(n,k) SOURCE-CHANNEL ERASURE CODES: CAN PARITY BITS ALSO REFINE QUALITY?

by

S. Sandeep Pradhan, Rohit Puri and Kannan Ramchandran

Memorandum No. UCB/ERL M01/3

16 January 2001

(n,k) SOURCE-CHANNEL ERASURE CODES: CAN PARITY BITS ALSO REFINE QUALITY?

by

S. Sandeep Pradhan, Rohit Puri and Kannan Ramchandran

Memorandum No. UCB/ERL M01/3

16 December 2001

ELECTRONICS RESEARCH LABORATORY

College of Engineering
University of California, Berkeley
94720

(n, k) Source-Channel Erasure Codes : Can Parity Bits also Refine Quality ?

S. Sandeep Pradhan Rohit Puri Kannan Ramchandran
{pradhan5, rpuri, kannanr}@eecs.berkeley.edu
EECS Department, University of California, Berkeley,
211 Cory Hall #1772,
Berkeley, CA-94720.
Ph: (510)-643-4034 Fax: (510)-642-2845

Abstract

We consider the problem of maximum distance separable $(n, k, n-k+1)$ erasure code construction in a source-channel coding framework and present an achievable rate-distortion region for this problem using information theoretic techniques. The region so obtained has the following interpretation : the reception of *any* k channel symbols at the decoder leads to a reconstruction quality commensurate with the reception of k source symbols; more interestingly, the reception of $m > k$ channel symbols leads to a strictly superior quality. Thus, in this source-channel coding framework, the role of conventional "parity bits" takes on an interesting interpretation. When the channel is as bad as advertised, i.e. when only a fraction k/n of the packets arrive intact, the parity bits play the conventional role of aiding the full recovery of k information packets. When the channel is better than advertised, however, while conventional parity information is wasted, in our framework they "moonlight" in aiding to strictly improve the reconstructed source quality.

In the special case of a Gaussian source, we have the following interesting result: when any k packets arrive, our solution leads to a source reconstruction whose distortion exactly matches the information-theoretic distortion-rate performance of a Gaussian source corresponding to the rate associated with k packets; with the reception of $m > k$ packets, the reconstructed quality is strictly better, with the improvement being asymptotically linear in the number of packets received.

Keywords : Source-Channel Codes, Maximum Distance Separable (MDS) codes.

1 Introduction

With the explosive growth of packet switched networks like the Internet, the transmission of information over unreliable channels has received considerable attention lately. Such networks can be efficiently modeled as packet erasure channels. An information sequence is encoded into a large number ($\gg 2$) of packets and transmitted over the network. The network randomly erases some of the packets and transmits the rest of them errorlessly. Current data transfer protocols such as User Datagram Protocol (UDP) assume such a model for the packet switched networks. Erasure Channel Codes [1], which enable reconstruction at the decoder with the reception of a subset of the packets transmitted, offer a solution to transmission over such channels.

An (n, k, d) erasure channel code [1] refers to a construction where k user symbols belonging to a finite field are encoded into n channel symbols (also belonging to the same finite field) such that with the reception of *any* $(n - d + 1)$ of the n channel symbols, the original k user symbols can be recovered. Channel codes for which $d = n - k + 1$, i.e., the k user symbols can be recovered when *any* k channel symbols are received, are referred to as Maximum Distance Separable (MDS) codes [1]. Reed Solomon Codes are a popular class of codes that possess this property.

This problem has also been studied in the name of multiple description source coding in the literature [2, 3] where an information-theoretic treatment of the achievable rate region has been quantified and constructive approaches using scalar quantizers have been proposed. Some protocols that have been considered in the literature for transmission of multimedia information over packet networks have used MDS codes. In [4] a priority-based robust transmission algorithm is suggested for lossy-packet based networks. Information content with higher priority is encoded using lower rate MDS codes while those with lower priority is encoded using a corresponding higher rate code. Algorithms for image and video transmission systems have been proposed in [5, 6, 7] that use progressive source coding techniques concatenated with MDS codes.

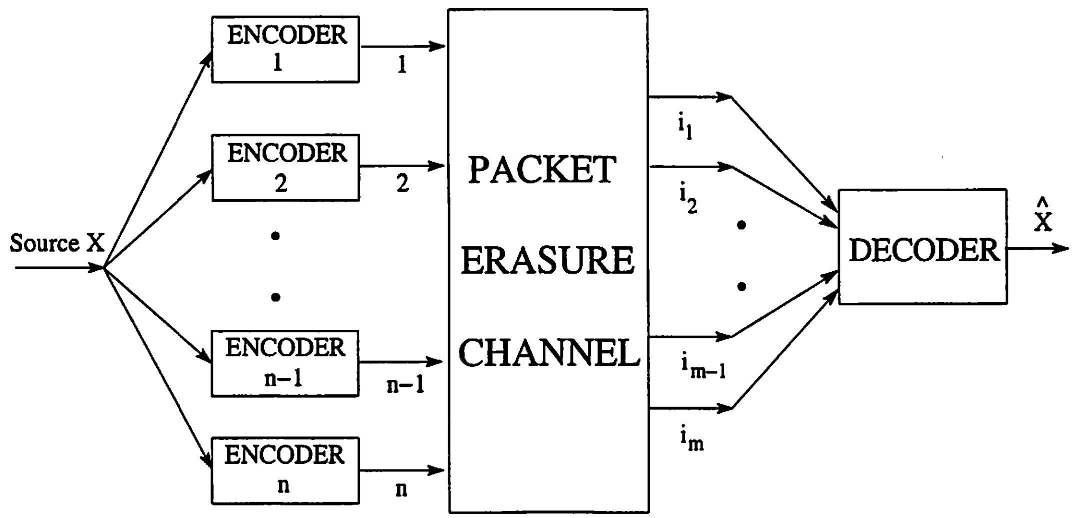

Figure 1: Maximum distance separable $(n, k, n-k+1)$ quantization code: A source is encoded into n descriptions or packets and transmitted over n errorless channels which are equally likely to breakdown. The decoder would like to reconstruct with the availability of any $m \geq k$ packets of information.

The construction of such codes has traditionally been considered in a discrete finite field setting. In this work, we present a new approach based on random binning arguments that extends these concepts into the source symbol domain. We consider direct mappings from the source symbol space to the erasure channel symbol space. While in this paper we undertake an information-theoretic analysis of the problem, the actual construction of such encoders and decoders based on group codes, inspired by these arguments, are of great interest and is a part of our ongoing work.

Our motivation stems from the following simple observation. Consider an information source, X, to be transmitted over n distinct channels each with transmission rate R bits/sample such that only one of these is guaranteed to reach the destination. One way of transmitting such a source is to construct a codebook, say C, with rate-distortion [8] performance with a rate of R bits/sample, i.e. attaining distortion $D(R)$ that lies on the rate-distortion function of the source. A bitstream characterizing the index of the codeword from this codebook used for quantizing the source X, can be encoded for transmission over the n channels using a conventional $(n, 1, n)$ MDS code which is a repetition code. All the n streams in such a system will be identical and the reception of *any* one of the n channels would enable the decoder to reconstruct the source with a distortion $D(R)$ which is the rate-distortion function of the source X. It is important to note that with the reception of any of the remaining $(n-1)$ packets, there is no improvement in performance.

Now consider the following alternate scenario. Here we generate n independent codebooks, say $\mathcal{C}_1, \mathcal{C}_2, \ldots \mathcal{C}_n$ each with the rate-distortion performance. The source is independently quantized using each of these codebooks. The quantized codeword index of the i^{th} codebook is transmitted over the i^{th} channel. In this scenario, when any one of the n channels is active, the decoded quality is identical to that in the first scenario. However, as the number of received packets increases, the decoded quality gets better because of the gain due to multiple independent looks at the source with independent quantizers! Thus, in this case the "parity" contributes to the decoded reconstruction fidelity unlike the first scenario. We have thus constructed an $(n, 1, n)$ "symbol" code as opposed to a "bit" code. Based on this motivation, we generalize these ideas to prove the existence of $(n, k, n-k+1)$ "symbol" codes in the following sections that strictly outperform the $(n, k, n-k+1)$ "bit" code (see Figure 1).

A key point to note in this problem is that there is an inherent uncertainty at the encoder about *which* packets have reached the decoder. This calls for a coding framework which deals with uncertainty at the encoder about the information available at the decoder. In [9], an information theoretic analysis of the problem of separate encoding of correlated distributed sources was studied. A class of coding techniques was developed based on group codes for this problem in a practical and constructive setting in [10]. In the above problem, the goal is to encode a set of multiple correlated distributed sources independently while still exploiting the joint correlation assuming the knowledge of the joint statistics. The encoder of each source sends only a carefully designed partial information such that all the sources of interest can be reconstructed with the availability of such partial information from all the sources at the joint decoder. In this paper we apply some of these techniques to the problem of the $(n, k, n-k+1)$ maximum distance-separable quantization code.

A key result of this paper suggests that using such an $(n, k, n-k+1)$ maximum distance separable quantization code, it is possible to encode a unit variance *i.i.d.* Gaussian source into n packets with each packet containing R bits/sample such that the reconstruction fidelity with the reception of any $(k+r)$ packets for $0 \leq r \leq (n-k)$ is given by

$$D_{k+r} = \frac{k}{2^{2kR}(k+r) - r}. \tag{1}$$

Note that when $r = 0$, $D_k = 2^{-2kR}$ which is on the rate-distortion function of the corresponding Gaussian source! Further, in the limit of high rate and large number of packets,

$$\frac{D_{k+r}}{D_k} \approx \frac{k}{k+r}. \tag{2}$$

Infact, for the specific case of a Gaussian source, we conjecture that the rate region obtained in this work is the absolute best and proving the converse theorem is a part of our ongoing work.

Though in this paper we restrict our attention to this problem, a bigger goal is to integrate such codes with different rates to build an efficient transmission protocol similar to [4, 5] where we have the flexibility of tuning the achievable reconstruction fidelity with the reception of any number of packets and this a part of our ongoing work.

2 Problem Formulation

2.1 Notation and Definitions

We first state some basic notation, definitions and properties that we will use in the sequel. Let $\mathcal{I}_l = \{1, 2, \ldots, l\}$. Let $|S|$ denote the cardinality of a set S. Bold-faced letters denote vectors. The elements of an l-vector are denoted as $\mathbf{x}_i = [x_{i1},\ x_{i2}, \ldots, x_{il}]$. For any subset S of $\mathcal{I}_l$, let a collection of l-sequence of vectors be denoted by $\mathbf{x}_S = \{\mathbf{x}_{i_1}, \mathbf{x}_{i_2}, \ldots, \mathbf{x}_{i_{|S|}}\}$. Upper case letters denote random variables/vectors. Let $H(.)$ and $I(.;.)$ denote Shannon entropy and mutual information as given in [8]. Let $X_i, i = 1, 2, \ldots$ be a sequence of independent identically distributed (*i.i.d.*) discrete random variables drawn according to some known probability mass function for random variable $X \sim q(x)$ whose alphabet takes values in $\mathcal{X}$. We are given a reconstruction alphabet $\hat{\mathcal{X}}$, with a *bounded* distortion measure d : $\mathcal{X} \times \hat{\mathcal{X}} \to \mathbb{R}^+$ s.t $d(\cdot,\cdot) \leq d_{max}$. The distortion measure on l-sequences in $\mathcal{X}^l \times \hat{\mathcal{X}}^l$ is defined by the average per-symbol distortion

$$d(\mathbf{x}, \hat{\mathbf{x}}) = \frac{1}{l} \sum_{i=1}^{l} d(x_i, \hat{x}_i). \tag{3}$$

We shall now briefly summarize the notion of strong typicality [8, 11] for discrete valued random variables. Let $\{Z_1, Z_2, \ldots, Z_k\}$ denote a finite collection of discrete random variables with alphabets in $\mathcal{Z}_1 \times \mathcal{Z}_2 \ldots \times \mathcal{Z}_k$, with some fixed joint distribution $p(z_1, z_2, \ldots, z_k)$ for $(z_1, z_2, \ldots, z_k) \in \mathcal{Z}_1 \times \mathcal{Z}_2 \times \ldots \times \mathcal{Z}_k$. Let S denote an arbitrary ordered subset of $\mathcal{I}_k$ and let $Z_S = \{Z_{i_1}, Z_{i_2}, \ldots, Z_{i_m}\}$ for $S = \{i_1, i_2, \ldots, i_m\}$, $m = |S|$. Consider l independent copies of Z_S. Then,

$$P\left[\mathbf{Z}_S = \mathbf{z}_{i_1}, \mathbf{z}_{i_2}, \ldots, \mathbf{z}_{i_{|S|}}\right] = \prod_{i=1}^{l} P\{Z_{S,i} = z_{i_1,i}, z_{i_2,i}, \ldots, z_{i_{|S|},i}\}, \qquad \mathbf{z}_{i_k} \in \mathcal{Z}^l_{i_k}. \tag{4}$$

For a given $\mathbf{a}_S \in \mathcal{Z}_S{}^l$, let for all $b_S \in \mathcal{Z}_S$, $N(b_S; \mathbf{a}_S)$ be the number of indices $i \in \mathcal{I}_l$ such that $a_{S,i} = b_S$. By the law of large numbers [8], for any $S \subseteq \mathcal{I}_k$ and for all $b_S \in \mathcal{Z}_S$,

$$\frac{1}{l} N(b_S; \mathbf{a}_S) \to p(b_S) \tag{5}$$

and

$$-\frac{1}{l} \log\ p(Z_{S,1}, Z_{S,2}, \ldots, Z_{S,l}) = -\frac{1}{l} \sum_{i=1}^{l} \log\ p(Z_{S,i}) \to H(Z_S). \tag{6}$$

Convergence in equations (5) and (6) takes place simultaneously with probability one for all the 2^k subsets $S \subseteq \mathcal{I}_k$. The set $T_\epsilon(Z_1, Z_2, \ldots, Z_k)$ of *strongly ϵ-typical* l-sequences $(\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_k)$ is defined by

$$T_\epsilon(Z_1, Z_2, \ldots, Z_k) = \left\{ \mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_k \in \mathcal{Z}_1^l \times \mathcal{Z}_2^l \times \ldots \times \mathcal{Z}_k^l : \left| \frac{1}{l} N(b_S; \mathbf{z}_S) - p(b_S) \right| < \epsilon, \forall b_S \in \mathcal{Z}_S, \forall S \subseteq \mathcal{I}_k \right\} \tag{7}$$

Let $T_\epsilon(Z_S)$ denote the restriction of T_ϵ to the coordinates of S. Some of the important properties of T_ϵ are given below. There exists a $\delta(\epsilon) \to 0$ as $\epsilon \to 0$, such that for sufficiently large l, and any $\epsilon \to 0$,

$$P\{T_\epsilon(Z_S)\} \geq 1 - \epsilon, \forall S \subseteq \mathcal{I}_k, \tag{8}$$

$$2^{l(H(Z_S)-\delta)} \leq |T_\epsilon(Z_S)| \leq 2^{l(H(Z_S)+\delta)}, \tag{9}$$

$$2^{-l(H(Z_S)+\delta)} \leq p(\mathbf{z}_S) \leq 2^{-l(H(Z_S)-\delta)}, \forall \mathbf{z}_S \in T_\epsilon(Z_S). \tag{10}$$

2.2 Problem Statement

We now formulate the $(n, k, n-k+1)$ code achievable region problem. We are given an information source X with probability mass function $q(x)$. The goal is to find n encoding functions of the source where each is a mapping from the source symbol space to a bit stream (description or packet of information) at a rate of R bits/symbol. These bitstreams are transmitted over n independent channels such that any decoder which has access to any $m \geq k$ of them should be able to get some reconstruction of the source with a fidelity criterion given by D_m. Let $M = \{J : J \subseteq \mathcal{I}_n, |J| \geq k\}$. A maximum distance separable quantization code, $(l, \Theta, \Delta_k, \Delta_{k+1}, \ldots, \Delta_n)$ is defined by a set of n encoding functions

$$F_i : \mathcal{X}^l \rightarrow \{1, 2, \ldots, \Theta\} \;\; \forall i \in \mathcal{I}_n \tag{11}$$

and a set of decoding functions

$$G_J : \bigotimes_J \{1, 2, \ldots, \Theta\} \rightarrow \hat{\mathcal{X}}^l \quad \forall J \in M \tag{12}$$

where $\bigotimes$ denotes the Cartesian product and $\forall k \leq h \leq n$

$$\Delta_h = \frac{1}{l} E\left[d\left(\mathbf{X}, G_J\left(F_{i_1}(\mathbf{X}), F_{i_2}(\mathbf{X}), \ldots, F_{i_h}(\mathbf{X})\right)\right)\right] \;\; \forall J \subset M, |J| = h, J = \{i_1, i_2, \ldots, i_h\}. \tag{13}$$

Note: For this code, l denotes the block-length in encoding, Θ denotes the size of the index set with rate $\frac{1}{l}\log\Theta$ of each packet or description, and Δ_m denotes the reconstruction distortion when any m packets are received. Note also that we have inherently assumed symmetry in the encoding functions in the sense that for a given $|J|$ the decoded quality depends only on $|J|$ and not on which $|J|$ bitstreams are used for decoding. The latter case only makes the problem notationally cumbersome and moreover is not of much interest. Henceforth, we will be assuming symmetry in the problem.

A tuple $(R, D_k, D_{k+1}, \ldots, D_n)$ is said to be achievable if for arbitrary $\nu > 0$ there exists, for sufficiently large l, a code $(l, \Theta, \Delta_k, \Delta_{k+1}, \ldots, \Delta_n)$ with

$$\Theta \leq 2^{l(R+\nu)} \text{ and } \Delta_h \leq D_h + \nu \;\; \forall h = k, k+1, \ldots, n. \tag{14}$$

Let $\overline{\mathcal{R}}(D_k, D_{k+1}, \ldots, D_n)$ be the set of achievable rates R for the distortion tuple $D_k, D_{k+1}, \ldots, D_n$. The goal is to determine this rate region. In other words, for a given R, the goal is to optimize the fidelity criteria $\mathbf{D} = \{D_k, D_{k+1}, \ldots, D_n\}$ for sufficiently large l.

2.3 Summary of results

We now summarize the main result of this work. Let the source X be as given above. Let $p(x, y_1, y_2, \ldots, y_n)$ be a probability mass function which defines the random variables $X, Y_1, Y_2, \ldots, Y_n$, where Y_i has some alphabet $\mathcal{Y}_i$ such that

$$\sum_{y_i \in \mathcal{Y}_i, i=1,2,\ldots,n} p(x, y_1, y_2, \ldots, y_n) = q(x), \tag{15}$$

and $Y_1, Y_2, \ldots, Y_n$ are identically distributed (assuming symmetry in the problem) and conditionally independent given X implying

$$p(x, y_1, y_2, \ldots, y_n) = q(x) \prod_{i=1}^{n} p_t(y_i|x). \tag{16}$$

Random variables $Y_1, Y_2, \ldots, Y_n$ are associated with codewords that are transmitted on channels $1, 2, \ldots, n$. The main result of this work states that if the transmission rate R on each channel is given by $\{R : R > \frac{1}{k}I(X; Y_J)\}$ then with the reception of *any* $m \geq k$ packets distortion Δ_m can be attained, if $Y_1, Y_2, \ldots, Y_n$ are chosen appropriately.

Stating more formally, let $\mathcal{A}(D_k, D_{k+1}, \ldots, D_n)$ be the set of all probability mass functions $p(x, y_1, y_2, \ldots, y_n)$ as given above such that there exists a set of functions $f_J : \bigotimes_{i \in J} \mathcal{Y}_i \longrightarrow \hat{\mathcal{X}}, \forall J \subset M$ such that, $\forall h \in \{k, k+1, \ldots, n\}$,

$$E\left[d\left(X, f_J(Y_{i_1}, Y_{i_2}, \ldots, Y_{i_h})\right)\right] \leq D_h, \quad \forall J \subset M \text{ such that } |J| = h, J = \{i_1, i_2, \ldots, i_h\}. \tag{17}$$

Now corresponding to a $p(.) \in \mathcal{A}(D_k, D_{k+1}, \ldots, D_n)$, let

$$\mathcal{R}^{(p)}(D_k, D_{k+1}, \ldots, D_n) = \left\{R : R > \frac{1}{k}I(X; Y_J)\right\} \tag{18}$$

for any $J \subset M$ such that $|J| = k$, $Y_J = Y_{i_1}, Y_{i_2}, \ldots, Y_{i_k}$ and $J = \{i_1, i_2, \ldots, i_k\}$. Note that $Y_1, Y_2, \ldots, Y_n$ are identically distributed. Let

$$\mathcal{R}^*(D_k, D_{k+1}, \ldots, D_n) = \bigcup_{p \in \mathcal{A}(D_k, D_{k+1}, \ldots, D_n)} \mathcal{R}^{(p)}(D_k, D_{k+1}, \ldots, D_n) \tag{19}$$

Theorem : $\mathcal{R}^*(D_k, D_{k+1}, \ldots, D_n) \subseteq \overline{\mathcal{R}}(D_k, D_{k+1}, \ldots, D_n) \quad \forall D_i \in \mathbb{R}^+, \quad \forall k \leq i \leq n$

Remarks: The rate region is given by $R \geq \frac{1}{k}I(X; Y_1, Y_2, \ldots, Y_k)$ since the Y_i's are identically distributed. This key result says that we can achieve the rate region given in equation (18) which is the same as the case when X is jointly quantized to $Y_1, Y_2, \ldots, Y_k$ and the bitstream characterizing the quantizer index is broken into k packets with rate R bits/sample. It is rather interesting to note that we can achieve this while receiving any k packets and maintaining the same distortion and further, the reconstruction distortion is monotone decreasing function of the number of any received packets.

3 An Example of a Gaussian Source

Although we restrict ourselves in this paper to the discrete alphabet case, the bounds presented here can be generalized to the case of a Gaussian source using techniques of [12]. We now take the example of a $(3, 2, 2)$ code construction for a Gaussian source and consider the rate-distortion tuples attainable.

Consider a Gaussian source X distributed as $N(0, 1)$. Define random variables $Y_i, i = 1, 2, 3$, given by

$$Y_i = X + q_i, \tag{20}$$

where the q_i's are *i.i.d.* and are distributed[1] as $N(0, \frac{2}{15})$. With the availability of any two descriptions $i_1, i_2 \in \mathcal{I}_3$, we will be able to recover the corresponding random variables Y_{i_1}, Y_{i_2} and the reconstruction fidelity is given by the Linear Minimum Mean Square Estimate (LMMSE) of X. The expected distortion with the availability of any two descriptions (using standard Wiener estimation) is given by

$$\frac{\sigma_q^2}{\sigma_q^2 + 2} = \frac{\frac{2}{15}}{\frac{2}{15} + 2} = \frac{1}{16}. \tag{21}$$

[1] In this example, we have chosen the variance of q_i as 2/15 to have integer rates of transmission.

The rate of transmission of each of the descriptions (packets) is given by

$$R = \frac{1}{2}I(X;Y_{i_1},Y_{i_2}) = \frac{I(X;Y_{i_1}) + I(X;Y_{i_2}) - I(Y_{i_1};Y_{i_2})}{2} \tag{22}$$

Hence, R is given by

$$R = \frac{1}{4}[\log\frac{1+\sigma_q^2}{\sigma_q^2} + \log\frac{1+\sigma_q^2}{\sigma_q^2} - \log\frac{(1+\sigma_q^2)^2}{(2+\sigma_q^2)\cdot(\sigma_q^2)}] = 1 \text{ bit/sample} \tag{23}$$

Thus, it can be seen that when any two descriptions get through (together contributing 2 bits/sample of information), the observed distortion is $\frac{1}{16}$, which lies *on* the rate-distortion function of the unit-variance Gaussian source encoded at 2 bits/sample ($D(R) = 2^{-2R}$). What is interesting is that when all the three packets are received, we get a distortion that equals $\frac{\sigma_q^2}{3+\sigma_q^2} = \frac{1}{23.5}$. Thus the "parity" ends up contributing to the decoded quality !

In general, for an n-packet system, with each packet transmitted at a rate of R bits/sample, the achievable distortions with the availability of any $(k+r)$ packets of information, can be shown to be given as follows:

$$D_{k+r} = \frac{k}{2^{2kR}(k+r) - r}, \tag{24}$$

for $0 \leq r \leq (n-k)$. Note that with the availability of any k packets we are operating on the rate-distortion function with the distortion given by 2^{-2kR}. The reconstruction distortion monotonically decreases as a function of the number of received packets. Further, the ratio of the distortion with the reception of any $m \geq k$ packets to that corresponding to the reception of any k packets is given by

$$\frac{k.2^{2kR}}{2^{2kR}.m - (m-k)} \approx \frac{k}{m} \tag{25}$$

in the limit of high rate and large number of packets.

4 Proof of Theorem

We now proceed with the outline of the proof of the main theorem. The key concepts involved in the achievability of the rate region are as follows: we independently construct n quantizers and encode the source independently using these quantizers. Each quantizer codebook is partitioned into bins and each encoder sends only the index of the bin containing the quantized outcome. The sizes of these bins are designed such that with the reception of such indices from any k channels say, $(i_1, i_2, \ldots, i_k)$, it will be possible to recover the index of the quantized outcomes of the corresponding codebooks $(i_1, i_2, \ldots, i_k)$. The quantizers are designed such that the distortion with the recovery of any k quantized outcomes is at D_k. Similarly with the reception of any m $(k \leq m \leq n)$ packets, it will be possible to recover m indices of codewords in the corresponding quantizer codebooks which results in a distortion of D_m, that is strictly lower than D_k.

4.1 Encoder and Decoder

Random Coding : Let l-vectors $(\mathbf{Y}_1(1), \mathbf{Y}_1(2), \ldots, \mathbf{Y}_1(2^{l \cdot R'}))$, for some R', be drawn independently and according to a uniform distribution over the set $T_\epsilon(Y_1)$ of ϵ-typical $\mathbf{Y}_1$ l-vectors. Call this codebook $\mathcal{C}_1$. Generate similar codebooks $\mathcal{C}_i$ for each $Y_i, i = 2, 3, \ldots, n$.

Random Binning : Let $\xi = 2^{[l \cdot (R' - R + \gamma)]}$ where γ will be specified later. From the codebook $\mathcal{C}_1$ extract ξ

codewords independently and with replacement and assign them to a bin say B_{11}. Repeat this for a total of $2^{l.R}$ bins. Similar $2^{l.R}$ random bins are constructed for all $Y_i, i = 2, 3, \dots, n$ (see Figure 2).

Encoding : Given an $\mathbf{X} \in \mathcal{X}^l$, find a codeword $\mathbf{Y}_i$ from the codebook $\mathcal{C}_i$, $\forall i \in \mathcal{I}_n$ such that $(\mathbf{X}, \mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_n)$ are ϵ-jointly typical. If successful, let j_i denote the index of the codeword $\mathbf{Y}_i$ in $\mathcal{C}_i$. If not, let $j_i = 0 \;\; \forall i \in \mathcal{I}_n$. For $i = 1, 2, \dots, n$ define index $t_i(\mathbf{Y}_i)$ to be transmitted over the channel i as follows. If the l-sequence $\mathbf{Y}_i$ corresponding to the non-zero j_i belongs to at least one of the corresponding $2^{l.R}$ bins, let $t_i(\mathbf{Y}_i)$ equals the least index of a bin containing this l-sequence. Otherwise or if $j_i = 0$, set $t_i(\mathbf{Y}_i) = 0$. Index $t_i(\mathbf{Y}_i)$ is sent over the channel i.

Decoding : The decoder receives some $m \geq k$ packets of information from some m channels. From the received m packets, it takes some k of them and then searches the corresponding bins (bins with index t_{i_u} in codebook $\mathcal{C}_{i_u}$ $\forall u \in \mathcal{I}_k$) to obtain a k-tuple $j_{i_1}, j_{i_2}, \dots j_{i_k}$ of indices such that $\mathbf{Y}_1(j_{i_1}), \mathbf{Y}_1(j_{i_2}), \dots, \mathbf{Y}_k(j_{i_k})$ are ϵ-jointly typical l-sequences. If there exists more than one such tuple, the decoder declares error and uses the reconstruction vector $\hat{\mathbf{X}}$ as any arbitrary $\hat{\mathbf{x}} \in \hat{\mathcal{X}}^l$.

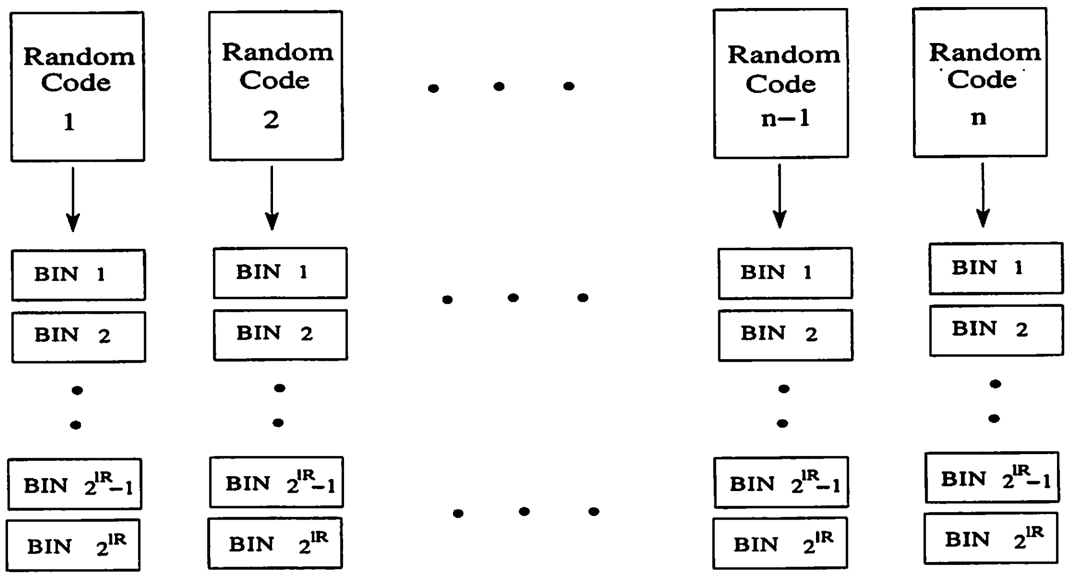

Figure 2: Random code construction: n independent random codebooks are constructed each with $2^{lR'}$ codewords. Each codebook is randomly partitioned into 2^{lR} bins each with $2^{l(R'-R)}$ codewords.

Let us make a small observation here. Assume that it is possible to guarantee that the corresponding codebook indices $j_{i_1}, j_{i_2}, \dots j_{i_k}$ indices can be recovered when some k out of n packets are received. Then if $m > k$ packets are received the corresponding codebook indices $j_{i_1}, j_{i_2}, \dots j_{i_m}$ will also be recovered exactly. This can be seen by repeatedly decoding the m codebook indices in sets of k at a time. We thus need to ensure that whenever some k packets get through the corresponding codebook indices can be recovered perfectly with very high probability and this would ensure with very high probability that $m > k$ indices can be decoded correctly in the event that $m > k$ indices are received.

4.2 Analysis of error events

Let us define the following error events which can lead to a decoding error E at the decoder.

Error Events:

1. E_0: X^l does not belong to $T_\epsilon(X)$.
2. E_1: There exists no indices $(j_1, j_2, \ldots j_n)$ such that $(\mathbf{X}, \mathbf{Y}_1(j_1), \mathbf{Y}_2(j_2), \ldots \mathbf{Y}_n(j_n))$ are jointly typical.
3. E_2: Not all channel indices t_i are greater than zero.
4. E_3: For some set of k received indices $\{t_{i_1}, t_{i_2}, \ldots, t_{i_k}\}$, there exist another set of l-vectors $\{\mathbf{Y}_{i_1}(j'_{i_1}), \mathbf{Y}_{i_2}(j'_{i_2}), \ldots, \mathbf{Y}_{i_k}(j'_{i_k})\}$ that are ϵ-jointly typical and belong to the same corresponding bins.

$E = \bigcup_{i=0}^{3} E_i$ and the probability of error is bounded above by :

$$P(E) \leq \sum_{i=0}^{3} P(E_i) \tag{26}$$

Bounding $P(E_0)$: By the well known property of typical sets (equation (8)), $P(E_0) \to 0$, as l is made sufficiently large.

Bounding $P(E_1)$: For any arbitrary randomly and independently chosen sequences $(\mathbf{X}, \mathbf{Y}'_1, \mathbf{Y}'_2, \ldots, \mathbf{Y}'_n) \in \mathcal{X}^l \times \mathcal{Y}_1^l \times \mathcal{Y}_2^l \times \ldots \times \mathcal{Y}_n^l$, the probability T that they are ϵ-jointly typical is given by :

$$T \geq \frac{2^{l(H(X,Y_1,Y_2,\ldots Y_n)-\delta_{X,Y_1,Y_2,\ldots Y_n})}}{2^{l(H(X)+\delta_X)}.2^{l(H(Y_1)+\delta_{Y_1})}.2^{l(H(Y_2)+\delta_{Y_2})}\ldots.2^{l(H(Y_n)+\delta_{Y_n})}} = \zeta \tag{27}$$

which follows from the property of typical sets and the constants $\delta_{X,Y_1,Y_2,\ldots,Y_n}, \delta_X, \delta_{Y_i}$ are chosen according to equation (9). Thus $P(E_1)$ can be bounded as

$$P(E_1) \leq (1-\zeta)^{2^{n \cdot l \cdot R'}} \tag{28}$$

Using the standard inequality [8] $(1-y)^n \leq e^{-y \cdot n}$, $\forall\ 0 \leq y \leq 1$, and $n > 0$, we can conclude that $P(E_1) \to 0$ when

$$n \cdot R' > \sum_{i=1}^{n} H(Y_i) - H(Y_1, Y_2, \ldots Y_n|X) - \delta' \tag{29}$$

where $\delta' = \delta_{X,Y_1,Y_2,\ldots Y_n} + \delta_X + \delta_{Y_1} + \delta_{Y_2} + \ldots + \delta_{Y_n}$ However, $\{Y_1, Y_2, \ldots, Y_n\}$ are conditionally independent given X implies that

$$H(Y_1, Y_2, \ldots Y_n|X) = \sum_{i=1}^{n} H(Y_i|X) \tag{30}$$

Since δ' can be chosen to be arbitrarily small, from equations (29) and (30) it follows that

$$n \cdot R' > \sum_{i=1}^{n} (H(Y_i) - H(Y_i|X)) = \sum_{i=1}^{n} I(X; Y_i) = n \cdot I(X; Y_h) \quad \forall h \in \mathcal{I}_n \tag{31}$$

Hence $R' > I(X; Y_h) \quad \forall h \in \mathcal{I}_n$ would suffice.

Bounding $P(E_2)$: To prove $P(E_2) \to 0$ we will show that $P(t_1(\mathbf{Y}_1) > 0) \to 1$. By symmetry it will follow that $P(t_i(\mathbf{Y}_i) > 0) \to 1\ \forall\ i = 1, 2, \ldots, n$. Since the intersection of a finite number of sets of probability 1 is a set with probability 1 it will follow that $P(E_2^c) \to 1$ as l is made sufficiently large. The probability of finding a $\mathbf{Y}_1$ in codebook $\mathcal{C}_1$ that is ϵ-jointly typical with the given $\mathbf{X}$ tends to one for the chosen codebook size. Hence, it would suffice to show that

$$G = P\left[t_1(\mathbf{Y}_1) > 0 | \mathbf{Y}_1 \in \mathcal{C}_1, (\mathbf{X}, \mathbf{Y}_1) \in T_\epsilon(X, Y_1)\right] \to 1 \tag{32}$$

Consider $Q = 1 - G$. Let A denote the event that $[\mathbf{Y}_1 \in \mathcal{C}_1, (\mathbf{X}, \mathbf{Y}_1) \in T_\epsilon(X, Y_1)]$. Then

$$Q = P[\mathbf{Y}_{1_j} \neq \mathbf{Y}_1, 1 \leq j \leq N | A] \tag{33}$$

where $N = \xi \cdot 2^{l \cdot R} = 2^{l \cdot (R' + \gamma)}$ is the total number of $\mathbf{Y}_1$ selected for all bins and $\mathbf{Y}_{1_j}$ is the j^{th} such selected $\mathbf{Y}_1$. Thus, $Q = (P[\mathbf{Y}_{1_1} \neq \mathbf{Y}_1 | A])^N$. This is because each $\mathbf{Y}_{1_j}$ has the same chance of equaling $\mathbf{Y}_1$. Since the cardinality of $\mathcal{C}_1$ is $2^{l \cdot R'}$, the desired probability $Q = (1 - 2^{-l \cdot R'})^N$. Using the inequality $\log z \leq (z - 1)$ we get

$$\log Q \leq -2^{l \cdot (R' + \gamma)} \cdot 2^{-l \cdot (R')} = -2^{l \cdot \gamma} \tag{34}$$

Hence if $\gamma > 0$, $\log Q \to -\infty$. Equivalently, $Q \to 0$ as desired.

Bounding $P(E_3)$: We will now find a bound on $P(E_3)$ using a constraint on R which guarantees with high probability that there is a unique set of l-vectors with one from each bin (whose index is received from the encoder) which are ϵ-jointly typical. Let $t_{i_1}, t_{i_2} \ldots t_{i_k}$ be the k indices received corresponding to channels $i_1, i_2 \ldots i_k \in \mathcal{I}_n$. Then the probability ($P$) that one or more k-tuple of l-vectors other than those used by the encoder are jointly typical is bounded by

$$P \leq (\xi^k - 1) \cdot P(\mathbf{Y}^*_{i_1}, \mathbf{Y}^*_{i_2}, \ldots \mathbf{Y}^*_{i_k} \in T_\epsilon(Y_{i_1}, Y_{i_2}, \ldots, Y_{i_k})) \tag{35}$$

where $\mathbf{Y}^*_{i_j}$ is a randomly chosen vector from $T_\epsilon(Y_{i_j})$. The second term on the right hand side of equation (35) is bounded by

$$P[\mathbf{Y}^*_{i_1}, \mathbf{Y}^*_{i_2}, \ldots, \mathbf{Y}^*_{i_k} \in T_\epsilon(Y_{i_1}, Y_{i_2}, \ldots, Y_{i_n})] \leq \frac{2^{l[H(Y_{i_1}, Y_{i_2}, \ldots Y_{i_k}) + \delta_{Y_{i_1}, Y_{i_2}, \ldots Y_{i_k}})}}{2^{l(H(Y_{i_1}) - \delta_{Y_{i_1}})} \cdot 2^{l(H(Y_{i_2}) - \delta_{Y_{i_2}})} \ldots 2^{l(H(Y_{i_k}) - \delta_{Y_{i_k}})}} \tag{36}$$

where $\delta_{Y_{i_1}, Y_{i_2}, \ldots, Y_{i_k}}$, $\delta_{Y_{i_h}}$ for $h \in \mathcal{I}_k$ are defined as before and let $\delta'' = \delta_{Y_{i_1}, Y_{i_2}, \ldots, Y_{i_k}} + \delta_{Y_{i_1}} + \delta_{Y_{i_2}} + \ldots + \delta_{Y_{i_k}}$. Thus we have,

$$P \leq 2^{k \cdot l(R' - R + \gamma)} \cdot 2^{l(H(Y_{i_1}, Y_{i_2}, \ldots Y_{i_k}) - H(Y_{i_1}) - H(Y_{i_2}) \ldots - H(Y_{i_k}) + \delta'')} \tag{37}$$

Thus, $P \to 0$ if

$$k(R' + \gamma) + H(Y_{i_1}, Y_{i_2}, \ldots Y_{i_k}) - H(Y_{i_1}) - H(Y_{i_2}) - \ldots - H(Y_{i_k}) + \delta'' < k.R \tag{38}$$

and using the fact that

$$\sum_{h \in J} I(X; Y_h) - \sum_{h \in J} H(Y_h) + H(Y_J) = I(X; Y_J), \tag{39}$$

we essentially obtained the condition mentioned in equation (18). This argument thus shows that $P(E_3) \to 0$ for l sufficiently large. Now for the case when the number of channel indices received $m > k$, we can similarly argue that with high probability we can recover all the corresponding m codebook indices.

We will now argue that the expected distortion is asymptotically unaffected. Let $\mathbf{d}$ be the observed distortion over the m channels. We note that $E\{\mathbf{d}|E^c\} \leq \mathbf{D} + \epsilon \mathbf{1}$ by the construction [11] of T_ϵ where $\mathbf{1}$ is the vector of (m) ones and $\mathbf{D} = (D_k, D_{k+1}, \ldots, D_n)$. Hence,

$$E(\mathbf{d}) \leq P(E^c)(\mathbf{D} + \epsilon \mathbf{1}) + P(E) \cdot \mathbf{1}.d_{max} \tag{40}$$

where d_{max} is the maximum value of the bounded distortion measure d and ^{c} denotes complementation. Since the expected distortion of this construction process is as desired we conclude that from amongst the ensemble of encoder-decoder sets there exists atleast one set of encoders-decoders for which the attained distortion is as desired.

Q. E. D.

5 Conclusion and Future Work

We thus have provided an information theoretic achievable rate region for the problem of $(n, k, n-k+1)$ maximum distance separable quantization code. The decoder starts decoding with the reception of any k packets. The reception of any of the remaining packets contributes to the improvements in the reconstruction fidelity. We also considered an example of Gaussian distribution where with such a code we are operating on the rate-distortion function with the reception of any k packets and further, as the number of received packets increases, there is a monotone increasing gain in the reconstruction fidelity. Ongoing and future work includes actual construction of such $(n, k, n-k+1)$ quantization codes based on group codes [1] such as trellis codes and lattice codes and integration of such codes with different rates to build an efficient protocol for the packet erasure networks.

Acknowledgements: We would like to thank V. Vaishampayan and S. Diggavi of AT&T for inspiring discussions.

References

[1] F. J. Macwilliams and N. J. A. Sloane, *The Theory of Error Correcting Codes.* Elseiver-North-Holland, 1977.

[2] A. E. Gamal and T. M. Cover, "Achievable Rates for Multiple Descriptions," *IEEE Transactions on Information Theory*, vol. 28, pp. 851–857, November 1982.

[3] V. A. Vaishampayan, "Design of Multiple Description Scalar Quantizers," *IEEE Transactions on Information Theory*, vol. 3, pp. 821–834, May 1993.

[4] J. Albanese, A. amd Blomer, J. Edmonds, M. Luby, and M. Sudan, "Priority Encoded Transmission," *IEEE Transactions on Information Theory*, vol. 42, pp. 1737–1744, November 1996.

[5] R. Puri and K. Ramchandran, "Multiple Description Source Coding Through Forward Error Correction Codes," *33rd Asilomar Conference on Signals, Systems and Computers*, October, Pacific Grove, CA, 1999.

[6] A. E. Mohr, E. A. Riskin, and R. E. Ladner, "Unequal Loss Protection: Graceful Degradation of Image Quality over Packet Erasure Channels through Forward Error Correction," *Proceedings of the Data Compression Conference (DCC)*, March, Snowbird, UT, 1999.

[7] R. L. Urbanke and A. D. Wyner, "Packetizing for the Erasure Broadcast Channel with an Internet Application," *Preprint*, July 1998.

[8] M. T. Cover and J. A. Thomas, *Elements of Information Theory.* John Wiley and Sons, 1991.

[9] D. K. Slepian and J. K. Wolf, "Noiseless Coding of Correlated Information Sources," *IEEE Transactions on Information Theory*, vol. 19, pp. 471–480, July 1973.

[10] S. S. Pradhan and K. Ramchandran, "Group Theoretic Construction and Analysis of Generalized Coset Codes for Asymmetric-Symmetric Distributed Source Coding," *Proceedings Conference on Information Sciences and Systems (CISS)*, March Princeton, NJ, 2000.

[11] T. Berger, "Multiterminal Source Coding," *Information Theory Approach to Communication, (CISM Courses and Lecture Notes No. 229)*, G. Longo, Ed., Wien and New York: Springer-Verlag, 1977.

[12] R. G. Gallager, *Information Theory and Reliable Communication.* New York: John Wiley and Sons, 1968.