1.** Even though we consider a single isolated erasure burst in (3), the results in Theorem 1 and 2 immediately apply when the channel introduces multiple bursts with a guard spacing of at least \( W + 1 \). Intuitively this property arises due to the Markov nature of the source. Given a source sequence at time \( i \), all the future source sequences \( \{s_t^n\}_{t \geq i} \) are independent of the past \( \{s_t^n\}_{t < i} \) when conditioned on \( s_i^n \). Thus when a particular source sequence is reconstructed at the destination, it becomes oblivious to past erasures. We note that this is not necessarily true for the lossy case where the Markov property does not apply to the reconstruction sequences.

B. Gauss-Markov Sources

We study the lossy rate recovery function when \( \{s_t^n\} \) is sampled i.i.d. \( \mathcal{N}(0, \sigma_e^2) \) along the spatial dimension and forms a first-order Markov chain across the temporal dimension i.e.,

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

where \( \rho \in (0, 1) \) and \( n_t \sim \mathcal{N}(0, \sigma_e^2(1-\rho^2)) \). Without loss of generality we assume \( \sigma_e^2 = 1 \). We consider the quadratic distortion measure \( d(s_t, \hat{s}_t) = (s_t - \hat{s}_t)^2 \) between the source symbol \( s_t \) and its reconstruction \( \hat{s}_t \). In this paper we focus on the special
case of $W = 0$, where the reconstruction must begin immediately after the erasure burst. We briefly remark about the case when $W > 0$ at the end of section VI. As stated before unlike the lossless case, the results of Gauss-Markov sources for single erasure burst channels do not readily extend to the multiple erasure burst case. This follows from the fact that the Markov property of the source sequences does not continue to hold among the reconstructed sequences at the decoder. Therefore, we treat the two cases separately.

1) Channels with Single Erasure Burst: In this channel model, as stated in (3), we assume that the channel can introduce a single erasure burst of length up to $B$ during the transmission period.

**Proposition 1.** (Lower Bound on Lossy Rate-Recovery for Gauss-Markov Sources) The lossy rate-recovery function of the Gauss-Markov source when $W = 0$ satisfies

$$ R(B, W = 0, D) \geq R_{GM}^-(B, D) \triangleq \frac{1}{2} \log \left( \frac{D\rho^2 + 1 - \rho^{2(B+1)} + \sqrt{\Delta}}{2D} \right) $$

where

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

**Proposition 2.** (Upper Bound on Lossy Rate-Recovery for Gauss-Markov Sources) The lossy rate-recovery function of the Gauss-Markov source when $W = 0$ satisfies

$$ R(B, W = 0, D) \leq R_{GM}^+(B, D) \triangleq I(s_t; u_t | \tilde{s}_{t-B}) $$

$$ = \lim_{t \to \infty} I(s_t; u_t | [u]_{0}^{t-B-1}) $$

where $\tilde{s}_{t-B} = s_{t-B} + e, e \sim N(0, \Sigma(\sigma^2_z)/(1 - \Sigma(\sigma^2_z)))$, and for any $t \geq 0$, $u_t \triangleq s_t + z_t$ and $z_t$ is sampled i.i.d. from $N(0, \sigma^2_z)$. The noise in the test channel, $\sigma^2_z$ satisfies

$$ E[(s_t - \tilde{s}_t)^2] \leq D $$

and $\tilde{s}_t$ denotes the minimum mean-squared estimate (MMSE) of $s_t$ from $\{\tilde{s}_{t-B}, u_t\}$. Also $\Sigma(\sigma^2_z)$ is the error covariance of estimating $s_t$ form $[u]_{0}^{t-1}$ in steady state when $t \to \infty$ as

$$ \Sigma(\sigma^2_z) \triangleq \frac{1}{2} \sqrt{(1 - \sigma^2_z)^2(1 - \rho^2) + 4\sigma^2_z(1 - \rho^2)} \leq \frac{1 - \rho^2}{2}(1 - \sigma^2_z) $$

The proof of Prop. 1 and Prop. 2 are presented in and Section VI-A and Section VI-B respectively. The lower bound in Prop. 1 extends the lower bound in Theorem 1 to the Gaussian case using the Entropy Power Inequality (EPI). The upper bound is based on a natural quantize and binning strategy. However the analysis of the achievable rate is significantly more involved than the lossless case since the reconstruction sequences do not form a Markov chain. The limit in (13) is also explicitly evaluated in Section VI-B through the steady-state analysis of the MMSE estimation error. A comparison of lower and upper bounds is illustrated in Fig. 2 and Fig. 3 as a function of the temporal correlation coefficient $\rho$, and the distortion $D$.  

---

**Fig. 2:** Lower and upper bounds of lossy rate-recovery function vs $\rho$ for $D = 0.2$, $D = 0.3$ and $B = 1$, $B = 2$.

**Fig. 3:** Lower and upper bounds of lossy rate-recovery function vs $D$ for $\rho = 0.9$, $\rho = 0.7$ and $B = 1$, $B = 2$. 

---
In order to study the effect of multiple erasure bursts on lossy rate-recovery function of Gauss-Markov sources, we extend the streaming problem as follows. For the new setup the encoder is defined similar to (2) and we define \( g_i \in \{*, f_i \} \) as output of the channel. The channel can introduce multiple erasure bursts each of length up to \( B \) and a guaranteed guard of length \( L \) separating the consecutive bursts. Upon observing the sequence \( \{g_i\}_{i \geq 0} \), the decoder is required to reconstruct each source sequence with zero delay i.e.,

\[
\hat{s}_i^n = g_i(\hat{s}_0, \ldots, \hat{s}_{i-1}), \quad \text{whenever } g_i \neq *
\]

such that the reconstructed source sequence \( \hat{s}_n \) is required to satisfy the quadratic average distortion constraint as in single erasure case. The destination is not required to produce the source vectors that appear during the erasure burst. Similarly, the rate \( R(L, B, D) \) is feasible if a sequence of encoding and decoding functions exists that satisfy the average distortion constraint. The minimum feasible rate \( R_{\text{GM-ME}}(L, B, D) \), is the lossy rate-recovery function.\(^1\)

**Proposition 3.** The lossy rate-recovery for Gauss-Markov sources over multiple-burst-erasure channel with separating guard bound, \( R_{\text{GM-ME}}(L, B, D) \) satisfies,

\[
R_{\text{GM-ME}}(L, B, D) \leq R^+_{\text{GM-ME}}(L, B, D) \triangleq I(u_i; \hat{s}_0; \hat{s}_{i-L-B-1}, [u]_{i-L-B+1}^{i-B-1})
\]

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

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

and \( \hat{s}_i \) denotes the MMSE of \( s_i \) from \( \{\hat{s}_{i-L-B}; [u]_{i-L-B+1}^{i-B-1}, u_i\} \).

The proof of Prop. 3 presented in Section VI-C is again based on quantize-and-binning technique and involves computing the worst-case erasure pattern by the channel. Note that using similar methods in deriving lower bounds on lossy rate-recovery function for single erasure burst, it is not hard to derive lower bounds on lossy rate-recovery function under the mentioned multiple-erasure channel constraints. However, for sake of simplicity we refrain to present the result in this paper as we do not think it brings novel insights into the problem. The comparison of different classes of achievable rates for various guard length \( L \) is illustrated in Fig. 4. As the guard length \( L \) increases, the achievable rate approaches to the rate for single erasure burst.

3) High Resolution regime: For both the single and multiple erasure burst models, the upper and lower bounds coincide in the high resolution limit as stated below.

**Corollary 2.** In the high resolution limit, the lossy rate-recovery function satisfies the following:

\[
R(B, W = 0, D) = \frac{1}{2} \log \left( \frac{1 - \rho^2(B+1)}{D} \right) + o(D).
\]

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

\(^1\)This model can be generalized for any \( W \), however we only consider the case \( W = 0 \) in this paper (See Remark 4).
C. Gaussian Sources with Sliding Window Recovery Constraints

So far our results for both the discrete sources in section III-A and the Gauss-Markov sources in section III-B involve bounds on the rate-recovery function. In this section we present a setup involving independent Gaussian sources where the lossy rate recovery function is completely characterized.

Consider a sequence of i.i.d. Gaussian sources 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. The encoder’s output at time $i$ is given by $f_i = F_i(s^n_{i-1}, \ldots, s^n_i) \in [1, 2^{nR}]$ as before. At time $i$ the decoder is interested in reproducing a collection of past $K+1$ sources\footnote{In this section it is sufficient to assume that any source sequence with a negative time index is a constant sequence.}:

$$t^n_i = (s^n_i, s^{n-1}_i, \ldots, s^{n-K}_i) \quad (21)$$

within a distortion vector $d = (d_0, d_1, \ldots, d_K)$. Thus for any $i \geq 0$ and $0 \leq j \leq K$, if $\hat{s}^{n-j}_i$ is the reconstruction sequence of $s^{n-j}_i$ at time $i$, we must have that $E[|\hat{s}^{n-j}_i - s^{n-j}_i|^2] \leq nd_j$. As before, the channel can introduce an erasure-burst of length $B$ in an arbitrary interval $[j, j + B - 1]$. The decoder is not required to output a reproduction of the sequences $t^n_i$ for $i \in [j, j + B + W - 1]$. The lossy rate-recovery function denoted by $R(B, W, d)$ is the minimum rate required to satisfy these constraints.

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

$$R(B, W, d) = \frac{1}{2} \log \left( \frac{1}{d_0} \right) + \frac{1}{W + 1} \sum_{k=1}^{\min(K-W, B)} \frac{1}{2} \log \left( \frac{1}{d_{W+k}} \right) \quad (22)$$

The proof of Theorem 3 is provided in Section VII. The coding scheme for the proposed model involves using a successively refinement codebook for each sequence $s^n_i$ 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 VII.

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

IV. GENERAL UPPER AND LOWER BOUNDS ON LOSSLESS RATE-RECOVERY FUNCTION

In this section we present the proof of Theorem 1. Our key contribution in Theorem 1 is the lower bound which is inspired by a connection to a multi-terminal source coding problem introduced in section IV-A. Based on this connection, the proof of the lower bound in Theorem 1 is presented in section IV-B. The upper bound, which is based on the random binning technique, is sketched in section IV-C.

A. Connection to Multi-terminal Source Coding Problem

We present the connection between the streaming setup and a multi-terminal source coding problem. We discuss this connection for the case when $B = 1$ and $W = 1$. At any given time $j$ the encoder output $f_j$ must satisfy two objectives simultaneously. If $j$ is outside the error propagation period then the decoder should use $f_j$ and the past sequences to reconstruct $s^n_j$. If $j$ is within the recovery period then $f_j$ must only help in the recovery of a future source sequence.

Fig. 5 illustrates a multi-terminal source coding problem with one encoder and two decoders that simultaneously captures both these constraints. The encoder is revealed $(s^n_i, s^n_{j+1})$ and produces outputs $f_j$ and $f_{j+1}$. Decoder 1 needs to recover $s^n_j$ given $f_j$ and $s^n_{j+1}$ while decoder 2 needs to recover $s^n_{j+1}$ given $s^n_{j+2}$ and $(f_j, f_{j+1})$. Thus decoder 1 corresponds to the steady state of the system when there is no loss while decoder 2 corresponds to the recovery immediately after an erasure when $B = 1$ and $W = 1$. For the above multi-terminal problem, we establish a lower bound on the symmetric rate $R$ as follows:

$$2nR \geq H(f_j, f_{j+1}) \geq H(f_j, f_{j+1} | s^n_{j-2})$$

$$= H(f_j, f_{j+1} | s^n_{j-1}, s^n_{j-2}) - H(s^n_{j-1} | f_j, f_{j+1}, s^n_{j-2}) \quad (23)$$

$$\geq H(f_j, s^n_{j+1}, s^n_{j-2}) - n\varepsilon_n$$

$$= H(s^n_{j+1} | s^n_{j-2}) + H(f_j | s^n_{j+1}, s^n_{j-2}) - n\varepsilon_n \quad (24)$$

$$\geq H(s^n_{j+1} | s^n_{j-2}) - 2n\varepsilon_n$$

$$= nH(s^n_j | s^n_{j-1}) + nH(s^n_{j+1} | s^n_{j+2}) - 2n\varepsilon_n \quad (25)$$

$$= nH(s^n_j | s^n_{j-1}) + nH(s^n_{j+1} | s^n_{j+2}) - 2n\varepsilon_n \quad (26)$$

$$= nH(s^n_j | s^n_{j-1}) + nH(s^n_{j+1} | s^n_{j+2}) - 2n\varepsilon_n \quad (27)$$

$$= nH(s^n_j | s^n_{j-1}) + nH(s^n_{j+1} | s^n_{j+2}) - 2n\varepsilon_n \quad (28)$$

$$= nH(s^n_j | s^n_{j-1}) + nH(s^n_{j+1} | s^n_{j+2}) - 2n\varepsilon_n \quad (29)$$
Fig. 5: A multi-terminal source coding problem related to the proposed streaming setup. The erasure at time $t = j - 1$ leads to two virtual decoders with different side information as shown.

where (25) follows from the fact that $s_{n+1}^n$ must be recovered from $\{f_j, f_{j+1}, s_{n+2}^n\}$ at decoder 2 hence Fano’s inequality applies and (26) follows from the fact that conditioning reduces entropy. Eq. (27) follows from Fano’s inequality applied to decoder 1 and finally (28) follows from the Markov chain associated with the source process. Dividing throughout by $n$ in (29) and taking $n \to \infty$ yields

$$R \geq \frac{1}{2}H(s_1|s_0) + \frac{1}{2}H(s_3|s_0).$$

(30)

The equivalence between (30) and (7) can be established easily as follows:

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

(31)

$$= H(s_1|s_0) + \frac{1}{2}H(s_3|s_0) - \frac{1}{2}H(s_3|s_1)$$

(32)

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

(33)

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

(34)

where both (33) and (34) follow from the first-order Markov Chain property $s_0 \to s_1 \to s_2 \to s_3$.

**Remark 2.** The setup in Fig. 5 reduces to the source coding problem in [21], [22] when $s_{n+2}^n = \phi$. It is also a successive refinement source coding problem with different side information at the decoders and a special distortion constraints at each of the decoders. However to the best of our knowledge the multi-terminal problem in Fig. 5 has not been addressed in the literature nor has the connection to our proposed streaming setup been made in earlier works.

The above lower bound does not directly apply to the streaming setup. It did not take into account the fact that the decoders have access to all the past encoder outputs and that the encoder has access to all the past source sequences. It turns out that even with this additional knowledge, the lower bound continues to hold. The formal proof provided in following section.

**B. 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

$$R \geq H(s_1|s_0) + \frac{1}{W+1}I(s_p; s_B|s_0) - (W + 1)\varepsilon_n$$

(35)

where throughout we let $p = B + W + 1$.

We consider a periodic erasure channel of period $p$ where the first $B$ packets are erased i.e., for each $k \geq 0$, suppose that an erasure happens in the interval $\{kp, kp+1, \ldots, kp+B-1\}$ and the interval $\{kp+B, \ldots, (k+1)p-1\}$ contains no erasures. Consider the following

$$(W + 1)n(t+1)R = H([f]_{t_B}^{p-1}, [f]_{p+B}^{2p-1}, \ldots, [f]_{(t-1)p+B}^{(t-1)p-1}, [f]_{tp+B}^{(t+1)p-1})$$

(36)

$$\geq H([f]_{t_B}^{p-1}, [f]_{p+B}^{2p-1}, \ldots, [f]_{(t-1)p+B}^{(t-1)p-1}, [f]_{tp+B}^{(t+1)p-1}, s_{n+1}^n)$$

(37)

$$= H([f]_{k_B}^{p-1}, [f]_{kp+B}^{2p-1}, \ldots, [f]_{(k-1)p+B}^{(k-1)p-1}, [f]_{(k+1)p+B}^{k_{n+1}^k})$$

(38)
As we take $n \to \infty$, we recover (35). This completes the proof of the lower bound in Theorem 1.
Upon receiving where (55) follows from the fact that the sequence of variables it immediately follows that (55) is also guaranteed by (53).

\[ s_n \]

This succeeds with high probability if \( R > H(\mathbf{s}_1|\mathbf{s}_0) \), which is guaranteed via (53).

Next suppose that \( s^n_{i-1} \) has not been recovered by the decoder simultaneously attempts to recover \((s^n_{i-B'-W-1}, \ldots, s^n_i)\) given \((s^n_{i-B'-W-1}, f_n-W, \ldots, f_1)\). This succeeds with high probability if \( R > H(\mathbf{s}_1|\mathbf{s}_0) \), which is guaranteed via (53).

This completes the justification of the upper bound.

C. 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 \( f_i \) in (2) is the bin-index of a Slepian-Wolf codebook [23], [24]. Following an erasure burst in \([j, 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

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

is achievable for any arbitrary \( \varepsilon > 0 \).

We use a codebook \( C \), which is generated by randomly partitioning the set of all typical sequences \( T^n_i(s) \) into \( 2^nR^+ \) bins. For each \( i \geq 0 \) the partitioning is done independently and all the partitions are revealed to the decoder ahead of time.

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

First, suppose that the sequence \( s^n_{i-1} \) has already been recovered. Then the destination attempts to recover \( s^n_i \) from \((f_i, s^n_{i-1})\). This succeeds with high probability if \( R > H(\mathbf{s}_1|\mathbf{s}_0) \), which is guaranteed via (53).

Next suppose that \( s^n_{i-1} \) has not been recovered by the destination but \( s^n_i \) needs to be recovered. This only happens when \( s^n_i \) is the first sequence to be recovered after the erasure burst. In particular the erasure burst must happen between \([i-B'-W, i-W-1]\) 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 \((s^n_{i-B'-W-1}, \ldots, s^n_i)\) given \((s^n_{i-B'-W-1}, f_{i-W}, \ldots, f_i)\). This succeeds with high probability if,

\[
(W+1)R = \sum_{j=i-W}^{i} H(f_j) > H(s_{i-W}, \ldots, s_{i-B'-W-1})
\]

where (55) follows from the fact that the sequence of variables \( s_i \) is a stationary, first order Markov process. Whenever \( B' \leq B \) it immediately follows that (55) is also guaranteed by (53).

This completes the justification of the upper bound.

V. Symmetric Sources: Proof of Theorem 2

The special case when \( W = 0 \) follows directly from (7). The achievability also follows from Theorem 1. We thus only need to prove the converse for \( W \geq 1 \). For simplicity in exposition we illustrate the case when \( W = 1 \). Then we need to show that

\[
R(B, W = 1) \geq \frac{1}{2} H(s_{B+1}, s_{B+2}|s_i)
\]  

The proof for general \( W > 1 \) will follow along similar lines and will be sketched thereafter.

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

\[
s_{j+1}^n = G_{j+1}\left([f]_{j-B+1}^{j-B-1}, f_j, f_{j+1}, s^i_{j-1}\right).
\]
Lemma 1. Suppose that the encoding function \( f_j = F_j(s^n_j) \) is memoryless. Suppose that there exist decoding functions \( \hat{s}^n_j = G_j([f]_0^n, s^n_{j-1}) \) and \( \hat{s}^n_{j+1} = G_{j+1}([f]_0^{j+1}, f_j, f_{j+1}, s^n_{j-1}) \) such that \( \Pr(\hat{s}^n_j \neq s^n_j) \) and \( \Pr(\hat{s}^n_{j+1} \neq s^n_{j+1}) \) both vanish to zero as \( n \to \infty \). Then

\[
\begin{align*}
H(s_j^n | s_{j-1}^n, f_j) &\leq n \varepsilon_n, \\
H(s_{j+1}^n | s_{j-B-1}^n, f_j, f_{j+1}) &\leq n \varepsilon_n
\end{align*}
\]

also hold.

Proof: To establish (65) we note that for the memoryless encoder the following Markov chain holds:

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

Hence we have that

\[
n \varepsilon_n \geq H(s_j^n | [f]_0^{j-1}, s_{j-1}^n) \geq H(s_j^n | [f]_0^{j-1}, s_{j-1}^n, f_j) = H(s_j^n | s_{j-1}^n, f_j),
\]

where the last step follows via (67). Similarly using \( (s_{j-1}^n, [f]_0^{j-B-1}) \rightarrow s_{j-B-1}^n \rightarrow (f_j, s_j^n, f_{j+1}) \), we have

\[
n \varepsilon_n \geq H(s_{j+1}^n | [f]_0^{j-B-1}, f_j, f_{j+1}, s_{j-1}^n) \geq H(s_{j+1}^n | [f]_0^{j-B-1}, s_{j-B-1}^n, f_j, f_{j+1}, s_{j-1}^n) = H(s_{j+1}^n | s_{j-B-1}^n, f_j, f_{j+1}).
\]

The conditions in (65) and (66) show that any rate that is achievable in the original problem is also achieved in the multi-terminal source network. Hence a lower bound to the 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. 6(a).
B. Lower Bound for Multi-terminal Source Coding Problem

Consider the source coding problem with side information illustrated in Fig. 6(a). In this setup there are four source sequences drawn i.i.d. from a joint distribution \( p(s_j, s_{j-1}, s_{j-B-1}) \). The two encoders \( j \) and \( j+1 \) are revealed source sequences \( s^n_j \) and \( s_{j+1}^n \) and the two decoders \( j \) and \( j+1 \) are revealed sources \( s_{j-1}^n \) and \( s_{j-B-1}^n \). 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 \((f_j, s_{j-1}^n)\) while decoder \( j+1 \) has access to \((f_{j+1}, s_{j-B-1}^n)\) and are required to reconstruct

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

(72)

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

Lemma 2. The set of all achievable rate-pairs \((R_j, R_{j+1})\) for the problem in Fig. 6(a) is identical to the set of all achievable rate-pairs for the problem in Fig. 6(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. 6(a) depends on the joint distribution \( p(s_j, s_{j-1}, s_{j-B-1}) \) only via the marginal distributions \( p(s_j, s_{j-1}) \) and \( p(s_j, s_{j-B-1}) \). 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 \( p(s_j, s_{j+B-1}) \) are identical and thus exchanging \( s_{j-1}^n \) with \( s_{j+B-1}^n \) 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. 6(b). First note that

\[
nR_{j+1} = H(f_{j+1})
\]

(73)

where (73) follows by applying Fano’s inequality since \( s_{j+1}^n \) can be recovered from \((s_{j-B-1}^n, f_j, f_{j+1})\). To bound \( R_j \)

\[
nR_j = H(f_j) \geq I(f_j; s_{j-B-1}^n | s_{j-B-1}^n) \geq H(s_{j-B-1}^n | s_{j-B-1}^n, f_j) - H(s_{j-B-1}^n, s_{j+B-1}^n, f_j) = nH(f_j) - nH(s_{j-B-1}^n, f_j) \geq nH(s_{j-B-1}^n, f_j) - nH(s_{j-B-1}^n, s_{j+B-1}^n, f_j) \geq nH(s_{j-B-1}^n, f_j) - nH(s_{j-B-1}^n, s_{j+B-1}^n, f_j) - n\varepsilon_n
\]

(74)

(75)

(76)

(77)

(78)

Remark 3. One way to interpret the lower bound in (64) is by observing that the decoder \( j+1 \) in Fig. 6(b) is able to recover not only \( s_{j+B-1}^n \) but also \( s_j^n \). In particular, the decoder \( j+1 \) first recovers \( s_{j+B-1}^n \). Then, similar to decoder \( j \), it also recovers \( s_j^n \) from \( f_j \) and \( s_{j+B-1}^n \) as side information. Hence, by only considering decoder \( j+1 \) when the two encoders are allowed to cooperate and following standard source coding argument, the lower bound on the sum-rate satisfies:

\[
R_j + R_{j+1} \geq H(s_j, s_{j+1}, | s_{j-B-1}^n).
\]

C. 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 \( f_k = F_k(s_k^n) \) for \( k \in \{j, \ldots, j+W\} \). Suppose that there exist decoding functions \( \hat{s}_k^n = \hat{G}_k(f_k^n, s_{k-1}^n) \) for \( k \in \{j, \ldots, j+W-1\} \) and \( \hat{s}_{j+W} = \hat{G}_{j+W}(f_{j+W}^n, s_j^n) \) such that \( \Pr(s_k^n \neq \hat{s}_k^n) \) vanishes to zero for all \( k \in \{j, \ldots, j+W-1\} \) as \( n \to \infty \). Then

\[
H(s_k^n | s_{k-1}^n, f_k) \leq n\varepsilon_n \quad k \in \{j, \ldots, j+W-1\}
\]

(77)

\[
H(s_{j+W}^n | s_{j-B-1}^n, f_{j+W}) \leq n\varepsilon_n
\]

(78)

Proof: The proof is an immediate extension of Lemma 1 and is included for completeness in Appendix B.

\[\square\]
The above 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 $y^n_t$ and compresses it into an index $f_t \in [1, 2^n R_t]$. The corresponding decoders are revealed $y^n_{t-1}$ for $t \in \{j, j+1\}$ and the decoder $j+2$ is revealed $y^n_{j+2}$ in Section VI-C. The proof of Corollary 2, is presented in section VI-A whereas the proof of the upper bound in Prop. 2 is presented in section VI-B. The proof of Prop. 3 using Lemma 3 for $W$ for each $i \geq 0$ we have

$$R = R_j + R_{j+1} + R_{j+2} \geq H(s_j, s_{j+1}, s_{j+2}|s_{j-B-1}).$$

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 any arbitrary $W$ is completely analogous.

VI. LOSSY RATE-RECOVERY FOR GAUSS-MARKOV SOURCES

We establish lower and upper bounds on the rate-recovery function of Gauss-Markov sources when an immediate recovery following the erasure burst is required i.e., $W = 0$. For the single burst erasure case, the proof of the lower bound in Prop. 1 is presented in section VI-A whereas the proof of the upper bound in Prop. 2 is presented in section VI-B. Th proof of Prop. 3 for the multiple burst erasure case with guard separating the bursts is presented in Section VI-C. The proof of Corollary 2, which establishes the rate-recovery function in the high resolution regime is presented in section VI-D.

A. Lower Bound: Single Burst Erasure

Consider any rate $R$ code that satisfies an average distortion of $D$ as stated in (5). For each $i \geq 0$ we have

$$nR \geq H(f_i) \geq H(f_i | f_{10}^{i-B-1}, s_{-1}^n)$$

where (80) follows from the fact that conditioning reduces the entropy.

We now present an upper bound for the second term and a lower bound for the first term. We first establish an upper bound for the second term in (82). Suppose that the erasure burst occurs in the interval the interval $[i-B, i-1]$. The reconstruction sequence $\hat{s}_i^n$ must be a function of $(f_i, f_{10}^{i-B-1}, s_{-1}^n)$. Thus we have

$$h(s_i^n | f_{10}^{i-B-1}, f_i, s_{-1}^n) = h(s_i^n - \hat{s}_i^n | f_{10}^{i-B-1}, f_i, s_{-1}^n)$$

where the last step uses the fact that the expected average distortion between $s^n$ and $\hat{s}_i^n$ is no greater than $D$, and standard arguments [25, Ch. 13].

To lower bound the first term in (82), we successively use the Gauss-Markov relation (11) 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 [25] we have

$$h(s_i^n | f_{10}^{i-B-1}, s_{-1}^n) \geq \frac{n}{2} \log \left( \rho^{2(B+1)} \frac{2^n h(s_{i-B-1}^n | f_{10}^{i-B-1}, s_{-1}^n) + 2\pi e (1 - \rho^{2(B+1)})} \right).$$

It remains to lower bound the entropy term in the right hand side of (85). We show the following in Appendix C.

**Lemma 4.** For any $k \geq 0$

$$2^{\frac{1}{2} h(s_i^n | f_{10}^{i-B-1}, s_{-1}^n)} \geq \frac{2\pi e (1 - \rho^2)}{2^R - \rho^2} \left( 1 - \left( \frac{\rho^2}{2^R} \right)^k \right).$$

Upon substituting, (86), (85), and (83) into (82) we obtain that for each $i \geq B + 1$
Lemma 5. The functions \( \lambda_t(k, B) \) and \( \gamma_t(k, B) \) satisfy the following properties:

1. For all \( t \geq B' \) and \( k \in [0, t - B') \), \( \lambda_t(k, B') \leq \lambda_t(0, B') \) and \( \gamma_t(k, B') \leq \gamma_t(0, B') \), i.e. the worse-case erasure pattern at time \( t \) includes the most recent erasure burst.
The above inequalities state that the conditional differential entropy of $B_i$.

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 of the system.

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

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

$$h(u_t|\{u_0^{t-B'-k-1},u_t^{t-1}_{t-k},s_{-1}\}) \leq h(u_t|\{u_0^{t-B'-k},u_t^{t-1}_{t-k+1},s_{-1}\}), \quad (98)$$

$$h(s_t|\{u_0^{t-B'-k-1},u_t^{t-1}_{t-k},s_{-1}\}) \leq h(s_t|\{u_0^{t-B'-k},u_t^{t-1}_{t-k+1},s_{-1}\}). \quad (99)$$

The above inequalities state that the conditional differential entropy of $u_t$ and $s_t$ is reduced if the variable $u_{t-B'-k}$ is replaced by $u_{t-k}$ in the conditioning and the remaining variables remain unchanged. Fig. 7 provides a schematic interpretation of the above inequalities.

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 $k \in [1 : t - B']$.

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

$$= h(u_t|\{u_0^{t-B'-k},u_t^{t-1}_{t-k+1},s_{-1}\}) - h(u_t|s_t) \leq h(u_t|\{u_0^{t-B'-k},u_t^{t-1}_{t-k+1},s_{-1}\}) - h(u_t|s_t)$$

$$= I(s_t; u_t^{t-B'-k}, [u_t^{t-1}_{t-k+1}], s_{-1}) - h(u_t|s_t) = \lambda_t(k - 1, B') \quad (100)$$

where (100) follows from using (98). In a similar fashion since

$$\gamma_t(k, B') = E[(s_t - \hat{s}_t^{(k)}|\{u_0^{t-B'-k},u_t^{t-1}_{t-k+1},s_{-1}\})^2]$$

is the MMSE estimation error of $s_t$ given $\{u_0^{t-B'-k},u_t^{t-1}_{t-k+1},s_{-1}\}$ we have

$$\frac{1}{2} \log (2\pi e \cdot \gamma_t(k, B')) = h(s_t|\{u_0^{t-B'-k},u_t^{t-1}_{t-k},s_{-1}\}) \leq h(s_t|\{u_0^{t-B'-k},u_t^{t-1}_{t-k+1},s_{-1}\})$$

$$= \frac{1}{2} \log (2\pi e \cdot \gamma_t(k - 1, B')) \quad (102)$$

where (102) follows from using (99). 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 (101) and (103) 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 simply 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 follows. For any $B' \leq B$ we have,

$$\lambda_t(0, B') = I(s_t; u_t^{t-B'-1}, s_{-1})$$

$$= h(u_t^{t-B'-1}, s_{-1}) - h(u_t|s_t)$$

$$= h(u_t^{t-B'-1}, s_{-1}) - h(u_t|s_t) \quad (104)$$

$$= \lambda_t(0, B) \quad (107)$$
where (104) and (106) follows from the Markov chain property
\[ u_t \to s_t \to \{[u_0]_{0}^{t-j-1}, s_{-1}\}, \quad j \in \{B, B'\} \]  \tag{108}
and (105) follows from the fact that conditioning reduces 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
\[ \lambda_{t+1}(0, B) = I(s_{t+1}; u_{t+1} | [u_0]_{0}^{t-B}, s_{-1}) \]
\[ = h(u_{t+1} | [u_0]_{0}^{t-B}, s_{-1}) - h(u_{t+1} | s_{t+1}) \]
\[ = h(u_{t+1} | [u_0]_{0}^{t-B}, s_{-1}) - h(u|s_t) \]
\[ \geq h(u_{t+1} | [u_0]_{0}^{t-B}, s_{-1}, s_0) - h(u|s_t) \]
\[ = h(u_{t+1} | [u_0]_{0}^{t-B}, s_0) - h(u|s_t) \]
\[ = h(u | [u_0]_{0}^{t-B-1}, s_{-1}) - h(u|s_t) \]
\[ = I(s_t; u_t | [u_0]_{0}^{t-B-1}, s_{-1}) \]
\[ = \lambda_t(0, B) \] \tag{113}
where (109) and (112) follow from time-invariant property of the source model and the test channel, (110) follows from the fact that conditioning reduces differential entropy and (111) uses the following Markov chain property.
\[ \{u_0, s_{-1}\} \to \{[u_0]_{0}^{t-B}, s_0\} \to u_{t+1}. \] \tag{114}
Similarly,
\[ \frac{1}{2} \log (2\pi e \cdot \gamma_{t+1}(0, B)) = h(s_{t+1} | [u_0]_{0}^{t-B}, u_{t+1}, s_{-1}) \]
\[ \geq h(s_{t+1} | [u_0]_{0}^{t-B}, u_{t+1}, s_0) \] \tag{115}
\[ = h(s_t | [u_0]_{0}^{t-B-1}, u_{t}, s_{-1}) \] \tag{116}
\[ = \frac{1}{2} \log (2\pi e \cdot \gamma_t(0, B)) \] \tag{117}
where (115) follows from the fact that conditioning reduces differential entropy and the following Markov chain property
\[ \{u_0, s_{-1}\} \to \{[u_0]_{0}^{t-B}, u_{t+1}, s_0\} \to s_{t+1}. \] \tag{118}
Since (113) and (117) 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
\[ \lambda_t(k, B') \leq \lambda_t(0, B') \] \tag{119}
\[ \leq \lambda_t(0, t) \]
\[ = h(u_t | s_{t-1}) - h(u_t | s_t) \]
\[ \leq h(u_B | s_{t-1}) - h(u_B | s_B) \] \tag{121}
\[ = \lambda_B(0, B) \] \tag{122}
where (119) follows from part 1 of the lemma, (120) is based on the fact that the worst-case erasure pattern contains most possible erasures and follows from the similar steps used in deriving (107) and using the fact that if \( t < B \), the erasure burst length is at most \( t \). Eq. (121) follows from the fact that whenever \( t < B \) the relation \( s_{-1} \to u_t \to u_B \) holds. In a similar fashion we can show that \( \gamma_t(k, B) \leq \gamma_B(0, B) \).

This completes the proof of lemma 5. ■

Following the four parts of Lemma 5, it can be seen that the worst-case erasure pattern happens at steady state i.e. \( t \to \infty \) when there is a burst of length \( B \) spans \([t-B, t-1]\). This implies that any pair \((R, D)\) is achievable if
\[ R \geq \lim_{t\to\infty} \lambda_t(0, B) \] \tag{123}
\[ D \geq \lim_{t\to\infty} \gamma_t(0, B) \] \tag{124}
To complete the achievability we need to show that any test channel satisfying (93) also implies (124) and any rate satisfying (94) implies (123). These relations can be established in a straightforward manner as shown below.

\[ R = \Lambda(B, \sigma_z^2) = \lim_{t \to \infty} I(s_t; u_t | [u]_{0}^{t-B-1}) \]
\[ = \lim_{t \to \infty} h(u_t | [u]_{0}^{t-B-1}) - h(u_t | s_t) \]
\[ \geq \lim_{t \to \infty} h(u_t | [u]_{0}^{t-B-1}, s_{t-1}) - h(u_t | s_t) \]
\[ = \lim_{t \to \infty} \Lambda_t(0, B) \]
(125)

and

\[ D \geq \Gamma(B, \sigma_z^2) = \lim_{t \to \infty} E \left[ \left( s_t - \hat{s}_t([u]_{0}^{t-B-1}, u_t) \right)^2 \right] \]
\[ \geq \lim_{t \to \infty} E \left[ \left( s_t - \hat{s}_t([u]_{0}^{t-B-1}, u_t, s_{t-1}) \right)^2 \right] \]
\[ = \lim_{t \to \infty} \gamma_t(0, B) \]
(128)

An Explicit Rate Expression: We derive the expression for \( R_{GM}^+(B, D) = \Lambda(B, \sigma_z^2) \). To this end it is helpful to consider the following single-variable discrete-time Kalman-Bucy filter for \( i \in [0, t - B - 1] \),

\[ s_i = \rho s_{i-1} + n_i, \quad n_i \sim N(0, 1 - \rho^2) \]
\[ u_i = s_t + z_t, \quad z_t \sim N(0, \sigma_z^2). \]
(131)

Note that \( s_t \) can be viewed as the state of the system updated according a Gauss-Markov model and \( u_t \) as the output of the system at each time \( t \), which is a noisy version of the state \( s_t \). 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., [27, Example V.B.2]):

\[ \Sigma(\sigma_z^2) \triangleq \lim_{t \to \infty} \sigma_{t-B}^2([u]_{0}^{t-B-1}) \]
\[ = \frac{1}{2} \sqrt{(1 - \sigma_z^2)^2(1 - \rho^2)^2 + 4\sigma_z^2(1 - \rho^2)} + \frac{1}{2} - \rho^2(1 - \sigma_z^2) \]
(133)

Also using the orthogonality principle for MMSE estimation we have

\[ [u]_{0}^{t-B-1} \to \hat{s}_{t-B}([u]_{0}^{t-B-1}) \to \hat{s}_{t-B} \to s_t \]
(135)

Thus we can express

\[ s_{t-B} = \hat{s}_{t-B}([u]_{0}^{t-B-1}) + \tilde{n} \]
(136)

where the noise \( e \sim N(0, \Sigma(\sigma_z^2)) \) is independent of the observation set \( [u]_{0}^{t-B-1} \). Furthermore since

\[ s_t = \rho^B s_{t-B} + \tilde{n} \]
where \( \tilde{n} \sim N(0, 1 - \rho^{2B}) \),

\[ \Gamma(B, \sigma_z^2) = \lim_{t \to \infty} \sigma_t^2([u]_{0}^{t-B-1}, u_t) \]
\[ = \lim_{t \to \infty} \sigma_t^2(\hat{s}_{t-B}([u]_{0}^{t-B-1}, u_t)) \]
\[ = \left[ \frac{1}{\sigma_z^2} + \frac{1 - \rho^{2B}}{1 - \Sigma(\sigma_z^2)} \right]^{-1} \]
(139)

where (139) follows from the Markov property in (135) and (140) follows from the application of MMSE estimator and using (137), (136) and the definition of the test channel in (90). Thus the noise \( \sigma_z^2 \) in the test channel (90) is obtained by setting

\[ \Gamma(B, \sigma_z^2) = D \]
(141)

Furthermore to compute the rate expression

\[ \Lambda(B, \sigma_z^2) = \lim_{t \to \infty} I(s_t; u_t | [u]_{0}^{t-B-1}) \]
\[ = \lim_{t \to \infty} I(s_t; \hat{s}_{t-B}([u]_{0}^{t-B-1})) \]
\[ = \lim_{t \to \infty} h(\hat{s}_t | \hat{s}_{t-B}([u]_{0}^{t-B-1})) - h(\hat{s}_t | \hat{s}_{t-B}([u]_{0}^{t-B-1}, u_t)) \]
\[ = \frac{1}{2} \log(2\pi e(1 - \rho^{2B}(1 - \Sigma(\sigma_z^2)))) - \frac{1}{2} \log(2\pi e\Gamma(B, \sigma_z^2)) \]
\[ = \frac{1}{2} \log \left( \frac{1 - \rho^{2B}(1 - \Sigma(\sigma_z^2))}{D} \right) \]
As the rate and the test channel have to satisfy (151) and (152) for all possible \( W \) with the test channel satisfying \( \lambda_t(\Omega_t) \) and \( \gamma_t(\Omega_t) \) among all possible erasure sets \( \Omega_t \) with total number of erasures \( n_e \). See Fig. 8a for an example of \( \Omega^*_t(n_e) \).

Remark 4. The analysis of the achievable rate for the Gauss-Markov sources is a non-trivial extension of the lossless setup even when \( W = 0 \). In extending this analysis to the case when \( W > 0 \) the generalization of Lemma 5 appears to involve a rate-region corresponding to the Berger-Tung inner bound [28]. This is a significant generalization and could not be treated in this paper. In addition to this, it appears that hybrid schemes may be required beyond the binning scheme treated in this paper. Thus the scope of this problem is well beyond the results in this paper.

C. Coding Scheme: Multiple Burst Erasures with Guard Bounds

As stated in Section III-B2, in order to study the effect of multiple erasure bursts during the communication period, we consider a scenario where the channel can introduce multiple erasure bursts each of length up to \( B \) such that there is a guaranteed guard of length at least \( L \) separating the bursts. In this section we investigate the quantization-and-binning based coding scheme of single burst erasure model for this new problem. Before going through the coding scheme, define the following notations.

\[
\begin{align*}
s_{\Omega} & = \{ s_i : i \in \Omega \} \\
\mathbf{u}_{\Omega} & = \{ u_i : i \in \Omega \}
\end{align*}
\]

The following lemma also is very useful is establishing the results of this section.

Lemma 6. Consider the two sets \( A, B \subseteq \mathbb{N} \) of the same 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 \( i \in \{ 1, \ldots, r \} \), \( a_i \leq b_i \). Also consider the Gauss-Markov source \( s^n_i \) and the test channel in (90). Then for any \( t \geq b_i \)

\[
\begin{align*}
& h(s_i | u_A, s_{-1}) \geq h(s_i | u_B, s_{-1}) \\
& h(u_i | u_A, s_{-1}) \geq h(u_i | u_B, s_{-1}).
\end{align*}
\]

The proof of Lemma 6 is available in Appendix E.

Now consider the system at time \( t \) that the channel packet of time \( t \) is not erased. Define the set \( \Omega_t \) denoting the set of time indices up to time \( t - 1 \) for which the channel packets are not erased by the channel and is available in the decoder, i.e.

\[
\Omega_t = \{ i : 0 \leq i \leq t - 1, s_i \neq \star \}.
\]

The decoder succeeds in reconstructing the source sequence \( s^n_i \) if

\[
R \geq \lambda_t(\Omega_t) \triangleq I(s_i; u_t | u_{\Omega_t}, s_{-1})
\]

with the test channel satisfying

\[
\gamma_t(\Omega_t) \triangleq E \left[ (s_i - \hat{s}_i(u_{\Omega_t}, s_{-1}))^2 \right] \leq D
\]

As the rate and the test channel have to satisfy (151) and (152) for all possible \( \Omega_t \) and \( t \geq 0 \), we are interested in the pair \( (\Omega_t, t) \) representing the worst-case erasure pattern. The following lemma characterizes the worst erasure pattern.

Lemma 7. Assume the system at time \( t \), then

1) Among all possible erasure patterns with the total number of erasures \( n_e \), the worst adversary channel erases the closest possible \( n_e \) packets to time \( t \). The set of non-erased indices in such an erasure pattern denoted by \( \Omega^*_t(n_e) \) maximizes \( \lambda_t(\Omega_t) \) and \( \gamma_t(\Omega_t) \) among all possible erasure sets \( \Omega_t \) with total number of erasures \( n_e \). See Fig. 8a for an example of \( \Omega^*_t(n_e) \).

2) The worst-case erasure pattern includes maximum number of erasures. The set of non-erased indices with maximum number of erasures denoted by \( \Omega^*_t \) maximizes \( \lambda_t(\Omega^*_t(n_e)) \) and \( \gamma_t(\Omega^*_t(n_e)) \). See Fig. 8b for an example of \( \Omega^*_t \).
3) The worst-case erasure pattern happens in steady state when $t \to \infty$, i.e. $\lambda_t(\Omega^*_t)$ and $\gamma_t(\Omega^*_t)$ are increasing functions with respect to $t$.

Proof:

1) For any set of indices $\Omega_t$ with size $t - n_e + 1$ which indicates the total number of $n_e$ erasures over the communication duration, we have

$$\lambda_t(\Omega_t) = I(s_t; u_t | u_{\Omega_t}, s_{-1})$$

$$= h(u_t | u_{\Omega_t}, s_{-1}) - h(u_t | s_t)$$

$$\leq h(u_t | u_{\Omega_t}((n_e), s_{-1}) - h(u_t | s_t)$$

$$= I(s_t; u_t | u_{\Omega_t}((n_e), s_{-1})$$

$$= \lambda_t(\Omega^*_t(n_e))$$

where (153) follows from the application of Lemma 6. Also note that

$$\frac{1}{2} \log (2\pi e \gamma_t(\Omega_t)) = h(s_t | u_t, u_{\Omega_t}, s_{-1})$$

$$\leq h(s_t | u_t, u_{\Omega_t}((n_e), s_{-1})$$

$$= \frac{1}{2} \log (2\pi e \gamma_t(\Omega^*_t(n_e)))$$

where (156) follows from the application of second result of Lemma 6 for the sets $A = \{\Omega^*_t(n_e), t\}$ and $B = \{\Omega_t, t\}$ which clearly satisfy the required property in Lemma 6. The monotonicity property of the function $f(x) = 1/2 \log(2\pi ex)$, completes the proof of the first part of the lemma.

2) We now show that the worst-case erasure pattern includes the maximum possible erasures. This is based on the fact that the decoder can always ignore the available channel packets and therefore if the decoder is successful in all possible erasure condition it will succeed in other erasure patterns. This concludes the proof of the second part of the lemma.

3) This property follows from the fact that in steady state the effect of knowing $s_{-1}$ vanishes. The formal proof of this part is similar to the proof of the third part of Lemma 5 as follows

$$\lambda_{t+1}(\Omega^*_t) = I(s_{t+1}; u_{t+1} | u_{\Omega^*_t+1}, s_{-1})$$

$$= h(u_{t+1} | u_{\Omega^*_t+1}, s_{-1}) - h(u_{t+1} | s_{t+1})$$

$$\geq h(u_{t+1} | u_{\Omega^*_t+1}, s_{-1}, s_t) - h(u_{t+1} | s_t)$$

$$= h(u_{t+1} | u_{\Omega^*_t+1}((0)), s_t) - h(u_{t+1} | s_t)$$

$$= h(u_t | u_{\Omega_t}, s_{-1}) - h(u_t | s_t)$$

$$= I(s_t; u_t | u_{\Omega_t}, s_{-1})$$

$$= \lambda_t(\Omega^*_t)$$

where (158) follows from the fact that conditioning reduces the differential entropy. Also in (159) the notation $\Omega^*_t+1 \setminus \{0\}$ indicates the set $\Omega^*_t+1$ when the index 0 is excluded if 0 $\in \Omega^*_t+1$. Then (159) follows from the following Markov property

$$\{u_0, s_{-1}\} \to \{u_{\Omega^*_t+1 \setminus \{0\}}, s_0\} \to u_{t+1}$$

Also (160) follows from the time-invariant property of source model and the test channel. Also note that

$$\frac{1}{2} \log (2\pi e \gamma_{t+1}(\Omega^*_t)) = h(s_{t+1} | u_{t+1}, u_{\Omega^*_t+1}, s_{-1})$$

$$\geq h(s_{t+1} | u_{t+1}, u_{\Omega^*_t+1}, s_{-1}, s_0)$$

$$= h(s_{t+1} | u_{t+1}, u_{\Omega^*_t+1 \setminus \{0\}}, s_0)$$

$$= h(s_t | u_t, u_{\Omega_t}, s_{-1})$$

$$= \frac{1}{2} \log (2\pi e \gamma_t(\Omega^*_t))$$

where (164) follows from the fact that conditioning reduces the differential entropy, (165) follows from the following Markov property

$$\{u_0, s_{-1}\} \to \{u_{\Omega^*_t+1 \setminus \{0\}}, u_{t+1}, s_0\} \to s_{t+1}$$

and (166) again follows from the time-invariant property of source model and the test channel.
Following the three steps in Lemma 7 implies that any pair \((R, D)\) is achievable if

\[
R \geq \lim_{t \to \infty} \lambda_t(\Omega_t^*) \tag{169}
\]

\[
D \geq \lim_{t \to \infty} \gamma_t(\Omega_t^*) \tag{170}
\]

Exploiting the periodic structure of the worst-case erasure pattern and the fact that this happens in steady state when \(t \to \infty\), we are able to compute the lossy rate-recovery function as stated in Prop. 3. First we compute the limit in (169) as follows.

\[
\lim_{t \to \infty} \lambda_t(\Omega_t^*) = \lim_{t \to \infty} I(s_t; u_t|\Omega_t^*, s_{-1})
\]

\[
= \lim_{t \to \infty} I(s_t; u_t|\Omega_{t-L-B}^*, [u]_{t-L-B+1}^{t-B-1}, s_{-1})
\]

\[
= \lim_{t \to \infty} I(s_t; u_t|\hat{s}_t-L-B(\Omega_{t-L-B}^*, s_{-1}), [u]_{t-L-B+1}^{t-B-1})
\]

\[
\leq I(s_t; u_t|\hat{s}_t-L-B + \hat{e}, [u]_{t-L-B+1}^{t-B-1})
\]

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

(176)

where in (173) \(\hat{a} = 1 - D\) and \(\hat{e} \sim N(0, D(1 - D))\). This follows from the fact that the estimate \(\hat{s}_t-L-B(\Omega_{t-L-B}^*, s_{-1})\) satisfies the distortion constraint and is within the average distortion at most \(D\). Thus the MMSE estimate \(\hat{s}_t-L-B(\Omega_{t-L-B}^*, s_{-1})\) in (172) can be replaced with \(\hat{s}_t-L-B\) in (175), where we define a virtual channel \(\hat{s}_t-L-B = s_t-L-B + e\) where \(e \sim N(0, D/(1 - D))\). Also (170) can be computed as

\[
\lim_{t \to \infty} \gamma_t(\Omega_t^*) = \lim_{t \to \infty} \sigma_t^2(u_t|\Omega_t^*, u_t, s_{-1})
\]

\[
= \lim_{t \to \infty} \sigma_t^2(u_{t-L-B}^*, [u]_{t-L-B+1}^{t-B-1}, u_t, s_{-1})
\]

\[
= \lim_{t \to \infty} \sigma_t^2(\hat{s}_t-L-B(u_{t-L-B}^*, s_{-1}), [u]_{t-L-B+1}^{t-B-1})
\]

\[
\leq \lim_{t \to \infty} \sigma_t^2(\hat{s}_t-L-B + \hat{e}, [u]_{t-L-B+1}^{t-B-1}, u_t)
\]

\[
= \lim_{t \to \infty} \sigma_t^2(s_t-L-B + e, [u]_{t-L-B+1}^{t-B-1}, u_t)
\]

\[
= \sigma_t^2(\hat{s}_t-L-B, [u]_{t-L-B+1}^{t-B-1}, u_t)
\]

(181)

where (179) follows from the following Markov property.

\[
\{u_{t-L-B}^*, s_{-1}\} \rightarrow \{\hat{s}_t-L-B(u_{t-L-B}^*, s_{-1}), [u]_{t-L-B+1}^{t-B-1}\} \rightarrow s_t
\]

(183)

and (180) again follows from the fact that the estimate \(\hat{s}_t-L-B(u_{t-L-B}^*, s_{-1})\) satisfies the distortion constraint. All the constants and variables in (180) and (181) are as defined before.

According to (176) and (182) if we choose the noise in the test channel \(\sigma_t^2\) to satisfy

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

(184)

then the test channel and the rate \(R_{\text{GM-ME}}^+(L, B, D)\) defined in (176) both satisfy rate and distortion constraints in (169) and (170) and therefore \(R_{\text{GM-ME}}^+(L, B, D)\) is achievable. This completes the proof of Prop 3.

**An Explicit Rate Expression:** We derive the expression for \(R_{\text{GM-ME}}^+(L, B, D)\) in (176). 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.

\[
\eta(\sigma_t^2) \triangleq \sigma_t^2(\hat{s}_t-L-B, [u]_{t-L-B+1}^{t-B-1}) = E[\hat{s}_t^2-L-B-1] - E[s_{t-B-1}U]E[U^TU]\]

\[
= 1 - A_1(A_2)^{-1}A_1^T
\]

(185)

(186)

where we define the notation \(U \triangleq (\hat{s}_t-L-B, u_{t-L-B+1}, u_{t-L-B+2}, \ldots, u_{t-B-1})\) and \((\cdot)^T\) denotes the transpose operation. Also note that \(A_1\) and \(A_2\) can be computed as follows.

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

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

(187)

(188)
According to (182) we can write
\[
D = \sigma^2_t(\hat{s}_L-L-B, [u]_{L-B+1}^{L-B-1}, u_t) = \sigma^2_t(\hat{s}_L-L-B, [u]_{L-B+1}^{L-B-1}, u_t)
\]
\[
= \left[ \frac{1}{\sigma^2_t} + \frac{1}{1 - \rho^{2(B+1)}(1 - \eta(\sigma^2_t))} \right]^{-1}
\]
(189)
And
\[
R_{GM-ME}^+(L, B, D) = I(s_t; u_t|\hat{s}_L-L-B, [u]_{L-B+1}^{L-B-1})
\]
\[
= h(s_t|\hat{s}_L-L-B, [u]_{L-B+1}^{L-B-1}) - h(s_t|\hat{s}_L-L-B, [u]_{L-B+1}^{L-B-1}, u_t)
\]
\[
= h(s_t|\hat{s}_L-L-B, [u]_{L-B+1}^{L-B-1}) - \frac{1}{2} \log(2\pi e D)
\]
(190)
\[
= \frac{1}{2} \log \left( \frac{1 - \rho^{2(B+1)}(1 - \eta(\sigma^2_t))}{D} \right)
\]
(191)
\[
= \frac{1}{2} \log \left( \frac{1 - \rho^{2(B+1)}(1 - \eta(\sigma^2_t))}{D} \right)
\]
(192)
Therefore by solving (189) for \(\sigma^2_t\) and then replacing it into (192), \(R_{GM-ME}^+(L, B, D)\) can be computed.

D. High Resolution Regime

In this section we investigate the behaviour of the lossy rate-recovery functions for Gauss-Markov sources for single and multiple burst-erasure channel models, i.e. \(R_{GM}(B, D)\) and \(R_{GM-ME}(L, B, D)\), in high resolution regime. As stated in Corollary 2 our upper and lower bounds coincide in this regime. In order to establish this result, first note the following inequalities.
\[
R_{GM}^-(B, D) \leq R_{GM}(B, D) \leq R_{GM-ME}(L, B, D) \leq R_{GM-ME}^+(L, B, D)
\]
(193)
The first and the last inequalities in (193) is by definition and the second inequality follows from the fact that the rate achievable for multiple erasure model is also achievable for single erasure bursts as the decoder can simply ignore the available codewords in reconstructing the source sequences. According to (193), it suffices to characterize the high resolution behaviour of \(R_{GM}(B, D)\) and \(R_{GM-ME}(L, B, D)\) by specializing Prop. 1 and Prop. 3 in the limit \(D \to 0\).

For the lower bound note that as \(D \to 0\) the expression for \(\Delta\) in (12) satisfies
\[
\Delta \triangleq (DP^2 + 1 - \rho^{2(B+1)})^2 - 4D\rho^2(1 - \rho^2) \to (1 - \rho^{2(B+1)})^2.
\]
Upon direct substitution in (12) we have that
\[
\lim_{D \to 0} \left\{ R_{GM}(B, D) - \frac{1}{2} \log \left( \frac{1 - \rho^{2(B+1)}}{D} \right) \right\} = 0,
\]
(194)
as required.

To establish the upper bound note that according to Prop. 3 we can write
\[
R_{GM-ME}^+(L, B, D) = I(s_t; u_t|\hat{s}_L-L-B, [u]_{L-B+1}^{L-B-1})
\]
\[
= h(s_t|\hat{s}_L-L-B, [u]_{L-B+1}^{L-B-1}) - h(s_t|\hat{s}_L-L-B, [u]_{L-B+1}^{L-B-1}, u_t)
\]
\[
= h(s_t|\hat{s}_L-L-B, [u]_{L-B+1}^{L-B-1}) - \frac{1}{2} \log(2\pi e D)
\]
(195)
\[
= \frac{1}{2} \log \left( \frac{1 - \rho^{2(B+1)}}{D} \right)
\]
(196)
where the last term follows from the definition of \(\hat{s}_L-L-B\). Also we have
\[
h(s_t|\hat{s}_L-L-B, [u]_{L-B+1}^{L-B-1}) \leq h(s_t|\hat{s}_L-L-B, [u]_{L-B+1}^{L-B-1}, u_t) \leq h(s_t|u_{t-1})
\]
(197)
where the left hand side inequality in (197) follows from the following Markov property,
\[
\{\hat{s}_L-L-B, [u]_{L-B+1}^{L-B-1}\} \to s_t \to s_{t-1}
\]
(198)
and the fact that conditioning reduces the differential entropy. Also, the right hand side inequality in (197) follows from the latter fact. By computing the upper and lower bounds in (197) we have
\[
\frac{1}{2} \log \left( 2\pi e (1 - \rho^{2(B+1)}) \right) \leq h(s_t|\hat{s}_L-L-B, [u]_{L-B+1}^{L-B-1}) \leq \frac{1}{2} \log \left( 2\pi e \left( 1 - \frac{\rho^{2(B+1)}}{1 + \sigma^2_t} \right) \right)
\]
(199)
Now note that
\[
D \geq \sigma^2_t(\hat{s}_L-L-B, [u]_{L-B+1}^{L-B-1}, u_t) \geq \sigma^2_t(u_t, s_{t-1})
\]
(200)
\[
= \left( \frac{1}{\sigma^2_t} + \frac{1}{1 - \rho^2} \right)^{-1}
\]
(201)
which equivalently shows that if $D \to 0$ we have that $\sigma^2_2 \to 0$. By computing the limit of the upper and lower bounds in (199) as $D \to \infty$, we can see that
\[
\lim_{D \to 0} \left\{ b(s_i|s_{i-L-B}, [u]^t_{i-L-B+1}) - \frac{1}{2} \log \left( 2\pi e(1 - \rho^2(B+1)) \right) \right\} = 0 \tag{202}
\]
Finally (202) and (196) results in
\[
\lim_{D \to 0} \left\{ R_{GM-ME}(L, B, D) - \frac{1}{2} \log \left( 1 - \frac{\rho^2(B+1)}{D} \right) \right\} = 0 \tag{203}
\]
as required. Equations (194), (203) and (193) establishes the results of Corollary 2.

VII. INDEPENDENT GAUSSIAN SOURCES WITH SLIDING WINDOW RECOVERY: PROOF OF THEOREM 3

In this section we study the Gaussian source model discussed in section III-C and establish the rate-recovery function stated in Theorem 3. We do this by presenting the coding scheme in section VII-B and the converse in section VII-C. We also study some baseline schemes and compare their performance with the rate-recovery function at the end of this section.

A. Sufficiency of $K = B + W$

First we argue that it is sufficient to consider the case $K = B + W$. In particular, 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
\[
d'_j = \begin{cases} d_j & \text{for } j \in \{1, 2, \ldots, K\} \\ 1 & \text{for } j \in \{K + 1, \ldots, K'\}, \end{cases} \tag{204}
\]
and thus this case is a special case of $K = B + W$.

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 that it is possible to simply reuse a better reproduction of $s^n_{i,j}$ from the past. In particular there are two possibilities during the recovery at time $i$. Either, $\bar{t}_{i-1}$ or, $\bar{t}_{i-B-W-1}$ are guaranteed to have been reconstructed. In the former case $\{s^n_{i,j}\}_{d_{i-1}}$ is\footnote{The notation $\{\bar{s}^n\}_d$ indicates the reconstruction of $\bar{s}^n$ within average distortion $d$.} available from time $i - 1$ and $d_{i-1} \leq d_j$. In the latter case $\{s^n_{i,j}\}_{d_{i-B-W-1}}$ is available from time $i - B - W - 1$ and again $d_{i-B-W-1} \leq d_j$. Thus the reconstruction of any layer $j \geq B + W$ does not require any additional rate and it again suffices to assume $K = B + W$.

B. Coding Scheme

Throughout the coding scheme, we assume the source sequences are of length $n \cdot t$ where both $n$ and $t$ will be assumed to be arbitrarily large. The block diagram of the scheme is shown in Fig. 9. As discussed in section VII-B1, we partition $s^n_{i,t}$ into $t$ blocks each consisting of $n$ symbols $(s^n_{i,l})$ and apply a successive refinement quantization codebook to each such block to generate $B + 1$ refinement layers. As discussed in section VII-B2, these layers are carefully assigned across each channel packet so that the distortion constraints can be satisfied. At each time we generate $t$ channel packets as shown in Fig. 9. Each channel packet is treated as a super-symbol. We finally bin such a sequence of $t$ symbols at each time and transmit the bin index over the channel. At the receiver the sequence $c^n_i$ is first reconstructed by the inner decoder. Thereafter upon rearranging the refinement layers in each packet, the required reconstruction sequences are produced.

1) Successive Refinement (SR) Encoder: The encoder at time $i$, first partitions the source sequence $s^n_{i,t}$ into $t$ source sequences $(s^n_{i,l})_l \in [1, t]$. As shown in Fig. 10, we encode each source signal $(s^n_{i,l})$ using a $(B + 1)$-layer successive refinement codebook [29], [30] to generate $(B + 1)$ codewords whose indices are given by $\{(m_{i,j})_l, (m_{i,1})_t, \ldots, (m_{i,B})_t\}$ where $(m_{i,j})_l \in \{1, 2, \ldots, 2^{nR_l}\}$ for $j \in \{0, 1, \ldots, B\}$ and
\[
\hat{R}_j = \begin{cases} 0 & \text{for } j = 0 \\ \frac{1}{2} \log \left( \frac{2^{w_{i,j+1}}}{2^{w_{i,j}}} \right) & \text{for } j \in \{1, 2, \ldots, B - 1\} \\ \frac{1}{2} \log \left( \frac{1}{2^{w_{i,B}}} \right) & \text{for } j = B, \end{cases} \tag{205}
\]

The $j$-th layer uses indices
\[
(M_{i,j})_l \triangleq \{(m_{i,j})_l, \ldots, (m_{i,B})_t\} \tag{206}
\]
for reproduction and the associated rate with layer $j$ is given by:
\[
R_j = \begin{cases} \sum_{k=0}^{B} \hat{R}_k & \text{for } j = 0 \\ \sum_{k=j}^{B} \hat{R}_k & \text{for } j \in \{1, 2, \ldots, B\}, \end{cases} \tag{207}
\]
Fig. 9: Schematic of encoder and decoder for i.i.d. Gaussian with sliding window recovery constraint. SR and LR indicate successive refinement and layer rearrangement, respectively.

![Schematic of encoder and decoder](image)

Fig. 10: \((B + 1)\)-layer coding scheme based on successive refinement.

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

Clearly, for recovering \(t_i^{nt} = (s_i^{nt}, \ldots, s_{i-B-W}^{nt})\) with a distortion tuple \((d_0, \ldots, d_{B+W})\), it suffices that the destination have access to:

\[
\mathcal{M}_t^i = \begin{cases} 
M_t^{i,0} \\
M_t^{i,-1,0} \\
\vdots \\
M_t^{i,-W,0} \\
M_t^{i,-W-1,1} \\
\vdots \\
M_t^{i,-W-B,1}
\end{cases}
\]  

(208)

where each \(M_{i,j}^t\) is defined as

\[
M_{i,j}^t = \{(M_{i,j})_1, \ldots, (M_{i,j})_t\}
\]

(209)

and where \((M_{i,j})_t\) is the \(j\)-th layer in the quantization of \((s_i^{nt})_t\) as defined in (206). In fact if the decoder can reconstruct \(\mathcal{M}_t^i\) at time \(i\), it is guaranteed to recover \((s_i^{nt}, \ldots, s_{i-B-W}^{nt})\) with distortions \(d_0\) and the sequences \(s_{i-W-1}^{nt}, \ldots, s_{i-W-B}^{nt}\) with distortions \(d_{W+1}, \ldots, d_{W+B}\) respectively, which clearly satisfies the original distortion constraint.
2) Layer Rearrangement (LR) and Binning: In order to transmit the source sequences $M_i$ through the burst erasure channel, the encoder first rearrange them to produce an auxiliary set of codewords as follows.

$$\mathbf{c}_i = \begin{pmatrix} c_{i,0} \\ c_{i,1} \\ \vdots \\ c_{i,B} \end{pmatrix} = \begin{pmatrix} M_{i,0} \\ M_{i-1,1} \\ \vdots \\ M_{i-B,B} \end{pmatrix}$$  \hspace{1cm} (210)

It can be verified from (210) and (207) that the rate associated with the $c_i^k$ is given by

$$R_C = \sum_{k=0}^{B} R_k = \frac{1}{2} \log \left( \frac{1}{d_0} \right) + \sum_{j=1}^{B} \frac{1}{2} \log \left( \frac{1}{d_{W+j}} \right)$$  \hspace{1cm} (211)

Thus compared to the achievable rate (22) in Theorem 3 we are missing the factor of $\frac{1}{W+1}$ in the second term. To reduce the rate we bin each the set of all sequences $c_i^k$ into $2^{nR}$ bins as in section IV-C. Upon observing $c_i^k$ we only transmit its bin index $f_i$ through the channel. We next describe the decoder and compute the minimum rate required to reconstruct $c_i^k$.

3) Decoding: 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 $c_i^k$ as follows.

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

$$nR \geq H(c_i | c_{i-1}^j)$$  \hspace{1cm} (212)

$$= H(M_{i,0}, M_{i-1,1}, \ldots, M_{i-B,B} | c_{i-1}^j)$$  \hspace{1cm} (213)

$$= H(M_{i,0}, M_{i-1,1}, \ldots, M_{i-B,B} | M_{i-1,0}, M_{i-2,1}, \ldots, M_{i-B,B-1}, M_{i-B-B,B})$$  \hspace{1cm} (214)

$$= H(M_{i,0})$$  \hspace{1cm} (215)

$$= nR_0$$  \hspace{1cm} (216)

where we use (210) in (213) and (214), and the fact that layer $j$ is a subset of layer $j-1$ i.e., $M_{i-j,j} \subset M_{i-j,j-1}$ in (215). Thus the reconstruction of $c_i^k$ follows since the choice of (22) satisfies (216).

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

$$n(W + 1)R \geq H(c_i^j, c_{i-W+1}^j, \ldots, c_i^j)$$  \hspace{1cm} (217)

$$= H(c_i^j, c_{i-W+1}^j, c_{i-W+2}^j, \ldots, c_i^j)$$  \hspace{1cm} (218)

$$= n \sum_{k=1}^{B} R_k + n(W + 1)R_0$$  \hspace{1cm} (219)

Where in (217) we explicitly identify the innovation sub-symbols in $(c_i^j, \ldots, c_i^j)$. Note that the sub-symbols satisfy $c_{i+1,j+1} \subset c_{i,j}$. In particular, in computing the rate in (217) all the sub-symbols in $c_i^j$ and the sub-symbols $c_{i,0}$ for $j \in [i - W + 1, i]$ need to be considered. From (216), (219) and (207), the rate $R$ is achievable if

$$R \geq R_0 + \frac{1}{W + 1} \sum_{k=1}^{B} R_k$$  \hspace{1cm} (220)

$$= \frac{1}{2} \log \frac{1}{d_0} + \frac{1}{2(W + 1)} \sum_{k=1}^{B} \log \frac{1}{d_{W+j}}.$$  \hspace{1cm} (221)

as required.

It only remains to show that the decoder is able to recover $M_i^j$ at each time $i$. From (208) and (210) it can be easily seen that the decoder is able to recover $M_i^j$ given the sequences $(c_i^j, c_{i-1}^j, \ldots, c_{i-W}^j)$. From the above cases it follows that for each time $i$, outside the error propagation window, the decoder can indeed reconstruct the sequence $(c_i^j, c_{i-1}^j, \ldots, c_{i-W}^j)$ with high probability and in turn recover $M_i^j$. Finally as noted before, upon recovering $M_i^j$, we can recover each source sequence in the sliding window within the specified distortion.
Remark 5. Our coding scheme builds upon the technique introduced in [31] for lossless recovery of deterministic sources. The example involving deterministic sources in [31] established that the lower bound in Theorem 1 can be attained for a certain class of deterministic sources. The binning based scheme is suboptimal in general. The present paper does not include this example, but the reader is encouraged to see [31].

Remark 6. In recent years there has been a growing interest finding deterministic approximations to Gaussian source and channel models (see e.g., [32]) using the binary expansion of real numbers. Such approximations have resulted in new constant-gap capacity results for some long-standing open problems in Gaussian multi-terminal information theory. Our result for the Gaussian case have also been motivated from our earlier work on the deterministic source model [31]. One could view the different sub-symbols in the diagonally correlated deterministic Markov source model [31] as the different resolution levels required during the reconstruction of each Gaussian sources. However we note that our result for the Gaussian case results in an exact rate-recovery function rather than a constant-gap approximation.

C. Converse 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 (221).

As in the proof of Theorem 1, we consider a periodic erasure channel of period \(p = B + W + 1\) and assume that the first \(B\) positions of each period are erased. Consider,

\[
(W + 1)n(t + 1)R = H \left( \left[ f_{B}^{p-1}, f_{p+B}^{p-1}, \ldots, f_{(t-1)p+B}^{p-1} \right] \right) + H \left( \left[ f_{0}^{p-1}, f_{p}^{p-1}, \ldots, f_{(t-1)p+B}^{p-1} \right] ; s_{-1} \right)
\]

which the last step follows from the fact that conditioning reduces entropy.

We next establish the following claim, whose proof is in Appendix F.

Claim 1. For each \(k \geq 1\) we have that

\[
H \left( \left[ f_{k}^{p-1} \right] | f_{0}^{p-1}, s_{-1} \right) \geq \frac{n}{2} \log \left( \frac{1}{d_{W+i}} \right) + \frac{n(W+1)}{2} \log \left( \frac{1}{d_{0}} \right)
\]

Proof: See Appendix F.

Substituting (225) into (224) and taking \(n \to \infty\) and then \(t \to \infty\), we recover

\[
R \geq \frac{1}{2} \log_2 \left( \frac{1}{d_0} \right) + \frac{1}{2(W+1)} \sum_{j=1}^{B} \log_2 \left( \frac{1}{d_{W+j}} \right)
\]
as required.

D. Illustrative Suboptimal Schemes

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

1) 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

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

In this scheme, the decoder is able to recover the source whenever its codeword is available, i.e. at all the times except when the erasure happens.
Wyner-Ziv Compression with Delayed Side Information: At time $i$ the encoder 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

$$R_{WZ}(B, d) = I(t_i; \hat{t}_i | \hat{t}_{i-B-1}) = \frac{B}{2} \log(\frac{1}{d_k})$$

(228)

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 $I(t_i; \hat{t}_i | \hat{t}_{i-B-1}) \geq I(t_i; \hat{t}_{i-1})$.

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.

$$R_{FEC}(B, W, d) = I(t_i; \hat{t}_i | \hat{t}_{i-B-1}) = \frac{B}{2(W+1)} \log(\frac{1}{d_0})$$

(229)

(230)

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

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

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

(231)

(232)

In Fig. 11, we compare the result in Theorem 3 with the described schemes. It can be observed from Fig. 11 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 $W = 0$. 

Fig. 11: 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 $d = (0.1, 0.25, 0.4, 0.55, 0.7, .85)$. 

VIII. Conclusions

We presented a real-time streaming scenario where a sequence of source vectors must be sequentially encoded, and transmitted over a burst erasure channel. The source vectors must be reconstructed with zero delay at the destination. However those sequences that occur during the erasure burst or a period of length $W$ following the burst need not be reconstructed. We assume that the source vectors are sampled i.i.d. across the spatial dimension and from a first-order, stationary, Markov process across the temporal dimension. We study the minimum achievable compression rate, which we define to be the rate-recovery function in our setup.

For the case of discrete sources and lossless recovery, we establish upper and lower bounds on the rate-recovery function and observe that they coincide for the special cases when $W = 0$ and $W \to \infty$. More generally both our upper and lower bound expressions can be expressed as the rate of predictive coding plus another term that decreases at-least inversely with $W$. For the restricted class of memoryless encoders and symmetric sources, we establish that a binning based scheme is optimal. For the case of Gauss-Markov sources and a quadratic distortion measure, we establish upper and lower bounds on the minimum rate when $W = 0$ and observe that these bounds coincide in the high-resolution regime. The achievability is based on a quantize and binning scheme, but the analysis is a non-trivial extension of the lossless case as the reconstruction sequences at the destination do not form a Markov chain. We also study another setup involving independent Gaussian sources and a sliding-window reconstruction constraint where the rate-recovery function is attained using a successively refinement coding scheme. Finally we relax the zero-delay condition at the destination to a finite decoding delay and establish the minimum required rate when $W = 0$.

We believe that the present work can be extended in a number of directions. The focus in this paper has been on a burst-erasure channels. It may be interesting to consider channels that introduce both burst erasures and isolated erasures as considered recently in the channel coding context [19]. We expect the characterization of the worst-case erasure patterns for such models to be more challenging than the case of burst erasures. Secondly our present setup assumes that within the recovery period, at the destination do not form a Markov chain. We also study another setup involving independent Gaussian sources and a quantize and binning scheme, but the analysis is a non-trivial extension of the lossless case as the reconstruction sequences

APPENDIX A

Proof of Corollary 1

According to the chain rule of entropies, the term in (8) can be written as

$$H(s_{B+1}, s_{B+2}, \ldots, s_{B+W+1}| s_0)$$

$$= H(s_{B+1}| s_0) + \sum_{k=1}^{W} H(s_{B+k+1}| s_0, s_{B+1}, \ldots, s_{B+k})$$

$$= H(s_{B+1}| s_0) + W H(s_1| s_0)$$

$$= H(s_{B+1}| s_0) - H(s_{B+1}| s_B, s_1) + H(s_{B+1}| s_B, s_0) + W H(s_1| s_0)$$

$$= H(s_{B+1}| s_0) - H(s_{B+1}| s_B, s_1) + H(s_{B+1}| s_B) + W H(s_1| s_0)$$

$$= I(s_{B+1}; s_B| s_0) + (W + 1) H(s_1| s_0)$$

$$= (W + 1) R^+(B, W)$$

where (234) follows from the Markov property

$$s_0, s_{B+1}, \ldots, s_{B+k-1} \rightarrow s_{B+k} \rightarrow s_{B+k+1}$$

for any $k$ and from the temporally independency and stationarity of the sources which for each $k$ implies that

$$H(s_{B+k+1}| s_{B+k}) = H(s_1| s_0).$$

Note that in (235) we add and subtract the same term and (236) also follows from the Markov property of (239) for $k = 0$.

APPENDIX B

Proof of Lemma 3

To establish (77) we note that for the memoryless encoders and $k \in \{j, \ldots, j + W - 1\}$ the following Markov chain holds:

$$(s^n_{n-1}, f^n_{k-1}^{k-1}) \rightarrow s^n_k \rightarrow (f_k, s^n_k).$$

Hence we have that

$$n \geq H(s^n_k | f^n_{k-1}^{k-1}) \geq H(s^n_k | f^n_{k-1}^{k-1}, s^n_{k-1}, f_k, s^n_k)$$

$$= H(s^n_k | s^n_{k-1}, f_k),$$

$$= H(s^n_{k-1}, f_k),$$

$$H(s^n_{k-1}, f_k),$$

$$H(s^n_{k-1}, f_k).$$

$$H(s^n_{k-1}, f_k).$$

$$H(s^n_{k-1}, f_k).$$
This completes the proof.

where the last step follows via (241). Similarly using \((s_1^n, [f]_0^{j}\rightarrow W^{-1}) \rightarrow (s_{j-B}^n, s_{j+W}^n, f_{j+W})\), we have
\[
\begin{align*}
\sum_{n \leq n} & H(s_j^n, f_j^{j-1}, s_{j-W}^n) \\
& \geq H(s_j^n, f_j^{j-1}, [f]_j^{j+W}, s_{j-W}^n) \\
& = H(s_j^n, f_j^{j-W}, [f]_j^{j+W}, s_{j-W}^n) + H(s_j^n, f_j^{j-W}, s_{j-W}^n) \\
& \geq H(s_j^n, f_j^{j-W}, s_{j-W}^n) + H(s_j^n) \\
& \geq n + \frac{2}{2} \log \left( \rho^2 \frac{h(s_{j-1}^n, [f]_j^{j-1}, s_{j-W}^n)}{2R} + 2\pi e(1 - \rho^2) \right) - nR
\end{align*}
\]  
(244)

This completes the proof.

**Appendix C**

**Proof of Lemma 4**

Define \(q_k \triangleq 2\pi h(s_k^n)^2\). Consider the following entropy term.
\[
\begin{align*}
& h(s_k^n)^2 \\
= & h(s_k^n)^2 - I(f_k; s_k^n)^2 \\
= & h(s_k^n)^2 - H(f_k[s_k^n]) \\
& + H(f_k) \\
\geq & n + 2\pi e(1 - \rho^2) - nR
\end{align*}
\]  
(245)

Where (245) follows from the fact that conditioning reduces entropy and (246) follows from the Entropy Power Inequality(246). Thus
\[
q_k \geq \frac{\rho^2}{22R} k + 2\pi e(1 - \rho^2).
\]  
(247)

By repeating the iteration in (247), we have
\[
q_k \geq \frac{\rho^2}{22R} k + 2\pi e(1 - \rho^2) \frac{\rho^2}{22R} k + \frac{\rho^2}{22R} k + \frac{\rho^2}{22R} k + \ldots
\]  
(248)

where (249) follows from the fact \(0 < \frac{\rho^2}{22R} < 1\) for any \(\rho \in (0, 1)\) and \(R > 0\). This completes the proof.

**Appendix D**

**Proof of (98) and (99)**

In order to establish (98) and (99) we first establish the following Lemmas.

**Lemma 8.** Consider random variables \(\{X_0, X_1, X_2, Y_1, Y_2\}\) that are jointly Gaussian, \(X_k \sim \mathcal{N}(0, 1)\), \(k \in \{0, 1, 2\}\), \(X_0 \rightarrow X_1 \rightarrow X_2\) and that for \(j \in \{1, 2\}\) we have:
\[
\begin{align*}
X_j &= \rho_j X_{j-1} + N_j, \\
Y_j &= X_j + Z_j.
\end{align*}
\]  
(250)

Assume that \(Z_j \sim \mathcal{N}(0, \sigma_j^2)\) are independent of all random variables and likewise \(N_j \sim \mathcal{N}(0, 1 - \rho_j^2)\) for \(j \in \{1, 2\}\) are also independent of all random variables. The structure of correlation is sketched in Fig. 12. Then we have that:
\[
\sigma_{X_2}^2(X_0, Y_2) \leq \sigma_{X_2}^2(X_0, Y_1)
\]  
(252)

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

\[
\sigma_{\hat{x},t}(X_0, Y_2) = E[X_2^2] - (E[X_2 Y_2]) \left( \begin{array}{cc}
E[Y_2^2] & E[X_2 Y_2] \\
E[X_2 Y_2] & E[X_2^2]
\end{array} \right)^{-1} \left( \begin{array}{c}
E[X_2 Y_2] \\
E[X_2^2]
\end{array} \right)
\]

(253)

\[
= 1 - \rho_2^2(1 + \sigma_2^2) \left( \frac{1 + \sigma_2^2}{\rho_1} \right) \left( \begin{array}{c}
1 \\
\rho_1
\end{array} \right)
\]

(254)

\[
= 1 - \rho_2^2(1 + \sigma_2^2) \frac{1 + \sigma_2^2}{\rho_1}
\]

(255)

where we use the fact that \( E[X_0^2] = 1, E[Y_1^2] = 1 + \sigma_2^2, E[X_0 Y_1] = \rho_1, E[X_2 X_0] = \rho_0 \rho_1 \) and \( E[X_2 Y_1] = \rho_2 \). In a similar fashion it can be shown that:

\[
\sigma_{\hat{x},t}(X_0, Y_2) = 1 - (1 + \rho_2 \rho_1) \left( \frac{1 + \sigma_2^2}{\rho_1 \rho_2} \right) \left( \begin{array}{c}
1 \\
\rho_1 \rho_2
\end{array} \right)
\]

(256)

\[
= 1 - \rho_2^2(1 + \sigma_2^2) \frac{1 + \sigma_2^2}{\rho_1 \rho_2}
\]

(257)

To establish (252) we only need to show that,

\[
\frac{\rho_1^2 \rho_2^2 \sigma_2^2 - \rho_1^2 \rho_2^2 + 1}{1 + \sigma_2^2 - \rho_1 \rho_2^2} \geq \frac{\rho_1^2 \rho_2^2 \sigma_2^2 - \rho_1^2 \rho_2^2 + \rho_2^2}{1 + \sigma_2^2 - \rho_1 \rho_2^2}
\]

(258)

It is equivalent to showing

\[
\frac{1 + \sigma_2^2 - \rho_1^2}{1 + \sigma_2^2 - \rho_1 \rho_2^2} \geq \frac{\rho_1^2 \rho_2^2 \sigma_2^2 - \rho_1^2 \rho_2^2 + \rho_2^2}{\rho_1^2 \rho_2^2 \sigma_2^2 - \rho_1 \rho_2^2 + 1}
\]

(259)

or equivalently

\[
1 - \rho_2^2(1 - \rho_2^2) \geq 1 - \frac{1 - \rho_2^2}{\rho_1^2 \rho_2^2 \sigma_2^2 - \rho_1 \rho_2^2 + 1}
\]

(260)

which is equivalent to showing

\[
\frac{\rho_1^2}{1 + \sigma_2^2 - \rho_1 \rho_2^2} \leq \frac{1}{\rho_1^2 \rho_2^2 \sigma_2^2 - \rho_1 \rho_2^2 + 1}
\]

(261)

However note that (261) 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 \( \rho_2^2 \in (0, 1) \). This completes the proof.

\[\square\]

Lemma 9. Consider the Gauss-Markov source model and the test channel defined before. For a fixed \( t, k \in [1, t] \) and the set \( \Omega \subseteq [t - k, t] \), consider two sets of random variables \( \mathcal{W}_1 \) and \( \mathcal{W}_2 \) such that the following Markov property holds:

\[
\mathcal{W}_1 \rightarrow s_{t-k} \rightarrow \{ s_t, u_t \}
\]

(262)

\[
\mathcal{W}_2 \rightarrow s_{t-k} \rightarrow \{ s_t, u_t \}
\]

(263)

Then, \( \sigma_{s_{t-k}}^2(\mathcal{W}_1) \leq \sigma_{s_{t-k}}^2(\mathcal{W}_2) \) implies that

\[
h(s_t | \mathcal{W}_1, u_t) \leq h(s_t | \mathcal{W}_2, u_t)
\]

(264)

\[
h(u_t | \mathcal{W}_1, u_t) \leq h(u_t | \mathcal{W}_2, u_t).
\]

(265)

Proof: Based on the fact that all the random variables are Gaussian, we can express the MMSE estimates of \( s_{t-k} \) from \( \mathcal{W}_j, j \in \{1, 2\} \) as follows.

\[
\hat{s}_{t-k}(\mathcal{W}_1) = \alpha_1 s_{t-k} + e_1
\]

(266)

\[
\hat{s}_{t-k}(\mathcal{W}_2) = \alpha_2 s_{t-k} + e_2
\]

(267)

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 (266) and (267) are given by

\[
\alpha_j = 1 - \sigma_{s_{t-k}}^2(\mathcal{W}_j)
\]

(268)

\[
E_j = \sigma_{s_{t-k}}^2(\mathcal{W}_j)(1 - \sigma_{s_{t-k}}^2(\mathcal{W}_j))
\]

(269)
for $j = 1, 2$. Then we have
\[
\begin{align*}
h(s_t | \mathcal{W}_1, u_{t1}) &= h(s_t | \hat{s}_{t-k}(\mathcal{W}_1), u_{t1}) \\
&= h(s_t | \sigma_1 \hat{s}_{t-k} + e_t, u_{t1}) \\
&\leq h(s_t | \sigma_2 \hat{s}_{t-k} + e_t, u_{t1}) \\
&= h(s_t | \hat{s}_{t-k}(\mathcal{W}_2), u_{t1}) \\
&= h(s_t | \mathcal{W}_1, u_{t1})
\end{align*}
\] (270)

where (270) and (274) follows from the following Markov property.
\[
\begin{align*}
\mathcal{W}_1 &\rightarrow \hat{s}_{t-k}(\mathcal{W}_1) \rightarrow \{s_t, u_{t1}\} \\
\mathcal{W}_2 &\rightarrow \hat{s}_{t-k}(\mathcal{W}_2) \rightarrow \{s_t, u_{t1}\}
\end{align*}
\] (275)

(271) and (273) follows from (266) and (267) and (272) follows from the fact that $\sigma^2_{t-k}(\mathcal{W}_1) \leq \sigma^2_{t-k}(\mathcal{W}_2)$ implies that
\[
\frac{E_1}{\sigma_1^2} \leq \frac{E_2}{\sigma_2^2}
\] (277)

Thus the only difference between (271) and (272) 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 (271), which justifies the inequality in (272).

Eq. 265 can be established using the fact that the noise $z_t$ in the test channel is independent noise and according to Shannon’s EPI we have for any $j \in \{1, 2\}$
\[
h(u_t | \mathcal{W}_j, u_{t1}) = \frac{1}{2} \left( 2^{h(s_t | \mathcal{W}_j, u_{t1})} + 2\pi e \sigma^2_t \right)
\] (278)

Based on (278), (264) implies (265). This completes the proof of the lemma.

We now proceed to establish (98) and subsequently establish (99) in a similar fashion. We proceed the proof in two steps.

1) First by applying Lemma 8 we show.
\[
\sigma^2_{t-k}(u_{t0}^{t-B'-k-1}, u_{t-k}, s_{-1}) \leq \sigma^2_{t-k}(u_{t0}^{t-B'-k-1}, u_{t-B'-k}, s_{-1})
\] (279)

Eq states that knowing $\{u_{t0}^{t-B'-k-1}, u_{t-k}, s_{-1}\}$ rather than $\{u_{t0}^{t-B'-k-1}, u_{t-B'-k}, s_{-1}\}$, improves the estimate of the source $s_{t-k}$.

To show (279) Let $\hat{s}_{t-B'-k}(u_{t0}^{t-B'-k-1}, s_{-1})$ be the MMSE estimator of $s_{t-B'-k}$ given $\{u_{t0}^{t-B'-k-1}, s_{-1}\}$. Note that $\hat{s}_{t-B'-k}(u_{t0}^{t-B'-k-1}, s_{-1})$ is a sufficient statistic of $s_{t-B'-k}$ given $\{u_{t0}^{t-B'-k-1}, s_{-1}\}$ and thus we have that:
\[
\{u_{t0}^{t-B'-k-1}, s_{-1}\} \rightarrow \hat{s}_{t-B'-k}(u_{t0}^{t-B'-k-1}, s_{-1}) \
\rightarrow s_{t-B'-k} \rightarrow s_{t-k}.
\] (280)

Therefore, by application of Lemma 8 for $X_0 = \hat{s}_{t-B'-k}(u_{t0}^{t-B'-k-1}, s_{-1})$, $X_1 = s_{t-B'-k}$, $Y_1 = u_{t-B'-k}$, $X_2 = s_{t-k}$ and $Y_2 = u_{t-k}$, we have
\[
\sigma^2_{t-k}(u_{t0}^{t-B'-k-1}, u_{t-k}, s_{-1}) = \sigma^2_{t-k}(\hat{s}_{t-B'-k}(u_{t0}^{t-B'-k-1}, s_{-1}), u_{t-k})
\] (281)

\[
\leq \sigma^2_{t-k}(\hat{s}_{t-B'-k}(u_{t0}^{t-B'-k-1}, s_{-1}), u_{t-B'-k})
\] (282)

\[
= \sigma^2_{t-k}(u_{t0}^{t-B'-k-1}, u_{t-B'-k}, s_{-1}).
\] (283)

where (281) and (283) both follow form (280). This completes the claim in (279).

2) Now by application of Lemma 9 for
\[
\begin{align*}
\mathcal{W}_1 &= \{u_{t0}^{t-B'-k-1}, u_{t-k}, s_{-1}\} \\
\mathcal{W}_2 &= \{u_{t0}^{t-B'-k-1}, u_{t-B'-k}, s_{-1}\} \\
\Omega &= [t-k+1, t-1]
\end{align*}
\] (284)

(285)

we have
\[
\begin{align*}
h(u_t | u_{t0}^{t-B'-k-1}, u_{t-k}^{t-1}, s_{-1}) &\leq h(u_t | u_{t0}^{t-B'-k}, u_{t-k}^{t-1+1}, s_{-1})
\end{align*}
\] (287)

And similarly by application of Lemma 9 for the $\mathcal{W}_1$ and $\mathcal{W}_2$ in (284) and (285) and $\Omega = [t-k+1, t]$, we have
\[
\begin{align*}
h(s_t | u_{t0}^{t-B'-k-1}, u_{t-k}^{t-1}, s_{-1}) &\leq h(s_t | u_{t0}^{t-B'-k}, u_{t-k}^{t-1+1}, s_{-1})
\end{align*}
\] (288)

This establish (98) and (99).
Appendix E

Proof of Lemma 6

We first prove (148) by induction as follows.

• First we show that (148) is true for \( r = 1 \), i.e. for some \( 0 \leq a_1 \leq b_1 \leq t \) we need to show

\[
  h(s_t \mid u_{a_1}, s_{-1}) \geq h(s_t \mid u_{b_1}, s_{-1})
\]

(289)

First we apply Lemma 8 in Appendix D 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

\[
  h(s_{a_1} \mid u_{a_1}, s_{-1}) \geq h(s_{b_1} \mid u_{b_1}, s_{-1})
\]

(290)

According to the source model we can write \( s_t = \rho^{t-b_1} s_{b_1} + \tilde{n} \) where \( \tilde{n} \sim \mathcal{N}(0, 1 - \rho^{2(t-b_1)}) \) and also we can express \( s_{b_1} = \hat{s}_{b_1}(u_{j}, s_{-1}) + w_j \) for \( j \in \{ a_1, b_1 \} \) where \( w_j \sim \mathcal{N}(0, \sigma^2_{b_1}(u_{j}, s_{-1})) \) is the MMSE estimation error. For \( j \in \{ a_1, b_1 \} \), we can write

\[
  s_t = \rho^{t-b_1} \hat{s}_{b_1}(u_{j}, s_{-1}) + \rho^{t-b_1}w_j + \tilde{n}.
\]

(291)

Then we have

\[
  \sigma^2_t(\hat{s}_{b_1}, s_{-1}) = \rho^{2(t-b_1)} \sigma^2_{b_1}(u_{j}, s_{-1}) + 1 - \rho^{2(t-b_1)}
\]

(292)

\[
  \geq \rho^{2(t-b_1)} \sigma^2_{b_1}(u_{b_1}, s_{-1}) + 1 - \rho^{2(t-b_1)}
\]

(293)

\[
  = \sigma^2_t(u_{b_1}, s_{-1})
\]

(294)

where (293) immediately follows from (290). Equivalently (294) concludes (289).

• Now assume that (148) 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 \( t \geq b_r \),

\[
  h(s_t \mid u_{a_1}, s_{-1}) \geq h(s_t \mid u_{b_1}, s_{-1})
\]

(295)

We are interested to show that assuming the induction hypothesis in (295), 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 \( a_{r+1} \leq b_{r+1} \leq t \).

We establish this in two steps.

1) We show that

\[
  h(s_t \mid u_{A_{r+1}}, s_{-1}) \geq h(s_t \mid u_{B_{r+1}}, s_{-1}).
\]

(296)

First note that by application of Lemma 8 for \( \{ X_0, X_1, X_2, Y_1, Y_2 \} = \{ \hat{s}_{b_1}(u_{A_r}, s_{-1}), s_{a_{r+1}}, s_{b_{r+1}}, u_{a_{r+1}}, u_{b_{r+1}} \} \) we have

\[
  h(s_{a_{r+1}} \mid \hat{s}_{b_1}(u_{A_r}, s_{-1}), u_{b_{r+1}}) \geq h(s_{b_{r+1}} \mid \hat{s}_{b_1}(u_{A_r}, s_{-1}), u_{b_{r+1}})
\]

(297)

Again according to the source model we can write \( s_t = \rho^{t-b_{r+1}} s_{b_{r+1}} + \tilde{n} \) where \( \tilde{n} \sim \mathcal{N}(0, 1 - \rho^{2(t-b_{r+1})}) \) and also we can express \( s_{b_{r+1}} = \hat{s}_{b_{r+1}}(\hat{s}_{b_1}(u_{A_r}, s_{-1}), u_j) + w_j \) for \( j \in \{ a_{r+1}, b_{r+1} \} \) where \( w_j \sim \mathcal{N}(0, \sigma^2_{b_1}(\hat{s}_{b_1}(u_{A_r}, s_{-1}), u_j)) \) is the MMSE estimation error. For \( j \in \{ a_{r+1}, b_{r+1} \} \), we can write

\[
  s_t = \rho^{t-b_{r+1}} \hat{s}_{b_{r+1}}(\hat{s}_{b_1}(u_{A_r}, s_{-1}), u_j) + \rho^{t-b_{r+1}}w_j + \tilde{n}.
\]

(298)

Then we have

\[
  \sigma^2_t(\hat{s}_{b_{r+1}}, s_{-1}) = \sigma^2_t(\hat{s}_{b_1}(u_{A_r}, s_{-1}), u_{b_{r+1}})
\]

(299)

\[
  = \rho^{2(t-b_{r+1})} \sigma^2_{b_{r+1}}(\hat{s}_{b_1}(u_{A_r}, s_{-1}), u_{b_{r+1}}) + 1 - \rho^{2(t-b_{r+1})}
\]

(300)

\[
  \geq \sigma^2_t(\hat{s}_{b_1}(u_{A_r}, u_{b_{r+1}}, s_{-1}))
\]

(301)

\[
  = \sigma^2_t(u_{b_{r+1}}, s_{-1})
\]

(302)

where (301) immediately follows from (297). Equivalently (302) concludes (296).

2) We show that

\[
  h(s_t \mid u_{A_r}, u_{b_{r+1}}, s_{-1}) \geq h(s_t \mid u_{B_{r+1}}, s_{-1}).
\]

(303)

First note that based on the induction hypothesis in (295) for \( t = b_{r+1} \) we have

\[
  h(s_{b_{r+1}} \mid u_{A_r}, s_{-1}) \geq h(s_{b_{r+1}} \mid u_{B_{r+1}}, s_{-1})
\]

(304)

and equivalently

\[
  \sigma^2_{b_{r+1}}(u_{A_r}, s_{-1}) \geq \sigma^2_{b_{r+1}}(u_{B_{r+1}}, s_{-1})
\]

(305)
Now by application of Lemma 9 for \( k = t - b_{r+1} \) and
\[
\mathcal{W}_1 = \{ u_{B_1}, s_{-1} \} \\
\mathcal{W}_2 = \{ u_{A_1}, s_{-1} \} \\
\Omega = \{ b_{r+1} \}
\]
we have that
\[
h(s_t | u_{A_1}, u_{B_{r+1}}, s_{-1}) \geq h(s_t | u_{B_1}, u_{B_{r+1}}, s_{-1})
\]
which is equivalent to (303).

Combining (296) and (303) we have \( h(u_t | u_{A_{r+1}}, s_t) \geq h(u_t | u_{B_{r+1}}, s_t) \) which shows that (148) is also true for \( r+1 \).

Finally note that (148) implies (149) as follows.
\[
h(u_t | u_{A_1}, s_{-1}) = \frac{1}{2} \log \left( 2^{2h(s_t | u_{A_1}, s_{-1})} + 2\pi e\sigma^2 \right)
\]
where (311) follows from the application of Shannon’s EPI and the fact that the noise in the test channel is independent. Also (311) follows from (148). This completes the proof.

**APPENDIX F**

**PROOF OF CLAIM 1**

We need to lower bound \( H(( f_0^{(k+1)p-1} | f_0^{k-1}, s_0^n)) \). Consider
\[
H(( f_0^{(k+1)p-1} | f_0^{k-1}, s_0^n)) = I(( f_0^{(k+1)p-1}, t_0^n) | f_0^{k-1}, s_0^n) + H(( f_0^{(k+1)p-1}, t_0^n) | f_0^{k-1}, s_0^n)
\]
where (315) follows since \( t_0^{(k+1)p-1} = (s_0^n, \ldots, s_0^n) \) is independent of \( ( f_0^{k-1}, s_0^n) \) as the source sequences \( s_0^n \) are generated i.i.d. By expanding \( t_0^{(k+1)p-1} \) we have that
\[
h(t_0^{(k+1)p-1}) = h(s_0^n, s_0^n, s_0^{n+1}) + h(s_0^n, s_0^n, s_0^n)
\]
and
\[
h(t_0^{(k+1)p-1}) = h(s_0^n, s_0^n, s_0^{n+1}) = h(s_0^n, s_0^n, s_0^n)
\]
We next show the following
\[
h(s_0^n, s_0^n, s_0^{n+1}) - h(s_0^n, s_0^n, s_0^{n+1}) \geq \sum_{i=1}^{B} \frac{n}{2} \log \left( \frac{1}{d_{W+i}} \right)
\]
and that
\[
h(s_0^{n+1}, s_0^{n+1}) \geq \sum_{i=0}^{B-1} \left( h(s_0^{n+1}) - h(s_0^{n+1}) \right)
\]
The proof of Claim 1 follows from (315), (316), (317), (318) and (319).
We show that for each $i = 0, 1, \ldots, B - 1$
\[
h(s^k_{kp+i}) - h(s^k_{kp+i})|f_0^{|k|p-1}, f_{kp+B}^{|k|p-1}, s_{i-1}^n) \geq \frac{n}{2} \log \left( \frac{1}{dB+W-i} \right),
\]
which then establishes (318).

Recall that since there is an erasure burst between time $t \in [kp, kp + B - 1]$ the receiver is required to reconstruct
\[
\hat{s}_{kp+i}^n = [\hat{s}_{kp+B+W}, \ldots, \hat{s}_{kp+i}^n]
\]
with a distortion vector $(d_0, \ldots, d_{B+W})$ i.e., a reconstruction of $\hat{s}_{kp+i}^n$ is desired with a distortion of $dB+W-i$ for $i = 0, 1, \ldots, B + W$ when the decoder is revealed $(f_{kp-1}^{kp-1}, f_{kp+B}^{kp+B})$. Hence
\[
h(s_{kp+i}^n) - h(s_{kp+i}^n|f_0^{|kp-1}, f_{kp+B}^{|kp+B}, s_{i-1}^n, \{\hat{s}_{kp+i}^n\}_{dB+W-i}) \geq h(s_{kp+i}^n) - h(s_{kp+i}^n|\{\hat{s}_{kp+i}^n\}_{dB+W-i}) \geq h(s_{kp+i}^n) - h(s_{kp+i}^n - \{\hat{s}_{kp+i}^n\}_{dB+W-i})
\]
Since we have that
\[
E \left[ \frac{1}{n} \sum_{i=1}^{n} (s_{kp+i,j} - \hat{s}_{kp+i,j})^2 \right] \leq dB+W-i
\]
It follows from standard arguments that [25, Chapter 13] that
\[
h(s_{kp+i}^n - \{\hat{s}_{kp+i}^n\}_{dB+W-i}) \leq \frac{n}{2} \log 2\pi e dB+W-i.
\]
Substituting (327) into (325) and the fact that $h(s_{kp+i}^n) = \frac{n}{2} \log 2\pi e$ establishes (321).

It finally remains to establish (319).
\[
h(s_{kp+B}^n, \ldots, s_{kp+B-1}^n|f_0^{|kp-1}, s_{kp+B}^n, \ldots, s_{kp+B-1}^n) - h(s_{kp+B}^n, \ldots, s_{kp+B-1}^n|f_0^{|kp-1}, f_{kp+B}^{|kp+B}, s_{kp+B}^n, \ldots, s_{kp+B-1}^n)
\]
\[
+ H(f_{kp+B}^{|kp+B}, f_0^{|kp-1}, s_{kp+B}^n, \ldots, s_{kp+B-1}^n, s_{i-1}^n)
\]
\[
= I(s_{kp+B}^n, \ldots, s_{kp+B-1}^n|f_0^{|kp-1}, s_{kp+B}^n, \ldots, s_{kp+B-1}^n, s_{i-1}^n) + H(f_{kp+B}^{|kp+B}, f_0^{|kp-1}, s_{kp+B}^n, \ldots, s_{kp+B-1}^n, s_{i-1}^n)
\]
\[
= I(f_{kp+B}^{|kp+B}, s_{kp+B}^n, \ldots, s_{kp+B-1}^n|f_0^{|kp-1}, s_{kp+B}^n, \ldots, s_{kp+B-1}^n, s_{i-1}^n)
\]
The above mutual information term can be bounded as follows:
\[
h(s_{kp+B}^n, \ldots, s_{kp+B-1}^n|f_0^{|kp-1}, s_{kp+B}^n, \ldots, s_{kp+B-1}^n, s_{i-1}^n) - h(s_{kp+B}^n, \ldots, s_{kp+B-1}^n|f_0^{|kp-1}, s_{kp+B}^n, \ldots, s_{kp+B-1}^n, s_{i-1}^n)
\]
\[
= h(s_{kp+B}^n, \ldots, s_{kp+B-1}^n|f_0^{|kp-1}, s_{kp+B}^n, \ldots, s_{kp+B-1}^n, s_{i-1}^n) - h(s_{kp+B}^n, \ldots, s_{kp+B-1}^n|f_0^{|kp-1}, s_{kp+B}^n, \ldots, s_{kp+B-1}^n, s_{i-1}^n)
\]
\[
= h(s_{kp+B}^n, \ldots, s_{kp+B-1}^n|f_0^{|kp-1}, s_{kp+B}^n, \ldots, s_{kp+B-1}^n, s_{i-1}^n)
\]
\[
= \sum_{i=0}^{W} h(s_{kp+B+i}^n - h(s_{kp+B+i}^n - \{\hat{s}_{kp+B+i}^n\}_{dB+i})
\]
\[
\geq \sum_{i=0}^{W} \frac{n}{2} \log \left( \frac{1}{d_0} \right) = \frac{n(W+1)}{2} \log \left( \frac{1}{d_0} \right)
\]
where (331) follows from the independence of $(s_{kp+B}^n, \ldots, s_{kp+B-1}^n)$ from the past sequences, and (332) follows from the fact that given the entire past $f_0^{|kp-1}$ 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 (319).

References


