Information Transmission over the Postal Channel with and without Feedback

Wenyi Zhang, Shivaprasad Kotagiri, and J. Nicholas Laneman
Department of Electrical Engineering, University of Notre Dame
Email: {wzhang1, skotagir, jnl}@nd.edu

***Abstract*— The postal channel models a postal system in which letters, each consisting of a number of characters, are sometimes lost. We study the postal channel with variable-length letters and variable-length coding over letters, both with and without letter-by-letter feedback. Without allowing letter lengths to encode information, we examine one feedback strategy consisting of automatic repeat-request (ARQ) with exponentially increasing letter lengths. For this strategy we investigate an alternative notion of information rate per character, based upon the total, random number of characters required to convey the messages instead of its expectation. This information rate exhibits a phase transition in its convergence as the number of messages becomes large: if the letter lengths increase by a factor less than the inverse of the probability that a letter is lost, it converges to the channel capacity; otherwise, it converges to a number strictly larger than channel capacity. More generally, when we allow both the characters and the length of a letter to convey information, we compute the corresponding channel capacity with and without feedback, and find that it is twice the channel capacity of the original postal channel without allowing letter lengths to encode information.**

I. Introduction

The postal channel was introduced by Wolf, Wyner, and Ziv to model a postal service that occasionally loses letters.[1] The description of the original channel model is as follows [1]:

"The postal channel considered here consists of two parts, a letter writer and a postal service. The letter writer can inscribe (say) binary digits at a rate of one per second, and cut the binary stream into letters of length m which are then mailed via the postal service. The postal service delivers the letters to the destination, occasionally losing letters. We assume that letters are lost independently with probability p ($0 \leq p \leq 1$). The user is allowed to choose the parameter m freely, and we assume that m is known to the destination, as well as to the transmitter."

For this model, the channel capacity is shown in [1] to be $C = 1 - p$, which is the same as the capacity of a memoryless binary erasure channel (BEC) with erasure probability p [2, Sec. 8.1.5]. As described above, the original postal channel model has three basic constraints: first, the letter length m is fixed and identical for all the letters; second, the receiver knows m in advance so the letter length does not convey information; and third, there is no feedback.

In this paper, we extend the model to study variable-length letters and variable-length coding over letters, with and without letter-by-letter feedback. As we will see, if we do not allow letter lengths to convey information, the channel capacity C does not increase with either variable-length coding over letters or letter-by-letter feedback. However, interpreting information rates for variable-length encoding is more subtle, as the example of *exponential automatic repeat request (ARQ)* illustrates in Section II. Finally, coding information into letter lengths increases channel capacity to $2C$, with or without feedback, as Section III details.

II. Exponential ARQ

In this section, we consider different interpretations of information rate in the context of a feedback strategy we call *exponential ARQ*, involving ARQ [3, Chap. 22] with exponentially increasing letter lengths.

To set the stage for our discussion, consider a situation in which there are Q separate trials of variable-length transmission over a channel, each transmitting a message $\mathsf{w} \in \{1, 2, \ldots, M\}$. Here we may have two ways to quantify the "average information rate per character":

$$\frac{\log_2 M}{\frac{1}{Q}\sum_{i=1}^{Q} \tilde{\mathsf{n}}_i}, \tag{1}$$

and

$$\frac{1}{Q}\sum_{i=1}^{Q} \frac{\log_2 M}{\tilde{\mathsf{n}}_i}, \tag{2}$$

where $\tilde{\mathsf{n}}_{i=1,\ldots,Q}$ is the realized number of channel uses for the i-th cycle. The former quantity corresponds to the usual definition of information rate (and channel capacity) as $M \to \infty$ and $Q \to \infty$[6]. The latter quantity, however, has received little attention in information theory.

For many scenarios the two quantities lead to the same asymptotic value as $M \to \infty$. Fixed-length coding is a trivial example. For variable-length coding, consider a BEC with erasure probability p, for which a simple but asymptotically optimal transmission scheme is persistently retransmitting each erased bit until it is successfully received [2, Sec. 8.1.5]. Each bit consumes a random number of channel uses to receive, and let us call these channel uses one "cycle". The length of a cycle for the i-th bit, k_i, is geometrically distributed, i.e., $\mathsf{k}_i = k$ with probability $(1-p)p^{k-1}$, $k = 1, 2, \ldots$. Thus, transmitting a message with $n = \log_2 M$ bits requires a total number of channel uses $\tilde{\mathsf{n}} = \sum_{i=1}^{n} \mathsf{k}_i$, i.e., the sum of n independent,

[1] Throughout the paper, "letter" always refers to the postal letter, rather than characters within a letter.

identically distributed (i.i.d.) k_i. As $n \to \infty$, by the strong law of large numbers, the above two definitions of information rate coincide with probability one. That is, for transmitting a long block of information bits, it is typical for the number of channel uses to be "roughly" its expectation.

As we develop in the remainder of this section, the situation outlined above changes dramatically for the postal channel with variable-length coding and letter-by-letter feedback. Section II-A describes the scheme, Section II-B develops an analysis of different information rates, Section II-C discusses subtleties of the analysis, and Section II-D presents numerical results.

A. Description

Here we describe one cycle of the exponential ARQ feedback strategy. The transmitter first transmits a letter of one character, containing a single message bit. If the transmitter is informed via feedback that this letter is erased, instead of simply retransmitting the erased bit, it concatenates $L-1$ new message bits such that its next letter has length $L \geq 1$. Recursively, when the k-th retransmission occurs, the letter has length L^k, with L^{k-1} previously transmitted message bits, and $L^k - L^{k-1}$ new message bits. Whenever a successful reception occurs, the cycle terminates. Figure 1 illustrates one such cycle for $L = 2$.

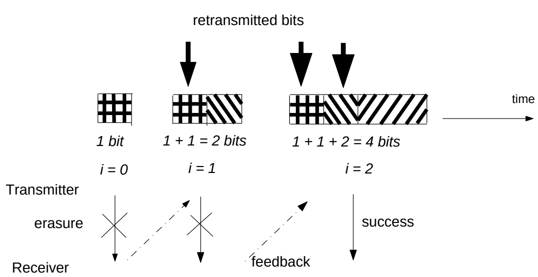


Fig. 1. Illustration of one cycle in the exponential ARQ feedback strategy for $L = 2$.

For transmitting a message with a certain finite length $n = \log_2 M$, the above exponential ARQ feedback strategy should be truncated accordingly when the number of remaining message bits are insufficient to fill in a long retransmission letter. In that situation, we simply stop increasing the letter length and retransmit the remaining message bits until they are successfully received.

B. Analysis

Let us analyze the behavior of the exponential ARQ feedback strategy. In the present analysis we assume that the message has an infinite number of bits, and we will soon see that this assumption gives rise to rather subtle convergence issues. Consider one cycle consisting of k retransmissions. That is, L^k message bits are successfully transmitted within

$$1 + L + L^2 + \ldots + L^k = \frac{L^{k+1} - 1}{L - 1}$$

channel uses.

For any erasure probability p strictly less than one, i.e., as long as the letters are not erased with probability one, since the postal channel is letter-by-letter memoryless, the number of retransmissions in a cycle is a geometric random variable k, with $P(\mathsf{k} = k) = p^k(1-p)$, $k = 0, 1, \ldots$. Consider J consecutive cycles, with their numbers of retransmissions denoted by k_j, $j = 1, \ldots, J$. Hence there are $\sum_{j=1}^{J} \frac{L^{\mathsf{k}_j+1}-1}{L-1}$ channel uses, and $\sum_{j=1}^{J} L^{\mathsf{k}_j}$ transmitted message bits. We define the empirical information rate as their ratio

$$\begin{aligned} \mathsf{r}(J, L) &= \frac{\sum_{j=1}^{J} L^{\mathsf{k}_j}}{\sum_{j=1}^{J} (L^{\mathsf{k}_j+1} - 1)/(L-1)} \\ &= \frac{L-1}{L - 1/\left(\frac{1}{J}\sum_{j=1}^{J} L^{\mathsf{k}_j}\right)}. \end{aligned} \quad (3)$$

Here note that $\mathsf{r}(J, L)$ is merely a random variable induced by $\{\mathsf{k}_j\}_{j=1}^{J}$. The behavior of $\mathsf{r}(J, L)$ as $J \to \infty$ is determined by the random variable L^{k}.

- If $L < 1/p$, L^{k} has bounded expectation, from the strong law of large numbers,

  $$\lim_{J\to\infty} \frac{1}{J}\sum_{j=1}^{J} L^{\mathsf{k}_j} = \mathbf{E}[L^{\mathsf{k}}] = \frac{1-p}{1-Lp} \quad \text{w.p. } 1. \quad (4)$$

  Consequently,

  $$\mathsf{r}(L) = \lim_{J\to\infty} \mathsf{r}(J, L) = 1 - p \quad \text{w.p. } 1, \quad (5)$$

  which is precisely C as we summarize in Section I.
- If $L \geq 1/p$, L^{k} has unbounded expectation. By a corollary of the strong law of large numbers (see, *e.g.*, [4, Chap. 4, Sec. 22]), for any $L \geq 1/p$ and any arbitrarily large $A > 0$,

  $$\lim_{J\to\infty} \frac{1}{J}\sum_{j=1}^{J} L^{\mathsf{k}_j} \geq A \quad \text{w.p. } 1. \quad (6)$$

  Consequently,

  $$\mathsf{r}(L) = \lim_{J\to\infty} \mathsf{r}(J, L) = \frac{L-1}{L} = 1 - \frac{1}{L} \quad \text{w.p. } 1. \quad (7)$$

  Here note that the random variable $\mathsf{r}(L)$ can actually be viewed as an almost surely deterministic quantity. Since $L \geq 1/p$ can be chosen arbitrarily large, for any small $\delta > 0$ we can choose $L = \lceil 1/\delta \rceil$ such that $1 - 1/L \geq 1 - \delta$. Therefore we have

  $$R := \lim_{L\to\infty} \mathsf{r}(L) = 1 \quad \text{w.p. } 1. \quad (8)$$

Remarks:

(1) The exponential ARQ feedback strategy can further be applied to the case where the letter erasure probability depends upon letter lengths. Assume that erasures remain independent among letters, and for a letter of length m, its erasure probability is a function of m denoted by $p(m), m = 1, \ldots$. The above analysis remains unchanged until (3). The subsequent steps differ since the empirical average $(1/J)\sum_{j=1}^{J} L^{\mathsf{k}_j}$ is now taken with respect to a non-geometric distribution. However,

from the strong law of large numbers we have that (8) holds if and only if one of the following two conditions holds:

1) $\lim_{L\to\infty} \mathbf{E}\left[L^{\mathsf{k}}\right] = \infty$;
2) $\lim_{L\to\infty} \mathbf{E}\left[L^{\mathsf{k}}\right] < \infty$, but $\lim_{L\to\infty} L\mathbf{E}\left[L^{\mathsf{k}}\right] = \infty$.

Since $\mathbf{E}\left[L^{\mathsf{k}}\right] \geq 1$, the above two conditions actually exhaust all possibilities. Thus, for any erasure probability $p(m) < 1, m = 1, \ldots$, *e.g.*, BEC, (8) always holds.

(2) Let us impose a restriction on the scheme such that, if after $\bar{k}$ retransmissions the letter is still erased, then the transmitter stops increasing the retransmitted letter length. It can be shown that, for any finite $\bar{k}$, the truncated exponential ARQ scheme only achieves

$$R = \lim_{L\to\infty} \mathsf{r}(L) = 1 - p = C \quad \text{w.p. } 1, \tag{9}$$

instead of (8).

C. Discussion

There lies a subtlety in the preceding analysis of the exponential ARQ feedback strategy. The asymptotic behavior of $\mathsf{r}(L) = \lim_{J\to\infty} \mathsf{r}(J, L)$ in fact indicates that there exists *a sequence* in the number of messages such that the information rate evaluated *for that sequence* approaches $\mathsf{r}(L)$ asymptotically. However, since such a sequence is *randomly* generated by the feedback decisions, there is no guarantee that for any *deterministic* number of messages M the rate $\mathsf{r}(L)$ can be approached with high probability. By contrast, the Shannon-theoretic channel capacity essentially requires that for *every* sufficiently large M the rate C is approached with high probability. The "empirical information rate" of the exponential ARQ feedback strategy therefore is not a Shannon-theoretic quantity hence does not contradict the channel capacity C, but it provides an alternative perspective of information transmission for channels with variable-length coding and feedback.

The seemingly unreasonable behavior of (8), *i.e.*, information transmission at one bit per channel use, stems from the key assumption that there is an infinite number of message bits. Transmission never stops, thus even if all the previously transmitted letters were erased, just one subsequent successful retransmission is sufficient to compensate for all those erasures. For the postal channel this transmission strategy does lead to some interesting phenomena even for transmitting finite-length messages, as will be illustrated numerically in the next section.

It may worth mentioning that there exists another alternative perspective of the channel capacity, for multiuser communication [5, Sec. V], that shares a certain similarity with the above behavior of exponential ARQ. Consider a memoryless broadcast channel with one input x and two outputs $\mathsf{y}_1, \mathsf{y}_2$. The transmitter uses its first $n_1^{(1)}$ channel uses to encode a message for y_1, so for sufficiently large $n_1^{(1)}$ an information rate of $C_1 = \max_{p(x)} I(\mathsf{x}; \mathsf{y}_1)$ can be arbitrarily closely approached. Then the transmitter uses the next $n_2^{(1)}$ channel uses to encode a message for y_2, so for sufficiently large $n_2^{(1)} \gg n_1^{(1)}$ an information rate of $C_2 = \max_{p(x)} I(\mathsf{x}; \mathsf{y}_2)$ can be arbitrarily closely approached within the total $n_1^{(1)} + n_2^{(1)}$ channel uses. Recursively, we simply keep increasing the coding block lengths such that for each receiver there exists a sequence of time instants at which its single-user channel capacity is achieved. This multiuser coding scheme and the preceding exponential ARQ feedback strategy are similar in that neither of them can achieve the information rate for every sufficiently large number of messages, as required by the Shannon-theoretic channel capacity.

D. Numerical Results

In this section we consider the transmission of a long, but finite-length, block of message bits by the exponential ARQ feedback strategy. Unfortunately, an analysis is less easy to develop because, for $L \geq 1/p$, the expected number of transmitted message bits in one single cycle, $\mathbf{E}[L^{\mathsf{k}}]$, is unbounded. Hence no matter how long the message is, it is always "short" compared to $\mathbf{E}[L^{\mathsf{k}}]$. In this section we therefore rely on numerical simulation to draw some preliminary observations.

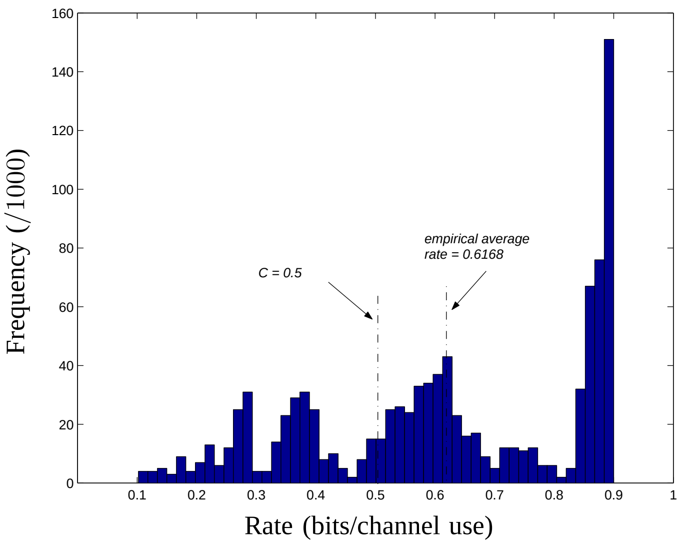
In the simulation, we specify the channel erasure probability p, the length of the message B, and the parameter L in the exponential ARQ scheme. Since $B < \infty$, it is possible that the remaining message bits are insufficient to fill in a long retransmitted letter. When that situation occurs, we follow the procedure described in Section II-A.

Figure 2 displays two typical histograms of the empirical distribution of the information rates achieved by exponential ARQ for $L > 1/p$ and $L < 1/p$, respectively.

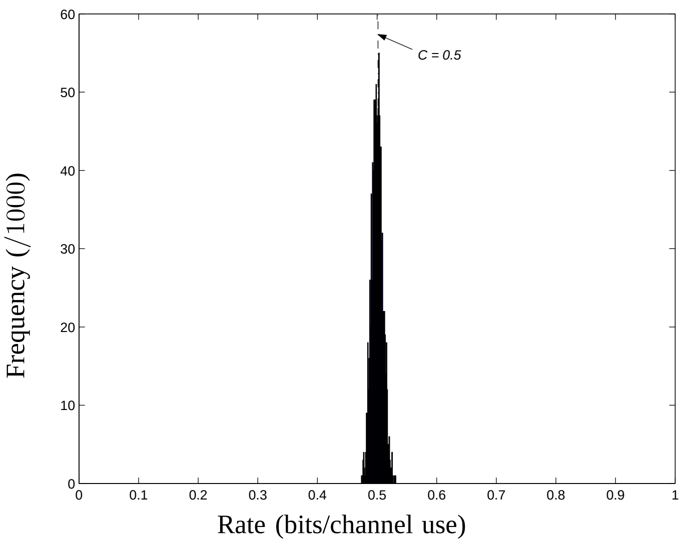
In Figure 2(a) for $L > 1/p$ we observe a spectrum of achieved rates, instead of a tight cluster around $C = 1 - p = 0.5$ bits per channel use as in Figure 2(b) for $L < 1/p$. As indicated in Figure 2(a), the discrepancy between the two definitions of information rate, (1) and (2), is clear. The empirical average rate corresponding to (2) is around 0.62 bits per channel use, and around 70% of runs outperform (1), which corresponds to C.

Another observation from the numerical simulation is that there is an approximate "scale-free" property of the achieved rates. That is, for a wide range of the message block size B, the statistics of the achieved rates are approximately comparable. This behavior is illustrated in Table I, in which we tabulate the mean and variance of the achieved rates, as well as the percentage that the achieved rates outperform C, for B ranging from 10^3 to 10^7. We can see that for each L, these statistics are relatively stable across such a wide range of B.

Finally let us turn to the simulation of changing the erasure probability p. We tabulate in Table II the statistics of the achieved rates for p ranging from 0.1 to 0.9. In the simulation we let $L = 10$, and the message block size B be according to $\text{Uniform}[1, 2] \times 10^5$. Again we run the simulation 1000 times to obtain these statistics. From Table II we observe that the empirical average rate always outperforms C. For lossier channels, *i.e.*, channels with larger p, such a "rate gain" becomes more noticeable. As p deceases, this discrepancy turns out to be only marginal.



(a) $L = 10 > 1/p$



(b) $L = 1 < 1/p$

Fig. 2. Typical histograms of the empirical distribution of the information rate achieved by exponential ARQ. We let $p = 0.5$, $B \sim \text{Uniform}[1, 2] \times 10^5$, and $L = 1$ or 10, and run the simulation $Q = 1000$ times.

III. The Extended Postal Channel

In the original postal channel model, the letter length m is fixed and does not convey information. In this section we extend the model to incorporate variable-length letters, variable-length coding blocks, and possibly feedback. With an appropriate formulation of the model, we are able to state the results as a special case of [6].

The extended channel model is described as follows. The i-th channel input x_i is a letter of length $\mathsf{s}_i \in \{1, 2, \ldots\}$ bits, i.e., x_i is a vector in $\{0, 1\}^{\mathsf{s}_i}$. Hence the channel input alphabet is the union of all non-null binary strings, denoted by $\{0, 1\}^*$. The corresponding channel output symbol y_i is taken from the alphabet $\{0, 1\}^* \bigcup \{E\}$, where E represents an erasure. The

L	B	Mean(rate)	Var(rate)	%(rate $\geq C$)
5	$\text{Uniform}[1, 2] \times 10^3$	0.5785	0.0341	68%
	$\text{Uniform}[1, 2] \times 10^5$	0.5814	0.0339	70%
	$\text{Uniform}[1, 2] \times 10^7$	0.5877	0.0351	69%
10	$\text{Uniform}[1, 2] \times 10^3$	0.6170	0.0501	71%
	$\text{Uniform}[1, 2] \times 10^5$	0.6249	0.0502	72%
	$\text{Uniform}[1, 2] \times 10^7$	0.5940	0.0500	67%
15	$\text{Uniform}[1, 2] \times 10^3$	0.6457	0.0554	63%
	$\text{Uniform}[1, 2] \times 10^5$	0.6126	0.0481	65%
	$\text{Uniform}[1, 2] \times 10^7$	0.6324	0.0598	70%
20	$\text{Uniform}[1, 2] \times 10^3$	0.6189	0.0491	65%
	$\text{Uniform}[1, 2] \times 10^5$	0.6584	0.0749	65%
	$\text{Uniform}[1, 2] \times 10^7$	0.6185	0.0530	61%

TABLE I

The behavior of the exponential ARQ for different L, $p = 0.5$, and $Q = 1000$ times of simulation.

p	C_{PF}	Mean(rate)	Var(rate)	%(rate $\geq C$)
0.1	0.9	0.9065	0.0024	84%
0.2	0.8	0.8258	0.0181	80%
0.3	0.7	0.7602	0.0317	68%
0.4	0.6	0.6882	0.0435	65%
0.5	0.5	0.6232	0.0512	73%
0.6	0.4	0.5165	0.0566	59%
0.7	0.3	0.4421	0.0566	65%
0.8	0.2	0.3317	0.0490	64%
0.9	0.1	0.2223	0.0373	67%

TABLE II

The behavior of exponential ARQ for different p, $L = 10$, $B \sim \text{Uniform}[1, 2] \times 10^5$, and $Q = 1000$ times of simulation.

channel is memoryless with transition probability

$$p(\mathsf{y}_i = y | \mathsf{x}_i = x) = \begin{cases} 1 - p & \text{if } y = x \\ p & \text{if } y = E \\ 0 & \text{otherwise} \end{cases}. \tag{10}$$

Let us define an (M, N, ϵ)-code as follows. There are M distinct messages, with each message w transmitted with uniform probability $1/M$. When there is no feedback, the encoder is a mapping

$$\{\mathsf{x}_i\}_{i=1}^{\mathsf{n}} = f(\mathsf{w}),$$

and when there is feedback, the encoder is a collection of n mappings

$$\mathsf{x}_i = f_i\left(\mathsf{w}, \{\mathsf{y}_j\}_{j=1}^{i-1}\right), \quad i = 1, 2, \ldots, \mathsf{n}.$$

For variable-length coding, n is a random variable, and we assume that it is known at the decoder. Either with or without feedback, the decoder is a mapping

$$\hat{\mathsf{w}} = g\left(\{\mathsf{y}_i\}_{i=1}^{\mathsf{n}}\right).$$

If $\hat{\mathsf{w}} \neq \mathsf{w}$, we say that a decoding error occurs. For an (M, N, ϵ)-code, we require the average error probability to satisfy

$$\frac{1}{M} \sum_{m=1}^{M} \text{Prob}\{\hat{\mathsf{w}} \neq \mathsf{w} | \mathsf{w} = m\} \leq \epsilon,$$

and the average total length, viewed as a general input cost, to satisfy

$$\frac{1}{M}\sum_{m=1}^{M}\mathbf{E}\left[\sum_{i=1}^{\mathsf{n}}\mathsf{s}_i\right]\leq N,$$

where s_i is the length of letter x_i. We say that $R\geq 0$ is an achievable rate if, for any $\epsilon>0$, there exists a sufficiently large $\bar{M}$ such that for every $M>\bar{M}$ there exists an (M,N,ϵ)-code with

$$\frac{\log M}{N}\geq R.$$

The channel capacity is the supremum of the achievable rates over all possible (M,N,ϵ)-codes.

Additional flexibility is available in the extended channel model through variable-length coding over letters. However, it is shown in [6] that even with this flexibility, the channel capacity per unit cost of the memoryless channel is

$$C_{\text{ext}}=\sup_{\mathsf{x}}\frac{I(\mathsf{x};\mathsf{y})}{\mathbf{E}[\mathsf{s}]}.$$

Furthermore, the capacity does not increase with feedback, which is a generalization of the result for fixed-length coding [7]. To evaluate C_{ext}, let us expand $I(\mathsf{x};\mathsf{y})$ as

$$\begin{aligned} I(\mathsf{x};\mathsf{y}) &\stackrel{(a)}{=} I(\mathsf{x},\mathsf{s};\mathsf{y}) \\ &\stackrel{(b)}{=} I(\mathsf{x};\mathsf{y}|\mathsf{s})+I(\mathsf{s};\mathsf{y}), \end{aligned} \tag{11}$$

where (a) follows from the fact that s is a deterministic function of x, and (b) follows from the chain rule for mutual information. The first term in (11) corresponds to the information conveyed by encoding letter characters, and can be upper bounded by

$$I(\mathsf{x};\mathsf{y}|\mathsf{s}) \quad\leq\quad (1-p)\mathbf{E}[\mathsf{s}], \tag{12}$$

with equality holding if, for any given s, the input x is chosen uniformly from $\{0,1\}^{\mathsf{s}}$. On the other hand, the second term in (11) corresponds to the additional information transferable by encoding letter lengths. To maximize $I(\mathsf{s};\mathsf{y})/\mathbf{E}[\mathsf{s}]$, we optimize the distribution of $\mathsf{s}\in\{1,2,\ldots\}$ and find that the maximizing distribution is $\Pr(\mathsf{s}=i)=2^{-i}$, which achieves $\max_{\mathsf{s}} I(\mathsf{s};\mathsf{y})/\mathbf{E}[\mathsf{s}]=1-p$. Therefore the channel capacity of the extended postal channel, both with and without feedback, is

$$C_{\text{ext}}=2(1-p), \tag{13}$$

which is twice the channel capacity of the original postal channel. Encoding the letter characters and encoding the letter lengths each contributes $(1-p)$ to the capacity.

IV. Concluding Remarks

The observations of this paper poses a question: "How to characterize the behavior of a channel in the presence of feedback?" Without feedback, there is little doubt that the channel capacity serves as the tightest rate upper bound for all possible reliable communication schemes. With feedback, however, there is room for different interpretations of information rate as illustrated by the analysis of exponential ARQ for the postal channel.

The channel capacity C_{ext} of the extended postal channel critically depends upon two assumptions implicitly made in the paper. First, there exists some packaging mechanism for the receiver to distinguish the starts and ends of consecutive letters. Second, there exists some marking mechanism for for the receiver to tell if an erasure has occurred. A future direction is to investigate the channel behavior if the two assumptions above are relaxed.

In a recent work studying general erasure (as well as deletion) channels [8], it is shown that, for very general erasure processes, the channel capacity is the same as for the memoryless erasure process with the same long-term erasure probability, and the channel capacity does not increase with feedback. However, those authors assume the channel symbols to be of fixed length and the erasure process to be independent of the input process, so their results do not automatically extend to the postal channel.

Acknowledgment

This work has been supported in part by the State of Indiana through the 21st Century Research Fund, by NSF through contract ECS03-29766, and by the CAM Fellowship of the University of Notre Dame.

References

[1] J. K. Wolf, A. D. Wyner, and J. Ziv, "The Channel Capacity of the Postal Channel," *Inform. Contr.*, 16, pp. 167–172, 1970.
[2] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, John Wiley & Sons, Inc., New York, 1991.
[3] S. Lin and D. J. Costello, *Error Control Coding*, 2nd Edition, Prentice Hall, Inc., 2004.
[4] P. Billingsley, *Probability and Measure*, 3rd Edition, John Wiley & Sons, Inc., New York, 1995.
[5] T. M. Cover, "Comments on Broadcast Channels," *IEEE Trans. Inform. Theory*, vol. 44, no. 6, pp. 2524–2530, Oct. 1998.
[6] I. Csiszár, "On the Capacity of Noisy Channels with Arbitrary Signal Costs," *Prob. Contr. Inform. Theory*, 2, pp. 283–304, 1973.
Information Theory: Coding Theorems for Discrete Memoryless Systems, Academic Press, 1981.
[7] C. E. Shannon, "The Zero Error Capacity of a Noisy Channel," *IRE Trans. Inform. Theory*, 2, pp. 8–19, 1956.
[8] S. Diggavi and M. Grossglauser, "On Information Transmission over a Finite Buffer Channel," *IEEE Trans. Inform. Theory*, vol. 52, no. 3, pp. 1226–1237, Mar. 2006.