ISIT'2013 1569717359

Multi-rate Sequential Data Transmission

Cheuk Ting Li
Dep. of Electrical Engr.
Stanford University
CA, USA
Email: ctli@stanford.edu

Shenghao Yang
Institute for Interdisciplinary
Information Sciences
Tsinghua University
Beijing, China
Email: shyang@tsinghua.edu.cn

Raymond W. Yeung
Dep. of Information Engr. and
Institute of Network Coding
The Chinese University of Hong Kong
Hong Kong, China
Email: whyeung@ie.cuhk.edu.hk

***Abstract*—We investigate the data transmission problem in which a sequence of data is broadcast to a number of receivers via erasure channels with different erasure probabilities. Accordingly, the receivers wish to decode the data sequentially at different rates. We present a formulation of the problem and propose an optimal coding scheme. Our results can be employed in the streaming of a video clip by broadcasting, so that receivers with different bandwidths can play the video at different speeds. Specifically, receivers with sufficiently large bandwidth can play the video at normal speed, while others can play the video with pauses, or at a slower speed using time-scale modification. Our results completely characterize the fundamental tradeoff between the available bandwidth and the playback speed of the video.**

***Index Terms*—Multi-rate, streaming, erasure channel.**

I. Introduction

Consider the scenario in which a long video clip has to be transmitted to a number of receivers having different packet loss rates. One approach is to divide the video into blocks of K packets, encode each block into $L \geq K$ encoding packets, and then transmit the blocks to the receivers sequentially. Using any capacity-achieving erasure code, a receiver can decode the block if about K out of L packets are received. This method, which we call a *blockwise code*, can only cater for the need of receivers with packet loss probability less than $1 - K/L$.

To suit the need of two receivers with different packet loss rates, we perform time multiplexing on two blockwise codes at different rates. This method will be referred to as a *multiplexed code*. Again divide the video data into blocks of K packets. Consider Blockwise Codes 1 and 2, which use random linear projections to encode each block into L_1 and L_2 packets respectively ($K \leq L_1 < L_2$). Denote the i-th packet generated using Blockwise Code k by $P_{k,i}$, $k = 1, 2$. We transmit the packets of the two codes in an interleaved manner, i.e., in the sequence $P_{1,1}, P_{2,1}, P_{1,2}, P_{2,2}, P_{1,3}, \ldots$. Receiver 1, which uses only the packets generated using Blockwise Code 1 (and discards those generated using Blockwise Code 2), can decode a block using K out of the L_1 packets encoded from the block, and therefore can tolerate a packet loss probability $1 - K/L_1$. As Blockwise Code 1 transmits a block of K packet per L_1 channel uses, taking interleaving into account, Receiver 1 can decode at a rate of $K/(2L_1)$ packets per channel use.

Receiver 2 uses packets generated by both codes. It can decode a block using K out of the $L_1 + L_2$ packets encoded from the block, and allows a higher packet loss probability $1 - K/(L_1 + L_2)$. However, Receiver 2 has to wait for the slower Blockwise Code 2, which transmits a block of K packet per L_2 channel uses. It can decode at a slower rate $K/(2L_2)$. If the playout speed of Receiver 1 is normal, then Receiver 2 will experience a playout speed slower than normal, as the trade-off of allowing a higher packet loss rate.

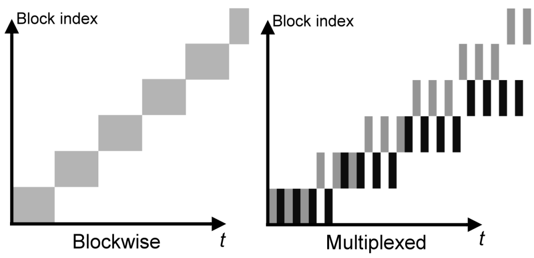

Figure I.1. Encoding process of blockwise and multiplexed codes

The reuse of packets from Blockwise Code 1 by Receiver 2 is the key of our method. If we split the bandwidth into two halves to serve Receivers 1 and 2 separately (Receiver 2 uses only packets from Blockwise Code 2), then the allowed packet loss probability for Receiver 2 is decreased to $1 - K/L_2$. This indicates the simple time sharing scheme is not optimal. Since Blockwise Code 1 has a faster transmission rate than Blockwise Code 2, Receiver 2 can also utilize the coded packets generated by Blockwise Code 1 without reducing its decoding rate. On the other hand, the faster Receiver 1 cannot use the coded packets generated by the slower Blockwise Code 2 without reducing its decoding rate.

Figure I.1 illustrates the transmission process of a blockwise code and a multiplexed codes described above. The subfigure on the left illustrates a blockwise code, where the blocks are encoded sequentially and the coded packets are transmitted accordingly. The subfigure on the right illustrates the multiplexing of two blockwise codes. The grey packets and the black packets are generated by Blockwise Code 1 (higher rate) and Blockwise Code 2 (lower rate), respectively. Note that the grey packets and the black packets are transmitted alternately.

In the second scenario studied above, the receivers with different channel conditions wish to decode the same sequence of data. Each receiver decodes the data sequentially at a different rate which depends on the channel condition. We call such a setting *multi-rate sequential data transmission*, which is the main subject to be studied in this paper.

Related problems have been considered in previous works.

This work was done when C. T. Li was with The Chinese University of Hong Kong.

This work was partially supported by a grant from University Grants Committee of the Hong Kong Special Administrative Region, China (Project No. AoE/E-02/08).

Multilevel diversity coding [1], [2] and priority encoding transmission [3] concern the scenario where a piece of message divided into several levels of importance is encoded into a number of packets. Each level requires a different number of packets to decode. The transmission model where the number of decoded bits depends on the channel conditions is encompassed in the study of variable-rate channel capacity [4]. Compared with these works, multi-rate sequential data transmission considers a continuous stream of data instead of a file with definite size.

In video streaming, when the decoding rate cannot catch up with the normal playout rate, the receiver may simply pause the video from time to time. Alternatively, a variable playout speed is sometimes employed to accommodate different channel conditions. Adaptive media playout (e.g., [5]) is a technique which allows the receiver to lower the playout speed when the amount of data buffered is low. Time-scale modification (e.g. [6]) may be performed at the receiver to adjust the playout speed without significant deterioration of audio quality. Other related approaches to this problem include layered multicast streaming and scalable video coding (e.g. [7], [8]), in which receivers with different bandwidths receive a different number of layers of the content corresponding to different levels of quality.

Multi-rate sequential data transmission is applicable in the broadcast of a piece of message which can be organized in a sequential manner. For instance, in the broadcast streaming of a piece of video, the transmitter may employ our code to cater for the need of receivers having different bandwidths, without the need of transmitting different data to different receivers. The ability of adapting the decoding rate for different packet loss rates automatically without feedback makes our code applicable also on point-to-point communications.

II. Problem Formulation

We consider transmitting data sequentially from a sender to a receiver through a memoryless erasure channel. We assume that the erasure channel is binary, and one bit is transmitted on the channel every unit time. Our results can readily be extended if the channel is instead a packet erasure channel.

Let the message sequence be a sequence of bits $M_1, M_2, ...$, where $M_i \overset{i.i.d.}{\sim} \mathrm{Bern}(1/2)$. The sender encodes the message sequence into a sequence of coded bits $X_1, X_2, ...$ and transmits them on the erasure channel. The coded bits are received at the receiver as $Y_1, Y_2, ...$, where $Y_i = X_i$ if no erasure occurs, otherwise $Y_i = e$. Based on these received symbols, the receiver tries to decode the message sequence as $\widetilde{M_1}, \widetilde{M_2}, \ldots$ Let $\mathcal{Y} = \bigcup_{n\in\mathbb{N}} \{0, 1, e\}^n$ be the set of all possible sequences of received symbols.

In our setting, the erasure probability of the channel, or equivalently the channel capacity, is unspecified. We want to design a coding scheme that can guarantee different decoding rates for different channel capacities. This formulation also applies when the sender broadcasts a sequence of data to a number of receivers through erasure channels with a common input but different capacities.

For simplicity, we use the notation $Y_a^b = (Y_a, Y_{a+1}, ..., Y_b)$.

Definition 1 (MRS code)**.** A *multi-rate streaming code* (MRS code) is specified by a pair of encoding and decoding functions. The encoding function

$$\mathrm{Enc} : \{0,1\}^{\mathbb{N}} \times \mathbb{N} \to \{0,1\},\ (\{M_i\}, j) \mapsto X_j$$

maps the message bits $\{M_i\}$ to the coded bits $\{X_j\}$. The decoding function

$$\mathrm{Dec} : \mathcal{Y} \times \mathbb{N} \to \{0,1\},\ (Y_1^n, i) \mapsto \widetilde{M_i}$$

maps the received symbols $\{Y_j\}$ to the decoded bits $\left\{\widetilde{M_i}\right\}$.

Note that the MRS code does not admit a fixed decoding rate like a block code. Instead its rate depends on the capacity of the erasure channel. Also note that in the above definition, the encoder can use the infinitely long message sequence for the encoding of every bit in the coded sequence. We will see in the next section that this assumption is actually superfluous, because there exist optimal codes for which the encoder uses only a finite segment of the message sequence to generate each bit, where the length of the segment increases with time.

Definition 2 (admissible pair)**.** For an MRS code, a capacity-rate pair (c, r) is *ϵ-admissible* if there exist N_0 such that when the receiver receives symbols from the sender through an erasure channel with capacity c (i.e., with erasure probability $1 - c$),

$$\mathbb{P}\left\{M_i \neq \widetilde{M_i}\left(Y_1^N\right)\right\} \leq \epsilon$$

for any $N \geq N_0$ and $i \leq N(r - \epsilon)$.

In other words, if (c, r) is ϵ-admissible, any receiver with channel capacity c can decode the first $N(r - \epsilon)$ message bits $M_1^{N(r-\epsilon)}$ with bit error probability less than ϵ when the first N symbols Y_1^N are received, for sufficiently large N, i.e., the receiver can decode the message causally at rate $r - \epsilon$ bits per channel use. Note that if $r \leq \epsilon$, then the capacity-rate pair (c, r) is always ϵ-admissible by any MRS code.

It is clear that for an MRS code, if (c, r) is ϵ-admissible, then all the pairs in $\{(c', r') | r' \leq r, c \leq c' \leq 1\}$ are ϵ-admissible. Therefore we can use a function to characterize all ϵ-admissible capacity-rate pairs of the code. We call $r : [0, 1] \to [0, \infty)$ a *rate-capacity function* if it is monotonically increasing, right continuous, and $r(0) = 0$.

Definition 3 (rate of MRS code)**.** A rate-capacity function $r(c)$ is called *ϵ-admissible* by a code if all of the pairs $(c, r(c))$ are ϵ-admissible by the code.

Definition 4 (achievable rate-capacity functions)**.** A rate-capacity function $r(c)$ is *achievable* if for any $\epsilon > 0$, there exist a code where $r(c)$ is ϵ-admissible by that code.

It can be shown that the rate-capacity functions shown in Figure II.1 are achievable, by considering the blockwise code and the multiplexed blockwise code described in the introduction, with $K \to \infty$ and the ratios between K and L, L_1 and L_2 kept constant.

III. Multiplexed Blockwise Codes

In this section, we will present the design of a class of MRS codes, called *multiplexed blockwise codes*, and give a sufficient condition for a rate-capacity function to be ϵ-admissible by a multiplexed blockwise code.

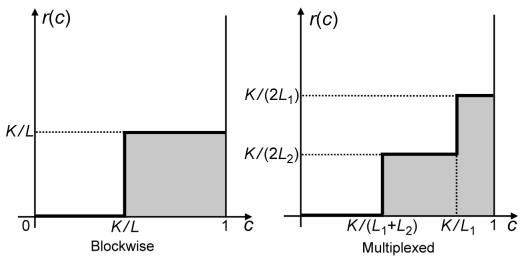

Figure II.1. Rate-capacity functions achieved by the blockwise and multiplexed blockwise code

Definition 5 (multiplexed blockwise codes)**.** A multiplexed blockwise code is characterized by the block size $K \in \mathbb{N}$, the period $\ell \in \mathbb{N}$ and the sequence of rates $\alpha_1, ..., \alpha_\ell \in (0, \infty)$, which may or may not be distinct. The encoding process is divided into two steps:

- **Blockwise encoding step:** The message $\{M_i\}$ is divided into blocks of K symbols, $B_i = M_{(i-1)K+1}^{iK}$, and each block B_i is encoded into a codeword B'_i of $K' \geq \sum_{m=1}^{\ell} \left\lceil \frac{K}{\ell \cdot \alpha_m} \right\rceil$ symbols using an erasure code.
- **Multiplexing step:** At time instance n, the sender generates an encoding symbol $X_n = B'_{i_n, j_n}$, where $i_n = \lceil n \cdot \alpha_m / K \rceil$ and $m \in \{1, ..., \ell\}, m \equiv n \,(\mathrm{mod}\, \ell)$, and j_n is the smallest positive integer such that the j_n-th symbol of B'_{i_n} was not used previously in the multiplexing step, i.e., $j_n = |\{k | k < n \text{ and } i_k = i_n\}| + 1$.

We can understand the multiplexing step in another way: The blockwise code with codewords $B'_1, B'_2 \ldots$ is split into ℓ streams, where the m-th stream contains around $\frac{K}{\ell \cdot \alpha_m}$ symbols of each codeword B'_i. The multiplexed blockwise code essentially performs time multiplex on the ℓ streams. The requirement $K' \geq \sum_{m=1}^{\ell} \left\lceil \frac{K}{\ell \cdot \alpha_m} \right\rceil$ ensures that there will be enough symbols for the encoding process.

On the receiver side, when a symbol is received at time instance n (the symbol originates from B'_{i_n}), the receiver would try to decode B_{i_n} using the current symbol and the previously received symbols from B'_{i_n}. If the symbols are insufficient, the current symbol is stored until enough symbols are received. We assume an infinite buffer size at the receiver.

A suitable erasure code is needed in the blockwise encoding step. We assume that for every sufficiently large K, we have a (K', K) erasure code where $K' = O(K)$ such that if the number of non-erased symbols is at least $K + o(K)$, then the receiver can decode the message with error probability tends to zero as K goes to infinity. Examples of such codes include random linear codes [9] and fountain codes [10], [11].

The codes described in the introduction are examples of multiplexed blockwise codes. For the first example, the parameters are taken to be $\ell = 1$, $\alpha_1 = K/L$. For the second example, $\ell = 2$, $\alpha_1 = K/(2L_1)$, $\alpha_2 = K/(2L_2)$.

The choice of the parameters is closely related to the rate-capacity function we would like to achieve. The following theorem describes their relation.

Theorem 6. *For any $\epsilon > 0$, the rate-capacity function $r(c)$ is ϵ-admissible by the multiplexed blockwise code with block size K and a given sequence of rates $\alpha_1, ..., \alpha_\ell$ for all sufficiently large K, if there exists $\xi > 0$ satisfying*

$$c \cdot g(r(c)) \geq 1 + \xi \textit{ for all } c > 0 \textit{ with } r(c) > 0,$$

where

$$g(r) = \frac{1}{\ell} \cdot \sum_{m=1,...,\ell,\ \alpha_m \geq r} \frac{1}{\alpha_m}.$$

Proof: Fix $\epsilon > 0$. Let $r(c)$ be a rate-capacity function, and suppose ℓ and $\alpha_1, ..., \alpha_\ell$ satisfies the condition $c \cdot g(r(c)) \geq 1 + \xi$ for some $\xi > 0$. We now consider the multiplexed blockwise code with block size K and sequence of rates $\alpha_1, ..., \alpha_\ell$, where the value of K will be determined later.

Let $\eta > 0$ such that $r(\eta) < \epsilon$ (η exists by $r(0) = 0$ and right-continuity of r). Fix any capacity c with $r(c) > \epsilon$ (we do not need to consider the case $r(c) \leq \epsilon$ as $(c, r(c))$ is always ϵ-admissible), then $c > \eta$. Let $r_1 = r(c)$ and $r_0 = r(c) - \epsilon$. Let $N \geq \frac{K}{\epsilon}$, and $k \leq \frac{N r_0}{K} + 1$. We will analyze whether the block B_k can be decoded using Y_1^N with error probability less than ϵ when the channel capacity is c. If so, then $M_1^{\lfloor r_0 N \rfloor}$ can be decoded using Y_1^N with bit error probability less than ϵ when $N \geq \frac{K}{\epsilon}$, and the capacity-rate pair (c, r_1) is ϵ-admissible.

Consider an $m \in \{1, ..., \ell\}$ with $\alpha_m \geq r_1$. For time instances $n = m, m+\ell, m+2\ell, ...$ and $n \leq N$, the encoder generates a symbol from $B'_{\lceil n \cdot \alpha_m / K \rceil}$. For any $k \geq 1$, the codeword B'_k is chosen at one of those time instance n, i.e., $i_n = k$, if

$$(k-1) \cdot K/\alpha_m < n \leq k \cdot K/\alpha_m.$$

For $k \leq \frac{N r_0}{K} + 1$, the upper bound above satisfies

$$k \cdot \frac{K}{\alpha_m} \leq \left(\frac{N(r_1 - \epsilon)}{K} + 1 \right) \cdot \frac{K}{r_1} = N - \left(\frac{N\epsilon - K}{r_1} \right) \leq N.$$

Therefore, the number of times B'_k is chosen at time instances $n = m, m + \ell, m + 2\ell, ...$ and $n \leq N$ is at least $\frac{K}{\alpha_m \ell} - 1$. Let the total number of times that B'_k is chosen be S. Summing through all m with $\alpha_m \geq r_1$, we obtain

$$S \geq \sum_{m=1,...,\ell,\ \alpha_m \geq r_1} \frac{K}{\alpha_m \ell} - \ell.$$

Using the condition $c \cdot g(r(c)) \geq 1 + \xi$, we have

$$cS \geq K(1+\xi) - c\ell \geq K(1+\xi) - \ell.$$

As each of those encoding symbols have probability $1 - c$ to be erased, the number of received symbols in B'_k follows a binomial distribution $B(S, c)$. Let the number of received symbols be X. Assume that the erasure code used requires $K + \delta$ received symbols to decode with error probability $\epsilon/2$, where $\delta = o(K)$. For K large enough, we can have

$$\mathbb{E}[X] = cS \geq K(1+\xi) - \ell > K + \delta.$$

Hoeffding's inequality gives

$$\mathbb{P}\{X < K + \delta\} < \exp\left(-2(cS - K - \delta)^2/S\right).$$

By the monotonicity of $(cS - K - \delta)^2/S$ with respect to S,

$$(cS - K - \delta)^2/S \geq \eta \left(K\xi - \ell - \delta\right)^2 / \left(K(1+\xi) - \ell\right)$$

which tends to infinity as $K \to \infty$. Therefore when K is large enough such that $K(1 + \xi) - \ell > K + \delta$ and $\exp\left(-2\eta \left(K\xi - \ell - \delta\right)^2 / \left(K(1+\xi) - \ell\right)\right) < \epsilon/2$, the probability of receiving insufficient symbols can be bounded by $\epsilon/2$.

Note that the choice of K is independent of c. Together with the error probability of the erasure code, the overall failure probability is bounded by ϵ. The result follows. ■

IV. Achievable Region of Rate-capacity Functions

The following theorem gives the achievable region for the rate-capacity function, and shows that multiplexed blockwise codes are optimal.

Theorem 7. *A rate-capacity function $r(c)$ is achievable if and only if*

$$\int_0^1 \frac{1}{c} dr(c) = \lim_{t\to 0^+} \int_t^1 \frac{1}{c} dr(c) \le 1,$$

and any achievable $r(c)$ can be achieved using multiplexed blockwise codes.

Proof of achievability: Assume $r(c)$ is a rate-capacity function satisfying $\int_0^1 \frac{1}{c} dr(c) \le 1$. Note that for any $0 < t < c \le 1$, $\left(r(c) - r(t)\right)/c \le \int_t^c \frac{1}{x} dr(x) \le 1$. Since $r(c)$ is right-continuous and $r(0) = 0$, we have $r(c) \le c$ for any $c \in (0, 1]$. Fix $\epsilon \in (0, r(1))$. Let $r_2(c) = \max(r(c) - \epsilon/2, 0)$, note that $r(c) - r_2(c)$ is monotonically increasing, and

$$\begin{aligned}
\int_0^1 \frac{1}{c} dr_2(c) &\le 1 - \left(\int_0^1 \frac{1}{c} dr(c) - \int_0^1 \frac{1}{c} dr_2(c)\right) \\
&= 1 - \int_0^1 \frac{1}{c} d\left(r(c) - r_2(c)\right) \\
&\le 1 - \left((r(1) - r_2(1)) - (r(0) - r_2(0))\right) \\
&= 1 - \epsilon/2. \qquad \text{(IV.1)}
\end{aligned}$$

As $r(\epsilon/2) \le \epsilon/2$, we have $r_2(c) = 0$ for $c \le \epsilon/2$. Let $\xi = \epsilon/8$. Let $N \in \mathbb{N}$ and $c_k = (1 + \xi k)^{-1}$ for $k = 0, ..., N$, such that $c_N < \epsilon/2$ (and therefore $r_2(c_N) = 0$). We construct a multiplexed blockwise code by assigning the sequence of rates by first fixing the period $\ell \ge 8N/\epsilon$, then setting the values of α_m to one of $\{r_2(c_i)\}_{i=0,...,N-1}$ such that

$$|\{m|\alpha_m = r_2(c_0)\}| \ge \left\lceil (1 + \xi)^2 \cdot \ell \cdot r_2(c_0)\right\rceil,$$

and

$$|\{m|\alpha_m = r_2(c_k)\}| \ge \lceil (1 + \xi) \cdot \xi\ell \cdot r_2(c_k)\rceil$$

for $k = 1, ..., N - 1$. This assignment is possible only when the sum of the right hand sides of the two inequalities above is not greater than ℓ. To check this, note that

$$\xi\ell \cdot r_2(c_k) = \ell \cdot \int_{c_{k-1}}^{c_k} r_2\left(c_k\right) d\frac{1}{c} \le \ell \cdot \int_{c_{k-1}}^{c_k} r_2\left(c\right) d\frac{1}{c}.$$

Therefore,

$$\begin{aligned}
&\left\lceil (1 + \xi)^2 \cdot \ell \cdot r_2(c_0)\right\rceil + \sum_{k=1}^{N-1} \lceil (1 + \xi) \cdot \xi\ell \cdot r_2(c_k)\rceil \\
&\le (1+\xi)\left((1 + \xi)\ell \cdot r_2(c_0) + \ell \cdot \sum_{k=1}^{N-1} \int_{c_{k-1}}^{c_k} r_2\left(c\right) d\frac{1}{c}\right) + N \\
&\le (1 + \xi)\left(\ell \cdot r_2(1) + \ell \cdot \int_1^0 r_2\left(c\right) d\frac{1}{c} + \xi\ell \cdot r_2(1) + N\right) \\
&= (1 + \xi)\left(\ell \cdot \int_0^1 \frac{1}{c} dr_2\left(c\right) + \xi\ell \cdot r_2(1) + N\right) \\
&\le \ell \cdot (1 + \epsilon/8)\left(1 - \epsilon/2 + \epsilon/8 + \epsilon/8\right) \\
&\le \ell
\end{aligned}$$

since $r_2(1) \le r(1) \le 1$. To check whether $c \cdot g\left(r_2(c)\right) \ge 1 + \xi$ is satisfied, note that when $c \in (c_k, c_{k-1}]$ (if c is in none of these ranges, then $c \le c_N$, which implies $r_2(c) = 0$),

$$\begin{aligned}
c \cdot g\left(r_2(c)\right) &\ge \frac{c_k}{\ell} \cdot \left(\frac{1}{r_2(c_0)} \cdot \left\lceil (1 + \xi)^2 \ell \cdot r_2(c_0)\right\rceil \right. \\
&\qquad \left. + \sum_{i=1}^{k-1} \frac{1}{r_2(c_i)} \cdot \lceil (1 + \xi)\xi\ell \cdot r_2(c_i)\rceil\right) \\
&\ge \frac{c_k}{\ell} \cdot \left((1 + \xi)^2 \ell + (1 + \xi)(k - 1)\xi\ell\right) \\
&= 1 + \xi.
\end{aligned}$$

By Theorem 6, the rate-capacity function $r_2(c) = \max(r(c) - \epsilon/2, 0)$ is $\epsilon/2$-admissible by the multiplexed blockwise code for sufficiently large block size, which implies that $r(c)$ is ϵ-admissible by the code (cf. Definition 2). ■

Proof of converse: Let $r(c)$ be an achievable rate-capacity function. For any $\epsilon > 0$, consider a code where $r(c)$ is ϵ-admissible. The message $\{M_i\}$ are encoded into binary symbols $\{X_i\}$, and sent through an erasure channel with capacity c (we call it Channel c) to give $\{Y_{c,i}\}$ for all $c > 0$. Assume we have the following for any c,

$$\max_{i \le n(r(c)-\epsilon)} \mathbb{P}\left\{M_i \ne \widetilde{M}_i\left(Y_{c,1}^n\right)\right\} < \epsilon \text{ for any } n \ge N_0.$$

Let $N \ge N_0$. Let $E_{c,i}$ be the indicator of the events that X_i is not erased in Channel c, i.e., $Y_{c,i} = X_i$ when $E_{c,i} = 1$, and $Y_{c,i} = e$ when $E_{c,i} = 0$. Let $r_2(c) = \max(r(c) - \epsilon, 0)$, and define

$$f(c) = \frac{1}{Nc} \cdot H(Y_{c,1}^N | M_1^{\lfloor N r_2(c)\rfloor}, E_{c,1}^N). \qquad \text{(IV.2)}$$

Let $\eta = \sqrt{\epsilon}$. Consider Channels c_0 and c_1 where $\eta \le c_0 < c_1 \le 1$. Let $k_0 = \lfloor N r_2(c_0)\rfloor$, and $k_1 = \lfloor N r_2(c_1)\rfloor$. Note that

$$\begin{aligned}
H(Y_{c_1,1}^N | M_1^{k_1}, E_{c_1,1}^N) &= H(Y_{c_1,1}^N, M_{k_0+1}^{k_1} | M_1^{k_0}, E_{c_1,1}^N) \\
&\quad - H(M_{k_0+1}^{k_1} | M_1^{k_0}, E_{c_1,1}^N), \qquad \text{(IV.3)}
\end{aligned}$$

where, due to the assumption that $M_i \overset{i.i.d.}{\sim} \text{Bern}(1/2)$,

$$\begin{aligned}
H(M_{k_0+1}^{k_1} | M_1^{k_0}, E_{c_1,1}^N) &= k_1 - k_0 \\
&\ge N(r_2(c_1) - r_2(c_0)) - 1. \qquad \text{(IV.4)}
\end{aligned}$$

As $M_1^{k_1}$ can be decoded using $Y_{c_1,1}^N$ with bit error probability less than ϵ, by Fano's inequality,

$$\begin{aligned}
&H(Y_{c_1,1}^N, M_{k_0+1}^{k_1} | M_1^{k_0}, E_{c_1,1}^N) \\
&= H(Y_{c_1,1}^N | M_1^{k_0}, E_{c_1,1}^N) + \sum_{i=k_0+1}^{k_1} H(M_i | Y_{c_1,1}^N, M_1^{i-1}, E_{c_1,1}^N) \\
&\le H(Y_{c_1,1}^N | M_1^{k_0}, E_{c_1,1}^N) + N(r_2(c_1) - r_2(c_0))H(\epsilon) + 1. \qquad \text{(IV.5)}
\end{aligned}$$

For any $S \subseteq \{1, ..., N\}$, the probability that $E_{c_1,i} = 1$ if and only if $i \in S$ can be given by $c_1^{|S|}\left(1 - c_1\right)^{N-|S|}$. We write X_S for the concatenation of X_i with $i \in S$. We can obtain

$$\begin{aligned}
&H(Y_{c_1,1}^N | M_1^{k_0}, E_{c_1,1}^N) \\
&= \sum_{t=0}^{N} c_1^t \left(1 - c_1\right)^{N-t} \sum_{|S|=t} H(Y_{c_1,1}^N | M_1^{k_0}, E_{c_1,1}^N : E_{c_1,i} = 1 \text{ iff } i \in S) \\
&\overset{\text{(i)}}{=} \sum\nolimits_{t=1}^{N} c_1^t \left(1 - c_1\right)^{N-t} \sum\nolimits_{|S|=t} H(X_S | M_1^{k_0})
\end{aligned}$$

$$\stackrel{\text{(ii)}}{=} \sum_{t=1}^{N} c_1^t (1-c_1)^{N-t} \frac{1}{t} \cdot \frac{c_1}{c_0}$$
$$\cdot \sum_{s=1}^{t} \left(\frac{c_0}{c_1}\right)^s \left(1-\frac{c_0}{c_1}\right)^{t-s} \binom{t}{s} \cdot s \sum_{|S|=t} H(X_S|M_1^{k_0})$$
$$\stackrel{\text{(iii)}}{\le} \sum_{t=1}^{N} c_1^t (1-c_1)^{N-t} \frac{1}{t} \cdot \frac{c_1}{c_0}$$
$$\cdot \sum_{s=1}^{t} \left(\frac{c_0}{c_1}\right)^s \left(1-\frac{c_0}{c_1}\right)^{t-s} \binom{t}{s} \cdot \frac{t\binom{N}{t}}{\binom{N}{s}} \sum_{|S|=s} H(X_S|M_1^{k_0})$$
$$= \frac{c_1}{c_0} \sum_{s=1}^{N} \sum_{t=s}^{N} (1-c_1)^{N-t} c_0^s (c_1-c_0)^{t-s} \binom{N-s}{t-s} \sum_{|S|=s} H(X_S|M_1^{k_0})$$
$$= \frac{c_1}{c_0} \sum_{s=1}^{N} c_0^s (1-c_0)^{N-s}$$
$$\cdot \sum_{t=s}^{N} \left(\frac{1-c_1}{1-c_0}\right)^{N-t} \left(\frac{c_1-c_0}{1-c_0}\right)^{t-s} \binom{N-s}{t-s} \sum_{|S|=s} H(X_S|M_1^{k_0})$$
$$= \frac{c_1}{c_0} \sum_{s=1}^{N} c_0^s (1-c_0)^{N-s} \sum_{|S|=s} H(X_S|M_1^{k_0})$$
$$= \frac{c_1}{c_0} H(Y_{c_0,1}^N | M_1^{k_0}, E_{c_0,1}^N), \tag{IV.6}$$

where (i) is because $Y_{c_1,1}^N$ and X_S are functions of each other given that $E_{c_1,i} = 1$ if and only if $i \in S$, $(X_S, M_1^{k_0})$ is independent of $E_{c_1,1}^N$, and the term is 0 when $t = 0$; (ii) is a result of the formula $\sum_{s=1}^{t} \alpha^s (1-\alpha)^{t-s} \binom{t}{s} \cdot s = t\alpha$; and (iii) follows from Han's inequality [12].

Hence by (IV.3), (IV.4), (IV.5) and (IV.6),

$$H(Y_{c_1,1}^N | M_1^{k_1}, E_{c_1,1}^N)$$
$$\le \frac{c_1}{c_0} \cdot H(Y_{c_0,1}^N | M_1^{k_0}, E_{c_0,1}^N) + (N(r_2(c_1) - r_2(c_0))) \cdot H(\epsilon)$$
$$+1 - (N(r_2(c_1) - r_2(c_0)) - 1),$$

Upon replacing the terms by $f(c)$ using (IV.2),

$$\frac{r_2(c_1) - r_2(c_0)}{c_1} \le \frac{f(c_0) - f(c_1) + \frac{2}{Nc_1}}{1 - H(\epsilon)} \le \frac{f(c_0) - f(c_1) + \frac{2}{N\eta}}{1 - H(\epsilon)}.$$

Due to the monotonicity of $r_2(c)$,

$$\int_{c_0}^{c_1} \frac{1}{c} dr_2(c) - \left(\frac{1}{c_0} - \frac{1}{c_1}\right)(r_2(c_1) - r_2(c_0)) \le \frac{1}{c_1}(r_2(c_1) - r_2(c_0))$$
$$\le \frac{f(c_0) - f(c_1) + \frac{2}{N\eta}}{1 - H(\epsilon)}.$$

Let $L = \lfloor \sqrt{N} \rfloor$, and $c_i' = \left(c_0^{-1} + \left(c_1^{-1} - c_0^{-1}\right) \cdot i/L\right)^{-1}$ for $i = 0, ..., L$. Note that for $i = 0, ..., L-1$,

$$\frac{1}{c_i'} - \frac{1}{c_{i+1}'} = \frac{1}{L}\left(\frac{1}{c_0} - \frac{1}{c_1}\right) \le \frac{1}{L\eta}$$

as $\eta \le c_0 < c_1$. Then we have

$$\int_{c_0}^{c_1} \frac{1}{c} dr_2(c)$$
$$= \sum_{i=0}^{m-1} \int_{c_i'}^{c_{i+1}'} \frac{1}{c} dr_2(c)$$
$$\le \sum_{i=0}^{m-1} \left(\frac{f(c_i') - f(c_{i+1}') + \frac{2}{N\eta}}{1 - H(\epsilon)} + \left(\frac{1}{c_i'} - \frac{1}{c_{i+1}'}\right)\left(r_2(c_{i+1}') - r_2(c_i')\right) \right)$$
$$\le \frac{f(c_0) - f(c_1) + \frac{2L}{N\eta}}{1 - H(\epsilon)} + \frac{1}{L\eta} r(1).$$

Note that $f(c) \le \frac{1}{Nc} \cdot H(Y_{c,1}^N | E_{c,1}^N) \le 1$. Thus we have, for any $\eta \le c_0 < c_1 \le 1$,

$$\int_{c_0}^{c_1} \frac{1}{c} dr_2(c) \le \frac{1 + \frac{2L}{N\eta}}{1 - H(\epsilon)} + \frac{1}{L\eta} r(1).$$

Therefore we can obtain an inequality on $r(c)$ by

$$\int_{\eta}^{1} \frac{1}{c} dr(c) \le \int_{\eta}^{1} \frac{1}{c} dr_2(c) + \frac{\epsilon}{\eta} \le \frac{1 + \frac{2L}{N\eta}}{1 - H(\epsilon)} + \frac{1}{L\eta} r(1) + \frac{\epsilon}{\eta},$$

where the proof of the first inequality above is similar to that of (IV.1). Letting $N \to \infty$, we have

$$\int_{\eta}^{1} \frac{1}{c} dr(c) \le \frac{1}{1 - H(\epsilon)} + \frac{\epsilon}{\eta} = \frac{1}{1 - H(\eta^2)} + \eta$$

for any $\eta > 0$. Hence $\int_0^1 \frac{1}{c} dr(c) = \lim_{\eta \to 0} \int_{\eta}^{1} \frac{1}{c} dr(c) \le 1$. ■

V. Conclusion

In this paper, we have presented a formulation for multi-rate sequential data transmission, proposed a coding scheme, and proved its optimality. One direction for future research is to extend the result to multiple transmitters, where each receiver may have different channel capacities for different transmitters, and the rate-capacity function takes k parameters instead of one, where k is the number of transmitters. It can be shown that multiplexed blockwise codes are no longer optimal in this setting [13]. The region of achievable rate-capacity functions for this setting is yet to be determined.

References

[1] J. R. Roche, R. W. Yeung, and K. P. Hau, "Symmetrical multilevel diversity coding," *IEEE Trans. Info. Theory*, pp. 1059–1064, 1997.
[2] R. W. Yeung and Z. Zhang, "On symmetrical multilevel diversity coding," *IEEE Trans. Info. Theory*, pp. 609–621, 1999.
[3] A. Albanese, J. Blomer, J. Edmonds, M. Luby, and M. Sudan, "Priority encoding transmission," *IEEE Trans. Info. Theory*, pp. 1737–1744, 1996.
[4] S. Verdú and S. Shamai, "Variable-rate channel capacity," *IEEE Trans. Info. Theory*, vol. 56, no. 6, pp. 2651–2667, Jun. 2010.
[5] E. Steinbach, N. Färber, and B. Girod, "Adaptive playout for low latency video streaming," in *IEEE Proc. International Conference on Image Processing*, October 2001, pp. 962–965.
[6] Y. J. Liang, N. Färber, and B. Girod, "Adaptive playout scheduling using time-scale modification in packet voice communications," in *Proc. ICASSP*, May 2001, pp. 1445–1448.
[7] N. Shacham, "Multipoint communication by hierarchically encoded data," in *Proceedings IEEE Infocom '92*, 1992, pp. 2107–2114.
[8] D. Taubman and A. Zakhor, "Multirate 3-D subband coding of video," *IEEE Transactions on Image Processing*, vol. 3, no. 5, pp. 572–588, September 1994.
[9] P. Elias, "Coding for two noisy channels," in *Proc. 3rd London Symp. Information Theory*, London, U.K., Sep. 1955, pp. 61–76.
[10] M. Luby, "LT codes," *Proc. IEEE Symp. Found. Comp. Sci.*, pp. 271–280, 2002.
[11] A. Shokrollahi, "Raptor codes," *IEEE Trans. Info. Theory*, vol. 52, no. 6, pp. 2551–2567, Jun. 2006.
[12] T. S. Han, "Nonnegative entropy measures of multivariate symmetric correlations," *Inform. Contr.*, vol. 36, pp. 133–156, 1978.
[13] C. T. Li, "On Multi-rate Sequential Data Transmission," *ArXiv e-prints*, Oct. 2012. [Online]. Available: http://arxiv.org/abs/1210.3427