Supplementary document for the paper “Real-Time Streaming of Gauss-Markov Sources over Sliding Window Burst-Erasure Channels” submitted to ISIT 2013

Farrokh Etzadi, Ashish Khisti

I. PROOF OF LEMMA 1

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

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

(1)

is an monotonically increasing function with respect to $x$, because

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

By applying Shannon’s EPI we have,

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

(3)

and thus,

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

(4)

$$\leq \frac{1}{2} \log (2\pi e r) - \frac{1}{2} \log \left( \rho^{2m} 2^{2h(s_k | f_a)} + 2\pi e (1 - \rho^{2m}) \right)$$

(5)

$$\leq \frac{1}{2} \log \left( \frac{r}{1 - (1 - \rho) \rho^{2m}} \right)$$

(6)

where (5) follows from the assumption that $h(s_k | f_a) \leq \frac{1}{2} \log (\rho^{2m} 2\pi e r)$ and the monotonicity property of $f(x)$. This completes the proof.

II. PROOF OF THEOREM 1: LOWER BOUND ON RATE RECOVERY FUNCTION

Define $p = L + B - 1$ and consider a periodic erasure channel with period $p$ where in $k$th period only the codewords $f_{k-1}^{k}$ are available and the rest of the codewords in that period is erased by the channel. Fig. 1 shows the $k$th period of the source model. Now consider the following inequality.

$$(W + 1)nR \geq H(f_{mp+L+1}^{mp+L+W+1})$$

$$\geq H(f_{mp+L+1}^{mp+L+W+1} | f_{k}^{k-1})$$

$$= I(s_{mp+L+W+1}^{mp+L+1} | f_{k}^{k-1}) + H(f_{mp+L+1}^{mp+L+W+1} | f_{k}^{k-1})$$

$$\geq h(s_{mp+L+W+1}^{mp+L+1} | f_{k}^{k-1}) - h(s_{mp+L+1}^{mp+L+1} | f_{k}^{k-1}) + H(f_{mp+L+1}^{mp+L+W+1} | f_{k}^{k-1})$$

(7)

(8)

(9)
where (7) follows from the fact that conditioning reduces the differential entropy and (9) is because the term $f_{mp+L+W+1}$ is removed from the third term in (9). According to the problem description the second term of (9) can be upper bounded as follows.

$$h(s_{mp+L+W+1} | f_{(k-1)p+L+1}^{mp+L} )_{k=-\infty}^{m} f_{mp+L+W+1} \leq \frac{n}{2} \log(2\pi e D)$$ \hspace{1cm} (10)

Now consider the first term in (9).

$$h(s_{mp+L+W+1} | f_{(k-1)p+L+1}^{mp+L-B} )_{k=-\infty}^{m} f_{mp+L+B} \geq \frac{n}{2} \log(\rho^2(B+L+W+1) + 2\pi e(1 - \rho^2(B+L+W+1)))$$ \hspace{1cm} (11)

$$\geq \frac{n}{2} \log \left( \rho^2(B+L+W+1) G(L, B, \rho, R) + 2\pi e(1 - \rho^2(B+L+W+1)) \right)$$ \hspace{1cm} (12)

where (11) follows from the Shannon’s EP and (12) follows from the following lemma.

**Lemma A. Define**

$$G(L, B, \rho, R) = \frac{2\pi e}{2^{2R} - \rho^2(B+L+W)(\rho^2 - 2R)(L-2)} \left( \frac{1 - \rho^2}{1 - \rho^2 - 2R} \right)$$

Then

$$2\pi h(s_{mp+L-B} | f_{(k-1)p+L+1}^{mp+L-B} )_{k=-\infty}^{m} f_{mp+L+B} \geq G(L, B, \rho, R)$$ \hspace{1cm} (14)

**Proof.** See Appendix A. \hfill \Box

It remains to lower bound the third term in (9).

$$H(f_{mp+L+W}^{m} s_{mp+L+W+1} | f_{mp+L+B}^{m} s_{mp+L+B+1}, f_{mp+L+B}^{m} s_{mp+L+B+1}, f_{mp+L+B}^{m} s_{mp+L+B+1})$$

$$\geq I(f_{mp+L+B}^{m} s_{mp+L+B+1}, s_{mp+L+B+1}, f_{mp+L+B}^{m} s_{mp+L+B+1}, f_{mp+L+B}^{m} s_{mp+L+B+1})$$

$$+ H(f_{mp+L+B}^{m} s_{mp+L+B+1}, s_{mp+L+B+1}, f_{mp+L+B}^{m} s_{mp+L+B+1}, f_{mp+L+B}^{m} s_{mp+L+B+1})$$

$$\geq h(s_{mp+L+B+1}, \cdots, s_{mp+L+W}^{m} f_{mp+L+B+1}, f_{mp+L+B}^{m} s_{mp+L+B+1})$$

$$- h(s_{mp+L+B+1}, \cdots, s_{mp+L+W}^{m} f_{mp+L+B+1}, f_{mp+L+B}^{m} s_{mp+L+B+1})$$ \hspace{1cm} (17)

where (15) follows from the fact that conditioning reduces the entropy and (17) is based on the positivity of the entropy term.
First consider the first term in (17).

\[
h(s^n_{mp+L+1}, \cdots, s^n_{mp+L+W}) | s^n_{mp+L+W+1} \{ f_{(k-1)p+L+1}^m \}_{k=-\infty}^{m} f_{mp+L-B+1}^m \]

\[
= h(s^n_{mp+L+1}, \cdots, s^n_{mp+L+W+1} | \{ f_{(k-1)p+L+1}^m \}_{k=-\infty}^{m} f_{mp+L-B+1}^m )
\]

\[
- h(s^n_{mp+L+W+1} | \{ f_{(k-1)p+L+1}^m \}_{k=-\infty}^{m} f_{mp+L-B+1}^m )
\]

\[
= h(s^n_{mp+L+1}, \cdots, s^n_{mp+L+B}) | m \}
\]

\[
- h(s^n_{mp+L+B+1} | \{ f_{(k-1)p+L+1}^m \}_{k=-\infty}^{m} f_{mp+L-B+1}^m )
\]

\[
\geq \frac{n}{2} \log \left( (2 - 2R \rho^2)^B \rho^2 G(L, B, \rho, R) + 2\pi e(1 - \rho^2) \right) - \frac{n}{2} \log \left( 2\pi e(1 - (1 - D)\rho^{2(W+1)}) \right) + nW h(s_2 | s_1)
\]

where (18) follows from the following Markov chain property

\[
\{ \{ f_{(k-1)p+L+1}^m \}_{k=-\infty}^{m} f_{mp+L-B+1}^m \} \rightarrow s^n_{mp+L+1} \rightarrow \{ s^n_{mp+L+2}, \cdots, s^n_{mp+L+W+1} \}
\]

and the fact that

\[
h(s^n_{mp+L+1}, \cdots, s^n_{mp+L+W+1} | s^n_{mp+L+1}) = nW h(s_2 | s_1).
\]

The first term in (19) also follows from the application of Shannon’s EPI as follows

\[
\frac{n}{2} \log \left( (2 - 2R \rho^2)^B \rho^2 G(L, B, \rho, R) + 2\pi e(1 - \rho^2) \right) - \frac{n}{2} \log \left( 2\pi e(1 - (1 - D)\rho^{2(W+1)}) \right) + nW h(s_2 | s_1)
\]

By repeating the same step for \( B \) steps and using (14), first term in (19) is resulted. The second term in (19) follows from the following.

\[
h(s^n_{mp+L+1}, \cdots, s^n_{mp+L+W+1} | f_{(k-1)p+L+1}^m m \}
\]

\[
= h(s^n_{mp+L+1}, \cdots, s^n_{mp+L+W+1} | f_{(k-1)p+L+1}^m m \}
\]

\[
- h(s^n_{mp+L+W+1} | f_{(k-1)p+L+1}^m m \}
\]

\[
\leq h(s^n_{mp+L+W+1} | f_{(k-1)p+L+1}^m m \}
\]

\[
= h(s^n_{mp+L+1}, \cdots, s^n_{mp+L+L}) + z^n | | s^n_{mp+L}
\]

\[
\leq h(s^n_{mp+L+1}, \cdots, s^n_{mp+L+L}) + z^n | | s^n_{mp+L}
\]

\[
\leq \frac{1}{2} \log \left( 2\pi e(\rho^{2(W+1)}) D + 1 - \rho^{2(W+1)}) \right) - \frac{1}{2} \log \left( 2\pi e(1 - (1 - D)\rho^{2(W+1)}) \right)
\]

where (26) is because the estimate \( \hat{s}^n_{mp+L} \) is a function of \( \{ f_{(k-1)p+L+1}^m m \}

\[
\text{and (27) follows from the fact that unconditioning increases the differential entropy. In (28), } z^n \sim \mathcal{N}(0, 1 - \rho^{2(W+1)}) \text{ and (30) follows from the fact that with the same variance, the Gaussian source has the highest entropy.}
For the second term in (17) we have

\[
q(W) \doteq h(s_{mp+L+1}^n, \ldots, s_{mp+L+W}^n) \bigg\{ f_{(k-1)p+L+1}^{kp+L-B} \bigg| k = -\infty, f_{mp+L+B+1}^{mp+L+B} \bigg\} \\
= h(s_{mp+L+1}^n, \ldots, s_{mp+L+W}^n) \bigg\{ f_{(k-1)p+L+1}^{kp+L-B} \bigg| k = -\infty, f_{mp+L+B+1}^{mp+L+B} \bigg\} \\
- h(s_{mp+L+W+1}^n) \bigg\{ f_{(k-1)p+L+1}^{kp+L-B} \bigg| k = -\infty, f_{mp+L+B+1}^{mp+L+B} \bigg\} \\
= h(s_{mp+L+W}^n) \bigg\{ f_{(k-1)p+L+1}^{kp+L-B} \bigg| k = -\infty, f_{mp+L+B+1}^{mp+L+B} \bigg\} \\
+ h(s_{mp+L+1}^n) \bigg\{ f_{(k-1)p+L+1}^{kp+L-B} \bigg| k = -\infty, f_{mp+L+B+1}^{mp+L+B} \bigg\} \\
\leq \frac{n}{2} \log \left( \frac{D}{1 - (1 - D)\rho^2} \right) + nW h(s_2|s_1) + q(0)
\]

where (33) follows from the following Lemma 1 in the paper whose proof is in Section I in this report. Furthermore (33) can be reduced as follows.

\[
q(W) \leq \frac{nW}{2} \log \left( \frac{D}{1 - (1 - D)\rho^2} \right) + nW h(s_2|s_1) + q(0)
\]

where \( q(0) = 0 \).

By replacing (34) and (19) into (17) we have

\[
H(f_{mp+L+W}^{mp+L+1}) \bigg\{ f_{(k-1)p+L+1}^{kp+L-B} \bigg| k = -\infty \bigg\} \\
\geq \frac{n}{2} \log \left( \frac{(2-2\rho^2)^B \rho^2 G(L, B, \rho, R) + 2\pi e(1 - \rho^2)^{1-(2-2\rho^2)^B+1}}{2\pi e(1 - (1 - D)\rho^2)^2(W+1)} \right) \\
\frac{(2-2\rho^2)^B}{2\pi e(1 - (1 - D)\rho^2)^2(W+1)} \left( 1 - (1 - D)\rho^2 \right)^{W}
\]

Finally By replacing (35), (12) and (10) into (9) the lower bound on rate can be derived.

\[
(W + 1)R \geq \frac{1}{2} \log \left( \frac{(2-2\rho^2)^B \rho^2 G(L, B, \rho, R) + 2\pi e(1 - \rho^2)^{2(B+W+1)}}{2\pi e D} \right) \\
+ \frac{1}{2} \log \left( \frac{(2-2\rho^2)^B \rho^2 G(L, B, \rho, R) + 2\pi e(1 - \rho^2)^{1-(2-2\rho^2)^B+1}}{2\pi e(1 - (1 - D)\rho^2)^2(W+1)} \right) \\
\frac{(2-2\rho^2)^B}{2\pi e(1 - (1 - D)\rho^2)^2(W+1)} \left( 1 - (1 - D)\rho^2 \right)^W
\]

By solving (36) for \( R \), the lower bound of the rate is derived. This completes the proof.

III. PROOF OF COROLLARY 1: HIGH RESOLUTION REGIME

We show the high resolution results of Corollary 1 by computing the limit of the lower and upper bounds of lossy rate-recovery function in Theorems 1 and 2 of the paper when \( D \) approaches the zero. First consider the upper bound \( R^+(L, B, W, D) \). Note that

\[
\frac{1}{2} \log(2\pi e D) \geq h(s_{2L+B+W}^n|s_{L+W+1}^n, u_{L+W+2}^{2L-1}, u_{2L+B}^{2L+B+W}) \\
\geq h(s_{2L+B+W}^n | s_{2L+B+W-1}^n, u_{2L+B+W}^n) \\
= \frac{1}{2} \log \left( \frac{2\pi e}{1/x + 1/(1 - \rho^2)} \right)
\]

where (37) follows from the Markov chain property

\[
\{ s_{L+W+1}^n, u_{L+W+2}^{2L-1}, u_{2L+B}^{2L+B+W-1} \} \rightarrow \{ s_{2L+B+W-1}^n, u_{2L+B+W}^n \} \rightarrow s_{2L+B+W}.
\]

From (38) we have

\[
x \leq \frac{D}{1 - D/(1 - \rho^2)^{2(n+1)}}
\]
and thus $D \to 0$ requires $x \to 0$. 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

$$h(s_{2L+B}^{2L+B+W}|s_{L+W+1}, u_{L+W+2}^{2L-1}) = h(s_{2L+B}|s_{L+W+1}, u_{L+W+2}^{2L-1}) + \sum_{j=1}^{W} h(s_{2L+B+j}|s_{2L+B+j-1})$$

(41)

$$\approx h(s_{2L+B}|s_{2L-1}) + Wh(s_2|s_1)$$

(42)

$$= \frac{1}{2} \log((2\pi e)^{W+1}(1-\rho^{2(B+1)})(1-\rho^2)^W)$$

(43)

and

$$h(s_{2L+B}^{2L+B+W}|s_{L+W+1}, u_{2L+B}^{2L+B}) \approx \sum_{j=0}^{W} h(s_{2L+B+j}|u_{2L+B+j})$$

$$= (W+1)h(s_1|u_1) = \frac{W+1}{2} \log(2\pi eD)$$

(44)

By combining (43) and (44) we can see

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

(45)

For the lower bound $R^-(L, B, W, D)$, note that in high resolution the rate $R$ becomes large and we need to consider equation (2) of the paper in the limit $R \to \infty$. From equation (3) of the paper, it is clear that

$$\lim_{R \to \infty} G(L, B, \rho, R) = 0.$$  

(46)

By taking the limit of equation (2) of the paper as $R \to \infty$ and $D \to 0$, it can be easily observed that

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

(47)

IV. PROOF OF LEMMA 2

Before going through consider the following lemmas whose proofs are available in Appendices B and C respectively.

**Lemma B.** Consider the seven jointly Gaussian random variables $X_0, X_1, X_2, X_3, Y_1, Y_2, Y_3$ drawn in Fig. 2 such that

$$X_{i+1} = \rho_i X_i + n_i \quad i \in \{0, 1, 2\}$$

(48)

$$Y_j = \tilde{\alpha} X_j + \tilde{n}_j \quad j \in \{1, 2, 3\}$$

(49)

where $X_i \sim N(0, 1)$. Also $n_i \sim N(0, 1-\rho_i^2)$ and $\tilde{n}_j \sim N(0, \tilde{\alpha}^2)$ are independent noises. Then

$$h(Y_3|X_0, Y_2) \leq h(Y_3|X_0, Y_1)$$

(50)

**Lemma C.** Consider the two sets $A, B \subseteq \mathbb{N}$ of the same size $r$ as $A = \{a_1, a_2, \ldots, a_r\}$, $B = \{b_1, b_2, \ldots, b_r\}$ such that $1 \leq a_1 < a_2 < \cdots < a_r$ and $1 \leq b_1 < b_2 < \cdots < b_r$ and for any $i \in \{1, \ldots, r\}$, $a_i \leq b_i$. Also consider
the Gauss-Markov source model and define \( u_t = \alpha s_t + n_t \) as the quantizes version of the source \( s_t \). Then for any \( t \geq b_r \)

\[
h(u_t | u_A, s_t) \geq h(u_t | u_B, s_t). \tag{51}
\]

Now in order to prove Lemma 2, Fig. 3 illustrates the source sequences. First define \( \hat{s}_i(\Omega) \) as the minimum mean square error estimate of \( s_i \) given \( \Omega \) and define \( h_i(\Omega) \) and \( \sigma_i^2(\Omega) \) as an independent noise representing corresponding mean square error and its variance, respectively. It is equivalent to write

\[
\hat{s}_i(\Omega) = a_i s_i + v_i(\Omega) \tag{52}
\]

where \( v_i \sim N(0, \sigma_i^2) \) and define \( \sigma_i^2(\Omega) \triangleq \frac{\sigma_i^2}{\nu_i^2} \). Set \( j = L + W + 1 \). Note that

\[
\sigma_i^2(\Omega) = \frac{\sigma_i^2(\Omega)}{1 + \sigma_i^2(\Omega)} \tag{53}
\]

According to the assumptions of Theorem 2 of the paper, we have

\[
h(s_j | \hat{s}_j) = \frac{1}{2} \log(2\pi eD) \tag{54}
\]

Assume that

\[
h(s_j | \hat{s}_j) < h(s_{j+1} | \hat{s}_j, u_{j+1}) \tag{55}
\]

This means that the estimation error of estimating \( \hat{s}_{j+1} | (\hat{s}_j, u_{j+1}) \) is strictly greater than the estimation error of estimating \( \hat{s}_j \), i.e.

\[
\sigma_{j+1}^2(\hat{s}_j, u_{j+1}) > \sigma_j^2(\hat{s}_j). \tag{56}
\]

which based on (53) results in

\[
\sigma_{j+1}^2(\hat{s}_j, u_{j+1}) > \sigma_j^2(\hat{s}_j). \tag{57}
\]

Now consider

\[
h(s_{j+2} | \hat{s}_j, u_{j+1}, u_{j+2}) = h(s_{j+2} | \hat{s}_{j+1} (\hat{s}_j, u_{j+1}), u_{j+2}) \tag{58}
\]

\[
= h(s_{j+2} | a_j s_{j+1} + v_{j+1} (\hat{s}_j, u_{j+1}), u_{j+1}) \tag{59}
\]

\[
= h(s_j | a_j s_j + v_{j+1} (\hat{s}_j, u_{j+1}), u_j) \tag{60}
\]

\[
> h(s_{j+1} | a_j s_j + v_j (\hat{s}_j, u_{j+1}), u_j) \tag{61}
\]

\[
= h(s_{j+1} | \hat{s}_j, u_{j+1}) \tag{62}
\]

where (58) follows from the fact that MMSE is the optimal estimator for jointly Gaussian random variables. (59) is based on the virtual channel model in (52). (60) follows from time-invariant property of the source model and (61) follows from (57). Note that from (62) it can be deduced that

\[
\sigma_{j+2}^2(\hat{s}_j, u_{j+1}) > \sigma_{j+1}^2(\hat{s}_j, u_{j+1}). \tag{63}
\]
By repeating the same method it can be seen that assuming (55) for any i ∈ [j + 1, j + p] we have
\[ h(s_{j+i} | \hat{s}_j, u_{j+i+1}^{i+j}) > h(s_{j+i-1} | \hat{s}_j, u_{j+i+1}^{i+j-1}) \] (64)

On the other hand we know
\[ \frac{1}{2} \log(2\pi eD) \geq h(s_{j+p} | \hat{s}_j, u_{j+1}^{i+j} + B, u_{j+p+1}^{i+j+p}) \] (65)
\[ \geq h(s_{j+p} | \hat{s}_j, u_{j+1}^{i+j+1}) \] (66)
\[ > h(s_j | \hat{s}_j) \] (67)

where (65) is according to the problem requirement. However (67) contradicts (54). Thus (55) is not correct and
\[ h(s_j | \hat{s}_j) \geq h(s_{j+1} | \hat{s}_j, u_{j+1}) \] (68)

Knowing this we have
\[ \sigma_j^{2}(\hat{s}_j, u_{j+1}) \leq \sigma_j^{2}(\hat{s}_j). \] (69)
and using the steps in (58)–(62) in the other direction it is not hard to show that
\[ h(s_{j+2} | \hat{s}_j, u_{j+1}, u_{j+2}) \leq h(s_{j+1} | \hat{s}_j, u_{j+1}). \] (70)

Similarly for any i ∈ [j + 1, j + p] we have
\[ h(s_{j+i} | \hat{s}_j, u_{j+i+1}) \leq h(s_{j+i-1} | \hat{s}_j, u_{j+i+1}^{i+j}) \] (71)

which proves the first inequality of the lemma.

For the other inequality set j = p + L. We need to show that for any i ∈ [1, W + 1]
\[ h(s_{j+i} | \hat{s}_j, u_{j+i+1}^{i+j}) \geq h(s_{j+i+1} | \hat{s}_j, u_{j+i+1}^{i+j+1}). \] (72)

To this end we have
\[ h(s_{j+i+1} | \hat{s}_j, u_{j+i+1}^{i+j}) \leq h(s_{j+i+1} | \hat{s}_j, u_{j+i+1}^{i+j+1}) \] (73)
\[ = h(s_{j+i+1} | \hat{s}_j, u_{j+i+1}^{i+j+1}) \] (74)
\[ = h(s_{j+i+1} | \hat{s}_j, u_{j+i+1}^{i+j+1}) \] (75)
\[ = h(s_{j+i+1} | \hat{s}_j, u_{j+i+1}^{i+j+1}) \] (76)
\[ = h(s_{j+i+1} | \hat{s}_j, u_{j+i+1}^{i+j+1}) \] (77)

where (73) follows from Lemma C. (75) follows from the fact that in (74) the noise \( v_{j-B+1} \) is independent and the source model in problem set up is time invariant. (76) follows from the fact that from what we proves earlier in this lemma.

\[ \sigma_j^{2}(\hat{s}_j, u_{j+i}^{i+j+1}) \leq \sigma_j^{2}(\hat{s}_j, u_{j+i}^{i+j+1}) \] (78)

This completes the proof of the lemma.
V. Coding Theorem

**Theorem A.** Consider the dominant term in equation (20) and equation (21) of the paper as

\[ R \geq I(u_{L+L+2}; s_{L+W+1} | \hat{s}_{L+W+1}) \]  

(79)

\[ R \geq \max_{0 \leq \Lambda \leq |\Omega|} \frac{1}{W+1} I(s_{\Omega}; u_{\Omega} | \hat{s}_{L+W+1; u_{L+W+2}, u_{\Omega^{c}}}) \]  

(80)

The two constraints in (79) and (80) are equivalent to

\[ R \geq \frac{1}{W+1} I(s_{\Lambda}; u_{\Lambda} | \hat{s}_{L+W+1; u_{L+W+2}, u_{\Omega^{c}}}) \]  

(81)

where \( \Lambda \triangleq [2L + B, 2L + B + W] \).

**Proof.** First consider (80) and define the operator \( L(\Omega, k) \) which applies on set \( \Omega \in \mathbb{Z} \) with \( k \leq |\Omega| \) and returns the set containing the \( k \) largest elements of \( \Omega \). Also define

\[ \Lambda_k \triangleq L(\Lambda, k) \]  

(82)

i.e. \( \Lambda_k \) is the set containing the \( k \) largest elements in \( \Lambda \). Consider the following lemmas whose proof is in Appendix D and Appendix E, respectively.

**Lemma D.**

\[ \arg \max_{0 \leq \Lambda \leq |\Omega|} I(s_{\Omega}; u_{\Omega} | \hat{s}_{L+W+1; u_{L+W+2}, u_{\Omega^{c}}}) = \Lambda_k \]  

(83)

**Lemma E.** For any \( k \leq W \)

\[ \frac{1}{k} I(s_{\Lambda_k}; u_{\Lambda_k} | u_{\Gamma(m+p L-B)}, u_{\Lambda_k(m)^{c}}; s_0) \leq \frac{1}{k+1} I(s_{\Lambda_{k+1}}; u_{\Lambda_{k+1}(m)} | u_{\Gamma(m+p L-B)}, u_{|\Lambda_{k+1}(m)^{c}}; s_0) \]  

(84)

According to Lemma D, (80) is equivalent to

\[ R \geq \max_{0 \leq \Lambda \leq |\Omega|} \frac{1}{W+1} I(s_{\Omega}; u_{\Omega} | \hat{s}_{L+W+1; u_{L+W+2}, u_{\Omega^{c}}}) \]  

(85)

and according to Lemma E, (85) is equivalent to

\[ R \geq \frac{1}{W+1} I(s_{\Lambda}; u_{\Lambda} | \hat{s}_{L+W+1; u_{L+W+2}, u_{\Omega^{c}}}) \]  

(86)

It only remains to show that (86) dominates (79). To see that note that according to the Theorem 2 of the paper, the test channel is such that

\[ h \left( s_{p+L+W+1} | \hat{s}_{L+W+1}, u_{L+W+2}, u_{p+L+W+1} \right) = \frac{1}{2} \log(2\pi e D) \]  

(87)

On the hand and using Lemma 2 of the paper it is clear that

\[ \frac{1}{W+1} h(u_{\Lambda} | \hat{s}_{L+W+1; u_{L+W+2}, u_{L+W+2}}) \geq \frac{1}{2} \log(2\pi e D) \]  

(88)

\[ \geq h(u_{L+W+2} | \hat{s}_{L+W+1}) \]  

(89)

where both (88) and (89) follow from Lemma 2 of the paper

\[ \frac{1}{W+1} I(s_{\Lambda}; u_{\Lambda} | \hat{s}_{L+W+1; u_{L+W+2}}) = \frac{1}{W+1} h(u_{\Lambda} | \hat{s}_{L+W+1; u_{L+W+2}}) - \frac{1}{W+1} h(u_{\Lambda} | s_{\Lambda}) \]  

(90)

\[ \geq h(u_{L+W+2} | \hat{s}_{L+W+1}) - h(u_{L+W+2} | s_{L+W+2}) \]  

(91)

\[ = I(u_{L+W+2}; s_{L+W+2} | \hat{s}_{L+W+1}) \]  

(92)

which completes the proof. \( \square \)
VI. CODING THEOREM: DISTORTION CONSTRAINTS

The distortion constraints for the time before the erasure starts can be written as follows. For any \( i \in [L + W + 2, p + L - B] \) we should have

\[
h(s_i | \hat{s}_{L+W+1}, u_{L+W+2}^i) \leq \frac{1}{2} \log(2\pi e D)
\]

(93)

Also according to Lemma 2 of the paper for

\[
h(s_i | \hat{s}_{L+W+1}, u_{L+W+2}^i) \leq h(s_{L+W+1} | \hat{s}_{L+W+1})
\]

(94)

\[
h(s_{L+B+W} | \hat{s}_{L+W+1}, u_{L+W+2}^p, u_{L+W+2}^{p+1}) = \frac{1}{2} \log(2\pi e D)
\]

(95)

\[
= \frac{1}{2} \log(2\pi e D)
\]

(96)

\[
= \frac{1}{2} \log(2\pi e D)
\]

(97)

which completes the proof.

APPENDIX A
PROOF OF LEMMA A

\[
g_{m,n} \triangleq h(s_{mp+L-B} | \{f_{(k-1)p+L+1}^{k=\infty} \} f_{k=\infty}^{m=\infty} f_{k=\infty}^{m=\infty})
\]

(98)

\[
= h(s_{mp+L-B} | \{f_{(k-1)p+L+1}^{k=\infty} \} f_{k=\infty}^{m=\infty} f_{k=\infty}^{m=\infty}) - I(s_{mp+L-B} ; f_{(k-1)p+L+1}^{k=\infty} f_{k=\infty}^{m=\infty} f_{k=\infty}^{m=\infty})
\]

(99)

\[
\geq h(s_{mp+L-B} | \{f_{(k-1)p+L+1}^{k=\infty} \} f_{k=\infty}^{m=\infty} f_{k=\infty}^{m=\infty}) - H(f_{mp+L-B} | \{f_{(k-1)p+L+1}^{k=\infty} \} f_{k=\infty}^{m=\infty} f_{k=\infty}^{m=\infty})
\]

(100)

\[
\geq h(s_{mp+L-B} | \{f_{(k-1)p+L+1}^{k=\infty} \} f_{k=\infty}^{m=\infty} f_{k=\infty}^{m=\infty}) - nR
\]

(101)

\[
\geq \frac{n}{2} \log \left( \rho^{2^{2R} g_{m,n}} \right) \geq \rho^{2^{2R} \frac{1}{2} h(s_{mp+L-B-1} | \{f_{(k-1)p+L+1}^{k=\infty} \} f_{k=\infty}^{m=\infty} f_{k=\infty}^{m=\infty}) + 2\pi e(1 - \rho^2)
\]

(102)

which results in

\[
2^{2R} \frac{1}{2} g_{m,n} \geq \rho^{2^{2R} \frac{1}{2} h(s_{mp+L-B-1} | \{f_{(k-1)p+L+1}^{k=\infty} \} f_{k=\infty}^{m=\infty} f_{k=\infty}^{m=\infty}) + 2\pi e(1 - \rho^2)
\]

(103)

\[
\geq \rho^{2^{2R} \frac{1}{2} h(s_{mp+L-B-1} | \{f_{(k-1)p+L+1}^{k=\infty} \} f_{k=\infty}^{m=\infty} f_{k=\infty}^{m=\infty}) + 2\pi e(1 - \rho^2)
\]

(104)

\[
\geq \rho^{2^{2R} \frac{1}{2} h(s_{mp+L-B-1} | \{f_{(k-1)p+L+1}^{k=\infty} \} f_{k=\infty}^{m=\infty} f_{k=\infty}^{m=\infty}) + 2\pi e(1 - \rho^2)
\]

(105)

\[
\geq \rho^{2^{2R} \frac{1}{2} h(s_{mp+L-B-1} | \{f_{(k-1)p+L+1}^{k=\infty} \} f_{k=\infty}^{m=\infty} f_{k=\infty}^{m=\infty}) + 2\pi e(1 - \rho^2)
\]

(106)

\[
\geq \rho^{2^{2R} \frac{1}{2} h(s_{mp+L-B-1} | \{f_{(k-1)p+L+1}^{k=\infty} \} f_{k=\infty}^{m=\infty} f_{k=\infty}^{m=\infty}) + 2\pi e(1 - \rho^2)
\]

(107)

By repeating the same method, it is not hard to see that

\[
2^{2R} \frac{1}{2} g_{m,n} \geq (\rho^{2^{2R} \frac{1}{2} h(s_{mp+L-B-1} | \{f_{(k-1)p+L+1}^{k=\infty} \} f_{k=\infty}^{m=\infty} f_{k=\infty}^{m=\infty}) + 2\pi e(1 - \rho^2)) + 2\pi e(1 - \rho^2) + L-3 \sum_{i=0}^{L-3} (\rho^{2^{2R} \frac{1}{2} h(s_{mp+L-B-1} | \{f_{(k-1)p+L+1}^{k=\infty} \} f_{k=\infty}^{m=\infty} f_{k=\infty}^{m=\infty}))
\]

(108)

\[
= \rho^{2^{2R} \frac{1}{2} h(s_{mp+L-B-1} | \{f_{(k-1)p+L+1}^{k=\infty} \} f_{k=\infty}^{m=\infty} f_{k=\infty}^{m=\infty}) + 2\pi e(1 - \rho^2)} + 2\pi e(1 - \rho^2) - \frac{1 - (\rho^{2^{2R} \frac{1}{2} h(s_{mp+L-B-1} | \{f_{(k-1)p+L+1}^{k=\infty} \} f_{k=\infty}^{m=\infty} f_{k=\infty}^{m=\infty}) + 2\pi e(1 - \rho^2))}{1 - \rho^{2^{2R} \frac{1}{2} h(s_{mp+L-B-1} | \{f_{(k-1)p+L+1}^{k=\infty} \} f_{k=\infty}^{m=\infty} f_{k=\infty}^{m=\infty}) + 2\pi e(1 - \rho^2))}
\]

(109)

\[
= \rho^{2^{2R} \frac{1}{2} h(s_{mp+L-B-1} | \{f_{(k-1)p+L+1}^{k=\infty} \} f_{k=\infty}^{m=\infty} f_{k=\infty}^{m=\infty}) + 2\pi e(1 - \rho^2)} + 2\pi e(1 - \rho^2) - \frac{1 - (\rho^{2^{2R} \frac{1}{2} h(s_{mp+L-B-1} | \{f_{(k-1)p+L+1}^{k=\infty} \} f_{k=\infty}^{m=\infty} f_{k=\infty}^{m=\infty}) + 2\pi e(1 - \rho^2))}{1 - \rho^{2^{2R} \frac{1}{2} h(s_{mp+L-B-1} | \{f_{(k-1)p+L+1}^{k=\infty} \} f_{k=\infty}^{m=\infty} f_{k=\infty}^{m=\infty}) + 2\pi e(1 - \rho^2))}
\]

(110)
By assuming that in steady state \( g_{m-1,n} = g_{m,n} \), a lower bound can be found as follows
\[
2\pi g_{m,n} \geq G(L, B, \rho, R)
\] (111)
where \( G(L, B, \rho, R) \) is defined in (13).

**Appendix B**

**Proof of Lemma B**

Note that all the random variables are jointly Gaussian. Thus
\[
h(Y_3|X_0, Y_i) = \frac{1}{2} \log(2\pi e\sigma_i^2) \quad i \in \{1, 2\}
\] (112)
where \( \sigma_i^2 \) is the mean square error of estimating \( Y_3 \) knowing \( \{X_0, Y_i\} \). Define \( y = \sigma^2 / \alpha^2 \), we have
\[
\sigma_1^2 = \tilde{\sigma}^2 + \sigma^2 - (\tilde{\alpha}\rho_0\rho_1\rho_2) \left( \frac{1}{\tilde{\alpha}\rho_0} \tilde{\alpha} \rho_0 \right) \left( \frac{\tilde{\alpha}^2 + \sigma^2}{\tilde{\alpha}^2 \rho_1 \rho_2} \right) \] (113)
\[
= \tilde{\sigma}^2 + \sigma^2 - \tilde{\alpha}^2 \rho_1 \rho_2 \rho_0^2 y - \rho_0^2 + \frac{1}{1 + y - \rho_0^2}
\] (114)
\[
\sigma_2^2 = \tilde{\alpha}^2 + \sigma^2 - (\tilde{\alpha}\rho_0\rho_1\rho_2) \left( \frac{1}{\tilde{\alpha}\rho_1} \tilde{\alpha} \rho_1 \right) \left( \frac{\tilde{\alpha}^2 + \sigma^2}{\tilde{\alpha}^2 \rho_2 \rho_0} \right) \] (115)
\[
= \tilde{\alpha}^2 + \sigma^2 - \tilde{\alpha}^2 \rho_2 \rho_0^2 \rho_1^2 y - \rho_0^2 \rho_1^2 + \frac{1}{1 + y - \rho_0^2 \rho_1^2}
\] (116)

Note that
\[
1 + \rho_0^4 \rho_1^2 \geq \rho_0^2(1 + \rho_1^2)
\] (117)
also
\[
1 + \rho_0^4 \rho_1^2 + y(1 - \rho_0^4 \rho_1^2) \geq 1 + \rho_0^2 \rho_1^2 \geq \rho_0^2(1 + \rho_1^2)
\] (118)

From (118) we have
\[
(1 - \rho_1^2)(1 + \rho_0^4 \rho_1^2 + y(1 - \rho_0^4 \rho_1^2)) = 1 + \rho_0^4 \rho_1^2 + y(1 - \rho_0^4 \rho_1^2) - \rho_1^2(1 + \rho_0^4 \rho_1^2 + y(1 - \rho_0^4 \rho_1^2))
\]
\[
\geq \rho_0^2(1 - \rho_1^2)(1 + \rho_1^2)
\] (119)
and from (119),
\[
1 + \rho_0^4 \rho_1^2 + y(1 - \rho_0^4 \rho_1^2) - \rho_0^2 \geq \rho_1^2(1 + \rho_0^4 \rho_1^2 + y(1 - \rho_0^4 \rho_1^2)) - \rho_1^2 \rho_0^2
\] (120)

It is not hard to see that (120) is equivalent to
\[
\frac{\rho_0^2 \rho_1^2 y - \rho_0^2 \rho_1^2 + 1}{1 + y - \rho_0^2 \rho_1^2} \geq \frac{\rho_0^2 \rho_1^2 y - \rho_0^2 \rho_1^2 + \rho_1^2}{1 + y - \rho_0^2}
\] (121)

which results in \( \sigma_1^2 \geq \sigma_2^2 \) and this completes the proof.
APPENDIX C
PROOF OF LEMMA C

We prove the lemma by induction. First we show that the lemma is true for \( r = 1 \). This is by direct application of the Lemma B when \( X_0 = s_0, X_1 = s_{a_1}, X_2 = s_0, \) and \( X_3 = s_1 \) and \( Y_i = u_i \) for \( i \in \{a_1, b_1, t\} \).

Now assume that the lemma is true for \( r \), i.e. for the sets \( A_r, B_r \) of size \( r \) satisfying \( a_i \leq b_i \) for \( i \in \{1, \ldots, r\} \),

\[
h(u_t | u_{A_r}, s_0) \geq h(u_t | u_{B_r}, s_0). \tag{122}
\]

In two steps, we show that the lemma is also true for the sets \( A_{r+1} = \{ A_r, a_{r+1} \} \) and \( B_r = \{ B_r, b_{r+1} \} \) where \( a_r \leq a_{r+1}, b_r \leq b_{r+1} \) and \( a_{r+1} \leq b_{r+1} \leq t \).

**Step 1:** We show that

\[
h(u_t | u_{A_{r+1}}, s_0) \geq h(u_t | u_{A_r}, u_{b_{r+1}}, s_0). \tag{123}
\]

\[
\text{Note that}
\]

\[
h(u_t | u_{A_{r+1}}, s_0) = h(u_t | \hat{s}_{a_r}, u_{A_r}, u_{b_{r+1}}, s_0) \tag{124}
\]

\[
\geq h(u_t | \hat{s}_{a_r}, u_{a_{r+1}}) \tag{125}
\]

\[
\geq h(u_t | \hat{s}_{a_r}, u_{b_{r+1}}) \tag{126}
\]

\[
\geq h(u_t | \hat{s}_{a_r}, u_{A_r}, u_{b_{r+1}}, s_0) \tag{127}
\]

\[
\geq h(u_t | u_{A_r}, u_{b_{r+1}}, s_0) \tag{128}
\]

where in (124) \( \hat{s}_{a_r} \) is the MMSE estimate of \( s_{a_r} \) knowing \( \{ u_{A_r}, s_0 \} \) and hence is the sufficient statistics for \( u_t \) and \( u_{a_{r+1}} \). (125) follows from the Markov chain property \( \{ u_{A_r}, s_0 \} \rightarrow \hat{s}_{a_r} \rightarrow u_t \). (126) follows from application of Lemma B when \( X_0 = s_{a_r}, X_1 = s_{a_{r+1}}, X_2 = s_0, \) and \( X_3 = s_1 \) and \( Y_i = u_i \) for \( i \in \{a_{r+1}, b_{r+1}, t\} \). Also (127) is based on the fact that conditioning reduces the differential entropy and (128) from the fact that \( \hat{s}_{a_r} \) is a function of \( \{ u_{A_r}, s_0 \} \).

**Step 2:** We show that

\[
h(u_t | u_{A_r}, u_{b_{r+1}}, s_0) \geq h(u_t | u_{B_{r+1}}, s_0). \tag{129}
\]

Define \( \hat{s}_{b_{r+1}}(\Omega) \) as the MMSE estimate of \( s_{b_{r+1}} \) knowing \( \{ \Omega, s_0 \} \) for \( \Omega \in \{ A_r, B_r \} \) and define \( \sigma^2(\Omega) \) as the associated MSE. In order to show (129) it suffices to show that

\[
h(u_t | \hat{s}_{b_{r+1}}(A_r), u_{b_{r+1}}) \geq h(u_t | \hat{s}_{b_{r+1}}(B_r), u_{b_{r+1}}) \tag{130}
\]

By assuming the hypothesis (the fact that the lemma works for \( r \)) it is not hard to observe that

\[
h(s_{b_{r+1}} | u_{A_r}, s_0) \geq h(s_{b_{r+1}} | u_{B_r}, s_0) \tag{131}
\]

and thus \( \sigma^2(A_r) \leq \sigma^2(B_r) \). This results in (130) indicating that the estimation error of estimating \( u_t \) by knowing two independent noisy versions of \( s_{b_{r+1}} \) when one of them is fixed, become lower if the other become less noisy.

Combining (123) and (129) we have \( h(u_t | u_{A_{r+1}}, s_0) \geq h(u_t | u_{B_{r+1}}, s_0) \) which completes the proof.

APPENDIX D
PROOF OF LEMMA D

For any set \( \Omega \subseteq \Lambda \) such that \( |\Omega| = k \), we have

\[
I(s_0, u_0 | u_0^\Omega, \hat{s}_{L+W+1}, u_{2L-1}^L) = h(u_0^\Omega | u_0^\Omega, \hat{s}_{L+W+1}, u_{2L-1}^L) - h(u_0^\Omega | \hat{s}_{L+W+1}, u_{2L-1}^L) \tag{132}
\]

\[
= h(u_0^\Omega, u_0 | \hat{s}_{L+W+1}, u_{2L-1}^L) - h(u_0^\Omega | \hat{s}_{L+W+1}, u_{2L-1}^L) - h(u_0^\Omega | \hat{s}_{L+W+1}, u_{2L-1}^L) - h(u_0 | \hat{s}_{L+W+1}, u_{2L-1}^L) \tag{133}
\]

\[
= h(u_0^\Omega | \hat{s}_{L+W+1}, u_{2L-1}^L) - h(u_0^\Omega | \hat{s}_{L+W+1}, u_{2L-1}^L) - h(u_0^\Omega | \hat{s}_{L+W+1}, u_{2L-1}^L) \tag{134}
\]

\[
= h(u_0^\Omega | \hat{s}_{L+W+1}, u_{2L-1}^L) - h(u_0^\Omega | \hat{s}_{L+W+1}, u_{2L-1}^L) - h(u_0^\Omega | \hat{s}_{L+W+1}, u_{2L-1}^L) \tag{135}
\]

\[
= I(s_0, u_0 | u_0^\Omega, \hat{s}_{L+W+1}, u_{2L-1}^L) \tag{136}
\]

11
where (132) and (134) follow the fact that for any $i$, $u_i$ given $s_i$ is independent of all the other random variables. Also in (134), we have used

$$h(u_{\Lambda_k} | \hat{s}_{L+W+1}, u_{L+W+2}^{2L-1}) \leq h(u_{\Omega} | \hat{s}_{L+W+1}, u_{L+W+2}^{2L-1})$$

and this is based on Lemma 2 of the paper. The intuition is that among all the sets $u_{\Omega}$ where $|\Omega| = k$, $u_{\Lambda_k}$ contains random variables closer to $\hat{s}_{L+W+1}, u_{L+W+2}^{2L-1}$ which results in less conditional differential entropy.

**APPENDIX E**

**PROOF OF LEMMA E**

First note that according to Lemma 2 of the paper, for any $k \leq W$

$$h(u_{\Lambda_k} | \hat{s}_{L+W+1}, u_{L+W+2}^{2L-1}, u_{|\Lambda_k|}) \leq k \cdot h(u_{\Lambda_k} | \hat{s}_{L+W+1}, u_{L+W+2}^{2L-1}, u_{|\Lambda_k|})$$

(135)

Now consider

$$\frac{1}{k} I(s_{\Lambda_k}, u_{\Lambda_k} | \hat{s}_{L+W+1}, u_{L+W+2}^{2L-1}, u_{|\Lambda_k|}) = \frac{1}{k+1} \left( k I(s_{\Lambda_k}, u_{\Lambda_k} | \hat{s}_{L+W+1}, u_{L+W+2}^{2L-1}, u_{|\Lambda_k|}) \right)$$

$$= \frac{1}{k+1} \left( k I(s_{\Lambda_k}, u_{\Lambda_k} | \hat{s}_{L+W+1}, u_{L+W+2}^{2L-1}, u_{|\Lambda_k|}) - h(u_i | s_i) \right)$$

(136)

$$\leq \frac{1}{k+1} \left( k I(s_{\Lambda_k}, u_{\Lambda_k} | \hat{s}_{L+W+1}, u_{L+W+2}^{2L-1}, u_{|\Lambda_k|}) - h(u_{\Lambda_k} | s_{\Lambda_k}) \right)$$

(137)

$$= \frac{1}{k+1} \left( I(s_{\Lambda_k}, u_{\Lambda_k} | \hat{s}_{L+W+1}, u_{L+W+2}^{2L-1}, u_{|\Lambda_k|}) \right)$$

(138)

$$= \frac{1}{k+1} \left( I(s_{\Lambda_k}, u_{\Lambda_k} | \hat{s}_{L+W+1}, u_{L+W+2}^{2L-1}, u_{|\Lambda_k|}) \right)$$

(139)

$$= \frac{1}{k+1} \left( I(s_{\Lambda_k}, u_{\Lambda_k} | \hat{s}_{L+W+1}, u_{L+W+2}^{2L-1}, u_{|\Lambda_k|}) \right)$$

(140)

$$= \frac{1}{k+1} \left( I(s_{\Lambda_k}, u_{\Lambda_k} | \hat{s}_{L+W+1}, u_{L+W+2}^{2L-1}, u_{|\Lambda_k|}) \right)$$

(141)
where (136) and (138) follow from the fact that knowing $s_{\Lambda_k}$, $u_{\Lambda_k}$ is i.i.d. vector Gaussian Noise. Also (137) follows from (135). (139) follows from the following Markov chain properties

\begin{align}
    s_{\lambda_{p+L+W-k+1}} &\rightarrow s_{\Lambda_k} \rightarrow u_{\Lambda_k} \\
    s_{\Lambda_k} &\rightarrow s_{\lambda_{p+L+W-k+1}} \rightarrow u_{\lambda_{p+L+W-k+3}}
\end{align}

and (140) is based on the chain rule of mutual information.