VARIABLE LENGTH CODING FOR CORRELATED INFORMATION SOURCES

by

David W. Bailey, B.Eng.

A thesis submitted to the Faculty of Graduate Studies and Research
in partial fulfillment of the requirements for the degree of
Master of Engineering.

Department of Electrical Engineering,
McGill University,
Montreal, Quebec.

March, 1975.

© David W. Bailey 1975

ABSTRACT

The theory of variable length coding for a discrete memoryless information source is extended to the problem of two correlated sources. It is well known that the output sequence from a single source X can be encoded and subsequently reconstructed by a decoder with zero probability of error if and only if the average codeword length $\overline{n}_x$ satisfies $\overline{n}_x \geq H(X)$. This familiar conclusion is generalized to cover correlated source coding under several different assumptions about the encoders and decoders. A method is developed to determine what minimum average codeword lengths $\overline{n}_x$ and $\overline{n}_y$ are needed in order to achieve zero-error communication for any pair of correlated sources X and Y. The results are presented as an admissible rate region in the $\overline{n}_x - \overline{n}_y$ plane.

RESUME

La théorie du codage à longueur variable pour une source d'information discrète sans mémoire est étendue au cas de deux sources correllées. Il est bien connu que la séquence de sortie d'une source unique X peut être codée et par la suite reconstruite par un décodeur avec une probabilité nulle d'erreur si et seulement si la longueur moyenne $\bar{n}_x$ du mot codé satisfait la relation $\bar{n}_x \geq H(X)$. Ce fait connu est généralisé au cas de codage de sources correllées, grâce à un certain nombre d'hypothèses concernant les codeurs - décodeurs. Une méthode, permettant de déterminer les longueurs moyennes minimales $\bar{n}_x$ et $\bar{n}_y$ des mots codés afin d'obtenir la communication sans erreur pour deux sources correllées X et Y, est développée. Les résultats sont présentés sous forme d'une région à taux admissible dans le plan $\bar{n}_x - \bar{n}_y$.

ACKNOWLEDGEMENT

The author expresses sincere thanks to Dr. Michael J. Ferguson who proposed and supervised the project. His patience and encouragement were greatly appreciated.

Ms. Hyland did an excellent job of typing. The author thanks her.

This research was supported by the National Research Council of Canada.

TABLE OF CONTENTS

CHAPTER I

INTRODUCTION

The purpose of this thesis is to extend the familiar theory of variable length coding for a discrete memoryless information source to the more general situation of two correlated sources.

One of the most interesting problems concerning correlated source coding results when the encoders and decoders are arranged as illustrated in Figure 1-1. Notice that although each encoder is restricted to see the output sequence from only one source, the decoder is allowed to observe both of the encoded message streams. Systems of this type (and other related configurations) are studied in detail in this thesis to determine what minimum average codeword lengths $\bar{n}_x$ and $\bar{n}_y$ are required by the encoders in order that the decoder can reconstruct the source output sequences with zero probability of error. The results are presented as an allowable rate region in the $\bar{n}_x - \bar{n}_y$ plane.

As will be shown in Chapter III, a typical problem having the form of Figure 1-1, might have an allowable rate region of the nature indicated in Figure 1-2. The important implication of such a rate region is that it is possible for the outputs of two correlated sources to be communicated to a decoder with zero distortion by using encoders whose average codeword lengths satisfy $\bar{n}_x < H(X)$ and $\bar{n}_y < H(Y)$. This is an improvement over the classical situation illustrated in Figure 1-3, in which the two sources are encoded and decoded independently, which

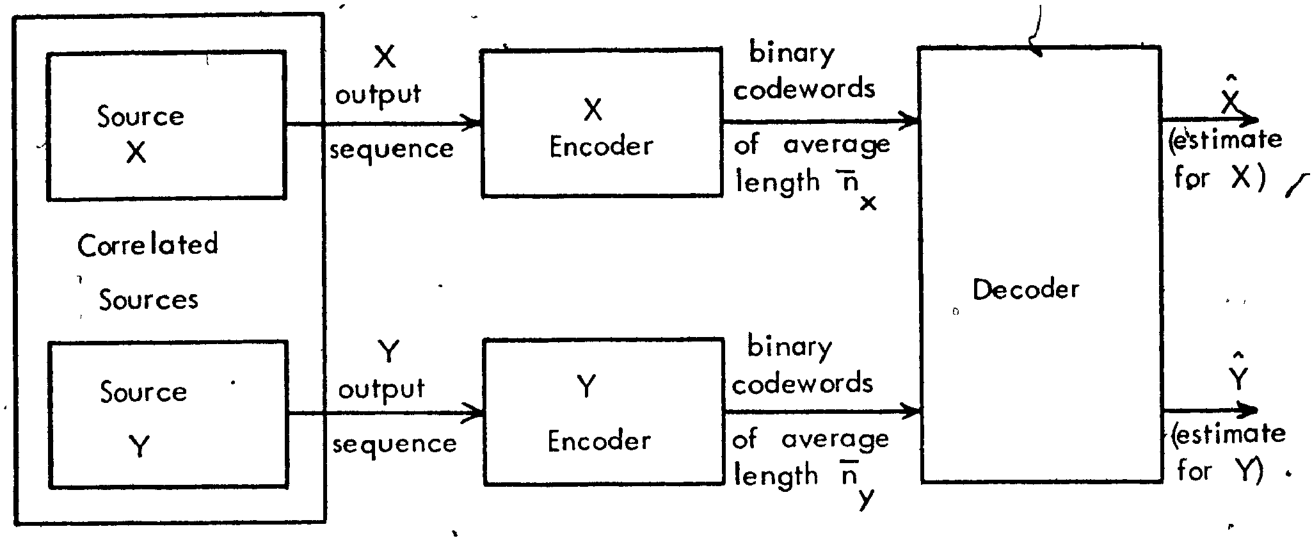


Figure 1-1 : A Correlated Source Coding Problem.

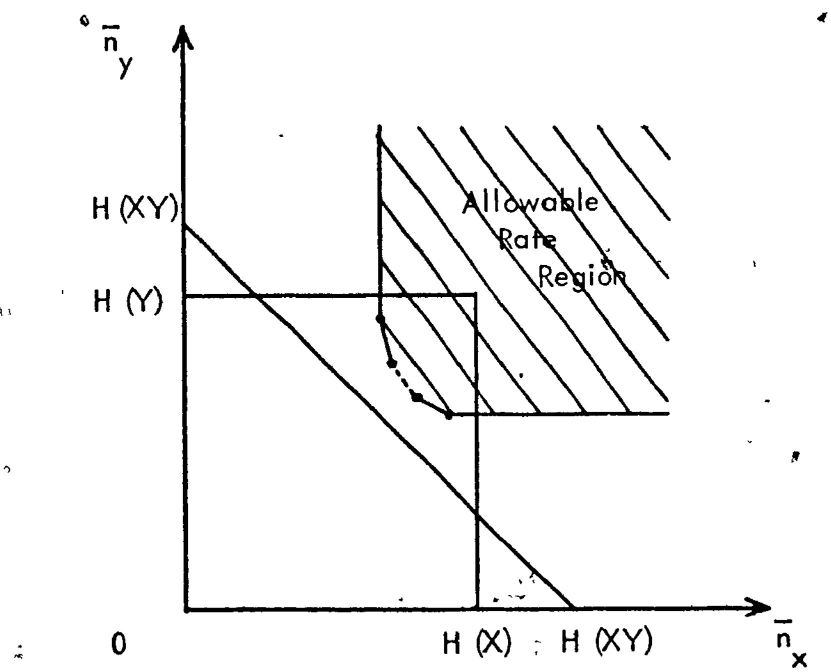


Figure 1-2 : A Typical Allowable Rate Region.

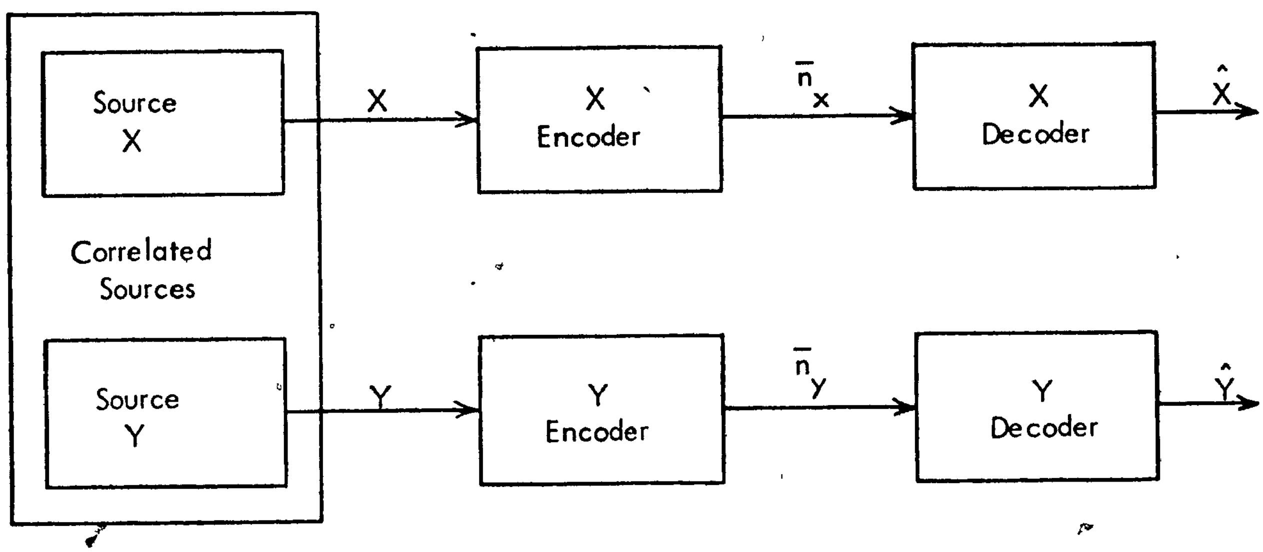


Figure 1-3 : Independent Coding for Correlated Sources .

requires that $\bar{n}_x \geq H(X)$ and $\bar{n}_y \geq H(Y)$. Therefore, there is a special interest in studying the problem of coding for correlated sources, the goal being to discover how to take advantage of this correlation between source output sequences in designing the best possible encoders and decoders. Since the well-known Huffman code is the procedure for constructing optimum codes for a single information source, the main theme of this thesis can be summarized as being the generalization of the Huffman code to the case of two correlated sources.

The correlated source coding problem illustrated in Figure 1-1 is only one of several related systems to be considered in this thesis. As indicated in Figure 1-4, there exist sixteen different arrangements for the encoders and decoders corresponding to all possible ways of positioning the four switches S_1, S_2, S_3, and S_4. Notice that the configuration of Figure 1-1 is just the situation which occurs when switches S_1 and S_2 are open with S_3 and S_4 closed.

The subject of this thesis, as introduced above, is one of several interesting topics concerning the joint coding of correlated sources. Although most of these problems still remain unsolved, two important contributions in this area have recently been reported in the literature. Slepian and Wolf [4] considered the problem of fixed length or block coding for correlated sources. For all of the configurations of Figure 1-4, they determined what minimum numbers of bits per character were needed in order to communicate the source output sequences to the decoder with arbitrarily small decoding error probabilities. Of course, this differs from the problem

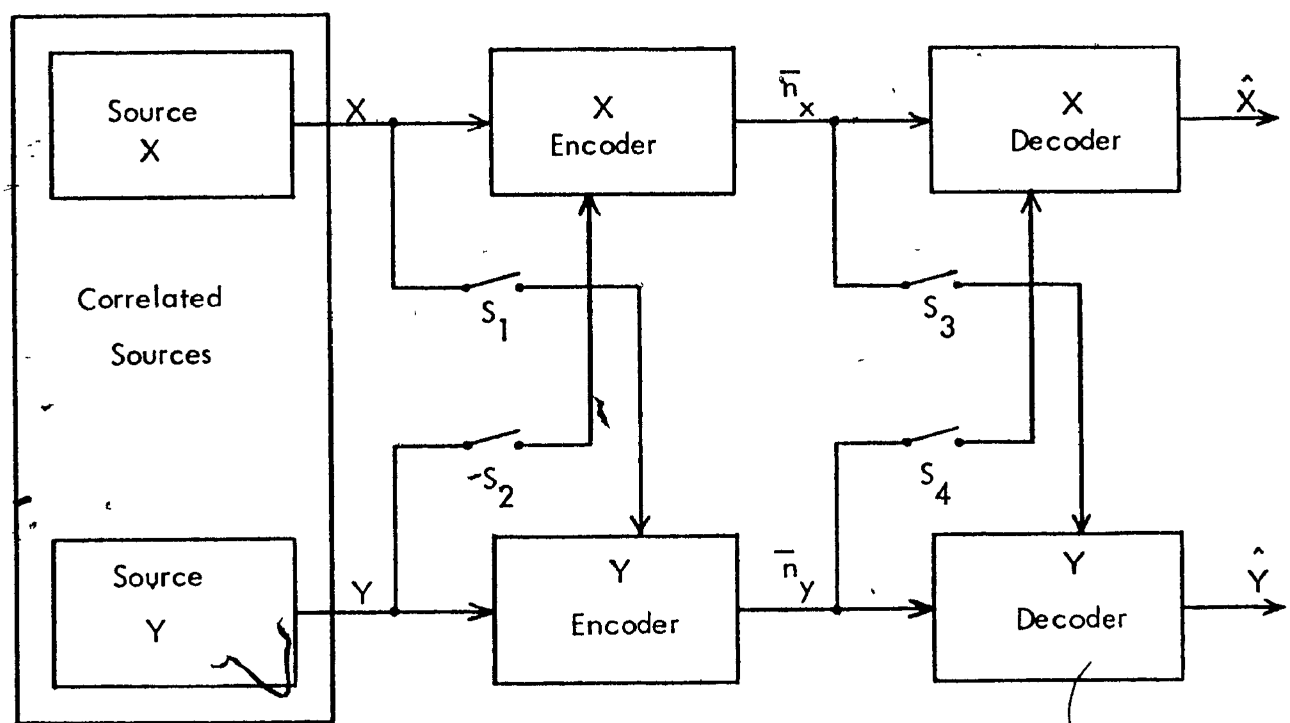


Figure 1-4 : Sixteen Correlated Source Coding Configurations.

of variable length coding which employs a zero probability of error criterion. A paper published by Wyner [5] established a similar result, again concerning fixed length coding for joint sources. The author of this thesis believes that the problem of variable length coding for correlated information sources has until this time been unsolved and that consequently the solutions which are presented in this thesis are contributions to original knowledge.

The material studied in this thesis is organized in the following manner. Chapter II contains a brief review of various fundamental results on variable length coding for a single source. Some useful quantities such as entropy and average code-word length are defined, followed by a statement of three well-known source coding theorems.

Chapter III is devoted to studying the correlated source coding system of Figure 1-1. First of all, the problem is defined precisely and then a theory is developed starting from first principles. Several examples of varying difficulty are presented to aid in illustrating many of the new ideas.

In Chapter IV, the results of Chapter III are exploited in devising a practical method for solving the problem of Figure 1-1 for any given pair of correlated sources. The form of this algorithm allows it to be implemented easily by a computer program. A report is given on how efficiently such a program performed when it was used to solve specific examples.

The purpose of Chapter V is to extend the results of Chapters III and IV, valid only for the system of Figure 1-1, to the other fifteen coding configurations of Figure 1-4. Fortunately, it turns out that only minor modifications to the methods of Chapter IV are necessary. Finally, Chapter VI is a summary of some of the more important results of this thesis, together with a mention of some related topics which might be areas of future research.

CHAPTER II

VARIABLE LENGTH CODING FOR A SINGLE SOURCE

The theory of variable length coding for a single information source is well known (see Gallager [2, pp. 43-55]). This chapter is devoted to reviewing some of the important results of this theory, results which will subsequently be applied when solving the problem of coding for correlated sources.

Therefore, consider the classical source coding problem illustrated in Figure 2-1. Here, source X is assumed to be a discrete memoryless source. This means that each unit of time, the source produces one of a finite set of source letters, say $x_1, x_2, \ldots, x_k$, with a fixed set of probabilities $Pr(x_1), Pr(x_2), \ldots,$ and $Pr(x_k)$. Of course, these probabilities must satisfy

$$\sum_{i=1}^{k} Pr(x_i) = 1 .$$

The information rate of source X is described by a very important quantity called the <u>entropy</u> of source X. It is defined by

$$H(X) \triangleq - \sum_{i=1}^{k} Pr(x_i) \log_2 Pr(x_i) ,$$

where H(X) is the entropy expressed in units called bits of information.

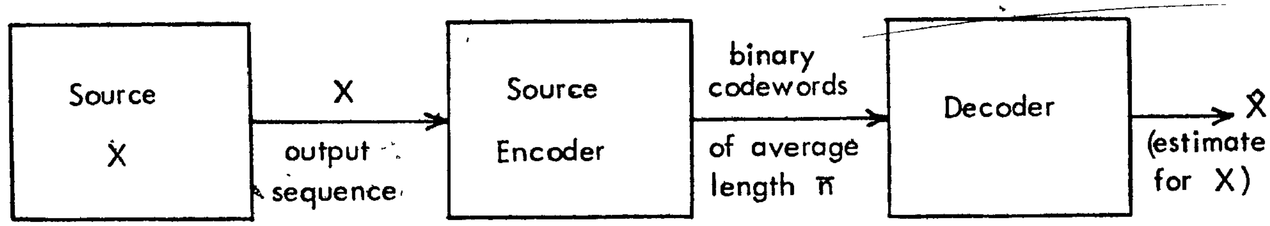


Figure 2-1 : The single source coding problem.

The function of the source encoder is to represent each source letter by a codeword consisting of a sequence of binary letters. More precisely, the encoder performs a one-to-one mapping from the k source letters $x_1, x_2, \ldots, x_k$, to a set of k binary codewords having lengths $m_1, m_2, \ldots,$ and m_k. The average codeword length $\bar{n}$ turns out to be a very useful measure of performance and is defined by

$$\bar{n} \overset{\Delta}{=} \sum_{i=1}^{k} m_i \Pr(x_i) .$$

The decoder for the system of Figure 2-1 performs the following operations. It observes the sequence of binary letters coming from the source encoder and based on this information produces $\hat{X}$, an estimate of the original source output X. It is desirable to design the source encoder in such a way that the decoder can reconstruct the source output sequence with zero probability of error. In order for this requirement to be met, it is necessary and sufficient to choose the set of k codewords to be uniquely decodable. This means that any finite sequence of binary symbols from the source encoder can be uniquely resolved into sequences of codewords.

The objective in studying the system of Figure 2-1, is to determine how to design the best possible source encoder. The optimum encoder is defined to be the one which has the minimum possible average codeword length $\bar{n}$ with the restriction that the code must be uniquely decodable. The following familiar theorem sheds some light on the subject of optimum encoders.

Theorem 2-1 : (for proof, see [2, pp. 50-51])

For the system illustrated in Figure 2-1, it is possible to assign codewords to the source letters such that the code is uniquely decodable and such that the average codeword length $\bar{n}$ satisfies

$$\bar{n} < H(X) + 1 .$$

Furthermore, for any uniquely decodable code of this type, it is necessary that

$$\bar{n} \geq H(X) .$$

Although Theorem 2-1 does not indicate exactly how to design an optimum source encoder, it does establish that the optimum system has an average codeword length somewhere in the range

$$H(X) \leq \bar{n} < H(X) + 1 .$$

A stronger theorem can be established by allowing the source encoder to assign codewords to sequences of L source letters. Specifically, the encoder can be redefined as being a one-to-one mapping from the set of k^L different source sequences of length L to a uniquely decodable set of k^L binary codewords. For this more general situation, the following theorem can be shown to apply.

Theorem 2-2 : (for proof, see [2, p. 51])

For the system shown in Figure 2-1, it is possible to assign codewords to sequences of L source letters such that the code is uniquely decodable and such that the average codeword length $\bar{n}$ satisfies

$$\bar{n} < H(X) + 1/L .$$

Furthermore, for any uniquely decodable code of this generalized type, it is necessary that

$$\bar{n} \geq H(X) .$$

This theorem establishes that in general, the optimum source encoder has an average codeword length $\bar{n}$ somewhere in the range

$$H(X) \leq \bar{n} < H(X) + 1/L .$$

The actual finding of this optimum code can be accomplished by applying a famed constructive procedure called the Huffman code (see Gallager [2, pp. 52-55]) . The key consequence of Theorem 2-2 is that by making L arbitrarily large (that is, by assigning codewords to arbitrarily long source sequences), it is possible to design a source encoder with an average codeword length $\bar{n}$ which is arbitrarily close to H (X). This result is summarized by the following theorem.

Theorem 2-3: The output sequence from source X for the system of Figure 2-1, can be communicated to the decoder with zero probability of error if and only if the average codeword length for the source encoder satisfies

$$\bar{n} \geq H(X) .$$

CHAPTER III

CODING FOR CORRELATED SOURCES

The theory of variable length coding for a single information source (as reviewed in Chapter II) will now be generalized to the correlated source coding problem illustrated in Figure 3-1. This is the same problem which was initially introduced in Chapter I (see Figure 1-1).

In the following discussions, it will be assumed that both source X and source Y are discrete memoryless sources. This implies that during each unit of time, source X produces one of a finite set of source letters, say $x_1, x_2, \ldots, x_k$, and simultaneously source Y produces one letter from the set $y_1, y_2, \ldots, y_q$. Successive occurrences of (X, Y) pairs are independent and are governed by the fixed set of probabilities

$$\{ Pr(x_i, y_j) : i = 1, 2, \ldots, k ; j = 1, 2, \ldots, q \},$$

where of course

$$\sum_{i=1}^{k} \sum_{j=1}^{q} Pr(x_i, y_j) = 1 .$$

The correlation between sources X and Y is best summarized by arranging the given set of probabilities into a q x k probability matrix P as follows :

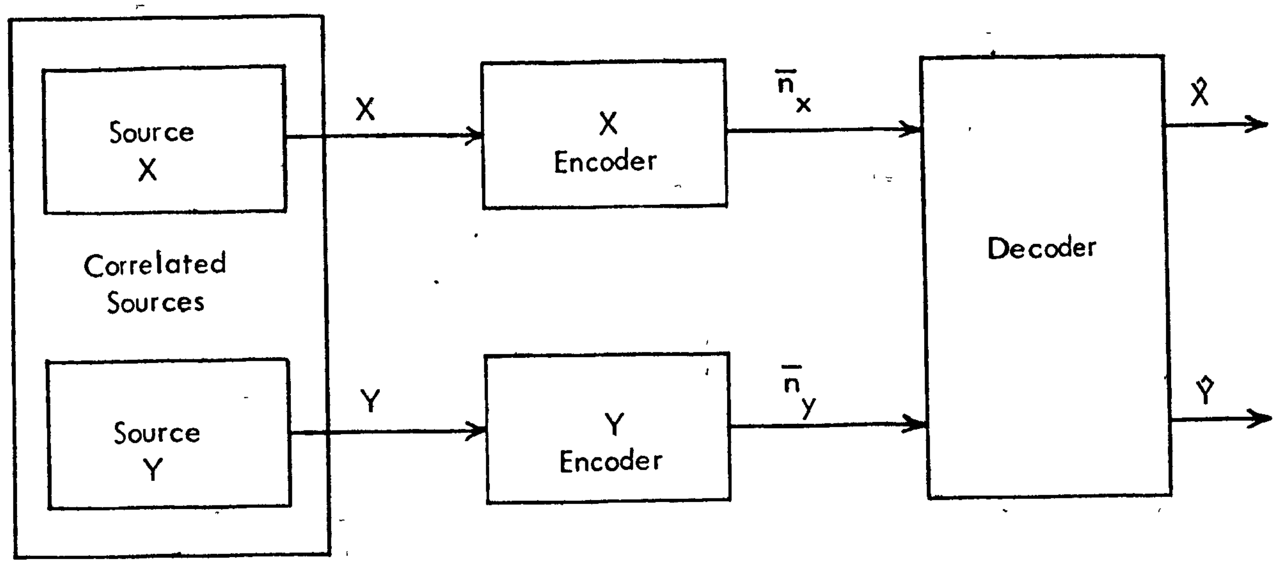


Figure 3-1 : A correlated source coding problem.

$$P \triangleq \begin{bmatrix} \Pr(x_1, y_1) & \Pr(x_2, y_1) & \cdots & \Pr(x_k, y_1) \\ \Pr(x_1, y_2) & \Pr(x_2, y_2) & \cdots & \Pr(x_k, y_2) \\ \vdots & \vdots & & \vdots \\ \Pr(x_1, y_q) & \Pr(x_2, y_q) & \cdots & \Pr(x_k, y_q) \end{bmatrix} .$$

Notice that the marginal probabilities for the X source letters are described by

$$\Pr(x_i) = \sum_{j=1}^{q} \Pr(x_i, y_j) \qquad \text{for } i = 1, 2, \ldots, k .$$

Similarly, the marginal probabilities for source Y are

$$\Pr(y_j) = \sum_{i=1}^{k} \Pr(x_i, y_j) \qquad \text{for } j = 1, 2, \ldots, q .$$

As in Chapter II, it is convenient to characterize sources X and Y by their entropies. The entropy of source X is defined to be

$$H(X) \triangleq - \sum_{i=1}^{k} \Pr(x_i) \log_2 \Pr(x_i) \text{ bits}$$

and similarly source Y has an entropy of

$$H(Y) \triangleq - \sum_{j=1}^{q} \Pr(y_j) \log_2 \Pr(y_j) \text{ bits} .$$

However, this is only a partial characterization because there is a dependence between the two sources. For this reason, it is necessary to introduce H (XY), the joint entropy of sources X and Y. This important quantity is defined by

$$H(XY) \triangleq -\sum_{i=1}^{k} \sum_{j=1}^{q} \Pr(x_i, y_j) \log_2 \Pr(x_i, y_j) \text{ bits .}$$

These entropies will appear often in subsequent derivations regarding the correlated source coding problem.

In order to facilitate the development of a clear and concise theory, it is advantageous to think of the X and Y encoders for the system of Figure 3-1 as being composed of two stages as illustrated in Figure 3-2. That this idea does not result in any loss of generality will become obvious from the following definitions for the precoders and the X' and Y' encoders.

Define the X precoder to be a single valued transformation from the individual source letters $x_1, x_2, \ldots, x_k$ to the new set of letters $x_1', x_2', \ldots, x_M'$, where $M \le k$. Similarly, let the Y precoder be a single valued transformation from the letters $y_1, y_2, \ldots, y_q$, to the new set $y_1', y_2', \ldots, y_N'$, where $N \le q$. In other words, the precoders perform mappings of the form illustrated in Figure 3-3, where the letters $x_{11}, x_{12}, \ldots, x_{Mr_M}$ and $y_{11}, y_{12}, \ldots, y_{Ns_N}$ are just relabelings and reorderings of the original source letters. (This situation will be generalized later to include coding for sequences of source letters.) Notice that $\sum_{i=1}^{M} r_i = k$ and $\sum_{j=1}^{N} s_j = q$. The net effect of the precoders in the

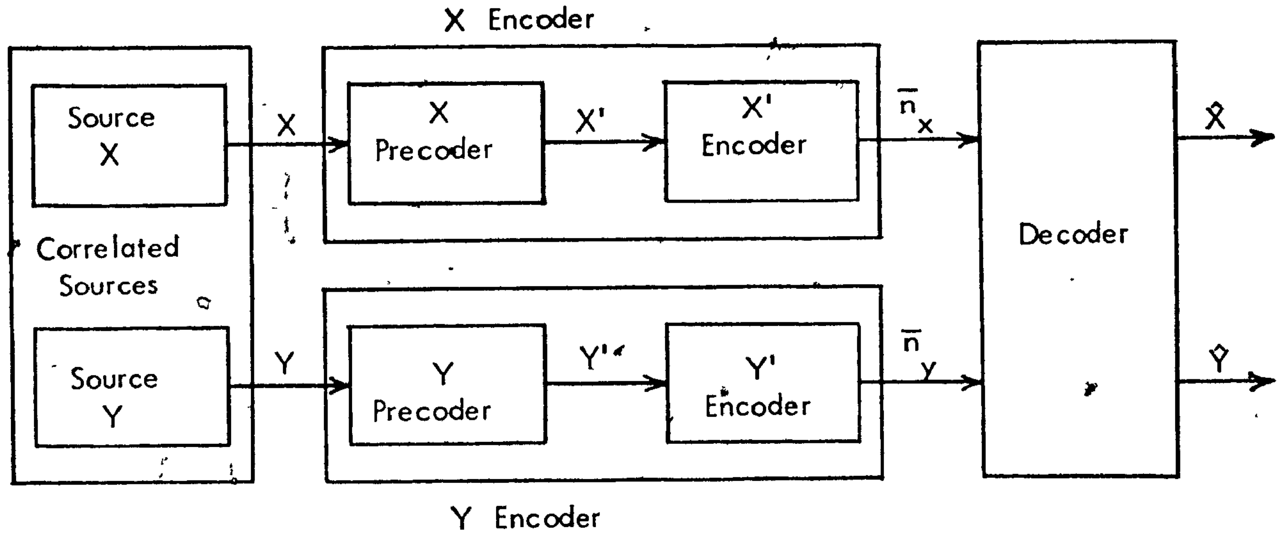


Figure 3-2 : The form of the encoders in the correlated source coding problem.

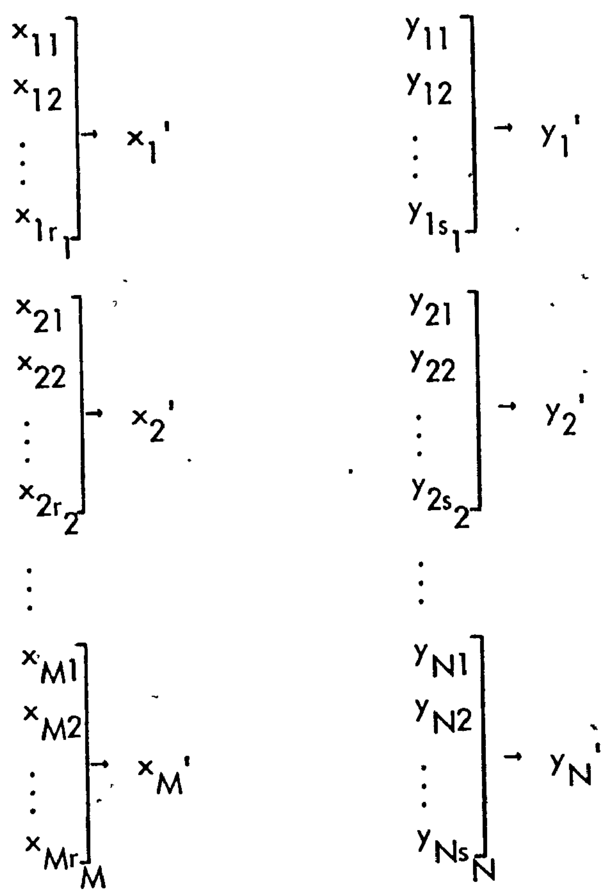


Figure 3-3 : The precoders for the system of Figure 3-2.

system of Figure 3-2 is to transform sources X and Y into simpler sources X' and Y' . These new sources can be described by entropies H (X') and H (Y') respectively where H (X') is defined by

$$H(X') \triangleq -\sum_{i=1}^{M} \Pr(x_i') \log_2 \Pr(x_i')$$
$$= -\sum_{i=1}^{M} \left[\sum_{j=1}^{r_i} \Pr(x_{ij}) \right] \log_2 \left[\sum_{j=1}^{r_i} \Pr(x_{ij}) \right] \text{ bits },$$

and H (Y') is defined in a similar fashion.

Now consider the second stages of the encoders of Figure 3-2, namely the X' and Y' encoders. Having average codeword lengths of $\bar{n}_x$ and $\bar{n}_y$ respectively, these encoders are defined to be uniquely decodable representations for the "transformed" sources X' and Y' . It should be noted that according to Theorem 2-3, the minimum possible values for $\bar{n}_x$ and $\bar{n}_y$ in this situation must be $\bar{n}_x = H(X')$ and $\bar{n}_y = H(Y')$.

The insistence on unique decodability for the X' and Y' encoders ensures that the outcomes X' and Y' can always be communicated to the decoder independently and with zero probability of error. The decoder must make use of this knowledge to produce $\hat{X}$ and $\hat{Y}$, estimates for the source outputs X and Y respectively. The only encoders of interest however, are those for which the decoder can produce $\hat{X} = X$ and $\hat{Y} = Y$ with probability one. This can only happen if the mapping performed by the precoders (see Figure 3-3) is reversible. That is, specifying letters X' and Y' must always uniquely determine the source outputs X and Y. The very special precoder schemes which satisfy this property are called admissible schemes.

By definition, then, distortionless communication is possible for the system of Figure 3-2 if and only if the precoder scheme is admissible. Consequently, it would be very useful to find the set of all admissible precoders. The theorem below gives the necessary and sufficient conditions for a coding strategy to be admissible. First, however, some new definitions are required.

Based on the reordering of X and Y source letters for the precoder shown in Figure 3-3, define a new $q \times k$ matrix $\overline{P}$ by interchanging rows and columns of the probability matrix P as follows :

$$\overline{P} \triangleq \left[\begin{array}{ccc|c|ccc}
Pr(x_{11}, y_{11}) & \cdots & Pr(x_{1r_1}, y_{11}) & \cdots & Pr(x_{M1}, y_{11}) & \cdots & Pr(x_{Mr_M}, y_{11}) \\
\vdots & & \vdots & & \vdots & & \vdots \\
Pr(x_{11}, y_{1s_1}) & \cdots & Pr(x_{1r_1}, y_{1s_1}) & \cdots & Pr(x_{M1}, y_{1s_1}) & \cdots & Pr(x_{Mr_M}, y_{1s_1}) \\
\hline
 & \vdots & & & & \vdots & \\
\hline
Pr(x_{11}, y_{N1}) & \cdots & Pr(x_{1r_1}, y_{N1}) & \cdots & Pr(x_{M1}, y_{N1}) & \cdots & Pr(x_{Mr_M}, y_{N1}) \\
\vdots & & \vdots & & \vdots & & \vdots \\
Pr(x_{11}, y_{Ns_N}) & \cdots & Pr(x_{1r_1}, y_{Ns_N}) & \cdots & Pr(x_{M1}, y_{Ns_N}) & \cdots & Pr(x_{Mr_M}, y_{Ns_N})
\end{array}\right]$$

Notice that the matrix $\overline{P}$ is of the form

$$\overline{P} = \left[\begin{array}{c|c|c|c} P_{11} & P_{21} & \cdots & P_{M1} \\ \hline P_{12} & P_{22} & \cdots & P_{M2} \\ \hline \vdots & \vdots & & \vdots \\ \hline P_{1N} & P_{2N} & \cdots & P_{MN} \end{array}\right]$$

where P_{ij} is the $s_j \times r_i$ matrix defined for $i = 1, 2, \ldots, M$ and $j = 1, 2, \ldots, N$ to be

$$P_{ij} \stackrel{\Delta}{=} \begin{bmatrix} \Pr(x_{i1}, y_{j1}) & \cdots & \Pr(x_{ir_i}, y_{j1}) \\ \vdots & & \vdots \\ \Pr(x_{i1}, y_{js_j}) & \cdots & \Pr(x_{ir_i}, y_{js_j}) \end{bmatrix} .$$

<u>Theorem 3-1</u> : A precoder scheme is admissible for the system of Figure 3-2 if and only if the corresponding matrix $\overline{P}$ as defined above has at most one non-zero element in each of its MN submatrices P_{ij}, for $i = 1, 2, \ldots,$ M and $j = 1, 2, \ldots, N$.

<u>Proof :</u> Investigate the precoder scheme illustrated in Figure 3-3 and its corresponding probability matrix $\overline{P}$. Consider any one of the submatrices P_{ij} which make up $\overline{P}$. The $r_i s_j$ entries of P_{ij} are associated with the following $r_i s_j$ (X, Y) pairs : (x_{i1}, y_{j1}) , (x_{i2}, y_{j1}) , ..., (x_{ir_i}, y_{j1}), (x_{i1}, y_{j2}) , (x_{i2}, y_{j2}), ..., (x_{ir_i}, y_{j2}), ..., (x_{i1}, y_{js_j}), (x_{i2}, y_{js_j}) , ..., and (x_{ir_i}, y_{js_j}) . But the precoder of Figure 3-3 maps all of these (X, Y) pairs onto the same (X', Y') pair, namely

(x_i', y_i'). It is therefore not possible for the letters (x_i', y_i') to uniquely determine which of the above $r_i s_j$ (X, Y) pairs occurred unless no more than one these pairs can occur with non-zero probability. This conclusion is true for each of the submatrices P_{ij}. Hence, the precoder scheme is admissible if and only if every submatrix P_{ij} has no more than one non-zero element.

Corollary (see Lu [3]): The only admissible precoder scheme for a system of the form of Figure 3-2 if its probability matrix P has no zero elements, is the trivial mapping $X' = X$ and $Y' = Y$. (This is equivalent to encoding and decoding the two sources independently which, according to Theorem 2-3, requires that $\bar{n}_x \geq H(X)$ and $\bar{n}_y \geq H(Y)$. This result is in marked contrast to the ϵ - error results of Slepian and Wolf [4].)

Example 3-1: Find all admissible precoder schemes for the system of Figure 3-2 if the correlated sources are described by the probability matrix

$$P = \begin{bmatrix} 1/3 & 0 & 1/6 \\ 0 & 1/3 & 1/6 \end{bmatrix} \begin{matrix} y_1 \\ y_2 \end{matrix}$$

$$\begin{matrix} x_1 & \quad x_2 & \quad x_3 \end{matrix}$$

There are two possible Y-precoder strategies to consider and they are

$$\text{(i)}\quad \begin{array}{l} y_1] \to y_1' \\ y_2] \to y_2' \end{array} \qquad \text{and} \qquad \text{(ii)}\quad \left.\begin{array}{l} y_1 \\ y_2 \end{array}\right] \to y_1'$$

Similarly, there are five candidates for X-precoder schemes and these are :

$$\text{(i)}\quad \begin{array}{l} \left.\begin{array}{l} x_1 \\ x_2 \end{array}\right] \to x_1' \\ x_3] \to x_2' \end{array} \qquad \text{(ii)}\quad \begin{array}{l} \left.\begin{array}{l} x_1 \\ x_3 \end{array}\right] \to x_1' \\ x_2] \to x_2' \end{array} \qquad \text{(iii)}\quad \begin{array}{l} \left.\begin{array}{l} x_2 \\ x_3 \end{array}\right] \to x_1' \\ x_1] \to x_2' \end{array}$$

$$\text{(iv)}\quad \left.\begin{array}{l} x_1 \\ x_2 \\ x_3 \end{array}\right] \to x_1' \qquad \text{and} \qquad \text{(v)}\quad \begin{array}{l} x_1] \to x_1' \\ x_2] \to x_2' \\ x_3] \to x_3' \end{array} .$$

Thus, there are ten precoder schemes of the form of Figure 3-3 corresponding to the ten possible ways of choosing one of the two Y-precoders and one of the five X-precoders.

To test these ten schemes for admissibility, it is only necessary to form the matrix $\overline{P}$ for each case and apply the test derived in Theorem 3-1 to each sub-matrix of $\overline{P}$. For example, consider the precoder

$$\begin{array}{l} \left.\begin{array}{l} x_1 \\ x_2 \end{array}\right] \to x_1' \\ x_3] \to x_2' \end{array} \qquad \begin{array}{l} y_1] \to y_1' \\ y_2] \to y_2' \end{array}$$

For this case,

$$\bar{P} \doteq \left[\begin{array}{cc|c} 1/3 & 0 & 1/6 \\ \hline 0 & 1/3 & 1/6 \end{array}\right] = \left[\begin{array}{c|c} P_{11} & P_{21} \\ \hline P_{12} & P_{22} \end{array}\right] .$$

Since each submatrix P_{ij} has only one non-zero element, this precoder is admissible. Similarly, by testing the other nine precoders, it is found that none of them are admissible except for the trivial case which is defined by $X' = X$ and $Y' = Y$.

Returning to the consideration of the general system depicted in Figure 3-2, it is useful to summarize the results obtained up to this point. It has been established that the outputs of the correlated sources X and Y can be communicated to the decoder with zero probability of error if and only if the precoder scheme is admissible. For any particular problem, it is possible to determine the entire set of admissible precoders, simply by applying the testing procedure of Theorem 3-1. It is known that for any precoder described by entropies $H(X')$ and $H(Y')$, the average codeword lengths for the X' and Y' encoders must satisfy $\bar{n}_x \geq H(X')$ and $\bar{n}_y \geq H(Y')$. Therefore, by calculating these lower bounds $H(X')$ and $H(Y')$ for each member of the set of all admissible precoders, it is possible to plot an allowed two-dimensional rate region. Specifically, the set of points $(H(X'), H(Y'))$ can be used to construct an admissible region R in the $\bar{n}_x - \bar{n}_y$ plane. Region R can be defined formally by stating that any point

(R_x, R_y) must lie inside this area if and only if there exist encoders with $\bar{n}_x = R_x$ and $\bar{n}_y = R_y$ which allow the decoder to reconstruct the source outputs with zero distortion.

The following two theorems are necessary for determining the admissible region R in any general problem.

<u>Theorem 3-2 Bit Stuffing</u> : If the point $(R_x, R_y) \in R$, then the point $(R_x + \delta_x, R_y + \delta_y) \in R$ for any $\delta_x, \delta_y \geq 0$.

<u>Proof</u> : By definition, since the point $(R_x, R_y) \in R$, there must exist an encoder having $\bar{n}_x = R_x$ and $\bar{n}_y = R_y$ which allows the decoder to reconstruct the source outputs with zero probability of error. Modify this encoder as follows : after every L_1 codewords sent out by the X encoder, send K_1 arbitrary binary characters ; similarly for the Y encoder, send K_2 arbitrary binary symbols after every L_2 codewords. For this new coding scheme, the average codeword lengths are $\bar{n}_x = R_x + K_1 / L_1$ and $\bar{n}_y = R_y + K_2 / L_2$. The decoder can still reconstruct the source outputs with zero distortion because it knows the numbers K_i, L_i (for i = 1, 2) and hence can count out sequences of L_i codewords and discard the following K_i meaningless binary symbols. Therefore, the point $(R_x + K_1 / L_1, R_y + K_2 / L_2) \in R$ where K_1, K_2, L_1, and L_2 are any positive integers. Any positive real number can be expressed as accurately as desired as the ratio of two positive integers by taking those integers to be sufficiently large. Con-

sequently, in the limit of large integers, K_1 / L_1 and K_2 / L_2 can be replaced by positive real numbers δ_x and δ_y. Thus, the point $(R_x + \delta_x, R_y + \delta_y) \in R$.

<u>Theorem 3-3 Time Sharing</u>: If $(R_{x1}, R_{y1}) \in R$ and $(R_{x2}, R_{y2}) \in R$, then $(\lambda R_{x1} + (1-\lambda) R_{x2}, \lambda R_{y1} + (1-\lambda) R_{y2}) \in R$ for any λ in the range $0 \le \lambda \le 1$.

<u>Proof</u>: Since (R_{x1}, R_{y1}) and (R_{x2}, R_{y2}) belong to R, there must exist the following two encoders which allow the source outputs to be communicated to the decoder with zero probability of error: Encoder I having $\bar{n}_x = R_{x1}$ and $\bar{n}_y = R_{y1}$, and Encoder II having $\bar{n}_x = R_{x2}$ and $\bar{n}_y = R_{y2}$. Consider the construction of an encoder which uses the mapping scheme of Encoder I u times and follows by using the strategy of Encoder II v times. That is, $u / (u + v)$ of the time, the encoder has $\bar{n}_x = R_{x1}$ and $\bar{n}_y = R_{y1}$ and the rest of the time, it has $\bar{n}_x = R_{x2}$ and $\bar{n}_y = R_{y2}$. This new encoder has average codeword lengths of $\bar{n}_x = (uR_{x1} + vR_{x2}) / (u + v)$ and $\bar{n}_y = (uR_{y1} + vR_{y2}) / (u + v)$. The decoder can still reconstruct X and Y with zero distortion because it knows the values for u and v and can thus keep track at all times of which of the two coding strategies is being used. Therefore, the rate point $((uR_{x1} + vR_{x2}) / (u + v), (uR_{y1} + vR_{y2}) / (u + v))$ must belong to the admissible region. By letting $\lambda = u / (u + v)$, an equivalent statement is that the point $(\lambda R_{x1} + (1-\lambda) R_{x2}, \lambda R_{y1} + (1-\lambda) R_{y2}) \in R$. The desired result follows by noting that the value for λ can be made to vary continuously from 0 to 1 by choosing the integers u and v to be sufficiently large and in the correct ratio.

The above discussions have indicated that the admissible rate region for any problem of the form illustrated in Figure 3-2 can be determined by carrying out the three steps summarized below :

(i) Determine the set of all admissible precoder schemes with the aid of Theorem 3-1,

(ii) For each member of this set, determine the lower bounds $H(X')$ and $H(Y')$ for $\bar{n}_x$ and $\bar{n}_y$ respectively. Plot all these points $(H(X'), H(Y'))$ on the $\bar{n}_x - \bar{n}_y$ plane,

(iii) Apply Theorems 3-2 and 3-3 to the series of points plotted in (ii) in order to discover the entire admissible region R .

This basic method is best illustrated by applying it to the solution of several examples.

Example 3-1 : (continued)

For the probability matrix $P = \begin{bmatrix} 1/3 & 0 & 1/6 \\ 0 & 1/3 & 1/6 \end{bmatrix}$, it was found that the only admissible precoders were

$$\text{(i)} \quad \left.\begin{matrix} x_1 \\ x_2 \end{matrix}\right] \to x_1' \qquad y_1] \to y_1' \qquad \text{and} \qquad \text{(ii)} \quad x_1] \to x_1' \qquad y_1] \to y_1'$$

$$x_3] \to x_2' \qquad y_2] \to y_2' \qquad\qquad x_2] \to x_2' \qquad y_2] \to y_2' .$$

$$x_3] \to x_3'$$

For scheme (i) it is necessary that $\bar{n}_x \geq H(X')$ and $\bar{n}_y \geq H(Y')$ where

$$\begin{aligned} H(X') &= -\Pr(x_1') \log_2 \Pr(x_1') - \Pr(x_2') \log_2 \Pr(x_2') \\ &= -(2/3) \log_2 (2/3) - (1/3) \log_2 (1/3) \\ &= 0.918 \end{aligned}$$

and $\quad H(Y') = H(Y) = -\frac{1}{2} \log_2 \frac{1}{2} - \frac{1}{2} \log_2 \frac{1}{2} = 1 .$

For scheme (ii), it is necessary that $\bar{n}_x \geq H(X) = 1.585$ and $\bar{n}_y \geq H(Y) = 1$.

These two admissible rate points (0.918, 1) and (1.585, 1) are plotted in Figure 3-4. By applying Theorem 3-2, the admissible region R is found to include all points of the form $(0.918 + \delta_x , 1 + \delta_y)$ for any $\delta_x , \delta_y \geq 0$. The resulting region shown in Figure 3-4 is actually the entire admissible region. It has such a simple shape that Theorem 3-3 does not yield any new information about R .

Notice that if sources X and Y were coded independently, the admissible region would be the double hatched region in Figure 3-4, the subset of region R described by $\bar{n}_x \geq H(X)$ and $\bar{n}_y \geq H(Y)$.

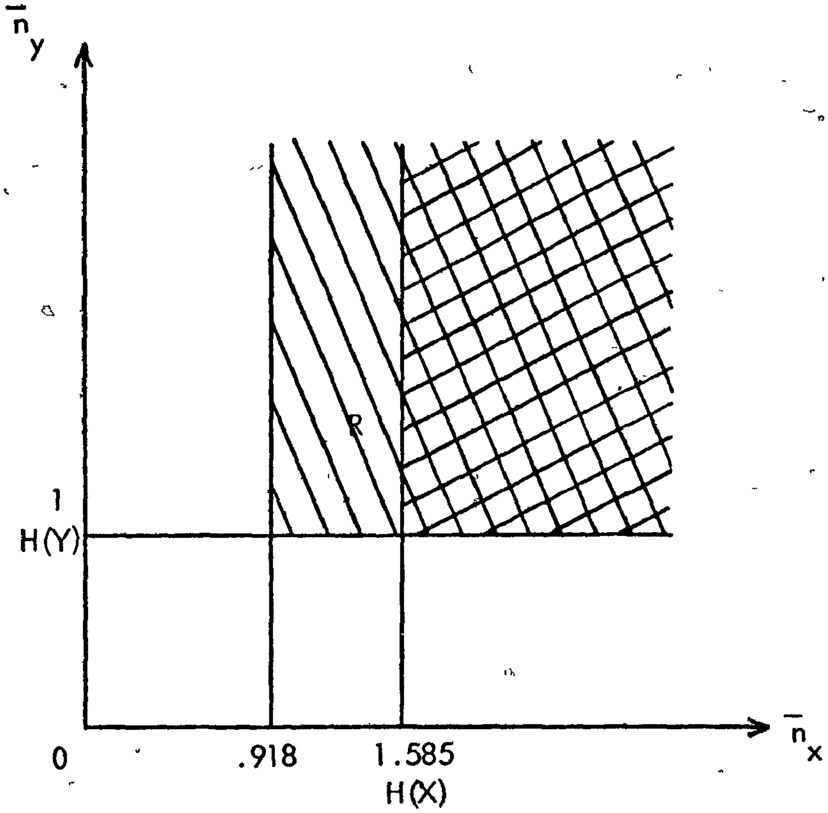


Figure 3-4 : The admissible region for Example 3-1 .

Example 3-2 : Suppose the system of Figure 3-2 is described by the following probability matrix :

$$P = \begin{array}{c} \\ \left[\begin{array}{cccc} \frac{1}{4} & 0 & 0 & 0 \\ 0 & \frac{1}{4} & \frac{1}{4} & 0 \\ 0 & 0 & 0 & \frac{1}{4} \end{array}\right] \begin{array}{c} y_1 \\ y_2 \\ y_3 \end{array} \\ \begin{array}{cccc} x_1 & x_2 & x_3 & x_4 \end{array} \end{array} .$$

Instead of searching randomly through a large number of possibilities to find the set of all admissible precoders, it is more efficient to first discover all admissible coding schemes of the following two special types :

(i) Type A ; precoders whose Y precoder is the one-to-one mapping $Y' = Y$, and

(ii) Type B ; precoders whose X precoder is the one-to-one mapping $X' = X$.

By applying the results of Theorem 3-1 to this example, it is easily found that there are ten admissible precoders of Type A as follows :

(1) $\left.\begin{array}{c} x_1 \\ x_2 \\ x_4 \end{array}\right] \rightarrow x_1' \qquad \left.x_3\right] \rightarrow x_2'$

(2) $\left.\begin{array}{c} x_1 \\ x_3 \\ x_4 \end{array}\right] \rightarrow x_1' \qquad \left.x_2\right] \rightarrow x_2'$

(3) $\left.\begin{array}{c} x_1 \\ x_2 \end{array}\right] \rightarrow x_1' \qquad \left.\begin{array}{c} x_3 \\ x_4 \end{array}\right] \rightarrow x_2'$

(4)

$$\left.\begin{matrix} x_1 \\ x_3 \end{matrix}\right] \rightarrow x_1' \qquad \left.\begin{matrix} x_2 \\ x_4 \end{matrix}\right] \rightarrow x_2'$$

(5)

$$\left.\begin{matrix} x_1 \\ x_4 \end{matrix}\right] \rightarrow x_1' \qquad x_2] \rightarrow x_2' \qquad x_3] \rightarrow x_3'$$

(6)

$$\left.\begin{matrix} x_1 \\ x_2 \end{matrix}\right] \rightarrow x_1' \qquad x_3] \rightarrow x_2' \qquad x_4] \rightarrow x_3'$$

(7)

$$x_1] \rightarrow x_1' \qquad x_2] \rightarrow x_2' \qquad \left.\begin{matrix} x_3 \\ x_4 \end{matrix}\right] \rightarrow x_3'$$

(8)

$$\left.\begin{matrix} x_1 \\ x_3 \end{matrix}\right] \rightarrow x_1' \qquad x_2] \rightarrow x_2' \qquad x_4] \rightarrow x_3'$$

(9)

$$x_1] \rightarrow x_1' \qquad \left.\begin{matrix} x_2 \\ x_4 \end{matrix}\right] \rightarrow x_2' \qquad x_3] \rightarrow x_3'$$

(10)

$$x_1] \rightarrow x_1' \qquad x_2] \rightarrow x_2' \qquad x_3] \rightarrow x_3' \qquad x_4] \rightarrow x_4' .$$

Of course, the Y precoder scheme in each of these cases is understood to be the trivial one $Y' = Y$.

Similarly, it is easily found that there are five admissible precoders of Type B as follows :

(1) $$\left.\begin{array}{l} y_1 \\ y_2 \\ y_3 \end{array}\right] \rightarrow y_1'$$

(2) $$\left.\begin{array}{l} y_1 \\ y_2 \end{array}\right] \rightarrow y_1' \qquad y_3 \,] \rightarrow y_2'$$

(3) $$\left.\begin{array}{l} y_1 \\ y_3 \end{array}\right] \rightarrow y_1' \qquad y_2 \,] \rightarrow y_2'$$

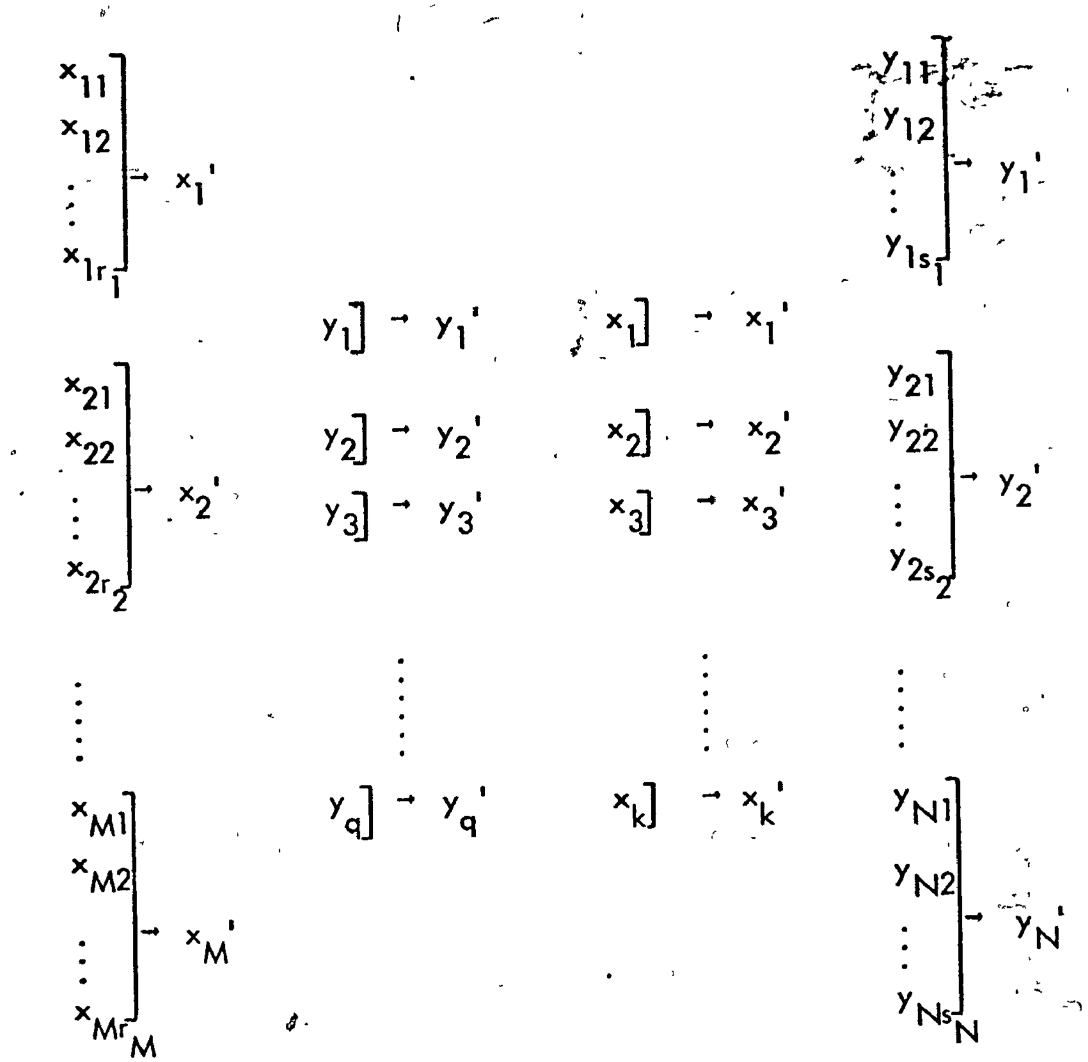
(4) $$y_1 \,] \rightarrow y_1' \qquad \left.\begin{array}{l} y_2 \\ y_3 \end{array}\right] \rightarrow y_2'$$

(5) $$y_1 \,] \rightarrow y_1' \qquad y_2 \,] \rightarrow y_2' \qquad y_3 \,] \rightarrow y_3' \,.$$

It is understood that the X precoder in each of these cases is defined by $X' = X$.

As illustrated by this example, the set of all admissible Type A precoders is actually a list of X precoder schemes. Similarly, finding all admissible Type B precoders gives a list of Y precoder strategies. The significance of these two lists is that any precoder of the general form of Figure 3-3 can only be admissible if its X precoder belongs to the Type A list and its Y precoder belongs to the Type B list. In other words, if the precoder of Figure 3-3 is admissible, the two precoders shown in Figure 3-5 must also be admissible. This fact becomes obvious according to Theorem 3-1 by inspecting the $\overline{P}$ matrices corresponding to the three precoders in question.

A method, then, to determine the set of all admissible precoders in any problem, is to consider all possible ways of choosing an X precoder from the Type A list and a Y precoder from the Type B list and to apply the test of Theorem 3-1 to each of these possibilities.



(a) A Type A precoder.

(b) A Type B precoder.

Figure 3-5 : Type A and Type B precoders.

In the present example, there are $10 \times 5 = 50$ ways of choosing one of the ten X precoders and one of the five Y precoders. Actually, only $9 \times 4 = 36$ of these need to be tested for admissibility because the fourteen schemes having either $X' = X$ or $Y' = Y$ are already known to be admissible.

By performing these 36 admissibility tests, it is easily concluded that there are a total of 26 admissible codes for this example. These precoders are arrayed in Table 3-1 together with their corresponding entropies. All of the resulting rate points of the form $(H(X'), (H(Y'))$ are plotted in Figure 3-6. According to Theorem 3-3 on time sharing, points on the line segments ab and bc must belong to R. By applying Theorem 3-2, the complete admissible region R is determined to be that drawn in Figure 3-6.

<u>Example 3-3</u>: Consider a system characterized by the probability matrix

$$P = \begin{bmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & d_k \end{bmatrix},$$

where $\sum_{i=1}^{k} d_i = 1$. First note that according to Theorem 3-1, the following two precoders are admissible:

Precoder No.	X Precoder No.	Y Precoder No.	H (X')	H (Y')
1	10	1	2	0
2	10	2	2	.811
3	10	3	2	1
4	10	4	2	.811
5	10	5	2	1.5
6	9	2	1.5	.811
7	9	3	1.5	1
8	9	5	1.5	1.5
9	8	3	1.5	1
10	8	4	1.5	.811
11	8	5	1.5	1.5
12	7	2	1.5	.811
13	7	3	1.5	1
14	7	5	1.5	1.5
15	6	3	1.5	1
16	6	4	1.5	.811
17	6	5	1.5	1.5
18	5	2	1.5	.811
19	5	4	1.5	.811
20	5	5	1.5	1.5
21	4	3	1	1
22	4	5	1	1.5
23	3	3	1	1
24	3	5	1	1.5
25	2	5	.811	1.5
26	1	5	.811	1.5

Table 3-1 : The admissible precoders for Example 3-2.

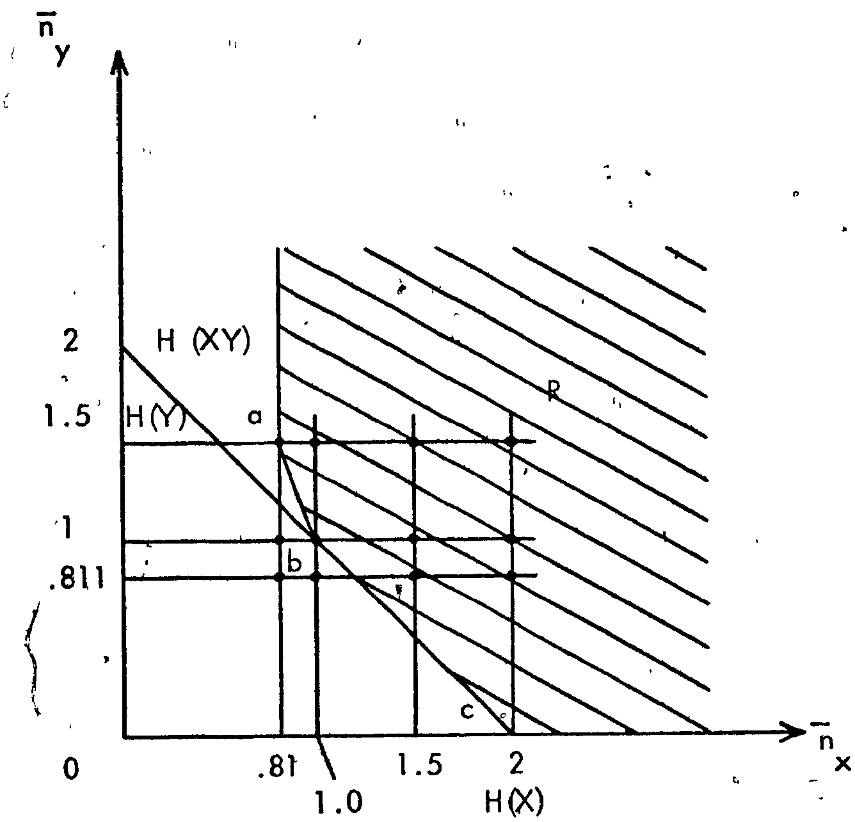


Figure 3-6 : The admissible region for Example 3-2 .

$$(1)\quad \left.\begin{array}{c} x_1 \\ x_2 \\ \vdots \\ x_k \end{array}\right] \to x_1' \qquad \begin{array}{c} y_1] \to y_1' \\ y_2] \to y_2' \\ \vdots \quad\quad \vdots \\ y_k] \to y_k' \end{array} \qquad \text{and} \qquad (2)\quad \begin{array}{c} x_1] \to x_1' \\ x_2] \to x_2' \\ \vdots \quad\quad \vdots \\ x_k] \to x_k' \end{array} \qquad \left.\begin{array}{c} y_1 \\ y_2 \\ \vdots \\ y_k \end{array}\right] \to y_1' \,.$$

Therefore the points $(0, H(Y))$ and $(H(X), 0)$ belong to the admissible region R. Notice that since P is diagonal, $H(X) = H(Y) = H(XY)$. The two points $(0, H(XY))$ and $(H(XY), 0)$ are plotted in Figure 3-7. The following theorem allows the admissible region for this example to be determined by inspection.

<u>Theorem 3-4</u>: For a system of the form of Figure 3-1 (or Figure 3-2) the source outputs can be communicated to the decoder with zero probability of error only if $\bar{n}_x + \bar{n}_y \geq H(XY)$.

<u>Proof</u>: Suppose a system exists which allows the source information to be communicated to the decoder with zero distortion and which has encoders such that $\bar{n}_x + \bar{n}_y < H(XY)$. Consider a combining of the X and Y encoders of this system to form an XY encoder for the joint source XY. For example, choose an XY encoder which alternates on every bit between the codeword sequences of the X and Y encoders. This encoder allows the decoder to reconstruct the output of source XY with zero distortion and furthermore, it has an average codeword length of $\bar{n}_{xy} = \bar{n}_x + \bar{n}_y$. Therefore, $\bar{n}_{xy} < H(XY)$. However, according to Theorem 2-3 for the single source XY, it is necessary that $\bar{n}_{xy} \geq H(XY)$ in order to have distortionless communication. This contradiction implies that the opening assump-

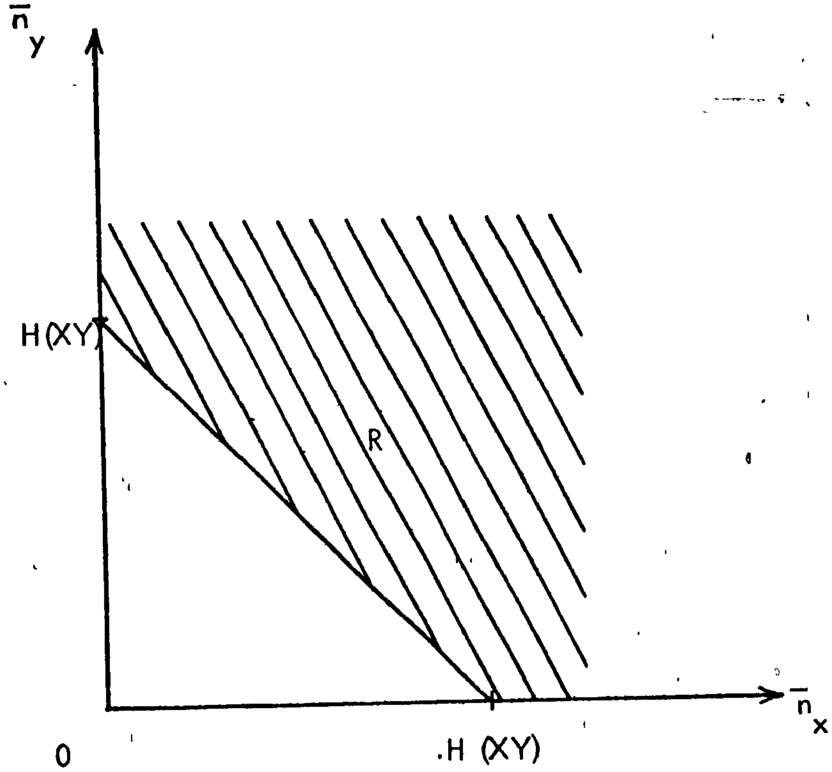


Figure 3-7: The admissible region for Example 3-3.

tion was wrong. Therefore, no system can exist with $\bar{n}_x + \bar{n}_y < H(XY)$ such that the source outputs can be reconstructed by a decoder with zero probability of error. That is, it is always necessary that $\bar{n}_x + \bar{n}_y \geq H(XY)$.

Example 3-3 (concluded) : In Figure 3-7, the line segment joining the points $(0, H(XY))$ and $(H(XY), 0)$ is the line $\bar{n}_x + \bar{n}_y = H(XY)$. According to the Time Sharing Theorem, points on this segment are admissible. Theorem 3-4 proves that no points below this line can belong to region R. Therefore, the entire admissible region R is that illustrated in Figure 3-7.

This chapter has developed some simple procedures for finding the admissible region R for problems of the form indicated in Figure 3-2. Until now, this has been a restricted class of problems because the precoders have been limited to performing transformations on the individual source letters. Fortunately, it is very easy to generalize this situation to allow coding for sequences of source letters. Indeed, by pretending that the X source letters $x_1, x_2, .., x_k$ and the Y source letters $y_1, y_2, \ldots, y_q$ are actually sequences of L letters from two simpler sources, the most general problem can be solved using exactly the same methods employed previously in this chapter. This fact is illustrated by the following concluding example.

Example 3-4 : Find the admissible region R for a system described by the probability matrix

$$P = \begin{bmatrix} 1/3 & 0 & 1/6 \\ 0 & 1/3 & 1/6 \end{bmatrix} \begin{matrix} y_1 \\ y_2 \end{matrix} ,$$
$$\begin{matrix} x_1 & x_2 & x_3 \end{matrix}$$

if coding is permitted for sequences of $L = 2$ source letters.

By assuming that source X actually has 9 outcomes $x_1 x_1$, $x_1 x_2$, $x_1 x_3$, ..., $x_3 x_3$ and source Y has four letters $y_1 y_1$, $y_1 y_2$, $y_2 y_1$, and $y_2 y_2$, the following equivalent problem can be set up : find the admissible region for a system with probability matrix

$$P = \begin{bmatrix} 1/9 & 0 & 1/18 & 0 & 0 & 0 & 1/18 & 0 & 1/36 \\ 0 & 1/9 & 1/18 & 0 & 0 & 0 & 0 & 1/18 & 1/36 \\ 0 & 0 & 0 & 1/9 & 0 & 1/18 & 1/18 & 0 & 1/36 \\ 0 & 0 & 0 & 0 & 1/9 & 1/18 & 0 & 1/18 & 1/36 \end{bmatrix} \begin{matrix} y_1 y_1 \\ y_1 y_2 \\ y_2 y_1 \\ y_2 y_2 \end{matrix}$$
$$\begin{matrix} x_1x_1 & x_1x_2 & x_1x_3 & x_2x_1 & x_2x_2 & x_2x_3 & x_3x_1 & x_3x_2 & x_3x_3 \end{matrix}$$

if coding is only permitted for individual source letters.

It has been established in this chapter how to solve such a problem. It turns out in this case that no admissible Type B precoders exist besides the trivial one but that the best (lowest entropy) admissible Type A precoder is the code

$$\left.\begin{matrix} x_1 x_1 \\ x_1 x_2 \\ x_2 x_1 \\ x_2 x_2 \end{matrix}\right] \rightarrow x_1' \qquad\qquad y_1 y_1] \rightarrow y_1'$$

$$\left.\begin{matrix} x_1 x_3 \\ x_2 x_3 \end{matrix}\right] \rightarrow x_2' \qquad\qquad y_1 y_2] \rightarrow y_2'$$

$$y_2 y_1] \rightarrow y_3'$$

$$\left.\begin{matrix} x_3 x_1 \\ x_3 x_2 \end{matrix}\right] \rightarrow x_3' \qquad\qquad y_2 y_2] \rightarrow y_4'$$

$$x_3 x_3] \rightarrow x_4'$$

which has $H(X') = 2(0.918)$ and $H(Y') = 2(1)$. Therefore, the point $\bar{n}_x = H(X')/2 = 0.918$ and $\bar{n}_y = H(Y')/2 = 1$ is an admissible rate point, the extra factor of two arising because the above precoder is for sequences of length two. Consequently, the admissible region R is the same as that plotted in Figure 3-4 in connection with Example 3-1 .

CHAPTER IV

AN ALGORITHM FOR DETERMINING THE ADMISSIBLE REGION

All the basic ideas and theorems necessary in understanding the system of correlated sources illustrated in Figure 3-2 have been established in Chapter III. In practice, however, the solution of problems described by large, sparse probability matrices can require an enormous amount of work. For example, for a system with a 10 x 10 probability matrix, there are many billions of different precoder mapping combinations of the form of Figure 3-3. To search randomly through this gigantic number of possibilities to find the admissible coding schemes is obviously impractical, if not impossible. The purpose of this chapter, then, is to develop an algorithm which allows such large problems to be solved efficiently with the aid of a digital computer. It should be kept in mind that even though the methods below assume that only coding for individual source letters is permitted, they can be applied equally well to problems allowing coding for sequences.

During the discussions of Chapter III and specifically in connection with Example 3-2, the following four step method was suggested for solving any correlated source coding problem of the form of Figure 3-2 :

<u>Step A</u> : Determine the set of all admissible precoders whose Y precoder is defined by the one-to-one mapping Y' = Y . Each admissible code of this restricted type, which we shall call Type A, has a different X precoder. The result of Step A, then, is a list of various X precoder mapping schemes.

Step B : Determine the set of all admissible precoders whose X precoder is defined by the one-to-one mapping X' = X . Each admissible code of this restricted type, called Type B, is characterized by a different Y precoder. Therefore, this step results in a list of various Y precoder mapping strategies.

Step C : Consider the set of all possible ways of choosing an X precoder from the list discovered in Step A and a Y precoder from the list found in Step B. Test each such precoder for admissibility. As proven in Example 3-2, this procedure determines the set of all admissible precoders.

Step D : For each admissible precoder revealed by Step C, calculate and plot the points (H (X'), H (Y')) on the $\bar{n}_x - \bar{n}_y$ plane. Find the entire admissible region R by applying Theorems 3-2 and 3-3 .

The above four steps form the basis for an algorithm which will be developed in the remainder of this chapter. First of all, a detailed strategy will be given for efficiently performing Step A . No new methods will be necessary for Step B because it only differs from Step A in that the roles of X and Y are reversed. An improved final step will then be worked out by combining Steps C and D . This procedure will take advantage of the fact that in most problems, it is not necessary to determine the complete set of admissible precoders (as in Step C above) in order to draw the admissible region R. For instance the admissible region drawn in Figure 3-6 in connection with Example 3-2 is defined by only three precoders, those corresponding to the points a, b, and c. None of the other admissible precoders appear in the final solution.

<u>Algorithm for Step A :</u> Consider a problem of the form depicted in Figure 3-2 which is characterized by a given probability matrix P. The goal of Step A is to use this given information to discover the set of all admissible precoders of the special type (Type A) shown in Figure 3-5a. Notice that the Y precoder is limited to be a one-to-one mapping whereas the X precoder is unrestricted. It is convenient to divide the set of all admissible Type A precoders into different classifications based on their X precoders. Specifically, catagorize the precoders according to how many groupings of two or more X source letters occur in the X precoder mapping scheme. For example, for the precoder of Figure 3-5a, if $r_1, r_2, \ldots, r_J$ are all greater than one and $r_{J+1}, r_{J+2}, \ldots, r_M$ are all equal to one, then this X precoder has J groupings of two or more source letters.

This classification for Type A precoders indicates that Step A can be carried out in several sequential steps as follows :

<u>Step A (1) :</u> Determine the set of all admissible Type A precoders whose X mapping schemes have at most one grouping of two or more source letters.

Step A (2) : Determine the set of all admissible Type A precoders whose X mapping schemes have two groupings of two or more source letters.

.
.
.

Step A (J) : Determine the set of all admissible Type A precoders whose X mapping schemes have J groupings of two or more source letters.

By choosing the number J to be large enough, the sets of admissible precoders found in Steps A (1) through A (J) together make up the set of all admissible Type A precoders. The advantage of performing Step A according to this sequence of J steps is that it turns out that once Step A (1) has been completed, Step A (2) can be performed directly from the results of Step A (1) without even looking at the probability matrix. Similarly, it will be shown below how any Step A (I) can be carried out directly from the results of Step A (I−1) by a simple procedure.

The key, then, to efficiently performing Step A according to the set of steps indicated above is to develop an algorithm to carry out Step A (1) . This involves discovering the set of all admissible precoders of the special form illustrated in Figure 4-1 . These codes are characterized by X mapping schemes having either one grouping of two or more source letters (when $r_1 > 1$) or having no such groupings (when $r_1 = 1$) . For any admissible precoder of this form we will henceforth refer to its only grouping of source letters $(x_{11}, x_{12}, \ldots, x_{1r_1})$ as being an admissible X-

$$\left.\begin{matrix} x_{11} \\ x_{12} \\ \vdots \\ x_{1r_1} \end{matrix}\right] \rightarrow x_1' \qquad y_1] \rightarrow y_1'$$

$$y_2] \rightarrow y_2'$$

$$x_{21}] \rightarrow x_2' \qquad y_3] \rightarrow y_3'$$

$$x_{31}] \rightarrow x_3' \qquad \vdots$$

$$\vdots \qquad y_q] \rightarrow y_q'$$

$$x_{M1}] \rightarrow x_M'$$

Figure 4-1 : A Type A precoder having only one X-grouping.

grouping. Using this terminology, the goal of Step A (1) can be restated as follows : find the set of all admissible X-groupings.

To achieve this objective, consider the forming of a matrix P_0 by performing column interchanges on the probability matrix P according to the following rules :

(i) Find the row of P which has the most non-zero entries : the J_0^{th} row. If two or more rows have the same maximum number of non-zero entries, choose any one of these rows. Define I_0 to be the number of zeroes in the J_0^{th} row.

(ii) Interchange columns of P so that the J_0^{th} row has all of its I_0 zeroes in the leftmost columns.

The resulting matrix P_0 has an appearance of the following form :

$$P_0 = \begin{matrix} y_1 \\ \vdots \\ y_{J_0} \\ \vdots \\ y_q \end{matrix} \left[\begin{array}{cccc|cccc} & & & & & & & \\ 0 & 0 & \cdots & 0 & x & x & \cdots & x \\ & & & & & & & \end{array} \right] \triangleq [P_{01} \mid P_{02}],$$
$$\qquad\qquad x_1^{01} \quad x_2^{01} \quad \cdots \quad x_{I_0}^{01} \quad x_1^{02} \quad x_2^{02} \quad \cdots \quad x_{k-I_0}^{02}$$

where "x" denotes a non-zero element. Notice that P_0 has been partitioned into two matrices P_{01} and P_{02} and the X source letters have been relabeled accordingly.

Since each column of P_{02} has a non-zero element in the J_0^{th} row, Theorem 3-1 implies that any precoder of the form of Figure 4-1 cannot be admissible if the X-grouping $(x_{11}, x_{12}, \ldots, x_{1r_1})$ contains two or more of the source letters $x_1^{02}, x_2^{02}, \ldots,$ and $x_{k-l_0}^{02}$. In other words, any admissible X-grouping for the system described by matrix P (or equivalently P_0), must contain only one or else none of the letters $x_1^{02}, x_2^{02}, \ldots, x_{k-l_0}^{02}$. Consequently, Step A (1) can be solved by combining the solutions of the following two simpler problems:

(I) Find the set S_1 of all admissible X-groupings which contain none of the source letters $x_1^{02}, x_2^{02}, \ldots,$ and $x_{k-l_0}^{02}$ and

(II) Find the set S_{12} of all admissible X-groupings which contain exactly one of the source letters $x_1^{02}, x_2^{02}, \ldots,$ and $x_{k-l_0}^{02}$.

It will now be shown how these two steps can be used as a basis for a recursive algorithm for carrying out Step A (1).

First of all, consider the first step, problem I. It involves the study of X-groupings which contain only source letters associated with matrix P_{01}, namely $x_1^{01}, x_2^{01}, \ldots, x_{l_0}^{01}$. But the set of all admissible X-groupings of this type is just the set of all admissible X-groupings for a system of the form of Figure 3-2 which is described by the probability matrix P_{01} instead of P_0. Therefore, the problem of finding all admissible X-groupings for matrix P_0 (P) is dependent on the solution of the same problem for a smaller matrix P_{01} .

The procedure of interchanging columns of P to get matrix P_0 in a special partitioned form, can be repeated for matrix P_{01} . Specifically, it is possible to form a matrix P_1 by interchanging columns of P_{01} according to the given rules to get $P_1 = [P_{11} \vdots P_{12}]$. As above, this step makes the problem of finding all admissible X-groupings for matrix P_{01} (or equivalently P_1) depend on the simpler problem of finding all admissible X-groupings for the smaller matrix P_{11} . By repeating this operation for matrix P_{11} and later for P_{21}, P_{31}, . . . etc., a sequence of problems of decreasing complexity is generated. Eventually a final stage must be reached, with $P_i = [P_{i1} \vdots P_{i2}]$, where the matrix P_{i1} will have a row containing no zero elements. It is obvious in this case that the only admissible X-groupings formed from the letters of P_{i1} will be trivial groupings consisting of one source letter.

Step A (1) can be completed by working backwards step by step. The set S_{i+1} of all admissible X-groupings as found for matrix P_{i1} can be used to determine S_i, the set of all admissible X-groupings for P_i or $P_{(i-1)1}$. Similarly, it is possible to progress all the way back, ending up with S_0 , the set of all admissible X-groupings for the original matrix P. To illustrate this procedure, suppose that the set S_1 of admissible groupings for P_{01} has already been determined and we want to use this information to find S_0 . Since it has been proven above that $S_0 = S_1 \cup S_{12}$, it remains only to specify S_{12} in order to be able to calculate S_0 .

The simplest kind of admissible X-groupings belonging to the set S_{12} are the single letter groupings (x_1^{02}), (x_2^{02}), . . . , and $(x_{k-l_0}^{02})$. All other groupings of S_{12} must consist of one of these letters, x_i^{02}, joined together with one or more letters from the set $x_1^{01}, x_2^{01}, \ldots, x_{l_0}^{01}$, thus resulting in groupings of the form $(x_i^{01}, \ldots, x_m^{01}, x_j^{02})$, where $1 \le i, m \le l_0$ and $1 \le j \le k-l_0$. But if such a grouping is admissible, Theorem 3-1 implies that its subset $(x_i^{01}, \ldots, x_m^{01})$ must also be an admissible X-grouping and hence must belong to the set S_1 (because it contains none of the x^{02} letters) . Consequently, all admissible X-groupings of the form $(x_i^{01}, \ldots, x_m^{01}, x_j^{02})$ are composed of an X-grouping from the known set S_1 annexed to one of the source letters $x_1^{02}, x_2^{02}, \ldots, x_{k-l_0}^{02}$. It is easy to determine all such groupings simply by trying all possible ways of annexing one of the x^{02} letters to each admissible grouping of the set S_1 . In summary, then, a practical method for determining S_{12} involves the following :

(i) Combine the single letter groupings $(x_1^{02}), \ldots, (x_{k-l_0}^{02})$ with

(ii) the set of all admissible groupings formed by annexing one of the letters $x_1^{02}, \ldots, x_{k-l_0}^{02}$ to the X-groupings of the set S_1 .

With S_1 and S_{12} known, it is a simple matter to calculate S_0 , the union of these sets.

A complete method has now been developed for performing Step A (1). The algorithm is best illustrated by using it to solve a simple example.

<u>Example 4-1:</u> Find the set of all admissible X-groupings for a system described by the matrix

$$P = \begin{bmatrix} \frac{1}{4} & 0 & 0 & 0 \\ 0 & \frac{1}{4} & \frac{1}{4} & 0 \\ 0 & 0 & 0 & \frac{1}{4} \end{bmatrix} \begin{matrix} y_1 \\ y_2 \\ y_3 \end{matrix}$$

$$\begin{matrix} \quad x_1 & x_2 & x_3 & x_4 \end{matrix}$$

The row with the most non-zero elements is the second row. By interchanging columns of P, the matrix P_0 is formed as follows:

$$P_0 = \left[\begin{array}{cc:cc} \frac{1}{4} & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{4} & \frac{1}{4} \\ 0 & \frac{1}{4} & 0 & 0 \end{array}\right] = [\, P_{01} \,\vdots\, P_{02} \,] .$$

$$\begin{matrix} \quad x_1 & x_4 & x_2 & x_3 \end{matrix}$$

By repeating this procedure for P_{01}, the matrix P_1 is found to be

$$P_1 = \left[\begin{array}{c:c} \frac{1}{4} & 0 \\ 0 & 0 \\ 0 & \frac{1}{4} \end{array}\right] = [\, P_{11} \,\vdots\, P_{12} \,] .$$

$$\begin{matrix} \quad x_1 & x_4 \end{matrix}$$

Since P_{11} is a column matrix, its set of admissible X-groupings S_2 consists only of the single letter grouping (x_1). The set S_1 of admissible X-groupings for matrix P_1 is formed by combining

(i) the set $S_2 = \{(x_1)\}$ with

(ii) the single letter grouping (x_4) and

(iii) the set of all admissible X-groupings formed by annexing the letter x_4 to each grouping in the set S_2. The only grouping falling into this category is $(x_1\, x_4)$. Therefore, $S_1 = \{(x_1), (x_4), (x_1\, x_4)\}$.

The set S_0 of admissible X-groupings for matrix P_0 is formed by combining

(i) the set $S_1 = \{(x_1), (x_4), (x_1\, x_4)\}$ with

(ii) the single letter groupings (x_2), (x_3), and

(iii) the set of all admissible X-groupings formed by annexing one of the letters x_2, x_3, to groupings from the set S_1. The admissible X-groupings of this latter type are $(x_1\, x_2)$, $(x_1\, x_3)$, $(x_4\, x_2)$, $(x_4\, x_3)$, $(x_1\, x_4\, x_2)$ and $(x_1\, x_4\, x_3)$.

Therefore, the set of all admissible X-groupings for this problem is $S_0 = \{(x_1), (x_4), (x_1\, x_4), (x_2), (x_3), (x_1\, x_2), (x_1\, x_3), (x_4\, x_2), (x_4\, x_3), (x_1\, x_4\, x_2), (x_1\, x_4\, x_3)\}$.

Now that an algorithm has been developed for Step A (1), let us turn our attention to Step A (2), the problem of finding all admissible precoders of the type illustrated in Figure 4-2 (where r_1, $r_2 > 1$). Notice that if the precoder of Figure 4-2 is admissible, then according to Theorem 3-1, the X-groupings $(x_{11}, x_{12}, \ldots, x_{1r_1})$ and $(x_{21}, x_{22}, \ldots, x_{2r_2})$ must be admissible by themselves and thus must belong to the set S_0. Consequently, it is possible to carry out Step A (2) directly from the results of Step A (1) as follows : take the set S_0 of admissible X-groupings found in Step A (1) and try all possible ways of combining two of these groupings to form precoders of the form shown in Figure 4-2.

Example 4-1 : (continued) It has been found above that the set of all admissible X-groupings is $S_0 = \{(x_1), (x_4), (x_1\, x_4), (x_2), (x_3), (x_1\, x_2), (x_1\, x_3), (x_4\, x_2), (x_4\, x_3), (x_1\, x_4\, x_2), (x_1\, x_4\, x_3)\}$. To find all precoders having the form shown in Figure 4-2, we need only investigate all possible ways of choosing two from the following X-groupings : $(x_1\, x_4)$, $(x_1\, x_2)$, $(x_1\, x_3)$, $(x_4\, x_2)$, $(x_4\, x_3)$, $(x_1\, x_4\, x_2)$, $(x_1\, x_4\, x_3)$.

Forming a precoder from the two groupings $(x_1\, x_4)$ and $(x_1\, x_2)$ is clearly unacceptable because the letter x_1 appears twice so that an X-mapping scheme is not well defined. Similarly, choosing the grouping $(x_1\, x_4)$ along with any other does not form an acceptable precoder. By trying all the other possibilities, it is easily found that the following are the only two admissible precoders of the required type :

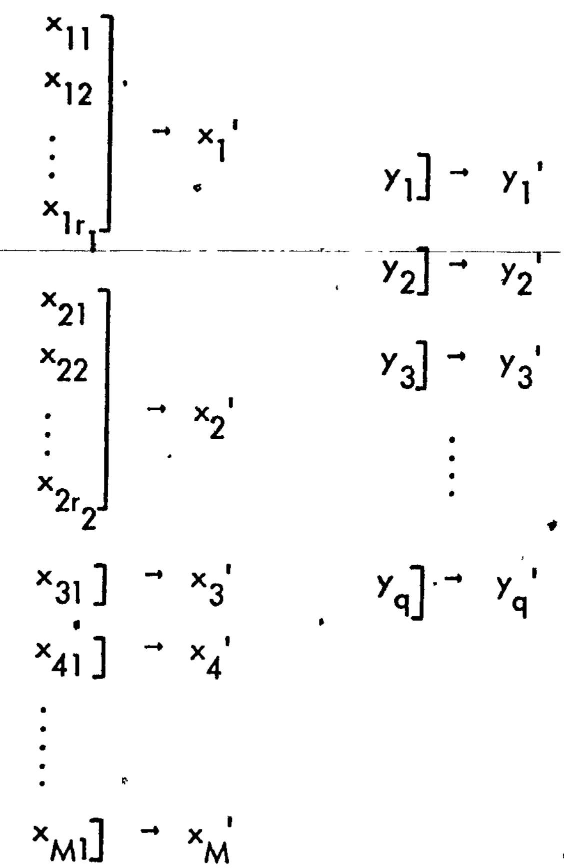


Figure 4-2 : A Type A precoder having two X-groupings.

$$\left.\begin{matrix} x_1 \\ x_2 \end{matrix}\right] \rightarrow x_1' \qquad \left.\begin{matrix} x_1 \\ x_3 \end{matrix}\right] \rightarrow x_1'$$

and

$$\left.\begin{matrix} x_3 \\ x_4 \end{matrix}\right] \rightarrow x_2' \qquad \left.\begin{matrix} x_2 \\ x_4 \end{matrix}\right] \rightarrow x_2' .$$

Consider now a method for performing Step A (J) given the results of Steps A (1), A (2), ..., and A (J-1). The goal is to determine all admissible precoders of the form shown in Figure 4-3, where $r_1, r_2, \ldots, r_J$ are all restricted to be larger than one. According to Theorem 3-1, if the precoder of Figure 4-3 is admissible, all of the X-groupings $(x_{11}, x_{12}, \ldots, x_{1r_1})$, $(x_{21}, x_{22}, \ldots, x_{2r_2}), \ldots,$ and $(x_{J1}, x_{J2}, \ldots, x_{Jr_J})$ must be admissible by themselves and therefore must belong to the list found in Step A (1). Furthermore, any precoder formed by choosing any (J-1) of the J X-groupings of Figure 4-3 must be admissible and thus must belong to the list of admissible precoders found in Step A (J-1).

Hence, a method for finding the set of all admissible precoders of the form in Figure 4-3 is to investigate all possible ways of annexing one of the admissible X-groupings found in Step A (1) to one of the admissible precoders found in Step A (J-1).

<u>Example 4-2</u>: Find the set of all admissible Type A precoders given that the outcome of Step A (1) (the set of all admissible X-groupings) is the set

$S_0 = \{(x_1 x_2), (x_3 x_4), (x_5 x_6), (x_1 x_2 x_3), (x_1 x_3), (x_2 x_3), (x_1), (x_2), (x_3), (x_4), (x_5), (x_6)\}$.

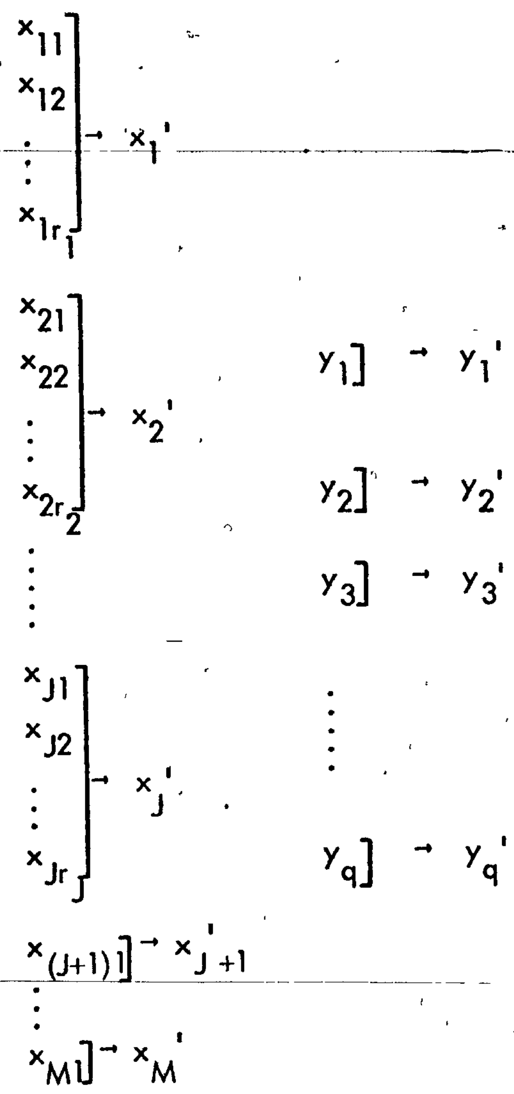


Figure 4-3: A Type A precoder having J X- groupings.

Step A (2) : By trying each possible way of choosing two of the X-groupings from the set S_0, (excluding the trivial groupings (x_1), (x_2), ..., (x_6)), the following precoders are found to be admissible :

$$\left.\begin{matrix}x_1\\x_2\end{matrix}\right] \to x_1' \qquad \left.\begin{matrix}x_1\\x_2\end{matrix}\right] \to x_1' \qquad \left.\begin{matrix}x_3\\x_4\end{matrix}\right] \to x_1'$$

$$\left.\begin{matrix}x_3\\x_4\end{matrix}\right] \to x_2' \qquad \left.\begin{matrix}x_5\\x_6\end{matrix}\right] \to x_2' \qquad \left.\begin{matrix}x_5\\x_6\end{matrix}\right] \to x_2'$$

$$x_5] \to x_3' \qquad x_3] \to x_3' \qquad x_1] \to x_3'$$

$$x_6] \to x_4' \qquad x_4] \to x_4' \qquad x_2] \to x_4'$$

$$\left.\begin{matrix}x_1\\x_2\\x_3\end{matrix}\right] \to x_1' \qquad \left.\begin{matrix}x_1\\x_3\end{matrix}\right] \to x_1' \qquad \left.\begin{matrix}x_2\\x_3\end{matrix}\right] \to x_1'$$

$$\left.\begin{matrix}x_5\\x_6\end{matrix}\right] \to x_2' \qquad \left.\begin{matrix}x_5\\x_6\end{matrix}\right] \to x_2' \qquad \left.\begin{matrix}x_5\\x_6\end{matrix}\right] \to x_2'$$

$$x_4] \to x_3' \qquad x_2] \to x_3' \qquad x_1] \to x_3'$$

$$\qquad x_4] \to x_4' \qquad x_4] \to x_4' .$$

Step A (3) : By trying each possible way of annexing one of the X-groupings of the set S_0 to one of the six admissible precoders found in Step A (2), the following precoder is found to be the only admissible one :

$$\left.\begin{matrix}x_1\\x_2\end{matrix}\right] \to x_1'$$

$$\left.\begin{matrix}x_3\\x_4\end{matrix}\right] \to x_2'$$

$$\left.\begin{matrix}x_5\\x_6\end{matrix}\right] \to x_3' .$$

There are no precoders containing more than three X-groupings of more than one source letter. Therefore, the set of all admissible Type A precoders is formed by combining the admissible precoders found in Steps A (1), A (2), and A (3).

The algorithm developed above for performing Steps A (1) through A (J), allows the determination of all admissible Type A precoders as required by Step A. The methods are very well suited for computer programming due to the simple step by step progression. Some more examples will be presented towards the end of this chapter to illustrate how efficiently the above method can be implemented by a computer program.

Algorithm for Step B: The object of Step B is to determine the set of all admissible Type B precoders. This is exactly the same problem as solved in Step A except that the roles of X and Y are reversed. Therefore, the identical method developed above can also be used for Step B simply by replacing the probability matrix P by its transpose P^T.

Algorithm for Steps C and D: The goal of this final step is to determine the admissible region R given the set of all admissible Type A and Type B precoders as found in Steps A and B respectively. The list of admissible Type A precoders is actually a list of K_A different X precoder mapping schemes and similarly, the set of admissible Type B precoders is a list of K_B Y precoder mapping schemes. The set of all admissible precoders is thus a subset of the $K_A K_B$ different possible ways of choosing an X precoder from the Type A list and a Y precoder from the Type B list. As shown in Figure 4-4, each of these $K_A K_B$ possibilities defines a point (H (X'), H (Y')) in the $\bar{n}_x - \bar{n}_y$ plane. Therefore, the problem of determining the admissible region R involves searching through a grid of $K_A K_B$ points to find which ones (out of those corresponding to admissible precoders) are the vertices defining region R.

Many of the grid points in Figure 4-4 can be eliminated from further consideration by inspection. For example, as proven by Theorem 3-4, none of the grid points below the line $\bar{n}_x + \bar{n}_y = H\ (XY)$ (which is the line FG in Figure 4-4) can represent admissible precoders. Furthermore, since points C and D are known to represent admissible precoders (of Type A and Type B respectively), all of the grid points above the line segment CD must lie inside the boundary of the admissible region R, as proven by Theorems 3-2 and 3-3. Consequently, it can be stated that no points other than those lying on or between lines CD and FG can be vertices for region R.

A method, then, for discovering the admissible region R, is to systematically choose points from the area between lines CD and FG, and to test the corresponding precoders for admissibility. Whenever a code is tested and found to be admissible, the region R can be increased accordingly. For example, if the code corresponding to point E of Figure 4-4 is found to be admissible, the triangular area CED can be added to R. Moreover, only the grid points lying above line FG and below triangle CED still need to be considered as possible vertices for region R.

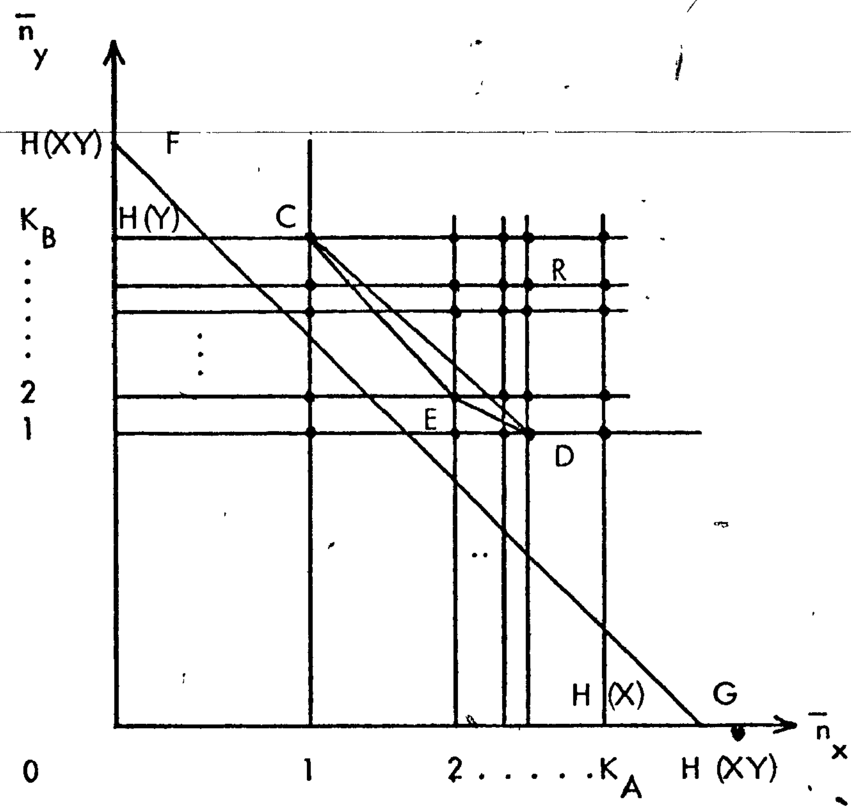


Figure 4-4 : Determination of the admissible region.

The method described above can be summarized by the following steps :

(i) Let R_0 be the admissible region defined by points C and D.

(ii) Set $i = 0$.

(iii) Continue (or start if $i = 0$) a systematic search through the set of all $K_A K_B$ grid points. Check each point to see if it lies in the area below region R_i and above line FG. If it does, test the corresponding precoder for admissibility. Continue this search only until an admissible point is discovered.

(iv) Form admissible region R_{i+1} by taking region R_i and annexing the area defined by the admissible point found in (iii) .

(v) Set $i = i + 1$.

(vi) Go to (iii) .

At some stage, when step (iii) results in no new admissible points, the problem is completed and the admissible region R is just region R_i .

Two important points that should be discussed concerning step (iii) of the above method are how to efficiently carry out the grid search and the admissibility tests. Rather than searching through the set of all grid points in a random order, it should be more efficient to search in some organized manner. One promising idea is to search through the grid points in the order of increasing $\bar{n}_x + \bar{n}_y$. This causes points which are on the average farthest below the region R_i to be tested first. This method seems to be very efficient because when an admissible point is discovered, it defines a comparatively large area to be annexed to region R_i. This not only causes region R_{i+1} to be a much better approximation to the entire admissible region R but it also greatly reduces the number of grid points (lying in the area between R_{i+1} and line FG) which remain to be tested.

Experience has shown that careful ordering of a large number of grid points according to $\bar{n}_x + \bar{n}_y$ is usually not practical but luckily a rough ordering of this nature is already available from the results of Steps A and B. During the method followed for these two steps, the precoders were arranged according to the number of groupings of two or more source letters. But, on the average, precoders having the larger number of such groupings tend to have the smaller entropies. Therefore, a practical method of performing the grid search can begin by searching through the precoders whose X and Y mapping schemes have the maximum numbers of groupings of two or more source letters. The search then continues by considering precoders which have diminishing numbers of source letter groupings.

An important operation which must be performed during the grid search is the testing of various precoder schemes for admissibility. According to Theorem 3-1, one admissibility testing procedure is to form the matrix $\bar{P}$ corresponding to the precoder under consideration, and to count the number of non-zero elements in certain submatrices of $\overline{P}$. Although this method seems simple enough, it is very inefficient. Counting elements in a matrix is very time consuming, especially for problems which have large probability matrices. Furthermore, in large problems, a great many admissibility tests have to be performed. An alternative method for performing admissibility tests is illustrated by the following example.

<u>Example 4-3</u>: Given a system described by a 5×5 probability matrix, it is desired to test the following precoder for admissibility :

$$\left.\begin{matrix} x_1 \\ x_2 \end{matrix}\right] \to x_1' \qquad \left.\begin{matrix} y_1 \\ y_2 \\ y_3 \end{matrix}\right] \to y_1'$$

$$\left.\begin{matrix} x_3 \\ x_4 \end{matrix}\right] \to x_2' \qquad \left.\begin{matrix} y_4 \\ y_5 \end{matrix}\right] \to y_2'$$

$$\left. x_5 \right] \to x_3'$$

It is easy to see from Theorem 3-1 that this code will be admissible if and only if the following four simpler precoders are all admissible :

(i)

$$\left.\begin{matrix} x_1 \\ x_2 \end{matrix}\right] \to x_1' \qquad \left.\begin{matrix} y_1 \\ y_2 \\ y_3 \end{matrix}\right] \to y_1'$$

$$\left. x_3 \right] \to x_2'$$

$$\left. x_4 \right] \to x_3' \qquad \left. y_4 \right] \to y_2'$$

$$\left. x_5 \right] \to x_4' \qquad \left. y_5 \right] \to y_3'$$

(ii)

$$\left.\begin{matrix} x_1 \\ x_2 \end{matrix}\right] \to x_1' \qquad \left. y_1 \right] \to y_1'$$

$$\left. y_2 \right] \to y_2'$$

$$\left. x_3 \right] \to x_2' \qquad \left. y_3 \right] \to y_3'$$

$$\left. x_4 \right] \to x_3'$$

$$\left. x_5 \right] \to x_4' \qquad \left.\begin{matrix} y_4 \\ y_5 \end{matrix}\right] \to y_4'$$

(iii)
$$\left.\begin{matrix}x_1\end{matrix}\right] \to x_1' \quad \left.\begin{matrix}x_2\end{matrix}\right] \to x_2' \quad \left.\begin{matrix}x_3\\x_4\end{matrix}\right] \to x_3' \quad \left.\begin{matrix}x_5\end{matrix}\right] \to x_4' \qquad \left.\begin{matrix}y_1\\y_2\\y_3\end{matrix}\right] \to y_1' \quad \left.\begin{matrix}y_4\end{matrix}\right] \to y_2' \quad \left.\begin{matrix}y_5\end{matrix}\right] \to y_3'$$

(iv)
$$\left.\begin{matrix}x_1\end{matrix}\right] \to x_1' \quad \left.\begin{matrix}x_2\end{matrix}\right] \to x_2' \quad \left.\begin{matrix}x_3\\x_4\end{matrix}\right] \to x_3' \quad \left.\begin{matrix}x_5\end{matrix}\right] \to x_4' \qquad \left.\begin{matrix}y_1\end{matrix}\right] \to y_1' \quad \left.\begin{matrix}y_2\end{matrix}\right] \to y_2' \quad \left.\begin{matrix}y_3\end{matrix}\right] \to y_3' \quad \left.\begin{matrix}y_4\\y_5\end{matrix}\right] \to y_4' .$$

As shown in the above example, any precoder can be tested for admissibility by testing several precoders of a simpler type, those having only one X-grouping and one Y-grouping of two or more source letters. Therefore, an efficient test method can begin by forming a table of all X-groupings versus all Y-groupings and entering into the elements of this table "admissible" or "not admissible" according to whether or not the corresponding precoders defined by one X-grouping and one Y-grouping are admissible. Consequently, any general precoder strategy can be tested for admissibility by looking up the correct entries in this table. Using this method, the code of Example 4-3 could be tested for admissibility by performing only four table look-ups. This is obviously very efficient compared to the alternative of searching through 25 elements of the probability matrix P to count up non-zero elements. It will be illustrated by the examples below that especially for large, sparse matrices, the cost of setting up the table as described above is small compared to the large savings which result during the performance of admissibility tests.

A Fortran computer program (380 cards long) has been written to apply the methods developed in this chapter to the problem of finding the admissible region R for correlated sources described by any given probability matrix. This program has been used to solve several examples, two of which will now be presented.

<u>Example 4-4</u> : Find the admissible region R for sources whose correlation is described by the following probability matrix :

$$P = \begin{bmatrix} 0.2 & 0 & 0 & 0 & 0 \\ 0 & 0.1 & 0 & 0.05 & 0.05 \\ 0 & 0 & 0.2 & 0 & 0 \\ 0 & 0 & 0 & 0.2 & 0 \\ 0 & 0 & 0 & 0 & 0.2 \end{bmatrix} .$$

The computer program found the admissible region R to be that shown in Figure 4-5, a region defined by three points C, D, and E. The execution time required to solve this problem was 0.74 seconds. Step A resulted in the finding of 17 admissible Type A precoders and Step B produced 27 admissible Type B precoders. The final step of the algorithm, then, was a search through 17 x 27 = 459 grid points. However, only 47 admissibility tests actually had to be performed since during the carrying out of the grid search, all other grid points were found to lie either below the line $\bar{n}_x + \bar{n}_y = H(XY)$ or inside what was already known to be part of the admissible region R. The 47 admissibility tests were accomplished as explained above by performing look-ups in a table, which in this case had a size of 10 x 14. While such tables may not result in great savings in computing time for simple problems such as Example 4-4, they save a tremendous amount of work in more complicated problems like the example which follows.

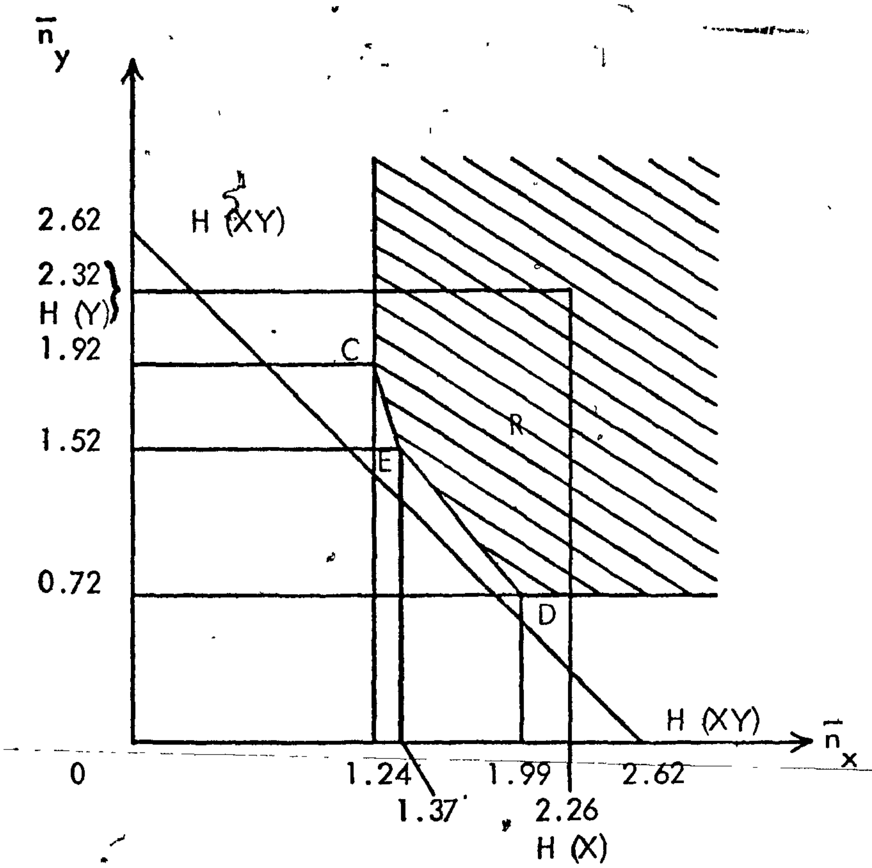


Figure 4-5: The admissible region for Example 4-4.

<u>Example 4-5</u>: Find the admissible region R if the probability matrix is

$$P = \begin{bmatrix} .05 & 0 & 0 & .025 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & .05 & 0 & 0 & .025 & 0 & .025 & 0 & 0 & 0 \\ .025 & 0 & .05 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & .05 & 0 & 0 & .025 & 0 & .025 & 0 \\ .025 & 0 & .025 & 0 & .05 & 0 & 0 & 0 & .025 & 0 \\ 0 & 0 & 0 & .025 & 0 & .05 & 0 & 0 & 0 & 0 \\ 0 & 0 & .025 & 0 & 0 & 0 & .05 & 0 & .025 & 0 \\ .025 & 0 & 0 & 0 & .025 & 0 & 0 & .05 & 0 & 0 \\ 0 & 0 & 0 & 0 & .025 & 0 & 0 & 0 & .05 & 0 \\ .025 & 0 & .025 & 0 & .025 & 0 & .025 & .025 & 0 & .05 \end{bmatrix} .$$

Notice that this matrix has only 30 non-zero entries out of 100. Steps A and B of the computer program resulted in the determination of 500 admissible Type A precoders and 205 Type B precoders. The execution of Steps A and B required five seconds computer time.

The final step of the algorithm, then, consisted of a grid search through 500 x 205 = 102,500 points. This was obviously the most time consuming step. The computer program used an additional execution time of 35 seconds to determine that the admissible region R is defined by four points as illustrated in Figure 4-6.

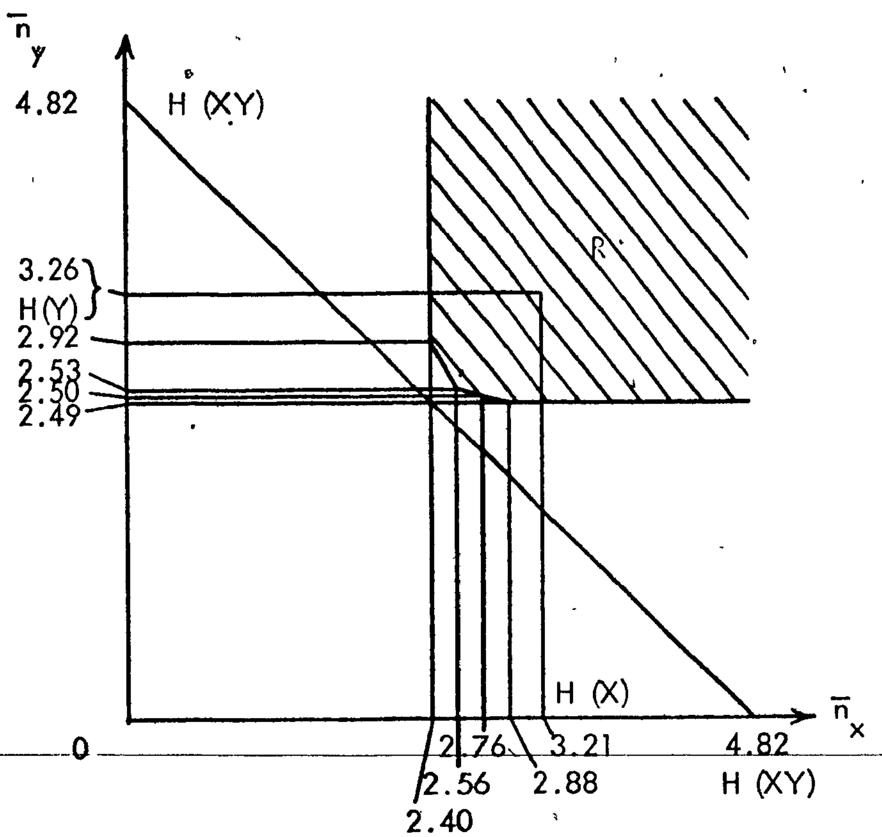


Figure 4-6 : The aumissible region for Example 4-5 .

The computer solution of this problem used a total storage area of 74,100 bytes.

It is interesting to note that out of the 102,500 grid points, far less than 10,000 admissibility tests had to be performed. These tests were carried out by referring to a table of 34 x 18 (612 elements). The great saving in computing time due to this table is obvious when one realizes that a precoder scheme such as

$$
\begin{aligned}
&\left.x_1\right] \rightarrow x_1' \\
&\left.x_3\right] \rightarrow x_2' \\
&\left.\begin{matrix} x_4 \\ x_5 \end{matrix}\right] \rightarrow x_3' \\
&\left.\begin{matrix} x_6 \\ x_7 \end{matrix}\right] \rightarrow x_4' \\
&\left.\begin{matrix} x_8 \\ x_9 \end{matrix}\right] \rightarrow x_5' \\
&\left.\begin{matrix} x_2 \\ x_{10} \end{matrix}\right] \rightarrow x_6'
\end{aligned}
\qquad\qquad
\begin{aligned}
&\left.\begin{matrix} y_1 \\ y_9 \end{matrix}\right] \rightarrow y_1' \\
&\left.\begin{matrix} y_2 \\ y_3 \\ y_6 \end{matrix}\right] \rightarrow y_2' \\
&\left.y_4\right] \rightarrow y_3' \\
&\left.y_5\right] \rightarrow y_4' \\
&\left.\begin{matrix} y_7 \\ y_8 \end{matrix}\right] \rightarrow y_5' \\
&\left.y_{10}\right] \rightarrow y_6'
\end{aligned}
$$

could be tested for admissibility by looking up just 12 entries in the table instead of the old way of partitioning matrix P and sorting through 100 entries to count the non-zero ones. Furthermore, the saving is magnified due to the fact that thousands of tests had to be done.

The successful solution by digital computer of the above examples and many others, has indicated that the methods developed in this chapter work very well. As expected, the method is fastest and most efficient for problems described by probability matrices containing very few zeroes. In such cases, the numbers of admissible Type A and Type B precoders are small. This means that only a small number of grid points are candidates for vertices of the admissible region and consequently that very few admissibility tests must be performed. Naturally, the method consumes more time for sparse probability matrices, but it is still an efficient method of solution as illustrated for the sparse 10 x 10 probability matrix of Example 4-5.

It should be pointed out, however, that although the above methods may be successful in solving problems of two correlated sources which are coded using sequences of length $L = 1$, they rapidly lose their usefulness with increasing L. The finding of admissible rate regions when codewords are provided for sequences of L source letters involves applying the same methods but to a new probability matrix of dimension $q^L \times k^L$ (see Example 3-4). Consequently, the amount of work required in solving these problems increases exponentially with L, thus quickly becoming impractical. Further study might help lessen this difficulty; it might, for example, be instructive to explore the progression of the sizes and shapes of admissible regions as L increases.

CHAPTER V

ALTERNATIVE CONFIGURATIONS IN CORRELATED SOURCE CODING

In this chapter, the theory of variable length coding will be extended to apply to all of the sixteen different coding configurations introduced in Chapter I (see Figure 1-4). It will be assumed here (just as in Chapters III and IV) that the encoders are constructed in two stages as illustrated in Figure 5-1. However, in order to deal with some of the new coding arrangements, it is necessary to generalize our definitions for the precoders. Consider, for example, the problem depicted in Figure 5-2, where a single precoder has knowledge of the outputs from two sources X and Y. We must redefine this precoder as being a transformation from the set of all (XY) outcomes $x_1 y_1, x_2 y_1, \ldots, x_k y_1, x_1 y_2, \ldots,$ and $x_k y_q$, to a new set of letters $z_1', z_2', \ldots, z_N'$. This definition can be further extended to allow coding for sequences of L (XY) outcomes.

It is revealing to determine for the system of Figure 5-2 what minimum average codeword length is required in order that the decoder can reproduce the output sequence from, for example, source Y with zero probability of error. Since Theorem 2-3 is not strong enough to handle this situation, a more general theorem will now be proven.

<u>Theorem 5-1</u> : For the system of Figure 5-2, the decoder can reconstruct the output sequence from source Y with zero probability of error if and only if $\bar{n}_y \geq H(Y)$.

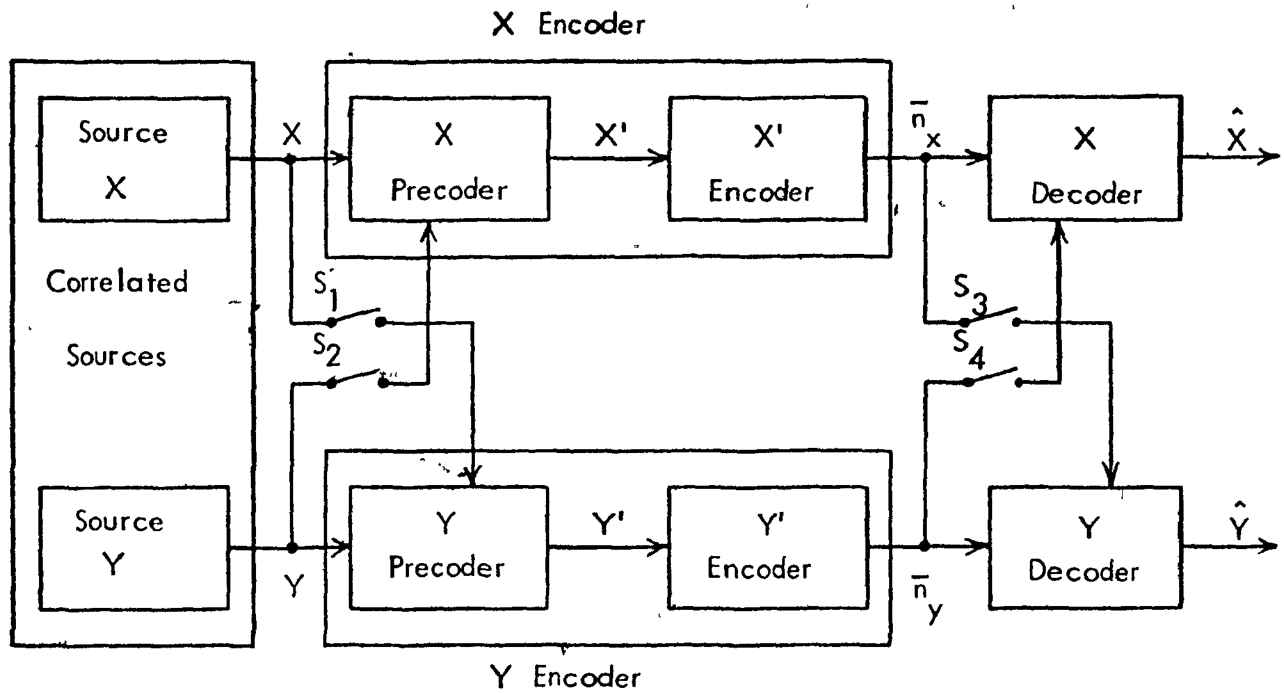


Figure 5-1 : Sixteen Correlated Source Coding Configurations.

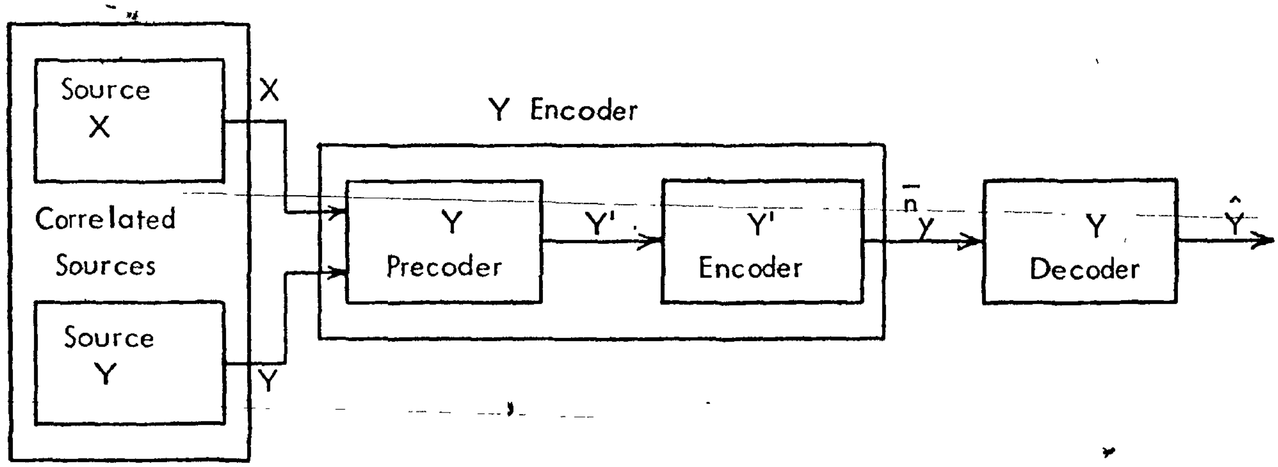


Figure 5-2 : The Coding Problem Studied in Theorem 5-1.

<u>Proof</u> : According to Theorem 2-3, the condition $\bar{n}_y \geq H(Y)$ is clearly sufficient to allow zero distortion communication from source Y to a decoder. An encoder can achieve $\bar{n}_y = H(Y)$ even if it entirely ignores the output from source X. It remains only for us to prove that $\bar{n}_y \geq H(Y)$ is also a necessary condition.

Assume initially that the probability matrix has no zero elements ; that is, $Pr(x_i, y_j) \neq 0$ for all i, j. The precoder must therefore satisfy the condition that only $x_i y_j$ pairs having the same Y letter y_j can be mapped onto the same Z' letter. For example, a situation having $x_1 y_1$ and $x_1 y_2$ both mapped onto z_1' would not be acceptable because if the decoder were to receive the codeword z_1', it would not be able to deduce whether y_1 or y_2 was actually transmitted by source Y . Of all precoders satisfying this restriction, the one having the lowest entropy is the following :

$$\left.\begin{matrix} x_1 y_1 \\ x_2 y_1 \\ \vdots \\ x_k y_1 \end{matrix}\right] \rightarrow z_1'$$

$$\vdots$$

$$\left.\begin{matrix} x_1 y_i \\ x_2 y_i \\ \vdots \\ x_k y_i \end{matrix}\right] \rightarrow z_i'$$

$$\vdots$$

$$\left.\begin{matrix} x_1 y_q \\ x_2 y_q \\ \vdots \\ x_k y_q \end{matrix}\right] \rightarrow z_q' ,$$

This scheme has an entropy of

$$H(Z') = -\sum_{j=1}^{q} \left[\sum_{i=1}^{k} \Pr(x_i, y_j)\right] \log_2 \left[\sum_{i=1}^{k} \Pr(x_i, y_j)\right]$$

$$= -\sum_{j=1}^{q} \Pr(y_j) \log_2 \Pr(y_j) = H(Y) .$$

Therefore, according to Theorem 2-3, the decoder of Figure 5-2 can only reconstruct the output of source Y with zero probability of error if the average codeword length satisfies $\bar{n}_y \geq H(Y)$. This conclusion is still valid for the case when zeroes are allowed in the probability matrix because if any outcome (x_i, y_j) occurs with probability zero, the precoder may map $x_i y_j$ onto any one of the z' letters without changing the probability of error. Moreover, the entropy calculation is unaffected since $\Pr(x_i, y_j) \log_2 \Pr(x_i, y_j) = 0$.

The above result also holds for the situation where the precoder is allowed to be a mapping for sequences of (XY) outcomes. This can be shown easily by pretending that the letters $x_1, x_2, \ldots, x_k$, and $y_1, y_2, \ldots, y_q$ actually represent sequences of L letters from simpler sources.

Let us first consider the four correlated source coding configurations in Figure 5-3, all of which are characterized by uncoupled decoders. For Case 1, sources X and Y are encoded and decoded independently and therefore the results of Chapter II for single sources can be applied. That is, the source outputs for Case 1 can be communicated to the decoders with zero distortion if and only if $\bar{n}_x \geq H(X)$ and $\bar{n}_y \geq H(Y)$.

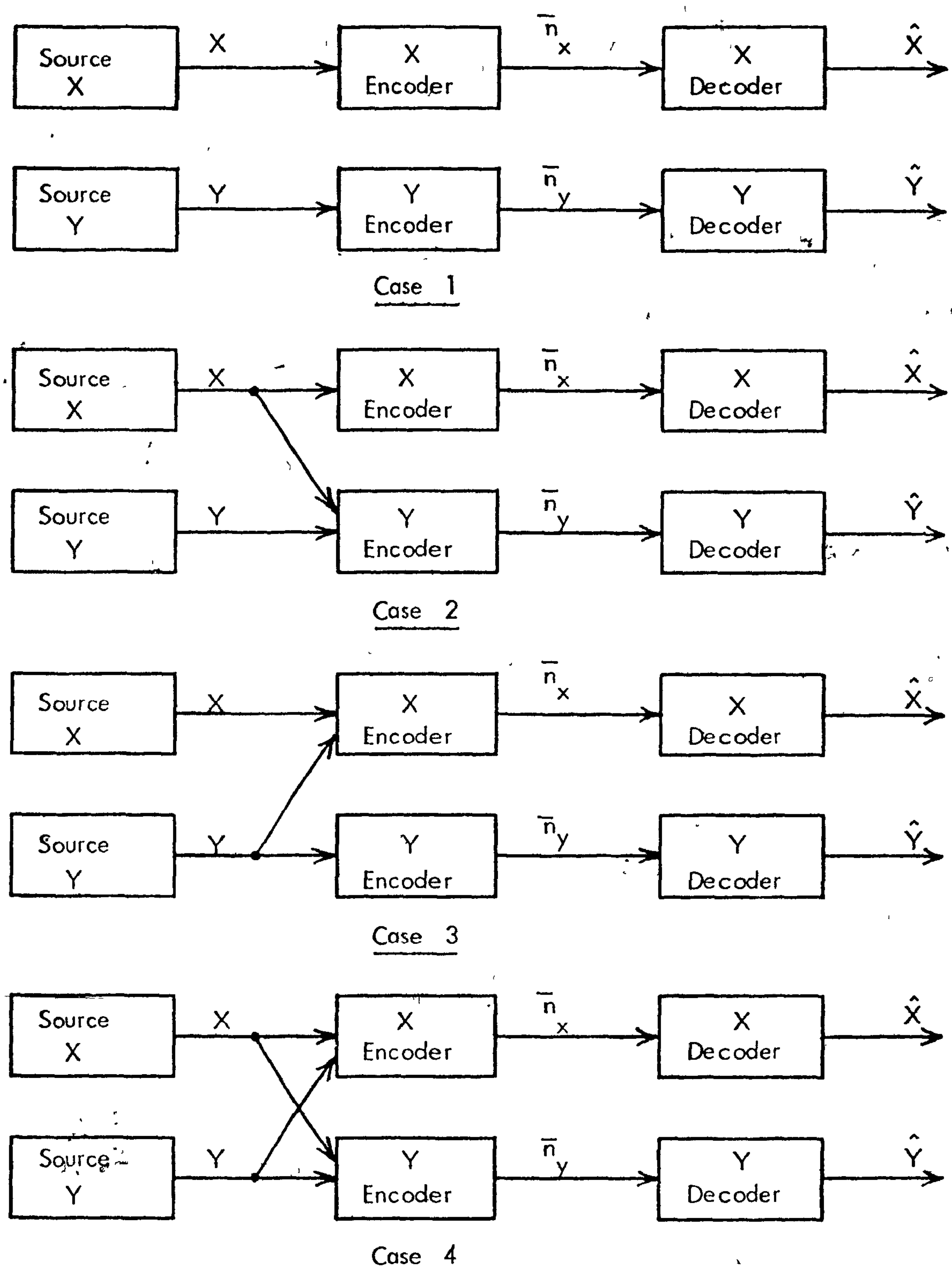


Figure 5-3: Correlated Source Coding with Uncoupled Decoders.

Cases 2, 3, and 4 appear to be slightly more complicated but by applying Theorems 5-1 and 2-3 to these problems, it is easy to see that the admissible regions are all identical to that for Case 1, namely the area $\bar{n}_x \geq H(X)$, $\bar{n}_y \geq H(Y)$. It should be emphasized that these results hold even when coding is permitted for sequences of source letters.

Consider next the four correlated source coding arrangements illustrated in Figure 5-4, those characterized by completely coupled decoders. Of these four problems, Case 8 has the simplest solution. Since both of its encoders are allowed to see the outputs from both sources X and Y, the entire source output information can be communicated to the decoder through either one of the encoders exclusively. Thus, (H (XY), 0) and (0, H (XY)) are admissible rate points and according to Theorems 3-2, 3-3, and 3-4, the admissible region R for Case 8 must be the area $\bar{n}_x + \bar{n}_y \geq H(XY)$.

Unfortunately, it is not possible to derive any such simple formula to describe the admissible region for Case 5. This fact has already been demonstrated very clearly, Case 5 being none other than the problem studied in such detail in Chapters III and IV. However, with the aid of the following theorem, it will be shown that the admissible region for Case 6 takes a very simple form. Note that Case 7 does not need to be studied separately because it is just a symmetric version of Case 6, formed by interchanging the roles of X and Y.

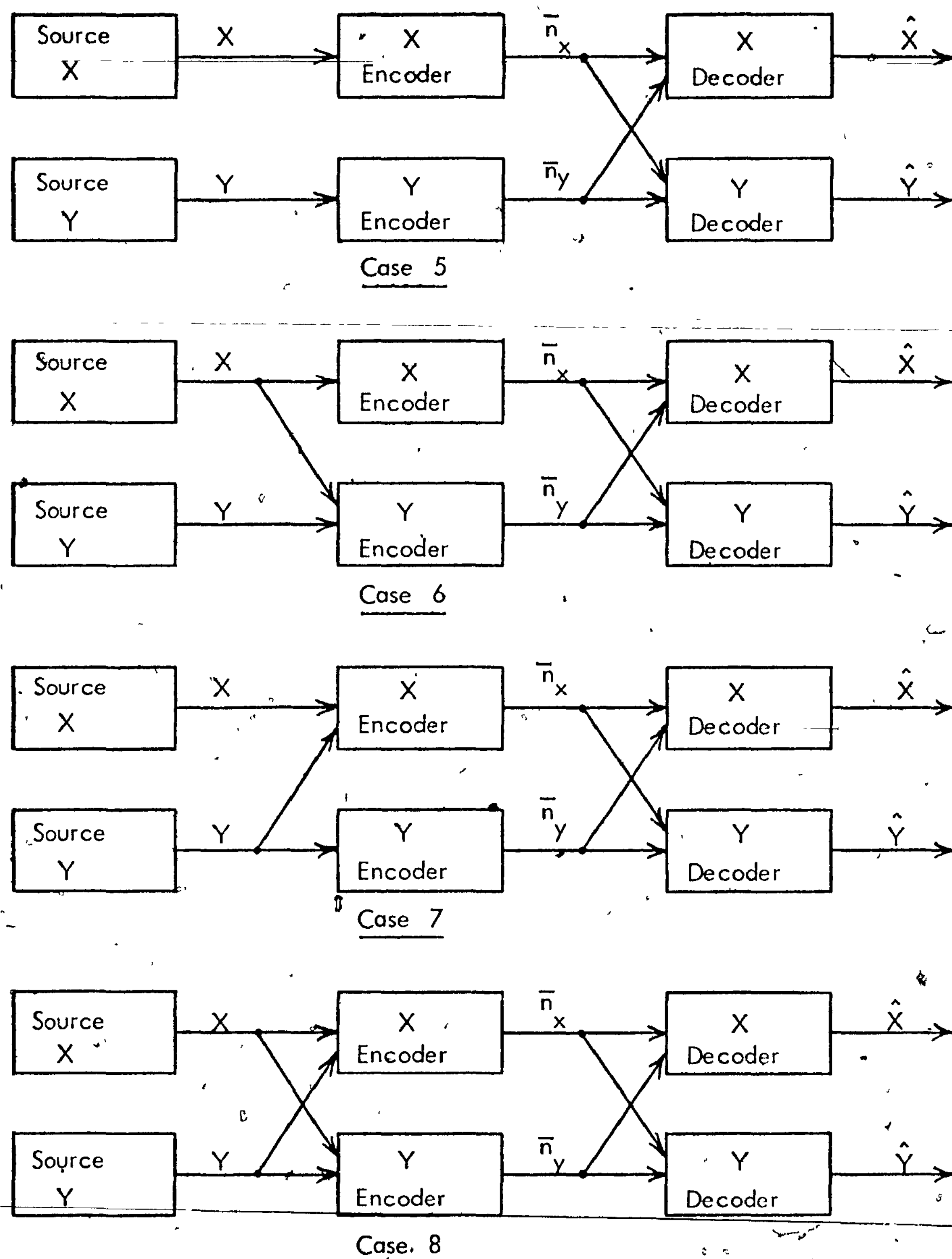


Figure 5-4: Correlated Source Coding with Coupled Decoders.

Theorem 5-2 : (see Lu [3]) For the system illustrated in Figure 5-5, the point $(\bar{n}_x, \bar{n}_y) = (H(X), H(Y \mid X))$ is an admissible rate point, where $H(Y \mid X)$ is called the conditional entropy of Y given X and is defined by

$$H(Y \mid X) \triangleq - \sum_{i=1}^{k} \sum_{j=1}^{q} Pr(x_i, y_j) \log_2 Pr(y_j \mid x_i) = H(XY) - H(X).$$

Proof : Since the output of source X is always known to the Y precoder, it is possible to design a Y precoder which employs several different coding strategies depending on what outcome is produced by source X. Consider, for example, a Y precoder which operates in the following manner : when the output of source X is x_i, the Y source letters are Huffman coded according to the set of probabilities $Pr(y_1 \mid x_i)$, $Pr(y_2 \mid x_i)$, ..., and $Pr(y_q \mid x_i)$. By defining $\bar{n}_{y \mid x_1}$, $\bar{n}_{y \mid x_2}$, ..., and $\bar{n}_{y \mid x_k}$ to be the average codeword lengths corresponding to these k different Y-codes, the overall average codeword length $\bar{n}_y$ can be expressed as follows :

$$\bar{n}_y = \sum_{i=1}^{k} Pr(x_i)\, \bar{n}_{y \mid x_i}.$$

Note that it would be possible for the above Y encoder to transmit the output sequence from source Y to the Y decoder with no errors if only the Y decoder ~~also had knowledge of the X output sequence~~ (because knowing X would enable the decoder to deduce which of the k coding schemes was actually used by the Y precoder). Distortionless communication can therefore be guaranteed by choosing the X precoder to be uniquely decodable (which it must be in any case).

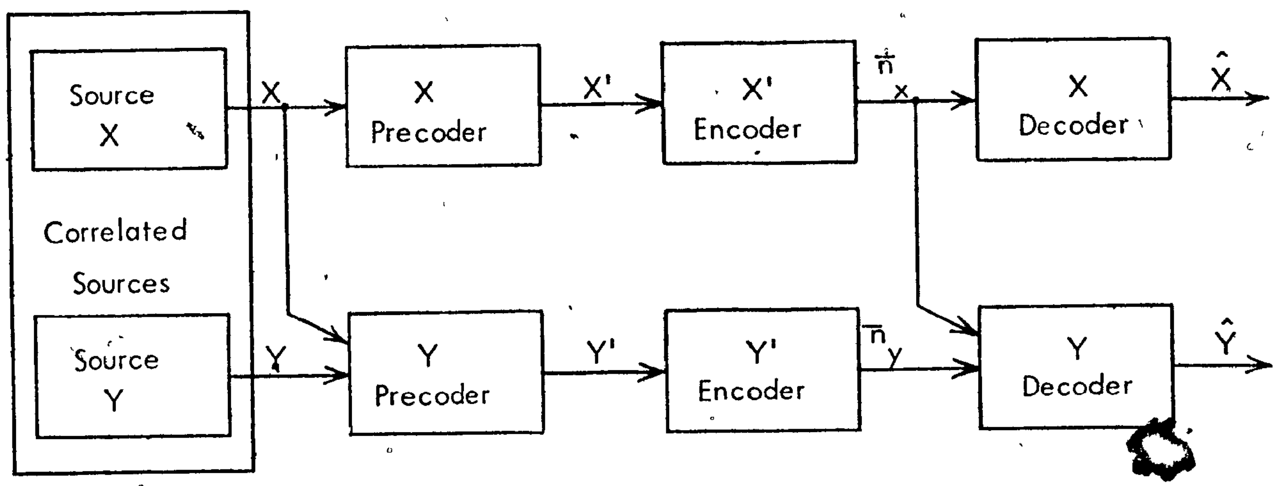


Figure 5-5 : The source coding system studied in Theorem 5-2.

Now let us calculate the minimum average codeword lengths for the above coding strategies. Since the X precoder is uniquely decodable, it is of course sufficient that $\bar{n}_x \geq H(X)$. Consider the coding scheme used by the Y encoder when $X = x_i$. Theorem 2-3 tells us that the Y output can always be reconstructed correctly by the Y decoder as long as

$$\bar{n}_{y \mid x_i} \geq -\sum_{j=1}^{q} \Pr(y_j \mid x_i) \log_2 \Pr(y_j \mid x_i) .$$

By combining this statement with the expression for $\bar{n}_y$, we see that zero-error communication is possible through the Y-channel of the system of Figure 5-5 provided that

$$\bar{n}_y \geq -\sum_{i=1}^{k} \Pr(x_i) \left[\sum_{j=1}^{q} \Pr(y_j \mid x_i) \log_2 \Pr(y_j \mid x_i)\right]$$

$$= -\sum_{i=1}^{k} \sum_{j=1}^{q} \Pr(x_i, y_j) \log_2 \Pr(y_j \mid x_i) = H(Y \mid X) .$$

Thus, the point $(\bar{n}_x, \bar{n}_y) = (H(X), H(Y \mid X))$ must be an admissible rate point for the system of Figure 5-5 .

Theorem 5-2 can be applied directly to any Case 6 problem and it implies that the point $(H(X), H(Y \mid X))$ must belong to its admissible region R . We can also assert that $(0, H(XY)) \in R$ because, as can be seen in Figure 5-4, it is possible for all source information to be conveyed to the decoders by way of the

Y channel alone. Since the points (H (X), H (Y | X)) and (0, H (XY)) both lie on the line $\bar{n}_x + \bar{n}_y$ = H (XY), Theorem 3-4 tells us that the entire admissible region R for any Case 6 system is simply that shown in Figure 5-6. Observe that no points below the horizontal line $\bar{n}_y$ = H (Y | X) can belong to R because even if $\bar{n}_x$ is increased above the value H (X), the X encoder can be no better than uniquely decodable and consequently no further improvements can be made to the Y encoder.

Even though the admissible region for Case 6 has now been discovered, it is still useful to mention how one might go about determining the set of admissible precoders (this problem will arise in connection with Case 10). A careful inspection of the Case 6 system reveals that it can be redrawn as in Figure 5-7 so as to make it appear very similar to a Case 5 problem, the most notable difference being that source Y has been replaced by the joint source XY. By relabeling source XY as a new source Z and by realizing that the decoder of Figure 5-7 can be thought of as producing estimates $\hat{X}$ and $\hat{Z}$ rather than $\hat{X}$ and $\hat{Y}$ (these two situations are equivalent because X and Z (X, Y) can be recovered with zero probability of error by a decoder if and only if X and Y can), it follows that the system of Figure 5-7 is mathematically identical to a Case 5 problem for two correlated sources X and Z. Therefore, by setting up a new probability matrix P_{new} to describe the correlation between sources X and XY instead of between sources X and Y, the methods of Chapters III and IV can be applied directly to find the admissible precoders for any Case 6 problem.

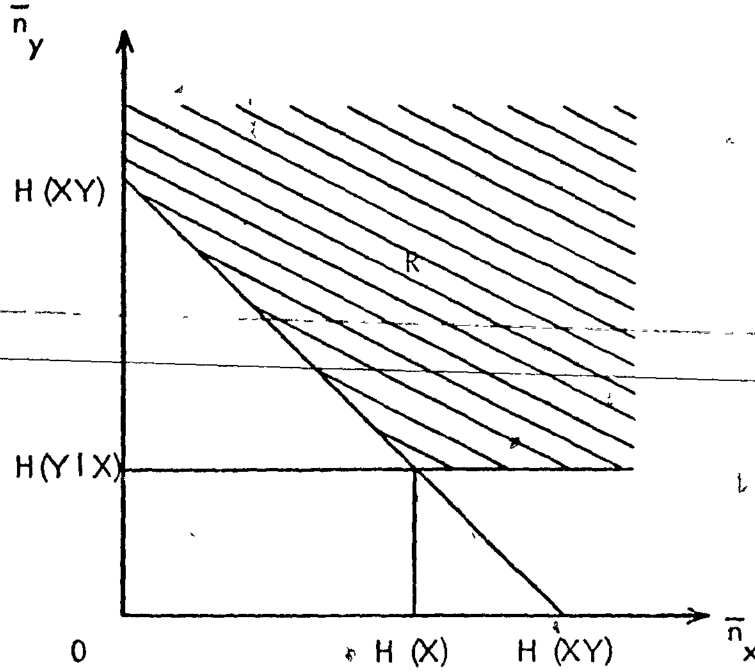


Figure 5-6 : The admissible region for Case 6 .

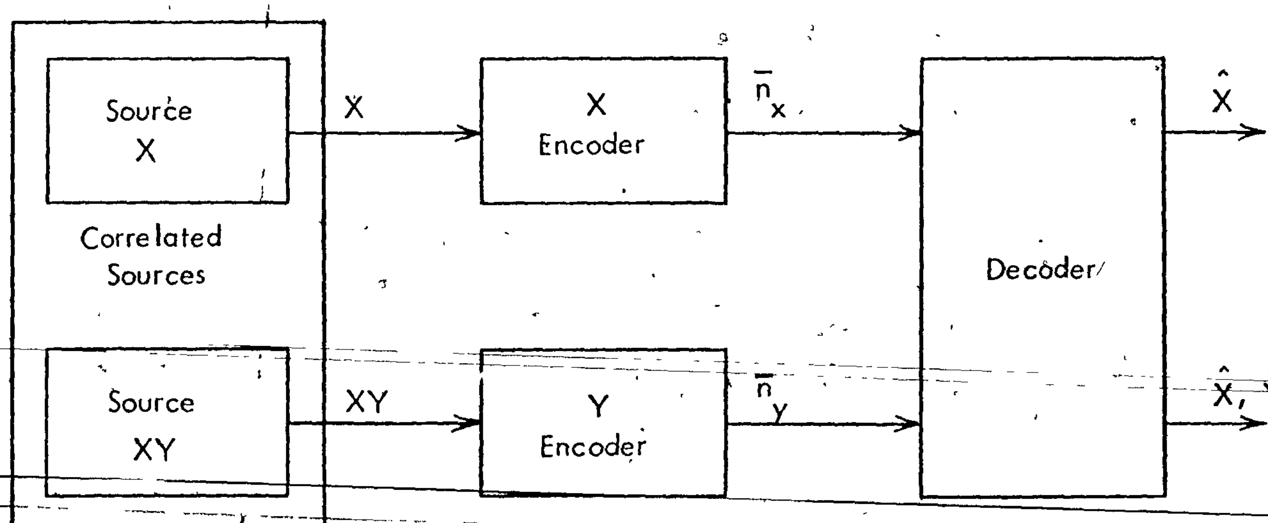


Figure 5-7 : An alternative view of the Case 6 system.

Consider now the four systems depicted in Figure 5-8. These cases are characterized by a Y decoder which operates without knowledge of the X encoded message stream even though the X decoder has access to both codeword sequences. While it is not trivial to determine the admissible regions for Cases 9 and 10, very simple regions exist for Cases 11 and 12. For Case 11, it follows directly from Theorem 5-2 (by interchanging X and Y) that (H (X | Y), H (Y)) is an admissible point. Since this point lies on the line $\bar{n}_x + \bar{n}_y = H(XY)$ and since the Y encoder for Case 11 must be uniquely decodable, the admissible region R is just the rectangular area shown in Figure 5-9.

Theorem 5-2 can also be applied to Case 12 to show that (H (X | Y), H (Y)) ϵ R. Since all source information for Case 12 can be transmitted to the decoders through the Y channel alone, it can also be stated that (0, H (XY)) ϵ R. By realizing that these two points lie on the line $\bar{n}_x + \bar{n}_y = H(XY)$ and that Theorem 5-1 requires that $\bar{n}_y \geq H(Y)$, the entire admissible region R for Case 12 is found to be that drawn in Figure 5-10.

In order to solve the Case 9 problem, it is necessary to make use of some of the results of Chapters III and IV. This is a logical approach because there turns out to be a close relationship between the Case 9 and Case 5 configurations. In fact, it is easy to recognize that any precoder scheme which is admissible for Case 9 must also be admissible for Case 5. However, since the Y encoder for Case 9 must always be uniquely decodable, the set of all admissible

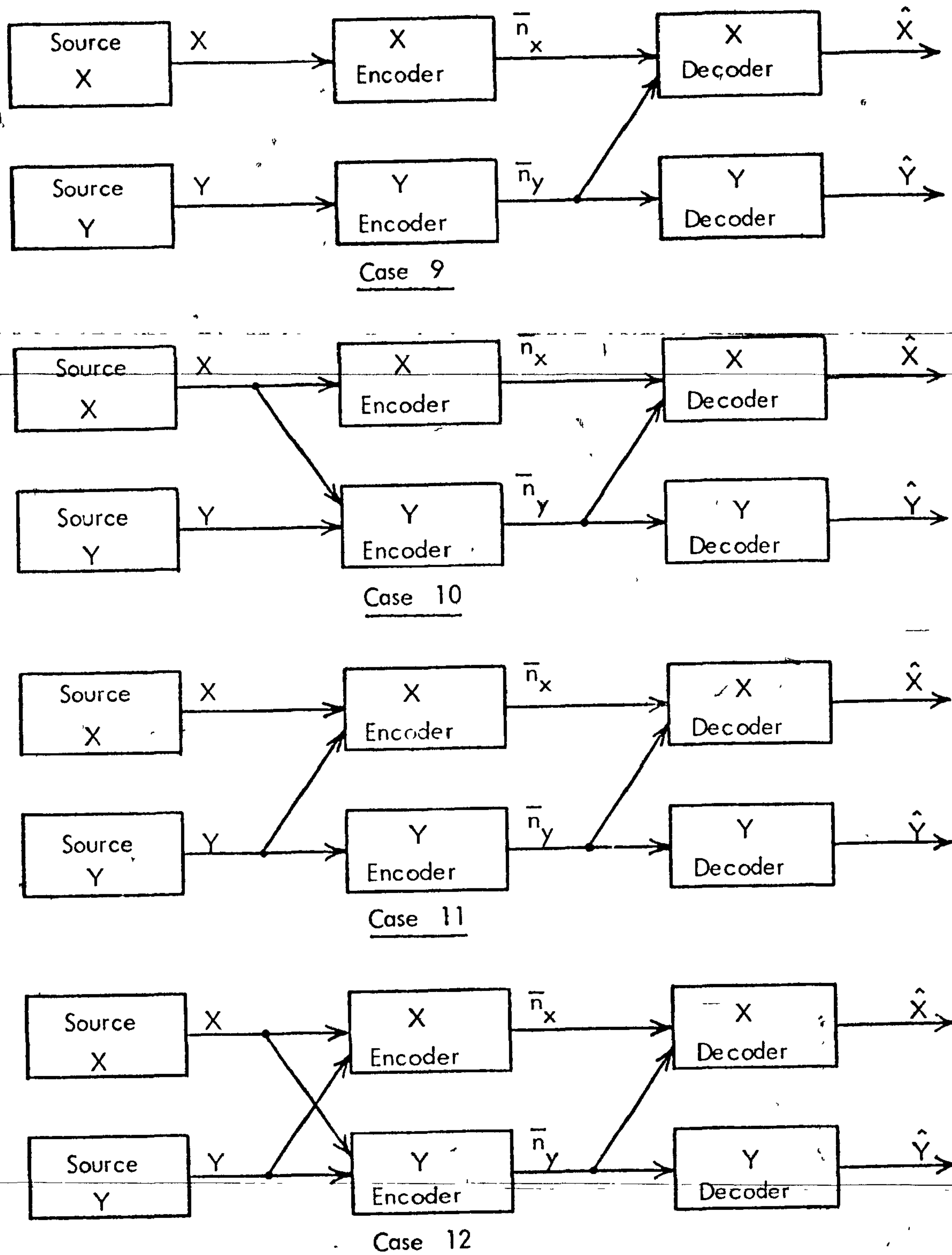


Figure 5-8 : Correlated Source Coding with Partially Coupled Decoders.

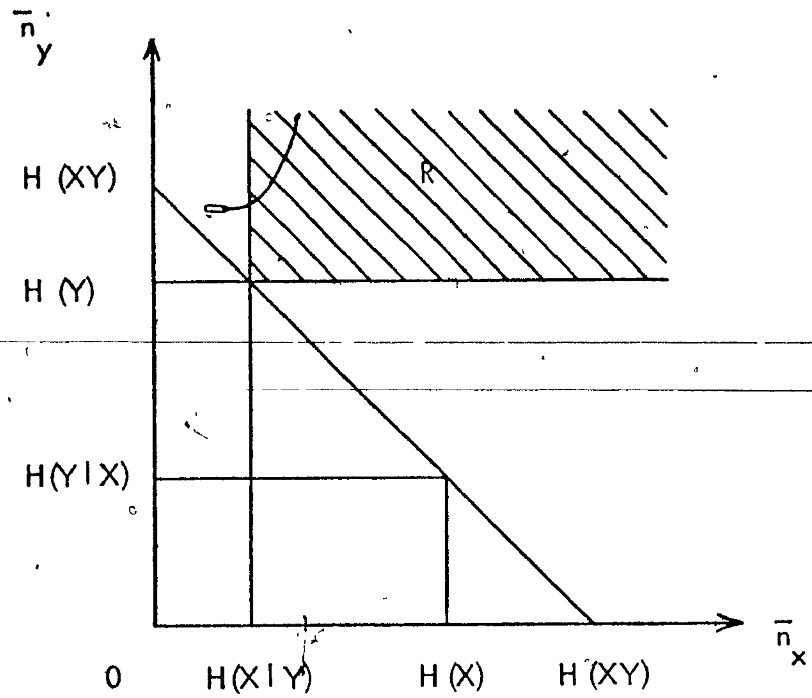


Figure 5-9 : The admissible region for Case 11 .

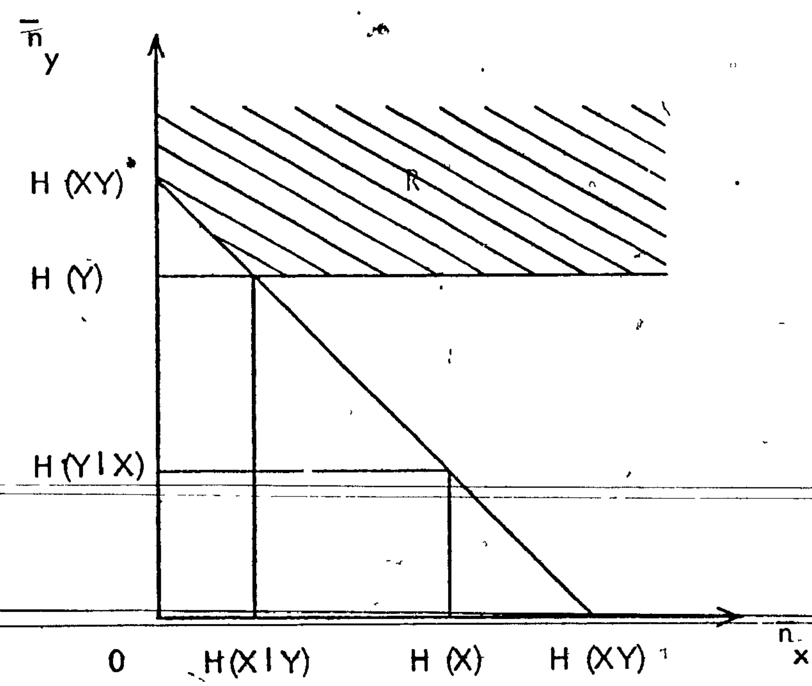


Figure 5-10 : The admissible region for Case 12 .

precoders for Case 9 must include only the admissible Type A precoders of Case 5. This result allows Case 9 to be solved simply by performing Step A, as described in Chapter. IV, for the corresponding Case 5 problem. The resulting admissible region R has a simple rectangular form as illustrated in Figure 5-11.

It is possible to solve Case 10 by exploiting its similarity to Case 6. It is a fact that any admissible precoder for Case 10 must be admissible for Case 6 also. However, the reverse is not true because, as seen in Figure 5-8, the Y encoder for Case 10 must be uniquely decodable, a restriction not present in the Case 6 problem. Indeed, a little thought will show that any admissible precoder for Case 6 will be admissible for Case 10 if and only if its Y precoder is such that Y can be decoded independently of X.

Therefore, a method of solving Case 10 can proceed as follows. Find the sets of all admissible Type A and Type B precoders for the corresponding Case 6 problem. From the list of Type B precoders, reject those for which Y cannot be uniquely decoded independently of the encoded X message stream. Consider all possible ways of choosing a Y precoder from this reduced list and an X precoder from the full list of admissible Type A precoders. Search through the resulting grid, using exactly the same method as followed in Chapter IV, to find the admissible region R.

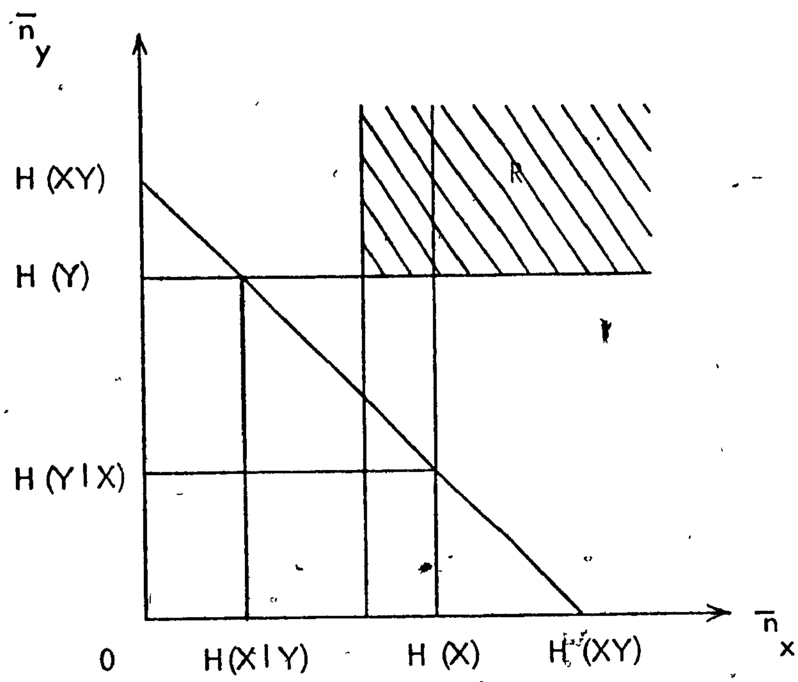
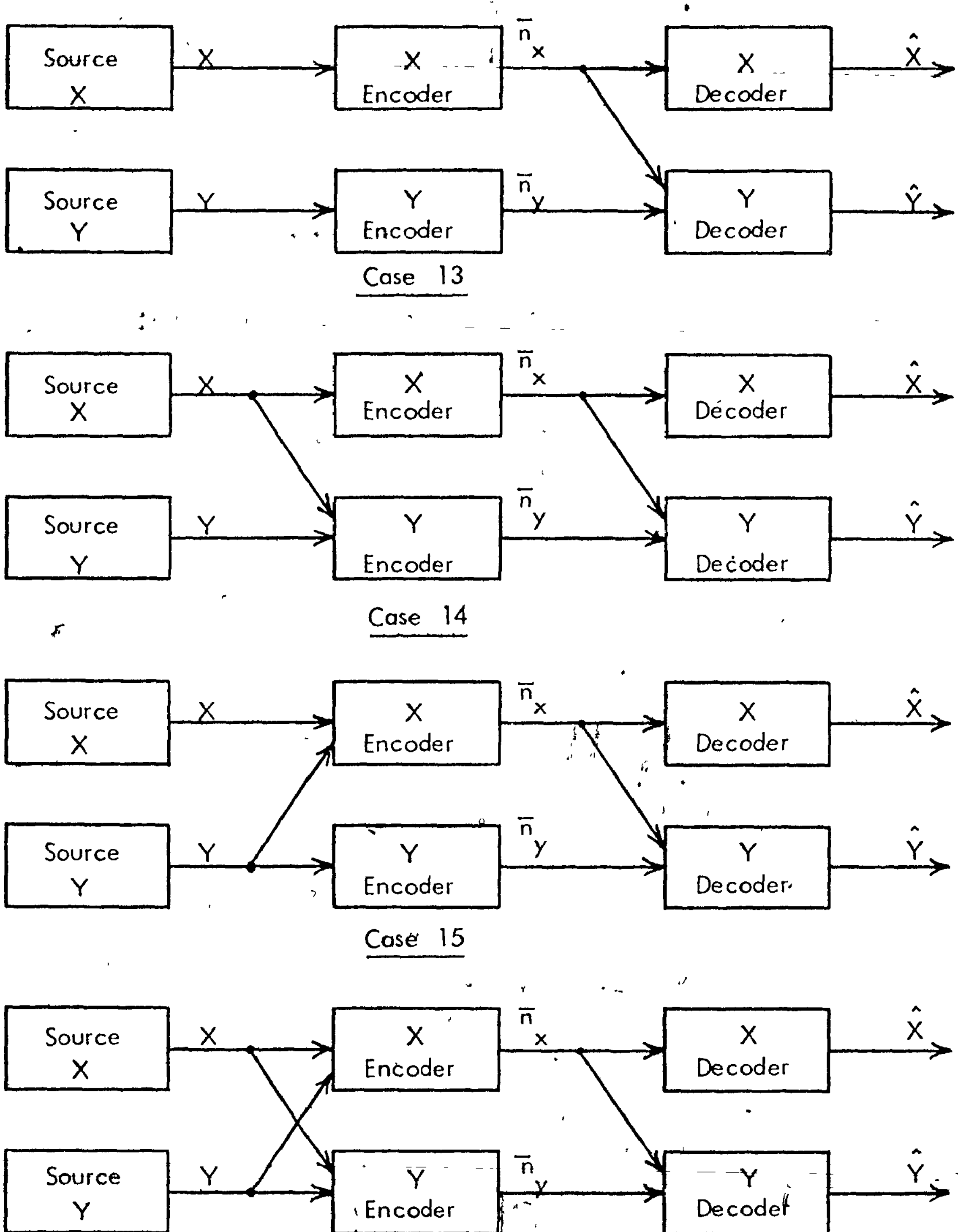


Figure 5-11 : The admissible region for Case 9 .

The only remaining correlated source coding configurations to be considered are the four cases illustrated in Figure 5-12. Cases 13, 14, 15, and 16 correspond exactly to Cases 9, 11, 10, and 12 respectively except that the roles of X and Y have been reversed. Consequently, the problems of Figure 5-12 can be solved using the methods developed above.



Case 13

Case 14

Case 15

Case 16

Figure 5-12: Correlated Source Coding with Partially Coupled Decoders.

CHAPTER VI

CONCLUSION

In summary, the principal achievement of this thesis has been the presentation of a compact theory regarding variable length coding for two correlated information sources. Methods have been successfully derived for the purpose of determining admissible rate regions for correlated sources which are encoded in any one of sixteen possible configurations. For simplicity, most of the results in preceding chapters were established by assuming that the precoders could only provide codewords for individual source letters. But in Example 3-4, it was shown how this situation can be easily generalized to include coding for sequences of L source letters. Actually, this approach in correlated source coding is analagous to that used in Huffman coding for single sources, because in both cases the finding of optimum codes depends upon an initial assumption as to the length of sequences of source letters to be encoded. In fact, the major contribution of this thesis may be considered to be the generalization of the Huffman code to more complicated source structures.

Now that the theory of joint coding for two sources has been thoroughly investigated, it is quite natural to inquire if the results of this thesis can be extended to the problem of variable length coding for N correlated sources (see Cover [1]). To try and answer this question, consider the arrangement of three sources illustrated in Figure 6-1. Just as in Chapter III, it is convenient as

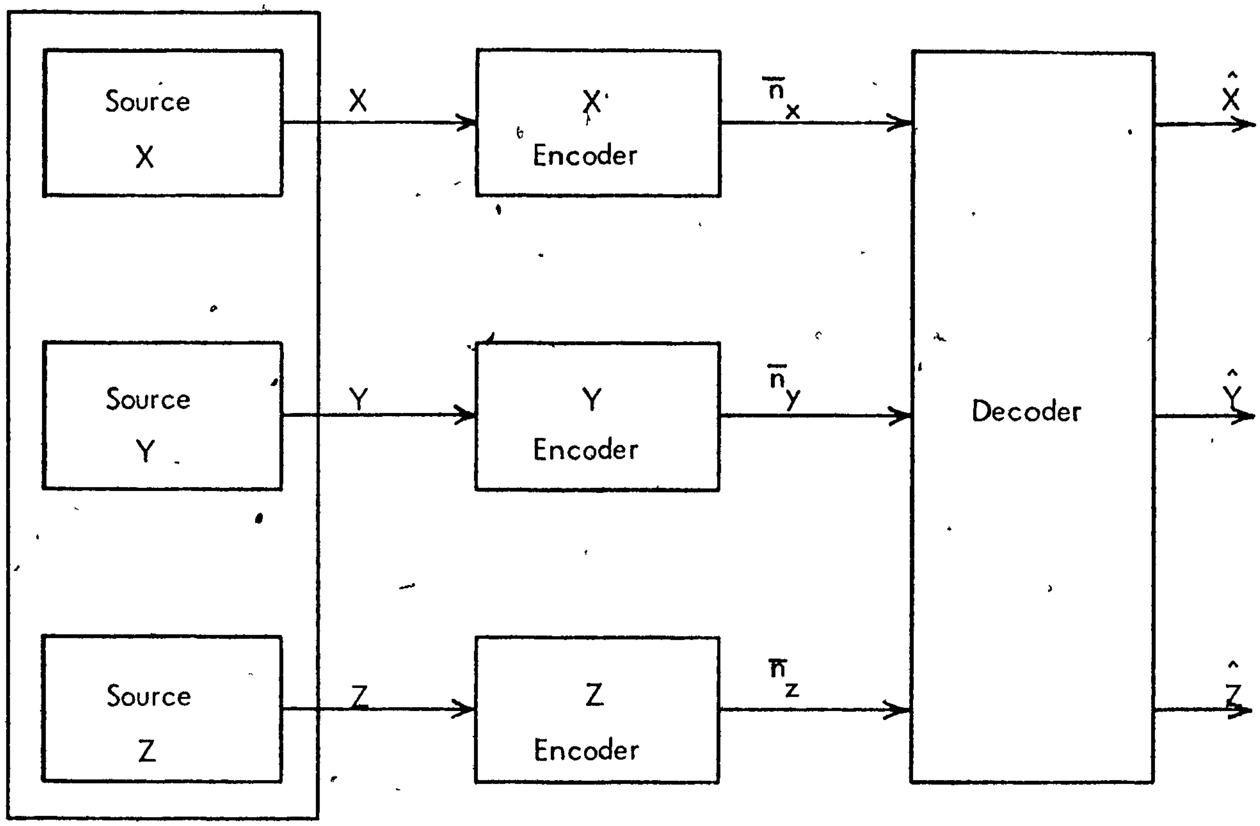


Figure 6-1 : Coding for Three Correlated Sources.

an initial step to set up a probability matrix P to describe the correlation between these sources. As indicated in Figure 6-2, however, a three-dimensional array is needed to store the complete set of joint probabilities. Proceeding as in Chapter III, it is possible to prove necessary and sufficient conditions concerning the admissibility of precoder schemes for the system of Figure 6-1. As before, the reordering of source letters defined by any specific precoder scheme suggests a corresponding way of subdividing the probability array of Figure 6-2 into a set of three-dimensional blocks. It turns out that if each such block contains at most one non-zero element, then and only then will the precoder scheme be admissible.

It can be shown that all the important concepts of Chapter III can be extended to apply not only to the above system of three sources, but more generally to the problem of N correlated sources. Unfortunately, however, the replacement of matrices by higher order tensors makes the theory much more difficult to visualize. Furthermore, even if it were possible to generalize the solution methods of Chapter IV and compose computer programs to tackle multi-source problems, the amount of computing work required would quickly become unmanageable with an increasing number of correlated sources. Besides this, there is even one more complication not mentioned until now, the fact that in undertaking a complete study of the problem of N correlated sources, it is necessary to consider $2^{2N(N-1)}$ different configurations for the encoders and decoders, a seemingly hopeless proposition for $N \geq 3$.

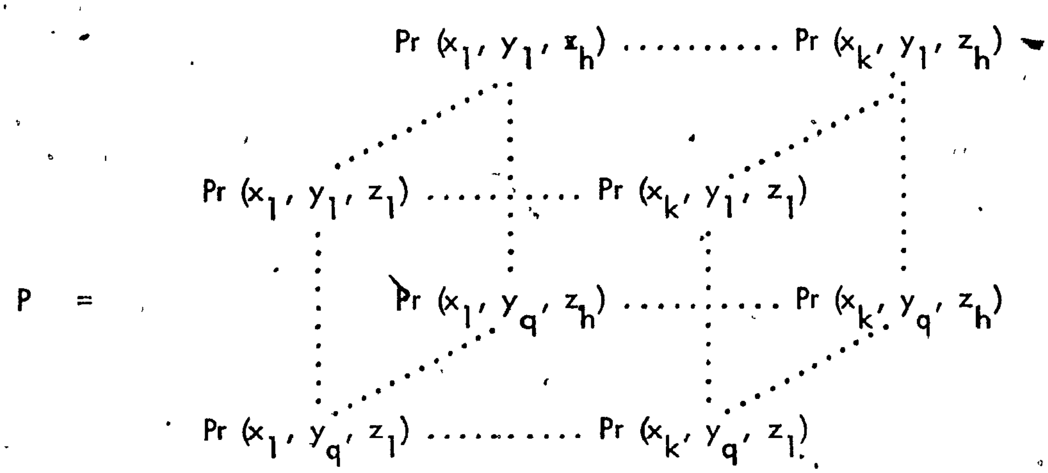


Figure 6-2 : A Probability Array for Three Correlated Sources.

In conclusion, it is hoped that this thesis has been successful in shedding much light on the subject of variable length coding for two or more correlated sources, a problem which has been unsolved for some time. Nevertheless, many questions still remain unanswered and several related problems remain to be explored.

BIBLIOGRAPHY

[1] Cover, T. M., "A Proof of the Data Compression Theorem of Slepian and Wolf for Ergodic Sources," submitted to IEEE Transactions on Information Theory.

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

[3] Lu, S. C., "Variable Length Source Coding Theorem for Correlated Information Sources," submitted to IEEE Transactions on Information Theory.

[4] Slepian, D., and Wolf, J., "Noiseless Coding of Correlated Information Sources," IEEE Transactions on Information Theory, Vol. IT-19, No. 4, pp. 471-480, July 1973.

[5] Wyner, A. D., "A Theorem on the Entropy of Certain Binary Sequences and Applications : Part II," IEEE Transactions on Information Theory, Vol. IT-19, No. 6, pp. 772-777, November 1973.