# Trade-off of lossless source coding error exponents

Cheng Chang and Anant Sahai

***Abstract*— We consider the lossless encoding of two simultaneous sources. The encoder may choose to discriminate against one source and hence the error exponents for the two sources can be different. The goal of this paper is to understand the region of achievable error-exponent pairs for lossless source coding. In the fixed-block-length case, the error exponent region is completely characterized and is found to be relatively trivial. However, in the streaming context, it is shown that there exists a non-trivial trade-off between the two error exponents. Both an inner bound and an outer bound are given for that case, but they do not match. The outer bound comes from a multi-stream version of the uncertainty-focusing bound.**

## I. Introduction and Problem Setup

Classical error exponents show the tradeoff between the amount of information communicated and the reliability of that communication [7]. In a multiuser setting, a new tradeoff is possible since different users can have different *error exponents* while sharing the same underlying communication *resources*. The vectors of achievable error exponents are known as the error exponent region. The error exponent region is studied for Gaussian broadcast and multiple-access channels in [11] where outer and inner bounds are derived.

In this paper, we simplify the problem further by considering only the case when the two users can jointly encode and jointly decode. In the context of streaming[1] messages, the feedback-channel-coding version of even this simplified problem is very important since [9] established an intimate link between multistream anytime-reliability regions and the stabilizability over noisy channels of unstable linear systems with vector-valued states. However, as was seen in [8], [2], the case of source coding error exponents is often simpler than that of channel coding. Therefore, the ideas are developed here in the source coding context.

In Section I-A and Section I-B, we review the point to point error exponent results for both the fixed-block-length and streaming contexts. Then, in Section I-C, we formally define the error-exponent region for the two sources one encoder problem. The main results are then stated in Section II with a numeric example in Section III. Abbreviated proofs follow in Section IV.

### A. Point-to-Point fixed length lossless source coding

Consider a discrete memoryless iid source with distribution $p_{\mathsf{x}}$ defined on finite alphabet $\mathcal{X}$. A rate-$R$ fixed-block-length source coding system for $n$ source symbols consists of an encoder-decoder pair $(\mathcal{E}_n, \mathcal{D}_n)$, where[2]

$$\mathcal{E}_n : \mathcal{X}^n \longrightarrow \{0,1\}^{nR}, \qquad \mathcal{E}_n(x_1^n) = b_1^{nR}$$
$$\mathcal{D}_n : \{0,1\}^{nR} \longrightarrow \mathcal{X}^n, \qquad \mathcal{D}_n(b_1^{nR}) = \widehat{x}_1^n$$

The probability of block decoding error is defined as

$$\Pr[x_1^n \neq \widehat{x}_1^n] = \Pr[x_1^n \neq \mathcal{D}_n(\mathcal{E}_n(x_1^n))].$$

In his seminal paper [10], Shannon proved that arbitrarily small error probabilities are achievable by letting $n$ get big as long as the encoder rate is larger than the entropy of the source, $R > H(p_{\mathsf{x}})$. Furthermore, it turns out that the error probability goes to zero exponentially in $n$.

*Theorem 1:* (From [5]) For a discrete memoryless source $\mathsf{x} \sim p_{\mathsf{x}}$ and encoder rate $R < \log|\mathcal{X}|$,

$\forall \epsilon > 0$, $\exists K_\epsilon < \infty$, s.t. $\forall n \geq 0$, $\exists$ a block encoder-decoder pair $\mathcal{E}_n, \mathcal{D}_n$ such that

$$\Pr[x_1^n \neq \widehat{x}_1^n] \leq K_\epsilon 2^{-n(E_b^{\mathsf{x}}(R)-\epsilon)}. \tag{1}$$

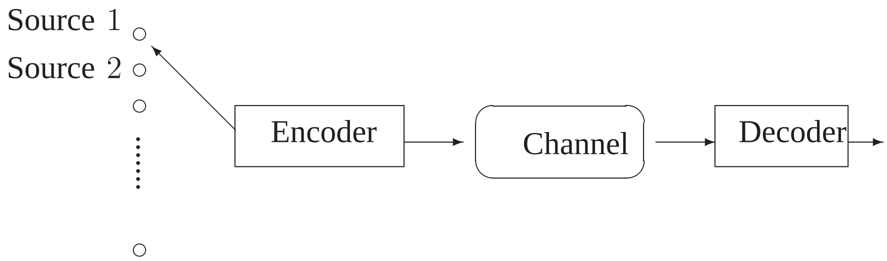
This result is asymptotically tight, in the sense that $\forall \epsilon > 0$, $\exists G_\epsilon > 0$, s.t. $\forall n \geq 0$, for all block encoder-decoder pairs $\mathcal{E}_n, \mathcal{D}_n$

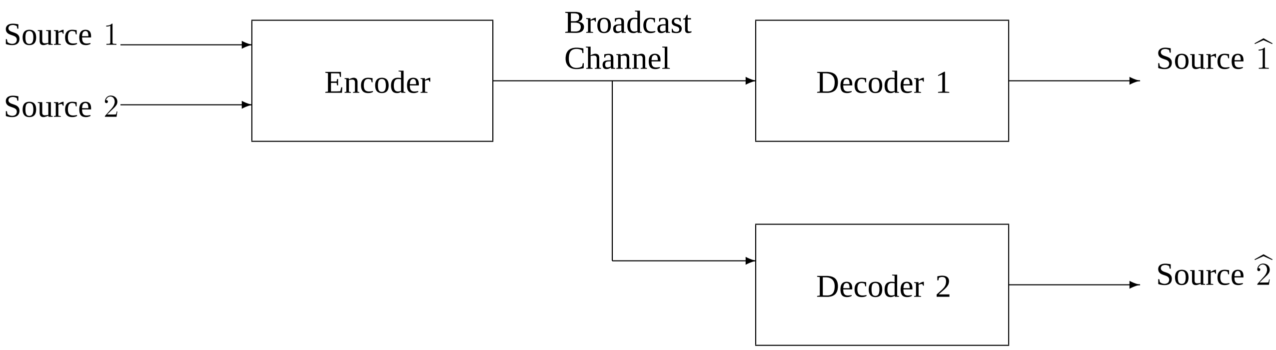
$$\Pr[x_1^n \neq \widehat{x}_1^n] \geq G_\epsilon 2^{-n(E_b^{\mathsf{x}}(R)+\epsilon)} \tag{2}$$

C. Chang is with HP Labs, Palo Alto. This work was done when the first author was doing his Phd study at UC Berkeley. A. Sahai is with the Wireless Foundations Center, Department of Electrical Engineering and Computer Science, University of California at Berkeley. Email: cchang@eecs.berkeley.edu, sahai@eecs.berkeley.edu

[1]The streaming context is where the message is revealed to the encoder gradually in real-time *and* is composed of finely-grained information that is incrementally useful in small chunks. The streaming context is distinguished from the hard real-time (or zero-delay) context in that the streaming context tolerates a substantial end-to-end delay between when the chunk of information enters the encoder and when it is needed at the destination. The streaming context is distinguished from the usual block-coding context in that the streaming end-to-end delay is considered to be much larger than the granularity of the information itself. The usual block-coding context can be considered one in which the tolerable end-to-end delay is equal in size to one information chunk. See [8] for a detailed discussion about the difference between the streaming and block contexts.

[2]We assume that $nR$ is an integer. It should be clear that this assumption is insignificant in the asymptotic regime where $n$ is big.




where $E_b^x(R)$ is defined as the block source coding error exponent with the form:

$$E_b^x(R) = \min_{q:H(q)\geq R} D(q\|p_x). \tag{3}$$

The error exponent $E_b^x(R)$ is monotonically increasing and convex [6] for $R \in [H(p_x), \log(|\mathcal{X}|)]$ as illustrated in Figure 1.

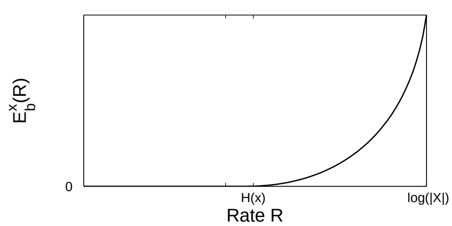



Fig. 1. Block coding error exponent $E_b^x(R)$

*B. Point-to-Point delay constrained streaming source coding*

In [1] and [8], delay constrained streaming source coding and streaming channel coding are studied. The delay-universal source coding model is illustrated in Figure 2. Rather than being known in advance, the source symbols enter the encoder in a streaming fashion. We assume that the discrete memoryless source generates one source symbol $x_i$ per second from finite alphabet $\mathcal{X}$. Where $x_i$'s are i.i.d from a distribution $p_x$. The $i^{th}$ source symbol $x_i$ is not known at the encoder until time $i$. This is the fundamental difference in the system model from the block source coding setup in Section I-A.

*Definition 1:* A delay-universal sequential encoder-decoder pair $\mathcal{E}, \mathcal{D}$ is a 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 $j$,

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

The outputs of the delay-universal decoder $\mathcal{D}_j$ are the decoding decisions of all the arrived source symbols at the encoder by time $j$ based on the received binary bits up to time $j$

$$\mathcal{D}_j : \{0,1\}^{\lfloor jR\rfloor} \rightarrow \mathcal{X}^j,\ \mathcal{D}_j(b_1^{\lfloor jR\rfloor}) = \widehat{x}_1^j(j)$$

where $\widehat{x}_1^j(j)$ is the estimation, at time $j$, of $x_1^j$ and thus the end-to-end delay of symbol $x_i$ at time $j$ is $j-i$ seconds for $i \leq j$. In a delay-universal scheme, the decoder emits revised estimates for all source symbols so far. The coding system is illustrated in Figure 2.

*Definition 2:* A delay-constrained error exponent $E_s^x(R)$ is said to be achievable if and only if $E_s^x(R)$ is the largest real number s.t. for all $\epsilon > 0$, there exists $K_\epsilon < \infty$, $\exists$ delay-universal encoder/decoder pairs $\mathcal{E}$, $\mathcal{D}$, s.t. $\forall 0 < i < j < \infty$:

$$\Pr[x_i \neq \widehat{x}_i(j)] \leq K_\epsilon 2^{-(j-i)(E_s^x(R)-\epsilon)}$$

This error exponent is derived in [1] and [3], [2].

*Theorem 2:* (From [1]) delay-constrained error exponent for streaming source coding

$$E_s^x(R) = \inf_{\alpha>0} \frac{1}{\alpha} E_b^x((\alpha+1)R) \tag{4}$$

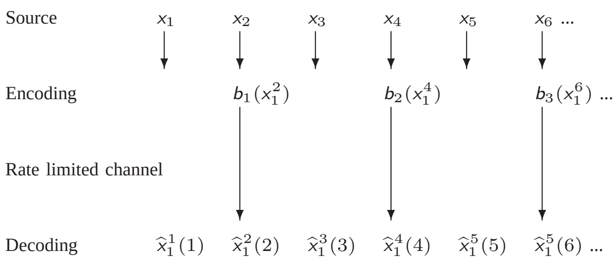



Fig. 2. Time line of delay-universal source coding: rate $R = \frac{1}{2}$

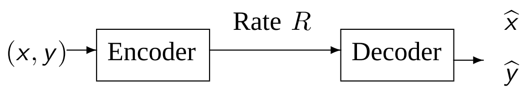



Fig. 3. Two sources one encoder source coding

where $E_b^x(R)$ is the block-coding error exponent defined in Theorem 1.

### C. Two sources and one encoder problem and its error exponent region

Instead of having one source, we study the problem illustrated in Figure 3. There are two sources, both of which enter the single encoder and must be reconstructed by the single decoder. If the decoding error is defined as $\Pr[(x, y) \neq (\widehat{x}, \widehat{y})]$, this problem was discussed in the previous two sections. However, in this paper, we study the *error vector*— $(\Pr[x \neq \widehat{x}], \Pr[y \neq \widehat{y}])$ and the asymptotic behaviors as the block-length or delay gets long — error-exponent vectors for both the block-coding and streaming contexts.

*Definition 3:* We denote by $(E(R, x), E(R, y))$ an achievable error-exponent pair for rate $R$. All the achievable pairs form an error-exponent region that is a subset of the first quadrant: $\{(X, Y) \in \mathcal{R}^+ \times \mathcal{R}^+ :$ $X$ and $Y$ are achievable error exponents for source $x$ and $y$ respectively $\}$

In this setup, the two sources share the total rate of $R$. The goal of this paper is to characterize the error-exponent region for both block and streaming source coding.

We say an error exponent pair $(E_1, E_2)$ **dominates** another pair $(F_1, F_2)$ iff $E_1 \geq F_1$ and $E_2 \geq F_2$, this gives a partial order on $\mathcal{R} \times \mathcal{R}$. Obviously, we only need to determine those exponent pairs that are not dominated by *any* other exponent pairs.

## II. Main Results

### A. Fixed-block-length coding

*Theorem 3:* Consider fixed-block-length source coding of iid data $(x_i, y_i) \sim p_{xy}$, and the error exponent region in Definition 3, then the error exponent region is an "L" shaped region:

$$E(R, x) \leq E_b^x(R) \quad \text{and} \quad E(R, y) \leq E_b^y(R)$$
$$E(R, x) \leq E_b^{xy}(R) \quad \text{or} \quad E(R, y) \leq E_b^{xy}(R).$$

### B. Delay-universal streaming coding

We summarize the outer bound result in Theorem 4 and the inner bound result in Theorem 5.

*Theorem 4:* (Outer bound) Consider the delay constrained source coding of iid sources $(x_i, y_i) \sim p_{xy}$, and the error exponent region in Definition 3, then the error exponent region is a **subset** of:

$$\{(X, Y) : X \leq E_s^x(R) \text{ and } Y \leq E_s^y(R)\} \bigcap$$
$$\{\bigcap_{\alpha \in [0,1], \beta > 0} \mathcal{A}_R^x(\alpha, \beta)\} \bigcap \{\bigcap_{\alpha \in [0,1], \beta > 0} \mathcal{A}_R^y(\alpha, \beta)\}$$

where $\mathcal{A}_R^x(\alpha, \beta)$ and $\mathcal{A}_R^y(\alpha, \beta)$ are "L" shaped regions, $\mathcal{A}_R^x(\alpha, \beta)$ is

$$\{(X, Y) : X \leq \frac{1}{\beta} F_R^x(\alpha, \beta) \text{ or } Y \leq \frac{1}{1 + \beta - \alpha} F_R^x(\alpha, \beta)\}$$

where $F_R^x(\alpha, \beta) =$

$$\min_{\theta \in [0,1]} \alpha E_b^{xy}(\frac{\theta(1 + \beta)R}{\alpha}) + (1 - \alpha) E_b^x(\frac{(1 - \theta)(1 + \beta)R}{1 - \alpha})$$

and similarly $\mathcal{A}_R^y(\alpha, \beta)$ is

$$\{(X, Y) : X \leq \frac{1}{1 + \beta - \alpha} F_R^y(\alpha, \beta) \text{ or } Y \leq \frac{1}{\beta} F_R^y(\alpha, \beta)\}$$

where $F_R^y(\alpha, \beta) =$

$$\min_{\theta \in [0,1]} \alpha E_b^{xy}(\frac{\theta(1 + \beta)R}{\alpha}) + (1 - \alpha) E_b^y(\frac{(1 - \theta)(1 + \beta)R}{1 - \alpha}).$$

*Theorem 5:* (inner bound) Consider the delay constrained source coding of iid sources $(x_i, y_i) \sim p_{xy}$, and the error exponent region in Definition 3, then the true error exponent region is a **superset** of:

$$\{\bigcup_{\alpha \in [0,1]} \mathcal{B}_R^x(\alpha)\} \bigcup \{\bigcup_{\alpha \in [0,1]} \mathcal{B}_R^y(\alpha)\}$$

where $\mathcal{B}_R^x(\alpha)$ and $\mathcal{B}_R^y(\alpha)$ are rectangular regions $\mathcal{B}_R^x(\alpha) =$

$$\bigcap_{\beta > 0, \theta \in [0,1]} \{(X, Y) :$$
$$X \leq \frac{\alpha}{\beta} E_b^{xy}(\frac{\theta(1+\beta)R}{\alpha}) + \frac{1-\alpha}{\beta} E_b^x(\frac{(1-\theta)(1+\beta)R}{1-\alpha}) \text{ and}$$
$$Y \leq \frac{1}{\beta} E_b^{xy}(\theta(1+\beta)R) + \frac{\min\{\beta, \frac{1}{\alpha} - 1\}}{\beta} E_b^x(\frac{(1-\theta)(1+\beta)R}{\min\{\beta, \frac{1}{\alpha} - 1\}})\}.$$

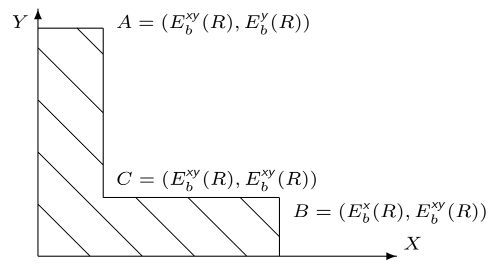



Fig. 4. Error exponent region for block source coding, the three corner points are marked as $A$,$B$ and $C$, where $A$ and $B$ are the dominant operation points.

Similarly for $\mathcal{B}_R^x(\alpha)$.

These two bounds are illustrated in Figure 5.

## III. Numerical Results

The source consists of two independent random variables $x$ and $y$ with marginals $p_x = p_y = (0.02, 0.98)$.

At rate $R = 0.5$, the fixed-block-length error exponents are $E_b^x(R) = E_b^y(R) = 0.1426$ and $E_b^{xy}(R) = 0.0253$. The error-exponent region is the shaded region in Figure 4. $A$ and $B$ are the only two operating points that is not dominated by other achievable error exponent pairs.

At rate $R = 0.5$, the delay-constrained error exponents are $E_s^x(R) = E_s^y(R) = 0.9019$ and $E_b^{xy}(R) = 0.2255$. The inner and outer bounds are plotted in Figure 5.

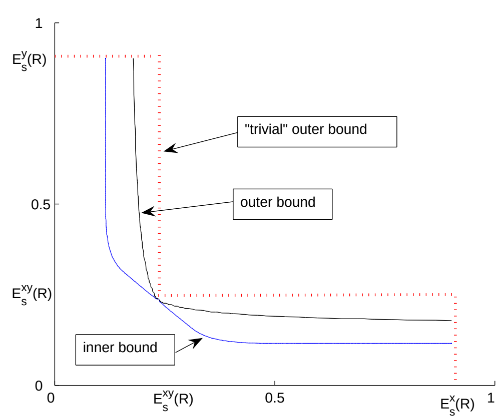



Fig. 5. Delay constrained streaming coding error exponent region

## IV. Proofs

In this section, we give the sketch of the proofs of Theorems 3, 4 and 5.

### A. Proof of Theorem 3

**Converse:** the proof of converse is trivial. By Theorem 1, we know that

$$E(R, x) \leq E_b^x(R) \text{ and } E(R, y) \leq E_b^y(R) \quad (5)$$

and for any $\epsilon > 0$, there exists a $G_\epsilon > 0$, s.t. the following is true for all $n$:

$$\Pr[(x_1^n, y_1^n) \neq (\widehat{x}_1^n, \widehat{y}_1^n)] \geq G_\epsilon 2^{-n(E_b^{xy}(R)+\epsilon)}. \quad (6)$$

Then by noticing that either $\Pr[x_1^n \neq \widehat{x}_1^n]$ or $\Pr[y_1^n \neq \widehat{y}_1^n]$ has to be at least half of $\Pr[(x_1^n, y_1^n) \neq (\widehat{x}_1^n, \widehat{y}_1^n)]$:

$$\Pr[x_1^n \neq \widehat{x}_1^n] \geq \frac{G_\epsilon}{2} 2^{-n(E_b^{xy}(R)+\epsilon)}$$

$$\text{or} \qquad \Pr[y_1^n \neq \widehat{y}_1^n] \geq \frac{G_\epsilon}{2} 2^{-n(E_b^{xy}(R)+\epsilon)}.$$

Taking logarithm at both sides, notice that this is true for all $\epsilon$ and $n$, by letting $\epsilon \to 0$ and $n \to \infty$, we have

$$E(R, x) \leq E_b^{xy}(R) \text{ or } E(R, x) \leq E_b^{xy}(R) \quad (7)$$

Combining (5) and (7), we prove the converse.

**Achievability:** By symmetry and the partial order established by dominance, we only need to show that the following error-exponent pair is achievable: $(E_b^x(R), E_b^{xy}(R))$. By Theorem 1, for all $\epsilon > 0$, there exists $K_\epsilon < \infty$, s.t. for all $n$ there exist a source coding system $(\mathcal{E}_n^x, \mathcal{D}_n^x)$ s.t.

$$\Pr[x_1^n \neq \mathcal{D}_n^x(\mathcal{E}_n^x(x_1^n))] \leq K_\epsilon 2^{-n(E_b^x(R)-\epsilon)}$$

**and** there exist a source coding system $(\mathcal{E}_n^{xy}, \mathcal{D}_n^{xy})$ s.t.

$$\Pr[(x_1^n, y_1^n) \neq \mathcal{D}_n^{xy}(\mathcal{E}_n^{xy}(x_1^n, y_1^n))] \leq K_\epsilon 2^{-n(E_b^{xy}(R)-\epsilon)}.$$

The new "biased" coding system $(\mathcal{E}_n^{x>y}, \mathcal{D}_n^{x>y})$ is as follows: for a source sequence pair $(x_1^n, y_1^n)$, if $(x_1^n, y_1^n) = \mathcal{D}_n^{xy}(\mathcal{E}_n^{xy}(x_1^n, y_1^n))$

$$\mathcal{E}_n^{x>y}(x_1^n, y_1^n) = \langle 0, \mathcal{E}_n^{xy}(x_1^n, y_1^n) \rangle$$

otherwise $\mathcal{E}_n^{x>y}(x_1^n, y_1^n) = \langle 1, \mathcal{E}_n^x(x_1^n) \rangle$, where $\langle \vec{a}_1, \vec{a}_2 \rangle$ concatenate two binary strings $\vec{a}_1$ and $\vec{a}_2$. We denote by $\vec{b}$ the output of the encoder. The length of $\vec{b}$ is $nR + 1$, to simply the notations we denote by $\vec{b}_{-1}$ the string of $\vec{b}$ with the first bit removed.

At the decoder side, if the first bit of the string $\vec{b}$ is 0,

$$\mathcal{D}_n^{x>y}(\vec{b}) = D_n^{xy}(\vec{b}_{-1}) = D_n^{xy}(E_n^{xy}(x_1^n, y_1^n)),$$

if the first bit is 1,

$$\mathcal{D}_n^{x>y}(\vec{b}) = (D_n^x(\vec{b}_{-1}), 0_1^n) = (D_n^x(E_n^x(x_1^n)), 0_1^n).$$

Obviously the new coding system $(\mathcal{E}_n^{x>y}, \mathcal{D}_n^{x>y})$ makes an decoding error on $x$ only if $(\mathcal{E}_n^x, \mathcal{D}_n^x)$ makes an decoding error on $x$, and $(\mathcal{E}_n^{x>y}, \mathcal{D}_n^{x>y})$ makes an decoding error on $y$ only if $(\mathcal{E}_n^{xy}, \mathcal{D}_n^{xy})$ makes an decoding error on $(x, y)$. The new coding system uses one extra bit that is insignificant asymptotically. This gives the desired result that the error-exponent pair $(E_b^x(R), E_b^{xy}(R))$ is achievable.

### B. Proof of Theorem 4

The proof is the multistream generalization of the proof for the single source streaming source coding case in [1], [3], [2]. The idea is to figure out the *dominant* error event for a particular delay. By using the method of types [4], we can give exponent of the dominant error event in the block coding context. Then, we translate these block coding errors to symbol errors and thus derive a bound on the delay-constrained error exponent for streaming source coding.

Now we can only give the sketch of the proof due to the space limitations. As shown in Figure 6, if the *total empirical* entropy of the random sequence[3] of $(x_1^n, y_1^{\alpha n})$ is higher than $(1+\beta)nR$, with high probability the coding system makes a block decoding error at time $(1+\beta)n$ for $(x_1^n, y_1^{\alpha n})$. We know that the true distribution of the source is $p_{xy}$, so the probability that the source behaves atypically such that the empirical entropy is higher than $(1+\beta)nR$ is as follows:

$$\begin{aligned} & \Pr[(x_1^n, y_1^{\alpha n}) \neq (\widehat{x}_1^n(1+\beta)n, \widehat{y}_1^{\alpha n}(1+\beta)n)] \\ \geq \quad & 2^{-n(\alpha D(q_{xy}\|p_{xy}) + (1-\alpha)D(r_x\|p_x) - \epsilon)}, \qquad (8) \\ \forall \quad & \alpha H(q_{xy}) + (1-\alpha)H(r_x) > (1+\beta)nR. \end{aligned}$$

The last line is equivalent to $\exists \theta \in [0,1]$, s.t.

$$\begin{aligned} & \alpha H(q_{xy}) > \theta(1+\beta)nR \\ \text{and} \qquad & (1-\alpha)H(r_x) > (1-\theta)(1+\beta)nR. \end{aligned}$$

Substitute the above two inequalities into (8) and by the definition of the source coding error exponents defined in Theorem 1, we have:

$$\begin{aligned} & \Pr[(x_1^n, y_1^{\alpha n}) \neq (\widehat{x}_1^n(1+\beta)n, \widehat{y}_1^{\alpha n}(1+\beta)n)] \\ \geq \; & 2^{-n \min_{\theta\in[0,1]} \left(\alpha E_b^{xy}(\frac{\theta(1+\beta)R}{\alpha}) + (1-\alpha)E_b^x(\frac{(1-\theta)(1+\beta)R}{1-\alpha}) - \epsilon\right)}. \qquad (9) \end{aligned}$$

Now assume the delay-constrained error exponent pair is $(X, Y)$, then the decoding error at time $(1+\beta)n$ can be bounded as follows:

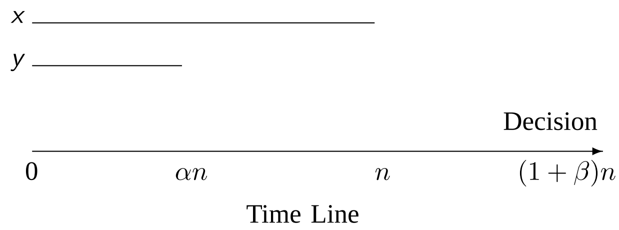



Fig. 6. Bounding the *atypicality* for the outer bound

$$\begin{aligned} & \Pr[(x_1^n, y_1^{\alpha n}) \neq (\widehat{x}_1^n(1+\beta)n, \widehat{y}_1^{\alpha n}(1+\beta)n)] \\ \leq \quad & \sum_{i=1}^{n} \Pr[x_i \neq \widehat{x}_i\,((1+\beta)n)] \\ & + \sum_{i=1}^{\alpha n} \Pr[y_i \neq \widehat{y}_i\,((1+\beta)n)] \\ \leq \quad & K_\epsilon \left(2^{-n\beta(X-\epsilon)} + 2^{-n(1+\beta-\alpha)(Y-\epsilon)}\right). \end{aligned}$$

Combining the above inequality with (9) and letting $(n, \epsilon)$ go to $(\infty, 0)$, we get the desired result:

$$X \leq \frac{1}{\beta} F_R^x(\alpha,\beta) \text{ or } Y \leq \frac{1}{1+\beta-\alpha} F_R^x(\alpha,\beta) \qquad (10)$$

where $F_R^x(\alpha,\beta)$ is defined in Theorem 4. We have the obvious bounds that

$$E(R, x) \leq E_s^x(R) \text{ and } E(R, y) \leq E_s^y(R). \qquad (11)$$

Combining (11) and (10) and noticing the symmetry, we prove Theorem 4. □

### C. Proof of Theorem 5

Due to the space limitations, we can only give the sketch of the coding scheme here and must omit the proof entirely. We describe the scheme that treats $X$ with higher priority, with a parameter $\alpha \in [0,1]$. The case of giving $Y$ priority is identical.

The coding system is illustrated in Figure 7. The encoder first chops the sequences $x_1, ....$ and $y_1, ....$ into blocks of size $N$, where $N$ is sufficiently big. Then the encoder converts each block $\vec{x}_i$ and $\vec{y}_i$ of length $N$ into prefix-free codes with length $N(H(\vec{x}_i) + \epsilon_N)$ and $N(H(\vec{y}_i|\vec{x}_i) + \epsilon_N)$ respectively , where $H(\vec{x}_i)$ is the empirical entropy of block $\vec{x}_i$ and $H(\vec{y}_i|\vec{x}_i)$ is the empirical conditional entropy, $\epsilon_N$ goes to 0 as $N$ goes to infinity.

The FIFO ($\alpha$) encoder buffer has *two* buffers, one for $x$, one for $y$. The buffer sends one *code word* for an $x$ block $\vec{x}_i$ or a $y$ block $\vec{y}_j$ based on the priority order described as follows. $\vec{x}_i$ has higher priority than any $\vec{x}_j$, $j > i$, in that the individual buffers are FIFO. Suppose

[3] We denote the entropy of the type of a sequence by *empirical entropy*. We ignore the integer effects since $n$ can be arbitrarily large.

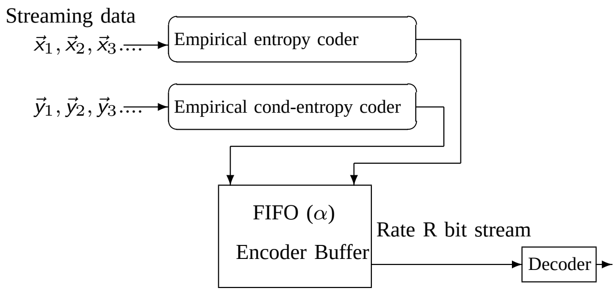



Fig. 7. Two sources one encoder source coding

at time $t$, $\vec{x}_i$ and $\vec{y}_j$ are the topmost blocks for source $x$ and source $y$ respectively. Denote by $l(t)$ the last time that both of the buffers were empty. Then $\vec{x}_i$ has higher priority if $\alpha(i - \frac{l(t)}{N}) \geq (j - \frac{l(t)}{N})$, otherwise $\vec{y}_j$ has higher priority. Notice that $\alpha < 1$, so $i \geq j$.

The decoding error is then converted into a buffer overflow problem, and then the inner bound in Theorem 5 is derived following the analysis in [1], [2].

## V. Streaming error exponent tradeoff for binary erasure channel (BEC) coding

Now consider a streaming channel coding problem shown in Figure 8 The two sources generate a bit stream according to Bernoulli $0.5$ every $\frac{1}{R_1}$ and $\frac{1}{R_2}$ seconds respectively. Then the two sources are fed into a feedback channel coding system where the channel is a binary erasure channel with erasure rate $\delta$. The *anytime* channel coding problem for a single data streaming is studied in [8]. Similar to the source coding case, the delay constrained error exponent tells how fast the bit error converges to zero exponentially with *delay*.

$$\Pr(a_i \neq \widehat{a}_i(j)) \simeq 2^{-(j-i)E(R)}$$

In [8], a "focusing bound" is derived for BEC's which says:

$$E_1(R) = \inf_{\beta>0} \frac{1+\beta}{\beta} D(1 - \frac{R}{1+\beta} \| \delta) \tag{12}$$

The achievability of this bound is proved by a simple send until through coding system. This can be done because the channel is a binary erasure channel with feedback, in a way, the encoder and the decoder are synchronized at every step.

Similar to the source coding problem in Definition 3, the anytime reliability region is defined as follows.

*Definition 4:* We denote by $(E_a(R_1, \delta), E_b(R_2, \delta))$ an achievable error-exponent pair for feedback BEC with erasure rate $\delta$ and two data streams that operate at rate $R_1$ and $R_2$ respectively. All the achievable pairs form an error-exponent region that is a subset of the first quadrant: $\{(E_a, E_b) \in \mathcal{R}^+ \times \mathcal{R}^+ : E_a$ and $E_b$ are achievable anytime reliability functions for two streams with rate $R_1$ and $R_2$ respectively $\}$

First we describe an outer and an inner bound on the anytime reliability region.

### A. Error exponent region for BEC with feedback

We summarize the outer bound result in Theorem 6 and the inner bound result in Theorem 7.

*Theorem 6:* (Outer bound) Consider the anytime channel coding for BEC ($\delta$) for two source with rate $R_1$ and $R_2$ respectively, and the error exponent region in Definition 4, then the error exponent region is a **subset** of:

$$\{(A,B) : A \leq E_1(R_1) \text{ and } B \leq E_1(R_2)\} \bigcap$$
$$\{\bigcap_{\alpha\in[0,1],\beta>0} \mathcal{R}_A(\alpha,\beta)\} \bigcap \{\bigcap_{\alpha\in[0,1],\beta>0} \mathcal{R}_B(\alpha,\beta)\}$$

where $\mathcal{R}_A(\alpha,\beta)$ and $\mathcal{R}_B(\alpha,\beta)$ are "L" shaped regions, $\mathcal{R}_A(\alpha,\beta)$ is

$$\{(A,B) : A \leq \frac{1+\beta}{\beta} G(\alpha,\beta) \text{ or } B \leq \frac{1+\beta}{1+\beta-\alpha} G(\alpha,\beta)\}$$

where

$$G(\alpha,\beta) = D(1 - \frac{R_1 + \alpha R_2}{1+\beta} \| \delta),$$

and similarly $\mathcal{R}_B(\alpha,\beta)$ is

$$\{(A,B) : A \leq \frac{1+\beta}{1+\beta-\alpha} G'(\alpha,\beta) \text{ or } B \leq \frac{1+\beta}{\beta} G'(\alpha,\beta)\}$$

where

$$G'(\alpha,\beta) = D(1 - \frac{R_2 + \alpha R_1}{1+\beta} \| \delta).$$

*Theorem 7:* (inner bound) Consider the anytime channel coding for BEC ($\delta$) for two source with rate $R_1$ and $R_2$ respectively, and the error exponent region in Definition 4, then the error exponent region is a **superset** of:

$$\{\bigcup_{\alpha\in[0,1]} \mathcal{S}_A(\alpha)\} \bigcup \{\bigcup_{\alpha\in[0,1]} \mathcal{S}_B(\alpha)\}$$

where $\mathcal{S}_A(\alpha)$ and $\mathcal{S}_B(\alpha)$ are rectangular regions:

$$\mathcal{S}_A(\alpha) = \left( \bigcap_{\beta>0} \bar{\mathcal{R}}(\alpha,\beta) \right) \bigcap \tag{13}$$
$$\{(A,B) : B < \inf_{\alpha-1<\beta<0} \frac{1+\beta}{1+\beta-\alpha} D(1 - R_1 - \frac{\alpha R_2}{1+\beta} \| \delta)\}$$

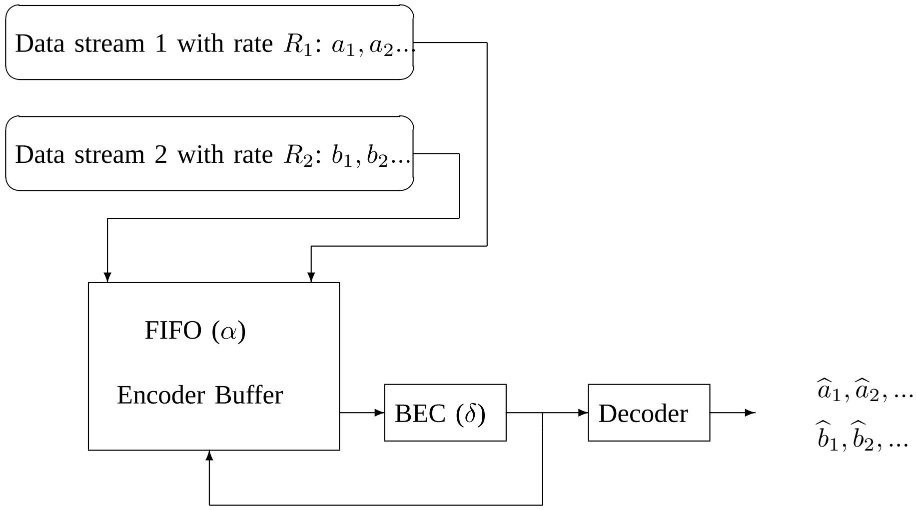



Fig. 8. Anytime channel coding for BEC with feedback, 2 streaming data

where $\bar{\mathcal{R}}(\alpha,\beta) \triangleq \{(A,B) : A \le \frac{1+\beta}{\beta} G(\alpha,\beta)$ and

$$B \le \frac{1+\beta}{1+\beta-\alpha} G(\alpha,\beta)\}.$$

Similarly for $\mathcal{S}_B(\alpha)$. The second line in (13) comes from the scenario where a decoding error might be made at time $(1+\beta)n$ for $b_{\alpha n}$, where $\alpha - 1 < \beta < 0$. A proper bounding of the error probability gives the extra term in the second line of (13). This gives the monotonic behavior of the inner bound boundary which is lacking in previous studies.

*Remark: Note that the corner point of $\bar{\mathcal{R}}(\alpha,\beta)$ for the inner bound and the corner point of $\mathcal{R}(\alpha,\beta)$ are the same. This is quite interesting because the BEC with feedback problem is simpler than the two stream source coding problem, because all the randomness comes from the single source–the channel. And it is also easy to see why the inner bound is smaller than the outer bound. The outer bound is the intersections of the "L" shaped regions with corners parameterized by $(\alpha,\beta)$, while the inner bound is a subset of the union (parameterized by $\alpha$)of the intersection of rectangles (parameterized by $\beta$) with the same corners parameterized by $(\alpha,\beta)$.*

### *B. Proofs*

*1) Outer bound:* Now we consider the most likely error event for those bits considered in Figure 9.

$$\Pr((a_1^{nR_1}, b_1^{\alpha nR_2}) \ne (\widehat{a}_1^{nR_1}((1+\beta)n), \widehat{b}_1^{\alpha nR_2}(1+\beta)n))$$

An error occurs if the number of bits through the BEC with feedback from time 1 to $(1+\beta)n$ is less than the total number of bits generated by the two sources. Or equivalently, the bits erased by the channel is more than

$$(1+\beta)n - nR_1 - \alpha nR_2$$

The empirical erasure rate is $1 - \frac{R_1+\alpha R_2}{1+\beta}$. The probability of that is lower bounded by

$$2^{-[(1+\beta)n]\left(D(1-\frac{R_1+\alpha R_2}{1+\beta}\|\delta)-\epsilon_n\right)} \tag{14}$$

where $\epsilon_n$ converges to 0 with $n$. Note that we are concerned the union bound on the errors for both the data streams. So similar to the streaming source coding case we have:

$$\begin{aligned}&\Pr((a_1^{nR_1}, b_1^{\alpha nR_2}) \ne (\widehat{a}_1^{nR_1}((1+\beta)n), \widehat{b}_1^{\alpha nR_2}(1+\beta)n))\\ &> 2^{-[(1+\beta)n]\left(D(1-\frac{R_1+\alpha R_2}{1+\beta}\|\delta)-\epsilon_n\right)}.\end{aligned} \tag{15}$$

This means that at time $(1+\beta)n$, at least one bit of $a_1^{nR_1}$ *or* one bit of $b_1^{\alpha nR_2}$ is not decoded correctly. The *minimum* effective delay is $\beta n$ and $(\beta+1-\alpha)n$ respectively. This implies that either

$$E_a \le \frac{1+\beta}{\beta} D(1 - \frac{R_1+\alpha R_2}{1+\beta}\|\delta) \tag{16}$$

*or*

$$E_b \le \frac{1+\beta}{1+\beta-\alpha} D(1 - \frac{R_1+\alpha R_2}{1+\beta}\|\delta). \tag{17}$$

Since $\beta$ and $\alpha$ are arbitrary, and by noticing the symmetry, we have the desired result in Theorem 6. ■

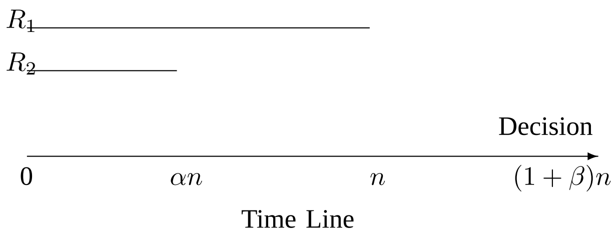



Fig. 9. Bounding the *atypicality* for the two stream BEC with feedback

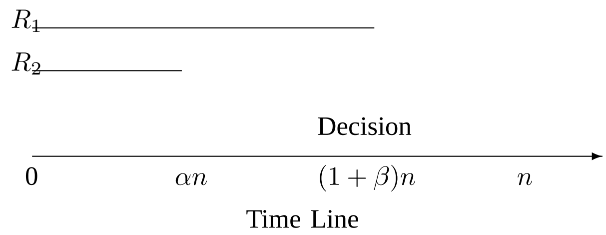



Fig. 10. Bounding the *atypicality* for the two stream BEC with feedback, case $\beta < 0$

*2) Inner boundd:* The encoder is similar to the single stream anytime channel coding for BEC's. The encoder simply sends the bit with the highest priority remained in the buffer. If the bit gets erased by the channel, the encoder resends it until the decoder receives the bit. If the buffer is empty the encoder sends a garbage bit to the channel. The encoder and the decoder are synchronized in the sense that the decoder is aware of when the encoder buffer is empty thus simply discards the garbage bit when the buffer is empty.

The only remaining issue is to give the priorities to the bits in the encoder buffer. Similar to the two stream source coding problem, we implement the FIFO($\alpha$) protocol. Within the same stream, older bits always have higher priority. Now we give the priority across the streams. Without loss of generality we assume that the first stream with rate $R_1$ has higher priority, this is parameterized by an $\alpha \in [0,1]$. Now suppose the last time when the buffer is empty is at time $0$ (if not, we can shift the time-line to make it so). Then the $nR_1$'th bit from stream $1$ which is generated at time $n$: $a_{nR_1}$ has the same priority as the $\alpha nR_2$'th bit from stream $2$ which is generated at time $\alpha n$: $b_{nR_2}$.

With the in-stream and cross-stream priorities defined, we can bound the dominant error event for any bit. Again, without loss of generality we assume that the last time the buffer is empty is at time $0$. As illustrated in Figure 9, a decoding error probability for $a_{nR_1}$ and $b_{nR_2}$ are the same at time $(1+\beta)n$ for $\beta > 0$, because they have the same priority. Exactly the same as the outer bound analysis, an error occurs if and only if there are too many erasures between time $0$ and time $(1+\beta)n$. Hence we have

$$\begin{aligned}&\Pr(a_{nR_1} \neq \widehat{a}_{nR_1}((1+\beta)n))\\ &< 2^{-[(1+\beta)n]\left(D(1-\frac{R_1+\alpha R_2}{1+\beta}\|\delta)+\epsilon_n\right)}\end{aligned} \tag{18}$$

and

$$\begin{aligned}&\Pr(b_{\alpha nR_2} \neq \widehat{b}_{\alpha nR_2}((1+\beta)n))\\ &< 2^{-[(1+\beta)n]\left(D(1-\frac{R_1+\alpha R_2}{1+\beta}\|\delta)+\epsilon_n\right)}\end{aligned} \tag{19}$$

The effective delay is $\beta n$ and $(\beta+1-\alpha)n$ respectively. Union bounding all the error probabilities for $\beta > 0$, we have

$$E_a \geq \inf_{\beta>0} \frac{1+\beta}{\beta} D(1-\frac{R_1+\alpha R_2}{1+\beta}\|\delta) \tag{20}$$

*and*

$$E_b \geq \inf_{\beta>0} \frac{1+\beta}{1+\beta-\alpha} D(1-\frac{R_1+\alpha R_2}{1+\beta}\|\delta). \tag{21}$$

Now let us consider the scenario when the decoder decodes $b_{\alpha n}$ before time $n$, this is illustrated in Figure 10. Following the same argument as before, we can easily get

$$\begin{aligned}&\Pr(b_{\alpha nR_2} \neq \widehat{b}_{\alpha nR_2}((1+\beta)n))\\ &< 2^{-[(1+\beta)n]\left(D(1-\frac{R_1(1+\beta)+\alpha R_2}{1+\beta}\|\delta)+\epsilon_n\right)}\end{aligned} \tag{22}$$

The effective delay is still $(1-\alpha+\beta)n$ for stream 2, thus we have[4]

$$\begin{aligned}E_b &> \inf_{\alpha-1<\beta<0} \frac{1+\beta}{1+\beta-\alpha} D(1-R_1-\frac{\alpha R_2}{1+\beta}\|\delta)\\ &= \inf_{\alpha<\lambda<1} \frac{1}{1-\lambda} D(1-R_1-\lambda R_2\|\delta).\end{aligned}$$

Note that the above object function does not have $\alpha$ in it. So there exists $\lambda^* \in [0,1]$ to minimize it, hence if $\alpha < \lambda^*$, the above bound does not change with $\alpha$. This is clearly illustrated in Figure 11.

The above three inequalities on $E_a$ and $E_b$ are true for a fixed $\alpha \in [0,1]$ which is predetermined by the coding system. Now the union of those rectangle regions and symmetry give the desired result. ■

These two bounds are illustrated in Figure 11.

[4] On the second line we replace $\frac{\alpha}{1+\beta}$ with $\lambda$ which is within $[\alpha, 1]$.

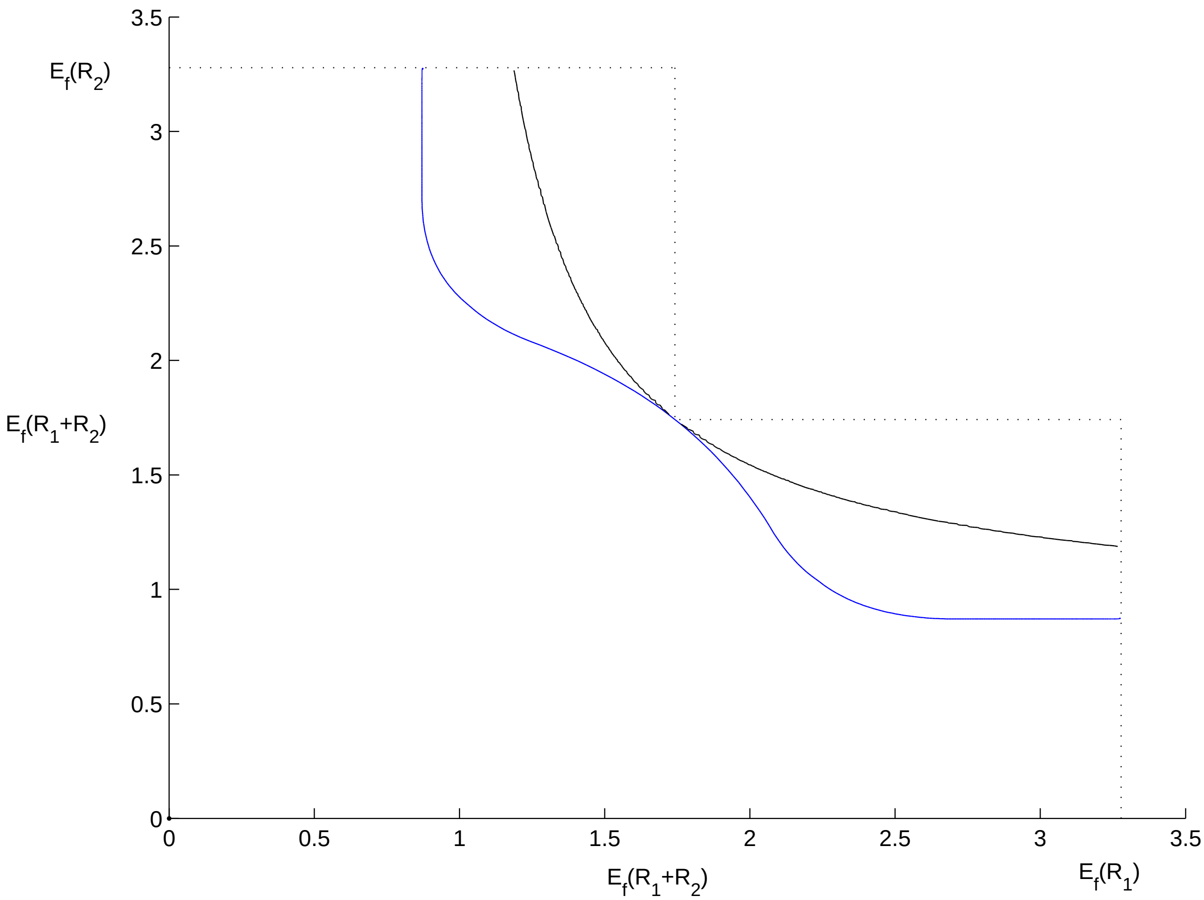



Fig. 11. Anytime error exponent region for BEC with erasure rate $0.1$, $R_1 = R_2 = 0.4$, $E_f(R)$ is the focusing bound.

## REFERENCES


[1] Cheng Chang. *Streaming source coding with delay, PhD thesis*. University of California, Berkeley, 2007.

[2] Cheng Chang and Anant Sahai. The price of ignorance: the impact on side-information for delay in lossless source coding. *submitted to IEEE Transactions on Information Theory*, 2007.

[3] Cheng Chang and Anant Sahai. The error exponent with delay for lossless source coding. *Information Theory Workshop, Punta del Este, Uruguay*, March 2006.

[4] Imre Csiszár. The method of types. *IEEE Transactions on Information Theory*, 44:2505–2523, 1998.

[5] Imre Csiszár and János Körner. *Information Theory*. Akadémiai Kiadó, Budapest, 1986.

[6] Robert Gallager. Source coding with side information and universal coding. Technical Report LIDS-P-937, Mass. Instit. Tech., 1976.

[7] Robert G. Gallager. *Information Theory and Reliable Communication*. John Wiley, New York, NY, 1971.

[8] Anant Sahai. Why block-length and delay behave differently if feedback is present. *IEEE Trans. Inform. Theory*, 54(5):1860–1886, 2008.

[9] Anant Sahai and Sanjoy K. Mitter. The necessity and sufficiency of anytime capacity for stabilization of a linear system over a noisy communication link. part II: vector systems. *IEEE Trans. Inform. Theory*, submitted.

[10] Claude Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27, 1948.

[11] Lihua Weng, Sandeep S. Pradhan, and Achilleas Anastasopoulos. Error exponent regions for gaussian broadcast and multiple access channels. *IEEE Transactions on Information Theory*, 2008 to appear.