Real-Time Coding of Gauss-Markov Sources over Burst Erasure Channels

Farrokh Etezadi and Ashish Khisti
University of Toronto
10 King's College Road
Toronto, ON, Canada
Email: {akhisti, fetezadi}@comm.utoronto.ca

***Abstract*—We study real-time coding of Gauss-Markov sources over burst-erasure channels. A time-invariant encoder sequentially compresses a sequence of vector Gaussian sources, which are spatially i.i.d. and temporally correlated according to a Gauss-Markov model. The channel is a burst erasure channel that erases up to B packets in a single burst. The decoder is interested in instantaneously reconstructing all the source sequences within a quadratic (mean square error) distortion D, except for packets lost in the erasure burst window. We study the minimum achievable rate for the encoder under these constraints and define it as the lossy rate-recovery function $R(B, D)$. We develop lower and upper bounds for the function and observe that the bounds coincide in the high resolution limit. Numerical comparisons indicate that the proposed joint source-channel coding scheme provides significant gains over a separation based scheme.**

I. Introduction

We study fundamental limits of real-time video streaming, where the encoder sequentially encodes temporally correlated video frames. When the channel is an ideal bit-pipe, it is well-known that predictive coding [1] is optimal. However in many video distribution systems, such as peer-to-peer and mobile systems, packet losses are unavoidable. Predictive coding can lead to a significant amounts of error propagation. Various methods have been suggested to develop robust video compression techniques in order to combat against the packet losses in such systems. Some examples include error control codes in (e.g., [2], [3]), leaky-DPCM (e.g. [4], [5]), distributed video coding in e.g. [6], etc. However fundamental limits of such systems are not fully understood.

Information theoretic analysis of video coding problem has been studied in [7], [8], [9]. All of these works consider the problem from a source coding prospective. The channel is an ideal bit-pipe. More recently in [10], we consider the case when the encoder sequentially encodes each source sequence, the channel introduces an erasure-burst and the decoder is interested in instantaneously recovering all of the source sequences in a lossless manner, except those that fall in an error propagation window following the erasure burst. The notion of *rate-recovery function* is introduced to capture the trade-off between compression rate and the error propagation at the decoder. Upper and lower bounds on this function are obtained that coincide in some special cases.

In this paper we extend our results in [10] to the case of Gauss-Markov sources. The decoder is required to reconstruct the source sequences within a certain fidelity measured according to the quadratic (mean square error) distortion measure. The minimum attainable rate for a time-invariant sequential encoder is studied and the lower and upper bounds are derived. The proposed scheme is compared to a separation based technique and observed to have significant gains over the latter. In the high resolution limit our bounds coincide yielding a tight result.

The rest of the paper is organized as follows. The problem setup is described in Section II. Section III includes the main results of the paper. The proof of the lower bound on lossy rate-recovery function and the achievability of the coding scheme are presented in Sections IV and V, respectively. The comparison of the performance with sub-optimal schemes are also presented in Section V. Section VI contains the asymptotic results of high resolution scheme. Conclusions are provided in Section VII.

II. Problem Statement

A. Source Model

We consider a semi-infinite stationary vector source process $\{s_t^n\}_{t\geq 0}$ whose real-valued symbols are drawn independently across the spatial dimension according to a marginal distribution equivalent to a zero-mean unit-variance Gaussian distribution and form a first-order Markov chain across the temporal dimension, i.e. the consequent sources are temporally correlated according to the following model.

$$s_t^n = \rho s_{t-1}^n + n_t^n, \tag{1}$$

where $0 \leq \rho \leq 1$ is the correlation coefficient and n_t^n are drawn i.i.d. according to zero-mean Gaussian distribution with the variance $1-\rho^2$, i.e. $N(0, 1-\rho^2)$, independent of s_{t-1}^n.

B. Encoder

A rate-R causal encoder maps the sequence $\{s_t^n\}_{t\geq 0}$ to an index $f_t \in [1, 2^{nR}]$ according to some function

$$f_t = \mathcal{F}_t\left(s_0^n, ..., s_t^n\right) \tag{2}$$

for each $t \geq 0$. In general, the decoder is time-variant, however in this paper we also consider the steady state case where the decoders are time-invariant and the system is in steady state at time $t = 0$ i.e., we assume that the system begins operation at $t = -\infty$ and there are no erasures until time $t = 0$ so that the system is in steady state up to that point. We will however suppress the availability of encoded packets for $t < 0$ for this case.

C. 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 $g_t = \star$ for $t \in \{j, j+1, ..., j+B-1\}$ and $g_t = f_t$ otherwise i.e.,

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

D. Lossy Rate-Recovery Function

Upon observing the sequence $\{g_t\}_{t\geq 0}$, the decoder is required to generate the reproduction of all the source sequences s_t^n, denoted by $\hat{s}_t^n \in \mathbb{R}^n$, using decoding functions

$$\hat{s}_t^n = \mathcal{G}_t(g_0, g_1, \ldots, g_t), \quad t \notin \{j, \ldots, j+B-1\}. \tag{4}$$

except for a window of length B from the time of channel erasure and within average quadratic distortion D, i.e. the reproduction sequence $\hat{s}_t^n$ need to satisfy

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} E\{(s_{t,k} - \hat{s}_{t,k})^2\} \leq D \tag{5}$$

for all $t \notin \{j, \ldots, j+B-1\}$.

The setup is shown in Fig. 1.

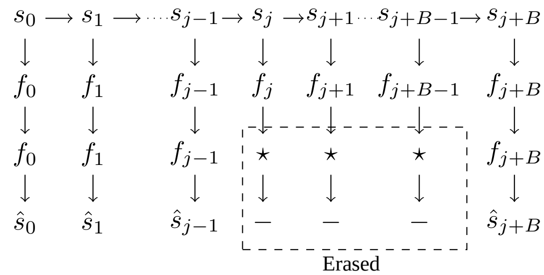

Fig. 1. Problem Setup: The encoder output f_j is a function of the source sequences up to time j i.e., $s_0^n, s_1^n, \ldots, s_j^n$. The channel introduces an erasure burst of length B. The decoder produces $\hat{s}_j^n$ upon observing the sequence g_0^j. The decoder is also required to recover the source sequence immediately after the channel erasure.

A rate R is feasible if such sequence of encoding and decoding functions exist. We seek the minimum feasible rate $R(B, D)$, which we define to be the *lossy rate-recovery* function.

III. Main Results

The following Theorems characterizes the lossy rate-recovery function for the Gauss-Markov source model.

Theorem 1. *The lossy rate-recovery function of the Gauss-Markov source satisfies* [1]

$$R(B, D) \geq R^-(B, D) \triangleq \left[\frac{1}{2}\log\left(\frac{1-\rho^{2(B+1)}}{D}\right)\right]^+ \tag{6}$$

where $[x]^+ = \max\{x, 0\}$. □

Theorem 2. *For time-invariant encoders where the system is in steady state, the lossy rate-recovery function of the Gauss-Markov source satisfies*

$$R(B, D) \geq R_s^-(B, D) \tag{7}$$

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

where

$$\Delta = (D\rho^2 + 1 - \rho^{2(B+1)})^2 - 4D\rho^2(1-\rho^{2B}). \tag{9}$$

[1]All the logarithms are taken to base 2.

$\square$

The proof of the Theorems 1 and 2 is presented in Section IV. Now define the following functions.

$$\frac{1}{y(x)} = \frac{1}{x} + \frac{1}{\Sigma_\infty(x)} - 1. \quad (10)$$

$$\Sigma_\infty(x) = \frac{1}{2}\sqrt{((x-1)(1-\rho^2))^2 + 4x(1-\rho^2)} - (1-\rho^2)(\frac{x}{2} - 1) \quad (11)$$

Note that, as it will become clear in sequel, $\Sigma_\infty(x)$ is the MMSE reconstruction error of estimating the source at time t when all the quantized sources till time $t-1$ are available and the signal-to-noise ratio in the test-channel equals $1/x$.

The following theorem provides an achievable lossy rate-recovery function.

Theorem 3. *For time-invariant encoders where the system is in steady state, the lossy rate-recovery function is upper bounded as follows.*

$$R(B,D) \leq R^+(B,D) = \frac{1}{2}\log\left(\frac{1}{D} - \frac{\rho^{2(B+1)}}{D(1+y(x^*))}\right) \quad (12)$$

where $x^* \geq 0$ *is the unique solution of the following equation.*

$$D(1+y(x))(1+x) - (1+y(x)-\rho^{2(B+1)})x - D\rho^{2(B+1)} = 0. \quad (13)$$

$\square$

The proof of the Theorems 3 is presented in Section V. In Fig. 2 the lower and upper bounds of lossy rate-recovery function for $D = 0.2, 0.3$ and $B = 1, 2$ are shown as a function of ρ (the correlation between the sources). Obviously decreasing distortion D and increasing channel erasure length B increases the lossy rate-recovery function. Fig. 3 shows the lower and upper bounds as a function of D (distortion).

The following Corollary characterizes the behavior of lossy rate-recovery function in high resolution regime.

Corollary 1. *For time-invariant encoders where the system is in steady state, in high resolution scheme the lossy rate-recovery function satisfies the following.*

$$R(B,D) = \frac{1}{2}\log\left(\frac{1-\rho^{2(B+1)}}{D}\right) + o(D). \quad (14)$$

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

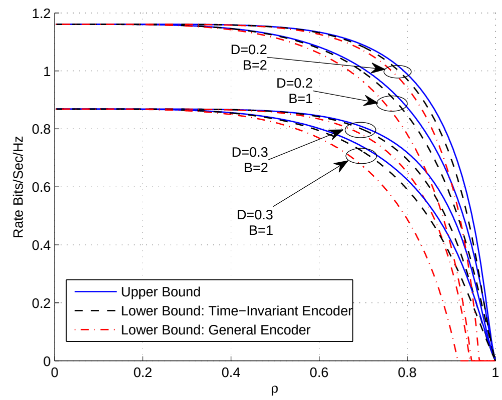

Fig. 2. Lower and Upper Bounds of Achievable Rate for $D = 0.2$, $D = 0.3$ and $B = 1$, $B = 2$.

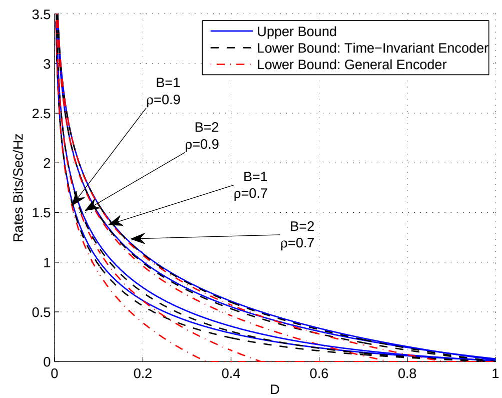

Fig. 3. Lower and Upper Bounds of Achievable Rate for $\rho = 0.9$, $\rho = 0.7$ and $B = 1$, $B = 2$.

Corollary 1 indicates that in high resolution the lower and upper bounds of Theorems 2 and 3 coincides which are equal to the lossy rate-recovery function.

IV. Lower Bounds for Lossy Rate-Recovery Function

A. General Encoders: Proof of Theorem 1

According to the problem description, in the case of channel erasure burst, the decoder is required to reconstruct the source sequences immediately after the erasure burst and all the future times within the required distortion. In order to derive a lower bound for the rate, we consider a sequence of encoded packets separated by $p = B + 1$ units i.e., we consider a periodic erasure channel of period p where the first B packets are erased and the last packet is received to the decoder.

By considering $(t+1)$ such periods note that:

$$n(t+1)R \geq H\left(f_B, f_{p+B}, \ldots, f_{(t-1)p+B}, f_{tp+B}\right)$$
$$= H(f_B) + \sum_{k=1}^{t} H\left(f_{kp+B}|f_B, f_{p+B}, \ldots, f_{(k-1)p+B}\right) \quad (15)$$

Now consider the joint entropy of

$$H(f_{kp+B}|f_B, f_{p+B}, \ldots, f_{(k-1)p+B})$$
$$\geq H(f_{pk+B}|f_0^{kp-1}) \quad (16)$$
$$= I(f_{kp+B}; s^n_{pk+B}|f_0^{kp-1}) + H(f_{pk+B}|s^n_{pk+B}, f_0^{kp-1})$$
$$\geq h(s^n_{kp+B}|f_0^{kp-1}) - h(s^n_{pk+B}|f_0^{kp-1}, f_{pk+B}) \quad (17)$$

where (16) follows from the fact that conditioning reduces the entropy.

For first term in (17) we can write

$$h(s^n_{kp+B}|f_0^{kp-1}) \geq h(s^n_{kp+B}|f_0^{kp-1}, s^n_{kp-1}) \quad (18)$$
$$\geq h(s^n_{kp+B}|s^n_{kp-1}) \quad (19)$$
$$= nh(s_{B+1}|s_0)$$
$$= \frac{n}{2}\log(2\pi e(1-\rho^{2(B+1)})) \quad (20)$$

where (18) follows from the fact that conditioning reduces the entropy, (19) follows from the Markov property

$$f_0^{kp-1} \to s^n_{kp-1} \to s^n_{kp+B} \quad (21)$$

and (20) is based on the fact that

$$s_{B+1} = \rho^{B+1} s_0 + \tilde{n} \quad (22)$$

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

For the second term in (17) note that according to the problem set up one can write

$$h(s^n_{pk+B}|f_0^{kp-1}, f_{pk+B}) =$$
$$h(s^n_{pk+B} - \hat{s}^n_{pk+B}|f_0^{kp-1}, f_{pk+B}) \leq h(s^n_{pk+B} - \hat{s}^n_{pk+B})$$
$$\leq \frac{n}{2}\log(2\pi e D). \quad (23)$$

Where the last step follows from the fact that Gaussian distribution attain highest entropy among all the probability distributions with similar variance.

By replacing (20) and (23) into (17) and replacing the result into (15) and letting $t \to \infty$, the lower bound of (6) in Theorem 1 is derived.

B. Time-Invariant Encoders: Proof of Theorem 2

Now consider the system in which encoders are time-invariant and the system is in steady state at time $t = 0$ i.e., we assume that the system begins operation at $t = -\infty$ and there are no erasures until time $t = 0$ so that the system is in steady state up to that point. Here we follow the same steps for deriving the lower bound as Section IV-A, except some slight modification which helps in improving the lower bound. The assumption that the system is in steady state allows to derive tighter lower bound on the first term in (17). The following lemma is helpful.

Lemma 1. *For $a > b \geq 0$, we have*

$$h(s^n_a|f_0^b) \geq$$
$$\frac{n}{2}\log\left(2\pi e\left(\rho^{2(a-b)}\frac{1-\rho^2}{2^{2R}-\rho^2} + 1 - \rho^{2(a-b)}\right)\right) \quad (24)$$

Proof: See Appendix A. ■

According to Lemma 1, we have

$$h(s^n_{kp+B}|f_0^{kp-1}) \geq \frac{n}{2}\log(2\pi e) +$$
$$\frac{n}{2}\log\left(\rho^{2(B+1)}\frac{(1-\rho^2)}{2^{2R}-\rho^2} + (1-\rho^{2(B+1)})\right). \quad (25)$$

By replacing (25) and (23) into (17) and replacing the result into (15) and letting $t \to \infty$, the following can be recovered.

$$D2^{4R} - (D\rho^2 + 1 - \rho^{2(B+1)})2^{2R} + \rho^2(1-\rho^{2B}) \geq 0 \quad (26)$$

and by solving (26) for R, (8) is derived. This completes the proof of Theorem 2.

V. Coding Scheme

A. Proof of Theorem 3

The coding scheme is based on Wyner-Ziv coding scheme [11]. In particular, the encoder at each time t quantizes the Gaussian source sequence s^n_t to generate an auxiliary random variable u^n_t jointly typical with the source. In addition the encoder randomly and independently assigns 2^{nR} bin indices to all possible quantized sequence u^n_t and while observing u^n_t transmits its bin index, denoted by f_t, through the channel.

The test channel associated with our quantizer is

$$u_t = s_t + z_t \quad (27)$$

where $z_t \sim N(0, x)$ is drawn independent of s_t. Clearly the Markov chain $u_t \to s_t \to s_{t+1} \to u_{t+1}$ holds for $t > 0$.

Assume that the erasure with the length B starts at time t. At any time $l \geq t + B$, the decoder has access to the codewords $\{f_0, f_1, \ldots, f_{t-1}, f_{t+B}, f_{t+B+1}, \ldots, f_l\}$ and is interested in recovering $\hat{s}^n_l$ as an estimate of s^n_l within distortion D. According to Wyner-Ziv scheme, the decoder succeeds in recovering u_l^n at time l from the bin indices $\{f_0, \ldots, f_{t-1}, f_{t+B}, f_{t+B+1}, \ldots, f_l\}$ with high probability if the rate R satisfies

$$R \geq I\left(s_l \; ; \; u_l \mid u_0^{t-1}, u_{t+B}^{l-1}\right) \triangleq f_l(x, B). \quad (28)$$

where $u_a^b = \{u_a, u_{a+1}, \ldots, u_b\}$. On the other hand, Define the estimation error as

$$\lambda_l(x, B) \triangleq \frac{1}{n}\sum_{k=1}^{n} E\{(s_{l,k} - \hat{s}_{l,k})^2 \mid u_{0,k}, \ldots, u_{t-1,k}, u_{t+B,k}, \ldots, u_{l,k}\}.$$

Then, for any $l \geq t+B$, the estimation errors have to satisfy $\lambda_l(x, B) \leq D$. In other words, the following conditions have to be satisfied

$$R \geq \max_{l \geq t+B} f_l(x, B), \quad (29)$$

$$D \geq \max_{l \geq t+B} \lambda_l(x, B). \quad (30)$$

Now consider the following lemma.

Lemma 2. *For any $l > t+B$,*

$$f_l(x, B) \leq f_{t+B}(x, B), \quad (31)$$

$$\lambda_l(x, B) \leq \lambda_{t+B}(x, B). \quad (32)$$

Proof: See Appendix B. ■

According to Lemma 2, (29) reduces to

$$R \geq f_{t+B}(x, B) = I\left(s_{t+B} \; ; \; u_{t+B} \mid u_0^{t-1}\right) \quad (33)$$

$$= h\left(s_{t+B} \mid u_0^{t-1}\right) - h\left(s_{t+B} \mid u_{t+B}, u_0^{t-1}\right) \quad (34)$$

This indicates that any rate R satisfying (34) also satisfies (28) for any $l > t+B$, which guarantees the success of the decoder in recovering u_l at time l. On the other hand, (32) indicates that if the decoder is able to recover $\hat{s}_{t+B}$ within distortion $\lambda_{t+B}(x, B) \leq D$, it will be able to recover $\hat{s}_l$ within distortion $\lambda_l(x, B) \leq D$, for $l > t+B$.

According to these arguments, the following rate is achievable for the problem.

$$R^+(B, D) = \min_x \quad f_{t+B}(x, B) \quad \text{subject to} \quad \lambda_{t+B}(x, B) \leq D \quad (35)$$

It is not hard to show that the function $f_0(x, B)$ attains its minimum when the inequality constraint satisfies equality, i.e. for $x = x^*$ where $\lambda_{t+B}(x^*, B) = D$. This equation is equivalent to (13) in Theorem 3. Thus, (35) reduces to

$$R^+(B, W) = f_0(x^*, B) \quad (36)$$

Now it remains to evaluate (36) in order to show (12) in Theorem 3.

To evaluate $f_{t+B}(x, B)$ in (34) we let $\tilde{s}_{t-1}$ to be the MMSE estimate of s_{t-1} given u_0^{t-1}. Since the decoder is in steady state at time $t-1$ we can express

$$\tilde{s}_{t-1} = \tilde{\alpha}_x s_{t-1} + \tilde{n}_{t-1} \quad (37)$$

where $\tilde{n}_{t-1} \sim \mathcal{N}(0, \tilde{\sigma}_x^2)$. Define $y(x) = \tilde{\sigma}_x^2/\tilde{\alpha}_x^2$. To compute the noise variance we note that the system until time $t-1$ is a Kalman filter that follows

$$s_i = \rho s_{i-1} + n_i, \qquad n_i \sim N(0, 1-\rho^2) \quad (38)$$

$$u_i = s_i + z_i, \qquad z_i \sim N(0, x) \quad (39)$$

According to Kalman filter model, s_i can be viewed as state of the system updated according a Gauss-Markov model and u_i as output of the system at each time i, which is a noisy version of state s_i. It is known that, the decoder at time $t-1$ while observing all the previous outputs upto time $\{t-1\}$, is able to estimate the source s_{t-1}^n with the following error

$$\sigma_{t-1|t-1} = \frac{\Sigma_\infty(x)}{\frac{1}{x}\Sigma_\infty(x) + 1} \quad (40)$$

where $\sigma_{t-1|t-1}$ is the error of estimating s_{t-1} at time $t-1$. The expression for $\Sigma_\infty(x)$ defined in (11) is well known.

For the test channel of (37) the MMSE error of estimating s_{t-1} equals $\sigma_T = y(x)/(1+y(x))$. By setting $\sigma_T = \sigma_{t-1|t-1}$ in (40), the expression (10) is derived.

By applying the MMSE estimator and using the fact that $\tilde{s}_{t-1}$ is a sufficient statistic for s_j for $j \geq t$,

$$h\left(s_{t+B} \mid u_0^{t-1}\right) = h\left(s_{t+B} \mid \tilde{s}_{t-1}\right) \quad (41)$$

$$= \frac{1}{2}\log\left(2\pi e\left(1 - \frac{\rho^{2(B+1)}}{1+y(x)}\right)\right). \quad (42)$$

Furthermore note that, the estimation error $\lambda_{t+B}(x, B)$ can be written as

$$\lambda_{t+B}(x, B) = \frac{1}{n}\sum_{k=1}^{n} \mathrm{E}\{(s_{t+B,k} - \hat{s}_{t+B,k})^2 \mid u_{0,k}, \ldots, u_{t-1,k}, u_{t+B,k}\}$$

$$= \frac{1}{n}\sum_{k=1}^{n} \mathrm{E}\{(s_{t+B,k} - \hat{s}_{t+B,k})^2 \mid \tilde{s}_{t-1,k}, u_{t+B,k}\} \quad (43)$$

$$= \frac{1 - \frac{\rho^{2(B+1)}}{1+y(x)}}{1 + \frac{1}{x} - \frac{\rho^{2(B+1)}}{x(1+y(x))}} \leq D, \quad (44)$$

where (44) follows from the fact that the decoder is required to reconstruct the source s_{t+B}^n, from $\{u_0^n, \ldots, u_{t-1}^n, u_{t+B}^n\}$ within distortion D.

Note that the MMSE estimator operates on jointly Gaussian signals and thus is the optimal estimator, i.e.

$$h\left(s_{t+B} \mid u_0^{t-1}, u_{t+B}\right) = \frac{1}{2}\log(2\pi e \cdot \lambda_{t+B}(x,B)). \tag{45}$$

Accordingly, from (34) we have that

$$f_{t+B}(x,B) = \frac{1}{2}\log\left(\frac{1}{\lambda_{t+B}(x,B)}(1-\frac{\rho^{2(B+1)}}{1+y(x)})\right) \tag{46}$$

$$= \frac{1}{2}\log\left(1+\frac{1}{x}-\frac{\rho^{2(B+1)}}{x(1+y(x))}\right) \tag{47}$$

By replacing $x = x^*$ into (46) and noting that $\lambda_{t+B}(x^*, B) = D$, the expression in (12) is derived which completes he proof of the theorem.

B. Comparison with Sub-Optimal Schemes

In this section, we compare the lower and upper bounds on optimal lossy rate-recovery function with sub-optimal schemes as follows.

1) Still Image Compression: In this scheme, the encoder ignores the decoder's memory and at time $t \geq 0$ and encodes the source s_i^n in a memoryless manner and sends the codewords through the channel. The rate associated to this scheme is

$$R_{\text{SI}} = I(s_t; u_t) = \frac{1}{2}\log_2(\frac{1}{D}). \tag{48}$$

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

2) Source-Channel Separation-Based Scheme: This scheme consists of predictive coding (DPC) 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 the available codeword, the rate of this scheme can be computed as follows.

$$R_{\text{FEC}} = (B+1)R^+(B=0,D). \tag{49}$$

Fig. 4 shows the rate performance of these sub-optimal systems as well as lower and upper bounds on optimal lossy rate-recovery function as a function of distortion D, for $\rho = 0.9$ and $B = 1$. In can be observe that both still image compression and source-channel separation based method are quite sub-optimal methods and the achievable scheme based on Wyner-Ziv outperforms the other two schemes.

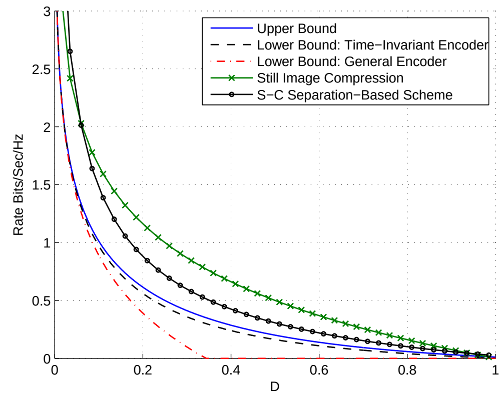

Fig. 4. Comparison of achievable rates of sub-optimal schemes with the Lower and Upper Bounds of lossy rate-recovery function for $\rho = 0.9$ and $B = 1$.

VI. High Resolution Scheme, Proof Sketch for Corollary 1

We show the high resolution results by computing the limit of the upper and lower bounds on lossy rate-recovery function in Theorems 2 and 3 when D approaches to 0 and showing that the two limits coincides.

The high resolution behavior of the lower bound function $R^-(B,D)$, in (8) is straightforward, i.e. it is not hard to observe that

$$\lim_{D\to 0}\left(R^-(B,D) - \frac{1}{2}\log(\frac{1-\rho^{2(B+1)}}{D})\right) = 0 \tag{50}$$

Maybe the more insightful way to investigate the high resolution behavior of the upper bound $R^+(B,D)$ is to note that

$$\frac{1}{2}\log(2\pi e D) \geq h(s_t|u_0^t) \geq h(s_t|u_0^t, s_{t-1})$$

$$= h(s_t|u_t, s_{t-1}) \tag{51}$$

$$= \frac{1}{2}\log\left(\frac{2\pi e}{1/x + 1/(1-\rho^2)}\right) \tag{52}$$

where (51) follows from the Markov chain property $u_0^{t-1} \to u_t, s_{t-1} \to s_t$. From (52) it can be seen that $D \to 0$ requires $x \to 0$. In particular

$$x \leq \frac{D}{1 - D/(1-\rho^2)}. \tag{53}$$

On the other hand when $x = \sigma^2/\alpha^2 \to 0$, the quantized version of the sources at each time become very close to the original source sequences. Thus, we have the following approximations

$$h(s_{t+B}|u_0^{t-1}) \approx h(s_{t+B}|s_{t-1})$$

$$= \frac{1}{2}\log\left(2\pi e(1-\rho^{2(B+1)})\right) \tag{54}$$

and

$$h(s_{t+B}|u_0^t, u_{t+B}) \approx h(s_{t+B}|u_{t+B}) \quad \approx \frac{1}{2}\log(2\pi e x) \tag{55}$$

$$\approx \frac{1}{2}\log(2\pi e D) \tag{56}$$

where (56) follows from replacing (53) into (55) for $D \to 0$. By computing the limit of the achievable rate ($R^+(B, W, D)$) of Theorem 3 when D approaches to 0, it can be obtained that

$$\lim_{D\to 0}\left(R^+(B, D) - \frac{1}{2}\log(\frac{1-\rho^{2(B+1)}}{D})\right)$$
$$= \lim_{D\to 0}\left(I(s_{t+B}; u_{t+B}|u_0^t) - \frac{1}{2}\log(\frac{1-\rho^{2(B+1)}}{D})\right) = 0 \tag{57}$$

(50) together with (57) characterizes the high resolution behavior of lossy rate-recovery function as in Corollary 1.

VII. Conclusion

In this paper, we investigated the real-time encoding of Gauss-Markov sources where the source sequences are spatially i.i.d. and temporally distributed according to a first order Gauss-Markov model. The encoder sequentially encodes the source sequences and sends the codewords through the channel which introduces single burst erasure of length B. The decoder aims to causally reconstruct the source sequences within a specific distortion except for erasure times. The minimum rate attainable for this problem is introduced as lossy rate-recovery function and the lower and upper bounds for this function is derived. It is also shown that in high resolution regime when the system is in steady state, the lower and upper bounds coincides.

Our future work will attempt to extend the results to the case when an error propagation window is allowed following the erasure burst.

Appendix A
Proof of Lemma 1

According to the Gauss-Markov source model for $a > b \geq 1$,

$$s_a^n = \rho^{a-b} s_b^n + \tilde{n}^n \tag{58}$$

where $\tilde{n}^n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1-\rho^{2(a-b)})$. By applying Shannon's EPI we have

$$h(s_a^n|f_0^b)$$
$$\geq \frac{n}{2}\log\left(\rho^{2(a-b)} 2^{\frac{2}{n}h(s_b^n|f_0^b)} + 2\pi e(1-\rho^{2(a-b)})\right). \tag{59}$$

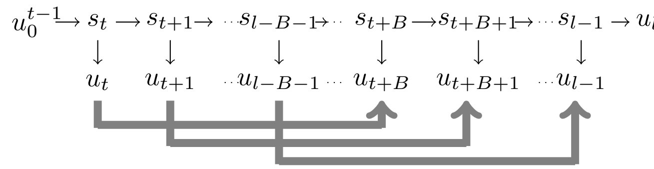

Fig. 5. Knowing u_{t+B}^{l-1} instead of u_t^{l-B-1}, reduces the entropy term in (64).

On the other hand,

$$h(s_b^n|f_0^b) = h(s_b^n|f_0^{b-1}) - I(f_b; s_b^n|f_0^{b-1})$$
$$= h(s_b^n|f_0^{b-1}) - H(f_b|f_0^{b-1}) + H(f_b|s_b^n, f_0^{b-1})$$
$$\geq h(s_b^n|f_0^{b-1}) - H(f_b) \tag{60}$$
$$\geq \frac{n}{2}\log\left(\rho^2 2^{\frac{2}{n}h(s_{b-1}^n|f_0^{b-1})} + 2\pi e(1-\rho^2)\right) - nR \tag{61}$$

Where (60) follows from the fact that conditioning reduces entropy and (61) again follows from Shannon's EPI. Assuming that the system is in steady state, we have $h(s_{b-1}^n|f_0^{b-1}) = h(s_b^n|f_0^b)$. Thus (61) reduces to

$$2^{\frac{2}{n}h(s_b^n|f_0^b)} \geq \frac{2\pi e(1-\rho^2)}{2^{2R}-\rho^2} \tag{62}$$

Replacing (62) into (59), (24) is derived.

Appendix B
Proof of Lemma 2

First note that, in the steady state when $t \gg 1$, we have

$$f_{t+B}(x, B) = I\left(s_{t+B}; u_{t+B} \mid u_0^{t-1}\right)$$
$$= I\left(s_l \;;\; u_l|\; u_0^{t-1}, u_t^{l-B-1}\right)$$
$$= h\left(u_l|\; u_0^{t-1}, u_t^{l-B-1}\right) - h\left(u_l|\; u_0^{t-1}, u_t^{l-B-1}, s_l\right)$$
$$= h\left(u_l|\; u_0^{t-1}, u_t^{l-B-1}\right) - h\left(u_l|s_l\right) \tag{63}$$
$$\geq h\left(u_l|\; u_0^{t-1}, u_{t+B}^{l-1}\right) - h\left(u_l|s_l\right) \tag{64}$$
$$= h\left(u_l|\; u_0^{t-1}, u_{t+B}^{l-1}\right) - h\left(u_l|\; u_0^{t-1}, u_{t+B}^{l-1}, s_l\right) \tag{65}$$
$$= I\left(s_l \;;\; u_l|\; u_0^{t-1}, u_{t+B}^{l-1}\right)$$
$$= f_l(x, B)$$

where (63) and (65) follow from the test channel model in (27). Fig. 5 clarifies the idea behind the step (64). As a result of the Markov chain in Fig. 5, any auxiliary random variable u_j which is farther from to u_l is a more noisier representation of u_l and consequently contains less information about it. Therefore, as shown in Fig. 5, replacing any conditioning element u_j in the first term of (63) with another element u_{j+B} which is closer to u_l

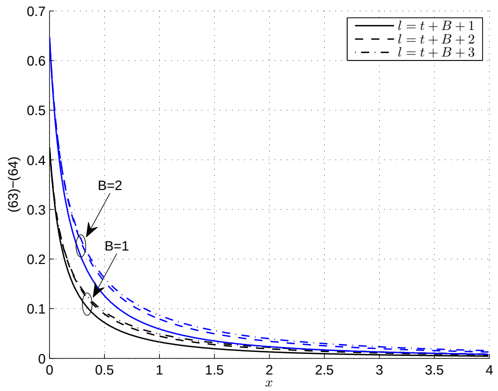

Fig. 6. The difference between the terms in (63) and (64) is positive for all values of x.

can only reduces the entropy. Fig 6, shows the difference between the terms in (63) and (64) as a function of x, for $B = 1, 2$ and for various ls, which confirms (64).

Now note that, as all the random variables are jointly Gaussian, we also have

$$\lambda_l(x, B) = \frac{2^{h\left(s_l \mid u_0^{t-1}, u_{t+B}^l\right)}}{2\pi e}. \tag{66}$$

Also note that

$$h\left(s_{t+B} \mid u_0^{t-1}, u_{t+B}\right) = h\left(s_l \mid u_0^{t-1}, u_t^{l-B-1}, u_l\right) \tag{67}$$

$$\geq h\left(s_l \mid u_0^{t-1}, u_{t+B}^l\right), \tag{68}$$

where (68) follows from the arguments similar to (64). From (68) and (66), and the fact that $2^{(\cdot)}/2\pi e$ is a monotonically increasing function, (32) is immediate. This completes the proof.

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] 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] 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.

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

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

[9] 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.

[10] F. Etezadi, A. Khisti, and M. Trott, "Sequential coding of markov sources over burst erasure channels," available at http://arxiv.org/abs/1202.5259, Feb. 2012.

[11] A. D. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Trans. Inform. Theory*, vol. 22, pp. 1–10, Jan. 1976.