MULTI-WAY COMMUNICATION CHANNELS

R. AHLSWELDE
The Ohio State University
Columbus, Ohio
The University of Illinois
Urbana, Illinois

ABSTRACT

In the first two sections we define and classify multi-way channels of various kinds. The study of these channels was started by Shannon in his basic paper Two-way Communication Channels. Multi-way channels are much more complex than one-way channels and some definitely new phenomena occur. For instance the difference between the concept of a code with maximal error and the concept of a code with average error, which is unessential for one-way channels, is relevant for multi-way channels as was shown in [1]. Closely connected with this is the fact that maximal code methods, which have been developed for one-way channels (Feinstein [4], Wolfowitz [10]), do not exist until now (or may not exist at all) for multi-way channels. The only method which has been used successfully in proving coding theorems for multi-way channels (see [8]) is Shannon's random coding method [7]. Using this method and a lemma of Fano [3] we determine the capacity regions of the following multi-way channels:

(a) A channel with two senders and one receiver (Theorem 1 in Section 3)
(b) A channel with three senders and one receiver (Theorem 2 in Section 4).

Results for channels with two receivers will be presented in a forthcoming paper The Capacity Region of a Channel with Two Senders and Two Receivers.

1. INTRODUCTION

In [8] Shannon mentions (pp. 636 and 641) different kinds of multi-way channels and he makes some remarks on their capacity regions. We shall comment on those remarks later in connection with the results we obtain.

In order to have a clear and unique notation for the present paper and also for further papers we give now a formal description of various types of multi-way channels and we state the coding problems for them. However, for reasons of conceptual and notational simplicity we restrict ourselves to the cases with three or fewer senders and receivers. The extensions of our definitions to more general cases are straightforward.
Let $X = \{1, \ldots, a\}$, $Y = \{1, \ldots, b\}$, $Z = \{1, \ldots, c\}$, $\overline{X} = \{1, \ldots, \overline{a}\}$, $\overline{Y} = \{1, \ldots, \overline{b}\}$ and $\overline{Z} = \{1, \ldots, \overline{c}\}$ be finite sets. $X$, $Y$, and $Z$ are the input alphabets for the senders $S_X$, $S_Y$ and $S_Z$, and $\overline{X}$, $\overline{Y}$ and $\overline{Z}$ are the output alphabets for the receivers $R_{\overline{X}}$, $R_{\overline{Y}}$ and $R_{\overline{Z}}$. Let $X' = X$, $Y' = Y$, $Z' = Z$, $\overline{X}' = \overline{X}$, $\overline{Y}' = \overline{Y}$ and $\overline{Z}' = \overline{Z}$ for $t = 1, \ldots, n$ and define

$$X_n = \prod_{i=1}^{n} X_i, \quad Y_n = \prod_{i=1}^{n} Y_i, \quad Z_n = \prod_{i=1}^{n} Z_i, \quad \overline{X}_n = \prod_{i=1}^{n} \overline{X}_i, \quad \overline{Y}_n = \prod_{i=1}^{n} \overline{Y}_i \quad \text{and} \quad \overline{Z}_n = \prod_{i=1}^{n} \overline{Z}_i \quad \text{for} \quad n = 1, 2, \ldots.$$

Let $\omega(\overline{x}, \overline{y}, \overline{z} \mid x, y, z)$ be a non-negative function, which is defined on $X \times Y \times Z \times \overline{X} \times \overline{Y} \times \overline{Z}$ and satisfies

$$\sum_{\overline{x} \in \overline{X}} \sum_{\overline{y} \in \overline{Y}} \sum_{\overline{z} \in \overline{Z}} \omega(\overline{x}, \overline{y}, \overline{z} \mid x, y, z) = 1 \quad (1.1)$$

for every $(x, y, z) \in X \times Y \times Z$. The transmission probabilities of a three-way channel are defined by

$$P(\overline{x}_n, \overline{y}_n, \overline{z}_n \mid x_n, y_n, z_n) = \prod_{i=1}^{n} \omega(\overline{x}_i, \overline{y}_i, \overline{z}_i \mid x_i, y_i, z_i) \quad (1.2)$$

for every $x_n = (x^1, \ldots, x^n) \in X_n$, $y_n = (y^1, \ldots, y^n) \in Y_n$, $z_n = (z^1, \ldots, z^n) \in Z_n$, $\overline{x}_n = (\overline{x}^1, \ldots, \overline{x}^n) \in \overline{X}_n$, $\overline{y}_n = (\overline{y}^1, \ldots, \overline{y}^n) \in \overline{Y}_n$ and every $\overline{z}_n = (\overline{z}^1, \ldots, \overline{z}^n) \in \overline{Z}_n$, $(n = 1, 2, \ldots)$.

In an actual communication situation several senders and receivers may be located at the same terminal and can exchange their "information". In order to give a complete description of the communication situation we have to specify at which terminal the senders and receivers are located. We therefore introduce a system of sets $T = (\overline{T}_1, \ldots, \overline{T}_6)$, where

$$\bigcup_{i=1}^{6} \overline{T}_i = \{S_X, S_Y, S_Z, R_{\overline{X}}, R_{\overline{Y}}, R_{\overline{Z}}\} \quad \text{and} \quad \overline{T}_i \cap \overline{T}_j = \emptyset \quad \text{for} \quad i \neq j. \quad (1.3)$$

We shall say that for instance $S_X$ is at terminal $\overline{T}_i$ if $S_X \in \overline{T}_i$. It may be that some of the $\overline{T}_i$'s are empty. (We exclude here the case where a sender or a receiver "is" at more than one terminal, which would mean that the "information" at several terminals is available to him.) A three-way channel is completely described by a pair $(P, T)$, where $P$ denotes the transmission probabilities — as defined in (1.2) — and $T = (\overline{T}_1, \ldots, \overline{T}_6)$ says at which terminals the senders and the receivers are located. $(P, T)$ is said to be of pure type if no set $\overline{T}_i (i = 1, \ldots, 6)$ contains an element of $\{R_{\overline{X}}, R_{\overline{Y}}, R_{\overline{Z}}\}$ and an element of $\{S_X, S_Y, S_Z\}$. Otherwise we say that $(P, T)$ is of mixed type. Shannon's two-way communication channels [8] and also the channels introduced by van der Meulen [9] are of mixed type. We limit ourselves throughout this paper to channels of pure type, since they are already very complex. We denote by $(P, T_{ij})$ a channel with $i$ senders and $j$ receivers, each at a different terminal. This is actually already the general case of a channel of pure type. If a channel of pure type has more than one sender
(or receiver) at one terminal, then we always can consider those senders (or receivers) as one sender (or receiver). However, a channel \((P, T_{ij})\) can be used in general in several different ways for communication between the senders and the receivers. We classify now the various communication problems. For the channel \((P, T_{12})\) we have the following possibilities:

I) \(S_X\) sends to \(R_X\) or to \(R_Y\)
II) \(S_X\) sends to \(R_X\) and to \(R_Y\).

In the first case we have just a discrete memoryless channel. In the second case we have a compound channel, where the "receiver knows the individual channel which governs the transmission". (See [10], paragraph 4.5.) In both cases the capacities are known.

We consider now the channel \((P, T_{21})\). Again there are two cases:

I) Both \(S_X\) and \(S_Y\) send to \(R_X\)
II) \(S_X\) sends to \(R_X\), whereas \(S_Y\) sends letters in order to help maximizing the transmission of information from \(S_X\) to \(R_X\).

We denote the first case by \((P, T_{21}, I)\) and the second case by \((P, T_{21}, II)\). The treatment of the coding problem for the case \((P, T_{21}, II)\) is easy. If \(S_Y\) keeps sending the same letter \(y\) then the transmission from \(S_X\) to \(R_X\) is governed by a discrete memoryless channel with a capacity \(C_y \max_{y \in Y} C_y\) is the capacity in case \((P, T_{21}, II)\). The case \((P, T_{21}, I)\) for \(y \in Y\) will be treated in Section 3. In paragraph 17 of [8], Shannon writes: "In another paper we will discuss the case of a channel with two or more terminals having inputs only and one terminal with an output only, a case for which a complete and simple solution of the capacity region has been found." The paper cited appeared in 1962 and till now Shannon did not publish his results. It is not clear what kind of a characterization for the capacity region he had in mind. If this characterization is similar to the one given for two-way channels in paragraph 15 of [8], then this characterization cannot be considered to be simple or even useful. (Compare [6] for a stronger result.) For the channel \((P, T_{22})\) we can think of the following possibilities:

I) \(S_X\) sends to \(R_X\) and \(S_Y\) sends to \(R_Y\)
II) \(S_X\) and \(S_Y\) send to \(R_X\) and to \(R_Y\)
III) \(S_X\) sends to \(R_X\) and to \(R_Y\), \(S_Y\) sends to \(R_Y\) only
IV) \(S_X\) sends to \(R_X\) and to \(R_Y\), \(S_Y\) sends letters in order to help maximizing the transmission of information from \(S_X\) to \(R_X\) and to \(R_Y\).

We denote these four cases by \((P, T_{22}, I)\), \((P, T_{22}, II)\), \((P, T_{22}, III)\) and \((P, T_{22}, IV)\). Very likely Shannon had case \((P, T_{22}, I)\) in mind when he mentioned in the bottom lines of page 636 of [8] that the capacity region can be obtained by using independent sources. But this conjecture has — as far as we know — never been proved or disproved. A partial result — a
characterization of the capacity region by using dependent sources — is given in Section 2. (It is perhaps surprising that in case \((P, T_{22}, \Pi)\) a characterization of the capacity region in terms of independent sources can be given. This result will be stated and proved in a forthcoming paper.) The capacity in case \((P, T_{22}, \text{IV})\) can easily be determined by using results for compound channels. One can give similar classifications in general for channels of type \((P, T_{ij})\).

We consider only the channel \((P, T_{31})\). There are two new cases:

I) \(S_X, S_Y\) and \(S_Z\) send to \(R_X\)

II) \(S_X, S_Y\) send to \(R_X\). \(S_Z\) helps \(S_X\) and \(S_Y\) to achieve long and good codes.

The capacity regions for both cases are determined in Section 4.

We give now the definitions of the codes which are appropriate in the various situations, which we consider in the following sections. These are the cases: \((P, T_{21}, \text{I})\), \((P, T_{31}, \text{I})\), \((P, T_{31}, \Pi)\) and \((P, T_{22}, \text{I})\). The transmission probabilities for the channel \((P, T_{21})\) are defined by

\[
P(x_n | x_n, y_n) = \prod_{i=1}^{n} \omega(x' | x', y')
\]  

(1.4)

for every \(x_n = (x^1, \ldots, x^n)\), \(y_n = (y^1, \ldots, y^n) \in Y_n\), and \(\bar{x}_n = (\bar{x}^1, \ldots, \bar{x}^n) \in \bar{X}_n\) \((n = 1, 2, \ldots)\). \(\omega(x | x, y)\) is a non-negative function defined on \(X \times Y \times \bar{X}\) and satisfies

\[
\sum_{\bar{x} \in \bar{X}} \omega(x | x, y) = 1 \text{ for every } (x, y) \in X \times Y.
\]

A code \((n, N_1, N_2)\) for the communication situation \((P, T_{21}, \text{I})\) is a system

\[
\{ (u_i, v_j, A_{ij}) \mid i = 1, \ldots, N_1; j = 1, \ldots, N_2 \},
\]

(1.5)

where \(u_i \in X_n\), \(v_j \in Y_n\), \(A_{ij} \subset \bar{X}_n\) for \(i = 1, \ldots, N_1; j = 1, \ldots, N_2\), and \(A_{ij} \cap A_{i'j'} = \emptyset\) for \((i, j) \neq (i', j')\).

A code \((n, N_1, N_2)\) is an \((n, N_1, N_2, \lambda)\) code for \((P, T_{21}, \text{I})\), if

\[
\frac{1}{N_1N_2} \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} P(A_{ij} | u_i, v_j) \geq 1 - \lambda.
\]

(1.6)

The transmission probabilities of the channel \((P, T_{31})\) are defined by

\[
P(\bar{x}_n | x_n, y_n, z_n) = \prod_{i=1}^{n} \omega(x' | x', y', z')
\]

(1.7)

for every \(x_n = (x^1, \ldots, x^n) \in X_n\), \(y_n = (y^1, \ldots, y^n) \in Y_n\), \(z_n = (z^1, \ldots, z_n) \in Z_n\) and \(\bar{x}_n = (\bar{x}^1, \ldots, \bar{x}^n) \in \bar{X}_n\) \((n = 1, 2, \ldots)\). \(\omega(x | x', y', z')\) is defined on \(X \times Y \times Z \times \bar{X}\) and stochastic as usual.
A code \((n, N_1, N_2, N_3)\) for the communication situation \((P, T_{31}, \mathcal{I})\) is a system

\[
\{(u_i, v_j, w_k, A_{ijk}) \mid i = 1, \ldots, N_1; j = 1, \ldots, N_2; k = 1, \ldots, N_3\}
\]  

(1.8)

where \(u_i \in \mathcal{X}_n, v_j \in \mathcal{Y}_n, w_k \in \mathcal{Z}_n, A_{ijk} \subset \mathcal{X}_n\) for \(i = 1, \ldots, N_1; j = 1, \ldots, N_2; k = 1, \ldots, N_3\); and \(A_{ijk} \cap A_{i'j'k'} = \emptyset\) for \((i, j, k) \neq (i', j', k')\). A code \((n, N_1, N_2, \mathcal{I})\) is an \((n, N_1, N_2, \mathcal{I})\) code for \((P, T_{31}, \mathcal{I})\), if

\[
\frac{1}{N_1 N_2 N_3} \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} \sum_{k=1}^{N_3} P(A_{ijk} \mid u_i, v_j, w_k) \geq 1 - \lambda .
\]

(1.9)

A code \((n, N_1, N_2)\) for the communication situation \((P, T_{31}, \mathcal{I})\) is a system

\[
\{(u_i, v_j, w, A_{ij}) \mid i = 1, \ldots, N_1; j = 1, \ldots, N_2\}
\]

(1.10)

where \(u_i \in \mathcal{X}_n, v_j \in \mathcal{Y}_n, w \in \mathcal{Z}_n, A_{ij} \subset \mathcal{X}_n\) for \(i = 1, \ldots, N_1; j = 1, \ldots, N_2\) and \(A_{ij} \cap A_{i'j'} = \emptyset\) for \((i, j) \neq (i', j')\). A code \((n, N_1, N_2)\) for \((P, T_{31}, \mathcal{I})\) is an \((n, N_1, N_2, \mathcal{I})\) code, if

\[
\frac{1}{N_1 N_2} \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} P(A_{ij} \mid u_i, v_j, w) \geq 1 - \lambda .
\]

(1.11)

The transmission probabilities for the channel \((P, T_{32})\) are defined by

\[
P(\bar{x}_n, \bar{y}_n \mid x_n, y_n) = \prod_{i=1}^{n} \omega(\bar{x}_i, \bar{y}_i \mid x_i, y_i)
\]

(1.12)

for every \(x_n = (x^1, \ldots, x^n) \in \mathcal{X}_n, y_n = (y^1, \ldots, y^n) \in \mathcal{Y}_n, \bar{x}_n = (\bar{x}^1, \ldots, \bar{x}^n) \in \mathcal{X}_n\) and every \(\bar{y}_n = (\bar{y}^1, \ldots, \bar{y}^n) \in \mathcal{Y}_n\) \((n = 1, 2, \ldots)\).

\(\omega(\bar{x}, \bar{y} \mid x, y)\) is defined on \(\mathcal{X} \times \mathcal{Y} \times \mathcal{X} \times \mathcal{Y}\), is non-negative, and such that

\[
\sum_{\bar{x} \in \mathcal{X}} \sum_{\bar{y} \in \mathcal{Y}} \omega(\bar{x}, \bar{y} \mid x, y) = 1\text{ for every } (x, y) \in \mathcal{X} \times \mathcal{Y}.
\]

A code \((n, N_1, N_2)\) for \((P, T_{32}, \mathcal{I})\) is a system

\[
\{(u_i, v_j, A_i, B_j) \mid i = 1, \ldots, N_1; j = 1, \ldots, N_2\}\}
\]

(1.13)

where \(u_i \in \mathcal{X}_n, v_j \in \mathcal{Y}_n, A_i \subset \mathcal{X}_n, B_j \subset \mathcal{Y}_n\) for \(i = 1, \ldots, N_1; j = 1, \ldots, N_2\); and \(A_i \cap A_i = \emptyset\) for \(i = i', j = j'\).

For \(A \subset \mathcal{X}_n, B \subset \mathcal{Y}_n\) define

\[
P(A \mid x_n, y_n) = \sum_{\bar{x}_n \in A} \sum_{\bar{y}_n \in B} P(\bar{x}_n, \bar{y}_n \mid x_n, y_n)
\]

(1.14)

and

\[
P(B \mid x_n, y_n) = \sum_{\bar{x}_n \in A} \sum_{\bar{y}_n \in B} P(\bar{x}_n, \bar{y}_n \mid x_n, y_n)\text{ for } (x_n, y_n) \in \mathcal{X}_n \times \mathcal{Y}_n.
\]
A code \((n, N_1, N_2)\) is an \((n, N_1, N_2, \lambda_1, \lambda_2)\) code for \((P, T_{22}, I)\) if
\[
\frac{1}{N_1 N_2} \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} P(A_i | u_i, v_j) \geq 1 - \lambda_1 \tag{1.15}
\]
and
\[
\frac{1}{N_1 N_2} \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} P(B_j | u_i, v_j) \geq 1 - \lambda_2.
\]

In (1.6), (1.9), (1.11) and (1.15) we have defined codes with average errors \(\leq \lambda\). Throughout this paper only average errors will be used. The coding theory for multi-way channels for maximal errors is entirely different. This has been demonstrated in Section 5 of [1]. In that paper we deal with Shannon's two-way communication channels, but the discussion in Section 5 of [1] is relevant for all multi-way channels. Till now no coding theorem for any nontrivial multi-way channel has been proved in case that one requires the maximal errors to be small.

We define now the capacity regions for \((P, T_{21}, I)\), \((P, T_{31}, I)\), \((P, T_{31}, \Pi)\) and \((P, T_{22}, I)\). A pair of non-negative real numbers \((R_1, R_2)\) is called a pair of achievable rates for \((P, T_{21}, I)\) or \((P, T_{31}, \Pi)\) if for any \(\lambda (0 < \lambda < 1)\) and any \(\epsilon > 0\) there exists a code \((n, N_1, N_2, \lambda)\) such that \(\frac{1}{n} \log N_1 \geq R_1 - \epsilon\) and \(\frac{1}{n} \log N_2 \geq R_2 - \epsilon\) for all sufficiently large \(n\).

\((R_1, R_2)\) is called a pair of achievable rates for \((P, T_{22}, I)\) if for any \((\lambda_1, \lambda_2), 0 < \lambda_1, \lambda_2 < 1, \) and any \(\epsilon > 0\) there exists a code \((n, N_1, N_2, \lambda_1, \lambda_2)\) such that \(\frac{1}{n} \log N_1 \geq R_1 - \epsilon\) and \(\frac{1}{n} \log N_2 \geq R_2 - \epsilon\) for all sufficiently large \(n\).

A triple of non-negative real numbers \((R_1, R_2, R_3)\) is called a triple of achievable rates for \((P, T_{31}, I)\) if for any \(\lambda (0 < \lambda < 1)\) and any \(\epsilon > 0\) there exists a code \((N_1, N_2, N_3, \lambda)\) such that \(\frac{1}{n} \log N_1 \geq R_1 - \epsilon, \frac{1}{n} \log N_2 \geq R_2 - \epsilon\) and \(\frac{1}{n} \log N_3 \geq R_3 - \epsilon\) for all sufficiently large \(n\).

Let \(G(P, T_{21}, I)\) be the set of all pairs of achievable rates for \((P, T_{21}, I)\), let \(G(P, T_{31}, \Pi)\) be the set of all pairs of achievable rates for \((P, T_{31}, \Pi)\) and let \(G(P, T_{22}, I)\) be the set of all pairs of achievable rates for \((P, T_{22}, I)\). The set of triples of achievable rates for \((P, T_{31}, I)\) is denoted by \(G(P, T_{31}, I)\).

In Section 3 we give a simple characterization for \(G(P, T_{21}, I)\) and in Section 4 we give simple characterizations for \(G(P, T_{31}, I)\) and \(G(P, T_{31}, \Pi)\). The problem to find a simple characterization for \(G(P, T_{22}, I)\) is still unsolved. We shall discuss this problem in Section 2.
2. ON A CHANNEL WITH TWO SENDER S AND TWO RECEIVERS

We consider now the channel \((P, T_{22})\) in case \((P, T_{22}, I)\) and we present first a characterization of \(G(P, T_{22}, I)\) in terms of dependent sources. This characterization was obtained in discussions with J. Wolfowitz, who raised the question, and with U. Augustin, who wrote down a formal proof, during a visit in Heidelberg in the summer of 1969. We state the result as Lemma 1 below. The result is theoretically and practically unsatisfactory, because it cannot be used for actually computing \(G(P, T_{22}, I)\). We give the result here, because it is interesting to compare it with the results, which we obtain in Sections 3 and 4 for \(G(P, T_{21}, I)\), \(G(P, T_{21}, I)\) and \(G(P, T_{31}, II)\). The ideas which led to those results may also be helpful in finding better characterizations for \(G(P, T_{22}, I)\).

Before we can state Lemma 1 we need some definitions. Let \(p_n\) be a probability distribution (p. d.) on \(X_n\) and let \(q_n\) be a p. d. on \(Y_n\). For every \(x_n \in X_n\) and \(\bar{x}_n \in \bar{X}_n\) we define now \(P(\bar{x}_n \mid x_n)\) by

\[
P(\bar{x}_n \mid x_n) = \sum_{y_n \in Y_n} q_n(y_n) P(\bar{y}_n \mid x_n, y_n).
\]  

(2.1)

For every \(y_n \in Y_n\) and \(\bar{y}_n \in \bar{Y}_n\) we define \(P(\bar{y}_n \mid y_n)\) by

\[
P(\bar{y}_n \mid y_n) = \sum_{x_n \in X_n} p_n(x_n) P(\bar{y}_n \mid x_n, y_n).
\]  

(2.2)

For every \(\bar{x}_n \in \bar{X}_n\), \(\bar{y}_n \in \bar{Y}_n\) we define \(p'_n(\bar{x}_n)\) and \(q'_n(\bar{x}_n)\) by

\[
p'_n(\bar{x}_n) = \sum_{x_n \in X_n} p_n(x_n) P(\bar{x}_n \mid x_n)
\]  

(2.3)

and

\[
q'_n(\bar{y}_n) = \sum_{y_n \in Y_n} q_n(y_n) P(\bar{y}_n \mid y_n).
\]  

(2.4)

The "rate" functions \(R_1(p_n, q_n)\) and \(R_2(p_n, q_n)\) are given by

\[
R_1(p_n, q_n) = \sum_{\bar{x}_n \in \bar{X}_n} \sum_{x_n \in X_n} p_n(x_n) P(\bar{x}_n \mid x_n) \log \frac{P(\bar{x}_n \mid x_n)}{p'_n(\bar{x}_n)}
\]  

(2.5)

and

\[
R_2(p_n, q_n) = \sum_{\bar{y}_n \in \bar{Y}_n} \sum_{y_n \in Y_n} q_n(y_n) P(\bar{y}_n \mid y_n) \log \frac{P(\bar{y}_n \mid y_n)}{q'_n(\bar{y}_n)}.
\]  

(2.6)

Consider the set of pairs

\[
G_n = \left\{ \frac{1}{n} R_1(p_n, q_n), \frac{1}{n} R_2(p_n, q_n) \right\} | p_n \text{ p. d. on } X_n, q_n \text{ p. d. on } Y_n
\]

and define \(G = \bigcup_{n=1}^{\infty} G_n\).
Lemma 1

a) $G(P, T_{22}, I) = G$.

b) $G(P, T_{22}, I)$ is a closed convex set in the Euclidean plane, which contains with every pair $(R_1, R_2)$ also the projections $(R_1, 0)$, $(0, R_2)$, and $(0, 0)$.

The proof is based on Shannon's random coding method [7] and on Fano's lemma [3].

Proof. We show first that $G \subset G(P, T_{22}, I)$. Let $\hat{u} = (u_1, \ldots, u_{N_1})$ be a vector with $N_1$ components, which are elements of $X_n$, and let $\hat{v} = (v_1, \ldots, v_{N_2})$ be a vector with $N_2$ components, which are elements of $Y_n$. Define for $\hat{u}$ the decoding sets $A_1, \ldots, A_{N_1}$, by

$$A_i = \{ \bar{x}_n | P(\bar{x}_n | u_i) > P(\bar{x}_n | u_j) \text{ for } j \neq i \} \quad (2.7)$$

for $i = 1, \ldots, N_1$.

Similarly we define the decoding sets $B_1, \ldots, B_{N_2}$ for $\hat{v}$ by

$$B_j = \{ \bar{y}_n | P(\bar{y}_n | v_j) > P(\bar{y}_n | v_k) \text{ for } k \neq j \} \quad (2.8)$$

for $j = 1, \ldots, N_2$.

Let $A_i^c$ be the complement of $A_i$ and $B_i^c$ be the complement of $B_i$. The average errors for the code $\{(u_i, v_j, A_i, B_j) | i = 1, \ldots, N_1; j = 1, \ldots, N_2\}$ are given by

$$\lambda_1(\hat{u}, \hat{v}) = \frac{1}{N_1 N_2} \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} P(A_i^c | u_i, v_j) \quad (2.9)$$

and

$$\lambda_2(\hat{u}, \hat{v}) = \frac{1}{N_1 N_2} \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} P(B_i^c | u_i, v_j) \quad (2.10)$$

We select now $\hat{u}$ and $\hat{v}$ independently at random according to the probability distributions $\hat{p}_n$ and $\hat{q}_n$, given by

$$\hat{p}_n(\hat{u}) = \prod_{i=1}^{N_1} p_n(u_i) \quad (2.11)$$

for all $u_i \in X_n$, $i = 1, \ldots, N_1$, and

$$\hat{q}_n(\hat{v}) = \prod_{j=1}^{N_2} q_n(v_j) \quad (2.12)$$

for all $v_j \in Y_n$, $j = 1, \ldots, N_2$.

We give now an upper bound on $l$, given by

$$l = \sum_{\hat{u}} \sum_{\hat{v}} \hat{p}_n(\hat{u}) \hat{q}_n(\hat{v}) \left[ \lambda_1(\hat{u}, \hat{v}) + \lambda_2(\hat{u}, \hat{v}) \right] \quad (2.13)$$

If $l$ is small, then there exists a pair $(\hat{u}_o, \hat{v}_o)$ with small decoding error probabilities.
In order to obtain an upper bound on \( l \) it suffices for symmetry reasons to give a bound on \( l_1 \), given by

\[
l_1 = \sum_u \sum_{\hat{v}} \hat{p}_n(\hat{u}) \hat{q}_n(\hat{v}) \lambda_1(\hat{u}, \hat{v})
\]

\[
= \sum_{\hat{u}} \sum_{\hat{v}} \hat{p}_n(\hat{u}) \hat{q}_n(\hat{v}) \left[ \frac{1}{N_1 N_2} \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} P(A_i^\xi | u_i, v_j) \right].
\]  

(2.14)

Since \( \hat{u}, \hat{v} \) are selected independently we obtain

\[
l_1 = \sum_{\hat{u}} \hat{p}_n(\hat{u}) \left[ \frac{1}{N_1} \sum_{i=1}^{N_1} P(A_i^\xi | u_i) \right].
\]  

(2.15)

To (2.15) we can apply the usual error bound which one obtains in connection with Shannon's random coding method in case of a one-way channel ([5], [7], [11]). Choosing \( p_n = px \ldots xp, q_n = qx \ldots xq \) yields (as for discrete memoryless channels) that \( G_1 \subseteq G(P, T_{22}, I) \). Replacing in our arguments \( x \) by \( X_i, Y \) by \( Y_i, \bar{X} \) by \( \bar{X}_i \) and \( \bar{Y} \) by \( \bar{Y}_i \) yields \( G_i \subseteq G(P, T_{22}, I) \) for \( t = 1, 2, \ldots \). We thus have proved: \( G \subseteq G(P, T_{22}, I) \).

Let now \( \{ (u_i, v_j, A_i, B_j) \mid i = 1, \ldots, N_1; j = 1, \ldots, N_2 \} \) be an \((n, N_1, N_2, \lambda_1, \lambda_2)\) code for \((P, T_{22}, I)\). Then we have

\[
\lambda_1 \geq \frac{1}{N_1 N_2} \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} P(A_i^\xi | u_i, v_j)
\]  

(2.16)

and

\[
\lambda_2 \geq \frac{1}{N_1 N_2} \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} P(B_j^\xi | u_i, v_j).
\]  

(2.17)

Define \( p_n \) on \( X_n \) by

\[
p_n(u_i) = \frac{1}{N_1}, \quad i = 1, \ldots, N_1,
\]  

(2.18)

and \( q_n \) on \( Y_n \) by

\[
q_n(v_j) = \frac{1}{N_2}, \quad j = 1, \ldots, N_2.
\]  

(2.19)

Then we obtain

\[
\lambda_1 \geq \frac{1}{N_1} \sum_{i=1}^{N_1} \left( \sum_{j=1}^{N_2} \frac{1}{N_2} P(A_i^\xi | u_i, v_j) \right) = \frac{1}{N_1} \sum_{i=1}^{N_1} P(A_i^\xi | u_i)
\]  

(2.20)

and similarly

\[
\lambda_2 \geq \frac{1}{N_2} \sum_{j=1}^{N_2} P(B_j^\xi | v_j).
\]  

(2.21)
Fano's lemma ([3], [5], [11]) yields

$$\log N_1 \leq \frac{R_1(p_n, q_n) + 1}{1 - \lambda_1} \quad (2.22)$$

and

$$\log N_2 \leq \frac{R_2(p_n, q_n) + 1}{1 - \lambda_2}. \quad (2.23)$$

One can show — by a similar argument as Shannon gave in paragraph 12 of [8], (see also Section 3) — that $G$ is convex and closed under projections. It follows therefore from the definition of $G(P, T_{22}, I)$ and from (2.22) and (2.23) that $G \supseteq G(P, T_{22}, I)$. This completes the proof of Lemma 1. The conjecture made in the last sentence of page 636 in [8] means in our terminology that $G(P, T_{22}, I)$ equals the convex hull of $G_1$.

In order to verify this conjecture one might think of the following possibilities:

1) One tries to find a new Fano type estimate for $(P, T_{22}, I)$, which then might yield the desired result.
2) One tries to prove directly that $G$ equals the convex hull of $G_1$.

We introduce now a special channel $(P, T_{22})$ which may be helpful in order to disprove the conjecture. Let $\bar{Y} = Y$ and let the transmission matrix be such that

$$w(\bar{y} | x, y) = 1 \text{ for } \bar{y} = y \in Y. \quad (2.24)$$

$w(\bar{x} | x, y)$ can be chosen as appropriate.

3. THE CAPACITY REGION OF A CHANNEL WITH TWO SENDERS AND ONE RECEIVER

We consider now the channel $(P, T_{21})$ in case $(P, T_{21}, I)$. We recall its definition — given in Section 1 — and also definitions (1.4), (1.5), (1.6) and (1.16). We need now some more definitions.

Let $p$ be a p. d. on $X$ and let $q$ be a p. d. on $Y$. We define now the functions $R(p, q), R_1(p, q), R_{21}(p, q), R_2(p, q)$ and $R_2(p, q)$ by

$$R(p, q) = \sum_{x \in X} \sum_{y \in Y} \sum_{\bar{x} \in \bar{X}} p(x) q(y) w(\bar{x} | x, y) \log \frac{w(\bar{x} | x, y)}{\sum_{x \in X} \sum_{y \in Y} p(x) q(y) w(\bar{x} | x, y)}, \quad (3.1)$$

$$R_1(p, q) = \sum_{x \in X} \sum_{y \in Y} \sum_{\bar{x} \in \bar{X}} p(x) q(y) w(\bar{x} | x, y) \log \frac{w(\bar{x} | x, y)}{\sum_{x \in X} p(x) w(\bar{x} | x, y)}. \quad (3.2)$$
Multi-way communication channels

\[ R_{12}(p, q) = \sum_{x \in X} \sum_{y \in Y} \sum_{\tilde{x} \in \tilde{X}} p(x) q(y) w(\tilde{x} \mid x, y) \log \frac{\sum_{\tilde{x}} p(x) w(\tilde{x} \mid x, y)}{\sum_{x \in X} \sum_{y \in Y} q(y) w(\tilde{x} \mid x, y)}, \]

(3.3)

\[ R_{21}^2(p, q)' = \sum_{x \in X} \sum_{y \in Y} \sum_{\tilde{x} \in \tilde{X}} p(x) q(y) w(\tilde{x} \mid x, y) \log \frac{\sum_{y} q(y) w(\tilde{x} \mid x, y)}{\sum_{x \in X} \sum_{y \in Y} p(x) q(y) w(\tilde{x} \mid x, y)}, \]

(3.4)

and

\[ R_2(p, q) = \sum_{x \in X} \sum_{y \in Y} \sum_{\tilde{x} \in \tilde{X}} p(x) q(y) w(\tilde{x} \mid x, y) \log \frac{w(\tilde{x} \mid x, y)}{\sum_{y} q(y) w(\tilde{x} \mid x, y)}. \]

(3.5)

One easily verifies that

\[ R(p, q) = R_1(p, q) + R_{12}(p, q) = R_2(p, q) + R_{21}^2(p, q). \]

(3.6)

Define the sets

\[ G = \{(R_1(p, q), R_{12}^2(p, q)) \mid p \text{ p. d. on } X, q \text{ p. d. on } Y\}, \]

\[ \tilde{G} = \{(R_{21}^2(p, q), R_2(p, q)) \mid p \text{ p. d. on } X, q \text{ p. d. on } Y\} \]

and \( G = G \cup \tilde{G} \). Finally we define \( G^* \) as the convex hull of \( G \).

Let \( C_1 = \max_p \max_q R_1(p, q) \) and let \( C_2 = \max_p \max_q R_2(p, q) \). It follows from the definitions (3.2) and (3.5) that

\[ C_1 \leq \max_q \sum_y q(y) \left[ \max_p \sum_{x, \tilde{x}} p(x) w(\tilde{x} \mid x, y) \log \frac{w(\tilde{x} \mid x, y)}{\sum_{x} p(x) w(\tilde{x} \mid x, y)} \right]. \]

Let \( y_0 \) be such that the term in brackets is maximal. The p. d. \( q_0 \) which assigns probability 1 to \( y_0 \) maximizes the right side of the inequality. We therefore have \( C_1 = \max_p R_1(p, q_0) \). It is easy to see that \( R_{12}^2(p, q_0) = 0 \) for all \( p \). Similarly one can find a \( p_0 \), which is concentrated on one point, such that \( C_2 = \max_R R_2(p_0, q) \) and \( R_{21}^2(p_0, q) = 0 \) for all \( q \). It is also easy to see that for \( (p_0, q_0) \), for instance, \( R_1(p_0, q_0) = R_{12}^2(p_0, q_0) = 0 \). It follows from a convexity argument (see Lemma 2 in Section 4) that \( R_{12}^2(p, q) \leq R_2(p, q) \) and \( R_{21}^2(p, q) \leq R_1(p, q) \). Hence, for any element \( (R_1, R_2) \) in \( G^* \) we have

\[ R_1 \leq C_1 \quad \text{and} \quad R_2 \leq C_2. \]

Since \((0, 0) \in G^* \) and \( G^* \) is convex, we obtain that \((R_1, 0)\) and \((0, R_2)\) are also in \( G^* \). We are now ready to state

**Theorem 1**

(a) The capacity region \( G(P, T_{21}, I) \) for \((P, T_{21}, I)\) equals \( G^* \).

(b) \( G(P, T_{21}, I) \) is a closed convex set in the Euclidean plane and contains with every point \((R_1, R_2)\) also the projections \((R_1, 0), (0, R_2),\) and \((0, 0)\).
Proof. We show first that $G^* \subseteq G(P, T_{21}, I)$. For this it suffices to show that for any pair $(p, q)$ $(R_1(p, q), R_1(p, q)) \in G(p, T_{21}, I)$. The proof that $(R_2(p, q), R_2(p, q))$ is contained in $G(P, T_{21}, I)$ is symmetrically the same. Once we know that $G \subseteq G(P, T_{21}, I)$ we can show that $G^* \subseteq G(P, T_{21}, I)$ by concatenation.

Let $\hat{u} = (u_1, \ldots, u_{N_1})$ be a vector with $N_1$ components, which are elements of $X_n$, and let $\hat{v} = (v_1, \ldots, v_{N_2})$ be a vector with $N_2$ components, which are elements of $Y_n$. Let $p_n$ be a p.d. on $X_n$ and let $q_n$ be a p.d. on $Y_n$. We select now $\hat{u}$ and $\hat{v}$ independently at random according to the probability distributions $\hat{p}_n$ and $\hat{q}_n$, given by

$$\hat{p}_n(\hat{u}) = \prod_{i=1}^{N_1} p_n(u_i)$$

for all $u_i \in X_n$, $i = 1, \ldots, N_1$, and

$$\hat{q}_n(\hat{v}) = \prod_{j=1}^{N_2} q_n(v_j)$$

for all $v_j \in Y_n$, $j = 1, \ldots, N_2$.

Define for every $y_n \in Y_n$, $\bar{x}_n \in \bar{X}_n$ $P(\bar{x}_n | y_n)$ by

$$P(\bar{x}_n | y_n) = \sum_{x_n \in X_n} p_n(x_n)P(\bar{x}_n | x_n, y_n).$$

We define for $\hat{v}$ the decoding sets $B_1, \ldots, B_{N_2}$ by

$$B_j = \{\bar{x}_n | P(\bar{x}_n | v_j) > P(\bar{x}_n | v_k) \text{ for } k \neq j\}$$

for $j = 1, \ldots, N_2$.

For the pair $(\hat{u}, \hat{v})$ we define the decoding sets $A^*_i(i = 1, \ldots, N_1; j = 1, \ldots, N_2)$ by

$$A^*_i = \{\bar{x}_n | P(\bar{x}_n | u_i, v_j) > P(\bar{x}_n | u_i, v_l) \text{ for } l \neq i\}$$

for $i = 1, \ldots, N_1; j = 1, \ldots, N_2$.

Obviously, for every fixed $j$, $A^*_i \cap A^*_{i'} = \emptyset$ for $i \neq i'$.

Define $A_{ij}$ by

$$A_{ij} = A^*_i \cap B_j.$$

We have $A_{ij} \cap A_{i'j} = \emptyset$ for $(i, j) \neq (i', j')$. Let $A^*_{ij}$ be the complement of $A_{ij}$. The average error for the code $\{(u_i, v_j, A_{ij}) | i = 1, \ldots, N_1; j = 1, \ldots, N_2\}$ is given by

$$\lambda(\hat{u}, \hat{v}) = \frac{1}{N_1 N_2} \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} P(A^*_i | u_i, v_j).$$

We want to give an upper bound on $\epsilon = \sum \sum \hat{p}_n(\hat{u}) \hat{q}_n(\hat{v}) \lambda(\hat{u}, \hat{v})$. We rewrite first (3.13) as

$$\lambda(\hat{u}, \hat{v}) = \frac{1}{N_1 N_2} \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} P(A^*_{ij} \cup B_j | u_i, v_j).$$

(3.14)
It follows from (3.14) that
\[ \lambda(\hat{u}, \hat{v}) \leq \frac{1}{N_1 N_2} \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} [P(A_{ij}^c \mid u_i, v_j) + P(B_j^c \mid u_i, v_j)]. \] (3.15)

(3.15) yields
\[ e \leq \sum_{\hat{u}} \sum_{\hat{v}} \hat{p}_n(\hat{u}) \hat{q}_n(\hat{v}) \frac{1}{N_1 N_2} \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} P(A_{ij}^c \mid u_i, v_j) + \]
\[ + \sum_{\hat{v}} \hat{q}_n(\hat{v}) \frac{1}{N_2} \sum_{j=1}^{N_2} P(B_j^c \mid v_j). \] (3.16)

We choose now \( p_n = px \ldots xp, q_n = qx \ldots xq \) and we denote the first term in the sum above by \( e_1 \) and the second term by \( e_2 \). \( e_1 \) is an error term which occurs by applying the random coding method to Shannon's two-way channels (see [8] or [12], p. 5). Shannon's estimates for \( e_1 \) yield that for any \( \delta > 0 \) we can choose \( N_1 = \beta(p, q) \), where \( R_1 \) equals \( R_1(p, q) \), and we can find an \( \epsilon > 0 \), such that \( e_1 = e_1(N_1) \leq e^{-\epsilon n} \) for all sufficiently large \( n \). \( e_2 \) is the usual error term occurring in applying the random coding method to the discrete memoryless channel whose transmission matrix \( \hat{w}(\cdot \mid \cdot) \) is given by
\[ \hat{w}(\hat{x} \mid y) = \sum_{x \in X} p(x)w(x \mid x, y) \text{ for every } y \in Y, \hat{x} \in \hat{X}. \] (3.17)

We therefore have that for any \( \eta > 0 \) and \( N_2 = \beta(p, q) \), where \( R_2 \) equals \( R_2(p, q) \), we can find an \( \epsilon > 0 \) such that \( e_3 = e_3(N_2) \leq e^{-\epsilon n} \) for all sufficiently large \( n \). We thus have proved that \( G^* \subseteq G(P, T_{21}, I) \). We shall prove now that \( G^* \subseteq G(P, T_{21}, I) \). Let \( \{(u_i, v_j, A_{ij}) \mid i = 1, \ldots, N_1; j = 1, \ldots, N_2\} \) be an \((n, N_1, N_2, \lambda)\) code for \((P, T_{21}, I)\). From the inequality
\[ \frac{1}{N_1 N_2} \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} P(A_{ij}^c \mid u_i, v_j) \leq \lambda \]
we obtain that
\[ \frac{1}{N_2} \sum_{j=1}^{N_2} P(A_{ij}^c \mid u_i, v_j) \leq 2\lambda \text{ for at least } \left\lceil \frac{N_1}{2} \right\rceil u_i's. \] (3.18)

We can change the numbering such that these \( u_i's \) are \( u_1, \ldots, u_{N_1^*} \) with \( N_1^* = \left\lceil \frac{N_1}{2} \right\rceil \). Similarly we can pick — after renumbering — \( v_1, \ldots, v_{N_2^*} \), with \( N_2^* = \left\lceil \frac{N_2}{2} \right\rceil \), such that
\[ \frac{1}{N_1} \sum_{i=1}^{N_1} P(A_{ij}^c \mid u_i, v_j) \leq 2\lambda \text{ for } j = 1, \ldots, N_2^*. \] (3.19)
It follows from (3.18) that
\[ \frac{1}{N_2^*} \sum_{i=1}^{N_2^*} P(A_{ij} \mid u_i, v_j) \leq 4\lambda \text{ for } i = 1, \ldots, N_1^* \] (3.20)
and it follows from (3.19) that
\[ \frac{1}{N_2^*} \sum_{i=1}^{N_2^*} P(A_{ij} \mid u_i, v_j) \leq 4\lambda \text{ for } j = 1, \ldots, N_2^*. \] (3.21)
(3.20) or (3.21) imply that
\[ \frac{1}{N_1^* \cdot N_2^*} \sum_{i=1}^{N_1^*} \sum_{j=1}^{N_2^*} P(A_{ij} \mid u_i, v_j) \leq 4\lambda. \] (3.22)

We write now \( \lambda^* \) instead of \( 4\lambda \). Henceforth we consider the \( (n, N_1^*, N_2^*, \lambda^*) \) code \( \{(u_i, v_j, A_{ij}) \mid i = 1, \ldots, N_1^*; j = 1, \ldots, N_2^*\} \). Upper bounds on \( N_1^* \) and \( N_2^* \) yield immediately upper bounds on \( N_1 \) and \( N_2 \). We define now a discrete memoryless channel \( \hat{P} \) with input alphabet \( Z = X \times Y \) and output alphabet \( \bar{X} \). The transmission matrix \( \hat{\omega} \) for \( \hat{P} \) is defined by \( \hat{\omega}(\bar{x} \mid z) = \hat{\omega}(\bar{x} \mid x, y) \) for all \( z = (x, y) \in X \times Y \). For every \( z_n \in Z_n \) and \( \bar{x}_n \in \bar{X}_n \) we define \( \hat{P}(\bar{x}_n \mid z_n) \) by
\[ \hat{P}(\bar{x}_n \mid z_n) = \prod_{i=1}^{n} \hat{\omega}(\bar{x}_i \mid z_i). \] (3.23)

Write \( u_i = (u_{i1}, \ldots, u_{in}) \) for \( i = 1, \ldots, N_1 \) and \( v_j = (v_{j1}, \ldots, v_{jn}) \) for \( j = 1, \ldots, N_2 \). If we define \( w_{ij} \) by
\[ w_{ij} = (u_{ij}, v_{ij}) \] (3.24)
for \( i = 1, \ldots, N_1; j = 1, \ldots, N_2 \) and \( t = 1, 2, \ldots, n \); and if we define \( w_{ij} \) by
\[ w_{ij} = (w_{ij1}, \ldots, w_{ijn}) \] (3.25)
for \( i = 1, \ldots, N_1; j = 1, \ldots, N_2 \); then we have that
\[ \hat{P}(A_{ij} \mid u_i, v_j) = \hat{P}(A_{ij} \mid w_{ij}) \] (3.26)
for \( i = 1, \ldots, N_1^*; j = 1, \ldots, N_2^* \). \( \{(w_{ij}, A_{ij}) \mid i = 1, \ldots, N_1^*; j = 1, \ldots, N_2^*\} \) is because of (3.26) an \( (n, N_1^*, N_2^*, \lambda^*) \) code for the channel \( \hat{P} \). Let \( N \) equal \( N_1^* \times N_2^* \) and define the probability distribution \( r_n \) on \( Z_n \) by
\[ r_n(w_{ij}) = \frac{1}{N} \] (3.27)
for \( i = 1, \ldots, N_1; j = 1, \ldots, N_2 \).

Fano's lemma yields that
\[ \log N \leq (R(r_n, \hat{P}) + 1)(1 - \lambda)^{-1}, \] where \( R(r_n, \hat{P}) \) equals
\[ \sum_{z_n \in Z_n} \sum_{\bar{x}_n \in \bar{X}_n} r_n(z_n) \hat{P}(\bar{x}_n \mid z_n) \log \frac{\hat{P}(\bar{x}_n \mid z_n)}{\sum_{z_n \in Z_n} r_n(z_n) \hat{P}(\bar{x}_n \mid z_n)}. \] (3.28)
Let $r^t$ be the $t$th 1-dimensional marginal distribution of $r_n$ and define $R(r^t, \bar{\omega})$ by
\[
R(r^t, \bar{\omega}) = \sum_{x^t} \sum_{z^t} r^t(z^t) \bar{\omega}(x^t | z^t) \log \frac{\bar{\omega}(x^t | z^t)}{\sum_{z^t} r^t(z^t) \bar{\omega}(x^t | z^t)} \quad (t = 1, 2, \ldots, n). \tag{3.29}
\]

It is well known (see [3] or [5], Theorem 4.2.1) that
\[
R(r_n, P) \leq \sum_{t=1}^{n} R(r^t, \bar{\omega}). \tag{3.30}
\]

Define now
\[
p^t(x) = \frac{|\{i \mid u^t_i = x, i \in \{1, \ldots, N^*_1\}\}|}{N^*_1} \tag{3.31}
\]
for $x \in X^t$, $t = 1, 2, \ldots, n$; and
\[
q^t(y) = \frac{|\{j \mid v^t_j = y, j \in \{1, \ldots, N^*_2\}\}|}{N^*_2} \tag{3.32}
\]
for $y \in Y^t$, $t = 1, 2, \ldots, n$.

$p^t(\ )$ is a p.d. on $X^t$ and $q^t(\ )$ is a p.d. on $Y^t$. It follows from our definitions that
\[
r^t(z) = p^t(x) \cdot q^t(y) \text{ for all } z = (x, y) \in Z^t, t = 1, 2, \ldots, n. \tag{3.33}
\]

It follows from (3.1), (3.29), (3.33) and the definition of $\bar{\omega}$ that
\[
R(r^t, \bar{\omega}) = R(p^t, q^t) \text{ for } t = 1, \ldots, n. \tag{3.34}
\]

(3.28), (3.30) and (3.34) imply that
\[
\log N \leq \left[ \sum_{t=1}^{n} R(p^t, q^t) + 1 \right] (1 - \lambda)^{-1}. \tag{3.35}
\]

We give now upper bounds on $N^*_1$ and $N^*_2$. For every $v_j = (v^1_j, \ldots, v^n_j)$, $j = 1, \ldots, N^*_2$, define $P(\cdot \mid v_j)$ by
\[
P(\bar{x}_n \mid x_n, v_j) = \prod_{t=1}^{n} w(\bar{x}_t \mid x^t, v^t_j) \tag{3.36}
\]
for every $\bar{x}_n = (\bar{x}^1, \ldots, \bar{x}^n) \in \bar{X}_n$, $x_n = (x^1, \ldots, x^n) \in X_n$. For every $u_i = (u^1_i, \ldots, u^n_i)$, $i = 1, \ldots, N^*_1$, define $P(\cdot \mid u_i, \cdot)$ by
\[
P(\bar{x}_n \mid u_i, y_n) = \prod_{t=1}^{n} w(\bar{x}_t \mid u^t_i, y^t) \tag{3.37}
\]
for every $\bar{x}_n \in \bar{X}_n$, $y_n \in Y_n$. Let $R(p^t, \omega(\cdot \mid v^t_j))$ and $R(q^t, \omega(\cdot \mid u^t_i, \cdot))$ be defined as $R(r^t, \omega)$ in (3.29), (3.20), (3.21) and Fano’s lemma yield
\[
\log N^*_1 \leq \left( \sum_{t=1}^{n} R(p^t, \omega(\cdot \mid v^t_j)) + 1 \right) (1 - \lambda^*)^{-1} \text{ for } j = 1, \ldots, N^*_2 \tag{3.38}
\]
and
\[
\log N^* \leq \left( \sum_{i=1}^{n} R(q', \omega(i, u'_i, \cdot) + 1)(1 - \lambda^*)^{-1} \right). \tag{3.39}
\]

It follows from the definitions (3.1), (3.5), (3.31) and (3.32) that
\[
\frac{1}{N^*_2} \sum_{j=1}^{N^*_2} \sum_{i=1}^{n} R(p', \omega(\cdot, v'_j, \cdot)) = \sum_{i=1}^{n} R_1(p', q') \tag{3.40}
\]
and that
\[
\frac{1}{N^*_1} \sum_{i=1}^{N^*_1} \sum_{j=1}^{n} R(q', \omega(\cdot, u'_i, \cdot)) = \sum_{i=1}^{n} R_2(p', q'). \tag{3.41}
\]

(3.38) and (3.40) imply
\[
\log N^*_1 \leq \left( \sum_{i=1}^{n} R_1(p', q') + 1 \right)(1 - \lambda^*)^{-1}. \tag{3.42}
\]

(3.39) and (3.41) imply
\[
\log N^*_2 \leq \left( \sum_{i=1}^{n} R_2(p', q') + 1 \right)(1 - \lambda^*)^{-1}. \tag{3.43}
\]

(3.6), (3.35), (3.42) and (3.43) suffice to prove the desired result. We consider two cases

a) \[
\log N^*_1 \leq \left\lfloor \sum_{i=1}^{n} R^3_{21}(p', q') + 1 \right\rfloor (1 - \lambda^*)^{-1}. \tag{3.44}
\]

Then the desired result follows by choosing \( \lambda \) sufficiently small from (3.43) and the definitions of \( G^* \) and \( G(p, T_{21}, I) \).

b) \[
\log N^*_1 > \left\lceil \sum_{i=1}^{n} R^3_{21}(p', q') + 1 \right\rceil (1 - \lambda^*)^{-1}. \tag{3.45}
\]

This and (3.42) imply the existence of an \( \alpha, 0 \leq \alpha \leq 1 \), such that
\[
\log N^*_1 = \left\lfloor \sum_{i=1}^{n} \alpha R_1(p', q') + (1 - \alpha) R^3_{21}(p', q') + 1 \right\rfloor (1 - \lambda^*)^{-1}. \tag{3.46}
\]

It follows from (3.6) that \( R(p', q') = \alpha R_1(p', q') + \alpha R^3_{12}(p', q') + (1 - \alpha) R_2(p', q') + (1 - \alpha) R^3_{21}(p', q') \). This, (3.35) and (3.44) imply that
\[
\log N^*_2 = \log N - \log N^*_1 \leq \left\lfloor \sum_{i=1}^{n} (\alpha R^3_{12}(p', q') + (1 - \alpha) R_2(p', q')) + 1 \right\rfloor (1 - \lambda^*)^{-1}. \tag{3.45}
\]

(3.44), (3.45) and the definition of \( G^* \) yield that \( \frac{1}{n} \log N_1, \frac{1}{n} \log N_2 \) is arbitrarily close to \( G^* \) for \( \lambda \) sufficiently small and all large \( n \). This proves that \( G(P, T_{21}, I) \subset G^* \) and thus completes the proof of the theorem.
We define first several functions in \( p, q \) and \( r \), which shall play a similar role to the functions \( R(p, q), R_1(p, q), R_2(p, q), R_{12}(p, q) \) and \( R_{23}(p, q) \) in Section 3.

\[
R(p, q, r) = \sum_{x,y,z,\bar{x}} p(x)q(y)r(z)w(\bar{x} \mid x, y, z) \log \frac{w(\bar{x} \mid x, y, z)}{\sum_{x,y,z} p(x)q(y)r(z)w(\bar{x} \mid x, y, z)} 
\]  
\( (4.1) \)

\[
R_1(p, q, r) = \sum_{x,y,z,\bar{x}} p(x)q(y)r(z)w(\bar{x} \mid x, y, z) \log \frac{w(\bar{x} \mid x, y, z)}{\sum_{x} p(x)w(\bar{x} \mid x, y, z)} 
\]  
\( (4.2) \)

\[
R_2(p, q, r) = \sum_{x,y,z,\bar{x}} p(x)q(y)r(z)w(\bar{x} \mid x, y, z) \log \frac{w(\bar{x} \mid x, y, z)}{\sum_{y} q(y)w(\bar{x} \mid x, y, z)} 
\]  
\( (4.3) \)

\[
R_3(p, q, r) = \sum_{x,y,z,\bar{x}} p(x)q(y)r(z)w(\bar{x} \mid x, y, z) \log \frac{w(\bar{x} \mid x, y, z)}{\sum_{z} r(z)w(\bar{x} \mid x, y, z)} 
\]  
\( (4.4) \)

\[
R_{12}(p, q, r) = \sum_{x,y,z,\bar{x}} p(x)q(y)r(z)w(\bar{x} \mid x, y, z) \log \frac{\sum_{x} p(x)w(\bar{x} \mid x, y, z)}{\sum_{x,y} p(x)q(y)w(\bar{x} \mid x, y, z)} 
\]  
\( (4.5) \)

\[
R_{13}(p, q, r) = \sum_{x,y,z,\bar{x}} p(x)q(y)r(z)w(\bar{x} \mid x, y, z) \log \frac{\sum_{x} p(x)w(\bar{x} \mid x, y, z)}{\sum_{x,z} p(x)q(y)w(\bar{x} \mid x, y, z)} 
\]  
\( (4.6) \)

Analogously we define functions \( R_{21}^2(p, q, r) \), \( R_{23}^2(p, q, r) \), \( R_{31}^2(p, q, r) \) and \( R_{32}^2(p, q, r) \).

\[
R_{123}^2(p, q, r) = \sum_{x,y,z,\bar{x}} p(x)q(y)r(z)w(\bar{x} \mid x, y, z) \log \frac{\sum_{x,y} p(x)q(y)w(\bar{x} \mid x, y, z)}{\sum_{x,y,z} p(x)q(y)r(z)w(\bar{x} \mid x, y, z)} 
\]  
\( (4.7) \)

\( R_{132}^2(p, q, r) \) and \( R_{231}^2(p, q, r) \) are defined analogously.

For reasons of brevity we omit now the arguments \( p, q \) and \( r \). The following six identities can easily be verified:

\[
R = R_1 + R_{12}^2 + R_{13}^2 
\]  
\( (4.8) \)

\[
R = R_1 + R_{13}^2 + R_{12}^3 
\]  
\( (4.9) \)

\[
R = R_2 + R_{21}^2 + R_{13}^2 
\]  
\( (4.10) \)

\[
R = R_2 + R_{23}^2 + R_{231}^2 
\]  
\( (4.11) \)

\[
R = R_3 + R_{31}^2 + R_{132}^3 
\]  
\( (4.12) \)

\[
R = R_3 + R_{32}^3 + R_{231}^2 
\]  
\( (4.13) \)
We define the sets of triples of real numbers
\[ G_1 = \{(R_1, R_{12}, R_{123}) \mid p \text{ p. d. on } X, q \text{ p. d. on } Y, r \text{ p. d. on } Z\}, \]
\[ G_2 = \{(R_1, R_{132}, R_{13}) \mid \text{for all } p, q, \text{ and } r\}, \]
\[ G_3 = \{(R_{21}, R_2, R_{213}) \mid \text{for all } p, q, \text{ and } r\}, \]
\[ G_4 = \{(R_{231}, R_2, R_{23}) \mid \text{for all } p, q, \text{ and } r\}, \]
\[ G_5 = \{(R_{31}, R_{312}, R_3) \mid \text{for all } p, q, \text{ and } r\}, \]
\[ G_6 = \{(R_{32}, R_{323}, R_3) \mid \text{for all } p, q, \text{ and } r\}, \]
and \[ G = \bigcup_{i=1}^{6} G_i. \] We denote the convex hull of \( G \) by \( G^{**} \).

**Lemma 2**

For every triple \((p, q, r)\) the following inequalities hold:

1) \[ R_1 \geq R_{21} \geq R_{213} \]
2) \[ R_1 \geq R_{31} \geq R_{231} \]
3) \[ R_2 \geq R_{12} \geq R_{132} \]
4) \[ R_2 \geq R_{32} \geq R_{132} \]
5) \[ R_3 \geq R_{13} \geq R_{123} \]
6) \[ R_3 \geq R_{32} \geq R_{123} \].

**Proof.** It is well known that the function \( R(\pi, \omega) \), given by
\[
R(\pi, \omega) = \frac{\sum_{i=1}^{a} \sum_{j=1}^{b} \pi_i \omega(j \mid i) \log \frac{\omega(j \mid i)}{\sum_{i=1}^{a} \pi_i \omega(j \mid i)}}{(4.14)}
\]
for every probability vector \( \pi = (\pi_1, \ldots, \pi_a) \) and every stochastic matrix \((\omega(j \mid i)) \) \( i = 1, \ldots, a, j = 1, \ldots, b \) is convex in \( \omega \), that is, for two stochastic matrices \( \omega_1 \) and \( \omega_2 \) and \( \alpha = (\alpha_1, \alpha_2) \), where \( 0 \leq \alpha_1, \alpha_2 \leq 1 \) and \( \alpha_1 + \alpha_2 = 1 \), we have
\[
R(\pi, \alpha \omega_1 + \alpha_2 \omega_2) \leq \alpha_1 R(\pi, \omega_1) + \alpha_2 R(\pi, \omega_2). \quad (4.15)
\]
The inequalities stated in Lemma 2 follow by the iterated applications of inequality (4.15).

There is a close relationship between the communication situations \((P, T_{31}, \Pi)\) and \((P, T_{21}, \Pi)\). Let \( \omega(\bar{x} \mid x, y) = \omega(\bar{x} \mid x, y, z) \) for all \( x \in X, y \in Y, z \in Z, \bar{x} \in \bar{X} \). Define \( \tilde{R}(p, q), \tilde{R}_1(p, q), \tilde{R}_2(p, q), \tilde{R}_{12} (p, q), \tilde{R}_{31} (p, q) \) and \( \tilde{G}^{*} \) for \( \omega(\cdot \mid \cdot \cdot \cdot) \) as \( R(p, q), R_1(p, q), \) etc. in Section 2. Finally, define \( G^{***} \) as the convex hull of \( \bigcup z \in \bar{Z} G^* \). We shall prove in paragraph 4 that \( G^{***} = G(P, T_{31}, \Pi) \).
Lemma 3

Let \( L_s, s = 1, \ldots, d \) be non-negative random variables, defined on the same probability space, such that \( E L_s \leq \alpha, s = 1, \ldots, d \). For any \( \varepsilon > 0 \) the probability of \( B^* = \{ L_s \leq d (\alpha + \varepsilon) \text{ for } s = 1, \ldots, d \} \) satisfies
\[
P(B^*) \geq \frac{\varepsilon}{\alpha + \varepsilon}.
\]

This is a trivial refinement of the Lemma in [8]. For a proof see [2, p. 467].

2. The main results

Theorem 2

a) The capacity region \( G(P, T_{31}, I) \) for \( (P, T_{31}, I) \) equals \( G^{**} \).

b) \( G(P, T_{31}, I) \) is a closed convex set in Euclidean 3-space and contains with every point \((R_1, R_2, R_3)\) also the projections \((R_1, 0, 0)\), \((0, R_2, 0)\), \((0, 0, R_3)\), and \((0, 0, 0)\).

Theorem 3

a) The capacity region \( G(P, T_{31}, II) \) for \( (P, T_{31}, I) \) equals \( G^{***} \).

b) \( G(P, T_{31}, II) \) is a closed convex set in Euclidean 3-space and contains with every point \((R_1, R_2, R_3)\) also the four projections \((R_1, 0, 0)\), \((0, R_2, 0)\), \((0, 0, R_3)\), and \((0, 0, 0)\).

3. The proof for \( G^{**} \subset G(P, T_{31}, I) \)

It can be shown by the arguments used in Section 3 that \( G^{**} \) is closed under projections.

In order to show that \( G^{**} \subset G(P, T_{31}, I) \) it suffices to prove that for any triple \((p, q, r) \subset (R_1(p, q, r), R_{12}(p, q, r), R_{123}(p, q, r)) \subset G(P, T_{31}, I)\).

The proof that \( G_i \) \((i = 2, \ldots, 6) \subset G(P, T_{31}, I) \) is symmetrically the same. Once we know that \( G \subset G(P, T_{31}, I) \) we can show that \( G^{**} \subset G(P, T_{31}, I) \) by concatenation.

Let \( \hat{u} = (u_1, \ldots, u_{N_3}) \) and \( \hat{v} = (v_1, \ldots, v_{N_3}) \) be defined as in Section 3, and let \( w = (w_1, \ldots, w_{N_3}) \) be a vector with \( N_3 \) components, which are elements of \( Z_n \). Let \( p_n \) be a p. d. on \( X_n \), \( q_n \) be a p. d. on \( Y_n \), and let \( r_n \) be a p. d. on \( Z_n \). Define \( \hat{p}_n \) and \( \hat{q}_n \) as in (3.7), (3.8) and define \( \hat{r}_n \) by
\[
\hat{r}_n(\hat{w}) = \prod_{k=1}^{N_3} r_n(w_k)
\]  

(4.16)

for all \( w_k \in Z_n, k = 1, \ldots, N_3 \).

Define for every \( z_n \in Z_n, \bar{x}_n \in X_n \) \( P(\bar{x}_n \mid z_n) \) by
\[
P(\bar{x} \mid z_n) = \sum_{x_n \in X_n} \sum_{y_n \in Y_n} p_n(x_n) q_n(y_n) P(\bar{x}_n \mid x_n, y_n, z_n)
\]  

(4.17)
and define for every \( z_n \in Z_n, y_n \in Y_n, \bar{x}_n \in \bar{X}_n \) \( P(\bar{x}_n \mid y_n, z_n) \) by

\[
P(\bar{x}_n \mid y_n, z_n) = \sum_{x_n \in X_n} p_n(x_n) P(\bar{x}_n \mid x_n, y_n, z_n).
\] (4.18)

We define for \( \hat{w} \) the decoding sets \( C_1, \ldots, C_{N_3} \) by

\[
C_k = \{ \bar{x}_n \mid P(\bar{x}_n \mid w_k) > P(\bar{x}_n \mid w_l) \text{ for } l \neq k \}
\] (4.19)

for \( k = 1, \ldots, N_3 \).

For the pair \((\hat{v}, \hat{w})\) we define the decoding sets \( B_{jk}^* (j = 1, \ldots, N_2; k = 1, \ldots, N_3) \) by

\[
B_{jk}^* = \{ \bar{x}_n \mid P(\bar{x}_n \mid v_j, w_k) > P(\bar{x}_n \mid v_m, w_k) \text{ for } m \neq j \}
\] (4.20)

for \( j = 1, \ldots, N_2; k = 1, \ldots, N_3 \).

Obviously, for every fixed \( k \), \( B_{jk}^* \cap B_{j'k}^* = \emptyset \) for \( j \neq j' \).

For the triple \((\hat{u}, \hat{v}, \hat{w})\) we define the decoding set \( A_{ijk}^* \) by

\[
A_{ijk}^* = \{ \bar{x}_n \mid P(\bar{x}_n \mid u_i, v_j, w_k) > P(\bar{x}_n \mid u_l, v_j, w_k) \text{ for } l \neq i \}
\] (4.21)

for \( i = 1, \ldots, N_1; j = 1, \ldots, N_2; k = 1, \ldots, N_3 \).

We have for every fixed pair \((j, k)\) that \( A_{ijk}^* \cap A_{i'j'k'}^* = \emptyset \) for \( i' \neq i \).

Define — for all index constellations — \( A_{ijk} \) by

\[
A_{ijk} = A_{ijk}^* \cap B_{jk}^* \cap C_k.
\] (4.22)

It follows from our definitions that \( A_{ijk} \cap A_{i'j'k'} = \emptyset \) for \((i, j, k) \neq (i', j', k')\).

Let \( A_{ijk}^c \) be the complement of \( A_{ijk} \).

The average error for the code \( \{(u_i, v_j, w_k, A_{ijk}) \mid (i = 1, \ldots, N_1; j = 1, \ldots, N_2; k = 1, \ldots, N_3)\} \) is given by

\[
\lambda(\hat{u}, \hat{v}, \hat{w}) = \frac{1}{N_1 N_2 N_3} \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} \sum_{k=1}^{N_3} P(A_{ijk}^c \mid u_i, v_j, w_k).
\] (4.23)

We want to give an upper bound on \( l \), given by

\[
l = \sum_{\hat{u}, \hat{v}, \hat{w}} \hat{p}_n(\hat{u}) \hat{q}_n(\hat{v}) \hat{r}_n(\hat{w}) \lambda(\hat{u}, \hat{v}, \hat{w}).
\] (4.24)

It follows from (4.22) that \( A_{ijk}^c = A_{ijk}^* \cup B_{jk}^{c*} \cup C_k^c \). This and (4.23) imply

\[
\lambda(\hat{u}, \hat{v}, \hat{w}) \leq \frac{1}{N_1 N_2 N_3} \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} \sum_{k=1}^{N_3} P(A_{ijk}^c \mid u_i, v_j, w_k) + P(B_{jk}^{c*} \mid u_i, v_j, w_k)
\]

\[
+ P(C_k^c \mid u_i, v_j, w_k).
\] (4.25)
It follows from (4.25) that
\[
I \leq \sum_{\hat{u}, \hat{v}} \hat{p}_n(\hat{u}) \hat{q}_n(\hat{v}) \hat{r}_n(\hat{w}) \cdot \frac{1}{N_1 N_2 N_3} \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} \sum_{k=1}^{N_3} P(A^*_{ijk} \mid u_i, v_j, w_k) \\
+ \sum_{\hat{r}, \hat{w}} \hat{q}_n(\hat{v}) \hat{r}_n(\hat{w}) \frac{1}{N_1 N_2} \sum_{j=1}^{N_2} \sum_{k=1}^{N_3} P(B^*_{jk} \mid v_j, w_k) \\
+ \sum_{\hat{w}} \hat{r}_n(\hat{w}) \frac{1}{N_3} \sum_{k=1}^{N_3} P(C^*_k \mid W_k).
\]

(4.26)

We choose now \(p_n = px \ldots xp, q_n = qx \ldots qx, r_n = rx \ldots rx\) and we denote in the sum above the first term by \(l_1\), the second term by \(l_2\) and the third term by \(l_3\). \(l_1\) is an error term which occurs by applying the random coding method to Shannon's two-way channels in case \(Y \times Z\) serves as one input alphabet. As in Section 3, we obtain the estimate:

for any \(\delta > 0\) we can choose \(N_1 = \exp \{R_1(p, q, r) - \delta n\}\) and we can find an \(\epsilon > 0\), such that \(l_1 = l_1(N_1) \leq e^{-\epsilon n}\) for all sufficiently large \(n\). \(l_2\) can be treated as the first term in (3.16). We obtain the result:

for any \(\delta > 0\), we can choose \(N_2 = \exp \{R_2^*(p, q, r) - \delta n\}\) and we can find an \(\epsilon > 0\), such that \(l_2 = l_2(N_2) \leq e^{-\epsilon n}\) for all sufficiently large \(n\). \(l_3\) can be treated as the second term in (3.16). Thus we have:

for any \(\delta > 0\), we can choose \(N_3 = \exp \{R_3^*(p, q, r) - \delta n\}\) and we can find an \(\epsilon > 0\), such that \(l_3 = l_3(N_3) \leq e^{-\epsilon n}\) for all large \(n\).

The existence of a pure code follows by means of Lemma 3 as in [8]. We thus have proved that \(G^{**} \subset G(P, T_{31}, I)\).

4. Results which are needed for proving \(G^{**} \supset G(P, T_{31}, I)\)

Let \(\{(u_i, v_j, w_k, A_{ijk}) \mid (i = 1, \ldots, N_1; j = 1, \ldots, N_2; k' = 1, \ldots, N_3)\}\)
be an \((n, N_1, N_2, N_3, \lambda)\) code for \((P, T_{31}, I)\). From
\[
\frac{1}{N_1 N_2 N_3} \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} \sum_{k=1}^{N_3} P(A^*_{ijk} \mid u_i, v_j, w_k) \leq \lambda
\]

(4.27)

we obtain — by a similar conclusion as the one which led to (3.20), (3.21) and (3.22) — that we can pick \(u_i\)’s, \(v_j\)’s and \(w_k\)’s such that after renumbering:
\[
\frac{1}{N_2 N_3} \sum_{j=1}^{N_2} \sum_{k=1}^{N_3} P(A^*_{ijk} \mid u_i, v_j, w_k) \leq 8\lambda \text{ for } i = 1, \ldots, N_1^*,
\]

(4.28)
\[
\frac{1}{N_1 N_3^*} \sum_{i=1}^{N_1} \sum_{k=1}^{N_3^*} P(A^*_{ijk} \mid u_i, v_j, w_k) \leq 8\lambda \text{ for } j = 1, \ldots, N_2^*,
\]

(4.29)
\[
\frac{1}{N_1^* N_2^*} \sum_{i=1}^{N_1^*} \sum_{j=1}^{N_2^*} P(A^*_{ijk} \mid u_i, v_j, w_k) \leq 8\lambda \text{ for } k = 1, \ldots, N_3^*,
\]

(4.30)
and (as a consequence)
\[
\frac{1}{N_1^* N_2^* N_3^*} \sum_{i=1}^{N_1^*} \sum_{j=1}^{N_2^*} \sum_{k=1}^{N_3^*} P(A_{ijk} \mid u_i, v_j, w_k) \leq 8\lambda. \tag{4.31}
\]

\(N_1^*, N_2^*,\) and \(N_3^*\) can be chosen as \(\left[ \frac{N_1}{2} \right], \left[ \frac{N_2}{2} \right]\) and \(\left[ \frac{N_3}{2} \right]\). We write \(\lambda^*\) instead of \(8\lambda\).

We establish now upper bounds on \(K_{12} = (1 - \lambda^*) \log (N_1^* N_2^*) - 1,\)
\(K_{13} = (1 - \lambda^*) \log (N_1^* N_3^*) - 1,\)
\(K_{23} = (1 - \lambda^*) \log (N_2^* N_3^*) - 1,\)
\(K_{123} = (1 - \lambda^*) \log (N_1^* N_2^* N_3^*) - 1,\)
\(K_1 = (1 - \lambda^*) \log N_1^* - 1,\)
\(K_2 = (1 - \lambda^*) \log N_2^* - 1\) and \(K_3 = (1 - \lambda^*) \log N_3^* - 1.\)

We derive first an upper bound on \(K_{12}.\) For this purpose we introduce for every \(w_k(k = 1, \ldots, N_2^*)\) a nonstationary discrete memoryless channel \(P_k\) with input alphabet \(X \times Y,\) output alphabet \(X\) and with transmission probabilities given by
\[
P(\bar{x}_n \mid x_n, y_n, w_k) = \prod_{i=1}^{n} \omega(\bar{x}^i \mid x^i, y^i, w^i_k) \tag{4.32}
\]
for every \(\bar{x}_n \in \bar{X}_n, x_n \in X_n\) and \(y_n \in Y_n.\)

\((u_i, v_j, A_{ijk}) \mid (i = 1, \ldots, N_1^*; j = 1, \ldots, N_2^*)\) is because of (4.30) an \((n, N_1^* \cdot N_2^*, \lambda^*)\) code for \(P_k.\)

Define \(p'(\cdot)\) and \(q'(\cdot)\) as in (3.31) and (3.32). We define \(r'(\cdot)\) by
\[
r'(z) = \frac{\left| \left\{ k \mid w_k^i = z, k \in \{1, \ldots, N_2^*\} \right\} \right|}{N_2^*} \tag{4.33}
\]
for \(z \in Z', t = 1, 2, \ldots, n.\) \(r'(\cdot)\) is a p. d. on \(z'.\)

Fano's lemma leads to the estimate:
\[
K_{12} \leq \left[ \sum_{i=1}^{n} \sum_{x, y, \bar{x}} p'(x) q'(y) w(\bar{x} \mid x, y, w_k^i) \cdot \log \frac{w(\bar{x} \mid x, y, w_k^i)}{\sum_{x, y} w(\bar{x} \mid x, y, w_k^i)} \right] \tag{4.34}
\]
for all \(k = 1, \ldots, N_2^*.\)

It follows from (4.34) and the definition of \(r'(\cdot)\) that
\[
K_{12} \leq \left[ \sum_{i=1}^{n} \sum_{x, y, \bar{x}} p'(x) q'(y) r'(z) w(\bar{x} \mid x, y, z) \log \frac{w(\bar{x} \mid x, y, z)}{\sum_{x, y} p'(x) q'(y) w(\bar{x} \mid x, y, z)} \right]. \tag{4.35}
\]

The right side of this inequality is equal to
\[
\sum_{i=1}^{n} [R_1(p', q', r') + R_{12}(p', q', r')] = \sum_{i=1}^{n} [R_2(p', q', r') + R_{21}(p', q', r')]. \tag{4.36}
\]
Similarly one can show that
\[ K_{13} \leq \sum_{i=1}^{n} [R_{1}(p', q', r') + R_{13}^{1}(p', q', r')] = \sum_{i=1}^{n} [R_{3}(p', q', r') + R_{31}^{3}(p', q', r')] \]
and that
\[ K_{23} \leq \sum_{i=1}^{n} [R_{2}(p', q', r') + R_{23}^{2}(p', q', r')] = \sum_{i=1}^{n} [R_{3}(p', q', r') + R_{32}^{3}(p', q', r')] . \]
(4.37) (4.38)

We give now an upper bound on \( K_{123} \). Let us consider the discrete memoryless channel \( P^{**} \) with transmission matrix \( \omega^{**} \) given by
\[ \omega^{**} (\bar{x} | z^*) = w(\bar{x} | x, y, z) \] for all \( z^* = (x, y, z) \in X \times Y \times Z \) and all \( \bar{x} \in \bar{X} \).
(4.39)

The results obtained for the channel \( P^{*} \) in Section 3 immediately generalize to the channel \( P^{**} \). Instead of inequality (3.35) we obtain now
\[ K_{123} \leq \sum_{i=1}^{n} R(p', q', r') . \]
(4.40)

In trying to give nontrivial upper bounds for \( K_{1}, K_{2} \) and \( K_{3} \) we are facing a difficulty, which is not present for the cases above or in the corresponding problems in Section 3. This difficulty comes from the fact that it is in general impossible to select \( u_i \)'s, \( v_j \)'s and \( w_k \)'s such that — for instance —
\[ \frac{1}{N^{*}} \sum_{i=1}^{N^{*}} P(A_{ijk} | u_i, v_j, w_k) \leq \lambda^{**} \]
for \( j = 1, \ldots, N^{*}_{2}; k = 1, \ldots, N^{*}_{3} \), where \( \lambda^{**} \) tends to 0 as \( \lambda \) tends to 0 and \( N^{**}_{i} \) equals \( cN_{i} \) for \( i = 1, 2, 3; c \) is a constant.

This follows from an asymptotic estimate concerning a problem of Zarkiewicz which we derived in [1]. The discussion in Section 5 of [1] is also valid in the present situation. We can resolve the difficulty here by an approximation argument.

Let \( (\Omega, m) \) be a probability space, where
\[ \Omega = \{v_1, \ldots, v^{*}_{N_2}\} \times \{w_1, \ldots, w^{*}_{N_3}\} \]
and \( m \) is the equaldistribution on \( \Omega \). Define a random variable \( L_{1} \) by
\[ L_{1}(v_j, w_k) = \frac{1}{N^{*}_{1}} \sum_{i=1}^{N^{*}_{1}} P(A_{ijk} | u_i, v_j, w_k) \] for all \( (v_j, w_k) \in \Omega \).
(4.41)

(4.31) implies that the expectation \( EL_{1} \leq \lambda^{*} \). Applying Lemma 3 for \( d = 1 \) we obtain that there is a subset \( B^{*} \) of \( \Omega \) such that
\[ L_{1}(v_j, w_k) = \frac{1}{N^{*}_{1}} \sum_{i=1}^{N^{*}_{1}} P(A_{ijk}^{*} | u_i, v_j, w_k) \leq \lambda^{*} + \varepsilon \] for \( (u_j, w_k) \in B^{*} \).
(4.42)
and
\[ |B^*| \geq \frac{\epsilon}{\lambda^* + \epsilon} N_2^* N_3^*. \] (4.43)

We derive now an upper bound on \( K_1 = K_1(\lambda^*) \). We introduce for every \((v_j, w_k) \in B^*\) a nonstationary discrete memoryless channel \( P_{jk} \) with input alphabet \( \bar{X} \), output alphabet \( \bar{Y} \) and with transmission probabilities given by
\[ P(\bar{x}_n | x_n, v_j, w_k) = \prod_{i=1}^{n} \omega(\bar{x}_i | x'_i, v'_j, w'_k) \quad \text{for every } \bar{x}_n \in \bar{X}_n \text{ and } x_n \in X_n. \] (4.44)

\[ \{(u_i, A_{ijk}) \mid i = 1, \ldots, N_1^*\} \] is because of (4.42) an \((n, N_1^*, \lambda^* + \epsilon)\) code for \( P_{jk} \). (4.42), (4.44) and Fano's lemma yield
\[ \log N_1^* \leq \left[ \sum_{i=1}^{n} R(p^i, \omega(\cdot | \cdot, v_j, w_k)) + 1 \right] (1 - \lambda^* - \epsilon)^{-1} \text{ for } (v_j, w_k) \in B^*. \] (4.45)

If (4.45) held for all \((v_j, w_k) \in \Omega\), then we would obtain that
\[ \log N_1^* \leq \frac{1}{N_2^* N_3^*} \left[ \sum_{i=1}^{n} R(p^i, \omega(\cdot | \cdot, v_j, w_k)) + 1 \right] (1 - \lambda^* - \epsilon)^{-1}. \]

The right side of this inequality is because of the definitions of \( p^i(\cdot), q^i(\cdot), r^i(\cdot) \) and definition (4.2) equal to
\[ \left[ \sum_{i=1}^{n} R_1(p^i, q^i, r^i) + 1 \right] (1 - \lambda^* - \epsilon)^{-1}. \]

Since (4.42) holds only for elements of \( B^* \) we obtain
\[ \log N_1^* \leq \left[ \sum_{i=1}^{n} R_1(p^i, q^i, r^i) + 1 \right] (1 - \lambda^* - \epsilon)^{-1} + E_n(\lambda^*, \epsilon), \] (4.46)

where \( E_n \) is an "error" term, which we bound now from above. Since \( \log N_1^* \leq n \log a \cdot (1 - \lambda^* - \epsilon)^{-1} \) and because of (4.43) we obtain
\[ E_n(\lambda^*, \epsilon) \leq \frac{\lambda^*}{\lambda^* + \epsilon} (1 - \lambda^* - \epsilon)^{-1} n \cdot \log a. \] (4.47)

Since \( \lim_{\lambda^* \to 0} \frac{1}{n} E_n(\lambda^*, \epsilon) = 0 \), uniformly in \( n \), and since we are interested

only in results for \( \lambda^* \) arbitrarily close to 0 we ignore the term \( E_n(\lambda^*, \epsilon) \) completely. Since \((1 - \lambda^*)(1 - \lambda^* - \epsilon)^{-1}\) also tends to 1 as \( \lambda^* \) and \( \epsilon \) tends to 0, we just write for convenience
\[ K_1 \leq \sum_{i=1}^{n} R_1(p^i, q^i, r^i). \] (4.48)
The proof that
\[ K_2 \leq \sum_{i=1}^{n} R_2(p^i, q^i, r^i) \]  \hspace{1cm} (4.49)
and
\[ K_3 \leq \sum_{i=1}^{n} R_3(p^i, q^i, r^i) \]  \hspace{1cm} (4.50)
is symmetrically the same.

We introduce now a convenient notation. We shall write \( \bar{R}_1, \bar{R}_2, \bar{R}_3, \bar{R}_{12} \) etc., where for instance \( \bar{R}_{12} \) is given by
\[ \bar{R}_{12} = \frac{1}{n} \sum_{i=1}^{n} R_{12}(p^i, q^i, r^i); \bar{R} \text{ stands for } \frac{1}{n} \sum_{i=1}^{n} R(p^i, q^i, r^i). \]

We also introduce \( \bar{K}_1, \bar{K}_2, \bar{K}_3, \bar{K}_{12} \) etc., where for instance \( \bar{K}_{12} = \frac{1}{n} K_{12} \).

Using this notation we summarize our results [(4.35), (4.36), (4.37), (4.38), (4.40), (4.48), (4.49) and (4.50)] in

**Lemma 4**

1) \( \bar{K}_1 \leq \bar{R}_1 \)
2) \( \bar{K}_2 \leq \bar{R}_2 \)
3) \( \bar{K}_3 \leq \bar{R}_3 \)
4) \( \bar{K}_{12} \leq \bar{R}_1 + \bar{R}_{12} = \bar{R}_2 + \bar{R}_{21} \)
5) \( \bar{K}_{13} \leq \bar{R}_1 + \bar{R}_{13} = \bar{R}_3 + \bar{R}_{31} \)
6) \( \bar{K}_{23} \leq \bar{R}_2 + \bar{R}_{23} = \bar{R}_3 + \bar{R}_{32} \)
7) \( \bar{K}_{123} \leq \bar{R} \)

**5. The proof for \( G^{**} \supset G(P, T_{31}, I) \)**

From (4.8)–(4.13) we obtain the system of equations
\[ \begin{align*}
\bar{R} &= \bar{R}_1 + \bar{R}_{12} + \bar{R}_{123} \\
\bar{R} &= \bar{R}_1 + \bar{R}_{13} + \bar{R}_{132} \\
\bar{R} &= \bar{R}_{21} + \bar{R}_2 + \bar{R}_{123} \\
\bar{R} &= \bar{R}_{23} + \bar{R}_2 + \bar{R}_{23} \\
\bar{R} &= \bar{R}_{31} + \bar{R}_3 + \bar{R}_{132} \\
\bar{R} &= \bar{R}_{31} + \bar{R}_{132} + \bar{R}_3.
\end{align*} \]  \hspace{1cm} (4.51)
We also make use of the matrix

\[ \mathbf{R} = \begin{pmatrix}
R_1 & R_{12} & R_{13}
R_1 & R_{12} & R_{13}
R_{21} & R_2 & R_{123}
R_{21} & R_2 & R_{13}
R_{31} & R_{13} & R_3
R_{31} & R_{13} & R_3
\end{pmatrix} \]

It follows from our definitions that the sth row vector of \( \mathbf{R} \) is contained in \( G^s \), the convex hull of \( G_s \), for \( s = 1, \ldots, 6 \). We shall prove now that \((\bar{K}_1, \bar{K}_2, \bar{K}_3) - \) and therefore also \( \left( \frac{1}{n} \log N_1, \frac{1}{n} \log N_2, \frac{1}{n} \log N_3 \right) \) - is contained in an arbitrary \( \varepsilon \) neighbourhood of \( G^{**} \) for \( \lambda \) sufficiently small.

We consider now two cases: a) \( \bar{K}_1 = \bar{R}_1 \) and b) \( \bar{K}_1 < \bar{R}_1 \). We begin with

a) \( \bar{K}_1 = \bar{R}_1 \).

Lemma 4, 7) and (4.51) imply

\[ R_{12} + R_{13}^1 \geq \bar{K}_2 + \bar{K}_3 \] (4.52)

and

\[ R_{13} + R_{13}^1 \geq \bar{K}_2 + \bar{K}_3 \] (4.53)

We can assume that for some \( \alpha \), \( 0 \leq \alpha \leq 1 \), \( \alpha \, R_{12} + (1 - \alpha) \, R_{13}^1 \geq \bar{K}_2 \).

(Otherwise we obtain a similar inequality for \( \bar{K}_3 \).) If there exists an \( \alpha_0 \), \( 0 \leq \alpha_0 \leq 1 \), such that actually

\[ \alpha_0 \, R_{12} + (1 - \alpha_0) \, R_{13}^1 = \bar{K}_2, \]

then we obtain from (4.52) and (4.53) that

\[ \bar{K}_3 \leq \alpha_0 \, R_{13}^1 + (1 - \alpha_0) \, R_{13}^1. \]

The result is proved in this case.

If such an \( \alpha_0 \) does not exist, then \( R_{12} \geq R_{13}^1 \) (Lemma 2) implies \( R_{13}^1 \geq \bar{K}_2 \). Since \( \bar{K}_1 + \bar{K}_3 = \bar{K}_{13} \leq \bar{R}_1 + \bar{R}_{13} \) (Lemma 4) and \( \bar{K}_1 = \bar{R}_1 \), we also have \( \bar{K}_3 \leq \bar{R}_{13} \). Thus

\[ (\bar{K}_1, \bar{K}_2, \bar{K}_3) \leq (\bar{R}_1, \bar{R}_{13}, \bar{R}_{13}) \in G_2 \subset G^{**}. \]

Since \( G^{**} \) is convex and contains the projections, we also obtain the desired result in this case.

b) \( \bar{K}_1 < \bar{R}_1 \).

We can assume that

\[ \bar{K}_1 > R_{31}^2, \] (4.54)
because otherwise we could conclude — using 2), 3) and 6) of Lemma 4 — that a suitable convex combination of row vectors 4 and 5 in $\overline{R}$ gives the solution.

$\overline{K}_1 < \overline{R}_1$ and (4.54) imply that there exists a $\beta, 0 < \beta < 1$, such that

$$\overline{K}_1 = \beta \overline{R}_1 + (1-\beta) \overline{R}_{231}^2 .$$

(4.55)

Choose $\beta_1 = \frac{\beta}{2}, \beta_2 = \frac{\beta}{2}, \beta_3 = 0, \beta_4 = \frac{1-\beta}{2}, \beta_5 = \frac{1-\beta}{2}, \beta_6 = 0$. We obtain from Lemma 4, 7) and (4.51) that

$$\beta_1 \overline{R}_{12} + \beta_2 \overline{R}_{132}^1 + \beta_4 \overline{R}_2 + \beta_5 \overline{R}_{32}^3 + \beta_1 \overline{R}_{123}^1 + \beta_2 \overline{R}_{13}$$

$$+ \beta_4 \overline{R}_{23} + \beta_5 \overline{R}_3 \geq \overline{K}_2 + \overline{K}_3 = \overline{K}_{123} - \overline{K}_1 .$$

(4.56)

For symmetry reasons we can assume w.l.o.g. that

$$\beta_1 \overline{R}_{12} + \beta_2 \overline{R}_{132}^1 + \beta_4 \overline{R}_2 + \beta_5 \overline{R}_{32}^3 \geq \overline{K}_2 .$$

(4.57)

Decrease $\beta_1$ and increase $\beta_2$, but such that their sum remains constant.

Since $\overline{R}_{12}^1 \geq \overline{R}_{132}^1$, the left side in (4.57) decreases under this operation.

Let $\beta'_1 = \beta_1 - \gamma, \beta'_2 = \beta_2 + \gamma$ for $0 \leq \gamma \leq \beta_1$. If for some $\gamma_0, \beta'_1 \overline{R}_{12}^1 + \beta'_2 \overline{R}_{132}^1 + \beta'_4 \overline{R}_2 + \beta'_5 \overline{R}_{32}^3 = \overline{K}_2$, then a convex combination of the row-vectors in $\overline{R}$ with coefficients $(\beta'_1, \beta'_2, 0, \beta'_4, \beta'_5, 0)$ is componentwise not smaller than $(\overline{K}_1, \overline{K}_2, \overline{K}_3)$. The result follows in this case.

If such a $\gamma_0$ does not exist, then we have

$$\beta \overline{R}_{132}^1 + \beta_4 \overline{R}_2 + \beta_5 \overline{R}_{32}^3 \geq \overline{K}_2 .$$

(4.58)

Now we apply the same trick once more. We replace $\beta_4$ by $\beta'_4 = \beta_4 - \delta$ and $\beta_5$ by $\beta'_5 = \beta_5 + \delta$ for some $\delta, 0 \leq \delta \leq \beta_4$. Since $\overline{R}_2 \geq \overline{R}_{32}^3$, the left side in (4.58) decreases under this replacement. If we can achieve equality, then we are finished as usual. Otherwise we have the following situation:

$$\overline{K}_1 = \beta \overline{R}_1 + (1-\beta) \overline{R}_{231}^2$$

$$\overline{K}_2 < \beta \overline{R}_{132}^1 + (1-\beta) \overline{R}_{32}^3 .$$

We can assume that

$$\overline{K}_2 > \overline{R}_{132}^1 .$$

(4.59)

because otherwise we could conclude — using 1), 3) and 5) of Lemma 4 — that a suitable convex combination of row vectors 2 and 6 in $\overline{R}$ gives the solution. (The argument is the same as the one given in (4.54).) We are left with

$$\overline{K}_2 < \overline{R}_{32}^3 .$$

(4.60)

Define now new coefficients $\alpha_2 = \beta, \alpha_5 = 1 - \beta - \eta, \alpha_6 = \eta$, where $0 < \eta < 1 - \beta$. 

REFERENCES


Wolfowitz, J., Two-way channels without feedback. Notes of a lecture given at Cornell University.