A bound for block codes with delayed feedback related to the sphere-packing bound

Harikrishna R Palaiyanur
Anant Sahai

Electrical Engineering and Computer Sciences
University of California at Berkeley

Technical Report No. UCB/EECS-2009-61
http://www.eecs.berkeley.edu/Pubs/TechRpts/2009/EECS-2009-61.html

May 12, 2009

Acknowledgement

NSF Grants CNS-403427 and CCF-729122 and an NSF Fellowship. This is partial work that benefits from ongoing discussions with Giacomo Como and Baris Nakiboglu at MIT.

A bound for block codes with delayed feedback related to sphere-packing bound

Hari Palaiyanur, Anant Sahai

This note gives an upper bound to the error exponent for block codes used over discrete memoryless channels (DMCs) with perfect feedback, where the feedback is delayed by some fixed number T of symbols. The result of this note is that the error exponent for rate R codes used with perfect feedback delayed by T symbols is upper bounded by (for T large enough)

$$E_{sp}(R - O(\log T/T)) + O(\log T/T), \tag{1}$$

where E_{sp} is the sphere-packing exponent and the constants depend on the channel transition matrix.

There are two reasons to consider this problem. The first is that in modern, high-rate communication systems, the number of symbols that must be encoded before the encoder receives a previous channel output (or more likely, a function of the channel output) can be potentially large. Two reasons for this gap between sending a channel input and receiving information about the channel output immediately come to mind: large propagation delays in wireless systems and inherent processing time for demodulation and other processing at the decoder. Consider communicating 20 symbols per microsecond on a 20 MHz channel over a distance of 1.5 km (round trip 3km). Even without accounting for processing time, the delay for a wave to travel to and fro would be 10 microseconds, meaning 200 symbols should have been transmitted in the meantime. Additionally, many communications systems have a half-duplex constraint, meaning they cannot listen and transmit at the same time and require some time to switch the context from talking to listening. Thus, feedback information may not return until the transmitter is finished transmitting some appropriate ‘block’ of symbols.

The second reason for examining delayed-feedback is strategic. It allows us to approach understanding the longstanding open problem of the feedback error-exponent for asymmetric DMCs.[1] principle (or trunking-principle in networks) holds that having access to parallel channels should scale up the capacity and error-exponents together. In other words, by encoding together T equal and independent flows together across T parallel channels, the error exponent should improve by a factor of T. However, T parallel channels with unit-delay feedback are clearly superior to 1 faster channel with T-step delayed feedback. Thus, if the parallel channel principle is true, then delaying the feedback should have no effect on the error exponent in the limit of large block-lengths that are much longer than the delay. So the result in this note establishes an interesting dichotomy: either the error-exponent for asymmetric DMCs is bounded by the classic sphere-packing bound (as everyone believes, but nobody can prove) or something about feedback interferes with the parallel channel principle. Feedback must be able to help a single channel alone in a way that it cannot help a group of channels working together. If this were true, it would be interesting indeed.

[1]In case this note finds its way into the hands of someone who has not already spent a great deal of time thinking about this problem, here's a brief recap of the story so far. For symmetric channels, we know that the error-exponent with feedback is limited by the sphere-packing bound. This was established by Berlekamp in his thesis, shows up as an exercise in Gallager's book for the BSC case, and was also proved by Dobrushin. Since the sphere-packing bound can be attained at high-rate without feedback, this established that for fixed block-lengths, even perfect feedback could only help at low rates, where indeed it was shown to help.

Haroutunian showed a bound that held with feedback for all channels, but while it equals the sphere-packing bound for symmetric channels, it is strictly greater for asymmetric channels. Haroutunian himself expressed mild frustration at not being able to close this gap. The form of Haroutunian's bound clearly suggests that it is loose, but it attempts to account for the fact that the encoder could change the input distribution in response to the channel feedback. Sheverdyaev had claimed a proof in PPI that finessed this problem, but it suffers from a “then a miracle occurs” type of step at a critical juncture.

This note grows out of an ongoing dialog between us at Berkeley and Baris Nakiboglu and Giacomo Como at MIT as we have attempted to resolve the issue.

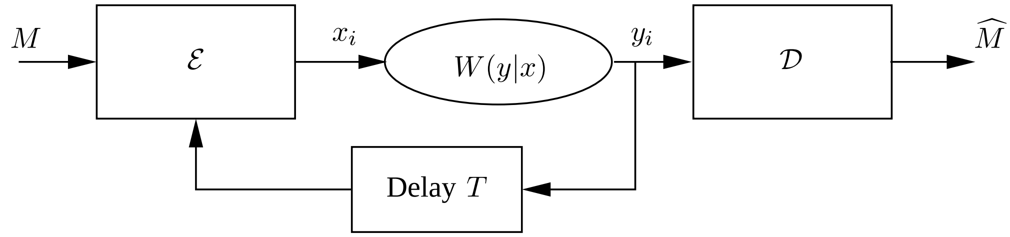

Fig. 1. Block coding for a DMC W with delayed feedback.

I. Problem Setup

The channel has finite input alphabet $\mathcal{X}$, finite output alphabet $\mathcal{Y}$ and known probability transition matrix $W(y|x)$. Fix a block length $n \geq 1$ and a feedback delay $T \geq 2$. The channel input at time $i \geq 1$ is $x_i \in \mathcal{X}$ and the channel output at time i is $y_i \in \mathcal{Y}$. A rate R, block length n coding system is an encoder-decoder pair $(\mathcal{E}, \mathcal{D})$.

Definition 1.1 (Type 1 Encoder - Delay T feedback): A rate R, block length n encoder $\mathcal{E}$ used with feedback delayed by T symbols is a sequence of maps $\{\phi_i\}_{i=1}^n$, with for $1 \leq i \leq T$,

$$\phi_i : \{1, 2, \ldots, 2^{nR}\} \to \mathcal{X}, \tag{2}$$

and for $i > T$,

$$\phi_i : \{1, 2, \ldots, 2^{nR}\} \times \mathcal{Y}^{i-T} \to \mathcal{X}. \tag{3}$$

Note that $T = 1$ is the usual perfect feedback setting where the encoder is aware of the channel output immediately before the next channel input must be selected.

Definition 1.2 (Type 2 Encoder - T symbol block feedback): This encoder is a more powerful class of encoding systems than the 'Delay T feedback' encoders. Here feedback is provided to encoder in blocks of T symbols at a time. That is, $(Y_1, \ldots, Y_T)$ is given to the encoder before the encoder chooses X_{T+1} and in general $(Y_{iT+1}, \ldots, Y_{(i+1)T})$ is provided to the encoder at time $(i+1)T$ before the encoder must choose $X_{(i+1)T+1}$. Hence a block length n type 2 encoder with rate R is a sequence of maps $\{\phi_i\}_{i=1}^n$

$$\phi_i : \{1, \ldots, 2^{nR}\} \times \mathcal{Y}^{\lfloor (i-1)/T \rfloor T} \to \mathcal{X} \tag{4}$$

Note that a type 1 encoder is also a type 2 encoder for a given $T \geq 1$.

A block length n decoder $\mathcal{D}$ is a map

$$\psi : \mathcal{Y}^n \to \{1, 2, \ldots, 2^{nR}\} \tag{5}$$

The decoding regions for each message are then $\psi^{-1}(m) = \{y^n : \psi(y^n) = m\}$ for $m \in \{1, \ldots, 2^{nR}\}$, where $y^n = (y_1, \ldots, y_n)$ represents a vector length n in $\mathcal{Y}^n$.

The random message M is selected from $\{1, 2, \ldots, 2^{nR}\}$ uniformly and we wish to calculate bounds on the average probability of error over all messages. The probability of error for a rate R, block length n, T-symbol feedback type 2 coding system $(\mathcal{E}, \mathcal{D})$ is defined as

$$\begin{aligned} P_e(n, \mathcal{E}, \mathcal{D}) &= \frac{1}{2^{nR}} \sum_{m=1}^{2^{nR}} Pr(\psi(Y^n) \neq m | M = m) \\ &= \frac{1}{2^{nR}} \sum_{m=1}^{2^{nR}} \sum_{y^n \notin \psi^{-1}(m)} \prod_{i=1}^{n} W\left(y_i \middle| \phi_i\left(m, y^{\lfloor (i-1)/T \rfloor T}\right)\right) \end{aligned} \tag{6, 7}$$

If we let $\mathcal{C}(n, R, T)$ be the set of block length n coding systems with rate at least equal to R used with delay T feedback type 2 encoders, we define the error exponent at rate R to be

$$E(R, T) = \limsup_{n \to \infty} -\frac{1}{n} \log \min_{(\mathcal{E}, \mathcal{D}) \in \mathcal{C}(n, R, T)} P_e(n, \mathcal{E}, \mathcal{D}) \tag{8}$$

Haroutunian showed the following result for channels with perfect feedback (for any $T \geq 1$),

$$E(R,T) \leq E_h(R) \triangleq \min_{V:C(V)\leq R} \max_p D(V||W|p) \tag{9}$$

In the above, the minimization is over channel transition matrices whose capacity is at most R, the maximization is over all probability mass functions p on the input alphabet $\mathcal{X}$, and $D(V||W|p)$ is the average divergence,

$$D(V||W|p) = \sum_{x\in\mathcal{X}} p(x) \sum_{y\in\mathcal{Y}} V(y|x) \log \frac{V(y|x)}{W(y|x)} \tag{10}$$

To ensure that continuity holds, we define $0\log 0 \triangleq 0$ and $1/0 \triangleq \infty$. Note that because the divergence is linear in p, the optimization in the $E_h(R)$ over p is attained at a corner point, which has no connection to the proportion with which an actual code uses that symbol over the channel.

The sphere packing bound, $E_{sp}(R)$ is defined as

$$E_{sp}(R) \triangleq \max_p \min_{V:I(p,V)\leq R} D(V||W|p) \tag{11}$$

where $I(p,V)$ is the mutual information of input distribution p across a channel V,

$$I(p,V) = \sum_{x\in\mathcal{X},y\in\mathcal{Y}} p(x)V(y|x)\log\frac{V(y|x)}{\sum_{x'\in\mathcal{X}} p(x')V(y|x')} \tag{12}$$

The sphere-packing bound serves as an upper bound to the error exponent for coding systems used with no feedback (i.e. $T=\infty$).

For suitably ‘output symmetric’ channels as defined by Haroutunian [1] and Gallager [2], $E_{sp}(R) = E_h(R)$ for all $R \geq 0$, however, in general, for asymmetric channels such as the binary ‘Z-channel’, $E_h(R) > E_{sp}(R)$. It is believed that the Haroutunian bound is loose and the sphere-packing bound should hold for block codes with perfect feedback (for any $T \geq 1$).

II. Result

Without loss of generality, we restrict attention to n that are multiples of T, that is $n = NT$ for some $N \geq 1$ (i.e. there are N total blocks of size T symbols each). Furthermore, we prove a bound on type 2 encoding systems, which immediately becomes a bound on type 1 encoders as well.

First, let $\mathcal{P}_T$ denote the set of types for $\mathcal{X}^T$ (as discussed in the book of Csiszar and Korner [3]). Second, for any given $p \in \mathcal{P}_T$, let $\mathcal{V}_T(p)$ be the set of V-shells associated with type p, i.e. the set of transition matrices for which $V(\cdot|x)$ is a type for $\mathcal{Y}^{Tp(x)}$ for each $x \in \mathcal{X}$.

Lemma 2.1: Define the channel independent constant

$$\alpha(T) \triangleq \frac{|\mathcal{X}|(2+|\mathcal{Y}|)\log(T+1)}{T} \tag{13}$$

Fix an $\epsilon > \alpha(T)$. Then, for any block length $n = NT$ (with $N \geq 1$) rate R coding system with a type 2 encoder,

$$-\frac{1}{NT}\log P_e(NT,\mathcal{E},\mathcal{D}) \leq \max_{p\in\mathcal{P}_T} \min_{V\in\mathcal{V}_T(p):I(p,V)\leq R-\epsilon} D(V||W|p) + \alpha(T) + \frac{1}{NT}\log\frac{1}{1-\exp(-NT(\epsilon-\alpha(T)))} \tag{14}$$

Proof: The argument begins, as with the sphere packing proof for codes without feedback, by showing that there is a roughly rate R code for which most codewords have the same input types. This is normally done by whittling down the messages to those whose codewords belong to the largest common type, by message population. We now have feedback every T symbols, so we need to carefully show in what way messages have the same input types. We do this by induction on N, the total number of T-length blocks.

First, for $N=1$, there has been no feedback. Let $p_1(m)$ denote the type of $x^T(m) \triangleq (\phi_1(m),\ldots,\phi_T(m))$. Now, group messages according to their type $p_1(m)$. Since $|\mathcal{P}_T|$ is at most $(T+1)^{|\mathcal{X}|}$, there exists a $p_1 \in \mathcal{P}_T$ such that

$$|\{m : p_1(m) = p_1\}| \geq \frac{2^{nR}}{(T+1)^{|\mathcal{X}|}} \tag{15}$$

This is the usual argument for fixing the composition of a high rate subcode in the proof of the sphere packing bound for codes used without feedback. After this, we choose a $V_1 \in \mathcal{V}_T(p_1)$ such that $I(p_1, V_1) \leq R - \epsilon$. The choice of V_1 is made so as to minimize $D(V_1||W|p_1)$ amongst those $V_1 \in \mathcal{V}_T(p_1)$ that have $I(p_1, V_1) \leq R - \epsilon$. The existence of a V_1 such that $I(p_1, V_1) \leq R - \epsilon$ is not immediately obvious for channels in which $E_{sp}(R)$ can be infinite, even if $E_{sp}(R)$ is not infinite for the given R. If no such V_1 exists, the result of the lemma is meaningless if we take the convention that $\min$ over the null set is ∞. Hence, if the optimization in the right hand side of (14) evaluates to something finite, we can safely assume the existence, for each p_1 of a V_1 with $I(p_1, V_1) \leq R - \epsilon$. For the rest of the proof, we assume we are in this case.

Without feedback, the selection at this point would be enough to show that for a high rate subcode with the same type, a substantial portion of the selected V_1-shells around the codewords for these messages overlap to cause a significant error. However, we now have feedback every T symbols, so we will iterate this selection process N times. We will prove a claim showing that the messages are not thinned too much and there are many y^n sequences which must 'overlap'. We will do this by selecting N input types $p_1, \ldots, p_N$ and N channel shells $V_1, \ldots, V_N$ sequentially and use induction. Let, for $1 \leq k \leq N$,

$$B^{(k)}(m) \triangleq \left\{ y^{kT} : y^{(k-1)T} \in B^{(k-1)}(m), p_k(m, y^{(k-1)T}) = p_k, y^{kT}_{(k-1)T+1} \in T_{V_k}\left(x^{kT}_{(k-1)T+1}(m, y^{(k-1)T})\right)\right\} \tag{16}$$

where $p_k(m, y^{(k-1)T})$ is the type of $x^{kT}_{(k-1)T+1}(m, y^{(k-1)T}) \triangleq \left(\phi_{(k-1)T+1}(m, y^{(k-1)T}), \ldots, \phi_{kT}(m, y^{(k-1)T})\right)$. Now, let for $1 \leq k \leq N$,

$$A^{(k)} \triangleq \left\{ m : |B^{(k)}(m)| \geq \frac{1}{(T+1)^{k(|\mathcal{X}|+|\mathcal{X}||\mathcal{Y}|)}} 2^{T\sum_{i=1}^{k} H(V_i|p_i)} \right\} \tag{17}$$

Note the dependence of both the $A^{(k)}$ and $B^{(k)}(m)$ sets on p_i and V_i. We drop the dependence in the notation for convenience.

Claim 2.1: For a block length $n = NT$, rate R type 2 encoder, there exists $p_1, \ldots, p_N \in \mathcal{P}_T$ and $V_1, \ldots, V_N$, with $V_i \in \arg\min_{V \in \mathcal{V}_T(p_i): I(p_i, V) \leq R-\epsilon} D(V||W|p_i)$ such that

$$|A^{(N)}| \geq \frac{2^{nR}}{(T+1)^{N|\mathcal{X}|}} \tag{18}$$

Proof: We proceed by induction with the base case of $N = k = 1$. As we have seen there is a $p_1 \in \mathcal{P}_T$ such that at least $2^{nR}/(T+1)^{|\mathcal{X}|}$ messages have $p_1(m) = p_1$. We then choose $V_1 \in \mathcal{V}_T(p_1)$ such that $I(p_1, V_1) \leq R - \epsilon$. It is clear that for those m with $p_1(m) = p_1$,

$$|B^{(1)}(m)| = |T_{V_1}(p_1)| \geq \frac{1}{(T+1)^{|\mathcal{X}||\mathcal{Y}|}} \exp_2(TH(V_1|p_1)) \tag{19}$$

where $|T_V(p)|$ denotes the number of vectors in a V-shell around a vector of type p (i.e. $|T_V(x^T)|$ if x^T is of type p). Hence, the claim is true for $N = 1$.

Now for $N = k > 1$, assume the claim is proved for $k-1$. For each $m \in A^{(k-1)}$, group the $y^{(k-1)T} \in B^{(k-1)}(m)$ according to $p_k(m, y^{(k-1)T})$. At least $|B^{(k-1)}(m)|/(T+1)^{|\mathcal{X}|}$ of the $y^{(k-1)T} \in B^{(k-1)}(m)$ have a common type $p_k(m, y^{(k-1)T}) = p_k(m)$. Now, group the messages in $A^{(k-1)}$ according to $p_k(m)$. At least $|A^{(k-1)}|/(T+1)^{|\mathcal{X}|}$ have a common type $p_k(m) = p_k$. Now select a $V_k \in \mathcal{V}_T(p_k)$ such that $I(p_k, V_k) \leq R - \epsilon$. It is now readily seen that for the $m \in A^{(k-1)}$ with $p_k(m) = p_k$,

$$|B^{(k)}(m)| \geq \frac{|B^{(k-1)}(m)|}{(T+1)^{|\mathcal{X}|}} |T_{V_k}(p_k)| \geq \frac{1}{(T+1)^{k(|\mathcal{X}|+|\mathcal{X}||\mathcal{Y}|)}} \exp_2\left(T\sum_{i=1}^{k} H(V_i|p_i)\right) \tag{20}$$

This holds for at least $|A^{(k-1)}|/(T+1)^{|\mathcal{X}|} \geq 2^{nR}/(T+1)^{k|\mathcal{X}|}$ messages, hence the claim is true. ■

Now, note that for all $y^n \in B^{(N)}(m)$ with $m \in A^{(N)}$, we also have $y^{iT}_{(i-1)T+1} \in T_{p_i V_i}$ for all $1 \leq i \leq N$. Hence,

$$\forall\, m,\; B^{(N)}(m) \subset T_{p_1 V_1} \times \cdots \times T_{p_N V_N} \tag{21}$$

$$\therefore |\cup_{m \in A^{(N)}} B^{(N)}(m)| \leq |T_{p_1 V_1}| \times \cdots \times |T_{p_N V_N}| \tag{22}$$

$$\leq \exp_2\left(T\sum_{i=1}^{N} H(p_i V_i)\right) \tag{23}$$

Finally, putting it all together,

$$
\begin{aligned}
P_e(n,\mathcal{E},\mathcal{D}) &= \frac{1}{2^{nR}}\sum_{m=1}^{2^{nR}}\sum_{y^n\notin\psi^{-1}(m)}\prod_{i=1}^{n} W\left(y_i\middle|\phi_i\left(m,y^{\lfloor i/T\rfloor T}\right)\right) && (24)\\
&\geq \frac{1}{2^{nR}}\sum_{m\in A^{(N)}}\sum_{y^n\in\overline{\psi^{-1}(m)}\cap B^{(N)}(m)}\prod_{i=1}^{n} W\left(y_i\middle|\phi_i\left(m,y^{\lfloor i/T\rfloor T}\right)\right) && (25)\\
&= \frac{1}{2^{nR}}\sum_{m\in A^{(N)}}\sum_{y^n\in\overline{\psi^{-1}(m)}\cap B^{(N)}(m)}\exp_2\left(-T\sum_{i=1}^{N}(D(V_i||W|p_i)+H(V_i|p_i))\right) && (26)\\
&= \frac{1}{2^{nR}}\sum_{m\in A^{(N)}}|\overline{\psi^{-1}(m)}\cap B^{(N)}(m)|\exp_2\left(-T\sum_{i=1}^{N}(D(V_i||W|p_i)+H(V_i|p_i))\right) && (27)\\
&= \frac{\exp_2\left(-T\sum_{i=1}^{N}(D(V_i||W|p_i)+H(V_i|p_i))\right)}{2^{nR}}\sum_{m\in A^{(N)}}(|B^{(N)}(m)|-|B^{(N)}(m)\cap\psi^{-1}(m)|) && (28)\\
&\geq \frac{\exp_2\left(-T\sum_{i=1}^{N}(D(V_i||W|p_i)+H(V_i|p_i))\right)}{2^{nR}}\left[\left(\sum_{m\in A^{(N)}}|B^{(N)}(m)|\right)-|\cup_{m\in A^{(N)}}B^{(N)}(m)|\right]\\
&\geq \frac{\exp_2\left(-T\sum_{i=1}^{N}(D(V_i||W|p_i)+H(V_i|p_i))\right)}{2^{nR}}\times\\
&\qquad\left(|A^{(N)}|\frac{1}{(T+1)^{N(|\mathcal{X}|+|\mathcal{X}||\mathcal{Y}|)}}2^{T\sum_{i=1}^{N}H(V_i|p_i)}-|\cup_{m\in A^{(N)}}B^{(N)}(m)|\right) && (29)\\
&\overset{(a)}{\geq} \exp_2\left(-T\sum_{i=1}^{N}D(V_i||W|p_i)\right)\times\\
&\qquad\left[\frac{1}{(T+1)^{N(2+|\mathcal{Y}|)|\mathcal{X}|}}-\exp_2\left(T\sum_{i=1}^{N}(H(p_iV_i)-H(V_i|p_i)-R)\right)\right] && (30)\\
&\geq \exp_2\left(-T\sum_{i=1}^{N}D(V_i||W|p_i)\right)\left[\frac{1}{(T+1)^{N(2+|\mathcal{Y}|)|\mathcal{X}|}}-\exp_2\left(-NT\epsilon\right)\right] && (31)\\
&= \frac{\exp_2\left(-T\sum_{i=1}^{N}D(V_i||W|p_i)\right)}{\exp_2(NT\alpha(T))}\left[1-\exp_2\left(-NT(\epsilon-\alpha(T))\right)\right] && (32)
\end{aligned}
$$

In inequality (a), we have used Claim 2.1 and the inequality of equation (23). In the selection process of the claim, for each $p_i\in\mathcal{P}_T$, we choose a $V_i\in\mathcal{V}_T(p)$ with $I(p_i,V_i)\leq R-\epsilon$ that minimizes the average divergence. Then, since we can't say anything about the p_i, we bound by the worst case p to take a max over all $p\in\mathcal{P}_T$. Taking logs and dividing by NT gives the result of the lemma. ∎

We now give an inequality relating the 'sphere packing bound for length-T' with the sphere packing bound.

Lemma 2.2: For any $T\geq 2|\mathcal{X}||\mathcal{Y}|$, for all $p\in\mathcal{P}_T$,

$$\min_{U\in\mathcal{V}_T(p):I(p,U)\leq R} D(U||W|p)\leq E_{sp}\left(p,R-\frac{2|\mathcal{X}||\mathcal{Y}|\log T}{T}\right)+\frac{\kappa|\mathcal{X}||\mathcal{Y}|}{T}+\frac{|\mathcal{X}||\mathcal{Y}|\log(T/|\mathcal{X}|)}{T} \quad (33)$$

where

$$E_{sp}(p,R) \triangleq \min_{V:I(p,V)\leq R} D(V||W|p), \quad (34)$$

$$\kappa \triangleq \max_{x,y:W(y|x)>0}\log\frac{1}{W(y|x)}. \quad (35)$$

Proof: First, write $p\in\mathcal{P}_T$ as $p(x)=k_x/T$ where k_x are nonnegative integers that sum to T.

Claim 2.2: Let U be an arbitrary channel for which $|U(y|x) - V(y|x)| \leq 1/k_x$ for all x, y, and $U(y|x) = V(y|x) = 0$ when $W(y|x) = 0$. Then,

$$|D(U||W|p) - D(V||W|p)| \leq \frac{\kappa|\mathcal{X}||\mathcal{Y}|}{T} + \frac{|\mathcal{X}||\mathcal{Y}|\log(T/|\mathcal{X}|)}{T} \tag{36}$$

Proof:

First, note that $|r\log r - s\log s| \leq -|r-s|\log|r-s|$ whenever $r, s \in [0,1]$. This can be seen by noting that the function $f(r) = -r\log r, r \in [0,1]$ is concave and maximal absolute slope at $r = 0$, where the derivative is unbounded above. Hence, $|f(r) - f(0)| = -r\log r, r \in [0,1]$ is a bound to the difference between two points on the curve at distance r. Now, keeping in mind that $U(y|x) = V(y|x) = 0$ whenever $W(y|x) = 0$,

$$\begin{aligned}
|D(U||W|p) - D(V||W|p)| &\leq \sum_x p(x) \sum_y \left| U(y|x)\log\frac{U(y|x)}{W(y|x)} - V(y|x)\log\frac{V(y|x)}{W(y|x)} \right| &(37)\\
&\leq \sum_x \frac{k_x}{T} \sum_y |U(y|x)\log U(y|x) - V(y|x)\log V(y|x)| + \\
&\qquad \sum_x \frac{k_x}{T} \sum_{y:W(y|x)>0} |U(y|x) - V(y|x)| \log\frac{1}{W(y|x)} &(38)\\
&\leq \sum_x \frac{k_x}{T} \sum_y \left[\frac{1}{k_x}\log k_x + \frac{1}{k_x}\kappa\right] &(39)\\
&\leq \frac{|\mathcal{X}||\mathcal{Y}|\kappa}{T} + \frac{|\mathcal{Y}|}{T}\sum_x \log k_x &(40)\\
&\overset{(a)}{\leq} \frac{|\mathcal{X}||\mathcal{Y}|\kappa}{T} + \frac{|\mathcal{X}||\mathcal{Y}|\log T/|\mathcal{X}|}{T} &(41)
\end{aligned}$$

In (a), we are using the fact that since $\log$ is a concave function,

$$\max_{k_x\in\mathbb{N}:\sum_x k_x = T} \sum_x \log k_x \leq \max_{k_x\in\mathbb{R}_{\geq 0}:\sum_x k_x = T} \sum_x \log k_x = |\mathcal{X}|\log\frac{T}{|\mathcal{X}|} \tag{42}$$

■

Now, for an $\epsilon > 0$, pick V to be in $\arg\min_{V':I(p,V')\leq R-\epsilon} D(V'||W|p)$. We will find a $U \in \mathcal{V}_T(p)$ such that $I(p,U) \leq I(p,V) + \epsilon \leq R$. First, we show that there exists a $U \in \mathcal{V}_T(p)$ such that $|U(y|x) - V(y|x)| \leq \frac{1}{k_x}$ for all x, y.

For each x, y, let $\widetilde{U}(y|x) = \lfloor k_x V(y|x) \rfloor / k_x$. Note that $\widetilde{U}$ is missing some mass to be a transition matrix if the entries of $V(\cdot|x)$ are not multiples of $1/k_x$. The missing mass can be bounded, for a fixed x,

$$\begin{aligned}
1 - \sum_y \widetilde{U}(y|x) &= \sum_y V(y|x) - \widetilde{U}(y|x) = \sum_y k_x V(y|x)\frac{1}{k_x} - \lfloor k_x V(y|x)\rfloor \frac{1}{k_x} &(43)\\
&\leq \sum_{y:k_x V(y|x)\notin\mathbb{Z}} 1/k_x = |\{y : k_x V(y|x) \notin \mathbb{Z}\}|/k_x &(44)
\end{aligned}$$

Now, the missing mass must be a multiple of $1/k_x$ for each x because $\widetilde{U}(\cdot|x)$ has terms that are multiples of $1/k_x$. Therefore, the missing mass can be distributed amongst the y that have $k_x V(y|x) \notin \mathbb{Z}$ in multiples of $1/k_x$ in such a way so that no y has more than $1/k_x$ mass added to it. We let $U(y|x)$ be the resulting transition matrix. Since $\widetilde{U}(y|x) = \lfloor k_x V(y|x)\rfloor / k_x$ and either 0 or $1/k_x$ is added to get to $U(y|x)$, it follows that $|V(y|x) - U(y|x)| \leq 1/k_x$. Also, $U(y|x) = 0$ when $V(y|x) = 0$.

Note that

$$\sum_x p(x) \sum_y |U(y|x) - V(y|x)| \leq \sum_x \frac{k_x}{T} \sum_y \frac{1}{k_x} \leq \frac{|\mathcal{X}||\mathcal{Y}|}{T}. \tag{45}$$

If $T \geq 2|\mathcal{X}||\mathcal{Y}|$, we can use the continuity lemma for entropy in Cover and Thomas [[4], Lemma 16.3.2], that is $|H(p) - H(q)| \leq -\|p-q\|_1 \log(\|p-q\|_1/|\mathcal{X}|)$ if $\|p-q\|_1 \leq 1/2$. By using this lemma twice after expanding the mutual information, we get

$$|I(p,U) - I(p,V)| \leq |H(pU) - H(pV)| + |H(p,U) - H(p,V)| \leq \frac{2|\mathcal{X}||\mathcal{Y}|}{T} \log T. \tag{46}$$

Hence,

$$I(p,U) \leq R - \epsilon + \frac{2|\mathcal{X}||\mathcal{Y}|}{T} \log T \leq R \tag{47}$$

provided

$$\frac{2|\mathcal{X}||\mathcal{Y}|}{T} \log T \leq \epsilon. \tag{48}$$

Therefore, there exists a $U \in \mathcal{V}_T(p)$, with $I(p,U) \leq R$ such that

$$D(U||W|p) = E_{sp}(p, R-\epsilon) + D(U||W|p) - D(V||W|p) \tag{49}$$

$$\leq E_{sp}\left(p, R - \frac{2|\mathcal{X}||\mathcal{Y}|}{T}\log T\right) + \frac{\kappa|\mathcal{X}||\mathcal{Y}|}{T} + \frac{|\mathcal{X}||\mathcal{Y}|}{T} \log \frac{T}{|\mathcal{X}|} \tag{50}$$

■

Putting the two lemmas together, we get

Theorem 2.1: Fix a $T \geq 2|\mathcal{X}||\mathcal{Y}|$, and consider a sequence of block length NT, rate $R_N \geq R$ type 2 coding systems $(\mathcal{E}_N, \mathcal{D}_N)_{N=1}^{\infty}$ with feedback update time T.

$$\lim_{N\to\infty} -\frac{1}{NT} \log P_e(NT, \mathcal{E}_N, \mathcal{D}_N) \leq E_{sp}\left(R - \alpha(T) - \frac{2|\mathcal{X}||\mathcal{Y}|}{T}\log T\right) + \frac{\kappa|\mathcal{X}||\mathcal{Y}|}{T} + \frac{|\mathcal{X}||\mathcal{Y}|}{T}\log\frac{T}{|\mathcal{X}|} + \alpha(T) \tag{51}$$

In other words, for type 1 and type 2 coding systems with feedback delay T and rate R,

$$E(R,T) \leq E_{sp}(R - O(\log T/T)) + O(\log T/T) \tag{52}$$

Proof: The result of lemma 2.1 is monotonic in the rate, so we need only that $R_N \geq R$ and bound using R. We have just combined the results of the two lemmas together and taken the limit as N tends to ∞. The only thing that needs to be checked is that the term

$$\frac{1}{NT} \log \frac{1}{1 - \exp_2(-NT(\epsilon - \alpha(T)))} \tag{53}$$

converges to 0 as $N \to \infty$ for any $\epsilon > \alpha(T)$. First, note that $\frac{1}{NT}$ converges to 0 as N tends to ∞ and $\frac{1}{1-\exp_2(-NT(\epsilon-\alpha(T)))}$ tends to 1 from above as N tends to ∞. Therefore, $\log \frac{1}{1-\exp_2(-NT(\epsilon-\alpha(T)))}$ is tending to 0 as well and the product in (53) is going to 0 with N. Therefore, can take ϵ arbitrarily close to $\alpha(T)$ and since E_{sp} is continuous for all R except possibly the R at which E_{sp} becomes infinite, we substitute $\alpha(T)$ for ϵ. ■

References

[1] E. Haroutunian, "Lower bound for error probability in channels with feedback," *Problemy Peredachi Informatsii*, pp. 36–44, 1977.

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

[3] I. Csiszar and J. Korner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*, 2nd ed. New York, NY: Academic Press, 1997.

[4] T. Cover and J. Thomas, *Elements of Information Theory*. New York, NY: John Wiley and Sons, 1991.