INFORMATION TO USERS

This material was produced from a microfilm copy of the original document. While the most advanced technological means to photograph and reproduce this document have been used, the quality is heavily dependent upon the quality of the original submitted.

The following explanation of techniques is provided to help you understand markings or patterns which may appear on this reproduction.

1. The sign or "target" for pages apparently lacking from the document photographed is "Missing Page(s)". If it was possible to obtain the missing page(s) or section, they are spliced into the film along with adjacent pages. This may have necessitated cutting thru an image and duplicating adjacent pages to insure you complete continuity.

2. When an image on the film is obliterated with a large round black mark, it is an indication that the photographer suspected that the copy may have moved during exposure and thus cause a blurred image. You will find a good image of the page in the adjacent frame.

3. When a map, drawing or chart, etc., was part of the material being photographed the photographer followed a definite method in "sectioning" the material. It is customary to begin photoing at the upper left hand corner of a large sheet and to continue photoing from left to right in equal sections with a small overlap. If necessary, sectioning is continued again—beginning below the first row and continuing on until complete.

4. The majority of users indicate that the textual content is of greatest value, however, a somewhat higher quality reproduction could be made from "photographs" if essential to the understanding of the dissertation. Silver prints of "photographs" may be ordered at additional charge by writing the Order Department, giving the catalog number, title, author and specific pages you wish reproduced.

5. PLEASE NOTE: Some pages may have indistinct print. Filmed as received.

Xerox University Microfilms
300 North Zeib Road
Ann Arbor, Michigan 48106
SEQUENTIAL CODING FOR CHANNELS WITH FEEDBACK
AND A CODING THEOREM FOR A CHANNEL WITH
SEVERAL SENDERS AND RECEIVERS

DISSERTATION

Presented in Partial Fulfillment of the Requirements for the
Degree Doctor of Philosophy in the Graduate School of
The Ohio State University

By
Michael L. Ulrey, A.B., M.Sc.

The Ohio State University
1973

Reading Committee:
Dr. Rudolf Ahlswede
Dr. Charles Saltzer
Dr. Thomas Schwartzbauer

Approved By
Rudolf Ahlswede
Adviser
Department of Mathematics
ACKNOWLEDGMENTS

I wish to express my deepest gratitude to Professor Rudolf Ahlswede, whose excellent teaching and indispensable advice and suggestions made the writing of this dissertation possible.
VITA

February 9, 1945 ........................ Born - Columbus, Ohio

1967 ........................................ A.B., Kenyon College, Gambier, Ohio

1967-1968 ................................. NDFA Title IV Fellow, The Ohio State University Columbus, Ohio

1968-1969 ................................. Teaching Assistant, The Ohio State University Columbus, Ohio

1969 ........................................ M.Sc., The Ohio State University Columbus, Ohio

1969-1972 ................................. NDFA Title IV Fellow, The Ohio State University Columbus, Ohio

1972-1973 ................................. Teaching Associate, The Ohio State University Columbus, Ohio

FIELDS OF STUDY

Major Field: Mathematics


Studies in Probability. Professor Fredos Papangelou.

Studies in Topology. Professor Norman Levine

Studies in Information Theory. Professor Rudolf Ahlswede.
TABLE OF CONTENTS

Acknowledgments .................................................. ii
Vita ........................................................................ iii
Introduction ........................................................... 1

Chapter 1 Basic Results in Probabilistic Coding Theory for the DMC ........................................... 3
1.1 Statement of the Problem and Some Fundamental Results ......................................................... 3
1.2 The Coding Theorem for the DMC and Fano's Lemma ................................................................. 6

Chapter 2 Sequential Coding for the DMC in the Presence of Noiseless Feedback ......................... 10
2.1 The DMC with Feedback and Sequential Coding ........................................................................ 10
2.2 The Weak Converse for a DMCF in the Case of Sequential Coding ........................................... 13
2.3 A Counterexample to the Strong Converse ................................................................................. 16

Chapter 3 A Coding Theorem for a Channel With Several Senders and Receivers ......................... 18
3.1 The Channel Model and Statement of the Coding Problem ......................................................... 18
3.2 A General Fano-Type Estimate ................................................................................................... 23
3.3 Capacity Region of a Channel with s Senders and One Receiver ............................................. 28
3.4 Capacity Region of a Channel with s Senders and r Receivers .............................................. 36

Bibliography .................................................................. 41
INTRODUCTION

In Chapter 2 of this paper we consider a discrete memoryless channel (DMC) in the case where noiseless feedback is present. Shannon [15] showed that the capacity of the DMC is not increased by feedback (in the case of block coding) by proving a weak converse. Later Kemperman [10] (see also [17], Chapter 4) and Kesten [17], Chapter 4) independently proved a strong converse.

The coding theorem for a DMC with feedback in the case of block coding is therefore a direct consequence of the coding theorem for a DMC. The known proofs of the coding theorem for a DMC use either a random coding method (Shannon [12]) or a maximal coding method (Feinstein [7], Wolfowitz [16]). The presence of feedback enabled Ahlswede [2] to give a proof of the coding theorem which is not based on either of these methods. His proof explicitly exhibits an encoding-decoding scheme which can perform with a probability of error below a previously given level while maintaining a rate arbitrarily close to capacity.

In the presence of feedback, it is natural to try sequential coding, which Norstein [9] did for the binary symmetric channel with feedback. In Chapter 2, we prove a weak converse for a DMC with feedback, and show that the strong converse does not hold in the case of sequential coding.

In Chapter 3, we consider a channel with several senders and
receivers. Multi-way channels were first studied by Shannon in his basic paper "Two-way communication channels." [14]. In [3], Ahlswede has defined and classified multi-way channels of various kinds and proved simple characterizations for the capacity regions of channels with a) two senders and one receiver and b) three senders and one receiver.

Recently, Ahlswede [4] has found a new approach to the coding problem for a channel with two senders and one receiver which led to an alternate characterization of the capacity region of this channel. This new approach seems to be more canonical than the earlier one, and was used successfully in determining the capacity region of a channel with two senders and two receivers in the case both senders send messages simultaneously to both receivers [4]. It was conjectured in [4] that this approach extends to the channel with several senders and receivers in the case all senders send independent messages simultaneously to all receivers.

In Chapter 3, we prove the conjecture to be true. Thus we obtain a characterization of the capacity regions of the described channels, which could be used for their numerical determination.
CHAPTER 1
BASIC RESULTS IN PROBABILISTIC CODING
THEORY FOR THE DMC

§1.1 Statement of the Problem and Some Fundamental Results

We give here a description of the discrete memoryless channel (DMC) and a brief sketch of some of the main results in order to introduce the notions and make needed results available for later use.

Let $X$ and $Y$ be finite sets which are called the input and output alphabets, respectively, of the channel. For each $t = 1, 2, 3, \ldots$ let $X^t = X$ and $Y^t = Y$. Define $X_n = \prod_{t=1}^{n} X^t$ and $Y_n = \prod_{t=1}^{n} Y^t$ for all $n = 1, 2, 3, \ldots$. The elements of $X_n$ are the input words of length $n$ and the elements of $Y_n$ the output words of length $n$.

Let $\omega(\cdot | \cdot)$ be a non-negative function defined on $X \times Y$ such that

\begin{equation}
\sum_{y \in Y} \omega(y | x) = 1 \text{ for all } x \in X.
\end{equation}

The number $\omega(y | x)$ is the probability that $y$ is received given that $x$ is sent over the channel. Then the transmission probabilities are defined by

\begin{equation}
P_n(y_n | x_n) = \prod_{t=1}^{n} \omega(y^t | x^t) \text{ for all } x_n = (x^1, \ldots, x^n) \in X_n,
\end{equation}

where $y_n = (y^1, \ldots, y^n) \in Y_n$ and $n = 1, 2, 3, \ldots$. 
Then the probability that the word \( y_n \) is received given that \( x_n \) is sent is given by \( P_n(y_n|x_n) \).

The channel is completely characterized by the input alphabet, output alphabet and transmission probabilities for all \( n = 1, 2, \ldots \), so we call the triple \( (X, Y, \{P_n(\cdot|\cdot)|n = 1, 2, 3, \ldots\}) \) a discrete memoryless channel.

If \( n \) and \( N \) are positive integers, a code \(- (n, N) \) for the DMC is a system

\[
(1.1.3) \quad ((u_i, A_i)|1 \leq i \leq N)
\]

such that

(i) \( u_i \in X_n \) for all \( i = 1, \ldots, N \)

(ii) \( A_i \subseteq Y_n \) for all \( i = 1, \ldots, N \)

(iii) \( A_i \cap A_j = \emptyset \) for all \( i \neq j \).

The \( u_i \)'s are called codewords, the \( A_i \)'s the decoding sets, \( n \) the word length, and \( N \) the code length of the code.

The code is used as follows. A codeword \( u_i \) is transmitted across the channel and the receiver receives a (chance) sequence \( y_n \). He checks to see in which decoding set \( y_n \) lies. If it is in \( A_j \), he decides \( u_j \) was sent. If \( y_n \) is in none of the \( A_i \)'s, the receiver may decide whatever he wishes. Of course the message is decoded correctly when \( y_n \in A_i \).

In order to decide how good a particular code is, we make the following definition. Let \( \lambda \) be a real number with \( 0 < \lambda < 1 \).
Then a code \((n,N)\) is also called a code \((n,N,\lambda)\) if it satisfies

\[
(1.1.5) \quad p_n(A_i|u_i) \geq 1 - \lambda \quad \text{for all } i = 1, \ldots, N.
\]

We call \(\lambda\) the maximal probability of error for the code. If \(u_i\) was sent, the receiver decides incorrectly only if the received sequence \(y_n\) does not lie in \(A_i\), which happens with probability less than \(\lambda\) by \((1.1.5)\), and this is true no matter which \(u_i\) was sent.

Then a fundamental problem of information theory can be formulated as follows. Let \(N(n,\lambda)\) denote the maximal code length of a code \((n,N,\lambda)\) for the DMC. The problem is then to determine \(N(n,\lambda)\) for all values of the parameters \(n\) and \(\lambda\). Of course it is too much to expect a closed form expression for \(N(n,\lambda)\).

However, Shannon conjectured in 1948 in his pioneering paper [11] that \(\lim_{n \to \infty} \frac{1}{n} \log N(n,\lambda)\) exists for all \(\lambda, 0 < \lambda < 1\), and that the limit is independent of \(\lambda\). In the same paper, he also conjectured a value for this limit, which he denoted by \(C\) and called the capacity of the channel.

In 1954, Feinstein [7] proved part of the conjecture, namely that \(\lim_{n \to \infty} \frac{1}{n} \log N(n,\lambda) \geq C\). An independent proof of this using the method of random codes was given by Shannon [12] in 1957. This result is now called the coding theorem for the DMC.

A partial result in the converse direction was given by Fano [6]...
in 1952, who showed \( \inf_{0<\lambda<1} \lim_{n \to \infty} \frac{1}{n} \log N(n,\lambda) \leq C \). Finally, in 1957 Wolfowitz [16] showed that \( \lim_{n \to \infty} \frac{1}{n} \log N(n,\lambda) \leq C \), independent of \( \lambda \). This latter result has come to be called the strong converse, while the former result of Fano is known as the weak converse.

Very good estimates on the error probability for given code lengths were given by Shannon, Gallager and Berlekamp ([8], [15]) in 1967.

§1.2 The Coding Theorem and Fano's Lemma

For later reference, we give here a statement of the coding theorem for the DMC with a brief sketch of Shannon's random coding argument and also a statement of Fano's Lemma. In the proofs of Theorems 3 and 4 (Chapter 3), we use both a random coding argument and a generalized Fano-type estimate.

Let \( A \) and \( B \) be finite sets, \( q(\cdot) \) a probability distribution on \( A \), and \( Q(\cdot|\cdot) \) a non-negative function on \( A \times B \) such that

\[ \sum_{b \in B} Q(b|a) = 1 \text{ for all } a \in A. \]

Then define the rate function \( R(\cdot,\cdot) \) by

\[ R(\cdot,\cdot) = \sum_{a \in A} \sum_{b \in B} q(a)Q(b|a) \log \frac{Q(b|a)}{\sum_{a \in A} q(a)Q(b|a)} \]

for all p.d.'s \( q(\cdot) \) on \( A \) and functions \( Q(\cdot|\cdot) \) as defined above.

Then let \( \Pi(\cdot) \) denote a p.d. on \( X \) and

\[ C = \max_{\Pi(\cdot)} R(\Pi,\omega(\cdot|\cdot)) \]
where the maximum is taken over all p.d.'s $\pi(\cdot)$ on $X$. This number $C$ is the capacity of the DMC, and we can now state the coding theorem.

**Theorem 1** (Coding theorem for the DMC). Let $\epsilon$ and $\lambda$ be real numbers with $\epsilon > 0$ and $0 < \lambda < 1$. Let $C$ be defined as in (1.2.2). Then for all $n$ sufficiently large, there is a code $- (n, N, \lambda)$ for the DMC such that $N > \exp\{nC - n\epsilon\}$.

Before sketching the proof, we make the following observation. Let $\lambda$ be a real number with $0 < \lambda < 1$. Then a code $- (n, N)$ is called a code $- (n, N, \lambda)$ with average probability of error $\lambda$ if

$$
(1.2.3) \quad \frac{1}{N} \sum_{i=1}^{N} P(A_i|u_i) \geq 1 - \lambda.
$$

It is easy to show that the existence of a code $- (n, 2N, \lambda)$ with average probability of error $\lambda$ implies the existence of a code $- (n, N, \lambda)$ (where $\lambda = 2\lambda$) with maximal probability of error $\lambda$. Since we are interested in asymptotic bounds on the code length, there is no essential loss in discussing codes with average error as opposed to those with maximal error (in the case of the DMC).

Now let $\mathcal{K}$ be the ensemble of all codes $- (n, N)$ for the DMC such that whenever $K = \{ (u_i, A_i) | i = 1, \ldots, N \} \in \mathcal{K}$, then

$$
(1.2.4) \quad A_i = \{ y_n \mid P_n(y_n | u_i) > P_n(y_n | u_j) \text{ for all } j \neq i \}
$$

for all $i = 1, \ldots, N$. 
The $A_1$'s are called maximum likelihood decoding sets. Also for each such $K \in \mathcal{K}$ define the average error probability $\overline{\lambda}(K)$ by

\[(1.2.5) \quad \overline{\lambda}(K) = \frac{1}{N} \sum_{i=1}^{N} P(A_i^c|u_i)\]

where $A_i^c$ denotes the complement of $A_i$.

If $\Pi(*)$ is a p.d. on $X$, let $\Pi_n(*)$ denote the $n$-th independent product distribution of $\Pi(*)$ on $X_n$. Choose code words $u_1, u_2, \ldots, u_N$ independently at random according to the p.d. $\Pi_n(*)$. Then let $A_1, A_2, \ldots, A_N$ be the maximum likelihood decoding sets determined by the codewords $u_1, u_2, \ldots, u_N$, respectively. Thus we have that $\{(u_i, A_i)[i=1, \ldots, N] \in \mathcal{K}\}$.

Let $\Pi(*)$ be a p.d. on $\mathcal{K}$ defined by $\Pi(K) = \Pi_n(u_i)$ for all $K \in \mathcal{K}$. Let $\lambda^*$ be a random variable which takes the value $\overline{\lambda}(K)$ whenever the code $K$ is chosen, and let $\lambda$ be a real number with $0 < \lambda < 1$.

Let $\varepsilon' > 0$. Then it is possible to show that, for all $n$ sufficiently large, $N < \exp[nC - n\varepsilon']$ implies

\[(1.2.6) \quad \mathbb{E}\lambda^* \leq \lambda.\]

If $\varepsilon'$ is chosen suitable small, then (1.2.6) implies that for all $n$ sufficiently large, there must actually exist a code $\mathcal{C}(n, N, \lambda)$ with average probability of error $\lambda$ satisfying $N > \exp[nC - n\varepsilon]$. By the remark following the statement of the theorem, we are done.
Now we state Fano’s Lemma (for the case of the DMC). Let
\( K = \{(u_i, A_i) | i = 1, \ldots, N\} \) be a code \( (n, N, \lambda) \) with maximal error
\( \lambda \) and define a p.d. \( p_0(\ast) \) on \( X_n \) by

\[
(1.2.7) \quad p_0(x_n) = \begin{cases} 
\frac{1}{N} & \text{if } x_n = u_i \text{ for some } i, \\
1 & \text{if } 1 \leq i \leq N, \\
0 & \text{otherwise}.
\end{cases}
\]

Lemma 1 (Fano). If \( \{(u_i, A_i) | 1 \leq i \leq N\} \) is a code \( (n, N, \lambda) \) and
\( p_0(\ast) \) is defined as in (1.2.7), then

\[
(1.2.8) \quad \log N \leq \frac{R(p_0, p_n(\ast | \ast)) + 1}{1 - \lambda}.
\]

We remark that the weak converse can be easily derived from
Fano’s Lemma.
CHAPTER 2

SEQUENTIAL CODING FOR THE DMC IN
THE CASE OF NOISELESS FEEDBACK

§2.1 The DMC with feedback and sequential coding

In order to describe a DMC with feedback (DMCF), the notion of an encoding function is needed.

Suppose the sender chooses the messages to be sent from a set of $N$ messages, which we denote for convenience by $M = \{1, \ldots, N\}$. For each $m \in M$, there is given an element $f_m^1 \in X$ and a sequence $f_m^2(\cdot), f_m^3(\cdot), f_m^t(\cdot), \ldots$ of functions where $f_m^t$ is defined on $Y_{t-1}$ (for $t > 1$) and takes values in $X$. Then for all $m \in M$ and $n = 1, 2, 3, \ldots$, a (vector-valued) encoding function $f_n(m, \cdot)$ is defined on $Y_{n-1}$ by

\begin{equation}
(2.1.1) \quad f_n(m, y_{n-1}) = [f_m^1, f_m^2(y^1), \ldots, f_m^t(y^1, \ldots, y^{t-1}), \ldots, \]
\begin{equation*}
\quad f_m^n(y^1, \ldots, y^{n-1})] \text{ for all } y_{n-1} = (y^1, \ldots, y^{n-1}) \in Y_{n-1}.
\end{equation*}

The sender makes use of the encoding function as follows. If he wants to send message $m$, the first letter he sends is $f_m^1$, which depends only on $m$. The receiver receives a letter $y^1$, say, which is a matter of chance. The sender knows that $y^1$ was received because...
of feedback, so the second letter he sends is $f_m^2(y^1)$. At any time $t > 1$, a sequence $y_1, y_2, \ldots, y_{t-1}$ of (chance) received letters is known to both sender and receiver. The sender's next transmitted letter is then $f_m^+(y_1, \ldots, y_{t-1})$.

Denote by $P(y_n | f_n(m))$ the probability that $y_n$ is received given that $m$ is sent and encoded by $f_n(m, \cdot)$.

A code - $(n, N)$ for a DMCF is a system
\begin{equation}
(f_n(m, \cdot), A_m) | m = 1, \ldots, N\end{equation}
such that
\begin{enumerate}
\item $(i)$ $f_n(m, \cdot)$ is an encoding function for all $m \in M$
\item $(ii)$ $A_m \subseteq Y_n$ for all $m \in M$
\item $(iii)$ $A_m \cap A_{m'} = \emptyset$ whenever $m \neq m'$.
\end{enumerate}

Let $\lambda$ be a real number with $0 < \lambda < 1$. Then a code - $(n, N)$ for a DMCF is also a code - $(n, N, \lambda)$ if it is true that
\begin{equation}
P(A_m | f_n(m)) \geq 1 - \lambda \text{ for all } m \in M.
\end{equation}

The following result, which is a strong converse for the DMCF, was proved by Kemperman [10] (see also [17], Chapter 4) and independently by Kesten [17]. It is needed in the proof of the weak converse for a DMCF in the case of sequential coding (§2.2).

**Theorem 2.** Any code - $(n, N, \lambda)$ for a DMCF satisfies
\[ N < \exp\{nC + K\sqrt{n}\} \]
where $C$ is the capacity of the corresponding DMC and $K$ is a constant depending on $\lambda$ but not on $n$. 
We now describe a sequential code for a DMCF. Let \( Y_\infty = \bigcup_{n=1}^{\infty} Y_n \) denote the set of all words of finite length with letters from \( Y \). Let \( s, t, \) and \( k \) be positive integers with \( k = s + t \). Then we say that a sequence \( y_t = (y^1, \ldots, y^t) \in Y_t \) is a prefix of a sequence \( y_k \in Y_k \) if there is a sequence \( y_s = (y^1, \ldots, y^s) \in Y_s \) such that \( y_k = y_t y_s = (y^1, \ldots, y^t y^s, \ldots, y^s) \). Also for each \( m \in M \), let \( \epsilon^*: Y_\infty \to \{0,1\} \) be a function such that \( \epsilon^*(y_t) = 0 \) implies \( \epsilon^*(y_k) = 0 \) whenever \( y_t \) is a prefix of \( y_k \).

For all \( m \in M \), a sequential encoding function \( f^*(m, \cdot) \) is defined on \( Y_\infty \) by

\[
(2.1.5) \quad f^*(m, y_{n-1}) = \left[f^1_m, f^2_m(y^1), \ldots, f^t_m(y^1, \ldots, y^{t-1}), \ldots, f^n_m(y^1, \ldots, y^{n-1}) \right] \quad \text{for all} \quad y_{n-1} = (y^1, \ldots, y^{n-1}) \in Y_\infty.
\]

The sender uses the sequential encoding function as follows. If he wants to send message \( m \), he transmits \( f^1_m \), which depends only on \( m \). A letter \( y^1 \), say, is received, which is known to the sender as well as the receiver because of feedback. Both sender and receiver calculate \( \epsilon^*(y^1) \). If it is 0, the transmission is ended. If it is 1, the sender continues by transmitting \( f^2_m(y^1) \). Assume that the sender has transmitted \( t - 1 \) letters \((t > 1)\), and that the (chance) received letters are \( y^1, \ldots, y^{t-1} \), known to both sender and receiver. They each calculate \( \epsilon^*(y^1, \ldots, y^{t-1}) \). If it is 0, the transmission is ended. If it is 1, the sender continues by
sending \( f_m^t(y^1, \ldots, y^{t-1}) \).

Denote by \( P(y_n | f^*(m)) \) the probability that \( y_n \) is received given that \( m \) is sent and encoded in the manner described. Henceforth we consider only sequential encoding rules such that  
\[
\sum_{k=1}^{\infty} kP(Y_k | f^*(m)) < \infty \quad \text{for all } m \in M.
\]

Then a sequential code \((L,N)\) for a DMCF is a system  
\[
((f^*(m,*), A_m) | m = 1, \ldots, N)
\]

such that  

(i) \( f^*(m,*) \) is a sequential encoding function for all \( m \in M \)

(ii) \( A_m \subseteq Y^\infty \) for all \( m \in M \)

(iii) \( A_m \cap A_{m'} = \emptyset \) whenever \( m \neq m' \)

(iv) No sequence in \( A_m \) is a prefix of a sequence in \( A_{m'} \), whenever \( m \neq m' \)

(v) \( L = \frac{1}{N} \sum_{m=1}^{N} \sum_{k=1}^{\infty} kP(Y_k | f^*(m)) \).

Let \( \lambda \) be a real number with \( 0 < \lambda < 1 \). Then a sequential code \((L,N)\) for a DMCF is also a sequential code \((L,N,\lambda)\) if  
\[
P(A_m | f^*(m)) > 1 - \lambda \quad \text{for all } m \in M.
\]

\section*{§2.2 The weak converse for a DMCF in the case of sequential coding}

Suppose \( ((f^*(m,*), A_m) | m = 1, \ldots, N) \) is a sequential code -
(L, N, λ) for a DMCF. Let ε > 0 and let s be the greatest integer smaller than L(1 + ε) and t the smallest integer larger than L(1 + ε). Define \( B = \bigcup_{m=1}^{N} A_{m} \) and \( B(\varepsilon) = \bigcup_{k=1}^{s} Y_{k} \). Then

\[
L = \frac{1}{N} \sum_{m=1}^{N} \sum_{k=1}^{\infty} kP(Y_{k} \mid f^{*}(m))
\]

\[
\geq \frac{1}{N} \sum_{m=1}^{N} \sum_{k=t}^{\infty} kP(Y_{k} \mid f^{*}(m))
\]

\[
\geq \frac{L(1 + \varepsilon)}{N} \sum_{m=1}^{N} P(B(\varepsilon)^{c} \mid f^{*}(m))
\]

where \( B(\varepsilon)^{c} \) denotes the complement of \( B(\varepsilon) \).

From (2.2.1) it follows that

\[
\frac{1}{N} \sum_{m=1}^{N} P(B(\varepsilon)^{c} \mid f^{*}(m)) \leq \frac{1}{1 + \varepsilon}.
\]

Also we have

\[
1 - \lambda \leq \frac{1}{N} \sum_{m=1}^{N} P(A_{m} \mid f^{*}(m))
\]

\[
= \frac{1}{N} \sum_{m=1}^{N} P(A_{m} \cap B(\varepsilon)^{c} \mid f^{*}(m))
\]

\[
+ \frac{1}{N} \sum_{m=1}^{N} P(A_{m} \cap B(\varepsilon) \mid f^{*}(m)).
\]

Therefore, for all \( \lambda \) suitably small,

\[
\frac{1}{N} \sum_{m=1}^{N} P(A_{m} \cap B(\varepsilon) \mid f^{*}(m)) \geq 1 - \lambda^{*} > 0
\]

where \( \lambda^{*} = \lambda + \frac{1}{1 + \varepsilon} \).

For each \( m \in M \), we modify \( f^{*}(m, \cdot) \) as follows. Suppose
\[ e^*(y_{n-1}) = 0 \] and that \[ e^*(y_k) = 1 \] for all \( k = 1, \ldots, n - 2 \), where \( y_k \) is the prefix of \( y_{n-1} \) of length \( k \). Then define an encoding function \( f'(m, \cdot) \) by

\[
(2.2.5) \quad f'(m, y_{n-1}) = \begin{cases} 
[f_m^1, f_m^2 (y^1), \ldots, f_m^s (y^1, \ldots, y^S)] & \text{if } n > L(1+\epsilon) \\
[f_m^1, f_m^2 (y^1), \ldots, f_m^n (y^1, \ldots, y_{n-1}), f_m^{n+1}, \ldots, f_m^s] & \text{if } n \leq L(1+\epsilon)
\end{cases}
\]

where \( f_m^{n+1}, \ldots, f_m^s \) are arbitrary elements of \( X \).

Similarly we modify each \( A_m \) to a decoding set \( A'_m \), which consists of all the sequences \( y_s \in Y_s \) which have the property that either \( y_s \in A_m \) or they can be written as \( y_s = y_k y_\ell \) for some \( y_k \in A_m \) and \( y_\ell \in Y_\ell \) where \( s = k + \ell \). Then if we let \( P(y_s | f'(m)) \) denote the probability that \( y_s \) is received given that \( m \) is sent and encoded by \( f'(m, \cdot) \), then (2.2.4) yields

\[
(2.2.6) \quad \frac{1}{N} \sum_{m=1}^{N} P(A'_m | f'(m)) \geq 1 - \lambda^*.
\]

Let \( \delta \) be a real number such that \( 0 < \lambda^* < \delta < 1 \), and let \( N^* \) be the greatest integer in \( (1 - \delta)N \). Then for at least \( N^* \) messages \( m \in M \),

\[
(2.2.7) \quad P(A'_m | f'(m)) \geq 1 - \frac{\lambda^*}{\delta}.
\]

For if not, then \( \frac{1}{N} \sum_{m=1}^{N} P(A'_m | f'(m)) < 1 - \lambda^* \), which contradicts (2.2.6).
Take those indices \( m = 1, \ldots, N \) which satisfy \((2.2.7)\) and form a block code \((s, N^*, \lambda^*)\) for the DMCF. The weak converse for a DMCF in the case of sequential coding then follows immediately from the strong converse for a DMCF (Theorem 2).

§2.3 A counterexample to the strong converse

Consider a binary symmetric channel with feedback (BSCF) with crossover probability \( p < 1/2 \). Let \( q = 1 - p \) and \( \epsilon \) and \( \lambda \) be real numbers satisfying \( \epsilon > 0 \) and \( 1 > 1 - \lambda > q \).

Let \( \{(u_i, A_i) | i = 1, \ldots, N\} \) be a block code \((n, N, \lambda)\) for the binary symmetric channel (BSC) with crossover probability \( p \); furthermore assume this code has rate \( C - \epsilon \), where
\[
C = 1 + p \log_2 p + q \log_2 q
\]
is the capacity of the BSC (and the BSCF, by Theorem 2).

Assume without loss of generality that \( \bigcup_{i=1}^{N} A_i = Y_n \). Let \( y_t \) be any sequence in \( Y_\infty \cup Y_0 \), where \( Y_0 \) contains only the empty sequence, and denote by \( [y_t] \) the collection of all sequences \( y_k \in Y_\infty \cup Y_0 \) such that either \( y_t = y_k \) or \( y_t \) is a prefix of \( y_k \).

If \( A_i^* = A_i = [0] \) for all \( i = 1, \ldots, N \), then \( \{(u_i, A_i^*) | i = 1, \ldots, N\} \) is a block code \((n, N, \lambda^*)\) for the BSC, where \( \lambda^* < \lambda + q < 1 \).

Now we construct a sequential code for the BSCF as follows. Define a function \( \epsilon^*(\cdot) \) on \( Y_\infty \) for each \( m = 1, \ldots, N \) by
(2.3.1) \[ \epsilon^*(y_{k-1}) = \begin{cases} 1 & \text{if } 1 \leq k - 1 < n \text{ and } y^1 \neq 0 \\ 0 & \text{otherwise} \end{cases} \]

for all \( y_{k-1} = (y^1, \ldots, y^{k-1}) \in Y_\infty \). Then the sequential encoding functions are given by

(2.3.2) \[ f^*(m, y_{k-1}) = \left[ f_m^1, f_m^2(y^1), \ldots, f_m^k(y^1, \ldots, y^{k-1}); \epsilon^*(y_{k-1}) \right] \]

for all \( m = 1, \ldots, N \), all \( y_{k-1} = (y^1, \ldots, y^{k-1}) \in Y_\infty \), and where

\( f_m^1 \) is the first component of \( u_m \) and \( f_m^t(y^1, \ldots, y^{t-1}) \) is the \( t \) - th component of \( u_m \) for all \( t = 2, \ldots, n \).

The sequential decoding sets are defined by

(2.3.3) \[ D_1 = \bigcup_{y_n \in A_1^*} [y_n] \cup [0] \]

\[ D_m = \bigcup_{y_n \in A_m^*} [y_n] \text{ for all } m = 2, \ldots, N. \]

Then \( \{(f^*(m, \cdot), D_m) | m = 1, \ldots, N\} \) is a sequential code \( (L, N, \lambda^*) \) for the BSCF where \( \lambda^* < 1 \) and \( L < q \cdot 1 + q \cdot n \). Hence the rate of the code satisfies

(2.3.4) \[ \frac{1}{L} \log N > \frac{n(C - \epsilon)}{q(n+1)} \]

\[ > C^* - \epsilon^* \]

where \( C^* = \frac{C}{q} > C \) and \( \epsilon^* \to 0 \) as \( \epsilon \to 0 \) and \( n \to \infty \). Hence the strong converse for a DMCF in the case of sequential coding cannot hold.
§3.1 The Channel Model and Statement of the Coding Problem

In this chapter we study a noisy memoryless channel with $s$ senders and $r$ receivers. We give now a mathematical description of this channel.

Let $X(1), X(2), \ldots, X(s)$ and $Y(1), Y(2), \ldots, Y(r)$ be finite sets; $X(1), \ldots, X(s)$ denote the input alphabets and $Y(1), \ldots, Y(r)$ the output alphabets of the channel to be described. For every $t = 1, 2, \ldots$, let $X_t^t = X(k)$ and $Y_j^t = Y(j)$ for all $k = 1, \ldots, s$ and $j = 1, \ldots, r$. Let $n$ be a positive integer and define

\begin{align*}
(3.1.1) \quad & X_k^n = \prod_{t=1}^{n} X_k^t \quad \text{and} \quad Y_j^n = \prod_{t=1}^{n} Y_j^t \quad \text{for all } k = 1, \ldots, s \quad \text{and} \quad j = 1, \ldots, r.
\end{align*}

For each $k = 1, \ldots, s$, $X_k^n$ is the set of words of length $n$ with letters from alphabet $X(k)$ which can be sent over the channel; and similarly, for each $j = 1, \ldots, r$, $Y_j^n$ is the set of words of length $n$ with letters from alphabet $Y(j)$ which can be received over the channel. Further define

\begin{align*}
(3.1.2) \quad & X_t^n = \prod_{k=1}^{s} X_k^t \quad \text{and} \quad Y_t^n = \prod_{j=1}^{r} Y_j^t \quad \text{for all } t = 1, \ldots, n.
\end{align*}
If $M$ is an $n \times s$ matrix, let $M_k^t$ be the element in the $t$-th row and $k$-th column of $M$, $M^t_k$ be the $t$-th row of $M$, and $M_t^k$ be the $k$-th column of $M$ for all $t = 1, \ldots, n$ and $k = 1, \ldots, s$. Similarly, define $\bar{M}_j^t$, $\bar{M}^t_j$, and $\bar{M}_j^t$ for an $n \times r$ matrix $\bar{M}$ for all $t = 1, \ldots, n$ and $j = 1, \ldots, r$. Then let

$$
\mathcal{M} = \{M|M\text{ is an } n \times s \text{ matrix where } M_k^t \in X_k^t \text{ for all } t = 1, \ldots, n \text{ and } k = 1, \ldots, s\}
$$

(3.1.3)

$$
\widehat{\mathcal{M}} = \{\bar{M}|\bar{M}\text{ is an } n \times r \text{ matrix where } \bar{M}_j^t \in Y_j^t \text{ for all } t = 1, \ldots, n \text{ and } j = 1, \ldots, r\}.
$$

The columns $M_k^t$ ($k = 1, \ldots, s$) of an $M \in \mathcal{M}$ represent words of length $n$ with letters from the input alphabet $X(k)$, while the rows $M^t_k$ represent $m$-tuples of letters from $X^t$ which are sent across the channel at instant $t$ ($t = 1, \ldots, n$). Similarly, a column $\bar{M}_j^t$ ($j = 1, \ldots, r$) of an $\bar{M} \in \widehat{\mathcal{M}}$ represents a word of length $n$ with letters from the output alphabet $Y(j)$, while the row $\bar{M}_j^t$ represents an $r$-tuple of letters received at time $t$ ($t = 1, \ldots, n$).

Let $\hat{X} = \prod_{k=1}^{s} X(k)$, $\hat{Y} = \prod_{j=1}^{r} Y(j)$, and $\omega(\cdot | \cdot)$ be a non-negative function defined on $\hat{X} \times \hat{Y}$ such that $\sum_{y \in Y} \omega(y | x) = 1$ for all $x \in \hat{X}$. Then the channel transmission probabilities are given by
The probability that the $r$ words $\overline{M}_1, \ldots, \overline{M}_r$ of length $n$ are received, given that the $s$ words $M_1, \ldots, M_s$ of length $n$ are sent is given by $P(\overline{M}|M)$. The channel with $s$ senders and $r$ receivers is then completely described by the input alphabets $X(1), \ldots, X(s)$, the output alphabets $Y(1), \ldots, Y(r)$, and the set of all transmission probabilities $P(\overline{M}|M)$ as $n$ varies over the set of positive integers.

Now a description of how this channel is actually used is needed. Throughout this paper, we assume that all of the $s$ senders send independent messages simultaneously to all of the $r$ receivers. This communication situation is denoted $(P, T_{sr})$.

We introduce now a code concept appropriate for the communication situation $(P, T_{sr})$. Let $N_1, \ldots, N_s$ be positive integers and

$$\overline{N} = (N_1, N_2, \ldots, N_s)$$

(3.1.5) \[ N = \prod_{k=1}^{s} N_k \]

$$\overline{I} = \{(i_1, i_2, \ldots, i_s) | i_k \text{ a pos. int., } 1 \leq i_k \leq N_k, k = 1, \ldots, s\}.$$

A code - $(n, \overline{N})$ for $(P, T_{sr})$ is a system
(3.1.6) \{(M(\vec{I}), A_j(\vec{I})) | \vec{I} \in \mathcal{I}, \ j = 1, \ldots, r\}

such that

(i) \(M(\vec{I}) \in \mathcal{M}\) for all \(\vec{I} \in \mathcal{I}\)

(ii) For all \(k = 1, \ldots, s\) there exist \(u_k(1) \in X_k, u_k(2) \in X_k, \ldots, u_k(N_k) \in X_k\) such that \(M_k(\vec{I}) = u_k(1)\)

for all \(\vec{I} = (i_1, \ldots, i_s) \in \mathcal{I}\) and \(k = 1, \ldots, s\)

(3.1.7)

(iii) \(A_j(\vec{I}) \subset Y_j\) for all \(\vec{I} \in \mathcal{I}, \ j = 1, \ldots, r\)

(iv) \(A_j(\vec{I}) \cap A_j(\vec{I}') = \emptyset\) whenever \(\vec{I} \neq \vec{I}'\), for all \(j = 1, \ldots, r\).

For each \(j = 1, \ldots, r\), define an auxiliary transmission probability \(P_j(\cdot|\cdot)\) by

(3.1.8) \[P_j(A|M) = \sum_{\overline{M} \in \mathcal{M}(A)} P(\overline{M}|M)\] for all \(M \in \mathcal{M}\) and \(A \subset Y_j\),

where \(\mathcal{M}(A) = (\overline{M} | \overline{M} \in \mathcal{M} \text{ and } \overline{M}_j \in A)\). Then if \(\lambda\) is a real number with \(0 < \lambda < 1\), a code \((n, \vec{N})\) is also a code \((n, \vec{N}, \lambda)\) if it be required that

(3.1.9) \[\frac{1}{N} \sum_{\vec{I} \in \mathcal{I}} \sum_{j=1}^{r} P_j(A_j(\vec{I})^c|M(\vec{I})) \leq \lambda\]
where $A_j(\bar{i})^C$ denotes the complement of $A_j(\bar{i})$.

An $s$-tuple $(R_1, \ldots, R_s)$ of real numbers is called an $s$-tuple of achievable rates for $(P, T_{sr})$ if for all $\epsilon > 0$ and $0 < \lambda < 1$, and for all $n$ sufficiently large, there is a code $(n, \hat{N}, \lambda)$ for $(P, T_{sr})$ such that

$$\frac{1}{n} \sum \log N_k \geq R_k - \epsilon \quad \text{for all} \quad k = 1, \ldots, s.$$ (3.1.10)

The set of all $s$-tuples of achievable rates is denoted by $G(P, T_{sr})$. Following the terminology of [14], we call $G(P, T_{sr})$ the capacity region.

The problem then is to find a simple ("single letter") characterization for $G(P, T_{sr})$. In §3.3 we obtain such a characterization for $G(P, T_{sl})$ and in §3.4 generalize this to $G(P, T_{sr})$. The characterization obtained is such that it can be used for determining $G(P, T_{sr})$ on a computer to within any desired accuracy.
§3.2 A General Fano-Type Estimate

In this section, a Fano-type lemma ([6], [8], [17]) is proved, which in §3.3 enables us to obtain an outer bound on the capacity region $G(P, T_{sl})$. We assume that $r = 1$ throughout §3.2 and §3.3.

Now we define the so-called "rate functions", which are useful in the proof of the lemma. Let $A$ and $B$ be finite sets, $q(\cdot)$ a p.d. on $A$ and $Q(\cdot|\cdot)$ a non-negative function on $A \times B$ such that $\sum_{b \in B} Q(b|a) = 1$ for all $a \in A$. Then define

$$R(q(\cdot), Q(\cdot|\cdot)) = \sum_{b \in B} \sum_{a \in A} q(a)Q(b|a) \log \frac{Q(b|a)}{\sum_{a \in A} q(a)Q(b|a)}$$  \hspace{1cm} (3.2.1)

Another collection of rate functions is defined as follows. For all $t = 1, \ldots, n$ and $k = 1, \ldots, s$, let $p^t_k(\cdot)$ be a p.d. on $X^t_k$. Then define p.d.'s $p^t(\cdot)$ and $p_k(\cdot)$ on $X^t$ and $X_k$, respectively, by

$$p^t(x^t) = \prod_{k=1}^s p^t_k(x^t_k) \text{ for all } t = 1, \ldots, n, \text{ and}$$

$$p_k(x_k) = \prod_{t=1}^n p^t_k(x^t_k) \text{ for all } k = 1, \ldots, s,$$  \hspace{1cm} (3.2.2)

for all $x^t = (x^t_1, \ldots, x^t_s) \in X^t$ and $x_k = (x^1_k, \ldots, x^n_k) \in X_k$. Then for all $D \subset \{1, \ldots, s\}$ with $D \neq \emptyset$, define
Note that for each code \( - (n, \bar{N}) \) for \((P, T_{sl})\) we are given a collection \( \{u_k(i_k) \mid 1 \leq i_k \leq N_k, k = 1, \ldots, s\} \) where \( u_k(i_k) \in X_k \) for all \( 1 \leq i_k \leq N_k \) and \( k = 1, \ldots, s \). Let \( u_k = (u_k(1), \ldots, u_k(N_k)) \) for each \( k = 1, \ldots, s \) and define

\[
(3.2.4) \quad \mathcal{U} = \{u \mid u = (u_1, \ldots, u_s) \text{ for some code } - (n, \bar{N}) \text{ for } (P, T_{sl})\}.
\]

Let \( u_k^t(i_k) \) denote the \( t \)-th component of \( u_k(i_k) \) and define a p.d. \( p_k^t(\cdot) \) on \( X_k^t \) by

\[
(3.2.5) \quad p_k^t(x_k^t) = \frac{|\{i_k \mid u_k^t(i_k) = x_k^t, i_k \in \{1, \ldots, N_k\}\}|}{N_k} \quad \text{for all } t = 1, \ldots, n \quad k = 1, \ldots, s \quad \text{and } x_k^t \in X_k^t.
\]

The following is a generalized Fano-type lemma. It was first stated and proved in [3] for the case \( s = 3, \ r = 1 \).

**Lemma 2.** Given a code \(- (n, \bar{N}, \lambda) \) for \((P, T_{sl})\); denote it\((\{M(\overline{i}), A(\overline{i})\} \mid \overline{i} \in \overline{I}) \) where \( A(\overline{i}) \subset Y(1) \) for all \( \overline{i} \in \overline{I} \). Let \( p_k^t(\cdot) \) be as defined in (3.2.5) and \( p^t(\cdot) \) as in (3.2.2). Let \( D \subseteq \{1, \ldots, s\}, \) \( D \neq \emptyset \). Then for all \( \varepsilon > 0 \) there is a number \( k_D(\lambda, \varepsilon, n) \) such that...
Proof. Our argument is an extension of the one used in [3]. For ease of notation it is assumed that $D = \{1, \ldots, d\}$ for some $d$, $1 \leq d \leq s$. The extension to arbitrary $D$ will be clear.

Consider the probability space $(\Omega, \overline{\mu})$ where

$$\Omega = \{\overline{u} | \overline{u} = (u_{d+1}(i_{d+1}), \ldots, u_s(i_s)) \text{ for some } i_k, 1 \leq i_k \leq n_k, \text{ for all } k = d + 1, \ldots, s\}$$

and $\overline{\mu}(\cdot)$ is the equidistribution on $\Omega$.

Fix a $\overline{\mu} = (u_{d+1}(j_{d+1}), \ldots, u_s(j_s)) \in \Omega$ and define a non-stationary discrete memoryless channel (depending on $\overline{\mu}$) as follows. The input alphabet is $\mathcal{X} = \prod_{t=1}^{d} X(k)$ and the output is $Y = Y(1)$. For each $t = 1, \ldots, n$, a non-negative function $\omega_{\overline{\mu}}^t(\cdot|\cdot)$ is defined on $\mathcal{X} \times \mathcal{Y}$ by

$$\omega_{\overline{\mu}}^t(y|x) = \omega(y|x(1), \ldots, x(d), u_{d+1}^t(j_{d+1}), \ldots, u_s^t(j_s)))$$

for all $x = (x(1), \ldots, x(d)) \in \mathcal{X}$ and $y \in \mathcal{Y}$. If we define

$$\tilde{M} = \{\tilde{M} | \tilde{M} \text{ is an } n \times d \text{ matrix where } \tilde{M}_k^t \in X_k^t \text{ for all } t = 1, \ldots, n \text{ and } k = 1, \ldots, d\}$$
then the transmission probabilities are given by

\[(3.2.10)\quad P(y_n | \bar{M}) = \prod_{t=1}^{n} \mu_t(y_t | \bar{M}_t)\]

for all \(y_n = (y^1, \ldots, y^n) \in \prod_{1}^{n} Y = Y_n\) and \(\bar{M} \in \bar{M}\).

Let \(\hat{N} = \prod_{k=1}^{T} N_k, \tilde{N} = \prod_{k=d+1}^{S} N_k,\) and \(\bar{I} = (\bar{i} | \bar{I}) = (i_1, \ldots, i_s)\) where \(i_k = j_k\) for \(k = d+1, \ldots, s\). Define a p.d. \(\hat{\mu}(\cdot)\) on \(\bar{M}\) by

\[(3.2.11)\quad \hat{\mu}(\bar{M}) = \begin{cases} \frac{1}{\hat{N}} & \text{if for each } k, 1 \leq k \leq s, \bar{M}_k = u_k(i_k) \text{ for some } i_k, 1 \leq i_k \leq N_k \\ 0 & \text{otherwise.} \end{cases}\]

Also, for all \(t = 1, \ldots, n\), we define a p.d. \(\tilde{\mu}(\cdot)\) on \(\bar{X}\) by

\[(3.2.12)\quad \tilde{\mu}(\bar{X}) = \sum_{\bar{M} | \bar{X}} \hat{\mu}(\bar{M}) \quad \text{for all } \bar{X} \in \bar{X}; \text{ then let } L \text{ be a random variable defined on } \Omega \text{ with p.d. }\]

\[(3.2.13)\quad \tilde{\mu}(L = \frac{1}{\hat{N}} \sum_{\bar{I} \in \bar{I}} P(A(\bar{I})^c | M(\bar{I})) = \tilde{\mu}(\bar{u}) = \frac{1}{\hat{N}}.\]

Then by (3.1.9) we have that

\[(3.2.14)\quad E L \leq \lambda\]

where the expectation is taken with respect to \(\tilde{\mu}(\cdot)\). By Chebyshev's inequality, the set \(B^* = \{ L < \lambda + \epsilon \}\) satisfies \(|B^*| \geq \frac{\epsilon N}{\lambda^*}\) where \(\lambda^* = \lambda + \epsilon\). Hence for every \(\bar{u} = (u_{d+1}(j_{d+1}), \ldots, u_{s}(j_s)) \in B^*,\)
is a code - (n, N, \lambda*) for the corresponding non-stationary channel. Therefore Fano's Lemma yields

\[ (3.2.16) \log N \leq \sum_{t=1}^{n} \frac{R(\mu^t, \omega^t(\cdot | \cdot)) + 1}{1 - \lambda*} \] for every \( \bar{u} \in B^* \).

Averaging (3.2.16) over all \( \bar{u} \in B^* \) gives

\[ (3.2.17) \log N \leq (1 - \lambda*)^{-1} \left[ \sum_{t=1}^{n} \frac{1}{N} \sum_{\bar{u} \in \Omega} \frac{1}{\bar{u}} R(\mu^t, \omega^t(\cdot | \cdot)) + 1 \right] + E_n(\lambda, \varepsilon) \]

where \( E_n(\lambda, \varepsilon) = (1 - \lambda*)^{-1} \left( \frac{\lambda}{\lambda*} \right) n B \) and \( B \) is an upper bound (uniform in \( t \)) on \( R(\mu^t, \omega^t(\cdot | \cdot)) \). Definitions (3.2.1), (3.2.3) and independence yield

\[ (3.2.18) \frac{1}{N} \sum_{\bar{u} \in \Omega} \frac{1}{\bar{u}} R(\mu^t, \omega^t(\cdot | \cdot)) = R_D(p^t). \]

Therefore

\[ (3.2.19) \log N \leq \sum_{t=1}^{n} R_D(p^t) + k_D(\lambda, \varepsilon, n) \]

where \( \frac{1}{n} k_D(\lambda, \varepsilon, n) \to 0 \) as \( n \to \infty \), \( \lambda \to 0 \), \( \varepsilon \to 0 \).
§3.3 Capacity Region of a Channel with \( s \) Senders and One Receiver

Order the \( \mathcal{L} = 2^s - 1 \) non-empty subsets of \( \{1, \ldots, s\} \) and call them \( D(1), \ldots, D(\mathcal{L}) \). Define

\[
(3.3.1) \quad F(Y) = \{ R_{D(1)}(q), \ldots, R_{D(\mathcal{L})}(q) \mid q(\cdot) = q_1(\cdot) \times \ldots \times q_s(\cdot) \]
for some \( q_1(\cdot), \ldots, q_s(\cdot) \) where \( q_k(\cdot) \) is a

\[\text{p.d. on } X(k) \text{ for } k = 1, \ldots, s\}.

Let \( F^*(Y) \) denote the convex hull of \( F(Y) \), and \( \tilde{\mathbf{R}} = (R_{1*}, \ldots, R_{\mathcal{L}*}) \) denote an arbitrary element of \( F^*(Y) \). Then let

\[
(3.3.2) \quad G(\tilde{\mathbf{R}}, Y) = \{ (R_1, \ldots, R_s) \mid \sum_{k \in D(j)} R_k \leq R_j* \text{ all } j = 1, \ldots, \mathcal{L} \}
\]

and

\[
(3.3.3) \quad G(Y) = \bigcup_{\tilde{\mathbf{R}} \in F^*(Y)} G(\tilde{\mathbf{R}}, Y).
\]

The set \( G(Y) \) has the following properties.

Lemma 3. \( G(Y) \) is convex, closed under projections, and compact in the usual topology of Euclidean \( s \)-space.

Proof. The facts that \( G(Y) \) is convex, closed under projections and bounded are immediate from the definition of \( G(Y) \). It only remains to show it is closed.

A p.d. \( q(\cdot) \) on a finite set \( A \) with \( |A| = a \) can be viewed as a "probability vector" \( q = (q_1, \ldots, q_a) \) where \( q_k \), for all \( k, 1 \leq k \leq a \), is the probability attached to the \( k \)-th element of \( A \) in some ordering. Thus \( q_k \geq 0 \) for all \( k, 1 \leq k \leq a \),
and $\sum_{k=1}^{a} q_k = 1$. Viewed in this sense, the set of all product
p.d.'s on $\hat{X}$ becomes a compact subset of Euclidean $|\hat{X}|$-space. Then
by the continuity of the rate functions, $F(Y)$ is a compact subset
of Euclidean $\ell$-space. Since the convex hull of a compact set in a
Euclidean space is also compact, $F^*(Y)$ is compact.

Let $\bar{R}(1), \bar{R}(2), \bar{R}(3), \ldots$ be a sequence of elements of $G(Y)$
where $\lim_{n \to \infty} \bar{R}(n)$ exists and equals $\bar{R}$, say. We will be done if we
show that $\bar{R} \in G(Y)$.

For all $n = 1, 2, 3, \ldots$ there exist $\tilde{R}(n) \in F^*(Y)$ such that
$\bar{R}(n) \in G(\tilde{R}(n), Y)$. Let $\bar{R} = (R_1, \ldots, R_\ell)$,
$\bar{R}(n) = (R_1(n), \ldots, R_\ell(n))$ and $\tilde{R}(n) = (R^*_1(n), \ldots, R^*_\ell(n))$ for all
$n = 1, 2, 3, \ldots$. By the boundedness of $F^*(Y)$ there is a $B > 0$
such that $R^*_j(n) \leq B$ for all $n = 1, 2, 3, \ldots$, and $j = 1, \ldots, \ell$.

Let $\epsilon > 0$. Then there is a positive integer $n(\epsilon)$ such
that $n \geq n(\epsilon)$ implies $\sum_{i \in D(j)} R_i - \epsilon \leq R^*_j(n) \leq B$ for all
$j = 1, \ldots, \ell$. Hence there is a subsequence $[n_t]_{t=1}^{\infty}$ of $[n]_{n=1}^{\infty}$ suc
that for all $j=1, \ldots, \ell$, $\lim_{t \to \infty} R^*_j(n_t)$ exists and equals $R^*_j$, say.
Since $F^*(Y)$ is closed, $\bar{R} = (R^*_1, \ldots, R^*_\ell) \in F^*(Y)$.

Furthermore, $\sum_{i \in D(j)} R_i - \epsilon \leq R^*_j$ for all $j = 1, \ldots, \ell$. Since
$\epsilon$ was arbitrary, $\sum_{i \in D(j)} R_i \leq R^*_j$ for all $j=1, \ldots, \ell$. Hence
$\bar{R} \in G(\bar{R}, Y)$, which implies $\bar{R} \in G(Y)$. 
Theorem 3. The capacity region $G(P, T_{sl}) = G(Y)$.

Proof. First we show $G(P, T_{sl}) \subseteq G(Y)$: Let $(R_1, \ldots, R_s) \in G(P, T_{sl})$. Let $\epsilon > 0$ and $0 < \lambda < 1$. Then there is a code - $(n, N, \lambda)$ such that $\frac{1}{n} \log N \geq R_k - \epsilon$ for all $k = 1, \ldots, s$. Using this fact and Lemma 2, it can be concluded that, for any $\delta > 0$ and for all $n$ sufficiently large and all $\epsilon$ and $\lambda$ sufficiently small,

$$\sum_{k \in D} R_k \leq \frac{1}{n} \sum_{t=1}^{n} R_D(p_t) + \delta \quad \text{for all } D \subseteq \{1, \ldots, s\}.$$  

Since

$$\frac{1}{n} \sum_{t=1}^{n} (R_D(1)(p_t), \ldots, R_D(s)(p_t)) \in F^*(Y), \quad (R_1-\delta, \ldots, R_s-\delta) \in G(Y)$$

for all $\delta > 0$. Thus $(R_1, \ldots, R_s)$ belongs to the closure of $G(Y)$, and hence to $G(Y)$, since it is closed.

Now we show $G(Y) \subseteq G(P, T_{sl})$: Let $(R_1, \ldots, R_s) \in G(Y)$. Then there is an $\ell$-tuple $\tilde{R} = (R_1^*, \ldots, R_s^*) \in F^*(Y)$ such that

$$\sum_{k \in D(j)} R_k \leq R_j^* \quad \text{for all } j = 1, \ldots, \ell.$$  

Define the set

$$\overline{F}(Y) = \{\mathbf{R} | \overline{R} = \frac{1}{n} \sum_{t=1}^{n} (R_D(1)(p_t), \ldots, R_D(s)(p_t)) \text{ where}$$

$$p_t = p_{1,t} \times \cdots \times p_{s,t} \text{ for some p.d.'s } p_k^{t}(\cdot) \text{ on } X_k^t$$

for all $k=1, \ldots, s, t=1, \ldots, n$ and $n = 1, 2, \ldots\}$

Let $\delta > 0$. Then it is possible to find a $\tilde{R}' = (R_1', \ldots, R_s') \in \overline{F}(Y)$ such that $|R_j^* - R_j'| < \delta$ for all $j = 1, \ldots, \ell$. Let $n$ be a
positive integer and \( p_k^t(\cdot) \) \((1 \leq k \leq s, 1 \leq t \leq n)\) a p.d. on 
\( X_k^t \) such that

\[
R_j^t = \frac{1}{n} \sum_{t=1}^{n} R_D(j)(p^t) \quad \text{for all } j = 1, \ldots, \ell.
\]

Let \( \epsilon > 0 \) and \( 0 < \lambda < 1 \). Define

\[
N_k = e^{n(R_k^t - \epsilon)} \quad \text{for } k = 1, \ldots, s.
\]

Then we will show the existence of a code \( - (n, N, \lambda) \) by using 
Shannon's random coding method.

For each \( \mathbf{i} \in \mathcal{I} \), a codeword \( M(\mathbf{i}) \) is chosen at random as follows. For all \( t = 1, \ldots, n \) and \( k = 1, \ldots, s \) let 
\( U_k^t(i_k) \) be a random variable taking values in \( X(k) \) with p.d. \( p_k^t(\cdot) \).

For all \( i_k = 1, \ldots, N_k \) and \( k = 1, \ldots, s \) let 
\( U_k(i_k) \) be a random vector taking values in \( X_k \) with p.d. \( p_k(\cdot) \). Finally let \( U \) be a random vector taking values in \( \mathcal{U} \) with p.d. \( p^*(\cdot) \), where

\[
p^*(u) = \prod_{k=1}^{s} \prod_{i_k=1}^{N_k} p_k(u_k(i_k)) \quad \text{for all } u \in \mathcal{U}.
\]

Whenever \( U = u \in \mathcal{U}, M(\mathbf{i}) \) is the matrix in \( \mathcal{M} \) such that 
\( M_k(i) = u_k(i_k) \) for \( k = 1, \ldots, s \). Once the codewords 
\( M(\mathbf{i}) \) \((\mathbf{i} \in \mathcal{I})\) have been chosen, define maximum likelihood decoding 
sets (depending on the \( M(\mathbf{i})'s \)) by

\[
A(\mathbf{i}) = \{ y_n | y_n \in Y_n = \prod_{t=1}^{n} X_t \quad \text{and} \quad P(y_n|M(\mathbf{i})) > P(y_n|M(\mathbf{j})) \}
\]

for all \( \mathbf{j} \neq \mathbf{i} \).
The average error for the code $\{(M(\overline{i}), A(\overline{i})) | \overline{i} \in \overline{I}\}$ is

\[(3.3.11) \lambda(u) = \frac{1}{N} \sum_{\overline{i} \in \overline{I}} P(A(\overline{i})^c | M(\overline{i})) . \]

The proof will be complete if an upper bound tending to zero as $n \to \infty$ can be found for

\[(3.3.12) \mathcal{E}\lambda(u) = \sum_{u \in \mathcal{U}} p^*(u) \lambda(u) . \]

Now $\sum_{u \in \mathcal{U}} p^*(u) P(A(\overline{i})^c | M(\overline{i})) = \sum_{u \in \mathcal{U}} p^*(u) P(A(\overline{j})^c | M(\overline{j}))$ for all $\overline{i}, \overline{j} \in \overline{I}$, so if $\lambda_{\perp}(u) = P(A(\overline{i})^c | M(\overline{i}))$, where $\overline{i} = (1,1,\ldots,1)$, then

\[(3.3.13) \mathcal{E}\lambda(u) = E \lambda_{\perp}(u) . \]

Let $M^*(\overline{i})$ be a random matrix with $M_k^*(\overline{i}) = U_k(i_k)$. Then

\[(3.3.14) \mathcal{E}\lambda_{\perp}(u) = \sum_{y_n \in Y_n} \sum_{M(\overline{i}) \in \mathcal{M}} p(M(\overline{i})) p(y_n | M(\overline{i}))
\times p^*\{p(y_n | M(\overline{i})) \leq p(y_n | M^*(\overline{i})) \text{ for some } \overline{i} \neq \overline{l}\} .
\]

Thus

\[(3.3.15) \mathcal{E}\lambda_{\perp}(u) \leq \sum_{i \neq l} \sum_{y_n \in Y_n} \sum_{M(\overline{i}) \in \mathcal{M}} p(M(\overline{i})) p(y_n | M(\overline{i}))
\times p^*\{p(y_n | M(\overline{i})) \leq p(y_n | M^*(\overline{i},\overline{l}))\} .
\]

where

\[(3.3.16) M_k^*(\overline{i},\overline{l}) = \begin{cases} U_k(i_k) & \text{if } i_k \neq l \\ U_k(l) & \text{if } i_k = l \end{cases} .
\]
The object then is to bound from above the term

\[(3.3.17) \sum_{y_n \in Y_n} \sum_{M(\mathcal{I}) \in \mathcal{M}} p(M(\mathcal{I})) \cdot p(y_n | M(\mathcal{I})) \times p^* \{p(y_n | M(\mathcal{I})) \leq p(y_n | M^*(\mathcal{I}, \mathcal{I})) \} \]

for each $\overline{\mathcal{I}} \neq \mathcal{I}$.

In order to carry out this objective, we make the following definitions. First let

\[(3.3.18) \mathcal{M}(M, D) = \{M' | M' \in \mathcal{M} \text{ and } M'_k = M_k \text{ for all } k \notin D \}

for all $M \in \mathcal{M}$ and $D \subseteq \{1, \ldots, s\}$. Then define the "information function" $I_D^*(\cdot, \cdot)$ by

\[(3.3.19) I_D^*(M, y_n) = \log \frac{p(y_n | M)}{\sum_{M' \in \mathcal{M}(M, D)} \prod_{k \in D} p_k(M'_k) p(y_n | M')}

for all $M \in \mathcal{M}$ and $y_n \in Y_n$. Now let $d$ be an integer with

$1 \leq d \leq s$ and

\[
\overline{\mathcal{U}} = \{\overline{u} | \overline{u} = (u_1, \ldots, u_d) \text{ for some } (u_1, \ldots, u_s) = u \in \mathcal{U} \}
\]

\[(3.3.20) \overline{\mathcal{U}} = \{\overline{u} | \overline{u} = (u_{d+1}, \ldots, u_s) \text{ for some } (u_1, \ldots, u_s) = u \in \mathcal{U} \}.

Also define p.d.'s $\overline{\rho}(\cdot)$ and $\overline{\mu}(\cdot)$ on $\overline{\mathcal{U}}$ and $\overline{\mathcal{U}}$, respectively, by

\[
\overline{\rho}(\overline{u}) = \prod_{k=1}^{d} \prod_{i_k=1}^{N_k} p_k(u_k(i_k)) \text{ for all } \overline{u} \in \overline{\mathcal{U}}
\]

\[(3.3.21) \overline{\mu}(\overline{u}) = \prod_{k=d+1}^{s} \prod_{i_k=1}^{N_k} p_k(u_k(i_k)) \text{ for all } \overline{u} \in \overline{\mathcal{U}}.
\]
Finally let

\[(3.3.22) \quad B(\tilde{u}) = \{ (y_n, \tilde{u}) | \tilde{u} \in \tilde{U}, y_n \in Y_n \text{ and } I_D(M(\tilde{I}), y_n) \geq \log \alpha \tilde{N} \}\]

for all \( \tilde{u} \in \tilde{U} \), where \( D = \{1, \ldots, d\} \) and \( \alpha \) is some positive real number.

Recall that we want to bound \( (3.3.17) \) from above for all \( \tilde{I} \neq \tilde{I} \).

For ease of notation, we restrict our attention to those \( \tilde{I}'s \) with
the following property: if \( \tilde{I} = (i_1, \ldots, i_s) \), there is an integer \( d \),
\( 1 \leq d \leq s \), such that \( i_k \neq 1 \) for \( k = 1, \ldots, d \) and \( i_k = 1 \) for
\( k = d+1, \ldots, s \). The proof for arbitrary \( \tilde{I} \in \tilde{I} \) will be clear. Thus
in the following we assume \( \tilde{I} \) has the above property for some \( d \)
with \( 1 \leq d \leq s \), that \( D = \{1, \ldots, d\} \), and that definitions \((3.3.18)
(3.3.22)\) are made in terms of the particular \( d \) and \( D \).

Then \((3.3.17)\) is bounded above by

\[(3.3.23) \quad \sum_{\tilde{u} \in \tilde{U}} \tilde{P}(\tilde{u}) \left[ \sum_{u \in U} \tilde{P}(u) P(y_n | M(\tilde{I})) \right]
+ \tilde{N} \sum_{B(\tilde{u})} \tilde{P}(\tilde{u}) P(y_n | M(\tilde{I})) \tilde{P}^* \left[ P(y_n | M(\tilde{I})) \leq P(y_n | M^*(\tilde{I}, \tilde{I})) \right] \]

where \( \tilde{2} = (2,2,\ldots,2) \). The second term of this expression comes
from the fact that, for any \( k=1, \ldots, s \), \( U_k(1), U_k(2), \ldots, U_k(N_k) \) are
identically distributed.

Now \( (y_n, \tilde{u}) \in B(\tilde{u}) \) implies

\[(3.3.24) \quad \sum_{\tilde{u} \in \tilde{U}} \tilde{P}(\tilde{u}) P(y_n | M(\tilde{I})) \leq \frac{P(y_n | M(\tilde{I}))}{\alpha \tilde{N}} .\]

Hence the second term in \((3.3.23)\) is bounded above by
As for the first term, note the fact that

\[(3.3.26) \quad \mathbb{E}_{D}(\cdot, \cdot) = \sum_{t=1}^{n} R_{D}(p^{t}) \quad \text{for all} \quad D \subseteq \{1, \ldots, s\}\]

where the expectation is taken with respect to the p.d. \(p(\cdot)P(\cdot|\cdot)\) on \(\mathbb{M} \times Y_{n}\). Thus we choose \(\alpha\) so that

\[(3.3.27) \quad \alpha \tilde{N} < \exp\left[ \sum_{t=1}^{n} R_{D}(p^{t}) - k\sqrt{n} \right]\]

for some positive constant \(k\). Note that we can choose \(\alpha\) arbitrarily large so as to make the bound \(\frac{1}{\alpha}\) on the second term small and at the same time have \(\alpha\) satisfy (3.3.27). The see this, recall (3.3.5), (3.3.7) and the fact that we chose \(\tilde{R}^{i} = (R_{1}^{i}, \ldots, R_{J}^{i})\) to satisfy \(|R_{j}^{i} - R_{j}^{j}| < \delta\) for all \(j = 1, \ldots, J\). For if \(j\) is the index such that \(D = D(j)\), we have

\[(3.3.28) \quad \alpha \tilde{N} = \alpha \exp\left[ n \sum_{k=1}^{d} (R_{k} - \epsilon) \right] \leq \alpha \exp\left\{ n \frac{R_{j}^{i} + n\delta - ns}{\epsilon} \right\} \leq \alpha \exp\left\{ n \sum_{t=1}^{n} R_{D}(p^{t}) + n\delta - ns \right\} = \exp\left[ \sum_{t=1}^{n} R_{D}(p^{t}) - k\sqrt{n} + (k\sqrt{n} + \log \alpha - n(\epsilon - \delta)) \right]\]

where \(k\sqrt{n} + \log \alpha - n(\epsilon - \delta)\) can always be made less than zero for fixed \(\alpha\) and \(0 < \delta < \epsilon\) by making \(n\) sufficiently large.
Then with \( \alpha \) chosen to satisfy (3.3.27), it follows that

\[
(3.3.29) \quad \sum_{\vec{u} \in \mathcal{U}} \vec{p}(\vec{u}) \sum_{\vec{B}(\vec{v})} \vec{p}(\vec{v}) p(y_n | M(\vec{v})) = \sum_T p(M)p(y_n | M)
\]

where \( T = \{(y_n, M) | I_D(y_n, M) < E I_D(\cdot, \cdot) - k\sqrt{n}\} \). By Chebyshev's inequality, (3.3.29) is bounded above by

\[
(3.3.30) \quad \frac{\text{Var}(I_D(\cdot, \cdot))}{k^2 n}
\]

where \( \text{Var}(I_D(\cdot, \cdot)) \) denotes the variance of \( I_D(\cdot, \cdot) \). It is known (see [17], Chap. 8, for example) that there is a constant \( k_0 \) (independent of \( n \)) such that \( \text{Var}(I_D(\cdot, \cdot)) \leq k_0 n \) for all \( n = 1, 2, 3, \ldots \). Thus (3.3.30) can be made arbitrarily small by making \( k \) sufficiently large.

§3.4 Capacity Region of a Channel with \( s \) Senders and \( r \) Receivers

For the characterization of the capacity region, the following rate functions are needed. For all \( j = 1, \ldots, r \) define a function \( q_j(\cdot, \cdot) \) on \( \hat{X} \times Y(j) \) by

\[
(3.4.1) \quad q_j(y|\hat{x}) = \sum_{\hat{y}(j, y)} w(\hat{y}|\hat{x}) \text{ for all } \hat{x} \in \hat{X} \text{ and } y \in Y(j)
\]
where \( \hat{\gamma}(j,y) = \{ \hat{\gamma} \mid \hat{\gamma} = (y(1), \ldots, y(r)) \in \hat{\gamma} \text{ and } y(j) = y \} \). Let \( q_k(*) \) be a p.d. on \( X(k) \) for all \( k, 1 \leq k \leq s \), and \( q(*) \) denote the independent product distribution \( q_1 \times q_2 \times \cdots \times q_s \) on \( \hat{\gamma} \). Then for every \( D \subseteq \{1, \ldots, s\}, \ D \neq \emptyset \), a rate function is given by

\[
(3.4.2) \quad R^j_D(q) = \sum_{y \in \hat{\gamma}(j)} \sum_{\hat{x} \in \hat{\gamma}} q(\hat{x}) q_j(y | \hat{x}) \times \log q_j(y | \hat{x})
\]

\[
\quad \times \log \left( \sum_{k \in D} \sum_{x(k) \in X(k)} \prod_{k \in D} q_k(x(k)) q_j(y | \hat{x}) \right)
\]

for all \( j = 1, \ldots, r \), where \( \hat{x} = (x(1), \ldots, x(s)) \in \hat{\gamma} \).

Let \( \rho \) denote a finite set of \( s \)-tuples \( (q_1, \ldots, q_s) \) where \( q_k(*) \) is a p.d. on \( X_k \) for all \( k = 1, \ldots, s \). Again let \( q(*) \) denote the independent product distribution \( q_1 \times q_2 \times \cdots \times q_s \) on \( \hat{\gamma} \). Also let \( \mu(*) \) be a p.d. on \( \rho \).

To each pair \( (\rho, \mu) \) is assigned a vector \( \tilde{R}(\rho, \mu) \) as follows.

Let \( \tilde{R}_j(q) = [R^j_D(1)(q), \ldots, R^j_D(s)(q)] \) for all \( j = 1, \ldots, r \), and define
(3.4.3) \( \tilde{R}(\rho, \mu) = \min \{ \sum_{(q_1, \ldots, q_s) \in \Phi} \mu(q_1, \ldots, q_s) \tilde{R}_1(q), \ldots, \sum_{(q_1, \ldots, q_s) \in \Phi} \mu(q_1, \ldots, q_s) \tilde{R}_r(q) \} \)

where the minimum is to be taken componentwise. Denote

(3.4.4) \( F(Y(1), \ldots, Y(r)) = [\tilde{R} | \tilde{R} = \tilde{R}(\rho, \mu) \text{ for some } (\rho, \mu)] \)

and write its elements as \( \tilde{R} = (\tilde{R}_1, \ldots, \tilde{R}_r) \). Then define

(3.4.5) \( G(\tilde{R}) = \{ (R_1, \ldots, R_s) | \sum_{k \in D(j)} R_k \leq \tilde{R}_j \text{ for all } j = 1, \ldots, r \} \)

and

(3.4.6) \( G = \bigcup_{\tilde{R} \in F(Y(1), \ldots, Y(r))} G(\tilde{R}) \).

We remark that \( G \) is convex, closed under projections, and compact in the usual topology of Euclidean s-space.

Theorem 4. The capacity region \( G(P, T_{sr}) = G \).

Proof. First we show \( G(P, T_{sr}) \subseteq G \): Let \( (R_1, \ldots, R_s) \in G(P, T_{sr}) \).

Then for all \( \epsilon > 0 \) and \( 0 < \lambda < 1 \) there exists a code \( - (n, \tilde{N}, \lambda) \) for \( (P, T_{sr}) \) such that \( \frac{1}{n} \log N_k \geq R_k - \epsilon \) for all \( k = 1, \ldots, s \).

Let \( D \subseteq \{ 1, \ldots, s \} \), \( p_k^t(\cdot) \) be defined as in (3.2.5) for all \( t = 1, \ldots, n \) and \( k = 1, \ldots, s \), and \( p^t(\cdot) \) be defined by (3.2.2). By Lemma 2, for
all \( \epsilon > 0 \), there is a number \( k_D(\lambda, \epsilon, n) \) such that

\[
(3.4.7) \quad \log \prod_{k \in D} N_k \leq \sum_{t=1}^{n} R_D^j(p^t) + k_D(\lambda, \epsilon, n)
\]

for all \( j = 1, \ldots, r \) where \( \frac{1}{n} k_D(\lambda, \epsilon, n) \to 0 \) as \( n \to \infty \), \( \epsilon \to 0 \), and \( \lambda \to 0 \).

Then let \( \rho = \{(p_1^t, \ldots, p_s^t) | t = 1, \ldots, n\} \) and \( \mu(\cdot) \) be the equidistribution on \( \rho \). By arguing as Theorem 3, it follows that for all \( \delta > 0 \), if \( \epsilon \) and \( \lambda \) are sufficiently small and \( n \) sufficiently large, \( (R_1 - \delta, \ldots, R_s - \delta) \in G(\hat{R}) \) where \( \hat{R} = \hat{R}(\rho, \mu) \). Since \( G \) is closed, \( (R_1, \ldots, R_s) \in G \).

Finally we show \( G \subseteq G(P, T_{\text{sr}}) \). Let \( (R_1, \ldots, R_s) \in G \). Then there exists \( \hat{R} = \hat{R}(\rho, \mu) = (\hat{R}_1, \ldots, \hat{R}_k) \) such that

\[
\sum_{k \in D(j)} R_k \leq \hat{R}_j \quad \text{for all } j = 1, \ldots, l.
\]

Let \( \delta > 0 \). Then it is possible to find a positive integer \( n \) and p.d.'s \( p_k^t(\cdot) \) on \( X_k^t \) for all \( t = 1, \ldots, n \) and \( k = 1, \ldots, s \) such that if

\[
(3.4.8) \quad \tilde{R}' = \min\left\{ \frac{1}{n} \sum_{t=1}^{n} R_1(p^t), \ldots, \frac{1}{n} \sum_{t=1}^{n} R_s(p^t) \right\}
\]

(\( \text{where again the minimum is taken componentwise} \)), then

\[
(3.4.9) \quad d(\tilde{R}, \tilde{R}') < \delta
\]
where $d(\cdot, \cdot)$ denotes the usual metric in Euclidean $L$-space.

Let $\epsilon > 0$ and $N_k = e^{n(R_k - \epsilon)}$ for $k = 1, \ldots, s$. Select code-words $M(i)$ at random as in the proof of Theorem 3 and define decoding sets

$$
(3.4.10) \quad A_j(i) = \{ y_j | y_j \in Y_j \text{ and } P_j(y_j | M(i)) > P_j(y_j | M(i')) \}
$$

for $i' \neq i$.

Let $\lambda$ be a real number with $0 < \lambda < 1$. For each $j = 1, \ldots, r$ argue as in Theorem 1 with $A(i)$ replaced by $A_j(i)$ and $P(\cdot | \cdot)$ replaced by $P_j(\cdot | \cdot)$. Then it is possible to show the existence of a system $\{(M(i), A_j(i)) | i \in I, 1 \leq j \leq r\}$ satisfying

$$
(3.4.11) \quad \frac{1}{N} \sum_{i \in I} P_j(A_j(i) | M(i)) \leq \frac{\lambda}{r} \text{ for all } j = 1, \ldots, r
$$

if only it be required that

$$
(3.4.12) \quad \sum_{k \in D} R_k \leq \min \{ \sum_{1 \leq j \leq r} \mu(q_1, \ldots, q_s) R_D^j(q) \}
$$

for all $D \subseteq \{1, \ldots, s\}$. But this is exactly what it means for $(R_1, \ldots, R_s)$ to be in $G$. Hence $(R_1, \ldots, R_s) \in G(P, T_{sx})$. 

BIBLIOGRAPHY


