TREE CODING OF SOURCES AND CHANNELS

by
TAKESHI HASHIMOTO

Department of Mechanical Engineering
Faculty of Engineering Science
Osaka University
January 1981

TREE CODING OF SOURCES AND CHANNELS

A DISSERTATION

SUBMITTED TO OSAKA UNIVERSITY

FOR THE DEGREE OF

DOCTOR OF ENGINEERING

by

Takeshi Hashimoto

January 1981

Abstract

Basic properties of block coding and tree coding are treated for both channels and sources. Hierarchical channel codes c_M^N of rates $R_M = (1/N)\log M$ for DMC's are constructed so that codes $c_1^N \subset c_2^N \subset \ldots$ satisfy the expurgated exponent and other exponent functions. Universality of channel and source codes is achieved that satisfy the ultimate limits on performance over all DMC's in channel coding and over all stationary ergodic sources in source coding. A simple concept of universality is introduced and is shown to give a simple proof to Gray-Davisson's source coding theorem for stationary nonergodic sources. The tree coding research concerns convolutional tree coding on DMC's with the Viterbi and sequential tree searching algorithms, and tree coding of stationary ergodic sources with several tree searching algorithms. In channel tree coding the computational problem associated with sequential decoding is fully investigated, and analytical confirmation is given to known experimental data for convolutional codes. In source coding the main topic is a proof of the tree coding theorem for

stationary ergodic sources using a new tree searching algorithm. Finally, as a practical application, tree coding of speech is investigated. It is shown that, in spite of the apparent nonstationarity of speech, tree codes yield satisfactory speech compression.

Acknowledgement

I am very grateful to the supervisor Prof. S. Arimoto for his persistent encouragement, and to prof. T. Kakutani, Prof. T. Kasami, Prof. T. Okamoto, and Prof. M. Sakaguchi for their useful advice to improve the thesis. Also, I would like to thank Prof. Han Te-Sun, Dr. Kobayashi, and H. Morita for their invaluable contribution to my studying information theory, and thank Prof. T. Berger for his kind suggestions in improving the style of Abstract and other parts of the thesis. Of course, I have been much owed so many people whom I can not mention here. My thanks are to be given them, too.

Finally, I would like to thank my family for their unconditional support, without which I could not have accomplished anything.

Contents

I. INTRODUCTION

In this thesis we explore problems of channel and source coding, especially, with tree or convolutional tree codes. Two subjects in Information Theory concern these problems: they are generally referred to as Shannon Theory and (Algebraic) Coding Theory. In a loose sense, Shannon Theory concerns the ultimate limits of communication, while (Algebraic) Coding Theory concerns the way in which efficient communication is realized. However we treat problems largely from the Shannon Theory side, and study how effectively the tree and convolutional tree codes, codes having tree-like and trellis-like geometric structure, can be used to code channels and sources.

Shannon Theory has assumed originally the use of arbitrary block codes which may not have any structure. It is asserted that, if appropriate block encoders and decoders are devised, we can ultimately achieve theoretical limits on rates, for error-free communication over noisy channels (channel coding), and on coding distortions, for efficiency-oriented transmission of data (source coding). However, as the theory

develops and communication engineers notice the information-theoretic approach to communication system design, it becomes apparent that ordinary block encoding and decoding are insufficient for practical use since they require impossibly much computation. Tree codes are thus frequently used with a hope that they might afford implementable encoding and decoding.

In this regard we study properties and ultimate performance of tree codes in channel and source coding. In Chapter II, we first see a standard approach to channel coding for a DMC, which is basic in other chapters. Chapter III concerns basic, rather mathematical, features that we can make codes possess, hierarchical structures between codes and universality of codes for channels. In Chapter IV we investigate two decoding schemes for convolutional codes, namely, sequential decoding and Viterbi decoding. Although Viterbi decoding is more suited to hardware implementation in the modern communication systems, they are complementary. The emphasis is placed on the computational aspects on sequential decoding with convolutional codes. In this chapter we see why communication engineers prefer tree codes in channel coding and how they work in a certain application.

In the later half of the thesis we consider source coding. Since our aim is the efficient coding of practically important sources which have, in general, correlations between output letters (sources with memory) , we first summarize, in Chapter V, necessary notations and well-known results. Chapter VI is devoted to the study of the existence of codes with universally good performance over sources, universal codes, which are particularly important in source coding. We introduce possibly the least restrictive class of universal codes, and discuss the practical meaning of it. The problem of tree encoding is treated in Chapter VII. The main topic in this chapter is a tree coding theorem for stationary ergodic sources, which is the first satisfactory tree coding theorem. The theorem asserts that tree codes can ultimately attain the theoretical limits on the coding distortions. Finally, in Chapter VIII we consider tree coding of a particular source, speech. There, we can see the difficulty in treating real sources which do not have uniform and purely stochastic characteristics. However our experimental data reveal that speech may be encoded more efficiently if suitable codes and source adaptation mechanisms are selected.

Chapter II

Channel Coding Preliminaries

1. Communication system and coding

A fundamental problem in communication is how efficiently one can send signals from an object to other distant point. The object may be speech or written articles and the emitted signal may be a speech signal converted to electric current by a microphone or an electric pulse train corresponding to letters in the written message such as in telegram. We call such an object an information source, or simply, a source. At the destination, the exact reproduction of the signal is indispensable in some cases. In other cases only an approximate reproduction with specified fidelity is sufficient. We call the medium that carries necessary data from the source to the destination a channel; e.g., atmosphere in radio communication, or copper wire and repeaters in cable communication.

Some limitations in communication often occur because of the effect of noise in the channel, such as the thermal noise in an electronic circuit or the disturbance in long distance radio communication. When high speed communication or high quality communication is required, these considerations become paramount. Direct connection of the source to the channel is not generally a good answer even when possible; communication engineers have invented various devices, commonly refered as encoders and decoders (see Fig. 2.1.1), to facilitate efficient communication

How well should they work, and how should they be constructed is the coding problem for the source and channel.

Information theory, originated by Shannon [1], provides a mathematical basis for studying the existence of a good encoder and good decoder. One of his main theorems is stated, in rather vague terminology, as follows:

> Desired communication is possible if the "rate" of the source relative to a given fidelity criterion is less than the "capacity" of the channel, and it is impossible if the "capacity" is less than the "rate".

Here we mentioned two notions; the "capacity" of the channel and the "rate" of the source, both of which will be defined rigrously later.

Let $A,B,\hat{A}$, and $\hat{B}$ be finite sets, and suppose that, at time i, the source emits x_i selected from A, the channel emits $\hat{y}_i$ selected from $\hat{B}$ when it receives x_i selected from $\hat{A}$, and the decoder, observing $\hat{y}_i$, emits y_i selected from B as a reproduction of x_i. Note that the encoder, the channel, and the decoder may use the data that have been received in the past or that will be received in the future to choose the symbol emitted that time. $A,B,\hat{A}$, and $\hat{B}$ are called alphabets and their elements are called letters. (Of course, the situation is too simplified; the source and channel are not necessarily synchronized with each other.)

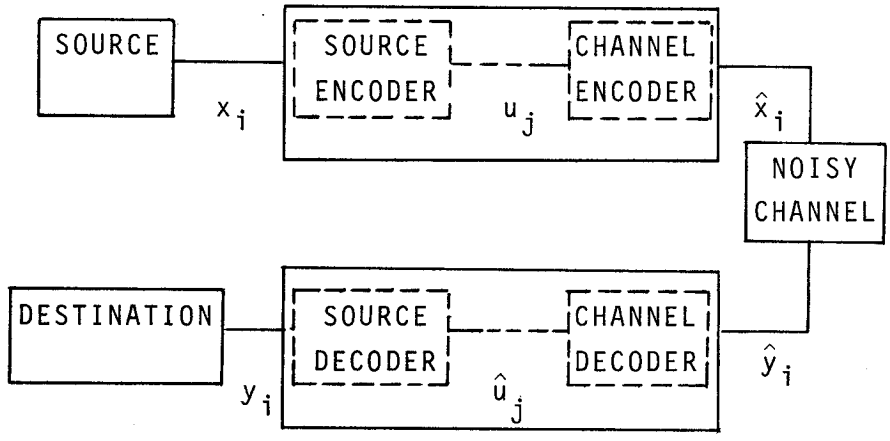


Fig. 2.1.1 — A Communication System

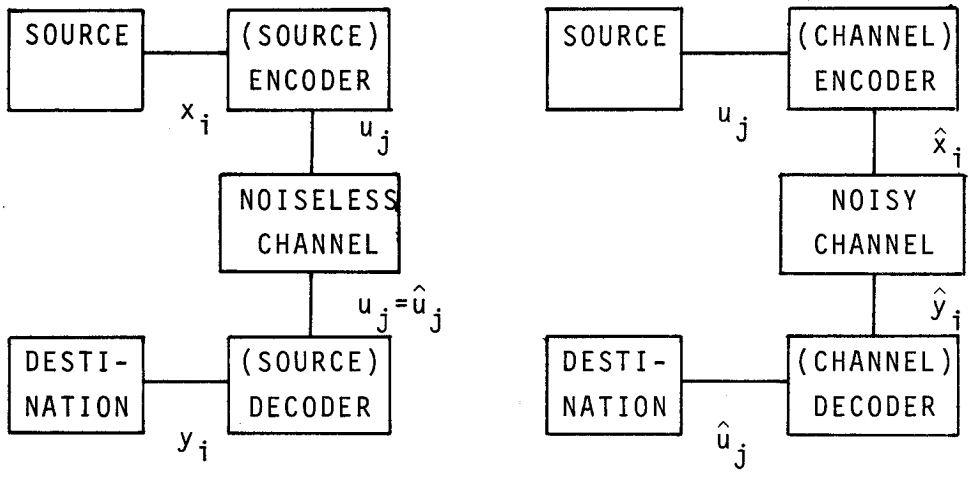


Fig. 2.1.2 — Coding Problems

The respective actions of the encoder and decoder are not merely inversions of each other, for the decoder should detect faithfully which $\underline{\hat{x}}^N = \hat{x}_1 \ldots \hat{x}_N$ is sent over the channel observing the channel output $\underline{\hat{y}}^N = \hat{y}_1 \ldots \hat{y}_N$ while the encoder should decide which $\underline{\hat{x}}^N$ yields a good reproduction $\underline{y}^N = y_1 \ldots y_N$ provided by decoder's faithful detection of $\underline{\hat{x}}^N$ from $\underline{\hat{y}}^N$. This idea will be well understood by introducing intermediate binary (or q-nary) sequences $\underline{u}^n = u_1 \ldots u_n$ and $\underline{\hat{u}}^n = \hat{u}_1 \ldots \hat{u}_n$ with u_i, $\hat{u}_i$ = 0 or 1 (or, 0, 1, ... ,or q-1) , as depicted by broken line in Fig.2.1.1. We call $\underline{u}$ and $\underline{\hat{u}}$, respectively, a message sequence and a decoded message sequence. Generally, the periods between digits in these sequences may not agree with those between letters in source output or channel output ($n \neq N$).

With these intermediate sequences, the coding problem for the source and channel splits into that for the channel and that for the source. Information theory assures that entirely separate attacks on respective problems are permissible under broad conditions[2]. The former is called the channel coding problem, which aims at reliable communication over the channel, and the latter is called the source coding problem, which aims at good reproduction at the destination when the message is sent over a noiseless channel, see Fig.2.1.2.

Especially, the source coding problem is said to be noiseless if the reproduction $\underline{y}^N$ is exact, $\underline{x}^N = \underline{y}^N$. and is said to have a fidelity criterion if $\underline{y}^N$ satisfies the condition $d(\underline{x}^N, \underline{y}^N) \leq ND$ for a given function d on $A^N \times B^N$ and a fidelity $D \geq 0$.

We again return to Fig. 2.1.1. We have been concerned only with the relationships between letters. However, since the source may continuously emit letters $x_1 x_2 \ldots$ indefinitely, the decoder and encoder should be active as long as the source is. A simple way accommodating the system to such situation is to partition the stream of source output into blocks of a given number of consecutive letters and to encode each block independently. This block-wise coding scheme is called block coding.

More precisely, the block source encoder encodes each block of N consecutive source letters $\underline{x}^N = x_1 \ldots x_N$ into a block of binary (q-nary) digits, $\underline{u}^n = u_1 \ldots u_n$; the block channel encoder encodes $\underline{u}^N$ into a block of channel input letters, $\hat{\underline{x}}^N = \hat{\underline{x}}_1 \ldots \hat{x}_N$. Then the channel decoder, observing the block of channel output letters, $\hat{\underline{y}}^N = \hat{y}_1 \ldots \hat{y}_N$, detects the transmitted $\hat{\underline{x}}^N$, and emits a block of binary (q-nary) digits, $\hat{\underline{u}}^n = \hat{u}_1 \ldots \hat{u}_n$, expressing $\underline{x}$; the source decoder simply converts $\hat{\underline{u}}^n$ into a block of reproduction letters, $\underline{y}^N = y_1 \ldots y_N$. See Fig.2.1.3.

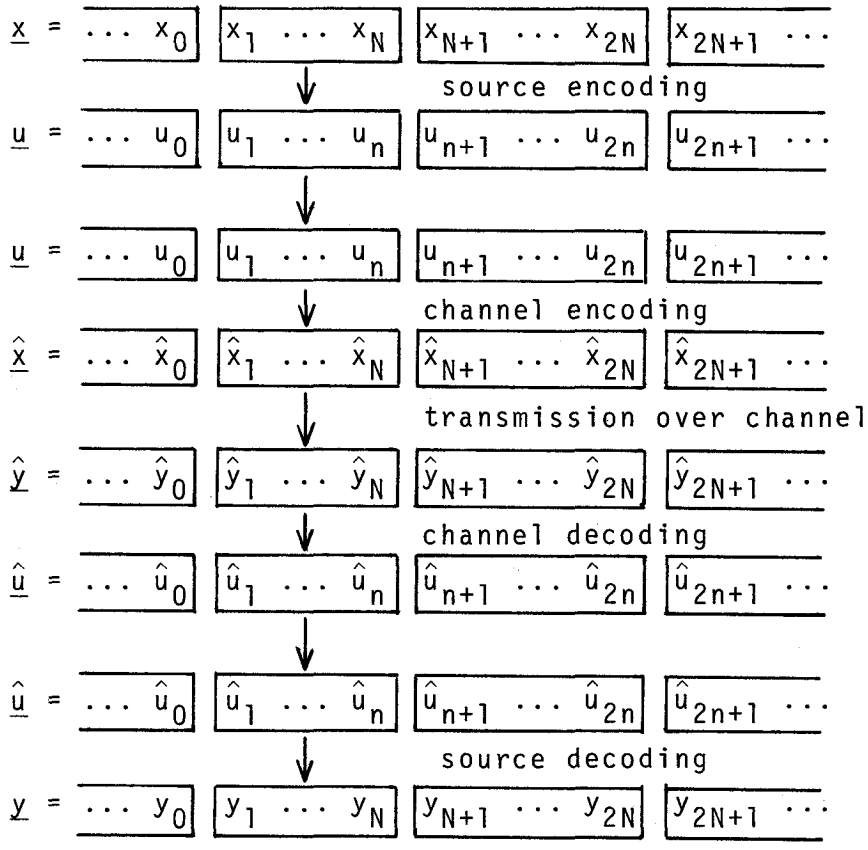


Fig. 2.1.3 — Block Coding

We call each $\underline{\hat{x}}^N$ and each $\underline{y}^N$ in this scheme a block channel codeword and block source codeword respectively, and we call the respective totalities of these codewords a block channel code and block source code. We note that, if codewords in respective codes are numbered from 1 to M, the intermediate sequences can be replaced with numbers m = 1, ... ,M. Then, the rates of codes are $(1/N)\log_e M$ nats/letter [or $(1/N)\log_2 M$ bits/letter].

We have so far discussed from a rather mathematical point, where any encoding and decoding computations are assumed possible. However a little reflection reveals that a block encoder should repeat an exponentially increasing number of operations such as distortion calculation or distortion comparison in an only linearly increasing time span as the block length gets longer. With this regard we need codes to have some structures between codewords that allow easier decoding.

Tree codes are important such codes, which have a tree-like structure between codewords. This geometric structure facilitates efficient encoding and decoding methods; some of them are systemtic versions of list-searching scheme, and others are different.

Lastly, we give several notations, which are assumed in this and the next two chapters where we are concerned with the simplest channels, discrete memoryless channels. In the later chapters, where we are concerned with more complex sources and channels, more subtle notations are given.

Given a set A, A^n denotes the set of all n-tuples $a_1 \ldots a_n$, $a_i \in A$, and, for each n, elements in A^n are identified by small letters with an under bar and a super script n, e.g., $\underline{x}^n$ or $\underline{a}^n$, keeping capital letters for random variables and sets, e.g., $\underline{X}^n$ and S. For any subset S in A^n, S^c denotes its complement. Let G be a statement which is true or false. Then we denote the indicater function of G by $\chi[G]$; $\chi[G] = 1$ if G is true and $\chi[G] = 0$ if otherwise. Although distinct symbols are used for source and channel alphabets in this section, we use the same symbols A and B for channel and source alphabets in the subsequent chapters. Finally, all logarithmic functions log(*) are assumed to have the natural base e – for the function with the base two, we write as $\log_2(*)$ explicitely.

2. Block coding of DMC

A discrete memoryless channel (DMC) is a channel with stochastic noise such that, for received channel input sequence $\underline{x}^N = x_1 \ldots x_N$, the channel emits channel output sequence $\underline{y}^N = y_1 \ldots y_N$ with the product probability

$$P(\underline{y}^N | \underline{x}^N) = \prod_{n=1}^{N} P(y_n | x_n) ,$$

where $P = \{P(b|a),\ a \in A,\ b \in B\}$ is a conditional pmf defined on $A \times B$. The DMC is identified, symbolically, by P. Obviously, the outcome from the DMC is a sequence of iid random variables for each input sequence. For a given pmf on A, we also write

$$p(\underline{x}^N) = \prod_{n=1}^{N} p(x_n),$$

for all $\underline{x}^N \in A^N$.

When a block code $c^N = \{ \underline{x}_m^N,\ m = 1, \ldots, M \}$ is given, the most popular decoder is a maximum likelihood decoder (MLD) which operates as : decode each channel output $\underline{y}^N$ into a message m if $\underline{y}^N$ is in the set, call it the decoding region for m,

$$Y_{N,m}(P) \triangleq \{ \underline{y}^N \in B^N : P(\underline{y}^N | \underline{x}_m^N) > P(\underline{y}^N | \underline{x}_{m'}^N),\ \text{all } m' \neq m \}$$

(of course, $\underline{y}^N$ may not be in any $Y_{N,m}(P)$, but we can

neglect such an event without any drawback.)

Let $P_{e,m}(c^N)$ be the probability that, when the massage m is sent, a decoder, not necessarily a MLD, fails to decode the channel output into m correctly, and let the average probability of error be

$$P_e(c^N) \triangleq \frac{1}{M} \sum_{m=1}^{M} P_{e,m}(c^N).$$

Then, as one can see in [2], the MLD minimizes $P_e(c^N)$ and gives

$$\begin{aligned} P_{e,m}(c^N) &= \sum_{\underline{y}^N \in Y^c_{N,m}(P)} P(\underline{y}^N | \underline{x}^N_m) \\ &= \sum_{\underline{y}^N \in B^N} P(\underline{y}^N | \underline{x}^N_m) \; \chi[\, P(\underline{y}^N | \underline{x}^N_m) \leq P(\underline{y}^N | \underline{x}^N_{m'}), \\ &\qquad \text{some } m' \neq m \,] \quad . \end{aligned} \tag{2.2.1}$$

The extreme right-hand side is further bounded by the following form with free parameter $0 \leq \rho \leq 1$

$$\sum_{\underline{y}^N \in B^N} P(\underline{y}^N | \underline{x}^N_m) \left\{ \sum_{m'(\neq m)} \left[\frac{P(\underline{y}^N | \underline{x}^N_{m'})}{P(\underline{y}^N | \underline{x}^N_m)} \right]^{\frac{1}{1+\rho}} \right\}^{\rho} .$$

However, it is quite difficult to obtain an analytically tractable approximation of it for a particular code. Nevertheless, if we let $\mathcal{C}^N$ be a random code

consisting of random codewords $\underline{X}_m^N$ with probabilities $\Pr\{ \underline{X}_m^N = \underline{x}^N \} = p(\underline{x}^N)$, all $\underline{x}^N \in A^N$, for a pmf p on A, and if we apply the bound to $\mathcal{C}^N$, the expectation of the bound with respect to $\mathcal{C}^N$ is further bounded by an elegant form. That is, letting $\mathcal{E}$ be the expectation operator with respect to $\mathcal{C}^N$, the following result is known [2].

Lemma 2.2.1: The MLD minimizes $P_e(c^N)$ for any c^N and gives

$$\mathcal{E} P_e(\mathcal{C}^N) \triangleq \mathcal{E} P_{e,m}(\mathcal{C}^N)$$

$$\leq \exp\{ -N[E_o(\rho,p,P) - \rho R]\} \quad ; \ 0 \leq \rho \leq 1,$$

for all $m = 1, \ldots, M$, where $R = (1/N)\log M$ and

$$E_o(\rho,p,P) \triangleq -\log \sum_{b\in B} \Big[\sum_{a\in A} p(a)\, P^{\frac{1}{1+\rho}}(b|a) \Big]^{1+\rho} .$$

Since $\mathcal{E} P_e(\mathcal{C}^N) = \Sigma_{\text{all codes}} \Pr\{ \mathcal{C}^N = c^N \} P_e(c^N)$, there exists at least one code c^N satisfying $P_e(c^N) \leq \mathcal{E} P_e(\mathcal{C}^N)$, and the bound in the lemma is really a bound on $P_e(c^N)$.

Theorem 2.2.1: There is a block code c^N of rate $R = (1/N)\log M$ such that the MLD yields

$$P_e(c^N) \leq \exp\{ -NE_r(P,R) \}$$

where

$$E_r(P,R) \triangleq \max_{p} \max_{0 \leq \rho \leq 1} [\, E_o(\rho,p,P) - \rho R \,]$$

The function $E_r(P,R)$ is called the random coding exponent function. (The term "random coding" comes from the argument above the theorem, which is called a random coding argument.) A typical curve of the function is illustrated in Fig.2.2.1: $E_r(P,R)$ has the slope -1 for rates less than a critical rate $R_0(P)$ and has positive values for rates less than C(P), where

$$C(P) \triangleq \max_{p} I(p,P) \qquad \text{and}$$

$$I(p,P) \triangleq \sum_{a \in A} \sum_{b \in B} p(a)P(b|a) \log \frac{P(b|a)}{\sum_{a' \in A} p(a')P(b|a')} .$$

I(p,P) is called the mutual information quantity for p and P, and C(P) is called the channel capacity of DMC P. The term "capacity" referring to C(P) is justified by the next theorem.

Theorem 2.2.2: For a DMC P and any $R > 0$, if $R < C(P)$, then, for any $\varepsilon > 0$, there exist a block encoder and block decoder with a rate larger than $R - \varepsilon$ and a probability of decoding error less than ε , and, conversely,

if $R > C(P)$, then there are no such encoders and decoders.

Since $(1/N)\log M$ is made close to any positive value for large N and large M, the first part of Theorem 2.2.2 is a consequence of Theorem 2.2.1, and the proof of the latter part is seen in [2].

Letting $R < C(P)$, the next interesting problem is how fast the error probability decreases as the block length gets longer. (From Theorem 2.2.1, we know that it is no longer slower than $\exp\{-NE_r(P,R)\}$.) We call the maximally attainable exponential rate,

$$E(P,R) = \varliminf_{N\to\infty} -\frac{1}{N} \log \inf_{c^N} P_e(c^N),$$

the reliability rate function of P. It is known [2],[3] that $E(P,R) = E_r(P,R)$ for $R_0(P) \leq R \leq C(P)$ where $R_0(P)$ is the critical rate defined before.

To obtain a stronger bound for $R < R_0(P)$, we let

$$Z(a,a') \triangleq \sum_{b\in B} \sqrt{P(b|a)P(b|a')}$$

for each $a, a' \in A$, and let

$$Z(\underline{x}^N,\underline{x}'^N) \triangleq \sum_{\underline{y}^N\in B^N} \sqrt{P(\underline{y}^N|\underline{x}^N)P(\underline{y}^N|\underline{x}'^N)}$$

for each $\underline{x}^N, \underline{x}'^N \in A^N$; they are called, respectively,

the Bhattacharyya distance between a and a' and the Bhattacharyya distance between $\underline{x}^N$ and $\underline{x}'^N$. Applying (2.2.1), with $\rho = 1$, to a code c'^N containing 2M codewords, we have

$$P_{e,m}(c'^N) \leq \sum_{m'(\neq m)} Z(\underline{x}_m^N, \underline{x}_{m'}^N) \quad ; \quad m = 1, \ldots, 2M \qquad (2.3.2)$$

Using the standard inequality $(\Sigma a_i)^s \leq \Sigma a_i^s$, $0 \leq s \leq 1$, we have

$$P_{e,m}^{1/\rho}(c'^N) \leq \sum_{m'(\neq m)} Z^{1/\rho}(\underline{x}_m^N, \underline{x}_{m'}^N)$$

for $1 \leq \rho$ and $m = 1, \ldots, 2M$. Let $\mathcal{C}'^N$ be a random block code with 2M codewords constructed in the same way as before. The following lemma is obtained from the arguments in [2].

Lemma 2.2.2: $\mathcal{E} P_{e,m}^{1/\rho}(\mathcal{C}'^N)$ is independent of m and

$$\{ \mathcal{E} P_{e,m}^{1/\rho}(\mathcal{C}'^N) \}^\rho$$

$$\leq \exp\{ -N[E_x(\rho,p,P) - \rho(1/N)\log 2M] \}$$

for $1 \leq \rho$ and $m = 1, \ldots, 2M$, where

$$E_x(\rho,p,P)$$

$$\triangleq -(\rho/N)\log \mathcal{E} Z^{1/\rho}(\underline{X}_1^N,\underline{X}_2^N)$$

$$= -\rho \log \sum_{a,a' \in A} p(a)p(a')Z^{1/\rho}(a,a')$$

for $1 \leq \rho$.

For $1 \leq \rho$, let ϕ_m be the indicater function of

$$P_{e,m} < 2^{\rho}[\ \mathcal{E} P_{e,m}^{1/\rho}\ ]^{\rho}$$

where we let

$$\mathcal{E} P_{e,m}^{1/\rho} \triangleq \mathcal{E} P_{e,1}^{1/\rho}(\mathcal{C}'^N) \quad .$$

Since, from Markov's inequality, we have

$$\mathcal{E} \sum_{m=1}^{2M} \phi_m \geq 1/2 \quad ,$$

there exists at least a code c'^N such that $\phi_m = 1$ for at least M m's. We renumber the codewords that are specified by m for which $\phi_m = 1$ in c'^N, and let c^N be the block code consisting of these M codewords. Since expurgation of codewords does not increase the average probability of error, we know

$$P_{e,m}(c^N) \leq \exp\{ -N[E_x(\rho,p,P) - \rho(1/N)\log 4M] \}$$

for m = 1, ... ,M and have a well-known theorem:

Theorem 2.2.3: For each N, there exists a block code c^N of rate R = (1/N)log M such that the MLD yields

$$P_e(c^N) \leq \exp\{ -N E_{ex}(P, R + (1/N)\log 4) \}$$

where

$$E_{ex}(P,R) \triangleq \max_p \sup_{\rho \geq 1} [E_x(\rho,p,P) - \rho R]$$

The function $E_{ex}(P,R)$ is called the expurgated exponent function, and is known to be [2],[3]:

$$E_{ex}(P,R) = E(P,R) \quad ; R = 0,$$

$$E_{ex}(P,R) > E_r(P,R) \quad ; 0 \leq R < R_1(P),$$

$$E_{ex}(P,R) = E_r(P,R) \quad ; R_1(P) \leq R \leq R_0(P),$$

where $R_1(P)$ is another critical rate. All of the curves are summarized and illustrated in Fig. 2.2.1.

In the figure, $E_{ex}(P,R)$ is depicted as if it is finite for all $R \geq 0$. However, certain channels have infinite values. Consider the channel P depicted in Fig.2.2.2. It is certain that the channel yields no error at $R = \log 2$, that is, $E(P, \log 2) = \infty$. We say that a DMC P has a zero-error-capacity R_∞ if $E(P,R) = \infty$ for $R < R_\infty$, and, conversely, say that P has zero zero-error-capacity if $R_\infty = 0$. It is known [2] that the quantity

$$R_{x,\infty} = \sup_{\rho \geq 1} \max_{p} \frac{1}{\rho} E_x(\rho,p,P)$$

is a lower bound of the zero-error-capacity. Therefore, if P has zero zero-error-capacity, then $\sup_{\rho \geq 1} [E_x(\rho,p,P) - \rho R]$ is attained with finite ρ for $R > 0$.

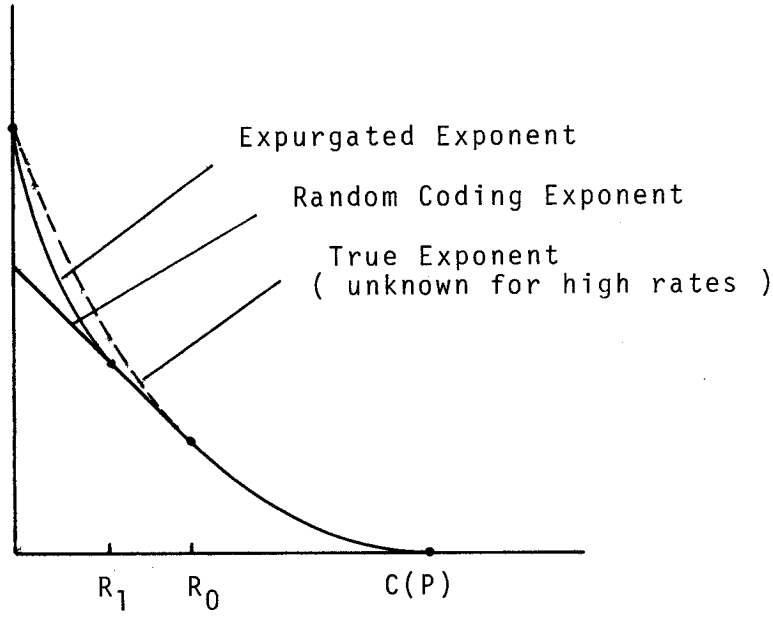


Fig. 2.2.1 — Several Exponents in Channel Coding

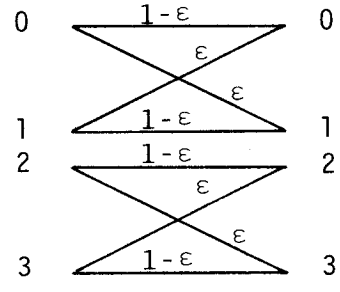


Fig. 2.2.2 — A Channel with Positive Zero-Error-Capacity

Chapter III

Properties of Block Codes for DMC

1. Hierarchical construction of good codes

In the previous chapter, good codes are shown for each rate, or for each code size M, and, in such an approach, a good code c_1^N with M codewords and another good code c_2^N with M codewords, both for the same channel, might be made of entirely different class of codewords each other. However, we can show two possible forms of hierarchical relationships between $c_1^N \subset c_2^N \subset \ldots$ where the inclusion means that every codeword of c_i^N is in c_{i+1}^N. Proofs of such hierarchical structures are virtually related to sequential construction of good codes; by an addition of a suitable codeword to a good code c_i^N with i members, enlarge it up to another good code c_{i+1}^N with i + 1 members. Of course, the hierarchy relies upon in which sense we say a code is good. If we mean it under the light of the expurgated exponent function, the structure takes quite an elegant form. So we start with this exponent.

Hierarchy For Expurgated Exponent Function

Suppose that the DMC P has alphabets A = {0,1, ... ,α-1} and B = {0,1, ... ,β-1}, and that it has zero zero-error-capacity(see Sec. 2.2). The basic bound in proving the expurgated bound (Theorem 2.2.3) is (2.2.2). We put as

$$F(c_M^N) \triangleq \frac{1}{M} \sum_{m=1}^{M} \sum_{m'(\neq m)} Z(\underline{x}_m^N, \underline{x}_{m'}^N).$$

Then, $F(c_M^N)$ is a bound on $P_e(c_M^N)$, where the subscript M in c_M^N means the M members in the code. For the later convenience, we let

$$E_{ex}(p,P,R) \triangleq \sup_{\rho \geq 1} [\, E_x(\rho,p,P) - \rho R \,]$$

for $R \geq 0$, and call it the expurgated exponent function (although it is a little abuse in terminology). The pmf p is fixed arbitrarily through the section. We note that, if $\mathcal{E}_X$ denotes the expectation operator with respect to a random codeword $\underline{X}^N = X_1 \ldots X_N$ assuming values $\underline{x}^N \in A^N$ with the product probability $p(\underline{x}^N)$, then $E_x(\rho,p,P)$ defined in Lemma 2.2.2 is recasted as

$$E_x(\rho,p,P) = -(\rho/N) \log \mathcal{E}_X \mathcal{E}_{X'} Z^{1/\rho}(\underline{X}^N, \underline{X}'^N)$$

for all $\rho \geq 1$.

The recursive selection for the exponent $E_{ex}(p,P,R)$ starts from showing a good two word code. Let γ be an integer, and let $0 < s(i) \leq 1$, $i = 1, \ldots, \gamma$, be numbers all of which are determined later. Then Markov's inequality implies the inequality

$$\Pr\{\ \mathcal{E}_{X,Z^{s(i)}}(\underline{X}^N,\underline{X}'^N) > 2\gamma \mathcal{E}_X\, \mathcal{E}_{X,Z^{s(i)}}(\underline{X}^N,\underline{X}'^N),$$

$$\text{some } i = 1, \ldots, \gamma \ \} < 1/2 \ , \qquad (3.1.1)$$

and the inequality shows the existence of $\underline{x}_1^N \in A^N$ such that

$$\mathcal{E}_{X,Z^{s(i)}}(\underline{x}_1^N,\underline{X}'^N)$$

$$\leq 2\gamma\, \mathcal{E}_X\, \mathcal{E}_{X,Z^{s(i)}}(\underline{X}^N,\underline{X}'^N) \ ; \ i = 1, \ldots, \gamma. \qquad (3.1.2)$$

On the other hand, using Markov's inequality again, we have

$$\Pr\{\ Z(\underline{x}_1^N,\underline{X}'^N) > \min_i [\ 2\, \mathcal{E}_{X,Z^{s(i)}}(\underline{x}_1^N,\underline{X}'^N)\ ]^{1/s(i)}\ \}$$

$$= \Pr\{\ Z^{s^*}(\underline{x}_1^N,\underline{X}'^N) > 2\, \mathcal{E}_{X,Z^{s^*}}(\underline{x}_1^N,\underline{X}'^N)\ \}$$

$$< 1/2, \qquad (3.1.3)$$

where s^* is the minimizing $s(i)$. From (3.1.1) and (3.1.3), there is $\underline{x}_2^N \in A^N$ such that

$$\mathcal{E}_{X,Z^{s(i)}}(\underline{x}_2^N, \underline{X}^{,N})$$

$$\leq 2\gamma \mathcal{E}_X \mathcal{E}_{X,Z^{s(i)}}(\underline{X}^N, \underline{X}^{,N}) \ ; \ i = 1, \ \ldots \ , \gamma, \text{ and} \qquad (3.1.4a)$$

$$Z(\underline{x}_1^N, \underline{x}_2^N)$$

$$\leq \min_i [\ 4\gamma \mathcal{E}_X \mathcal{E}_{X,Z^{s(i)}}(\underline{X}^N, \underline{X}^{,N})\ ]^{1/s(i)} \ . \qquad (3.1.4b)$$

Let $c_2^N = \{ \underline{x}_1^N, \underline{x}_2^N \}$. Next, we suppose that we have $c_M^N = \{ \underline{x}_1^N, \ \ldots \ , \underline{x}_M^N \}$, $\underline{x}_i^N \in A^N$, such that

$$\mathcal{E}_{X,Z^{s(i)}}(\underline{x}_m^N, \underline{X}^{,N})$$

$$\leq 2\gamma \mathcal{E}_X \mathcal{E}_{X,Z^{s(i)}}(\underline{X}^N, \underline{X}^{,N}) \ ; \ i = 1, \ \ldots \ , \gamma \ ; \qquad (3.1.5a)$$

$$m = 1, \ \ldots \ , M, \text{ and}$$

$$F(c_M^N)$$

$$\leq \min_i [\ 4\gamma(M-1) \mathcal{E}_X \mathcal{E}_{X,Z^{s(i)}}(\underline{X}^N, \underline{X}^{,N})\ ]^{1/s(i)} \ . \qquad (3.1.5b)$$

Then, Markov's inequality implies

$$\Pr\{ \sum_{m=1}^{M} Z(\underline{x}_m^N, \underline{X}^{,N})$$

$$< \min_i [\ 2 \sum_{m=1}^{M} \mathcal{E}_{X,Z^{s(i)}}(\underline{x}_m^N, \underline{X}^{,N})\ ]^{1/s(i)} \ \}$$

$$\leq \Pr\{ \sum_{m=1}^{M} Z^{s^*}(\underline{x}_m^N, \underline{X}^{\prime N}) > 2 \sum_{m=1}^{M} \mathcal{E}_{X^{\prime}} Z^{s^*}(\underline{x}_m^N, \underline{X}^{\prime N}) \}$$

$$< 1/2 ,$$

where s^* is the minimizing $s(i)$ and we used $0 < s^*$ in the first inequality. The above inequalities imply, together with (3.1.1) and(3.1.5), that there is $\underline{x}_{M+1}^N \in A^N$ such that

$$\mathcal{E}_{X^{\prime}} Z^{s(i)}(\underline{x}_{M+1}^N, \underline{X}^{\prime N})$$

$$\leq 2\gamma \mathcal{E}_X \mathcal{E}_{X^{\prime}} Z^{s(i)}(\underline{X}^N, \underline{X}^{\prime N}) \ ; \ i = 1, \ \ldots \ , \gamma, \qquad (3.1.6a)$$

and

$$\sum_{m=1}^{M} Z(\underline{x}_m^N, \underline{x}_{M+1}^N) \leq \min_i [\ 4\gamma M \mathcal{E}_X \mathcal{E}_{X^{\prime}} Z^{s(i)}(\underline{x}^N, \underline{X}^{\prime N}) \ ]^{1/s(i)} .$$

From the latter inequality,

$$F(c_M^N \cup \{\underline{x}_{M+1}^N\})$$

$$= \frac{1}{M+1} \{ MF(c_M^N) + 2 \sum_{m=1}^{M} Z(\underline{x}_m^N, \underline{x}_{M+1}^N) \}$$

$$\leq \min_i [\ 4\gamma M \mathcal{E}_X \mathcal{E}_{X^{\prime}} Z^{s(i)}(\underline{X}^N, \underline{X}^{\prime N}) \ ]^{1/s(i)}, \qquad (3.1.6b)$$

where we used $M(M-1)^{1/s(i)} \leq M^{1/s(i)}(M-1)$ and $s(i) \leq 1$.

Since we already have c_2^N, we now conclude that there is a series of code enlargements $c_{M+1}^N = c_M^N \cup \{\underline{x}_{M+1}^N\}$ such that each c_M^N satisfies

$$F(c_M^N)$$

$$\leq \exp\{ -N \max_i [E_x(\rho(i),p,P) - (\rho(i)/N)\log[4\gamma(M-1)]] \}$$

where we put $\rho(i) = 1/s(i)$. Let $\overline{R}_i = (i/N)\log \alpha$, let $\gamma = N$, and let $\rho = \rho(i)$ maximize $E_x(\rho,p,P) - \rho\overline{R}_i$. Then, for each M such that $\overline{R}_j \geq (1/N)\log[4N(M-1)] > \overline{R}_{j-1}$, we have

$$F(c_M^N) \leq \exp\{ -N \max_i [E_x(\rho(i),p,P) - \rho(i)\overline{R}_j] \}$$

$$\leq \exp\{ -N E_{ex}(p,P,R_M+\delta_N) \}$$

where $R_M = (1/N)\log(M-1)$ and $\delta_N = (1/N)\log 4\alpha N$. Thus we proved the following theorem.

Theorem 3.1.1: For any N, there is a sequence of block codes $c_{M+1}^N = c_M^N \cup \{\underline{x}_{M+1}^N\}$ such that, for each M, the MLD yields

$$P_e(c_M^N) \leq \exp\{ -N E_{ex}(p,P,R_M+\delta_N) \}$$

where $R_M = (1/N)\log(M-1)$ and $\delta_N = (1/N)\log 4\alpha N$.

$E_{ex}(p,P,R)$ is not the true expurgated exponent function. If we intend to obtain a hierarchy with respect to $E_{ex}(P,R)$, we have to vary p for each M. However such adaptation for rates will destroy our basic argument.

Omura [4] shows a recursive argument for $E_{ex}(P,R)$. However, his argument leads to a hierarchical series of codes relative to the expurgated exponent function only if a pmf p achieves $E_{ex}(P,R)$ simultaneously for all R, which is satisfied only for symmetric channels.

To complete our argument, we specialize the channel into a binary symmetric channel (BSC) and codes into linear codes. The BSC P is characterized by A = B = {0,1} and, for $0 \leq \varepsilon \leq 1$,

$$P(b|a) = \begin{cases} 1 - \varepsilon \ ; & \text{if } a = b \\ \varepsilon \ ; & \text{if } a \neq b \ . \end{cases}$$

A (K,N)-linear code is a block code whose codewords are genrated by K N-vectors $\underline{g}_k^N \in \{0,1\}^N$ according to the linear combination

$$\underline{x}^N = \sum_{k=1}^{K} \gamma_k \, \underline{g}_k^N \ ,$$

where $\gamma_k \in \{0,1\}$ and the summation and multiplication are arithmetic in GF(2), the Galois field with elements 0 and 1. It is easy to see that $E_{ex}(p,P,R)$ is maximized by the symmetric pmf p_s, $p_s(0) = p_s(1) = 1/2$, and that

$$E_x(\rho,p_s,P) = -\rho \log\{ (1/2)[Z^{1/\rho}(0) + Z^{1/\rho}(1)] \}$$

for all $\rho \geq 1$ where $Z(0) = Z(0,0)$ and $Z(1) = Z(0,1)$. In the remainder of this section, we assume $p = p_s$.

First, we consider $P_e(c_1^*)$ for (1,N)-linear codes $c_1^* = \{ \underline{0}^N, \underline{g}^N \}$. The argument goes in almost the same way as in the previous proof. We can see the existence of a good c_1^* such that

$$F(c_1^*) \leq \inf_{\rho \geq 1} [\mathcal{E}_X Z^{1/\rho}(\underline{X}^N)]^\rho$$

where $Z(\underline{x}^N) = Z(\underline{0}^N, \underline{x}^N)$ for all $\underline{x}^N \in \{0,1\}^N$. Thus, we have

$$F(c_1^*) \leq \exp\{ -N E_{ex}((1/N)\log 1 , P)) \}.$$

Next, we suppose that there is a good (K,N)-linear code c_K^* which satisfies

$$F(c_K^*) \leq \inf_{\rho \geq 1} [(2^K - 1) \mathcal{E}_X Z^{1/\rho}(\underline{X}^N)]^\rho .$$

We note that the right-hand side is precisely $\exp\{ -N E_{ex}((1/N)\log(2^K-1) ,P) \}$. Then, for a non-trivial enlargement $c^*_{K+1} = \{ \underline{x}^N = \underline{x}'^N + \gamma \underline{g}^N_{K+1}, \underline{x}'^N \in c^*_K, \gamma = 0,1 \}$, we have

$$F(c^*_{K+1}) = \sum_{\substack{\underline{x}^N \in c^*_K \\ \underline{x}^N \neq \underline{0}^N}} Z(\underline{x}^N) + \sum_{\underline{x}^N \in c^*_K} Z(\underline{x}^N + \underline{g}^N_{K+1}) \quad .$$

By a random coding argument, there is $\underline{g}^N_{K+1}$ such that the last term is bounded by $\inf_{\rho \geq 1} [2^K \mathcal{E}_X Z^{1/\rho}(\underline{X}^N)]^\rho$. Therefore, for any $\rho \geq 1$, we have the bound

$$F(c^*_{K+1}) \leq \inf_{\rho \geq 1} [(2^K-1)^\rho + 2^{\rho K}][\mathcal{E}_X Z^{1/\rho}(\underline{X}^N)]^\rho$$

$$\leq \exp\{ -N E_{ex}((1/N)\log(2^{K+1}-1) ,P) \}.$$

We summarize the result in a theorem:

Theorem 3.1.2: For a BSC P, there is a sequence of binary N-vectors $\underline{g}^N_1, \underline{g}^N_2, \ldots$ such that, for each K, the (K,N)-linear code c^*_K generated by $\{\underline{g}^N_1, \ldots, \underline{g}^N_K\}$ has the average error probability

$$P_e(c^*_K) \leq \exp\{ -N E_{ex}(P,R^*_K) \}$$

for the MLD, where $R^*_K = (1/N)\log(2^K-1)$.

Hierarchy Towards Channel Capacity

An important exponent function other than $E_{ex}(p,P,R)$ is

$$E_r(p,P,R) = \max_{0 \leq \rho \leq 1} [\, E_o(\rho,p,P) - \rho R \,] \quad .$$

We call $E_r(p,P,R)$ the random coding exponent function, although the so-called random coding exponent function is the maximum $\max_p E_r(p,P,R)$. From random coding arguments, it is relatively easy to see that, for any fixed ρ, $0 \leq \rho \leq 1$, there exists a code c_M^N such that $P_e(c_M^N) \leq \exp\{ -N[E_o(\rho,p,P) - \rho R_M] \}$ and that, given such code c_M^N, we can always select a subcode c_{M-1}^N from it so that $P_e(c_{M-1}^N) \leq \exp\{ -N[E_o(\rho,p,P) - \rho R_{M-L}] \}$. However we can never vary ρ in this argument. Non-exixtence of any fine hierarchical structure may be a feature of random coding exponent function that gives the exact exponent at high rates. Nevertheless, it seems desirable to show a hierarchy under an exponent related to the random coding exponent function since the structure of codes of rates near the channel capacity is also interesting in connection with channel coding theorems.

To see such a hierarchy, we assign new decoding regions to a given c_M^N such that, for each m,

$$\{ \log \frac{P(\underline{y}^N|\underline{x}_m^N)}{q(\underline{y}^N)} \geq NR_{m+1} \quad \text{and} \quad \log \frac{P(\underline{y}^N|\underline{x}_{m'}^N)}{q(\underline{y}^N)} < NR_{m'+1} \ ;$$

$$\text{all } m' = 1, \ \ldots \ , m-1 \ \} \ , \tag{3.1.7}$$

where q is a pmf on B given by $q(b) = \Sigma_{a\in A} P(b|a)p(a)$ for each $b \in B$. We call a decoder with this decoding rule a modified MLD. This decoding rule is not maximum likelihood decoding, and the probability of decoding error for m in this scheme depends only on the first m codewords;namely $P_{e,m}(c_M^N) = P_{e,m}(c_m^N)$ where $c_m^N = \{ c_1^N, \ \ldots \ , c_m^N \}$. Using standard bounds, we see

$$P_{e,m}(c_m^N)$$

$$\leq \sum_{\underline{y}^N \in B^N} P(\underline{y}^N|\underline{x}_m^N) \ \chi[\log \frac{P(\underline{y}^N|\underline{x}_m^N)}{q(\underline{y}^N)} \leq NR_{m+1} \ ]$$

$$+ \sum_{\underline{y}^N \in B^N} P(\underline{y}^N|\underline{x}_m^N) \quad [\log \frac{P(\underline{y}^N|\underline{x}_{m'}^N)}{q(\underline{y}^N)} \geq NR_{m'+1} \ ,$$

$$\text{some } m' = 1, \ \ldots \ , m-1 \ ]$$

$$\leq m^{\rho/(1+\rho)} \sum_{\underline{y}^N \in B^N} q^{\rho/(1+\rho)}(\underline{y}^N) \ P^{1/(1+\rho)}(\underline{y}^N|\underline{x}_m^N)$$

$$+ \sum_{m'=1}^{m-1} m'^{-1/(1+\rho)} \sum_{\underline{y}^N \in B^N} P(\underline{y}^N | \underline{x}_m^N)\ q^{-1/(1+\rho)}(\underline{y}^N)$$

$$\times\ P^{1/(1+\rho)}(\underline{y}^N | \underline{x}_{m'})$$

for all $m = 1, \ldots, M$ and all $0 \leq \rho \leq 1$. To simplify the arguments we fix ρ for a while. Suppose that c_M^N satisfies

$$\sum_{\underline{y}^N \in B^N} q^{\rho/(1+\rho)}(\underline{y}^N)\ P^{1/(1+\rho)}(\underline{y}^N | \underline{x}_m^N)$$

$$\leq 2\, \xi_X \sum_{\underline{y}^N \in B^N} q^{\rho/(1+\rho)}(\underline{y}^N)\ P^{1/(1+\rho)}(\underline{y}^N | \underline{X}^N) \quad .$$

Denote the expectation in the right-hand side by $\phi(\rho)$. Then, with the aid of Hölder's inequality, we see that $\phi(\rho) \leq \exp\{ -N\, E_o(\rho,p,P)/(1+\rho) \}$. By the same argument as that used for the transition from (3.1.5) to (3.1.6), we can show that there exists $\underline{x}_{M+1}^N$ such that

$$\sum_{\underline{y}^N \in B^N} q^{\rho/(1+\rho)}(\underline{y}^N)\ P^{1/(1+\rho)}(\underline{y}^N | \underline{x}_{M+1}^N) \leq 2\phi(\rho)$$

and

$$\sum_{m'=1}^{m-1} m'^{-1/(1+\rho)} \sum_{\underline{y}^N \in B^N} P(\underline{y}^N | \underline{x}_{M+1}^N)\ q^{-1/(1+\rho)}(\underline{y}^N)\ P^{1/(1+\rho)}(\underline{y}^N | \underline{x}_{m'}^N)$$

$$\leq 2 \sum_{m'=1}^{m-1} m'^{-1/(1+\rho)} \sum_{\underline{y}^N \in B^N} q^{\rho/(1+\rho)}(\underline{y}^N)\, P^{1/(1+\rho)}(\underline{y}^N | \underline{x}_m^N)$$

In this way we can show the existence of $\underline{x}_1^N, \underline{x}_2^N, \ldots$ such that the code $c_M^N = \{ \underline{x}_1^N, \ldots, \underline{x}_M^N \}$ satisfies

$$P_{e,M}(c_M^N) \leq 2[\, m^{1/(1+\rho)} + 2 \sum_{m'=1}^{m-1} m'^{-1/(1+\rho)} \,]\phi(\rho)$$

for each M. Thus, if we note the inequality $\sum_{m'=1}^{m-1} m'^{-1/(1+\rho)} \leq (1+\rho^{-1})[\, m^{\rho/(1+\rho)} - 1 \,] \leq m^{\rho/(1+\rho)}$, and if we apply the technique used in the proof of Theorem 3.1.1, we have the following theorem.

Theorem 3.1.3: For any N, there is a sequence of block codes $c_{M+1}^N = c_M^N \cup \{\underline{x}_{M+1}^N\}$ such that, for each M, the modified MLD yields

$$P_{e,m} \leq \exp\{ -N[\, E_r^\circ(p,P,R_{m+1}) - \delta_N' \,] \}$$

for all $m = 1, \ldots, M$ where $\delta_N = (1/N)\log 6\alpha N$ and

$$E_r^\circ(p,P,R) \triangleq \max_{0 \leq \rho \leq 1} [\, E_o(\rho,p,P) - \rho R \,]/(1+\rho) \quad .$$

It is almost evident that $E_r(p,P,R) \geq E_r^\circ(p,P,R) \geq (1/2)E_r(p,P,R)$ and $E_r^\circ(p,P,R)/E_r(p,P,R) \to 1$ as $R \to C(P)$, the channel capacity of P. Thus, Theorem 3.1.3 gives

a channel coding theorem for DMC's.
Indeed, this approach serves as a version of
Feinstein's argument on his fundamental lemma,
an important lemma in classical Shannon theory [5],[6].

2. Universal performance of good codes

Up to here, all good codes are obtained for any, but fixed, DMC. However, due to the lack of consistency in channel characteristics, sometimes we have to make a code not knowing about channel identity in each communication. Csiszár, Körner, and Marton [7] (also see [8])show a surprising answer to coding DMC's under such a situation.

Let S be the totality of DMC's with the input alphabet $A = \{ 0, \ldots, \alpha-1 \}$ and the output alphabet $B = \{ 0, \ldots, \beta-1 \}$, and define the mutual information function between $\underline{x}^N \in A^N$ and $\underline{y}^N \in B^N$ by

$$I(\underline{x}^N, \underline{y}^N)$$

$$\triangleq \sum_{a \in A} \sum_{b \in B} \frac{1}{N} N(a,b|\underline{x}^N, \underline{y}^N) \log \frac{N(a,b|\underline{x}^N, \underline{y}^N)}{N(a|\underline{x}^N) N(b|\underline{y}^N)} \quad ,$$

where $N(a,b|\underline{x}^N, \underline{y}^N)$ is the number of (a,b) in $(\underline{x}^N, \underline{y}^N)$, and $N(a|\underline{x}^N)$ and $N(b|\underline{y}^N)$ are its respective marginal sums over $a \in A$ and $b \in B$. We say that a code c^N has a fixed composition p_N if p_N is a pmf on A and

$$N(a|\underline{x}_m^N) = N\, p_N(a) \quad ; \text{ all } m = 1, \ldots, M \text{ and } a \in A.$$

A maximum mutual information decoder (MMID) is a decoder which decodes every $\underline{y}^N$ into m if $I(\underline{x}_m^N, \underline{y}^N) \geq I(\underline{x}_{m'}^N, \underline{y}^N)$ for all $m' \neq m$. The following is one of their main results.

Theorem 3.2.1: For each N, there is a block code c^N of fixed composition p_N and rate $R = (1/N)\log M$ such that the MMID yiels

$$P_e(c^N) \leq \exp\{ -N[E_r^*(p_N, P, R) - o(N)] \}$$

where $o(N)/N \to 0$ as $N \to \infty$.

The function $E_r^*(p,P,R)$, which is not defined here, is called the random coding exponent function for fixed composition codes and is such that:

$$E_r^*(p,P,R) \geq E_r(p,P,R) \qquad \text{and}$$

$$\max_p E_r^*(p,P,R) = E_r(P,R) \quad .$$

Despite of these strong mathematical implications, however, the codewords are to have a fixed composition, and the mutual information function is not as cumulative as the log-likelihood function $\log P(\underline{y}^N | \underline{x}^N)$ is; both keep the theorem away from application to tree codes.

In this section, we treat the problem in a somewhat different way, and extend the result to tree codes in Section 6.2 .

From Lemma 2.1.1, we have, for any $Q,P \in S$,

$$P_e(c^N) = \frac{1}{M} \sum_{m=1}^{M} \sum_{\underline{y}^N \in Y^c_{N,m}(P)} P(\underline{y}^N | \underline{x}^N_m)$$

$$\leq \frac{1}{M} \sum_{m=1}^{M} \sum_{\underline{y}^N \in Y^c_{N,m}(Q)} P(\underline{y}^N | \underline{x}^N_m)$$

$$= \frac{1}{M} \sum_{m=1}^{M} \sum_{\underline{y}^N \in Y^c_{N,m}(Q)} Q(\underline{y}^N | \underline{x}^N_m) \exp\left[\log \frac{P(\underline{y}^N | \underline{x}^N_m)}{Q(\underline{y}^N | \underline{x}^N_m)} \right]$$

where $Q_e(c^N)$ symbolizes that Q is used. Thus, if we let

$$d(P|Q) = \max_{a \in A,\ b \in B} \log \frac{P(b|a)}{Q(b|a)}$$

with the convention that $\log(0/0) = 0$, then we have the channel mismatch relation:

$$P_e(c^N) \leq Q_e(c^N)\ e^{Nd(P|Q)} \quad ; \text{ all } P,Q \in S \qquad (3.2.1)$$

The next lemma is proved later.

Lemma 3.2.1: For any $0 \leq \varepsilon \leq 1/2\alpha^2$, there is a

subset $S(\varepsilon)$ of S with at most $\varepsilon^{-2\alpha\beta}$ DMC's such that, for each $P \in S$, a DMC $Q \in S(\varepsilon)$ satisfies the followings:

1) $d(P|Q) \leq \varepsilon$, and

2) $E_o(\rho,p,P) - E_o(\rho,p,P) \leq 2\beta^3\varepsilon$ for $0 \leq \rho \leq 1$.

For an arbitrary, but fixed, $\varepsilon > 0$ and pmf p on A, let $S(\varepsilon)$ be the subset given in the lemma, and let $\mathcal{C}^N$ be a random block code each of whose random codewords has the probability $p(\underline{x}^N)$ for each $\underline{x}^N \in A^N$. Then, from Markov's ineqality, we have

$$\Pr\{ Q_e(\mathcal{C}^N) > 3\varepsilon^{-2\alpha\beta}\mathcal{E}Q_e \text{ , some } Q \in S(\varepsilon) \}$$

$$\leq \sum_{Q\in S(\varepsilon)} \Pr\{ Q_e(\mathcal{C}^N) > 3\varepsilon^{-2\alpha\beta}\mathcal{E}Q_e \}$$

$$< \sum_{Q\in S(\varepsilon)} \varepsilon^{2\alpha\beta}/3$$

$$\leq 1/3 \ ,$$

where $\mathcal{E}$ denotes the expectation operator relative to $\mathcal{C}^N$, and we put

$$\mathcal{E}Q_e \triangleq \mathcal{E}Q_e(\mathcal{C}^N) \ .$$

Therefore we have the probability

$$\Pr\{\, Q_e(\mathcal{C}^N) \leq 3\varepsilon^{-2\alpha\beta}\mathcal{E}Q_e \quad , \text{ every } Q \in S(\varepsilon) \,\} > 2/3 \ .$$

From this probability we see the existence of a code c^N such that $Q_e(c^N) \leq 3\varepsilon^{-2\alpha\beta}\mathcal{E}Q_e$ holds for all $Q \in S(\varepsilon)$. Thus, from the channel mismatch relation, Lemma 2.2.1, and Lemma 3.2.1, the error probability for this code is

$$P_e(c^N)$$

$$\leq \exp\{\, -N[\, E_o(\rho,p,Q) - \rho R + (1/N)\log(\varepsilon^{2\alpha\beta}/3) - \varepsilon \,] \,\}$$

$$\leq \exp\{\, -N[\, E_o(\rho,p,P) - \rho R + (1/N)\log(\varepsilon^{2\alpha\beta}/3) - \varepsilon - 2\beta^3\varepsilon \,] \,\}$$

for all $0 \leq \rho \leq 1$ and all $P \in S$. And we have a theorem:

Theorem 3.2.2: For suffciently large N, there exists a block code c^N of rate $R = (1/N)\log M$ such that MLD yields

$$P_e(c^N) \leq \exp\{\, -N[\, E_r(p,P,R) - o(N) \,] \,\}$$

for any $P \in S$ where $o(N)/N \to 0$ as $N \to \infty$.

We note that, whereas MMID never needs the exact description of channels, MLD does. However this

approach is simple and flexible enough for application to tree codes.

Next we show another universality with respect to the expurgated exponent function. Let $\mathcal{C}^N$ be a random block code consisting of 2M independent random codewords each of which assumes $\underline{x}^N \in A^N$ with the probability $p(\underline{x}^N)$, and let

$$\mathcal{E}Q_e^s \triangleq \mathcal{E}Q_{e,m}^s(\mathcal{C}^N)$$

for $0 < s \leq 1$ and each m. (It is well-defined for we can see that the right-hand side is independent of m in the same manner as $\mathcal{E}Q_{e,m}(\mathcal{C}^N)$) From Markov's inequality, we have

$$\Pr\{ Q_{e,m}(\mathcal{C}^N) > \inf_{0<s\leq 1} (3\varepsilon^{-2\alpha\beta}\mathcal{E}Q_e^s)^{1/s} \text{ , some } Q \in S(\varepsilon) \}$$

$$< \inf_{t>0} \{ \varepsilon^{-2\alpha\beta}\mathcal{E}Q_e^t / [\inf_{0<s\leq 1} (3\varepsilon^{-2\alpha\beta}\mathcal{E}Q_e^s)^{t/s}] \}$$

$$\leq 1/3$$

for each m = 1, ... ,2M, where the last inequality follows by letting t equal s .
In view of this inequality, if we let ϕ_m be the indicator function of the event

$$Q_{e,m}(\mathcal{C}^N) \leq \inf_{0<s\leq 1} (3\varepsilon^{-2\alpha\beta}\mathcal{E}Q_e^s)^{1/s} \text{ ; all } Q \in S(\varepsilon),$$

then we have

$$\Pr\{ \frac{1}{M} \sum_{m=1}^{2M} \phi_m \geq 1 \} > 2/3 .$$

Therefore, there exists c'^N with 2M members such that $\phi_m = 1$ for at least a half of them, and hence there exists c^N, a subset of c'^N consisting of M codewords, such that

$$Q_{e,m}(c^N) \leq \inf_{0<s\leq 1} (3\varepsilon^{-2\alpha\beta} Q_e^s)^{1/s}$$

for all $Q \in S(\varepsilon)$ and all $m = 1, \ldots, M$. Thus, letting $\rho = 1/s$, we have

$$P_e(c^N)$$

$$\leq Q_e(c^N)\, e^{\varepsilon N}$$

$$\leq \exp\{ -N \sup_{\rho\geq 1} [E_x(\rho,p,Q) - \rho R - (\rho/N)\log 9\varepsilon^{-2\alpha\beta} - \varepsilon] \}.$$

The following lemma is proved later.

Lemma 3.2.2: For each $P \in S$, there exists $Q \in S(\varepsilon)$ that satisfies the conditions 1) and 2) in Lemma 3.2.1 as well as the additional one:

3) $E_x(\rho,p,P) - E_x(\rho,p,Q) \leq \rho\beta(2\beta^2)^{1/\rho}$; $\rho \geq 1$.

From the lemma, the above inequality is further upper bounded:

$$P_e(c^N)$$

$$\leq \exp\{ -N \sup_{\rho \geq 1} [E_x(\rho,p,P) - \rho R - \rho\beta(2\beta^2\varepsilon)^{1/\rho} - \rho\varepsilon_1] \}$$

where, from $\rho \geq 1$, we put $\varepsilon_1 = (1/N)\log 9\varepsilon^{-2\alpha\beta} + \varepsilon$. Here we must note that the term $(2\beta^2\varepsilon)^{1/\rho}$ crucially depends on $\rho \geq 1$. Now let S_γ be the set of all $P \in S$ such that $E_{ex}(p,P,R)$ is attained by $\rho \leq \gamma$ in $\sup_{\rho \geq 1} [E_x(\rho,p,P) - \rho R]$. Then, from the all arguments above, for any $\varepsilon^* > 0$ and any $\gamma > 0$, there exists a code c^N such that

$$P_e(c^N) \leq \exp\{ -N E_{ex}(p,P, R + \varepsilon_2) \}$$

for all $P \in S_\gamma$, where $R = (1/N)\log M$ and

$$\varepsilon_2 = \beta(2\beta^2\varepsilon)^{1/\gamma} + (1/N)\log 9\varepsilon^{-2\alpha\beta} + \varepsilon.$$

Therefore, for any increasing positive numbers γ_i and decreasing positive numbers ε_i, there exist increasing positive integers N_i, M_i and block codes c_i with M_i codewords having block length N_i such that $(1/N_i)\log M_i \geq R - \varepsilon_i$ and

$$P_e(c_i) \leq \exp\{ -N_i E_{ex}(p,P,R+\varepsilon_i) \}$$

for all $P \in S_{\gamma_i}$. Let S_∞ be the set of all $P \in S$ with zero zero-error-capacity. It is evident that $S_\gamma \to S$ as $\gamma \to \infty$ for $R > 0$. Therefore we have proved the following theorem.

Theorem 3.2.3: For any $R > 0$, there exist block codes c_i each having M_i codewords of block length N_i such that, for any $P \in S_\infty$, the MLD yields

$$P_e(c_i) \leq \exp\{ -N_i[E_{ex}(p,P,R) - \varepsilon_i] \}$$

for all sufficiently large i, where $\varepsilon_i \to 0$ as $i \to \infty$.

This theorem is weaker than the previous one; one block code does not necessarily possess a uniform bound over all $P \in S$. Finally, to complete our argument we combine Theorem 3.2.2 and Theorem 3.2.3 in a single form:

Theorem 3.2.4: For any $R > 0$, there exist block codes c_i each having M_i codewords of block length N_i such that, for any $P \in S_\infty$, the MLD yields

$$P_e(c_i) \leq \exp\{ -N_i[E_c(p,P,R) - \varepsilon_i] \}$$

for all sufficiently large i, where $\varepsilon_i \to 0$ as $i \to \infty$, and

$$E_c(p,P,R) \triangleq \max\{ E_r(p,P,R), E_{ex}(p,P,R) \}.$$

The proof is seen in the latter part of this section.

Finally, we briefly mention to a recent result due to Csiszár and Körner [9]. They show, in a different framework, that, for any $R > 0$ and any $\varepsilon > 0$, there exists a block code c^N with a sufficiently large block length N and a rate larger than $R - \varepsilon$ such that "MLD" decodes the code in the error probability

$$P_e(c^N) \leq \exp\{ -N[E_c^*(p_N,P,R) - \varepsilon] \}$$

where p_N is the fixed composition of the code and

$$E_c^*(p,P,R) \triangleq \max\{ E_r^*(p,P,R), E_{ex}^*(p,P,R) \}.$$

The function $E_{ex}^*(p,P,R)$ is a counterpart, for fixed composition codes, of the expurgated exponent function. It is shown in [8] that $\max_p E_c^*(p,P,R) = E_c(p,P,R)$. It is interesting to observe that MLD attains the expurgated exponent function universally over channels, but MMID may not.

Appendix to Chapter III

Proof of Lemma 3.2.1

Let n be an integer such that $n + 1 \leq 1/\varepsilon^2 \leq n + 2$, and let $S(\varepsilon)$ be the set of all $Q \in S$ taking rational values $Q(b|a) = k/n$ ($k = 0,1, \ldots ,n$) for each $a \in A$ and each $b \in B$. Clearly the size of $S(\varepsilon)$ is less than $(n+1)^{\alpha\beta}$ ($\leq \varepsilon^{-2\alpha\beta}$). First, fix $P \in S$ and $a \in A$ arbitrarily, and let $b^* \in B$ be the letter such that $P(b^*|a) \geq P(b|a)$ for all $b \in B$. Then there exists $Q \in S(\varepsilon)$ that satisfies

$$P(b|a) + 1/n > Q(b|a) \geq P(b\ a) \quad ; \text{ all } b(\neq b^*), \text{ and}$$

$$| Q(b^*|a) - P(b^*|a) | \leq (\beta-1)/n.$$

For this channel Q, it follows that

$$\log \frac{P(b|a)}{Q(b|a)} \leq 0 \qquad ; \text{ all } b(\neq b^*), \text{ and}$$

$$\log \frac{P(b^*|a)}{Q(b^*|a)} < \log \frac{P(b^*|a)}{P(b^*|a)-(\beta-1)/n} = -\log\left[1 - \frac{(\beta-1)/n}{P(b^*|a)} \right] .$$

Since $P(b^*|a) \geq \beta^{-1}$, $\beta \geq 2$, and $\beta^2\varepsilon < 1/2$ by assumption,

$$\frac{(\beta-1)/n}{P(b^*|a)} \leq \frac{\beta^2 - \beta}{n} \leq \frac{\beta^2 - \beta + 1}{n + 2} \leq \beta^2\varepsilon^2 \leq \varepsilon/2.$$

Hence, for $\varepsilon < 1/8$,

$$\log \frac{P(b^*|a)}{Q(b^*|a)} \leq -\log(1 - \varepsilon/2).$$

Thus statement 1) is proved since $P \in S$ and $a \in A$ are arbitrary. To prove 2), note that

$$\begin{aligned}
&\Big| \sum_{b\in B} \Big\{ \sum_{a\in A} p(a)P(b|a)^{1/(1+\rho)} \Big\}^{1+\rho} \\
&\qquad - \sum_{b\in B} \Big\{ \sum_{a\in A} p(a)Q(b|a)^{1/(1+\rho)} \Big\}^{1+\rho} \Big| \\
&\leq \sum_{b\in B} \Big| \Big\{ \sum_{a\in A} p(a)P(b|a)^{1/(1+\rho)} \Big\}^{1+\rho} \\
&\qquad - \Big\{ \sum_{a\in A} p(a)Q(b|a)^{1/(1+\rho)} \Big\}^{1+\rho} \Big| \\
&\leq \sum_{b\in B} (1+\rho) \sum_{a\in A} p(a) \, | \, P(b|a)^{1/(1+\rho)} - Q(b|a)^{1/(1+\rho)} | \\
&\leq \sum_{b\in B} (1+\rho) \sum_{a\in A} p(a) \, | \, P(b|a) - Q(b|a) \, |^{1/(1+\rho)} \\
&\leq (1+\rho)\beta\{(\beta-1)/n\}^{1/(1+\rho)} \\
&\leq 2\varepsilon\beta^2.
\end{aligned}$$

On the other hand,

$$\sum_{b\in B} \Big\{ \sum_{a\in A} p(a)P(b|a)^{1/(1+\rho)} \Big\}^{1+\rho} \geq \beta^{-1/\rho}.$$

Thus, from $\log x \leq x-1$,

$$E_o(\rho,p,P) - E_o(\rho,p,P) \leq 2\varepsilon\beta^3 .$$

Proof of Lemma 3.2.2.

Let Q be the channel in the above proof for the given P. First note that, for $a, a' \in A$,

$$\left| \{ \sum_{b\in B} \sqrt{P(b|a)P(b|a')} \}^{1/\rho} - \{ \sum_{b\in B} \sqrt{Q(b|a)Q(b|a')} \}^{1/\rho} \right|$$

$$\leq \{ \sum_{b\in B} \left| \sqrt{P(b|a)P(b|a')} - \sqrt{Q(b|a)Q(b|a')} \right| \}^{1/\rho}$$

$$\leq \{ \sum_{b\ B} \left| P(b|a)P(b|a') - Q(b|a)Q(b|a') \right|^{1/2} \}^{1/\rho}$$

$$\leq \{ \beta\sqrt{2(\beta-1)/n} \}^{1/\rho}$$

$$\leq (2\varepsilon\beta^2)^{1/\rho}$$

for $0 < \varepsilon \leq 1/2\beta^2$. Hence, by the inequality $\log x \leq x-1$, it follows that

$$E_x(\rho,p,P) - E_x(\rho,p,P) \leq (2\varepsilon\beta^2)^{1/\rho} \exp\{ E_x(\rho,p,P)/\rho \}$$

$$\leq \beta(2\varepsilon\beta^2)^{1/\rho}$$

for $0 < \varepsilon \leq 1/2\beta^2$, where the last inequality follows since $E_x(\rho,p,P)/$ is decreasing in ρ and $E_x(1,p,P) = E_o(1,p,P) \leq \log \beta$. The lemma follows from $E_x(\rho,p,P)/\rho \leq R$.

Proof of Theorem 3.2.4

Let $\mathcal{C}^N$ be the random code used in the proof of Theorem 3.2.2. Then, for any $\varepsilon > 0$, we have

$$\Pr\{ Q_{e,m}(\mathcal{C}^N) \leq 3\varepsilon^{-2\alpha\beta} \mathcal{E} Q_e, \text{ all } Q \in S(\varepsilon) \} \geq 2/3$$

and

$$\Pr\{ Q_{e,m}(\mathcal{C}^N) \leq \inf_{0<s\leq 1} (3\varepsilon^{-2\alpha\beta} \mathcal{E} Q_e^s)^{1/s}, \text{ all } Q \in S(\varepsilon) \} \geq 2/3.$$

Let ϕ_m be the indicator function of the joint event

$$Q_{e,m}(\mathcal{C}^N) \leq 3\varepsilon^{-2\alpha\beta} \mathcal{E} Q_e, \quad \text{and}$$

$$Q_{e,m}(\mathcal{C}^N) \leq \inf_{0<s\leq 1} (3\varepsilon^{-2\alpha\beta} \mathcal{E} Q_e^s)^{1/s} \; ; \quad \text{all } Q \in S(\varepsilon).$$

Then, we have

$$\Pr\{ \frac{1}{2M} \sum_{m=1}^{2M} \phi_m \geq \frac{1}{2} \} > 1/3.$$

Therefore there exists a code c^N such that $\phi_m = 1$ for a half of m's. Now let c^N be the code consisting of those codewords corresponding to such m's with appropriate renumbering. Since expurgation of codewords does not increase the error probability, we obtain

a code such that

$$Q_{e,m}(c^N) \leq 3\varepsilon^{-2\alpha\beta} \xi Q_e, \quad \text{and}$$

$$Q_{e,m}(c^N) \leq \inf_{0<s\leq 1} (3\varepsilon^{-2\alpha\beta} \xi Q_e^s)^{1/s} \quad ; \quad \text{all } Q \in S(\varepsilon).$$

Thus the proof is completed by constructing those codes for different values of ε.

Chapter IV

Convolutional Tree Coding of DMC

1. Tree and convolutional tree codes

The block codes that we have treated are codes made up from arbitrarily selected codewords, except the linear codes in Section 3.1. However, from the practical side, algebraic or geometric structure between codewords are often indispensable to facilitate highly reliable and practically implementable encoding-decoding.

For example consider a binary block code having the rate 1/3 bits-per-letter,

111101000	111101111
000111010	111010101
111010010	000111101
000000000	000000111 .

For this code, the decoder needs 72 bits memory other than calculation of probabilities. On the other hand, if we rearrange the ensemble as sited in Fig. 4.1.1, the tree-like systematic structure reduces the memory requirement from 72 to 42 bits. Such a code which has a tree-like skelton is called a tree code; the terminologies, nodes, root node, and branches, are also used to indicate individual elements of the skelton of the code tree. The encoder's action is simply to trace branches and to emit sequences attached to branches called branch sequences in order. In the

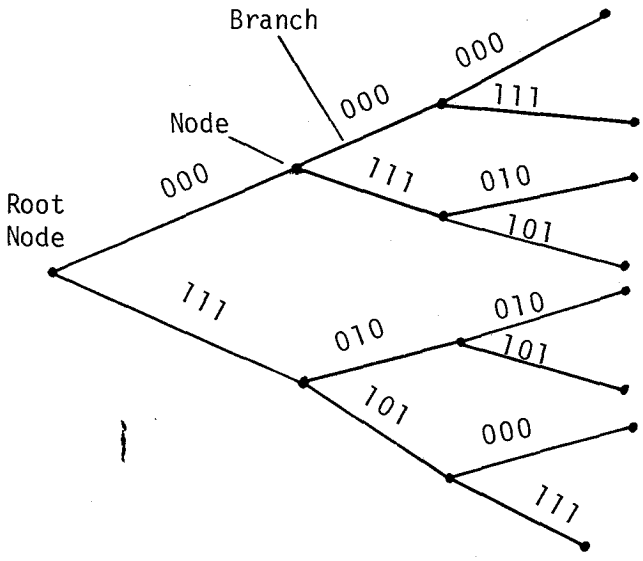


Fig. 4.1.1 — A Binary Tree Code

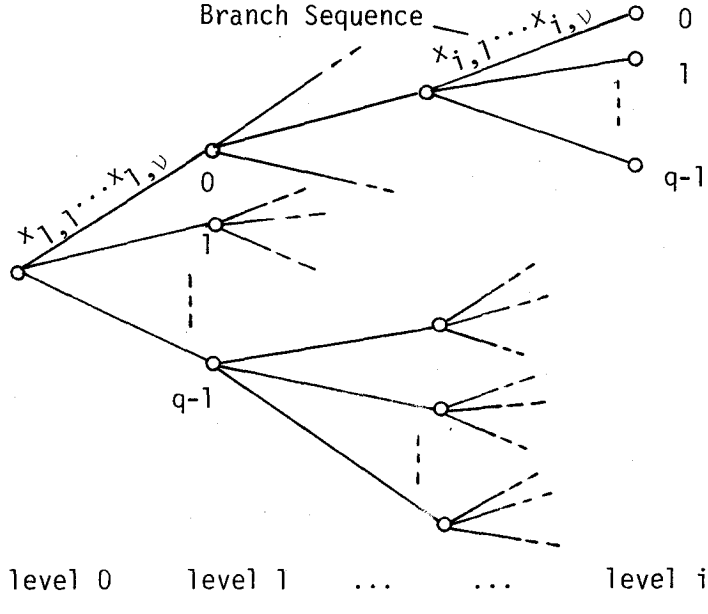


Fig. 4.1.2 — A q-nary Tree Code of Rate $\frac{1}{\nu}$ log q

example, assigning 0 to an upward move and 1 to a downward move on a node, the message 100 specifies the codeword 111010010. From the example, we see that the rate is almost determined by the number of branches growing on a node, q, and the number of letters assigned to each branch, ν.

Suppose that the channel input and output alphabets are $A = \{0,\ldots,\alpha-1\}$ and $B = \{0,\ldots,\beta-1\}$ respectively, and that message symbols are q-nary digits. A general code tree (or tree code) with relevant notations are depicted in Fig. 4.1.2. We say that a node is at the i-th level if the node is connected to the root node through i branches. The root node has the level 0. Any nodes on a path connecting the root node and a node are called antecedents of that node; conversely, the node is called a descendant of those nodes. On each node at the i-th level, there are emanating q branches, which are numbered from 0 to q-1, and each of which has a branch sequence consisting of letters from A, $\underline{x}_i = x_{i,1}\cdots x_{i,\nu}$. A message sequence $\underline{u}^L = u_1\ldots u_L$ specifies both a node in the tree and sequence $\underline{x}_1^\nu\ldots\underline{x}_L^\nu$, called the codeword for $\underline{u}^L$. Then, we also denote the node by $\underline{u}^L$. If the tree code has the maximum level L, then the block length is νL, and the rate is $(1/\nu)\log q$, which is virtually

independent of L.

However, the tree-like structure alone is insufficient to save decoder's memory. See the system in Fig. 4.1.3, where all the additions are modulo-2 sum . If, for each shift-register content, we arrange the outputs in a row, $v_1 v_2 v_3$, then we immediately obtain the previous example. Since the major operations are convolutions, such a code is called a <u>convolutional tree code</u>, or simply a convolutional code, and the shift-register length is called the <u>constraint length</u> of the code, K (= 3). The algebraic structure reduces memory requirement from 42 to approximately 0 bit, if the circuit is invariant throughout encoding (<u>time-invariant convolutional code</u>).

A more general convolutional encoder consists of a K-stage shift-register, adders and multipliers over GF(q), the Galois field with elements 0,1, ... ,q-1, and a channel letter selector, as shown in Fig. 4.1.4. A message sequence $\underline{u} = u_1 u_2 \ldots$ is fed into the shift-register one digit a time from the left; for each content $u_i u_{i-1} \cdots u_{i-K+1}$, ν q-nary digits $s_{i,j}$ are given by the linear operation over GF(q),

$$s_{i,j} = \sum_{k=0}^{K-1} u_{i-K} \, g_{k+1,j}^{(i)} \quad ; \; j = 1, \ldots, \nu \, . \tag{4.1.1}$$

The commutator arranges these q-nary digits into a

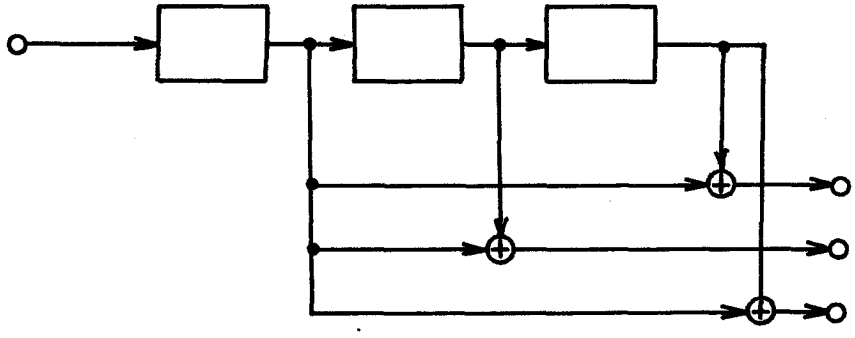


Fig. 4.1.3 — The Generator of the Binary Tree Code in Fig. 4.1.1

sequence $s_{i,1} \cdots s_{i,\nu}$ which is added componentwise to a bias sequence $v_{i,1} \cdots v_{i,\nu}$ consisting of digits from GF(q) to form $z_{i,1} \cdots z_{i,\nu}$ by

$$z_{i,j} = s_{i,j} + v_{i,j} \ .$$

Finally the channel letter selecter converts $z_{i,1} \cdots z_{i,\nu}$ letter-wise into the branch sequence $\underline{x}_i = x_{i,1} \cdots x_{i,\nu}$, according to the rule:

$$z \to \begin{cases} x \in GF(q), \text{ if } \sum_{a=1}^{x-1} n_a \leq z < \sum_{a=1}^{x} n_a \ , \\ \\ 0 \in GF(q), \text{ if } \quad z < n_0 \ , \end{cases} \tag{4.1.2}$$

where $n_0, n_1, \cdots n_{\alpha-1}$ are positive integers whose sum is q and z is interpreted as an integer in the "if" statements.

If $v_{i,j}$ and $g_{k,j}^{(i)}$ assume values in GF(q) independently with an equal probability, then, from the property of GF(q) arithmetic, all $z_{i,j}$ also assume values in GF(q) independently with an equal probability. Thus, from (4.1.2), $x_{i,j}$ are iid random variables with the pmf

$$p = \{ p(a) = \frac{n_a}{q} \ , \ a = 0,1, \ \cdots \ ,\alpha-1 \ \}. \tag{4.1.3}$$

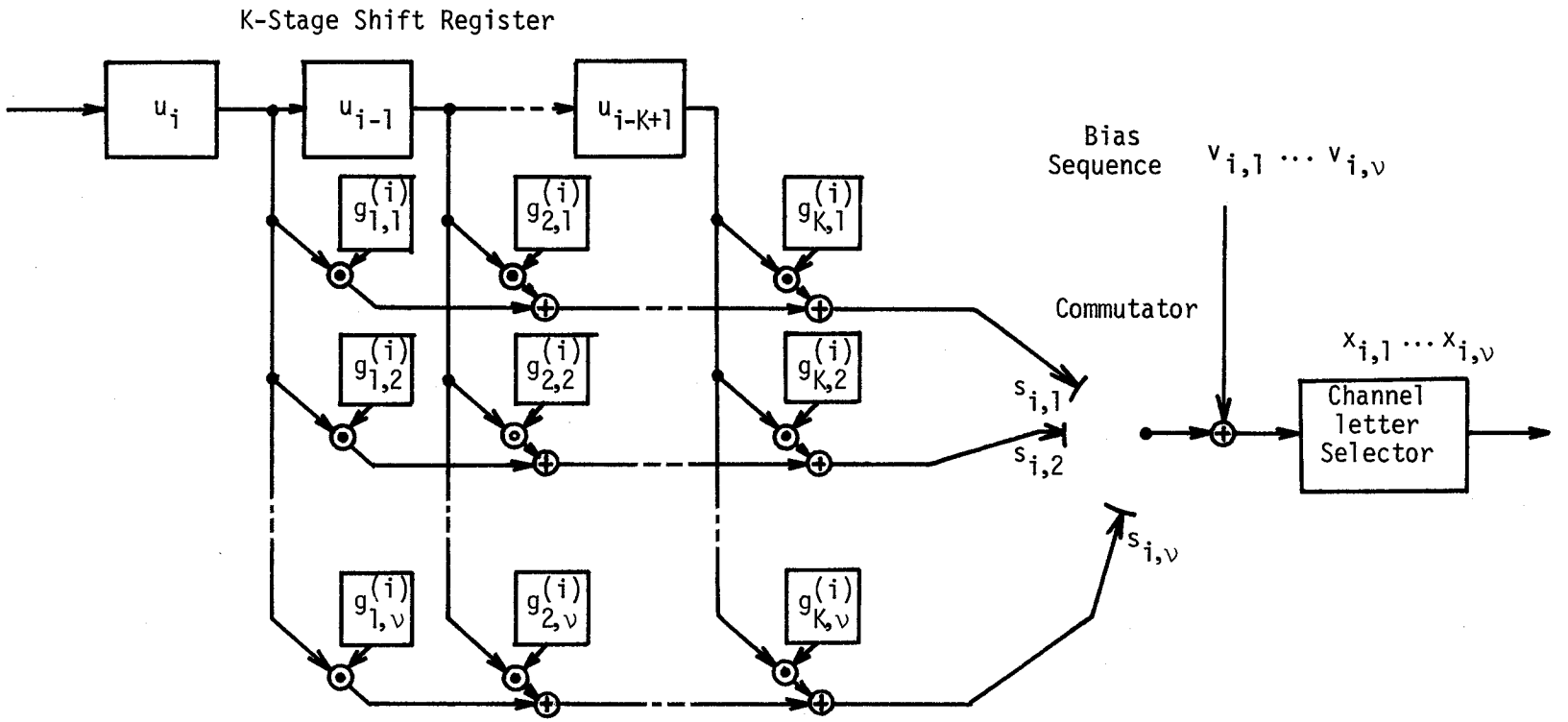


Fig. 4.1.4 — Convolutional Tree Encoder

The totality gives a random convolutional codes. When we refer to (random) convolutional codes later, (4.1.3) is always assumed. For sufficiently large q, any pmf's are approximated with this form arbitrarily well. We summarize several properties of the random convolutional code for the later use (cf. [2]).

Lemma 4.1.1: 1) Succesive letters in a random codeword are iid random variables: 2) If two paths $\underline{u}^{\ell}$ and $\underline{u}'^{\ell}$ differ in every K consecutive symbols, $u_{i+1} \cdots u_{i+K} \neq u'_{i+1} \cdots u'_{i+K}$, $i = 0, \ldots, \ell-K$, then the specified random codeowrds $\underline{X}_1^{\nu} \cdots \underline{X}_{\ell}^{\nu}$ and $\underline{X}_1'^{\nu} \cdots \underline{X}_{\ell}'^{\nu}$ are independent.

The lemma is an easy consequence of GF(q)-arithmetic and the configuration for the encoder.

Unfortunately, since the random convolutional code is generated by multiplication coefficients and a bias sequence selected randomly each time, we can not exclude, from our view, time-varying convolutional codes, codes with varying coefficients time after time, as far as random coding arguments are used. In contrast to time-invariant codes, time-varying ones require linearly increasing memory as block length increases.

As a summary, encoders genrate convolutional tree codes of rate $R = (1/\nu)\log q$ and block length νL, for message sequences $\underline{u}^{L}$. Each codeword is sometimes written as $\underline{x}^{n} = \underline{x}_1^{\nu} \cdots \underline{x}_n^{\nu}$, and corresponding channel outputs are as $\underline{y}^{n} = \underline{y}_1^{\nu} \cdots \underline{y}_n^{\nu}$.

2. Universal performance of convolutional codes

In this section we see the error probability for convolutional codes , and investigate universality of codes over DMC's . From the last purpose, this section is complementary to Section 3.2. First we see an efficient maximum-likelihood decoding algorithm called Viterbi algorithm. Notations in the previous section are maintained.

Viterbi Decoding Algorithm

From fig. 4.1.4, we see that, given a node $\underline{u}^i$, the branch sequence $\underline{x}_i^\nu$ depends only on the K latest message simbols $u_i \ldots u_{i-K+1}$; the K-1 latest history $u_{i-1} \cdots u_{i-K+1}$, called the state of the node $\underline{u}^{i-1}$, and u_i indicating which branch on $\underline{u}^{i-1}$ leads to the node $\underline{u}^i$. Therefore, the convolutional code is completely specified if we know the branch sequences on q branches emanating from nodes having respective states and respective levels. The diagram representing these minimally necessary specifications has a structure like a trellis as shown in Fig. 4.2.1, which is another representation of the example in the previous section. Because of such a trellis-like configuration for codes, convolutional tree codes are often called trellis codes.

Given an channel output $\underline{y} = \underline{y}_1^{\nu} \ldots \underline{y}_L^{\nu}$, we can assign the weight (log-likel

$$\log P(\underline{y}_i^{\nu} | \underline{x}_i^{\nu})$$

to each branch with the branch sequence $\underline{x}_i^{\nu}$. Then, MLD searches over the trellis for the path that maximizes the cumulative weight (up to the L-th level)

$$\sum_{i=1}^{L} \log P(\underline{y}_i^{\nu} | \underline{x}_i^{\nu}).$$

The Viterbi decoding algorithm [10] implements maximum likelihood decoding of trellis codes. It is described as follows (see Fig. 4.2.1): 1) At the first step, the decoder searches all q^K paths $\underline{u}^K$ and, for each $\underline{a}^{K-1} \in A^{K-1}$, retains the path,

$$\underline{u}^K(\underline{a}^{K-1}) = * \, a_1 \ldots a_{K-1},$$

that has the maximum weight among all $\underline{u}^K$ with the state $\underline{a}^{K-1}$: 2) At the i-th step, the q^{K-1} previously retained paths $\underline{u}^{K+i-2}(\underline{a}^{K-1})$, called the survivors, are extended one branch to give q^K candidates, and, for each $\underline{a}^{K-1} \in A^{K-1}$, the decoder selects a new survivor

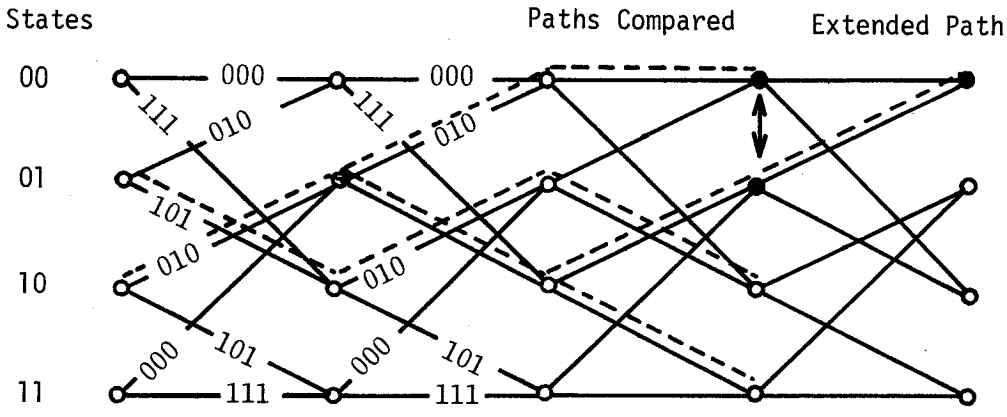


Fig. 4.2.1 — Trellis Diagram for the Tree Code in Fig. 4.1.1 and the Viterbi Algorithm

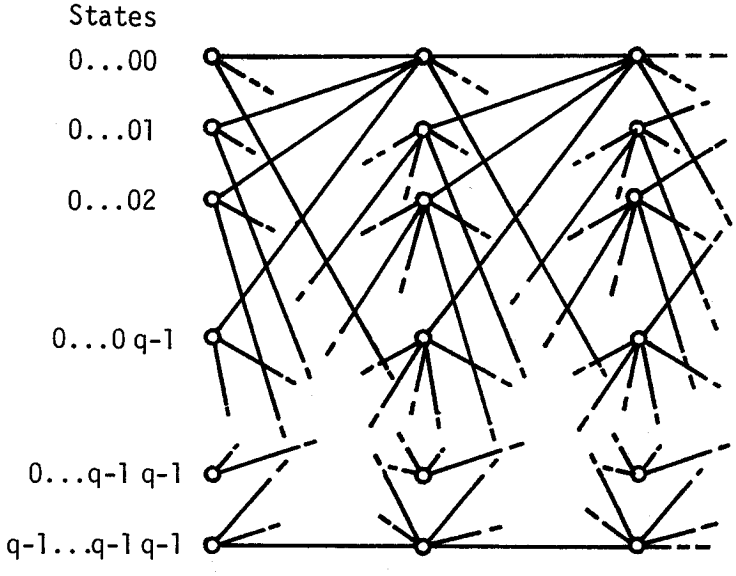


Fig. 4.2.2 — Trellis Diagram for q-nary Codes

$$\underline{u}^{K+i-1}(\underline{a}^{K-1}) = * \ \ldots \ * \ a_1 \ \ldots \ a_{K-1}$$

that has the maximum weight among q candidates with the state $\underline{a}^{K-1}$: 3) When the topmost nodes of the trellis are reached, the decoder selects the path $\underline{u}^{*L}$ that has the largest weight between the latest survivors, and emits it as the decoded message sequence.

Readers with an interest in operations research will soon recognizes that the procedure is not anything else, but is just a version of algorithms for the shortest path problem, as noted in [11] and [12].

From the trellis structure, it is easy to see that K branches are sufficient for a transition from a branch to any branch stemming from a node of any state. This suggests that the effective block length of the codes is νK, just the constraint length times ν. In fact, modifying the message format by additional K-1 consecutive 0's as

$$\underline{u}^{L+K-1} = u_1 \ \ldots \ u_L \overbrace{0 \ \ldots \ 0}^{K-1}, \qquad (4.2.1)$$

Viterbi [10] shows the following theorem.

Theorem 4.2.1: For any $0 \leq \rho \leq 1$ satisfying

$E_o(\rho,p,P) - \rho R > 0$, there exists a convolutional tree code c such that the Viterbi decoder yields

$$P_e(c) \leq \frac{L(q-1)}{1-e^{-\varepsilon\nu}} \exp\{ -\nu K E_o(\rho,p,P) \}$$

where $R = (1/\nu)\log q$ and $\varepsilon = E_o(\rho,p,P) - \rho R$.

The proof is done by estimationg each probability that the message sequence is purged from comparison at a step, and it is clearly visible in the proof of the next theorem.

Universality of Good Convolutional Codes

First suppose that the transmitted message is an all-zero sequence $\underline{0}^{L+K-1} = 0 \ldots 0$. Then, a decoding error occurs if, at some step, say the (j+1)th step, the comparison between paths

$$\begin{array}{cccccc} & & & j+1) & & j+K) \\ 0 & \ldots & 0 & 0 & 0 \ldots & 0 \\ * & \ldots & * & 1 & 0 \ldots & 0 \\ & \ldots & & \ldots & \ldots & \\ * & \ldots & * & q-1 & 0 \ldots & 0 \end{array}$$

discards the path $\underline{0}^{j+K}$. This occurs only if a path of the form

$$\underline{u}^{j+K}(\ell) = 0 \ldots 0 \underset{j+1-\ell)}{\text{nonzero}} * \ldots * \underset{j+1)}{\text{nonzero}} 0 \ldots 0 \underset{j+K)}{} ,$$

for some $\ell = 0,1, \ldots j$, has a weight larger than that of $\underline{0}^{j+K}$. The number of such potential adversaries is $(q-1)^2 q^{\ell-1}$ [or $q-1$ if $\ell = 0$].

Each $\underline{u}^{j+K}(\ell)$ specifies a codeword having the same first $j-\ell$ branch sequences as the one for $\underline{0}^{j+K}$. We denote, by $\underline{x}^*\underline{x}^{\ell+K}$ and $\underline{x}^*\underline{x}^{\ell+K}(\ell)$, the codewords that correspond to $\underline{0}^{j+K}$ and $\underline{u}^{j+K}(\ell)$ respectively. Then the probability that $\underline{0}^{j+K}$ is eliminated at the $(j+1)$th step, write $P_e(\underline{0}^{j+K})$, is

$$P_e(\underline{0}^{j+K}) \leq \sum_{\ell=0}^{j} P_e(\underline{0}^{j+K}, \ell)$$

where

$$P_e(\underline{0}^{j+K}, \ell)$$

$$\triangleq \sum_{\underline{y} \in B^{\nu(\ell+K)}} P(\underline{y}|\underline{x}^{\ell+K}) \; \chi[\, P(\underline{y}|\underline{x}^{\ell+K}) \geq P(\underline{y}|\underline{x}^{\ell+K}(\ell)) \;,$$

$$\text{some } \underline{u}^{j+K}(\ell) \,].$$

The same argument equally applies to all message sequences $\underline{u}^{j+K}$. Let

$$P_e(j+1,\ell) = \frac{1}{q^{j+K}} \sum_{\text{all } \underline{u}^{j+K}} P_e(\underline{u}^{j+K}, \ell) \;.$$

Then, $P_e(\underline{u}^{j+K}, \ell)$ has the form very similar to the

probability of decoding error in Section 2.2 if we let $N = \nu(\ell+K)$. The similarity is strengthened by the following lemma:

Lemma 4.2.1: Any two survivors at each step of Viterbi algorithm are different in any K consecutive symbols, and the corresponding random codewords in the random convolutional code are independent each other.

The first assertion is verified by a reflection on the algorithm, and the second is a consequence of the former and Lemma 4.1.1.

From the lemma, we see (cf. [2]) that, under the operation of expectation relative to the random code,

$$\mathcal{E}P_e(\ell) \triangleq \mathcal{E}P_e(j+1,\ell)$$

$$\leq \exp\{ -\nu(\ell+K)[E_o(\rho,p,P) - \rho R_\ell] \} \; ; \; 0 \leq \rho \leq 1, \tag{4.2.2a}$$

where

$$R_\ell = \begin{cases} \frac{1}{\nu(K+\ell)} \log[(q-1)^2 q^{\ell-1}] & ; \ell > 0, \\ \frac{1}{\nu K} \log (q-1) & ; \ell = 0. \end{cases} \tag{4.2.2b}$$

For any $\varepsilon > 0$, let S and $S(\varepsilon)$ be the set given in Section 3.2 for alphabets A and B. Then, the argument

therin yields the probability

$$\Pr\{\, Q_e(j+1,\ell) \leq \gamma^{-1}\varepsilon^{-2\alpha\beta}\, \boldsymbol{\mathcal{E}} Q_e(\ell),\ \text{all } Q \in S(\varepsilon) \,\} \geq 1 - \gamma$$

for any $\gamma > 0$, where $Q_e(j+1,\ell)$ and $\boldsymbol{\mathcal{E}} Q_e(\varepsilon)$ symbolize the channel $Q \in S(\varepsilon)$. Thus, if we put $\gamma = 1/L(L+1)$, we obtain a convolutional code c such that

$$Q_e(j+1,\ell) \leq L(L+1)\varepsilon^{-2\alpha\beta}\, \boldsymbol{\mathcal{E}} Q_e(\ell) \ ;\ \text{all } Q \in S(\varepsilon);$$

$$\text{all } \ell = 0, \ldots, j;$$

$$\text{all } j = 0, \ldots, L-1. \quad (4.2.3)$$

From (4.2.2), (4.2.3), and Lemma 3.2.1, we know that the convolutional code gives, for each $P \in S$, the average probability of error (cf. [10])

$$P_e(c) \leq \sum_{j=0}^{L-1} \sum_{\ell=0}^{j} P_e(j+1,\ell) \leq \sum_{j=0}^{L-1} \sum_{\ell=0}^{j} Q_e(j+1,\ell)\, e^{\varepsilon(K+\ell)}$$

$$\leq \sum_{j=0}^{L-1} \sum_{\ell=0}^{j} L(L+1)\varepsilon^{-2\alpha\beta}$$

$$\times \exp\{\, -\nu(K+\ell)[\, E_o(\rho,p,Q) - \rho R - \varepsilon/\nu \,] \,\}$$

$$\leq \frac{(L+1)^3(q-1)\varepsilon^{-2\alpha\beta}}{1 - e^{-\delta\nu}} \exp\{\, -\nu K[\, E_o(\rho,p,P) - \varepsilon^* \,] \,\}$$

provided that

$$\delta = E_o(p,p,P) - \rho R - \varepsilon^* > 0, \text{ and} \qquad (4.2.4a)$$

$$\varepsilon^* = \varepsilon(1/\nu + 2\beta^3), \qquad (4.2.4b)$$

where $Q \in S(\varepsilon)$ is the channel given in Lemma 3.2.1. Thus we have proved the following result.

Theorem 4.2.2: For any $\varepsilon^* > 0$, there exists a convolutional code c of constraint length K and rate $R = (1/\nu)\log q$ such that the Viterbi decoder yields

$$P_e \leq \frac{(L+1)^3(q-1)\varepsilon^{-2\alpha\beta}}{1 - e^{-\delta\nu}} \exp\{ -\nu K[E_o(\rho,p,P) - \varepsilon^*] \}$$

for every $P \in S$ and all $0 \leq \rho \leq 1$ satisfying (4.2.4) where ε is given by (4.2.4b).

From this theorem we see that the exponent $E_o(\rho,p,P)$ is attained universally, if maximum likelihood decoding is used. For each rate R, the exponent can be optimized, and an asymptotic form is illustrated in Fig. 4.2.3. We note that the actual block length is νL, while the effective block length is νK. In the figure, we can see that the reliability exponent for convolutional codes is much greater than the reliability exponent for block codes with the same effective block length νK.

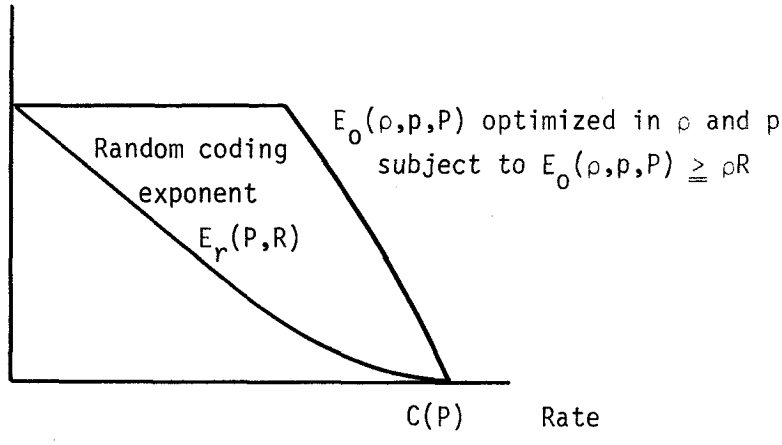


Fig. 4.2.3 — Comparison of Exponents for Block Codes and Convolutional Codes

Why tree codes ?

As we have seen, properly constructed convolutional codes have large reliability exponents if the effective block length νK is identified with the block length of ordinary block codes. We can see that such a comparison is completely reasonable; both codes with the same effective block length require approximately the same computations in decoding.

Consider two channel coding systems, one with a block code $c^{\nu K}$ with q^K block codewords and the other with a convolutional code $c^{\nu L}$ of rate $(1/\nu)\log q$ having constraint length K (cf. Fig. 2.1.3). The block decoding with $c^{\nu K}$ is performed, for each channel output $\underline{y}^{\nu K}$, by the combination of parallel weight enumeration and heirarchical parallel comparison as depicted in Fig. 4.2.4 (a). Using this parallel processing, the decoder should possess computational speed that enable one weight enumeration and K weight comparisons for a codeword. On the other hand, decoding for the convolutional code, using the Viterbi algorithm, will comprise of alternating weight enumeration for branch sequences and path comparison as shown in Fig. 4.2.4 (b). From the inspection of both schemes, we can see that they require almost the same computational loads and speed. High reliability with realizable

computational requirment makes the convolutional code with Viterbi decoding practically significant. Indeed, Viterbi decoders implemented by hardware are sold for the use in practice.

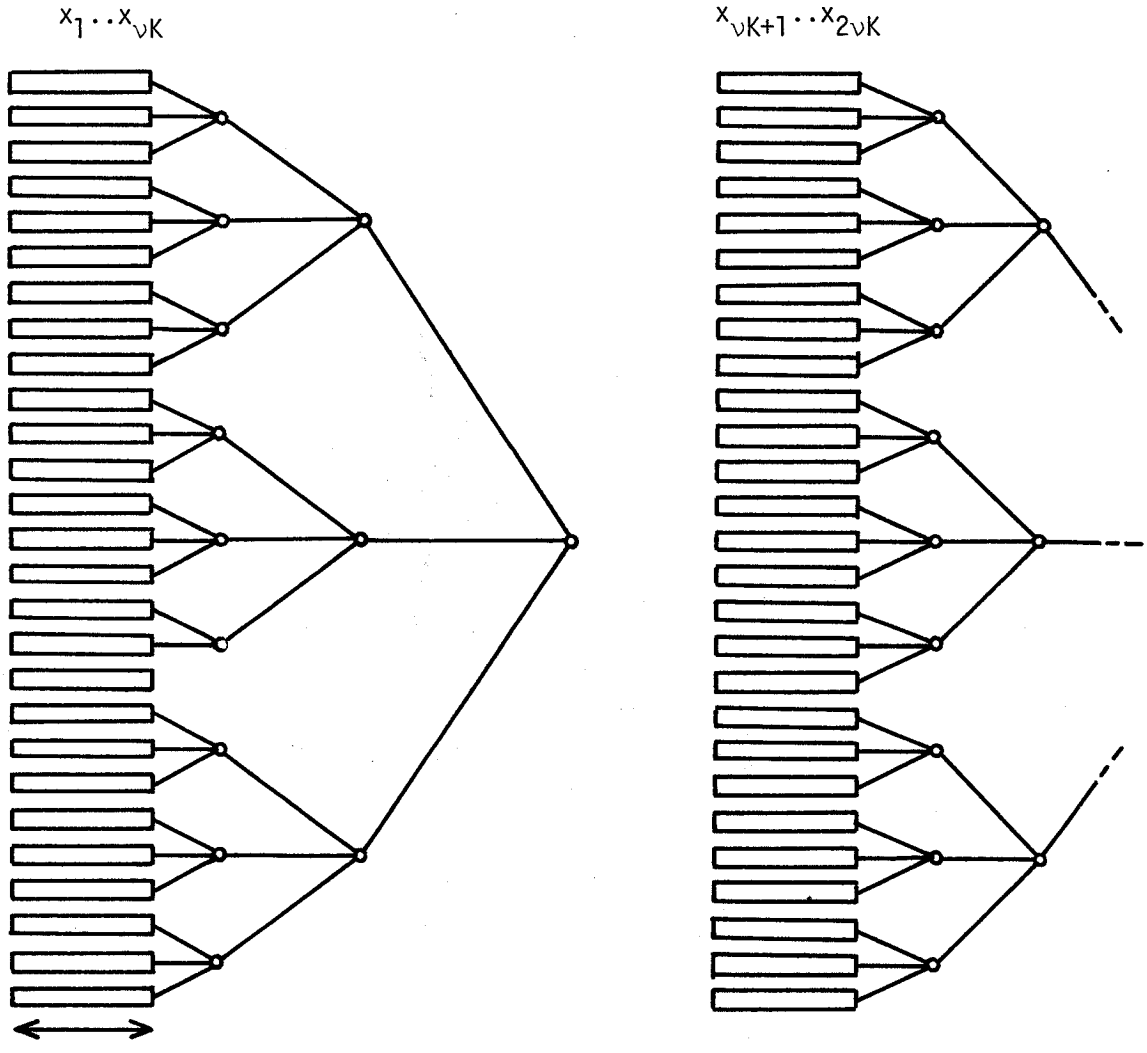


Fig. 4.2.4(a) — Computations in the block decoder
(k = q =3)

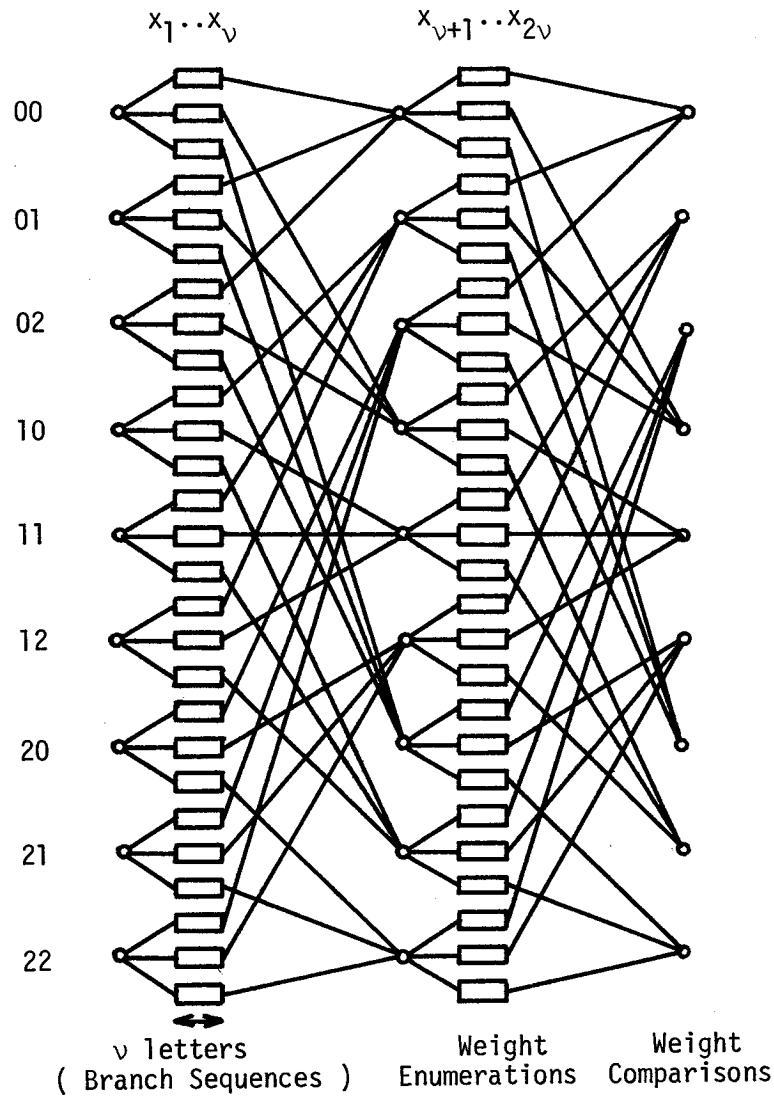


Fig. 4.2.4 (b) — Computations in the Viterbi Decoder
(K = q = 3)

3. Sequential decoding of convolution codes: Computational moments

In the previous section, we have seen the good performance of convolutional codes. However, as the constraint length increases, computational burden in Viterbi decoding becomes great. This is quite discouraging if we need an extremely small probability of error. Sequential decoding is a substitute for Viterbi decoding under such a requirement [13], although the former has a rather old origin; it is invented by Wozencraft [14] and is almost completed by Fano [15].

For a DMC P with input and output alphabets A = { 0, ... ,α-1 } and B = { 0, ... ,β-1 } respectively and a pmf p on A satisfying (4.1.3), let q be the pmf on B given by

$$q(b) = \sum_{a \in A} p(a)P(b|a)$$

for all b ε B. Once the channel out put $\underline{y} = \underline{y}_1^{\nu}\underline{y}_2^{\nu} \cdots$ is accepted, we can assign to each node (or path) $\underline{u}^i$ a weight by the function

$$\Gamma(\underline{u}^i) = \sum_{j=1}^{i} \left[\log \frac{P(\underline{y}_j|\underline{x}_j)}{q(\underline{y}_j)} - \nu R \right] \qquad (4.3.1)$$

where $R = (1/\nu)\log q$ is the rate of codes and $\underline{x}^i = \underline{x}_1^{\nu} \cdots \underline{x}_i^{\nu}$ is the codeword specified by $\underline{u}^i$. Maximum

likelihood decoding suggests exhaustive searching for the node with the largest weight as Viterbi decoding, which, however, is sometimes costly. Instead, sequential decoding algorithms search nodes selectively.

We use the modified version by Gallager [2]. Chiefly it consists of three moves on nodes: forward, lateral, and backward moves (see Fig. 4.3.1). A forward move on a node is a move to the immediate descendant that is numbered 0. A lateral move on a node is a move to the next neighbouring node between q nodes having the same antecedent. And, a backward move on a node is a move to the immediate antecedent of the node. Shift on a node is controled by comparison of weights and a threshold value T which is renewed after each shift. The precise rule is described in Table 4.3.1, where Δ is the size of threshold increment-decrement. If $I(p,P) > R$, then the low of large numbers implies that $\Gamma(\underline{u}^i)$ tends to increase on the path specified by the message. On the other hand, $\Gamma(\underline{u}^i)$ would lastly fall below zero on the other paths.

An example is depicted in Fig. 4.3.2, where we suppose 1000 to be the message. As the example shows, the decoder output has no synchronization with the channel input; the message blocks are sometimes decoded quickly, and other times are not decoded even when the next message bloks are to be treated. Therefore,

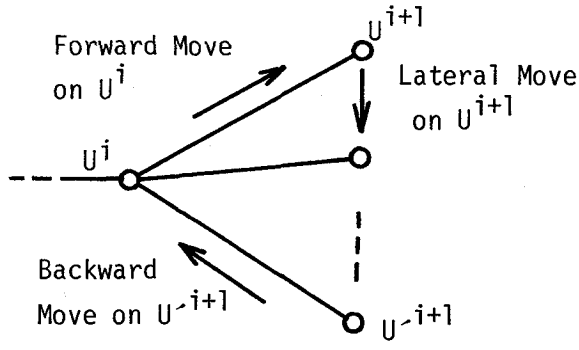


Fig. 4.3.1 — Three Basic Moves on Nodes

Condition on Nodes			Action to be Taken	
Previous Move	Comparison of $\Gamma(\underline{u}^{i-1})$ and $\Gamma(\underline{u}^{i})$ with Initial Threshold		Final Treshold	Move
F or L	$\Gamma(\underline{u}^{i-1}) < T+\Delta$	$\Gamma(\underline{u}^{i}) \geq T$	Raise*	F
F or L	$\Gamma(\underline{u}^{i-1}) \geq T+\Delta$	$\Gamma(\underline{u}^{i}) < T$	No Change	F
F or L	any $\Gamma(\underline{u}^{i-1})$	$\Gamma(\underline{u}^{i}) < T$	No Change	L
B	$\Gamma(\underline{u}^{i-1}) < T$	any $\Gamma(\underline{u}^{i})$	Lower by	F
B	$\Gamma(\underline{u}^{i-1}) \geq T$	any $\Gamma(\underline{u}^{i})$	No Change	L

* Add $j\Delta$ to threshold, where j is chosen to satisfy $\Gamma(\underline{u}^{i})-\Delta < T+j\Delta \leq \Gamma(\underline{u}^{i})$.

Table 4.3.1 — Sequential Decoding Algorithm

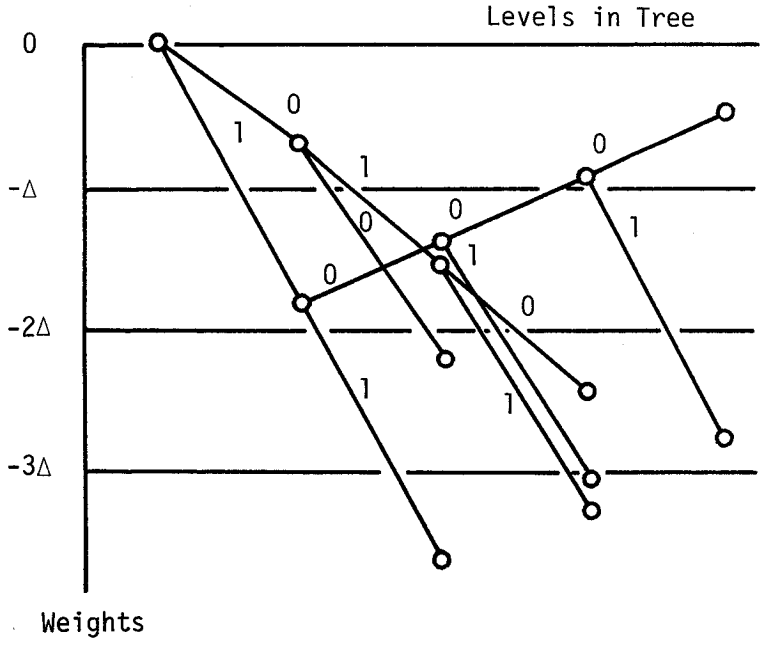


Fig. 4.3.2 (a) — Sequential Decoding: Weights

Path	Threshold	Move	Path	Threshold	Move
—	0	F	0	-2Δ	F
0	0	L	00	-2Δ	L
1	0	B	01	-2Δ	F
—	-Δ	F	010	-2Δ	L
0	-Δ	F	011	-2Δ	B
00	-Δ	L	01	-2Δ	B
01	-Δ	B	0	-2Δ	L
0	-Δ	L	1	-2Δ	F
1	-Δ	B	10	-2Δ	F
—	-2Δ	F	100	-Δ	F

Fig. 4.3.2 (b) — Sequential Decoding: Decoder Movements

the sequential decoder must be accommodated with a buffer to smooth these occasional delay.

One of the problems, other than such a computational one, is the capability of reliable encoding and decoding. If we suppose the message format (4.2.1), we have a theorem from the arguments due to Jelinek [16].

Theorem 4.3.1: For $\rho > 0$ satisfying $E_o(\rho,p,P) - \rho R > 0$, there exists a convolutional code c such that the sequential decoder yields

$$P_e \leq \frac{Le^{\Delta/(1+\rho)}}{(1-e^{-\Delta/(1+\rho)})(1-e^{-\varepsilon\nu})(1-e^{-\varepsilon\nu\rho})} \exp\{-\nu(K-1)[E_o(\rho,p,P)-\varepsilon]\}$$

where $\varepsilon = [\, E_o(\rho,p,) - \rho R \,]/(1+\rho)$.

From Theorem 4.3.1, we see that the sequential algorithm implemented independently of the constraint length K, shows approximately the same performance as the Viterbi algorithm which is heavily dependent on K. However, the advantage is sometimes lost; a long burst of severe channel noise forces the decoder to stray into wrong paths through the code tree while channel output letters continuously accumulate in the buffer and overflow. Hence, assessment of such failure is indispensable in sequential decoding.

We assume that the sent message is an all zero

sequence. We say that an F-hypothesis occurs on a node if the sequential decoder searches the node and makes a forward move on it. We denote, by W_i, the number of F-hypotheses which occurred to decode the (i+1)th message symbol correctly. (This rather vague definition will be made rigid later.)

Now consider a node $\underline{u}^i$ and its immediate descendents $\underline{u}^{i+1}(j)$ in Fig. 4.3.1. For lateral moves on $\underline{u}^{i+1}(j)$ to be made, the node $\underline{u}^i$ must be first F-hypothesized, and a backward move occurs only at the last descedent $\underline{u}^{i+1}(q-1)$. Therefore, for W F-hypotheses, there are made at most (q+1)W basic moves in the code tree. Suppose that the decoder is capable of σ basic moves while ν channel letters come. Then, if a buffer overflow occurs, we necessarily have $(q+1)\sum_{i=1}^{T} W_i > \sigma T$ for the buffer capable of storing νT letters, and the buffer-over flow probability will be

$$\Pr\{ (q+1) \sum_{i=1}^{T} W_i > \sigma T \} .$$

It is known [2],[16] that this probability is intimately related with boundedness of the ρ-th moments of W_i, EW^{ρ}.

The Number of F-Hypotheses

For a transmitted message sequence, say, $\underline{0} = 00\ldots$, we say that a path, or a node, $\underline{u}^i$ is correct if $\underline{u}^i = 0 \ldots 0$, and, on the other hand, say that $\underline{u}^i$ is incorrect if $\underline{u}^i \neq 0 \ldots 0$. Let D_i be the set of nodes $\underline{u}^j = 0 \ldots 0\, u_{i+1} \ldots u_j$, $j \geq i$ and $u_{i+1} \neq 0$, and call it the i-th set of incorrect nodes. It also contains the i-th correct node $\underline{0}^i$ (see Fig. 4.3.3). We define W_i, the number of F-hypotheses to decode the i-th correct node, as the number of all F-hypotheses occurred in D_i. The following lemma shows a necessary condition for a node in D_i to be F-hypothesized (cf. [2, p.275]).

Lemma 4.3.1: A node $\underline{u}^j$ in D_i is F-hypothesized for the h-th time only if

$$\Gamma(\underline{u}^j) \geq \Gamma_{\min,i} + (h - 2)\Delta$$

where

$$\Gamma_{\min,i} \triangleq \min_{m=i,\ldots,L} \Gamma_m .$$

Now let D_i^j be the set of all nodes in D_i and at the j-th level, and let $\rho > 0$. From the lemma, we have the following bounds on W_i:

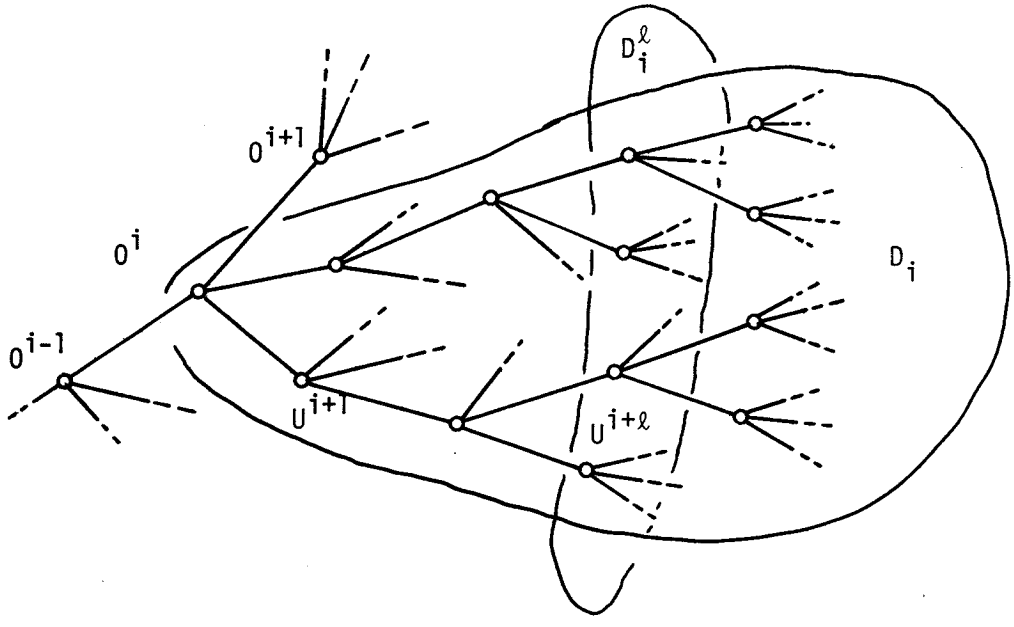


Fig. 4.3.3 — The Sets of Incorrect Nodes

$$W_i \leq \sum_{h=1}^{\infty} \sum_{\ell=0}^{L-i} \sum_{\underline{u}\in D_i^{\ell}} \chi[\, \Gamma(\underline{u}) \geq \Gamma_{\min,i} + (h-2)\Delta \,]$$

$$\leq \sum_{h=1}^{\infty} \sum_{\ell=0}^{L-i} \sum_{m=i}^{L} \sum_{\underline{u}\in D_i^{\ell}} \chi[\, \Gamma(\underline{u}) \geq \Gamma_m + (h-2)\Delta \,]$$

$$\leq \sum_{h=1}^{\infty} \sum_{\ell=0}^{L-i} \sum_{m=i}^{L} \sum_{\underline{u}\in D_i^{\ell}} \exp\{ \frac{1}{1+\rho}[\, \Gamma(\underline{u}) - \Gamma_m - (h-2)\Delta \,] \} ;$$

$$\rho > 0. \qquad (4.3.2)$$

To make the arguments straightforward and simple, we deal only with W_0, and use the abbreviations $\underline{x}_i = \underline{x}_1^{\nu} \ldots \underline{x}_i^{\nu}$ and $\underline{y}_i = \underline{y}_1^{\nu} \ldots \underline{y}_i^{\nu}$. Under this convention,

$$W_0 \leq \gamma_0 \sum_{\ell=0}^{L} \sum_{m=0}^{L} \sum_{\underline{u}\in D_0} \left\{ \frac{P(\underline{y}_\ell|\underline{x}_\ell) q(\underline{y}_m)}{q(\underline{y}_\ell) P(\underline{y}_m|\underline{x}_m)} \right\}^{\frac{1}{1+\rho}} \exp\{ \frac{\nu}{1+\rho}(m-\ell)R \} ;$$

$$\rho > 0, \quad (4.3.3)$$

where $\underline{x}_\ell$ and $\tilde{\underline{x}}_m$ mean the codewords specified by $\underline{u}^{\ell} \in D_0^{\ell}$ and the m-th correct node $\underline{0}^m$ respectively, and

$$\gamma_0 = \frac{e^{2\Delta/(1+\rho)}}{e^{\Delta/(1+\rho)} - 1} .$$

Henceforce we use distinct notations for expectation

operators with respect to the random code, use $\mathcal{E}$, and with respect to the channel and any particular code, use E. The operator $\mathcal{E}$ is a product of two expectation operators $\mathcal{E}_C$ and $\mathcal{E}_I$; The former is for the random codewords specified by correct nodes and the latter is for the ones specified by incorrect nodes (cf. [2] and [16]).

Suppose that convolutional codes have infinite constraint length. According to random coding arguments, our task is to bound the average ρ-th moments $\mathcal{E}EW^{\rho}$. Remember that R is the rate of codes and that p is given by (4.1.3). Falconer [17] shows that $\mathcal{E}EW^{\rho} < \infty$ if $\rho R < E_0(\rho,p,P)$ and $0 \leq \rho \leq 1$. For all poitive integers ρ, Savage [18] and Zigangirov [19] show that $\mathcal{E}EW^{\rho} < \infty$ if $\rho R < E_0(\rho,p,P)$ for tree codes possibly without algebraic structures. And their result is extended to all positive ρ by Jelinek [20]. Though several simulation results suggest that these should be also true for convolutional codes, however, no proof has been known. The difficulty has its root in the algebraic structure that makes the codes feasible.

For finite constraint length , it is known [21] that, with slight modification in the algorithm, the computational burden is lessened by decreasing constraint length K. Therefore we always suppose infinite constraint length, $K = \infty$.

L–Independence

We say that a set of incorrect nodes $\{ \underline{u}^i(1), \ldots, \underline{u}^i(n) \}$ has rank k if, there exist maximally k nodes $\underline{u}^i(n_1), \cdots, \underline{u}^i(n_k)$ in the set such that the all nodes are expressed by linear combination as

$$\underline{u}^i(j) = \sum_{\ell=1}^{k} \alpha_\ell \, \underline{u}^i(n_\ell), \qquad \text{for } \alpha_\ell \in GF(q),$$

where all $\underline{u}^i(j)$ are interpreted as i-vectors over GF(q). The algebraic dependence of nodes implies another structure between corresponding random codewords :

Lemma 4.3.2: If a set of nodes $\{ \underline{u}^i(1), \ldots, \underline{u}^i(n) \}$ has rank k, then, between the corresponding i-th random branch sequences in the random convolutional code, $\underline{X}_i^\nu(1), \ldots, \underline{X}_i^\nu(n)$, there exist k mutually independent $\underline{X}_i^\nu(n_1), \ldots, \underline{X}_i^\nu(n_k)$.

Corollary: If $\{ \underline{u}^i(1), \ldots, \underline{u}^i(n) \}$ has rank k, then there exist n´ subsets with the following properties:

1) Each subset consists of k mutually independent i-th random branch sequences $\underline{X}_i^\nu(n_{j,1}), \ldots, \underline{X}_i^\nu(n_{j,k})$, for each $j = 1, \ldots, n'$:

2) Their union is $\{ \underline{X}_i^\nu(1), \ldots, \underline{X}_i^\nu(n) \}$, the set of all corresponding random branch sequences:

3) $n' \leq n$.

The proof of Lemma 4.3.2 is given later, but the proof of Corollary is relatively easy and is omitted.

As an example, consider nodes (1110),(1011),(1101), and (1000), for q = 2. Their rank is 3, and

$$\{ \underline{X}_4(1110), \underline{X}_4(1011), \underline{X}_4(1101) \},$$

$$\{ \underline{X}_4(1110), \underline{X}_4(1011), \underline{X}_4(1000) \},$$

$$\{ \underline{X}_4(1110), \underline{X}_4(1101), \underline{X}_4(1000) \}$$

give the subsets assured by Corollary.

Given a set of incorrect nodes, $U = \{ \underline{u}^L(1) \ \ldots \ \underline{u}^L(n) \} \subset D_0^L$, we write the antecedens at the ℓ-th level of $\underline{u}^L(j)$ as $\underline{u}^\ell(j)$. Let $\underline{L} = (L_1, \ \ldots \ ,L_n)$ be any vector with integral components. We say that U is $\underline{L}$-independent if $\{ \underline{u}^\ell(1), \ \ldots \ ,\underline{u}^\ell(n) \}$ has rank k wnenever $L_{k-1} < \ell \leq L_k$, where $L_0 = 0$. Nodes in the above example are (1,2,4,4)-independent. The following lemma gives an upper bound on the number of $\underline{L}$-independent sets [66]:

Lemma 4.3.3: For integers $L_0 (= 0) \leq L_1 \leq \ \ldots \ \leq L_n$, the number $M(\underline{L})$ of distinct $\underline{L}$-independent sets $\{ \underline{u}^L(1), \ \ldots \ ,\underline{u}^L(n) \}$ is bounded by

$$M(\underline{L}) \leq \exp\{ \sum_{k=1}^{n} k(L_k - L_{k-1})\log q + \frac{n^2(n+1)}{2} \log q \}.$$

Bounding $\mathcal{E}EW_0^\rho$

Applying Minkovsky's inequality to the bound (4.3.3), we have a bound on the ρ-th moment.

$$
\begin{aligned}
&\left(\mathcal{E}EW_0^\rho \right)^{1/\rho} \\
&\leq \gamma_0 \sum_{\ell=0}^{L} \sum_{m=0}^{L} \left\{ \mathcal{E}_C E \left(\frac{q(\underline{Y}_m)}{P(\underline{Y}_m | \tilde{\underline{X}}_m)} \right)^{\frac{\rho}{1+\rho}} \mathcal{E}_I \left[\sum_{\underline{u} \in D_0^\ell} \frac{P(\underline{Y}_\ell | \underline{X}_\ell)}{q(\underline{Y}_\ell)} \right]^{\frac{\rho}{1+\rho}} \right] \\
&\qquad \times \exp\left[\frac{\nu\rho}{1+\rho} (m - \ell) R \right] \Bigg\}^{\frac{1}{\rho}} \qquad (4.3.4)
\end{aligned}
$$

First note the identity

$$
\begin{aligned}
&\mathcal{E}_C E \left(\frac{q(\underline{Y}_m)}{P(\underline{Y}_m | \tilde{\underline{X}}_m)} \right)^{\frac{\rho}{1+\rho}} \mathcal{E}_I \left[* \right]^\rho \\
&= \sum_{\underline{x}_L \in A^{\nu L}} \sum_{\underline{y}_L \in B^{\nu L}} p(\underline{x}_L) P(\underline{y}_L | \underline{x}_L) \left(\frac{q(\underline{y}_m)}{P(\underline{y}_m | \underline{x}_m)} \right)^{\frac{\rho}{1+\rho}} \mathcal{E}_I \left[* \right]^\rho \\
&= \sum_{\underline{y}_L \in A^{\nu L}} \left[\sum_{\underline{x}_m \in A^{\nu m}} p(\underline{x}_m) \left(\frac{P(\underline{y}_m | \underline{x}_m)}{q(\underline{y}_m)} \right)^{\frac{1}{1+\rho}} \right] \mathcal{E}_I \left[* \right]^\rho .
\end{aligned}
$$

Then, from Hölder's inequality, we have

$$
\left(\mathcal{E}EW_0^\rho \right)^{1/\rho} \leq
$$

$$\gamma_0 \sum_{m=0}^{L} \left\{ \sum_{\underline{y}_m \in B^{\nu m}} q(\underline{y}_m) \left[\sum_{\underline{x}_m \in A^{\nu m}} p(\underline{x}_m) \left(\frac{P(\underline{y}_m | \underline{x}_m)}{q(\underline{y}_m)} \right)^{\frac{1}{1+\rho}} \right]^{1+\rho} \right\}^{\frac{1}{\rho(1+\rho)}}$$

$$\times \sum_{\ell=0}^{L} \left\{ \sum_{\underline{y}_\ell \in B^{\nu \ell}} q(\underline{y}_\rho) \left[\mathcal{E}_I \left(\sum_{\underline{u} \in D_0^\ell} \left[\frac{P(\underline{y}_\ell | \underline{X}_\ell)}{q(\underline{y}_\ell)} \right]^{\frac{1}{1+\rho}} \right)^{\rho} \right]^{\frac{1+\rho}{\rho}} \right\}^{\frac{1}{1+\rho}}$$

$$\times \exp\{ \frac{\nu}{1+\rho}(m - \ell)R \} \quad . \qquad (4.3.5)$$

Next we bound the expectation $\mathcal{E}_I(*)$ with respect to incorrect codewords in (4.3.5). Let n be an integer such that $n-1 < \rho \leq n$. From Jensen's inequality, the expectation is

$$\mathcal{E}_I \left(\sum_{\underline{u} \in D_0^\ell} \left[\frac{P(\underline{y}_\ell | \underline{X}_\ell)}{q(\underline{y}_\ell)} \right]^{\frac{1}{1+\rho}} \right)^{\rho}$$

$$\leq \left\{ \mathcal{E}_I \left(\sum_{\underline{u} \in D_0^\ell} \left[\frac{P(\underline{y}_\ell | \underline{X}_\ell)}{q(\underline{y}_\ell)} \right]^{\frac{1}{1+\rho}} \right)^{n} \right\}^{\frac{\rho}{n}}$$

$$= \left\{ \sum_{\substack{\text{all} \\ \underline{u}^\ell(1)\ldots\underline{u}^\ell(n) \\ \text{in } D_0^\ell}} \mathcal{E}_I \prod_{j=1}^{n} \left[\frac{P(\underline{y}_\ell | \underline{X}_\ell(j))}{q(\underline{y}_\ell)} \right]^{\frac{1}{1+\rho}} \right\}^{\frac{\rho}{n}}$$

where each $\underline{X}_\ell(j)$ corresponds to each $\underline{u}^\ell(j)$. The following lemma is crucial in bounding the expectation in the extreme right-hand side .

Lemma 4.3.4: If $\{ \underline{u}^\ell(1), \ldots, \underline{u}^\ell(n) \}$ is $(\ell_1, \ldots, \ell_n)$-independent, then the expectation is

$$\mathcal{E}_I \prod_{j=1}^{n} \left[\frac{P(\underline{y}_\ell | \underline{X}_\ell(j))}{q(\underline{y}_\ell)} \right]^{\frac{1}{1+\rho}}$$

$$\leq \prod_{k=1}^{n} \prod_{i=\ell_{k-1}+1}^{\ell_k} \left(\sum_{\underline{x}^\nu \in A^\nu} p(\underline{x}^\nu) \left[\frac{P(\underline{y}_i^\nu | \underline{x}^\nu)}{q(\underline{y}_i^\nu)} \right]^{\frac{1}{1+\rho}} \right)^{\frac{(1+\rho)k-n}{\rho}},$$

where $\ell_0 = 0$ and $\prod_{i=\ell}^{m}(*) = 1$ if $\ell > m$.

According to the lemma, the bound above continues as follows.

$$\leq \left\{ \sum_{\underline{\ell}} \prod_{k=1}^{n} \prod_{j=\ell_{k-1}+1}^{\ell_k} \right.$$

$$\left. \times \left(\sum_{\underline{x}^\nu \in A^\nu} p(\underline{x}^\nu) \left[\frac{P(\underline{y}_j^\nu | \underline{x}^\nu)}{q(\underline{y}_j^\nu)} \right]^{\frac{1}{1+\rho}} \right)^{\frac{(1+\rho)k-n}{\rho}} \right\}^{\frac{\rho}{n}}$$

where the first summation in the right-hand side is over all n-vectors with positive and nondecreasing integral components less than ℓ, the number of which

is at most $(\ell+1)^n$. Averaging the both sides of the above bounds with respect to all $\underline{y}_\ell$, we obtain

$$\left\{\sum_{\underline{y}_\ell \in B^{\nu\ell}} q(\underline{y}_\ell)\left[\mathcal{E}_I\left(\sum_{\underline{u}\in D_0^\ell}\left[\frac{P(\underline{y}_\ell|\underline{X}_\ell)}{q(\underline{y}_\ell)}\right]^{\frac{1}{1+\rho}}\right)^\rho\right]^{\frac{1+\rho}{\rho}}\right\}^{\frac{\rho}{1+\rho}}$$

$$\leq \left\{\sum_{\underline{y}_\ell \in B^{\nu\ell}} q(\underline{y}_\ell)\left[\sum_{\underline{\ell}} M^n(\underline{\ell}) \prod_{k=1}^{n} \prod_{j=\ell_{k-1}+1}^{\ell_k} \left(*\right)^{\frac{(1+\rho)k-n}{\rho n}}\right]^{\frac{1+\rho}{\rho}}\right\}^{\frac{\rho}{1+\rho}}$$

$$\leq \sum_{\underline{\ell}} M^n(\underline{\ell})\left[\sum_{\underline{y}_\ell \in B^{\nu\ell}} q(\underline{y}_\ell) \prod_{k=1}^{n} \prod_{j=\ell_{k-1}+1}^{\ell_k} \left(*\right)^{\frac{(1+\rho)[(1+\rho)k-n]}{\rho n}}\right]^{\frac{1}{1+\rho}}$$

$$\leq \sum_{\underline{\ell}} M^n(\underline{\ell}) \prod_{k=1}^{n} \prod_{j=\ell_{k-1}+1}^{\ell_k}$$

$$\times\left[\sum_{\underline{y}^\nu \in B^\nu} q(\underline{y}_\ell)\left(\sum_{\underline{x}^\nu \in A^\nu} p(\underline{x}^\nu)\left[\frac{P(\underline{y}^\nu|\underline{x}^\nu)}{q(\underline{y}^\nu)}\right]^{\frac{1}{1+\rho}}\right)^{1+\rho}\right]^{\frac{(1+\rho)k-n}{\rho n}}$$

where the second inequality follows from Minkvsky's inequality and the third from Jensen's inequality using

$$0 < \frac{(1+\rho)k - n}{\rho n} \leq 1 .$$

By substitution of the above bound into (4.3.5) and using Lemma 4.3.3 and the identity

$$\sum_{\underline{y}^\nu \in B^\nu} q(\underline{y}^\nu) \left(\sum_{\underline{x}^\nu \in B^\nu} p(\underline{x}^\nu) \left[\frac{P(\underline{y}^\nu | \underline{x}^\nu)}{q(\underline{y}^\nu)} \right]^{\frac{1}{1+\rho}} \right)^{1+\rho}$$

$$= \exp\{ -\nu E_o(\rho,p,P) \},$$

we have

$$\left(\mathcal{E} EW_0^\rho \right)^{1/\rho}$$

$$\leq \gamma_0 \sum_{m=0}^{L} \exp\{ -\frac{\nu m}{(1+\rho)} [E_o(\rho,p,P) - \rho R] \}$$

$$\times \sum_{\ell=0}^{L} (\ell + 1)^{n/\ell}$$

$$\times \max_{\underline{\ell}} \exp\{ -\frac{1}{\rho} [\frac{\nu}{n} \sum_{k=1}^{n} k(\ell_k - \ell_{k-1}) - \frac{\nu \ell}{1+\rho}]$$

$$\times [E_o(\rho,p,P) - \rho R] + \frac{n(n+1)}{2} \log q \} .$$

Since $\sum_{k=1}^{n} k(\ell_k - \ell_{k-1}) \geq \ell$ for any $\underline{\ell}$, we finally have the bound:

$$\left(\mathcal{E} EW_0^\rho \right)^{1/\rho}$$

$$\leq \gamma_0 \sum_{m=0}^{L} \exp\{ -\frac{\nu m}{(1+\rho)} [E_o(\rho,p,P) - \rho R] \}$$

$$\times \sum_{\ell=0}^{L} (\ell + 1)^{n/\ell} \exp\{ - \frac{\nu\ell}{\rho}(\frac{1}{n} - \frac{1}{1+\rho})[E_o(\rho,p,P) - \rho R] \}$$

$$\times \exp\{ \frac{n(n+1)}{2} \log q \} \quad .$$

From $n-1 < \rho \leq n$, all summations converge as $L \to \infty$ if $\rho R < E_o(\rho,p,P)$. Therefore we have a theorem.

Theorem 4.3.2: For any $\rho > 0$ satisfying $E_o(\rho,p,P) > \rho R$, the average of the ρ-th moment EW_0^ρ with respect to the random convolutional code ($K = \infty$), $\mathcal{E} EW_0^\rho$, is bounded by a finite constant which is independent of the message length L, where $R = (1/\nu)\log q$.

Corollary 1: For any $\rho > 0$ satisfying $E_o(\rho,p,P) > \rho R$, and any L, there exists a convolutional code of rate $R = (1/\nu)\log q$ with block length L such that the ρ-th moment EW_0^ρ is bounded by a finite constant which is independent of L.

Corollary 2: Under the same condition as Corollary 1, there exists a convolutional code of rate $R = (1/\nu)\log q$ with block length L that has the probability distribution

$$\Pr\{ W_0 \geq w \} \leq \frac{\overline{W}(\rho)}{w^\rho}$$

for $w \geq 0$, where $\overline{W}(\rho)$ is independent of L.

We note that all the above arguments equally apply to W_i and that these statements also hold for each W_i.

A probability distribution $F(w) = \text{const.} \times w^{-\gamma}$, $w > 0$, is called a Pareto distribution, and, from the upper bound in Corollary 2, we may think that W_0 has the same tail probabilities as a Pareto distribution. In fact, this is generally true: we see in the next section that the tail probabilities of W_0 are also bounded below by a Pareto distribution. Historically, several simulation data (cf. [22]) have predicted such a observation, which is now analytically proved for convolutional codes. In Fig. 4.3.4, we see an example of computer simulation by Jordan*[22] over BSC's with crossover probabilities ε. The binary convolutional code used in this simulation has the finite constrain length K = 60. Since Pareto distribution F(w) decreases algebraically as $w \to \infty$, an extreme number of incorrect F-hypotheses are likely to happen and tend to accumulate in the decoder. We can see an illustrative example in the same literature, which is reproduced in Fig. 4.3.5. The position in the tree indicates the highest level that the decoder has ever reached, and the waiting line indicates the number of data that stored in the memory and waiting for decoding. The computational critical rate R_{comp} and

---

* By permission of K.L. Jordan, Jr.

computational speed σ are defined in the next section. Intuitively, R_{comp} corresponds to the extent of noise such that $\Pr\{ W_i \geq w \} \sim w^{-1}$ and the waiting data accumulate indefinitely (since $\int^{\delta} \Pr\{ W_i \geq w \} dw \to \infty$ as $\delta \to \infty$). In the next section we study the effect of such accumulation more thoroughly.

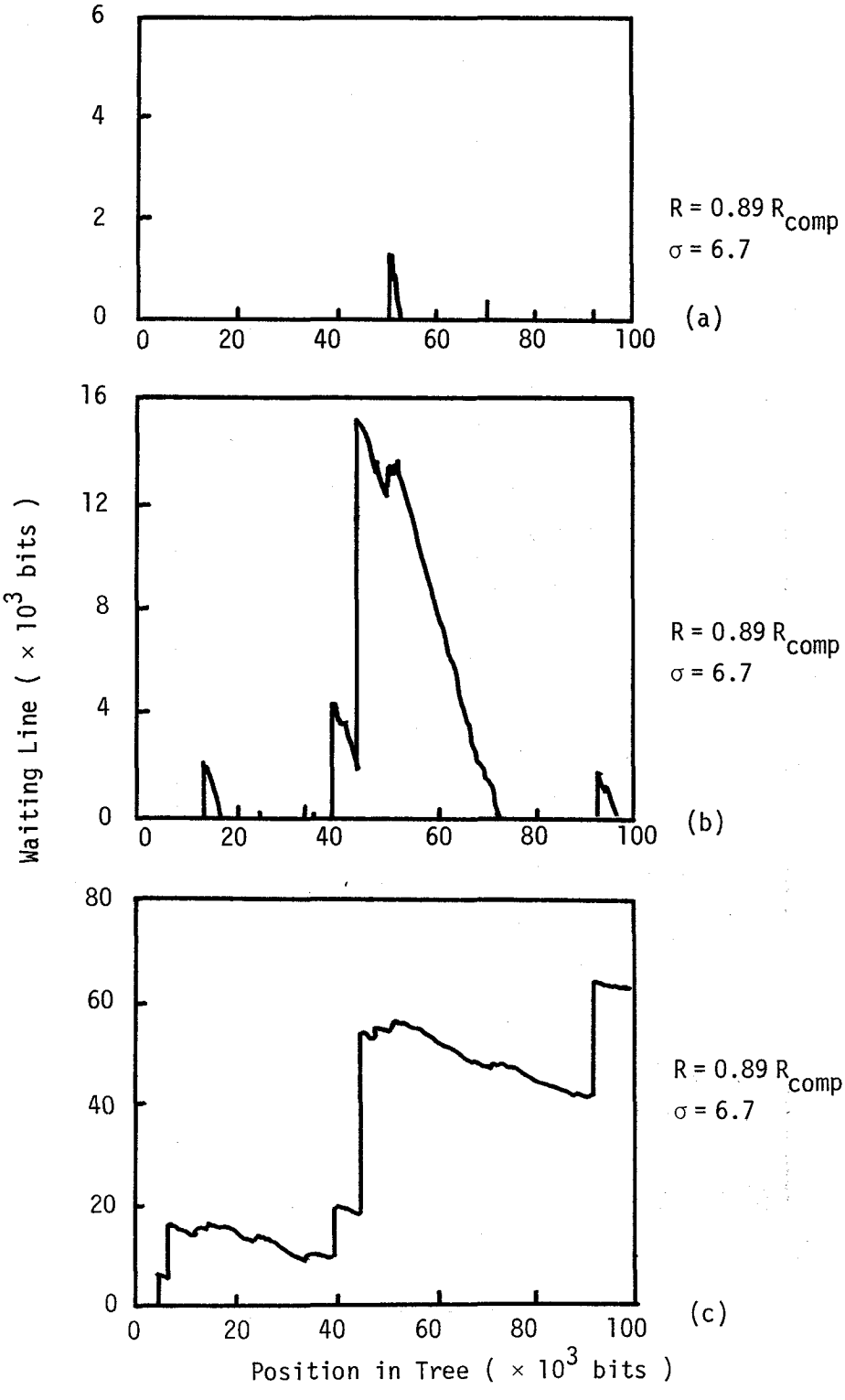


Fig. 4.3.5 — Sample Waiting Lines

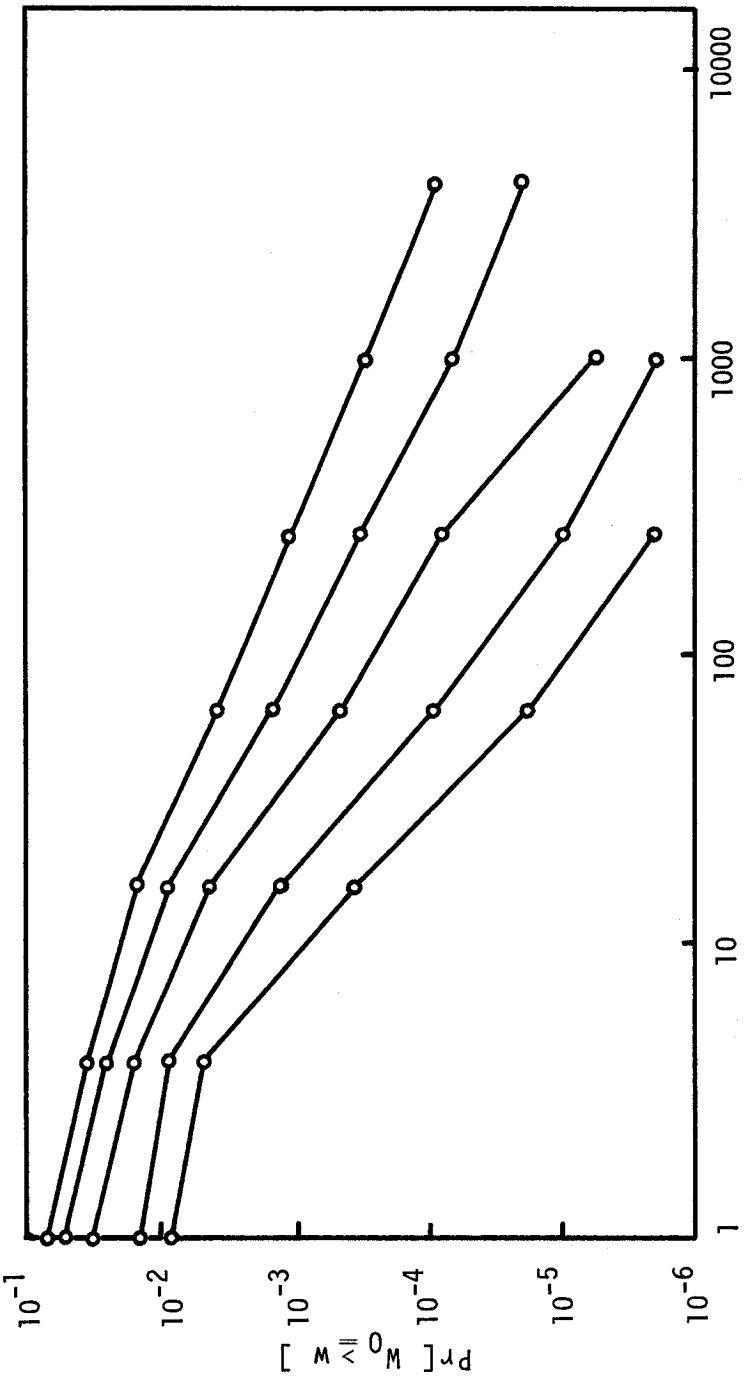


Fig. 4.3.4 – Measured Cumulative Distribution for the Number of Computations

4. Probability of deficient decoding

Implementable Decoder and Deficient decoding

In the previous section we observed that the ρ-th moment of the number of F-hypotheses per a node remains finite when block length increases, if $E_o(\rho,p,P) > \rho R$ for $\rho > 0$. In this section we see how the occasional heavy computational loads affect decoder, using a simple and practically meaningful decoder model.

We study a decoder consisting of three main units: a buffer which has storage capacity of νT channel output letters, a searching unit which retains a tree having (T+S) levels, and a control unit which controls node searching process according to a modified sequential decoding algorithm with a fixed search length S. The whole system is depicted in Fig. 4.4.1.

The sequential decoders that have concerned us search nodes sequentially, but emit decoded sequence in a block. On the other hand, the modified sequential decoding algorithm makes each decision on decoder output letters sequentially as follows: 1) The decoder searches the code tree, and, when a first F-hypothesis is made on a node at the S-th level, the decoder decides that the correct path is through the first antecedent of that node, call it a decoded node (see Fig. 4.4.2): 2) In general, the decoder searches all descendants of the previously decoded node, say, at the i-th level,

and, when an F-hypothesis is first made on a node at the (i+S)th level, the decoder makes the next descendant, directing toward the just reached node, of the previously decoded node a new decoded node. We call S the search length and call T the buffer length.

Using this modified sequential decoding algorithm, the decoder proceeds repeating successive decoding cycles, in each of which a reproduced part of the original code tree is decoded until T more nodes are decoded, and, after the completion, all of the buffer content is shifted into the searching unit for another T cycles of node searching on a sub-tree.
We suppose that each message has the format

$$\overbrace{* \ \ldots \ *}^{T}\overbrace{0 \ \ldots \ 0}^{S} \ .$$

We say that a deficient decoding has occurerred for the first time in the k-th decoding cycle and denote the event by G_k, if the first error is found between symbols decoded in this cycle. The cause of a deficient decoding in the k-th decoding cycle is two-fold: an inevitable error inherent in the decoding algorithm, with a fixed search length S, and a buffer overflow caused by severe computational requirements for correct decoding. We call the

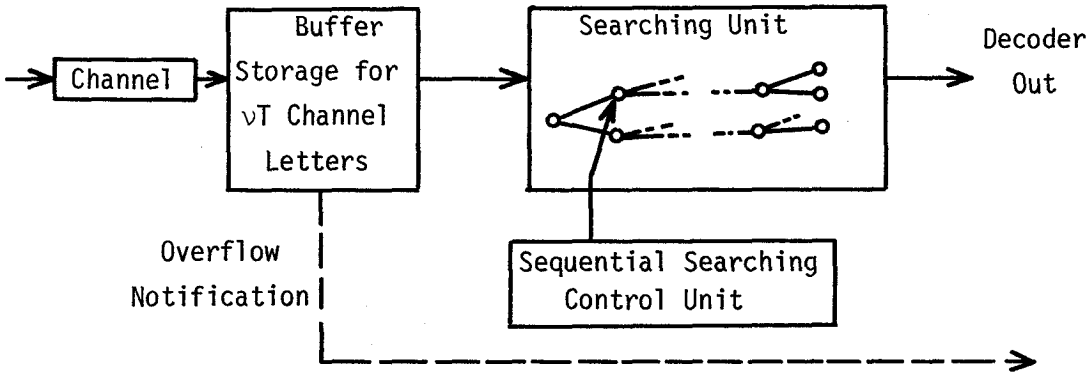


Fig. 4.4.1 — Implementable Sequential Decoder

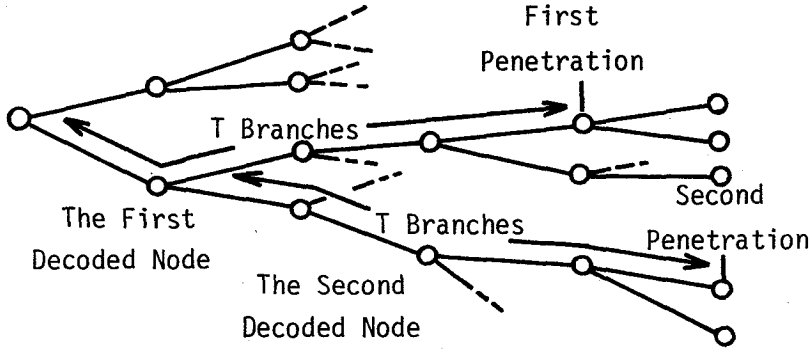


Fig. 4.4.2 — The Modified Sequential Decoding
with the Fixed Search Length S

former an errorneous decoding, denote it by E_k, and call the latter a buffer overflow, denote it by B_k. Thus,

$$\Pr\{ G_k \} = \Pr\{ B_k \} + \Pr\{ E_k \} ,$$

and the average probability of deficient decoding per a decoding cycle is

$$\overline{P}_G \triangleq \frac{1}{n_d} \sum_{k=1}^{n_d} P\{ G_k \},$$

where n_d is the number of decoding cycles needed to decode message sequences.

Let σ be the maximum number of basic moves that the decoder can carry out while ν channel output symbols come, and call it the computational speed. The following lower bound is shown in Jelinek [16] using the result of Jacobs and Berlekamp [23].

Theorem 4.1.1: For $\rho > 0$ satisfiying $\hat{E}_{sp}(\rho,P) < \rho R$,

$$\overline{P}_G \geq \frac{\exp\{ -O(\sqrt{\log \sigma[S+T]}) \}}{\sigma^{\rho}(S + T)^{\rho-1}}$$

where $\hat{E}_{sp}$ is the convex hull of E_{sp}, in ρ, given by

$$E_{sp}(\rho,P) \triangleq \max_{p} E_o(\rho,p,P)$$

and $O(*)$ is any function such that $O(\delta)/\delta$ is bounded for all large δ.

Thus the probability of deficient decoding never decreases faster than algebraic convergence for large S, T, and σ. This, perhaps, surprisingly slow convergence is balanced with decoder's moderate computation. For almost DMC's, $E_{sp}(\rho,P)$ is convex in ρ, and hence is equal to $\hat{E}_{sp}(\rho,P)$. A pathological exception is seen in Gallager [2, p.147].

The interest of this section is on upper bounds of $\overline{P}_G$; namely, we show the existence of good convolutional codes that allow us tight bounding of $\overline{P}_G$. Again the proof is through a random coding argument using the random convolutional code. Writing the expectation operator relative to the random convolutional code as $\mathcal{E}$, we have

$$\mathcal{E}\overline{P}_G = \mathcal{E}\Pr\{ B_1 \} + \mathcal{E}\Pr\{ E_1 \}.$$

As for the probability of errorneous decoding, we have a bound from the results in [16].

Lemma 4.4.1: For ε, $\rho > 0$ satisfying $E_o(\rho,p,P) - \rho R = \varepsilon/(1+\rho)$,

$$\Pr\{ E_1 \} \leq \frac{T}{1 - e^{-\varepsilon\nu}} e^{\rho\Delta/(1+\rho)} e^{-\varepsilon\nu\rho S},$$

where the constraint length is $K = \infty$.

Therefore our major task is to show a sufficiently tight upper bound on the probability of buffer overflow.

The Probabilities of Deficient Decoding and Buffer Overflow

A buffer overflow occurs if, and only if, the number of basic moves in a decoding cycle exceeds $\sigma(S+T)$ As noted in the previous section, we know that, for each F-hypothesis on a node, at most q-1 lateral moves and a backward move can occur between the immediate descendants of the node. Therefore, we have the bound

$$\Pr\{ B_1 \} \leq \Pr\{ (q+1) \sum_{j=0}^{S+T-1} W_j > \sigma(S+T) \}. \quad (4.4.1)$$

Jelinek [16] shows the following bounds on $\mathcal{E}\Pr\{B_1\}$ for tree codes (not necessarily convolutional codes):

Theorem 4.4.2: For $\rho > 0$ satisfying $E_o(\rho,p,) > \rho R$,

$$\mathcal{E}\Pr\{B_1\} \leq \begin{cases} \dfrac{\gamma_1}{\sigma^{\rho}(S + T)^{\rho-1}} , & \text{if } 0 < \rho \leq 1, \\ \dfrac{\gamma_2}{(\sigma - \gamma_3) (T/S + 1)^{\rho-1}} , & \text{if } 1 < \rho \leq 2, \\ \dfrac{\gamma_4}{(\sigma - \gamma_3) (T/S + 1)^{\rho/2}} , & \text{if } 2 < \rho , \\ \dfrac{SM_1(\sigma)}{(T/S + 1)^{\rho-1}} , & \text{if } 2 < \rho , \end{cases}$$

where $\gamma_1 - \gamma_4$ are constants and $M_1(*)$ is a decreasing finite valued function of σ, all of which are independent of σ(except $M_1(*)$), S, and T.

The theorem gives an asymptotically tight bound for $0 < \rho \leq 1$. But it is obviously uninteresting case for P_G increases as S and T increase from Theorem 4.1.1. In this sense, the rate $R_{comp} = E_{sp}(1,P)$ is regarded as the limit for meanigful sequential decoding, and is said to be the computational critical rate. For $\rho > 1$, the bounds are rather loose and disunited, and, even worse, give no answer for convolutional codes.

To derive a more general bound, we use an additional notation: for a node $\underline{u}^{j+\ell}$ specifying a codeword $\underline{x}_1^{\nu} \ldots \underline{x}_{j+\ell}^{\nu}$, let

$$\Gamma_{\ell,j}(\underline{u}^{j+\ell}) \triangleq \sum_{k=1}^{\ell} \left[\log \frac{P(\underline{y}_{j+k}^{\nu} | \underline{x}_{j+k}^{\nu})}{q(\underline{y}_{j+k}^{\nu})} - \nu R \right] .$$

If $\underline{u}^{j+m}$ is a correct node, we write it simply as $\Gamma_{m,j}$. With this notation, we see

$$\Gamma_{j+\ell}(\underline{u}^{j+\ell}) - \Gamma_{j+\ell} = \Gamma_{\ell,j}(\underline{u}^{j+\ell}) - \Gamma_{\ell,j}$$

for any $\underline{u}^{j+\ell} \in D_j^{\ell}$.

Therefore, if we put as

$$W_{\ell,m,j} \triangleq \sum_{h=1}^{\infty} \sum_{\underline{u}\in D_j^{\ell}} \chi[\, \Gamma_{\ell,j}(\underline{u}) > \Gamma_{m,j} + (h-2)\Delta \,],$$

then, from (4.4.1) and the next to the last expression in (4.3.2), we have

$$\Pr\{ B_1 \}$$

$$\leq \Pr\{ (q+1) \sum_{\ell=0}^{S} \sum_{m=0}^{S} \sum_{j=0}^{S+T-1} W_{\ell,m,j} > \sigma(S + T) \}$$

$$\leq \sum_{\ell=0}^{S} \sum_{m=0}^{S} \Pr\{ (q+1) \sum_{j=0}^{S+T-1} W_{\ell,m,j} > \sigma_{\ell,m}(S + T) \}$$

where $\sigma_{\ell,m}$ are positive numbers satisfying

$$\sum_{\ell=0}^{S} \sum_{m=0}^{S} \sigma_{\ell,m} = \sigma$$

and are determined later. The right-hand side summation in the above bound is divided into three:

$$\mathcal{E}\Pr\{ B_1 \} \leq \left(\sum_{\ell=m=0}^{S} + \sum_{\ell=1}^{S} \sum_{m=0}^{\ell-1} + \sum_{m=1}^{S} \sum_{\ell=0}^{m-1} \right)$$

$$\times \mathcal{E}\Pr\{ (q+1) \sum_{j=0}^{S+T-1} W_{\ell,m,j} > \sigma_{\ell,m}(S + T) \}$$

$$\triangleq P_{B,1} + P_{B,2} + P_{B,3} ,$$

where we put the respective summations as $P_{B,1}$, $P_{B,2}$, and $P_{B,3}$.

The first summation is further bounded as follows.

$$P_{B,1} \leq \sum_{m=0}^{S} \sum_{k=0}^{m} \xi \Pr\left\{ (q+1) \sum_{\substack{mj+k \leq S+T-1 \\ j \geq 0}} W_{m,m,mj+k} > \sigma_{m,m}(S+T)/m \right\}.$$

Now we note that $W_{\ell,m,j}$ and $W_{\ell,m,j+k}$ are independent random variables under the probability of the random code and channel if $k > \max(\ell,m)$; an easy consequence of the memoryless property. Therefore the right-hand side of the bound on $P_{B,1}$ is exactly the sum of tail probabilities of accumulated iid random variables. And the next bound on the tail probability of the sum of iid random variables is vital, which is proved later.

Theorem 4.4.3: For iid nonnegative random variables $X_1, X_2, \ldots$ with finite ρ-th moments ($\rho \geq 1$) and $EX_1 < \sigma - 1$,

$$\Pr\{ \sum_{j=1}^{N} X_j \geq \sigma N \} \leq \begin{cases} \dfrac{2^{1+\rho} EX_1^{\rho}}{(\sigma - EX_1)^{\rho} N^{\rho-1}}, & \text{if } 1 \leq \rho < 2, \end{cases}$$

$$\left\{ \begin{array}{l} \dfrac{2^{1+\rho} EX_1^{\rho}}{(\sigma - EX_1)^{\rho} N^{\rho-1}} \\ \\ + \dfrac{2^{2\rho+5}(\rho+2)^{3\rho+3}(EX_1^{\rho})^{\xi}}{(\sigma - EX_1)^{\rho} N^{\rho}} \end{array} \right. , \text{ if } 2 \leq \rho,$$

for all N, where E is the expectation operator and $\xi = 3$ if $EX_1 \geq 1$ and $\xi = 1$ if $EX_1 < 1$.

To make arguments simple, we temporarily assume $1 \leq \rho < 2$. Then, in view of the lemma, a little calculation reveals

$$P_{B,1}$$

$$\leq \sum_{m=0}^{S} \frac{(q+1)^{\rho} 2^{1+\rho} m^{\rho} \xi EW_{m,m,0}^{\rho}}{[\sigma_{m,m} - (q+1)\xi EW_{m,m,0}]^{\rho}(S+T)^{\rho-1}} .$$

Almost in the same way, the other terms are bounded as

$$P_{B,2}$$

$$\leq \sum_{\ell=1}^{S} \sum_{m=0}^{\ell-1} \frac{(q+1)^{\rho} 2^{1+\rho} \ell^{\rho} \xi EW_{\ell,m,0}^{\rho}}{[\sigma_{\ell,m} - (q+1)\xi EW_{\ell,m,0}]^{\rho}(S+T)^{\rho-1}} , \text{ and}$$

$$P_{B,3}$$

$$\leq \sum_{\ell=1}^{S} \sum_{m=0}^{\ell-1} \frac{(q+1)^{\rho} 2^{1+\rho} m^{\rho} \mathcal{E} EW_{\ell,m,0}^{\rho}}{[\ \sigma_{\ell,m} - (q+1)\mathcal{E} EW_{\ell,m,0}\ ]^{\rho}(\ S + T\ )^{\rho-1}} .$$

Therefore, we have

$$\mathcal{E} Pr\{\ B_1\ \}$$

$$\leq \sum_{\ell=0}^{S} \sum_{m=0}^{S} \frac{(q+1)^{\rho} 2^{1+\rho} \ell^{\rho} m^{\rho} \mathcal{E} EW_{\ell,m,0}^{\rho}}{[\ \sigma_{\ell,m} - (q+1)\mathcal{E} EW_{\ell,m,0}\ ]^{\rho}(\ S + T\ )^{\rho-1}} .$$

The convergence of summations in the right-hand side is assured by the following lemma, which is obtained by a slight modification of the bounds in the previous section as shown later.

Lemma 4.4.2: For any $\rho > 0$,

$$\mathcal{E} EW_{\ell,m,0}^{\rho} \leq \gamma_0(\ \ell + 1\ )^{n} q^{\rho n(n+1)/2}$$

$$\times \exp\{ - \frac{\nu}{1+\rho} [\ m + (1+\rho-n)\ell/n\ ][\ E_o(\rho,p,P) - \rho R\ ] .$$

where $n-1 < \rho \leq n$ and γ_0 is given below (4.3.3).

According to this lemma, if we let

$$\sigma_{\ell,m} - (q+1)\,\xi EW^{\rho}_{\ell,m,0}$$

$$= \frac{e^{-\gamma\ell-\delta m}}{\sum_{\ell=0}^{S}\sum_{m=0}^{S} e^{-\gamma\ell-\delta m}} \left[\, \sigma - (q+1) \sum_{\ell=0}^{S}\sum_{m=0}^{S} \xi EW_{\ell,m,0} \,\right]$$

and, if we let γ and δ be sufficiently small positive constants, then the bound on $\Pr\{B_1\}$ converges as $S \to \infty$, and we have

$$\Pr\{ B_1 \} \leq \frac{W_2^*}{[\, \sigma - W_1^* \,]^{\rho}(S + T)^{\rho-1}}$$

where we put

$$W_1^* = (q+1) \sum_{\ell=0}^{\infty}\sum_{m=0}^{\infty} \xi EW_{\ell,m,0} \,, \qquad \text{and}$$

$$W_2^* = \frac{(q+1)\; 2^{1+\rho}}{(1 - e^{-\gamma})(1 - e^{-\delta})}$$

$$\times \sum_{\ell=0}^{\infty}\sum_{m=0}^{\infty} \ell^{\rho} m^{\rho} e^{-\gamma\ell-\delta m} \xi EW^{\rho}_{\ell,m,0} \,.$$

Note that $W_1^* < \infty$ is implied by $W_2^* < \infty$ because $\rho \geq 1$. For the case $2 \leq \rho$, the arguments also hold with a little

modification. We state the result in a theorem:

Theorem 4.4.4: For any $\rho \geq 1$ satisfying $E_o(\rho,p,P) > \rho R$,

$$\mathcal{E}\Pr\{ B_1 \} \leq \begin{cases} \dfrac{W_2^*}{(\sigma - W_1^*)^{\rho}(S + T)^{\rho-1}} , & \text{if } 1 \leq \rho < 2, \\ \dfrac{W_2^*}{(\sigma - W_1^*)^{\rho}(S + T)^{\rho-1}} \left(1 + \dfrac{W_3^*}{S + T} \right) , & \text{if } 2 \leq \rho , \end{cases}$$

where W_1^*, W_2^*, and W_3^* are finite constants independent of σ, S, and T.

Now let $S \geq (1/\varepsilon)\log[(\sigma - W_1^*)(S+T)]$ in Lemma 4.4.1. Then the following is an immediate consequence of the lemma and Theorem 4.4.4.

Theorem 4.4.5: Suppose that $\rho \geq 1$ and $E_o(\rho,p,P) > \rho R$. Then, for sufficiently large S and T, there exists a convolutional code such that

$$\overline{P}_G \leq \frac{W_0^*}{(\sigma - W_1^*)^{\rho}(S + T)^{\rho-1}}$$

where W_0^* and W_1^* are finite constants independent of

σ, S, and T.

Corollary: For $\rho > 1$, suppose that $E_{sp}(\rho,P) = \hat{E}_{sp}(\rho,P)$. Then, for any $\varepsilon > 0$, the best attainable $\overline{P}_G$ satisfies

$$\frac{\delta_1}{\sigma^{\rho+\varepsilon}(S + T)^{\rho+\varepsilon-1}} \leq \inf_{\substack{\text{conv. codes} \\ (K=\infty)}} \overline{P}_G$$

$$\leq \frac{\delta_2}{\sigma^{\rho-\varepsilon}(S + T)^{\rho-\varepsilon-1}}$$

for sufficiently large σ, S, and T, where the infimum is over all convolutional codes ($K = \infty$) and δ_1 and δ_2 are positive constants independent of σ, S, and T.

Corollary gives a complete answer to the asymptotic behavior of the probability of deficient decoding, when $K = \infty$. For finite constraint length, a similar result will be shown with more elaborate analysis.

Finally we note that all results derived here aplly to time-varying convolutional codes. Since codes used in practice are of time-invariant, another problem thus seems to exist.

5. Convolutional codes in practice

Before ending this chapter, we briefly mention to space communication as a promising field for convolutional codes and Viterbi or sequenctial decoding. Space communication includes satellite-to-ground communication (relatively short distance) and space-probe-to-earth communication (long distance): The former requires high speed transmission and the latter requires extremely large ability to overcome severe circumstances.

Consider satellite-to-ground transmission of data (see Fig. 4.5.1), which may be messages from other ground stations or data about the weather of a district. Typically, the data are binary digits and to be transmitted one bit each τ seconds in the form $a(t) = a_i$ ($= \pm 1$) for $i\tau \leq t < (i+1)\tau$. A standard modulation technique is phase-shift keying (PSK), where the modulated carrier signal is

$$x(t) = \alpha a(t) \cos \omega_c t \ .$$

The parameter α indicates the power of the transmitted signal. In an ideal situation, the ground station demodulates the received signal $y(t)$ through a correlator

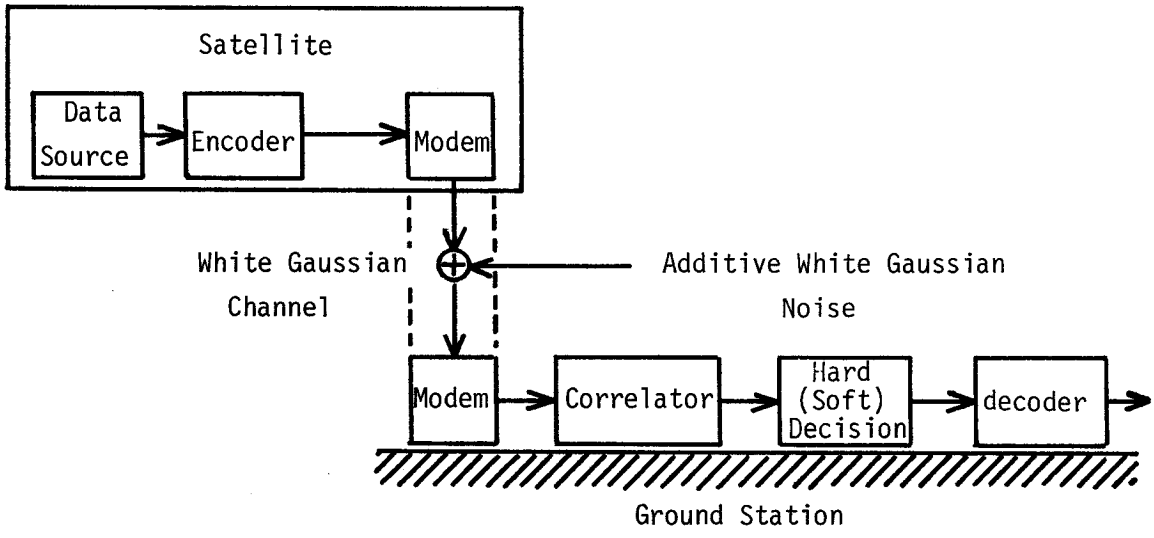


Fig. 4.5.1 — Satellite-To-Groung Communication

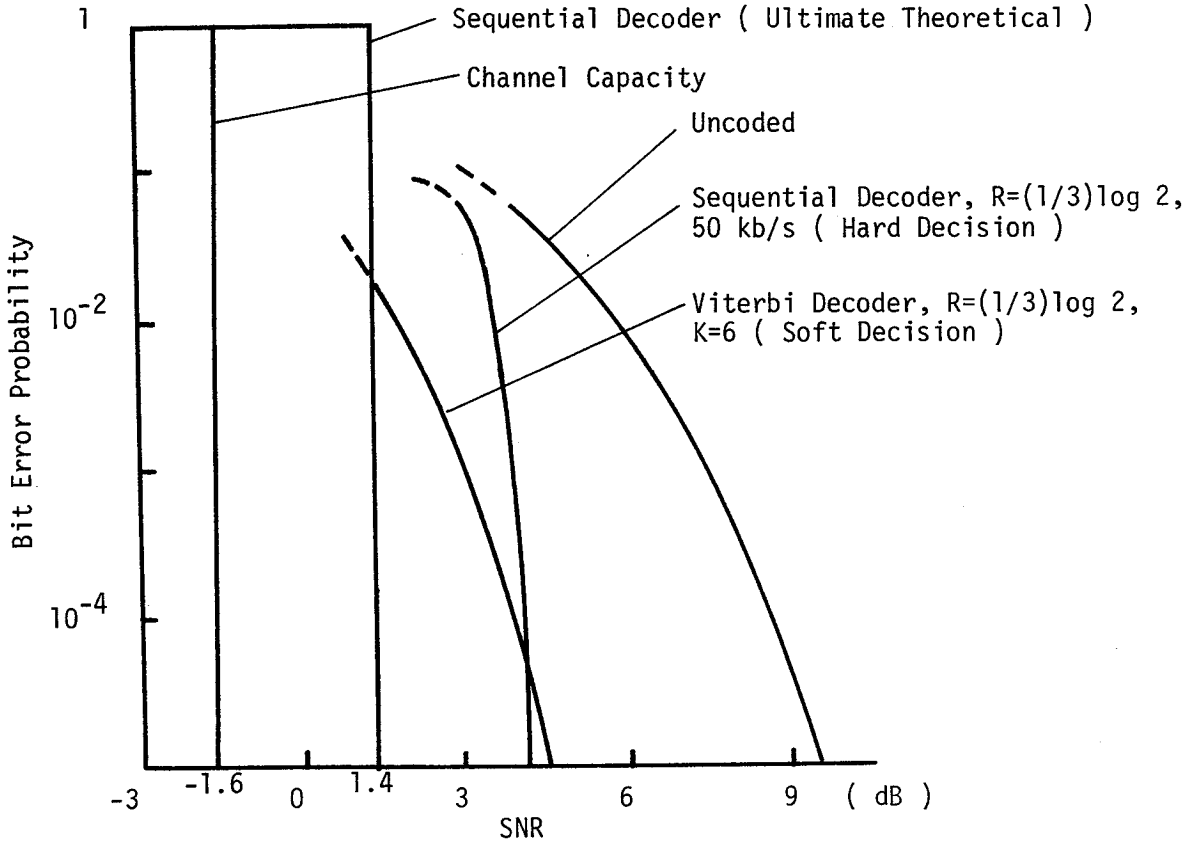


Fig. 4.5.2 — Performance of Decoders

as

$$z_i = \int_{i\tau}^{(i+1)\tau} y(t) \cos \omega_c t \, dt$$

$$= \alpha a_i + n_i \quad ,$$

where n_i are the noise. Let $(1/2\tau)N_o$ be the variance of the noise n_i. The ratio of the signal power α^2 to N_o is called the signal-to-noise ratio (SNR) per a bit and is expressed in 10 log α^2/N_o (dB).

In space communication, the noise is surprisingly well approximated by iid zero-mean Gaussian random variables. In literatures such a channel is called a white Gaussian channel. Generally the greater the nominal bandwidth $1/2\tau$ is the more the capacity of the channel increases, and, as $\tau \to 0$, the limit SNR necessary for efficient communication approaches to -1.6 dB, called the Shannon limit.

In this scheme, though the original data are binary, the correlator outputs are not binary. We can convert these analog data into binary data $\hat{a}_i$ by hard decision, $\hat{a}_i = 1$ (-1) for $z_i \geq$ ($<$) 0, and decode $\hat{a}_1\hat{a}_2$ .... Contrary to hard decision, we can decode the correlator output z_1z_2... directly. Then we say that

the decoder uses soft decision. Only soft decision decoding attains the Shannon limit in the limit.

In Fig. 4.5.2, typical performance curves are shown (see [64],[65]). Convolutional codes used with Viterbi decoders have maximally K = 7 or 8, while convolutional codes used with sequential decoders have generally larger constraint lengths. We can see that Viterbi decoders are generally superior to other decoders for moderate error probabilities. On the other hand, because of the larger constraint lengths, sequential decoders exhibit sharp reduction of the error probability by increasing coding gain (SNR). Thus sequential decoders will find applications in fields where extremely small error probabilities are needed.

Appendix to Chapter IV

Proof of Lemma 4.3.4

First note

$$\mathcal{E}_I \prod_{j=1}^{n} \left[\frac{P(\underline{y}_\ell | \underline{X}_\ell(j))}{q(\underline{y}_\ell)} \right]^{\frac{1}{1+\rho}}$$

$$= \mathcal{E}_I \prod_{j=1}^{n} \prod_{i=1}^{\ell} \left[\frac{P(\underline{y}_i | \underline{X}_i(j))}{q(\underline{y}_i)} \right]^{\frac{1}{1+\rho}}$$

$$= \prod_{k=1}^{n} \prod_{i=\ell_{k-1}+1}^{\ell_k} \mathcal{E}_I \prod_{j=1}^{n} \left[\frac{P(\underline{y}_i | \underline{X}_i(j))}{q(\underline{y}_i)} \right]^{\frac{1}{1+\rho}}$$

where $\prod_{i=\ell}^{m} (*) = 1$ for $\ell > m$. Suppose that $\ell_{k-1} < i \leq \ell_k$. Then, from Corollary to Lemma 4.3.2, there are n' ($\leq n$) subsets consisting of k independent random i-th branch sequences:

$$U(\beta) = \{ \underline{X}^{\nu}(\beta,1), \ldots, \underline{X}^{\nu}(\beta,k) \} , \beta = 1, \ldots, n' ,$$

in the set $\{ \underline{X}_i^{\nu}(1), \ldots, \underline{X}_i^{\nu}(n) \}$. Note that we may have $\underline{X}^{\nu}(\alpha,\beta) = \underline{X}^{\nu}(\alpha',\beta')$ for distinct (α,β) and (α',β'). Let $D_{\alpha,\beta}$ denote the number of $U(\beta)$'s that contain $X(\alpha,\beta)$. Obviously,

$$\sum_{\alpha=1}^{k} \sum_{\beta=1}^{n'} \frac{1}{D_{\alpha,\beta}} = n.$$

Therefore

$$\mathcal{E}_I \prod_{j=1}^{n} \left[\frac{P(\underline{y}_i^\nu | \underline{X}_i^\nu(j))}{q(\underline{y}_i^\nu)} \right]^{\frac{1}{1+\rho}}$$

$$= \mathcal{E}_I \prod_{\alpha=1}^{n'} \prod_{\beta=1}^{k} \left[\frac{P(\underline{y}_i^\nu | \underline{X}^\nu(\alpha,\beta))}{q(\underline{y}_i^\nu)} \right]^{\frac{1}{(1+\rho)D_{\alpha,\beta}}}$$

$$\leq \prod_{\alpha=1}^{n'} \left\{ \mathcal{E}_I \prod_{\beta=1}^{k} \left[\frac{P(\underline{y}_i^\nu | \underline{X}^\nu(\alpha,\beta))}{q(\underline{y}_i^\nu)} \right]^{\frac{1}{(1+\rho)D_{\alpha,\beta}}} \right\}^{\frac{1}{n'}}$$

where the last inequality follows from Hölder's inequality for n' random variables. Since each $U(\beta)$ consists of independent random branch sequences, the bound continues as:

$$\prod_{\alpha=1}^{n'} \left\{ \prod_{\beta=1}^{k} \sum_{\underline{x}^\nu \in A^\nu} p(\underline{x}^\nu) \left[\frac{P(\underline{y}_i^\nu | \underline{x}_i^\nu)}{q(\underline{y}_i^\nu)} \right]^{\frac{n}{(1+\rho)D_{\alpha,\beta}}} \right\}^{\frac{1}{n'}}.$$

For any, but fixed, α and β, let $n'/D_{\alpha,\beta} = \xi$. Since $n \geq n' \geq D_{\alpha,\beta} \geq 1$, we have $1+\rho > n \geq \xi \geq 1$. Thus, the summation over $\underline{x}^\nu$ in the extreme right-hand side is bounded by Jensen's inequality as follows:

$$\sum_{\underline{x}^\nu \in A^\nu} p(\underline{x}^\nu) \left(\frac{P(\underline{y}^\nu | \underline{x}^\nu)}{q(\underline{y}^\nu)} \right)^{\frac{\xi}{(1+\rho)}}$$

$$= \sum_{\underline{x}^\nu \in A^\nu} p(\underline{x}^\nu) \left(\frac{Q(\underline{x}^\nu|\underline{y}^\nu)}{p(\underline{x}^\nu)} \right)^{\frac{\xi}{1+\rho}}$$

$$= \sum_{\underline{x}^\nu \in A^\nu} Q(\underline{x}^\nu|\underline{y}^\nu) \left(\frac{p(\underline{x}^\nu)}{Q(\underline{x}^\nu|\underline{y}^\nu)} \right)^{\frac{1+\rho-\xi}{1+\rho}}$$

$$\leq \left[\sum_{\underline{x}^\nu \in A^\nu} Q(\underline{x}^\nu|\underline{y}^\nu) \left(\frac{p(\underline{x}^\nu)}{Q(\underline{x}^\nu|\underline{y}^\nu)} \right)^{\frac{\rho}{1+\rho}} \right]^{\frac{1+\rho-\xi}{\rho}}$$

$$= \left[\sum_{\underline{x}^\nu \in A^\nu} p(\underline{x}^\nu) \left(\frac{P(\underline{y}^\nu|\underline{x}^\nu)}{q(\underline{y}^\nu)} \right)^{\frac{1}{1+\rho}} \right]^{\frac{1+\rho-\xi}{\rho}}$$

where Q is the inverse conditional pmf given by $Q(\underline{x}^\nu|\underline{y}^\nu) = P(\underline{y}^\nu|\underline{x}^\nu)p(\underline{x}^\nu)/q(\underline{y}^\nu)$ for each $\underline{x}^\nu \in A^\nu$ and each $\underline{y}^\nu \in B^\nu$. Therefore we have

$$\xi_I \prod_{i=1}^{n} \left(\frac{P(\underline{y}_i^\nu|\underline{x}_i^\nu(j))}{q(\underline{y}_i^\nu)} \right)^{\frac{1}{1+\rho}}$$

$$\leq \prod_{\alpha=1}^{n'} \left\{ \prod_{\beta=1}^{k} \left[\sum_{\underline{x}^\nu \in A^\nu} p(\underline{x}^\nu) \left(\frac{P(\underline{y}_i^\nu|\underline{x}^\nu)}{q(\underline{y}_i^\nu)} \right)^{\frac{1}{1+\rho}} \right]^{\frac{1+\rho-n}{\rho D_{\alpha,\beta}}} \right\}^{\frac{1}{n'}}$$

$$= \left[\sum_{\underline{x}^\nu \in A^\nu} p(\underline{x}^\nu) \left(\frac{P(\underline{y}_i^\nu|\underline{x}^\nu)}{q(\underline{y}_i^\nu)} \right)^{\frac{1}{1+p}} \right]^{\sum_{\alpha=1}^{n'} \sum_{\beta=1}^{k} \left(\frac{1+\rho}{\rho n'} - \frac{1}{\rho D_{\alpha,\beta}} \right)}$$

$$= \left[\sum_{\underline{x}^{\nu} \in A^{\nu}} p(\underline{x}^{\nu}) \left(\frac{P(\underline{y}_i^{\nu} | \underline{x}^{\nu})}{q(\underline{y}_i^{\nu})} \right)^{\frac{1}{1+\rho}} \right]^{\frac{(1+\rho)k-n}{\rho}} .$$

This proves the lemma.

Proof of Lemma 4.3.2

Without loss of generality suppose that the set of first k sequences, $\underline{u}^i(1), \ldots, \underline{u}^i(k)$, has rank k. Let [U], [S], and [G] be an n×i matrix with (j,ℓ) entry $u_\ell(j)$, n×ν matrix with (j,ℓ) entry $s^j_{i,\ell}$, and ν×i matrix with (j,ℓ) entry $g^{(i)}_{j,\ell}$, respectively. Then (4.1.1) is written as

$$[S] = [U][G] .$$

Since the set of messages has rank k, there is a nonsingular linear transformation [T] such that

$$[U][T] = \begin{bmatrix} u_1(1) & & & \\ u_1(2) & u_2(2) & & 0 \\ \vdots & & \ddots & \\ u_1(k) & \ldots & u_k(k) & \\ \vdots & \ldots & \vdots & 0 \\ u_1(n) & \ldots & u_k(n) & \end{bmatrix}_{n\times n}$$

This transformation yields

$$s^j_{i,\ell} = u_j(j)g^{(i)}_{i-j+1,\ell} + \sum_{m=i-j+1}^{i-1} u_{i-m}(j)g^{(i)}_{m+1,\ell}$$

for j = 1, ... ,k and ℓ = 1, ... ,ν, where $g^{(i)}_{j,\ell}$ are elements of the matrix $[T]^{-1}[G]$. Since [G] is a matrix with independent equi-probably distributed random components, $[T]^{-1}[G]$ has the same statistical

property. Therefore $s^{j}_{i,\ell}$,j = 1, ... ,k, ℓ = 1, ... ,ν, are also distributed independently and equiprobably. Therefore, from (4.1.2), the assertions follow.

Proof of Lemma 4.3.3

First consider the subsequences, $\underline{u}^{(1)}_{L_0,L_1}$, ... ,$\underline{u}^{(n)}_{L_0,L_1}$ of the sequences $\underline{u}^{L}(1)$, ... ,$\underline{u}^{L}(n)$, where we let

$$\underline{u}_{L_{k-1},L_k}(i) = u_{L_{k-1}+1}(i) \ldots u_{L_k}(i) ,$$

for i = 1, ... ,n and k = 1, ... ,n. Since the set of these subsequences has rank 1,

$$\underline{u}^{(i)}_{L_0,L_1} = \alpha_i \, \underline{u}^{(1)}_{L_0,L_1} ; \; i = 2, \ldots ,n,$$

for some nonzero α_i in GF(q). Since the number of distinct $\underline{u}^{(1)}_{L_0,L_1}$ is $q^{L_1-L_0}$, the number of all subsequemces $\{\underline{u}^{(1)}_{L_0,L_1}, \ldots ,\underline{u}^{(n)}_{L_0,L_1}\}$ is bounded by $q^{n-1}q^{L_0-L_1}$. Next, consider subsequences, $\underline{u}^{(1)}_{L_1,L_2}$, ... ,$\underline{u}^{(n)}_{L_1,L_2}$, of rank 2. For these subsequences these exist α_i and β_i in GF(q) such that (assuming the first two subsequences are linearly independent)

$$\underline{u}^{(i)}_{L_1,L_2} = \alpha_i \underline{u}^{(i)}_{L_1,L_2} + \beta_i \underline{u}^{(i)}_{L_1,L_2} \quad ; \; i = 3, \ldots, n.$$

Since the number of pairs $\underline{u}^{(1)}_{L_1,L_2}$ and $\underline{u}^{(2)}_{L_1,L_2}$ of rank 2 is bounded by $(q-1)q^{L_1-L_2-1}q^{L_2-L_1}q^{2(L_2-L_1)}$, the number of sets $\{\underline{u}^{(1)}_{L_1,L_2}, \ldots, \underline{u}^{(n)}_{L_1,L_2}\}$ of rank 2 is bounded by $q^{2n}q^{2(L_2-L_1)}$. In general, for $L_k > L_{k-1}$, the number of sets $\{\underline{u}^{(1)}_{L_{k-1},L_k}, \ldots, \underline{u}^{(n)}_{L_{k-1},L_k}\}$ of rank k is bounded by $q^{kn}q^{k(L_k-L_{k-1})}$. Since this bound is also valid for $L_k = L_{k-1}$, the total number of sets $\{\underline{u}^{\underline{L}}(1), \ldots, \underline{u}^{\underline{L}}(n)\}$ which is $\underline{L}$-independent is bounded by

$$M(\underline{L}) \leq \sum_{k=1}^{n} q^{kn+k(L_k-L_{k-1})}$$

$$= q^{\sum_{k=1}^{n} [\, kn + k(L_k-L_{k-1}) \,]}$$

$$= q^{n^2(n+1)/2 + \sum_{k=1}^{n} k(L_k-L_{k-1})} .$$

Proof of Theorem 4.4.3

Since the first bound, for $1 \leq \rho < 2$, is already proved in Jelinek [16] with the aid of the inequality in [24], we assume $\rho \geq 2$. Let the random variables

$$U_j \triangleq X_j \, \chi[\, X_j < \beta N \,] \qquad \text{and}$$

$$V_j \triangleq X_j - U_j \,,$$

where β is a positive constant determined later. Note that $V_j \geq \beta N$ whenever $V_j > 0$. First, we have

$$\Pr\{ \sum_{j=1}^{N} X_j \geq \sigma N \}$$

$$\leq \Pr\{ \sum_{j=1}^{N} V_j \geq N \, EV_1 \}$$

$$+ \Pr\{ \sum_{j=1}^{N} (U_j - EU_1) \geq (\sigma - EX_1)N \} \,,$$

where E denotes the expectation operator. From Markov's inequality and the note above, the first term in the right-hand side is

$$\Pr\{ \sum_{j=1}^{N} V_j \geq \sigma N \} \leq N \Pr\{ V_j \geq \beta N \}$$

$$\leq \frac{EV_1^{\rho}}{\beta^{\rho} N^{\rho-1}} \,.$$

Therefore, if we let $Z_j = U_j - EU_1$ and $\sigma_0 = \sigma - EX_1$, then the bound on the tail probability is

$$\Pr\{ \sum_{j=1}^{N} X_j \geq \sigma N \}$$

$$\leq \frac{EV_1^\rho}{\beta^\rho N^{\rho-1}} + \Pr\{ \sum_{j=1}^{N} Z_j \geq \sigma_0 N \} .$$

The last term is the tail probability of the sum of zero-mean random variables, which we approximate next. For any positive integer n, the n-th power of the sum is

$$\left\{ \sum_{j=1}^{N} Z_j \right\}^n$$

$$= \sum_{k=1}^{n} \sum_{0 \leq n_1 \leq \ldots \leq n_k \atop \Sigma_j n_j = n} \frac{n!}{\prod_{j=1}^{k} n_j !} \sum_{i_1, \ldots, i_k} \prod_{j=1}^{k} Z_{i_j}^{n_j}$$

where the second summation is over all k-tuples (n_1, ... ,n_k) consisting of increasing positive integers whose sum is n, and the last summation is over all ordered k-tuples (i_1, ... ,i_k) consisting of distinct integers from 1 to N. From this time on, all summations over n_j's and i_j's are to be understood in these respective meanings, and are denoted simply by Σ_n^* and Σ_i^* respectively.

With this expression for the n-th power, we have a bound on the last term, the tail probability of the sum of Z_j,

$$\Pr\{ \sum_{j=1}^{N} Z_j \geq \sigma_0 N \}$$

$$= \Pr\left\{ \left(\sum_{j=1}^{N} Z_j \right)^n \geq \sigma_0^n N^n \right\}$$

$$\leq \Pr\{ \sum_{j=1}^{N} Z_j^n \geq (1-s)\sigma_0^n N^n \}$$

$$+ \sum_{k=2}^{n} \Sigma_n^* \Pr\{ \Sigma_i^* \prod_{j=1}^{k} Z_{i_j}^{n_j} \geq \sigma_1^n N^n \} ,$$

for $0 < s < 1$, where

$$\sigma_1^n = s\sigma_0^n / \sum_{i=2}^{n} i^n .$$

Now let $n-1 < \rho \leq n$. Then the first term of this bound is

$$\Pr\{ \sum_{j=1}^{N} Z_j^n \geq (1-s)\sigma_0^n N^n \}$$

$$\leq \Pr\{ \sum_{j=1}^{N} |Z_j|^\rho \geq (1-s)^{\rho/n} \sigma_0^\rho N^\rho \}$$

$$\leq \frac{2^\rho \, EU_1^\rho}{(1-s)^{\rho/n} \, \sigma_0^\rho \, N^{\rho-1}} ,$$

where the first inequality follows from $\Sigma_j Z_j^n \leq \Sigma_j |Z_j|^n \leq (\Sigma_j |Z_j|^\rho)^{n/\rho}$ and $n \geq \rho$, and the last one follows from $(E|U_1 - EU_1|^\rho)^{1/\rho} \leq (EU^\rho)^{1/\rho} + EU \leq 2(EU^\rho)^{1/\rho}$. Let $\beta = \sigma_0$ [$= \sigma - EX_1$] and let $(1-s)^{-\rho/n} = 2$ [then, $s \geq 1/4$]. Then, we have shown

$$\Pr\{ \sum_{j=1}^{N} X_j \geq \sigma N \}$$

$$\leq \frac{EV_1^\rho}{\sigma_0^\rho N^{\rho-1}} + \frac{2^{1+\rho} EU_1^\rho}{\sigma_0^\rho N^{\rho-1}}$$

$$+ \sum_{k=2}^{n} \Sigma_n^* \Pr\{ \Sigma_i^* \prod_{j=1}^{k} Z_{i_j}^{n_j} \geq \sigma_1^n N^n \}$$

$$\leq \frac{2^{1+\rho} EX_1^\rho}{(\sigma - EX_1)^\rho N^{\rho-1}}$$

$$+ \sum_{k=2}^{n} \Sigma_n^* \frac{1}{\sigma_1^{2n} N^{2n}} \Sigma_i^* \Sigma_{i'}^* \prod_{j=1}^{k} E Z_{i_j}^{n_j} Z_{i'_j}^{n_j}$$

where, in the last inequality, we used Markov's inequality, and note that

$$\sigma_1^n \geq s\sigma_0^n / (4 \sum_{i=2}^{n} i^n) \geq \sigma_0^n / 4(n + 1)^n .$$

Therefore, in the remainder of the proof, we bound the expectation

$$\frac{1}{\sigma_1^{2n} N^{2n}} \Sigma_i^* \Sigma_{i'}^* \sum_{j=1}^{k} E Z_{i_j}^{n_j} Z_{i'_j}^{n_j} \quad .$$

For $n_1 \leq \ldots \leq n_k$, let k_i be the number of j's such that $n_j = i$. The final step is carried out for three distinct cases respectively.

[Case I: $2n_k \leq \rho$] Then, the expectation contains no moments of order higher than ρ. Moreover, $i_j \neq i'_j$ ($0 < j \leq k_1$) implies $E Z_{i_j}^{n_j} Z_{i'_j}^{n_j} = 0$ and the number of non-zero moments is at most

$$\sum_{i=0}^{k-k_1-1} (N - i)^2 \sum_{i=k-k_1}^{k-1} (N - i) \leq N^{2k-k_1} \quad .$$

Therefore

$$\frac{1}{\sigma_1^{2n} N^{2n}} \Sigma_i^* \Sigma_{i'}^* \sum_{j=1}^{k} E Z_{i_j}^{n_j} Z_{i'_j}^{n_j}$$

$$\leq \frac{(E|Z_1|^\rho)^{2n/\rho}}{\sigma_1^{2n} N^{2n-2k+k_1}}$$

$$\leq \frac{2^{2n}(EU^\rho)^{2n/\rho}}{\sigma_1^{2n} N^n} \quad ,$$

where the last inequality follows since $n = \Sigma_i ik_i$.

[Case II: $2n_k > \rho$ and $2n_{k-1} \leq \rho$] Then, the summation splits into two terms:

$$\frac{1}{\sigma_1^{2n} N^{2n}} \Sigma_i^* \Sigma_{i'}^* \chi[\, i_k \neq i_k' \,] \prod_{j=1}^{k} E\, Z_{i_j}^{n_j} Z_{i_j'}^{n_j}$$

$$+ \frac{1}{\sigma_1^{2n} N^{2n}} \Sigma_i^* \Sigma_{i'}^* \chi[\, i_k = i_k' \,] \prod_{j=1}^{k} E\, Z_{i_j}^{n_j} Z_{i_j'}^{n_j}$$

In the first term, there is no mement with order higher than ρ, and hence it is bounded as Case I. On the other hand, the second term contains only n^{2k-k_1-1} non-zero summands, each of which has only a moment with order higher than ρ, $EZ_{i_k}^{2n_k}$. The effect of other moments is at most $(E|Z_1|^{\rho})^{2(n-n_k)/\rho}$. Therefore, the second term is

$$\frac{1}{\sigma_1^{2n} N^{2n}} \Sigma_i^* \Sigma_{i'}^* \chi[\, i_k = i_k' \,] \prod_{j=1}^{k} E\, Z_{i_j}^{n_j} Z_{i_j'}^{n_j}$$

$$\leq \frac{1}{\sigma_1^{2n} N^{2n}} N^{2k-k_1-1} E|Z_1|^{2n_k} (E|Z_1|^{\rho})^{2(n-n_k)/\rho}$$

$$\leq \frac{2^{2n}}{\sigma_1^{2n} N^{2n-2k+k_1+1}} EU_1^{2n_k} (EU_1^{\rho})^{2(n-n_k)/\rho}.$$

Now note that, for $2n_k \geq \rho$,

$$EU_1^{2n_k} \leq (\beta N)^{2n_k-\rho} EU_1^{\rho} = (\sigma_0 N)^{2n_k-\rho} EU_1^{\rho} .$$

Thus we have the bound

$$\frac{1}{\sigma_1^{2n} N^{2n}} \Sigma_i^* \Sigma_{i'}^* \prod_{j=1}^{k} E Z_{i_j}^{n_j} Z_{i'_j}^{n_j}$$

$$\leq \frac{2^{2n}}{\sigma_1^{2n} N^{n}} (EU_1^{\rho})^{2n/\rho}$$

$$+ \frac{2^{2n}}{\sigma_0^{2n_k-\rho} \sigma_1^{2n} N^{2n-2n_k+\rho-2k+k_1+1}} (EU_1^{\rho})^{1+2(n-n_k)/\rho}$$

$$\leq \frac{2^{2n} \sigma_0^{n}}{\sigma_1^{2n} N^{\rho}} (EU_1^{\rho})^{\xi}$$

where we used the inequality $2n-2n_k-2k+k_1+1 \geq 0$ and $1 + 2(n-n_k)/\rho \leq 2n/\rho$, $\sigma_0 \geq 1$, and we put $\xi = 2n/\rho$ if $EU_1 \geq 1$ and $\xi = 1$ if $EU_1 < 1$.

[Case III: $2n_k > \rho$ and $2n_{k-1} > \rho$] Then, it should be that $k = 2$ and $n_1 = n_2 = n/2$, and there are only two combinations; $i_1 = i_1'$ and $i_2 = i_2'$ or $i_1 = i_2'$ and $i_2 = i_1'$. Therefore

$$\frac{1}{\sigma_1^{2n} N^{2n}} \Sigma_i^* \Sigma_{i'}^* \prod_{j=1}^{k} E Z_{i_j}^{n_j} Z_{i'_j}^{n_j}$$

$$= \frac{1}{\sigma_1^{2n} N^{2n}} 2N^2 \left(E|Z_1|^n \right)^2$$

$$\leq \frac{1}{\sigma_1^{2n} N^{2n}} 2^{2n+1} N^2 \left(\sigma_0 N \right)^{2(n-\rho)} \left(EU_1^\rho \right)^2$$

$$\leq \frac{2^{2n+1} \sigma_0^n}{\sigma_1^{2n} N^\rho} \left(EU_1^\rho \right)^\xi ,$$

where $\rho > 2$ is used.

By combination of these three results, we have

$$\sum_{k=2}^{n} \Sigma_n^* \frac{1}{\sigma_1^{2n} N^{2n}} \Sigma_i^* \Sigma_{i'}^* \prod_{j=1}^{k} E Z_{i_j}^{n_j} Z_{i'_j}^{n_j}$$

$$\leq \sum_{k=2}^{n} \Sigma_n^* \frac{2^{2n+1} \sigma_0^n}{\sigma_1^{2n} N^\rho} \left(EU_1^\rho \right)^\xi$$

$$\leq \frac{2^{2n+3} (n + 1)^{3n}}{\sigma_0^n N^\rho} \left(EU_1^\rho \right)^\xi$$

where we note $\Sigma_{k=2}^{n} \Sigma_n^* \leq (n+1)^n$. Therefore we have proved the theorem since $2n/\rho \leq 3$.

Proof of Lemma 4.4.2

We use the convention in Section 4.3. According to the arguments there, we have

$$\mathcal{E}EW_{\ell,m,0}^{\rho}$$

$$\leq \mathcal{E}_C E \exp\left[- \frac{\rho}{1+\rho} \Gamma_m \right] \mathcal{E}_I \left\{ \sum_{\underline{u}\in D_0^{\ell}} \exp\left[\frac{1}{1+\rho} \Gamma_{\ell} \right] \right\}^{\rho}$$

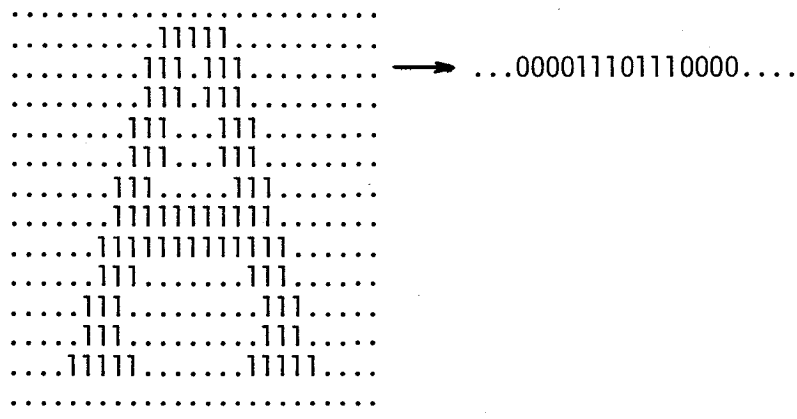
$$= \mathcal{E}_C E \left(\frac{q(\underline{Y}_m)}{P(\underline{Y}_m|\underline{X}_m)} \right)^{\frac{\rho}{1+\rho}} \mathcal{E}_I \left\{ \sum_{\underline{u}\in D_0^{\ell}} \left(\frac{P(\underline{Y}_{\ell}|\underline{X}_{\ell})}{q(\underline{Y}_{\ell})} \right)^{\frac{1}{1+\rho}} \right\}^{\rho}$$

$$\times \exp\left[\frac{\rho\nu}{1+\rho} (m - \ell) R \right] .$$

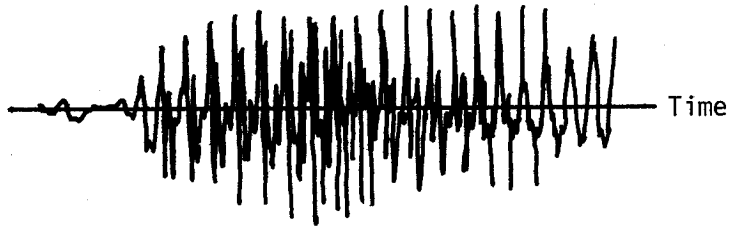
Note that the extreme right-hand side is just the ρ-th power of the summand in the right-hand side of (4.3.4). Thus the lemma is immediate.

Chapter V

Source Coding Preliminaries



(a) Facsimile



(b) Speech

Fig. 5.1.1 — Sources With Memory

1. Notations and preliminaries

In this and the subsequent chapters, we discuss source coding problems especially for sources with memory. In contrast to channel coding problems where DMC's play dominant roles, memoryless sources, sources whose outputs are iid random variables, are less significant in source coding. Important sources like speech signals or facsimile signals are never considered to be memoryless. Because of the memory in sources, we sometimes require complicated mathematics. Thus we start with notations and some preliminaries.

Let A and B be any finite alphabets. Denote n-length sequences $a_1 \ldots a_n$ consisting of letters from A by $\underline{a}_1^n$, and let A^n be the set of all $\underline{a}_1^n$. If we write simply as $\underline{x}$, we mean a doubly infinite sequence $\ldots x_{-1}x_0x_1 \ldots$, $x_n \in A$, and denote the set of all $\underline{x}$ by $\underline{A}$. For each $\underline{x} \in \underline{A}$ and each $m \leq n$, let $\underline{x}_m^n$ and $\underline{x}^n$ be subsequences $\underline{x}_m \ldots x_n$ and $\ldots x_{n-1}x_n$ respectively. And, for each $\underline{a}_1^n \in A^n$ and each m, a cylinder set $c_{m+1}^{m+n}(\underline{a}_1^n)$ is a subset of $\underline{A}$ such that $\underline{x}_{m+1}^{m+n} = \underline{a}_1^n$. (Of course occasional deviations are made to avoid tedious expressions if there is no ambiguity.) These definitions and notations equally apply to sequences with alphabet B, and then symbols b or w are used instead of a or x.

Let $\mathcal{A}$ and $\mathcal{B}$ be the Borel fields over $\underline{A}$ and $\underline{B}$, respectively, generated by all cylinder sets. Then $(\underline{A},\mathcal{A})$, $(\underline{B},\mathcal{B})$, and $(\underline{A}\times\underline{B},\mathcal{A}\times\mathcal{B})$ are all measurable spaces. Thus, given probability measures μ, η, and ω, respective spaces $(\underline{A},\mathcal{A},\mu)$, $(\underline{B},\mathcal{B},\eta)$, and $(\underline{A}\times\underline{B},\mathcal{A}\times\mathcal{B},\omega)$ are all probability spaces, where measurablity of all sets in $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{A}\times\mathcal{B}$ by μ, η, and ω, respectively, is always assumed. Given the probability spaces, a statement in $(\underline{A},\mathcal{A})$ is said to hold for μ-almost every $\underline{x}$ (with symbolic expression μ-a.e. $\underline{x}$) if the subset of $\underline{A}$ consisting of all $\underline{x}$ that make the statement invalid has μ-measure zero (completeness, namely, that all subsets contained in a measurable set having zero measure are also measurable and have zero measure , is assumed). Respective notations η-a.e. $\underline{w}$ and ω-a.e. $(\underline{x},\underline{w})$ are defined in the same manner.

An n-th coordinate function X_n is an $\mathcal{A}$-measurable function defined by $X_n(\underline{x}) = x_n$ for all $\underline{x} \in \underline{A}$, and let $\underline{X} = \ldots X_{-1}X_0X_1 \ldots$ and $\underline{X}_m^n = X_m \ldots X_n$ for $m \leq n$. (The coordinate functions are termed random variables in the previous chapters.) The quadruplet $(\underline{A},\mathcal{A},\mu,\underline{X})$ specifies a stochastic process, which is called a source and is denoted by $[\underline{X},\mu]$. For the sequence of coordinate functions $\underline{W}$ in $(\underline{B},\mathcal{B})$, we also have two processes $[\underline{W},\eta]$ and $[(\underline{X},\underline{W}),\omega]$, which are called a code generation process and a joint source respectively. For later

convenience, we sometimes use the notation $[\underline{Y},\eta]$ for the code generation process instead of $[\underline{W},\eta]$: the latter is preferred if the code generation process is used by itself, not as a part of the joint source.

For each measure μ and each n, a pmf is given by $\mu(\underline{a}_1^n) = \mu[c_1^n(\underline{a}_1^n)]$ and a conditional pmf is given by $\mu(a_n|\underline{a}_1^{n-1}) = \mu(\underline{a}_1^n)/\mu(\underline{a}_1^{n-1})$, for $\underline{a}_1^n \in A^n$ such that $\mu(\underline{a}_1^{n-1}) > 0$, where $\underline{a}_m^n$ is interpreted as void whenever $n < m$. A conditional pmf $\mu(X_1|\underline{X}^0)$ is suitably defined for μ-almost every $\underline{x}$ as well. These definitions are also valid for η and ω. The distinction between pmf's and measures will be clear from the situation they appear.

A shift T in $(\underline{A},\mathcal{A})$ is an operation that shifts coordinates as $\underline{X}_n(T\underline{x}) = x_{n+1}$ for $\underline{x} \in \underline{A}$ and as $TE = \{ T\underline{x}, \underline{x} \in E \}$ for $E \in \mathcal{A}$. The same notation T is also used for shifts in $(\underline{B},\mathcal{B})$ and $(\underline{A}\times\underline{B},\mathcal{A}\times\mathcal{B})$.

We say that $[\underline{X},\mu]$ is a <u>stationary</u> if $\mu(TE) = \mu(E)$ for all $E \in \mathcal{A}$, and say that $[\underline{X},\mu]$ is <u>ergodic</u> if $\mu(E) = 1$ or 0 for every invariant set E, $TE = E$. A simple example of stationary ergodic sources is <u>discrete memoryless sources</u> (DMS's), whose pmf's are given by the product probabilities $\mu(\underline{a}_1^n) = \prod_{i=1}^n p(a_i)$ for all $\underline{a}_1^n \in A^n$ using pmf's p on A. Each DMC is denoted by the symbol p, which expresses the characterizing pmf is p.

A channel $[\underline{X},\nu,\underline{W}]$ is a class of probability measures $\nu_{\underline{x}}$ such that $\nu_{\underline{x}}$ is a probability measure in $(B,\mathcal{B})$ for each $\underline{x} \in \underline{A}$ and $\nu_{\underline{x}}(F)$ is an $\mathcal{A}$-measurable function of $\underline{x}$ for each $F \in \mathcal{B}$. Since $\nu_{\underline{x}}$ is a probability measure for each $\underline{x} \in \underline{A}$, we denote respective probabilities by $\nu_{\underline{x}}(w_m^n)$, $\nu_{\underline{x}}(w_n|\underline{w}_m^{n-1})$, and $\nu_{\underline{x}}(w_n|\underline{w}^{n-1})$ for each $(\underline{x},\underline{w}) \in \underline{A}\,\underline{B}$ and each $m \leq n$. If the channel is a DMC P, then $\nu_{\underline{x}}(\underline{w}_m^n) = \prod_{i=m}^{n} P(w_i|x_i)$ for all $(\underline{x},\underline{w})$. We say that $[\underline{X},\nu,\underline{W}]$ is stationary if $\nu_{\underline{x}}(F) = \nu_{T\underline{x}}(TF)$ for all $\underline{x} \in \underline{A}$ and all $F \in \mathcal{B}$. Moreover, we say that the channel is ergodic if, for any ergodic source $[\underline{X},\mu]$, the joint source $[(\underline{X},\underline{W}),\mu\nu]$ is ergodic, where $\mu\nu$ is the measure given by

$$\mu\nu(E\times F) = \int_E \nu_{\underline{x}}(F)\, d\mu(\underline{x})$$

for all $E \in \mathcal{A}$ and all $F \in \mathcal{B}$. If $[\underline{X},\mu]$ and $[\underline{X},\nu,\underline{W}]$ are stationary, then $[(\underline{X},\underline{W}),\mu\nu]$ is always stationary, but may not be ergodic even if the source is ergodic. A sufficient condition for the ergodicity is the output strongly mixing property (cf. Berger [25]):

$$\lim_{n\to\infty} \left| \nu_{\underline{x}}(T^n E \cap F) - \nu_{\underline{x}}(T^n E)\nu_{\underline{x}}(F) \right| = 0$$

for all cylinder sets $E, F \in \mathcal{B}$ and all $\underline{x} \in \underline{A}$.

Apparently, DMC satifies this condition and hence is ergodic as well as stationary.

For a probability measure ω in $(\underline{A}\times\underline{B}, \mathcal{A}\times\mathcal{B})$ with marginals μ and η in $(\underline{A},\mathcal{A})$ and $(\underline{B},\mathcal{B})$ respectively, let

$$H_{\mu}(\underline{X}_1^n) \triangleq - \sum_{\underline{a}_1^n \in A^n} \mu(\underline{a}_1^n) \log \mu(\underline{a}_1^n)$$

$$H_{\omega}(\underline{X}_1^n,\underline{W}_1^n) \triangleq - \sum_{\underline{a}_1^n \in A^n} \sum_{\underline{b}_1^n \in B^n} \omega(\underline{a}_1^n,\underline{b}_1^n) \log \omega(\underline{a}_1^n,\underline{b}_1^n)$$

$$I_{\omega}(\underline{X}_1^n;\underline{W}_1^n) \triangleq - \sum_{\underline{a}_1^n \in A^n} \sum_{\underline{b}_1^n \in B^n} \omega(\underline{a}_1^n,\underline{b}_1^n) \log \frac{\omega(\underline{a}_1^n,\underline{b}_1^n)}{\mu(\underline{a}_1^n)\eta(\underline{b}_1^n)}$$

and, for a code generation process $[\underline{Y},\tilde{\eta}]$, let

$$I_{\omega|\tilde{\eta}}(\underline{X}_1^n;\underline{W}_1^n) \triangleq - \sum_{\underline{a}_1^n \in A^n} \sum_{\underline{b}_1^n \in B^n} \omega(\underline{a}_1^n,\underline{b}_1^n) \log \frac{\omega(\underline{a}_1^n,\underline{b}_1^n)}{\mu(\underline{a}_1^n)\tilde{\eta}(\underline{b}_1^n)} .$$

Moreover, we write

$$i_{\omega}(\underline{x}_1^n;\underline{w}_1^n) = \log \frac{\omega(\underline{x}_1^n,\underline{w}_1^n)}{\mu(\underline{x}_1^n)\eta(\underline{w}_1^n)} \qquad \text{and}$$

$$i_{\omega|\tilde{\eta}}(\underline{x}_1^n;\underline{w}_1^n) = \log \frac{\omega(\underline{x}_1^n,\underline{w}_1^n)}{\mu(\underline{x}_1^n)\tilde{\eta}(\underline{w}_1^n)} ,$$

for each $(\underline{x},\underline{w}) \in \underline{A}\times\underline{B}$. $H_{\mu}(\underline{X}_1^n)$ and $H_{\omega}(\underline{X}_1^n,\underline{W}_1^n)$ are known as

the entropy of $\underline{X}_1^n$ and the entropy of $(\underline{X}_1^n,\underline{W}_1^n)$ respectively, and $I_\omega(\underline{X}_1^n;\underline{W}_1^n)$ is known as the mutual information quantity between $\underline{X}_1^n$ and $\underline{W}_1^n$. We tentatively call $I_{\omega|\tilde{\eta}}(\underline{X}_1^n;\underline{W}_1^n)$ the mutual information quantity between $\underline{X}_1^n$ and $\underline{W}_1^n$ relative to $[\underline{Y},\tilde{\eta}]$. And we call $i_\omega(\underline{x}_1^n;\underline{w}_1^n)$ and $i_{\omega|\tilde{\eta}}(\underline{x}_1^n;\underline{w}_1^n)$ the information densities.

When we write as $\omega(\underline{a}_1^n,\underline{b}_1^n) = \mu(\underline{a}_1^n)P^n(\underline{b}_1^n|\underline{a}_1^n)$ for some conditional pmf P^n defined on $A^n\times B^n$, then we say that $\underline{X}_1^n$ and $\underline{W}_1^n$ are connected by $[\underline{X}_1^n,P^n,\underline{W}_1^n]$ and prefer $I(\mu^n,P^n)$ to $I_\omega(\underline{X}_1^n;\underline{W}_1^n)$. If the source is a DMS p and the channel is a DMC P, then $I(p^n,P^n) = n\, I(p,P)$ for all $n \geq 1$ (cf. Section 2.2).

For these processes, let

$$H_\mu(\underline{X}) \triangleq \lim_{n\to\infty} \frac{1}{n} H_\mu(\underline{X}_1^n) \ ,$$

$$H_\omega(\underline{X},\underline{W}) \triangleq \lim_{n\to\infty} \frac{1}{n} H_\omega(\underline{X}_1^n,\underline{W}_1^n) \ ,$$

$$I_\omega(\underline{X};\underline{W}) \triangleq \lim_{n\to\infty} \frac{1}{n} I_\omega(\underline{X}_1^n;\underline{W}_1^n)$$

$$I_{\omega|\tilde{\eta}}(\underline{X};\underline{W}) \triangleq \liminf_{n\to\infty} \frac{1}{n} I_{\omega|\tilde{\eta}}(\underline{X}_1^n;\underline{W}_1^n)$$

provided that the respective limits exist. It is easy to see that, if $H_\mu(\underline{X})$ and $H_\omega(\underline{X},\underline{W})$ exist, then $I_\omega(\underline{X};\underline{W})$ also exists.

Lemma 5.1.1: If $[(\underline{X},\underline{W}),\omega]$ is stationary, then $H_\mu(\underline{X})$, $H_\omega(\underline{X},\underline{W})$, and $I_\omega(\underline{X};\underline{W})$ are well defined, and if $[\underline{Y},\tilde{\eta}]$ is a stationary finite-order Markov process, then $I_{\omega|\tilde{\eta}}(\underline{X};\underline{W})$ is also given as a limit (, which may be infinite). Moreover, if we let

$$I_\omega(\underline{X};\underline{W}_{-1}^{n}) \triangleq \lim_{\ell,m\to\infty} I_\omega(\underline{X}_{-\ell}^{m};\underline{W}_{-1}^{n}),$$

then

$$I_\omega(\underline{X};\underline{W}) = \lim_{n\to\infty} \frac{1}{n} I_\omega(\underline{X};\underline{W}_{-1}^{n}) ,$$

where all limits exist.

Corollary: For stationary $[(\underline{X},\underline{W}),\omega]$, let

$$I_\omega(\underline{X};W_1|\underline{W}_{-n}^{0}) \triangleq I_\omega(\underline{X};\underline{W}_{-n}^{1}) - I_\omega(\underline{X};\underline{W}_{-n}^{0}).$$

Then, we have

$$\lim_{n\to\infty} I_\omega(\underline{X};W_1|\underline{W}_{-n}^{0}) = I_\omega(\underline{X};\underline{W}) .$$

Proof of Lemma 5.1.1. The first half follows from Theorem 2.5.1. of Gallager [2], and the second half follows from the proof of Theorem 6.1.1 of Pinsker [26].

Remark: It is easy to see that, for stationary $[\underline{X},\mu]$ and $[\underline{X},\nu,\underline{W}]$, we always have

$$I_{\mu\nu}(\underline{X};\underline{W}_1^n) = E_\mu \sum_{\underline{b}_1^n \in B^n} \nu_{\underline{X}}(\underline{b}_1^n) \log \frac{\nu_{\underline{X}}(b_n|\underline{b}_1^{n-1})}{\eta(b_n|\underline{b}_1^{n-1})}$$

and, for an (n-1)th order stationary Markov $[\underline{Y},\tilde{\eta}]$,

$$I_{\mu\nu|\tilde{\eta}}(\underline{X};\underline{W}) = \sum_{\underline{b}_1^n \in B^n} \eta(\underline{b}_1^n) \log \frac{1}{\tilde{\eta}(b_n|\underline{b}_1^{n-1})} - H_{\mu\nu}(\underline{W}|\underline{X}),$$

if $I_{\mu\nu|\tilde{\eta}}(\underline{X};\underline{W}) < \infty$, where $H_{\mu\nu}(\underline{W}|\underline{X}) = H_{\mu\nu}(\underline{X},\underline{W}) - H_\mu(\underline{X})$ and E_μ denotes the expectation operator relative to μ.

The importance of these information-theoretic quantities comes from Shannon-McMillan-Breiman Theorem (see Billingsley [27]):

Lemma 5.1.2: If $[\underline{X},\mu]$ is stationary ergodic, then

$$\lim_{n\to\infty} \frac{1}{n} \log \mu(\underline{X}_1^n) = H_\mu(\underline{X}) \qquad \mu\text{-a.e. } \underline{x} .$$

Corollary: If $[(\underline{X},\underline{W}),\omega]$ is stationary ergodic, and $[\underline{Y},\tilde{\eta}]$ is stationary finite-order Markov, then

$$\lim_{n\to\infty} \frac{1}{n} i_\omega(\underline{X}_1^n;\underline{W}_1^n) = I_\omega(\underline{X};\underline{W}) \quad \omega\text{-a.e. } (\underline{x},\underline{w}) \text{ , and}$$

$$\lim_{n\to\infty} \frac{1}{n}\, i_{\omega|\tilde{\eta}}(\underline{X}_1^n;\underline{W}_1^n) = I_{\omega|\tilde{\eta}}(\underline{X};\underline{W}) \qquad \omega\text{-a.e. } (\underline{x},\underline{w}),$$

where μ and η are marginals of ω on $(\underline{A},\mathcal{A})$ and $(B,\mathcal{B})$ respectively, providede that $I_{\omega|\tilde{\eta}}(\underline{X};\underline{W}) < \infty$.

Now we define distortion. Let d(a,b) be any, but fixed throughout the remainder, nonnegative finite-valued function on A×B with the maximum value d_o. The distortion bewteen $\underline{a}_1^n \in A^n$ and $\underline{b}_1^n \in B^n$ is then given by

$$d(\underline{a}_1^n,\underline{b}_1^n) \triangleq \sum_{i=1}^{n} d(a_i,b_i) ,$$

and the average distortion induced by $[(\underline{X},\underline{W}),\omega]$ is written as

$$d_\omega(\underline{X},\underline{W}) = \limsup_{n\to\infty} \frac{1}{n}\, E_\omega\, d(\underline{X}_1^n,\underline{W}_1^n) ,$$

where E_ω denotes the expectation operator with respect to ω. Apparently, for stationary $[(\underline{X},\underline{W}),\omega]$, $d_\omega(\underline{X},\underline{W}) = E_\omega d(X_1,W_1)$. When we write as $\omega(\underline{a}_1^n,\underline{b}_1^n) = \mu(\underline{a}_1^n)P^n(\underline{b}_1^n|\underline{a}_1^n)$ for some conditional pmf P^n on $A^n\times B^n$, then we use $d(\mu^n,P^n) = E_\omega d(\underline{X}_1^n,\underline{W}_1^n)$.

A block code c^N of rate R with block length N is a set $\{\, \underline{y}(m)_1^N,\ m = 1, \ldots, M \,\}$ of M sequences in B^N with $R = (1/N)\log M$, and the distortion in

coding $\underline{x}_1^N \in A^N$ by the code is

$$d(\underline{x}_1^N, c^N) \triangleq \min_{m=1,\ldots,M} d(\underline{x}_1^N, \underline{y}(m)_1^N) \ .$$

For any $S \in A^N$, we write

$$d(S, c^N) \triangleq \max_{\underline{x}_1^N \in S} d(\underline{x}_1^N, c^N) \ .$$

For every $R > 0$ and every $D > 0$, we say that (D,R) is achievable for the source $[\underline{X},\mu]$, if, for each $\varepsilon > 0$, there exists a code c^N of rate less than $R + \varepsilon$ that yields the average distortion

$$d_\mu(c^N) \triangleq E_\mu \frac{1}{N} d(\underline{x}_1^N, c^N) \leq D + \varepsilon d_o .$$

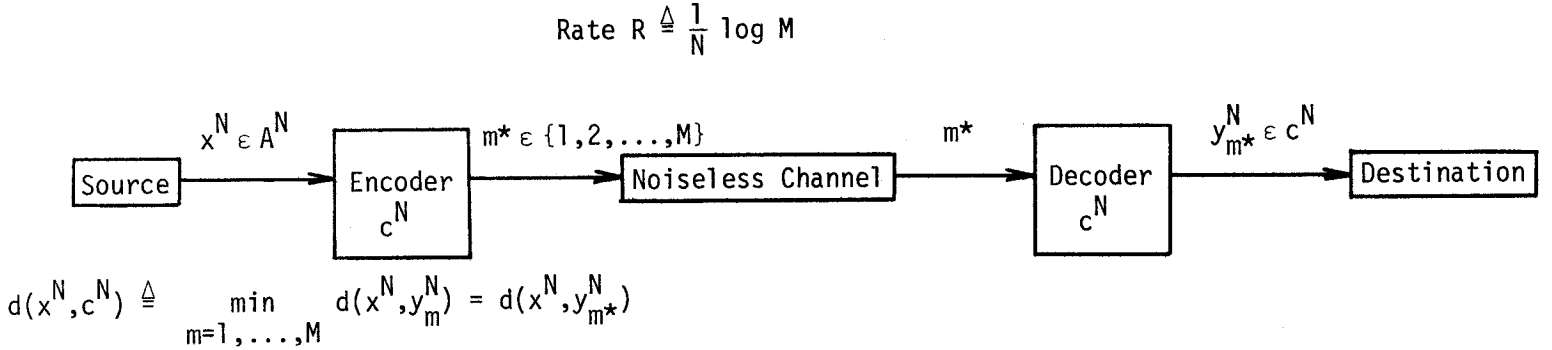


Fig. 5.2.1 — Transmission Of Source Outputs Subject To
A Single Letter Fidelity Criterion

2. Coding theorem for stationary ergodic sources

Source coding with a fidelity criterion concerns how efficiently one can transmit outcomes from a source through a channel capable of carrying them at rates up to its channel capacity (see Fig. 5.2.1). Shannon's source coding theorem suggests that sources have their own effective rates relative to a given fidelity criterion. And, when sources are mathematically described, the relation between the effective rates R and the fidelity D is known to have definite functional forms, called the distortion-rate functions D(R) or rate-distortion functions R(D).

The distortion-rate function $D_\mu(R)$ of a stationary source $[\underline{X},\mu]$ is given by the limit

$$D_\mu(R) \triangleq \lim_{n\to\infty} D_{\mu,n}(R) \ ,$$

where the n-th order distortion-rate function $D_{\mu,n}(R)$ is the minimum of the following information-theoretic optimization over test channels $[\underline{X}_1^n, P^n, \underline{W}_1^n]$:

$$D_{\mu,n}(R) \triangleq \min_{(1/n)I(\mu^n,P^n) \le R} \frac{1}{n} d(\mu^n,P^n).$$

The function $D_\mu(R)$ is a convex, decreasing, continuous function of R, and the limit in the definition is actually the infimum in n. Moreover, for a DMS p,

the distortion-rate function is $D_\mu(R) = D_{\mu,1}(R)$ [$\triangleq D_p(R)$]; the distortion-rate function is obtained by a single minimization over DMC's P.

For an example, let $A = B = \{0,1\}$, and suppose that the source is a binary symmetric source (BSS), $p(0) = p(1) = 1/2$, and the distortion measure is Hamming, $d(a,b) = 0$ if $a = b$, $d(a,b) = 1$ if $a \neq b$. Then the rate-distortion function of the source, $R_p(D)$, is simply $R_p(D) = \log 2 - H(D)$ for $0 \leq D \leq 1/2$, where $H(D) = -D \log D - (1-D)\log(1-D)$.

The above definition well reflects the history of the development of source coding with a fidelity criterion. Historically, a source coding theorem with a fidelity criterion is first proved by Shannon [28], and it asserts that $D_{\mu,1}(R)$ is achievable for any stationary ergodic source $[\underline{X},\mu]$. Given that $D_{\mu,1}(R)$ is achievable, the extension of the achievability to $D_{\mu,n}(R)$, for $n > 1$, seems immediate since the n-th order super source μ', each of whose letters $X_i = X_{i(n-1)+1} \ldots X_{in}$ is n successive letters from the source, has its first order distortion-rate function $D_{\mu',1}(R) = D_{\mu,n}(R)$. however, this trick does not work well; the super source does not necessarily ergodic if the original source is ergodic. Gallager [2] tides over this theoretical difficulty using so-called Nedoma decomposition of stationary ergodic processes:

Nedoma Decomposition: Let $[\underline{X},\mu]$ be a stationary ergodic source and let $[\underline{X}',\mu']$ be the n-th order super source obtained from $[\underline{X},\mu]$. Then there are at most n stationary ergodic sources $[\underline{X}',\mu_i']$, i = 1, ... ,n, with the super alphabet A^n such that

$$\mu'(S) = \frac{1}{n} \sum_{i=1}^{n} \mu_i'(S), \qquad \text{and}$$

$$\mu_i'(TS) = \mu'_{[i+1]}(S) \ , \text{ for } i = 1, \ldots ,n$$

for all $S' \subset \underline{A}'$, where $[i+1] = i+1$ for $1 \leq i < n$ and $[n+1] = 1$, and $\underline{A}'$ is the set of all super sequences $\underline{x}'$.

That is, every output $\underline{x}' \in \underline{A}'$ from the super source comes, with equal probability, from one of ergodic super sources.

In view of this decomposition theorem, a nice trick [2, p.498] allows a coding theorem for stationary ergodic sources with a fidelity criterion:

Theorem 5.2.1: Let $[\underline{X},\mu]$ be a stationary ergodic source. Then, for any $R > 0$, any D satisfying $D \geq D_\mu(R)$, and any $\varepsilon > 0$, there exists a block code c^N of rate at most $R + \varepsilon$ such that

$$d_\mu(c^N) \leq D + \varepsilon d_o.$$

conversely, there is no such a code for $D < D_{\mu}(R)$.

The theorem just asserts that $R_{\mu}(D)$ is the effective rate of the source relative to the fidelity D. However, the definition of $R_{\mu}(D)$ or $D_{\mu}(R)$ assumes the block coding on super sources, which is troublesome hypothesis when we consider tree coding as we discuss in Section 7.2. In the next section, instead, we study a more useful definition of the distortion-rate function.

Chapter VI

Process Approach to Coding Theorems

1. Process definition of $D_\mu(R)$ and a source coding theorem

In the preceeding chapter, we see that $D_\mu(R)$ gives the boundary of achievable distortion-rate region. However, the argument up to the coding theorem for stationary ergodic sources is indirect; Nedoma decomposition is used to extend the coding theorem for memoryless sources up to the one for stationary ergodic sources.

Recently, Gray, Neuhoff, and Omura [29] propose a more direct approach to the coding theorem through a different definition of the distortion-rate function:

$$D_\mu^{(P)}(R) = \inf d_\mu(\underline{X},\underline{W})$$

where the infimum is taken over all stationary ergodic $[(\underline{X},\underline{W}),\omega]$ with the marginal $[\underline{X},\mu]$ on $(\underline{A},\mathcal{A})$ such that

$$I_\omega(\underline{X};\underline{W}) \leq R.$$

Under this definition, proof of coding theorems is greatly simplified, as they show for ergodicity is already a part of the definition. Most importantly, they prove the equivalence of both definitions:

Theorem 6.1.1 – Process Definition : For a stationary ergodic source $[\underline{X},\mu]$ and each $R > 0$,

$$D_\mu(R) = D_\mu^{(P)}(R) \ .$$

In view of the theorem, their definition is called the process definition.

However, the original proof of the process definition assumes known coding theorems and is involved; they require quite mathematical evidences such as the sliding-block codes(see Section 7.1). In this section we give a more elementary proof to the process definition theorem and a more natural proof to the source coding theorem. The arguments contained in the latter enable us to see several features of good source codes; especially they lead to a tree coding theorem in Chapter VII.

Proof of Theorem 6.1.1.

Our proof is based on the argument used to prove the following weak statement due to Gray, Neuhoff, and Omura [29] and Marton [30]:

Theorem 6.1.2: For a stationary source $[\underline{X},\mu]$ and any $R > 0$,

$$D_\mu(R) = \inf d_\omega(\underline{X},\underline{W})$$

where the infimum is taken over all stationary $[(\underline{X},\underline{W}),\omega]$ with the marginal on $(\underline{A},\mathcal{A})$ such that

$$I_{\omega}(\underline{X};\underline{W}) \leq R.$$

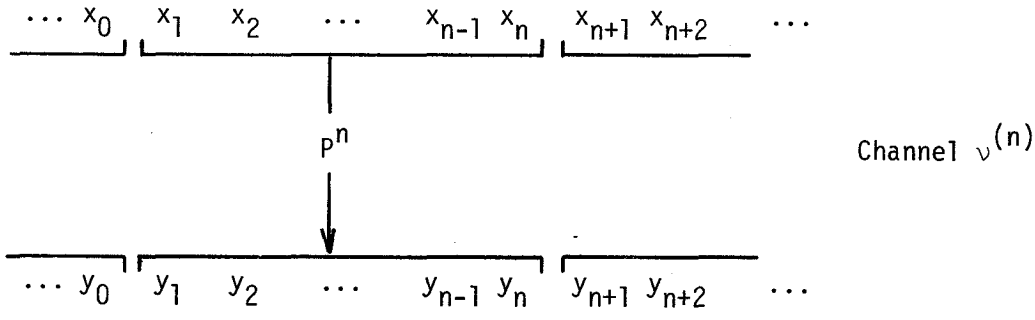
Let $[\underline{X}_1^n,P^n,\underline{W}_1^n]$ be a test channel, and let $[\underline{X},\nu^{(n)},\underline{W}]$ be a test channel defined by independent application of P^n to each $\underline{X}_{in+1}^{(i+1)n}$, $i = \ldots ,-1,0,1, \ldots$ (see Fig. 6.1.1). Generally, $\nu^{(n)}$ is not stationary. On the other hand, if we let $[\underline{X},\nu,W]$ be such that

$$\nu_{\underline{x}}(F) = \frac{1}{n} \sum_{\theta=0}^{n-1} \nu_{\underline{x}}^{\theta}(F) \qquad (6.1.1)$$

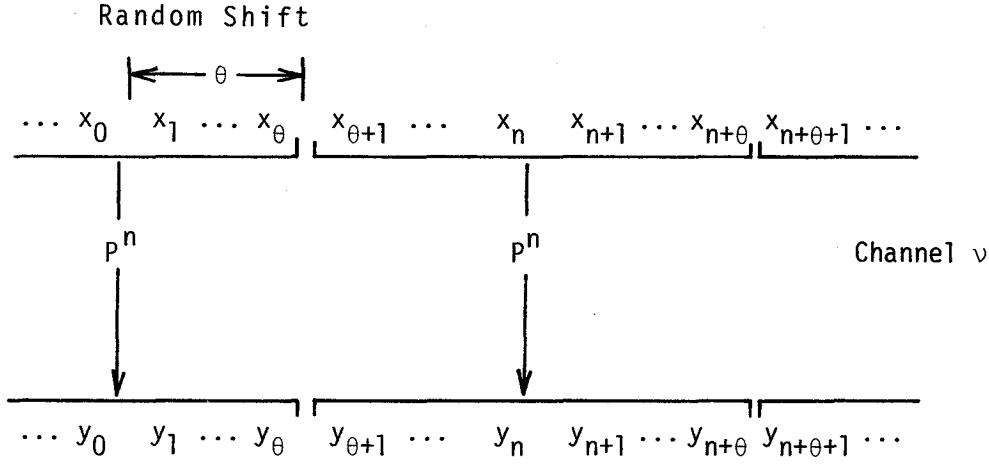
for each $F \in \mathcal{F}$, where $\nu_{\underline{x}}^{\theta}(F) = \nu_{\underline{x}'}^{(n)}(TF)$ for $\underline{x}' = T^{\theta}\underline{x}$, then $[\underline{X},\nu,\underline{W}]$ is stationary (although it may not be ergodic). Intuitively, the channel ν consists of n channels which operate block-wise, and a channel selected from them with the probability 1/n determines the real input-output relationship for each input $\underline{x}$ (cf. Fig. 6.1.1). For ν, it is shown in Appendix that

$$I_{\mu\nu}(\underline{X};W_1|\underline{W}_{-N+2}^{0}) \leq \frac{1}{n} I(\mu^n,P^n) + \frac{1}{N-n} \log n \qquad (6.1.2a)$$

$$d_{\mu\nu}(\underline{X},\underline{W}) = \frac{1}{n} d(\mu^n,P^n) \ . \qquad (6.1.2b)$$



(a)



(b)

Fig. 6.1.1 — Test Channels Acting Block-Wise

In view of Corollary to Lemma 5.1.1, Theorem 6.1.2 is immediate from (6.1.2).

Now we consider a new stationary channel $\tilde{\nu}$ such that

$$\tilde{\nu}_{\underline{x}}(w_{i+1}|\underline{w}^i) = \nu_{\underline{x}}(w_{i+1}|\underline{w}_{i-N+2}^i)$$

for all i and all $\underline{w} \in \underline{B}$. With proper stationary probabilities, the output of the channel becomes an (N-1)th order Markov process for given $\underline{x}$. We can determine its stationary probabilitites so that

$$\tilde{\nu}_{\underline{x}}(\underline{w}_{i-N+2}^i) = \nu_{\underline{x}}(\underline{w}_{i-N+2}^i)$$

for all i. To show it, it is enough to prove that, if the identity holds for some i, then it holds also for i+1: indeed,

$$\begin{aligned}
\tilde{\nu}_{\underline{x}}(\underline{w}_{i-N+3}^{i+1}) &= \sum_{w_{i-N+2} \in B} \tilde{\nu}_{\underline{x}}(w_{i+1}|\underline{w}_{i-N+2}^i)\tilde{\nu}_{\underline{x}}(\underline{w}_{i-N+2}^i) \\
&= \sum_{w_{i-N+2} \in B} \nu_{\underline{x}}(w_{i+1}|\underline{w}_{i-N+2}^i)\nu_{\underline{x}}(\underline{w}_{i-N+2}^i) \\
&= \nu_{\underline{x}}(\underline{w}_{i-N+3}^{i+1}) \ .
\end{aligned}$$

Because of the inequality

$$I_{\mu\tilde{\nu}}(\underline{X};W_1|\underline{W}^0) \leq I_{\mu\tilde{\nu}}(\underline{X};W_1|\underline{W}^0_{-N+2}) = I_{\mu\nu}(\underline{X};W_1|\underline{W}^0_{-N+2}) \ ,$$

Corollary to Lemma 5.1.1 and (6.1.2) imply that

$$I_{\mu\tilde{\nu}}(\underline{X};\underline{W}) \leq \frac{1}{n} I(\mu^n,P^n) + \frac{1}{N-n} \log n \quad , \text{ and} \qquad (6.1.3a)$$

$$d_{\mu\tilde{\nu}}(\underline{X},\underline{W}) = \frac{1}{n} d(\mu^n,P^n) \ . \qquad (6.1.3b)$$

Next, let $\hat{P}^n$ be a channel which achieves $D_{\mu,n}[\ R - (N-n)^{-1}\log n\ ]$, and let

$$P^n(\underline{b}_1^n|\underline{a}_1^n) = (1-\delta)\hat{P}^n(\underline{b}_1^n|\underline{a}_1^n) + \delta\beta^{-n} \quad ,$$

for arbitrary $\delta > 0$, where β is the size of B. Then, by the inequality $(x+y)\log[(x+y)/(u+v)] \leq x \log (x/u) + y \log (y/v)$, for $x,y,u,v > 0$, we have

$$I(\mu^n,P^n) \leq (1-\delta)\ I(\mu^n,\hat{P}^n) \quad , \text{ and}$$

$$d(\mu^n,P^n) \leq (1-\delta)\ d(\mu^n,\hat{P}^n) + n\delta d_o .$$

These inequalities and (6.1.3) imply

$$I_{\mu\tilde{\nu}}(\underline{X};\underline{W}) \leq (1-\delta)R + \frac{1}{N-n} \log n \ , \text{ and} \qquad (6.1.4a)$$

$$d_{\mu\tilde{\nu}}(\underline{X},\underline{W}) \leq (1-\delta)\, D_{\mu,n}\left(R - \frac{1}{n} \log n \right) + \delta d_o \quad . \quad (6.1.4b)$$

From the choice of P^n and $\tilde{\nu}$, we know that

$$\tilde{\nu}_{\underline{x}}(w_i | \underline{w}^{i-1}) \geq \delta\beta^{-n} \quad , \qquad (6.1.5)$$

for all i and all $(\underline{x},\underline{w}) \in A\times B$.

Lemma 6.1.1: Suppose that $[\underline{X},\tilde{\nu},\underline{W}]$ statifies, for some $\rho > 0$,

$$\tilde{\nu}_{\underline{x}}(w_i | \underline{w}^{i-1}) = \tilde{\nu}_{\underline{x}}(w_i | \underline{w}_{i-N+1}^{i-1}) \geq \rho \quad ,$$

for all i and all $(\underline{x},\underline{w}) \in \underline{A}\times\underline{B}$. Then $[\underline{X},\tilde{\nu},\underline{W}]$ is output strongly mixing, and hence it is ergodic.

In view of Lemma 6.1.1, (6.1.4), and (6.1.5), we obtain a theorem:

Theorem 6.1.3: For any stationary ergodic source $[\underline{X},\mu]$ and any $R > 0$, let

$$D_{\mu,n}^{(P)}(R) = \inf d_{\mu\nu}(\underline{X},\underline{W})$$

where the infimum is taken over all stationary ergodic test channels $[\underline{X},\nu,\underline{W}]$ such that

$$I_{\mu\nu}(\underline{X};\underline{W}) \leq R \qquad \text{and}$$

$$\nu_{\underline{x}}(w_i|\underline{w}^{i-1}) = \nu_{\underline{x}}(w_i|\underline{w}_{i-2n+1}^{i-1})$$

for all i and all $(\underline{x},\underline{w}) \in A\times B$. Then, we have

$$D_{\mu,n}^{(P)}(R) \leq D_{\mu,n}\left(R - \frac{1}{n} \log n \right) .$$

Now Theorem 6.1.1 follows from Theorem 6.1.3, the continuity of $D_\mu(R)$, and the inequality

$$D_\mu^{(P)}(R) \leq D_{\mu,n}^{(P)}(R) \leq D_{\mu,n}\left(R - \frac{1}{n} \log n\right)$$

$$\leq D_{\mu,n}(R - \varepsilon) \leq D (R - \varepsilon) + \varepsilon,$$

for any $\varepsilon > 0$ and a sufficiently large n.

Proof of The Source Coding Theorem (Theorem 5.2.1)

According to the process definition, we give a new and simple proof to the source coding theorem for stationary ergodic sources.

For the stationary ergodic source $[\underline{X},\mu]$, we arbitrarily fix $\varepsilon > 0$, and let $[(\underline{X},\underline{W}),\omega]$ be a stationary ergodic joint source with marginals $[\underline{X},\mu]$ and $[\underline{Y},\eta]$ on $(\underline{A},\mathcal{A})$

and $(\underline{B},\mathcal{B})$ respectively such that

$$d_\omega(\underline{X},\underline{W}) \leq D_\mu(R) + \frac{\varepsilon d_o}{12} \qquad \text{and}$$

$$I_\omega(\underline{X};\underline{W}) \leq R + \frac{\varepsilon}{12} \quad .$$

Then, for $\mu(\underline{x}_1^N) > 0$, we have

$$\sum_{\underline{w}_1^N \in B^N} \chi\left[\frac{1}{N} d(\underline{x}_1^N,\underline{w}_1^N) \leq D_\mu(R) + \frac{\varepsilon d_o}{3} \right] \eta(\underline{w}_1^N)$$

$$= \sum_{\underline{w}_1^N \in B^N} \chi\left[\frac{1}{N} d(\underline{x}_1^N,\underline{w}_1^N) \leq D_\mu(R) + \frac{\varepsilon d_o}{3} \right] \frac{\omega(\underline{x}_1^N,\underline{w}_1^N)}{\mu(\underline{x}_1^N)}$$

$$\cdot \exp\left[- i_\omega(\underline{x}_1^N,\underline{w}_1^N) \right]$$

$$\geq \sum_{\underline{w}_1^N \in B^N} \phi(\underline{x}_1^N,\underline{w}_1^N) \frac{\omega(\underline{x}_1^N,\underline{w}_1^N)}{\mu(\underline{x}_1^N)} \; e^{-N(R + \varepsilon/3)} \; ,$$

where

$$\phi(\underline{x}_1^N,\underline{w}_1^N) \triangleq \chi\left[\frac{1}{N} d(\underline{x}_1^N,\underline{w}_1^N) \leq D_\mu(R) + \frac{\varepsilon d_o}{3} \quad \text{and} \right.$$

$$\left. \frac{1}{N} i_\omega(\underline{x}_1^N,\underline{w}_1^N) \leq I_\omega(\underline{X};\underline{W}) + \frac{\varepsilon}{6} \right] .$$

From Corollary to Lemma 5.1.2 and the ergodic theorem,

there exists an integer N_o such that

$$\sum_{\underline{x}_1^N \in A^N} \sum_{\underline{w}_1^N \in B^N} \phi(\underline{x}_1^N, \underline{w}_1^N)\ \omega(\underline{x}_1^N, \underline{w}_1^N) \geq 1 - \frac{\varepsilon}{6} \quad ,$$

for all $N \geq N_o$. Therefore, there exists $S^N \subset A^N$ such that $\mu(S^N) \geq 1 - \varepsilon/3$ and the inequality

$$\sum_{\underline{w}_1^N \in B^N} \phi(\underline{x}_1^N, \underline{w}_1^N)\ \frac{\omega(\underline{x}_1^N, \underline{w}_1^N)}{\mu(\underline{x}_1^N)} \geq \frac{1}{2}$$

holds for all $\underline{x}_1^N \in S^N$. Thus, the inequality

$$\sum_{\underline{w}_1^N \in B^N} \chi\left[\ \frac{1}{N}\, d(\underline{x}_1^N, \underline{w}_1^N) \leq D_\mu(R) + \frac{\varepsilon d_o}{3}\ \right]\ \eta(\underline{w}_1^N)$$

$$\geq \frac{1}{2} \exp\left[\ -N\left(\ R + \frac{\varepsilon}{3}\ \right)\ \right]$$

holds for all $\underline{x}_1^N \in S^N$ and $\mu(S^N) \geq 1 - \varepsilon/3$.

Let $\mathcal{C}^N = \{\ \underline{Y}_1^N(m)\ ,\ m = 1, \ \ldots\ , M\ \}$ be a random block code having block length N and consisting of random codewords independent of each other with the pmf $\{\ \eta(\underline{w}_1^N)\ \}$. Then we have

$$E_{\mathcal{C}}\ d_\mu(\mathcal{C}^N)$$

$$\leq D_\mu(R) + \frac{\varepsilon d_o}{3} + d_o E_\mu E_{\mathcal{C}}\ \chi\left[\ \frac{1}{N}\, d(\underline{X}_1^N, \mathcal{C}^N) \geq D_\mu(R) + \frac{\varepsilon d_o}{3}\ \right]$$

$$= D_\mu(R) + \frac{\varepsilon d_o}{3}$$

$$+ d_o \sum_{\underline{x}_1^N \in A^N} \mu(\underline{x}_1^N)\{ 1 - \sum_{\underline{w}_1^N \in B^N} \chi[\frac{1}{N} d(\underline{x}_1^N, \underline{w}_1^N) \leq D_\mu(R) + \frac{\varepsilon d_o}{3}] \eta(\underline{w}_1^N) \}^M$$

$$\leq D_\mu(R) + \frac{\varepsilon d_o}{3} + \frac{\varepsilon d_o}{3} + d_o \exp[- e^{N(\tilde{R} - R - \varepsilon/2)}] ,$$

for $N \geq N_o$, where we used $1 - x \leq e^{-x}$ for $1 \geq x \geq 0$, and $\tilde{R} = (1/N)\log M$.

Finally, let N abd M sufficiently large so that $\tilde{R} \leq R + \varepsilon$ and $E_{\mathcal{C}} d(\mathcal{C}^N) \leq D_\mu(R) + \varepsilon d_o$. Then we have a code c^N of rate less than $R + \varepsilon$ with distortion less than $D_\mu(R) + \varepsilon d_o$, which proves the theorem for ε is arbitrary.

2. Universal properties of good codes

In Section 3.2 we have seen a universal performance of good channel codes. Analogous properties are also seen in this section. They seem especially useful in source coding since signals through communication link such as the telephony link are seldom stationary or ergodic, rather are varying, for example, from one speaker to another and from one consonant to another in continuous speech.

The first step towards encoding sources without specific knowledge about them is made by Sakrison [31] and Ziv [32]. Ziv [32] shows that, for any rate R, there is a sequence of block codes c_i such that each stationary ergodic source is encoded by c_j, a block code of rate arbitrarily close to R, so that the coding distortion is arbitrarily close to its distortion-rate function. Because of this universal optimality, the sequence is termed as a universal sequence. Later, Davisson [33] classifies these universal sequences into three groups for noiseless source coding, and Neuhoff, Gray, and Davisson [34] extend the classification to sequences in source coding with a fidelity criterion. They are (fixed rate) weighted, weakly-minimax, and strongly-minimax universal sequences respectively.

Let Λ be a class of sources, and let $\{ c_i \}$ be any sequence of block codes, each of which has a rate R_i, such that $R_i \to R$, any positive number, as $i \to \infty$. Then, the sequence is a weighted universal sequence if

$$\int_{\Lambda} d_{\mu}(c_i)\, d\lambda(\mu) \overset{i\to\infty}{\to} \int_{\Lambda} D_{\mu}(R)\, d\lambda(\mu) ,$$

for a given measure λ defined on Λ. The sequence is a weakly-minmax universal sequence if

$$d_{\mu}(c_i) \overset{i\to\infty}{\to} D_{\mu}(R) ,$$

for each $\mu \in \Lambda$. And, the sequence is a strongly-minimax universal sequence if

$$d_{\mu}(c_i) \overset{i\to\infty}{\to} D_{\mu}(R) \quad ; \quad \text{uniformly over } \Lambda.$$

Apparently, they are ordered in increasing significance. In general, strong universality requires the strongest conditions on the class. These universal sequences are sometimes explicitly refered to as fixed rate universal sequences for rates of codes converge to a fixed number the rate of the sequences.

These universal sequences are typically built up

from smaller codes, and there are known two construction methods, one due to Ziv [32] (see also [34]) and the other due to Neuhoff, Gray, and Davisson [34]. Both methods use the notion of code concatenation: the K-th concatenation of a code c^N is a code $\hat{c}^{KN}$ consisting of successions of K codewords from c^N.

If c^N has M codewords, then $\hat{c}^{KN}$ has M^K codewords and the rate $(1/KN)\log M^K = (1/N)\log M$, the rate of c^N. Now, given J codes c_j^N of rate $R = (1/N)\log M$, let c* be the code consisting of all codewords from all concatenated codes $\hat{c}_j^{KN}$. Then c* has JM^K codewords and has the rate $(1/KN)\log JM^K$ which becomes approximately R for sufficiently large K. Therefore, if each subcode c_j^N achieves the distortion-rate bound of a source μ_j approximately, then c* achieves the distortion-rate bounds of all sources μ_j approximately. That is, c* is universally good over μ_j, $j = 1, \ldots, J$.

Ziv argues that, since reproduction alphabet B has only β letters, $J = M^{\beta N}$ codes are sufficient to construct c* for there are only J codes. Neuhoff et al. argue that, since source alphabet A has only α letters, we can approximate all sources with sufficiently many, say J, particular sources and hence that J codes are sufficient.

However, these construction methods are indirect and usually require a lot of subcodes: Ziv's method needs virtually all the possible block codes. In this section, we pursuit different universality.

Quasi-Universal Sequences

For each stationary ergodic source $[\underline{X},\mu]$, each stationary code generation process $[\underline{Y},\eta]$, and each $R > 0$, let

$$D_{\mu,\eta}(R) = \inf d_{\omega}(\underline{X},\underline{W})$$

where the infimum is taken over all stationary ergodic joint sources $[(\underline{X},\underline{W}),\omega]$ with the marginals $[\underline{X},\mu]$ on $(\underline{A},\mathcal{A})$ such that

$$I_{\omega|\eta}(\underline{X};\underline{W}) \leq R \ .$$

We call $D_{\mu,\eta}(R)$ the distortion-rate function of $[\underline{X},\mu]$ relative to $[\underline{Y},\eta]$.

Apparently, the infimum of $D_{\mu,\eta}(R)$ in code generation processes is the distortion-rate function $D_{\mu}(R)$ of the source. However, the continuity or the convexity of $D_{\mu,\eta}(R)$ with respect to R is not generally obvious since the behavior of the relative mutual information quantity $I_{\omega|\eta}(\underline{X};\underline{W})$ is involved

for inappropriate $[\underline{Y},\eta]$. We are not concerned with such geometrical properties of $D_{\mu,\eta}(R)$ here, but show its meanings.

Definition: For $R > 0$, we say that $\{ c^N \}$ is a quasi-universal sequence of rate R (relative to a code generation process $[\underline{Y},\eta]$) if, for each $\varepsilon > 0$ and each stationary ergodic source $[\underline{X},\mu]$, there exists an integer N_o such that each c^N in the sequence has a rate less than $R + \varepsilon$ and satisfies $d_\mu(c^N) \leq D_{\mu,\eta}(R-0) + \varepsilon$, for all $N \geq N_o$, where $D_{\mu,\eta}(R-0) = \lim_{\varepsilon \downarrow 0} D_{\mu,\eta}(R-\varepsilon)$.

The core of the proof of a quasi-universal source coding theorem is the following lemma:

Lemma 6.2.1: For any $[\underline{Y},\eta]$, let $S_{\eta,N}(R,D,\delta)$ be the set of those $\underline{x}_1^N \in A^N$ that satisfy

$$\sum_{\underline{w}_1^N \in B^N} \chi[\frac{1}{N} d(\underline{x}_1^N,\underline{w}_1^N) \leq D + \delta d_o] \eta(\underline{w}_1^N) \geq e^{-NR},$$

where δ, R, and D are any positive numbers and N is any positive integer. Then, for any $\varepsilon > 0$, there exists a block code c^N of rate less than $R + \varepsilon$ having sufficiently large block length N such that

$$d(S_{\eta,N}(R,D,\delta) , c^N) \leq D + 4\delta d_o ,$$

for any $\delta \geq \varepsilon$.

The intuitive meaning of the lemma is simple: all points $\underline{x}_1^N$ such that spheres of radius $D + \delta d_o$ centered at these ponints have η-probability at least e^{-NR} are encoded by the code with distortions at most $D + 4\delta d_o$.

Proof of Lemma 6.2.1. Let $\mathcal{C}^N = \{ \underline{Y}_1^N(m), m = 1, \ldots, M \}$ be a random block code generated using η as in the proof of Theorem 5.2.1 in the prevous section. Let h and k be integers such that $2 > \varepsilon h > 1$ and $2 > \varepsilon k > 1$, and let $D_i = id_o/k$ and $\varepsilon_j = j\varepsilon$ for $i = 0,1, \ldots, k$ and $j = 1, \ldots, h$ respectively. Then, from the union bound, we have

$$E_{\mathcal{C}} \chi\Big[\frac{1}{N} d\big(S_{\eta,N}(R,D,\delta) , \mathcal{C}^N \big) > D + 3\delta d_o ,$$

$$\text{some } \delta \geq \varepsilon \text{ and some } d_o \geq D > 0 \Big]$$

$$\leq \sum_{j=1}^{h} \sum_{i=0}^{k} \sum_{\underline{x}_1^N \in S_{i,j}} E_{\mathcal{C}} \chi\Big[\frac{1}{N} d(\underline{x}_1^N, \mathcal{C}^N) > D_i + \varepsilon_j d_o \Big]$$

$$\leq \sum_{j=1}^{h} \sum_{i=0}^{k} \sum_{\underline{x}_1^N \in S_{i,j}} \exp\big[-Me^{-NR} \big]$$

$$\leq \big(4\alpha^N/\varepsilon^2 \big) \exp\big[-e^{(\hat{R} - R)} \big] ,$$

where α is the size of A, $R = (1/N)\log M$, $S_{i,j} = S_{\eta,N}(R,D_i,\varepsilon_j)$, and we used $D + 3\delta d_o \geq D_i + \varepsilon_j d_o$ whenever $D_i \geq D > D_{i-1}$ and $\varepsilon_j \geq \varepsilon > \varepsilon_{j-1}$. Thus, for sufficiently large N and M, there exists a block code c^N of rate less than $R + \varepsilon$ such that

$$\frac{1}{N} d(S_{\eta,N}(R,D,\delta), c^N) \leq D + 3\delta d_o + \varepsilon d_o,$$

which proves the lemma.

In this lemma we show the existence of a code based only on knowledge about the source output sequences; every $\underline{x}_1^N$ in $S_{\eta,N}(R,D,\delta)$ is encoded with distortion approximately D. We can see that each stationary ergodic source $[\underline{X},\mu]$ emits $\underline{x}_1^N$ contained in $S_{\eta,N}(R,D,\delta)$ with large probability if $D_{\mu,\eta}(R) < D$ and N is sufficiently large. We show the following theorem.

Theorem 6.2.1 (Quasi-universal source coding theorem): Let $[\underline{Y},\eta]$ be any stationary finite-order Markov process, and let $R > 0$. Then, there exists a quasi-universal sequence of rate R relative to $[\underline{Y},\eta]$.

Proof. Let $\{ \varepsilon_i \}$ be a decreasing sequence of positive numbers such that $\lim_{i\to\infty} \varepsilon_i = 0$, and let c_i be codes with block length N_i obtained from Lemma

6.2.1 for $\varepsilon = \varepsilon_i$. For the proof, it is enough to show that, for any $\delta > 0$ and any $[\underline{X},\mu]$, there exists an integer N_o such that $\mu(S_{\eta,N}[R,D_{\mu,\eta}(R),\delta]) \geq 1 - \delta$ for $N \geq N_o$, since $d_\mu(c_i) \leq D_{\mu,\eta}(R) + 5\delta d_o$ holds for $\varepsilon_i \leq \delta$ and $N_i \geq N_o$ then.

For an arbitrary $\delta > 0$, let δ' and $[(\underline{X},\underline{W}),\omega]$ be, respectively, a positive number and a stationary ergodic joint source such that

$$I_{\omega|\eta}(\underline{X};\underline{W}) \leq R - 2\delta' \qquad \text{and}$$

$$d_\omega(\underline{X},\underline{W}) \leq D_{\mu,\eta}(R - 0) + \delta d_o/2 \ .$$

First we have, for $\mu(\underline{x}_1^N) > 0$,

$$\sum_{\underline{w}_1^N \varepsilon B^N} \chi[\frac{1}{N} d(\underline{x}_1^N,\underline{w}_1^N) \leq D + \delta d_o] \, \eta(\underline{w}_1^N) \, e^{NR}$$

$$\geq \sum_{\underline{w}_1^N \varepsilon B^N} \psi(\underline{x}_1^N,\underline{w}_1^N) \, \frac{\omega(\underline{x}_1^N,\underline{w}_1^N)}{\mu(\underline{x}_1^N)} \, e^{\delta' N}$$

where $D = D_{\mu,\eta}(R-0)$ and

$$\psi(\underline{x}_1^N,\underline{w}_1^N) = \chi[\frac{1}{N} d(\underline{x}_1^N,\underline{w}_1^N) \leq d_\omega(\underline{X},\underline{W}) + \delta d_o/2 \quad \text{and}$$

$$\frac{1}{N} i_{\omega|\eta}(\underline{x}_1^N;\underline{w}_1^N) \leq I_{\omega|\eta} + \delta'] \ .$$

In view of Corollary to Lemma 5.1.2 and the ergodic theorem, there exists $S^N \in A^N$ such that $\mu(S^N) \geq 1 - \delta$ and

$$\sum_{\underline{w}_1^N \in B^N} \psi(\underline{x}_1^N, \underline{w}_1^N) \frac{\omega(\underline{x}_1^N, \underline{w}_1^N)}{\mu(\underline{x}_1^N)} \geq \frac{1}{2}$$

for all $\underline{x}_1^N \in S^N$ and all sufficiently large N (so large that $e^{\delta' N} \geq 2$ as well). Therefore

$$\mu(S_{\eta,N}(R,D,\delta))$$

$$\geq \sum_{\underline{x}_1^N \in A^N} \chi[\sum_{\underline{w}_1^N \in B^N} \psi(\underline{x}_1^N, \underline{w}_1^N) \frac{\omega(\underline{x}_1^N, \underline{w}_1^N)}{\mu(\underline{x}_1^N)} e^{\delta' N} \geq 1] \mu(\underline{x}_1^N)$$

$$\geq 1 - \delta,$$

which completes the proof.

In Theorem 6.2.1, the code generation process $[\underline{Y},\eta]$ is assumed finite-order Markov. We generalize the result in the next theorem.

Theorem 6.2.2: For each stationary $[\underline{Y},\eta]$ and any $R > 0$, there exists a sequence of stationary finite-order Markov processes $[\underline{Y},\eta(n)]$ such that

$$\liminf_{n\to\infty} D_{\mu,\eta(n)}(R) \leq D_{\mu,\eta}(R - 0)$$

for each stationary ergodic $[\underline{X},\mu]$. Moreover, all $[\underline{Y},\eta(n)]$ can be made ergodic.

Corollary: For each stationary ergodic source $[\underline{X},\mu]$ and each $R > 0$, there exists a sequence of stationary (ergodic) finite-order Markov $[\underline{Y},\eta(n)]$ such that

$$\liminf_{n\to\infty} D_{\mu,\eta(n)}(R) = D_{\mu}(R) .$$

Proof of Theorem 6.2.2. For each $[\underline{X},\mu]$, let $\delta > 0$ be arbitrary, and let $[(\underline{X},\underline{W}),\omega]$ be a stationary ergodic joint source with the marginal $[\underline{X},\mu]$ on $(\underline{A},\mathcal{A})$ such that

$$d_{\omega}(\underline{X},\underline{W}) \leq D_{\mu,\eta}(R - \delta) + \delta \quad \text{and}$$

$$I_{\omega|\eta}(\underline{X};\underline{W}) \leq R .$$

Let each $\eta(n)$ be an n-th order Markov process with stationary probabilities $\eta(\underline{b}_1^{n+1})$, $\underline{b}_1^{n+1} \in B^{n+1}$ for each n. Then, from the remark below Lemm 5.1.1, we have

$$I_{\omega|\eta}(\underline{X};\underline{W})$$

$$\geq \liminf_{n\to\infty} \sum_{\underline{a}_1^n \in A^n} \sum_{\underline{b}_1^n \in B^n} \omega(\underline{a}_1^n, \underline{b}_1^n) \log \frac{\omega(b_n | \underline{b}_1^{n-1}, \underline{a}_1^n)}{\eta(b_n | \underline{b}_1^{n-1})}$$

$$= \liminf_{n\to\infty} I_{\omega|\eta(n)}(\underline{X};\underline{W}) \ .$$

Thus, for the set $\mathcal{N}$ of integers on which the above limit-supremum is a limit, we have

$$\liminf_{n\to\infty} D_{\mu,\eta(n)}(R)$$

$$\leq \liminf_{n\in\mathcal{N}} D_{\mu,\eta(n)}(I_{\omega|\eta(n)}(\underline{X};\underline{W}) - \delta)$$

$$\leq D_{\mu,\eta}(R - \delta) + \delta.$$

This proves the first half since δ is arbitrary. To prove the latter half , let each $\tilde{\eta}(n)$ be an n-th order stationary ergodic Markov process with transition probabilities $\tilde{\eta}(b_{n+1}|\underline{b}_1^n) = (1-\varepsilon)\eta(b_{n+1}|\underline{b}_1^n) + \varepsilon\beta^{-1}$ where ε is any positive number and β is the size of B. It is easy to see that [see below (6.1.3)]

$$I_{\omega|\eta(n)}(\underline{X};\underline{W}) \geq (1-\varepsilon) I_{\omega|\eta(n)}(\underline{X};\underline{W}) \ .$$

Therefore the above arguments also hold for $\tilde{\eta}$ with a slight modification.

Discussion

Finally, we give several remarks on the universal and quasi-universal sequences. The statements in the quasi-universal coding theorems are weaker versions of weakly-minimax universal coding theorems in [34]. However, if a practical problem, code construction, is involved, the situation seems to change; we have to consider the performance in moderate circumstances, moderate block length and moderate encoder complexity. Elementary code generation units, such as a convolutional encoder, linear block encoder (using linear codes), or more likely speech encoder which is treated later, generate codewords having particular characteristics: statistical independence between letters in linear codes and autoregressive characteristics in speech encoders. Thus ordinary universal encoders are best constructed by assembling suitably selected code generation units.

On the other hand, practical encoders can not possess so much sub-units because of cost performance balance. Thus, theorems in this section will be useful – a code generation process corresponds to a code generation unit.

3. Coding of stationary nonergodic sources

In this section the quasi-universal source coding theorems are used to prove directly the source coding theorem for stationary nonergodic sources [35]. First we see what are stationary nonergodic sources.

For each $\underline{x} \in \underline{A}$, we denote, by $\mu_{\underline{x}}$, a proability measure induced by $\underline{x}$ as the limit of relative frequencies $f^N(\underline{a}_1^n)$ of $\underline{a}_1^n \in A^n$ in subsequences $\underline{x}_{-N}^N$ as $N, n \to \infty$. Of course, they may not be well defined for some $\underline{x}$. However, $\mu_{\underline{x}}$ is well defined for μ-almost every $\underline{x}$, if the source is stationary, and $\mu_{\underline{x}} = \mu$ if the source is ergodic as well. Nonergodic sources give measures $\mu_{\underline{x}}$ which vary also randomly;

Theorem 6.3.1 (Ergodic Decomposition Theorem [36]): There exists an invariant set $G \in \mathcal{A}$, and, for each $\underline{x} \in G$, there associates a stationary ergodic measure $\mu_{\underline{x}}$ such that, for any bounded $\mathcal{A}$-measureable function h on $\underline{A}$, the integral $\int_{\underline{A}} h(\underline{x})\, d\mu_{\underline{x}}(\underline{x})$ is an $\mathcal{A}$-measurable function of $\underline{x}$ on G, and

$$\int_{\underline{A}} h(\underline{x})\, d\mu(\underline{x}) = \int_G \left[\int_{\underline{A}} h(\underline{x})\, d\mu_{\underline{x}}(\underline{x}) \right] d\mu(\underline{x})$$

for any stationary $[\underline{X},\mu]$. Moreover, if $[\underline{X},\mu]$ is stationary and ergodic, then $\mu_{\underline{x}} = \mu$ for μ-almost every $\underline{x}$.

For each $\underline{x} \in G$, we denote the distortion-rate function and the distortion-rate function relative to the process $[\underline{Y},\eta]$, respectively, of $[\underline{X},\mu_{\underline{x}}]$ by $D_{\underline{x}}(R)$ and $D_{\underline{x},\eta}(R)$. Moreover, we denote the expected distortion, by $d_{\underline{x}}(c)$, when a code c is used to encode a stationary ergodic source $[\underline{X},\mu_{\underline{x}}]$. We show the following theorem.

Theorem 6.3.2: For any stationary source $[\underline{X},\mu]$ and any $R > 0$,

$$\int_G D_{\underline{x}}(R)\, d\mu(\underline{x}) = \inf \int_G D_{\underline{x},\eta}(R)\, d\mu(\underline{x})$$

where the infimum is over all stationary ergodic finite-order Markov $[\underline{Y},\eta]$, and the right-hand side is achievable by block codes.

Proof The first statement is a consequence of $D_{\underline{x}}(R) \leq D_{\underline{x},\eta}(R)$ and the next lemma.

Lemma 6.3.1: For any stationary $[\underline{X},\mu]$, $R > 0$, and $\varepsilon > 0$, there exists a stationary ergodic finite-order Markov $[\underline{Y},\eta]$ such that

$$\int_G D_{\underline{x},\eta}(R)\, d\mu(\underline{x}) \leq \int_G D_{\underline{x}}(R - \varepsilon)\, d\mu(\underline{x}) + \varepsilon.$$

To prove the last statement, let $\varepsilon > 0$ be arbitrary, and let $[\underline{Y},\eta]$ be a stationary ergodic finite-order Markov

process given in Lemma 6.3.1. Then, the quasi-universal source coding theorem (Theorem 6.2.1) implies that there exists a quasi-universal sequence of block codes of rate R, $\{ c_i \}$, such that

$$\limsup_{i\to\infty} d_{\underline{x}}(c_i) \leq D_{\underline{x},\eta}(R) \leq D_{\underline{x}}(R - \varepsilon) + \varepsilon \quad .$$

for μ-almost every $\underline{x} \in \underline{A}$. Therefore, from the bounded convergence theorem and the ergodic decomposition theorem (Theorem 6.3.1), we have

$$\begin{aligned} \limsup_{i\to\infty} d_{\mu}(c_i) &= \limsup_{i\to\infty} \int_G d_{\underline{x}}(c_i)\, d\mu(\underline{x}) \\ &\leq \int_G \limsup_{i\to\infty} d_{\underline{x}}(c_i)\, d_{\mu}(\underline{x}) \\ &\leq \int_G D_{\underline{x}}(R - \varepsilon)\, d\mu(\underline{x}) + \varepsilon, \end{aligned}$$

which proves the theorem since the right-hand side is continuous in R and ε is arbitrary.

Conclusion

Gray and Davisson have shown in their noted paper [35] that $\int_G D_{\underline{x}}(R)\, d\mu(\underline{x})$ is achievable by block codes and that, if the noiseless channel (in Fig. 5.2.1) can transmit exactly one out of e^{NR} different codewords

in each block transmission, then any block codes never serve with strictly less coding distortion than this integral. In this sense, $\int_G D_{\underline{x}}(R)\, d\mu(\underline{x})$ is said to be the fixed-rate distortion-rate function of the source. If the source is ergodic, then the integral agrees with the ordinary distortion-rate function.

Originally, the source coding theorem for stationary nonergodic sources is considered as the consequences of universal coding theorems. We have shown, in this section, that the quasi-universal coding theorems also afford a coding theorem for stationary nonergodic sources.

Appendix to Chapter VI

Proof of The Inequalities (6.1.2)

First note the following bound

$$I_{\mu\nu}(\underline{X};W_0|\underline{W}_{-N+1}^{-1})$$

$$\leq I_{\mu\nu}(\ (\underline{X},\Theta)\ ;W_0|\underline{W}_{-N+1}^{-1})$$

$$= H_{\mu\nu}(W_0|\underline{W}_{-N+1}^{-1}) - H_{\mu\nu}(W_0|\underline{W}_{-N+1}^{-1},\underline{X},\Theta)\ ,$$

where Θ is a random variable assuming values $\theta = 0, \ldots, n-1$ with equal probability, the first inequality follows from a generalization of the equation (2.3.17) in Gallager [2, p.26], and (from the choice of ν)

$$H_{\mu\nu}(W_0|\underline{W}_{-N+1}^{-1},\underline{X},\Theta)$$

$$= \frac{1}{n}\sum_{\theta=0}^{n-1} H_{\mu\nu^\theta}(W_0|\underline{W}_{-N+1}^{-1},\underline{X})$$

$$= \frac{1}{n}\sum_{\theta=0}^{n-1} H_{\mu\nu^\theta}(W_0|\underline{W}_{\theta-n+1}^{-1},\underline{X}_{\theta-n+1}^{\theta}).$$

Then it is easy to see that the right-hand side of the above bound is decreasing in N for $N \geq n$.
To bound further this bound, we note the followings:

$$\frac{1}{n}\sum_{\theta=0}^{n-1}\left[\, H_{\mu\nu\theta}(W_0|\underline{W}_{\theta-n+1}^{-1}) - H_{\mu\nu\theta}(W_0|\underline{W}_{\theta-n+1}^{-1},\underline{X}_{\theta-n+1}^{\theta})\,\right]$$

$$= \frac{1}{n}\sum_{\theta=0}^{n-1} I_{\mu\nu\theta}(\underline{X}_{\theta-n+1}^{\theta};W_0|\underline{W}_{\theta-n+1}^{-1})$$

$$= \frac{1}{n}\sum_{\theta=0}^{n-1} I_{\mu\nu 0}(\underline{X}_{-n+1}^{0};W_{-\theta}|\underline{W}_{-n+1}^{-\theta-1})$$

$$= \frac{1}{n} I_{\mu\nu 0}(\underline{X}_{-n+1}^{0};\underline{W}_{-n+1}^{0})$$

$$= \frac{1}{n} I(\mu^n,P^n),$$

$$H_{\mu\nu}(W_0|\underline{W}_{-N+1}^{-1}) - H_{\mu\nu}(W_0|\underline{W}_{-N+1}^{-1},\Theta)$$

$$= I_{\mu\nu}(\Theta;W_0|\underline{W}_{-N+1}^{-1}),$$

and, for $N > n$,

$$\frac{1}{n}\sum_{\theta=0}^{n-1} H_{\mu\nu\theta}(W_0|\underline{W}_{\theta-n+1}^{-1}) - H_{\mu\nu}(W_0|\underline{W}_{-N+1}^{-1},\Theta)$$

$$= \frac{1}{n}\sum_{\theta=0}^{n-1}\left[\, H_{\mu\nu\theta}(W_0|\underline{W}_{\theta-n+1}^{-1}) - H_{\mu\nu\theta}(W_0|\underline{W}_{-N+1}^{-1})\,\right]$$

$$\geq 0 .$$

Therefore we have a bound on $I_{\mu\nu}(\underline{X};W_0|\underline{W}_{-N+1}^{-1})$:

$$I_{\mu\nu}(\underline{X};W_0|\underline{W}_{-N+1}^{-1}) \leq \frac{1}{n} I(\mu^n,P^n) + I_{\mu\nu}(\Theta;W_0|\underline{W}_{-N+1}^{-1}) .$$

Since

$$\sum_{N=1}^{\infty} I_{\mu\nu}(\Theta;W_0|\underline{W}_{-N+1}^{-1}) = \sum_{N=1}^{\infty} I_{\mu\nu}(\Theta;W_N|\underline{W}_1^{N-1}) \leq \log n ,$$

and since $I_{\mu\nu}(\underline{X};W_0|\underline{W}_{-N+1}^{-1})$ is decreasing for $N \geq n$, we have shown the inequality (6.1.2a) [if we shift the coordinates]. (6.1.2b) is obvious.

Proof of Lemma 6.1.1

For $S = B^{N-1}$, let $P_{n,t,s}$ be a transition pmf such that $P_{n,t,s} = \underline{x}(w_n \underline{w}_{n-N+1}^{n-1})$ for $t = (w_{n-N+2} \cdots w_n) \in S$ and $s = (w_{n-N+1} \cdots w_{n-1}) \in S$ and $P_{n,t,s} = 0$ otherwise

For $S = B^{N-1}$, let P_n be a transition pmf such that $t = (w_{n-N+2} \cdots w_n) \in S$ and $s = (w_{n-N+1} \cdots w_{n-1}) \in S$ and $P_{n,t,s} = 0$ otherwise, and let

$$q_{n,t} = \sum_{s \in S} P_{n,t,s} \, q_{n-1,s} \quad ,$$

for $n = 1,2, \ldots$ and arbitrary pmf's p_0 and q_0 on S. Let $v_{n,t} = p_{n,t} - q_{n,t}$ for each $t \in S$. Then $\sum_{s \in S} v_{n,s} = 0$ for $n = 0,1, \ldots$, and

$$v_{n,t} = \sum_{s \in S} P_{n,t,s} \, v_{n-1,s} \, .$$

Our purpose is to show that $v_{n,t}$ eventually converges to 0 for all $t \in S$ as $n \to \infty$. To show this, let S_n be the set of all $s \in S$ that satisfy $v_{n,s} \geq 0$, and let

$$v_n^+ = \sum_{s \in S_n} v_{n,s} \, ,$$

$$P_{n,s}^+ = \sum_{t \in S_n} P_{n,t,s} \, , \quad \text{and}$$

$$P_{n,s}^- = \sum_{t \notin S_n} P_{n,t,s} \, .$$

Then we have

$$v_n^+ = \sum_{s\in S_{n-1}} P_{n,s}^+ \, v_{n-1,s} + \sum_{s\notin S_{n-1}} P_{n,s}^+ \, v_{n-1,s}$$

$$\leq \left(\max_{s\in S_{n-1}} P_{n,s}^+ \right) v_{n-1}^+ - \left(\min_{s\notin S_{n-1}} P_{n,s}^+ \right) v_{n-1}^+ .$$

We can see that the condition in the lemma implies

$$\max_{s\in S_{n-1}} P_{n,s}^+ - \min_{s\notin S_{n-1}} P_{n,s}^+ \leq 1 - \rho\beta ,$$

where β is the size of B. Therefore we have $v_n^+ \leq (1 - \rho\beta) v_{n-1}^+$, and

$$\lim_{n\to\infty} \sum_{s\in S} | p_{n,s} - q_{n,s} |$$

$$= \lim_{n\to\infty} 2v_n^+ = 0 ,$$

which proves the lemma, since we have $v_{n,t} = p_{n,t} - q_{n,t}$ for all $t \in S$ whatever the initial conditions p_0 and q_0 are.

Proof of Lemma 6.3.1

Let $S \in \mathcal{B}$ and $\eta_{\underline{w}}$, respectively, be an invariant subset of $\underline{B}$ and stationary ergodic measures corresponding to $\underline{w} \in S$ that are assured by the ergodic decompostion theorem on $(\underline{B},\mathcal{B})$. Let $\delta > 0$ be arbitrary. For each $\underline{x} \in G$, the invariant subset of $\underline{A}$, let $[(\underline{X},\underline{W}),\omega(\underline{x})]$ be a stationary ergodic joint source with the marginal $[\underline{X},\mu_{\underline{X}}]$ on $(\underline{A},\mathcal{A})$ such that

$$I_{\omega(\underline{x})}(\underline{X};\underline{W}) \leq R - \delta \quad \text{and}$$

$$d_{\omega(\underline{x})}(\underline{X},\underline{W}) \leq D_{\underline{X}}(R - 2\delta) + \delta.$$

Let $[\underline{Y},\eta(\underline{x})]$ be a marginal of $[(\underline{X},\underline{W}),\omega(\underline{x})]$ on $(\underline{B},\mathcal{B})$, and let $\tilde{\eta}$ be the measure on $(\underline{B},\mathcal{B})$ given by

$$\tilde{\eta}(F) = \int_G \eta(F \,||\, \underline{x})\, d\mu(\underline{x})$$

for every $F \in \mathcal{B}$ where $\eta(F\,||\,\underline{x})$ is the $\eta(\underline{x})$-measure of F. First, we show that $I_{\omega(\underline{x})|\tilde{\eta}}(\underline{X};\underline{W}) = I_{\omega(\underline{x})}(\underline{X};\underline{W}) = I_{\omega(\underline{x})|\eta(\underline{x})}(\underline{X};\underline{W})$. To see it, we note the following lemma due to Parthasarathy [37, Theorem 2.6].

Lemma 6.3.2: We have, for $\tilde{\eta}$-almost every $\hat{\underline{w}}$,

$$\tilde{\eta}(Y_1|\underline{Y}^0) = \eta_{\hat{\underline{w}}}(Y_1|\underline{Y}^0) \qquad \text{for } \eta_{\hat{\underline{w}}}\text{-a.e. } \underline{w}.$$

In view of the stationarity of $\tilde{\eta}$, we have

$$\int_G \left[\liminf_{N\to\infty} -\frac{1}{N} \sum_{\underline{w}_1^N \in B^N} \eta(\underline{w}_1^N \| \underline{x}) \log \tilde{\eta}(\underline{w}_1^N) \right] d\mu(\underline{x})$$

$$\leq \lim_{N\to\infty} -\frac{1}{N} \sum_{\underline{w}_1^N \in B^N} \tilde{\eta}(\underline{w}_1^N) \log \tilde{\eta}(\underline{w}_1^N)$$

$$= \int_{\underline{B}} \left[-\log \tilde{\eta}(Y_1 | \underline{Y}^0) \right] d\tilde{\eta}(\underline{w})$$

$$= \int_S \int_{\underline{B}} \left[-\log \tilde{\eta}(Y_1 | \underline{Y}^0) \right] d\eta_{\hat{\underline{w}}}(\underline{w})\, d\tilde{\eta}(\hat{\underline{w}})$$

$$= \int_G \int_S \int_{\underline{B}} \left[-\log \eta_{\hat{\underline{w}}}(Y_1 | \underline{Y}^0) \right] d\eta_{\hat{\underline{w}}}(\underline{w})\, d\eta(\hat{\underline{w}} \| \underline{x})\, d\mu(\underline{x}) ,$$

where we use Lemma 6.3.2 in the last equality. Furthermore, if we let $\eta(Y_1 | \underline{Y}^0 \| \underline{x})$ be the conditional pmf induced by $\eta(\underline{x})$, the last term is calculated, using the ergodic decomposition, as follows:

$$\int_G \int_{\underline{B}} \left[-\log \eta(Y_1 | \underline{Y}^0 \| \underline{x}) \right] d\eta(\underline{w} \| \underline{x})\, d\mu(\underline{x})$$

$$= \int_G \lim_{N\to\infty} -\frac{1}{N} \sum_{\underline{w}_1^N \in B^N} \eta(\underline{w}_1^N \| \underline{x}) \log \eta(\underline{w}_1^N \| \underline{x})\, d\mu(\underline{x}) ,$$

where the last equality follows from the stationarity

of $\eta(\underline{x})$. Therefore

$$\int_G \liminf_{N\to\infty} \frac{1}{N} \sum_{\underline{w}_1^N \in B^N} \eta(\underline{w}_1^N \| \underline{x}) \log \frac{\eta(\underline{w}_1^N \| \underline{x})}{\tilde{\eta}(\underline{w}_1^N)} \, d\mu(\underline{x}) = 0,$$

and this implies that

$$\liminf_{N\to\infty} \frac{1}{N} \sum_{\underline{w}_1^N \in B^N} \eta(\underline{w}_1^N \| \underline{x}) \log \frac{\eta(\underline{w}_1^N \| \underline{x})}{\tilde{\eta}(\underline{w}_1^N)} = 0 \; ; \; \mu\text{-a.e. } \underline{x}.$$

Thus, for μ-almost every $\underline{x}$, $I_{\omega(\underline{x})|\tilde{\eta}}(\underline{X};\underline{W})$ is equal to $I_{\omega(\underline{x})}(\underline{X};\underline{W})$. It is immediate that

$$D_{\underline{x},\tilde{\eta}}(R - \delta) \leq D_{\underline{x}}(R - 2\delta) + \delta; \; \mu\text{-a.e. } \underline{x}.$$

Now let $[\underline{Y},\eta(n)]$ be stationary finite-order Markov process such that

$$\liminf_{n\to\infty} D_{\underline{x},(n)}(R) \leq D_{\underline{x},\tilde{\eta}}(R - \delta) \; ; \; \mu\text{-a.e. } \underline{x},$$

whose existence is shown by Theorem 6.2.2.

Then we obtain

$$\liminf_{n\to\infty} \int_G D_{\underline{x},\eta(n)}(R) \, d\mu(\underline{x})$$

$$\leq \int_G \liminf_{n\to\infty} D_{\underline{x},\eta(n)}(R)\, d\mu(\underline{x})$$

$$\leq \int_G D_{\underline{x}}(R - 2\delta)\, d\mu(\underline{x}) + \delta.$$

Therefore, for sufficiently large n, it holds that

$$\int_G D_{\underline{x},\eta(n)}(R)\, d\mu(\underline{x}) \leq \int_G D_{\underline{x}}(R - 2\delta)\, d\mu(\underline{x}) + 2\delta,$$

and the lemma is proved if we let $\varepsilon = 2\delta$.

Chapter VII

Tree Encoding of Sources

I. Introduction

Although source coding theorems assure efficient coding of sources, information-theoretic ideas have not used fully in real data compression systems yet. This is partly because real sources seldom have well defined characteristics, and because mathematical description of coding distortions is usually very defficult. Besides these obstacles, application of block coding so far discussed is sometimes avoided because of much computation in coding (especially in source coding with a fidelity criterion). For application of source coding theory to real situation we need codes with good performance and efficient coding algorithms. In this regard tree codes consitute an important class of source codes as in channel coding. We first develop mathematical basis for tree coding.

Tree coding theorems are well-known for discrete memoryless sources (DMS's) that emit iid output according to pmf's p on the source alphabet A. The first tree coding theorem is due to Jelinek [38] (which has a flaw, and is subsequently corrected by Davis and Hellman [39]). However, the most important of all such coding theorems is the trellis coding theorem due to Viterbi and Omura [40]. Trellis codes assumed therein are tree codes which have a trellis-like

structure as seen in Fig. 4.2.1 (but their codes do not necessarily have an algebraic structure like convolutional codes). They show the theorem:

Theorem 7.1.1: Let p be a DMS and let $R > 0$. Then there exists a (time-varying) trellis code c^N with q branches per a node, ν letters per a branch, and a constraint length K such that

$$d_p(c^N) \leq D_p(R) + \frac{d_1 e^{-\nu K E(R)}}{[\ 1 - e^{\nu \varepsilon E(R)}\ ]^2}$$

where d_1 is a constant, $E(R)$ and ε are positive numbers for $R = (1/\nu)\log q > R_p(R)$, the rate-distortion function, and the block length N is assumed sufficiently large.

In the proof the decoder is supposed to use Viterbi algorithm, an optimal searching algorithm on trellis.

These two theorems show the performance of tree codes when the best paths or codewords are found out. However exhaustive searching, searching the best one by inspection of all codewords, generally suffers from heavy computational loads in source coding, in contrast to those in channel coding. Many alternatives are devised; some enable us to find source coding theorems [41], [42], others do not, but afford efficient coding [43],

[44].

Despite of all these results for DMS's, however, it is rather surprising that no satisfactory coding theorem with a fidelity criterion has been known yet for more general sources, stationary ergodic sources. An exception may be Tan's result [45]:

Theorem 7.1.2: Let $[\underline{X},\mu]$ be a stationary ergodic source and let $R > 0$. Then, for any $\varepsilon > 0$, there exists a tree code c^N with sufficiently many branches per a node (q branches), sufficiently long branch sequences (ν letters), and a sufficiently large block length N such that

$$d_\mu(c^N) \leq D_\mu(R) + \varepsilon$$

where $R = (1/\nu)\log q$ is the rate of the code.

In this theorem, however, large q and ν are indispensable, and make the theorem less interesting for large q and ν generally increase encoder's computational task. There seems to exist nothing of notable advantage for tree codes having long branches over block codes. In the next section we prove a tree coding theorem for satationary ergodic sources using tree codes having

fixed branch length .

Concerning tree codes, we mention here that there exists an elegant mathematical formulation, called sliding-block coding, proposed by Gray [46]. However this formulation for source coding appears to hardly give us sufficient insights into coding at this stage of theory; finding good sliding-block encoders is only carried out by exhaustive simulations [47].

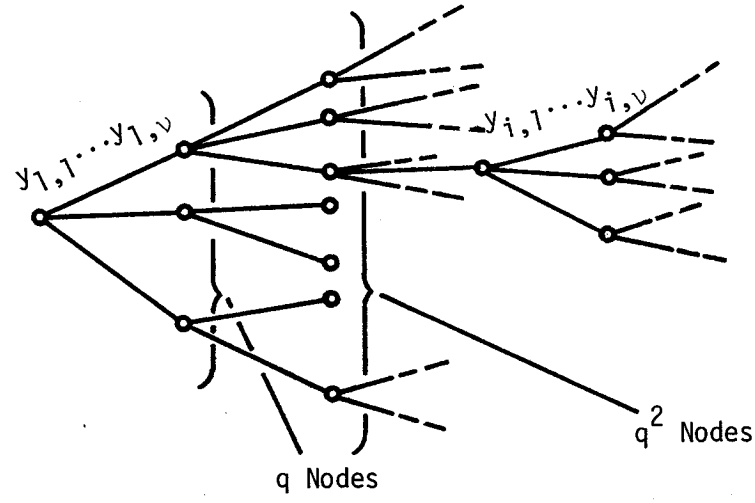


Fig. 7.2.1 — A q-nary Tree Code

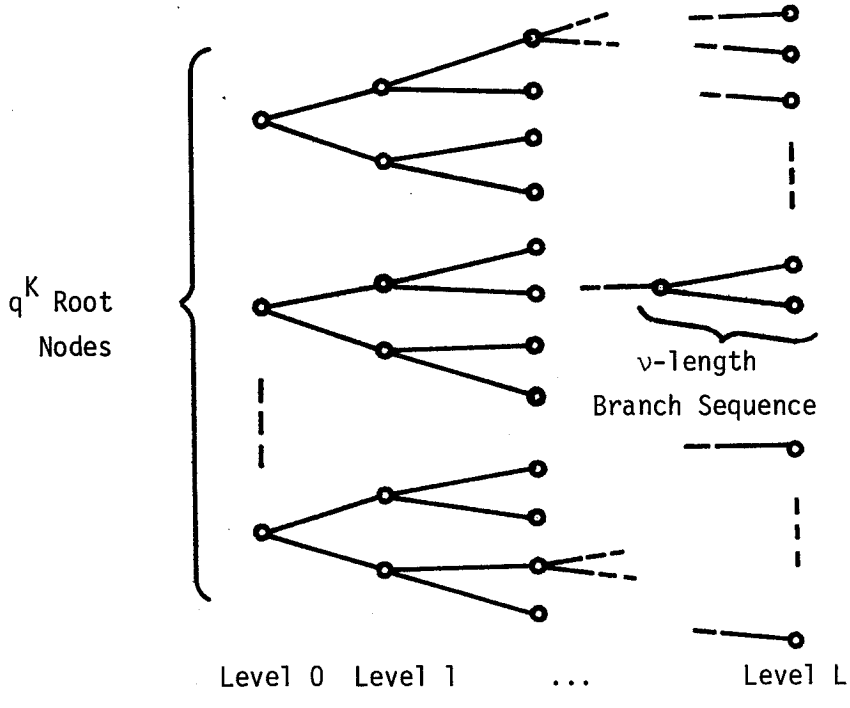


Fig. 7.2.2 — A (K,L)-Tree code

2. Tree encoding of stationary ergodic sources

As we have already seen in the previous chapter, there is a change in the notations; letter x for sources and letter y for codes. Therefore tree codes used in this chapter are represented as shown in Fig. 7.2.1.

Seeing the code tree, the first thought may be how they can encode sources in spite of rather poor number of codewords at first a few branchings; only q branches at the first branching and only q^2 of them even in the next branching. Thus it is a quite natural idea that branches in these first part do not have a significant role in coding.

We say that a code is a (K,L)-tree code if the code has block length N (= νL) and has q^K root nodes (see Fig. 7.2.2), where the length of branch sequences, ν, and the number of branches per a node, q, are all fixed throughout this chapter. Obviously, the (K,L)-tree code is a truncated tree code at the K -th level.

We suppose that N = νL and L = KL*, and divide the code tree into L* parts each having K levels. Then, the first part consists of those from the 0-th level to the K-th level, and the

second one consists of those from the K-th level to the 2K-th level, and so on. The i-th part contains $q^{(i+1)K}$ subsequences, concatenations of K branch sequences connecting the lowest nodes and the highest nodes in this part; one lowest node is connected with q^K highest nodes, the lowest nodes in the next part of the partition. For each lowest node, we number these highest nodes, and hence corresponding subsequences, from 1 to q^K in any order. Then each source output $\underline{x}_1 \ldots \underline{x}_L$ is partitioned accordingly as $\underline{x}^{(1)} \ldots \underline{x}^{(L^*)}$.

Each subsequence in each part of the partition has a distortion relative to the corresponding part of the source output. We call it the weight of the subsequence. The distortion of each path through the tree is then the sum of these weights along it. We use the following searching algorithm: 1) At the first step, q^{2K} candidates in the first part of the partition are classified into q^K groups $c_j^{(1)}$, $j = 1, \ldots, q^K$, so that $c_j^{(1)}$ consists of all subsequences numbered j, and the decoder retains q^K subsequences, call them survivors, each having the smallest weight in $c_j^{(1)}$, $j = 1, \ldots, q^K$: 2) At the ℓ-th step, in general, q^{2K} candidates in the ℓ-th part connected with the previous survivors are classified into q^K groups $c_j^{(\ell)}$, $j = 1, \ldots, q^K$ so that $c_j^{(\ell)}$ consists of all

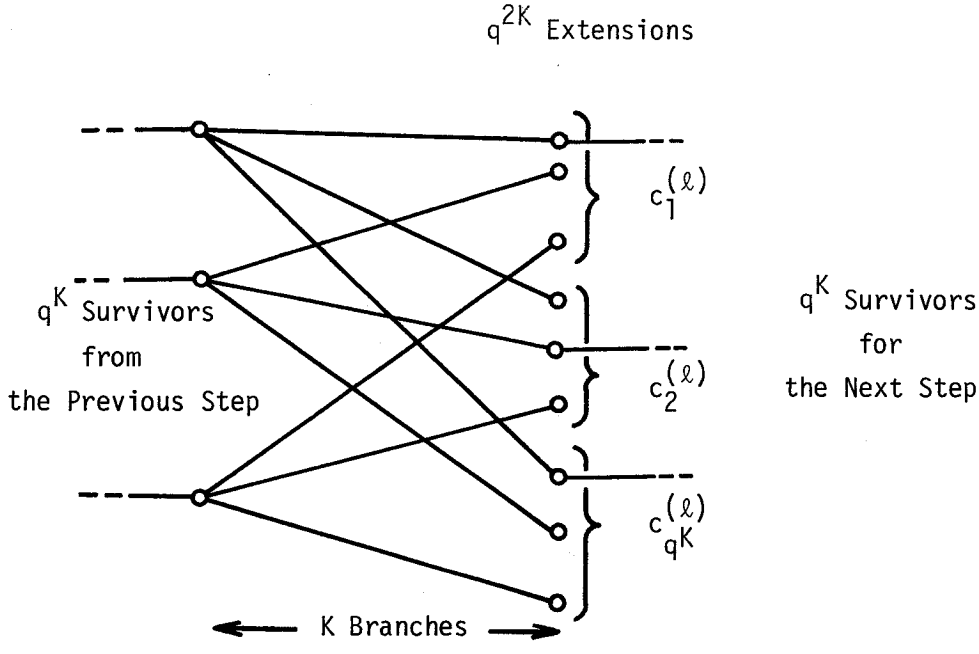


Fig. 7.2.3 — Searching on the (K,L)-Tree Code

subsequences numbered j, and the decoder retains q^K new survivors, each having the smallest weight in a group $c_j^{(\ell)}$, $j = 1, \ldots, q^K$: 3) When the decoder obtained q^K survivors at the last step, then each survivor uniquely specifies a path or a codeword with length N, and the decoder selects the best codeword and emits it. This algorithm resembles to the Viterbi algorithm, although this one is not exhaustive. (See Fig. 7.2.3.)

From the above description of the algorithm, the selected codeword is a concatenation of survivors. Thus the distortion is less than the sum of the respective maximum weights at steps; for a stationary ergodic source $[\underline{X},\mu]$,

$$
\begin{aligned}
N d_\mu(c^N) &\leq \sum_{\ell=1}^{L^*} E \max_{j=1,\ldots,q^K} d(\underline{X}^{(\ell)}, c_j^{(\ell)}) \\
&\leq ND^* + \nu d_o K \sum_{\ell=1}^{L^*} E\, \chi[\, d(\underline{X}^{(\ell)}, c_j^{(\ell)}) > \nu K D^*, \\
&\qquad\qquad \text{some } j = 1, \ldots, q^K \,] \\
&\leq ND^* + d_o K \sum_{j=1}^{q^K} \sum_{\ell=1}^{L^*} E\, \chi[\, d(\underline{X}^{(\ell)}, c_j^{(\ell)}) > \nu K D^* \,].
\end{aligned}
$$

For $R > 0$ and any $\varepsilon > 0$, let $[(\underline{X},\underline{W}),\omega]$ be a stationary ergodic joint source with marginals $[\underline{X},\mu]$ and $[\underline{Y},\eta]$ on $(\underline{A},\mathcal{A})$ and $(\underline{B},\mathcal{B})$ respectively such that

$$d_\omega(\underline{X},\underline{W}) \leq D_\mu(R) + \frac{\varepsilon d_o}{12} \qquad \text{and}$$

$$I_\omega(\underline{X};\underline{W}) \leq R + \frac{\varepsilon}{12} \ .$$

Let $\mathcal{C}^N$ be a random (K,N)-tree code constructed as: 1) Branch sequences on branches emanating from nodes at the ℓK-th level are assigned randomly and independently each other by the pmf $\{ \eta(\underline{w}_1^\nu),\ \underline{w}_1^\nu \in A^\nu \}$: 2) Branch sequences on branches after consecutive i branches ($i \leq K$), each assigned with a branch sequence $\underline{b}_{(j)}^\nu$, $j = 1,\ldots i$, are assigned randomly and independently each other by the conditional pmf $\{ \eta[\underline{w}_{i+1}^\nu | \underline{b}_{(1)}^\nu,\ldots,\underline{b}_{(i)}^\nu] \}$: 3) After K successive branches, branch sequences are selected independently of the previous assignment.

We denote the expectation operator relative to this random tree code by $\mathcal{E}$, and denote the respective groups appeared in the searching by $\mathcal{C}_j^{(\ell)}$.
Then we have

$$\mathcal{E}\, N\, d_\mu(\mathcal{C}^N)$$

$$\leq ND^* + \nu d_o K \sum_{j=1}^{q^K} \sum_{\ell=1}^{L^*} \mathcal{E} E\, \chi[\, d(\underline{X}^{(\ell)}, \mathcal{C}_j^{(\ell)}) > \nu K D^* \,] \,.$$

We note that, though all $\mathcal{C}_j^{(\ell)}$ are not necessarily independent, the random sequences in each $\mathcal{C}_j^{(\ell)}$ are independent of each other and have the same probabilities as $\underline{Y}_1^{\nu K}$. Indeed, each $\mathcal{C}_j^{(\ell)}$ is a random block code with q^K members having block length νK. Therefore we can use, for each $\mathcal{C}_j^{(\ell)}$, one of arguments in Section 6.1, which is stated as follows:

Lemma 7.2.1: For any R, $\varepsilon > 0$ and all sufficiently large K, there exists $S^{\nu K} \in A^{\nu K}$ such that $\mu(S^{\nu K}) \geq 1 - \varepsilon/3$, and

$$\mathcal{E}\, \chi[\, \frac{1}{\nu K} d(\underline{x}^{\nu K}, \mathcal{C}_j^{(\ell)}) > D_\mu(R) + \frac{\varepsilon d_o}{3} \,]$$

$$\leq \exp[\, - q^K e^{-\nu K(R + \varepsilon/2)} \,]$$

for all $\underline{x}^{\nu K} \in S^{\nu K}$, all $j = 1, \ldots, q^K$, and all $\ell = 1, \ldots, L^*$.

Now let $D^* = D_\mu(R) + \varepsilon d_o/3$ and $R + \varepsilon > (1/\nu)\log q > R + 3\varepsilon/4$. Then, according to the lemma, we

obtain, for sufficiently large K,

$$\mathcal{E}\, d_\mu(\mathcal{C}^N)$$

$$\leq D_\mu(R) + \frac{\varepsilon d_o}{3} + \frac{\varepsilon d_o}{3}$$

$$+ d_o q^K \cdot \exp[\, - q^K e^{-K(R+\varepsilon/2)}\,]$$

$$\leq D_\mu(R) + \frac{2\varepsilon d_o}{3} + d_o \exp[\quad K \log q - e^{\varepsilon\nu K/4}\,]$$

$$\leq D_\mu(R^* - \varepsilon) + \varepsilon d_o,$$

where $R^* = (1/\nu)\log q$. Since $D_\mu(R)$ is continuous in R, we have $D_\mu(R^*-\varepsilon) + \varepsilon d_o \leq D_\mu(R^*) + \varepsilon^* d_o$ for any $\varepsilon^* > 0$ if ε is sufficiently small. Therefore we have proved a lemma.

Lemma 7.2.2: Let $[\underline{X},\mu]$ be a stationary ergodic source. Then, for any $\varepsilon^* > 0$, there is a (K,L)-tree code c^N such that

$$d_\mu(c^N) \leq D_\mu(R^*) + \varepsilon^* d_o\ ,$$

where $R^* = (1/\nu)\log q$.

We note that the (K,L)-tree code is a truncated tree at the level K and its true rate is $[(N+\nu K)/N]\log q$, which is greater than R^*. If we use ordinary tree codes with single root nodes, then we have to consider the distortion caused by the first several branchings where only a poor number of codewords exist.

Theorem 7.2.1: For a stationary ergodic source $[\underline{X},\mu]$ and any $\varepsilon^* > 0$, there exists a tree code c^N of rate R with sufficiently large block length N such that

$$d_\mu(c^N) \leq D_\mu(R) + \varepsilon^* d_o + \frac{\nu K d_o}{N} .$$

As the block length gets large, the final term in the bound becomes arbitrarily small. Thus we have shown a tree encoding theorem for a stationary ergodic source with a bounded distortion measure and discrete alphabets.

These results are proved using a result in a random block coding argument, and do not necessarily tell us the superiority of tree codes to block codes. However, once the source coding capability of tree codes is known, we can appropriately modify the searching algorithm to make encoding computation feasible.

In the last chapter, the algorithm used to prove Theorem 7.2.1 is modified, and it is shown that the modified algorithm gives an efficient way of tree coding.

3. Encoding BSS with Hamming distortion measure

An interesting problem, theoretically as well as practically, is the speed of the convergence of distortions to the distortion-rate functions of sources. For a BSS p and Hamming distortion measure, Omura and Shōhara [48] argue that, if optimal codes c^N, either block codes or tree codes, are allowed to maintain rates at least R larger than $R_p(D^*)$ in a positive amount ε, then the convergence should be as fast as doubly exponentials,

$$d_p(c^N) - D^* \leq \exp[- e^{-\varepsilon N}], \qquad (7.3.1)$$

as $N \to \infty$. This assertion is proved for block codes, but only has a simulation evidence for tree codes, although it seems quite probable (also see [49]). In this short section, we observe that this conjecture is true.

For the combination of a BSS p and Hamming distortion measure d, the distortion-rate function is attained by test BSC's for all rates (see Section 5.2). Hence the optimal code generation processes $[\underline{Y},\eta]$ are iid sequences of random variables with the symmetric pmf, $\eta(0) = \eta(1) = 1/2$. From the argument in Section II of [48], Lemma 7.2.1 is strengthened

as follow.

Lemma 7.3.1: For any $\underline{x}^{\nu K} \in A^{\nu K}$,

$$\mathcal{E}\chi\left[\frac{1}{\nu K} d(\underline{x}^{\nu K}, \mathcal{C}_j^{(\ell)}) > D_p(R)\right]$$

$$\leq \exp\{ - q^K e^{- K[R + \delta(\nu K)]} \}$$

for all $j = 1, \ldots, q^K$ and all $\ell = 1, \ldots, L^*$, where $\delta(*)$ is a function such that $\delta(\gamma) \to 0$ as $\gamma \to \infty$.

In view of Lemma 7.3.1, Corollary to Theorem 7.2.1 is replaced by the next theorem, whose proof is omitted for it is almost a repetition.

Theorem 7.3.1: For a BSS p with Hamming distortion measure, any N, and any $R^* > 0$, there is a tree code c^N of rate $R = (1/\nu)\log q$ such that

$$d_p(c^N) \leq D_p(R^*) + \frac{\nu K}{N}$$

$$+ q^K \exp\{ - e^{\nu K[R - R^* - \delta(\nu K)]} \}$$

for any K, where $\delta(*)$ is a function such that $\delta(\gamma) \to 0$ as $\gamma \to \infty$.

If we let $\varepsilon = \nu K/N$ sufficiently small for large N, then the continuity of $D_p(R)$ assures

$D_p(R^*) + \varepsilon \leq D_p(R^*-\varepsilon^*)$ for any positive ε^*. Thus, letting dummy rate R^* sufficiently close to R, we have a corollary.

Corollary: For a BSS p with Hamming distortion measure, every sufficiently large N, and any $\varepsilon^* > 0$, there is a tree code c^N of rate $R = (1/\nu)\log q$ such that

$$d_p(c^N) \leq D_p(R - \varepsilon^*) + \exp[\varepsilon NR - e^{\varepsilon\varepsilon^* N}]$$

where ε is a positive constant.

Therefore the convergence has the higher-than-exponential (almost doubly exponential) speed if a small, but fixed, amount of the excess in rate, ε^*, is allowed.

However, such a positive excess ε^* can not be isolated in real situation. Instead, we ask, for a given rate R and given source, what is the ultimate distortion theoretically attainable, $D_p(R)$, and what is the distortion achievable by practical source encoders, $d_p(c^N)$; we want to know how fast the error

$$d_p(c^N) - D_p(R)$$

converges to zero. This error is bounded by the

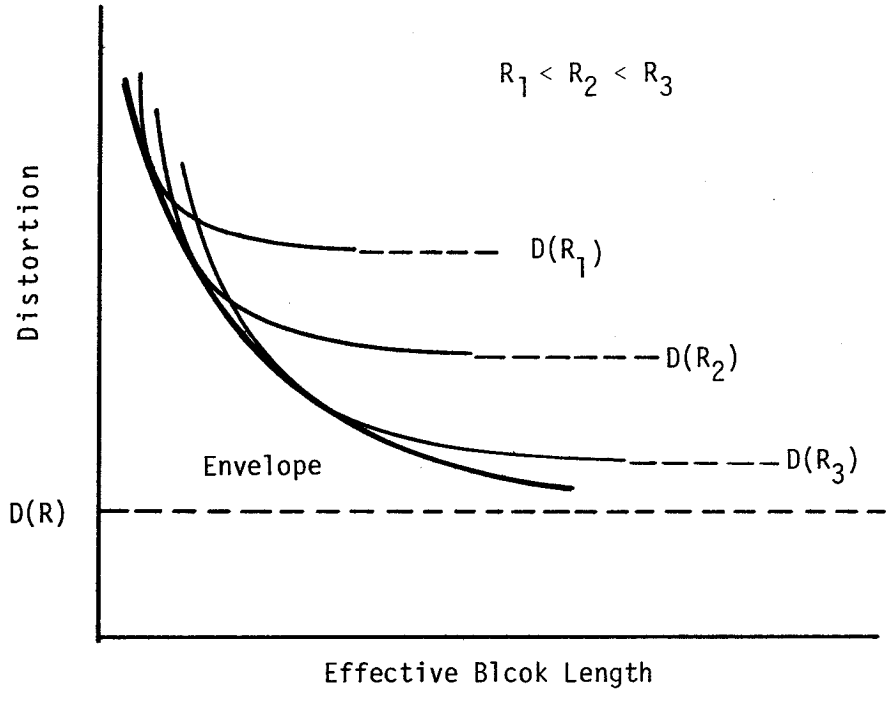


Fig. 7.3.1 — Nominal Convergence Curves and a Practically Meaningful Convergence Curve

envelope of $d_p(c^N) - D_p(R^*)$, $R^* > R$ (see Fig. 7.3.1). We show experimental data* [74] in Fig. 7.3.2 when trellis codes of rate $R = (1/2)\log 2$ with constraint length K are used to encode BSC. Codes are generated randomly, and each plot shows the meanvalue of several tens simulation data. As discussed in Section 4.2, 2K is the effective block length. In this figure, we also depicted the envelope of the doublly exponential convergence

$$d_p(c^{2K}) - D_p(R^*) = \exp[\, -e^{2K(R^* - R)} \,].$$

Though we do not discuss details, the convergence of the error in block coding can not be faster than $(1/2K)\log 2K$ (cf. [25, p.197]). From Fig. 7.3.2, we know that tree codes are really superior to block codes.

---

* By permission of Hiroyoshi Morita

σ, S, and T.

Corollary: For $\rho > 1$, suppose that $E_{sp}(\rho,P) = \hat{E}_{sp}(\rho,P)$. Then, for any $\varepsilon > 0$, the best attainable $\overline{P}_G$ satisfies

$$\frac{\delta_1}{\sigma^{\rho+\varepsilon}(S + T)^{\rho+\varepsilon-1}} \leq \inf_{\substack{\text{conv. codes} \\ (K=\infty)}} \overline{P}_G$$

$$\leq \frac{\delta_2}{\sigma^{\rho-\varepsilon}(S + T)^{\rho-\varepsilon-1}}$$

for sufficiently large σ, S, and T, where the infimum is over all convolutional codes ($K = \infty$) and δ_1 and δ_2 are positive constants independent of σ, S, and T.

Corollary gives a complete answer to the asymptotic behavior of the probability of deficient decoding, when $K = \infty$. For finite constraint length, a similar result will be shown with more elaborate analysis.

Finally we note that all results derived here aplly to time-varying convolutional codes. Since codes used in practice are of time-invariant, another problem thus seems to exist.

Chapter VIII

Tree Encoding of Speech and Speech-Like Sources

1. Introduction

In the previous chapters we have seen that tree codes can encode stationary ergodic sources up to their distortion-rate bounds, bounds on attainable distortion and rate given by distortion-rate or rate-distortion functions. In this section we apply tree codes to encoding speech, a practically important source.

In the practical field, source coding is referred to, from its analog-to-digital conversion, as data compression, and so is speech coding, which is sometimes called speech compression. Speech compression, or speech coding, has long been studied by many engineers. Indeed, it is a major factor in Shannon's developing noted theoretical idea about communication; then he has been with Bell System Laboratry where first rate communication problems have been worked out, capacity of telegram wire, Vocoder, etc. And, among other prblems, speech coding is a serious problem which is continuously increasing its significance. There are so many literatures that we can not list them up all here (cf. [58], [51], and references therein).

From statistical and mechanical evidence [50], speech is best described as the autoregressive-moving average (ARMA) source satisfying

$$X_t = - \sum_{k=1}^{m} a_k X_{t-k} + \sum_{k=1}^{n} b_k X_{t-k} ,$$

where the process $\underline{V}$ is a pulse train, for voiced speech, or is a white noise, a process consisting of iid random variables, for unvoiced speech. However, because of difficulty in identification and analysis, a simpler model, the autoregressive (AR) source satisfying

$$X_t = - \sum_{k=1}^{m} a_k X_{t-k} + V_t$$

is often preferred. (For example, PARCOR Vocoder [52] is based on this model.)

Speech coders, the terminology for speech encoding-decodeing machinery, are divided into principally two classes [51]: waveform coders and source coders. The former coders, as seen from their name, essentially strive for facsimile reproduction of the signal waveform, which are designed, in principle, to be source independent. PCM (Pulse Code Modulation) and DM (Delta Modulation) are in this class. They are used for many data compression systems not confined to speech compression. The latter coders, on the other hand, make use of the knowledge about speech generation mechanism. The idea is that respective fractions of speech offer numerical data on the actions of human vocal tarct and vocal code which turn out to advantage in efficiency describing the signal. Therefore,

the signal must be fitted into a specific mold and parametrized accordingly. This class includes PARCOR Vocoder.

Coders in these two classes have quite distinct features each other in efficiency and human reception. For example, the rates of waveform coders are no less than about 10 kbits/sec, while the rates of source coders are never more than several kbits/sec because of instrument complexity and cost. Moreover, the former coders have relatively natural quality while the latter coders produce sounds less natural and their quality is talker-, or even sentence-, dependent. Therefore, generally speaking, source coders with their extremely low rates can not be good substitutes for coders at higher rates.

Tree coders, speech coders using tree codes first systematically proposed by Anderson [54], constitue a class of efficient waveform coders [59], [60],[61], and are capable of encoding speech at relatively low rates, about 10 $\sim$ 20 kbits/sec. Especially, with sufficient instrumentation [60],[61], tree coders can encode speech at 8 kbits/sec yielding moderate quality. Since the data speeds 7.2 kbits/sec and 9.6 kbits/sec may be avaiable through the conventional telephony link [53] (the recommendation is under investigation in CCITT), tree coders

will be important speech coders.

Speech waveform sampled each unit time has discrete time coordinate, but has continuous magnitudes. In the source coding terminology, speech, as a source, has a continuous alphabet $\mathcal{R}^1$, the real line, in contrast to the sources that we have dealt with in the preceding chapters. Thus we briefly discuss how sources with continuous alphabets treated.

For these sources, source alphabets A and reproduction alphabets B are $\mathcal{R}^1$, and each string $\underline{x}$ from the sources is a point in an infinite-dimensional Euclidean space $\underline{A} = \mathcal{R}^\infty$. Let $[\underline{X},\mu]$ and $[\underline{X},\nu,\underline{W}]$ be a source and channel with continuous alphabets. To avoid unnecessary mathematical subtlety, we suppose that the measure μ and conditional measure ν have, respectively, the density p_μ and conditional density $p_\nu(*|\underline{x})$ for all $\underline{x} \in \underline{A}$ with respect to Lebesgue measures on $\underline{A}$ and $\underline{B}$. Then, according to general definitions of information quantities (cf. [25] and [55]), the mutual information quantity between X_1 and W_1 is the supremum

$$I_{\mu\nu}(X_1;W_1)$$

$$\triangleq \sup \sum_{i,j} \mu\nu(E_i^1 \times F_j^1) \log \frac{\mu\nu(E_i^1 \times F_j^1)}{\mu(E_i^1)\eta(F_j^1)}$$

over all finite partitions $\{E_i^1\}$ and $\{F_j^1\}$ of A and B (they are $\mathcal{R}^1$), where η is the marginal of $\mu\nu$ on $(\underline{B},\mathcal{B})$. Since each finite partition of $\mathcal{R}^1$ induces a finite partition of $\mathcal{R}^n$ in an obvious manner, and since the measures have densities (η automatically possesses its density), the mutual information quantity between $\underline{X}_1^n$ and $\underline{W}_1^n$ is then

$$I_{\mu\nu}(\underline{X}_1^n;\underline{W}_1^n)$$

$$\triangleq \sup \sum_{i,j} \mu\nu(E_i^n\times F_j^n) \log \frac{\mu\nu(E_i^n\times F_j^n)}{\mu(E_i^n)\eta(F_j^n)} \qquad (8.1.1)$$

where the supremum is over all partitions $\{E_i^n\}$ and $\{F_j^n\}$ of A^n and B^n induced by partitions of A and B respectively. Moreover, the right-hand side is actually the integral

$$\int_{A^n}\int_{B^n}\left[\log \frac{p_\mu(\underline{x}_1^n)p_\nu(\underline{w}_1^n|\underline{x}_1^n)}{p_\mu(\underline{x}_1^n)p_\eta(\underline{w}_1^n)}\right]p_\mu(\underline{x}_1^n)p_\nu(\underline{w}_1^n|\underline{x}_1^n)\,d\underline{x}_1^n\,d\underline{w}_1^n .$$

Now let

$$h_\mu(\underline{W}_1^n) \triangleq \int_{B^n}[\log p_\mu(\underline{w}_1^n)]\,p_\mu(\underline{w}_1^n)\,d\underline{w}_1^n \quad \text{and}$$

$$h_{\mu\nu}(\underline{W}_1^n|\underline{X}_1^n) \triangleq$$

$$\int_{A^n} \int_{B^n} [\ \log p_\nu(\underline{w}_1^n|\underline{x}_1^n)\ ]\ p_\mu(\underline{x}_1^n) p_\nu(\underline{w}_1^n|\underline{x}_1^n)\ d\underline{x}_1^n\ d\underline{w}_1^n\ ,$$

and call them the <u>differential entropy</u> and <u>conditional differential entropy</u> respectively. Then the following continuous alphabet analog is obtained for the mutual information quantity:

$$I_{\mu\nu}(\underline{X}_1^n;\underline{W}_1^n) = h_\eta(\underline{W}_1^n) - h_{\mu\nu}(\underline{W}_1^n|\underline{X}_1^n).$$

If we use the backward channel given by

$$p_\xi(\underline{x}_1^n|\underline{w}_1^n) = \frac{p_\mu(\underline{x}_1^n) p_\nu(\underline{w}_1^n|\underline{x}_1^n)}{p_\mu(\underline{x}_1^n) p_\eta(\underline{w}_1^n)}\ ,$$

for each $\underline{x}_1^n \in A^n$ and each $\underline{w}_1^n \in B^n$, then we have another form

$$I_{\mu\nu}(\underline{X}_1^n;\underline{W}_1^n) = h_\mu(\underline{X}_1^n) - h_{\eta\xi}(\underline{X}_1^n|\underline{W}_1^n).$$

And, if we define the differential entropy of the process $\underline{X}$ by the limit (it exists for stationary $\underline{X}$)

$$h_\mu(\underline{X}) \triangleq \lim_{n\to\infty} \frac{1}{n} h_\mu(\underline{X}_1^n)\ ,$$

we eventually obtain formulae for the information quantity between processes $\underline{X}$ and $\underline{W}$:

$$I_{\mu\nu}(\underline{X};\underline{W}) = h_{\mu}(\underline{X}) - h_{\eta\xi}(\underline{X}|\underline{W})$$

$$= h_{\eta}(\underline{W}) - h_{\mu\nu}(\underline{W}|\underline{X}) \ .$$

Again consider the general definition (8.1.1). For each pair of finite partitions $\{E_i^1\}$ and $\{F_j^1\}$, the summation in the right-hand side is regarded as the mutual information quantity across the system depicted in Fig. 8.1.1. Passing through the quantizer in the figure, all identification of $\underline{x}_1^n$ is lost except that $\underline{x}_1^n$ is in some E_i^n. Given the quantizer output $\hat{\underline{x}}_1^n$ (there are only finite number of them), the channel emits an output $\hat{\underline{w}}_1^n$ in F_j^n with the conditional probability $\mu\nu(E_i^n \times F_j^n)/\mu(E_i^n)$, and the average distortion across the channel is

$$\sum_{i,j} \mu\nu(E_i^n \times F_j^n)\ d(\underline{x}_1^n, \underline{w}_1^n) \ .$$

If we suitably choose numeric letters a_i corresponding to E_i^1 and numeric letters b_j corresponding to F_j^1 for all i and j, and let the partition sufficiently fine, then the mutual information quantity across the system is made arbitrarily close to $I_{\mu\nu}(\underline{X};\underline{W})$ and the average distortion is made arbitrarily close to the integral

$$\int_{A^n} \int_{B^n} d(\underline{x}_1^n, \underline{w}_1^n)\ p_\mu(\underline{x}_1^n) p_\nu(\underline{w}_1^n | \underline{x}_1^n)\ d\underline{x}_1^n\ d\underline{w}_1^n$$

$$= d_{\mu\nu}(\underline{X}_1^n, \underline{W}_1^n) \ .$$

Since stationarity and ergodicity are preserved through quantization, we can see that coding problems for continuous alphabets are approximated by those for discrete alphabets arbitrarily well.

In this chapter, we are concerned largely with the practical side of coding rather than the mathematical properties. Thus we avoid measure theoretical terminologies. Instead, we consider coordinate functions X_t as random variables, and distinguish random variables with distinct distributions as X_t and $\hat{X}_t$.

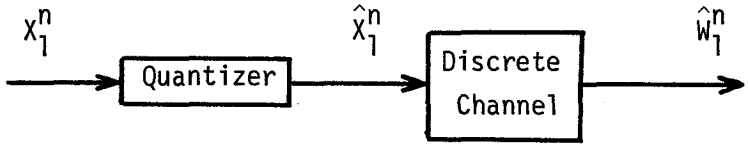


Fig. 8.1.1 — A Discrete Channel Approximation

2. Rate-distortion function of speech-like sources

As seen in Introduction, mathematical models of speech, called speech-like sources, are coded by tree codes up to rate-distortion bounds if the sources are stationary and ergodic. In this section, we discuss the rate-distortion function of AR sources as speech-like sources.

Gaussian AR Sources

AR sources are continuous amplitutde process $\underline{X}$ satisfying the difference equation

$$X_t = - \sum_{k=1}^{m} a_k X_{t-k} + V_t \quad , \text{ for } t = 1,2, \ldots, \tag{8.2.1a}$$

$$X_0 = \ldots = X_{1-m} = 0 \quad , \tag{8.2.1b}$$

where $\underline{V}$, the driving process, consists of iid zero-mean random variables with variance σ^2. When V_t are Gaussian random variables, we call the source the Gaussian AR source. As we see subsequently, the Gaussian source is not necessarily a suitable speech-like source for its probability dencity function does not have as sharp peak at zero amplitude as relative frequencies obtained from actual speech. However, investigation on this source is an important

step toward distortion-rate bounds of speech-like sources and the construction of tree codes for speech.

The behavior of the Gaussian AR source crucially depends on the location of zeros ρ_k, $k = 1, \ldots, m$, of the characteristic polynomial (cf. [25])

$$A(\rho) = 1 + a_1\rho^{-1} + \ldots + a_m\rho^{-m} .$$

Let ρ^* be the maximum magnitude

$$\rho^* \triangleq \max_{k=1,\ldots,m} |\rho_k| ,$$

and denote covariances as

$$\gamma_{t,s} = E\, X_t X_s ,$$

where E is the expectation operator. Then, the source is asymptotically stationary,

$$\gamma_{t,s} \to \gamma_k \quad \text{as } t,s \to \infty ,$$

for each $k = |\, t - s \,|$, if $\rho^* < 1$. On the other hand, the source is nonstationary and has exponentially diverging variances,

$$\gamma_{t,t} = O(\rho^{*2t}) \text{ as } t \to \infty ,$$

if $\rho^* > 1$, and the source is nonstationary and has algebraically diverging variances,

$$\gamma_{t,t} = O(t^{\alpha}) \qquad \text{as } t \to \infty \ ,$$

for some $\alpha > 0$, if $\rho^* = 1$, where $O(*)$ is any function such that $O(\delta)/\delta$ is bounded for large δ. An example of the last case is a Wiener sequence satisfying

$$X_t = X_{t-1} + V_t \ , \quad \text{and} \quad X_{-1} = 0$$

for all $t \geq 0$, with covariances $\gamma_{t,s} = \delta^2 \min(t,s)$. Of course, the most important is asymptotically stationary sources, which are ergodic as well.

Let $R_n(D)$ be the n-th order rate-distortion function of the Gaussian AR source relative to the squared-error distortion measure

$$d(x,w) = (x - w)^2$$

for each x and each w. Then, the rate-distortion function of the source is the limit

$$R(D) = \lim_{n \to \infty} R_n(D) \ ,$$

for $D > 0$, provided that the limit exists. Apparently, $R(D)$ is well-defined if the source is asymptotically stationary. The following block source coding theorem is known [56]:

Theorem 8.2.1: $R(D)$ is achievable (if it exists) using block codes.

The theorem asserts that $R(D)$ is achievable (see Section 5.1 for the terminology) even if the source is nonstationary with $\rho^* \geq 1$. This is a quite exceptional statement among source coding theorems; most of them concern stationary ergodic sources. Therefore we deduce a general formula for $R(D)$ here.

Let A_n be the $n \times n$ matrix

$$A_n = \begin{bmatrix} 1 & & & & \\ a_1 & 1 & & & \\ \vdots & \ddots & \ddots & & \\ a_m & \cdots & a_1 & 1 & \\ & \ddots & & \ddots & \ddots \\ & & a_m & \cdots & a_1 \quad 1 \end{bmatrix}_{n \times n} .$$

Then the covariance matrix $\Gamma_n = [\, \gamma_{s,t} \,]_{n \times n}$ is given by $\sigma^2 [\, A_n^T A_n \,]^{-1}$, and

$$A_n^T A_n = \begin{bmatrix} \alpha_0 & \cdots & \alpha_m & & & \\ \vdots & \ddots & & \ddots & & \\ \alpha_m & & \ddots & & \ddots & \\ & \ddots & & \ddots & & \\ & & \ddots & & \alpha_0 \cdots \alpha_m & \\ & & & & \vdots \quad * & \\ & & & & \alpha_m & \end{bmatrix} \quad n \times n$$

where, putting $a_0 = 1$,

$$\alpha_\ell = \sum_{k=0}^{m-\ell} a_k \, a_{k+\ell} \quad ,$$

for $\ell = 1, \ldots, m$, and A_n^T is the transpose of A_n. $R_n(D)$ is given, in terms of the eigenvalues of $A_n^T A_n$, $\lambda_{n,1} \leq \ldots \leq \lambda_{n,n}$, as

$$D_\theta = \frac{1}{n} \sum_{k=1}^{n} \min \left[\theta \, , \, \frac{\sigma^2}{\lambda_{n,k}} \right] \quad \text{and} \qquad (8.2.2a)$$

$$R_n(D_\theta) = \frac{1}{n} \sum_{k=1}^{n} \max \left[\frac{1}{2} \log \frac{\sigma^2}{\theta \lambda_{n,k}} \, , \, 0 \right] , \qquad (8.2.2b)$$

where θ is a parameter. To calculate R(D), we have to know the asymptotic behavior of eigenvalues $\lambda_{n,k}$.

Let Φ_n be a matrix with the same entries ϕ_ℓ on its upper and lower ℓ-th diagonal, $\ell = 0, \ldots, n-1$ (ϕ_0 on the main diagonal). Such a matrix is called a (finite) Toeplitz matrix. $A_n^T A_n$ is almost

Toeplitz matrix except the lower right m×m corner. Given ϕ_ℓ, $\ell = 0,1, \ldots$, Toeplitz matrices Φ_n have a nice asymptotic eigenvalue distribution:

Theorem 8.2.2 — Toeplitz Distribution Theorem:
Let δ and Δ be the essential infimum and supremum, respectively, of the real-valued function on $[-\pi,\pi]$

$$\Phi(\omega) = \sum_{k=-\infty}^{\infty} \phi_k \, e^{-jk\omega}$$

where $\phi_k = \phi_{-k}$ and $j^2 = -1$. Then, for any function $G(*)$ continuous in $[\delta,\Delta]$,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} G(\xi_{n,k}) = \frac{1}{2\pi} \int_{-\pi}^{\pi} G[\, \Phi(\omega) \,] \, d\omega$$

holds, where $\xi_{n,1} \leq \ldots \leq \xi_{n,n}$ are eigenvalues of Φ_n.

Remark: The theorem implies that the integral

$$\frac{1}{2\pi} \int_{\{\, \Phi(\omega) \leq \delta \,\}} d\omega$$

gives the asymptotic fraction of eigenvalues less than δ.

Now let Ψ_n be the Toeplitz matrix with α_ℓ on the upper and lower ℓ-th diagonal, $\ell = 1, \ldots, n$, for each n, and let $\tilde{\lambda}_{n,1} \leq \ldots \leq \tilde{\lambda}_{n,n}$ be its eigenvalues. From the Sturmian separation theorem [57],

we have

$$\lambda_{n,k} \geq \tilde{\lambda}_{n-m,k-m} \quad ; \quad k = m+1, \ \ldots \ , m \ .$$

Moreover, putting $\alpha_{\ell} = \alpha_{-\ell}$,

$$\lambda_{n,n} \leq \sum_{\ell=-m}^{m} \alpha_{\ell} \ .$$

Therefore we see that all eigenvalues, except the smallest m eigenvalues, have the same asymptotic distribution as $\tilde{\lambda}_{n,k}$.

If the source is asymptotically stationary, it is easy to see that all eigenvalues of Γ_n, $\lambda_{n,k}^{-1} \sigma^2$, are bounded above for large n. This implies that, for large n, the least m eigenvalues $\lambda_{n,k}$ are bounded away from zero, and their contributions to (8.2.2) are negligible for large n. Therefore, for the asymptotically stationary source, we have the parametric representation

$$D_{\theta} = \frac{1}{2\pi} \int_{-\pi}^{\pi} \min \left[\ \theta \ , \ \frac{\sigma^2}{g(\omega)} \ \right] d\omega \qquad \text{and}$$

$$R(D_{\theta}) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \max \left[\ \frac{1}{2} \log \frac{\sigma^2}{\theta g(\omega)} \ , \ 0 \ \right] d\omega \ ,$$

where $g(\omega) = | \ A(e^{-j\omega}) \ |^2$. The function $\sigma^2/g(\omega)$ is a spectral distribution of the source,

Fourier transform of the limit covariance sequence $\{\gamma_k\}$.

However, if the source is nonstationary and $\rho^* \geq 1$, then the speed of the convergence of $\lambda_{n,1} \leq \ldots \leq \lambda_{n,m}$ becomes significant in taking the limit of (8.2.2). We can show the following lemma .

Lemma 8.2.1: Suppose that

$$A(\rho) = \sum_{k=1}^{s} (1 - \rho_k \rho^{-1})^{\ell_k}$$

(ℓ_k is the multiplicity of ρ_k) and that

$$|\rho_1| > \ldots > |\rho_k| \geq 1 > |\rho_{r+1}| > \ldots > |\rho_s| .$$

Then,

$$\lambda_{n,\ell} = \begin{cases} \alpha_{n,\ell,k} \, |\rho_k|^{-2n} \; ; \; \sum_{i=1}^{k-1} \ell_i < \ell \leq \sum_{i=1}^{k} \ell_i \; ; \\ \qquad\qquad\qquad\qquad k \leq r \; , \\ \alpha_{n,\ell,k} \quad ; \text{ otherwise,} \end{cases}$$

where $\alpha_{n,\ell,k}$ are positive numbers decreasing at most algebraically as $n \to \infty$.

From the lemma, the following theorem is immediate .

Theorem 8.2.3: The rate-distortion function R(D) of the Gaussian AR source is represented parametrically as

$$D_\theta = \frac{1}{2\pi} \int_{-\pi}^{\pi} \min \left[\theta , \frac{\sigma^2}{g(\omega)} \right] d\omega \qquad \text{and}$$

$$R(D_\theta) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \max \left[\frac{1}{2} \log \frac{\sigma^2}{\theta g(\omega)} , 0 \right] d\omega$$

$$+ \sum_{k=1}^{m} \max \left[\log |\rho_k| , 0 \right]$$

where all ρ_k are zeros of the characteristic polynomial $A(\rho)$ and

$$g(\omega) = \left| \sum_{k=0}^{m} a_k e^{-jk\omega} \right|^2 , \quad a_0 = 1.$$

The theorem implies that nonstationary sources require additional rates corresponding to their exponential rates of diverging variances. As an illustrative example, consider a nonstationary source

$$X_t = \rho X_{t-1} + V_t \quad ; \; t = 1,., \ldots , \text{ and}$$

$$X_0 = 0,$$

where $\rho > 1$ and $\sigma^2 = 1$. X_t has variances $\gamma_{t,t} = (\rho^{2t}-1)/(\rho^2-1)$. Given X_n, the process proceeds backward as

$$X_{t-1} = - \frac{(\rho^{2t}-1)\rho}{\rho^{2(t+1)}-1} X_t + \sqrt{\frac{\rho^{2t}-1}{\rho^{2(t+1)}-1}} U_t$$

where all U_t are iid Gaussian random variables with unit variance. Let $\tilde{X}_t$ ($t \leq n$) be the conditional expectation of X_t when X_n is known, and let $\hat{X}_t = X_t - \tilde{X}_t$. Then $\underline{\hat{X}}$ approximately satisfies, for large n,

$$\hat{X}_{t-1} = - \frac{1}{\rho} \hat{X}_t + \frac{1}{\rho} U_t \text{ , and}$$

$$\hat{X}_n = 0 .$$

Thus, we see that the output of the source is decomposed into exponentially diverging random variable X_n and an approximately stationary backward process $\underline{\hat{X}}$. Since the rate-distortion function of X_n is $(1/2)\log \gamma_{n,n}/d$ for each $d > 0$, the average contribution of X_n to the total rate-distortion function is $(1/2n)\log \gamma_{n,n}/d \to \log \rho$ as $n \to \infty$. Therefore, if we let d sufficiently small, we have

$$D_\theta \simeq \frac{1}{2\pi} \int_{-\pi}^{\pi} \min \left[\theta , \frac{\rho^{-2}}{|1+\rho^{-1}e^{-j\omega}|^2} \right] d\omega \quad \text{and}$$

$$R(D_\theta) \simeq \frac{1}{2\pi} \int_{-\pi}^{\pi} \max \left[\frac{1}{2} \log \frac{\rho^{-2}}{\theta |1+\rho^{-1} e^{-j\omega}|^2} , 0 \right] d\omega$$

$$+ \log \rho$$

This coincides with the formula in the theorem.

Concerning Theorem 8.2.1 and Theorem 8.2.2, we give a brief comment. For stationary Gaussian AR sources, the rate-distortion function has been known [25]. Berger [58] shows a coding theorem for Wiener sequences and gives a formula of R(D) which is eventually equal to the integral in Theorem 8.2.3, since $\rho^* = 1$. Subsequently Gray proves a coding theorem (Theorem 8.2.1) for general Gaussian AR sources; however the formula for R(D) therein is misled by a wrong argument.

Stationary AR sources

We can see that whether the source is stationary or asymptotically stationary is not essential; asymptotically stationary sources turn out to be stationary sources when infinite time has passed by or suitable initial conditions are selected. Thus, in the latter part, we make no distinction between both sources and simply call them stationary sources.

For a stationary Gaussian AR source, suppose that

$$D_\theta \leq \min_{-\pi \leq \omega \leq \pi} \frac{\sigma^2}{g(\omega)} \triangleq D_0 .$$

Then, Theorem 8.2.3 gives

$$D_\theta = \theta \quad \text{and}$$

$$R(D_\theta) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{1}{2} \log \frac{\sigma^2}{\theta g(\omega)} \, d\omega .$$

By calculation of the integral, we have

$$R(D_\theta) = \frac{1}{2} \log \frac{\sigma^2}{\theta} ,$$

which is exactly the rate-distortion function of $\underline{V}$ for the fidelity θ (cf. [25]). Since $D_\theta \leq \theta$ and $R(D_\theta) \geq (1/2)\log(\sigma^2/\theta)$, for $\theta > 0$, we have a corollary.

Corollary to Theorem 8.2.2: For the stationary Gaussian AR source, the rate-distortion function R(D) of the source is

$$R(D) \geq R_V(D)$$

for all $D \geq 0$, and the equality holds for $D \leq D_0$ where $R_V(D)$ is the rate-distortion function of $\underline{V}$.

This corollary gives us a useful lower bound of R(D): this is a special form of Shannon lower bound (cf. [25]).

Theorem 8.2.3 — Shannon Lower Bound: For a stationary source $\underline{X}$, the rate-distortion function R(D) relative to the squared-error distortion measure has the lower bound

$$\underline{R}(D) \triangleq h(\underline{X}) - \max h(Z)$$

where the maximum is over all random variables Z satisfying $EZ^2 \leq D$.

Remark: This theorem holds for more general distortion measures, the difference distortion measures $d(x,w) = d(x-w)$.

We first note that the maximizing Z does not

depend on sources, and that it is necessarily a Gaussian random variable. For a stationary Gaussian source, the maximizing Z is a zero-mean Gaussian random variable with variance D, and the inequality in Corollary to Theorem 8.2.2 is a direct consequence of Theorem 8.2.3.

It proves to be useful later to investigate what the equality for $D \leq D_0$ means. Let $\underline{Z}$ be a process consisting of iid random variables distributed as the maximizing Gaussian Z, and suppose that there is a process $\underline{W}$ independent of $\underline{Z}$ such that $\underline{X} = \underline{W} + \underline{Z}$. Then we have

$$d(\underline{X},\underline{W}) = D$$

$$I(\underline{X};\underline{W}) = h(\underline{X}) - h(\underline{X}|\underline{W}) = R(D) \ ,$$

where the last equality follows from $h(\underline{X}|\underline{W}) = h(\underline{Z})$ and the corollary. These identities imply that $\underline{W}$ is just an optimal code generation process (for the terminology see Section 5.1). In more elaborate analysis, it is shown [25] that the Shannon lower bound $\underline{R}(D)$ equals R(D) if, and only if, the source output is obtained through the backward channel $\underline{X} = \underline{W} + \underline{Z}$. Corollary to Theorem 8.2.2 means that such an expression is possible if $D \leq D_0$. A similar

statement holds for general AR sources:

Theorem 8.2.4: For a stationary AR source, the rate-distortion function R(D) of the source is

$$R(D) \geq R_V(D)$$

for all $D \geq 0$, and the equality holds for all $D \leq (\delta^*/\sigma^2)D_0$ if $R_V(D)$ equals its Shannon lower bound for $D \leq \delta^*$ ($\leq \sigma^2$), where $R_V(D)$ is the rate-distortion function of $\underline{V}$.

For Gaussian AR source, $R_V(D)$ equals its Shannon lower bound for all $D \leq \sigma^2$, and hence Theorem 8.2.4 coincides with Corollary to Theorem 8.2.3. However, as we have noted below Theorem 8.2.3, for $R_V(D)$ to equal its Shannon lower bound, each V_t should be the sum of a Gaussian random variable and another random variable independent of it: This is quite improbable. In fact, almost speech-like sources do not satisfy this condition, and hence Theorem 8.2.4 is useless for most cases. Instead the following simple theorem is useful .

Theorem 8.2.5: Given a stationary AR source, let R(D) be the rate-distortion function of the source, and let $R_G(D)$ be the rate-distortion function of the stationary Gaussian AR source with the same parameters.

Then,

$$R_G(D) \geq R(D) \geq R_G(D) - [\, h_G(\underline{X}) - h(\underline{X}) \,]$$

for $D \leq D_0$, where $h(\underline{X})$ and $h_G(\underline{X})$ are, respectively, the differential entropy of the original source and the differential entropy of the Gaussian source.

Proof The left-most inequality is a consequence of Theorem 4.6.3 in Berger [25]. On the other hand, from Theorem 8.2.3 and the Gaussianity, there exists a random variable Z such that

$$R(D) \geq h(\underline{X}) - h(Z) \quad \text{and}$$

$$R_G(D) = h_G(\underline{X}) - h(Z) \,,$$

which proves the other inequality.

The bound given by this theorem is tight for most speech-like sources.

3. Tree encoding of speech and speech-like sources

In this section we show some results in coding speech and speech-like sources ..

Tree Codes

Below Theorem 8.3.4, we have noted that, for $D \leq D_0$, the rate-distortion function $R(D)$ of the (asymptotically) stationary Gaussian AR source is attained by the backward channel

$$X_t = W_t + Z_t \quad ,$$

for all t, where $\underline{Z}$ is a process independent of $\underline{W}$ and consisting of iid zero-mean Gaussian random variables with variance σ^2. The optimal code generation process $\underline{W}$ then has the spectral density

$$f_W(\omega) = \frac{\sigma^2}{g(\omega)} - D$$

$$= \sigma^2 \left| \sum_{k=0}^{m} b_k e^{-jk\omega} \right|^2 / \left| \sum_{k=0}^{m} a_k e^{-jk\omega} \right|^2 \quad ,$$

for $-\pi \leq \omega \leq \pi$, where the coefficients b_k are given by the factorization (it is possible for $D \leq D_0$ [25])

$$(D/\sigma^2) \left| \sum_{k=0}^{m} a_k e^{-jk\omega} \right|^2 - 1 = \left| \sum_{k=0}^{m} b_k e^{-jk\omega} \right|^2 \quad (8.3.1)$$

with $a_0 = 1$. Therefore $\underline{W}$ satisfies the ARMA model

$$W_t = - \sum_{k=1}^{m} a_k W_{t-k} + \sum_{k=0}^{m} b_k \tilde{V}_{t-k}$$

for all t where all $\tilde{V}_t$ are iid zero-mean Gaussian random variables with variance σ^2.

A straightforward construction of tree codes for the Gaussian sources is to simulate the random generation of tree codes described in Section 7.2 by computer-generated random process $\underline{W}$. However any tree code obtained in this way bears no superiority to simpler tree codes discussed subsequently; it requires large memory area in spite of the convergence of distortion which is rather slow. This is partly because ideal codes which perform at rates and distortions very near to the rate-distortion bound are not necessarily also good at moderate tree searching capability of encoders. Sometimes a better result is obtained by substitution of fixed number of numeric values $v_1, \ldots, v_q$, instead of randomly generated numbers, for the driving V_t; the totality of available sequences $\underline{w}_1^n$ constitutes, by itself, a tree codes of rate log q having block length n (assuming $w_0 = \ldots = w_{1-m} = 0$).

To obtain simple code design, we first note the

following equivalent definition of the code generation process $\underline{W}$.

$$W_t = \sum_{k=0}^{\infty} c_k \tilde{V}_{t-k}$$

where the coefficients c_k are given by the formal expansion

$$\left(\sum_{k=0}^{m} b_k \rho^{-k} \right) / \left(\sum_{k=0}^{m} a_k \rho^{-k} \right) = \sum_{k=0}^{\infty} c_k \rho^{-k} \quad .$$

We say that a tree code has <u>B-coefficients</u> or is a <u>B-code</u> if the code is generated by

$$y_t = \sum_{k=0}^{K} c_k v_{t-k} \qquad (8.3.2)$$

where K is a finite integer called the constraint length of the code and all v_t assume only fixed number of levels $s_1, \ldots, s_q$. The coefficients b_k can be replaced by other coefficients f_k determined according to different reasoning. For such selection $b_k = f_k$, for all k, we say that the tree code has <u>F-coefficients</u> or is an <u>F-code</u>. Especially, if $f_0 = 1$ and $f_k = 0$ for all $k > 0$, then the code is called a <u>No Smoothing (NS)-code</u>, and its coefficients are called an <u>NS-coefficients</u>, since f_k are generally selected to possess smoothing (filtering) effects

on the behavior of y_t. In this regard, another code design is possible by smoothing NS-coefficients c_k by $F(\rho) = \sum_{k=0}^{K'} f_k \rho^{-k}$ as

$$\sum_{k=0}^{K} c_k \rho^{-k} = F(\rho) \sum_{k=0}^{K''} c_k \rho^{-k}$$

where $K = K' + K'' - 1$. We call the code thus constructed an $\hat{F}$-code, and call its coefficients $\hat{F}$-coefficients. Since the NS-coefficients tend to zero rapidly for large k, F-codes and F-codes eventually become almost equal to each other for large constraint length K, for most cases.

M-algorithm [or (M,L)-algorithm]

Given these tree codes, the next problem is an efficient way of coding capable of finding codewords, or paths, that make the distortion small. Viterbi algorithm which is optimal for convolutional codes or trellis codes is no longer useful for these codes. And sequential algorithms have a significant drawback because of their buffer overflow problems, though they seem attractive in conceivably cheap instrumentation cost [62].

(M,L)-algorithm [43] is then a standard algorithm [54],[59]-[61], which is known to outperform the Viterbi algorithm in distortion when the number of computations

per an encoded node at the encoder is limited [44]. Suppose $q^{\ell} > M \geq q^{\ell-1}$, and call the root node the 0-th encoded node. It is described as follows (see Fig. 8.3.1): 1) First investigate all paths up to the ℓ-th level to find out M paths minimizing distortions between the codewords and the corresponding portion of the source output: 2) Investigate qM (or at most qM) paths extended one branch from these previously retained paths, and sort out M (or at most M) paths with the least distortions: 3) Whenever the decoder reaches nodes higher in L levels than the previously encoded node , sort out a path having the least distortion, let the immediate descendant of the previous encoded node on this path a new encoded node, abondon all nodes except those descendants of the new encoded node, and return to 2).

Due to the last operation that makes the selected path never jump transversally on the tree, the number of retained nodes is occasionaly less than M. Fig. 8.3.2 shows the combined scheme of the tree code generation and (M,L)-algorithm.

Coding of Speech-Like Sources

Two AR sources are chosen as speech-like sources: a Gaussian AR source and a Laplacian AR source whose

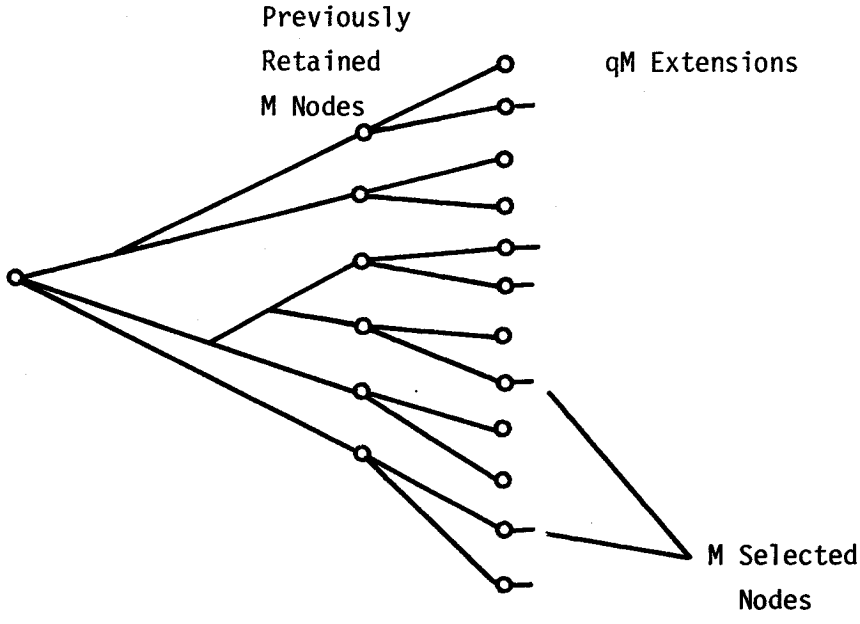


Fig. 8.3.1 — The (M,L)-Algorithm

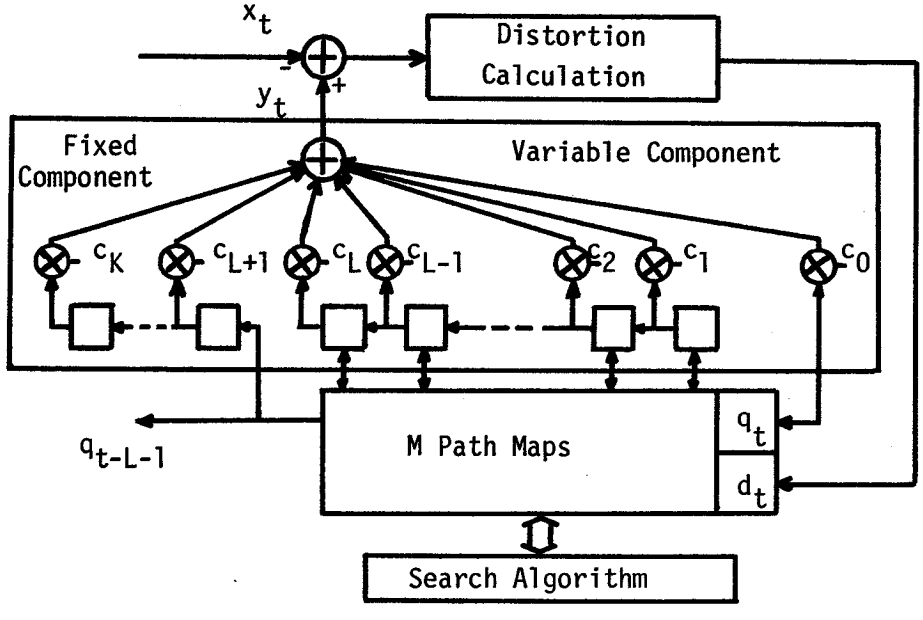


Fig. 8.3.2 — Tree Encoder

iid driving random variables V_t have the Laplacian (Two-side exponential) probability density

$$f(v) = \frac{1}{\sqrt{2}\sigma} \exp(- \sqrt{2}\, \sigma^{-1} |v|).$$

The former is selected bacause of observed analytical evidence above. The latter is selected because of its probability density having a sharp peak at zero, which is typically seen in relative frequencies for speech signals as shown in Fig. 8.3.3, where the horizontal axis gives the prediction error

$$X_t - \sum_{k=1}^{m} a_k X_{t-k}$$

for the AR coefficients a_k.

In Fig. 8.3.4, we show several rate-distortion curves. Rates are expressed in bits/sample (bits/letter) and distortions are expressed in terms of SNR (dB) given by

$$\text{SNR} = -10 \log_{10} \frac{D}{[\text{variance of } X_t]} ,$$

where the base of the logarithmic function is 10. In the fugure, $R_{1,G}(D)$ and $R_G(D)$ are the first order rate-distortion function and rate-distortion

function, respectively of the Gaussian AR source, whose AR coefficients are obtained from sampled speech. While $R_{L,V}(D)$ and $R_{G,V}(D)$ denote respective rate-distortion functions of the Laplacian driving process and Gaussian driving process. $R_{L,V}(D)$ is calculated numerically using Blahud's algorithm [63]. The difference between $R_{L,V}(D)$ and $R_{G,V}(D)$ is small compared with that between $R_G(D)$ and $R_{G,1}(D)$; from Theorem 8.2.3, the former difference is less than $(1/2)\log_2(\pi/e)$, the difference of respective values of differential entropy which is about 0.1 bits/sample. Theorem 8.2.5 also implies the difference between $R_G(D)$ and $R_L(D)$, the rate-distortion function of the Laplacian source, is less than 0.1 bits/sample for higher SNR than D_0 (about 20 dB). Indeed, the numerically obtained plots denoted by R_2 which are their second order rate-distortion functions show no noticeable difference, where the deviation from linearity at high rates is due to coarse quantization of coordinates to make Blahut's algorithm feasible for this two-dimensional case. Therefore we see that the shape of distribution is less relevant than the memory of sources in rate-distortion function; much reduction in rates comes from dependence between samples.

In the simulation we let q = 4 (2 bits/sample)

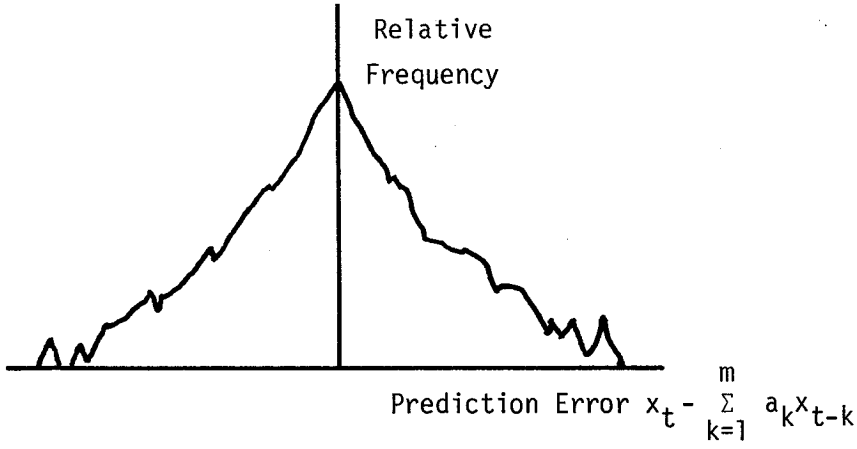


Fig. 8.3.3 — A Typical Relative Frequency Distrbution of the Prediction Error

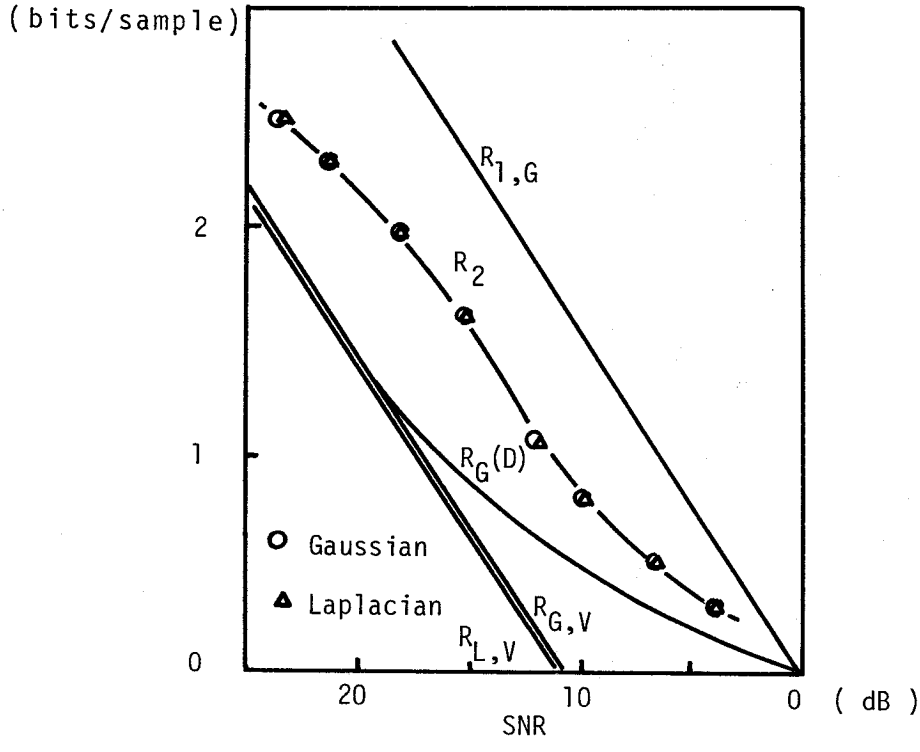


Fig. 8.3.4 — Rate-Distortion Functions

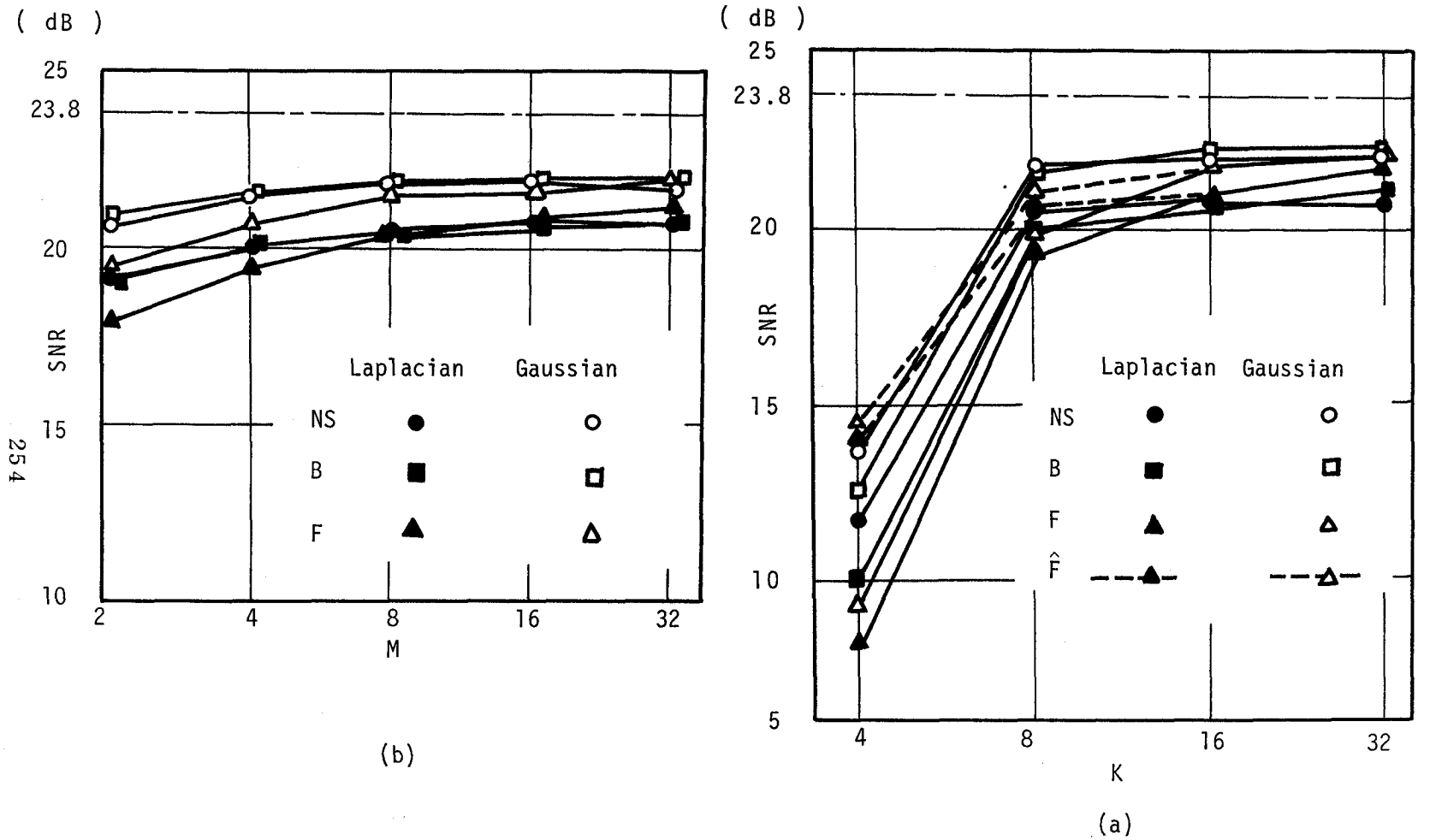


Fig. 8.3.5 — Tree Encoding on Speech-like Sources

and use the same AR coefficients as above. At this rate we have D(2) = 23.8 dB which is higher than D_0 in decibel and the reasoning for B-code is valid. The quantizer levels $s_1 = s_4$ and $s_2 = s_3$ are selected by another simulation with (M,L,K) = (4,8,8) for each of three code coefficient sets (the same level set is used for both F-code and $\hat{F}$-code). The smoothing filter $F(\rho) = (1+\rho^{-1})/2$ is selected to eliminate the noise typically observed for waveform coders with spectrum around a half the sampling frequency [54]. The simulation results are visualized in Fig. 8.3.5. The figures show universally good performance of the $F(\hat{F})$-code for large K and large M. While the F-code perform rather poorly at small encoding intensity M as seen in Fig. 8.3.5 (b), which is also observed in [60] for Gaussian sources. On the other hand, the NS-code and B-code both of which perform well for the Gaussian source no longer show good SNR for the Laplacian source. This is a clear contrast with the $F(\hat{F})$-code.

Coding of Speech

Though we have used stationary or asymptotically stationary source as speech-like sources, real speech signals are seldom considered as stationary, or even asymptotically stationary signals. Rather they are

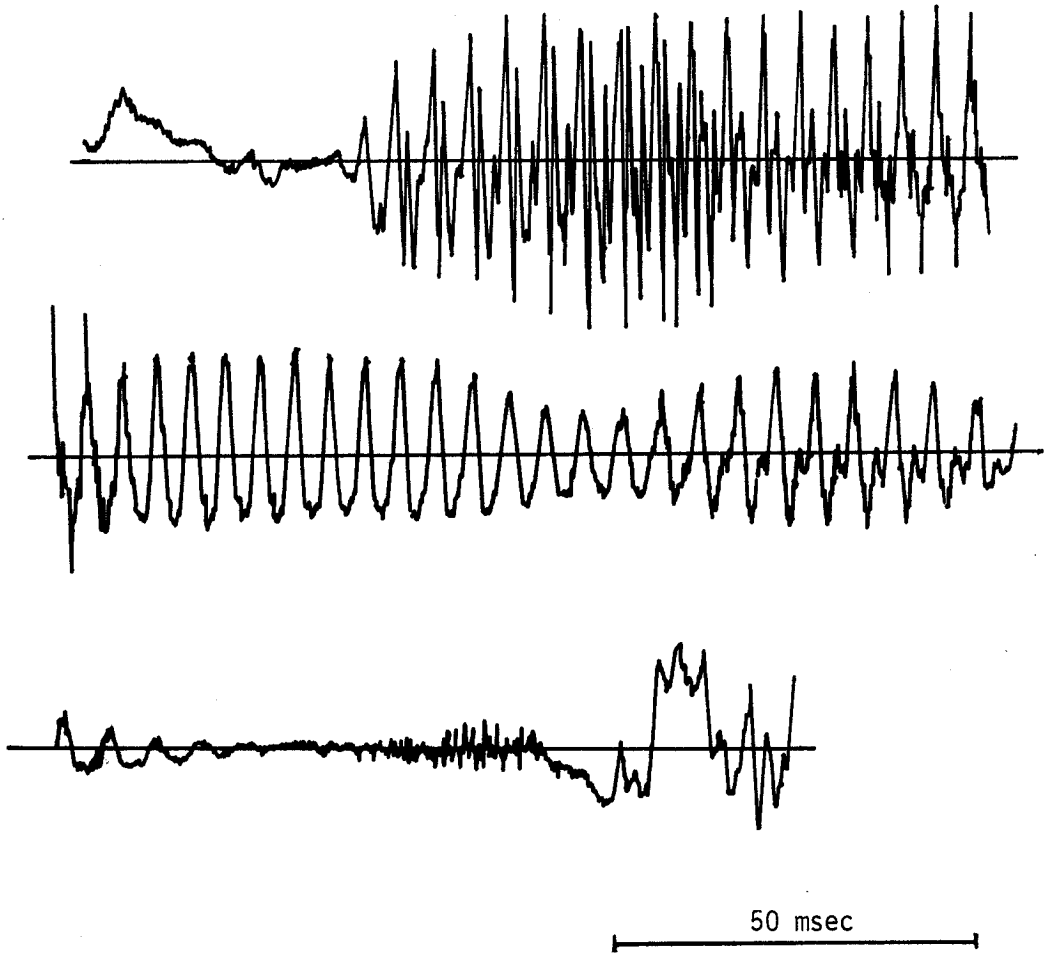


Fig. 8.3.6 — A Typical Speech Waveform

succession of relatively uniform waveforms lasting 30 ∿ 50 msec as seen in Fig 8.3.6. These fractions have respective power and spectrum, while statistical properties of codes are almost determined by code coefficients selected.

There are two directions in code-source adaptation: the one is to adapt codes only to power variation [54], [59] and the other is to adapt codes, varying code coefficients, to both power and spectrum variation [60],[61]. The former seems to fit for relatively high rates and the latter seems to fit for low rates where the cost needed to set up adaptation mechanisms is permissible.

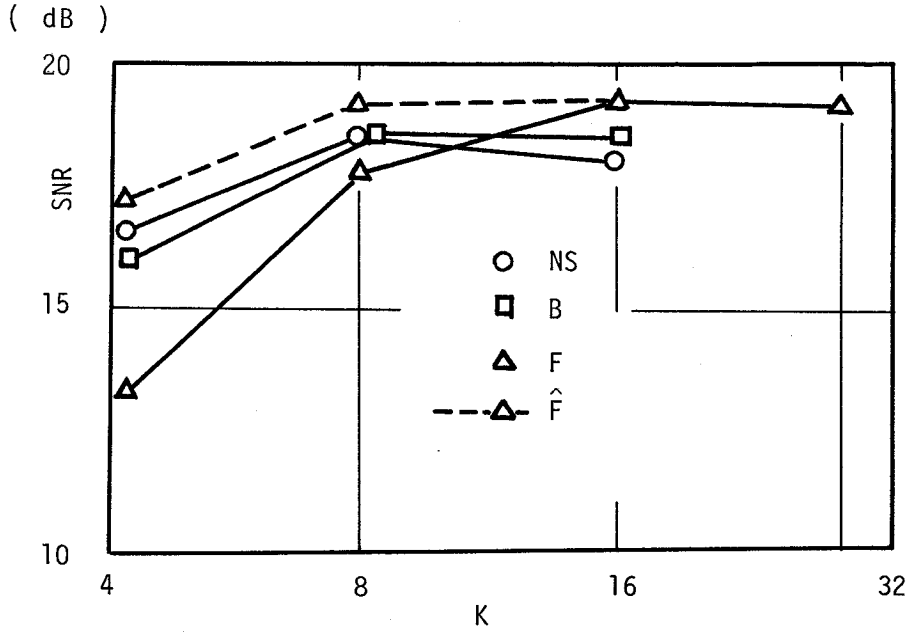
In this section, we apply two power-adaptation methods, AGC and AQ, to encode speech, sampled at 10 kHz, using tree codes of 2 bits/sample (i.e., 20 kbits/sec). AGC(Auto Gain Control) is the simplest mechanism that adjusts discontinuously its gain over the sampled data string partitioned into blocks of equal size so that sample power in each block remains about at a fixed level. We call the length of each block the AGC length and call the standard level the AGC level. With AGC, AGC gain has to be sent to reproduce signals at the decoder. However, the increase

in rate to send the additional signal is usually small, and it is neglected here. In contrast to AGC which adapts codes blockwise, AQ(Adaptive Quantizer) adapts the levels s_k, k = 1, ... ,q on each path in code trees according to the rule [59]

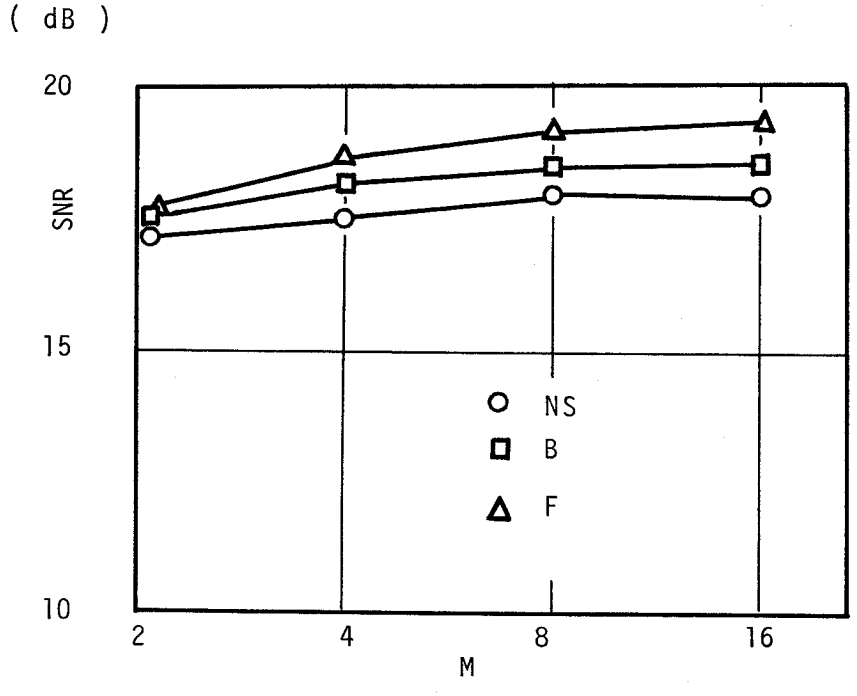
$$s_k^{(t)} = h_{k_{t-1}} s_k^{(t-1)}$$

for all k = 1, ... ,q , where $s_k^{(t)}$ are levels at t, k_t indicate the levels assumed at t, and h_k are AQ coefficients. AQ needs no additional messages at the decoder.

We first describe the results of sppech encoding using AGC. The data used to determine a_k are taken from a Japanese sentence "HONJITSU WA SEITEN NARI" sampled at 10 kHz and have about 16000 samples each represented by 12 bits. (These coefficients are used in all experiments.) Encoding is performed over the fraction "HONJITSU", about 4100 samples. AGC length is 10 msec (1000 samples) and AGC Level is fixed at a constant value for all codes according to preliminary expreiments which show approximately the same optimal AGC levels for all codes.
Results are reproduced in Fig 8.3.7, where each plot indicates the maximum SNR for L = 4,8,16. Two



(a)



(b)

Fig. 8.3.7 – Tree Encoding on Speech with AGC

Figures show uniform superiority of $\hat{F}$-code; F-code work only poorly for small K conceivably due to truncation effect. In Fig. 8.3.7(b), the superiority of F($\hat{F}$)-code becomes clear for large M.

Next we show the results of speech coding using AQ, which are given in Fig. 8.3.8. We put K = L in Fig. 8.3.8(a) and L = 8 in Fig. 8.3.8(b). Moreover we let AQ coefficients $h_1 = h_4$ and $h_2 = h_3$, and h_1 and h_2 are determined for NS-, B-, and F-codes (for F-coefficients and F-coefficients the same AQ coefficients are used) through preliminary experiments on (M,L,K) = (4,8,8). They are observed best or nearly best in coding on (M,L,K) = (4,16,16), too. We first notice that AQ inverts the order between codes: F($\hat{F}$)-code performs poorly for all K and M, and, on the other hand , B-code and NS-code work at SNR approximately 20 dB, much improvement over the gain obtained by AGC. However, F($\hat{F}$)-code tends to show better SNR for large K and M, whereas B-code and NS-code show no significant improvement in SNR by increase of K and M.

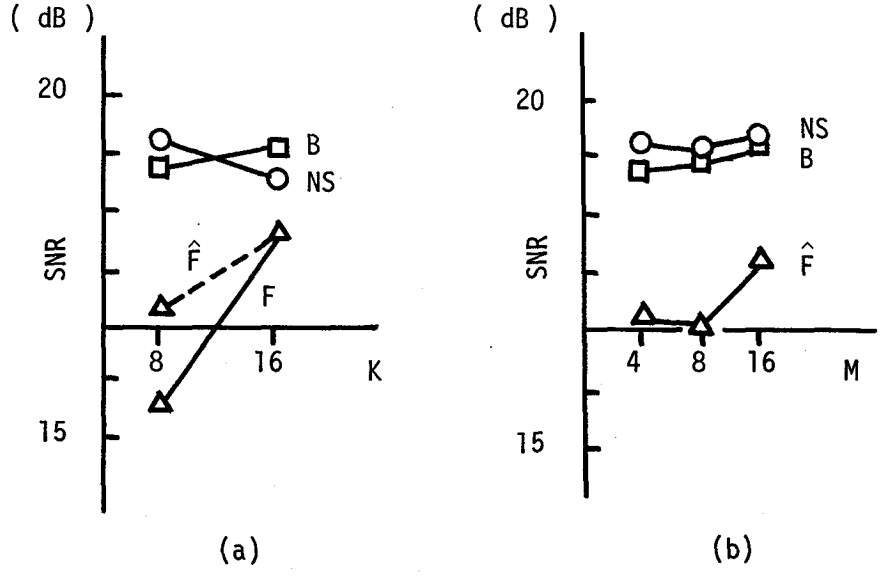


Fig. 8.3.8 — Tree Coding on Speech with AQ

Remarks on Encoding Algorithms– M-algorithm and Parallel Sorting M-algorithm

In this section, all tree encoders have used M-algorithm since it is simple and is easily implemented in software. However, this algorithm is not an efficient algorithm to attain high SNR [62]. One reason for it is that, as M gets large, sorting requires more computation. Another reason is that encoders have to sort out the best M paths from qM extensions. Generally, sorting the best M paths from qM extensions requires much computation than sorting out M ones by selecting the best M/2 paths from a half of the extensions and selectiong another set of the best M/2 paths from the other half of them separately.

Here we propose an efficient tree searching algorithm (in comparison with the M-algorithm), which is termed as the parallel sorting M-algorithm. This algorithm is obtained from the block-wise tree searching algorithm used in Section 7.2. The operation is simple. First consider the code tree in Fig. 8.3.10, and suppose that we are retaining M paths (or nodes). Since, from each retained path, we have q extensions numbered from 1 to q, all qM extensions are classified into q groups according to the attached numbers. In

the new algorithm, M paths are obtained by sorting out M/q paths from each group. If speech data sampled at 10 kHz are to be encoded, only M = 4 (qM^2 = 64 comparisons in 0.1 msec) is realizable using the M-algorithm, while M = 16 (M^2/q = 64 comparisons in 0.1 msec) is realizable using the proposed algorithm. In Fig. 8.3.9, the computational time per a node and coding distortion are shown for the two algorithms, where the source is speech and plots of the computational time for the parallel sorting M-algorithm show the overall computational time per a node. For M larger than 4, the new algorithm performs better than the M-algorithm. If we use a parallel sorting in the sorting, the computational time for the parallel sorting M-algorithm is reduced to approximately one q-th of the plots.

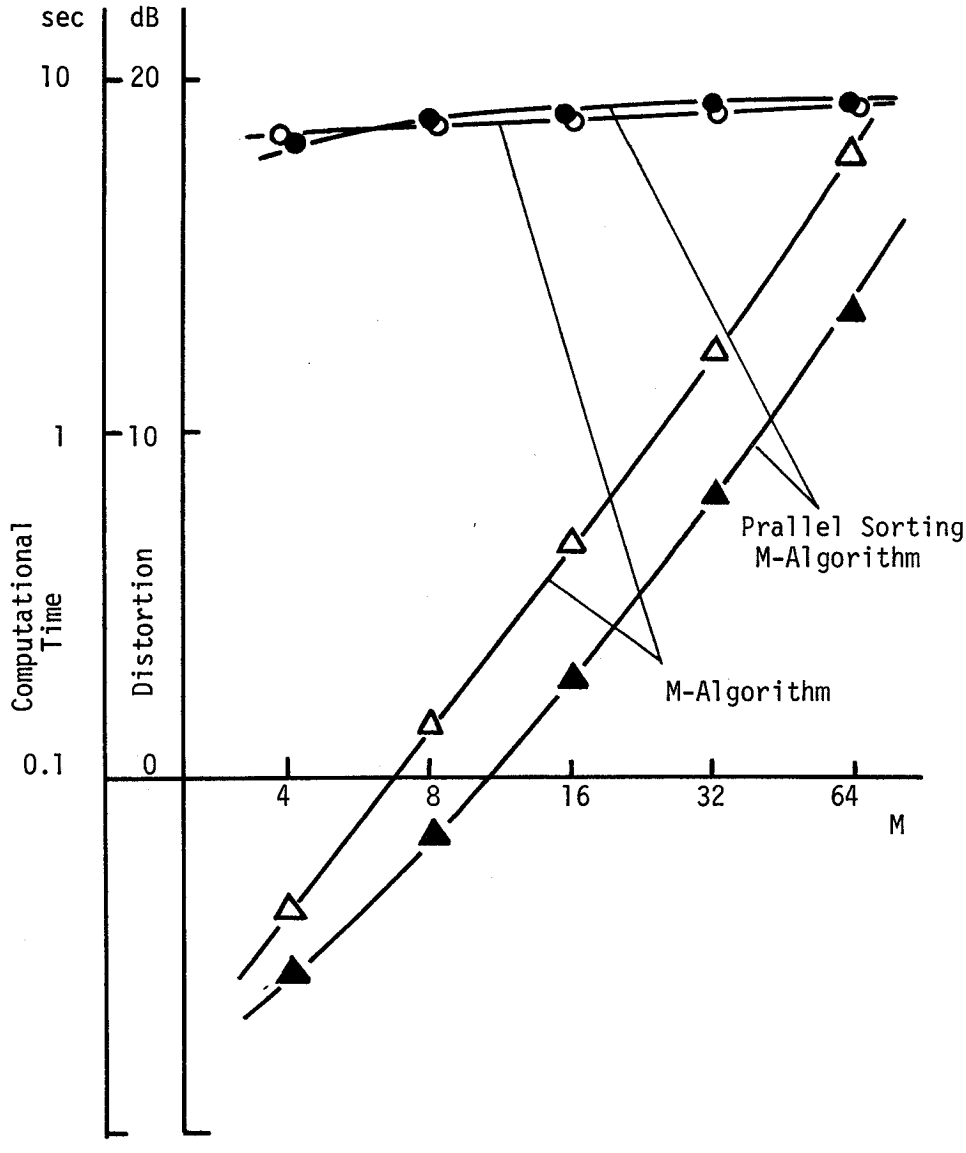
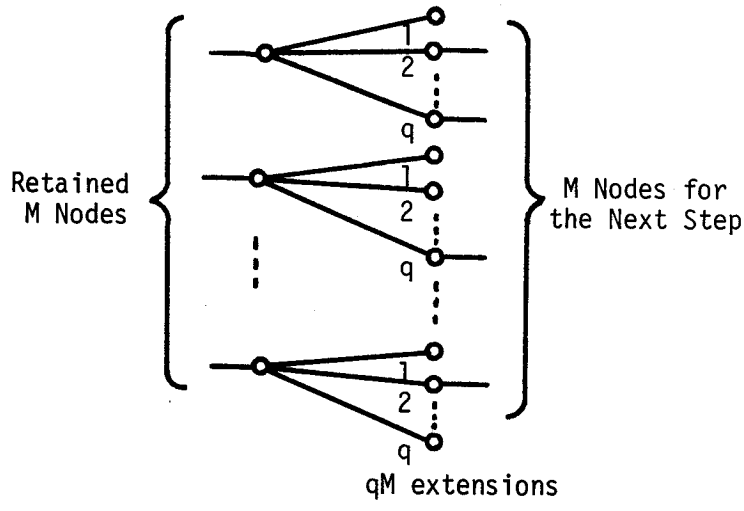
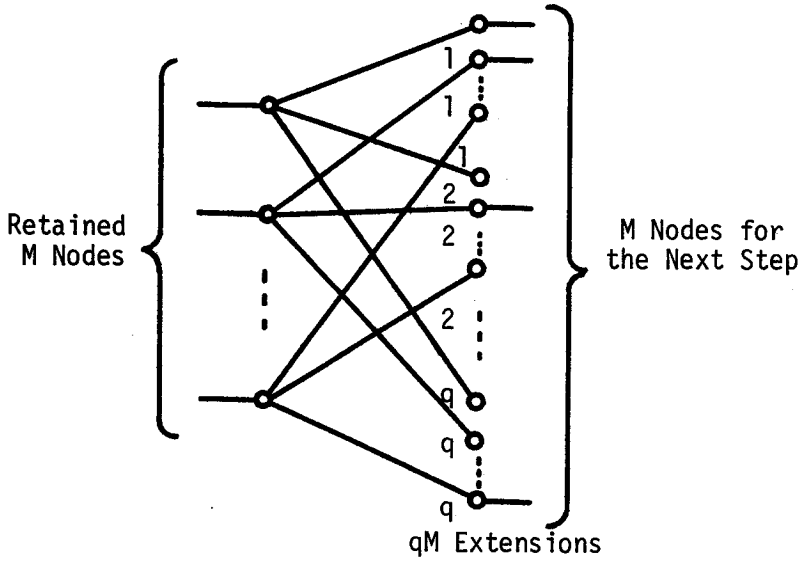


Fig. 8.3.9— Comparison of Computational Time and Distortion



(a) M-Algorithm



(b) Parallel Sorting M-Algorithm

Fig. 8.3.10 — The M-Algorithm and Parallel Sorting M-Algorithm

Supplementary page to p.262

Appendix to Chapter VIII

Proof of Lemma 8.2.1

Consider the difference equation

$$\xi_n + a_1\xi_{n-1} + \ldots + a_m\xi_{n-m} = 0 \ ; \ n = 1,2, \ldots ,$$

$$\xi_0 = \ldots = \xi_{1-m} = 0 \ ,$$

with the associated characteristic polynomial $A(\rho)$.

We first suppose that all zeros are distinct and satisfy $|\rho_1| \geq \ldots \geq |\rho_m| \geq 1 \geq |\rho_{m+1}| \geq \ldots \geq |\rho_m|$ (nondegenerate case). Then any solution (real) n-vector $\underline{\xi} = [\xi_1 \ \ldots \ \xi_n]^T$ of the difference equation is a linear combination of m linearly independent n-vecotrs $\underline{u}_k^n = [1 \ \rho_k \ \ldots \ \rho_k^{n-1}]^T$. Let $\underline{v}_{k+1}, \ \ldots \ , \underline{v}_n$ be any linearly independent n-vectors orthogonal to $\underline{u}_1^n, \ \ldots \ , \underline{u}_k^n$. And let $\|\underline{x}\|^2 = \underline{x}^*\underline{x}$ where $\underline{x}^*$ is the adjoint of $\underline{x}$. Then, from the Courant-Fisher theorem [57], we have

$$\lambda_{n,k} \leq \max_{\underline{x}:\text{orthogonal to } \underline{v}_{k+1},\ldots,\underline{v}_n} \|A_n\underline{x}\|^2/\|\underline{x}\|^2$$

$$= \max_{\beta_1,\ldots,\beta_k} \Big\| \sum_{\ell=1}^{k} \beta_\ell A_m \underline{u}^m \Big\|^2 / \Big\| \sum_{\ell=1}^{k} \beta_\ell \underline{u}^n \Big\|^2$$

$$\leq \max_{\beta_1,\ldots,\beta_k} \Big\| \sum_{\ell=1}^{k} \beta_\ell A_m \underline{u}^m \Big\|^2 / \Big\| \sum_{\ell=1}^{k} \beta_\ell \rho_\ell^{n-k} \underline{u}^k \Big\|^2$$

$$= 1/O(\ |\rho_k|^{2n}\ )$$

where $O(*)$ is any function such that $O(\delta)/\delta$ is bounded for large δ. The second inequality above follows from $\underline{x}^T(G+H)\underline{x} \geq \underline{x}^T G\underline{x}$ for any nonnegative summetric matrices G and H, and the last equality follows from positive definiteness of the matrix $[\ (\underline{u}_i^k)^* u_j^k\ ]$.

To bound the eigenvalues from lower, we note the matrix identity $A_n = B_1 B_2 \ldots B_m$ where

$$B_k = \begin{bmatrix} 1 & & & \\ -\rho_k & 1 & & \\ & \ddots & \ddots & \\ & & -\rho_k & 1 \end{bmatrix}_{n\times n} .$$

Moreover, any n-vector $\underline{x}$ orthogonal to $\underline{u}_1^n, \ldots, \underline{u}_{k-1}^n$ is writen as $\underline{x} = B_{k-1} \ldots B_1 \underline{y}$ where $\underline{y}$ is any n-vector such that $\underline{y} = [y_1 \ldots y_{n-k+1}\ 0 \ldots 0]^T$. Therefore, from the Courant-Fisher theorem,

$$1/\lambda_{n,k} \leq \max_{\substack{\underline{x}:\text{orthogonal to} \\ \underline{u}_1^n,\ldots,\underline{u}_{k-1}^n}} \|A_n^{-1}\underline{x}\|^2/\|\underline{x}\|^2$$

$$\leq \max_{\underline{x}} \|A_n^{-1}B_1 \ldots B_{k-1}\underline{x}\|^2/\|\underline{x}\|^2$$

$$\leq \text{trace } [B_m^{-1} \ \ldots \ B_k^{-1}]^*[B_m^{-1} \ \ldots \ B_k^{-1}]$$

$$= \begin{cases} O(\ |\rho_k|^{2n}\ ) & ;\ k = 1, \ \ldots, m', \\ \\ O(n) & ;\ k = m'+1, \ \ldots\ , m. \end{cases}$$

Therefore, the lemma has been proved for the non-degenerate case.

When some zeros degenerate as $\rho_i = \rho_{i+1} = \ldots = \rho_{i+j}$, then vectors $\underline{u}_i^n, \ \ldots\ , \underline{u}_{i+j}^n$ are given as

$$\underline{u}_{i+\ell}^n = [1 \ \ 2^{\ell}\rho_i \ \ldots \ n^{\ell}\rho_i^{n-1}],$$

for all $\ell = 0,1, \ \ldots\ ,j$, and the argument goes in almost the same way since $\underline{u}_1^k, \ \ldots\ , \underline{u}_k^k$ are again linearly independent.

IX. Conclusion

We have discussed channel and source coding, chiefly, with tree codes, and have seen that the tree codes are useful in designing practical communication systems. In channel coding, the theory well approximates the real system, and the practice well supports the assertions of the theory. In source coding, however, we do not have many succesful applications of the tree coding theory. This is partly because, in contrast to channels such as the white Gaussian channel in space communication, sources are active by itself and tend to alter their temporal characteristics according to the contents that should be transmitted, like speech that alters its power and spectrum from one consonent to another. Therefore robustness or universality of codes is an important factor in the coding system design, and should be explored thoroughly in conjunction with implementable tree encoding algorithms.

Reference

[1] Shannon, C.E.,"A mathematical theory of communication," *Bell System Tech. J.*, vol.27, pp.379-423 (Part I), pp.623-656 (Part II), 1948

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

[3] Shannon, C.E., R.G. Gallager, and E.R. Berlekamp, "Lower bounds to error probability for coding on discrete memoryless channels," *Inform. and Control*, vol.10, pp.65-103 (Part I), pp.522-552 (Part II), 1967

[4] Omura, J.K.,"Expurgated bounds, Bhattacharyya distance, and rate distortion functions," *Inform. and Control*, vol.24, pp.358-383, 1974

[5] Feinstein, A.,*Foundations of Information Theory*. New york: McGraw-Hill, 1958

[6] Wolfowitz, J.,*Coding Theorems of Information Theory*. Engelwood Criffs, N.J.: Prentice-Hall and Springer-Verlag, 1961

[7] Csiszar, I., J. Korner, and K. Marton,"A new look at the exponential error bounds for memoryless channels," in *Proc. 1977 IEEE Int. Symp. Inform. Theory*, Cornel Univ., Ithaca, NY, 1977

[8] Csiszar, I. and J. Korner, *Information Theory: Coding Theorems for Discrete Memoryless Systems.* New York: Academic, to appear

[9] Csiszar, I. and J. Korner,"Graph decomposition: A new key to coding theorems," in *Proc. 1979 IEEE Int. Symp. Inform. Theory,* Grignano, Italy, 1979

[10] Viterbi, A.J.,"Error bounds for convolutional codes and an asymptotically optimum decoding algorithm," *IEEE Trans. Inform. Theory,* vol.IT-13, pp.260-269, 1967

[11] Omura, J.K.,"On the Viterbi decoding algorithm," *IEEE Trans. Inform. Theory,* vol.IT-15, pp.177-179, 1969

[12] Viterbi, A.J. and J.P. Odenwalder,"Further results on optimal decoding of convolutional codes," *IEEE Trans. Inform. Theory,* vol.IT-15, pp.723-734, 1969

[13] Forney, G.D.,Jr.,"Coding and its application in space communications,"*IEEE Spectrum,* June 1970, pp.47-58

[14] Wozencraft, J.M. and B. Reiffen,*Sequential Decoding,* Cambridge, Mass.: MIT Press, 1961

[15] Fano, R.M.,"A heuristic discussion of probabilistic decoding," *IEEE Trans. Inform. Theory,* vol.IT-9, pp.64-74, 1963

[16] Jelinek, F., *Probabilistic Information Theory.* New York: McGraw-Hill, 1968

[17] Falconer, D.D.,"A hybrid sequential and algebraic decoding scheme," Ph. D. Thesis, Dept. of E.E., MIT, Cambridge, Mass.

[18] Savage, J.E.,"Sequential decoding– The computation problem," *Bell System Tech. J.*, vol.44, pp.149-175, 1966

[19] Zigangirov, K.Sh.,"Some sequential decoding procedures," *Problems in Information Transmission*, vol.2, pp.13-25, 1966 (in Russian)

[20] Jelinek, F.,"An upper bound on moments of sequential decoding effort," *IEEE Trans. Inform. Theory*, vol.IT-15, pp.140-149, 1969

[21] Haccoun, D. and M.J. Ferguson,"Generalized stack algorithms for decoding convolutional codes," *IEEE Trans. Inform. Theory*, vol.IT-21, pp.638-651, 1975

[22] Jordan, K.L., Jr.,"The performance of sequential decoding in conjunction with efficient modulation," *IEEE Trans. Commun.*,vol.COM-14, pp.283-297, 1966

[23] Jacobs, I.M. and E.R. Berlekamp,"A lower bound to the distribution of computation for sequential decoding," *IEEE Trans. Inform. Theory*, vol.IT-13, pp.167-174, 1967

[24] von Bahr, B. and Esseen, C.,"Inequalities for the r the absolute moment of a sum of random variables, $1 \leq r \leq 2$," *Ann. Math. Statist.*, vol.36 pp.299-303, 1965

[25] Berger, T., *Rate Distortion Theroy*. Englewood Cliffs, N.J.: Prentice-hall, 1971

[26] Pinsker, M.S., *Information and Information Stability of Random Variables and Processes*. San Francisco: Holden Day, 1965

[27] Billingsley, P.,*Ergodic Theory and Information*. New York: Wiley, 1960

[28] Shannon, C.E.,"Coding theorems for a discrete source with a fidelity criterion," *IRE Nat. Conv. Rec.*, pt.4, pp.142-163, 1959

[29] Gray, R.M., D.L. Neuhoff, and J.K. Omura, "Process definition of distortion-rate function and source coding theorems," *IEEE Trans. Inform. Theory*, vol.IT-21, pp.524-532, 1975

[30] Marton, K.,"On the rate distortion function of stationary sources," *Problems of Control and Information Theory*, vol.4, pp.289-297, 1975

[31] Sakrison, J.D.,"The rate-distortion function for a class of sources," *Information and Control*, vol.15, pp.165-195, 1969

[32] Ziv, J.,"Coding of sources with unknown statistics. Part I:Probability of encoding error, Part II:Distortion relative to a fidelity criterion," *IEEE Trans. Inform. Theory*, vol.IT-18, pp.384-389 and pp.389-394, 1972

[33] Davisson, L.D.,"Universal noiseless coding," *IEEE Trans. Inform. Theory*, vol.IT-19. pp.783-795, 1973

[34] Neuhoff, D.L., R.M. Gray, and L.D. Davisson, "Fixed rate universal block source coding with a fidelity criterion," *IEEE Trans. Inform. Theory*, vol.IT-21, pp.511-523, 1975

[35] Gray, R.M. and L.D. Davisson,"Source coding theorems without ergodic assumption," *IEEE Trans. Inform. Theory*, vol.IT-20, pp.512-516, 1974

[36] Oxtoby, J.C.,"Ergodic sets," *Bull. Amer. Math. Soc.*, vol.58, pp.116-136, 1952

[37] Parthasarathy, K.R.,"On the integral representation of the rate of transmission of a stationary channel," *Illi. J. Math.*,vol.5, pp.684-656, 1961

[38] Jelinek, F.,"Tree encoding of memoryless time-discrete sources with a fidelity criterion," *IEEE Trans. Inform. Theory*, vol.IT-15, pp. 584-590, 1969

[39] Davis, C.R. and M.E. Hellman,"On tree coding with a fidelity criterion," *IEEE Trans. Inform. Theory*, vol.IT-21, pp.373-378, 1975

[40] Viterbi, A.J. and J.K. Omura,"Trellis encoding of memoryless discrete-time sources with

a fidelity criterion," *IEEE Trans. Inform. Theory*, vol.IT-20, pp.325-332, 1974

[41] Anderson, J.B. and F. Jelinek,"A two-cycle algorithm for sources with a fidelity criterion," *IEEE Trans. Inform. Theory*, vol.IT-19, pp.77-92, 1973

[42] Gallager, R.G.,"Tree encoding for sources with a distortion measure," *IEEE Trans. Inform. Theory*, vol.IT-20, pp.65-72, 1974

[43] Jelinek, F. and J.B. Anderson,"Instrumentable tree encoding of information sources," *IEEE Trans. Inform. Theory*, vol.IT-22, pp.82-83, 1971

[44] Wilson, S.G. and D.W. Lytle,"Trellis encoding of continuous-amplitude memoryless sources," *IEEE Trans. Inform. Theory*, vol.IT-23, pp.404-409, 1977

[45] Tan, H.H.,"Tree coding of discrete-time abstract alphabet stationary block-ergodic sources with a fidelity criterion," *IEEE Trans. Inform. Theory*, vol.IT-22, pp.671-681, 1976

[46] Gray, R.M.,"Sliding-block source coding," *IEEE Trans. Inform. Theory*, vol. IT-21, pp.357-368, 1975

[47] Berger, T and Joseph Ka-Yin Lau,"On binary

sliding-block codes," *IEEE Trans Inform. Theory*, vol.IT-23, pp.343-353, 1977

[48] Omura, J.K. and A. Shōhara,"On convergence of distortion for block and tree encoding of symmetric sources," *IEEE Trans. Inform. Theory*, vol.IT-19, pp.573-577, 1973

[49] Anderson, J.B.,"A stack algorithm for source coding with a fidelity criterion," *IEEE Trans. Inform. Theory*, vol.IT-20, pp.211-216, 1974

[50] Rabiner, L.R. and R.W. Schafer, *Digital Processing of Speech Signals*. Englewood Cliffs, N.J.: Prentice-Hall, 1978

[51] Flanagan, J.L. et al."Speech Coding," *IEEE Trans. Commun.*, vol.COM-27, no.4, pp.710-737, 1979

[52] Itakura, F. and S. Saito,"Analysis-synthesis telephony based upon the maximu likelihood method," in *Proc. 6th Int. Congress on Acoustics*, pp.C17-20, Tokyo, 1968

[53] Gold, B."Digital speech networks," *Proc. IEEE*, vol.65, pp.1636-1658, 1977

[54] Anderson, J.B. and J.B. Bodie,"Tree encoding of speech," *IEEE Trans. Inform. Theory*, vol.IT-21, pp.376-387, 1975

[55] Dobrushin, R.L.,"A general formulation of the fundamental Shannon theorem in information theory," *Uspehi Mat. Akad. Nauk. SSSR*, vol.14, pp.3-104,

1959 (Also, *Trans. Amer. Math. Soc.*, series 2, vol.33, 323-438.)

[56] Gray, R.M.,"Information rates of autoregressive processes," *IEEE Trans. Inform. Theory*, vol.IT-16, pp.412-421, 1970

[57] Bellman, R., *Introduction to Matrix Analysis*. 2nd ed. New york:McGraw-Hill, 1970

[58] Berger, T.,"Information rates of Wiener proesses," *IEEE Trans. Inform. Theory*, vol.IT-16, pp.134-139, 1970

[59] Jayant, N.S. and S.W. Christensen,"Tree encoding of speech using the (M,L)-algorithm and adaptive quantization," *IEEE Trans. Commun.*, vol.COM-26, pp.1376-1379, 1978

[60] Wilson, S.G. and S. Husain,"Adaptive tree encoding of speech at 8000 bits/s with a frequency-weighted error criterion," *IEEE Trans. Commun.*, vol.COM-27, pp.165-170, 1979

[61] Matsuyama, Y. and R.M. Gray,"Universal tree encoding for speech," *IEEE Trans. Inform. Theory*, to appear

[62] Anderson, J.B. and S. Mohan,"A systematic analysis of cost for sequential coding algorithms," in *Proc. 1979 IEEE Int. Symp. Inform. Theory*, Grignano, Italy, 1979

[63] Blahut, R.E.,"Computation of channel capacity and rate-distortion functions," *IEEE Trans. Inform. Theory*, vol.IT-18, pp.460-473, 1972

[64] Forney, G.D.,Jr.,"Coding and its application in space communications," *IEEE Spectrum*, June, pp.47-58, 1970

[65] Jacobs, I.M.,"Practical application of coding," *IEEE Trans. Inform. Theory*, vol.IT-20, pp.305-310, 1974

[66] Hiroyoshi Morita,"On source coding using trellis codes," M.E. thesis, Osaka Univ., Osaka, Japan, 1980

Glossary of Abbreviations

AGC	auto gain control
AQ	adaptive quantizer
AR	autoregressive
ARMA	autoregressive moving-average
BSC	binary symmetric channel
BSS	binary symmetric source
DMC	discrete memoryless channel
DMS	discrete memoryless source
iid	independently and identically distributed
pmf	probability mass function
SNR	signal-to-noise power ratio