Error Exponents for Joint Source-Channel Coding with Delay-Constraints

Cheng Chang
Wireless Foundations, Dept. of EECS
University of California at Berkeley
Email: cchang@eecs.berkeley.edu

Anant Sahai
Wireless Foundations, Dept. of EECS
University of California at Berkeley
Email: sahai@eecs.berkeley.edu

***Abstract*— Traditionally, joint source-channel coding is viewed in the block coding context — all the source symbols are known in advance by the encoder. Here, we consider source symbols to be revealed to the encoder in real time and require that they be reconstructed at the decoder within a certain fixed end-to-end delay. We derive upper and lower bounds on the reliability function with delay for cases with and without channel feedback. For erasure channels with feedback, the upper-bound is shown to be achievable and the resulting scheme shows that from a delay perspective, *nearly* separate source and channel coding is optimal.**

I. Introduction

The block-length story for error exponents is particularly seductive when upper and lower bounds agree — as they do for both lossless source coding and for point-to-point channel coding in the high-rate regime [1], [2]. However, the block-code setting conflates a particular style of implementation with the problem statement itself. Recently, it has become clear that fixed-block codes may indeed incur unnecessarily poor performance with respect to end-to-end delay, even when the required delay is fixed.

In [3], we show that despite the block channel coding reliability functions not changing with feedback in the high rate regime, the reliability function with respect to fixed-delay can in fact improve dramatically with feedback.[1] For fixed-rate lossless source-coding, [5] showed similarly that the reliability function with fixed delay is much better than the reliability with fixed block-length. An example given in [5], [6] illustrated how sometimes an extremely simple and clearly suboptimal nonblock code can dramatically outperform the best possible fixed-length block-code when considering the tradeoff with fixed delay.

These results suggest that a more systematic examination of the tradeoff between delay and probability of error is needed in other contexts as well. The main results in this paper are upper and lower bounds on the error exponents with delay for lossless joint source-channel coding, both with and without feedback.

Without feedback, the upper bound strategy parallels the one used in [3] for channel coding alone. We also derive a lower bound (achievability) on the fixed-delay error exponents by combining variable length source coding and sequential random channel coding. These two bounds are in general not the same.[2] We then study the joint source-channel coding problem with noiseless feedback, giving a "focusing"[3] type of upper bound similar to the pure channel coding result in [3]. We then give a matching lower bound on the joint source-channel reliability for the special case of binary erasure channels. The scheme is interesting because it is essentially a separation based architecture except that it uses a variable-rate interface between the source and channel coding layers.

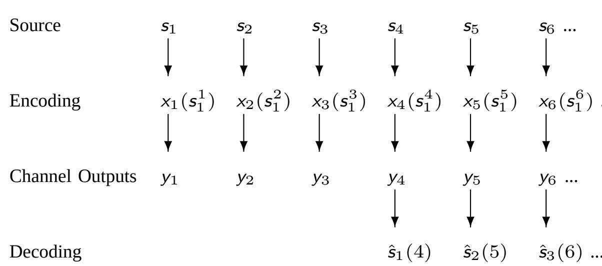

Fig. 1. Sequential joint source-channel coding, delay $\Delta = 3$

A. Review of block joint source-channel coding

The setup is shown in Figure 2. For convenience, there is a common clock for the iid source s_1^n, $s_i \sim Q_s$ [3] from a finite alphabet $\mathcal{S}$ and the DMC channel W. Without loss of generality, the source distribution $Q_s(s) > 0$, $\forall s \in \mathcal{S}$. The DMC has transition matrix $W_{y|x}$ where the input and output alphabets are finite sets $\mathcal{X}$ and $\mathcal{Y}$ respectively. A block joint source-channel coding system for n source symbols consists

[1] It had long been known that the reliability function with respect to *average* block-length can improve, but there was a mistaken assertion by Pinsker in [4] that the fixed-delay exponents do not improve with feedback.

[2] This parallels what was reported in [7] for block-coding. There, separate source-channel coding does not achieve the upper bound of the error exponent on joint source-channel coding.

[3] In this paper, s is random variable, s is the realization of the random variable.

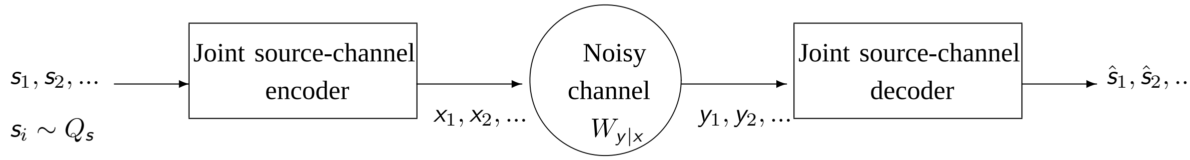

Fig. 2. Joint source-channel coding with a fixed delay requirement

of a encoder-decoder pair $(\mathcal{E}_n, \mathcal{D}_n)$. Where

$$\mathcal{E}_n : \mathcal{S}^n \to \mathcal{X}^n, \qquad \mathcal{E}_n(s_1^n) = x_1^n$$
$$\mathcal{D}_n : \mathcal{Y}^n \to \mathcal{S}^n, \mathcal{D}_n(y_1^n) = \hat{s}_1^n$$

The error probability is $\Pr(s_1^n \neq \hat{s}_1^n) = \Pr(s_1^n \neq \mathcal{D}_n(\mathcal{E}_n(s_1^n)))$. The block source-channel exponent $E_{b,sc}$ is achievable if $\exists$ a family of $\{(\mathcal{E}_n, \mathcal{D}_n)\}$, s.t.[4]

$$\lim_{n\to\infty}\inf -\frac{1}{n}\log_2 \Pr(s_1^n \neq \hat{s}_1^n) = E_{b,sc} \qquad (1)$$

The relevant results of [7] are summarized into the following theorem.

Theorem 1: $E_{b,sc}^{(1)} \leq E_{b,sc} \leq E_{b,sc}^{(2)}$ where

$$E_{b,sc}^{(1)} = \min_{H(Q_s)\leq R\leq \log|\mathcal{S}|}\{e(R) + E_r(R)\}$$
$$E_{b,sc}^{(2)} = \min_{H(Q_s)\leq R\leq \log|\mathcal{S}|}\{e(R) + E(R)\}$$

$e(R) = \min_{U:H(U)\geq R} D(U\|Q)$ is the block source coding error exponent for source Q_s. $E(R)$ is the block channel coding error exponent for channel $W_{y|x}$ and

$$E_r(R) = \max_P \min_V [D(V\|W|P) + |I(P,V) - R|^+]$$

is the random coding error exponent for channel $W_{y|x}$. Thus as shown in [1],

$$E(R) \leq E_{sp}(R) = \max_P \min_{V:I(V,P)\leq R} D(V\|W|P)$$

the sphere packing bound, and $E(R) = E_r(R) = E_{sp}(R)$ for $R_{cr} \leq R \leq C$. As shown in [7], if the minimum of $e(R)+E_r(R)$ is attained for an $R \geq R_{cr}$ then $E_{b,sc} = e(R)+E_r(R)$. This error exponent is in general better than the obvious separate source channel coding error exponent $\max_R\{\min\{e(R), E_r(R)\}\}$ [7].

B. Source coding and channel coding with delay constraint

We review some related results from [5] on sequential lossless source coding and from [3] on sequential channel coding.

[4]We use bits and $\log_2$ in this paper.

For lossless source coding, [5] derives a tight[5] bound on the error exponent with a hard delay constraint: $\forall i$, $\lim\inf_{\Delta\to\infty}\frac{-1}{\Delta}\log_2\Pr(s_i \neq \hat{s}_i(i+\Delta)) = E_{s,s}(R)$, where $E_{s,s}(R)$ is the sequential source coding error exponent:

$$E_{s,s}(R) = \inf_{\alpha>0, U_s:H(U_s)\geq(1+\alpha)R}\frac{1}{\alpha}D(U_s\|Q_s)$$
$$= \inf_{\alpha>0}\frac{1}{\alpha}e((1+\alpha)R) \qquad (2)$$

Where $e(R) = \inf_{U_s:H(U_s)\geq R} D(U_s\|Q_s)$ is the block source coding error exponent [2].

For channel coding without feedback but facing a hard delay constraint, [9] reviews how the random block-coding error exponent $E_r(R)$ governs how the probability of bit error decays with delay[6] for random tree codes. Formally, for all i[7] and Δ, there exists a sequential channel coding system, s.t.

$$\lim_{\Delta\to\infty}\inf\frac{-1}{\Delta}\log_2\Pr(b_i \neq \hat{b}_i(\frac{i}{R}+\Delta)) = E_r(R) \qquad (3)$$

where $\hat{b}_i(\frac{i}{R}+\Delta)$is the estimate of b_i after in total $\frac{i}{R}+\Delta$ channel uses, i.e. Δ seconds after b_i enters the encoder.

In [3], we generalize Pinsker's result from [4] to show that the Haroutunian Bound [10] $E^+(R)$ serves as an upper bound on the error exponent for channel coding with delay. Formally, for channel $V_{y|x}$, for any channel coding scheme, there exists a finite function $i(\Delta)$ on $\Delta, s.t.$ the error exponent with delay:

$$\lim_{\Delta\to\infty}\inf\frac{-1}{\Delta}\log_2\Pr(b_{i(\Delta)} \neq \hat{b}_{i(\Delta)}(\frac{i(\Delta)}{R}+\Delta)) \leq E^+(R) \qquad (4)$$

where $E^+(R)$ is the Haroutunian bound.

$$E^+(R) = \inf_{V_{y|x}:I(P_V,V_{y|x})<R}\sup_P D(V_{y|x}\|W_{y|x}|P)$$
$$= \inf_{V_{y|x}:I(P_V,V_{y|x})<R}\sup_{x\in\mathcal{X}} D(V_{y|x}(.|x)\|W_{y|x}(.|x))$$

[5]This bound is achieved by using a finite delay prefix code in [5].

[6]For channel coding with delay constraints, we measure the delay in terms of channel uses(seconds), not number of information bits.

[7]the i'th information bit b_i enters the encoder at the $\frac{i}{R}$'th channel use(second).

$I(P_V, V_{y|x})$ is the mutual information of input distribution P_V and channel $V_{y|x}$ where P_V is the input distribution to maximize $I(P_V, V_{y|x})$. The Haroutunian bound is the same as the random coding error exponent for symmetric channels in high rate regime [3].

In [3], an upper bound is given for the error exponent for delay-constrained channel coding with causal noiseless feedback problem:

$$\lim \inf_{\Delta\to\infty} \frac{-1}{\Delta} \log_2 \Pr(b_{i(\Delta)} \neq \hat{b}_{i(\Delta)}(\frac{i(\Delta)}{R} + \Delta)) \leq \inf_{\lambda\in[0,1]} \frac{E^+(\lambda R)}{1-\lambda} \quad (5)$$

The resulting "focusing" bound $E_a(R) \triangleq \inf_{\lambda\in[0,1]} \frac{E^+(\lambda R)}{1-\lambda}$ is strictly larger than the Haroutunian bound. This bound is achievable for binary erasure channels as well as sufficiently symmetric DMCs with strictly positive zero-error capacity.

C. Sequential Joint Source-Channel Coding

Rather than being known in advance, the source symbols enter the encoder in a real-time fashion. We assume that the source generates 1 source symbol s per second from the finite alphabet $\mathcal{S}$. The i'th source symbol s_i is not known at the encoder until time i at the decoder. Without loss of generality, the encoder uses 1 channel use per second.[8]

Definition 1: A sequential encoder-decoder pair $\mathcal{E}, \mathcal{D}$ are sequences of maps. $\{\mathcal{E}_i\}, i = 1, 2, ...$ and $\{\mathcal{D}_i\}, i = 1, 2, ...$ The output of $\mathcal{E}_i$ are the input to the channel at time i.

$$\mathcal{E}_i : \mathcal{S}^i \longrightarrow \mathcal{X}$$
$$\mathcal{E}_i(s_1^i) = x_i$$

The outputs of $\mathcal{D}_i$ are the decoding decisions of the i'th source symbol based on the channel outputs up to time $i + \Delta$.

$$\mathcal{D}_i : \mathcal{Y}^{i+\Delta} \longrightarrow \mathcal{S}$$
$$\mathcal{D}_i(y_1^{i+\Delta}) = \widehat{s}_i(i + \Delta)$$

where $\widehat{s}_i(i + \Delta)$ is the estimation of s_i at time $i + \Delta$ thus has end-to-end delay of Δ seconds. A sequential source-channel coding system is illustrated in Figure 1.

For sequential source-channel coding, it is important to study how symbol-wise decoding error decays with delay. This is parallel to the study of decoding error with respect to block length in block coding.

[8] If the source and channel are not synchronized, we can always synchronize them by grouping multiple source symbols and channels uses to a super source symbol and a super channel use respectively.

Definition 2: A joint source-channel coding delay-reliability(error exponent) $E_{s,sc}$ is achievable if and only if there exists a family of sequential source-channel codes $\{(\mathcal{E}, \mathcal{D})\}$ s.t. for all i,

$$\lim \inf_{\Delta\to\infty} \frac{-1}{\Delta} \log_2 \Pr(s_i \neq \widehat{s}_i(i + \Delta)) = E_{s,sc}$$

Write the delay-reliability with causal noiseless feedback as E_{s,sc_f}. In the rest of the paper, we study upper and lower bounds on $E_{s,sc}$ and E_{s,sc_f}.

II. Upper bound on error exponents with fixed delay

In this section we derive an upper bound on the sequential joint source channel error exponent $E_{s,sc}$.

Theorem 2: For the source-channel coding problem, if the source is iid $\sim Q_s$ from a finite alphabet and the channel is a DMC with transition probability $W_{y|x}$, then the error exponents $E_{s,sc}$ with fixed delay must satisfy $E_{s,sc} \leq E^{(2)}_{s,sc}$, where

$$\begin{aligned} E^{(2)}_{s,sc} &= \inf_R \{ \inf_{V_{y|x}: I(P_V, V_{y|x}) < R} \sup_P D(V_{y|x} \| W_{y|x} | P) \\ &+ \inf_{\alpha>0, U_s: H(U_s) > (1+\alpha)R} \frac{1}{\alpha} D(U_s \| Q_s) \} \\ &= \inf_R \{ E^+(R) + E_{s,s}(R) \} \end{aligned} \quad (6)$$

The theorem is proved using a variation of the bounding technique used in [3] (and originating in [4]) for the fixed-delay channel coding problem. Lemmas 1-6 are the joint source-channel coding counterparts to Lemmas 4.1-4.5 in [3]. The idea of the proof is to first build a feed-forward sequential source-channel decoder which has access to the previous source symbols in addition to the channel outputs. The second step is to construct a block source-channel coding scheme from the optimal feed-forward sequential decoder and showing that if the source-channel pair behave atypically enough, then the decoding error probability will be large for at least one of the source symbols. The next step is to prove that neither the atypicality of the channel before that particular source symbol nor the atypicality of the source after that particular source symbol causes the error because of the feed-forward information. Thus the cause of the decoding error for that particular symbol is the atypical behavior of the future channel or the atypical behavior of the past source symbols. The last step is to lower bound the probability of the atypical behavior and upper bound the error exponents. The proof spans into the next several subsections.

A. Feed-forward decoders

Definition 3: A delay Δ decoder $\mathcal{D}^\Delta$ with feed-forward is a decoder $\mathcal{D}^\Delta_j$ that also has access to the

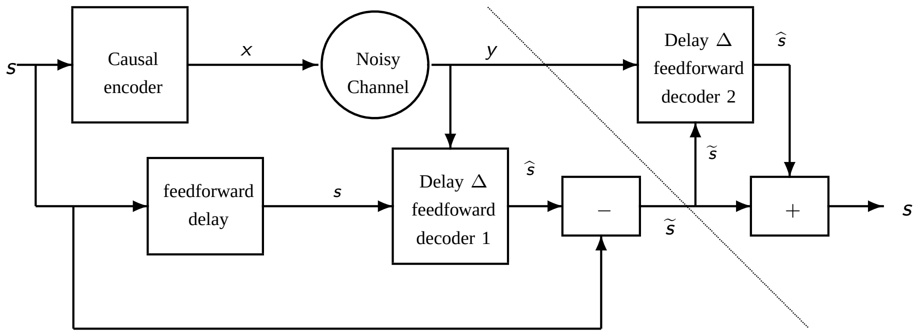

Fig. 3. A cutset illustration of the Markov Chain $s_1^n \; - \; (\tilde{s}_1^n, y_1^{n+\Delta}) \; - \; s_1^n$. Decoder 1 and decoder 2 are type I and II delay Δ feed-forward decoders respectively. They are equivalent.

past source symbols s_1^{j-1} in addition to the channel outputs $y_1^{j+\Delta}$.
Using this feed-forward decoder:

$$\hat{s}_j(j+\Delta) = \mathcal{D}_j^{\Delta}(y_1^{j+\Delta}, s_1^{j-1}) \tag{7}$$

Lemma 1: For any source-channel encoder $\mathcal{E}$, the optimal delay Δ decoder $\mathcal{D}^{\Delta}$ with feed-forward only needs to depend on $y_j^{j+\Delta}, s_1^{j-1}$.

Proof: The source s_i is an iid random process and the channel inputs $x_1^{j+\Delta}$ are functions of $s_1^{j+\Delta}$, so we have the Markov chain: $y_1^{j-1} \; - \; x_1^{j-1} \; - \; (s_1^{j-1}, y_j^{j+\Delta}) \; - \; s_j^{j+\Delta}$. Conditioned on the past source symbols, the past channel outputs are completely irrelevant for estimation. □

Write the error sequence of the feed-forward decoder as[9] $\tilde{s}_i = s_i - \hat{s}_i$. Then we have the following property for the feed-forward decoders.

Lemma 2: Given a joint source-channel encoder $\mathcal{E}$, the optimal delay Δ decoder $\mathcal{D}^{\Delta}$ with feed-forward for symbol j only needs to depend on $y_1^{j+\Delta}, \tilde{s}_1^{j-1}$

Proof: Proceed by induction. It holds for $j = 1$ since there are no prior source symbols. Suppose that it holds for all $j < k$ and consider $j = k$. By the induction hypothesis, the action of all the prior decoders j can be simulated using $(y_1^{j+\Delta}, \tilde{s}_1^{j-1})$ giving $\hat{s}_1^{k-1}$. This in turn allows the recovery of s_1^{k-1} since we also know $\tilde{s}_1^{k-1}$. Thus the decoder is equivalent. □

We call the feed-forward decoders in Lemmas 1 and 2 type I and II delay Δ feed-forward joint source-channel decoders respectively. Lemma 1 and 2 tell us that feed-forward decoders can be thought of in three ways: having access to all channel outputs and all past source symbols, $(y_1^{j+\Delta}, s_1^{j-1})$, having access to a recent window of channel outputs and all past source symbols, $(y_j^{j+\Delta}, s_1^{j-1})$, or having access to all channel outputs and all past decoding errors, $(y_1^{j+\Delta}, \tilde{s}_1^{j-1})$.

B. *Constructing a block code*

To encode a block of n source symbols, just run the joint source-channel encoder $\mathcal{E}$ and terminate with the encoder run using some random source symbols drawn according to the distribution of Q_s. To decode the block, just use the delay Δ decoder $\mathcal{D}^{\Delta}$ with feed-forward, and then use the feedforward error signals to correct any mistakes that might have occurred. As a block coding system, this hypothetical system never makes an error from end to end. As shown in Figure 3, the data processing inequality implies:

Lemma 3: If n is the block-length, the channel inputs are x_1^n, then

$$H(\tilde{s}_1^n) \geq H(s_1^n) - I(x_1^{n+\Delta}, y_1^{n+\Delta}) \tag{8}$$

Proof:

$$\begin{aligned} H(s_1^n) \quad &= \quad I(s_1^n; s_1^n) \\ &=_{(a)} \quad I(s_1^n; \tilde{s}_1^n, y_1^{n+\Delta}) \\ &=_{(b)} \quad I(s_1^n; y_1^{n+\Delta}) + I(s_1^n; \tilde{s}_1^n | y_1^{n+\Delta}) \\ &\leq_{(c)} \quad I(x_1^{n+\Delta}; y_1^{n+\Delta}) + I(s_1^n; \tilde{s}_1^n | y_1^{n+\Delta}) \\ &\leq \quad I(x_1^{n+\Delta}; y_1^{n+\Delta}) + H(\tilde{s}_1^n) \end{aligned}$$

(a) is true because of the data processing inequality considering the following Markov chain: $s_1^n \; - \; (\tilde{s}_1^n, y_1^{n+\Delta}) \; - \; s_1^n$ and the fact that $I(s_1^n; s_1^n) = H(s_1^n) \geq I(s_1^n; \tilde{s}_1^n, , y_1^{n+\Delta})$. (b) is the chain rule for mutual information. (c) is true because of the the data processing inequality considering the following Markov chain: $s_1^n \; - x_1^{n+\Delta} \; - y_1^{n+\Delta})$. □

C. *Lower bound the symbol-wise error probability*

Now suppose this block-code were to be run with the source distribution U_s from time 1 to n and were to be run with the distribution Q_s from time $n+1$ to $n+\Delta$, *and* the channel transition matrix $V_{y|x}$ from time 1 to $n+\Delta$ s.t. $nH(U_s) > I(x_1^{n+\Delta}; y_1^{n+\Delta})$,

[9] For any finite $|\mathcal{S}|$, we can always define a group $Z_{|\mathcal{S}|}$ on $\mathcal{S}$, where the operators $-$ and $+$ are indeed $-, + \mod |\mathcal{S}|$

where $x_1^{n+\Delta}$ is the input random vector to the channel. We first examine $I(x_1^{n+\Delta}; y_1^{n+\Delta})$ under the discrete memoryless channel law[10] $V_{y|x}$. By Lemma 8.9.2 in [11]:

$$I(x_1^{n+\Delta}; y_1^{n+\Delta}) \leq \sum_{i=1}^{n+\Delta} I(x_i; y_i)$$

Now assume P_V is the distribution on the input to maximize $I(x; y)$ given the channel law $V_{y|x}$. Then:

$$I(x_1^{n+\Delta}; y_1^{n+\Delta}) \leq (n+\Delta) I(P_V, V_{y|x})$$

If $nH(U_s) > (n+\Delta)I(P_V, V_{y|x}) \geq I(x_1^{n+\Delta}; y_1^{n+\Delta})$ then the block coding scheme constructed in the previous section will with probability 1 make a block error. Moreover, many individual symbols will also be in error often:

Lemma 4: If the source is coming from U_s from time 1 to n, *and* the channel law is $V_{y|x}$ from time 1 to time $n+\Delta$, such that $H(U_s) > (1+\frac{\Delta}{n})I(P_V, V_{y|x})$, then there exists a $\delta > 0$ so that for n large enough (here we fix the ratio $\frac{\Delta}{n}$), the feed-forward decoder will make at least $\frac{H(U_s) - \frac{n+\Delta}{n} I(P_V, V_{y|x})}{2\log_2|\mathcal{S}| - (H(U_s) - \frac{n+\Delta}{n} I(P_V, V_{y|x}))} n$ symbol errors with probability δ or above. δ satisfies [11] $h_\delta + \delta \log_2(|\mathcal{S}|-1) = \frac{1}{2}(H(U_s) - \frac{n+\Delta}{n} I(P_V, V_{y|x}))$.

Proof: Lemma 3 implies:

$$\begin{aligned}\sum_{i=1}^{n} H(\tilde{s}_i) &\geq H(\tilde{s}_1^n) \\ &\geq nH(U_s) - (n+\Delta)I(P_V, V_{y|x}) \quad (9)\end{aligned}$$

The average entropy per source symbol for $\tilde{s}$ is at least $H(U_s) - \frac{n+\Delta}{n} I(P_V, V_{y|x})$. Now suppose that $H(\tilde{s}_i) \geq \frac{1}{2}(H(U_s) - \frac{n+\Delta}{n} I(P_V, V_{y|x}))$ for t positions. By noticing that $H(\tilde{s}_i) \leq \log_2|\mathcal{S}|$, we have

$$\sum_{i=1}^{n} H(\tilde{s}_i) \leq t\log_2|\mathcal{S}| + (n-t)\frac{1}{2}(H(U_s) - \frac{n+\Delta}{n} I(P_V, V_{y|x}))$$

With Eqn. 9, we derive the desired result:

$$t \geq \frac{(H(U_s) - \frac{n+\Delta}{n} I(P_V, V_{y|x}))}{2\log_2|\mathcal{S}| - (H(U_s) - \frac{n+\Delta}{n} I(P_V, V_{y|x}))} n \quad (10)$$

Where $2\log_2|\mathcal{S}| - (H(U_s) - \frac{n+\Delta}{n} I(P_V, V_{y|x})) \geq 2\log_2|\mathcal{S}| - H(U_s) \geq 2\log_2|\mathcal{S}| - \log_2|\mathcal{S}| > 0$

Now for t positions $1 \leq j_1 < j_2 < \ldots < j_t \leq n$ the individual entropy $H(\tilde{s}_j) \geq \frac{1}{2}(H(U_s) - \frac{n+\Delta}{n} I(P_V, V_{y|x}))$. By the property of the binary entropy function[12], $\Pr(\tilde{s}_j \neq s_0) = \Pr(s_j \neq \hat{s}_j) \geq \delta$.$\square$

We can pick $j^* = j_t \geq t$, by the previous lemma we know that $j^* \geq \frac{1}{2}\frac{(H(U_s) - \frac{n+\Delta}{n} I(P_V, V_{y|x}))}{2\log_2|\mathcal{S}| - (H(U_s) - \frac{n+\Delta}{n} I(P_V, V_{y|x}))} n$, so if we fix $\frac{\Delta}{n}$ and let n go to infinity, then j^* goes to infinity as well.

At this point, Lemma 1 and 4 together imply that even if the channel only behaves like it came from the channel law $V_{y|x}$ from time j^* to $j^*+\Delta$ *and* the source behaves like it came from a distribution U_s from time 1 to j^*-1, s.t. $H(U_s) > (1+\frac{\Delta}{n})I(P_V, V_{y|x})$, whatever the source distribution from time $n+1$ to time $n+\Delta$ is, the same minimum error probability δ still holds. So we assume the source from time $n+1$ to time $n+\Delta$ is from distribution Q_s.

Now define the "bad source-channel-sequence" set E_{j^*} as the set of source and channel output sequence pairs so that the type I delay Δ joint source-channel decoder makes a decoding error at j^*. Formally[13]

$E_{j^*} = \{(\vec{s}, \bar{\bar{y}}) | s_{j^*} \neq \mathcal{D}^{\Delta}_{j^*}(\bar{s}, \bar{\bar{y}})\}$. By Lemma 4, $\Pr(E_{j^*}) \geq \delta$. Notice that E_{j^*} does not depend on the distribution of the source *or* the channel behavior but only on the encoder-decoder pair. Define $J = \min\{n, j^*+\Delta\}$, and $\bar{\bar{\bar{s}}} = s_1^J$.Now we write the typical set for distribution U_s: $A^{\epsilon_1}_{j^*}(U_s) = \{\vec{s} | \forall s \in \mathcal{S} : \frac{n_s(\bar{\bar{\bar{s}}})}{J} \in (U_s(s) - \epsilon_1, U_s(s) + \epsilon_1)\}$ and for each $\vec{s}$, write the strongly typical set for channel $V_{y|x}$: $A^{\epsilon_1,\epsilon_2}_{j^*}(V_{y|x}|\vec{s}) = \{\bar{\bar{y}} | \forall x \in \mathcal{X} \;\; either \;\; \frac{n_x(\bar{\bar{x}}(\vec{s}))}{\Delta} < \epsilon_2 \;\; or \;\; \forall y \in \mathcal{Y}, \frac{n_{x,y}(\bar{\bar{x}}(\vec{s}), \bar{\bar{y}})}{n_x(\bar{\bar{x}}(\vec{s}))} \in (v_{y|x} - \epsilon_1, v_{y|x} + \epsilon_1)\}$. Finally we have the joint strongly typical set for source U_s and channel $V_{y|x}$: $A^{\epsilon_1,\epsilon_2}_{j^*}(U_s, V_{y|x}) = \{(\vec{s}, \bar{\bar{y}}) | \vec{s} \in A^{\epsilon_1}_{j^*}(U_s) \;\; and \;\; \bar{\bar{y}} \in A^{\epsilon_1,\epsilon_2}_{j^*}(V_{y|x}|\vec{s})\}$

Lemma 5: $\Pr_{U_s, V_{y|x}}(E_{j^*} \cap A^{\epsilon_1,\epsilon_2}_{j^*}(U_s, V_{y|x})) \geq \frac{\delta}{2}$ for large n and Δ.

Proof: First, we define $A^{\epsilon_1}_{j^*}(U_s^C) = \{(\vec{s}, \bar{\bar{y}}) | \exists s \in \mathcal{S} : \frac{n_s(\bar{\bar{\bar{s}}})}{J} \notin (U_s(s) - \epsilon_1, U_s(s) + \epsilon_1)\}$ and $A^{\epsilon_1,\epsilon_2}_{j^*}(\vec{s}, V^C_{y|x}) = \{(\vec{s}, \bar{\bar{y}}) | \forall x \in \mathcal{X} \;\; either \;\; \frac{n_x(\bar{\bar{x}}(\vec{s}))}{\Delta} < \epsilon_2 \;\; or \;\; \forall y \in \mathcal{Y}, \frac{n_{x,y}(\bar{\bar{x}}(\vec{s}), \bar{\bar{y}})}{n_x(\bar{\bar{x}}(\vec{s}))} \in (v_{y|x} - \epsilon_1, v_{y|x} + \epsilon_1)\}$

So the total set $\{(\vec{s}, \bar{\bar{y}})\}$ can be partitioned as

$$\{(\vec{s}, \bar{\bar{y}})\} = A^{\epsilon_1,\epsilon_2}_{j^*}(U_s, V_{y|x}) \cup (A^{\epsilon_1}_{j^*}(U_s^C) \cup [\cup_{\vec{s}} A^{\epsilon_1,\epsilon_2}_{j^*}(\vec{s}, V^C_{y|x})]) \quad (11)$$

Fix $\frac{\Delta}{n}$, let n go to infinity, then $J = \min\{n, j^*+\Delta\}$ and Δ go to infinity. By strong typicality in Lemma 13.6.1 in [11], we know that $\forall \epsilon_1 > 0$, if J is large enough, then $\Pr_{U_s, V_{y|x}}(A^{\epsilon_1}_{j^*}(U_s^C)) \leq \frac{\delta}{4}$. By the same lemma: $\forall \vec{s} : V_{y|x}(A^{\epsilon_1,\epsilon_2}_{j^*}(\vec{s}, V^C_{y|x}) | \bar{\bar{x}}(\vec{s})) \leq \frac{\delta}{4}$. Because the channel behavior is independent with the source,

[10]We write the transition probability $V_{y|x}(x, y) = v_{y|x}$

[11]Write $h_\delta = -\delta\log_2\delta - (1-\delta)\log_2(1-\delta)$

[12]s_0 is the zero element in the finite group $Z_{|\mathcal{X}|}$.

[13]To simplify the notation, write: $\vec{s} = s_1^{j^*+\Delta}$, $\bar{s} = s_1^{j^*-1}$, $\bar{\bar{s}} = s_{j^*}^{j^*+\Delta}$, $\bar{\bar{y}} = y_{j^*}^{j^*+\Delta}$

we have:

$$\begin{aligned}
& \Pr_{U_s,V_{y|x}}(A^{\epsilon_1,\epsilon_2}_{j^*}(U_s,V_{y|x})) \\
=_{(a)}\ & 1-\Pr_{U_s,V_{y|x}}(A^{\epsilon_1}_{j^*}(U_s^C)) \\
& -\Pr_{U_s,V_{y|x}}(\cup_{\vec{s}}A^{\epsilon_1,\epsilon_2}_{j^*}(\vec{s},V^C_{y|x})) \\
\geq_{(b)}\ & 1-\frac{\delta}{4}-\sum_{\vec{s}}U_s(\vec{s})V_{y|x}(A^{\epsilon_1,\epsilon_2}_{j^*}(\vec{s},V^C_{y|x})|\bar{\bar{x}}(\vec{s})) \\
\geq\ & 1-\frac{\delta}{4}-\sum_{\vec{s}}U_s(\vec{s})\frac{\delta}{4} \\
=\ & 1-\frac{\delta}{2} \qquad (12)
\end{aligned}$$

(a) is true because of (11). (b) is true because the source and channel behavior are independent. From this we know:

$$\begin{aligned}
& \Pr_{U_s,V_{y|x}}(E_{j^*}\cap A^{\epsilon_1,\epsilon_2}_{j^*}(U_s,V_{y|x})) \\
\geq_{(a)}\ & \Pr_{U_s,V_{y|x}}(E_{j^*}) \\
& -(1-\Pr_{U_s,V_{y|x}}(A^{\epsilon_1,\epsilon_2}_{j^*}(U_s,V_{y|x}))) \\
\geq\ & \delta-\frac{\delta}{2} \\
=\ & \frac{\delta}{2}
\end{aligned}$$

(a) is true because $\Pr(A_1\cap A_2)\geq\Pr(A_1)-\Pr(A_2^C)=\Pr(A_1)-(1-\Pr(A_2))$ □

For source channel output pair $(\vec{s},\bar{\bar{y}})\in A^{\epsilon_1,\epsilon_2}_{j^*}(U_s,V_{y|x})$, we can bound the ratio of the probability of $(\vec{s},\bar{\bar{y}})$ under source U_s, channel rule $V_{y|x}$ and the probability of $(\vec{s},\bar{\bar{y}})$ under the true source Q_s and true channel rule $W_{y|x}$ as follows.

Lemma 6: $\forall\epsilon>0$, we can pick ϵ_1,ϵ_2 small enough, s.t. for large enough n ($\frac{\Delta}{n}$ is fixed): $\forall(\vec{s},\bar{\bar{y}})\in A^{\epsilon_1,\epsilon_2}_{j^*}(U_s,V_{y|x})$,

$$\frac{\Pr_{Q_s,W_{y|x}}((\vec{s},\bar{\bar{y}}))}{\Pr_{U_s,V_{y|x}}((\vec{s},\bar{\bar{y}}))} \geq 2^{-\Delta\sup_P D(V_{y|x}\|W_{y|x}|P)-j^*D(U_s\|Q_s)-(j^*+\Delta)\epsilon}$$

Proof: The source and the channel behavior are independent, so:

$$\frac{\Pr_{Q_s,W_{y|x}}((\vec{s},\bar{\bar{y}}))}{\Pr_{U_s,V_{y|x}}((\vec{s},\bar{\bar{y}}))}=\frac{Q_s(\vec{s})W_{y|x}(\bar{\bar{y}}|\bar{\bar{x}}(\vec{s}))}{U_s(\vec{s})V_{y|x}(\bar{\bar{y}}|\bar{\bar{x}}(\vec{s}))}$$

$\vec{s}\in A^{\epsilon_1}_{j^*}(U_s)$, from Lemma 6 in [12], we know that for n large enough:

$$\frac{Q_s(\vec{s})}{U_s(\vec{s})}\geq 2^{-j^*D(U_s\|Q_s)-j^*\epsilon}$$

Meanwhile, $\bar{\bar{y}}\in A^{\epsilon_1,\epsilon_2}_{j^*}(\vec{s},V_{y|x})$, from Lemma 4.4 in [3], we know that for n large enough:

$$\frac{W_{y|x}(\bar{\bar{y}}|\bar{\bar{x}}(\vec{s}))}{V_{y|x}(\bar{\bar{y}}|\bar{\bar{x}}(\vec{s}))}\geq 2^{-\Delta\sup_P D(V_{y|x}\|W_{y|x}|P)-\Delta\epsilon}$$

Combining the above two inequalities, we get the desired result. □

We are ready to bound the error probability of the j^*'th source symbol under the true source distribution Q_s and channel rule $W_{y|x}$.

Lemma 7: $\forall\epsilon>0$, and large enough n ($\frac{\Delta}{n}$ fixed):

$$\Pr_{Q_s,W_{y|x}}(E_{j^*})\geq \frac{\delta}{2}2^{-\Delta\sup_P D(V_{y|x}\|W_{y|x}|P)-j^*D(U_s\|Q_s)-(j^*+\Delta)\epsilon}$$

Proof: Combining Lemma 5 and 6.

$$\begin{aligned}
& \Pr_{Q_s,W_{y|x}}(E_{j^*}) \\
& \geq \Pr_{Q_s,W_{y|x}}(E_{j^*}\cap A^{\epsilon_1,\epsilon_2}_{j^*}(U_s,V_{y|x})) \\
& = \Pr_{U_s,V_{y|x}}(E_{j^*}\cap A^{\epsilon_1,\epsilon_2}_{j^*}(U_s,V_{y|x})) \\
& \quad \frac{\Pr_{Q_s,W_{y|x}}(E_{j^*}\cap A^{\epsilon_1,\epsilon_2}_{j^*}(U_s,V_{y|x}))}{\Pr_{U_s,V_{y|x}}(E_{j^*}\cap A^{\epsilon_1,\epsilon_2}_{j^*}(U_s,V_{y|x}))} \\
& \geq_{(a)} \frac{\delta}{2}\frac{\Pr_{Q_s,W_{y|x}}(E_{j^*}\cap A^{\epsilon_1,\epsilon_2}_{j^*}(U_s,V_{y|x}))}{\Pr_{U_s,V_{y|x}}(E_{j^*}\cap A^{\epsilon_1,\epsilon_2}_{j^*}(U_s,V_{y|x}))} \\
& \geq_{(b)} \frac{\delta}{2}2^{-\Delta\sup_P D(V_{y|x}\|W_{y|x}|P)-j^*D(U_s\|Q_s)-(j^*+\Delta)\epsilon}
\end{aligned}$$

(a) is true because of Lemma 5 and (b) is true because of Lemma 6. □

Now we are ready to prove Theorem 2. For U_s and $V_{y|x}$, as long as $H(U_s)>\frac{n+\Delta}{n}I(P_V,V_{y|x})$, we have a constant $\delta>0$, by letting ϵ go to 0, Δ and n go to infinity proportionally, we have:

$$\begin{aligned}
& \frac{-1}{\Delta}\log_2\Pr_{Q_s,W_{y|x}}(s_{j^*}(j^*+\Delta)\neq s_{j^*}) \\
& = \frac{-1}{\Delta}\log_2\Pr_{Q_s,W_{y|x}}(E_{j^*}) \\
& \leq \epsilon+\sup_P D(V_{y|x}\|W_{y|x}|P)+\frac{j^*}{\Delta}D(U_s\|Q_s) \\
& \leq \epsilon+\sup_P D(V_{y|x}\|W_{y|x}|P)+\frac{n}{\Delta}D(U_s\|Q_s)
\end{aligned}$$

where ϵ can be arbitrarily small. Notice that $H(U_s)>\frac{n+\Delta}{n}I(P_V,V_{y|x})$ is equivalent to $\exists R$, s.t. $\frac{1}{1+\alpha}H(U_s)>R>I(P_V,V_{y|x})$, where $\alpha=\frac{\Delta}{n}$. Then the upper bound on the error exponent is the minimum of the above error exponents over all $\alpha>0$, i.e:

$$\begin{aligned}
E^{(2)}_{s,sc} &= \inf_R \inf_{\alpha>0,U_s,V_{y|x}:I(P_V,V_{y|x})<R,H(U_s)>(1+\alpha)R} \\
& \quad \{\sup_P D(V_{y|x}\|W_{y|x}|P)+\frac{1}{\alpha}D(U_s\|Q_s)\} \\
&= \inf_R\{\inf_{V_{y|x}:I(P_V,V_{y|x})<R}\sup_P D(V_{y|x}\|W_{y|x}|P) \\
& \quad +\inf_{\alpha>0,U_s:H(U_s)>(1+\alpha)R}\frac{1}{\alpha}D(U_s\|Q_s)\} \\
&= \inf_R\{E^+(R)+E_{s,s}(R)\} \qquad (13)
\end{aligned}$$

■

Both the upper bound for sequential error exponent and the upper bound for the block source channel

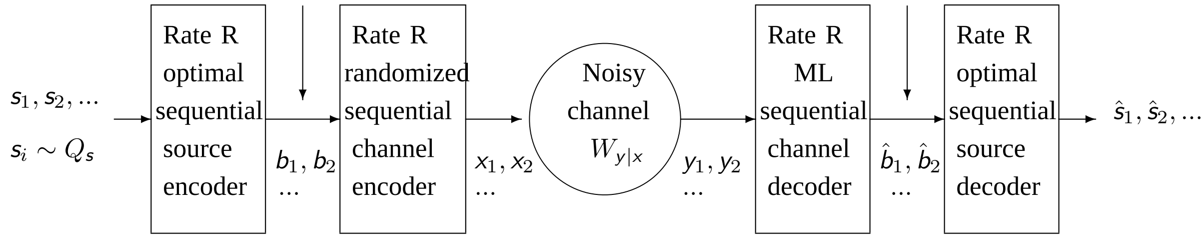

Fig. 4. Separate source-channel coding

coding error exponent have the form of a minimization over the sum of the source coding and the channel coding error exponents over an imaginary rate R. However, the sequential source coding error exponent $E_{s,s}(R)$ is strictly larger than the block source coding error exponent $e(R)$ in general. Thus the upper bound for sequential joint source-channel coding error exponent is generally strictly larger than its block coding counterpart.

III. Lower bound on achievable error exponents with delay

An obvious lower bound on the error exponent can be obtained by implementing an optimal sequential source code [5] and randomized sequential channel code [13] separately. Both the source coding and channel coding are committed at the same rate R, $R \in (H(Q_s), C_W)$, where C_W is the capacity of the channel. This separate source channel coding system works as shown in Figure 4.

Theorem 3: For the iid source $\sim Q_s$ and a discrete memoryless channel with transition probability $W_{y|x}$, the delay-reliability defined in Definition 2 is lower bounded by $E^{(1)}_{s,sc}$, where $E^{(1)}_{s,sc} = \max_R\{\min\{E_r(R), E_{s,s}(R)\}\}$. i.e. there exists a separate source channel coding system committed to rate R^*, where R^* maximizes $\min\{E_r(R), E_{s,s}(R)\}$ s.t. for all $\epsilon > 0$, there exists $K < \infty$, s.t. for all i, Δ,

$$\Pr(s_i \neq \hat{s}_i(i+\Delta)) \leq K 2^{-\Delta(E^{(1)}_{s,sc} - \epsilon)}$$

Proof: For this coding scheme, a decoding error on s_i at time $i+\Delta$ occurs only if the first channel error occurs before the last information bit describing s_i or the last information bit about s_i is not fed into the channel encoder at time $i+\Delta$ yet. The union bound on the error events is as follows. For all $\epsilon > 0$, there exists $K < \infty$ s.t.

$$\begin{aligned}
\Pr(s_i \neq \hat{s}_i(i+\Delta)) &\leq_{(a)} \sum_{k=1}^{iR} \Pr_W(b_1^k \neq \hat{b}_1^k(i+\Delta)) \\
&+ \sum_{k=iR}^{(i+\Delta)R} \Pr_W(b_1^k \neq \hat{b}_1^k(i+\Delta)) \Pr_s(s_i \neq \hat{s}_i(\frac{k}{R})) \\
&+ \Pr_s(s_i \neq \hat{s}_i(i+\Delta)) \\
&\leq_{(b)} \sum_{k=1}^{iR} K_1 2^{-(i+\Delta-\frac{k}{R})E_r(R)} \\
&+ \sum_{k=iR}^{(i+\Delta)R} K_1 2^{-(i+\Delta-\frac{k}{R})E_r(R)} 2^{-(\frac{k}{R}-i)(E_{s,s}(R)-\epsilon_1)} \\
&+ K_1 2^{-\Delta E_{s,s}(R)} \\
&\leq K_2 \Delta 2^{-\Delta(\min\{E_{s,s}(R), E_r(R)\}-\epsilon_1)} \\
&\leq K 2^{-\Delta(\min\{E_{s,s}(R), E_r(R)\}-\epsilon)}
\end{aligned} \tag{14}$$

where K_1, K_2 and ϵ_1 are properly chosen finite real numbers. (a) is the union bound on all possible error events and the fact that channel coding and source coding are independent as shown in Figure 4. (b) is by (2) and (3). $E^{(1)}_{s,sc}$ The above coding scheme works for all $R \in (H(Q_s), C_W)$. The optimal rate R^* is chosen for the source coder ***and*** the channel coder to achieve the error exponent $E^{(1)}_{s,sc}$. ■

Because $E_r(R) \leq E^+(R)$, in general the lower bound $E^{(1)}_{s,cs}$ is strictly smaller than the upper bound $E^{(2)}_{s,cs}$ in Eqn.13. This is comparable to the obvious non-optimal achievable error exponent for block source-channel coding.

IV. Joint source-channel coding with feedback

In this section, we study the joint source-channel coding with feedback problem under delay constraint. As shown in Figure 5, the output of the channel is fed back to the joint source-channel encoder noiselessly. Recall that E_{s,sc_f} is the error exponent with feedback with decoding delay.

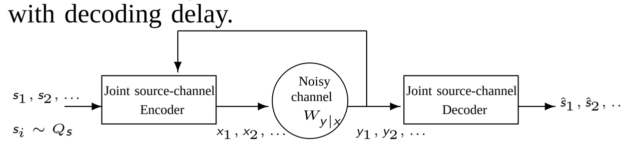

Fig. 5. joint source-channel coding with feedback

A. Upper bound on E_{s,sc_f}

Theorem 4: For the iid source $\sim Q_s$ and a discrete memoryless channel with transition probability $W_{y|x}$ and noiseless feedback, then

$$E_{s,sc_f} \leq \min_{\lambda \in [0,1], R}\{\frac{\lambda}{1-\lambda} e(R) + \frac{1}{1-\lambda} E^+(\lambda R)\}$$

Proof: Suppose a sequential joint source channel coding with feedback error exponent E_{s,sc_f} can be achieved. i.e. for all $\epsilon > 0$, there exists finite K_1, s.t. for all i, Δ:

$$\Pr(s_i \neq \hat{s}_i(i+\Delta)) \leq K_1 2^{-\Delta(E_{s,sc_f}-\epsilon)} \tag{15}$$

Thus we can give an upper bound on the block error at decoding time $n+\Delta$:

$$\Pr(s_1^n \neq \hat{s}_1^n(n+\Delta)) \leq \sum_{i=1}^{n} \Pr(s_i \neq \hat{s}_i(n+\Delta))$$
$$\leq \sum_{i=1}^{n} K_1 2^{-(n+\Delta-i)(E_{s,sc_f}-\epsilon)}$$
$$= K 2^{-\Delta(E_{s,sc_f}-\epsilon)} \tag{16}$$

where K is finite.

On the other hand, we also have a lower bound on the block error. For a message set of size 2^{mR} and allowing m channel uses, the channel coding error probability for the memoryless channel with causal feedback is lower bounded by $K2^{-m(E^+(R)-\epsilon_n)}$, where $\epsilon_n \to 0$. This gives the following lower bound on the block error probability for joint source-channel coding with feedback, where the length of the source block is n and allowing $n+\Delta$ channel uses.

$$\Pr(s_1^n \neq \hat{s}_1^n) \tag{17}$$
$$\geq_{(a)} \sum_{T_P^n \in T^n} \Pr_{Q_s}(T_P^n) 2^{-(n+\Delta)(E^+(\frac{nH(P)}{n+\Delta})+\epsilon_n^{(1)})}$$
$$\geq_{(b)} \sum_{T_P^n \in T^n} 2^{-n(D(P\|Q_s)+\epsilon_n^{(2)})}$$
$$2^{-(n+\Delta)(E^+(\frac{nH(P)}{n+\Delta})+\epsilon_n^{(1)})}$$
$$\geq 2^{-\min_P\{nD(P\|Q_s)+(n+\Delta)(E^+(\frac{nH(P)}{n+\Delta}))\}-n\epsilon_n}$$

where T_n is the set of types on $\mathcal{S}^n$ and $E^+(R) = 0$ for $R > C_W$, ϵ_n, $\epsilon_n^{(1)}$ and $\epsilon_n^{(2)}$ all converge to 0 as n goes to infinity. (a) is true because there are $2^{nH(P)}$ equally likely sequences in a type T_P^n, meanwhile $n+\Delta$ channel uses with feedback are available, thus we have the lower bound on the error probability by the Haroutunian bound. (b) is by Theorem 12.1.4 in [11].

Combining (16) and (17), we have:

$$E_{s,sc_f} \leq \epsilon\frac{\log_2 K}{\Delta} + \frac{\epsilon_n}{\Delta}$$
$$+\min_P\{\frac{n}{\Delta}D(P\|Q_s) + \frac{n+\Delta}{\Delta}E^+(\frac{nH(P)}{n+\Delta})\}$$

This is true for all Δ and n, write $\lambda = \frac{n}{n+\Delta}$ and let n goes to infinity:

$$E_{s,sc_f} \leq \min_{\lambda\in[0,1]P}\{\frac{\lambda}{1-\lambda}D(P\|Q_s)+\frac{1}{1-\lambda}E^+(\lambda H(P))\}$$

For any λ, the right hand side is minimized by P s.t. $H(P) \geq H(Q_s)$ because for those P, s.t. $H(P) < H(Q_s)$, we have

$$\frac{\lambda}{1-\lambda}D(P\|Q_s) + \frac{1}{1-\lambda}E^+(\lambda H(P))$$
$$\geq_{(a)} \frac{1}{1-\lambda}E^+(\lambda H(Q_s))$$
$$= \frac{\lambda}{1-\lambda}D(Q_s\|Q_s) + \frac{1}{1-\lambda}E^+(\lambda H(Q_s))$$

(a) is true because $E^+(\cdot)$ is monotonically decreasing.

$$E_{s,sc_f}$$
$$\leq \min_{\lambda\in[0,1],P:H(P)\geq H(Q_s)}\{\frac{\lambda D(P\|Q_s)+E^+(\lambda H(P))}{1-\lambda}\}$$
$$= \min_{\lambda\in[0,1],R\geq H(Q_s)}\{\min_{P:H(P)=R}\{\frac{\lambda D(P\|Q_s)+E^+(\lambda H(P))}{1-\lambda}\}\}$$
$$= \min_{\lambda\in[0,1],R\geq H(Q_s)}\{\frac{\lambda}{1-\lambda}e(R) + \frac{1}{1-\lambda}E^+(\lambda R)\}$$
$$= \min_{\lambda\in[0,1],R}\{\frac{\lambda}{1-\lambda}e(R) + \frac{1}{1-\lambda}E^+(\lambda R)\}$$

The last equality is because $e(R) = 0$, for $R < H(Q_s)$ and $E^+(\cdot)$ is monotonically decreasing.

■

B. E_{s,sc_f} for binary erasure channels (BEC)

We do not have a general scheme for joint source channel coding for arbitrary DMC's. But for binary erasure channels, we can apply an optimal universal source code[5] followed by an optimal "repeat until received" channel code[3] for a BEC. In [3], a "focusing" bound is derived for BEC.

Our optimal joint source-channel coding with feedback scheme for binary erasure channel is as follows. A block-length N is chosen that is much smaller than the target end-to-end delays[14], while still being large enough. For a discrete memoryless source and large block-lengths N, the best possible variable-length code is given in [2] and consists of two stages: first describing the type of the block $\vec{s}_i$ using $O(|\mathcal{S}|\log_2 N)$ bits and then describing which particular realization has occurred by using a variable $NH(\vec{s}_i)$ bits. The overhead $O(|\mathcal{S}|\log_2 N)$ is asymptotically negligible and the code is also universal in nature. It is easy to verify that $\lim_{N\to\infty}\frac{E_{Q_s}(l(\vec{s}))}{N} = H(Q_s)$ This code is obviously a prefix-free code. Write $l(\vec{s}_i)$ as the length of the codeword for $\vec{s}_i$, then:

$$l(\vec{s}_i) \leq |\mathcal{S}|\log_2(N+1) + NH(\vec{s}_i) \tag{18}$$

The binary sequences describing the source is fed to the optimal "repeat until received" channel coding system described in [3].

[14] We are interested in the performance with asymptotically large delays Δ

Theorem 5: For the iid source $\sim Q_s$ and a binary erasure channel with error rate δ with causal noiseless feedback, then using the code described above:

$$E_{s,sc_f} = \min_{\lambda\in[0,1],R}\{\frac{\lambda e(R) + E^+_{BEC}(\lambda R)}{1-\lambda}\} \triangleq E^*$$

where[15] $E^+_{BEC}(R) = D(1-R\|\delta)$ is the Haroutunian bound for BEC.

Before the proof, we have the following lemmas to bound the probabilities of atypical channel behavior and atypical source behavior respectively.

Lemma 8: (Channel atypicality) For a binary erasure channel with erasure rate δ, the probability of more than n erasures in t channel uses is upper bounded by $(t-n)2^{-tD(\frac{n}{t}\|\delta)}$ if $\frac{n}{t} > \delta$ and upper bounded by 1 if $\frac{n}{t} \leq \delta$.

Proof: The proof is trivial by applying Theorem 12.1.4 in [11]. Thus for all $\epsilon > 0$, there exists $K < \infty$, s.t. the above probability is upper bounded by:

$$K2^{-t(E^+_{BEC}(1-\frac{n}{t})-\epsilon)} \tag{19}$$

□

Lemma 9: (Source atypicality) for all $\epsilon > 0$, N large enough, there exists $K < \infty$, s.t. for all Δ, n :

$$\Pr(\sum_{i=1}^{n} l(\vec{\mathbf{s}}_i) > nNr) \leq K2^{-nN(e(r)-\epsilon)} \tag{20}$$

Proof: Only need to show the case for $r > H(Q_s)$. By the Cramér's theorem[14]:

$$\Pr(\sum_{i=1}^{n} l(\vec{\mathbf{s}}_i) > nNr) = \Pr(\frac{1}{n}\sum_{i=1}^{n} l(\vec{\mathbf{s}}_i) > Nr)$$
$$\leq (n+1)^{|\mathcal{S}|^N} 2^{-n\inf_{x\geq Nr} I(x)} \tag{21}$$

where the rate function $I(x)$ is [14]:

$$I(x) = \sup_{\rho\in\mathcal{R}}\{\rho x - \log_2(\sum_{\vec{s}\in\mathcal{S}^N} Q_s(\vec{s})2^{\rho l(\vec{s})})\}$$

Write $I(x,\rho) = \rho x - \log_2(\sum_{\vec{s}\in\mathcal{S}^N} Q_s(\vec{s})2^{\rho l(\vec{s})})$, $I(x,0) = 0$. $x \geq Nr > NH(Q_s)$, for large N:

$$\frac{\partial I(x,\rho)}{\partial \rho}|_{\rho=0} = x - \sum_{\vec{s}\in\mathcal{S}^N} Q_s(\vec{s})l(\vec{s}) \geq 0$$

By the Hölder inequality, for all ρ_1, ρ_2, and for all $\theta\in(0,1)$:

$$(\sum p_i 2^{\rho_1 l_i})^\theta(\sum p_i 2^{\rho_2 l_i})^{(1-\theta)}$$
$$\geq \sum (p_i^\theta 2^{\theta\rho_1 l_i})(p_i^{(1-\theta)}2^{(1-\theta)\rho_2 l_i})$$
$$= \sum p_i 2^{(\theta\rho_1+(1-\theta)\rho_2)l_i}$$

This shows that $\log_2(\sum_{\vec{s}\in\mathcal{S}^N} Q_s(\vec{s})2^{\rho l(\vec{s})})$ is a convex function on ρ, thus $I(x,\rho)$ is a concave function on ρ for fixed x. Then $\forall x > 0$, $\forall \rho < 0$, $I(x,\rho) < 0$. , which means that the ρ to maximize $I(x,\rho)$ is positive. This implies that $I(x)$ is monotonically increasing with x. Thus $\inf_{x\geq Nr} I(x) = I(Nr)$

For $\rho \geq 0$, using the upper bound on $l(\vec{x})$ in (18):

$$\log_2(\sum_{\vec{s}\in\mathcal{S}^N} Q_s(\vec{s})2^{\rho l(\vec{s})})$$
$$\leq \log_2(\sum_{T_P^N\in T^N} 2^{-ND(P\|Q_s)}2^{\rho(|\mathcal{S}|\log_2(N+1)+NH(P))})$$
$$\leq \log_2((N+1)^{|\mathcal{S}|}2^{-N\min_P\{D(P\|Q_s)-NH(P)\}+\rho|\mathcal{S}|\log_2(N+1)})$$
$$= N\big(-\min_P\{D(P\|Q_s) - \rho H(P)\} + \epsilon_N\big)$$

where $\epsilon_N = \frac{(1+\rho)|\mathcal{S}|\log_2(N+1)}{N}$ goes to 0 as N goes to infinity. Substitute the above equality to $I(Nr)$:

$$I(Nr) \geq N\big(\sup_{\rho>0}\{\min_P \rho(r - H(P)) + D(P\|Q_s)\} - \epsilon_N\big) \tag{22}$$

First fix ρ, by a simple Lagrange multiplier argument, with fixed $H(P)$, we know that the distribution to minimize $D(P\|Q_s)$ is a tilted distribution of Q_s^α. It can be verified that $\frac{\partial H(Q_s^\alpha)}{\partial\alpha} \geq 0$ and $\frac{\partial D(Q_s^\alpha\|Q_s)}{\partial\alpha} = \alpha\frac{\partial H(Q_s^\alpha)}{\partial\alpha}$. Thus the distribution to minimize $D(P\|Q_s) - \rho H(P)$ is Q_s^ρ. Using some algebra, we have

$$D(Q_s^\rho\|Q_s) - \rho H(Q_s^\rho) = -(1+\rho)\log_2\sum_{s\in\mathcal{S}} Q_s(s)^{\frac{1}{1+\rho}}$$

Substitute this into (22):

$$I(Nr) \geq N\big(\sup_{\rho>0}\rho r - (1+\rho)\log_2\sum_{s\in\mathcal{S}} Q_s(s)^{\frac{1}{1+\rho}} - \epsilon_N\big)$$
$$= N\big(e(r) - \epsilon_N\big) \tag{23}$$

The last equality can again be proved by a simple Lagrange multiplier argument. Substitute (23) into (21) and by letting N be big enough, we get the the desired bound in (20). □

Now we are ready to prove Theorem 5.

Proof: We give an upper bound on the decoding error on $\vec{\mathbf{s}}_t$ at time $(t+\Delta)N$. At time $(t+\Delta)N$, the decoder can *not* decode $\vec{\mathbf{s}}_t$ with 0 error probability iff the binary strings describing $\vec{\mathbf{s}}_t$ are *not* all out of the buffer. Since the encoding buffer is FIFO, this means that the number of non-erasures from some time $t_1 < tN$ to $(t+\Delta)N$ is less than the number of the bits in the buffer at time t_1 plus the number of bits coming into the encoder from time t_1 to time tN. Suppose the buffer is last empty at time $tN - nN$ where $0 \leq n \leq t$. Given this condition, the decoding error occurs only if $\sum_{i=0}^{n-1} l(\vec{\mathbf{s}}_{t-i}) > (n+\Delta)N - n_E(tN-nN, tN+\Delta N)$ where $n_E(t_1,t_2)$ is the number of erasures from time

[15] We write $D(a\|b)$ as the Kullback-Liebler divergence of two binary distributions $(a, 1-a)$ and $(b, 1-b)$.

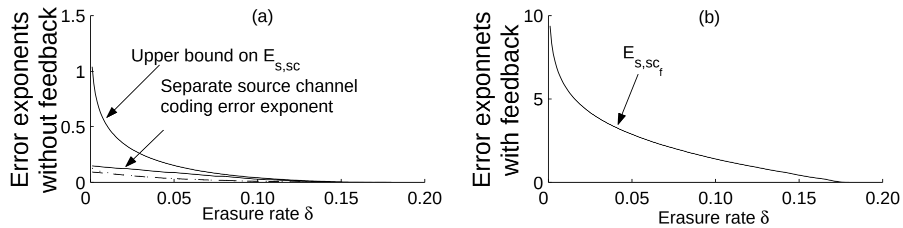

Fig. 6. (a)Upper/lower bounds on $E_{s,sc}$ as functions of erasure rate. Upper(dotted)/lower(dashed) bounds on block coding error exponent $E_{b,sc}$ (b) Joint source-channel coding error exponent with feedback E_{s,sc_f}

t_1 to time t_2. Then[16]

$$\begin{aligned}
&\Pr(\vec{s}_t \neq \vec{s}_t((t+\Delta)N) \\
&\leq \sum_{n=0}^{t} \Big[\sum_{m=0}^{nl_{max}} \Pr(\sum_{i=0}^{n-1} l(\vec{\mathsf{s}}_{t-i}) > m) \\
&\Pr((n+\Delta)N - n_E(tN - nN, tN+\Delta N) < m) \Big] \\
&\leq_{(a)} \sum_{n=0}^{t} \sum_{m=0}^{nl_{max}} K_1 2^{-nN(e(\frac{m}{nN})-\epsilon_1)} \\
&K_2 2^{-(n+\Delta)N(E^+_{BEC}(1-\frac{(n+\Delta)N-m}{(n+\Delta)N})-\epsilon_2)} \\
&\leq_{(b)} \sum_{n=0}^{t} K_3 2^{-N(\min_R\{ne(R)+(n+\Delta)E^+_{BEC}(\frac{nR}{n+\Delta})\}-\epsilon_3)} \\
&\leq_{(c)} \sum_{n=0}^{\gamma\Delta} K_3 2^{-\Delta N(\min_{R,\lambda\in[0,1]}\{\frac{\lambda e(R)+E^+_{BEC}(\lambda R)}{1-\lambda}\}-\epsilon_3)} \\
&+ \sum_{n=\gamma\Delta}^{\infty} K_3 2^{-nN(\min_R\{e(R)+E^+_{BEC}(R)\}-\epsilon_3)} \\
&\leq_{(d)} K 2^{-\Delta N(E^*-\epsilon)}
\end{aligned}$$

where, $K_i's$ and $\epsilon_i's$ are properly chosen real numbers. (a) is true because the source and memoryless channel are independent and Lemma 8 and 9. In (b), we write $R = \frac{m}{nN}$ and take the R to minimize the error exponents. Define $\gamma = \frac{E^*}{\min_R\{e(R)+E^+_{BEC}(R)\}}$. The first term of (c) comes from defining $\lambda = \frac{n}{n+\Delta}$, the second term is by noticing that $E^+(\cdot)$ is monotonically decreasing. (d) is by definitions of γ and E^*. ■

V. Example: Error exponents for BECs

For a Bernoulli (0.25) source followed by a binary erasure channel (δ) joint source-channel coding system, the entropy rate of the source is $H(\mathsf{s}) = 0.81$. We plot both the upper bound and the lower bound on the joint source channel coding error exponent with delay $E_{s,sc}$ and the bounds on block error exponents $E_{b,sc}$ in Figure 6(a). The feedback error exponent E_{s,sc_f} is plotted in Figure 6(b).

[16] l_{max} is the longest code length, $l_{max} \leq |\mathcal{S}|\log_2(N+1) + N|\mathcal{S}|$.

VI. Conclusions and Future Work

We studied sequential joint source-channel coding and defined the error exponents with delay. By applying a variation of the feed-forward channel coding analysis in [3], we derived an upper bound on error exponents with delay for lossless source-channel coding. A lower bound on the error exponent based on separate sequential source and channel coding is given. There is, in general, a gap between the lower and upper bounds. It is an open problem on how to implement joint source channel coding with delay constraint. For joint source channel coding with feedback, we only show the achievability result for erasure channels. General results for general DMCs are unknown. Moreover, no result on lossy source-channel coding with delay-constrained problem is known.

References

[1] R. G. Gallager, *Information Theory and Reliable Communication*. New York, NY: John Wiley, 1971.
[2] I. Csiszár and J. Körner, *Information Theory*. Budapest: Akadémiai Kiadó, 1986.
[3] A. Sahai, "Why block length and delay are not the same thing," *IEEE Transactions on Information Theory*, submitted. [Online]. Available: http://www.eecs.berkeley.edu/~sahai/Papers/FocusingBound.pdf
[4] M. Pinsker, "Bounds of the probability and of the number of correctable errors for nonblock codes," *Translation from Problemy Peredachi Informatsii*, vol. 3, pp. 44–55, 1967.
[5] C. Chang and A. Sahai, "The error exponent with delay for lossless source coding," *Information Theory Workshop*, 2006.
[6] S. Draper, C. Chang, and A. Sahai, "Sequential random binning for streaming distributed source coding," *ISIT*, 2005.
[7] I. Csiszar, "Joint source-channel error exponent," *Problems of Control and Information Theory*, vol. 9(5), pp. 315–328, 1980.
[8] M. V. Burnashev, "Data transmission over a discrete channel with feedback. Random transmission time," *Problems of Information Transmission*, vol. 12, no. 4, pp. 10–30, 1976.
[9] G. Forney, "Convolutional codes ii. maximum-likelihood decoding," *Inform. and Control*, vol. 25, pp. 222–266, 1974.
[10] E. A. Haroutunian, "Lower bound for error probability in channels with feedback," *Problemy Peredachi Informatsii*, vol. 13, no. 2, pp. 36–44, 1977.
[11] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: John Wiley and Sons Inc., 1991.
[12] C. Chang and A. Sahai, "Upper bound on error exponents with delay for lossless source coding with side-information," *IEEE International Symposium on Information Theory*, 2006.
[13] A. Sahai, *Anytime Information Theory, PhD thesis*. Massachusetts Institute Technology, 2001.
[14] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications*. Springer, 1998.