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

S. Sandeep Pradhan, Rohit Puri
and Kannan Ramchandran[1]
EECS Department
Univ. of California, Berkeley, USA
e-mail:
{pradhan5,rpuri,kannanr}@eecs.berkeley.edu

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.**

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.

I. 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 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 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.

[1]This work was supported by ONR Grant number N00014-97-1-0864.

The problem of transmission of information sources over unreliable channels 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 distortion region has been considered, and constructive approaches have been proposed. Some of the 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 packet-erasure networks. Information content with higher priority is encoded using lower rate MDS codes while those with lower priority is encoded using a correspondingly higher rate code. Algorithms for image and video transmission systems have been proposed in [5] that use progressive source coding techniques concatenated with MDS codes.

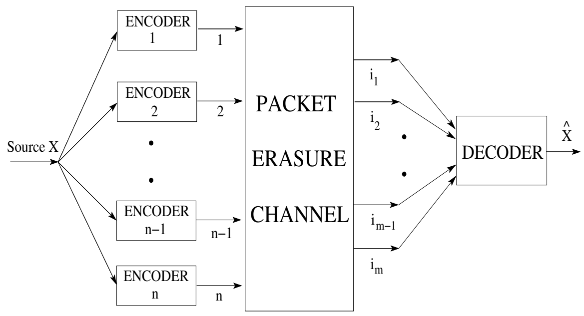

Figure 1: $(n, k, n-k+1)$ source-channel erasure code: A source is encoded into n descriptions or packets and transmitted over n errorless channels which are equally likely to breakdown. The decoder would like to reconstruct with the availability of any $m \geq k$ packets of information.

The construction of erasure channel codes has traditionally been considered in a discrete finite field setting [1]. A typical information source transmission scenario that invokes the principles of the source-channel separation theorem [6] involves taking a source coded stream of rate kR (k packets of rate R) and converting it to a channel-coded stream of rate nR for a desired k, n and R, using an $(n, k, n-k+1)$ code. Such a scheme exhibits the "cliff" effect - the reception of fewer than k packets is not sufficient to recover any data, while the reception of *any* k packets enables the recovery of the original k source packets and any further reception of packets do not aid any further recovery. In this work, we present a new approach based on random binning arguments that extends these concepts into the source symbol domain and at the same time serves to soften the "cliff" effect i.e., 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 any $m > k$ 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 fact, the proposed framework is very general in the sense that it subsumes the performance of the approaches based on standard MDS codes as a special case [1]. While in this paper we undertake an information-theoretic analysis of this joint source-channel coding problem, the actual construction of such encoders and decoders based on group codes, inspired by these arguments, is of great interest and is a part of our ongoing work.

A key result of this paper implies that using such an $(n, k, n-k+1)$ 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, for $2^{2kR} >> r$, the improvement is linear in the number of packets received, i.e.,

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

Thus for the specific case of a Gaussian source the proposed joint source-channel coding paradigm strictly outperforms the source-channel separation theorem based approach.

A Motivation

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 $\mathcal{C}$, having rate-distortion [6] performance with rate R bits/sample, i.e. attaining distortion $D(R)$ that lies on the rate-distortion function of the source. The output bitstream at this stage will consist of indices of codewords from this codebook and 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 $D(R)$. 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 achieving 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 $(n, 1, n)$ "source-channel" code as opposed to a traditional channel code. Based on this motivation, we generalize these ideas to prove the existence of $(n, k, n-k+1)$ source-channel codes in the following sections that outperform the conventional $(n, k, n-k+1)$ code constructed on finite alphabets (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 [7], 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 [8]. The goal is to encode a set of multiple correlated distributed sources independently (remotely) while exploiting the joint correlation. The encoder of each source sends only carefully designed partial information to facilitate reconstruction at the decoder with the availability of collections of such partial information from all the sources. In this paper we apply some of these techniques to the problem of $(n, k, n-k+1)$ source-channel erasure code.

The rest of the paper is organized as follows. Section II introduces the problem and states the main result. Section III presents the implication of these results for the Gaussian sources while Section IV sketches the outline of the proof of the main result, details of which are presented in the Appendix. In Section V we draw conclusions and offer directions for future work.

II. Problem Formulation

A 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 [6]. 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 : $\mathcal{X} \times \hat{\mathcal{X}} \rightarrow \mathcal{R}^+$ such that $d(\cdot, \cdot) \leq d_{max}$. The distortion measure on l-sequences in $\mathcal{X}^l \times \hat{\mathcal{X}}^l$ is defined by the average per-symbol distortion

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

[1] The simple modifications required for the same are presented at the end of Section IV.)

We shall now briefly summarize the notion of strong typicality [6, 9] 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,

$$P\left[\mathbf{Z}_S = \mathbf{z}_{i_1}, \ldots, \mathbf{z}_{i_{|S|}}\right] = \prod_{i=1}^{l} P\{Z_{S,i} = z_{i_1,i}, \ldots, z_{i_{|S|},i}\} \quad (4)$$

where $\mathbf{z}_{i_k} \in \mathcal{Z}_{i_k}^l$. 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 [6], for any $S \subseteq \mathcal{I}_k$ and for all $b_S \in \mathcal{Z}_S$,

$$\frac{1}{l} N(b_S; \mathbf{a}_S) \rightarrow p(b_S), \quad (5)$$

$$-\frac{1}{l} \log\, p(Z_{S,1}, \ldots, Z_{S,l}) = -\frac{1}{l} \sum_{i=1}^{l} \log\, p(Z_{S,i}) \rightarrow H(Z_S) \quad (6)$$

are true. Convergence in (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

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

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 $\delta(\epsilon) \rightarrow 0$ as $\epsilon \rightarrow 0$, such that for sufficiently large l, and any $\epsilon \rightarrow 0$,

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

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

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

B Problem Statement

We now formulate the problem of achievable region associated with the $(n, k, n-k+1)$ code. Assume an information source X with probability mass function $q(x)$. The goal is to find a set of encoding functions (n in number) and decoding functions for the source, where the former is a mapping from the source symbol space to a bit stream (description or packet of information) at a rate of R bits/symbol, and the latter a mapping from the space of bitstream indices to the source reconstruction space. The 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. Note that we have considered a symmetric system such that with the reception of any m packets, the reconstruction fidelity is at the same level, D_m. Thus the decoded quality is a function of only the *number* of descriptions received. Hence, from a distortion performance point of view there are $n-k+1$ decoding functions.

More formally, let $M = \{J : J \subseteq \mathcal{I}_n,\ |J| \geq k\}$. A source channel erasure code, $(l, \Theta, \Delta_k, \Delta_{k+1}, \ldots, \Delta_n)$ is defined by a set of n encoding functions

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

and a set of decoding functions

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

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

$$\Delta_h = \frac{1}{l} E\left[d\left(\mathbf{X}, G_J\left(F_{i_1}(\mathbf{X}), \ldots, F_{i_h}(\mathbf{X})\right)\right)\right], \quad J = (i_1, \ldots, i_h). \quad (13)$$

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 that for a given J, the decoded quality depends only on $|J|$ and not on which $|J|$ bitstreams are received.

Definition: 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 block length l, a code $(l, \Theta, \Delta_k, \Delta_{k+1}, \ldots, \Delta_n)$ with

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

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.

C Summary of results

Let the source $X \sim q(x)$ be as given above. Let $Y_1, Y_2, \ldots, Y_n$ be n identically distributed random variables where Y_i has some alphabet $\mathcal{Y}_i$ such that $p(x, y_1, y_2, \ldots, y_n) = q(x) \cdot p(y_1, y_2, \ldots, y_n | x)$ denotes their joint probability mass function with X. We now summarize an information-theoretic characterization of the achievable rate region.

Theorem: *An achievable rate region for a given distortion tuple $(D_k, D_{k+1}, \ldots, D_n)$ is given by*

$$R > \frac{1}{k} H(Y_1, Y_2, \ldots, Y_k) - \frac{1}{n} H(Y_1, Y_2, \ldots, Y_n | X). \quad (15)$$

for some probability mass functions $p(x, y_1, y_2, \ldots, y_n) = q(x) \cdot p(y_1, y_2, \ldots, y_n | x)$, and a set of decoding functions $g_J : \bigotimes_{i \in J} \mathcal{Y}_i \longrightarrow \hat{\mathcal{X}}$, $\forall J \in M$, such that $\forall h \in \{k, k+1, \ldots, n\}$,

$$E\left[d\left(X, g_J(Y_{i_1}, Y_{i_2}, \ldots, Y_{i_h})\right)\right] \leq D_h, \quad J = (i_1, \ldots, i_h). \quad (16)$$

That is, if R is as given by (15) then with the reception of *any* k packets distortion D_k is achieved and with the reception of $m > k$ packets $D_m \leq D_k$ is achieved. The reception of extra parity bits thus serves to enhance the received quality !

Note for the special case when $Y_1, Y_2, \ldots, Y_n$ are conditionally independent given X, we have

$$\frac{1}{k} H(Y_1, \ldots, Y_k) - \frac{1}{n} H(Y_1, \ldots, Y_n | X) = \frac{1}{k} I(X; Y_1, \ldots, Y_k). \quad (17)$$

This can be interpreted as follows: we can achieve the rate region given in (17) 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 a monotonic decreasing function of the number of any received packets.

Before we prove the main result of this work we evaluate the rate region described above for the special case of a Gaussian source in the following section.

III. The Case 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 [10]. We now take the example of a $(n, k, n-k+1)$ 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 $Y_i, i = 1, \ldots, n$, given by

$$Y_i = X + q_i, \tag{18}$$

where the q_i's are *i.i.d.* and are distributed as $N(0, \frac{k}{2^{2\cdot k\cdot R'}-1})$ for some $R' > 0$. In this example, we have chosen the variance of q_i as $\frac{k}{2^{2\cdot k\cdot R'}-1}$ so that with the reception of *any* k descriptions corresponding to k Y_is, the optimum minimum mean squared error which is given by $\frac{\sigma_q^2}{\sigma_q^2+k}$ equals $2^{-2\cdot k\cdot R'}$, the distortion value on the distortion-rate curve for the Gaussian source corresponding to the rate kR'.

The rate of transmission of each of the descriptions (packets) is given by (15). Let us calculate the individual terms in (15).

$$H(Y_1, Y_2, \ldots, Y_n|X) = \frac{1}{2}\log_2(2\pi e)^n(\sigma_q^2)^n \tag{19}$$

Let C denote the correlation matrix for $Y_1, Y_2, \ldots, Y_n$. It can be shown that $det(C) = (k+\sigma_q^2)(\sigma_q^2)^{k-1}$. Hence,

$$H(Y_1, \ldots, Y_k) = \frac{1}{2}\log_2(2\pi e)^k(k+\sigma_q^2)(\sigma_q^2)^{k-1} \tag{20}$$

Substituting these values in (15) we observe that $R = R'$! That is, when the transmission rate on each channel is R, then with the reception of *any* k packets, the reconstructed distortion is that which corresponds to a source coding rate of kR.

Furthermore, the achievable distortions with the availability of any $(k+r)$ packets of information, can be shown to be given as follows:

$$D_{k+r} = \frac{k}{2^{2kR}(k+r)-r}, \tag{21}$$

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 $k+r \geq k$ packets to that corresponding to the reception of any k packets is given by

$$\frac{k.2^{2kR}}{2^{2kR}.(k+r)-r} \approx \frac{k}{k+r} \tag{22}$$

for $2^{2kR} >> r$.

This is strictly superior to an encoding scheme where a conventional source coded bit stream of rate kR is encoded using an (n,k) MDS code into n packets each with rate R. Thus we have shown the potential for significant improvements in performance of media transmission over unreliable channels by espousing joint source-channel coding techniques over techniques based on conventional source-channel separation principles.

IV. Outline of Proof

We now proceed with a brief outline of the proof of the main result, the details of which are presented in the Appendix.

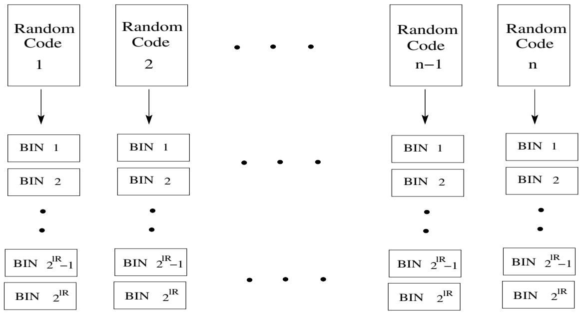

Figure 2: Random code construction: n independent random codebooks are constructed each with $2^{lR'}$ codewords. Each codebook is randomly partitioned into 2^{lR} bins each with $2^{l(R'-R)}$ codewords.

Given source $X \sim q(x)$, fix $p(y_1, \ldots, y_n|x)$ such that (16) holds. The random variables $Y_1, \ldots, Y_n$ are identically distributed.

Random Coding: Using the marginal densities $p(y_i)$, $i = 1, \ldots, n$ generate n codebooks each containing $2^{l\cdot R'}$ codewords of blocklength l with each of the symbols being drawn independently. We need the size R' to be sufficiently big so that for any source word of length l that occurs, we are able to find at least one codeword in each of the n codebooks such that these n codewords are jointly typical with the source word. This happens if

$$n \cdot R' > \sum_{i=1}^{n} H(Y_i) - H(Y_1, \ldots, Y_n|X) \tag{23}$$

Random Binning: Each of the codebooks is further independently and randomly partitioned into $2^{l\cdot R}$ bins, each having approximately $2^{l\cdot(R'-R)}$ codewords (see Figure 2).

Encoding: Given a source word $\mathbf{X} \in \mathcal{X}^l$ has occurred, find a codeword $\mathbf{Y}_i$ from the codebook $\mathcal{C}_i$, such that ($\mathbf{X}$, $\mathbf{Y}_1$, $\mathbf{Y}_2$, $\ldots$, $\mathbf{Y}_n$) are jointly typical. The index of the bin in which the codeword $\mathbf{Y}_i$ lies is transmitted over the ith channel.

Decoding: The decoder receives some $m \geq k$ indices. It takes some k of them and searches through the corresponding bins to identify a k tuple of codewords that is jointly typical. By virtue of the encoding strategy, there is at least one such k tuple, namely the one that was desired to be transmitted. To ensure that there are no other such k tuples we need to ensure that the size of each bin is small enough so that the probability of finding another such k tuple is very small. The decoder has to search over $2^{k\cdot(R'-R)}$ possible combinations. The probability p that a random k tuple of codewords is jointly typical is roughly given by $p = \frac{2^{l\cdot(H(Y_1,\ldots,Y_k))}}{2^{l\cdot(H(Y_1))}.2^{l\cdot(H(Y_2))}\ldots2^{l\cdot(H(Y_k))}}$. Intuitively the desired probability goes to zero if $2^{k\cdot(R'-R)} < 1/p$, i.e.,

$$k \cdot (R' - R) < \sum_{i=1}^{k} H(Y_i) - H(Y_1, \ldots, Y_k) \qquad (24)$$

Combining (23) and (24) we get the desired result in (15). We make another small note here. Suppose with very high probability it can be guaranteed that with the reception of k indices the correct codewords can be recovered. Then in the event of receiving more than k codewords we can use the above decoding procedure repeatedly on sets of k indices at a time and recover all the indices correctly. Furthermore because of the choice of the probability mass function (see (16)) we will be able to guarantee all the desired distortions. A more detailed treatment of the proof is presented in the Appendix.

Remark: The framework presented above can be modified slightly to subsume the case of the separation theorem based encoding strategy. For a given source $X \sim q(x)$ fix $p(y|x)$ and using the marginal density $p(y)$ generate a codebook with $2^{lR'}$ codewords with blocklength l with each symbol drawn independently. From the standard rate-distortion theorem [6] it can be shown that with very high probability we can find a codeword that is jointly typical with any observed source codeword if $R' > I(X;Y)$. This codebook is now repeated n times to generate n identical codebooks. Each of these identical codebooks is randomly and independently partitioned into 2^{lR} bins as done in Section IV. The encoder sends the index of the bin containing the intended codeword in the ith codebook over the ith channel. The decoder with the reception of any $m \geq k$ packets searches the corresponding bins to identify the codeword that is present in all of them. Using arguments similar to that presented in Section VI, it can be shown that the probability of decoding failure can be made arbitrarily small if $kR > R'$. This essentially means that if $kR > I(X;Y)$ then with the reception of any k packets we are operating on the distortion-rate function of the source corresponding to rate kR, thus achieving the performance of the conventional MDS code.

V. Conclusion and Future Work

We have provided an information theoretic achievable rate region for the problem of $(n, k, n-k+1)$ maximum distance separable source-channel code. The decoder starts decoding with the reception of any k packets. The reception of any more packets contributes to improvements in the reconstruction fidelity. We have also considered the interesting special case of Gaussian sources where with such a code we can strictly outperform the approach based on the source-channel separation principle. Our ongoing work includes practical construction of such $(n, k, n-k+1)$ source-channel codes based on group codes [1] such as trellis codes and lattice codes and integration of such codes in a framework similar to [5] to build efficient protocols for multimedia transmission over packet erasure networks.

VI. Appendix

A 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 $Y_i, i = 2, 3, \ldots, n$ (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 $(\mathbf{X},\ \mathbf{Y}_1,\ \mathbf{Y}_2,\ \ldots, \mathbf{Y}_n)$ are ϵ-jointly typical. If successful, let j_i denote the index of the codeword $\mathbf{Y}_i$ in C_i. If not, let $j_i = 0 \quad \forall i \in \mathcal{I}_n$. For $i = 1, 2, \ldots, n$ 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 $j_{i_1}, j_{i_2}, \ldots j_{i_k}$ of indices such that $\mathbf{Y}_{i_1}(j_{i_1}), \mathbf{Y}_{i_2}(j_{i_2}), \ldots, \mathbf{Y}_{i_k}(j_{i_k})$ 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 $\hat{\mathbf{x}} \in \hat{\mathcal{X}}^l$.

Let us make a small observation here. Assume that it is possible to guarantee that the corresponding codebook indices $j_{i_1}, j_{i_2}, \ldots j_{i_k}$ indices can be recovered when some k out of n packets are received. Then if $m > k$ packets are received the corresponding codebook indices $j_{i_1}, j_{i_2}, \ldots j_{i_m}$ will also be recovered exactly. This can be seen by repeatedly decoding the m codebook indices in sets of k at a time.

B 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: $\mathbf{X}$ 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:

$$P(E) \leq \sum_{i=0}^{3} P(E_i) \qquad (25)$$

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

Bounding $P(E_1)$: Using the properties of strongly typical sets and standard arguments as given in [9, 2], it can be shown [2] that $P(E_1)$ can be made arbitrarily small for large l if $\forall J \subset \mathcal{I}_n$

$$R' > H(Y_1) - \frac{1}{|J|} H(Y_1, Y_2, \ldots, Y_{|J|}|X). \qquad (26)$$

[2] We thank V. A. Vaishampayan and S. Diggavi of AT & T for their help in developing the proof.

In the symmetric case as considered in this paper, it turns out that all of the above constraints can be replaced by the following single constraint:

$$n \cdot R' > \sum_{i=1}^{n} H(Y_i) - H(Y_1, Y_2, \ldots, Y_n|X) \tag{27}$$

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

$$G = P\left[t_1(\mathbf{Y}_1) > 0 | \mathbf{Y}_1 \in \mathcal{C}_1, (\mathbf{X}, \mathbf{Y}_1) \in T_\epsilon(X, Y_1)\right] \to 1 \tag{28}$$

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

$$Q = P[\mathbf{Y}_{1_j} \neq \mathbf{Y}_1, 1 \leq j \leq N|A] \tag{29}$$

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

$$\log Q \leq -2^{l \cdot (R' + \gamma)} \cdot 2^{-l \cdot (R')} = -2^{l \cdot \gamma} \tag{30}$$

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

$$P \leq (\xi^k - 1) \cdot P(\mathbf{Y}^*_{i_1}, \mathbf{Y}^*_{i_2}, \ldots \mathbf{Y}^*_{i_k} \in T_\epsilon(Y_{i_1}, Y_{i_2}, \ldots, Y_{i_k})) \tag{31}$$

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 (31) is bounded by

$$P[\mathbf{Y}^*_{i_1}, \mathbf{Y}^*_{i_2}, \ldots, \mathbf{Y}^*_{i_k} \in T_\epsilon(Y_{i_1}, Y_{i_2}, \ldots, Y_{i_n})] \leq \frac{2^{l[H(Y_{i_1}, Y_{i_2}, \ldots Y_{i_k}) + \delta_{Y_{i_1}, Y_{i_2}, \ldots Y_{i_k}})}}{2^{l(H(Y_{i_1}) - \delta_{Y_{i_1}})} \ldots 2^{l(H(Y_{i_k}) - \delta_{Y_{i_k}})}} \tag{32}$$

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,

$$P \leq 2^{k \cdot l(R' - R + \gamma)} \cdot 2^{l(H(Y_{i_1}, \ldots Y_{i_k}) - H(Y_{i_1}) - H(Y_{i_2}) \ldots - H(Y_{i_k}) + \delta'')} \tag{33}$$

Thus, $P \to 0$ if

$$k(R' + \gamma) + H(Y_{i_1}, \ldots Y_{i_k}) - H(Y_{i_1}) \ldots - H(Y_{i_k}) + \delta'' < k.R \tag{34}$$

We obtain the condition mentioned in (15). 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 [9] of T_ϵ where $\mathbf{1}$ is the vector of (m) ones and $\mathbf{D} = (D_k, D_{k+1}, \ldots, D_n)$. Hence,

$$E(\mathbf{d}) \leq P(E^c)(\mathbf{D} + \epsilon\mathbf{1}) + P(E) \cdot \mathbf{1}.d_{max} \tag{35}$$

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 at least one set of encoders-decoders for which the attained distortion is as desired.

Acknowledgments

We would like to thank V. A. Vaishampayan and S. Diggavi of AT&T for inspiring discussions and useful comments.

References

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

[2] A. El 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] A. Albanese, J. Blomer, J. Edmonds, M. Luby and M. Sudan, "Priority Encoded Transmission",*IEEE Transactions on Information Theory*, vol. 42, no. 6, 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*, Pacific Grove, CA, October 1999.

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

[7] 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.

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

[9] 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.

[10] R. G. Gallager, "Information Theory and Reliable Communication", John Wiley and Sons, New York 1968.