A Truncated Prediction Framework for Streaming over Erasure Channels

Farrokh Etezadi, Ashish Khisti, Jun Chen

Abstract

We propose a new coding technique for sequential transmission of a stream of Gauss-Markov sources over erasure channels under a zero decoding delay constraint. Our proposed scheme is a combination (hybrid) of predictive coding with truncated memory, and quantization & binning.

We study the optimality of our proposed scheme using an information theoretic model. In our setup, the encoder observes a stream of source vectors that are spatially i.i.d. and temporally sampled from a first-order stationary Gauss-Markov process. The channel introduces an erasure burst of a certain maximum length B, starting at an arbitrary time, not known to the transmitter. The reconstruction of each source vector at the destination must be with zero delay and satisfy a quadratic distortion constraint with an average distortion of D. The decoder is not required to reconstruct those source vectors that belong to the period spanning the erasure burst and a recovery window of length W following it. We study the minimum compression rate $R(B, W, D)$ in this setup.

As our main result we establish upper and lower bounds on the compression rate. The upper bound (achievability) is based on our hybrid scheme. It achieves significant gains over baseline schemes such as predictive coding, memoryless binning, a separation-based scheme and a group of pictures (GOP) based scheme. The lower bound is established by observing connection to network source coding problem. The bounds simplify in the high resolution regime, where we provide explicit expressions whenever possible, and identify conditions when the proposed scheme is close to optimal.

We finally discuss the interplay between the parameters of our burst erasure channel and the statistical channel models and explain how the bounds in the former model can be used to derive insights into the simulation results involving the latter. In particular our proposed scheme outperforms baseline schemes over the i.i.d. erasure channel and the Gilbert-Elliott channel, and achieve performance close to a lower bound in some regimes.

I. Introduction

Real-time streaming applications require communication systems that are both efficient in compression, yet resilient to channel errors. In practice compression efficiency and error resilience are conflicting objectives, and one has to strike a balance between them. Linear predictive techniques such as Differential Pulse Code Modulation (DPCM) have long been used to exploit the source memory in streaming applications. However such techniques are known to be sensitive to channel errors. When packets are lost over the network, a mismatch in the information available at the encoder and decoder arises. This effect, known as *prediction drift*, can lead to a significant degradation in the performance. Commonly used video compression formats such as H.264/MPEG and HEVC use a combination of intra-coded (I) and predictively-coded (P) frames. The intra-coded frames that can be encoded and decoded in a stand alone manner are used to limit the effect of error propagation. The predictively-coded frames, which depend on the past frames, can be compressed at a much higher efficiency. The ratio of I and P frames determines the tradeoff between error resilience and compression efficiency. Other approaches that strike a similar tradeoff include error correction coding [1], [2], leaky predictive coding [3]–[5], and distributed video coding [6], [7]. The main objective of this paper is to study the tradeoff between error propagation and compression efficiency for Gauss-Markov sources in an information theoretic framework. Through our study we also develop a new coding technique that exhibits significant gains over baseline schemes.

We consider an information theoretic model for streaming that captures the trade-off between compression rate and error propagation over burst erasure channels. In particular we assume that the channel introduces an erasure burst of maximum length B. A stream consisting of source sequences that are sampled from a Gauss-Markov distribution, must be sequentially transmitted by the encoder, and reconstructed under a quadratic distortion measure. The decoder is required to reconstruct each source sequence with zero delay and within an average distortion D, except those sequences observed during the error propagation window following the start of the erasure burst. We assume that this period spans the duration of the erasure burst, as well as an interval of length W, called the recovery period, immediately following it. In practice the receiver will invoke a frame-freeze or use other type of error concealment techniques in this period. We study the minimum achievable rate $R(B, W, D)$, and define it as the rate-recovery function. The formal setup is defined in Section II.

As our main result we establish upper and lower bounds on the rate-recovery function. Our achievability, discussed in Sections III and IV, is based on a new coding scheme that can be described as a *hybrid* between predictive coding and quantization & binning. In the first step we apply generalized predictive coding, albeit with truncated memory at the encoder.[1]

---

[1]In general, the prediction coefficients will not be the optimal MMSE estimator as will be discussed in Section III.

Unlike ideal predictive coding, the resulting output sequences will not be independent of the past. In the second step we further reduce the rate using quantization and binning. We develop an expression for the achievable rate for our proposed scheme and demonstrate that it results in significant gains over ideal predictive coding and memoryless binning. We also establish a lower bound on the rate-recovery function in Section V. The key insight in our lower bounding technique is to consider a network source coding problem that captures the tradeoff between error propagation and compression rate, and establish lower bounds in such a setup. While our general expressions for the upper and lower bounds are rather involved they simplify considerably in the high resolution regime. We obtain several explicit bounds in this regime. The analysis of some baseline schemes is presented in Section VI for comparisons.

While our primary focus is on burst erasure channels, the coding schemes we study are not dependent on such a model. The only consideration of the channel model is in determining the rate R. As such decoder can resume zero-delay recovery following any erasure sequence after a sufficiently long waiting time. In Section VII we discuss how such a waiting time at the decoder can be computed for the hybrid scheme as well as some baseline schemes for mismatched burst lengths. We also present several simulation results over statistical channel models such as the i.i.d. erasure channel and the Gilbert-Elliott channel. These results demonstrate that the residual loss probability of our proposed hybrid scheme can be significantly lower than commonly used schemes such as predictive coding, memoryless binning, GOP-type compression and separation based schemes. Furthermore in certain cases the hybrid scheme achieves a performance close to a lower bound on the loss probability.

A. Related Works

The information theoretic model for sequential coding of video was introduced in [8] and followed up in a number of works, see e.g., [9], [10] and references therein. These works only consider source coding aspects and assume an ideal transmission channel. The source model used in the present paper follows this line of work. In [11] the authors study a variation of the setup in [8] where the decoder at each time has either the entire past or just the present output, with a different distortion imposed in each case. However this model too does not capture fact that the decoder has access to a subset of channel outputs in the streaming setup. In reference [12] the authors consider the case of a single packet loss on sequential coding and compare predictive coding and Wyner-Ziv coding in this setup. In reference [3] the authors study real-time transmission of Gauss-Markov sources over i.i.d. erasure channels. A leaky prediction based scheme is proposed and the coefficients are optimized under the assumption that the erasures are sparse enough that the effect of each erasure remains isolated. To the best of our knowledge both the coding techniques and lower bounds in the present work are new.

Our earlier work [13] considers a similar information theoretic model for streaming over burst erasure channels, and establishes upper and lower bounds on the rate-recovery function for the case of discrete memoryless sources under lossless recovery. The case of Gauss-Markov sources is also considered in the special case when immediate recovery is required following the erasure burst. The optimality of memoryless quantize and binning is established in the high resolution regime in this special case. In contrast, the present work focuses on Gauss-Markov sources in the general case when a non-zero recovery period is allowed. We show that the memoryless coding scheme in [13] is sub-optimal in this case. Our proposed coding scheme gives significant improvements, not only in the proposed burst-erasure model, but also in simulations involving i.i.d. erasure channels. The analysis of the achievable rate is considerably more challenging for the the lossy case as the reconstruction sequences do not satisfy the Markov condition of the source sequences. Finally we remark that while the present paper only considers the case of zero-delay recovery at the destination, extensions to the case of non-zero decoding delay can be considered and will be reported elsewhere; see [14], [15]. For a comprehensive treatment of all the results, see also [16].

Notations: Throughout this paper we represent the expectation operator by $E[\cdot]$ and the variance by $Var(\cdot)$. The notation "log" is used for the binary logarithm, and rates are expressed in bits. The operations $H(\cdot)$ and $h(\cdot)$ denote the entropy and the differential entropy, respectively. The "slanted sans serif" font a and the normal font a represent random variables and their realizations respectively. The notation $a_i^n = \{a_{i,1}, a_{i,2}, \ldots, a_{i,n}\}$ represents a length-n sequence of symbols at time i. The notation $[f]_i^j$ for $i < j$ represents $f_i, f_{i+1}, \ldots, f_j$.

II. PROBLEM STATEMENT

At each time t, the Gaussian source vector $\{s_t^n\}$ is sampled i.i.d. from a zero-mean Gaussian distribution $\mathcal{N}(0, \sigma_s^2)$ along the spatial dimension, and forms a first-order Gauss-Markov chain across the temporal dimension:

$$s_{t,i} = \rho s_{t-1,i} + n_{t,i},$$

where the first index $t \in [1 : T]$ denotes the temporal index and the second index $i \in [1 : n]$ denotes the spatial index. We let $\rho \in (0, 1)$ and $n_{t,i} \sim \mathcal{N}(0, (1 - \rho^2)\sigma_s^2)$. Without loss of generality we assume $\sigma_s^2 = 1$. The sequence s_0^n which is also sampled from the same distribution is revealed to both the encoder and decoder before the start of the communication. It plays the role of a synchronization frame. Without loss of generality, one can consider the source sequences $\{x_t^n\}$, instead of the original source sequences $\{s_t^n\}$, where
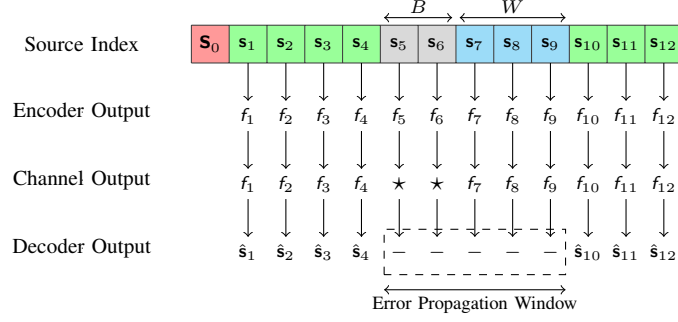
$$x_{t,i} \triangleq s_{t,i} - \rho^t s_{0,i}. \tag{1}$$

Fig. 1. Problem Setup: Consider the example of $B = 2$ and $W = 3$. The encoder output f_j is a function of the source sequences up to time j i.e., $s_0^n, s_1^n, \ldots, s_j^n$. The channel introduces an erasure burst of length B. The decoder produces $\hat{s}_j^n$ upon observing the sequence $\{g_0, g_1, \ldots, g_j\}$. The decoder is not required to produce those source sequences that fall in a window of length $B + W$ following the start of an erasure burst. However, the decoder recovers the rest of the source sequences within zero-delay and average distortion constraint $D_t \leq D$ where D_t is defined at (5).

Note that $x_{t,i} \sim \mathcal{N}\left(0, 1 - \rho^{2t}\right)$ and the source sequences $\{x_t^n\}$ inherits the first order Markov property from the original source sequences $\{s_t^n\}$, i.e., for any $i \in [1 : n]$ and $t \geq 1$

$$x_{t,i} = \rho x_{t-1,i} + n_{t,i}.$$

Throughout the paper, based on equivalence of the two models and in order to present the results in the simplest form, we will interchangeably use the two source models. We will also suppress the dependence of the encoding and decoding functions on s_0^n.

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

$$f_t = \mathcal{F}_t\left(x_1^n, x_2^n, \ldots, x_t^n\right), \qquad 1 \leq t \leq T. \tag{2}$$

Note that the encoder in (2) is a causal function of the source sequences.

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

$$g_t = \begin{cases} \star, & t \in \{j, j+1, \ldots, j+B-1\} \\ f_t, & \text{otherwise.} \end{cases} \tag{3}$$

To keep the setup simple, we consider a channel model with a single erasure burst in an unknown location during the entire communication period. However, as it will become clear in the sequel, the proposed coding schemes are not specific to this channel model.

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

$$\hat{x}_t^n = \mathcal{G}_t(g_1, g_2, \ldots, g_t), \qquad t \notin \{j, j+1, \ldots, j+B+W-1\}, \tag{4}$$

where $\hat{x}_t^n$ denotes the reconstruction sequence and j denotes the time at which burst erasure starts in (3). The destination is not required to produce the source vectors that appear either during the burst erasure or in the period of length W following it. We call this period the error propagation window. The average distortion at time t outside the error propagation window, denoted by D_t, is computed according to the quadratic distortion measure, i.e.,

$$D_t \triangleq E\left[\frac{1}{n}\sum_{i=1}^{n}(x_{t,i} - \hat{x}_{t,i})^2\right], \qquad t \notin \{j, j+1, \ldots, j+B+W-1\}. \tag{5}$$

We consider the case where the reconstruction in (4) satisfies the zero-delay and average distortion constraint of D. A tuple $(R_1, R_2, \ldots, R_T, D)$ is achievable if there exists a sequence of encoding and decoding functions and a sequence ϵ_n approaching to zero as $n \to \infty$ such that, for any burst erasure pattern in (3) and for any time t outside the error propagation window, $D_t \leq D + \epsilon_n$. Define $\mathfrak{R}_T$ to be the closure of the achievable tuples $(R_1, R_2, \ldots, R_T, D)$. We define the *rate-recovery function* as

$$R_T(B, W, D) \triangleq \inf_{(R_1, R_2, \ldots, R_T, D) \in \mathfrak{R}_T} \left\{\max_{k \in [1:T]} R_k\right\} \tag{6}$$
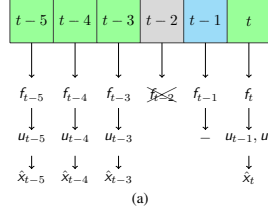
Fig. 2. Problem Setup for $B = W = 1$ and high resolution. The channel packet of time $t - 2$ is erased and $\hat{x}_t$ is required to be reconstructed at time t.

In particular we are interested in the rate-recovery function in the large T regime, i.e.,

$$R(B, W, D) \triangleq \limsup_{T \to \infty} R_T(B, W, D), \tag{7}$$

which we will simply call the rate-recovery function.

Furthermore, for convenience, we consider a normalization of the rate-recovery function. We first define $R_I(D)$ to be the steady state rate-distortion function of the erasure-free channel. This rate-distortion is achieved by a DPCM scheme and is computed as [8]:

$$R_I(D) \triangleq \frac{1}{2} \log \left(\frac{1 - \rho^2}{D} + \rho^2 \right). \tag{8}$$

We define the excess rate as follows.

Definition 1. *The excess rate associated with any rate-recovery function $R(B, W, D)$ is*

$$R_E(B, W, D) \triangleq R(B, W, D) - R_I(D), \tag{9}$$

where $R_I(D)$ is the rate associated with ideal DPCM (8).

In this paper, high resolution regime is defined as follows.

Definition 2. *The high resolution regime corresponds to the limit of $R_E(B, W, D)$ when $D \to 0$.*

Remark 1. *Our problem set up considers worst case metrics rather than the average packet loss rate and average distortion across time. Such worst case metrics may be more relevant to streaming applications where the atypical behavior is important. For example in interactive conferencing applications, it is meaningful to design the system so that the quality of the call remains acceptable up to a certain maximum burst length [17]. We will nevertheless also validate our results for statistical channels in Section VII.*

Remark 2. *We note that our setting involves a peak rate constraint in (6) since we take the maximum across all the rates. One can also consider average rate constraints. Indeed when we consider time-varying schemes such as GOP based coding, it is important to consider average rate constraints. However all the other schemes treated in this work are time-invariant in the steady state operation. Hence it suffices to consider the peak rate constraint for these schemes. We however note upfront that the peak rate constraint is used in deriving our lower bound, as it seems to simplify certain steps. The lower bound under an average rate constraint is not addressed in this paper.*

Remark 3. *The coding schemes proposed in this work are not specific to the burst erasure channel model in Fig 1. As will be apparent, they provide significant gains in simulations over i.i.d. erasure channels. Nevertheless the study of the simplified setup yields useful insights.*

Remark 4. *The source model considered in this paper, i.e., spatially i.i.d. and temporally Gauss-Markov process, has been also used in earlier works (see, e.g., [3], [8], [11]). The spatial independence assumption holds approximately if we consider the video innovation process generated by applying suitable transform on original video frames. While the symbols themselves might not be identically distributed, we note that our model provides a first step towards establishing the information theoretic tradeoffs. For further justification of the source and channel models, see [13, Sec. III-B].*

III. PRELIMINARY INSIGHTS

In this section we will discuss the implications of the truncated prediction step on the resulting test channels in the simpler setting when $B = W = 1$. We assume that the system operates in the steady state, the erasure happens at time $t - 2$ and zero-delay reconstruction must result starting at time t. The setup is shown in Fig 2. In analyzing the achievable rate we will also focus on high resolution regime, where the decoder is interested in reconstructing the source sequences with vanishing average distortion $D \to 0$.
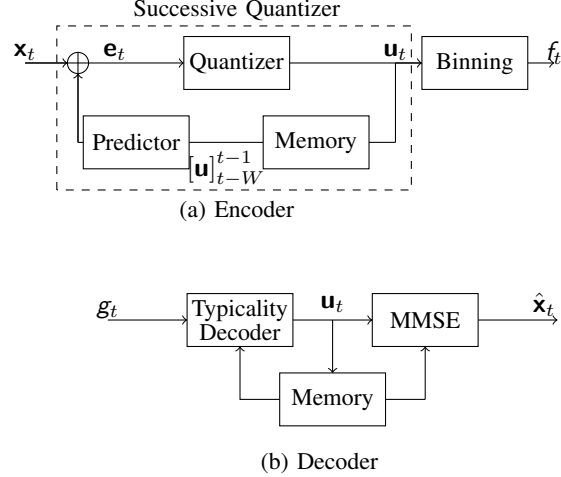
Fig. 3. Block diagram of the proposed coding scheme: a) Encoder, b) Decoder.

A. Hybrid Coding

Our proposed scheme, as illustrated in Fig. 3 encompasses two steps:

1) **Limited Prediction**: Predicting the current source sequence x_t^n, from the past W outputs at the decoder: $\{u_{t-1}^n, \ldots, u_{t-W}^n\}$, and quantizing the residual error e_t^n into a codeword u_t^n
2) **Binning**: We apply random binning to the sequence u_t^n in a memoryless fashion to output the index $f_t \in [1, 2^{nR}]$, that is sent over the channel

Throughout we will assume the test channels to be jointly Gaussian. In the special case when $W = 1$ we can represent the test channel associated with u_t as follows:

$$u_t = x_t - w \cdot u_{t-1} + z_t \tag{10}$$

where the scalar w is a free parameter that can be optimized over and $z_t \sim \mathcal{N}(0, \sigma_z^2)$ is the noise associated with the quantization step. We next bin the sequences u_t^n into indices $f_t \in [1, 2^{nR}]$. The rate is selected such that the decoder can reconstruct u_t^n and u_{t-1}^n simultaneously at time t when an erasure happens at time $t - 2$ (c.f. Fig. 2). From standard results in multi-terminal source coding the rate R must satisfy the following inequalities for all t (see e.g., [18]):

$$R \geq h(u_t | [u]_1^{t-3}, u_{t-1}) - \frac{1}{2} \log(2\pi eD), \tag{11}$$

$$R \geq h(u_{t-1} | [u]_1^{t-3}, u_t) - \frac{1}{2} \log(2\pi eD), \tag{12}$$

$$R_{\text{sum}} = 2R \geq h(u_{t-1}, u_t | [u]_1^{t-3}) - \log(2\pi eD). \tag{13}$$

Using the test channel (10), the decoder can reconstruct $\hat{x}_t$ using u_t and u_{t-1}:

$$\hat{x}_t = u_t + w \cdot u_{t-1} \tag{14}$$

Fig. 4 illustrates the operations at the encoder and decoder discussed above in the interval $[t-3, t]$. Note that by construction the truncated prediction scheme is such that the decoder always has the required codewords during the reconstruction process despite losses over the channel, except at times $\{t-2, t-1\}$ when reconstruction is not required.

Specializing to the high resolution regime, we must have $x_t \approx \hat{x}_t$, and the noise in the test channel $\sigma_z^2 \to 0$. In this case the resulting test channel (c.f. Fig. 3) can be expressed as:

$$u_t \approx n_t + (\rho - w)n_{t-1} + (\rho^2 - w\rho + w^2)n_{t-2} + \ldots, \tag{15}$$

where recall that the variables $n_t \sim \mathcal{N}(0, 1 - \rho^2)$ are the independent innovation (noise) symbols. The reader can verify that

$$u_t + w \cdot u_{t-1} \approx n_t + \rho n_{t-1} + \rho^2 n_{t-2} + \ldots = x_t.$$

Upon using (13) the sum rate can be expressed as:

$$R_{\text{sum}}(w) \approx h\left(n_t + (\rho - w)n_{t-1} + (\rho^2 - w\rho + w^2)n_{t-2}, n_{t-1} + (\rho - w)n_{t-2}\right) - \log(2\pi eD) \tag{16}$$
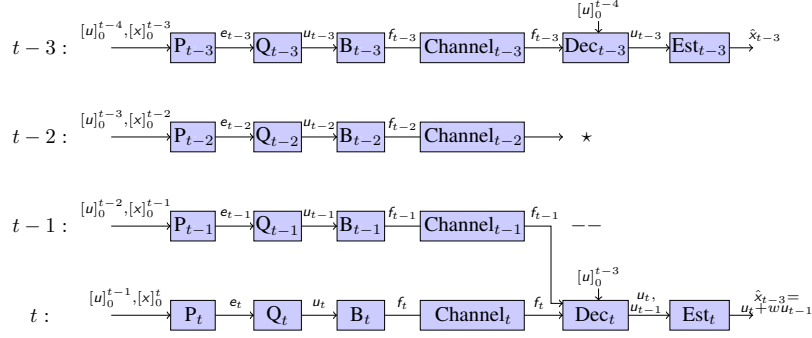
Fig. 4. Illustration of the hybrid coding scheme in Fig. 3. The notations P, Q and B represent the pre-processor, quantizer and binning blocks respectively.
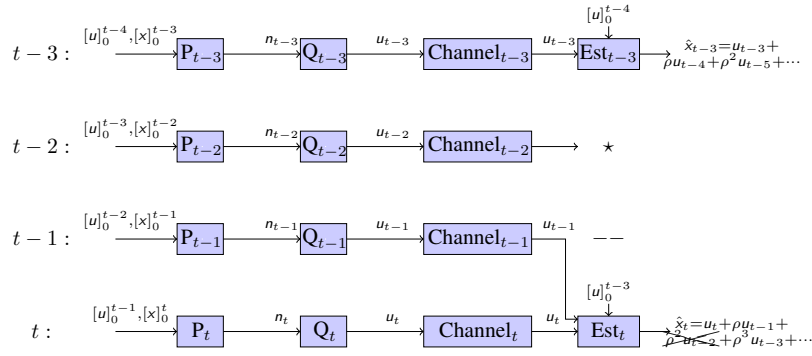

Fig. 5. Illustration of the predictive coding scheme in the high resolution regime, which fails in reconstructing x_t^n for the setup in Fig. 2. The notations P and Q represent the pre-processor and quantizer respectively.

Through simple algebra it can be shown that the first term in (16) is minimized when we select $w^\star = \frac{\rho}{1+\rho^2}$ and the associated rate satisfies:

$$\lim_{D \to 0} \left[R_{\text{sum}}(w^\star) - \log\left(\frac{1 - \rho^2}{D}\right) \right] = \frac{1}{2} \log\left(1 + \frac{\rho^2}{1 + \rho^4}\right), \tag{17}$$

where the term inside the bracket corresponds to excess rate over an erasure-free channel. In contrast the choice $w = 0$ corresponds to the memoryless quantization and binning, where the prediction loop in Fig. 3 is removed. It can be shown that in this case

$$\lim_{D \to 0} \left[R_{\text{sum}}(w = 0) - \log\left(\frac{1 - \rho^2}{D}\right) \right] = \frac{1}{2} \log\left(1 + \rho^2\right), \tag{18}$$

which is higher than the limit in (17). Another interesting choice is to set $w = \rho$, so that each term in (15) only consists of alternating noise symbols:

$$u_t = n_t + \rho^2 n_{t-2} + \rho^4 n_{t-4} + \ldots. \tag{19}$$

It can be shown that this choice leads to:

$$\lim_{D \to 0} \left[R_{\text{sum}}(w = \rho) - \log\left(\frac{1 - \rho^2}{D}\right) \right] = \frac{1}{2} \log\left(1 + \rho^4\right), \tag{20}$$

which improves upon Memoryless Quantize and Binning, but is still higher than (17).

B. Predictive Coding

A commonly used transmission technique for first order Gauss-Markov sources is predictive coding. The encoder generates an estimate of the source at time t, given all the past encoder outputs and subtracts it from the current source sample. The difference is quantized and transmitted over the channel. The setup is similar to Fig. 3, however the entire past $\mathbf{u}_1^{t-1}$ is used in the prediction loop and optimal MMSE estimation is applied when computing $\mathbf{e}_t$.

Unfortunately in the presence of packet losses over the channel, predictive coding can lead to a significant loss. To illustrate this effect we will consider the setup in Fig. 2, where a single loss happens at time $t-2$. We will also assume high resolution regime so that $x_t \approx \hat{x}_t$. In this regime we will have: $u_t \approx e_t \approx n_t$. Furthermore the reconstruction before time $t-2$ can be approximated as: $\hat{x}_k \approx x_k = n_k + \rho n_{k-1} + \rho^2 n_{k-2} + \cdots$. At time t, since u_{t-2} is not available to the decoder (see Fig 5), the reconstruction sequence is expressed as $\hat{x}_t \approx n_t + \rho n_{t-1} + \rho^3 n_{t-3} + \cdots$. Note that

$$x_t - \hat{x}_t \approx \rho^2 n_{t-2}.$$

Thus, the reconstruction at time t, as well as subsequent times cannot satisfy the high resolution constraint. More generally, such a mismatch between the encoder and decoder due to packet losses is referred to as prediction drift and can significantly degrade the performance in practice.

A commonly used modification to predictive coding is the so-called leaky predictive coding approach, see e.g., [3]–[5]. Such schemes allow the flexibility to sweep between the extremes of ideal prediction and no-prediction depending on the channel loss rate, typically by varying a single parameter. Our proposed hybrid scheme is different as it involves a truncated prediction approach based on limited past. We will discuss leaky predictive coding scheme in Section VI and provide numerical comparisons throughout.

IV. HYBRID CODING SCHEME

Our coding scheme consists of a prediction step followed by quantization-and-binning as shown in Fig. 3(a).

Codebook Generation:

- Fix the conditional multivariate Gaussian distribution $p([u]_1^T | [x]_1^T)$ such that for any $t \in [1 : T]$

$$u_t = x_t - \sum_{k=1}^{W} w_k u_{t-k} + z_t, \tag{21}$$

  where z_t is drawn i.i.d. from $\mathcal{N}(0, \sigma_z^2)$.[2]
- For each time $t \in [1 : T]$, randomly and independently generate $2^{n\tilde{R}_t}$ sequences $u_t^n(l_t)$ for $l_t \in [1 : 2^{n\tilde{R}_t}]$, each according to $\prod_{k=1}^{n} p_{u_t}(u_{t,k})$.
- For each time $t \in [1 : T]$, partition the set of indices $l_t \in [1 : 2^{n\tilde{R}_t}]$ into 2^{nR_t} equal-size bins with the bin index $f_t \in [1 : 2^{nR_t}]$.

The choice of the parameters σ_z^2, R_t and $\tilde{R}_t$ will be apparent in the analysis. The codebook is revealed to the encoder and decoder.

Encoding: The encoding consists of two steps:

Step 1 (Successive Quantization): The source sequences $\{x_1^n, x_2^n, \ldots, x_T^n\}$ are successively quantized into $\{u_1^n, u_2^n, \ldots, u_T^n\}$. In particular, the encoder, at each time t, identifies a quantization sequences u_t^n jointly typical [19] with $\{x_t^n, u_{t-W}^n, u_{t-W+1}^n, \ldots, u_{t-1}^n\}$ using the test channel (21).

Step 2 (Random Binning): The encoder, upon finding a particular quantization sequence u_t^n, sends its bin index $f_t \in [1 : 2^{nR_t}]$ through the channel.

Remark 5. *As sketched in Fig. 3(a), our proposed scheme consists of a prediction step followed by binning. Unlike DPCM, the predictor does not use all the past u_k^n, but only W most recent ones. These particular sequences are guaranteed to be available to the decoder since we assume that W sequences after the erasure burst are not supposed to be reconstructed. Thus there is no mismatch between the encoder and decoder in the prediction block. Furthermore since the codewords u_k^n are correlated, the binning block is used to exploit this correlation. In this paper we will refer to this approach of combining predictive coding and binning as* hybrid coding.

Decoding: The block diagram of the decoder is shown in Fig. 3(b). The decoder consists of two steps.

Step 1 (Decoding with Side-Information): The decoder, while receiving the channel outputs, applies the joint typicality decoding to recover the quantization codewords. Note that in this stage, all the previously recovered quantization sequences can be used by the decoder as the side information. In particular, for an erasure burst spanning $[j : j + B - 1]$, we have

$$\hat{u}_t^n = \begin{cases} \mathcal{T}_t([f]_1^t) & \text{for } t < j \\ \mathcal{T}_t([f]_1^{j-1}, [f]_{j+B}^{j+B+W}) & \text{for } j + B \leq t \leq j + B + W \\ \mathcal{T}_t([f]_1^{j-1}, [f]_{j+B}^t) & \text{for } t > j + B + W \end{cases},$$

where $\mathcal{T}_t(\cdot)$ denotes the typicality-based decoder [19] at time t.

Step 2 (Minimum Mean Squared Error (MMSE) Estimation): The decoder applies the MMSE estimation using all the recovered quantization sequences:

---

[2] u_i is zero for non-positive index i.

$$\hat{x}_t^n = \begin{cases} \mathcal{M}_t([\hat{u}^n]_1^t) & \text{for } t < j \\ \mathcal{M}_t([\hat{u}^n]_1^{j-1}, [\hat{u}^n]_{j+B}^t) & \text{for } t \geq j + B + W \end{cases},$$

where $\mathcal{M}_t(\cdot)$ denotes the MMSE estimator at the decoder at time t. This completes the description of the encoder and the decoder.

A. Analysis of Achievable Rate

We begin with a restatement of the test channel (21) used in the quantization step as follows:

$$\mathbf{A_w}[u]_1^T = [x]_1^T + [z]_1^T, \tag{22}$$

where $\mathbf{A_w}$ is a $T \times T$ lower triangular Topelitz matrix with (i,j)-th element $a_{i,j}$ defined as

$$a_{i,j} \triangleq \begin{cases} w_{i-j} & \text{if } 0 \leq i - j \leq W \\ 0 & \text{otherwise,} \end{cases}$$

where without loss of generality we assume $w_0 = 1$. Equivalently, (22) can be represented as

$$[u]_1^T = \mathbf{Q_w} \left([x]_1^T + [z]_1^T \right), \tag{23}$$

where $\mathbf{Q_w}$, i.e., the inverse of the matrix $\mathbf{A_w}$, is a $T \times T$ lower triangular matrix with (i,j)-th element $q_{i,j}$ as

$$q_{i,j} \triangleq \begin{cases} v_{i-j} & \text{if } i \geq j \\ 0 & \text{otherwise,} \end{cases}$$

where

$$v_k = -\sum_{j=0}^{k-1} w_{k-j} v_j. \tag{24}$$

In order to develop an expression for the achievable rate, we will focus on the steady state behavior. Assume that the erasure burst spans the interval $[t+1, t+B]$ for some $t \gg 1$. As will be shown later it suffices to focus on the reconstruction at time $t + B + W + 1$ as this term will dominate the rate expression. At this time the decoder has access to the following collection of variables $[u_1^t, u_{t+B+1}^{t+B+W+1}]$. In order to provide a computable expression for the achievable rate, we can simplify the dependence on these variables as follows:

First it turns out that for $t \gg 1$, the history u_1^t can be summarized using the scalar variable:

$$\tilde{s}_1 \triangleq s_1 + e, \tag{25}$$

where $e \sim \mathcal{N}\left(0, \Sigma(\sigma_z^2)/(1 - \Sigma(\sigma_z^2))\right)$ and

$$\Sigma(\sigma_z^2) \triangleq \frac{1}{2}\sqrt{(1 - \sigma_z^2)^2(1 - \rho^2)^2 + 4\sigma_z^2(1 - \rho^2)} + \frac{1 - \rho^2}{2}(1 - \sigma_z^2). \tag{26}$$

The term $\Sigma(\sigma_z^2)$ is the estimation error of s_t given the history u_1^{t-1} i.e.,

$$\Sigma(\sigma_z^2) = Var\left(s_t - E\{s_t | u_1^{t-1}\}\right)$$

in the steady state.

Next, the random variables $u_{t+B+1}^{t+B+W+1}$ can be replaced by their steady-state counterparts as:

$$\begin{pmatrix} \tilde{u}_{B+1} \\ \tilde{u}_{B+2} \\ \vdots \\ \tilde{u}_{B+W+1} \end{pmatrix} \triangleq \mathbf{Q}_{\text{eff}} \left(\begin{pmatrix} s_1 \\ s_2 \\ \vdots \\ s_{B+W+1} \end{pmatrix} + \begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_{B+W+1} \end{pmatrix} \right), \tag{27}$$

where $z_1, \ldots, z_{B+W+1}$ are drawn i.i.d. according to $\mathcal{N}(0, \sigma_z^2)$ and

$$\mathbf{Q}_{\text{eff}} \triangleq \begin{pmatrix} v_B & v_{B-1} & \cdots & 1 & 0 & \ldots & 0 \\ v_{B+1} & v_{B+2} & \cdots & v_1 & 1 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ v_{B+W} & v_{B+W-1} & \cdots & v_W & v_{W-1} & \ldots & 1 \end{pmatrix}. \tag{28}$$
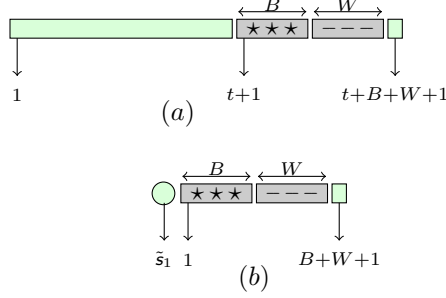
Fig. 6. Visualization of the results of Theorem 1: a) The actual worst-case burst erasure happens at steady state, i.e., $[t+1 : t+B]$ for $t \gg 1$. The decoder at time $t+B+W+1$ simultaneously recovers $\{u_{t+B+1}, \ldots, u_{t+B+W+1}\}$ and reconstruct $x_{t+B+W+1} \approx s_{t+B+W+1}$ from $\{u_1, \ldots, u_t\}$ and $\{u_{t+B+1}, \ldots, u_{t+B+W+1}\}$ b) Shows an equivalent formulation with the contribution of $[u]_1^t$ replaced by $\tilde{s}_1$ defined Theorem 1 and assuming steady state.

with the definition of v_i in (24). Note that $\mathbf{Q}_{\text{eff}}$ is a $(W+1) \times (B+W+1)$ matrix that contains rows $[B+1 : B+W+1]$ and columns $[1 : B+W+1]$ of $\mathbf{Q_w}$. [3]

The following theorem characterizes the achievable rate by the hybrid coding scheme.

Theorem 1. *For any given choice of $(\sigma_z^2, \mathbf{w})$, the rate R is achievable by the hybrid coding scheme provided that the constraint*

$$R \geq R_H(\sigma_z^2, \mathbf{w}) \triangleq \max_{\substack{\mathcal{M} \subseteq \mathcal{L} \\ \mathcal{M} \neq \phi}} \left\{ \frac{1}{|\mathcal{M}|} h\left([\tilde{u}]_{\mathcal{M}} | [\tilde{u}]_{\mathcal{M}^c}, \tilde{s}_1\right) \right\} - \frac{1}{2} \log(2\pi e \sigma_z^2), \tag{29}$$

is satisfied with the test channel noise σ_z^2 satisfying

$$\Delta_H(\sigma_z^2, \mathbf{w}) \triangleq Var(s_{B+W+1} | [\tilde{u}]_{B+1}^{B+W+1}, \tilde{s}_1) \leq D, \tag{30}$$

where $\mathcal{L} \triangleq \{B+1, \ldots, B+W+1\}$, and for any $\mathcal{M} \subseteq \mathcal{L}$, $\mathcal{M}^c$ represents complement of $\mathcal{M}$ with respect to $\mathcal{L}$.

$\square$

The proof of Theorem 1 is provided in Appendix B. The key step in our analysis involves showing that (i) the recovery window immediately after the erasure dominates the rate expression and (ii) the worst case erasure is in the steady state, i.e., as $t \to \infty$.

Since $\mathbf{w}$ and σ_z^2 are free parameters in Theorem 1, any rate $R \geq R_H^+(B, W, D)$ is achievable, where

$$R_H^+(B, W, D) = \min_{\substack{\mathbf{w}, \sigma_z^2 \\ \Delta_H(\sigma_z^2, \mathbf{w}) \leq D}} R_H(\sigma_z^2, \mathbf{w}). \tag{31}$$

B. Numerical Results

In this section we present numerical evaluation of the rate-recovery function.

Fig. 7 plots the excess rate function (c.f. Definition 1) for different coding schemes, as a function of ρ, the temporal correlation coefficient among the source symbols. Fig. 7(a) and Fig. 7(b) consider the case $D = 0.3$, and the remaining two consider the case $D = 0.6$. The baseline schemes and their achievable rate-recovery function are studied in detail in Section VI. Here we provide a brief overview of each scheme.

- **Predictive Coding:** The excess rate of predictive coding is represented by the dash-dot (red) line in Fig. 7. The overall distortion during reconstruction consists of two terms, i) the quantization noise of the current time and ii) the effect of erasure burst. To compensate for the latter, the quantization noise must be small enough. In general this requires considerably high rate, unless the value of ρ is small.
- **Leaky Prediction:** A simple modification of predictive coding scheme, known as *Leaky Predictive Coding*, makes the predictive coding scheme more robust to channel erasures. In this scheme the single prediction parameter is optimized to strike the trade-off between a predictive coding scheme and memoryless quantization scheme with no prediction at the encoder. The excess rate of leaky predictive coding in Fig. 7 shows performance improvements over predictive coding scheme.
- **Memoryless Quantization-and-Binning:** The excess rate of this scheme is represented by the dashed (blue) line. The encoder at each time quantizes the current source sequence in a memoryless fashion. This scheme is a special case of the

---

[3]Due to the structure of $\mathbf{Q_w}$, any set of rows $[t+B+1, t+B+W+1]$ and corresponding columns $[t+1 : t+B+W+1]$ of $\mathbf{Q_w}$ can be selected to generate $\mathbf{Q}_{\text{eff}}$.
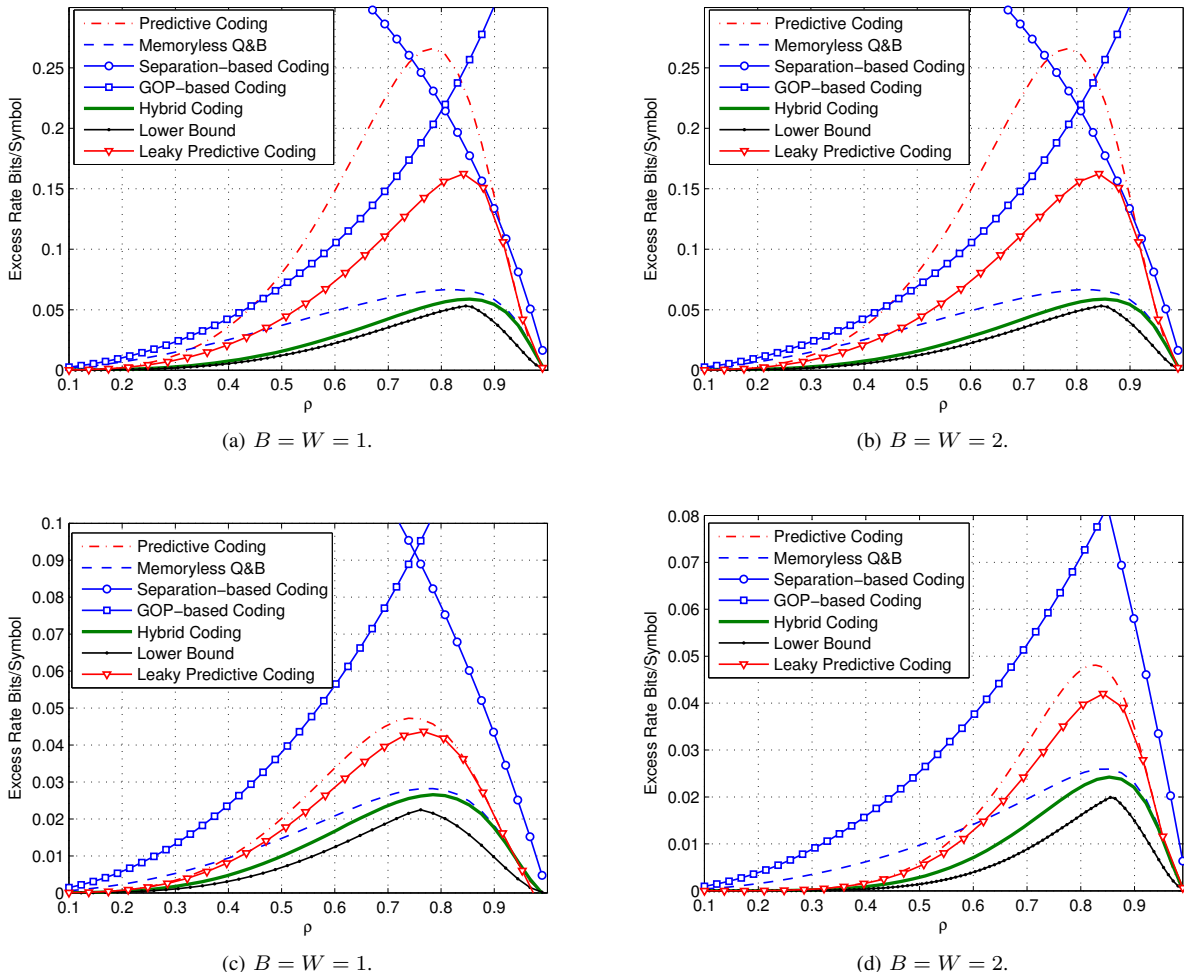
Fig. 7. Excess rates based on hybrid coding in comparison with the baseline schemes and the lower bound for (a) $B = W = 1$, $D = 0.3$ (b) $B = W = 2$, $D = 0.3$, (c) $B = W = 1$, $D = 0.6$ and (d) $B = W = 2$, $D = 0.6$.

hybrid coding scheme described in Section IV, with parameter $w_0 = 1$ and $w_k = 0$ for $k \in \{1, \ldots, W\}$. While the scheme outperforms the predictive coding except for small values of ρ, it is still far from the lower bound. The penalty of this scheme arises from the fact that it forces all the $W + 1$ quantized sources in the recovery period to be simultaneously recovered.

- **Hybrid Coding**: This scheme, illustrated by the solid green line, corresponds to the rate in Theorem 1.
- **Separation Based-Coding:** This scheme involves ideal prediction followed by an error correction code to recover from channel erasures. It also forces the recovery of all the source frames and is very inefficient in general, unless $\rho \approx 1$.
- **GOP-Based Coding:** In this approach, the source sequences occurring at multiples of $W + 1$ are encoded and decoded independently of the past. This corresponds to the presence of *I-frames* in video compression. The remaining sequences are compressed using predictive coding. From Fig. 7, the GOP-based coding scheme is also inefficient in general.
- **Lower Bound:** The lower bound on the excess rate, which is the lowermost curve in all plots in Fig. 7 is studied in Section V. It is numerically close to the hybrid-coding scheme.

C. High Resolution Regime

The achievable rate expression in Theorem 1 is rather complicated. In this section we consider the high resolution regime where the expression can be simplified in certain cases. We begin by stating optimal choice of the prediction coefficients w_k that minimize the sum rate constraint (i.e., $\mathcal{M} = \mathcal{L}$ in (29)) in the high resolution regime when $B = 1$.

Corollary 1. *In the high resolution regime, for $B = 1$ and any W, the excess sum-rate constraint of the hybrid coding scheme is minimized by:*

$$w_k^\star \triangleq \rho^k \frac{1 - \rho^{2(W-k+1)}}{1 - \rho^{2(W+1)}} \quad \text{for } k \in \{1, \ldots, W\}. \tag{32}$$

Furthermore, the achievable sum-rate by the hybrid coding scheme with this choice of coefficients coincides with the lower bound on the rate-recovery function (see Corollary 4) in the high resolution regime. □

The proof of Corollary 1 is presented in Appendix C.

Finding conditions under which the sum rate dominates (29) appears complicated in general. We have observed numerically that in the high resolution regime for $B = 1$ there exists a $\rho^\star \in (0, 1)$ such that for $\rho \geq \rho^\star$, the choice of (32) makes the sum-rate constraint dominant. Furthermore under this condition the resulting rate coincides with the lower bound in Section V, and hence is optimal. Table I shows some examples of $\rho^\star$. Unfortunately finding an analytical expression for $\rho^\star$ appears intractable at this point.

W	1	2	3	4
$\rho^\star$	0.9220	0.9604	0.9764	0.9845

TABLE I
NUMERICAL VALUES OF $\rho^\star$ FOR $W \in \{1, 2, 3, 4\}$.

The next two results pertain to upper bounds on the excess rate in the high resolution regime for $W = 1$ and any $\rho \in (0, 1)$.

Corollary 2. *In the high resolution regime, when $W = 1$ and $B = 1$, the excess rate of the hybrid coding scheme (See Definition 1), denoted by $R_{E,HR}(\rho, B = 1)$ is upper bounded as follows.*

$$R_{E,HR}(\rho, B = 1) \leq \frac{1}{4} \log \left(1 + \frac{2\rho^4}{(1 + \rho)^2} \right). \tag{33}$$

The proof of Corollary 2 is provided in Appendix D.

Corollary 3. *In the high resolution regime, when $W = 1$ and $B \to \infty$, we have*

$$R_{E,HR}(\rho, B \to \infty) = \frac{1}{4} \min_w \left\{ \log \left(f(w)^2 - g(w)^2 \right) \right\} \tag{34}$$

where

$$f(w) \triangleq \left(\frac{\rho^2}{1 - \rho^2} + \frac{1}{1 - w^2} \right) \frac{1}{(1 + w\rho)^2} \tag{35}$$

$$g(w) \triangleq \rho f(w) - \frac{w}{(1 + w\rho)(1 - w^2)}. \tag{36}$$

□

The proof of Corollary 3 is provided in Appendix E. The bounds presented in the above results are generally conservative. As illustrated in Fig. 8 the hybrid coding scheme performs very close to the lower bound for a wide range of ρ at least for $W = 1$ or $W = 2$. Fig. 8(a), (b) and (c) consider the examples of $B = W = 1$, $B = W = 2$ and $B \to \infty$, $W = 1$, respectively. The lower bound on the rate-recovery function and the baseline schemes will be discussed in the next two sections.

(a) $B = W = 1$. (b) $B = W = 2$. (c) $W = 1$, $B \to \infty$.
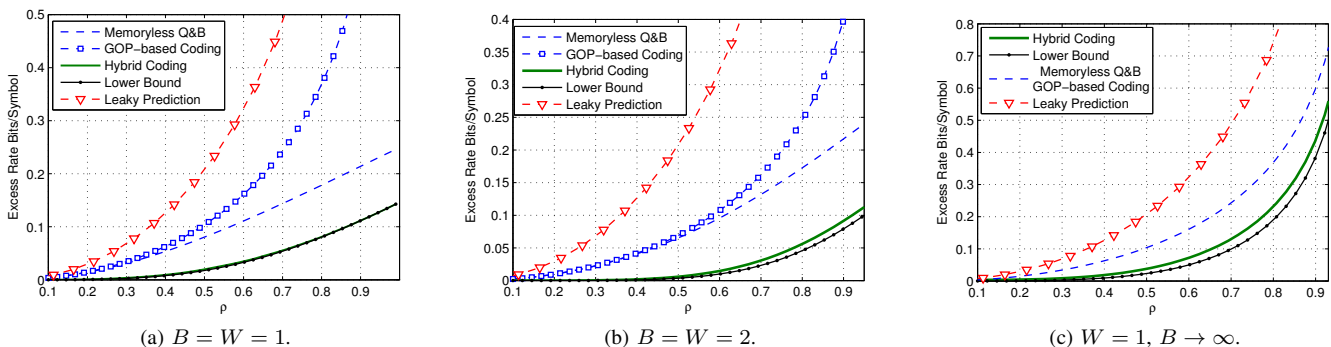
Fig. 8. Excess rates in high resolution regime based on hybrid coding in comparison with the baseline schemes and the lower bound, where (a) $B = W = 1$, (b) $B = W = 2$, and (c) $W = 1$, $B \to \infty$.

V. LOWER BOUND ON RATE-RECOVERY FUNCTION

A. Connection to the Network Source Coding Problem

Before stating the general lower bound on $R(B, W, D)$, we consider a special case of $B = W = 1$. For this case, we propose a lower bound by exploiting a connection between the streaming setup and the network source coding problem illustrated in
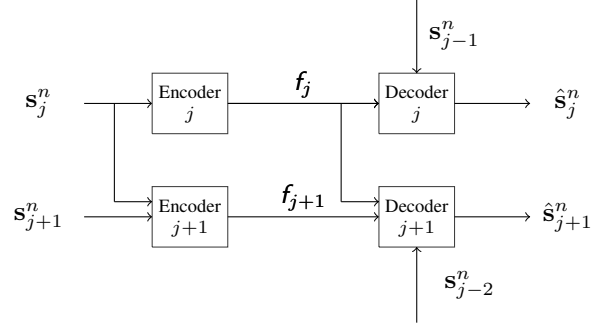
Fig. 9. A network source coding problem as an enhanced version of original streaming problem.

Fig. 9. Encoder 1 observes the source sequence s_j^n and Encoder 2 observes the two sources s_j^n and s_{j+1}^n. Decoder j is required to reconstruct s_j^n within distortion D while knowing s_{j-1}^n whereas decoder $j+1$ requires s_{j+1}^n within distortion D while knowing s_{j-2}^n and having access to the codewords $\{f_j, f_{j+1}\}$. Decoder j resembles a steady state decoder when the previous source sequence has been reconstructed whereas decoder $j+1$ resembles the decoder following an erasure and the associated recovery period. The proposed setting is slightly different from the original one in that the decoders are revealed actual source sequences rather than the past encoder outputs. Nevertheless the study of this model captures one source of tension inherent in the streaming setup. When encoding s_j^n we need to simultaneously satisfy two requirements: The sequence s_j^n must be reconstructed within a distortion of D at encoder j. It can also be used as a helper by decoder $j+1$. In general these requirements can be conflicting. If we set $s_{j-2}^n = \phi$ then the setup is reminiscent of zig-zag source coding problem [20].

We establish a lower bound on the sum-rate with symmetric rates R. In particular we show that for any $D \in (0, 1 - \rho^2)$ the inequality

$$2R \geq \frac{1}{2} \log \left(\frac{1 - \rho^6}{D} \right) + \frac{1}{2} \log \left(\frac{(1 - \rho^2)(1 - (1 - D)\rho^2)}{D(1 - \rho^4)} \right) \tag{37}$$

holds. To show (37), note that

$$\begin{aligned} 2nR &\geq H(f_j, f_{j+1}) \\ &\geq H(f_j, f_{j+1}|s_{j-2}^n) \\ &= I(f_j, f_{j+1}; s_{j+1}^n|s_{j-2}^n) + H(f_j, f_{j+1}|s_{j-2}^n, s_{j+1}^n) \\ &\geq h(s_{j+1}^n|s_{j-2}^n) - h(s_{j+1}^n|f_j, f_{j+1}, s_{j-2}^n) + H(f_j|s_{j-2}^n, s_{j+1}^n) \\ &\geq \frac{n}{2} \log \left(\frac{1 - \rho^6}{D} \right) + H(f_j|s_{j-2}^n, s_{j+1}^n), \end{aligned} \tag{38}$$

where (38) follows from the fact that s_{j+1}^n must be reconstructed from $(f_j, f_{j+1}, s_{j-2}^n)$ within distortion D at decoder $j+1$. The first term is the minimum rate associated with decoder $j+1$. We next lower bound the second term by using the fact that f_j must also be used by decoder j. In particular,

$$\begin{aligned} H(f_j|s_{j-2}^n, s_{j+1}^n) &\geq H(f_j|s_{j-2}^n, s_{j-1}^n, s_{j+1}^n) \\ &\geq I(f_j; s_j^n|s_{j-2}^n, s_{j-1}^n, s_{j+1}^n) \\ &= h(s_j^n|s_{j-1}^n, s_{j+1}^n) - h(s_j^n|s_{j-2}^n, s_{j-1}^n, s_{j+1}^n, f_j) \\ &= nh(s_1|s_0, s_2) - h(s_j^n|s_{j-2}^n, s_{j-1}^n, s_{j+1}^n, f_j) \\ &\geq \frac{n}{2} \log \left(2\pi e \frac{(1 - \rho^2)^2}{(1 - \rho^4)} \right) - h(s_j^n|s_{j-2}^n, s_{j-1}^n, s_{j+1}^n, f_j). \end{aligned} \tag{39}$$

One direct way to upper bound the last term in (39) is to use the fact that s_j can be reconstructed within distortion D using (f_j, s_{j-1}). Thus by ignoring the fact that s_{j+1} is also available, one can find the upper bound as

$$\begin{aligned} h(s_j^n|s_{j-2}^n, s_{j-1}^n, s_{j+1}^n, f_j) &\leq h(s_j^n|s_{j-1}^n, f_j) \\ &\leq \frac{n}{2} \log \left(2\pi e D \right). \end{aligned} \tag{40}$$

However knowing s_{j+1} can provide an extra observation to improve the estimation of s_j as well as the upper bound in (40). In particular, we can show that

$$h(s_j^n|s_{j-2}^n, s_{j-1}^n, s_{j+1}^n, f_j) \leq \frac{n}{2} \log \left(\frac{D(1 - \rho^2)}{1 - (1 - D)\rho^2} \right). \tag{41}$$

$$\Gamma(\vartheta, B, W) \triangleq \frac{1}{2(W+1)} \log \left(\frac{(1-(1-\vartheta)\rho^{2(B+W+1)})(1-(1-\vartheta)\rho^2)^{W+1}}{D\vartheta^W(1-(1-D)\rho^{2(W+1)})} \right) \tag{45}$$

Note that the upper bound in (41) is strictly tighter than (40), as the inequality

$$\frac{D(1-\rho^2)}{1-(1-D)\rho^2} \leq D,$$

always holds. To show (41), note that

$$
\begin{aligned}
&h(s_j^n|s_{j-2}^n, s_{j-1}^n, s_{j+1}^n, f_j) \\
&= h(s_j^n, s_{j+1}^n|s_{j-2}^n, s_{j-1}^n, f_j) - h(s_{j+1}^n|s_{j-2}^n, s_{j-1}^n, f_j) \\
&= h(s_j^n|s_{j-2}^n, s_{j-1}^n, f_j) - h(s_{j+1}^n|s_{j-2}^n, s_{j-1}^n, f_j) + h(s_{j+1}^n|s_j^n) \\
&= h(s_j^n|s_{j-2}^n, s_{j-1}^n, f_j) - h(s_{j+1}^n|s_{j-2}^n, s_{j-1}^n, f_j) + \frac{n}{2} \log\left(2\pi e(1-\rho^2)\right) \\
&\leq \frac{n}{2} \log\left(\frac{D}{1-(1-D)\rho^2}\right) + \frac{n}{2} \log\left(2\pi e(1-\rho^2)\right),
\end{aligned}
\tag{42}
$$

where the first term in (42) follows from the fact that at decoder j, s_j^n is reconstructed within distortion D knowing $\{s_{j-1}^n, f_j\}$ and hence

$$h(s_j^n|s_{j-2}^n, s_{j-1}^n, f_j) = h(s_j^n|s_{j-1}^n, f_j) \leq \frac{n}{2} \log(2\pi eD),$$

and using Lemma 1 stated below. By combining (39) and (42), we have

$$H(f_j|s_{j-2}^n, s_{j+1}^n) \geq \frac{n}{2} \log \left(\frac{(1-\rho^2)(1-(1-D)\rho^2)}{(1-\rho^4)D} \right). \tag{43}$$

Eq. (37) follows from (38) and (43).

Lemma 1. *Assume $s_a \sim N(0,1)$ and $s_b = \rho^m s_a + n$ for $n \sim N(0, 1-\rho^{2m})$. Also assume the Markov chain property $f_a \to s_a \to s_b$. If $h(s_a|f_a) \leq \frac{1}{2}\log(2\pi er)$, then*

$$h(s_a|f_a) - h(s_b|f_a) \leq \frac{1}{2} \log \left(\frac{r}{1-(1-r)\rho^{2m}} \right). \tag{44}$$

Proof. See Appendix F. □

In the following we present the general lower bound on the rate-recovery function.

B. General Lower Bound

In our original streaming setup, the bound derived in the previous section for the special case $B = W = 1$ can be tightened by noting that the side information to the decoders in Fig. 9 are actually encoder outputs rather than the true source sequences. The following theorem characterizes the general lower bound on the rate-recovery function.

Theorem 2. *The rate recovery function satisfies $R(B,W,D) \geq \Gamma(\vartheta, B, W)$ for some ϑ satisfying $\frac{1-\rho^2}{2^{2R}-\rho^2} \leq \vartheta \leq D$, where $\Gamma(\vartheta, B, W)$ is defined at the top of the next page.* □

The proof of Theorem 2 is provided in Appendix G. The lower bound in (45) can be viewed as summation of two terms as

$$(W+1)\Gamma(\vartheta, B, W) = \frac{1}{2} \log \left(\frac{(1-(1-\vartheta)\rho^{2(B+W+1)})}{D} \right) + \frac{1}{2} \log \left(\frac{(1-(1-\vartheta)\rho^2)^{W+1}}{\vartheta^W(1-(1-D)\rho^{2(W+1)})} \right), \tag{46}$$

where the terms in (46) are strengthened versions of (37).

The function $\Gamma(\vartheta, B, W)$ defined in (45) approaches to ∞ as $\vartheta \to 0$ and $\vartheta \to \infty$. It also has a global minimum at $\vartheta = \vartheta_m$. The expression (46) is however not an explicit lower bound on the rate as the interval for ϑ depends on the rate R. The lower bound in Theorem 2 is computed as follows. First Γ_λ is computed by solving the fixed point equation

$$\Gamma_\lambda = \Gamma \left(\frac{1-\rho^2}{2^{2\Gamma_\lambda}-\rho^2}, B, W \right).$$

Then the lower bound on $R^-(B,W,D)$ is computed as

$$R^-(B,W,D) \triangleq \begin{cases} \Gamma_\lambda & \text{if } \vartheta_m \leq \frac{1-\rho^2}{2^{2\Gamma_\lambda}-\rho^2} \leq D \\ \Gamma(\vartheta_m, B, W) & \text{if } \frac{1-\rho^2}{2^{2\Gamma_\lambda}-\rho^2} \leq \vartheta_m \leq D \\ \Gamma(D, B, W) & \text{if } \frac{1-\rho^2}{2^{2\Gamma_\lambda}-\rho^2} \leq D \leq \vartheta_m \end{cases}.$$
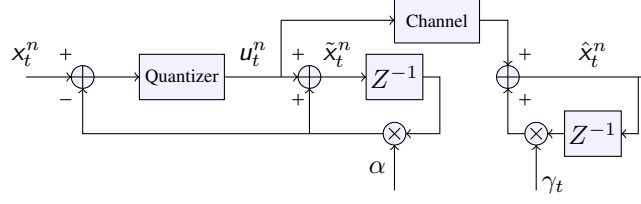
Fig. 10. Block diagram of Encoder and Decoder of Predictive Coding.

Corollary 4. *In the high resolution regime, the excess rate (see Def. 1) for any feasible scheme satisfies $R_E \geq R_E^-(B,W)$, where*

$$R_E^-(B,W) \triangleq \frac{1}{2(W+1)} \log \left(\frac{1 - \rho^{2(B+W+1)}}{1 - \rho^{2(W+1)}} \right). \tag{47}$$

Proof. In the high resolution $D \to 0$ and thus, $D \leq \vartheta_m$. Therefore the lower bound in Theorem 2 is computed as

$$R^-(B,W,D) = \Gamma(D,B,W).$$

The proof is completed by computing the limit $\lim_{D \to 0} \Gamma(D,B,W)$ and computing the excess rate. $\square$

VI. ANALYSIS OF BASELINE SCHEMES

In this section, we discuss the rate associated with various baseline schemes.

A. Predictive Coding

The block diagram of the encoder and decoder of the predictive coding is shown in Fig. 10 where $\alpha_{PC} = \gamma_{PC,t} = \rho$.
The following theorem characterizes the achievable rate of the predictive coding.

Theorem 3. *The predictive coding scheme achieves any rate $R \geq R_{PC}^+(B,W,D)$ for $D \geq \rho^{2(W+1)}(1-\rho^{2B})$ where,*

$$R_{PC}^+(B,W,D) \triangleq \frac{1}{2} \log \left(\frac{1 - \rho^{2(W+1)}(1-\rho^{2B}) - (1-D)\rho^2}{D - \rho^{2(W+1)}(1-\rho^{2B})} \right). \tag{48}$$

$\square$

The proof of Theorem 3 is presented in Appendix H.

Remark 6. *The minimum distortion achieved by predictive coding is $D_{\min}^\star = \rho^{2(W+1)}(1-\rho^{2B})$. This corresponds to the contribution of encoder outputs during the burst erasure. In particular predictive coding cannot be used in the high resolution when $D \to 0$.*

Leaky predictive coding is a modification of predictive coding with the similar encoder and decoder, shown in Fig. 10 with $\alpha_{LPC} \in [0,\rho]$ and

$$\gamma_{LPC,t} = \begin{cases} \alpha_{LPC}, & g_t = \star \\ \rho, & \text{otherwise} \end{cases}. \tag{49}$$

The encoder consists of
1) Calculating the prediction error of time t as

$$e_t = x_t - \alpha_{LPC}\tilde{x}_{t-1} \tag{50}$$

2) Quantizing the prediction error as $e_t = u_t + \tilde{z}_t$ with the rate $R = I(e_t; u_t)$ where $\tilde{z}_t$ is i.i.d. Gaussian noise.
3) Updating the encoders memory as $\tilde{x}_t = u_t + \alpha_{LPC}\tilde{x}_{t-1}$.
The decoder is able to reconstruct u_t from the latest channel output if and only if $g_t \neq \star$. It generates the source reconstruction at any time t as

$$\hat{x}_t = \begin{cases} u_t + \alpha_{LPC}\hat{x}_{t-1}, & g_t \neq \star \\ \rho\hat{x}_{t-1}, & g_t = \star \end{cases} \tag{51}$$

The leaky predictive coding is more robust to channel erasures than the predictive coding scheme. In high resolution excess rate, this scheme degenerates into to memoryless quantization with $\alpha_{LPC} = 0$. The excess rate can be shown to be:
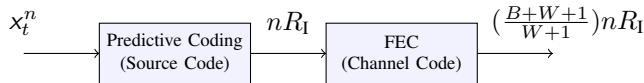
Fig. 11. Block diagram of Separation-Based Coding: A zero-delay predictive source code followed by a FEC channel code.

$$R_{\text{E,LPC}}^+ = \frac{1}{2} \log \left(\frac{1}{1-\rho^2} \right).$$

In this regime, the encoder of the leaky predictive coding scheme simply sends the quantized source sequences without any prediction at the encoder. As discussed next, the memoryless quantization-and-binning scheme will outperform this rate at high resolution regime via the binning step.

B. Memoryless Quantization-and-Binning

The memoryless quantization-and-binning scheme can be viewed as the special case of the hybrid coding scheme described in Section IV with

$$w_k = \begin{cases} 1 & k = 1 \\ 0 & \text{otherwise.} \end{cases}$$

Therefore the achievability result of Corollary 1 also holds for this scheme.

Corollary 5. *The memoryless quantization-and-binning coding scheme achieves the rate*

$$R_{QB}^+(B,W,D) = \min_{\sigma_z^2 : \Delta_{QB}(\sigma_z^2) \leq D} R_{QB}(\sigma_z^2) \tag{52}$$

where we define

$$R_{QB}(\sigma_z^2) \triangleq \frac{1}{W+1} h([\tilde{u}]_{B+1}^{B+W+1}|\tilde{s}_1) - \frac{1}{2} \log \left(2\pi e \sigma_z^2 \right), \tag{53}$$

$$\Delta_{QB}(\sigma_z^2) \triangleq Var(s_{B+W+1}|[\tilde{u}]_{B+1}^{B+W+1}, \tilde{s}_1), \tag{54}$$

where for any $\tilde{u}_i \triangleq s_i + z_i$ and z_i is sampled i.i.d. from $\mathcal{N}(0, \sigma_z^2)$. Also $\tilde{s}_1 \triangleq s_1 + e$ and $e \sim \mathcal{N}\left(0, \Sigma(\sigma_z^2)/(1 - \Sigma(\sigma_z^2))\right)$, with $\Sigma(\sigma_z^2)$ defined in (26). □

The proof of Corollary 5 is presented in Appendix I-A. The proof follows by specializing the proof of the hybrid coding scheme in Theorem 1 to the memoryless quantization-and-binning. In particular we show that the sum-rate constraint of simultaneously recovering $W+1$ sequences after the error propagation window, is the dominant constraint in this case.

The following corollary characterizes the high resolution performance of the memoryless quantization-and-binning.

Corollary 6. *In the high resolution regime, the excess rate of the memoryless quantization-and-binning scheme satisfies:*

$$\lim_{D \to 0} R_{E,QB}^+(B,W,D) = \lim_{D \to 0} \left(R_{QB}^+(B,W,D) - \frac{1}{2} \log \left(\frac{1-\rho^2}{D} + \rho^2 \right) \right)$$

$$= \frac{1}{2(W+1)} \log \left(\frac{1 - \rho^{2(B+1)}}{1 - \rho^2} \right). \tag{55}$$

□

The proof of Corollary 6 is presented in Appendix I-B. The proof is based on the observation that in the high resolution regime we have $u_t \approx x_t$. Therefore, the Markov chain property among the original source sequences also approximately holds for u_t.

C. Separation-Based Coding

The separation based coding scheme is based on the principle of using an optimal source code followed by an optimal channel code. Thus we apply ideal predictive coding for the source sequences, followed by a suitable forward-error-correcting (FEC) code to protect the channel packets from channel erasures. The following theorem characterizes the achievable rate.

Theorem 4. *The separation based coding scheme achieves any rate $R \geq R_{SC}^+(B,W,D)$ where*

$$R_{SC}^+(B,W,D) \triangleq \frac{B+W+1}{2(W+1)} \log \left(\frac{1-\rho^2}{D} + \rho^2 \right). \tag{56}$$

Coding Scheme	Memoryless Quantization-and-Binning	Leaky Predictive Coding	Hybrid Coding[4]	GOP Coding	Lower Bound
$B = W = 1$	$\frac{1}{4}\log\left(1+\rho^2\right)$	$\frac{1}{2}\log\left(\frac{1}{1-\rho^2}\right)$	$\leq \frac{1}{4}\log\left(1+\frac{2\rho^4}{(1+\rho)^2}\right)$	$\frac{1}{4}\log\left(\frac{1}{1-\rho^2}\right)$	$\frac{1}{4}\log\left(1+\frac{\rho^4}{1+\rho^2}\right)$
$B \to \infty$ $W = 1$	$\frac{1}{4}\log\left(\frac{1}{1-\rho^2}\right)$	$\frac{1}{2}\log\left(\frac{1}{1-\rho^2}\right)$	$\frac{1}{4}\log\left(f(w^\star)^2 - g(w^\star)^2\right)$	$\frac{1}{4}\log\left(\frac{1}{1-\rho^2}\right)$	$\frac{1}{4}\log\left(\frac{1}{1-\rho^4}\right)$

TABLE II

HIGH RESOLUTION EXCESS RATES FOR DIFFERENT CODING SCHEMES.

□

The block diagram of the coding scheme is shown in Fig.11. The n-length source vector at each time is first encoded via predictive coding scheme which results in nR_I bits per source vector. A rate-$(W+1)/(B+W+1)$ FEC code is then applied over the predictive encoder's outputs which enables the recovery of all $B+W+1$ codewords from $W+1$ available channel outputs after the erasure. In particular, the channel code consists of a $(B+W+1, W+1)$ Maximum Distance Separable (MDS) systematic convolutional code which is capable of correcting B erasures in arbitrary locations (including erasure bursts) [21]. The formal proof is omitted as it is rather straightforward. The excess rate of separation based coding scheme in high resolution is given by:

$$R_{E,SC}^+(B,W,D) = \lim_{D\to 0}\left(R_{SC}^+(B,W,D) - \frac{1}{2}\log\left(\frac{1-\rho^2}{D}+\rho^2\right)\right)$$
$$= \lim_{D\to 0}\frac{B}{2(W+1)}\log\left(\frac{1-\rho^2}{D}\right) = \infty. \tag{57}$$

D. GOP-Based Coding

In its simplest form a zero-delay GOP, used in many practical systems, contains the following structure:

- *I-frame* (intra coded frame) a picture that is coded independently of all other pictures. Each GOP begins (in decoding order) with this type of picture.
- *P-frame* (predictive coded frame) contains motion-compensated difference information relative to previously decoded pictures.

In order to control the error-propagation, the I-frames, which require higher rates, are transmitted periodically and the P-frames, with smaller rates, are transmitted in between. When the decoder fails in recovery of any frame during a GOP, the rest of the frames of that GOP are not recovered. However the decoder gets back to recovery of the frames only after the next I-frame.

In order to analyze the performance of the GOP based scheme in our framework, we assume that the source sequence occurring at times as multiples of $W+1$ are encoded as I-Frames using memoryless encoding and decoding. This guarantees that recover window at the destination will always be no greater than W. Predictive coding is applied to the remaining source sequences. The following theorem characterizes the achievable rate of such a scheme.

Theorem 5. *The GOP-based coding scheme achieves the average rate $\bar{R}_{GOP}(W,D)$ for any $B \geq 0$ where*

$$\bar{R}_{GOP}(W,D) = \frac{1}{2}\log\left(\frac{1-\rho^2}{D}+\rho^2\right) + \frac{1}{2(W+1)}\log\left(\frac{1}{1-(1-D)\rho^2}\right). \tag{58}$$

□

Remark 7. *The GOP-based coding scheme is a time-variant scheme and the rate associated with the I-frames and P-frames are not the same. In this paper we compare the average rate of the GOP-based scheme, rather than its peak rate, with other schemes.*

The proof of Theorem 5 is presented in Appendix J. It can be readily observed that in the high resolution regime when $D \to 0$, the average excess rate of the GOP-based scheme scales as

$$\lim_{D\to 0}\bar{R}_{E,GOP}(W,D) = \frac{1}{2(W+1)}\log\left(\frac{1}{1-\rho^2}\right). \tag{59}$$

Table II summarizes the high resolution results of the coding schemes discussed above. Note that the gap associated with predictive coding and separation based scheme is ∞ and hence not included in the comparison.

VII. MISMATCHED AND STATISTICAL CHANNELS

Although our analysis so far has focussed on the burst erasure channel, with a maximum burst-length of B, the coding schemes themselves are universal — following any erasure sequence, the decoder will resume zero-delay reconstruction after a sufficiently long recovery period. In this section we first discuss how such waiting time can be computed for the case of mismatched burst erasure channels. Thereafter we discuss simulations over statistical channels.

A. Mismatched Channel

In this section we investigate the performance of different coding schemes over mismatched channels. In particular consider a coding scheme that is designed for the parameters (B, W) where B is the channel erasure burst length and W is the waiting time after the erasure burst, before resuming zero-delay reconstruction of the source sequences. Assume that the underlying channel introduces an erasure burst of length β. Obviously if $\beta < B$, the decoder will have to wait for no more that W symbols after the erasure burst. Now consider the case $\beta > B$. As we discuss below, after a suitable waiting period, say $\Omega(\beta, R, D)$, the decoder will be able to resume zero-delay reconstruction.

1) Memoryless Quantization-and-Binning: For a rate R memoryless quantization-and-binning scheme, with test channel $u_t = s_t + z_t$, where $z_t \sim \mathcal{N}(0, \sigma_z^2)$, the minimum waiting period following an erasure burst of length β is given by:

$$\Omega_{\mathrm{QB}}(\beta, R, D) \triangleq \min_{\substack{r_{\mathrm{QB}}(\beta, W, \sigma_z^2) \leq R \\ d_{\mathrm{QB}}(\beta, W, \sigma_z^2) \leq D}} W,$$

where

$$r_{\mathrm{QB}}(\beta, W, \sigma_z^2) \triangleq \lim_{t \to \infty} \frac{1}{W+1} h([u]_{t-W}^t | u_1^{t-\beta-W-1}) - \frac{1}{2} \log(2\pi e \sigma_z^2),$$

and

$$d_{\mathrm{QB}}(\beta, W, \sigma_z^2) \triangleq \lim_{t \to \infty} \mathrm{Var}(s_t | u_1^{t-\beta-W-1}, [u]_{t-W}^t).$$

In particular the decoder will keep collecting the channel outputs until it accumulates sufficient rate to recover all the source sequences. We refer to Appendix I-A for details.

2) Hybrid Coding: By following the analysis similar to memoryless quantization-and-binning, the waiting time for the hybrid coding scheme can be expressed as:

$$\Omega_{\mathrm{H}}(\beta, R, D) \triangleq \min_{\substack{r_{\mathrm{H}}(\beta, W, \sigma_z^2) \leq R \\ d_{\mathrm{H}}(\beta, W, \sigma_z^2) \leq D}} W.$$

where the rate term $r_{\mathrm{H}}(\beta, W, \sigma_z^2)$ is equivalent to the rate expression in (29) of Theorem 1 for erasure burst of length β rather than B and

$$d_{\mathrm{H}}(\beta, W, \sigma_z^2) \triangleq \lim_{t \to \infty} \mathrm{Var}(s_t | u_1^{t-\beta-W-1}, [u]_{t-W}^t).$$

3) Predictive Coding: In the predictive coding, the waiting time is given by:

$$\Omega_{\mathrm{PC}}(\beta, R, D) \triangleq \min_{d_{\mathrm{PC}}(\beta, W, R) \leq D} W.$$

where

$$d_{\mathrm{PC}}(\beta, W, R) = \rho^{2(W+1)} \sigma_u^2 \sum_{l=0}^{\beta-1} \rho^{2l} + \tilde{\sigma}_z^2. \tag{60}$$

denotes the distortion at the destination in the first sequence to be recovered following the erasure burst. Through standard analysis it can be shown that $\tilde{\sigma}_z^2 = \frac{1-\rho^2}{2^{2R}-\rho^2}$ and $\sigma_u^2 = (1-\tilde{\sigma}_z^2)(1-\rho^2)$. Note that the first term in (60) decreases exponentially in W. Thus after a sufficiently long waiting time the effect of the erasure burst will become negligible and only the quantization noise will prevail.

Similar arguments of predictive coding can be applied to define $\Omega_{\mathrm{LPC}}(\beta, R, D)$ for the leaky predictive coding. As this scheme is not the main focus of this paper we only numerically calculate the waiting time function for this scheme.

4) GOP-based Coding: By fixing the operational rate to be R, the period of I-frame transmissions, i.e., $W^\star + 1$, is specified by solving

$$\bar{R}_{\mathrm{GOP}}(W^\star, D) = R,$$

where $\bar{R}_{\mathrm{GOP}}(W, D)$ is defined in (58). In this scheme whenever a burst of channel packets are erased, the decoder declares loss up to the time of the next non-erased I-frame, when the decoder gets back to recovery. Therefore $\Omega_{\mathrm{GOP}}(\beta, R, D)$ always belong to the set $\{0, 1, \ldots, W^\star\}$ with equal probability depending on the time when the erasure burst ends. For instance, if the erasure ends right before the I-frame, the waiting time will be 0, as the next frame will be recovered immediately. If the erasure burst ends while at the I-frame, the decoder need to wait for the entire period. Thus the average waiting time for this scheme is

$$\Omega_{\mathrm{GOP}}(\beta, R, D) = \frac{W^\star}{2}.$$

Note that for the GOP-based scheme, unlike the other schemes, the average rate is considered rather than the peak rate.

5) Lower Bound: In order to derive the lower bound on the waiting time for any burst length, we invoke the lower bound on the rate-recovery function (see Section V) to find the minimum required W for any burst length β and fixed operational rate R. In particular, the minimum possible waiting time after an erasure of length β, has to satisfy $R^-(\beta, W, D) \leq R$. It gives a lower bound on the waiting function as

$$\Omega^-(\beta, R, D) \triangleq \min_{R^-(\beta, W, D) \leq R} W,$$

where $R^-(\beta, W, D)$ is the lower bound on the rate recovery function derived in Theorem 2.

Burst Length	Memoryless Q&B	Predictive Coding GOP-based Coding	Leaky Predictive Coding	Hybrid Coding	Lower Bound
$\beta = 1$	4	3	2	1	1
$\beta \geq 2$	5	3	3	2	1

TABLE III

WAITING TIME $\Omega(\beta, R, D)$ FOR DIFFERENT ERASURE BURST LENGTHS β AND $R = 1.50$, $D = 0.1$ AND $\rho = 0.5$.

Burst Length	Memoryless Q&B	GOP-based Coding	Predictive Coding /Leaky Predictive Coding	Hybrid Coding	Lower Bound
$\beta = 1$	11	14.5	10 / 8	3	3
$\beta = 2$	17	14.5	11 / 9	4	3
$\beta = 3$	20	14.5	11 / 9	6	4
$\beta = 4$	22	14.5	12 / 9	8	4
$\beta = 5$	24	14.5	12 / 9	8	4
$\beta = 6$	24	14.5	12 / 9	9	4
$\beta \in [7 : 12]$	25	14.5	12 / 9	9	4
$\beta \in [13 : 14]$	26	14.5	12 / 9	9	4
$\beta \geq 15$	26	14.5	12 / 10	9	4

TABLE IV

WAITING TIME $\Omega(\beta, R, D)$ FOR DIFFERENT ERASURE BURST LENGTHS β AND $R = 1.063$, $D = 0.1$ AND $\rho = 0.8$

Tables III and IV illustrate $\Omega(\beta, R, D)$ as a function of burst length β, when the average distortion is $D = 0.1$, and correlation coefficients are $\rho = 0.5$ and $\rho = 0.8$, respectively. The communication rate is selected to be

$$R = 1.02 R_{\mathrm{I}} = 1.02 \times \frac{1}{2} \log \left(\frac{1 - \rho^2}{D} + \rho^2 \right),$$

which is a 2% rate-overhead over the ideal channel rate. We note that for $\beta = 1$, the hybrid coding scheme attains a wait time that coincides with the lower bound. This observation will be useful when we discuss simulations over i.i.d. erasure channels, where the isolated erasure can be dominant error patterns.

B. Statistical Channels

In this section we apply the coding schemes designed in the previous section over different statistical channel models.

1) Channel with i.i.d. erasures: We first consider the case of i.i.d. erasures. The channel at each time may introduce an erasure with probability of γ. The decoder declares a source sequence s_t^n as lost if it cannot reproduce it at time t within required distortion D. Naturally due to the zero delay constraint, all the source sequences corresponding to the channel erasures will have to be lost for any coding scheme. The probability of *excess* loss is the fraction of source sequences corresponding to non-erasures that are also declared as lost.

When the erasures are sparse, so that the effect of each erasure is locally isolated, we can approximate the excess loss rate P_{EL} using the function $\Omega(\beta, R, D)$ with $\beta = 1$. In particular

$$P_{\mathrm{EL}} \approx \gamma \cdot \Omega(1, R, D).$$

Fig. 13(a) and Fig. 14(a) illustrate the simulated excess loss rate of different schemes over i.i.d. erasure channel as a function of erasure probability for the case of $\rho = 0.5$ and $\rho = 0.8$, respectively, $D = 0.1$ and a 2% rate overhead. It can be observed that the hybrid coding scheme outperforms the other schemes. From Table III it can be observed that, for isolated erasure, i.e, $\beta = 1$, the hybrid coding scheme attains the same waiting time as the lower bound, i.e., $\Omega_{\mathrm{H}}(1, R, D) = \Omega^-(1, R, D)$. This is the main reason that the hybrid coding scheme performs very close to optimal in this scenario. The predictive coding and GOP-based coding schemes require three times longer waiting time and the memoryless quantization-and-binning requires four time longer waiting time after an isolated erasure. Note that as γ, i.e, the packet erasure probability, increases, the chance of consecutive erasures increases. This is the reason that the hybrid coding shows a loss probability slightly higher than the lower bound as γ increases.
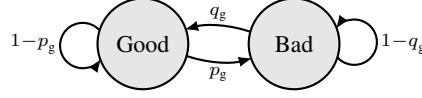
Fig. 12. Gilbert channel model: In the "Good" state, the channel perfectly reveals the encoder output to the decoder, while in the "Bad" state, the channel erases the decoder output.

2) Gilbert-Elliott Channel Model: We further consider the two-state Gilbert-Elliott channel model [22]–[24] (Fig. 12) in which all channel packets are erased in "bad state". In the "good state" a packet is erased with a probability of ϵ. Let p_g and q_g denote the probability of transition from "good" to "bad" state and vice versa. In steady state, the probability of being in "bad state" and thus the erasure probability is $p_g/(p_g + q_g)$. It is not hard to verify that the mean burst length is equal to $1/q_g$.
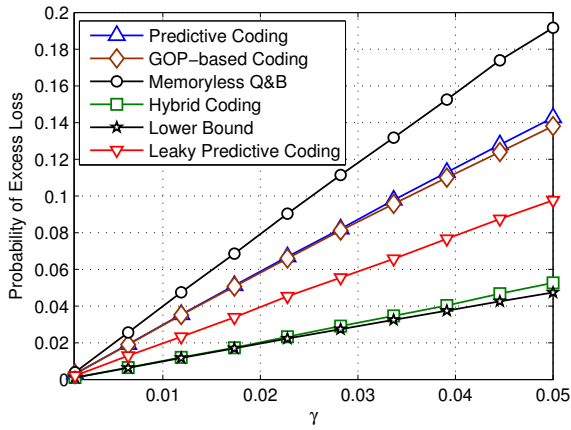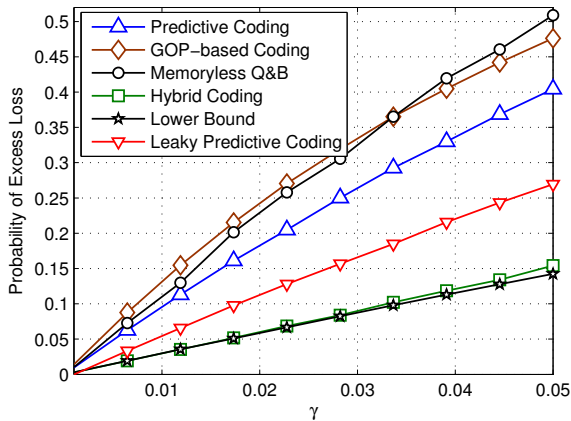

(a) Probability of excess loss versus probability of erasure (γ) of a Channel with independent erasures.
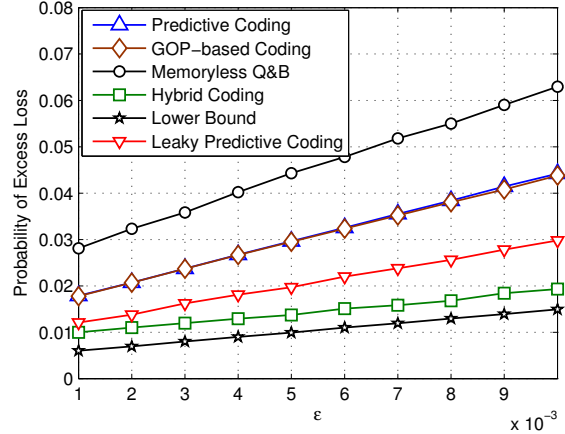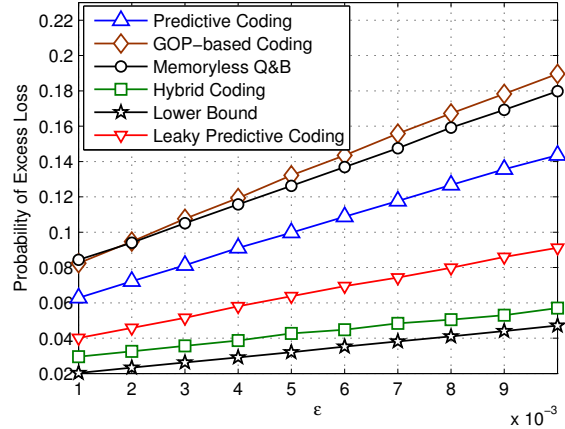
(b) Probability of excess loss versus probability of erasure in good states (ϵ) for Gilbert-Elliott Channel with $p_G = 5 \times 10^{-3}$, $q_G = \frac{1}{3}$.

Fig. 13. Comparison of different schemes $\rho = 0.5$, $D = 0.1$, and $R = 1.02R_I$

(a) Probability of excess loss versus probability of erasure (γ) of a Channel with independent erasures.

(b) Probability of excess loss versus probability of erasure in good states (ϵ) for Gilbert-Elliott Channel with $p_G = 5 \times 10^{-3}$, $q_G = \frac{1}{3}$.

Fig. 14. Comparison of different schemes $\rho = 0.8$, $D = 0.1$, and $R = 1.02R_I$.

Knowing that the function $\Omega(\beta, R, D)$ characterizes the waiting time for erasure bursts of different lengths, and assuming that the consecutive bursts are separated with long enough guard interval, the probability of excess loss, P_{EL} can be approximated as

$$P_{EL} \approx \epsilon \cdot \Omega(1, R) + p_g q_g \sum_{\beta=1}^{\infty} (1 - q_g)^{\beta-1} \cdot \Omega(\beta, R, D).$$

In particular the first term is the average number of losses resulted by isolated erasures in good state and the second term is the summation of average number of losses resulted by burst erasures of length β in bad state, for different burst lengths $\beta \in [1 : \infty)$.

In Fig. 13(b) and Fig. 14(b), we illustrate the performance of different schemes for a Gauss-Markov source with $\rho = 0.5$ and $\rho = 0.8$ respectively for average distortion $D = 0.1$ and 2% rate overhead over the Gilbert-Elliott channel. In both plots we fixed the channel parameters $(p_{\mathrm{g}}, q_{\mathrm{g}}) = (5 \times 10^{-3}, \frac{1}{3})$. The performance of different schemes are illustrated as a function of the parameter ϵ. We verify again that the hybrid coding scheme provides significant gains over other baseline schemes.

VIII. CONCLUSIONS

We study real-time coding of Gauss-Markov sources under a zero decoding delay constraint over an erasure channel. We first consider a simplified information theoretic model involving a burst erasure channel and define the rate-recovery function in this setup. We explain the weakness in traditional schemes such as predictive coding and memoryless quantization-and-binning in this setup. We propose a new coding scheme that can be viewed as a hybrid of these schemes that achieves significant gains. We also develop a lower bound on the rate-recovery function and show that in certain regimes our proposed scheme can be close to optimal. Several numerical computations and simulations are presented to illustrate the performance gains of our proposed technique over baseline schemes in a variety of channel models.

Our present setup requires that every sequence outside of the error propagation period be recovered with distortion D, while sequences inside this period need not be recovered. One natural extension is that the sequences within the error propagation period must be recovered, albeit with a higher distortion. Another extension is to relax the zero decoding delay constraint at the receiver, and allow a non-zero decoding delay. Some effort in that direction is presented in [14], [16] however a more complete characterization remains to be developed.

APPENDIX A
SUM-RATE OF THE CASE $B = W = 1$ IN THE HIGH RESOLUTION REGIME

In this section we establish the rate expressions stated in Section III. In all of our analysis we start with the sum-rate expression (13).

A. Memoryless quantization-and-binning

Since $u_t = x_t + z_t$ we have that:

$$
\begin{aligned}
\lim_{D \to 0} \{R_{\mathrm{sum}} + \log(2\pi e D)\} &= \lim_{D \to 0} \quad h\left(\begin{pmatrix} n_{t-1} + \rho n_{t-2} + \cdots + \rho^{t-2} n_1 + z_{t-1} \\ n_t + \rho n_{t-1} + \rho^2 n_{t-2} + \cdots + \rho^{t-1} n_1 + z_t \end{pmatrix} \Big| [n]_1^{t-3} \right) \\
&= \lim_{D \to 0} h\left(\begin{pmatrix} n_{t-1} + \rho n_{t-2} + z_{t-1} \\ n_t + \rho n_{t-1} + \rho^2 n_{t-2} + z_t \end{pmatrix} \right) \\
&= \lim_{D \to 0} \{h(n_{t-1} + \rho n_{t-2} + z_{t-1}) + h(n_t + \rho n_{t-1} + \rho^2 n_{t-2} + z_t | n_{t-1} + \rho n_{t-2} + z_{t-1})\} \\
&= \left\{ \frac{1}{2} \log(2\pi e(1 - \rho^2)(1 + \rho^2)) + \frac{1}{2} \log(2\pi e(1 - \rho^2)) \right\},
\end{aligned}
$$

where we applied the fact that $n_t \sim \mathcal{N}(0, 1 - \rho^2)$ and $z_t, z_{t-1} \sim \mathcal{N}(0, \sigma_z^2)$ with $\sigma_z \to 0$ as $D \to 0$. We have

$$
\lim_{D \to 0} \left\{ R_{\mathrm{sum}} - \log\left(\frac{1 - \rho^2}{D} \right) \right\} = \frac{1}{2} \log(1 + \rho^2),
$$

as required.

B. Imperfect Prediction-and-Binning

We have

$$
\begin{aligned}
\lim_{D \to 0} \{R_{\mathrm{sum}} + \log(2\pi e D)\} &= \lim_{D \to 0} h(u_{t-1}, u_t | [n]_1^{t-3}) \\
&= \lim_{D \to 0} h\left(\begin{pmatrix} n_{t-1} + \rho^2 n_{t-3} + \cdots + z_{t-1} - \rho z_{t-2} + \rho^2 z_{t-3} - \cdots \\ n_t + \rho^2 n_{t-2} + \cdots + z_t - \rho z_{t-1} + \rho^2 z_{t-2} - \cdots \end{pmatrix} \Big| [n]_1^{t-3} \right) \\
&= \lim_{D \to 0} \{h(n_{t-1}) + h(n_t + \rho^2 n_{t-2} | n_{t-1})\} \\
&= \frac{1}{2} \log(2\pi e(1 - \rho^2)) + \frac{1}{2} \log(2\pi e(1 + \rho^4)(1 - \rho^2)).
\end{aligned}
\tag{61}
$$

Finally (61) reduces to

$$
\lim_{D \to 0} \left\{ R_{\mathrm{sum}} - \log\left(\frac{1 - \rho^2}{D} \right) \right\} = \frac{1}{2} \log(1 + \rho^4),
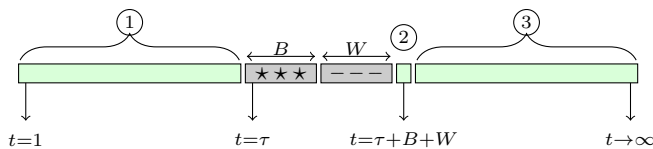$$

as required.

Fig. 15. Burst erasure channel model parametrized by starting time of erasure, τ.

APPENDIX B
ACHIEVABILITY OF HYBRID CODING: PROOF OF THEOREM 1

A. General Rate and Distortion Constraints

In order to study the rate of the hybrid coding scheme, consider the channel with an erasure burst spanning $[\tau : \tau + B - 1]$. Fig. 15 illustrates a burst erasure channel model parametrized by τ, i.e., the time where the burst erasure of length B starts. We identify three different time regions.

- *Region 1*: $t < \tau$, where there is no previous erasure by the channel. The decoder recovers $\mathbf{u}_t$ given $\{\mathbf{u}_1, \ldots, \mathbf{u}_{t-1}\}$. This succeeds with high probability if

$$R \geq R_{1,\tau}(t, \sigma_z^2) \triangleq h(u_t | [u]_1^{t-1}) - \frac{1}{2} \log \left(2\pi e \sigma_z^2 \right). \tag{62}$$

Furthermore, the decoder reconstructs the source sequence $\mathbf{x}_t$ within the distortion

$$D_{1,\tau}(t, \sigma_z^2) \triangleq \mathrm{Var}(x_t | [u]_1^t). \tag{63}$$

- *Region 2*: $t = \tau + B + W$, right after the erasure burst of length B spanning $[\tau, \tau + B - 1]$ and a window of length W after that. The decoder simultaneously recovers all the codewords $[\mathbf{u}]_{t-W}^t$ given $\{\mathbf{u}_1, \ldots, \mathbf{u}_{\tau-1}\}$. This succeeds with high probability if

$$R \geq R_{2,\tau}(\sigma_z^2)$$
$$\triangleq \max_{\substack{\mathcal{M} \subseteq \mathcal{L}_\tau \\ \mathcal{M} \neq \phi}} \frac{1}{|\mathcal{M}|} h([u]_{\mathcal{M}} | [u]_1^{\tau-1}, [u]_{\mathcal{M}^c}) - \frac{1}{2} \log \left(2\pi e \sigma_z^2 \right), \tag{64}$$

where

$$\mathcal{L}_\tau \triangleq \{\tau + B, \ldots, \tau + B + W\} \tag{65}$$

and $\mathcal{M}^c$ denotes the compliment of $\mathcal{M}$ with respect to the set $\mathcal{L}_\tau$. Furthermore, the decoder reconstructs the source sequence $\mathbf{x}_{\tau+B+W}$ within the distortion

$$D_{2,\tau}(\sigma_z^2) \triangleq \mathrm{Var}(x_{\tau+B+W} | [u]_1^{\tau-1}, [u]_{\tau+B}^{\tau+B+W}).$$

- *Region 3*: $t > \tau + B + W$, the time after Region 2. The decoder recovers $\mathbf{u}_t$ given

$$\{\mathbf{u}_1, \ldots, \mathbf{u}_{\tau-1}, \mathbf{u}_{\tau+B+W}, \ldots, \mathbf{u}_{t-1}\}.$$

This succeeds with high probability if

$$R \geq R_{3,\tau}(t, \sigma_z^2) \triangleq h(u_t | [u]_1^{\tau-1}, [u]_{\tau+B}^{t-1}) - \frac{1}{2} \log \left(2\pi e \sigma_z^2 \right). \tag{66}$$

Furthermore, the decoder reconstructs the source sequence $\mathbf{x}_t$ within the distortion

$$D_{3,\tau}(t, \sigma_z^2) \triangleq \mathrm{Var}(x_t | [u]_1^{\tau-1}, [u]_{\tau+B}^t).$$

For any parameter τ, define

$$R_\tau(t, \sigma_z^2) \triangleq \begin{cases} R_{1,\tau}(t, \sigma_z^2), & t < \tau \\ R_{2,\tau}(\sigma_z^2), & t \in [\tau, \tau + B + W] \\ R_{3,\tau}(t, \sigma_z^2), & t > \tau + B + W, \end{cases} \tag{67}$$

$$D_\tau(t, \sigma_z^2) \triangleq \begin{cases} D_{1,\tau}(t, \sigma_z^2), & t < \tau \\ D_{2,\tau}(\sigma_z^2), & t \in [\tau, \tau + B + W] \\ D_{3,\tau}(t, \sigma_z^2), & t > \tau + B + W. \end{cases} \tag{68}$$

The rate and distortion constraints have to be satisfied for all possible parameters τ. In particular, the rate

$$R \geq \max_{\tau \in [1:T-B]} \max_t R_\tau(t, \sigma_z^2),$$

is achievable, for any test channel noise satisfying

$$\max_{\tau \in [1:T-B]} \max_t D_\tau(t, \sigma_z^2) \leq D.$$

B. Worst-case Characterization of Burst Erasure

In this section we show that for any test channel noise σ_z^2, the worst-case rate constraint (67), is

$$\sup_{\tau, t} R_\tau(t, \sigma_z^2) = \lim_{\tau \to \infty} R_{2, \tau}(\sigma_z^2)$$

$$= \lim_{\tau \to \infty} \max_{\substack{\mathcal{M} \subseteq \mathcal{L}_\tau \\ \mathcal{M} \neq \phi}} \frac{1}{|\mathcal{M}|} h([u]_\mathcal{M}|[u]_1^{\tau-1}, [u]_{\mathcal{M}^c}) - \frac{1}{2} \log(2\pi e \sigma_z^2), \tag{69}$$

where $\mathcal{L}_\tau$ is defined in (65). In addition the test channel noise σ_z^2 has to satisfy the worst-case distortion constraint

$$\sup_{\tau, t} D_\tau(t, \sigma_z^2) = \lim_{\tau \to \infty} D_{2, \tau}(\sigma_z^2)$$

$$= \lim_{\tau \to \infty} \mathrm{Var}(x_{\tau+B+W}|[u]_1^{\tau-1}, [u]_{\tau+B}^{\tau+B+W}) \leq D. \tag{70}$$

Before the proof consider the following lemmas.

Lemma 2. *(Time-Shifting Lemma) For the hybrid coding scheme and for any $k < t$,*

$$h(x_t|[u]_\mathcal{M}, [x]_1^k, [z]_1^k) = h(x_{t-k}|[u]_{\mathcal{M}-k})$$
$$\text{for} \quad \mathcal{M} \subseteq \{1, 2, , \ldots, t\} \tag{71}$$

$$h(u_t|[u]_\mathcal{M}, [x]_1^k, [z]_1^k) = h(u_{t-k}|[u]_{\mathcal{M}-k})$$
$$\text{for} \quad \mathcal{M} \subseteq \{1, 2, , \ldots, t-1\}, \tag{72}$$

where $\mathcal{M} - k \triangleq \{m - k | m \in \mathcal{M}, m > k\}$.

Remark 8. *Similar equality holds for estimation error function rather than differential entropy. In particular,*

$$\mathrm{Var}(x_t|[u]_\mathcal{M}, [x]_1^k, [z]_1^k) = \mathrm{Var}(x_{t-k}|[u]_{\mathcal{M}-k})$$
$$\text{for} \quad \mathcal{M} \subseteq \{1, 2, , \ldots, t-1\}. \tag{73}$$

This follows from the fact that for a Gaussian variable X we have $h(X) = \frac{1}{2} \log(2\pi e \mathrm{Var}(X))$.

Proof. First consider (71) and note that for any $k < j \leq t$, we have

$$x_j = \rho^{j-k} x_k + \sum_{l=k+1}^{j} \rho^{j-l} n_l.$$

Now for any $\mathcal{M} \subseteq \{1, 2, , \ldots, t\}$ we have

$$h(x_t|[u]_\mathcal{M}, [x]_1^k, [z]_1^k) = h\left(\rho^{t-k} x_k + \sum_{l=k+1}^{t} \rho^{t-l} n_l \left| \left\{ \sum_{l=1}^{j} q_{j,l}(x_l + z_l) \right\}_{j \in \mathcal{M}, j > k}, \{u_j\}_{j \in \mathcal{M}, j \leq k}, [x]_1^k, [z]_1^k \right. \right)$$

$$= h\left(\sum_{l=k+1}^{t} \rho^{t-l} n_l \left| \left\{ \sum_{l=k+1}^{j} q_{j,l}(x_l + z_l) \right\}_{j \in \mathcal{M}, j > k}, [x]_1^k, [z]_1^k \right. \right) \tag{74}$$

$$= h\left(\sum_{l=1}^{t-k} \rho^{t-k-l} n_l \left| \left\{ \sum_{l=1}^{j} q_{j,l}(x_l + z_l) \right\}_{j \in \mathcal{M}-k} \right. \right)$$

$$= h(x_{t-k}|[u]_{\mathcal{M}-k}),$$

where (74) follows from the fact that u_j for $j \leq k$ are function of $\{[x]_1^k, [z]_1^k\}$. Also (72) can be verified using similar methods. $\square$

Lemma 3. *For any test channel noise σ_z^2 and any τ and t, we have the inequalities*

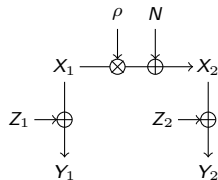$$R_{\tau+1}(t+1, \sigma_z^2) \geq R_\tau(t, \sigma_z^2) \tag{75}$$

Fig. 16. Schematic of the variables in Lemma 4.

$$D_{\tau+1}(t+1, \sigma_z^2) \geq D_\tau(t, \sigma_z^2). \tag{76}$$

Proof. First consider the rate inequality in (75). It suffices to show that the inequality holds for all the rate expressions in (67). For instance using (62) for $R_{1,\tau}(t, \sigma_z^2)$, we have

$$R_{1,\tau+1}(t+1, \sigma_z^2) \triangleq h(u_{t+1}|[u]_1^t) - \frac{1}{2}\log\left(2\pi e \sigma_z^2\right)$$

$$\geq h(u_{t+1}|[u]_1^t, x_1, z_1) - \frac{1}{2}\log\left(2\pi e \sigma_z^2\right) \tag{77}$$

$$= h(u_t|[u]_1^{t-1}) - \frac{1}{2}\log\left(2\pi e \sigma_z^2\right) \triangleq R_{1,\tau}(t, \sigma_z^2) \tag{78}$$

where (77) follows from the fact that conditioning reduces the differential entropy, and (78) follows from the application of second equality in Lemma 2 at time $t+1$ for $\mathcal{M} = \{1, \ldots, t\}$ and $k = 1$. The similar inequalities can be derived for $R_{2,\tau}(\sigma_z^2)$ and $R_{3,\tau}(t, \sigma_z^2)$. This verifies (75).

The same method can be applied for the distortion constraints to show (76). For example for $D_{1,\tau+1}(t+1, \sigma_z^2)$ from (63) we have

$$D_{1,\tau+1}(t+1, \sigma_z^2) \triangleq \text{Var}(x_{t+1}|[u]_1^{t+1})$$

$$\geq \text{Var}(x_{t+1}|[u]_1^{t+1}, x_1) \tag{79}$$

$$= \text{Var}(x_t|[u]_1^t) \triangleq D_{1,\tau}(t, \sigma_z^2) \tag{80}$$

where (79) follows from the fact that revealing the additional information x_1 can only reduces the distortion and (80) follows from Remark 8. Similar inequalities can be derived for $D_{2,\tau}(\sigma_z^2)$ and $D_{3,\tau}(t, \sigma_z^2)$. This verifies (76). $\square$

Lemma 4. *Consider jointly Gaussian random variables $\{X_1, X_2, Y_1, Y_2\}$ as shown in Fig. 16, such that for $k \in \{1, 2\}$*

$$X_k \sim \mathcal{N}(0, 1)$$
$$Z_k \sim \mathcal{N}(0, e_k)$$
$$Y_k = X_k + Z_k.$$

Also $X_2 = \rho X_1 + N$. Define

$$\delta(e_1, e_2, \rho) \triangleq \text{Var}(X_1|Y_1) - \text{Var}(X_2|Y_1, Y_2).$$

For any $e_2, \rho \geq 0$,

$$\frac{d\delta(e_1, e_2, \rho)}{de_1} \geq 0. \tag{81}$$

Proof. Note that

$$\delta(e_1, e_2, \rho) \triangleq \text{Var}(X_1|Y_1) - \text{Var}(X_2|Y_1, Y_2)$$

$$= 1 - \frac{1}{1+e_1} - 1 + \begin{pmatrix} \rho & 1 \end{pmatrix} \begin{pmatrix} 1+e_1 & \rho \\ \rho & 1+e_2 \end{pmatrix} \begin{pmatrix} \rho \\ 1 \end{pmatrix}$$

$$= \frac{1 + e_1 - \rho^2(1-e_2)}{(1+e_1)(1+e_2) - \rho^2} - \frac{1}{1+e_1}.$$

We have

$$\frac{d\delta(e_1, e_2, \rho)}{de_1} = \frac{1}{(1+e_1)^2} - \frac{\rho^2 e_2^2}{((1+e_1)(1+e_2) - \rho^2)^2}. \tag{82}$$

It can be readily seen that (82) is non-negative, by simple manipulation of the inequality

$$(1 + e_1)(1 + (1 - \rho)e_2) \geq 1 \geq \rho^2.$$

This completes the proof. □

Remark 9. *The condition (81) can be interpreted as follows. Suppose that $Z \sim \mathcal{N}(0, \sigma_1^2)$ and $Z' \sim \mathcal{N}(0, \sigma_2^2)$ be independent of all other variables and $\sigma_1^2 \leq \sigma_2^2$. Then if $Var(X_1|X_1 + Z) - Var(X_2|X_1 + Z, Y_2) \geq 0$ we also have that $Var(X_1|X_1 + Z') - Var(X_2|X_1 + Z', Y_2) \geq 0$ must be satisfied.*

Lemma 5. *In the hybrid coding scheme, for any test channel noise σ_z^2 and any $t \geq \tau + B + W$ we have*

$$\lim_{\tau \to \infty} \left[Var(x_{t-1}|[u]_1^{\tau-1}, [u]_{\tau+B}^{t-1}) - Var(x_t|[u]_1^{\tau-1}, [u]_{\tau+B}^{t}) \right] \geq 0. \tag{83}$$

Proof. We first note the following for $t \geq \tau + B + W$:

$$\mathrm{Var}(x_t|[u]_1^{\tau-1}, [u]_{\tau+B}^{t-1}, u_t) = \mathrm{Var}(x_t|[u]_1^{\tau-1}, [u]_{\tau+B}^{t-1}, u_t + \sum_{k=1}^{W} w_k u_{t-k})$$

$$= \mathrm{Var}(x_t|[u]_1^{\tau-1}, [u]_{\tau+B}^{t-1}, x_t + z_t).$$

Next, if $\hat{x}_t([u]_1^{\tau-1}, [u]_{\tau+B}^{t-1})$ denotes the estimate of x_t given $([u]_1^{\tau-1}, [u]_{\tau+B}^{t-1})$, such that we have

$$\hat{x}_{t-1}([u]_1^{\tau-1}, [u]_{\tau+B}^{t-1}) = x_{t-1} + \tilde{n}_1$$

where $\tilde{n}_1 \sim \mathcal{N}(0, \tilde{\sigma}_1^2)$ is independent of x_{t-1}. From the principle of sufficient statistics, we have that

$$\mathrm{Var}(x_{t-1}|[u]_1^{\tau-1}, [u]_{\tau+B}^{t-1}) - \mathrm{Var}(x_t|[u]_1^{\tau-1}, [u]_{\tau+B}^{t}) = \mathrm{Var}(x_{t-1}|x_{t-1} + \tilde{n}_1) - \mathrm{Var}(x_t|x_{t-1} + \tilde{n}_1, x_t + z_t). \tag{84}$$

It suffices to show the non-negativity of the right hand side of (84) as $t \to \infty$. To establish this we instead exhibit a $\tilde{n}_2 \sim \mathcal{N}(0, \tilde{\sigma}_2^2)$ with $\tilde{\sigma}_2^2 \leq \tilde{\sigma}_1^2$ such that

$$\lim_{t \to \infty} \left[\mathrm{Var}(x_{t-1}|x_{t-1} + \tilde{n}_2) - \mathrm{Var}(x_t|x_{t-1} + \tilde{n}_2, x_t + z_t) \right] = 0. \tag{85}$$

The result then follows from Remark 9. In particular we can express:

$$\hat{x}_{t-1}([u]_1^{t-1}) = x_{t-1} + \tilde{n}_2$$

where $\tilde{n}_2 \sim \mathcal{N}(0, \tilde{\sigma}_2^2)$ is independent of x_t and clearly $\tilde{\sigma}_2^2 \leq \tilde{\sigma}_1^2$. By following the steps similar to (84), it can be shown that

$$\mathrm{Var}(x_{t-1}|[u]_1^{t-1}) - \mathrm{Var}(x_t|[u]_1^{t}) = \mathrm{Var}(x_{t-1}|x_{t-1} + \tilde{n}_2) - \mathrm{Var}(x_t|x_{t-1} + \tilde{n}_2, x_t + z_t).$$

Finally note that in steady state we have:

$$\lim_{t \to \infty} \mathrm{Var}(x_{t-1}|[u]_1^{t-1}) = \lim_{t \to \infty} \mathrm{Var}(x_t|[u]_1^{t}), \tag{86}$$

which in turn establishes (85), and completes the proof. □

Having established the necessary auxiliary results, we now turn to the proof of (69) and (70). Towards this end we show that the terms $R_{2,\tau}$ and $D_{2,\tau}$ will dominate the rate and distortion expressions. This is accomplished through several steps outlined below.

Step 1: For any fixed τ, and any $t \leq \tau$ we have

$$R_{2,\tau}(\sigma_z^2) \geq R_{1,\tau}(t, \sigma_z^2) \tag{87}$$

$$D_{2,\tau}(\sigma_z^2) \geq D_{1,\tau}(t, \sigma_z^2). \tag{88}$$

Proof. To show (87), note that from the definition of $R_{2,\tau}(\sigma_z^2)$ (see (64)), we have

$$R_{2,\tau}(\sigma_z^2) \geq h(u_{\tau+B+W}|[u]_1^{\tau-1}, [u]_{\tau+B}^{\tau+B+W-1}) - \frac{1}{2} \log 2\pi e \sigma_z^2$$

$$\geq h(u_{\tau+B+W}|[u]_1^{\tau+B+W-1}, [x]_1^{\tau+B+W-t}, [z]_1^{\tau+B+W-t}) - \frac{1}{2} \log 2\pi e \sigma_z^2 \tag{89}$$

$$= h(u_t|[u]_1^{t-1}) - \frac{1}{2} \log 2\pi e \sigma_z^2 \tag{90}$$

$$= R_{1,\tau}(t, \sigma_z^2),$$

where (89) follows from the fact that conditioning reduces the differential entropy and (90) follows from the application of Lemma 2 at time $\tau + B + W$ for $\mathcal{M} = \{1, \dots, \tau + B + W\}$ and $k = \tau + B + W - t$. The distortion constraint in (88) can be verified through the similar steps. □

Step 2: The functions $R_\tau(\sigma_z^2)$ and $D_\tau(\sigma_z^2)$, defined as,

$$R_\tau(\sigma_z^2) \triangleq \max_t R_\tau(t, \sigma_z^2) \tag{91}$$

$$D_\tau(\sigma_z^2) \triangleq \max_t D_\tau(t, \sigma_z^2) \tag{92}$$

are increasing functions with respect to τ.

Proof. Note that

$$
\begin{aligned}
R_\tau(\sigma_z^2) &= \max_t R_\tau(t, \sigma_z^2) \\
&\leq \max_t R_{\tau+1}(t+1, \sigma_z^2) \\
&= R_{\tau+1}(\sigma_z^2),
\end{aligned}
\tag{93}
$$

where (93) follows from (75) in Lemma 3. The proof for $D_\tau(\sigma_z^2)$ follows similarly. $\qquad\square$

This step shows that the dominant term in the rate and distortion constraints corresponds to the steady state behaviour i.e., $\tau \to \infty$.

Step 3: For any $t > \tau + B + W$, we have that

$$R_{2,\tau}(\sigma_z^2) \geq R_{3,\tau}(t, \sigma_z^2) \tag{94}$$

$$D_{2,\tau}(\sigma_z^2) \geq D_{3,\tau}(t, \sigma_z^2) \tag{95}$$

are satisfied in the limit $\tau \to \infty$.

Proof. We first focus on the rate constraint (94). Note that by the definition of $R_{2,\tau}(\sigma_z^2)$ in (64), we have

$$R_{2,\tau}(\sigma_z^2) \geq h\big(u_{\tau+B+W} \big| [u]_1^{\tau-1}, [u]_{\tau+B}^{\tau+B+W-1}\big) - \frac{1}{2}\log\big(2\pi e \sigma_z^2\big). \tag{96}$$

According to (96) and the definition of $R_{3,\tau}(t, \sigma_z^2)$ in (66), in order to show (94), it suffices to show the following for any $t \geq \tau + B + W$,

$$h\big(u_t \big| [u]_1^{\tau-1}, [u]_{\tau+B}^{t-1}\big) \geq h\big(u_{t+1} \big| [u]_1^{\tau-1}, [u]_{\tau+B}^{t}\big). \tag{97}$$

Note that according to the definition of hybrid test channel, we have

$$x_t + z_t = u_t + \sum_{k=1}^{W} w_k u_{t-k},$$

i.e., $x_t + z_t$ is only the function of the current and the past W test channel outputs. Thus for any $t \geq \tau + B + W$,

$$h\big(u_t \big| [u]_1^{\tau-1}, [u]_{\tau+B}^{t-1}\big) = h\big(x_t + z_t \big| [u]_1^{\tau-1}, [u]_{\tau+B}^{t-1}\big),$$

and hence (97) can be written as

$$h\big(x_t + z_t \big| [u]_1^{\tau-1}, [u]_{\tau+B}^{t-1}\big) \geq h\big(x_{t+1} + z_{t+1} \big| [u]_1^{\tau-1}, [u]_{\tau+B}^{t}\big), \tag{98}$$

and based on the fact that z_t and z_{t+1} are i.i.d., to show (98) it suffices to show that

$$h\big(x_t \big| [u]_1^{\tau-1}, [u]_{\tau+B}^{t-1}\big) \geq h\big(x_{t+1} \big| [u]_1^{\tau-1}, [u]_{\tau+B}^{t}\big),$$

which is equivalent to show,

$$\mathrm{Var}\big(x_t \big| [u]_1^{\tau-1}, [u]_{\tau+B}^{t-1}\big) \geq \mathrm{Var}\big(x_{t+1} \big| [u]_1^{\tau-1}, [u]_{\tau+B}^{t}\big). \tag{99}$$

In addition note that for any $t \geq \tau + B + W$,

$$
\begin{aligned}
\mathrm{Var}\big(x_t \big| [u]_1^{\tau-1}, [u]_{\tau+B}^{t-1}\big) &= \mathrm{Var}\big(\rho x_{t-1} + n_t \big| [u]_1^{\tau-1}, [u]_{\tau+B}^{t-1}\big) \\
&= \rho^2 \mathrm{Var}\big(x_{t-1} \big| [u]_1^{\tau-1}, [u]_{\tau+B}^{t-1}\big) + (1 - \rho^2).
\end{aligned}
$$

Thus to show (99), it suffice to show

$$\mathrm{Var}\big(x_{t-1} \big| [u]_1^{\tau-1}, [u]_{\tau+B}^{t-1}\big) \geq \mathrm{Var}\big(x_t \big| [u]_1^{\tau-1}, [u]_{\tau+B}^{t}\big), \tag{100}$$

and we only need to show (100) when $\tau \to \infty$, which follows from Lemma 5.

Now consider the distortion constraint in (95). By definition, it suffices to show that for any $t \geq \tau + B + W$,

$$\mathrm{Var}\big(x_t \big| [u]_1^{\tau-1}, [u]_{\tau+B}^{t}\big) \geq \mathrm{Var}\big(x_{t+1} \big| [u]_1^{\tau-1}, [u]_{\tau+B}^{t+1}\big),$$

which is readily justified according to Lemma 5 in the limit $\tau \to \infty$. This proves the distortion constraint in (70). $\qquad\square$

According to Step 1, for any t, the rate and distortion constraints of region 2, i.e., $R_{2,\tau}(\sigma_z^2)$ and $D_{2,\tau}(\sigma_z^2)$, always dominate the constraints of region 1. According to step 2, we only need to focus on the case where the erasure burst happens at $\tau \to \infty$. Finally according to step 3, as $\tau \to \infty$, the rate and distortion constraints of region 2 also dominate the constraints of region 3. By combining these results, it can be concluded that $\lim_{\tau\to\infty} R_{2,\tau}(\sigma_z^2)$ and $\lim_{\tau\to\infty} D_{2,\tau}(\sigma_z^2)$ are the dominating rate and distortion constraints as required in (69) and (70).

C. Rate Computation

In this step we show how the limit $\tau \to \infty$ in the rate and distortion constraints in (69) and (70) can be replaced with the steady-state expressions in Theorem 1. In order to establish (29), we need to show that for any $\mathcal{M}_\tau \subset \mathcal{L}_\tau$, we have that:

$$\lim_{\tau\to\infty} h([u]_{\mathcal{M}_\tau}|[u]_1^{\tau-1}, [u]_{\mathcal{M}^c_\tau}) = h\left([\tilde{u}]_{\mathcal{M}}|[\tilde{u}]_{\mathcal{M}^c}, \tilde{s}_1\right), \tag{101}$$

where the variables $\tilde{u}_j$ and $\tilde{s}_1$ are defined in (27) and (25) respectively and $\mathcal{M}_\tau$ is the set $\mathcal{M}$ shifted by time τ.

In a similar fashion for the distortion constraint we must show that:

$$\lim_{\tau\to\infty} \mathrm{Var}(x_{\tau+B+W}|[u]_1^{\tau-1}, [u]_{\tau+B}^{\tau+B+W}) = \mathrm{Var}(s_{B+W+1}|[\tilde{u}]_{B+1}^{B+W+1}, \tilde{s}_1). \tag{102}$$

In what follows we establish (102). The derivation of (101) is analogous and hence will be omitted. Consider the $(\tau-1) \times (\tau-1)$ matrix $\mathbf{Q}_{\mathrm{H},\tau}$ which consists of the first $\tau-1$ rows and columns of $\mathbf{Q_w}$ in Section IV-A. This matrix is lower triangular and thus invertible. We can express:

$$\mathbf{Q}_{\mathrm{H},\tau}^{-1}\begin{pmatrix} u_1 \\ \vdots \\ u_{\tau-1} \end{pmatrix} = \begin{pmatrix} x_1 \\ \vdots \\ x_{\tau-1} \end{pmatrix} + \begin{pmatrix} z_1 \\ \vdots \\ z_{\tau-1} \end{pmatrix} \triangleq \begin{pmatrix} v_1 \\ \vdots \\ v_{\tau-1} \end{pmatrix}. \tag{103}$$

Now consider the left hand side of (102):

$$\mathrm{Var}(x_{\tau+B+W}|[u]_1^{\tau-1}, [u]_{\tau+B}^{\tau+B+W})$$
$$= \mathrm{Var}(x_{\tau+B+W}|\mathbf{Q}_{\mathrm{H},\tau}^{-1}[u]_1^{\tau-1}, [u]_{\tau+B}^{\tau+B+W})$$
$$= \mathrm{Var}(x_{\tau+B+W}|[v]_1^{\tau-1}, [u]_{\mathcal{L}_\tau}), \tag{104}$$

where the last step uses $\mathcal{L}_\tau \triangleq \{\tau+B, \ldots, \tau+B+W\}$.

Now suppose that the matrix $\mathbf{Q}_1$ is of size $(W+1) \times (\tau+B+W)$ consisting of rows with index $\mathcal{L}_\tau$ and columns with index $[1, \tau+B+W]$ of matrix $\mathbf{Q_w}$. In addition suppose that the matrices $\tilde{\mathbf{Q}}_1$ and $\mathbf{Q}_{\mathrm{eff}}$ are of sizes $(W+1) \times (\tau-1)$ and $(W+1) \times (B+W+1)$, respectively such that

$$\mathbf{Q}_1 = [\tilde{\mathbf{Q}}_1, \mathbf{Q}_{\mathrm{eff}}].$$

Then we can express:

$$[u]_{\mathcal{L}_\tau} = \mathbf{Q}_1 \left([x]_1^{\tau+B+W} + [z]_1^{\tau+B+W}\right)$$
$$= [\tilde{\mathbf{Q}}_1, \mathbf{Q}_{\mathrm{eff}}] \left([x]_1^{\tau+B+W} + [z]_1^{\tau+B+W}\right)$$
$$= \tilde{\mathbf{Q}}_1 \left([x]_1^{\tau-1} + [z]_1^{\tau-1}\right) + \mathbf{Q}_{\mathrm{eff}} \left([x]_\tau^{\tau+B+W} + [z]_\tau^{\tau+B+W}\right)$$
$$= \tilde{\mathbf{Q}}_1 [v]_1^{\tau-1} + \mathbf{Q}_{\mathrm{eff}}[v]_\tau^{\tau+B+W}. \tag{105}$$

Thus we can express (104) as:

$$\mathrm{Var}(x_{\tau+B+W}|[v]_1^{\tau-1}, [u]_{\mathcal{L}_\tau})$$
$$= \mathrm{Var}(x_{\tau+B+W}|[v]_1^{\tau-1}, \mathbf{Q}_{\mathrm{eff}}[v]_\tau^{\tau+B+W})$$
$$= \mathrm{Var}(x_{\tau+B+W}|\tilde{x}_\tau([v]_1^{\tau-1}), \mathbf{Q}_{\mathrm{eff}}[v]_\tau^{\tau+B+W}), \tag{106}$$

where $\tilde{x}_\tau([v]_1^{\tau-1})$ the MMSE estimate of x_τ given $[v]_1^{\tau-1}$, which can be done since $\{x_t\}$ also constitute a Markov chain. Now consider the limit $\tau \to \infty$:

$$\lim_{\tau\to\infty} \mathrm{Var}(x_{\tau+B+W}|[v]_1^{\tau-1}, [u]_{\mathcal{L}_\tau})$$
$$= \lim_{\tau\to\infty} \mathrm{Var}(x_{\tau+B+W}|\tilde{x}_\tau([v]_1^{\tau-1}), \mathbf{Q}_{\mathrm{eff}}([x]_\tau^{\tau+B+W}+[z]_\tau^{\tau+B+W}))$$
$$= \lim_{\tau\to\infty} \mathrm{Var}(s_{\tau+B+W}|\tilde{s}_\tau([v]_1^{\tau-1}), \mathbf{Q}_{\mathrm{eff}}([s]_\tau^{\tau+B+W}+[z]_\tau^{\tau+B+W})),$$

where we use the fact that as $\tau \to \infty$, $x_\tau \to s_\tau$ and furthermore $\tilde{s}_\tau([v]_1^{\tau-1})$ is the MMSE estimate of s_τ given $[v]_1^{\tau-1}$. Using the analysis of a recursive one-dimensional Kalman filter [25]:

$$s_i = \rho s_{i-1} + n_i \quad n_i \sim \mathcal{N}(0, 1 - \rho^2)$$
$$v_i = s_i + z_i, \quad z \sim \mathcal{N}(0, \sigma_z^2),$$

we can express the estimate of s_τ in the steady state as:

$$\tilde{s}_\tau = s_\tau + e$$

where $e \sim \mathcal{N}\left(0, \Sigma(\sigma_z^2)/(1 - \Sigma(\sigma_z^2))\right)$, and $\Sigma(\sigma_z^2)$ is defined in (26).

Finally according to the Toeplitz property of the matrix $\mathbf{Q_w}$, and therefore $\mathbf{Q}_{\text{eff}}$, we can write (106) as

$$\max_{\tau, t} D_\tau(t, \sigma_z^2)$$
$$= \lim_{\tau \to \infty} \text{Var}(s_{\tau+B+W} | \tilde{s}_\tau, \mathbf{Q}_{\text{eff}}([s]_\tau^{\tau+B+W} + [z]_\tau^{\tau+B+W}))$$
$$= \text{Var}(s_{B+W+1} | \tilde{s}_1, \mathbf{Q}_{\text{eff}}([s]_1^{B+W+1} + [z]_1^{B+W+1}))$$
$$= \text{Var}(s_{B+W+1} | \tilde{s}_1, [\tilde{u}]_{B+1}^{B+W+1}),$$

as required.

Appendix C
Hybrid Coding in the High Resolution Regime: Proof of Corollary 1

According to the rate analysis of the hybrid coding scheme, the excess sum-rate constraint associated with an erasure burst spanning $[t : t + B - 1]$ in steady state, can be expressed as

$$R_{E,\text{sum}}(D) \triangleq R_{\text{sum}}(D) - R_{\text{I}}(D)$$
$$= \lim_{t \to \infty} h([u]_{t+B}^{t+B+W} | [u]_1^{t-1}) - \frac{(W+1)}{2} \log\left(2\pi e \sigma_z^2\right) - (W+1)R_{\text{I}}(D), \tag{107}$$

where $R_{\text{I}}(D)$ is defined in (1). The test channel noise σ_z^2 depends on the distortion D and the choice of hybrid coding coefficients $\mathbf{w}$, and has to satisfy

$$\lim_{t \to \infty} \text{Var}(x_{t+B+W} | [u]_1^{t-1} [u]_{t+B}^{t+B+W}) = D. \tag{108}$$

We show that the choice of weights in (32) minimizes the asymptotic behaviour of the excess sum-rate in (107) in the high resolution regime when $D \to 0$.

We first consider the problem for any distortion D. Following similar argument as in the proof of Theorem 1, according to (105), we have

$$\lim_{t \to \infty} h\left([u]_{t+B}^{t+B+W} | [u]_1^{t-1}\right)$$
$$= \lim_{t \to \infty} h\left(\mathbf{Q}_{\text{eff}}([x]_t^{t+B+W} + [z]_t^{t+B+W}) | [x]_1^{t-1} + [z]_1^{t-1}\right)$$
$$= \lim_{t \to \infty} h\left(\mathbf{Q}_{\text{eff}}([s]_t^{t+B+W} + [z]_t^{t+B+W}) | \hat{s}_{t-1}([x]_1^{t-1} + [z]_1^{t-1})\right)$$
$$= \lim_{t \to \infty} h\left(\mathbf{Q}_{\text{eff}}([s]_t^{t+B+W} + [z]_t^{t+B+W}) | s_{t-1} + e\right) \tag{109}$$
$$= h\left(\mathbf{Q}_{\text{eff}}([s]_1^{B+W+1} + [z]_1^{B+W+1}) | s_0 + e\right) \tag{110}$$
$$= h\left([\tilde{u}]_{B+1}^{B+W+1} | s_0 + e\right), \tag{111}$$

where in (109) we replaced the estimate of s_{t-1} with a noisy version of the source s_{t-1}. The noise e is related to the estimation error of the source s_{t-1} from $[u]_1^{t-1}$. Also (110) follows from the stationarity property of the sources and is independent of t. The variables $\tilde{u}_i$ in (111) are also defined is Theorem 1.

According to (111), the excess sum-rate in (107) can be expressed as

$$R_{E,\text{sum}}(D) = h\left([\tilde{u}]_{B+1}^{B+W+1} | s_0 + e\right) - (W+1)\left(\frac{1}{2} \log\left(2\pi e \sigma_z^2\right) + R_{\text{I}}(D)\right). \tag{112}$$

We now consider a generalization of the test channel in (27) as

$$\begin{pmatrix} \tilde{u}_1 \\ \vdots \\ \tilde{u}_{B+W+1} \end{pmatrix} \triangleq \tilde{\mathbf{Q}}_{\text{eff}}\left(\begin{pmatrix} s_1 \\ \vdots \\ s_{B+W+1} \end{pmatrix} + \begin{pmatrix} z_1 \\ \vdots \\ z_{B+W+1} \end{pmatrix}\right),$$

where

$$\tilde{\mathbf{Q}}_{\text{eff}} \triangleq \begin{pmatrix} 1 & 0 & \cdots & 0 \\ v_1 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ v_{B+W} & v_{B+W-1} & \cdots & 1 \end{pmatrix}.$$

Note that $\mathbf{Q}_{\text{eff}}$ consists of rows $[B+1 : B+W+1]$ of $\tilde{\mathbf{Q}}_{\text{eff}}$. The following lemma is valid for any distortion D.

Lemma 6. *For any B and W, the choice of the hybrid coding scheme parameters $\mathbf{w}$ minimizing $h\left([\tilde{u}]_{B+1}^{B+W+1}|s_0+e\right)$ also minimizes*

$$I([\tilde{u}]_1^B; [\tilde{u}]_{B+1}^{B+W+1}|s_0+e, s_{B+W+1}). \tag{113}$$

Proof. Note that,

$$\begin{aligned} &I([\tilde{u}]_1^B; [\tilde{u}]_{B+1}^{B+W+1}|s_0+e, s_{B+W+1}) \\ &= h([\tilde{u}]_{B+1}^{B+W+1}|s_0+e, s_{B+W+1}) \\ &\quad - h([\tilde{u}]_{B+1}^{B+W+1}|s_0+e, [\tilde{u}]_1^B, s_{B+W+1}) \\ &= h([\tilde{u}]_{B+1}^{B+W+1}|s_0+e) + h(s_{B+W+1}|s_0+e, [\tilde{u}]_{B+1}^{B+W+1}) \\ &\quad - h(s_{B+W+1}|s_0+e) - h([\tilde{u}]_{B+1}^{B+W+1}|s_0+e, [\tilde{u}]_1^B) \\ &\quad - h(s_{B+W+1}|s_0+e, [\tilde{u}]_1^{B+W+1}) \\ &\quad + h(s_{B+W+1}|s_0+e, [\tilde{u}]_1^B). \end{aligned} \tag{114}$$

The third term in (114) is clearly independent of $\mathbf{w}$. The second term in (114) is equal to $\frac{1}{2}\log(2\pi eD)$ according to the distortion constraint, and is independent of $\mathbf{w}$. The fifth and sixth terms are also independent of $\mathbf{w}$, because of the invertibility of any upper-left square sub-matrix of $\tilde{\mathbf{Q}}_{\text{eff}}$. For instance, the sixth term can be written as

$$h(s_{B+W+1}|s_0+e, [\tilde{u}]_1^B) = h(s_{B+W+1}|s_0+e, [s]_1^B + [z]_1^B).$$

Also the fourth term is independent of $\mathbf{w}$, because

$$h([\tilde{u}]_{t+1}^{t+W+1}|s_0+e, [\tilde{u}]_1^t) = \sum_{j=t+1}^{t+W+1} h(\tilde{u}_j|s_0+e, [\tilde{u}]_1^{j-1}),$$

which is independent of the choice of $\mathbf{w}$, because, for any j, $h(\tilde{u}_j|s_0+e, [\tilde{u}]_1^{j-1})$ is independent of the choice of $\mathbf{w}$, i.e.,

$$\begin{aligned} h(\tilde{u}_j|s_0+e, [\tilde{u}]_1^{j-1}) &= h(\tilde{u}_j|s_0+e, [s]_1^{j-1} + [z]_1^{j-1}) \\ &= h\left((s_j+z_j) + \sum_{k=1}^{j-1} q_{j,k}(s_k+z_k)\Big|s_0+e, [s]_1^{j-1} + [z]_1^{j-1}\right) \\ &= h(s_j+z_j|s_0+e, [s]_1^{j-1} + [z]_1^{j-1}). \end{aligned}$$

Thus the choice of $\mathbf{w}$ which minimizes the first term in (114), minimizes the mutual information. This completes the proof. $\square$

The following lemma characterizes the high resolution behaviour of the second term in the sum-rate (112).

Lemma 7. *In the high resolution regime, the asymptotic behaviour of the second and third terms in (112), is independent of $\mathbf{w}$. In particular, for any choice of $\mathbf{w}$,*

$$\lim_{D\to 0}\left(\frac{1}{2}\log(2\pi e\sigma_z^2) + R_I(D)\right) = \frac{1}{2}\log(2\pi e(1-\rho^2)). \tag{115}$$

Proof. First note that the distortion constraint in (108), is in fact the distortion constraint after the error propagation window for a erasure burst at steady state, i.e.,

$$\text{Var}(x_{B+W+1}|[u]_{B+1}^{B+W+1}, \tilde{s}_1) = \lim_{t\to\infty} \text{Var}(x_{t+W+1}|[u]_1^{t-B}, [u]_{t+1}^{t+W+1}).$$

We now derive lower and upper bounds on the test channel noise σ_z^2. According to the hybrid test channel, we have

$$x_{t+W+1} + z_{t+W+1} = \sum_{j=0}^{W} w_j u_{t+W-j+1}.$$

$$
\begin{pmatrix} \widehat{u}_1 \\ \vdots \\ \widehat{u}_{W+2} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ v_1 & 1 & 0 & \cdots & 0 \\ v_2 & v_1 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ v_{W+1} & v_W & v_{W-1} & \cdots & 1 \end{pmatrix} \left(\begin{pmatrix} x_1 \\ \vdots \\ x_{W+2} \end{pmatrix} + \begin{pmatrix} z_1 \\ \vdots \\ z_{W+2} \end{pmatrix} \right)
$$

$$
= \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ v_1 & 1 & 0 & \cdots & 0 \\ v_2 & v_1 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ v_{W+1} & v_W & v_{W-1} & \cdots & 1 \end{pmatrix} \times \left(\begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ \rho & 1 & 0 & \cdots & 0 \\ \rho^2 & \rho & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{W+1} & \rho^W & \rho^{W-1} & \cdots & 1 \end{pmatrix} \begin{pmatrix} n_1 \\ \vdots \\ n_{W+2} \end{pmatrix} + \begin{pmatrix} z_1 \\ \vdots \\ z_{W+2} \end{pmatrix} \right). \tag{122}
$$

Therefore at least $x_{t+W+1} + z_{t+W+1}$ is available at the decoder while reconstructing x_{t+W+1}. So

$$
\begin{aligned}
D &= \lim_{t \to \infty} \mathrm{Var}(x_{t+W+1} | [u]_1^{t-B}, [u]_{t+1}^{t+W+1}) \\
&\leq \lim_{t \to \infty} \mathrm{Var}(x_{t+W+1} | x_{t+W+1} + z_{t+W+1}) \\
&= \frac{\sigma_z^2}{1 + \sigma_z^2}.
\end{aligned} \tag{116}
$$

Also note from the Markov property among the source sequences that

$$
\begin{aligned}
D &= \lim_{t \to \infty} \mathrm{Var}(x_{t+W+1} | [u]_1^{t-B}, [u]_{t+1}^{t+W+1}) \\
&\geq \lim_{t \to \infty} \mathrm{Var}(x_{t+W+1} | [x]_1^{t+W}, [u]_1^{t-B}, [u]_{t+1}^{t+W+1}) \\
&\geq \lim_{t \to \infty} \mathrm{Var}(x_{t+W+1} | x_{t+W}, x_{t+W+1} + z_{t+W+1}) = \frac{\sigma_z^2}{1 + \frac{\sigma_z^2}{1-\rho^2}}.
\end{aligned} \tag{117}
$$

From (116) and (117), we have

$$
\frac{1}{1-D} \leq \frac{\sigma_z^2}{D} \leq \frac{1}{1 - \frac{D}{1-\rho^2}}. \tag{118}
$$

From (118),

$$
\lim_{D \to 0} \frac{\sigma_z^2}{D} = 1. \tag{119}
$$

Thus

$$
\begin{aligned}
&\lim_{D \to 0} \left(\frac{1}{2} \log(2\pi e \sigma_z^2) + \frac{1}{2} \log\left(\frac{1-\rho^2}{D} + \rho^2 \right) \right) \\
&= \frac{1}{2} \log(2\pi e (1-\rho^2)).
\end{aligned} \tag{120}
$$

which implies (115). $\qquad \square$

According to Lemma 7, in the high resolution regime, the contribution of the second and third terms of the excess sum rate in (107) is independent of the choice of the hybrid coding coefficient $\mathbf{w}$. So, it suffices to focus on the high resolution behaviour of the first term in the excess sum-rate $R_{\mathrm{sum}}(D)$ in (112). Also according to Lemma 6, it suffices to characterize the choice of $\mathbf{w}$ that minimizes the mutual information term in (113). We specialize the result of Lemma 6 to the case of $B = 1$ and the high resolution case. It is not hard to verify that in the high resolution regime, the variance of the noise e vanishes to zero. By subtracting out the effect of s_0, we define

$$
[\widehat{u}]_1^{W+2} = \widehat{\mathbf{Q}} \left([x]_1^{W+2} + [z]_1^{W+2} \right), \tag{121}
$$

where $\widehat{\mathbf{Q}}$ is the square matrix of size $W + 2$. In particular, (121) can be written as (122) at the top of next page. Thus the mutual information term in (113) for the case of $B = 1$ reduces to

$$
\begin{aligned}
&\lim_{D \to 0} I(\tilde{u}_1; [\tilde{u}]_2^{W+2} | s_0 + e, s_{W+2}) \\
&= \lim_{D \to 0} I(\tilde{u}_1; [\tilde{u}]_2^{W+2} | s_0, s_{W+2}) \\
&= \lim_{D \to 0} I(\widehat{u}_1; [\widehat{u}]_2^{W+2} | x_{W+2}).
\end{aligned} \tag{123}
$$

$$\begin{pmatrix} \widehat{u}_1 \\ \vdots \\ \widehat{u}_{W+2} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ v_1 & 1 & 0 & \cdots & 0 \\ v_2 & v_1 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ v_{W+1} & v_W & v_{W-1} & \cdots & 1 \end{pmatrix} \times \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ \rho & 1 & 0 & \cdots & 0 \\ \rho^2 & \rho & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{W+1} & \rho^W & \rho^{W-1} & \cdots & 1 \end{pmatrix} \begin{pmatrix} n_1 \\ \vdots \\ n_{W+2} \end{pmatrix} \tag{126}$$

$$= \widehat{\mathbf{Q}} \Lambda \mathbf{n} \tag{127}$$

The following lemma shows that the choice of $\mathbf{w}$ in (124) minimizes (123).

Lemma 8. *The vector $\mathbf{w}^\star$, with the elements*

$$w_k^\star = \rho^k \frac{1 - \rho^{2(W-k+1)}}{1 - \rho^{2(W+1)}} \quad \text{for } k \in \{1, \ldots, W\}, \tag{124}$$

satisfies:

$$\lim_{D \to 0} I(\widehat{u}_1; [\widehat{u}]_2^{W+2} | x_{W+2}) = 0. \tag{125}$$

Thus minimizes the sum-rate.

Proof. Since $D \to 0$ requires $\sigma_z^2 \to 0$, we will only establish (125) when $\sigma_z^2 = 0$. The proof when $D \to 0$ will follow from the continuity of the mutual information in σ_z^2. Thus we will consider the test channel at (126) at the top of this page. when establishing (125). In (127), $\widehat{\mathbf{Q}}$ and Λ represent the first and second matrices of size $W+2 \times W+2$ in (126) and $\mathbf{n}$ represents the vector of the $W+2$ i.i.d. innovation random variables $[n]_1^{W+2}$. We prove that $I(\widehat{u}_1; [\widehat{u}]_2^{W+2} | x_{W+2}) = 0$ by establishing the following Markov property

$$\widehat{u}_1 \to x_{W+2} \to [\widehat{u}]_2^{W+2}. \tag{128}$$

Based on the fact the random variables are jointly Gaussian, in order to show (128), it suffices to show that the MMSE estimator of $\widehat{u}_1$ from $[\widehat{u}]_2^{W+2}$ is a scaled version of x_{W+2}. This is equivalent to show that a scalar κ exists such that

$$(x_{W+2} - \kappa \widehat{u}_1) \perp \{\widehat{u}_2, \widehat{u}_3, \ldots, \widehat{u}_{W+2}\}.$$

This is equivalent to

$$E\{(x_{W+2} - \kappa \widehat{u}_1)(\widehat{\mathbf{Q}}\Lambda n)^T\} = \xi \begin{pmatrix} 1 & 0 & \cdots & 0 \end{pmatrix}, \tag{129}$$

i.e, $x_{W+2} - \kappa \widehat{u}_1$ is orthogonal to all the rows of $\widehat{\mathbf{Q}}\Lambda n$ except the first row. Note that

$$x_{W+2} - \kappa \widehat{u}_1 = \begin{pmatrix} \rho^{W+1} - \kappa & \rho^W & \cdots & \rho & 1 \end{pmatrix} \mathbf{n} \tag{130}$$

Using (130) and from the fact that $E\{nn^T\} = (1 - \rho^2)I$, the left hand side of (129), can be written as

$$E\{(x_{W+2} - \kappa \widehat{u}_1)(\widehat{\mathbf{Q}}\Lambda n)^T\}$$
$$= (1 - \rho^2)\begin{pmatrix} \rho^{W+1} - \kappa & \rho^W & \cdots & \rho & 1 \end{pmatrix} \Lambda^T \widehat{\mathbf{Q}}^T \tag{131}$$

First note that according to the definition of $\widehat{\mathbf{Q}}$, we have

$$\widehat{\mathbf{A}} \triangleq \widehat{\mathbf{Q}}^{-1} = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 & 0 \\ w_1 & 1 & 0 & \cdots & 0 & 0 \\ w_2 & w_1 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ w_W & w_{W-1} & w_{W-2} & \cdots & 1 & 0 \\ 0 & w_W & w_{W-1} & \cdots & w_1 & 1 \end{pmatrix}$$

From (131), by multiplying $\widehat{\mathbf{Q}}^{-T}$ from the right, it can be observed that showing (129) is equivalent to show

$$\begin{pmatrix} \rho^{W+1} - \kappa & \rho^W & \cdots & \rho & 1 \end{pmatrix} \Lambda^T = \frac{\xi}{1 - \rho^2} \begin{pmatrix} 1 & 0 & \cdots & 0 \end{pmatrix} \widehat{\mathbf{A}}^T \tag{132}$$

It can be written as

$$\begin{pmatrix} \rho^{W+1} - \kappa & \rho^W & \cdots & \rho & 1 \end{pmatrix} \begin{pmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{W+1} \\ 0 & 1 & \rho & \cdots & \rho^W \\ 0 & 0 & 1 & \cdots & \rho^{W-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{pmatrix} = \frac{\xi}{1 - \rho^2}\begin{pmatrix} 1 & w_1 & w_2 & \cdots & w_W & 0 \end{pmatrix}. \tag{133}$$

First from the first element of (133), we have

$$\rho^{W+1} - \kappa = \frac{\xi}{1 - \rho^2}. \tag{134}$$

From the second element of (133), we have

$$(\rho^{W+1} - \kappa)\rho + \rho^W = \frac{\xi}{1 - \rho^2} w_1. \tag{135}$$

Thus, from (134) and (135), we have

$$\rho^{W+1} - \kappa = \frac{\xi}{1 - \rho^2} = \frac{\rho^W}{w_1 - \rho}. \tag{136}$$

Now we show that the choice of coefficients $w_k^\star$ in (124), for any $k \in \{2, \ldots, W\}$, satisfies

$$(\rho^{W+1} - \kappa)\rho^k + \sum_{j=0}^{k-1} \rho^{k+W-2j} = \frac{\xi}{1 - \rho^2} w_k^\star. \tag{137}$$

By using (136) and noting that

$$w_1^\star - \rho = \frac{-\rho^{2W}(1 - \rho^2)}{1 - \rho^{2(W+1)}},$$

equation (137) can be verified through basic steps. Finally, as the last element of (133), we need to show that

$$(\rho^{W+1} - \kappa)\rho^{W+1} + \sum_{j=0}^{W} \rho^{2(W-j)+1} = 0,$$

which is readily verified. This completes the proof. $\qquad \square$

Finally we show that the sum-rate constraint in (107) coincide with the high resolution lower bound in Corollary 4, and thus is optimal. In particular we want to show that, with the choice of hybrid coding weights in (32), we have

$$\lim_{D \to 0} \left\{ R_{\text{sum}} - \frac{1}{2} \log\left(\frac{1 - \rho^2}{D}\right) \right\} = \lim_{D \to 0} \frac{1}{2(W+1)} \log\left(\frac{1 - \rho^{2(W+2)}}{1 - \rho^{2(W+1)}}\right). \tag{138}$$

We have

$$\lim_{D \to 0} R_{\text{sum}} = \lim_{D \to 0} \frac{1}{W+1} h([\tilde{u}]_{B+1}^{B+W+1} | s_0 + e) - \frac{1}{2} \log(2\pi e D). \tag{139}$$

First note that by similar argument used before, (139) can be written as

$$\lim_{D \to 0} R_{\text{sum}} = \lim_{D \to 0} \frac{1}{W+1} h([\hat{u}]_2^{W+2}) - \frac{1}{2} \log(2\pi e D) \tag{140}$$

where $\hat{u}$ are defined in (121). Now note that

$$\begin{aligned}
h([\hat{u}]_2^{W+2}) &= I(x_{W+2}; [\hat{u}]_2^{W+2}) + h([\hat{u}]_2^{W+2} | x_{W+2}) \\
&= h(x_{W+2}) - h(x_{W+2} | [\hat{u}]_2^{W+2}) + h([\hat{u}]_2^{W+2} | x_{W+2}) \\
&= \frac{1}{2} \log\left(2\pi e(1 - \rho^{2(W+2)})\right) - h(x_{W+2} | [\hat{u}]_2^{W+2}) \\
&\quad + I(\hat{u}_1; [\hat{u}]_2^{W+2} | x_{W+2}) + h([\hat{u}]_2^{W+2} | \hat{u}_1, x_{W+2}).
\end{aligned} \tag{141}$$

When $D \to 0$, the mutual information term in (141) approaches to zero according to Lemma 8. Now consider the last term in (141), we have

$$\begin{aligned}
\lim_{D \to 0} h([\hat{u}]_2^{W+2} | \hat{u}_1, x_{W+2}) &= \lim_{D \to 0} h([\hat{u}]_1^{W+1} | x_{W+1}) \\
&= \lim_{D \to 0} \left(h([\hat{u}]_1^{W+1}) - I([\hat{u}]_1^{W+1}; x_{W+1})\right) \\
&= h([x]_1^{W+1}) - h(x_{W+1}) + \lim_{D \to 0} h(x_{W+1} | [u]_1^{W+1}) \\
&= \frac{1}{2} \log\left((2\pi e)^W \frac{(1 - \rho^2)^{W+1}}{1 - \rho^{2(W+1)}}\right) + \lim_{D \to 0} h(x_{W+1} | [u]_1^{W+1}).
\end{aligned} \tag{142}$$

Note that the second term in (141) and the last term in (142) cancel each other. Thus (141) can be written as

$$h([\hat{u}]_2^{W+2}) = \frac{1}{2} \log\left(\frac{1 - \rho^{2(W+2)}}{1 - \rho^{2(W+1)}}\right) + \frac{W+1}{2} \log\left(2\pi e(1 - \rho^2)\right). \tag{143}$$

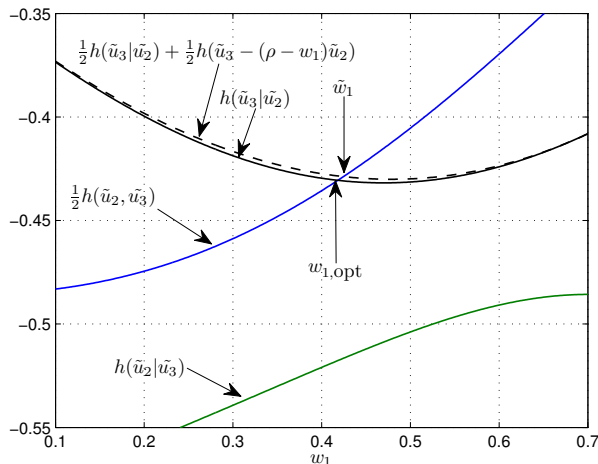Finally by replacing (143) into (140), (138) is verified. This completes the proof.

Fig. 17. The sum-rate and marginal rates as a function of coefficient w_1, for $B = W = 1$ and $\rho = 0.7$. The marginal rate is upper bounded for approximation.

APPENDIX D
PROOF OF COROLLARY 2

We need to show that the high resolution excess rate for $B = W = 1$ is upper bounded by the expression

$$R_{\text{E,HR}}(\rho, B = 1) \leq \frac{1}{4} \log \left(1 + \frac{2\rho^4}{(1+\rho)^2} \right).$$

We prove this part through the following steps.

1) Specializing Theorem 1 for $B = W = 1$ and the high resolution regime, we have that the achievable rate can be expressed as:

$$R_{\text{H}}(D, w_1) = \max \left\{ \frac{1}{2} h(\tilde{u}_2, \tilde{u}_3), h(\tilde{u}_2 | \tilde{u}_3), h(\tilde{u}_3 | \tilde{u}_2) \right\} - \frac{1}{2} \log(2\pi e D), \tag{144}$$

where

$$\begin{pmatrix} \tilde{u}_3 \\ \tilde{u}_2 \end{pmatrix} \triangleq \begin{pmatrix} 1 & -w_1 & w_1^2 \\ 0 & 1 & -w_1 \end{pmatrix} \begin{pmatrix} 1 & \rho & \rho^2 \\ 0 & 1 & \rho \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} n_3 \\ n_2 \\ n_1 \end{pmatrix}$$

$$= \begin{pmatrix} 1 & \rho - w_1 & \rho^2 - w_1\rho + w_1^2 \\ 0 & 1 & \rho - w_1 \end{pmatrix} \begin{pmatrix} n_3 \\ n_2 \\ n_1 \end{pmatrix}.$$

2) For any choice of $w_1 \in [0, \rho]$, we have

$$\text{Var}(\tilde{u}_3) = (1 - \rho^2)(1 + (\rho - w_1)^2 + (\rho^2 - \rho w_1 + w_1^2)^2)$$
$$\geq (1 - \rho^2)(1 + (\rho - w_1)^2) = \text{Var}(\tilde{u}_2)$$

and therefore,

$$h(\tilde{u}_3 | \tilde{u}_2) = h(\tilde{u}_2, \tilde{u}_3) - h(\tilde{u}_2)$$
$$\geq h(\tilde{u}_2, \tilde{u}_3) - h(\tilde{u}_3) = h(\tilde{u}_2 | \tilde{u}_3).$$

Thus (144) reduces to

$$R_{\text{H}}(D, w_1) = \max \left\{ \frac{1}{2} h(\tilde{u}_2, \tilde{u}_3), h(\tilde{u}_3 | \tilde{u}_2) \right\} - \frac{1}{2} \log(2\pi e D)$$
$$= \frac{1}{2} h(\tilde{u}_3 | \tilde{u}_2) + \frac{1}{2} \max \left\{ h(\tilde{u}_2), h(\tilde{u}_3 | \tilde{u}_2) \right\} - \frac{1}{2} \log(2\pi e D). \tag{145}$$

3) Fig. 17 shows an example of terms $h(\tilde{u}_2)$ and $h(\tilde{u}_3 | \tilde{u}_2)$ for $\rho = 0.7$ and $w_1 \in [0, \rho]$. Finding the close form expression of the value of $w_{1,\text{opt}}$ at the intersection is not straightforward. We apply the following approximation.

$$h(\tilde{u}_3 | \tilde{u}_2) \leq h(\tilde{u}_3 - (\rho - w_1)\tilde{u}_2)$$

$$= h(n_3 + \rho w_1 n_1)$$

$$= \frac{1}{2} \log \left(2\pi e (1 - \rho^2)(1 + \rho^2 w_1^2) \right).$$

It can be easily observed that the upper bound of $h(\tilde{u}_3|\tilde{u}_2)$ and $h(\tilde{u}_2)$, intersect at $\tilde{w}_1 = \rho/(1+\rho)$. By replacing this value to compute

$$\frac{1}{2} h(\tilde{u}_2, \tilde{u}_3)|_{w_1 = \tilde{w}_1}$$

$$= \frac{1}{2} h \left(\begin{pmatrix} 1 & \rho - \tilde{w}_1 & \rho^2 - \tilde{w}_1 \rho + \tilde{w}_1^2 \\ 0 & 1 & \rho - \tilde{w}_1 \end{pmatrix} \begin{pmatrix} n_3 \\ n_2 \\ n_1 \end{pmatrix} \right)$$

$$= \frac{1}{4} \log \left(2\pi e (1 - \rho^2)^2 \left(1 + 2\frac{\rho^4}{(1+\rho)^2} \right) \right).$$

Thus the rate expression in (145) is upper bounded as

$$R_{\mathrm{H}}(D, w_1) \leq \frac{1}{2} \log \left(\frac{1 - \rho^2}{D} \right) + \frac{1}{4} \log \left(1 + 2\frac{\rho^4}{(1+\rho)^2} \right).$$

This completes the proof.

APPENDIX E
PROOF OF COROLLARY 3

We consider the case where $W = 1$ and $B \to \infty$. Consider the system at time t where $t \to \infty$ and the erasure burst spans the interval $[1 : t - 2]$. We have,

$$\begin{pmatrix} u_t \\ u_{t-1} \end{pmatrix} = \begin{pmatrix} 1 & -w_1 & w_1^2 & -w_1^3 & \cdots \\ 0 & 1 & -w_1 & w_1^2 & \cdots \end{pmatrix} \begin{pmatrix} s_t \\ s_{t-1} \\ s_{t-2} \\ \vdots \end{pmatrix}. \tag{146}$$

Now consider the following lemma.

Lemma 9. *For the random variables defined in* (146), *we have*

$$E\{|u_t|^2\} = E\{|u_{t-1}|^2\}$$

$$= (1 - \rho^2) \left(\frac{\rho^2}{1 - \rho^2} + \frac{1}{1 - w_1^2} \right) \frac{1}{(1 + w_1\rho)^2}$$

$$= (1 - \rho^2) f(w_1)$$

$$E\{u_t u_{t-1}\} = (1 - \rho^2) \left(\rho f(w_1) - \frac{w_1}{(1 + w_1\rho)(1 - w_1^2)} \right)$$

$$= (1 - \rho^2) g(w_1)$$

where $f(.)$ and $g(.)$ are defined in (35) *and* (36), *respectively.*

Proof. From (146), consider the following definition.

$$\begin{pmatrix} u_t \\ u_{t-1} \end{pmatrix} = \begin{pmatrix} 1 & -w_1 & w_1^2 & -w_1^3 & \cdots \\ 0 & 1 & -w_1 & w_1^2 & \cdots \end{pmatrix} \begin{pmatrix} s_t \\ s_{t-1} \\ s_{t-2} \\ \vdots \end{pmatrix}$$

$$\triangleq \begin{pmatrix} a_0 & a_1 & a_2 & \cdots \\ 0 & a_0 & a_1 & \cdots \end{pmatrix} \begin{pmatrix} n_t \\ n_{t-1} \\ \vdots \end{pmatrix}.$$

It is not hard to observe that, for any $m \geq 0$,

$$a_m = \sum_{k=0}^{m} (-w_1)^k \rho^{m-k}.$$

We have

$$E\{|u_t|^2\} = E\{|u_{t-1}|^2\} = (1 - \rho^2) \sum_{m=0}^{\infty} a_m^2$$

$$E\{u_t u_{t-1}\} = (1 - \rho^2) \sum_{m=0}^{\infty} a_m a_{m+1}.$$

Note that

$$a_m^2 = \sum_{l=0}^{m} \sum_{k=0}^{m} (-w_1)^{k+l} \rho^{2m-k-l}$$

$$= \sum_{l=0}^{m} \sum_{k=0}^{m} \rho^{2m} (\frac{-w_1}{\rho})^{k+l}$$

$$= \sum_{j=0}^{m-1} \rho^{2m} (j+1) \left((\frac{-w_1}{\rho})^j + (\frac{-w_1}{\rho})^{2m-j} \right) + \rho^{2m} (m+1)(\frac{-w_1}{\rho})^m,$$

and therefore,

$$\sum_{m=0}^{\infty} a_m^2 = \sum_{m=0}^{\infty} (m+1)(-w_1\rho)^m + \sum_{m=0}^{\infty} \sum_{j=0}^{m-1} \rho^{2m} (j+1) \left((\frac{-w_1}{\rho})^j + (\frac{-w_1}{\rho})^{2m-j} \right)$$

$$= \frac{1}{(1+w_1\rho)^2} + \sum_{j=0}^{\infty} \sum_{m=j+1}^{\infty} \rho^{2m} (j+1) \left((\frac{-w_1}{\rho})^j + (\frac{-w_1}{\rho})^{2m-j} \right)$$

$$= \frac{1}{(1+w_1\rho)^2} + \sum_{j=0}^{\infty} \left((j+1)(\frac{-w_1}{\rho})^j \sum_{m=j+1}^{\infty} \rho^{2m} \right) + \sum_{j=0}^{\infty} \left((j+1)(\frac{-w_1}{\rho})^{-j} \sum_{m=j+1}^{\infty} \rho^{2m} (\frac{-w_1}{\rho})^{2m} \right)$$

$$= \frac{1}{(1+w_1\rho)^2} + \sum_{j=0}^{\infty} (j+1) \left((\frac{-w_1}{\rho})^j \frac{\rho^{2(j+1)}}{1-\rho^2} + (\frac{-w_1}{\rho})^{-j} \frac{w_1^{2(j+1)}}{1-w_1^2} \right)$$

$$= \frac{1}{(1+w_1\rho)^2} + \frac{\rho^2}{1-\rho^2} \frac{1}{(1+w_1\rho)^2} + \frac{w_1^2}{1-w_1^2} \frac{1}{(1+w_1\rho)^2}$$

$$= \left(\frac{1}{1-\rho^2} + \frac{w_1^2}{1-w_1^2} \right) \frac{1}{(1+w_1\rho)^2}$$

$$= \left(\frac{\rho^2}{1-\rho^2} + \frac{1}{1-w_1^2} \right) \frac{1}{(1+w_1\rho)^2} \triangleq f(w_1).$$

Similarly,

$$\sum_{m=0}^{\infty} a_m a_{m+1} = \sum_{m=0}^{\infty} \sum_{l=0}^{m} \sum_{k=0}^{m+1} \rho^{2m+1} (\frac{-w_1}{\rho})^{k+l}$$

$$= \rho f(w_1) + \sum_{m=0}^{\infty} \sum_{l=0}^{m} \rho^{2m+1} (\frac{-w_1}{\rho})^{m+l+1}$$

$$= \rho f(w_1) + \sum_{l=0}^{\infty} \sum_{m=l}^{\infty} \rho^{2m+1} (\frac{-w_1}{\rho})^{m+l+1}$$

$$= \rho f(w_1) - \frac{w_1}{(1+w_1\rho)(1-w_1^2)} \triangleq g(w_1).$$

$\square$

By application of Lemma 9, the sum-rate constraint is

$$2R \geq \frac{1}{2} \log \left(\frac{(2\pi e)^2 \det \begin{pmatrix} E\{|u_t|^2\} & E\{u_t u_{t-1}\} \\ E\{u_t u_{t-1}\} & E\{|u_{t-1}|^2\} \end{pmatrix}}{(2\pi e D)^2} \right)$$

$$= \log \left(\frac{1-\rho^2}{D} \right) + \frac{1}{2} \log \left(f(w_1)^2 - g(w_1)^2 \right).$$

Now it suffices to show that the sum-rate in indeed the dominant constraint. In particular, note that

$$h(u_{t-1}|u_t) = h(u_{t-1}, u_t) - h(u_t)$$
$$= h(u_{t-1}, u_t) - h(u_{t-1}) = h(u_t|u_{t-1}),$$

i.e., the two marginal constraints are the same, and

$$\frac{1}{2}h(u_{t-1}, u_t) = \frac{h(u_t) + h(u_{t-1}|u_t)}{2}$$
$$\geq \frac{h(u_{t-1}|u_t) + h(u_{t-1}|u_t)}{2} = h(u_{t-1}|u_t),$$

i.e., the sum-rate constraint dominates the marginal rate constraints. This completes the proof.

APPENDIX F
PROOF OF LEMMA 1

First note that for any $\rho \in (0,1)$ and $x \in \mathbb{R}$ the function

$$f(x) = x - \frac{1}{2}\log\left(\rho^{2m}2^{2x} + 2\pi e(1-\rho^{2m})\right)$$

is an monotonically increasing function with respect to x, because

$$f'(x) = \frac{2\pi e(1-\rho^{2m})}{\rho^{2m}2^{2x} + 2\pi e(1-\rho^{2m})} > 0.$$

By applying Shannon's EPI we have.

$$h(s_b|f_a) \geq \frac{1}{2}\log\left(\rho^{2m}2^{2h(s_a|f_a)} + 2\pi e(1-\rho^{2m})\right)$$

and thus,

$$h(s_a|f_a) - h(s_b|f_a)$$
$$\leq h(s_a|f_a) - \frac{1}{2}\log\left(\rho^{2m}2^{2h(s_a|f_a)} + 2\pi e(1-\rho^{2m})\right)$$
$$\leq \frac{1}{2}\log(2\pi er) - \frac{1}{2}\log\left(\rho^{2m}2\pi er + 2\pi e(1-\rho^{2m})\right) \tag{147}$$
$$= \frac{1}{2}\log\left(\frac{r}{1-(1-r)\rho^{2m}}\right),$$

where (147) follows from the assumption that $h(s_a|f_a) \leq \frac{1}{2}\log(2\pi er)$ and the monotonicity property of $f(x)$. This completes the proof.

APPENDIX G
GENERAL LOWER BOUND ON THE RATE-RECOVERY FUNCTION

Before providing the proof of the lower bound, define the notations

$$\theta_\tau \triangleq \frac{1}{2\pi e}2^{\frac{2}{n}h(s_\tau^n|[f]_1^\tau, s_0^n)},$$

$$F_a(\tau) \triangleq 2\pi e\left(\rho^{2a}\theta_\tau + 1 - \rho^{2a}\right),$$

$$G \triangleq (1 - (1-D)\rho^{2(W+1)}).$$

In order to derive a lower bound on the rate-recovery function in general case, consider the case where the burst erasure of length B spans the interval $[t-B-W : t-W-1]$ and the decoder is interested in reconstructing the source sequence s_t^n within distortion D at time t. Then the inequality

$$n\sum_{k=t-W}^{t} R_k \geq H([f]_{t-W}^t)$$

$$\geq H([f]_{t-W}^t|[f]_1^{t-B-W-1}, s_0^n) \tag{148}$$
$$= I(s_t^n; [f]_{t-W}^t|[f]_1^{t-B-W-1}, s_0^n) + H([f]_{t-W}^t|s_t^n, [f]_1^{t-B-W-1}, s_0^n)$$
$$\geq h(s_t^n|[f]_1^{t-B-W-1}, s_0^n) - h(s_t^n|[f]_1^{t-B-W-1}[f]_{t-W}^t, s_0^n) + H([f]_{t-W}^t|s_t^n, [f]_1^{t-B-W-1}, s_0^n) \tag{149}$$

holds, where (148) follows from the fact that conditioning never increases the differential entropy.

The first term in (149) can be lower bounded using $s_t = \rho^{B+W+1} s_{t-B-W-1} + n'_t$ with $n'_t \in \mathcal{N}(0, 1 - \rho^{2(B+W+1)})$, and using the entropy power inequality as:

$$h(s_t^n | [f]_1^{t-B-W-1}, s_0^n) \geq \frac{n}{2} \log \left(F_{B+W+1}(t - B - W - 1) \right). \tag{150}$$

The second term in (149) that can be lower bounded based on the fact that the decoder is able to reconstruct the source sequence s_t^n within distortion D knowing $\{[f]_1^{t-B-W-1} [f]_{t-W}^t, s_0^n\}$ and follows from the standard source coding arguments, i.e.,

$$h(s_t^n | [f]_1^{t-B-W-1} [f]_{t-W}^t, s_0^n) \leq \frac{n}{2} \log(2\pi e D). \tag{151}$$

To derive a lower bound on the third term in (149) consider the following:

$$
\begin{aligned}
& H([f]_{t-W}^t | s_t^n, [f]_1^{t-B-W-1}, s_0^n) \\
& \geq H([f]_{t-W}^{t-1} | s_t^n, [f]_1^{t-W-1}, s_0^n) \\
& = I([f]_{t-W}^{t-1}; [s^n]_{t-W}^{t-1} | s_t^n, [f]_1^{t-W-1}, s_0^n) + H([f]_{t-W}^{t-1} | [s^n]_{t-W}^t, [f]_1^{t-W-1}, s_0^n) \\
& \geq h([s^n]_{t-W}^{t-1} | s_t^n, [f]_1^{t-W-1}, s_0^n) - h([s^n]_{t-W}^{t-1} | s_t^n, [f]_1^{t-1}, s_0^n).
\end{aligned}
\tag{152}
$$

The first term in (152) can be written as

$$
\begin{aligned}
& h([s^n]_{t-W}^{t-1} | s_t^n, [f]_1^{t-W-1}, s_0^n) \\
& = h([s^n]_{t-W}^t | [f]_1^{t-W-1}, s_0^n) - h(s_t^n | [f]_1^{t-W-1}, s_0^n) \\
& = h(s_{t-W}^n | [f]_1^{t-W-1}, s_0^n) + nWh(s_1|s_0) - h(s_t^n | [f]_1^{t-W-1}, s_0^n) \\
& \geq \frac{n}{2} \log(F_1(t - W - 1)) + nWh(s_1|s_0) - \frac{n}{2} \log(2\pi e G),
\end{aligned}
\tag{153}
$$

where the first term in (153) follows from Shannon's EPI with $s_{t-W}^n = \rho s_{t-W-1}^n + n_{t-W}^n$, and the last term follows from

$$h(s_t^n | [f]_1^{t-W-1}, s_0^n) \leq h(s_t^n - \hat{s}_t^n([f]_1^{t-W-1}, s_0^n)) \tag{154}$$

$$\leq \frac{n}{2} \log \left(2\pi e \rho^{2(W+1)} D + 2\pi e(1 - \rho^{2(W+1)}) \right) \tag{155}$$

$$= \frac{n}{2} \log \left(2\pi e(1 - (1 - D)\rho^{2(W+1)}) \right),$$

where (154) follows from the fact that knowing $\{[f]_1^{t-W-1}, s_0^n\}$ the decoder is able to reproduce an estimate of s_t as

$$\hat{s}_t^n([f]_1^{t-W-1}, s_0^n) = \rho^{W+1} \hat{s}_{t-W-1}^n([f]_1^{t-W-1}, s_0^n) + \tilde{n}$$

where $\tilde{n} \sim \mathcal{N}(0, 1 - \rho^{2(W+1)})$. Eq. (155) also follows from the fact that the Gaussian distribution maximizes the differential entropy. The second term in (152) can be written as

$$
\begin{aligned}
q(W) \triangleq\ & h([s^n]_{t-W}^{t-1} | s_t^n, [f]_1^{t-1}, s_0^n) \\
= & h([s^n]_{t-W}^t | [f]_1^{t-1}, s_0^n) - h(s_t^n | [f]_1^{t-1}, s_0^n) \\
= & h(s_{t-1}^n | [f]_1^{t-1}, s_0^n) + h(s_t^n | s_{t-1}^n, [f]_1^{t-1}, s_0^n) + h([s^n]_{t-W}^{t-2} | s_t^n, s_{t-1}^n, [f]_1^{t-1}, s_0^n) - h(s_t^n | [f]_1^{t-1}, s_0^n) \\
\leq & h(s_{t-1}^n | [f]_1^{t-1}, s_0^n) + h(s_t^n | s_{t-1}^n) + h([s^n]_{t-W}^{t-2} | s_{t-1}^n, [f]_1^{t-1}, s_0^n) - h(s_t^n | [f]_1^{t-1}, s_0^n) \tag{156} \\
\leq & \frac{n}{2} \log(2\pi e \theta_{t-1}) + nh(s_1|s_0) + h([s^n]_{t-W}^{t-2} | s_{t-1}^n, [f]_1^{t-2}, s_0^n) - \frac{n}{2} \log(F_1(t-1)) \tag{157} \\
= & \frac{n}{2} \log(2\pi e \theta_{t-1}) + nh(s_1|s_0) + q(W - 1) - \frac{n}{2} \log(F_1(t-1)), \tag{158}
\end{aligned}
$$

where the last term in (157) follows from EPI, the second and the third terms in (156) follow from the Markov chain properties

$$\{[f]_0^{t-1}, s_{-1}^n\} \to s_{t-1}^n \to s_t^n,$$

$$s_t^n \to \{s_{t-1}^n, [f]_0^{t-1}, s_{-1}^n\} \to [s^n]_{t-W}^{t-2}.$$

By repeating these steps steps W times and using $q(0) = 0$:

$$q(W) \leq \frac{n}{2} \log \left(\frac{(2\pi e)^W \prod_{k=1}^W \theta_{t-k}}{\prod_{k=1}^W F_1(t-k)} \right) + nWh(s_1|s_0). \tag{159}$$

Substituting (153) and (159) into (152):

$$H([f]_{t-W}^t | s_t^n, [f]_1^{t-B-W-1}, s_0^n) \geq \frac{n}{2} \log \left(\frac{\prod_{k=1}^{W+1} F_1(t-k)}{(2\pi e)^{W+1} G \prod_{k=1}^{W} \theta_{t-k}} \right). \tag{160}$$

Finally by replacing (150), (151) and (160) into (149), we have

$$\sum_{k=t-W}^{t} R_k \geq \frac{1}{2} \log \left(\frac{F_{B+W+1}(t-B-W-1) \prod_{k=1}^{W+1} F_1(t-k)}{(2\pi e)^{W+2} GD \prod_{k=1}^{W} \theta_{t-k}} \right). \tag{161}$$

Using (161) we have

$$\left((2\pi e)^{W+2} GD \right)^K 2^{\sum_{t=K+1}^{2K} \sum_{k=t-W}^{t} R_k} \geq \prod_{t=K+1}^{2K} \frac{F_{B+W+1}(t-B-W-1) \prod_{k=1}^{W+1} F_1(t-k)}{\prod_{k=1}^{W} \theta_{t-k}}. \tag{162}$$

Now note that

$$\left((2\pi e)^{W+2} GD \right)^K 2^{K(W+1)R} \geq \left((2\pi e)^{W+2} GD \right)^K 2^{\sum_{t=K+1}^{2K} \sum_{k=t-W}^{t} R_k} \tag{163}$$

$$\geq F_x \prod_{\tau=K}^{2K-B-W-1} \frac{F_{B+W+1}(\tau)(F_1(\tau))^{W+1}}{\theta_\tau^W}, \tag{164}$$

where (163) follows from the fact that peak rate is greater than the average rate, and (164) follows from (162) after simple manipulations, also F_x is defined as

$$F_x \triangleq \left(\prod_{k=K+1}^{K+B+W} F_{B+W+1}(k-B-W-1) \prod_{k'=k-W-1}^{K} F_1(k') \right) \left(\prod_{k=2K-B-W}^{2K} \prod_{k'=2K-B-W-1}^{k-1} F_1(k') \right),$$

and is bounded as $K \to \infty$. From (162) we can write,

$$R \geq \lim_{K \to \infty} \frac{1}{2K(W+1)} \sum_{\tau=K}^{2K-B-W-1} \log \left(\frac{F_{B+W+1}(\tau)(F_1(\tau))^{W+1}}{(2\pi e)^{W+2} GD \theta_\tau^W} \right)$$

$$\geq \lim_{K \to \infty} \frac{1}{K} \sum_{\tau=K}^{2K-B-W-1} \Gamma(\theta_\tau, B, W), \tag{165}$$

where the function $\Gamma(\vartheta, B, W)$ is defined in (45) in Theorem 2. Note that $\Gamma(\vartheta, B, W)$ is a continuous function with respect to $\vartheta \in (0, \infty)$. It approaches $+\infty$ when $\vartheta \to 0^+$ and $\vartheta \to \infty$. In addition the expression inside the logarithm has a positive second derivative and thus, $\Gamma(\vartheta, B, W)$ has a global optimum at ϑ_m. The variable θ_τ satisfies the upper and lower bounds

$$\lambda(R) \triangleq \frac{1-\rho^2}{2^{2R} - \rho^2} \leq \theta_\tau \leq D \tag{166}$$

when $\tau \to \infty$. The upper bound follows from the fact the the decoder will recover the source sequence within average distortion D provided that all the past channel packets are available. To derive the lower bound note that,

$$h(s_\tau^n | [f]_1^\tau, s_0^n) = h(s_\tau^n | [f]_1^{\tau-1}, s_0^n) - I(f_\tau; s_\tau^n | [f]_1^{\tau-1}, s_0^n)$$
$$= h(s_\tau^n | [f]_1^{\tau-1}, s_0^n) - H(f_\tau | [f]_1^{\tau-1}, s_0^n) + H(f_\tau | s_\tau^n, [f]_1^{\tau-1}, s_0^n)$$
$$\geq h(s_\tau | [f]_1^{\tau-1}, s_0^n) - H(f_\tau) \tag{167}$$
$$\geq \frac{n}{2} \log \left(\rho^2 2^{\frac{n}{2} h(s_{\tau-1}^n | [f]_1^{\tau-1}, s_0^n)} + 2\pi e(1-\rho^2) \right) - nR, \tag{168}$$

where (167) follows from the fact that the conditioning never increases the entropy and (168) follows from EPI. Thus,

$$\theta_\tau \geq \frac{\rho^2}{2^{2R}} \theta_{\tau-1} + \frac{1-\rho^2}{2^{2R}},$$

which results in

$$\theta_\tau \geq \left(\frac{\rho^2}{2^{2R}} \right)^{\tau-1} \theta_1 + \frac{1-\rho^2}{2^{2R}} \sum_{l=0}^{\tau-2} \left(\frac{\rho^2}{2\pi e 2^{2R}} \right)^l$$

$$\geq \frac{1-\rho^2}{2^{2R} - \rho^2} \left(1 - \left(\frac{\rho^2}{2^{2R}} \right)^{\tau-1} \right), \tag{169}$$

where (169) follows from the fact that $0 < \frac{\rho^2}{2^{2R}} < 1$ for any $\rho \in (0,1)$ and $R > 0$. The lower bound in (166) refers to the tightest lower bound in (169) at $\tau \to \infty$.

Applying the lower bound on ϑ in (165) results in a lower bound in rate as $R \geq \Gamma_\lambda$. This is because $R - \Gamma(\lambda(R), B, W)$ is an increasing function with respect to R. Also $R \geq \Gamma(\vartheta_m, B, W)$ and $R \geq \Gamma(D, B, W)$. Therefore,

- If $\vartheta_m \leq \lambda(\Gamma_\lambda) \leq D$, then $R \geq \Gamma_\lambda$.
- If $\lambda(\Gamma_\lambda) \leq \vartheta_m \leq D$, then $R \geq \Gamma(\vartheta_m, B, W)$.
- If $\lambda(\Gamma_\lambda) \leq D \leq \vartheta_m$, then $R \geq \Gamma(D, B, W)$.

APPENDIX H
PREDICTIVE CODING: PROOF OF THEOREM 3

In predictive coding, the encoder at each time t, computes the MMSE estimation error of the source $\mathbf{x}_t$ from all the previous codewords $\mathbf{u}_i, i \leq t-1$. Based on the optimality of the MMSE estimator for jointly Gaussian sources, the estimation error e_t, and thus u_t, is independent of the random variables $u_i, i \leq t-1$. In the analysis, it is more convenient to use a backward test channel:

$$e_t = u_t + \tilde{z}_t,$$

where $\tilde{z}_t \sim \mathcal{N}(0, \tilde{\sigma}_z^2)$ is independent of $u_i, \forall i \leq t$. Using the orthogonality principle, one can show that

$$e_t = \rho \tilde{z}_{t-1} + n_t,$$

and furthermore we can show that

$$x_t = \rho^{t-1} u_1 + \rho^{t-2} u_2 + \ldots + \rho u_{t-1} + u_t + \tilde{z}_t.$$

Furthermore, the encoder at each time t quantizes e_t where the quantization rate satisfies,

$$
\begin{aligned}
R \geq R_{\text{PC}}(\sigma_e^2) &\triangleq I(e_t; u_t) \\
&= \frac{1}{2} \log\left(\frac{\sigma_e^2}{\tilde{\sigma}_z^2}\right) \\
&= \frac{1}{2} \log\left(\frac{1 - (1 - \tilde{\sigma}_z^2)\rho^2}{\tilde{\sigma}_z^2}\right).
\end{aligned}
\tag{170}
$$

The value of $\tilde{\sigma}_z^2$ will be specified in the sequel.

For the analysis of the burst erasure channel model observe that the decoder at anytime t, when the channel output f_t is not erased, recovers $\mathbf{u}_t$. Thus the reconstruction at time $t = \tau + B + W$, following an erasure burst in $[\tau, \tau + B - 1]$ is

$$
\begin{aligned}
\hat{x}_t &= \rho^t u_0 + \rho^{t-1} u_1 + \ldots + \rho^{B+W+1} u_{\tau-1} \\
&\quad + \rho^W u_{\tau+B} + \rho^{W-1} u_{\tau+B+1} + \ldots + u_{\tau+B+W}.
\end{aligned}
$$

One can show that this corresponds to the worst case distortion which is

$$
\begin{aligned}
E\left[(x_t - \hat{x}_t)^2\right]\big|_{t=\tau+B+W} &\triangleq \Sigma_{\text{PC}}(\sigma_e^2) \\
&= \tilde{\sigma}_z^2 + \sigma_u^2 \sum_{k=W+1}^{W+B} \rho^{2k} \\
&= \tilde{\sigma}_z^2 + \sigma_u^2 \rho^{2(W+1)} \frac{1 - \rho^{2B}}{1 - \rho^2} \\
&= \tilde{\sigma}_z^2 + (1 - \tilde{\sigma}_z^2)\rho^{2(W+1)}(1 - \rho^{2B}).
\end{aligned}
$$

By setting $\Sigma_{\text{PC}}(\sigma_e^2) = D$, we have

$$\tilde{\sigma}_z^2 = \frac{D - \rho^{2(W+1)}(1 - \rho^{2B})}{1 - \rho^{2(W+1)}(1 - \rho^{2B})},$$

for $D \geq \rho^{2(W+1)}(1 - \rho^{2B})$. By replacing $\tilde{\sigma}_z^2$ into the rate expression (170), we can observe that, for $D \geq \rho^{2(W+1)}(1 - \rho^{2B})$, any rate R satisfying

$$
\begin{aligned}
R &\geq R_{\text{PC}}^+(B, W, D) \\
&\triangleq \frac{1}{2} \log\left(\frac{1 - \rho^{2(W+1)}(1 - \rho^{2B}) - (1 - D)\rho^2}{D - \rho^{2(W+1)}(1 - \rho^{2B})}\right)
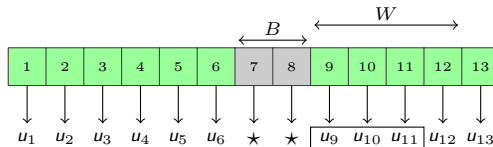\end{aligned}
$$

is achievable.

Fig. 18. Example of Lemma 10 with $B = 2$, $W = 4$, $\tau = 7$. In this case $k = 3$, $\mathcal{L}_\tau = \{9, 10, 11, 12, 13\}$ and $\mathcal{K}^\star = \{9, 10, 11\}$. According to Lemma 10, among any subset of $\mathcal{L}_\tau$ with size $k = 3$, differential entropy of $[u]_{\mathcal{K}^\star}$ give $[u]_1^{\tau-1}$ is the minimum.

APPENDIX I
MEMORYLESS QUANTIZATION-AND-BINNING SCHEME

A. Proof of Corollary 5

The memoryless quantization-and-binning scheme can be viewed as the special case of hybrid coding scheme with $\mathbf{A_w} = \mathbf{I}_T$. Therefore the achievability result of Theorem 1 also holds for this scheme. In particular from (69), the rate

$$R \geq \lim_{\tau \to \infty} \max_{\substack{\mathcal{M} \subseteq \mathcal{L}_\tau \\ \mathcal{M} \neq \phi}} \frac{1}{|\mathcal{M}|} h([u]_\mathcal{M}|[u]_1^{\tau-1}, [u]_{\bar{\mathcal{M}}}) - \frac{1}{2} \log(2\pi e \sigma_z^2) \tag{171}$$

is achievable, where $\mathcal{L}_\tau$ is defined in (65), and the test channel is defined as $u_i = s_i + z_i$. Recall that $z_i \sim \mathcal{N}(0, \sigma_z^2)$ and the test channel noise σ_z^2 satisfies the distortion constraint in (70). In order to prove Theorem 5, it suffices to show that the sum-rate constraint is the dominant constraint of (171), i.e., as $\tau \to \infty$, we have

$$\operatorname*{argmax}_{\substack{\mathcal{M} \subseteq \mathcal{L}_\tau \\ \mathcal{M} \neq \phi}} \frac{1}{|\mathcal{M}|} h([u]_\mathcal{M}|[u]_1^{\tau-1}.[u]_{\bar{\mathcal{M}}}) = \mathcal{L}_\tau. \tag{172}$$

We prove the (172) through the following steps.

Step 1: We first show that, for any fixed $m \in \{1, \ldots, W+1\}$, among all $\mathcal{M} \subseteq \mathcal{L}_\tau$ such that $|\mathcal{M}| = m$, the maximum rate is attained by the subset $\{\tau + B + W - m + 1, \ldots, \tau + B + W\}$, i.e.,

$$\operatorname*{argmax}_{\substack{\mathcal{M} \subseteq \mathcal{L}_\tau \\ |\mathcal{M}| = m}} \frac{1}{|\mathcal{M}|} h([u]_\mathcal{M}|[u]_1^{\tau-1}, [u]_{\bar{\mathcal{M}}})$$
$$= \{\tau + B + W - m + 1, \ldots, \tau + B + W\}. \tag{173}$$

To show (173) note that, for any $\mathcal{M} \subseteq \mathcal{L}_\tau$ such that $|\mathcal{M}| = m$, we have

$$h([u]_\mathcal{M}|[u]_0^{\tau-1}, [u]_{\bar{\mathcal{M}}})$$
$$= h([u]_M, [u]_{\bar{\mathcal{M}}}|[u]_0^{\tau-1}) - h([u]_{\bar{\mathcal{M}}}|[u]_0^{\tau-1})$$
$$= h([u]_{\tau+B}^{\tau+B+W}|[u]_0^{\tau-1}) - h([u]_{\bar{\mathcal{M}}}|[u]_0^{\tau-1}). \tag{174}$$

The first term in (174) is independent of $\mathcal{M}$. Thus we are looking for the set $\mathcal{M} \subseteq \mathcal{L}_\tau$ of size m that minimizes the second term in (174). Consider the following lemma.

Lemma 10. *For any set $\mathcal{K} \subseteq \mathcal{L}_\tau$ such that $|\mathcal{K}| = k$, we have:*

$$h([u]_\mathcal{K}|[u]_1^{\tau-1}) \geq h([u]_{\mathcal{K}^\star}|[u]_1^{\tau-1}) \tag{175}$$

where $\mathcal{K}^\star \triangleq \{\tau + B. \ldots, \tau + B + k - 1\}$.

Fig 18 schematically illustrates an example of Lemma 10. The proof follows from the application of [13, Lemma. 8] and is omitted here. According to Lemma 10, we have

$$\operatorname*{argmin}_{\mathcal{M} \subseteq \mathcal{L}_\tau, \mathcal{M} \neq \phi} h([u]_{\bar{\mathcal{M}}}|[u]_0^{\tau-1}) \tag{176}$$
$$= \{\tau + B + W - m + 1, \ldots, \tau + B + W\},$$

as required in (173). According to step 1,

$$\max_{\substack{\mathcal{M} \subseteq \mathcal{L}_\tau \\ \mathcal{M} \neq \phi}} \frac{1}{|\mathcal{M}|} h([u]_\mathcal{M}|[u]_1^{\tau-1}, [u]_{\bar{\mathcal{M}}}) = h([u]_{\tau+B+W-m+1}^{\tau+B+W}|[u]_0^{\tau-1}, [u]_{\tau+B}^{\tau+B+W-m}). \tag{177}$$

It remains to show that the term in (177) is an increasing function of m.

Step 2: For $\tau \to \infty$ and any $m \leq W$,

$$\frac{1}{m+1} h([u]_{\tau+B+W-m}^{\tau+B+W}|[u]_0^{\tau-1}, [u]_{\tau+B}^{\tau+B+W-m-1}) \geq \frac{1}{m} h([u]_{\tau+B+W-m+1}^{\tau+B+W}|[u]_0^{\tau-1}, [u]_{\tau+B}^{\tau+B+W-m}). \tag{178}$$

To show (178), it suffices to show that

$$mh([u]_{\tau+B+W-m}^{\tau+B+W}|[u]_0^{\tau-1}, [u]_{\tau+B}^{\tau+B+W-m-1}) \geq (m+1)h([u]_{\tau+B+W-m+1}^{\tau+B+W}|[u]_0^{\tau-1}, [u]_{\tau+B}^{\tau+B+W-m}),$$

or equivalently,

$$mh(u_{\tau+B+W-m}|[u]_0^{\tau-1}, [u]_{\tau+B}^{\tau+B+W-m-1}) \geq h([u]_{\tau+B+W-m+1}^{\tau+B+W}|[u]_0^{\tau-1}, [u]_{\tau+B}^{\tau+B+W-m}). \tag{179}$$

To show (179), note that

$$h([u]_{\tau+B+W-m+1}^{\tau+B+W}|[u]_0^{\tau-1}, [u]_{\tau+B}^{\tau+B+W-m}) = \sum_{k=0}^{m-1} h(u_{\tau+B+W-m+k+1}|[u]_0^{\tau-1}, [u]_{\tau+B}^{\tau+B+W-m+k})$$

$$= \sum_{k=0}^{m-1} h(u_{\tau+B+W-m}|[u]_0^{\tau-k-2}, [u]_{\tau+B-k-1}^{\tau+B+W-m-1}) \tag{180}$$

$$\leq mh(u_{\tau+B+W-m}|[u]_0^{\tau-1}, [u]_{\tau+B}^{\tau+B+W-m-1}), \tag{181}$$

where (180) follows from the time invariant property among the random variables at steady state when $\tau \to \infty$, (181) again follows from the application of [13, Lemma 8]. According to this lemma if the random variables u with indices closer to a particular time are erased, the conditional entropy in the largest.

According to step 2, (177) is an increasing function of m, thus is maximized with $m = W$. This proves (172) as required.

B. High Resolution Regime: Proof of Corollary 6

In order to analyze the high resolution behavior of the memoryless quantization-and-binning scheme, it suffices to study the rate expression in (53), in the limit $D \to 0$. In particular we need to show that,

$$\lim_{D\to 0} \left\{ R_{\mathrm{QB}}^{+}(B, W, D) - \frac{1}{2}\log\left(\frac{1-\rho^2}{D}\right) \right\} = \frac{1}{2(W+1)}\log\left(\frac{1-\rho^{2(B+1)}}{1-\rho^2}\right) + o(D), \tag{182}$$

where $\lim_{D\to 0} o(D) = 0$. First we set $\sigma_z^2 = D$, which satisfies the distortion constraint, i.e.,

$$\mathrm{Var}(s_{t+W}|[u]_t^{t+W}, \tilde{s}_{t-B}) \leq \mathrm{Var}(s_{t+W}|u_{t+W})$$

$$= \frac{D}{1+D} \leq D.$$

Note that when $\sigma_z^2 = D \to 0$, random variable u_t becomes asymptotically close to s_t. Thus the Markov property among the sources s_t approximately holds among u_t. Based on this observation, the high resolution limit of the first differential entropy term in (53) can be calculated as

$$\lim_{D\to 0}\lim_{\tau\to\infty} h([u]_{\tau+B}^{\tau+B+W}|[u]_1^{\tau-1}) = \lim_{\tau\to\infty} h([s]_{\tau+B}^{\tau+B+W}|[s]_1^{\tau-1})$$

$$= h([s]_{B+2}^{B+W+2}|s_1)$$

$$= \frac{1}{2}\log\left((2\pi e)^{W+1}(1-\rho^{2(B+1)})(1-\rho^2)^W\right). \tag{183}$$

Finally by replacing (183) into (53) with $\sigma_z^2 = D$, the expression in (182) is derived. This completes the proof.

APPENDIX J
GOP-BASED CODING: PROOF OF THEOREM 5

The GOP-Based coding scheme for the zero-delay streaming setup periodically transmits the *I-frames* as the *intra-coded* pictures that can be decoded at the decoder without the use of any other frame. Between the two consecutive I-frames, the *P-frames* as the *predicted* pictures are transmitted which require the use of previous frames to be decoded at the decoder.

According to the problem setup, in case of burst erasure of the channel, the decoder is required to start decoding the source vectors at most $W + 1$ times after the erasure ends. It is not hard to observe that in the GOP-based scheme the worst erasure pattern erases the I-frame and reveals the packets right after the I-frame. This suggest that in order to guarantee the recovery after $W + 1$ times, the I-frames have to be sent at least with a period of $W + 1$.

Let us define v_t as the quantization of the source vector s_t as the I-frame. Using the Gaussian test channel, the quantization can be modeled as

$$s_t = v_t + z_t,$$

Note that $z_t \sim \mathcal{N}(0, D)$ which guarantees the average distortion constraint. The decoder succeeds in reconstructing the source by only using the encoder output at time t if the rate satisfy

$$R_t \geq \frac{1}{2} \log \left(\frac{1}{D} \right). \tag{184}$$

For the time interval $[t+1 : t+W]$ the encoder sends $\mathbf{u}_t$ as the output of the predictive encoder, i.e., the P-frame. Using the similar notation for the predictive coding it is not hard to observe that the source s_i for any $i \in [t+1 : t+W]$ can be represented as

$$s_i = \rho^{i-t} v_t + \rho^{i-t-1} u_{t+1} + \ldots + \rho u_{i-1} + u_i + z_i.$$

At each time $i \in [t+1 : t+W]$, the decoder succeeds in recovering u_i if the rate satisfy

$$R_i \geq \frac{1}{2} \log \left(\frac{1 - \rho^2}{D} + \rho^2 \right). \tag{185}$$

From (184) and (185) it can be observed that the scheme requires the average rate

$$\begin{aligned} \bar{R}_{\text{GOP}}(W, D) &= \frac{1}{W+1} \sum_{i=t}^{t+W} R_i \\ &= \frac{1}{2} \log \left(\frac{1 - \rho^2}{D} + \rho^2 \right) + \frac{1}{2(W+1)} \log \left(\frac{1}{1 - (1-D)\rho^2} \right). \end{aligned} \tag{186}$$

Note that the rate expression in (186) is independent of the burst length B.

REFERENCES

[1] W. Tan and A. Zakhor, "Video multicast using layered FEC and scalable compression," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 11, pp. 373–386, 2001.
[2] Z. Li, A. Khisti, and B. Girod, "Forward error correction for low-delay packet video," in *Packet Video Workshop*, 2010.
[3] Y. Huang, Y. Kochman, and G. Wornell, "Causal transmission of colored source frames over a packet erasure channel," in *DCC*, 2010, pp. 129–138.
[4] K. Chang and R. Donaldson, "Analysis, optimization, and sensitivity study of differential PCM systems operating on noisy communication channels," *IEEE Trans. Comm.*, vol. 20, no. 3, pp. 338–350, 1972.
[5] D. Connor, "Techniques for reducing the visibility of transmission errors in digitally encoded video signals," *IEEE Trans. Comm.*, vol. 21, pp. 695–706, 1973.
[6] S. S. Pradhan and K. Ramchandran, "Distributed source coding using syndromes (DISCUS): Design and construction," *IEEE Trans. Inform. Theory*, vol. 49, pp. 626–643, Mar. 2003.
[7] B. Girod, A. M. Aaron, S. Rane, and D. Rebollo-Monedero, "Distributed video coding," *Proceedings of the IEEE*, vol. 93, pp. 71–83, Jan. 2005.
[8] H. Viswanathan and T. Berger, "Sequential coding of correlated sources," *IEEE Trans. Inform. Theory*, vol. 46, pp. 236–246, Jan. 2000.
[9] J. Wang and X. Wu, "Information flows in video coding," in *Data Compression Conference (DCC), 2010*, 2010, pp. 149–158.
[10] J. Nayak and E. Tuncel, "Successive coding of correlated sources," *IEEE Trans. Inform. Theory*, vol. 55, no. 9, pp. 4286–4298, 2009.
[11] L. Song, J. Chen, J. Wang, and T. Liu, "Gaussian robust sequential and predictive coding," *IEEE Trans. Inform. Theory*, vol. 59, no. 6, pp. 3635–3652, 2013.
[12] J. Wang, V. Prabhakaran, and K. Ramchandran, "Syndrome-based robust video transmission over networks with bursty losses," in *ICIP*, Atlanta, GA, 2006.
[13] F. Etezadi, A. Khisti, and M. Trott, "Zero-delay sequential transmission of markov sources over burst erasure channels," *IEEE Trans. Inform. Theory*, vol. 60, pp. 4584–4613, Aug. 2014.
[14] F. Etezadi and A. Khisti, "Delay-constrained streaming of Gauss-Markov sources over erasure channel," in *International Symposium on Information Theory (ISIT), Hong Kong*, 2015.
[15] ——, "Sequential transmission of Markov Sources over burst erasure channels," in *International Zurich Seminar (IZS), Zurich, Switzerland*, 2014.
[16] F. Etezadi, "Streaming of markov sources over burst erasure channels," Ph.D. dissertation, University of Toronto, 2015.
[17] A. Badr, A. Khisti, W. Tan, and J. Apostolopoulos, "Forward error correction for low-delay interactive applications," *Signal Processing Magazine*, 2016, To Appear.
[18] S. Tung, "Multiterminal source coding," Ph.D. dissertation, Cornell University, 1978.
[19] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. John Wiley and Sons, 1991.
[20] I. Csiszàr and J. Korner, "Towards a general theory of source networks," *IEEE Trans. Inform. Theory*, vol. 26, no. 2, pp. 155 – 165, mar 1980.
[21] J. Rosenthal and R. Smarandache, "Maximum distance separable convolutional codes," *Appl. Algebra Engrg. Comm. Comput.*, vol. 10, no. 1, pp. 15–32, 1999.
[22] E. N. Gilbert, "Capacity of a burst-noise channel," *Bell System Technical Journal*, vol. 39, pp. 1253–1265, Sep. 1960.
[23] A. Konrad, B. Y. Zhao, A. D. Joseph, and R. Ludwig, "A markov-based channel model algorithm for wireless networks," *Wireless Networks*, vol. 9, no. 3, pp. 189–199, 2003.
[24] E. O. Elliot, "Estimates of error rates for codes on burst-noise channels," *Bell Syst. Tech. J.*, vol. 42, pp. 1977–1997, Sep. 1963.
[25] H. V. Poor, *An Introduction to Signal Detection and Estimation, 2nd Ed.* NewYork: Springer Verlag.