HUFFMAN ENCODING UTILIZING THE KRAFT - MCMILLAN INEQUALITY

A thesis submitted

by

Gerald Shea

to

the Faculty of Pure and Applied Science

of the University of Ottawa

in partial fulfillment of the requirements

for the degree of

Master of Science

in the subject of

Mathematics

1970

ACKNOWLEDGMENTS

I wish to express my appreciation to my supervisor, Professor Edward L. Cohen, for his invaluable guidance, his constant encouragement and his constructive suggestions made during the research work and the writing of this thesis. I would also like to thank Professor S. G. S. Shiva, for his suggestion of the problem contained herein and for his consultations with me.

Thanks are due to the National Research Council of Canada and to the Provincial Government of Ontario for their financial aid.

ABSTRACT

Huffman encoding is studied as a logical consequence of the Shannon code for discrete noiseless channels. By means of the Kraft-McMillan inequality, a new algorithm for Huffman encoding is given. In many cases, the algorithm proves to be considerably shorter than previously known methods. An upper bound is found for the lengths of the individual Huffman code words. In conclusion, the efficiency of Huffman codes is discussed.

TABLE of CONTENTS

		page
Introduction		i
Chapter I	Instantaneous Codes and the Kraft-McMillan Inequality	1
	Entropy of the Source and Shannon Codes	6
	Huffman Encoding	10
Chapter II	Another Algorithm for Huffman Encoding	19
Chapter III	Examples and Applications of the Algorithm	31
Chapter IV	Entropy and Huffman Encoding	49
List of Definitions		52
Bibliography		53
Errata		56

INTRODUCTION

Information theory has its origins in the engineering problem of transmitting data accurately and efficiently through telegraph lines, telephone cable, radio bands, etc. The data may vary from the thousands of words in the English vocabulary to a small set of numbers like $\{0, 1, \ldots, 9\}$; however, the transmission system is often limited to a few symbols. Binary symbols such as positive or negative charge and electrical pulse or no pulse are common.

In this paper accuracy is not of concern, since an error-free system is assumed. It is also assumed that the encoder produces a finite positive number D of distinct signals that can be transmitted and then processed by the decoder. Thus, this paper restricts itself to a discrete, noiseless channel--a system that delivers to the decoder a perfect copy of each distinct symbol produced by the encoder.

One measure of efficiency is the time taken to send and receive each message. Let $\{\alpha_i\}_{i=1}^{K}$ be a list of possible messages, p_i the probability that α_i will be transmitted at any given moment, and n_i the number of symbols needed to encode α_i. Then the average length of the code for these messages equals $\sum_{i=1}^{K} n_i p_i$. To save transmission time, this average should be as short as possible.

A desirable property of an encoding is that it be instantaneous; that is, each code word can be immediately

recognized and uniquely decoded. D. Huffman [15] gave a procedure for finding a minimum average length instantaneous code for any finite set of messages $\{\alpha_i\}_{i=1}^{K}$ with fixed probabilities of transmission $\{p_i\}_{i=1}^{K}$. The purpose of this paper is to derive another algorithm for Huffman codes. The method given here both shortens in many cases the number of steps in the procedure and aids to elucidate the nature of Huffman codes.

Chapter I presents commonly accepted definitions and lemmas basic to the study of instantaneous codes. The sources from which these details are drawn are given in square brackets; the numbers therein refer to the bibliography.

Since inductive arguments are frequently employed, no mention of this fact is made in the proofs. The form of the argument will reveal when induction is being used.

Chapter II develops a new algorithm for Huffman encoding. For any fixed set of probabilities $\{p_i\}_{i=1}^{K}$, a unique instantaneous code is constructed. The algorithm repeatedly reduces the average length of this code. The process ceases when further reductions would destroy the instantaneous property of the encoding. Theorem 1 states that the algorithm has relatively few steps; Theorem 3 shows that the resultant encoding is Huffman.

Chapter III gives examples of codes derived by applying the algorithm in various forms. Some results that simplify computation are illustrated.

Chapter IV describes a Huffman code with reference to the lengths of the individual code words and in terms of its overall efficiency.

CHAPTER I

INSTANTANEOUS CODES and the KRAFT-McMILLAN INEQUALITY

DEFINITION 1: Let Z be a source of data and let $A = \{\alpha_i\}_{i=1}^{K}$ be the total vocabulary of symbols used by Z. Then each α_i is said to be a *word* and A a *word list* of K words.

DEFINITION 2: Let T be a channel consisting of encoder, transmission line, and decoder. Let $D \geq 2$ be the number of symbols acceptable to T. Then the set $\{0, 1, \ldots, D-1\}$ is said to be the *coding alphabet* and an integer x, $0 \leq x \leq D-1$, a *code letter*.

DEFINITION 3 [1, p. 46]: A *block code*, or simply a *code*, relates each word α_i of a word list A into a fixed sequence of code letters. These fixed sequences are called *code words* and are denoted by $X_i = (x_1\ x_2\ \ldots\ x_{n_i})$; n_i is the *length* of the code word X_i.

Hence a code may be denoted by $\{(\alpha_i, X_i)\}_{i=1}^{K}$.

DEFINITION 4: A code $\{(\alpha_i, X_i)\}_{i=1}^{K}$ is said to be *uniquely decodable* [1, p. 48] if every finite sequence $X_1\ X_2\ \ldots\ X_\mu = x_1\ x_2\ \ldots\ x_\lambda$ of μ code words and λ code letters is distinct from every different finite sequence of ν code words $Y_1\ Y_2\ \ldots\ Y_\nu = y_1\ y_2\ \ldots\ y_\lambda$ of λ code letters.

Therefore a uniquely decodable code can decode any code letter sequence $x_1\ x_2\ \ldots\ x_\lambda$ at most one way.

DEFINITION 5: A code $\{(\alpha_i, X_i)\}_{i=1}^{K}$ is said to be *instantaneous* [1, p. 50] if each word received through a discrete noiseless channel can be immediately and uniquely decoded.

An instantaneous code is a uniquely decodable code with the property that $X_i = x_1 x_2 \ldots x_{n_i}$ is recognized as soon as it is received. That is, no investigation of the subsequent code letters is needed to assure the decoder that $x_1 x_2 \ldots x_{n_i}$ is a complete code word, namely, X_i .

DEFINITION 6: Let $X_i = x_1 x_2 \ldots x_{n_i}$ be a code word. The sequence of code letters $(x_1 x_2 \ldots x_\lambda)$ with $\lambda \leq n_i$ is called a *prefix* of X_i [1, p. 50].

LEMMA 1: A code $\{(\alpha_i, X_i)\}_{i=1}^{K}$ is instantaneous iff no X_j is a prefix of X_i , for all i and all j not equal to i . [1, p. 51]

Proof: The sufficiency. Let the sequence

$$X_i X_j \ldots = x_1 x_2 \ldots x_{n_i} y_1 y_2 \ldots y_{n_j} \ldots$$

be sent through the channel. By the prefix property $x_1 x_2 \ldots x_{n_i}$ is distinct from all other code words. Also, $x_1 x_2 \ldots x_m$ is not a code word for $m < n_i$ as no prefix of X_i is, and hence $x_1 x_2 \ldots x_m$ will not be decoded. Again $x_1 x_2 \ldots x_{n_i} y_1$, $x_1 x_2 \ldots x_{n_i} y_1 y_2$, ... are not code words as X_i is not a prefix of any other code word. Therefore X_i is uniquely decipherable as soon as $x_1 x_2 \ldots x_{n_i}$ is received.

The necessity. Assume there exist code words X_i and X_j such that X_i is a prefix of X_j . Consider the sequence

$$X_i X_j \ldots = x_1 x_2 \ldots x_{n_i} y_1 y_2 \ldots y_{n_j} \ldots .$$

If $n_i = n_j$, $X_i = X_j$, and the sequence cannot be uniquely decoded. Assume $n_i < n_j$. Then $x_1 x_2 \ldots x_{n_i}$ could be either X_i or the first n_i code letters of X_j . No decision could be made, if at all, until at least one more code letter is examined. In any case the sequence cannot be both uniquely and immediately decoded. QED

LEMMA 2: [17, 20, 22; 1, pp. 53 - 61]

THE KRAFT - McMILLAN INEQUALITY

Let $\{n_i\}_{i=1}^{K}$ be a set of positive integers. Then there exists a uniquely decodable code $\{(\alpha_i, X_i)\}_{i=1}^{K}$ with these word lengths if, and only if,

$$\sum_{i=1}^{K} D^{-n_i} \leq 1 .$$

Proof: The sufficiency. Let $\{n_i\}_{i=1}^{K}$ be a set of positive integers such that $\sum_{i=1}^{K} D^{-n_i} \leq 1$. It is enough to show that an instantaneous code with these word lengths can be constructed.

Let $N = \max \{n_i\}_{i=1}^{K}$, and let q_j , $j = 1, 2, \ldots, N$ be the number of n_i equal to j . That is, $\sum_{j=1}^{N} q_j = K$. Then $\sum_{j=1}^{N} q_j D^{-j} \leq 1$ or $\sum_{j=1}^{N} q_j D^{N-j} \leq D^N$. Hence

$$q_N \leq D^N - q_1 D^{N-1} - q_2 D^{N-2} - \ldots - q_{N-1} D .$$

Since $q_j \geq 0$, $j = 1, 2, \ldots, N$ and $D > 0$, the following sequence of inequalities is formed by successively dividing by D and rearranging the terms.

$$q_{N-1} \leq D^{N-1} - q_1 D^{N-2} - q_2 D^{N-3} - \ldots - q_{N-2} D$$

. . . . . .

$$q_2 \leq D^2 - q_1 D$$

$$q_1 \leq D \quad .$$

By means of Lemma 1, an instantaneous code is constructed as follows:

Since $q_1 \leq D$, q_1 distinct code words of length equal to 1 are formed from $\{0, 1, 2, \ldots, q_1 - 1\}$. $D - q_1$ distinct prefixes, namely $\{q_1, q_1 + 1, \ldots, D - 1\}$, can be used for longer code words. Thus $(D - q_1)D$ code words of two code letters may be formed with distinct prefixes. As $q_2 \leq (D - q_1)D$, the necessary code words are constructed.

Let $Y_j = D^j - q_1 D^{j-1} - q_2 D^{j-2} - \ldots - q_{j-1} D$, for some j, $1 \leq j \leq N$. Assume q_j code words of length j have been formed satisfying the prefix property. As $q_j \leq Y_j$, $Y_j - q_j$ distinct prefixes of length j remain. In fact, $(Y_j - q_j)D$ code words of length $j + 1$ may be formed under the prefix property. As $q_{j+1} \leq (Y_j - q_j)D$, these code words can be constructed.

The necessity. Let $I = \{(\alpha_i, X_i)\}_{i=1}^{K}$ be a uniquely decodable code with code word lengths $\{n_i\}_{i=1}^{K}$. That is, each finite sequence of r code letters can be decoded at most one way. Let Q_r be the number of sequences of q code words that can be formed so that each sequence has exactly r code letters. Since each sequence must be distinct, $Q_r \leq D^r$, as there are only D^r distinct r-tuples.

Consider $\left(\sum_{i=1}^{K} D^{-n_i} \right)^q = (D^{-n_1} + D^{-n_2} + \ldots + D^{-n_K})^q$.

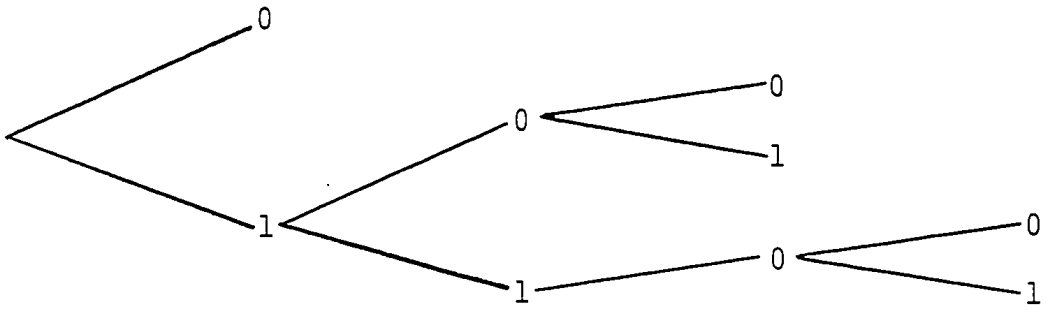
Let $N = \max \{n_i\}_{i=1}^{K}$. Since each of the K^q terms of the expansion is of the form $D^{-n_{i_1}} D^{-n_{i_2}} \ldots D^{-n_{i_q}}$ and there are Q_r terms such that $n_{i_1} + n_{i_2} + \ldots + n_{i_q} = r$, implying that $q \leq r \leq qN$, then

$$\left(\sum_{i=1}^{K} D^{-n_i} \right)^q = \sum_{r=q}^{qN} Q_r D^{-r} \leq \sum_{r=q}^{qN} D^r D^{-r} \leq qN - q + 1 \leq qN .$$

Since N is constant, and the inequality holds for all q , $\sum_{i=1}^{K} D^{-n_i} \leq 1$. QED

By means of the prefix property, given the $\{n_i\}_{i=1}^{K}$ of an instantaneous code, a representation of the code may be easily constructed [1, pp. 52 - 53] .

One method is to construct "trees": Let $D = 2$. Note $2^{-1} + 2^{-3} + 2^{-3} + 2^{-4} + 2^{-4} \leq 1$. Procede as follows.



Then

$n_1 = 1$ :: $X_1 = 0$

$n_2 = 3$ :: $X_2 = 1\ 0\ 0$

$n_3 = 3$ :: $X_3 = 1\ 0\ 1$

$n_4 = 4$ :: $X_4 = 1\ 1\ 0\ 0$

$n_5 = 4$ :: $X_5 = 1\ 1\ 0\ 1$

DEFINITION 7: Let $I = \{(\alpha_i, X_i)\}_{i=1}^{K}$ be a code with code word lengths $\{n_i\}_{i=1}^{K}$. Define

$$S_I = \sum_{i=1}^{K} D^{-n_i} .$$

Hence by the prefix property and Definition 7 with the Kraft-McMillan Inequality, an instantaneous code may be redefined by $I = \{(\alpha_i, n_i)\}_{i=1}^{K}$ such that $S_I \leq 1$. Then $\{(\alpha_i, X_i)\}_{i=1}^{K}$ may be constructed as illustrated on the previous page.

ENTROPY of the SOURCE and SHANNON CODES

DEFINITION 8: Let Z be a discrete source and $\{\alpha_i\}_{i=1}^{K}$ the word list of all possible words used by Z. Let $\{p_i\}_{i=1}^{K}$ be given such that p_i is the probability that α_i is emitted ($i = 1,2,\ldots,K$) with $\sum_{i=1}^{K} p_i = 1$. If successive words $\alpha_{i_1}\ \alpha_{i_2}\ \ldots$ are statistically independent, Z is said to be a *zero-memory source* [1, p. 14].

A zero-memory source is completely described by $\{\alpha_i\}_{i=1}^{K}$ and $\{p_i\}_{i=1}^{K}$ and shall be denoted $Z = \{(\alpha_i, p_i)\}_{i=1}^{K}$. For the statistical independence of Z gives the probability of any sequence of words $\alpha_{i_1}\ \alpha_{i_2}\ \ldots\ \alpha_{i_n}$ as $p_{i_1}\ p_{i_2}\ \ldots\ p_{i_n}$.

DEFINITION 9: The *entropy* $H_D(Z)$ [1, p. 14] of a zero-memory source $Z = \{(\alpha_i, p_i)\}_{i=1}^{K}$ with respect to a code with D code letters is defined by

$$H_D(Z) = -\sum_{i=1}^{K} p_i \log_D p_i .$$

Shannon [30] indicates that $-\log_D p_i$ serves to measure the uncertainty of the event α_i. Hence it gives the value of the information that α_i did occur. It also justifies the use of approximately $-\log_D p_i$ code letters to encode α_i.

DEFINITION 10: Let $Z = \{(\alpha_i, p_i)\}_{i=1}^{K}$ be a zero-memory source. The *n-th extension* Z^n of Z [1, p. 20] is defined by $Z^n = \{(\beta_j, \pi_j)\}_{j=1}^{K^n}$ where the β_j are distinct sequences of code words $\beta_j = (\alpha_{i_1} \alpha_{i_2} \ldots \alpha_{i_n})$ for $i = 1, 2, \ldots, K^n$.

Therefore $\pi_j = p_{i_1} p_{i_2} \ldots p_{i_n}$. Also

$$\sum_{j=1}^{K^n} \pi_j = \sum_{i_1=1}^{K} p_{i_1} \sum_{i_2=1}^{K} p_{i_2} \cdots \sum_{i_n=1}^{K} p_{i_n} = 1 .$$

Thus Z^n is a zero-memory source.

Z^n is also called a "finite delay source". In practice, the coder waits until a sequence $\alpha_{i_1} \alpha_{i_2} \ldots \alpha_{i_n}$ of n words is received from Z and then encodes this sequence as a single word from the zero-memory source Z^n.

LEMMA 3: Let $Z = \{(\alpha_i, p_i)\}_{i=1}^{K}$ be a zero-memory source and $Z^n = \{(\beta_j, \pi_j)\}_{j=1}^{K^n}$ its n-th extension. Then

$$H_D(Z^n) = nH_D(Z) .$$

Proof [1, p. 21] : Let each $\beta_j = (\alpha_{i_1} \alpha_{i_2} \ldots \alpha_{i_n})$. Then each $\pi_j = p_{i_1} p_{i_2} \ldots p_{i_n}$, $j = 1, 2, \ldots, K^n$. Thus

$$H_D(Z^n) = -\sum_{j=1}^{K^n} \pi_j \log_D \pi_j$$

$$= -\sum_{j=1}^{K^n} \pi_j \log_D p_{i_1} - \sum_{j=1}^{K^n} \pi_j \log_D p_{i_2} - \cdots - \sum_{j=1}^{K^n} \pi_j \log_D p_{i_n}$$

For any r, $1 \leq r \leq n$,

$$-\sum_{j=1}^{K^n} \pi_j \log_D p_{i_r} = -\sum_{i_r=1}^{K} p_{i_r} \log_D p_{i_r} \sum_{i_1=1}^{K} p_{i_1} \cdots \sum_{i_n=1}^{K} p_{i_n}$$

$$= -\sum_{i_r=1}^{K} p_{i_r} \log_D p_{i_r} = H_D(Z) .$$

Therefore $H_D(Z^n) = nH_D(Z)$. QED

DEFINITION 11: Let $Z = \{(\alpha_i, p_i)\}_{i=1}^{K}$ be a zero-memory source and $I = \{(\alpha_i, n_i)\}_{i=1}^{K}$ an instantaneous code on Z. The *average length* of I, denoted by L_I, equals $\sum_{i=1}^{K} n_i p_i$ [1, p. 66].

For a source $Z = \{(\alpha_i, p_i)\}_{i=1}^{K}$ and an instantaneous code $I = \{(\alpha_i, n_i)\}_{i=1}^{K}$ on Z, $H_D(Z) = -\sum_{i=1}^{K} p_i \log_D p_i$, $S_I = \sum_{i=1}^{K} D^{-n_i}$, and $L_I = \sum_{i=1}^{K} n_i p_i$ are independent of the word list $\{\alpha_i\}_{i=1}^{K}$. Henceforth the interests of this paper centre on the probabilities $\{p_i\}_{i=1}^{K}$ characterizing the source Z and the code word lengths $\{n_i\}_{i=1}^{K}$ of I. For this purpose, the code I shall be denoted by the relation $\{(p_i, n_i)\}_{i=1}^{K}$ with the set of probabilities $\{p_i\}_{i=1}^{K}$ fixed. To return to the $\{(\alpha_i, n_i)\}_{i=1}^{K}$ form of I, let $Z = \{(\alpha_i, p_i)\}_{i=1}^{K}$. Then $\alpha_i \to p_i \to n_i$ $(i = 1,2,\ldots,K)$ gives the customary form, aside from trivial differences in the case of equiprobable words.

DEFINITION 12: Let $\{p_i\}_{i=1}^{K}$ be a fixed set of probabilities. Then the *Shannon code* [30; 1, p. 72] $E = \{(p_i, n_i)\}_{i=1}^{K}$ is the unique code defined by:

$$-\log_D p_i \leq n_i < 1 - \log_D p_i \quad (i = 1,2,\ldots,K) .$$

By the left-hand inequality, $D^{-n_i} \leq p_i \quad (i = 1,2,\ldots K)$.

Therefore $S_I = \sum_{i=1}^{K} D^{-n_i} \leq \sum_{i=1}^{K} p_i = 1$. Hence, E is an instantaneous code.

LEMMA 4 [30; 1, pp. 72-73] :

SHANNON'S THEOREM on NOISELESS CODING

Let $Z = \{(\alpha_i, p_i)\}_{i=1}^{K}$ be a zero-memory source and $Z^n = \{(\beta_j, \pi_j)\}_{j=1}^{K^n}$ be the n-th extension. Let E and E^n be the Shannon codes $\{(p_i, n_i)\}_{i=1}^{K}$ and $\{(\pi_j, \nu_j)\}_{j=1}^{K^n}$ for Z and Z^n, respectively. Let L_n be the average length of E^n . Then

$$H_D(Z) \leq L_n / n < H_D(Z) + 1/n .$$

Proof: By Definition 12, $E^n = \{(\pi_j, \nu_j)\}_{j=1}^{K^n}$ is defined by:

$-\log_D \pi_j \leq \nu_j < 1 - \log_D \pi_j \quad (j = 1,2,\ldots,K^n)$. Then

$$-\sum_{j=1}^{K^n} \pi_j \log_D \pi_j \leq \sum_{j=1}^{K^n} \pi_j \nu_j < \sum_{j=1}^{K^n} \pi_j - \sum_{j=1}^{K^n} \pi_j \log_D \pi_j .$$

That is, $H_D(Z^n) \leq L_n < 1 + H_D(Z^n)$. By Lemma 3, $H_D(Z^n) = nH_D(Z)$. Hence $H_D(Z) \leq L_n / n < H_D(Z) + 1/n$. QED

L_n is the average length of the code words in E^n per word $\beta_j = (\alpha_{i_1} \alpha_{i_2} \ldots \alpha_{i_n})$ in Z^n . Hence, L_n / n is the average length of E^n per α_i emitted by Z .

Therefore by increasing the complexity of the encoder and decoder to transmit K^n words, the average length of the code per word from Z can be made arbitrarily close to the entropy of the source.

However Shannon's code, while instantaneous, is often inefficient. In the next section, an encoding is introduced which is, in a sense, the best instantaneous code.

HUFFMAN ENCODING

DEFINITION 13: Let $\{p_i\}_{i=1}^{K}$ be a fixed set of probabilities. A *Huffman (or optimum) code* [15] $C = \{(p_i, \tilde{n}_i)\}_{i=1}^{K}$ is an instantaneous code of minimum average length L_C.

Huffman's procedure for constructing an optimum code for a fixed set of probabilities $\{p_i\}_{i=1}^{K}$ is in two parts:

Part I:

(I : 1) List all probabilities such that $p_1 \geq p_2 \geq \cdots \geq p_K$.

(I : 2) Let $A \equiv K \pmod{D-1}$ such that $2 \leq A \leq D$.

(I : 3) Define a new probability list $\{p_i'\}_{i=1}^{K'}$ such that

$$p_i' = p_i \ , \quad i = 1, 2, \ldots, K - A$$

$$p_{K'}' = \sum_{i=K-A+1}^{K} p_i \ . \quad \text{Hence} \quad K' = K - A + 1 \ .$$

Repeat steps (I: 1, 2, 3) for $\{p_i'\}_{i=1}^{K'}$ until a list is formed with D or fewer probabilities.

LEMMA 5: In the second (or subsequent) list, let K' be the number of probabilities. Then $K' \equiv D \pmod{D-1}$; so $A \equiv K' \pmod{D-1}$, $2 \leq A \leq D$, implies $A = D$.

Proof: Let $A \equiv K \pmod{D-1}$. Then by (I : 3) A of the probabilities are summed to one new probability in the second list and $K' = K - A + 1 \equiv 1 \equiv D \pmod{D-1}$.

Next let K' be the number of probabilities in any list after the first and assume $K' \equiv D \pmod{D-1}$. Then for this list $A = D$. Therefore the succeeding list has $K'' = K' - D + 1 \equiv 1 \equiv D \pmod{D-1}$. QED

There are $[(K - A) / (D - 1)] + 1$ probability lists in this form of encoding. Let K_j, $j = 0, 1, \ldots, (K-A)/(D-1)$ represent the number of probabilities in each list, starting with the shortest list. Denote these lists by $P_j = \{p_i^j\}_{i=1}^{K_j}$, $j = 0, 1, \ldots, (K-A)/(D-1)$. [The j of p_i^j is a superscript.] By Lemma 5 and step (I : 3) $K_0 = D$. Thus $K_j = D + j(D-1)$, $j = 0, 1, \ldots, [(K-A)/(D-1)] - 1$. For $j' = (K-A)/(D-1)$, $K_{j'} = K$.

Using this terminology, (I : 3) becomes:

For P_{j+1}, $(K-A)/(D-1) > j \geq 1$, define a new probability list $P_j = \{p_i^j\}_{i=1}^{K_j}$ such that:

$$p_i^j = p_i^{j+1}, \quad i = 1, 2, \ldots, K_{j+1} - A = K_j - 1$$

$$p_{K_j}^j = \sum_{i=K_j}^{K_{j+1}} p_i^{j+1} .$$

Part II:

(II : 1) For $j = 0$, assign the code words $0, 1, \ldots, (D-1)$ to $p_1^0, p_2^0, \ldots, p_D^0$ respectively.

(II : 2) For $1 \leq j < (K-A)/(D-1)$, let $p_\alpha^j = \sum_{i=K_j}^{K_{j+1}} p_i^{j+1}$ as in (I : 3), $1 \leq \alpha \leq K_j$. Denote the code word for p_i^j of length n_i^j by X_i^j, $i = 1, 2, \ldots, K_j$. Then assign code words for P_{j+1} as follows: p_i^{j+1} is encoded by X_i^j for $i = 1, 2, \ldots, \alpha-1$ and by X_{i+1}^j for $i = \alpha, \alpha+1, \ldots, K_j-1$.

The last D probabilities of P_{j+1}, that is $p^{j+1}_{K_j+1}, p^{j+1}_{K_j+2}, \ldots, p^{j+1}_{K_{j+1}}$ are encoded by adding the code letters $0, 1, \ldots, (D-1)$ to the prefix X^j_α, respectively. Hence the code word lengths n^{j+1}_i, $i = 1, 2, \ldots, K_j-1$, are the same as the n^j_i, $i = 1, 2, \ldots, \alpha-1, \alpha+1, \ldots, K_{j+1}$. But $n^{j+1}_i = n^j_\alpha + 1$, $i = K_j, K_j+1, \ldots, K_{j+1}$.

EXAMPLE: For $D = 4$ and $K = 14 \equiv 2 \pmod 3$.

P_4 : C_4	P_3 : C_3	P_2 : C_2	P_1 : C_1	P_0 : C_0
				→1.00
				→.42 0
.20 1	.20 1	.20 1	.20 1	.20 1
			→.20 2	.20 2
.18 3	.18 3	.18 3	.18 3	.18 3
		→.12 00	.12 00	
.10 01	.10 01	.10 01	.10 01	
.10 02	.10 02	.10 02	.10 02	
.10 03	.10 03	.10 03	.10 03	
.06 20	.06 20	.06 20		
.06 21	.06 21	.06 21		
.04 22	.04 22	.04 22		
.04 23	.04 23	.04 23		
.04 000	.04 000			
.03 001	.03 001			
	→.03 002			
.02 003	.02 003			
.02 0020				
.01 0021				

$S_{C_i} = 1$, $i = 0, 1, 2, 3$ and $S_{C_4} = 254 / 256$. Also $L_{C_4} = 1.77$.

In showing that the final encoding is Huffman, it is seen that each code C_j , $j = 0, 1, \ldots, (K-A)/(D-1)$ is instantaneous and of minimum average length; that is, all are Huffman codes.

LEMMA 6: Let $C_j = \{(p_i^j, n_i^j)\}_{i=1}^{K_j}$, $j = 0, 1, \ldots, (K-A)/(D-1)$ be the encoding associated with the probability list P_j . Then $S_{C_j} = 1$, $j = 0, 1, \ldots, [(K-A)/(D-1)] - 1$. If $K \equiv D \pmod{D-1}$, then for $j' = (K-A)/(D-1)$, $S_{C_{j'}} = 1$; if not, $S_{C_{j'}} < 1$.

Proof: Let $j = 0$. Then $n_i^0 = 1$, $i = 1, 2, \ldots, D$ and $S_{C_0} = \sum_{i=1}^{D} D^{-1} = 1$.

Assume for some j , $1 \leq j < [(K-A)/(D-1)] - 1$, that $S_{C_j} = 1$. Then, by Lemma 5 $K_j \equiv K_{j+1} \equiv D \pmod{D-1}$ and $K_{j+1} = K_j + D - 1$ by (I : 3). Let p_α^j denote some probability in P_j such that $p_\alpha^j = \sum_{i=K_j}^{K_{j+1}} p_i^{j+1}$. Then encode P_{j+1} such that $n_i^{j+1} = n_\alpha^j + 1$, $i = K_j, K_j+1, \ldots, K_{j+1}$. Thus

$$S_{C_{j+1}} = \sum_{i=1}^{K_{j+1}} D^{-n_i^{j+1}} = \sum_{\substack{i=1 \\ i\neq\alpha}}^{K_j} D^{-n_i^j} + \sum_{i=K_j}^{K_{j+1}} D^{-(n_\alpha^j+1)}$$

$$= \sum_{\substack{i=1 \\ i\neq\alpha}}^{K_j} D^{-n_i^j} + D^{-n_\alpha^j} = 1 \text{ , since } S_{C_j} = \sum_{i=1}^{K_j} D^{-n_i^j} = 1 .$$

If $K \equiv D \pmod{D-1}$, the above holds for $C_{j'}$, $j' = (K-A)/(D-1)$. If $A \equiv K \pmod{D-1}$ is less than D , $\sum_{i=K_{j'-1}}^{K} D^{-n_i^{j'}} < D^{-n_\alpha^j}$ implying that $S_{C_{j'}} < 1$. QED

LEMMA 7: The encoding $C = \{(p_i^j, n_i^j)\}_{i=1}^{K_j}$ assigned to P_j, $j = 1, 2, \ldots, (K-A)/(D-1)$ by (II : 2) is a Huffman code.

Proof: For all $j = 1, 2, \ldots, (K-A)/(D-1)$, $S_{C_j} \leq 1$, by Lemma 6; so, C_j is instantaneous by Lemma 2--Huffman encoding is a "tree" method of of assigning code words with the prefix property. Hence it is sufficient to show that L_{C_j} is minimum, $j = 1, 2, \ldots, (K-A)/(D-1)$.

For $j = 0$, $C_0 = \{(p_i^0, n_i^0)\}_{i=1}^{D}$ has $n_i^0 = 1$, $i = 1, 2, \ldots, D$. Thus, no n_i^0 can be reduced, and $L_{C_0} = \sum_{i=1}^{D} n_i^0 p_i^0 = \sum_{i=1}^{D} p_i^0 = 1$ is minimum.

Assume for some C_j, $1 \leq j < (K-A)/(D-1)$, that L_{C_j} is minimum. Let p_α^j be a probability in P_j such that $p_\alpha^j = \sum_{i=K_j}^{K_{j+1}} p_i^{j+1}$. Then $n_i^{j+1} \geq n_\alpha^j + 1$, $i = K_j, K_j+1, \ldots, K_{j+1}$. To see this, assume $n_\beta^{j+1} \leq n_\alpha^j$, $K_j \leq \beta \leq K_{j+1}$. Thus

$$S_{C_{j+1}} = \sum_{i=1}^{K_j-1} D^{-n_i^j} + D^{-n_\beta^{j+1}} + \sum_{\substack{i=K_j \\ i\neq\beta}}^{K_{j+1}} D^{-n_i^{j+1}}$$

$$= \sum_{i=1}^{K_j} D^{-n_i^j} + \sum_{\substack{i=K_j \\ i\neq\beta}}^{K_{j+1}} D^{-n_i^{j+1}} > 1 \text{, since } K_{j+1} \geq K_j + 2 .$$

But this contradicts Lemma 6.

It is enough if $n_i^{j+1} = n_\alpha^j + 1$, $i = K_j, K_j+1, \ldots, K_{j+1}$. Then

$$L_{C_{j+1}} = \sum_{i=1}^{K_{j+1}} p_i^{j+1} n_i^{j+1} = \sum_{\substack{i=1 \\ i\neq\alpha}}^{K_j} p_i^j n_i^j + \sum_{i=K_j}^{K_{j+1}} p_i^{j+1} n_i^{j+1}$$

$$= \sum_{\substack{i=1 \\ i\neq\alpha}}^{K_j} p_i^j n_i^j + n_\alpha^j p_\alpha^j + \sum_{i=K_j}^{K_{j+1}} p_i^{j+1} = L_{C_j} + \sum_{i=K_j}^{K_{j+1}} p_i^{j+1} .$$

Since $\{p_i^{j+1}\}_{i=K_j}^{K_{j+1}}$ is the set of lowest probabilities in P_{j+1} by (I : 1) and (I : 3), and L_{C_j} is minimum, then $L_{C_{j+1}}$ is minimum. QED

Huffman [15] pointed out that the code word length $\tilde{n}_i$ of the encoding equals the number of times p_i was summed to form a larger probability and these new probabilities were in turn combined. Neumann [23, 9] gave a simple algorithm for finding how often these combinations were performed. An example taking the probability list previously encoded with $D = 4$ and $K = 14 \equiv 2 \pmod 3$ follows on the next page.

Neumann's method is actually a shorthand for noting in every list how each probability was formed. Just as in Huffman's procedure, first $A \equiv K \pmod{D-1}$, $2 \leq A \leq D$, probabilities are summed and then D more are successively combined. This continues until the sum 1.00 is reached. But with each summation the construction of the new probability from the original probabilities is described in the second column of table 2.

With respect to the example, the first step sums the two lowest probabilities .02 and .01 to .03, and denotes this by .03 | *2 in table 2, step 2. In the second step the four lowest probabilities are summed, namely .04, .03, and .02 from table 1 and .03 from table 2. But since .03 | *2 represents an implied summation of two probabilities from the original list, this implication is maintained by shifting *2 one place to the right. Thus .12 | *3*2 notes that .12 is the summation of five probabilities, three

	Table 1	Table 2	
Step 1	.20		
	.18		
P_4	.10		
	.10		
	.10		
	.06		
	.06		
	.04		
	.04		
	.04		
	.03		
	.02		
	(.02)		
	(.01)		
Step 2	.20	(.03)	*2
	.18		
P_3	.10		
	.10		
	.10		
	.06		
	.06		
	.04		
	.04		
	(.04)		
	(.03)		
	(.02)		
Step 3	.20	.12	*3*2
	.18		
P_2	.10		
	.10		
	.10		
	(.06)		
	(.06)		
	(.04)		
	(.04)		
Step 4	.20	(.12)	*3*2
	.18	.20	*4
P_1	(.10)		
	(.10)		
	(.10)		
Step 5	(.20)	(.20)	*4
P_0	(.18)	(.42)	*3*3*2
Step 6		1.00	*2*7*3*2

of which have been combined once and two twice. Note that in step 5, .20 | *4 and .42 | *3*3*2 are summed by first adding the corresponding summation indicators, shifting this total one place to the right and then registering the number of probabilities summed from table 1.

Continuing this way, eventually the K probabilities are summed to 1.00 . Then the implied summations may be used to assign Huffman code word lengths $\tilde{n}_i$. There will be two $\tilde{n}_i = 1$, seven $\tilde{n}_i = 2$, three $\tilde{n}_i = 3$, and two $\tilde{n}_i = 4$. Since $p_1 \geq p_2 \geq \cdots \geq p_K$, then $\tilde{n}_1 \leq \tilde{n}_2 \leq \cdots \leq \tilde{n}_K$, and so $\{(p_i, \tilde{n}_i)\}_{i=1}^{K}$ is formed. Finally by the prefix property, an encoding such as the one previously illustrated can be constructed.

Neumann's method thus has one more step then Huffman's, but the last step merely reflects the assigning of code words of length $n_i^0 = 1$ to the probability list P_0 in Huffman's procedure.

However, both Huffman's and Neumann's procedure have at least $(K-A)/(D-1)$ steps. As K increases, this becomes more and more unwieldy. In the next section another algorithm will be discussed, based on a method suggested by Shiva and Sheng[32]. There, two properties of the probability set $\{p_i\}_{i=1}^{K}$ will be assumed.

First, let $K \equiv D \pmod{D-1}$ or equivalently $K \equiv 1 \pmod{D-1}$. For if $K \equiv A \pmod{D-1}$, $2 \leq A < D$, then substitute $p_{K'} = \sum_{i=K-A+1}^{K} p_i$ as in (I : 3) of the Huffman encoding. Thus $K' \equiv D \pmod{D-1}$ as desired (Lemma 5). Having found

the Huffman encoding $\{(p_i, \tilde{n}_i)\}_{i=1}^{K'}$, by Lemma 7 the last step is to assign $\tilde{n}_i = \tilde{n}_\alpha + 1$, $i = K-A+1, K-A+2, \ldots, K$ as in (II : 2).

Note that in the binary case with $D = 2$, $K \equiv 1 \pmod{D-1}$ automatically.

Secondly, without loss of generality, take $\{p_i\}_{i=1}^{K}$ such that $p_1 \geq p_2 \geq \ldots \geq p_K$. Hence in a Huffman encoding $\{(p_i, \tilde{n}_i)\}_{i=1}^{K}$, $\tilde{n}_1 \leq \tilde{n}_2 \leq \ldots \leq \tilde{n}_K$.

CHAPTER II

ANOTHER ALGORITHM for HUFFMAN ENCODING

DEFINITION 14: Let $I = \{(p_i, n_i)\}_{i=1}^{K}$ be a code. Then define

$$R_I = 1 - S_I = 1 - \sum_{i=1}^{K} D^{-n_i} .$$

LEMMA 8: Let $E = \{(p_i, n_i)\}_{i=1}^{K}$ be a Shannon code. Then

$$R_E = \sum_{\mu=1}^{n_K} a_\mu (D - 1) D^{-\mu} \text{ ; for integers } a_\mu , \ 0 \leq a_\mu < D .$$

Proof: Since a Shannon code is instantaneous, $S_E \leq 1$, giving $R_E \geq 0$. Define $S_1 = \sum_{i=1}^{K} D^{n_K - n_i}$, $S_2 = D^{n_K}$. That is, $S_E = S_1 / S_2$. Now $D \geq 2$ implies that for any integer $a \geq 0$,

$$D^a - 1 = 0 \text{ or } D^a - 1 = (D - 1)(D^{a-1} + D^{a-2} + \ldots + 1) .$$

Thus $(D - 1) \mid (D^{n_K - n_i} - 1)$, $i = 1, 2, \ldots, K$. Hence

$$S_1 \equiv K \equiv 1 \pmod{D-1} \text{ and } S_2 \equiv 1 \pmod{D-1} .$$

Therefore $S_2 - S_1 \equiv 0 \pmod{D-1}$ and

$$R_E = (S_2 - S_1) / S_2 = B(D - 1) / D^{n_K}$$

for some integer $B \geq 0$. By repeated application of the Euclidean algorithm,

$$B = \sum_{\mu=1}^{n_K} a_\mu D^{n_K - \mu} \text{ , for integers } a_\mu , \ 0 \leq a_\mu < D .$$

Finally $R_E = (D - 1) D^{-n_K} \left(\sum_{\mu=1}^{n_K} a_\mu D^{n_K - \mu} \right) = \sum_{\mu=1}^{n_K} a_\mu (D - 1) D^{-\mu}$.

QED

DEFINITION 15: Let $\{p_i\}_{i=1}^K$ be a fixed set of probabilities. An instantaneous code $M = \{(p_i, n_i^o)\}_{i=1}^K$ is said to be *minimal* if L_M is a minimum such that each n_i^o is less than or equal to n_i of the Shannon code E.

By definition, a minimal code is not necessarily a Huffman code, and vice versa. However, Theorem 3 will show the two codes to be equivalent.

LEMMA 9: For any fixed set of probabilities $\{p_i\}_{i=1}^K$, there exists a minimal code $M = \{(p_i, n_i^o)\}_{i=1}^K$. Also $S_M = 1$.

Proof: Let $E = \{(p_i, n_i)\}_{i=1}^K$ be the Shannon code for $\{p_i\}_{i=1}^K$. Then there are at most $\Pi_{i=1}^K n_i$ distinct codes with each code word length less than or equal to n_i. Since the Shannon code is instantaneous, instantaneous codes form a non-empty subclass of this set of codes. Hence $M = \{(p_i, n_i^o)\}_{i=1}^K$ can be chosen from the finite subclass such that L_M is minimum.

By Lemma 2, $S_M \leq 1$. Assume $S_M < 1$. Define $S_1 = \sum_{i=1}^K D^{n_K^o - n_i^o}$ and $S_2 = D^{n_K^o}$. Then $S_M = S_1 / S_2$ and $S_1 < S_2$. As in the proof of Lemma 8, $S_1 \equiv S_2 \equiv 1 \pmod{D-1}$. For $0 \leq b \leq D-2$, $S_1 + b \not\equiv S_2 \pmod{D-1}$ giving $S_1 + b < S_2$. Thus $S_1 + D - 1 \leq S_2$. And therefore

$$S' = S_M + (D-1)D^{-n_K^o} = (S_1 + D - 1) / S_2 \leq 1 .$$

So, there exists a code $M' = \{(p_i, n_i')\}_{i=1}^K$ with $n_i' = n_i^o$, $i = 1, 2, \ldots, K-1$, $n_K' = n_K^o - 1$ and $S_{M'} = S' \leq 1$. By Lemma 2, M' is instantaneous and since $p_K > 0$, $L_{M'} < L_M$. This contradicts the minimality of M. QED

DEFINITION 16: Let $E = \{(p_i, n_i)\}_{i=1}^{K}$ be a Shannon code. Define

$$v_i = p_i D^{n_i} , \quad i = 1, 2, \ldots, K .$$

Then v_i is said to be the *valuation* of n_i .

THEOREM 1: For a fixed set of probabilities $\{p_i\}_{i=1}^{K}$, a minimal code $M = \{(p_i, n_i^o)\}_{i=1}^{K}$ can be found in at most $[-\log_D p_K]$ steps, where p_K is the lowest probability.

Before considering the algorithm needed to prove Theorem 1, some machinery must be introduced.

DEFINITION 17: The codes considered in the algorithm are called the *μ-th outer codes*, $\mu = n_K, n_K-1, \ldots, 2, 1$ and are denoted $E_\mu = \{(v_i, p_i, n_i)\}_{i=1}^{K}$. E_{n_K}, *the original outer code*, is composed of $E = \{(p_i, n_i)\}_{i=1}^{K}$, the Shannon code of $\{p_i\}_{i=1}^{K}$, plus the valuations v_i of E . E_μ , $\mu = n_K-1, n_K-2, \ldots, 2, 1$ represent codes evolved from E_{n_K} at successive stages of the algorithm. They are said to be *reduced outer codes*.

LEMMA 10: Let μ , $n_K \geq \mu \geq 1$, represent any distinct stage of the algorithm and let E_μ be the μ-th outer code. Then for integers a_ν , $R_{E_\mu} = \sum_{\nu=1}^{\mu} a_\nu (D - 1) D^{-\nu}$, $0 \leq a_\nu < D$.

Proof: For $\mu = n_K$, this is immediate from Lemma 8.

Assume the hypothesis for some μ , $n_K \geq \mu \geq 1$. Let $\mu' = \mu - 1$. Then

$$R_{E_\mu} = \sum_{\nu=1}^{\mu'} a_\nu (D - 1) D^{-\nu} + a_\mu (D - 1) D^{-\mu} .$$

Let M be the minimal code for $\{p_i\}_{i=1}^{K}$ of Theorem 1. Then by Lemma 9, $R_M = 0$. Also by Lemma 8, $0 \leq a_\mu < D$. Hence $a_\mu(D-1)D^{-\mu}$ can be subtracted from R_{E_μ} by, and only by, reducing sufficient $n_i \geq \mu$. As all such n_i are listed in E_μ, they are then reduced, giving, for $0 \leq a_\nu < D$

$$R_{E_{\mu'}} = R_{E_\mu} - a_\mu(D-1)D^{-\mu} = \sum_{\nu=1}^{\mu'} a_\nu(D-1)D^{-\nu} . \qquad \text{QED}$$

LEMMA 11: Let μ, $n_K \geq \mu \geq 1$, be a distinct stage of the algorithm and $E_\mu = \{(v_i, p_i, n_i)\}_{i=1}^{K}$ the μ-th outer code. Then each $n_i \geq \mu$, $1 \leq i \leq K$, is in a set of the form $\{n_{i_j} \geq \mu : j = 1,2,\ldots,F\}$ such that

$$\sum_{j=1}^{F} D^{-n_{i_j}} = D^{-\mu} .$$

If m is the number of such sets in E_μ, then $D|(m - a_\mu)$.

Proof: First let $\mu = n_K$. Then as $n_i \leq n_K$ $(i = 1,2,\ldots,K)$, $n_i \geq \mu = n_K$ implies that $n_i = \mu$. Thus each $n_i \geq \mu$ is in a singleton set $\{n_{i_1} = \mu\}$ and $D^{-n_{i_1}} = D^{-\mu}$. Let m be the number of the sets $\{n_{i_1} = n_K\}$ in E_{n_K}.

(*) Let b_ν, $\nu = 1, 2, \ldots, \mu-1$ be the number of sets $\{n_{i_1} = \nu\}$ and let $b_\mu = m$. Then

$$S_{E_\mu} = \sum_{i=1}^{K} D^{-n_i} = \sum_{\nu=1}^{\mu} b_\nu D^{-\nu} .$$

By the Euclidean algorithm, $b_\mu = qD + b$, $0 \leq b < D$ and $b, q \geq 0$ integers. Now

$$b_\mu D^{-\mu} = (qD + b)D^{-\mu} = qD^{1-\mu} + bD^{-\mu} .$$

Therefore $D^{1-\mu} \mid (S_{E_\mu} - bD^{-\mu})$. Also by Lemma 10, $R_{E_\mu} = \sum_{\nu=1}^{\mu} a_\nu (D-1)D^{-\nu}$ so that $D^{1-\mu} \mid (R_{E_\mu} - a_\mu (D-1)D^{-\mu})$. Since $\mu \geq 1$, $R_{E_\mu} + S_{E_\mu} = 1$ implies $D^{1-\mu} \mid (R_{E_\mu} + S_{E_\mu})$. Thus $D^{1-\mu} \mid \{[b + a_\mu (D-1)]D^{-\mu}\}$. Then $D \mid [b + a_\mu (D-1)]$ or $D \mid (b - a_\mu)$. As $0 \leq b, a_\mu < D$, $b = a_\mu$. Now $m = b_\mu = qD + b = qD + a_\mu$ and so $D \mid (m - a_\mu)$.

Secondly, assume the hypothesis holds for some arbitrary μ , $n_K \geq \mu \geq 1$. Let $\mu' = \mu - 1$. Since $m \geq 0$, and $D \mid (m - a_\mu)$, there are at least a_μ sets $\{n_{i_j} \geq \mu : j = 1,2,\ldots,F\}$ with $\sum_{j=1}^{F} D^{-n_{i_j}} = D^{-\mu}$. In these a_μ sets, the n_{i_j} are replaced by $n'_{i_j} = n_{i_j} - 1$. These sets are then in the form $\{n'_{i_j} \geq \mu' : j = 1,2,\ldots,F\}$ and $\sum_{j=1}^{F} D^{-n'_{i_j}} = \sum_{j=1}^{F} D^{1-n_{i_j}} = D^{-\mu'}$.

By the hypothesis, $D \mid (m - a_\mu)$. Therefore the remaining $(m - a_\mu)$ sets are combined D at a time giving $(m - a_\mu) / D$ sets of the form $\{n'_{i_j} \geq \mu > \mu' : j = 1,2,\ldots,F'\}$ where n'_{i_j} is unchanged from n_{i_j} and $\sum_{j=1}^{F'} D^{-n'_{i_j}} = D \sum_{j=1}^{F} D^{-n_{i_j}} = D^{-\mu'}$. Finally any sets $\{n_{i_1} = \mu'\}$ unreduced from the original outer code E_{n_K} have $D^{-n_{i_1}} = D^{-\mu'}$. Let m now be the number of sets $\{n'_{i_j} \geq \mu' : j = 1,2,\ldots,F'\}$. By the previous argument (*), $D \mid (m - a_{\mu'})$ QED

DEFINITION 18: Let μ , $n_K \geq \mu \geq 1$ be a distinct stage of the algorithm with μ-th outer code E_μ . Let m_μ be the number of sets in E_μ of the form given in Lemma 10. Define $N_{\mu,\lambda} = \{n_{i_j} \geq \mu : j = 1,2,\ldots,F_\lambda\}$ such that $\sum_{j=1}^{F_\lambda} D^{-n_{i_j}} = D^{-\mu}$; $\lambda = 1, 2, \ldots, m_\mu$. These $N_{\mu,\lambda}$ are said to be the *μ-blocks* of E_μ .

Thus each μ-block may have one element $n_{i_1} = \mu$ or several $n_{i_j} > \mu$, $j = 1, 2, \ldots, F_\lambda$.

DEFINITION 19: Let E_μ , $n_K \geq \mu \geq 1$ be a μ-th outer code with μ-blocks $N_{\mu,\lambda}$, $\lambda = 1, 2, \ldots, m_\mu$. Let $E_{\mu'}$ be a reduced outer code such that $\mu' = \mu - 1$ with μ'-blocks $N_{\mu',\lambda'}$, $\lambda' = 1, 2, \ldots, m_{\mu'}$. V_λ , the *common valuation* of $N_{\mu,\lambda}$ is defined inductively as follows:

For every singleton μ- or μ'-block, say $N_{\mu,\lambda} = \{n_{i_1} = \mu\}$, $V_\lambda = v_{i_1} = p_i D^{n_{i_1}}$ (hence for $\mu = n_K$, V_λ , $\lambda = 1, 2, \ldots, m_\mu$ is well defined as each $N_{\mu,\lambda}$ is a singleton set). For the reduced outer code $E_{\mu'}$, a_μ common valuations $V_{\lambda'}$ are found by dividing V_λ by D , and $(m_\mu - a_\mu) / D$ of the $V_{\lambda'}$ are derived by averaging the V_λ , taking D of them at a time.

To be specific,

$$V_{\lambda'} = V_\lambda / D \ , \quad \lambda = \lambda' = 1, 2, \ldots, a_\mu \ .$$

Let $A_{\lambda'} = \{\lambda : a_\mu + D(\lambda' - a_\mu - 1) + 1 \leq \lambda \leq a_\mu + D(\lambda' - a_\mu)\}$, $\lambda' = a_\mu+1, a_\mu+2, \ldots, a_\mu+(m_\mu-a_\mu)/D$. Then $V_{\lambda'} = \sum_{A_{\lambda'}} V_\lambda / D$.

The remaining μ'-blocks are singleton sets $\{n_{i_1} = \mu'\}$ unreduced from E_{n_K} with $V_{\lambda'} = v_{i_1}$.

For programming convenience, each n_{i_j} in $N_{\mu,\lambda}$ is given the valuation $v_{i_j} = V_\lambda$.

ALGORITHM: Repeat steps (i) and (ii) for $\mu = n_K, n_K-1, \ldots$ successively until step (iii) can be applied.

(i) List all (v_i, p_i, n_i) so that $v_1 \geq v_2 \geq \ldots \geq v_K$.

(ii) If for the first listed n_g , say $g = 1, 2, \ldots, G$

$$\sum_{g=1}^{G-1} D^{-n_g} < R_{E_\mu} < \sum_{g=1}^{G} D^{-n_g} ,$$

a reduced outer code is formed as follows--otherwise proceed to step (iii):

Consider the μ-blocks $N_{\mu,\lambda} = \{n_{i_j} \geq \mu : j = 1,2,\ldots,F_\lambda\}$, $\lambda = 1, 2, \ldots, m_\mu$. If $a_\mu \neq 0$, replace each n_{i_j} in $N_{\mu,\lambda}$ by $n'_{i_j} = n_{i_j} - 1$, $\lambda = 1, 2, \ldots, a_\mu$. Hence there are new common valuations $V_{\lambda'} = V_\lambda / D$ and $\sum_{j=1}^{F_\lambda} D^{-n'_{i_j}} = D^{1-\mu}$, $\lambda = \lambda' = 1, 2, \ldots, a_\mu$.

Now for any value of a_μ , $D|(m_\mu - a_\mu)$ by Lemma 11. Combine the μ-blocks $N_{\mu,\lambda}$, $\lambda = a_\mu+1, a_\mu+2,\ldots, m_\mu$ taking them D at a time and averaging their common valuations. That is, $N_{\mu,\lambda}$, $N_{\mu,\lambda+1}$, $\ldots$, $N_{\mu,\lambda+D-1}$ (where $\lambda = a_\mu+1$, $a_\mu+D+1$, $a_\mu+2D+1$, $\ldots$, $m_\mu-D+1$) are replaced by $N_{\mu,\lambda'} = \{n'_{i_j} \geq \mu : j = 1,2,\ldots,F_{\lambda'}\}$ such that $\sum_{j=1}^{F_{\lambda'}} D^{-n'_{i_j}} = D^{1-\mu}$ and $a_\mu + 1 \leq \lambda' \leq a_\mu + (m_\mu - a_\mu) / D$. Here the n'_{i_j} are unchanged from the n_{i_j} , $F_{\lambda'} = F_\lambda + F_{\lambda+1} + \ldots + F_{\lambda+D-1}$ and $V_{\lambda'} = (V_\lambda + V_{\lambda+1} + \ldots + V_{\lambda+D-1}) / D$.

Let $\mu' = \mu - 1$. There is a reduced outer code $E_{\mu'}$ with μ'-blocks $N_{\mu',\lambda'}$, common valuations $V_{\lambda'}$, and $R_{E_{\mu'}} = R_{E_\mu} - a_\mu D^{-\mu} = \sum_{\nu=1}^{\mu'} a_\nu D^{-\nu}$. Return to step (i).

(iii) Here, for the first listed n_g , $g = 1, 2, \ldots, G$, $\sum_{g=1}^{G} D^{-n_g} = R_{E_\mu}$. A final reduced outer code M is given by replacing the n_g by $n_g - 1$, $g = 1, 2, \ldots, G$. As $R_M = 0$, no further reductions to an instantaneous code are possible (Lemma 2). END OF ALGORITHM

LEMMA 12: Let E_μ , $n_K \geq \mu \geq 1$, be a μ-th outer code. Then for each μ-block $N_{\mu,\lambda} = \{n_{i_j} \geq \mu : j = 1, 2, \ldots, F_\lambda\}$ with common valuation V_λ , $\lambda = 1, 2, \ldots, m_\mu$

$$\sum_{j=1}^{F_\lambda} p_{i_j} = V_\lambda D^{-\mu} .$$

Proof: For $\mu = n_K$, this is immediate from Definitions 17 and 19. Assume the lemma holds for some μ , $n_K \geq \mu \geq 1$, and let $\mu' = \mu - 1$. There are three cases to be considered in the formation of μ'-blocks and common valuations in $E_{\mu'}$.

(a) If $a_\mu \neq 0$, for $\lambda' = \lambda = 1, 2, \ldots, a_\mu$ the μ'-blocks are constructed such that $N_{\mu',\lambda'} = \{n'_{i_j} \geq \mu' : j = 1,2,\ldots,F_{\lambda'}\}$ with $n'_{i_j} = n_{i_j} - 1$, for all n_{i_j} in $N_{\mu,\lambda}$. Hence

$$V_{\lambda'} D^{-\mu'} = (V_\lambda / D)(D^{1-\mu}) = V_\lambda D^{-\mu} = \sum_{j=1}^{F_\lambda} p_{i_j} = \sum_{j=1}^{F_{\lambda'}} p_{i_j} .$$

(b) For a fixed λ' , $a_\mu + 1 \leq \lambda' \leq a_\mu + (m_\mu - a_\mu)/D$, let $A_{\lambda'}$ again be defined by:

$$A_{\lambda'} = \{\lambda : a_\mu + D(\lambda' - a_\mu - 1) + 1 \leq \lambda \leq a_\mu + D(\lambda' - a_\mu)\} .$$

Then $N_{\mu',\lambda'} = \{n'_{i_j} \geq \mu : j = 1,2,\ldots,F_{\lambda'}\}$ such that

$F_{\lambda'} = \sum_{A_{\lambda'}} F_\lambda$, $\sum_{j=1}^{F_{\lambda'}} D^{-n'_{i_j}} = \sum_{A_{\lambda'}} \sum_{j=1}^{F_\lambda} D^{-n_{i_j}} = D \cdot D^{-\mu} = D^{-\mu'}$,

and $V_{\lambda'} = \sum_{A_{\lambda'}} V_\lambda / D$. Then

$$D^{-\mu'} V_{\lambda'} = (D^{1-\mu})\left(\sum_{A_{\lambda'}} V_\lambda / D\right) = D^{-\mu} \sum_{A_{\lambda'}} V_\lambda$$

$$= \sum_{A_{\lambda'}} \sum_{j=1}^{F_\lambda} p_{i_j} = \sum_{j=1}^{F_{\lambda'}} p_{i_j} .$$

(c) For $\lambda' > a_\mu + (m_\mu - a_\mu) / D$, $N_{\mu',\lambda'} = \{n_{i_1} = \mu'\}$ and by Definitions 17 and 19 $p_{i_1} = v_{i_1} D^{-\mu'} = V_{\lambda'} D^{-\mu'}$. QED

Since, for all n_{i_j} in $N_{\mu,\lambda}$, v_{i_j} is set equal to V_λ , then $V_\lambda D^{-\mu} = v_{i_j} \sum_{j=1}^{F_\lambda} D^{-n_{i_j}}$, $j = 1, 2, \ldots, F_\lambda$. Consequently $V_\lambda D^{-\mu} = \sum_{j=1}^{F_\lambda} v_{i_j} D^{-n_{i_j}}$. By Definition 17, for all $n_i < \mu$, $v_i = p_i D^{n_i}$ or $p_i = v_i D^{-n_i}$.

DEFINITION 20: Let E be the original outer code. Let M be the final reduced outer code and μ_o the final value of μ in step (iii) of the algorithm. The *change in the average length of* E_μ , the μ-th outer code, is dented by ΔL_μ and is defined as:

$$\Delta L_\mu = L_{E_\mu} - L_{E_{\mu-1}} \quad ; \quad \mu = n_K, n_K - 1, \ldots, \mu_o + 1$$

$$\Delta L_{\mu_o} = L_{E_{\mu_o}} - L_M .$$

DEFINITION 21: The *change in the average length*, denoted by ΔL, is given by

$$\Delta L = \sum_{\mu=\mu_o}^{n_K} \Delta L_\mu .$$

Proof of Theorem 1: For the original outer code $E = E_{n_K}$, $n_K \leq [-\log_D p_K] + 1$, where p_K is the lowest non-zero probability and $[x]$ is the integer part of x. Since $\mu_o \geq 1$, the number of reduced outer codes in the algorithm $(n_K - \mu_o) \leq [-\log_D p_K]$.

Since for a fixed set $\{p_i\}_{i=1}^{K}$ the original outer code E is unique and $\Delta L = L_E - L_M$, to show M is minimal it is sufficient that ΔL be maximum.

For $\mu = n_K, n_K - 1, \ldots, \mu_o + 1$, $a_\mu = 0$ implies that $\Delta L_\mu = 0$. If $a_\mu \neq 0$, by Lemma 12,

$$\Delta L_\mu = \sum_{\lambda=1}^{a_\mu} \sum_{j=1}^{F_\lambda} p_{i_j} = \sum_{\lambda=1}^{a_\mu} V_\lambda D^{-\mu}$$

Since V_λ, $\lambda = 1, 2, \ldots, a_\mu$, is maximum by step (i) of the algorithm, ΔL_μ is maximum.

For μ_o, there exists some G, and $g = 1, 2, \ldots, G$ such that $\sum_{g=1}^{G} D^{-n_g} = R_{E_{\mu_o}}$.

Hence $\Delta L_{\mu_o} = \sum_{g=1}^{G} p_g = \sum_{g=1}^{G} v_g D^{-n_g} \geq v_G \sum_{g=1}^{G} D^{-n_g} = v_G R_{E_{\mu_o}}$.

By step (i), $v_G \geq v_i$, $i \geq G + 1$. Also $v_{G+1} \geq v_1/D \geq v_g/D$, $g = 2, 3, \ldots, G$. For assume $v_{G+1} < v_1/D$. Since $n_1 < 1 - \log_D p_1$, $v_1 < D$ and so $v_{G+1} < 1$. Thus all n_i^o in M have been reduced at least once from the n_i in E.

Then $R_E \geq (D - 1)\sum_{i=1}^{K} D^{-n_i} = (D - 1)S_E$. That is, $R_E + S_E \geq D > 1$. But by Definition 14, this is impossible. Therefore all alternative sets $\{n_{i_h} : h = 1,2,\ldots,H\}$ such that $\sum_{h=1}^{H} D^{-n_{i_h}} = R_{E_{\mu_o}}$ have valuations $v_{i_h} \leq v_G$; so ΔL_{μ_o} is maximum. By Definition 21, ΔL is maximum. Since each code word of M has length bounded by the n_i of the Shannon code E , M is a minimal code. QED

THEOREM 2: Let $E = \{(p_i, n_i)\}_{i=1}^{K}$ be the Shannon code for a fixed set of probabilities $\{p_i\}_{i=1}^{K}$. Let $A = \{(p_i, n_i') : n_i' \geq n_i\}_{i=1}^{K}$. Then the code $M = \{(p_i, n_i^o)\}_{i=1}^{K}$ of minimum average length such that $n_i^o \leq n_i'$, $i = 1, 2, \ldots, K$ is a minimal code with $n_i^o \leq n_i \leq n_i'$.

Proof: By Theorem 1, this is trivial when $A = E$. So assume for some j , $1 \leq j \leq K$, that $n_j' > n_j$. Since $n_i' \geq n_i$, $i = 1, 2, \ldots, K$, then $S_A < S_E \leq 1$. By Definition 12, $n_j < 1 - \log_D p_j \leq n_j'$ implying that $v_j = p_j D^{n_j} < D \leq v_j' = p_j D^{n_j'}$. Hence n_j' is shortened to n_j with valuation always greater than or equal to D . That is, A is reduced to E . By the algorithm, E is reduced to M . As the valuation has always been maximum, the proof of Theorem 1 shows M to be a minimal code. QED

THEOREM 3: Let $E = \{(p_i, n_i)\}_{i=1}^{K}$ be the Shannon code for a fixed set of probabilities $\{p_i\}_{i=1}^{K}$. Let $C = \{(p_i, \tilde{n}_i)\}_{i=1}^{K}$ be a Huffman encoding of $\{p_i\}_{i=1}^{K}$. Then for all i , $\tilde{n}_i \leq n_i$. That is, every minimal code is a Huffman code.

Proof: Assume for some j, $1 \le j \le K$, that $\tilde{n}_j > n_j$. Let $C' = \{(p_i, n_i') : n_i' = \max[n_i, \tilde{n}_i]\}_{i=1}^{K}$. By this definition, C' can be returned to C by reducing the n_i' of the form $n_{i_q}' = n_{i_q} > \tilde{n}_{i_q}$ (i.e. $v_{i_q}' < D$). By Theorem 2, C' can also be reduced to a minimal code M. Now there exists an $n_j' = \tilde{n}_j > n_j$, such that $v_j' \ge D$. Since $n_i' \ge n_i$, $i = 1, 2, \ldots, K$, then $S_{C'} < S_E \le 1$; so $(D - 1)D^{-n_j'} \le S_E - S_{C'} \le S_M - S_{C'} = S_C - S_{C'}$ as $S_M = S_C = 1$ by Lemmas 6 and 9.

Therefore C' can be reduced to either C° or M° where $S_{M^\circ} = S_{C^\circ} = S_{C'} + (D - 1)D^{-n_j'} \le 1$, and $L_{M^\circ} = L_{C'} - p_j$, $L_{C^\circ} = L_{C'} - p_q$ ($p_q = \sum_{q=1}^{Q} p_{i_q}$ is defined to be the maximum possible decrease in $L_{C'}$ by reducing $n_{i_q}' = \tilde{n}_{i_q} > n_{i_q}$ and $D^{-n_j'} = \sum_{q=1}^{Q} D^{-n_{i_q}'}$). But $v_j' \ge D > v_{i_q}$, $q = 1, 2, \ldots, Q$, hence $p_{i_q} D^{n_{i_q}'} < p_j D^{n_j'}$ or $p_{i_q} < p_j D^{n_j' - n_{i_q}'}$. Then $\sum_{q=1}^{Q} p_{i_q} < p_j D^{n_j'} \sum_{q=1}^{Q} D^{-n_{i_q}'}$ or $p_q < p_j$. That is, $L_{M^\circ} < L_{C^\circ}$. By Theorem 2, the minimal code M is found such that $L_{M^\circ} - L_M \ge L_{C^\circ} - L_C$. Therefore $0 < L_{C^\circ} - L_{M^\circ} \le L_C - L_M$. Hence $L_M < L_C$, contradicting L_C is minimum. QED

CHAPTER III

EXAMPLES and APPLICATIONS of the ALGORITHM

The algorithm as presented in Chapter II is in a form suitable for computer programming. Examples will be given, some of which show methods that can simplify computation.

EXAMPLE 1: For $D = 2$. The original encoding $\{(p_i, \tilde{n}_i)\}_{i=1}^{K}$ is taken from Huffman [15]. Twelve steps were needed in his procedure.

Table 1

p_i	n_i	v_i	$\tilde{n}_i$
.20	3	1.60	2
.18	3	1.44	3
.10	4	1.60	3
.10	4	1.60	3
.10	4	1.60	3
.06	5	1.92	4
.06	5	1.92	5
.04	5	1.28	5
.04	5	1.28	5
.04	5	1.28	5
.04	5	1.28	5
.03	6	1.92	6
.01	7	1.28	6

In Table 1, $\sum_{i=1}^{K} p_i = 1$. Also $-\log_2 p_i \leq n_i < 1 - \log_2 p_i$ (Definition 12), $v_i = p_i 2^{n_i}$ (Definition 16), and $\tilde{n}_i$ is the length of the Huffman code words.

$S_E = 83/128$ and so $R_E = 2^{-7} + 2^{-5} + 2^{-4} + 2^{-2}$.

In Table 2, only the valuations v_i and the code word lengths n_i of the outer codes E_μ are necessary; the probabilities p_i are illustrative. Common valuations are found

by halving or averaging valuations and the final Huffman encoding arranges the probabilities in descending order with the resultant code word lengths in ascending order, as in $\{(p_i, \tilde{n}_i)\}_{i=1}^{K}$ of Table 1.

Table 2

E_7			E_6			E_5			E_4		
v_i	p_i	n_i	v_i	p_i	n_i	v_i	p_i	n_i	v_i	p_i	n_i
1.92	.06	5	1.92	.06	5	1.92	.06	(5)4	1.60	.20	(3)2
1.92	.06	5	1.92	.06	5	1.92	.06	5	1.60	.10	(4)3
1.92	.03	6	1.92	.03	6	1.60	.20	3	1.60	.10	(4)3
1.60	.20	3	1.60	.20	3	1.60	.10	4	1.60	.10	(4)3
1.60	.10	4	1.60	.10	4	1.60	.10	4	1.60	.06	5
1.60	.10	4	1.60	.10	4	1.60	.10	4	1.60	.04	5
1.60	.10	4	1.60	.10	4	1.44	.18	3	1.44	.18	3
1.44	.18	3	1.44	.18	3	1.28	.04	5	1.28	.04	5
1.28	.04	5	1.28	.04	5	1.28	.04	5	1.28	.04	5
1.28	.04	5	1.28	.04	5	1.28	.04	5	1.28	.04	5
1.28	.04	5	1.28	.04	5	1.28	.04	5	1.28	.03	6
1.28	.04	5	1.28	.04	5	1.28	.03	6	1.28	.01	6
1.28	.01	(7)6	.64	.01	6	1.28	.01	6	.96	.06	4

E_7: (1.28, .01, 7) is the only case to be considered. As 2^{-7} is a term of R_E, 7 is reduced to 6, giving the new valuation .64 = 1.28 / 2.

E_6: (1.92, .03, 6) and (.64, .01, 6) are combined, as 2^{-6} is not a term of R_E. They appear with common valuation 1.28 = (1.92 + .64) / 2 as a 5-block in E_5 since $2^{-6} + 2^{-6} = 2^{-5}$.

Multiple element μ-blocks are bracketed in E_5 and E_4.

E_5: (1.92, .06, 5) is reduced, for it is the first 5-block on the valuation list and 2^{-5} is a term of R_E. The remaining 5-blocks $N_{5,\lambda}$, $\lambda = 2, 3, \ldots, 7$ are combined in order of their appearance on the list, and their common valuations are averaged.

E_4: The first four code word lengths are reduced by 1 since $\sum_{i=1}^{4} 2^{-n_i} = 2^{-4} + 2^{-2}$, and these are the terms still to be accounted for in R_E. Hence, the final code word lengths have been found ($S_C = 1$ and $L_C = 3.42$).

EXAMPLE 2: In this and subsequent examples, the probabilities are not used after the original valuations have been computed. However, since the $\{p_i\}_{i=1}^{K}$ are in descending order, the $\{\tilde{n}_i\}_{i=1}^{K}$ when found are written in ascending order to give $C = \{(p_i, \tilde{n}_i)\}_{i=1}^{K}$.

The example is tabled on the following page. As $D = 6$, and $K = 29 \equiv 4 \pmod 5$, the last four probabilities, namely, .004, .002, .002, and .002, are summed to $p_\alpha = .01$ [denoted by parentheses]. When the Huffman code word length $\tilde{n}_\alpha$ for this substitute probability is found, set $\tilde{n}_i = \tilde{n}_\alpha + 1$, i= 26, 27, 28, 29. The substitute probability p_α and its code word length $\tilde{n}_\alpha$ is then removed from the final encoding.

The n_i and v_i are defined by the Shannon code: $-\log_6 p_i \leq n_i < 1 - \log_6 p_i$ and $v_i = p_i 6^{n_i}$, $i = 1, 2, \ldots, 25$ and $i = \alpha$. $S_E = 101 / 216$ or $R_E = 5(5)6^{-3} + 3(5)6^{-2}$.

Note that $(D - 1)D^{-n_1} = (5)6^{-1} > R_E$ and hence $n_1 = 1$ cannot be reduced. Therefore to assist calculations, v_1 is given the value 0.0 so that it will appear at the bottom

D = 5 and K = 29 ≡ 4 (mod 5)

	Shannon Code F			Huffman Code C	Valuation List ν_3	
No.	p_i	n_i	v_i	$\tilde{n}_i$	v_i	n_i
1	.20	1	1.20=0.0	1	5.40	(2) 1
2	.18	2	5.40	1	4.68	(2) 1
3	.13	2	4.68	1	4.68	(2) 1
4	.13	2	4.68	1	4.32	(3) 2
5	.05	2	1.80	2	4.32	(3) 2
6	.04	2	1.44	2	4.32	(3) 2
7	.04	2	1.44	2	4.32	(3) 2
8	.04	2	1.44	2	4.32	(3) 2
9	.03	2	1.08	2	2.16	3
10	.02	3	4.32	2	2.16	3
11	.02	3	4.32	2	2.16	3
12	.02	3	4.32	2	2.16	3
13	.02	3	4.32	2	2.16	3
14	.02	3	4.32	2	2.16	3
15	.01	3	2.16	3	1.80	2
16	.01	3	2.16	3	1.44	2
17	.01	3	2.16	3	1.44	2
18	.01	3	2.16	3	1.44	2
19	.01	3	2.16	3	1.08	2
(α)	→(.01)	3	2.16	(3)	1.08	3
20	.005	3	1.08	3	1.08	3
21	.005	3	1.08	3	1.08	3
22	.005	3	1.08	3	1.08	3
23	.005	3	1.08	3	1.08	3
24	.005	3	1.08	3	1.08	3
25	.005	3	1.08	3	0.0	1
26	.004			4		
27	.002			4		
28	.002			4		
29	.002			4		

$S_C = 1294 / 1296$ and $L_C = 1.49$.

of the valuation list.

In E_3 , step (iii) of the algorithm applies immediately, reducing the first eight n_i by 1 as indicated by the parentheses. As illustrated here, step (iii) may apply at an early stage of the algorithm; in particular, this often happens when there are a number of equal probabilities.

Finally, $C = \{(p_i, \tilde{n}_i)\}_{i=1}^{29}$ is found by ordering $\tilde{n}_1 \leq \tilde{n}_2 \leq \ldots \leq \tilde{n}_\alpha \leq \ldots \leq \tilde{n}_{25}$ and setting $\tilde{n}_i = 3 + 1$, $i = 26, 27, 28, 29$.

Because $K \not\equiv 1 \pmod{D-1}$, $S_C < 1$. END OF EXAMPLE.

Three results are given which can assist in the application of the algorithm. It is assumed that $K \equiv 1 \pmod{D-1}$.

RESULT 1: The following are equivalent:

(I) $-\log_D p_i \leq n_i < 1 - \log_D p_i$

(II) $D^{1-n_i} > p_i \geq D^{-n_i}$

(III) $1 \leq v_i < D$ where $v_i = p_i D^{n_i}$

By (II), for a given set of probabilities $\{p_i\}_{i=1}^{K}$, a sequence of inequalities is formed with a parallel sequence for code word lengths $\{n_i\}_{i=1}^{K}$:

$$p_a \geq D^{-1} > p_b \geq D^{-2} > p_c \geq D^{-3} > \ldots$$

$$n_a = 1 \; ; \quad n_b = 2 \; ; \quad n_c = 3 \; ; \quad \ldots$$

By (III), a system for finding the n_i and v_i for $\{p_i\}_{i=1}^K$ is constructed:

Let $r = 1$, and $v_i^r = p_i D^r$. If $v_i^r < 1$, increase r until $1 \leq v_i^r < D$; then $n_i = r$ and $v_i = v_i^r$.

If $p_1 \geq p_2 \geq \ldots \geq p_K$ for a Shannon code $\{(p_i, n_i)\}_{i=1}^K$, then $n_1 \leq n_2 \leq \ldots \leq n_K$ and either of the above suggested methods may be used successively for p_i, $i = 1, 2, \ldots, K$, to find n_i.

RESULT 2: Let $E = \{(p_i, n_i)\}_{i=1}^K$ be a Shannon code with $R_E = \sum_{\mu=1}^{n_K} a_\mu (D - 1) D^{-\mu}$, $0 \leq a_\mu < D$. Let Q_μ be the number of $n_i = \mu$, $\mu = n_K, n_K - 1, \ldots, 2, 1$. Then for all μ, $n_K \geq \mu \geq 1$, there exists an integer b_μ such that

$$Q_{n_K} = b_{n_K} \cdot D + a_{n_K}$$

and $$Q_\mu + b_{\mu+1} + a_{\mu+1} = b_\mu \cdot D + a_\mu \; ; \; n_K > \mu \geq 1 .$$

In fact, since $0 \leq a_\mu < D$, then $b_{n_K} = [Q_{n_K} / D]$ and $b_\mu = [(Q_\mu + b_{\mu+1} + a_{\mu+1}) / D]$, $n_K > \mu \geq 1$.

Proof: As in Lemma 12, there are three cases to be considered:

(a) $a_{\mu+1}$ is the number of μ-blocks formed by reducing $(\mu+1)$-blocks, $n_K > \mu \geq 1$.

(b) $b_{\mu+1}$ is the number of μ-blocks formed by combining $(\mu+1)$-blocks D at a time, $n_K > \mu \geq 1$.

(c) Q_μ is the number of μ-blocks $\{n_{i_1} = \mu\}$ not reduced from E, $n_K \geq \mu \geq 1$.

Therefore, let m_μ be the total number of μ-blocks. Then $m_{n_K} = Q_{n_K}$, and $m_\mu = Q_\mu + b_{\mu+1} + a_{\mu+1}$, $1 \leq \mu < n_K$. By Lemma 11, $D|(m_\mu - a_\mu)$, $1 \leq \mu \leq n_K$. QED

By means of this result, an algorithm can be constructed to find the a_μ, $n_K \geq \mu \geq 1$, of R_E.

RESULT 3: Let $M = \{(p_i, \tilde{n}_i)\}_{i=1}^K$ be a Huffman code with $K \equiv 1 \pmod{D-1}$. Then

$$\tilde{n}_{K-D+1} = \tilde{n}_{K-D+2} = \ldots = \tilde{n}_K \leq \tilde{n}_{K-D} + 1 .$$

Proof: This is immediate from Huffman's procedure [15]. QED

This result permits a defining of a new original outer code. Let $E = \{(p_i, n_i)\}_{i=1}^K$ be a Shannon code, and let $n = n_{K-D} + 1$. Define a code $E' = \{(p_i, n_i')\}_{i=1}^K$ by:

$$n_i' = n_i , \quad i = 1, 2, \ldots, K-D$$

$$n_i' = \min\{n_{K-D+1}, n\} , \quad i = K-D+1, K-D+2, \ldots, K .$$

By Theorem 3 and Result 3, $\tilde{n}_i \leq n_i'$, $i = 1, 2, \ldots, K$; so E' can serve as the original outer code.

EXAMPLE 3: (The example is tabled on the following page.)

The $\{p_i\}_{i=1}^K$ are listed in descending order. By means of Result 1, a sequence is set up:

$$.25 > p_a \geq .125 > p_b \geq .0625 > p_c \geq .03125 > \ldots$$

$$n_a = 3 ; \quad n_b = 4 ; \quad n_c = 5 ; \quad \ldots$$

Result 3 permits (.003, 8) and (.0015, 8) to be used in the original outer code. Then $v_i = p_i 2^8$, $i = K-1, K$.

	Shannon Code			Huffman Code	Valuation List	
	p_i	n_i	v_i	$\tilde{n}_i$	v_i	n_i
D = 2	.19	3	1.520	2	1.920	(4) 3
	.16	3	1.280	3	1.920	(5) 4
	.12	4	1.920	3	1.920	(6) 5
	.10	4	1.600	3	1.920	(6) 5
	.06	5	1.920	4	1.792	(7) 6
	.045	5	1.440	5	1.600	(4) 3
	.04	5	1.280	5	1.520	(3) 2
	.04	5	1.280	5	1.472	7
	.04	5	1.280	5	1.440	5
	.035	5	1.120	5	1.280	3
	.03	6	1.920	5	1.280	5
	.03	6	1.920	5	1.280	5
	.02	6	1.280	6	1.280	5
	.02	6	1.280	6	1.280	6
	.014	7	1.792	6	1.280	6
	.0115	7	1.472	7	1.280	7
	.01	7	1.280	7	1.280	7
	.01	7	1.280	7	1.280	7
	.01	7	1.280	7	1.280	7
	.01	7	1.280	7	1.120	5
	.003	9 ≃ 8	.768	8	.768	8
	.0015	10 ≃ 8	.384	8	.384	8

μ	Q_μ	$(b_{\mu+1} + a_{\mu+1})$		$b_\mu \cdot D$		a_μ
8	2		=	1·2	+	0
7	6	+ 1	=	3·2	+	1
6	4	+ 4	=	4·2	+	0
5	6	+ 4	=	5·2	+	0
4	2	+ 5	=	3·2	+	1
3	2	+ 4	=	3·2	+	0
2	0	+ 3	=	1·2	+	1
1	0	+ 2	=	1·2	+	0

$$R_E = 2^{-7} + 2^{-4} + 2^{-2}$$

By Result 2, a_μ , $n_K \geq \mu \geq 1$, is found by means of an algorithm. Here, Q_μ is the number of singleton sets $\{n_{i_1} = \mu\}$ in the original outer code, $b_\mu = [Q_\mu + (b_{\mu+1} + a_{\mu+1})]$, and $a_\mu \equiv (Q_\mu + b_{\mu+1} + a_{\mu+1}) \pmod 2$, $0 \leq a_\mu < 2$. Then

$$R_E = \sum_{\mu=1}^{n_K} a_\mu (D - 1) D^{-\mu} .$$

Then the valuation list $\{(v_i, n_i)\}_{i=1}^{K}$ is ordered such that $v_1 \geq v_2 \geq \ldots \geq v_K$. Step (iii) of the algorithm applies immediately and since $R_E = \sum_{i=1}^{7} 2^{-n_i}$ of the valuation list, the first seven n_i are reduced by 1. The Huffman code is formed as before, ordering the $\{\tilde{n}_i\}_{i=1}^{K}$ with $\tilde{n}_1 \leq \tilde{n}_2 \leq \ldots \leq \tilde{n}_K$.

EXAMPLE 4: $D = 5$. $K = 17 \equiv 1 \pmod 4$

Shannon Code			Huffman Code
p_i	n_i		$\tilde{n}_i$
.23	1		1
.18	2		1
.09	2		2
.08	2		2
.07	2		2
.0625	2		2
.058	2		2
.03	3		2
.03	3		2
.0275	3	(.1365, 2)	2
.0265	3		2
.0225	3		2
.0216	3		2
.021	3		2
.0186	3	(.0930, 2)	2
.018	3		2
.0138	3		2

μ	Q_μ		$(b_{\mu+1} + a_{\mu+1})$		$b_\mu \cdot D$		a_μ
3	15			=	3·5	+	0
2	5	+	3	=	1·5	+	3
1	1	+	4	=	1·5	+	0

$$R_E = 3(4)5^{-2}$$

In this example, it was not necessary to take valuations. As R_E is a simple expression, and the 2-blocks can easily be found, inspection of the Shannon code immediately gives the Huffman code. The three 2-blocks with highest total probability are reduced. END OF EXAMPLE.

Another form of the algorithm does not calculate the common valuations of all μ-blocks, but restricts itself to μ-blocks of high valuation, $\mu = n_K, n_K - 1, \ldots, 2, 1$. It investigates the original outer code E_{n_K}, and selects sufficient $n_i \geq \mu$ to form the a_μ μ-blocks of highest valuation. By Lemma 12, these are also the a_μ μ-blocks of highest total probability. All the n_{i_j} in the selected μ-blocks are replaced by $n'_{i_j} = n_{i_j} - 1$, written one space to the right of n_{i_j}. The moving of n'_{i_j} one column to the right is equivalent to setting its valuations $v'_{i_j} = v_{i_j} / D$.

The n_i which have not been reduced have in the first column $v_i \geq 1$, in the second $1 > v_i \geq D^{-1}$, in the third $D^{-1} > v_i \geq D^{-2}$, etc. Thus the n_i remain in order of valuation if one proceeds down the first column, then the second, third, etc., ignoring the reduced n_i.

Any n_i reduced by means of Result 3 is found in the original outer code and hence appears in the first column. Since the valuations of such n_i are less than 1, these n_i properly belong in the second, third, or following columns. But $v_i < 1$ places them at the bottom of the first column. Therefore they are reduced only after the n_{K-D} of the Shannon code is reduced. By Result 3, this is a legitimate procedure. Thus these n_i may remain in the first column until further reduced.

EXAMPLE 5:

Huffman Encoding for the Dewey Distribution of the English Alphabet

(The example is tabled on the two following pages.)

Phase 1: Since 2^{-11} is a term of R_E, an 11-block must be reduced. (1.6384, .0008, 11) is the first 11-block on the list and so has the highest valuation of all 11-blocks. Writing (11) 10 represents the division of $v_i = 1.6384$ by 2.

To subtract 2^{-10} from R_E, a 10-block is reduced. The first listed $n_i \geq 10$ are marked with a * . Only the first three $n_i \geq 10$ are so marked, as (1.6384, .0008, 11) and (1.0240, .0005, 11) form a 10-block which includes the $n_i \geq 10$ of highest valuation. The unmarked 10-blocks then have valuations no greater than this one. Also (1.3312, .0013, 10) is a 10-block in itself and needs only to be compared to 10-blocks composed of one or more $n_i \geq 10$ of higher valuation.

D = 2 .

	Shannon Code			Huffman Code
	p_i	n_i	v_i	$\tilde{n}_i$
Space	.1859	3	1.4872	3
E	.1031	4	1.6496	3
T	.0796	4	1.2736	4
A	.0642	4	1.0272	4
O	.0632	4	1.0112	4
I	.0575	5	1.8400	4
N	.0574	5	1.8368	4
S	.0514	5	1.6448	4
R	.0484	5	1.5488	4
H	.0467	5	1.4944	4
L	.0321	5	1.0272	5
D	.0317	5	1.0144	5
U	.0228	6	1.4592	5
C	.0218	6	1.3952	5
F	.0208	6	1.3312	6
M	.0198	6	1.2672	6
W	.0175	6	1.1200	6
Y	.0164	6	1.0496	6
G	.0152	7	1.9456	6
P	.0152	7	1.9456	6
B	.0127	7	1.6256	6
V	.0083	7	1.0624	7
K	.0049	8	1.2544	8
X	.0013	10	1.3312	10
J	.0008	11	1.6384	10
Q	.0008	11	1.6384	10
Z	.0005	11	1.0240	10

$$R_E = 2^{-11} + 2^{-10} + 2^{-7} + 2^{-6} + 2^{-2} .$$

Valuation List

v_i	p_i	n_i		
		Phase 1	Phase 2	Phase 3
1.9456	.0152	7	(7) 6	(7) 6
1.9456	.0152	7	(7) 6	(7) 6
1.8400	.0575	5	5 *	(5) 4
1.8368	.0574	5	5 *	(5) 4
1.6496	.1031	4	4 *	(4) 3
1.6448	.0514	5	5 *	(5) 4
1.6384	.0008	(11) 10	(11) 10	(11) 10
1.6384	.0008	11 *	(11) 10	(11) 10
1.6256	.0127	7	(7) 6	(7) 6
1.5488	.0484	5	5 *	(5) 4
1.4944	.0467	5	5 *	(5) 4
1.4872	.1859	3	3 *	3
1.4592	.0228	6	6 *	(6) 5
1.3952	.0218	6	6 *	(6) 5
1.3312	.0208	6	6	6
1.3312	.0013	10 *	10	10
1.2736	.0796	4	4	4
1.2672	.0198	6	6	6
1.2544	.0049	8	8	8
1.1200	.0164	6	6	6
1.0624	.0083	7	7	7
1.0496	.0164	6	6	6
1.0272	.0642	4	4	4
1.0272	.0321	5	5	5
1.0240	.0005	11 *	(11) 10	(11) 10
1.0144	.0317	5	5	5
1.0112	.0632	4	4	4

From these marked $n_i \geq 10$, two 10-blocks can be formed:

1.3312	.0008	11	1.3312	.0013	10
1.3312	.0005	11			

Since the common valuations and total probabilities are equal, the choice between the two 10-blocks is arbitrary. It will prove advantageous to reduce the one with the most $n_i \geq 10$.

Phase 2: Here 2^{-7} can be immediately subtracted from R_E by reducing (1.9456, .0152, 7) , the first 7-block on the list.

Similarly 2^{-6} is subtracted as

(1.9456, .0152, 7)
(1.6256, .0127, 7)

is the first 6-block on the list.

For 2^{-2} , a choice must be made from the $n_i \geq 2$ marked by a * . There are two 2-blocks to be considered since the first 3-block on the list is obvious.

.0575	5	.0575	5
.0574	5	.0574	5
.1031	4	.1031	4
.0514	5	.1859	3
.0484	5		
.0467	5		
.0228	6		
.0218	6		

The left is equivalent to (.4091, 2) and the right to (.4039, 2) . Hence each $n_i \geq 2$ in the right-hand 2-block is reduced by 1 . As $R_E = 0$, the Huffman encoding is given in Phase 3. END OF EXAMPLE.

Schwartz [29] has pointed out the desirability of having the largest Huffman code word length $\tilde{n}_K$ and the sum of the code word lengths $\sum_{i=1}^{K} \tilde{n}_i$ as short as possible.

RESULT 4: Let $\{(v_i, p_i, n_i)\}_{i=1}^{K}$ be an outer code. To minimize $\tilde{n}_K$ and $\sum_{i=1}^{K} \tilde{n}_i$ of the Huffman code $C = \{(p_i, \tilde{n}_i)\}_{i=1}^{K}$, list all (v_i, p_i, n_i) of equal valuation in descending order of n_i. Then apply the algorithm.

Proof: Though the Shannon code, and hence the original outer code, is unique for a fixed set of probabilities $\{p_i\}_{i=1}^{K}$, the Huffman encoding need not be. This occurs when a choice must be made to reduce n_i of equal valuations. Since $n_K \geq n_{K-1} \geq \ldots \geq n_1$, by listing the n_i of equal valuation in descending order, n_K will be reduced whenever its place on the valuation list requires it. That is, $\tilde{n}_K$ will be as short as possible.

Also, reducing the largest n_i gives the smallest increase D^{-n_i} in S_E; so the total number of reductions is maximized. Therefore $\sum_{i=1}^{K} \tilde{n}_i$ is a minimum for a Huffman encoding C. QED

Let $\{(p_i, n_i)\}_{i=1}^{K}$ be a Shannon code with $p_1 \geq p_2 \geq \ldots \geq p_K$. In practice, Result 4 is implemented by listing the valuations of E_{n_K} such that $v_1 \geq v_2 \geq \ldots \geq v_K$ by some method that examines in turn (v_i, p_i, n_i), $i = K, K-1, \ldots, 2, 1$. Since $n_K \geq n_{K-1} \geq \ldots \geq n_2 \geq n_1$, the (v_i, p_i, n_i) of equal valuation are listed in order from the largest n_i to the smallest.

By this ordering of E_{n_K}, the μ-blocks with the most members also have the largest values of n_i; hence, either criterion can be used in listing the μ-blocks with equal common valuations.

EXAMPLE 6: (The $\{p_i\}_{i=1}^{K}$ of Example 1 are encoded using Results 1, 2, 3, and 4.)

D = 2 .

Shannon Code			Huffman Code
p_i	n_i	v_i	$\tilde{n}_i$
.20	3	1.60	2
.18	3	1.44	3
.10	4	1.60	3
.10	4	1.60	3
.10	4	1.60	4
.06	5	1.92	4
.06	5	1.92	4
.04	5	1.28	5
.04	5	1.28	5
.04	5	1.28	5
.04	5	1.28	5
.03	6	1.92	5
.01	7 ≃ 6	.64	5

μ	Q_μ		$(b_{\mu+1} + a_{\mu+1})$		$b_\mu \cdot D$		a_μ
6	2			=	1•2	+	0
5	6	+	1	=	3•2	+	1
4	3	+	4	=	3•2	+	1
3	2	+	4	=	3•2	+	0
2	0	+	3	=	1•2	+	1
1	0	+	2	=	1•2	+	0

$$\therefore R_E = 2^{-5} + 2^{-4} + 2^{-2}$$

Valuation List

v_i	p_i	n_i			
		Phase 1	Phase 2	Phase 3	Phase 4
1.92	.03	6 *	6 *	(6) 5	(6) 5
1.92	.06	5 *	(5) 4	(5) 4	(5) 4
1.92	.06	5	5 *	(5) 4	(5) 4
1.60	.10	4	4 *	4 *	(4) 3
1.60	.10	4	4	4 *	(4) 3
1.60	.10	4	4	4 *	4
1.60	.20	3	3	3 *	(3) 2
1.44	.18	3	3	3	3
1.28	.04	5	5	5 *	5
1.28	.04	5	5	5 *	5
1.28	.04	5	5	5	5
1.28	.04	5	5	5	5
.64	.01	6 *	6 *	(6) 5	(6) 5

In Example 1: $\tilde{n}_K = 6$; $\sum_{i=1}^{K} \tilde{n}_i = 55$.

In Example 6: $\tilde{n}_K = 5$; $\sum_{i=1}^{K} \tilde{n}_i = 53$.

In both examples: $S_C = 1$, $L_C = 3.42$.

CODING WITH CONSTRAINTS

Karp [16] suggests the problem of constructing instantaneous codes $A = \{(p_i, \hat{n}_i)\}_{i=1}^{K}$ subject to "constraints" and having least average length. These constraints normally are dictated by physical limitations of the encoder and the decoder. One case is an upper bound, $N \geq 1$, placed on the code word lengths; that is, $\hat{n}_i \leq N$, $i = 1, 2, \ldots, K$.

To find A, a procedure analogous to the algorithm for minimal encoding is used. Assume that $p_1 \geq p_2 \geq \cdots \geq p_K$, and that $\log_D K \leq N < n_K$, for n_K from the Shannon code, $E = \{(p_i, n_i)\}_{i=1}^{K}$. Let the original outer code be defined by $E_N = \{(p_i, N)\}_{i=1}^{K}$. Then $S_{E_N} \leq 1$ (or $R_{E_N} \geq 0$), and the code word lengths with highest valuation are reduced until $R_A = 0$ is reached.

Usually this procedure can be shortened by defining the original outer code as $E' = \{(p_i, n_i') : n_i' = \min[n_i, N]\}_{i=1}^{K}$. For if $R_{E'} \geq 0$, then E_N reduces to E' in a manner similar to Theorem 2. Finally, reduce the n_i' in E' until $R_{E'} = 0$.

EXAMPLE 7: $D = 3$. $K = 17 \equiv 1 \pmod 2$. $N = 3$.

p_i	n_i	n_i'	$\hat{n}_i$	$\tilde{n}_i$
.18	2	2	2	2
.12	2	2	2	2
.12	2	2	2	2
.08	3	3	2	2
.075	3	3	2	2
.07	3	3	3	2
.06	3	3	3	3
.05	3	3	3	3
.05	3	3	3	3
.045	3	3	3	3
.04	3	3	3	3
.03	4	3	3	3
.03	4	3	3	3
.02	4	3	3	3
.01	5	3	3	4
.01	5	3	3	4
.01	5	3	3	4

$L_{E'} = 2.425$ $L_C = 2.385$

CHAPTER IV

ENTROPY and HUFFMAN ENCODING

Shannon's "Theorem on Noiseless Coding" (Lemma 4) states that the capacity of a channel can be made arbitrarily close to, though not less than, the entropy of a source, $Z = \{(\alpha_i, p_i)\}_{i=1}^{K}$. This is accomplished by encoding sequences of two or more source symbols rather than encoding each symbol emitted. Let $C = \{(p_i, \tilde{n}_i)\}_{i=1}^{K}$ be a Huffman code for Z, and C^n a Huffman encoding of Z^n, the n-th extension of Z. Abramson [1, p.87] notes that L_C, $L_{C^2} / 2$, $L_{C^3} / 3$, ... rapidly approach $H_D(Z)$. Typically L_C or at most $L_{C^2} / 2$ closely approximate $H_D(Z)$. We wish to illustrate why Huffman encoding is efficient in this respect.

THEOREM 4: Let $C = \{(p_i, \tilde{n}_i)\}_{i=1}^{K}$ and $E = \{(p_i, n_i)\}_{i=1}^{K}$ be the Huffman and Shannon codes, respectively, for a zero-memory source $Z = \{(\alpha_i, p_i)\}_{i=1}^{K}$ with entropy $H_D(Z)$. Let $r_i = n_i - \tilde{n}_i$, $i = 1, 2, \ldots, K$. Then

$$L_C - H_D(Z) = \sum_{i=1}^{K} p_i(\log_D v_i - r_i) .$$

Proof: By Theorem 3, $\tilde{n}_i \leq n_i$; hence $r_i \geq 0$, $i = 1, 2, \ldots, K$. Note the identity:

$$n_i = \log_D[(p_i D^{n_i}) / p_i] = \log_D v_i - \log_D p_i , \quad i = 1, 2, \ldots, K .$$

Then
$$L_C - H_D(Z) = \sum_{i=1}^{K} \tilde{n}_i p_i - (-\sum_{i=1}^{K} p_i \log_D p_i)$$
$$= \sum_{i=1}^{K} p_i(\tilde{n}_i + \log_D p_i) = \sum_{i=1}^{K} p_i(n_i - r_i + \log_D p_i)$$
$$= \sum_{i=1}^{K} p_i(\log_D v_i - r_i) . \qquad \text{QED}$$

Hence, if r_i is approximately equal to $\log_D v_i$, then $L_C - H_D(Z)$ is small, and the Huffman encoding C is efficient.

Suppose for some j, $1 \leq j \leq K$, that r_j does not "approximate" $\log_D v_j$. Since $0 \leq \log_D v_j < 1$, assume that $r_j \geq 2$. Then $\tilde{n}_j = n_j - r_j$; thus $n_j' = n_j - 1$ with valuation $v_j' = p_j / D^{1-n_j} < 1$ has been reduced. This implies that the decrease p_j in L_E when we reduce n_j' is less than the increase D^{1-n_j} in S_E. Should this occur many times, the Huffman code proves to be inefficient.

Thus an efficient Huffman encoding occurs whenever $n_i - \tilde{n}_i$ approximately equals $\log_D v_i$, $i = 1, 2, \ldots, K$. Since $\log_D v_i = n_i + \log_D p_i$, then $\tilde{n}_i$ approximately equals $-\log_D p_i$. Unlike the Shannon code, though, $\tilde{n}_i$ can be greater than, equal to, or less than, $-\log_D p_i$.

One cause of inefficient Huffman encoding is illustrated in the following example. Here, for a source Z with probabilities $\{p_i\}_{i=1}^{6}$, $p_1 = .80$, $v_1 = 1.60$, $\log_2 v_1 = .678$, but n_1 cannot be reduced. Consequently, n_i, $i = 2, 3, 4, 5, 6$ are reduced twice giving an inefficient Huffman code C. However, by taking the Huffman encoding, C^2, of Z^2, the second extension of Z, $p_1 = .64$, $v_1 = 1.28$, $\log_2 v_1 = .356$. Then $\tilde{n}_1 - n_1 = 0$ more closely approximates $\log_2 v_1$, and only three of the n_i, $i = 2, 3, \ldots, 36$ are reduced twice. C^2 can be considered efficient.

EXAMPLE 8: $D = 2$.

Z	E	C	Z^2	E^2	C^2
p_i	n_i	$\tilde{n}_i$	p_i	n_i	$\tilde{n}_i$
.80	1	1	.64	1	1
.06	5	3	.048	5	4
.05	5	3	.048	5	4
.04	5	3	.04	5	4
.03	6	4	.04	5	4
.02	6	4	.032	5	5
			.032	5	5
			.024	6	5
			.024	6	5
			.016	6	5
			.016	6	5
			.0036	9	8
			.003	9	8
			.003	9	8
			.0025	9	8
			.0024	9	8
			.0024	9	8
			.002	9	8
			.002	9	8
			.0018	10	8
			.0018	10	8
			.0016	10	9
			.0015	10	9
			.0015	10	9
			.0012	10	9
			.0012	10	9
			.0012	10	9
			.0012	10	9
			.001	10	9
			.001	10	9
			.0009	11	10
			.0008	11	10
			.0008	11	10
			.0006	11	10
			.0006	11	10
			.0004	12	10

$$L_C = 1.45$$

$$L_{C^2} / 2 \simeq 1.20$$

$$H_D(Z) \simeq 1.17$$

$$\frac{H_D(Z)}{L_C} \simeq .807$$

$$\frac{H_D(Z)}{L_{C^2} / 2} \simeq .975$$

DEFINITION LIST

	Definition No.	page
average length	11	8
block code	3	1
change in the average length	21	28
change in the average length(E_μ)	20	27
code	3	1
code letter	2	1
code word	3	1
coding alphabet	2	1
common valuation	19	24
entropy	9	6
Huffman code	13	10
instantaneous code	5	1
length of a code word	3	1
minimal code	15	20
n-th extension of a source	10	7
optimum code	13	10
original outer code	17	21
prefix	6	2
reduced outer code	17	21
Shannon code	12	9
uniquely decodable code	4	1
valuation	16	21
word	1	1
word list	1	1
zero-memory source	8	6
μ-blocks	18	24
μ-th outer code	17	21
R_I	14	19
S_I	7	6

BIBLIOGRAPHY

1. N. Abramson (1963): *Information Theory and Coding*, McGraw - Hill Book Company, Inc., New York.

2. P. Beckmann (1957): *Probability in Communication Engineering*, Harcourt, Brace, and World, Inc., New York.

3. P. Billingsley (1961): "On the Coding Theorem for the Noiseless Channel", *The Annals of Mathematical Statistics*, Vol. 32, No. 2, pp. 594 - 601.

4. N. M. Blachman (1954): "Minimum Cost Encoding of Information", *Transactions of the I.R.E. on Information Theory*, Vol. PGIT - 3, pp. 139 - 149.

5. D. L. Blackwell, L. Breiman, and A. J. Thomasian (1958): "Proof of Shannon's Transmission Theorem for Finite-state Indecomposible Channels", *The Annals of Mathematical Statistics*, Vol. 29, No. 4, pp. 1209 - 1220, December.

6. E. L. Blokh and A. A. Kharkevich (1956): "Geometrical Proof of Shannon's Theorem", *Radiotekh*, Vol. 11, pp. 5 - 16, November.

7. R. C. Bose and R. R. Kuebler, Jr. (1960): "A Geometry of Binary Sequences Associated with Group Alphabets in Information Theory", *The Annals of Mathematical Statistics*, Vol. 31, pp. 113 - 139, March.

8. L. Brillouin (1951): "Information Theory and Entropy", *The Journal of Applied Physics*, Vol. 22, No. 3, pp. 334 - 341.

9. E. L. Cohen (1964): "An Account of Two New Results on Coding for the Discrete Noiseless Channel", W - 7280, The MITRE Corporation, Bedford, Mass., September.

10. R. M. Fano (1949): "The Transmission of Information, I", Technical Report No. 65, Research Laboratory of Electronics, M.I.T., Cambridge, Mass.

11. R. M. Fano (1961): *The Transmission of Information*, John Wiley and Sons, Inc., New York.

12. E. N. Gilbert and E. F. Moore (1959): "Variable Length Binary Encodings", *Bell System Technical Journal*, Vol.38, pp. 933 - 968, July.

13. R. V. L. Hartley (1928): "Transmission of Information", *Bell System Technical Journal*, Vol. 7, pp. 535 - 563.

14. I. I. Hirschman (1957): "A Note on Entropy", *American Journal of Mathematics*, Vol. 79, pp. 152-156.

15. D. A. Huffman (1952): "A Method for the Construction of Minimum Redundancy Codes", *Proceedings of the I.R.E.*, Vol. 40, pp. 1098 - 1101.

16. R. M. Karp (1961): "Minimum Redundancy Coding for the Discrete Noiseless Channel", *I.R.E. Transactions on Information Theory*, Vol. IT - 7, pp. 27 - 38, January.

17. J. Karush (1961): "A Simple Proof of an Inequality of McMillan", *I.R.E. Transactions on Information Theory*, Vol. IT - 7, No. 2, p. 118, April.

18. J. L. Kelly, Jr. (1956): "A New Interpretation of Information Rate", *Bell System Technical Journal*, Vol. 35, pp. 917 - 927.

19. A. A. Kharkevich (1956): "On Optimum Codes", *Electrosyvaz*, Vol. 10, No. 2, pp. 65 - 70.

20. L. G. Kraft (1949): "A Device for Quantizing, Grouping, and Coding Amplitude Modulated Pulses", M. S. Thesis, Electrical Engineering Department, M. I. T., Cambridge, March.

21. B. McMillan (1953) "The Basic Theorems of Information Theory", *The Annals of Mathematical Statistics*, Vol. 24, pp. 196 - 219.

22. B. McMillan (1956): "Two Inequalities Implied by Unique Decipherability", *I.R.E. Transactions on Information Theory*, Vol. IT - 2, pp. 115 - 116, December.

23. P. G. Neumann (1960): "Funktionale Prefixcodes als Grundlage der praktischen Verschüsselung", Dr. rer. nat. thesis, Institut für Praktische Mathematik, Technische Hochschule, Darmstadt, W. Germany, June.

24. P. G. Neumann (1961): "Codes auf der Grundlage von Schaltfunktionen und ihre Anwendung in der Praxis der Verschüsselung", *Nachrichtentechnische Zeitschrift*, Vol. 14, No. 5 and 6, pp. 254 - 261, 307 - 312, May and June.

25. P. G. Neumann (1962): "Efficient Error-limiting Variable-length Codes", Ph. D. dissertation, Harvard University, Cambridge; also in *I.R.E. Transactions on Information Theory*, Vol. IT-8, No. 4, July.

26. P. G. Neumann (1962): "On a Class of Efficient Error Limiting Codes", *I.R.E. Transactions on Information Theory*, Vol. IT - 8, pp. S260 - S266.

27. F. M. Reza (1961): *An Introduction to Information Theory*, McGraw - Hill Book Company, Inc., New York.

28. A. A. Sardinas and G. W. Patterson (1953): "A Necessary and Sufficient Condition for the Unique Decomposition of Coded Messages", *I.R.E. Convention Record*, Pt. 8, pp. 104 - 108.

29. E. S. Schwartz (1964): "An Optimum Encoding with Minimum Longest Code and Total Number of Digits", *Information and Control*, Vol. 7, pp. 37 - 44.

30. C. E. Shannon (1948): "A Mathematical Theory of Communication", *Bell System Technical Journal*, Vol. 27, pp. 379 - 423.

31. C. E. Shannon and W. Weaver (1949): *The Mathematical Theory of Communication*, The University of Illinois Press, Urbana.

32. S. G. S. Shiva and C. L. Sheng (1967): "A Note on Optimum Coding", Technical Report No. 67 - 4, Department of Electrical Engineering, University of Ottawa, Ottawa, Canada.

ERRATA

Page 5, line 6T*: Replace second last "$\leq$" by "=" .

Page 7, line 3T: Replace the line by: "of the information one receives if α_i did occur. It also justifies..." .

Page 7, line 8T: Replace "i = 1, 2, ..." by "j = 1, 2, ..." .

Page 14, line 5B: Replace "=" by "$\geq$" .

Page 20, line 6T: Replace the second sentence by: "Theorem 3 states a sufficient condition for the two codes to be equivalent" .

Page 20, line 7B: Replace "$0 \leq b$ ..." by "$0 < b$ ..." .

Page 24, line 3T: Replace "Lemma 10" by "Lemma 11" .

Page 25, line 7T: Replace the line by:

$$(D-1)\sum_{g=1}^{G-1} D^{-n_g} < R_{E_\mu} < (D-1)\sum_{g=1}^{G} D^{-n_g} .$$

Page 26, line 5T: Replace the equation by:

$$(D-1)\sum_{g=1}^{G} D^{-n_g} = R_{E_\mu} .$$

Page 27, line 3B: Replace "dented" by "denoted" .

---

* nT: n-th line from top

mB: m-th line from bottom

Page 29, line 12T: Replace the line by: "is a minimal code with $n_i^o \leq n_i \leq n_i^j$ if the following sufficiency conditions hold:

(a) $n_i^o \geq n_i - 1$, $i = 1, 2, \ldots, K$.

(b) If $n_{i_j}^o < n_{i_j}$, $j = 1, 2, \ldots, J$, then

$v_{i_j} \geq v_{i_r}$, for any r such that $r > J$."

Page 29, lines 5B to 1B: Replace these lines by:

"Theorem 3: Let $E = \{(p_i, n_i)\}_{i=1}^K$ be the Shannon code for a fixed set of probabilities $\{p_i\}_{i=1}^K$. Let $M = \{(p_i, n_i^o)\}_{i=1}^K$ be a minimal code satisfying conditions (a) and (b) of Theorem 2. Let $C = \{(p_i, \tilde{n}_i)\}_{i=1}^K$ be a Huffman eocoding of $\{p_i\}_{i=1}^K$. Then for all i, $\tilde{n}_i \leq n_i$. That is, every minimal code formed by reducing only the maximum valuations is a Huffman code."

Page 34, line 7T: Under p_i , replace ".18" by ".15" .

Page 43, line 8B: Under p_i , replace ".0164" by ".0175"

Page 49, line 6B: Delete: "By Theorem 3, $\tilde{n}_i \leq n_i$; hence $r_i \geq 0$, $i = 1, 2, \ldots, K$.