Prospicient Real-Time Coding of Markov Sources over Burst Erasure Channels: Lossless Case

Farrokh Etezadi*, Ashish Khisti* and Mitchell D. Trott⋆

*University of Toronto
ECE Department
Toronto, ON, M1B 5P1, Canada
{fetezadi,akhisti}@comm.utoronto.ca

⋆HP Labs
MCNL Lab
Palo Alto, USA
michell.trott@hp.com

Abstract- We introduce a framework to study fundamental limits of sequential coding of Markov sources under an error propagation constraint. An encoder sequentially compresses a sequence of vector-sources that are spatially i.i.d. but temporally correlated according to a Markov process. The channel erases up to B packets in a single burst, but reveals all other packets to the destination. The destination is required to reproduce all the source-vectors instantaneously and in a lossless manner, except those sequences that occur in a window of length $B+W$ following the start of the erasure burst.

We define a rate-recovery function $R(B,W)$, the minimum compression rate that can be achieved in this framework, and develop upper and lower bounds for first-order Markov sources. For the special class of *linear diagonally correlated deterministic sources*, we propose a new coding technique — prospicient coding — that achieves the rate-recovery function. Finally, a lossy extension to the rate-recovery function is also studied for a class of Gaussian sources where the source is temporally and spatially i.i.d. and the decoder aims to recover a collection of past K sources with a quadratic distortion measure. The optimal rate-recovery function is compared with the sub-optimal techniques including forward error correction coding (FEC) and Wyner-Ziv coding, and performance gains are quantified.

1 Introduction

A tradeoff between compression efficiency and error resilience is fundamental to any video compression system. In live video streaming, an encoder observes a sequence of correlated video frames and produces a compressed bit-stream that is transmitted to the destination. If the underlying channel is an ideal bit-pipe, it is well known that predictive coding [1] achieves the optimum compression rate. Unfortunately in many emerging video distribution networks, such as peer-to-peer systems and mobile systems, packet losses are unavoidable. Predictive coding is highly sensitive to such packet losses and can lead to a significant amount of error propagation. Various techniques are used in practice to prevent such losses. Commonly used video coding techniques use a group of picture (GOP) architecture, where intra-frames are periodically inserted to limit the effect of error propagation. Forward error correction codes can also be applied to compressed bit-streams to recover from missing packets [2, 3]. Modifications to predictive coding, such as leaky-DPCM [4], have been proposed in the literature to deal with packet losses. The robustness of distributed video coding techniques in presence of packet losses has been studied in e.g., [5].

Information theoretic analysis of video coding has received significant attention in recent times, see e.g., [6, 7, 8] and the references therein. These works focus primarily on the source coding aspects of video. The *source process* is a sequence of vectors, each of which is spatially i.i.d. and temporally correlated. Each source vector is sequentially compressed into a bit stream. The destination is required to recover the source vectors in a sequential manner as well. However all of these works assume an ideal channel with no packet losses. To our knowledge even the effect of a single isolated packet loss is not fully understood [9].

In this work we introduce an information theoretic framework to characterize the tradeoff between error propagation and compression rate. An encoder is revealed source vectors in a sequential manner and compresses them sequentially into channel packets that are then transmitted over a channel. The concept of a *rate-recovery function* is introduced and information theoretic upper and lower bounds are obtained. The upper bound is obtained using a binning based scheme. The lower bound is obtained by drawing an interesting connection to a multi-terminal source coding problem. We introduce a special class of *diagonally correlated deterministic Markov sources* and propose a new coding scheme that establishes the optimality of the lower bound. We also study a family of i.i.d. Gaussian sources with sliding window recovery constraint. The coding scheme for such sources naturally maps to the deterministic source model and enables us to completely characterize the (lossy) rate-recovery function for this class. Finally performance of different sub-optimal systems are also compared to the optimal tradeoff.

The rest of this paper is organized as follows. Section 2 includes the problem description. Section 3 presents the main results of the paper including 1) The lower and upper bound for minimum rate-recovery function for general Markov sources, whose proof sketch is provided in Section 4, 2) The rate-recovery function for two class of sources, i.e. the diagonally correlated deterministic Markov which is explained in details in Section 5 and i.i.d. Gaussian sources with sliding window recovery constraint. Section 6 includes the comparison of the rate-recovery function of optimal and different sub-optimal systems for the Gaussian case. Section 7 concludes the paper.

2 Problem Statement

Source Model: We consider a semi-infinite stationary vector source process $\{s_t^n\}_{t\geq 0}$ whose symbols (defined over some finite alphabet $\mathcal{S}$) are drawn independently across the spatial dimension and from a first-order Markov chain across the temporal dimension, i.e., for each $t \geq 1$,

$$\Pr(\; s_t^n = s_t^n \mid s_{t-1}^n = s_{t-1}^n,\; s_{t-2}^n = s_{t-2}^n, \ldots) = \prod_{j=1}^{n} p_{s_1|s_0}(s_{t,j}|s_{t-1,j}), \quad \forall t \geq 1. \quad (1)$$

We assume that the underlying random variables $\{s_t\}_{t\geq 0}$ constitute a time-invariant and first-order stationary Markov chain with a common marginal distribution denoted by $p_s(\cdot)$.

Channel Model: The channel introduces an erasure burst of size B, i.e. for some particular $j \geq 0$, it introduces an erasure burst such that $g_i = \star$ for $i \in \{j, j+1, \ldots, j+B-1\}$

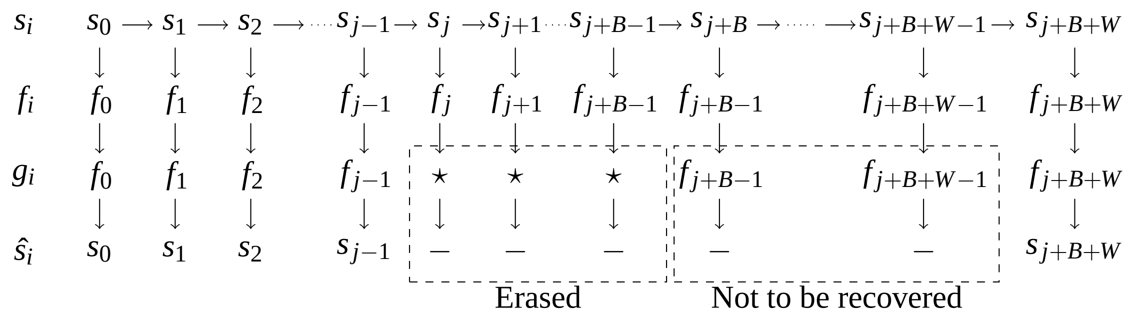

Figure 1: Problem Setup

and $g_i = f_i$ otherwise i.e.,

$$g_i = \begin{cases} \star, & i \in [j, j+1, \ldots, j+B-1] \\ f_i, & \text{else.} \end{cases} \tag{2}$$

Rate-Recovery Function: A rate-R causal encoder maps the sequence $\{s_i^n\}_{i\geq 0}$ to an index $f_i \in [1, 2^{nR}]$ according to some function $f_i = \mathcal{F}_i\left(s_0^n, \ldots, s_i^n\right)$ for each $i \geq 0$. A *memoryless* encoder satisfies $\mathcal{F}_i\left(s_0^n, \ldots, s_i^n\right) = \mathcal{F}_i(s_i^n)$ i.e., the encoder does not use the knowledge of the past sequences.

Upon observing the sequence $\{g_i\}_{i\geq 0}$ the decoder is required to perfectly recover all the source sequences using decoding functions

$$\hat{s}_i^n = \mathcal{G}_i(g_0, g_1, \ldots, g_i), \quad i \notin \{j, \ldots, j+B+W-1\}. \tag{3}$$

where j denotes the time at which the erasure burst starts in (2). It is however not required to produce the source sequences in the window of length $B+W$ following the start of an erasure burst. We call this period the error propagation window. The setup is shown in Fig. 1.

A rate $R(B,W)$ is feasible if there exists a sequence of encoding and decoding functions and a sequence ε_n that approaches zero as $n \to \infty$ such that, $\Pr(s_i^n \neq \hat{s}_i^{\,n}) \leq \varepsilon_n$ for all $i \notin \{j, \ldots, j+B+W-1\}$. We seek the minimum feasible rate $R(B,W)$, which we define to be the *rate-recovery* function.

3 Main Results

In this section we discuss the main results of this paper.

3.1 Upper and Lower Bounds

Theorem 1 *For any stationary first-order Markov source process the rate-recovery function satisfies $R^-(B,W) \leq R(B,W) \leq R^+(B,W)$ where*

$$R^+(B,W) = H(s_1|s_0) + \frac{1}{W+1} I(s_B\,;\, s_{B+1}|s_0) = \frac{1}{W+1} H(s_{B+1}, s_{B+2}, \ldots, s_{B+W+1}|s_0), \tag{4}$$

$$R^-(B,W) = H(s_1|s_0) + \frac{1}{W+1} I(s_B; s_{W+B+1}|s_0). \tag{5}$$

Notice that the upper and lower bound coincide for $W = 0$ and $W \to \infty$, yielding the rate-recovery function in these cases.

The upper bound is obtained via a binning based scheme and a memoryless encoder. At each time the encoding function f_i is the bin-index of a Slepian-Wolf codebook [10]. Following an erasure burst in $[j, j+B-1]$, the decoder collects $f_{j+B}, \ldots, f_{j+W+B}$ and attempts to jointly recover all the underlying sources at $t = j+W+B$. (It can be shown in a straightforward way that the two expressions of the upper bound in (4) are equivalent [11].)

The converse is based on connecting the problem to a source recovery over a periodic erasure channel and a multi-terminal source coding problem as elaborated in Sec. 4.

3.2 Diagonally Correlated Deterministic Markov Sources

We propose a special class of source models for which the lower bound in (5) is tight.

Definition 1 *(Diagonally Correlated Deterministic Sources) The alphabet of a* diagonally correlated deterministic source *consists of K sub-symbols i.e.,*

$$\mathbf{s}_i = (\mathbf{s}_{i,0}, \ldots, \mathbf{s}_{i,K}) \in \mathcal{S}_0 \times \mathcal{S}_1 \times \ldots \times \mathcal{S}_K, \tag{6}$$

where each $\mathcal{S}_i = \{0,1\}^{N_i}$ is a binary sequence. Suppose that the sub-sequence $\{\mathbf{s}_{i,0}\}_{i\geq 0}$ is an i.i.d. sequence sampled uniformly over $\mathcal{S}_0$ and for $1 \leq j \leq K$, the sub-symbol $\mathbf{s}_{i,j}$ is a linear deterministic function[1] *of $\mathbf{s}_{i-1,j-1}$ i.e.,*

$$\mathbf{s}_{i,j} = \mathbf{R}_{j,j-1} \cdot \mathbf{s}_{i-1,j-1}, \qquad 1 \leq j \leq K. \tag{7}$$

for fixed matrices $\mathbf{R}_{1,0}, \mathbf{R}_{2,1} \ldots, \mathbf{R}_{K,K-1}$ each of full row-rank i.e., $\mathrm{rank}(R_{j,j-1}) = N_j$.

Theorem 2 *For the class of Diagonally Correlated Deterministic Sources in Def. 1 the rate-recovery function is given by:*

$$R(B,W) = R^{-}(B,W) = H(s_1|s_0) + \frac{1}{W+1} I(s_B; s_{B+W+1}|s_0). \tag{8}$$

In particular the above class of sources establishes the sub-optimality of the binning based scheme in general. We propose a coding scheme that carefully exploits the *non-causal* knowledge of some future sub-symbols to achieve a lower rate than the binning based scheme. We call this scheme *prospicient coding* which is explained in Section 5.

3.3 Gaussian Sources

Our proposed framework can be easily extended to a continuous valued source process with a fidelity measure. While a complete treatment of the lossy case is beyond the scope of the present paper, we study one natural extension of the Diagonally Correlated Source Model in Def. 1 to a Gaussian sources.

Consider a Gaussian source process that is i.i.d. both in temporal and spatial dimensions, i.e., at time i, a sequence consisting of n symbols s_i^n, is sampled i.i.d. according to a zero mean unit variance Gaussian distribution $N(0,1)$.

[1] All multiplication is over the binary field.

The encoder output at time i is denoted by the index $f_i = \mathcal{F}(s_0^n, \ldots, s_i^n) \in [1, 2^{nR}]$ as before. At time i, upon receiving the channel outputs until time i, the decoder is interested in reproducing a collection of past K sources[2] $\mathbf{t}_i^n = \begin{pmatrix} s_i^n & s_{i-1}^n & \cdots & s_{i-K}^n \end{pmatrix}^T$ within a distortion vector $\mathbf{d} = (d_0, d_1, \cdots, d_K)^T$.

Thus for any $i \geq 0$ and $0 \leq j \leq K$, if $\hat{s}_{i-j}^n$ is the reconstruction sequence of s_{i-j}^n at time i, we must have that $E\left[||s_{i-j}^n - \hat{s}_{i-j}^n||^2\right] \leq nd_j$. We will assume that $d_0 \leq d_1 \leq \cdots \leq d_K$ holds.

As before, the channel can introduce an erasure-burst of length B in an arbitrary interval $[k, k+B-1]$. The decoder is not required to output a reproduction of the sequences $\mathbf{t}_i^n$ for $i \in [k, k+B+W-1]$. A *lossy rate-recovery function* denoted by $R(B, W, \mathbf{d})$ is the minimum rate required to satisfy these constraints.

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

$$R(B, W, \mathbf{d}) = \frac{1}{2}\log\left(\frac{1}{d_0}\right) + \frac{1}{W+1}\sum_{k=1}^{\min\{K-W, B\}} \frac{1}{2}\log\left(\frac{1}{d_{W+k}}\right). \tag{9}$$

The proof of Theorem 3 is available in [11]. The coding scheme involves mapping the Gaussian source to a deterministic source in the previous section via a successively refine-able code.

4 Proof of Theorem 1

To highlight the main idea behind the converse we consider the case when $W = 1$ and $B = 1$. The formal proof is provided [11]. Only the key ideas are described in this paper due to space constraints. Using the first-order Markov Chain property $s_0 \to s_1 \to s_2 \to s_3$ the lower bound (5) can be reduced to:

$$R^-(B = 1, W = 1) = \frac{1}{2}H(s_1|s_0, s_2) + \frac{1}{2}H(s_3|s_0) \tag{10}$$

We interpret the two terms in (10). Consider a *periodic erasure channel* where every third packet gets erased i.e., $g_k = \star$ for $t = 3k$, $k = 0, 1, 2, \ldots$. The destination, upon receiving $g_1 = f_1$ and $g_2 = f_2$ must recover s_2^n at $t = 2$. At this point, because of the first-order Markov nature of the source process, it becomes synchronized with the encoder i.e., the effect of earlier erasures is no longer relevant. Thus it treats the erasure at time $t = 3$ as a fresh erasure. Upon receiving f_4 and f_5 it must recover s_5^n at $t = 5$. More generally, it is able to recover s_{3k+2}^n at $t = 3k+2$ upon sequentially observing $\{f_{3i+1}, f_{3i+2}\}_{0 \leq i \leq k}$ and missing $\{f_{3i}\}_{0 \leq i \leq k}$. From the source coding theorem we must have

$$2kR \geq H(f_1, f_2, f_4, f_5, \ldots, f_{3k-2}, f_{3k-1}) \geq H(s_2^n, s_5^n, \ldots, s_{3k-1}^n) \tag{11}$$

$$\geq n(k-1)H(s_3|s_0) \tag{12}$$

which, upon taking $k \to \infty$ yields $R \geq \frac{1}{2}H(s_3|s_0)$.

[2] If the index of any source sub-sequence is negative, it is treated as an all-zero sequence.

[3] All logarithms are taken to base 2.

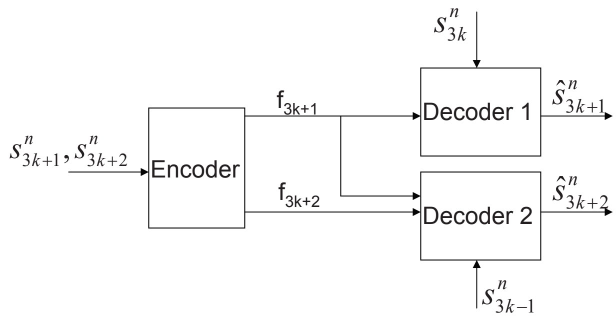

Figure 2: A Multi-terminal Source Coding Problem related to the proposed streaming setup. The erasure at time $t = 3k$ leads to two virtual decoders with different side-information.

The above argument only takes into account one constraint — when there is an erasure, the decoder needs to recover with $W = 1$. As the lower bound (10) suggests, this approach alone is not tight. The additional term of $\frac{1}{2}H(s_1|s_0, s_2)$ is not captured by this simplistic argument.

In Fig. 2 we illustrate a multi-terminal source coding problem with one encoder and two decoders. The encoder is revealed (s^n_{3k+1}, s^n_{3k+2}) and produces outputs f_{3k+1} and f_{3k+2}. Decoder 1 needs to recover s^n_{3k+1} given f_{3k+1} and s^n_{3k} while decoder 2 needs to recover s^n_{3k+2} given s^n_{3k-1} and (f_{3k+1}, f_{3k+2}). As we show in the formal proof, a lower bound for this system constitutes a lower bound to the streaming problem. In particular,

$$
\begin{aligned}
2nR &\geq H(f_{3k+1}, f_{3k+2}) \geq H(f_{3k+1}, f_{3k+2}|s^n_{3k-1}) && (13)\\
&= H(f_{3k+1}, f_{3k+2}, s^n_{3k+2}|s^n_{3k-1}) - H(s^n_{3k+2}|f_{3k+1}, f_{3k+2}, s^n_{3k-1}) && (14)\\
&\geq H(f_{3k+1}, s^n_{3k+2}|s^n_{3k-1}) - n\varepsilon_n && (15)\\
&\geq H(s^n_{3k+2}|s^n_{3k-1}) + H(f_{3k+1}|s^n_{3k+2}, s^n_{3k-1}) - n\varepsilon_n\\
&\geq H(s^n_{3k+2}|s^n_{3k-1}) + H(f_{3k+1}|s^n_{3k+2}, s^n_{3k}, s^n_{3k-1}) - n\varepsilon_n && (16)\\
&\geq H(s^n_{3k+2}|s^n_{3k-1}) + H(s^n_{3k+1}|s^n_{3k+2}, s^n_{3k}, s^n_{3k-1}) - 2n\varepsilon_n && (17)\\
&\geq H(s^n_{3k+2}|s^n_{3k-1}) + H(s^n_{3k+1}|s^n_{3k+2}, s^n_{3k}) - 2n\varepsilon_n && (18)\\
&= nH(s_3|s_0) + nH(s_1|s_2, s_0) - 2n\varepsilon_n && (19)
\end{aligned}
$$

where (15) follows from the fact that s^n_{3k+2} must be recovered from $(f_{3k+1}, f_{3k+2}, s^n_{3k-1})$ at decoder 2 hence Fano's inequality applies and (16) follows from the fact that conditioning reduces entropy. (17) follows from Fano's inequality applied to decoder 1 and finally (18) follows from the Markov chain associated with the source process. Dividing throughout by n in (19) and taking $n \to \infty$ recovers (10). The extension to arbitrary W and B uses similar ideas and the formal proof is provided in [11].

5 Diagonally Correlated Deterministic Sources

In this section we study a deterministic source model with a special diagonal correlation structure. While our results can be extended to a larger class of sources [11] the coding scheme is most natural for this class. Furthermore this class of deterministic sources also

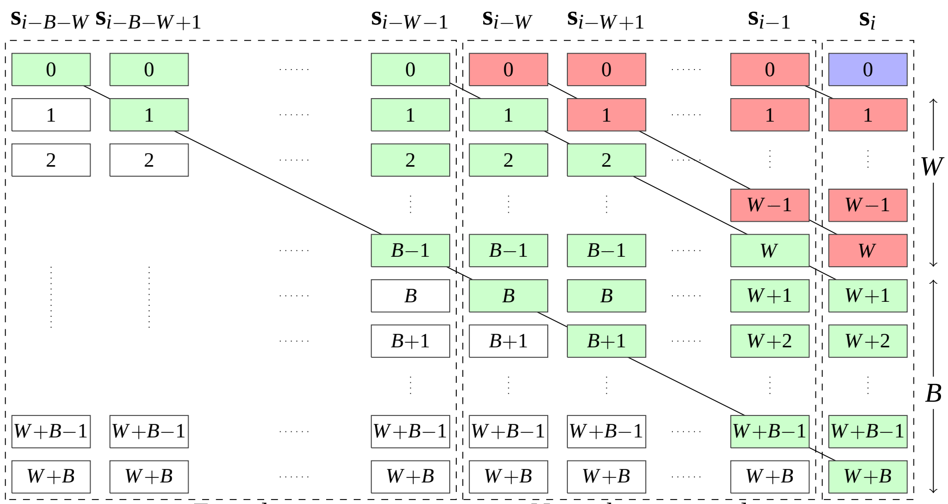

Figure 3: Schematic of Diagonally Correlated Deterministic Markov Source for $K = B + W$. The first row constitutes the innovation sub-symbols whereas the remaining rows are deterministic sub-symbols that follow a diagonal relation as shown in Definition 1.

provides a solution to the Gaussian source model that we will consider subsequently, thus yielding a new coding scheme in that case.

Fig. 3 shows the structure of the source for $K = B+W$. Any diagonal in Fig. 3 consists of linear combinations of the same source sub-symbols. In particular the innovation bits are introduced on the upper-left most entry of the diagonal. As we traverse down, each sub-symbol consists of some fixed linear combinations of these innovation bits. Furthermore the sub-symbol $\mathbf{s}_{i,j}$ is completely determined given the sub-symbol $\mathbf{s}_{i-1,j-1}$ for each $j \in \{1,\ldots,K\}$. It can be easily shown that we can take $K = B+W$ without loss of generality.

A complete proof of Theorem 2 is provided in the full paper [11]. We only sketch the main ideas in this short paper. In particular our code construction consists of two steps as discussed below.

1) Source Re-arrangement: The source symbols $\mathbf{s}_i$ consisting of innovation and deterministic sub-symbols as in Definition. 1 are first rearranged to produce an auxiliary set of codewords

$$\mathbf{c}_i = \begin{pmatrix} \mathbf{c}_{i,0} \\ \mathbf{c}_{i,1} \\ \mathbf{c}_{i,2} \\ \vdots \\ \mathbf{c}_{i,B} \end{pmatrix} = \begin{pmatrix} \mathbf{s}_{i,0} \\ \mathbf{s}_{i+W,W+1} \\ \mathbf{s}_{i+W,W+2} \\ \vdots \\ \mathbf{s}_{i+W,W+B} \end{pmatrix} = \begin{pmatrix} \mathbf{s}_{i,0} \\ \mathbf{R}_{W+1,1}\ \mathbf{s}_{i,1} \\ \mathbf{R}_{W+2,2}\ \mathbf{s}_{i,2} \\ \vdots \\ \mathbf{R}_{W+B,B}\ \mathbf{s}_{i,B} \end{pmatrix}, \tag{20}$$

where the last relation follows from definition. Note that the codeword $\mathbf{c}_i$ consists of the innovation symbol $\mathbf{s}_{i,0}$, as well as symbols $\mathbf{s}_{i+W,W+1},\ldots,\mathbf{s}_{i+W,W+B}$ that enable the recovery of symbols in $\mathbf{s}_{i+W}$.

2) Slepian-Wolf Coding: There is a strong temporal correlation between the auxiliary vector sequences $\{\mathbf{c}_i^n\}$ in (20). Hence we bin codeword sequences $\mathbf{c}_i^n$ into 2^{nR} bins where R is as given in (6) and only transmit the bin index of the associated codeword i.e., $\mathbf{f}_i = \mathcal{F}(\mathbf{c}_i^n) \in \{1, 2, \dots, 2^{nR}\}$, where R is selected to satisfy

$$(W+1)R \geq H(\mathbf{c}_i, \mathbf{c}_{i-1}, ..., \mathbf{c}_{i-W} | \mathbf{s}_{i-B-W-1}) = \sum_{k=0}^{W} H(\mathbf{c}_{i-k,0}) + \sum_{k=1}^{B} H(\mathbf{c}_{i-W,k}) \tag{21}$$

$$= (W+1)N_0 + \sum_{k=1}^{B} N_{W+k}, \tag{22}$$

The above expression is equivalent to Theorem 2 as established in [11].

For analysis of the decoder first assume that a burst-erasure happens between the interval $t \in [i-B-W, i-W-1]$ and the decoder is interested in recovering $\mathbf{s}_i$. The decoder has access to $\mathbf{f}_j$ for $j \in \{i-W, i-W+1, \dots, i\}$ as well as the last decoded source symbol $\mathbf{s}_{i-B-W-1}$. The decoder first recovers all the auxiliary codeword symbols $\{\mathbf{c}_j\}$ for $j \in [i-W, i]$ from the corresponding bin indices. The constraint in (21) guarantees that this step succeeds with high probability. Next by construction of $\mathbf{c}_j$ in (20), the decoder recovers the last B sub-symbols i.e., $\mathbf{s}_{i,W+1}, \dots, \mathbf{s}_{i,B+W}$ from $\mathbf{c}_{i-W}$. Finally the remaining sub-symbols $\mathbf{s}_{i,0}, \mathbf{s}_{i,1}, \dots, \mathbf{s}_{i,W}$ are recovered from the innovation part of $\mathbf{c}_i, \mathbf{c}_{i-1}, \dots, \mathbf{c}_{i-W}$ respectively.

If the erasure burst does not happen in $t \in [i-B-W, i-W-1]$ and if the receiver needs to recover $\mathbf{s}_i$ then observe that $\mathbf{s}_{i-1}$ is guaranteed to be available. In this case the codeword $\mathbf{c}_i$ can be recovered directly from the bin index $\mathbf{f}_i$ and $\mathbf{s}_{i-1}$ due to (21), and in turn the innovation part of $\mathbf{s}_i$ can be recovered. We refer the reader to the full paper [11] for complete details. The converse follows from Thoerem 1 and is also provided in the full paper.

6 Theorem 3 (Discussion)

The complete proof of Theorem (3) is available at [11]. This section contains the comparison of the optimal performance of Theorem 3 with the following sub-optimal systems.

- **Still-Image Compression:** In this scheme, the encoder ignores the decoder's memory and at time $i \geq 0$ encodes the source $\mathbf{t}_i$ in a memoryless manner and sends the codewords through the channel. The rate associated to this scheme is

$$R_{\text{SI}}(\mathbf{d}) = I(\mathbf{t}_i; \hat{\mathbf{t}}_i) = \sum_{k=0}^{K} \frac{1}{2} \log\left(\frac{1}{d_k}\right) \tag{23}$$

  In this scheme, the decoder is able to recover the source whenever its codeword is available, i.e. at all the times except the erasure period.

- **Wyner-Ziv Compression:** At time i the encoders assumes that $\mathbf{t}_{i-B-1}$ is already reconstructed at the receiver within distortion $\mathbf{d}$. With this assumption, it compresses the source $\mathbf{t}_i$ according to Wyner-Ziv scheme with $\hat{\mathbf{t}}_i$ as the side-information and

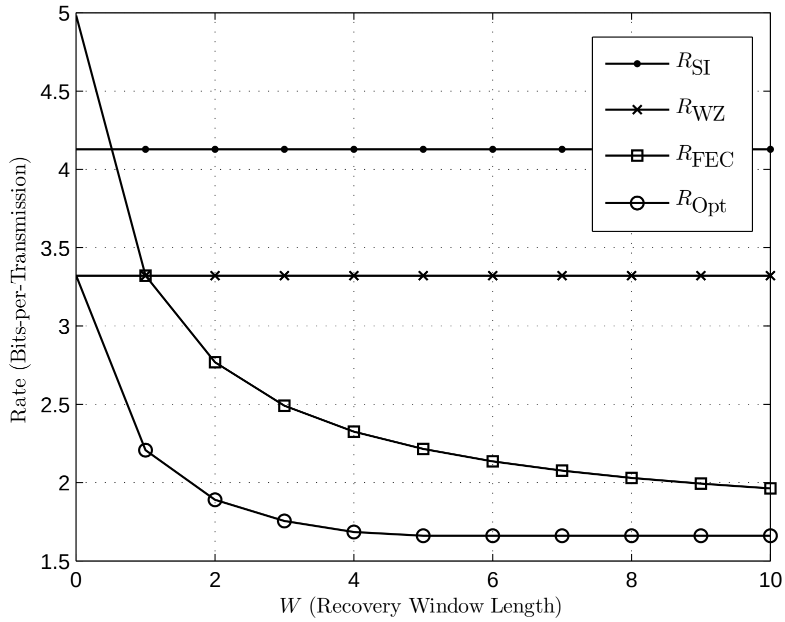

Figure 4: Comparison of rate-recovery of sub-optimal systems to minimum possible rate-recovery function for different recovery window length W.

transmits the codewords through the channel. The rate of this scheme is

$$R_{\text{WZ}}(B,\mathbf{d}) = I(\mathbf{t}_i;\hat{\mathbf{t}}_i|\hat{\mathbf{t}}_{i-B-1}) = \sum_{k=0}^{B} \frac{1}{2}\log\left(\frac{1}{d_k}\right) \tag{24}$$

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

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

$$R_{\text{FEC}}(B,W,\mathbf{d}) = \frac{B+W+1}{W+1} I(\mathbf{t}_i;\hat{\mathbf{t}}_i|\hat{\mathbf{t}}_{i-1}) = \frac{B+W+1}{2(W+1)}\log\left(\frac{1}{d_0}\right) \tag{25}$$

In Fig 4, the rate-recovery functions of explained sub-optimal schemes are compared to the minimum rate-recovery function. We assume $K=5$, $B=2$ and the distortion vector $\mathbf{d} = \{0.1, 0.25, 0.4, 0.55, 0.7, 0.85\}^T$. It can be observed from Fig 4 that for $W=0$, Wyner-Ziv schemes, as expected, is optimal. Note that Predictive Coding plus FEC scheme is commonly used in practice. Fig 4, exhibits the sub-optimality of the scheme even for reasonably large W.

7 Conclusion

We introduce an analytical framework to characterize a fundamental tradeoff between error propagation and compression rate for real-time coding of Markov sources over erasure channels. A new concept of a rate-recovery function is introduced and novel lower and upper bounds are developed.

Optimum rate-recovery function for two families of Markov sources — the deterministic diagonally correlated Markov source and i.i.d. Gaussian source with sliding window reconstruction constraint were obtained. The optimal coding scheme for such sources involves a pre-selection step that improves the performance over a binning-based scheme. For the Gaussian case, a number of sub-optimal systems were compared to the optimum rate-recovery function and performance gains were quantified.

References

[1] T. Berger, *Rate Distortion Theory, A Mathematical Basis for Data Compression.* NJ: Prentice-Hall Englewood Cliffs, 1971.

[2] 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.

[3] Z. Li, A. Khisti, and B. Girod, "Forward error correction for low-delay packet video," in *Packet Video Workshop*, 2010.

[4] Y. Huang, Y. Kochman, and G. Wornell, "Causal transmission of colored source frames over a packet erasure channel," in *DCC*, 2010, pp. 129–138.

[5] A. Sehgal, A. Jagmohan, and N. Ahuja, "Wyner-ziv coding of video: An error-resilient compression framework," *IEEE Trans. on Multimedia*, vol. 6, pp. 249–258, 2004.

[6] H. Viswanathan and T. Berger, "Sequential coding of correlated sources," *Information Theory, IEEE Transactions on*, vol. 46, no. 1, pp. 236 –246, jan 2000.

[7] N. Ma and P. Ishwar, "On delayed sequential coding of correlated sources," *IEEE Trans. Inform. Theory*, vol. 57, pp. 3763–3782, 2011.

[8] E. Yang, L. Zheng, D. He, and Z. Zhang, "Rate distortion theory for causal video coding: Characterization, computation algorithm, and comparison," *IEEE Trans. Inform. Theory*, vol. 57, pp. 5258–5280, 2011.

[9] J. Wang, V. Prabhakaran, and K. Ramchandran, "Syndrome-based robust video transmission over networks with bursty losses," in *ICIP*, Atlanta, GA, 2006.

[10] D. Slepian and J. K. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. Inform. Theory*, vol. 19, pp. 471–480, July 1973.

[11] F. Etezadi, A. Khisti, and M. D. Trott, "Prospicient Real-Time Coding of Markov Sources over Burst Erasure Channels," *Under Preparation*.