Upper Bound on Error Exponents with Delay for Lossless Source Coding with Side-Information

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*— The traditional view of source coding with side information is in the block coding context in which all the source symbols are known in advance by the encoder. We instead consider a sequential setting in which source symbols are revealed to the encoder in real time and need to be reconstructed at the decoder within a certain fixed delay. We derive an upper bound on the reliability function with delay that considers the errors induced by atypically strange side-information. It is shown to be tight for certain "symmetric" sources in low rate regime.**

I. Introduction

Ever since Shannon, block-coding has been the favored paradigm for studying codes. The block-length serves as a proxy for both implementation complexity and end-to-end delay in systems and the corresponding error exponents (or reliability functions) are used to study the tradeoffs involved. The story is particularly attractive 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].

Recently, it has become clear that fixed block-length need not be a good proxy for fixed-delay in all settings. 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] In addition, the nature of the dominant error events changes. Without feedback, errors are usually caused by *future*[2] channel atypicality. When feedback is present, it is a *combination of past and future* atypicality that forces errors.

[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]Past and future are considered relative to the time that the symbol in question enters the encoder.

For point-to-point fixed-rate lossless source-coding, [5] shows that the reliability function with fixed delay is much better than the reliability with fixed block-length and behaves analogous to channel coding with feedback. Furthermore, this can be asymptotically achieved by using a fixed-to-variable length code and smoothing the rate through a queue. An example given in [5], [6] illustrated how sometimes even 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. In addition, the errors are dominated by events involving the *past* atypicality of the source. The future does not matter.

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 result in this paper is an upper bound on the error exponents with delay for lossless source coding with side-information known only at the decoder. While the errors could in general be caused by atypical behavior of the joint source in both the past and the future, this paper's bound captures only the impact of the unknown and uncontrollable future — the potentially atypical behavior of future side-information symbols.

The problem turns out to be analogous to the case of channel coding without feedback, and the strategy used parallels the one used in [3] for that problem. We also quote the lower bound (achievability results) on the fixed-delay error exponents for which the proofs can be found in [6], [7]. The upper and lower bounds agree in the low-rate regime for sources with a uniform marginal and symmetric connections to the side-information.

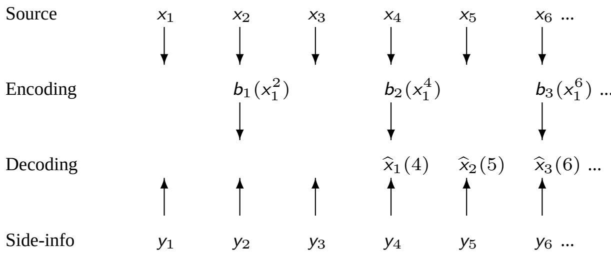

Fig. 1. Sequential source coding with side-information: rate $R = \frac{1}{2}$, delay $\Delta = 3$

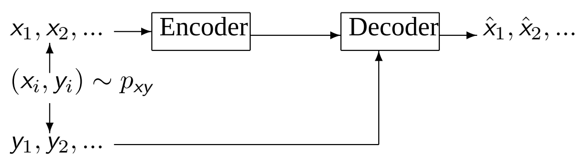

Fig. 2. Lossless source coding with side-information

A. Review of source coding with side information

As shown in Figure 2, the sources are iid random variables x_1^n, y_1^n [3] from a finite alphabet $\mathcal{X} \times \mathcal{Y}$. Without loss of generality, $p_x(x) > 0$, $\forall x \in \mathcal{X}$. x_1^n is the source known to the encoder and y_1^n is the side-information known only to the decoder. A rate R block source coding system for n source symbols consists of a encoder-decoder pair $(\mathcal{E}_n, \mathcal{D}_n)$. Where

$$\mathcal{E}_n : \mathcal{X}^n \rightarrow \{0,1\}^{\lfloor nR \rfloor}, \qquad \mathcal{E}_n(x_1^n) = b_1^{\lfloor nR \rfloor}$$
$$\mathcal{D}_n : \{0,1\}^{\lfloor nR \rfloor} \times \mathcal{Y}^n \rightarrow \mathcal{X}^n, \mathcal{D}_n(b_1^{\lfloor nR \rfloor}, y_1^n) = \hat{x}_1^n$$

The error probability is $\Pr(x_1^n \neq \hat{x}_1^n) = \Pr(x_1^n \neq \mathcal{D}_n(\mathcal{E}_n(x_1^n)))$. The exponent $Er_b(R)$ is achievable if $\exists$ a family of $\{(\mathcal{E}_n, \mathcal{D}_n)\}$, s.t.[4]

$$\lim_{n\to\infty} -\frac{1}{n} \log_2 \Pr(x_1^n \neq \hat{x}_1^n) = Er_b(R) \tag{1}$$

The relevant results of [2], [8] are summarized into the following theorem.

Theorem 1: $E_b^{(1)}(R) \leq Er_b(R) \leq E_b^{(2)}(R)$ where

$$E_b^{(1)}(R) = \min_{q_{xy}} \{D(q_{xy}\|p_{xy}) + \max\{0, R - H(q_{x|y})\}\}$$
$$E_b^{(2)}(R) = \min_{q_{xy}:H(q_{x|y})\geq R} \{D(q_{xy}\|p_{xy})\}$$

As shown in [2], the two bounds are the same in low rate regime.

[3]In this paper, x and y are random variables, x and y are realizations of the random variables.

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

B. Sequential Source Coding

Rather than being known in advance, the source symbols enter the encoder in a real-time fashion. We assume that the source $\mathcal{S}$ generates a pair of source symbols (x, y) per second from the finite alphabet $\mathcal{X} \times \mathcal{Y}$. The j'th source symbol x_j is not known at the encoder until time j and similarly for y_j at the decoder. Rate R operation means that the encoder sends 1 binary bit to the decoder every $\frac{1}{R}$ seconds. For obvious reasons, we focus on cases with $H_{x|y} < R < \log_2 |\mathcal{X}|$.

Definition 1: A sequential encoder-decoder pair $\mathcal{E}, \mathcal{D}$ are sequence of maps. $\{\mathcal{E}_j\}, j = 1, 2, \ldots$ and $\{\mathcal{D}_j\}, j = 1, 2, \ldots$. The outputs of $\mathcal{E}_j$ are the outputs of the encoder $\mathcal{E}$ from time $j-1$ to j.

$$\mathcal{E}_j : \mathcal{X}^j \longrightarrow \{0,1\}^{\lfloor jR \rfloor - \lfloor (j-1)R \rfloor}$$
$$\mathcal{E}_j(x_1^j) = b_{\lfloor (j-1)R \rfloor + 1}^{\lfloor jR \rfloor}$$

The outputs of $\mathcal{D}_j$ are the decoding decisions of all the arrived source symbols by time j based on the received binary bits up to time j as well as the side-information.

$$\mathcal{D}_j : \{0,1\}^{\lfloor jR \rfloor} \times \mathcal{Y}^j \longrightarrow \mathcal{X}$$
$$\mathcal{D}_j(b_1^{\lfloor jR \rfloor}, y_1^j) = \widehat{x}_{j-d}$$

Where $\widehat{x}_{j-d}$ is the estimation of x_{j-d} and thus has end-to-end delay of d seconds. A rate $R = \frac{1}{2}$ sequential source coding system is illustrated in Figure 1.

For sequential source coding, it is important to study the symbol by symbol decoding error probability instead of the block coding error probability.

Definition 2: A family of rate R sequential source codes $\{(\mathcal{E}^d, \mathcal{D}^d)\}$ are said to achieve delay-reliability $E_s(R)$ if and only if: $\forall i$

$$\lim_{d\to\infty} \frac{-1}{d} \log_2 P(x_i \neq \widehat{x}_i(i+d)) \leq E_s(R)$$

II. Lower bound on achievable error exponents with delay

We state the relevant lower bound (achievability) results without proof. More general results and the details of the proof can be found in [6], [7].

Theorem 2: Sequential random source coding theorem: Using a random sequential coding scheme and the ML decoding rule using side-information:

$$\Pr[\widehat{x}_{n+\Delta}(n) \neq x_n] \leq K 2^{-\Delta E_s^{(1)}(R)} \tag{2}$$

Where K is a constant, and $E_s^{(1)}(R) = E_b^{(1)}(R)$ as defined in Theorem 1.

The random sequential coding scheme can be realized using an infinite constraint-length time-varying random convolutional code and ML decoding at the receiver.

III. Upper bound on error exponents with fixed delay

Factor the joint probability to treat the source as a random variable x and consider the side-information y as the output of a discrete memoryless channel (DMC) $p_{y|x}$ with x as input. This model is shown in Figure 3.

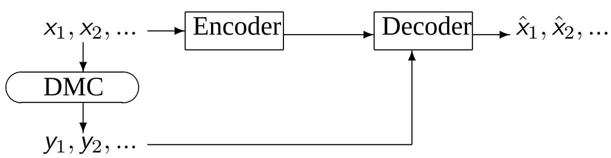

Fig. 3. Lossless source coding with side-information

Theorem 3: For the source coding with side-information problem in Figure 3, if the source is iid $\sim p_{xy}$ from a finite alphabet, then the error exponents $E_s(R)$ with fixed delay must satisfy $E_s(R) \leq E_s^{(2)}(R)$, where
$E_s^{(2)}(R) = \min\{\inf_{q_{xy}, \alpha \geq 1: H(q_{x|y}) > (1+\alpha)R}\{\frac{1}{\alpha} D(q_{xy} \| p_{xy})\}, \inf_{q_{xy}, 1 \geq \alpha \geq 0: H(q_{x|y}) > (1+\alpha)R}\{\frac{1-\alpha}{\alpha} D(q_x \| p_x) + D(q_{xy} \| p_{xy})\}\}$

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 source coding counterparts to Lemmas 4.1-4.5 in [3]. The idea of the proof is to first build a feed-forward sequential source decoder which has access to the previous source symbols in addition to the encoded bits and the side-information. The second step is to construct a block source-coding scheme from the optimal feed-forward sequential decoder and showing that if the side-information behaves 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 the atypicality of the side-information before that particular source symbol does not cause the error because of the feed-forward information. Thus, cause of the decoding error for that particular symbol is the atypical behavior of the future side-information only. 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 rate R decoder $\mathcal{D}^{\Delta,R}$ with feed-forward is a decoder $\mathcal{D}_j^{\Delta,R}$ that also has access to the past source symbols x_1^{j-1} in addition to the encoded bits $b_1^{\lfloor (j+\Delta)R \rfloor}$ and side-information $y_1^{j+\Delta}$.

Using this feed-forward decoder:

$$\widehat{x}_j(j+\Delta) = \mathcal{D}_j^{\Delta,R}(b_1^{\lfloor (j+\Delta)R \rfloor}, y_1^{j+\Delta}, x_1^{j-1}) \qquad (3)$$

Lemma 1: For any rate R encoder $\mathcal{E}$, the optimal delay Δ rate R decoder $\mathcal{D}^{\Delta,R}$ with feed-forward only needs to depend on $b_1^{\lfloor (j+\Delta)R \rfloor}, y_j^{j+\Delta}, x_1^{j-1}$

Proof: The source and side-information (x_i, y_i) is an iid random process and the encoded bits $b_1^{\lfloor (j+\Delta)R \rfloor}$ are functions of $x_1^{j+\Delta}$ so obeys the Markov chain: $y_1^{j-1} - (x_1^{j-1}, b_1^{\lfloor (j+\Delta)R \rfloor}, y_j^{j+\Delta}) - x_j^{j+\Delta}$. Conditioned on the past source symbols, the past side-information is completely irrelevant for estimation. □

Write the error sequence of the feed-forward decoder as[5] $\widetilde{x}_i = x_i - \widehat{x}_i$. Then we have the following property for the feed-forward decoders.

Lemma 2: Given a rate R encoder $\mathcal{E}$, the optimal delay Δ rate R decoder $\mathcal{D}^{\Delta,R}$ with feed-forward for symbol j only needs to depend on $b_1^{\lfloor (j+\Delta)R \rfloor}, y_1^{j+\Delta}, \widetilde{x}_1^{j-1}$

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 $(b_1^{\lfloor (j+\Delta)R \rfloor}, y_1^{j+\Delta}, \widetilde{x}_1^{j-1})$ giving $\widehat{x}_1^{k-1}$. This in turn allows the recovery of x_1^{k-1} since we also know $\widetilde{x}_1^{k-1}$. Thus the decoder is equivalent. □

We call the feed-forward decoders in Lemmas 1 and 2 type I and II delay Δ rate R feed-forward decoders respectively. Lemma 1 and 2 tell us that feed-forward decoders can be thought in three ways: having access to all encoded bits, all side information and all past source symbols, $(b_1^{\lfloor (j+\Delta)R \rfloor}, y_1^{j+\Delta}, x_1^{j-1})$, having access to all encoded bits, a recent window of side information and all past source symbols, $(b_1^{\lfloor (j+\Delta)R \rfloor}, y_j^{j+\Delta}, x_1^{j-1})$, or having access to all encoded bits, all side information and all past decoding errors, $(b_1^{\lfloor (j+\Delta)R \rfloor}, y_1^{j+\Delta}, \widetilde{x}_1^{j-1})$.

B. Constructing a block code

To encode a block of n source symbols, just run the rate R encoder $\mathcal{E}$ and terminate with the encoder run using some random source symbols drawn according to the distribution of p_x with matching side-information on the other side. To decode the block, just use the delay Δ rate R decoder $\mathcal{D}^{\Delta,R}$ with feed-forward, and then use the fedforward 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 4, the data processing inequality implies:

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

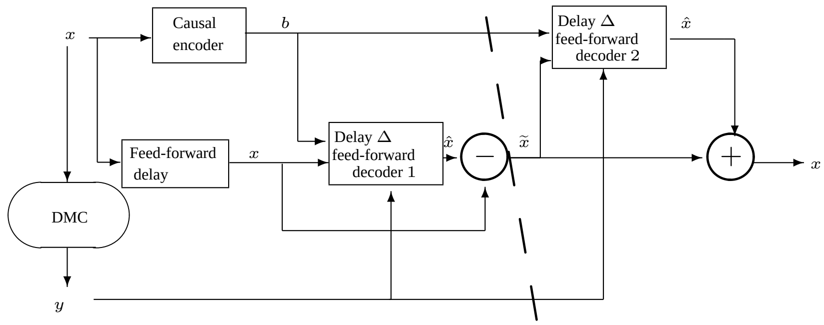

Fig. 4. A cutset illustration of the Markov Chain $x_1^n \;-\; (\widetilde{x}_1^n, b_1^{\lfloor (n+\Delta)R \rfloor}, y_1^{n+\Delta}) \;-\; x_1^n$. Decoder 1 and decoder 2 are type I and II delay Δ rate R feed-forward decoder respectively. They are equivalent.

Lemma 3: If n is the block-length, the block rate is $R(1+\frac{\Delta}{n})$, then

$$H(\widetilde{x}_1^n) \geq -(n+\Delta)R + nH(x|y) \tag{4}$$

Proof:

$$\begin{aligned} nH(x) &=_a H(x_1^n) = I(x_1^n; x_1^n) \\ &=_b I(x_1^n; \widetilde{x}_1^n, b_1^{\lfloor (n+\Delta)R \rfloor}, y_1^{n+\Delta}) \\ &=_c I(x_1^n; y_1^{n+\Delta}) + I(x_1^n; \widetilde{x}_1^n | y_1^{n+\Delta}) \\ &\quad + I(x_1^n; b_1^{\lfloor (n+\Delta)R \rfloor} | y_1^{n+\Delta}, \widetilde{x}_1^n) \\ &\leq nI(x, y) + H(\widetilde{x}_1^n) + H(b_1^{\lfloor (n+\Delta)R \rfloor}) \\ &\leq nH(x) - nH(x|y) + H(\widetilde{x}_1^n) \\ &\quad + (n+\Delta)R \end{aligned}$$

(a) is true because the source is i.i.d. (b) is true because of the data processing inequality considering the following Markov chain: $x_1^n \;-(\widetilde{x}_1^n, b_1^{\lfloor (n+\Delta)R \rfloor}, y_1^n) \;-\; x_1^n$ and the fact that $I(x_1^n; x_1^n) = H(x_1^n) \geq I(x_1^n; \widetilde{x}_1^n, b_1^{\lfloor (n+\Delta)R \rfloor}, y_1^{n+\Delta})$. (c) is the chain rule for mutual information. Other inequalities are obvious. □

C. Lower bound the symbol-wise error probability

Now suppose this block-code were to be run with the distribution q_{xy}, s.t. $H(q_{x|y}) > (1+\frac{\Delta}{n})R$, from time 1 to n, and were to be run with the distribution p_{xy} from time $n+1$ to $n+\Delta$. Write the hybrid distribution as Q_{xy}. 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 and side-information is coming from q_{xy}, then there exists a $\delta > 0$ so that for n large enough, the feed-forward decoder will make at least $\frac{H(q_{x|y}) - \frac{n+\Delta}{n}R}{2\log_2|\mathcal{X}| - (H(q_{x|y}) - \frac{n+\Delta}{n}R)} n$ symbol errors with probability δ or above. δ satisfies [6] $h_\delta + \delta \log_2(|\mathcal{X}|-1) = \frac{1}{2}(H(q_{x|y}) - \frac{n+\Delta}{n}R)$.

Proof: Lemma 3 implies:

$$\sum_{i=1}^{n} H(\widetilde{x}_i) \quad \geq \quad H(\widetilde{x}_1^n) \geq -(n+\Delta)R + nH(q_{x|y}) \tag{5}$$

The average entropy per source symbol for $\widetilde{x}$ is at least $H(q_{x|y}) - \frac{n+\Delta}{n}R$. Now suppose that $H(\widetilde{x}_i) \geq \frac{1}{2}(H(q_{x|y}) - \frac{n+\Delta}{n}R)$ for A positions. By noticing that $H(\widetilde{x}_i) \leq \log_2|\mathcal{X}|$, we have

$$\sum_{i=1}^{n} H(\widetilde{x}_i) \leq A\log_2|\mathcal{X}| + (n-A)\frac{1}{2}(H(q_{x|y}) - \frac{n+\Delta}{n}R)$$

With Eqn. 5, we derive the desired result:

$$A \geq \frac{(H(q_{x|y}) - \frac{n+\Delta}{n}R)}{2\log_2|\mathcal{X}| - (H(q_{x|y}) - \frac{n+\Delta}{n}R)} n \tag{6}$$

Where $2\log_2|\mathcal{X}| - (H(q_{x|y}) - \frac{n+\Delta}{n}R) \geq 2\log_2|\mathcal{X}| - H(q_{x|y}) \geq 2\log_2|\mathcal{X}| - \log_2|\mathcal{X}| > 0$

Now for A positions $1 \leq j_1 < j_2 < ... < j_A \leq n$ the individual entropy $H(\widetilde{x}_j) \geq \frac{1}{2}(H(q_{x|y}) - \frac{n+\Delta}{n}R)$. By the property of the binary entropy function[7], $P(\widetilde{x}_j \neq x_0) = P(x_j \neq \widehat{x}_j) \geq \delta$. □

We can pick $j^* = j_{\frac{A}{2}}$, by the previous lemma, we knot that $\min\{j^*, n-j^*\} \geq \frac{1}{2}\frac{(H(q_{x|y}) - \frac{n+\Delta}{n}R)}{2\log_2|\mathcal{X}| - (H(q_{x|y}) - \frac{n+\Delta}{n}R)} n$, so if we fix $\frac{\Delta}{n}$ and let n go to infinity, then $\min\{j^*, n-j^*\}$ goes to infinity as well.

At this point, Lemma 1 and 4 together imply that even if the source and side-information only behaves like it came from the hybrid distribution Q_{xy} from time j^* to $j^* + \Delta$ *and* the source behaves like it came from a distribution q_x from time 1 to $j^* - 1$, the same minimum error probability δ still holds. Now define the

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

[7] x_0 is the zero element in the finite group $Z_{|\mathcal{X}|}$.

"bad sequence" set E_{j^*} as the set of source and side-information sequence pairs so the type I delay Δ rate R decoder makes an decoding error at j^*. Formally[8]

$E_{j^*} = \{(\vec{x},\bar{\bar{y}})|x_{j^*} \neq \mathcal{D}^{\Delta,R}_{j^*}(\mathcal{E}(\vec{x}), y_j^{j^*+\Delta}, \bar{x})\}$. By Lemma 4, $Q_{xy}(E_{j^*}) \geq \delta$. Notice that E_{j^*} does not depend on the distribution of the source but only on the encoder-decoder pair. Define $J = \min\{n, j^*+\Delta\}$, and $\bar{\bar{x}} = x^J_{j^*}$, $\bar{\bar{y}} = y^J_{j^*}$.Now we write the strongly typical set[9] $A^\epsilon_J(q_{xy}) = \{(\vec{x},\bar{\bar{y}})|\forall(x,y), s.t. q_{xy}(x,y) > 0 : r_{\bar{\bar{x}},\bar{\bar{y}}}(x,y) \in (q_{xy}(x,y)-\epsilon, q_{xy}(x,y)+\epsilon), \forall(x,y), s.t. q_{xy}(x,y) = 0 : r_{\bar{\bar{x}},\bar{\bar{y}}}(x,y) = 0 \ \ and \ \ \forall x, r_{\bar{x}}(x) \in (q_x(x)-\epsilon, q_x(x)+\epsilon)\}$

Lemma 5: $Q_{xy}(E_{j^*} \cap A^\epsilon_J(q_{xy})) \geq \frac{\delta}{2}$ for large n and Δ.

Proof: Fix $\frac{\Delta}{n}$, let n go to infinity, then $\min\{j^*, n-j^*\}$ goes to infinity. By the definition of J, $\min\{j^*, J-j^*\}$ goes to infinity as while. By Lemma 13.6.1 in [9], we know that $\forall\epsilon > 0$, if $J-j^*$ and j^* are large enough, then $Q_{xy}(A^\epsilon_J(q_{xy})^C) \leq \frac{\delta}{2}$. By Lemma 4, $Q_{xy}(E_{j^*}) \geq \delta$. So

$$Q_{xy}(E_{j^*} \cap A^\epsilon_J(q_{xy})) \geq Q_{xy}(E_{j^*}) - Q_{xy}(A^\epsilon_J(q_{xy})^C) \geq \frac{\delta}{2}$$

□

Lemma 6: $\forall\epsilon < \min_{x,y:p_{xy}(x,y)>0}\{p_{xy}(x,y)\}$, $\forall(\vec{x},\bar{\bar{y}}) \in A^\epsilon_J(q_{xy})$,

$$\frac{p_{xy}(\vec{x},\bar{\bar{y}})}{Q_{xy}(\vec{x},\bar{\bar{y}})} \geq 2^{-(J-j^*+1)D(q_{xy}\|p_{xy})-(j^*-1)D(q_x\|p_x)-JG\epsilon}$$

where $G = \max\{|\mathcal{X}||\mathcal{Y}| + \sum_{x,y:p_{xy}(x,y)>0} \log_2(\frac{q_{xy}(x,y)}{p_{xy}(x,y)}+1), |\mathcal{X}| + \sum_x \log_2(\frac{q_x(x)}{p_x(x)}+1)\}$

Proof: For $(\vec{x},\bar{\bar{y}}) \in A^\epsilon_J(q_{xy})$, by definition of the strong typical set, it can be easily shown by algebra: $D(r_{\bar{\bar{x}},\bar{\bar{y}}}\|p_{xy}) \leq D(q_{xy}\|p_{xy}) + G\epsilon$ and $D(r_{\bar{x}}\|p_x) \leq D(q_x\|p_x) + G\epsilon$.

By Eqn. 12.60 in [9], we have:

$$\begin{aligned}
&\frac{p_{xy}(\vec{x},\bar{\bar{y}})}{Q_{xy}(\vec{x},\bar{\bar{y}})} = \frac{p_{xy}(\bar{x})}{q_{xy}(\bar{x})}\frac{p_{xy}(\bar{\bar{x}},\bar{\bar{x}})}{q_{xy}(\bar{\bar{x}},\bar{\bar{y}})}\frac{p_{xy}(x^{j^*+\Delta}_{J+1}, y^{j^*+\Delta}_{J+1})}{p_{xy}(x^{j^*+\Delta}_{J+1}, y^{j^*+\Delta}_{J+1})}\\
&= \frac{2^{-(J-j^*+1)(D(r_{\bar{\bar{x}},\bar{\bar{y}}}\|p_{xy})+H(r_{\bar{\bar{x}},\bar{\bar{y}}}))}}{2^{-(J-j^*+1)(D(r_{\bar{\bar{x}},\bar{\bar{y}}}\|q_{xy})+H(r_{\bar{\bar{x}},\bar{\bar{y}}}))}}\\
&\frac{2^{-(j^*-1)(D(r_{\bar{x}}\|p_x)+H(r_{\bar{x}}))}}{2^{-(j^*-1)(D(r_{\bar{x}}\|q_x)+H(r_{\bar{x}}))}}\\
&\geq 2^{-(J-j^*+1)(D(q_{xy}\|p_{xy})+G\epsilon)-(j^*-1)(D(q_x\|p_x)+G\epsilon)}\\
&= 2^{-(J-j^*+1)D(q_{xy}\|p_{xy})-(j^*-1)D(q_x\|p_x)-JG\epsilon} \square
\end{aligned}$$

Lemma 7: $\forall\epsilon < \min_{x,y}\{p_{xy}(x,y)\}$, and large Δ, n:

$$p_{xy}(E_{j^*}) \geq \frac{\delta}{2}2^{-(J-j^*+1)D(q_{xy}\|p_{xy})-(j^*-1)D(q_x\|p_x)-JG\epsilon}$$

,

Proof: Combining Lemma 5 and 6.

$$\begin{aligned}
&p_{xy}(E_{j^*}) \geq p_{xy}(E_{j^*} \cap A^\epsilon_J(q_{xy}))\\
&\geq q_{xy}(E_{j^*} \cap A^\epsilon_J(q_{xy}))2^{-(J-j^*+1)D(q_{xy}\|p_{xy})-(j^*-1)D(q_x\|p_x)-JG\epsilon}\\
&\geq \frac{\delta}{2}2^{-(J-j^*+1)D(q_{xy}\|p_{xy})-(j^*-1)D(q_x\|p_x)-JG\epsilon} \square
\end{aligned}$$

Now we are finally ready to prove Theorem 3. Notice that as long as $H(q_{x|y}) > \frac{n+\Delta}{n}R$, we know $\delta > 0$ by letting ϵ go to 0, Δ and n go to infinity proportionally. We have: $\Pr[\widehat{x}_{j^*}(j^*+\Delta) \neq x_{j^*}] = p_{xy}(E_{j^*}) \geq K2^{-(J-j^*+1)D(q_{xy}\|p_{xy})-(j^*-1)D(q_x\|p_x)}$.

Notice that $D(q_{xy}\|p_{xy}) \geq D(q_x\|p_x)$ and $J = \min\{n, j^*+\Delta\}$, then for all possible $j^* \in [1,n]$, we have: for $n \geq \Delta$

$$\begin{aligned}
&(J-j^*+1)D(q_{xy}\|p_{xy}) + (j^*-1)D(q_x\|p_x)\\
&\leq (\Delta+1)D(q_{xy}\|p_{xy}) + (n-\Delta-1)D(q_x\|p_x)\\
&\approx \Delta(D(q_{xy}\|p_{xy}) + \frac{n-\Delta}{\Delta}D(q_x\|p_x))
\end{aligned}$$

For $n < \Delta$

$$\begin{aligned}
&(J-j^*+1)D(q_{xy}\|p_{xy}) + (j^*-1)D(q_x\|p_x)\\
&\leq nD(q_{xy}\|p_{xy})\\
&= \Delta(\frac{n}{\Delta}D(q_{xy}\|p_{xy}))
\end{aligned}$$

Write $\alpha = \frac{\Delta}{n}$, then the upper bound on the error exponent is the minimum of the above error exponents over all $\alpha > 0$, i.e:

$$\begin{aligned}
E_s^{(2)}(R) = \min\{&\inf_{q_{xy},\alpha\geq 1:H(q_{x|y})>(1+\alpha)R}\{\frac{1}{\alpha}D(q_{xy}\|p_{xy})\},\\
&\inf_{q_{xy},1\geq\alpha\geq 0:H(q_{x|y})>(1+\alpha)R}\{\frac{1-\alpha}{\alpha}D(q_x\|p_x) + D(q_{xy}\|p_{xy})\}\}
\end{aligned}$$

[8] To simplify the notation, write: $\vec{x} = x_1^{j^*+\Delta}$, $\bar{x} = x_1^{j^*-1}$, $\bar{\bar{x}} = x_{j^*}^{j^*+\Delta}$, $\bar{\bar{y}} = y_{j^*}^{j^*+\Delta}$

[9] Write the empirical distribution of $(\bar{\bar{x}},\bar{\bar{y}})$ as $r_{\bar{\bar{x}},\bar{\bar{y}}}(x,y) = \frac{n_{x,y}(\bar{\bar{x}},\bar{\bar{y}})}{\Delta+1}$. Write the empirical distribution of $\bar{x}$ as $r_{\bar{x}}(x) = \frac{n_x(\bar{x})}{j^*-1}$.

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 Trans. Inform. 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 to appear.

[6] S. Draper, C. Chang, and A. Sahai, "Sequential random binning for streaming distributed source coding," *ISIT*, 2005.

[7] C. Chang, S. Draper, and A. Sahai, "Sequential random binning for streaming distributed source coding," *In preparation*, 2006.

[8] R. Gallager, "Source coding with side information and universal coding," *LIDS-P-937*, 1979.
[9] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: John Wiley and Sons Inc., 1991.