Why delay and block length are not the same thing for channel coding with feedback

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

***Abstract*— Classically, information theorists have understood the tradeoff between delay and system error performance by using block-length as a proxy. Although this turns out to be correct when feedback is absent, it is shown to be very misleading when feedback is available. For symmetric channels, fixed block-length communication systems show no improvements in their error exponents while fixed delay systems can show dramatic gains. We present a new upper bound ("the focusing bound") on the error performance possible in a fixed-delay system with feedback. Finally, we sketch how this upper bound can be asymptotically achieved with feedback for erasure channels as well as any symmetric channel that has strictly positive feedback zero-error capacity.**

I. Introduction

The channel coding theorems studied in information theory are not just interesting as mathematical results, they also provide insights into the underlying tradeoffs in communication systems. The two most fundamental parameters when it comes to reliable data transport are end-to-end system delay and the probability of error. Error probability is fundamental because a low probability of bit error lies at the heart of the digital revolution justified by the source/channel separation theorem. Shannon's contributions in [1], [2] justified an architecture based on abstracting all communication tasks as involving the transport of messages reliably. When delay does not matter, all communication tasks are essentially equivalent. In this light, delay is important because it is the most basic cost that a system must pay in exchange for reliability — it allows the laws of large numbers to be harnessed to smooth out the variability introduced by random communication channels. And yet, [3] points out that the architectural guidance provided by information theory into dealing with delay has not been nearly as successful as the basic message regarding the desirability of reliable communication.

Traditionally, block-length has been used as a proxy for end-to-end delay since block-codes are easier to understand than non-block codes. This turns out to be correct when there is no feedback since the dominant error events turn out to involve the channel's behavior between when the message is made available to the encoder and when it is needed at the decoder. When feedback is allowed, block-codes are unable to really exploit it. For memoryless symmetric channels, not only is the capacity not increased with feedback, but the sphere-packing bound remains essentially unchanged and so no significant improvements in error probability are possible in the high-rate regime of greatest interest [4]. When the latency is only constrained on average, then the variable block-length story of [5], [6] shows that feedback can significantly lower the probability of error without much impact on the average delay. Either way, the architectural message from block-codes seems to remain: with or without feedback, to get the best probability of error, aggregate your messages and use the longest block-length you can afford given your latency constraint.

Yet even when *fixed* end-to-end delay is desired, this paper shows that nonblock codes can provide a tremendous reduction in the probability of bit error if feedback is allowed. The role of the sphere-packing (volume) bound is played by the uncertainty focusing bound in giving the fundamental limit on what is asymptotically possible in the limit of large delays. The dominant error events involve a mixture of past and future channel behavior and the resulting bound is also achievable with feedback for erasure channels and any channel with strictly positive feedback zero-error capacity. The codes that achieve the focusing bound hint at an architectural message that is different from that provided by traditional fixed-length block-coding. When feedback is available, the "messages" should be of "moderate size" — long enough, but not a length comparable to the target end-to-end latency itself. Feedback should be used to adapt the block-lengths as needed and do flow-control on the instantaneous rate of information transfer.

This short paper is a summary of the results contained in [7] and [10]. The reader is referred to [7] for more details, motivation, as well as a technical perspective on the existing results in the literature concerning feedback and reliability. An interpretation of the new results in the context of remote stabilization problems [8] is found in [9] where it is shown how to leverage noiseless, but low rate, feedback to make more effective use of a higher-capacity noisy feedback channel in order to stabilize an unstable system in closed loop. The basic code construction used to show achievability of the focusing bound for channels with strictly positive zero-error capacity is then extended in [10] to generic channels. While not attaining the focusing bound itself, the codes of [10] do beat the sphere-packing bound with fixed delay in the high-rate regime.

The core ideas of [7] are also explored in the context of lossless source coding in [11], [12]. The point-to-point setting is considered in [11] and the source-coding counterpart of the uncertainty focusing bound is developed. Feedback is irrelevant and the dominant error events turn out to involve the source behavior *before* the symbol in question arrived at the encoder. This is used to show that optimal block-codes are quite bad from the perspective of fixed end-to-end delay constraints when used over fixed-rate noiseless channels. Instead, "moderate sized" fixed-to-variable length codes with their output rate smoothed using a FIFO queue with deterministic service times asymptotically achieve the best possible tradeoff between fixed deadlines and the probability of error. The case of side-information at the decoder is considered in [12] and an upper bound on the reliability function with delay is derived by considering error events due to atypically bad side information between the time of symbol arrival and when it is required at the decoder. This bound turns out to be tight for certain symmetric cases.

Put together, these results make precise Shannon's intriguing comment at the close of [2]:

> "[the duality between source and channel coding] can be pursued further and is related to a duality between past and future and the notions of control and knowledge. Thus we may have knowledge of the past and cannot control it; we may control the future but have no knowledge of it."

II. Review of Block Coding

A. Fixed-length

The fundamental lower-bound on error probability comes from the sphere-packing or volume bound, and this bound is also known to be achievable at high rates by random-coding [13]. Reliable communication is not possible if during the block, the channel acts like one whose capacity is less than the target rate. Following [14] and [15], for block codes this idea immediately gives the following bound on the exponential error probability:

$$E^{+}(R) = \inf_{G:C(G)<R} \max_{\vec{r}} D\left(G||P|\vec{r}\right) \tag{1}$$

where $D(G||P|\vec{r})$ is the divergence term that governs the exponentially small probability of the true channel P behaving like channel G when facing the input distribution coming from the codeword composition $\vec{r}$.

Even with causal noiseless feedback, there is no way around this bound because channel capacity does not increase with feedback for memoryless channels. Without feedback, the bound can be tightened to the form traditionally known as the sphere-packing bound.

$$E_{sp}(R) = \max_{\vec{r}} \min_{G:I(\vec{r},G)\leq R} D\left(G||P|\vec{r}\right) \tag{2}$$

For symmetric channels, the optimizing codeword composition $\vec{r}$ is always uniform and $E_{sp}(R) = E^{+}(R)$.

An alternate form for $E_{sp}(R)$ is given by:

$$E_{sp}(R) = \max_{\rho>0} \left[E_0(\rho) - \rho R\right] \tag{3}$$

with the Gallager function $E_0(\rho)$ defined as:

$$E_0(\rho) = \max_{\vec{q}} -\ln \sum_{y} \left[\sum_{x} q_x p_{x,y}^{\frac{1}{1+\rho}}\right]^{(1+\rho)} \tag{4}$$

Note that for symmetric channels, it suffices to use a uniform $\vec{q}$ while optimizing (4). Also, since the random-coding error exponent is given by:

$$E_r(R) = \max_{0<\rho\leq 1} \left[E_0(\rho) - \rho R\right] \tag{5}$$

It is clear that the sphere-packing bound is achievable, even without feedback, at rates close to C since for those rates, $\rho < 1$ optimizes both expressions [13]. The points on the sphere-packing bound where $\rho > 1$ are also achievable by random coding if the sense of "correct decoding" is slightly relaxed. Rather than forcing the decoder to emit a single estimated codeword, list-decoding allows the decoder to emit a list of guessed codewords. The decoding is considered correct if the true codeword is on the list. Decoding a random code with list size ℓ has exponent:[13]

$$E_{r,\ell}(R) = \max_{0<\rho\leq \ell} \left[E_0(\rho) - \rho R\right] \tag{6}$$

is achievable. At high rates (where the maximizing ρ is small), there is no benefit from relaxing to list-decoding, but it makes a difference at low rates.

B. Variable-length

Without feedback, a variable-length mode of operation is impossible since the encoder has no way to know if the channel is behaving typically or atypically. With noiseless feedback, the length of the codeword can be made to vary based on what the channel has done so far — as long as that variation depends only on the received channel symbols. The proposed error exponent for variable-length channel codes divided the negative log of the probability of block error ϵ by the expected length $E[N_\epsilon]$ of an average rate $\bar{R}$ variable-length code that attains the probability of error ϵ.

$$E_{vl}(\bar{R}) = \lim_{\epsilon\to 0} -\frac{\ln(\epsilon)}{E[N_\epsilon]}$$

Burnashev gave the upper bound to this exponent by using martingale arguments treating the ending of a block as a stopping time [5].

$$E_v(\bar{R}) = C_1\left(1 - \frac{\bar{R}}{C}\right) \tag{7}$$

where C is the Shannon capacity of the channel and

$$C_1 = \max_{i,k} D(P(\cdot|i)||P(\cdot|k)) = \max_{i,k} \sum_{l} p_{il} \ln \frac{p_{il}}{p_{kl}}. \tag{8}$$

represents the maximum divergence possible between channel output distributions given choice of two input letters.

This bound is also achievable by using a repeat-until-success scheme as described in [6]. The relevant "flow-control" information is carried by confirm/deny messages sent by the encoder telling the decoder whether or not it is going to repeat the block or not. It turns out that the rate just determines how many channel uses are left over for the flow-control messages and the variable block-coding reliability is purely a function of that proportion.

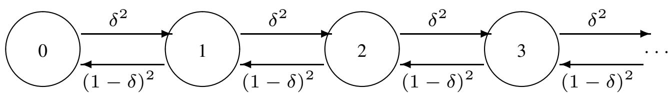

Fig. 1. The birth-death Markov chain governing the rate $\frac{1}{2}$ feedback communication system over an erasure channel.

III. NON-BLOCK CODES AND AN EXAMPLE

Another classical approach to the problem of reliable communication is to consider codes without a block structure. Convolutional and tree codes represent the prototypical examples. It was realized early on that in an infinite constraint length convolutional code under ML decoding, all bits will eventually be decoded correctly [13]. However, if the end-to-end delay is forced to be bounded, then the bit error probability with delay is governed by $E_r(R)$ for random convolutional codes, even when the constraint lengths are unbounded [16]. This performance with delay is also achievable using an appropriately biased sequential decoder [17]. A nice feature of sequential decoders is that they are not tuned to any target delay — they can be prompted for estimates at any time and they will give the best estimate that they have. Thus an infinite constraint-length convolutional code with appropriate sequential decoding achieves the exponent $E_r(R)$ delay universally over all (sufficiently long) delays.

Pinsker claimed in [18] that the sphere-packing bound continued to bound the performance of nonblock codes both with and without feedback. He had proofs for the BSC case, but asserted that the result held more generally. While he was right for the without feedback case, it turns out that there is a subtle flaw in his argument regarding the case with feedback.

A. The BEC example

The binary erasure channel with erasure probability $\delta < \frac{1}{2}$ used at bit-rate $R' = \frac{1}{2}$ gives a counterexample to Pinsker's conjecture. The BEC is so simple that everything can be understood with a minimum of overhead.

The sphere-packing bound in this case corresponds to the probability that the channel erases more than $\frac{1}{2}$ of the inputs during the block:

$$E_{sp}(\frac{1}{2}) = D(\frac{1}{2}||\delta) = -\frac{\ln(4\delta(1-\delta))}{2} \tag{9}$$

For $\delta = 0.4$, this corresponds to an error exponent of about 0.02. Even with feedback, there is no way for a fixed block-length code to beat this exponent. If the channel lets fewer than $\frac{n}{2}$ bits through the channel, it is impossible to reliably communicate an $\frac{n}{2}$ bit message!

If causal noiseless feedback is available, the natural non-block code just retransmits a bit until it is correctly received. As bits arrive steadily at the rate $R' = \frac{1}{2}$, they enter a FIFO queue of bits awaiting transmission. If we look at the queue state every two channel uses, it can be modeled (see Figure 1) as a birth-death Markov chain with a δ^2 probability of birth and a $(1-\delta)^2$ probability of death. Converting that into an error exponent with delay d gives:

$$E_f^{bec}(\frac{1}{2}) = \ln(1-\delta) - \ln(\delta) \tag{10}$$

Plugging in $\delta = 0.4$ gives an exponent of more than 0.40. This is about twenty times higher than the sphere-packing bound!

IV. THE FOCUSING BOUND

Restricting attention to symmetric channels, the BEC case can be abstracted to get a general bound on the probability of error with delay. We call this bound the "focusing bound" because it is based on the idea of having the encoder focus as much of the decoder's uncertainty as possible onto bits whose deadlines are not pending. To bound what is possible, a fixed delay code is translated into a fixed block-length code. Each different block-length provides its own bound at all rates, with the final bound at any given rate coming from the worst case block-length.

Definition 4.1: A *rate R encoder with noiseless feedback* is a sequence of maps $\mathcal{E}_t$. The range of each map is the discrete set $\mathcal{X}$. The t-th map takes as input the available data bits $B_1^{\lfloor R't \rfloor}$, as well as all the past channel outputs Y_1^{t-1}.

Randomized encoders with noiseless feedback also have access to a continuous uniform random variable W_t denoting the common randomness available in the system.

Definition 4.2: A *delay d rate R decoder* is a sequence of maps $\mathcal{D}_i$. The range of each map is just an estimate $\widehat{B}_i$ for the i-th bit taken from $\{0,1\}$. The i-th map takes as input the available channel outputs $Y_1^{\lceil \frac{i}{R'} \rceil + d}$ which means that it can see d time units beyond when the bit to be estimated first had a chance to impact the channel inputs.

Randomized decoders also have access to all the continuous uniform random variables W_t.

Definition 4.3: The fixed-delay error exponent α is asymptotically *achievable* at rate R across a noisy channel if for every delay d_j in some increasing sequence $d_j \to \infty$ there exist rate R encoders and delay d_j decoders $\mathcal{E}^{d_j}, \mathcal{D}^{d_j}$ that satisfy the following properties when used with input bits B_i drawn from iid fair coin tosses.

1) For every j, there exists an $\epsilon_j < 1$ so that $P(B_i \neq \widehat{B}_i(d_j)) \leq \epsilon_j$ for every $i \geq 1$. The $\widehat{B}_i(d_j)$ represents the delay d_j estimate of B_i produced by the $\mathcal{E}^j, \mathcal{D}^j$ pair connected to the input B and the channel in question.
2) $\lim_{j\to\infty} \frac{-\ln \epsilon_j}{d_j} \leq \alpha$

The exponent α is asymptotically *achievable universally over delay* or in an *anytime fashion* if a single encoder $\mathcal{E}$ can be used above for all d_j above.

Theorem 4.4: Focusing bound: For a discrete memoryless channel, no delay exponent $\alpha > E_a(R)$ is asymptotically achievable even if the encoders are allowed access to noiseless feedback.

$$E_a(R) = \inf_{0<\lambda<1} \frac{E^+(\lambda R)}{1-\lambda} \tag{11}$$

where E^+ is the Haroutunian exponent from (1). For a symmetric DMC, $E_a(R)$ can be expressed parametrically:

$$E_a(R) = E_0(\eta) \; ; \; R = \frac{E_0(\eta)}{\eta} \tag{12}$$

where $E_0(\eta)$ is the Gallager function from (4), and η ranges from 0 to ∞.

Proof: See [7]. The key idea is to construct a block-code of length $n = \frac{d}{1-\lambda}$ where d is the target latency. The rate is just λR since we can throw out any bits that arrive during the last d channel uses since their deadlines are after the end of the block. The focusing bound is obtained by translating a lower-bound on the error-exponent in the fixed-block paradigm to be relative to the delay d instead. □

V. THE (n, c, l) FAMILY OF CODES

The focusing bound is attained for the BEC with feedback using the natural "repeat bits until successful" code and can also be asymptotically attained for any noisy channel provided we have access to a low-rate channel that can deliver perfectly noiseless flow-control bits from the encoder to the decoder. The code is described below.

Call $c \geq 1$ the chunk length, 2^l the list length, and $n > l$ the data block length. The (n, c, l) scheme is:

- Queue up incoming bits and assemble them into blocks of size $\frac{ncR}{\ln 2}$ bits. If there are fewer than $\frac{ncR}{\ln 2}$ bits still awaiting transmission, just idle by transmitting an arbitrary input letter.
- At every noisy channel use, the encoder transmits the next position in an infinite-length random codeword [1] associated with the current data block.
- If the time is an integer multiple of c, use the noiselessly fedback channel outputs to simulate the decoder's attempt to decode the current codeword to within a list of the top 2^l items. If the true data-block is one of the 2^l items, send a 1 over the noiseless forward link. Also send the disambiguating l bits representing the true block's index within the decoder's list. Remove the current block of $\frac{ncR}{\ln 2}$ bits from the main data queue as well. If the true block is not in the decoder's list, just send a 0 over the noiseless forward link.
- At the decoder, the encoder queue length is known perfectly since it can only change by the deterministic arrival of data bits or when a noise-free confirm or deny bit has been sent. Thus the decoder always knows which input block a given channel output Y_t or fortified symbol S_t corresponds to.
- If the time is an integer multiple of c and the decoder receives a 1 noiselessly, then it decodes what it has seen to a list of the top 2^l possibilities for this block. It will use the next l noisefree bits to disambiguate this list and will use the result as its estimate for the block.

[1]Drawn according to the $E_0(\eta)$ maximizing distribution for the η such that the data rate $R = \frac{E_0(\eta)}{\eta}$.

Such schemes are shown in [7] to be asymptotically optimal:

Theorem 5.1: By appropriate choice of (n, c, l), it is possible to asymptotically achieve all delay exponents $\alpha < E_a(R)$ where $E_a(R)$ is the focusing bound for the fortified system built around a symmetric DMC by adding a rate $\frac{1}{k}$ noisefree forward link where k can be made as small as we like.

Key Proof Idea: [7] Consider each codeword as a point-message. Pick a ρ so that the target $\alpha = E_0(\rho)$. Let $\widetilde{n} = n - n\frac{R}{\widetilde{C}}$ where $\widetilde{C} = \frac{E_0(\rho)}{\rho}$. It turns out that the problem can be lifted to be viewed as the transmission of one point message for every $\widetilde{n}$ chunks where the probability of not making it across is only $\exp(-cE_0(\rho))$ for each chunk. Since $\widetilde{n}$ can be made large enough, this is like the low-rate erasure case. □

A. Channels with positive zero-error capacity

The fortified communication scheme is easily adapted to channels with strictly positive zero-error capacity by just using the feedback zero-error capacity to carry the flow-control information [7]. There is no k. Instead, let θ be block-length required to realize feedback zero-error transmission of at least $l+1$ bits. As illustrated in Figure 2, terminate each chunk with a length θ feedback zero-error code and use it to transmit the flow-control information. If the chunk size is c, then it is as though we are operating with only a fraction $(1 - \frac{\theta}{c})$ of the channel uses. The overhead tends to zero by making the chunk sizes long giving the following corollary to Theorem 5.1:

Corollary 5.1: By appropriate choice of (n, c, l), it is possible to asymptotically achieve all delay exponents $\alpha < E_a(R)$ where $E_a(R)$ is the focusing bound for any symmetric channel with $C_{0,f} > 0$. Furthermore, such schemes achieve the delay exponent α in a delay universal or "anytime" sense.

B. Channels without zero-error capacity

The flow control information is carried during its θ channel uses per chunk using an infinite constraint-length time-varying random convolutional code.[2] Unlike a zero-error code, all that such a code can guarantee is that the probability of error in the flow-control stream is exponentially small in the number of channel uses that have occurred in the code since that message.

As a result, the θ must be kept proportional to the chunk length c to avoid having the flow-control messages cause too many errors. The flow control effective rate therefore goes to zero and the relevant error exponent is about $E_0(1)$. Balancing all the error probabilities and optimizing the choice of θ gives the following theorem: [10]

Theorem 5.2: By appropriate choice of (n, c, l, θ), it is possible to asymptotically achieve all delay exponents $\alpha < E'(R)$ where the tradeoff curve is given parametrically by varying $\rho \in (0, \infty)$:

$$E'(\rho) = \left(\frac{1}{E_0(\rho)} + \frac{1}{E_0(1)} \right)^{-1} \tag{13}$$

$$R(\rho) = \frac{E'(\rho)}{\rho} \tag{14}$$

[2]By using feedback, [7] shows how to implement such a code with bounded expected computational complexity at low rates.

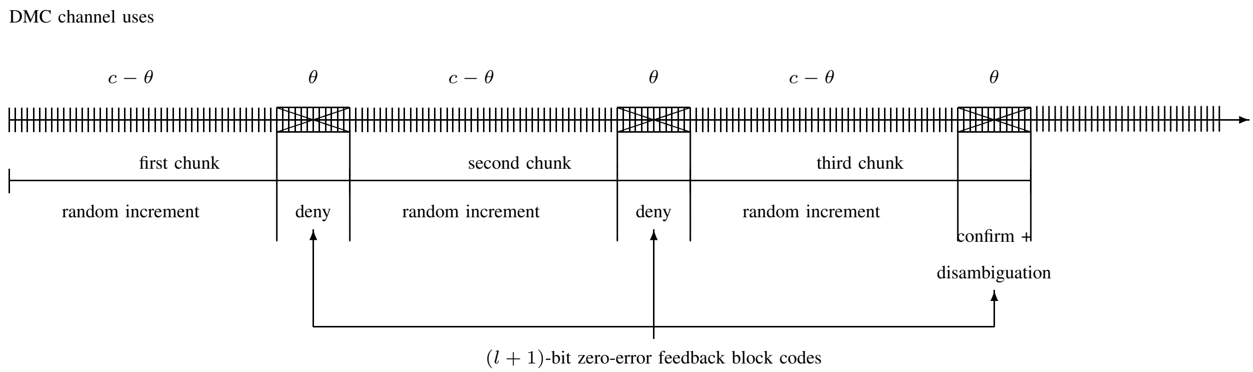

Fig. 2. One block's transmission in the (n, c, l, θ) code for noisy channels. The θ is used to carry $l+1$ flow-control bits reliably. If the channel has a strictly positive feedback zero-error capacity, θ does not scale with c. If it does not, θ is proportional to c.

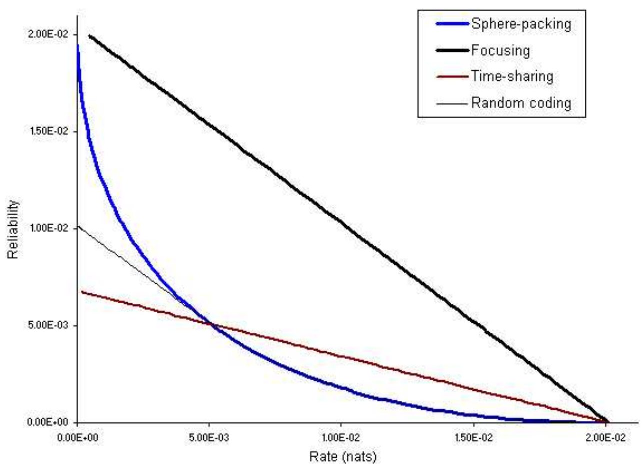

Fig. 3. The reliability functions for the binary symmetric channel with crossover probability $\delta = 0.4$. The sphere-packing bound approaches capacity quadratically flat while the focusing bound and the new scheme both approach the capacity point linearly.

The superiority of these exponents to the sphere-packing bound in the high rate regime is immediately clear since they are basically like the focusing bound in form. The reliability drops linearly in the neighborhood of capacity rather than quadratically flat. Some algebra and simple calculus reveals that the focusing bound has slope $2C/\frac{\partial^2 E_0(0)}{\partial \rho^2}$ in the vicinity of the $(C, 0)$ point, while the $E'(R)$ curve achieved by Theorem 5.2 has the lower slope $-2E_0(1)/(C - \frac{E_0(1)}{2C}\left(\frac{\partial^2 E_0(0)}{\partial \rho^2}\right))$. Figure 3 illustrates the bounds for a BSC with crossover probability 0.4.

ACKNOWLEDGMENTS

The author thanks his student Tunc Simsek for many productive discussions. This work essentially builds on the line of investigation that we opened up in Tunc's doctoral thesis.

REFERENCES

[1] C. E. Shannon, "A mathematical theory of communication," *Bell System Technical Journal*, vol. 27, pp. 379–423, 623–656, July/Oct. 1948.

[2] ——, "Coding theorems for a discrete source with a fidelity criterion," *IRE National Convention Record*, vol. 7, no. 4, pp. 142–163, 1959.

[3] A. Ephremides and B. Hajek, "Information theory and communication networks: An unconsummated union," *IEEE Trans. Inform. Theory*, vol. 44, pp. 2416–2434, Oct. 1998.

[4] R. L. Dobrushin, "An asymptotic bound for the probability error of information transmission through a channel without memory using the feedback," *Problemy Kibernetiki*, vol. 8, pp. 161–168, 1962.

[5] M. V. Burnashev, "Data transmission over a discrete channel with feedback, random transmission time," *Problemy Perdachi Informatsii*, vol. 12, no. 4, pp. 10–30, Oct./Dec. 1976.

[6] H. Yamamoto and K. Itoh, "Asymptotic performance of a modified Schalkwijk-Barron scheme for channels with noiseless feedback," *IEEE Transactions on Information Theory*, vol. 25, no. 6, pp. 729–733, Nov. 1979.

[7] 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

[8] A. Sahai and S. K. Mitter, "The necessity and sufficiency of anytime capacity for stabilization of a linear system over a noisy communication link. part I: scalar systems," *IEEE Trans. Inform. Theory*, submitted for publication. [Online]. Available: http://www.eecs.berkeley.edu/˜sahai/Papers/control-part-I.pdf

[9] A. Sahai, "Stabilization using both noisy and noiseless feedback," submitted to MTNS, 2006. [Online]. Available: http://www.eecs.berkeley.edu/˜sahai/Papers/Sugar.pdf

[10] ——, "How to beat the sphere-packing bound with feedback," submitted to ISIT, 2006. [Online]. Available: http://www.eecs.berkeley.edu/˜sahai/Papers/SphereBeat.pdf

[11] C. Chang and A. Sahai, "The error exponent with delay for lossless source coding," *Information Theory Workshop*, Mar. 2006 to appear. [Online]. Available: http://www.eecs.berkeley.edu/˜sahai/Papers/ITW06.pdf

[12] ——, "Upper bound on error exponents with delay for lossless source coding with side-information," submitted to ISIT, 2006. [Online]. Available: http://www.eecs.berkeley.edu/˜sahai/Papers/ChengISIT06.pdf

[13] R. G. Gallager, *Information Theory and Reliable Communication*. New York, NY: John Wiley, 1971.

[14] I. Csiszar and J. Korner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. Akademiai Kiado, 1981.

[15] E. A. Haroutunian, "Lower bound for error probability in channels with feedback," *Problemy Peredachi Informatsii*, vol. 13, no. 2, pp. 36–44, 1977.

[16] G. D. Forney, "Convolutional codes II. maximum-likelihood decoding," *Information and Control*, vol. 25, no. 3, pp. 222–266, July 1974.

[17] F. Jelinek, "Upper bounds on sequential decoding performance parameters," *IEEE Trans. Inform. Theory*, vol. 20, no. 2, pp. 227–239, Mar. 1974.

[18] M. S. Pinsker, "Bounds on the probability and of the number of correctable errors for nonblock codes," *Problemy Peredachi Informatsii*, vol. 3, no. 4, pp. 44–55, Oct./Dec. 1967.