RICE UNIVERSITY

Joint Source/Channel Coding for Discrete Memoryless Channels : Lessons To Learn

by

Vinay K. Bharadwaj

A Thesis Submitted
in Partial Fulfillment of the
Requirements for the Degree

Master of Science

Approved, Thesis Committee:

Behnaam Aazhang, Chair
Professor of Electrical and Computer Engineering

Joseph R. Cavallaro
Associate Professor of Electrical and Computer Engineering

Rob Nowak
Assistant Professor of Electrical and Computer Engineering

Houston, Texas

April, 2000

Joint Source/Channel Coding for Discrete Memoryless Channels : Lessons To Learn

Vinay K. Bharadwaj

Abstract

The design of optimal joint source/channel coding and decoding is examined for discrete memoryless channels with end-to-end distortion as the criterion for reliable communication. Joint source/channel encoders which map sequences of source symbols directly to sequences of channel symbols without any intermediate "bit" representation of source are considered. Optimum joint source/channel decoder that minimizes end-to-end distortion for a given encoder mapping is derived. The encoder mapping can be many to one, in the sense that many source sequences can be mapped to one sequence of channel symbols. To begin with, as an exercise, random coding bound on end-to-end distortion is derived for a general Maximum A Posteriori (MAP) decoder which has some estimate on the apriori probabilities of source symbols. It is shown that, the KL distance of the actual apriori probabilities with the estimated ones plays an important role. Then, a random coding bound on end-to-end distortion is derived with our optimal minimum distortion decoder mentioned above for the case when all source symbols are equally likely. It is shown that the performance increase with minimum distortion decoding as opposed to MAP (same as Maximum Likelihood (ML) decoding in this case when all source symbols are equally likely) is characterized by the faster decay of end-to-end distortion with respect to channel use.

Acknowledgments

I express my gratitude to my advisor, Prof. Behnaam Aazhang for his invaluable support in promoting independent thinking and freedom in research work. I also would like to thank Dr. Ashutosh Sabharwal(Postdoctoral Fellow, Rice University) and Prof. Mike Orchard (Professor, Princeton University) for all the discussions we had during the course of research, without which this work would not be a possibility.

Vinay K. Bharadwaj

Contents

Illustrations

Chapter 1

Introduction

According to classical theory of communication, source coders and channel coders can be designed separately to achieve reliable communication [7]. A discrete source modeled as a discrete random process can be compressed by assigning longer codewords(mostly binary in practice) to less likely source symbol sequences and shorter codewords to more likely sequences. Using this idea, the average number of bits required to represent a random source is given by the entropy of the source. And by taking larger and larger input symbols at a time, it is known that one can design binary codewords, such that the average length of the codewords is as as close as possible to the entropy of the source. Here, the mappings from source sequences to binary codewords are considered to be lossless, i.e., for every possible source sequence there exists a unique binary codeword. And the ideal source coder will produce a bitstream that is iid, i.e., 0 and 1 occur equally likely in the the output of the ideal source coder.

Now, the stream of bits are to be sent over a noisy channel by adding redundancy in a controlled fashion through channel coding. As Shannon's famous result says, when source entropy is less than channel capacity, then the stream of bits can be sent "reliably" over the channel by appropriate design of channel encoders and decoders. Here the notion of reliable communication is : as channel use increases, there is an exponential decay in average probability of error between the input and the decoded stream of bits. And the converse of Shannon's result says, when the source entropy is greater than channel capacity, reliable communication is not possible. Implicit in these results is the separate design of the source and channel coders. Recently in [9], an information-stable source/channel pair is cited which does not satisfy the separation theorem. Moreover, new necessary and sufficient conditions called domination and strict domination are derived for reliable communication with joint source and channel coding for a general class of sources and channels. But still, the criteria of reliable communication is average probability of error.

The research literature on joint source/channel(JSC) coding when criteria of reliable communication is end-to-end distortion, can be classified into 3 classes[5]. The first class is denoted as "Concatenated" JSC coding. Here known source coders and channel coders are placed in cascade, with bits allocated for source coder and channel coder can vary with overall rate constraint. For example, in [4] a Pearlman and Said progressive image coder is cascaded with rate compatible convolutional coder (RCPC) along with a bit-allocation algorithm optimized for end-to-end distortion. Also Hochwald[3], has considered the problem of finding channel code rate that optimally allocates the available transmission rate between lossy source coding and block channel coding over a noisy discrete channel. In second class denoted as "Constrained" JSC coding, the source coder and/or receiver are modified to account for the presence of a noisy channel. For example, Vaishampayan and Farvardin[8] derive optimal block cosine transform for noisy channels. Works that use some knowledge of the source or source coder properties to detect and compensate for channel errors are also included in this class of constrained JSC coding. In [5], a constrained JSC coder is described, wherein, the output of the source coder is not considered to be iid, but instead the correlations are captured with a markov assumption through training. These markov transition probabilities are then used in the channel decoder to improve system performance.

The third class is the "True" JSC coding, wherein the source and channel coding operations have been truly integrated. There are very few papers in this research area. In 1978, Massey[6], showed that for distortionless transmission of the source under the constraint of linear source and channel coders, a significant reduction in complexity with equivalent performance can be achieved with JSC coding. However, hardly any optimal decoders other than for Guassian noise are linear. So our goal here is to study the design of true joint source/channel coders and decoders for discrete memoryless channels. In [1], it is shown that differential pulse code modulation (DPCM) of video signals results in a non-symmetric source, and an optimal minimum error and minimum end-to-end distortion 2D signal set design has been proposed. Here our effort, will be to be to generalize the result when channel codewords are picked from more than two dimensions.

In this work, we address the general topic of the design of "True" JSC encoders and decoders with end-to-end distortion as the criteria for reliable communication. Chapter 1 discusses the problem we wish to address. Chapter 2 derives the optimal decoder that minimizes end-to-end distortion and demonstrates it with an example of two codewords. Chapter 3, then generalizes it to many codeword case, using random coding [2] and comparison is made with MAP/ML decoding.

Chapter 2

Problem Formulation

Consider the block diagram in Fig 2.1. The source produces one of A symbols - $(a_0, \dots, a_{A-1})$ with probabilities $\Omega(a_0), \dots, \Omega(a_{A-1})$. We wish to communicate L source symbols taken at a time with N uses of the channel. If M is the number of codewords (of dimensionality N), for the lossless mapping from source sequences to codewords $M = A^L$ and for the lossy case $M < A^L$. The number of possible received sequences for v^L is M and let this set be $\mathcal{V}_M$. The metric for reliable communication is the end-to-end distortion

$$D_e = \sum_{u^L} \sum_{v^L \neq u^L, v^L \in \mathcal{V}_M} D(u^L, v^L) P(u^L, v^L) \tag{2.1}$$

where $D(.,.) \to \mathcal{R}$ be a bounded distortion function defined on two possible sequences of L symbols. (We will use z^L and $\mathbf{z}$ interchangeably to represent a vector $(z_1, z_2, \dots, z_L)$). Let the number of channel input and output alphabets be K and J

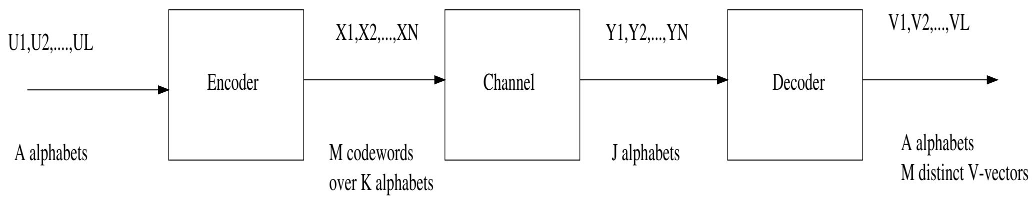

Figure 2.1 Overall Block Diagram

respectively. So any encoder is a choice of M codewords from the space of K^N. Any decoder is a partition of the space of J^N into M decision regions. The rate of the code is given by

$$R_s = \frac{L}{N} \text{ source symbols(sy) per channel use(cu)} \tag{2.2}$$

where R_s can be directly related to the the rate at which source is producing symbols. If the time duration of transmission of each channel alphabet is τ_c seconds, (which is directly limited by the bandwidth constraint on the channel), the source rate is

$$R_{abs} = \frac{R_s}{\tau_c} = \frac{L}{N\tau_c} \text{ source symbols/second.} \tag{2.3}$$

If $H(U)$ is the source entropy rate, then the number of bits required to represent this is source on an average is $H(U)$ bits per source symbol. So the source data rate at which we are communicating is

$$R_{bps} = R_{abs}H(U) = \frac{LH(U)}{N\tau_c} \text{ bits/second.} \tag{2.4}$$

Yet another, the most common way of representation of source rate is

$$R = \frac{LH(U)}{N} \text{ bits/cu.} \tag{2.5}$$

A rate R_s (in sy/cu) is achievable if end-to-end distortion (2.1) decays exponentially with N. Our immediate goal will be to find the set of achievable data rates and the maximum possible rate of decay for a given channel. And our long-term goal will be to design practical joint source/channel encoders, decoders at these achievable rates. In the next chapter, we first derive the optimal joint source/channel decoder and demonstrate it with a simple example.

Chapter 3

Optimal Minimum Distortion Decoding

In this chapter, we derive the optimal decoder which minimizes end-to-end distortion. Given that M codewords are sent over the channel, the decoder has to divide the space of received vector $\mathbf{y} = (y_1, \ldots, y_N)$ into M regions $R_0, R_1, \ldots, R_{M-1}$, such that the end-to-end distortion function defined in (2.1) is minimized. To consider the problem in its generality, we allow many to one mappings from source sequences to codewords i.e., the encoder can be lossy. For a source sequence $\mathbf{u}$, let $\mathcal{M}(\mathbf{u})$, be the codeword associated with it and let S_j, be the set of $\mathbf{u}$ mapped to the codeword $\mathbf{x_j}$. Also at the decoder, $\mathbf{x_j}$ is decoded to a unique source sequence $\mathbf{u} \in S_j$, denoted as $\mathcal{M}^{-1}(\mathbf{x_j})$. So number of possible sequences for $\mathbf{v}$ is M and let this set be $\mathcal{V}_M$. So now the end-to-end distortion function in (2.1) will be

$$D_e = \sum_{\mathbf{u}} \sum_{\mathbf{v} \in \mathcal{V}_M} D(\mathbf{u}, \mathbf{v}) P(\mathbf{v} \mid \mathbf{u}) P(\mathbf{u}) \tag{3.1}$$

Theorem 3.1 The optimal decoder which minimizes (3.1) decodes the received vector $\mathbf{y}$ into codeword m, i.e $\mathbf{y} \in R_m$, if

$$m = \arg[\min_{0 \le i \le (M-1)} \sum_{j=0, j \ne i}^{M-1} c_{ij} P(\mathbf{y} \mid \mathbf{x_j})] \tag{3.2}$$

where

$$c_{ij} = \sum_{\mathbf{u} \in S_j} D(\mathbf{u}, \mathcal{M}^{-1}(\mathbf{x_i})) P(\mathbf{u}) \quad 0 \le j \le M-1 \tag{3.3}$$

Proof :

The theorem is very general its valid for any function $D(.,.) \to \mathcal{R}$. The idea of proof is very similar to Maximum A-Posteriori(MAP) decoder minimizing probability of error. The c_{ij}'s form the weighting of the likelihood probabilities.

Given $\mathbf{u} \in S_j$, $P(\mathbf{v} \mid \mathbf{u}) = P(\mathbf{v} \mid \mathbf{x_j})$. So from (3.1),

$$\begin{aligned} D_e &= \sum_{\mathbf{v} \in \mathcal{V}_M} \sum_{j=0}^{M-1} \sum_{\mathbf{u} \in S_j} D(\mathbf{u}, \mathbf{v}) P(\mathbf{u}) P(\mathbf{v} \mid \mathbf{x_j}) \\ &= \sum_{i=0}^{M-1} \sum_{R_i} \sum_{j=0}^{M=1} (\sum_{\mathbf{u} \in S_j} D(\mathbf{u}, \mathcal{M}^{-1}(\mathbf{x_i}))) P(\mathbf{u}) P(\mathbf{y} \mid \mathbf{x_j}) \\ &= \sum_{i=0}^{M-1} \sum_{R_i} (\sum_{j=0}^{M=1} c_{ij} P(\mathbf{y} \mid \mathbf{x_j})) \end{aligned}$$

Now a given received vector $\mathbf{y}$ should be placed in the region R_i, which has the minimum expression within its sum and hence the result.

3.1 Two Codewords-An Example

In this section, the optimal minimum distortion decoding is illustrated for a simple example of two codewords over a power constrained Guassian channel. If the source has four symbols $\{a_0, a_1, a_2, a_3\}$, then more than one mapping exists to map the source symbols on to two codewords. To start with lets say, a_0 and a_1 are mapped to codeword $\mathbf{x_0}$, a_2 and a_3 to $\mathbf{x_1}$. Also let $\mathcal{M}^{-1}(\mathbf{x_0}) = a_0$ and $\mathcal{M}^{-1}(\mathbf{x_1}) = a_2$. From (3.1), replacing sums by integrals

$$\begin{aligned} D_e &= \int_{R_0} [d(a_1, a_0) P(a_1) P(\mathbf{y} \mid \mathbf{x_0}) && (3.4) \\ &\quad + \{d(a_2, a_0) P(a_2) + d(a_3, a_0) P(a_3)\} P(\mathbf{y} \mid \mathbf{x_1})] d\mathbf{y} && (3.5) \\ &\quad + \int_{R_1} [\{d(a_0, a_2) P(a_0) + d(a_1, a_2) P(a_1)\} P(\mathbf{y} \mid \mathbf{x_0}) && (3.6) \\ &\quad + d(a_3, a_2) P(a_3) P(\mathbf{y} \mid \mathbf{x_1})] d\mathbf{y} && (3.7) \end{aligned}$$

So the decision rule for minimum distortion decoding, in terms of coefficients c_ij

$$c_{00}P(\mathbf{y} \mid \mathbf{x_0}) + c_{01}P(\mathbf{y} \mid \mathbf{x_1}) \underset{\mathbf{x_0}}{\overset{\mathbf{x_1}}{\gtrless}} c_{10}P(\mathbf{y} \mid \mathbf{x_0}) + c_{11}P(\mathbf{y} \mid \mathbf{x_1}) \tag{3.8}$$

$$\frac{P(\mathbf{y} \mid \mathbf{x_1})}{P(\mathbf{y} \mid \mathbf{x_0})} \underset{\mathbf{x_0}}{\overset{\mathbf{x_1}}{\gtrless}} \frac{c_{10} - c_{00}}{c_{01} - c_{11}} = \gamma \tag{3.9}$$

Since its a Guassian channel, the sufficient statistic is linear and its a matched filter.

$$\Upsilon = < \mathbf{y}, \mathbf{x_1} > - < \mathbf{y}, \mathbf{x_0} > \underset{\mathbf{x_0}}{\overset{\mathbf{x_1}}{\gtrless}} \sigma^2 \log\gamma + \frac{\| \mathbf{x_1} \|^2 - \| \mathbf{x_0} \|^2}{2} \tag{3.10}$$

The sufficient statistic is $\sum_{i=0}^{M-1}(x_{1i} - x_{0i})y_i$ which is a linear sum of Guassians. The minimum distortion will be

$$\begin{aligned} D_e &= c_{00}\int_{R_0} P(\mathbf{y} \mid \mathbf{x_0})\mathbf{dy} + c_{01}\int_{R_0} P(\mathbf{y} \mid \mathbf{x_1})\mathbf{dy} && (3.11) \\ &\quad + c_{10}\int_{R_1} P(\mathbf{y} \mid \mathbf{x_0})\mathbf{dy} + c_{11}\int_{R_1} P(\mathbf{y} \mid \mathbf{x_1})\mathbf{dy} && (3.12) \\ \int_{R_0} P(\mathbf{y} \mid \mathbf{x_0})\mathbf{dy} &= P[\Upsilon < \sigma^2\log\gamma + \frac{\| \mathbf{x_1} \|^2 - \| \mathbf{x_0} \|^2}{2} \mid \mathbf{x_0} \text{ sent}] && (3.13) \\ &= 1 - Q(\frac{\sigma^2\log\gamma + \frac{\|\mathbf{x_1} - \mathbf{x_0}\|^2}{2}}{\sigma \| \mathbf{x_1} - \mathbf{x_0} \|}) && (3.14) \end{aligned}$$

Similarly calculating other integrals and using (3.9), $Q(-x) = 1 - Q(x)$ and $\mathbf{x_d} = \mathbf{x_1} - \mathbf{x_0}$

$$\begin{aligned} D_e &= c_{00}[1 - Q(\frac{\sigma^2\log\gamma + \| \mathbf{x_d} \|^2 /2}{\sigma \| \mathbf{x_d} \|})] + c_{01}[1 - Q(\frac{\sigma^2\log\gamma - \| \mathbf{x_d} \|^2 /2}{\sigma \| \mathbf{x_d} \|})] \\ &\quad c_{10}Q(\frac{\sigma^2\log\gamma + \| \mathbf{x_d} \|^2 /2}{\sigma \| \mathbf{x_d} \|}) + c_{11}Q(\frac{\sigma^2\log\gamma - \| \mathbf{x_d} \|^2 /2}{\sigma \| \mathbf{x_d} \|}) \\ &= c_{00} + c_{11} + (c_{10} - c_{00})Q(\frac{\sigma^2\log\gamma + \| \mathbf{x_d} \|^2 /2}{\sigma \| \mathbf{x_d} \|}) \\ &\quad + (c_{01} - c_{11})[1 - Q(\frac{\sigma^2\log\gamma - \| \mathbf{x_d} \|^2 /2}{\sigma \| \mathbf{x_d} \|})] \\ &= c_{00} + c_{11} + (c_{01} - c_{11})[\gamma Q(\frac{\| \mathbf{x_d} \|}{2\sigma} + \frac{\log\gamma}{\| \mathbf{x_d} \| /\sigma}) \\ &\quad + \; Q(\frac{\| \mathbf{x_d} \|}{2\sigma} - \frac{\log\gamma}{\| \mathbf{x_d} \| /\sigma})] && (3.15) \end{aligned}$$

Now in (3.15), $c_{00} + c_{11} = d(a_1, a_0)P(a_1) + P(a_3, a_2)P(a_3)$ is the distortion due to lossy, many to one encoding. The other part is the distortion due to the channel. $\parallel \mathbf{x_d} \parallel /\sigma$ represents the signal to noise ratio(SNR). Its infact the distance between the signals divided by the noise standard deviation. So after optimal decoding, the end-to-end distortion is a joint function of SNR and γ, which is a function of source symbol distortions, symbol apriori probabilities and the mapping from source symbols to channel codewords.

Our objective now is given that our signals are constrained by power P i.e $\parallel \mathbf{x_1} \parallel^2 < P$ and $\parallel \mathbf{x_0} \parallel^2 < P$, how do we further minimize end-to-end distortion with respect to the choice of codewords $\mathbf{x_1}$, $\mathbf{x_0}$ and the mapping which determines the c_{ij}'s and γ. Denoting $s = \parallel \mathbf{x_d} \parallel$, let

$$\begin{aligned} f(s,\gamma) &= \gamma Q(\frac{s}{2\sigma} + \frac{\log\gamma}{s/\sigma}) + Q(\frac{s}{2\sigma} - \frac{\log\gamma}{s/\sigma}) \\ \frac{\partial f(s,\gamma)}{\partial s} &= \gamma Q'(\frac{s}{2\sigma} + \frac{\log\gamma}{s/\sigma})(\frac{1}{2\sigma} - \frac{\sigma\log\gamma}{s^2}) + Q'(\frac{s}{2\sigma} - \frac{\log\gamma}{s/\sigma})(\frac{1}{2\sigma} + \frac{\sigma\log\gamma}{s^2}) \end{aligned}$$

Using $Q'(x) = -1/\sqrt{2\pi}e^{-x^2}$

$$\frac{\partial f(s,\gamma)}{\partial s} = -\sqrt{\gamma/2\pi\sigma^2}\exp^{-(\frac{s^2}{4\sigma^2} + \frac{(\sigma\log\gamma)^2}{s^2})} \leq 0 \tag{3.16}$$

This says that the function $f(s,\gamma)$ is non-increasing function with respect to s. So the expression of D_e in (3.15) is a non-increasing function of $\parallel \mathbf{x_d} \parallel$, i.e more the distance between the two codewords lesser will be D_e, independent of the value of γ. This happens for the two codeword case only as the only parameter to design about the signal set is $\parallel \mathbf{x_d} \parallel$. As one can visualize, if there are more than two codewords, its not true that the distances between codewords independently affect D_e.

Similarly,

$$\begin{aligned}\frac{\partial f(s,\gamma)}{\partial\gamma} &= Q(\frac{s}{2\sigma}+\frac{\sigma\log\gamma}{s})+\gamma Q'(\frac{s}{2\sigma}+\frac{\sigma\log\gamma}{s})\frac{\sigma}{s\gamma}+Q'(\frac{s}{2\sigma}-\frac{\sigma\log\gamma}{s})\frac{-\sigma}{s\gamma}\\ &= Q(\frac{s}{2\sigma}+\frac{\sigma\log\gamma}{s})\ \geq 0\end{aligned}$$

So $f(s,\gamma)$ is non-decreasing with respect γ. But $(c_{01}-c_{11})$ and $(c_{00}+c_{11})$ and γ are jointly determined by the source symbol to channel codeword mapping. So D_e has to be optimized over all possible mappings from source symbol to channel codewords. This is usually hard, but this is overcome by using random channel codewords, which is dealt in the next chapter.

3.2 Set of Non-Achievable Rates

In this section, we re-assure the fact that for rates greater than capacity the end-to-end distortion cannot be made arbitrary small. For lossless encoding, $\mathcal{V}_M$ will represent all possible input sequences and hence

$$P_e=\sum_{u^L}\sum_{v^L\neq u^L}P(u^L,v^L) \tag{3.17}$$

then,

$$D_e\geq D_{min}P_e \tag{3.18}$$

where $D_{min}\neq 0$ is the minimum distortion between two different sequences of L source symbols. From the converse to the coding theorem, we know that P_e cannot be made arbitrarily small for rates above channel capacity of the channel. Equivalently, the maximum achievable rate is

$$R_{s(max)}=\frac{C}{H(U)}\ \text{ sy/cu} \tag{3.19}$$

where C is the per-use capacity of the channel in bits and for R (bits/cu) $> C$, end-to-end distortion is bounded away from zero. So no matter what form of joint or separate source and channel coders we have, the end-to-end distortion cannot be made arbitrarily small for rates above capacity. So for optimal minimum distortion decoding too, all achievable rates are upto capacity of the channel. But, we will show that, the reason its optimal because it has a faster rate of decay with respect to channel use and we will quantify it in comparison with the rate of decay of MAP/ML decoders.

Chapter 4

Many Codewords - Random Coding Bounds

In this section, we consider the problem of expressing the minimum distortion after optimal decoding in terms of channel coding parameters and source coding parameters in the case of more than two codewords. If we adopt the approach used for the two codeword case (section 3.1), even for simple channels like Gaussian and Binary Symmetric Channel (BSC), evaluating exact expression for minimum distortion is not possible. This is due to the fact that, closed form expressions of probabilities over decision regions specified by optimal decoding cannot be obtained. The problem at hand is very similar to the case of obtaining minimum P_e expression with Maximum Likelihood (ML) or Maximum A Posteriori (MAP) decoding. The solution is to use random codes and bound P_e [2].

4.1 MAP Decoding

In this section we derive the random coding bound on end-to-end distortion for MAP decoding in a general setup. This will give us the understanding and tools to derive random coding bound for minimum distortion decoding that we derived in chapter 3. Let $q_0, q_1, \ldots, q_{M-1}$ be the probabilities of occurrence of M source vectors of L symbols. If source is iid and the ith sequence is $(i_0, \ldots, i_{L-1})$ (each i_j can be one of the A source symbols $a_0, \ldots, a_{A-1}$)

$$q_i = \prod_{j=0}^{L-1} \Omega(i_j) \tag{4.1}$$

where $\Omega(a_0), \ldots, \Omega(a_{A-1})$ are the probabilities of occurrence of the source symbols. If the decoder suppose does not know the exact apriori probabilities of source symbols, but instead uses some estimated probabilities $\Omega_r(.)$. In [2] a simple case where $\Omega_r(.) = \Omega(.)$, is setup as an exersize problem. Let $r_0, \ldots, r_{M-1}$ be the corresponding probabilities on L sequences. The source uncertainity per alphabet is

$$H(U) = \sum_{k=0}^{A-1} \Omega(a_k) \log(\frac{1}{\Omega(a_k)}) \text{ bits} \tag{4.2}$$

MAP decoder decodes the received $\mathbf{y}$ to the message m that has maximum $P(\mathbf{y}|\mathbf{x_m})r_m$.

The rate R of the encoder in bits per channel use is given by (2.5). Hence the encoder has to choose $M = 2^{NR}$ codewords from the space of K^N vectors. As suggested by Shannon, its hard to bound P_e (or D_e) for a particular set of codewords, but one can bound P_e (or D_e) for an ensemble of randomly chosen codewords [2]. If codewords are picked independently from a distribution $Q(\mathbf{x})$, following the notation of [2], probability of error when a message m enters the encoder is

$$\overline{P}_{e,m} = \sum_{\mathbf{x_m}} \sum_{\mathbf{y}} Q(\mathbf{x_m}) Pr[\text{error} \mid m, \mathbf{x_m}, \mathbf{y}] \tag{4.3}$$

Similarly for end-to-end distortion,

$$\overline{D}_{e,m} = \sum_{\mathbf{x_m}} \sum_{\mathbf{y}} Q(\mathbf{x_m})[\text{Expected distortion} \mid m, \mathbf{x_m}, \mathbf{y}] \tag{4.4}$$

If $D(.,.)$ is bounded, then its easy to see that

$$\overline{D}_{e,m} \leq D_{max} \overline{P}_{e,m} \tag{4.5}$$

where $D_{max} = max_{i \neq j} D(i,j)$. So the rate of decay of average distortion over the ensemble of random codes is bounded above by the rate of decay of average probability of error. So $\overline{D}_{e,m}$ is upper bounded if $\overline{P}_{e,m}$ is upper bounded.

Theorem 4.1 Let a DMC have transition probabilities $P(j \mid k)$. Let the source be iid having A alphabets and probability measure $\Omega(.)$. Let L sequences of the source be transmitted over N uses of DMC. Consider the ensemble of $M = A^L$ codewords in which each alphabet of each codeword is independently selected with probability assignment $Q(k)$. Then, for all $0 \leq \rho \leq 1$, the ensemble average distortion with MAP decoding which uses apriori probabilities $\Omega_r(.)$ satisfies

$$\overline{D}_e \leq 2^{-NE_0(\rho,Q)+LE_s(\rho)} = 2^{-N(E_0(\rho,Q)-E_s(\rho)R_s)} \tag{4.6}$$

$$E_0(\rho, Q) = -\log \sum_{j=0}^{J-1} [\sum_{k=0}^{K-1} Q(k) P(j \mid k)^{1/(\rho+1)}]^{\rho+1} \tag{4.7}$$

$$E_s(\rho) = \rho \log(\sum_{i=0}^{A-1} \Omega_r(a_i)^{1/(\rho+1)}) + \log(\sum_{i=0}^{A-1} \frac{\Omega(a_i)}{\Omega_r(a_i)} \Omega_r(a_i)^{1/(\rho+1)}) \tag{4.8}$$

Proof :

For a given m,$\mathbf{x_m}$,$\mathbf{y}$, define the event $A_{m'}$ for each $m' \neq m$, as the event that codeword $\mathbf{x_{m'}}$ is selected in such a way that $P(\mathbf{y}|\mathbf{x_{m'}})r_{m'} > P(\mathbf{y}|\mathbf{x_m})r_m$. Then

$$Pr[\text{error} \mid m, \mathbf{x_m}, \mathbf{y}] \leq P(\bigcup_{m' \neq m} A_{m'}) \tag{4.9}$$

$$\leq [\sum_{m' \neq m} P(A_{m'})]^\rho; \quad \text{any } \rho, \ 0 < \rho < 1 \tag{4.10}$$

where the second inequality is the union bound [2]. Using

$$P((A_{m'}) \leq \sum_{\mathbf{x_{m'}}} Q(\mathbf{x_{m'}}) \frac{P(\mathbf{y}|\mathbf{x_{m'}})^s r_{m'}^s}{P(\mathbf{y}|\mathbf{x_m})^s r_m^s}; \quad \text{any } s > 0 \tag{4.11}$$

Since $\mathbf{x_m}$'s are iid, $\mathbf{x_{m'}}$ is a dummy variable in the above summation. Using this fact in equation (4.10),

$$Pr[\text{error} \mid m, \mathbf{x_m}, \mathbf{y}] \leq \frac{1}{P(\mathbf{y}|\mathbf{x_m})^{s\rho} r_m^{s\rho}} [\sum_{m' \neq m} r_{m'}^s]^\rho \ [\sum_{\mathbf{x}} Q(\mathbf{x}) P(\mathbf{y}|\mathbf{x})^s]^\rho \tag{4.12}$$

Using this in (4.3), yields

$$\overline{P}_{e,m} \leq \frac{[\sum_{m' \neq m} r_{m'}^{s}]^{\rho}}{r_m^{s\rho}} \sum_{\mathbf{y}} [\sum_{\mathbf{x_m}} Q(\mathbf{x_m}) P(\mathbf{y}|\mathbf{x_m})^{1-s\rho}] [\sum_{\mathbf{x}} Q(\mathbf{x}) P(\mathbf{y}|\mathbf{x})^{s}]^{\rho} \tag{4.13}$$

To obtain the tightest bound, it can be shown that $s = 1/(\rho + 1)$ [2]. So the average probability of error for the ensemble of random codes

$$\overline{P_e} = \sum_m \overline{P}_{e,m} q_m \tag{4.14}$$

$$\leq [\sum_m \frac{q_m}{r_m} r_m^{1/(\rho+1)}][\sum_m r_m^{1/(\rho+1)}]^{\rho} \sum_{\mathbf{y}} [\sum_{\mathbf{x}} Q(\mathbf{x}) P(\mathbf{y}|\mathbf{x})^{1/(\rho+1)}]^{\rho+1} \tag{4.15}$$

Using the form of (4.1),

$$\sum_m r_m^{1/(\rho+1)} = \sum_{i_0, \ldots, i_{L-1}} [\Omega_r(i_0) \ldots \Omega_r(i_{L-1})]^{1/(\rho+1)} \tag{4.16}$$

$$= [\sum_{i=0}^{A-1} \Omega_r(a_i)^{1/(\rho+1)}]^{L} \tag{4.17}$$

Similarly, writing out first expression in (4.15), the first two expressions give rise to $2^{LE_s(\rho)}$ in (4.6). And for DMC's it can shown that the third expression in (4.15) yields $2^{-NE_0(\rho,Q)}$ in (4.6) [2].

Theorem 4.1, upper bounds distortion over the ensemble of random codes. From (4.6), if R_s is such that $E_0(\rho, Q) - E_s(\rho)R_s$ is positive, then there exists at least one code in the ensemble of random codes picked independently from the distribution $Q(.)$, for which distortion is exponentially decaying with channel use. In order to find the tightest bound, we have to maximize $E_0(\rho, Q) - E_s(\rho)R_s$ with respect to ρ, Q. Given $Q(.)$, it can be shown that $E_0(\rho, Q)$ is non-negative,non-decreasing,concave function of ρ with slope at $\rho = 0$ as mutual information [2]. Also, by differntiating (4.8), $E_s(\rho)$ can be shown to be non-negative, non-decreasing, convex function of ρ, with slope at zero as $H(U) + D_{\mathcal{K}}(\Omega \parallel \Omega_r)$. $D_{\mathcal{K}}(.,.)$ is the KL distance between two probability measures over the same source alphabet. So for exponent to be decaying

$$0 < R_s < \frac{C}{H(U) + D_{\mathcal{K}}(\Omega \parallel \Omega_r)} \quad \text{sy/cu} \tag{4.18}$$

Comparing the above to (3.19), larger the KL distance between apriori probabilities used by the MAP decoder and the original ones, set of achievable rates are less than $C/H(U)$. Fig 4.1 illustrates all cases for a binary source and a BSC at rate 0.3 bits/cu. The curve for 'MAP' uses exact apriori probabilities $(0.1, 0.9)$, 'Good MAP' uses $(0.2, 0.8)$ and 'Bad MAP' uses $(0.8, 0.2)$. 'ML' curve in Fig 4.1 corresponds to Maximum Likelihood decoder, wherein $\Omega(.) = \Omega_r(.) = 1/A$. The maximum vertical distance between 'E_0' and 'Good MAP' curves is almost two times the maximum vertical distance between 'E_0' and 'ML' curves. This implies that by exactly knowing the apriori probabilities of source symbols, the rate of decay of distortion can be doubled. Also the closeness of 'Good MAP' and 'MAP' curves, suggests that if the estimated apriori probabilities are good (in the sense of KL distance from the actual probabilities) then we are not losing much in terms of rate of decay of distortion. In the next section, we would like to similarly quantify the gains we can get in the rate of decay of distortion by knowing distortions between source symbols as opposed to apriori probabilities..

4.2 Optimal Minimum Distortion Decoding

In this section we derive the random coding bound on end-to-end distortion for discrete memoryless channels with optimal minimum distortion decoding (Chapter 3). If we assume all source symbols are equally likely, then minimum distortion decoding depends on distortions between two source sequences and the channel transition probabilities. So our objecctive of Theorem 4.2 will be to show that, the rate of decay of ensemble average end-to-end distortion will be faster than for ML (same as MAP)

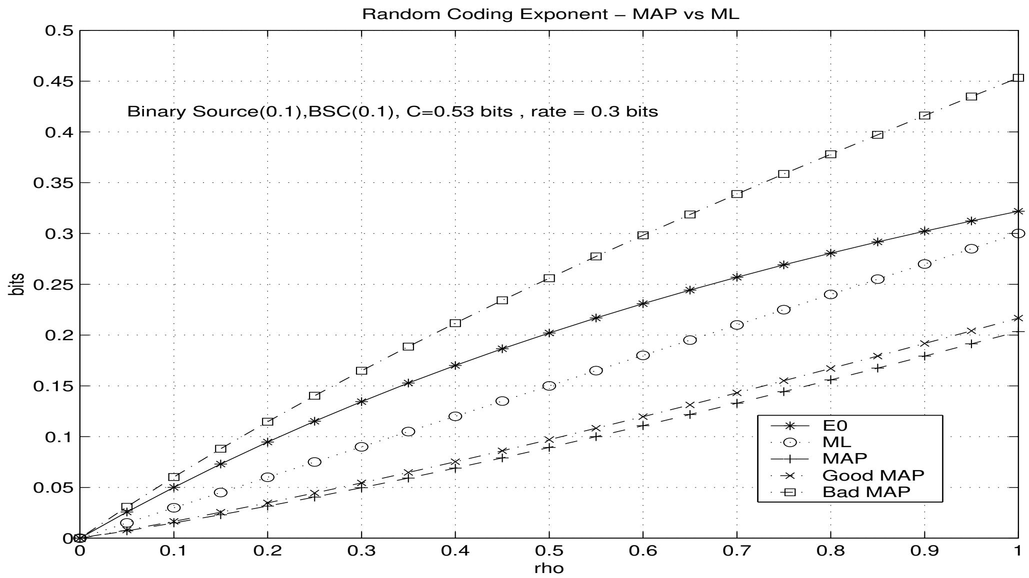

Figure 4.1 Random Coding Exponent-MAP vs ML

decoding, which from (4.6) is

$$\overline{D}_e \leq 2^{-NE_0(\rho,Q)+\rho L \log A} = 2^{-N(E_0(\rho,Q)-\rho R)}$$

Theorem 4.2 Let a DMC have transition probabilities $P(j \mid k)$. Let the source be iid having A equiprobable alphabets. Also let $D(u^L, v^L)$ defined on two L sequences be the sum of L individual distortion between corresponding alphabets. Let L sequences of the source be transmitted over N uses of DMC. Consider the ensemble of $M = A^L$ codewords in which each alphabet of each codeword is independently selected with probability assignment $Q(k)$. Then, for all $0 \leq \rho \leq 1$, the ensemble average distortion with Minimum Distortion Decoding satisfies

$$\overline{D}_e \leq L! \binom{A}{2} 2^{-NE_2(\rho,Q)} = 2^{-N(E_2(\rho,Q)-\rho E_d(A,L)R)} \tag{4.19}$$

where

$$E_2(\rho, Q) = -\log[\sum_{j=0}^{J-1}\sum_{k=0}^{K-1}(p(j)Q(k))^{\rho}p(j,k)^{1-\rho}] \quad (4.20)$$

$$E_d(A, L) = \frac{\log[L!\begin{pmatrix} A \\ 2 \end{pmatrix}]}{\log[A^L]} \quad (4.21)$$

Proof :

From representation defined in chapter 3, every L sequence of alphabets u^L has a codeword $\mathcal{M}(u^L) = \mathbf{x_i}$. For simplicity lets denote $D(\mathcal{M}^{-1}(\mathbf{x_i}), \mathcal{M}^{-1}(\mathbf{x_j}))$, the distortion between two source sequences of L alphabets mapped to codewords $\mathbf{x_i}$ and $\mathbf{x_j}$ respectively by D_{ij}. The distortion between two individual source alphabets a_l and a_m is denoted by d_{lm}. From (4.5), we know it suffices to upper bound $\overline{P}_{e,m}$ to upper bound $\overline{D}_{e,m}$. From (4.3),

$$\overline{P}_{e,m} = \sum_{\mathbf{x_m}}\sum_{\mathbf{y}} Q(\mathbf{x_m})Pr[\text{error} \mid m, \mathbf{x_m}, \mathbf{y}] \quad (4.22)$$

From (4.10), the union bound, we have

$$Pr[\text{error} \mid m, \mathbf{x_m}, \mathbf{y}] \leq [\sum_{m' \neq m} P(A_{m'})]^{\rho}; \quad \text{any}\,\rho, \quad 0 < \rho < 1 \quad (4.23)$$

For a given m,$\mathbf{x_m}$,$\mathbf{y}$, define the event $A_{m'}$ for each $m' \neq m$, as the event that codeword $\mathbf{x_{m'}}$ is selected in such a way that

$$\sum_{l=0, l\neq m'}^{M-1} D_{m'l}P(\mathbf{y} \mid \mathbf{x}_l) < \sum_{l=0, l\neq m}^{M-1} D_{ml}P(\mathbf{y} \mid \mathbf{x}_l) \quad (4.24)$$

Given, $\mathbf{y}$, $\mathbf{x_m}$, for every $l \neq m$, $P(\mathbf{y} \mid \mathbf{x}_l)$ is a random variable which is a function of $\mathbf{X}_l$. $\mathbf{X}_l$ is the random codeword picked from a distribution $Q(\mathbf{x})$. So the above inequality, specifies a region $\mathcal{G}_m$ in the space governed by $(\mathbf{x_0}, \ldots, \mathbf{x_{m-1}}, \mathbf{x_m}, \ldots, \mathbf{x_{M-1}})$ over which $P(A_{m'})$ has be summed.

$$P(A_{m'}) = \sum_{\mathcal{G}} Q(\mathbf{x_0}, \ldots, \mathbf{x_{m-1}}, \mathbf{x_m}, \ldots, \mathbf{x_{M-1}}) \tag{4.25}$$

Using the relation in (4.24)

$$\begin{aligned} P(A_{m'}) \;\; \leq & \sum_{\mathbf{x_0},\cdots,\mathbf{x_{m-1}},\mathbf{x_m},\cdots,\mathbf{x_{M-1}}} Q(\mathbf{x_0}, \ldots, \mathbf{x_{m-1}}, \mathbf{x_m}, \ldots, \mathbf{x_{M-1}}) \;\; . \\ & [\frac{P(\mathbf{y} \mid \mathbf{x_{m'}})}{P(\mathbf{y} \mid \mathbf{x_m})} + \sum_{l=0,l\neq m,m'}^{M-1} (\frac{D_{ml} - D_{m'l}}{D_{mm'}}) \frac{P(\mathbf{y} \mid \mathbf{x}_l)}{P(\mathbf{y} \mid \mathbf{x_m})}] \end{aligned} \tag{4.26}$$

Since $\mathbf{x}_l$ is a dummy variable of summation,

$$P(A_{m'}) \leq \frac{1}{P(\mathbf{y} \mid \mathbf{x_m})}[1 + \sum_{l=0,l\neq m,m'}^{M-1} \frac{D_{ml} - D_{m'l}}{D_{mm'}}] \sum_{\mathbf{x}} Q(\mathbf{x}) P(\mathbf{y} \mid \mathbf{x}) \tag{4.27}$$

$$\begin{aligned} Pr[\text{error} \mid m, \mathbf{x_m}, \mathbf{y}] \;\; \leq & \;\; \frac{1}{P(\mathbf{y} \mid \mathbf{x_m})^{\rho}}[\sum_{m'\neq m} D_{mm'} + \sum_{l=0,l\neq m,m'}^{M-1} (D_{ml} - D_{m'l})]^{\rho} \\ & [\sum_{\mathbf{x}} Q(\mathbf{x}) P(\mathbf{y} \mid \mathbf{x})]^{\rho} \end{aligned} \tag{4.28}$$

where $O(1)$ terms which does not depend on N,A, or L are neglected in the expressions.Hence

$$\begin{aligned} \overline{D}_{e,m} \;\; \leq & \;\; [\sum_{m'\neq m} D_{mm'} + \sum_{l=0,l\neq m,m'}^{M-1} (D_{ml} - D_{m'l})]^{\rho} \\ & \sum_{\mathbf{y}} [\sum_{\mathbf{x}} Q(\mathbf{x}) P(\mathbf{y} \mid \mathbf{x})^{1-\rho}] [\sum_{\mathbf{x}} Q(\mathbf{x}) P(\mathbf{y} \mid \mathbf{x})]^{\rho} \end{aligned} \tag{4.29}$$

After careful observation, one can convince oneself that

$$\sum_{l=0,l\neq j}^{M-1} D_{jl} = L! \sum_{i\neq k,\forall i,k} d_{i,k} \leq d_{max} L! \binom{A}{2} \tag{4.30}$$

i.e the right hand side of the equation is true $\forall l = 0, \ldots, M-1$. For DMC's the second line of equation (4.29) can be expressed as $2^{-NE_2(\rho,Q)}$. And hence the result.

$E_2(\rho, Q)$ can be thought of as some kind of distance between the joint density $p(j,k)$ and the product of marginals $p(j)Q(k)$. By differentiating the expression one can show that it is non-negative, non-decreasing (for $J > 2$), concave function with respect to ρ, with slope at $\rho = 0$ as the mutual information. From (4.2), the best rate

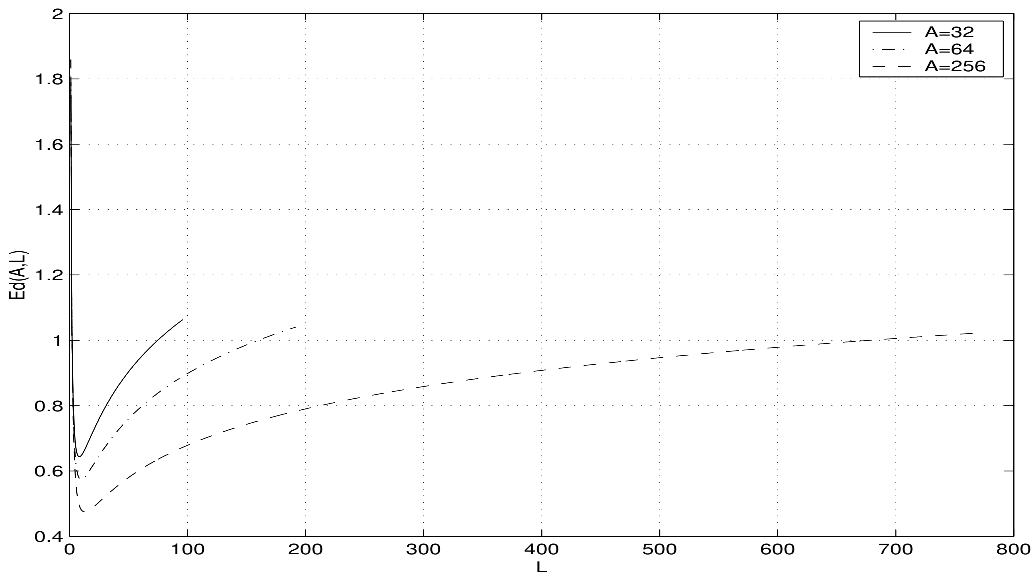

Figure 4.2 $E_d(A, L)$ vs L (equation 4.21)

of decay for ML(or MAP) decoding is

$$\max_{\rho,Q}[E_0(\rho, Q) - \rho R]$$

whereas from (4.19), the best rate of decay for minimum distortion decoding is

$$\max_{\rho,Q}[E_2(\rho, Q) - \rho E_d(A, L)R]$$

where $R = L \log A/N$ bits. Given A, the number of alphabets, to keep R constant L/N have to be maintained constant. As we increase L(and hence N proportionately), $E_d(A, L) < 1$ for L approximately less than eA and $E_d(A, L) > 1$ for larger L. Fig

4.2 plots $E_d(A, L)$ for different values of A. This can also be shown using Stirling approximation

$$L! \approx \sqrt{2\pi L}(\frac{L}{e})^L$$

We know that minimum distortion decoding is the optimal decoding when end-to-end distortion is the criteria. $E_d(A, L)$, roughly quantifies the gain with minimum distortion decoding over ML decoding for equally likely alphabet iid source over DMC channels. The bound is useless for when $E_d(A, L) \gg 1$. Fig 4.3 explains this for a BSC at rate $= 0.3$ bits with $A = 128$. 'Good MDD' corresponds to a case of $L = 24, N = 560$. 'Bad MDD' corresponds to $L = 600, N = 14000$. Hence for most practical cases, the gains in minimum distortion decoding is quantified by $E_d(A, L)$.

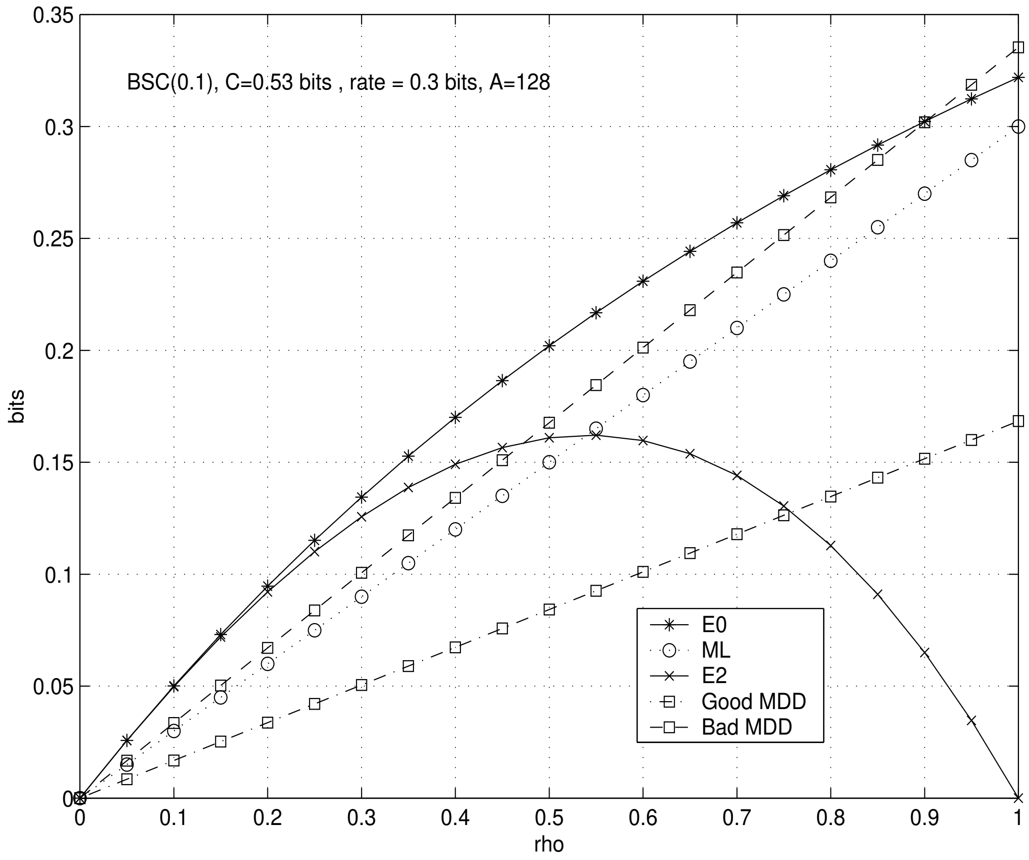

Figure 4.3 Random Coding Exponent-Minimum Distortion Decoding

Chapter 5

Conclusions and Future Directions

We showed that minimum distortion decoding (Chapter 2) is the optimal decoding rule when end-to-end distortion is our criteria (Fig 2.1). We quantified the gains of minimum distortion decoding, by deriving random coding exponent for the rate of decay of end-to-end distortion for lossless encoding. The gains are dependent on the parameter $E_d(A, L)$, which is less than one for most practical encoders. This we compared with the random coding bound of ML (and general MAP) decoding. So the result says, there exists one code in the ensemble of random codes, wherein with minimum distortion decoding can achieve the same end-to-end distortion with less channel use than the ML decoding, when $E_d(A, L) < 1$.

Future Directions include

- A stronger random coding bound on distortion similar to (4.19), where $E_d(A, L) \to 1$ as $L \to \infty$.
- Random coding bounds with lossy encoding and also for general class of sources which have correlation.
- Design of practical encoders which achieve the performance predicted by random coding bounds.
- Extension of the above results to continuous sources and waveform channels

Bibliography

[1] S. Emami and S.L. Miller. Nonsymmetric sources and optimum signal selection. *IEEE Trans. Commun.*, 44(4):440–447, April 1996.

[2] R. Gallager. *Information Theory and Reliable Communication*. New York: Wiley, 1968.

[3] B. Hochwald and K. Zeger. Tradeoff between source and channel coding. *IEEE Trans. Info. Th.*, 43(5):1412–1424, Sept. 1997.

[4] A. Nosratinia J. Lu and B. Aazhang. Progressive source-chanel coding of images over bursty error channels. In *Internation Conference on Image Proc. (ICIP)*, volume 2, pages 127–131, 1998.

[5] J.D.Gibson K.Sayood, F.Liu. A constrained joint source/channel coder design. *IEEE Journal on Selected Areas in Communications*, 12(9):1584–1593, Dec. 1994.

[6] J.L Massey. Joint source and channel coding. In J.K. Skwirzynski, editor, *Communication Systems and Random Process Theory*, pages 279–293. The Netherlands: Sijthoff and Nordhoff, 1978.

[7] C.E. Shannon. The mathematical theory of communication. *BSTJ*, 28:379–423, Oct. 1949.

[8] V.A. Vaishampayan and N. Farvardin. Optimal block cosine transform image coding for noisy channels. *IEEE Trans. Commun.*, 38:327–336, Mar. 1990.

[9] S. Vembu and S. Verdu. The source-channel separation theorem revisited. *IEEE Trans. Inf. Th.*, 41(1):44–54, Jan. 1995.