THE STRUCTURE OF CAPACITY FUNCTIONS FOR COMPOUND CHANNELS[1)]

R. Ahlswede and J. Wolfowitz
Ohio State University, Columbus, Ohio, USA
and
Cornell University, Ithaca, N.Y., USA

1. Definitions and introduction of the capacity functions $C(\bar{\lambda})$, $C(\lambda_R)$, $C(\bar{\lambda}_R)$.

Let $X = \{1,\dots,a\}$ and $Y = \{1,\dots,b\}$ be, respectively, the input and output alphabets which will be used for transmission over a channel (or a system of channels). Any sequence of n letters $x_n = (x^1,\dots,x^n) \in \prod_1^n X$ is called a transmitted or sent n-sequence, any sequence $y_n = (y^1,\dots,y^n) \in \prod_1^n Y$ is called a received n-sequence.

Let $S = \{1,\dots,k\}$, and

$$\mathcal{C} = \{w(\cdot|\cdot|s)\,|\,s \in S\},$$

where each $w(\cdot|\cdot|s)$ is an $(a\times b)$ stochastic matrix, also called a channel probability function (c.p.f.). For each $x_n = (x^1,\dots,x^n) \in X_n = \prod_1^n X$ we define a probability distribution (p.d.) on $Y_n = \prod_1^n Y$ by $P_n(y_n|x_n|s) = \prod_{t=1}^n w(y^t|x^t|s)$, $(y_n \in Y_n)$. $P_n(y_n|x_n|s)$ is the probability that, when the n-sequence x_n is sent, the (chance) sequence received is y_n. The sequence $(P_n(\cdot|\cdot|s))$ $n = 1,2,\dots$ describes a discrete channel without memory (d.m.c.).

1) Research of both authors supported by the U.S. Air Force under Grant AF-AFOSR-68-1472 to Cornell University.

Thus we have assigned to each $s \in S$ a d.m.c. We call the system of channels

$$\mathcal{C}^* = \{(P_n(\cdot|\cdot|s)),\ n = 1,2,..|s \in S\}$$

a <u>compound</u> (or simultaneous) channel (cf. [6]), if the transmission is governed as follows: each n-sequence x_n is transmitted according to some channel in $\mathcal{C}^*$ and the channel may vary <u>arbitrarily</u> in $\mathcal{C}^*$ from one such n-sequence to another.

We define a code (n,N,λ) for the compound channel as a system

$$\{(u_i,A_i)|u_i \in X_n,\ A_i \subset Y_n,\ A_i \cap A_j = \emptyset \text{ for } i \neq j,\ i = 1,\dots,N\}$$

which satisfies

$$P_n(A_i|u_i|s) \geq 1 - \lambda,\ i = 1,\dots N;\ s \in S.$$

As usual the entropy of a probability vector $\pi = (\pi_1,\dots,\pi_t)$ is defined to be $H(\pi) = -\sum_{i=1}^{t} \pi_i \log_2 \pi_i$. Denote the rate for the (row) probability vector π on X and c.p.f. $w(\cdot|\cdot|s)$ by $R(\pi,s) = H(\pi'(s)) - \sum_i \pi_i H(w(\cdot|i|s)$, where $\pi'(s) = \pi \cdot w(\cdot|\cdot|s)$. Let $N(n,\lambda)$ be the maximal length of an (n,N,λ) code for $\mathcal{C}^*$. It is an easy consequence of Theorem 1 in [4], that

$$\text{(1.1)} \qquad \lim \frac{1}{n} \log N(n,\lambda) = C$$

where C is a constant, independent of λ, given by

$$C = \max_{\pi} \inf_{s \in S} R(\pi,s).$$

(1.1) means that the coding theorem and strong converse of the coding theorem hold. C is called the capacity.

A code $(n, N, \bar{\lambda})$ with average error $\bar{\lambda}$ is a system

$$\{(u_i,A_i)\,|\,u_i \in X_n,\ A_i \subset Y_n, A_i \cap A_j = \phi \ \text{ for } i \neq j,\ i = 1,\ldots,N\}$$

which satisfies

$$\frac{1}{N}\sum_{i=1}^{N} P_n(A_i|u_i|s) \geq 1 - \bar{\lambda} \quad , \qquad s \in S.$$

Let $N(n,\bar{\lambda})$ be the maximal length of an $(n,N,\bar{\lambda})$ code for $\mathcal{C}^*$. It was proved in [3], that

$$\inf_{\bar{\lambda}>0} \lim_{n\to\infty} \frac{1}{n} \log N(n,\bar{\lambda}) = C.$$

(The coding theorem and weak converse for average error.)

When $|S| = 1$ it is immaterial whether we use maximal or average error (cf. [6], Ch. 3.1, Lemma 3.11). This has led to the belief - widespread among engineers - that this is true even for more complex channel systems. However, already for compound channels with $|S| = 2$ one has to distinguish carefully between these errors, as was shown in [1], example 1. In fact,

$$\underline{\lim}_{n\to\infty} \frac{1}{n} \log N(n, \bar{\lambda})$$

is in general greater than C. This means that, when we use average errors for codes for $\mathcal{C}^*$, we can achieve longer code lengths. The following questions are therefore of interest:

1) For which $\bar{\lambda}$ does $\lim_{n\to\infty} \frac{1}{n} \log N(n,\bar{\lambda})$ exist?

2) What can we say about the capacity function $C(\bar{\lambda})$, where

$$C(\bar{\lambda}) = \lim_{n\to\infty} \frac{1}{n} \log N(n,\bar{\lambda})$$

whenever the latter exists?

3) When $C(\bar{\lambda}) > C$, which encoding procedure gives the longest codes?

We shall also study channel $\mathcal{C}^*$ under randomized encoding. A random code (n,N,λ_R) is a system of pairs

$$\{(p^i,A_i) | p^i \text{ p.d. on } X_n, A_i \text{ disjoint}, \ i = 1,\dots,N\}$$

which satisfy

$$(1.2) \qquad \sum_{x_n \in X_n} p^i(x_n)\, P_n(A_i|x_n|s) \geq 1 - \lambda_R \qquad (i = 1,\dots,N)\ .$$

If we allow average error instead of maximal error we have to replace (1.2) by

$$(1.3) \qquad \frac{1}{N} \sum_{i=1}^{N} \sum_{x_n \in X_n} p^i(x_n)\, P_n(A_i|x_n|s) \geq 1 - \bar{\lambda}_R$$

in order to define a random (randomized) $(n,N,\bar{\lambda}_R)$ code.

The use of a random code is as follows: A set of messages $N = \{1,\dots N\}$ is given in advance. If messages i is to be sent the sender performs a random experiment according to p^i, and the outcome of the experiment is sent. The receiver, after receiving the n-sequence $y_n \in A_j$, decides that message j was intended. [This code concept was described in [2] under 2.1].

Questions of interest to us are:

1) For which values of λ_R, $\overline{\lambda}_R$ does $\lim_{n\to\infty} \frac{1}{n} \log N(n,\lambda_R)$, respectively $\lim_{n\to\infty} \frac{1}{n} \log N(n,\overline{\lambda}_R)$, exist?

2) What is the structure of the <u>capacity functions</u>

$$C(\overline{\lambda}_R) = \lim_{n\to\infty} \frac{1}{n} \log N(n,\overline{\lambda}_R)$$

and

$$C(\lambda_R) = \lim_{n\to\infty} \frac{1}{n} \log N(n,\lambda_R)$$

where these are well defined?

All our results will be obtained under the restriction that $\mathcal{C}$ contains only finitely many, say k, c.p.f's.

A word about notation. The functions $C(\overline{\lambda})$, $C(\lambda_R)$, and $C(\overline{\lambda}_R)$ are distinguished only by their arguments; these will always appear explicitly. The result is that all our results have to be interpreted with this understanding. For example, one of our theorems says that

$$C(\lambda_R) = C(\overline{\lambda}) = C(\overline{\lambda}_R)$$

under certain conditions when $\lambda_R = \overline{\lambda} = \overline{\lambda}_R$. Taken literally this is a trivial statement. In the light of our notation it means that the three functions coincide for certain values of the argument. This notation will result below in no confusion or ambiguity, and has the advantages of suggestiveness and typographical simplicity.

Throughout this paper λ, $\overline{\lambda}$, λ_R, and $\overline{\lambda}_R$ take values only in the open interval $(0,1)$. This assumption avoids the trivial and will not be stated again.

2. Auxiliary results.

1) In the following we need:

Lemma 1:

Let $S = \{1,\dots,d\}$ and let $\{(u_i,A_i) | i = 1,\dots,N\}$ be a code with $\inf_{s\in S} \frac{1}{N} \sum_{i=1}^{N} P_n(A_i|u_i|s) \geq 1 - \bar{\lambda}$. There exist sequences $\{u_{i_v} | v = 1,\dots,N_1\} \subset \{u_i | i = 1,\dots,N\}$ such that

$$P_n(A_{i_v}|u_{i_v}|s) \geq 1 - (\bar{\lambda} + \epsilon)d \text{ for } v = 1,\dots,N_1 = [\tfrac{\epsilon}{1+\epsilon} N]$$

$$\text{and for } s = 1,\dots,d.$$

Proof of Lemma 1: Define the probability distribution P^* on $\{1,\dots,N\}$ by $P^*(i) = \frac{1}{N}$ for $i = 1,\dots,N$. Define the random variables $\{X_s | s = 1,\dots,d\}$ by $X_s(i) = 1 - P(A_i|u_i|s)$ for $i = 1,\dots N$. Thus $X_s(i) \geq 0$ and

$$EX_s = 1 - \frac{1}{N} \sum_{i=1}^{N} P(A_i|u_i|s) \leq \bar{\lambda} .$$

Hence

$$P^*\{X_s \leq d\cdot EX_s \quad \text{for } s = 1,\dots,d\}$$

$$\leq P^*\{X_s \leq d(\bar{\lambda} + \epsilon) \quad \text{for } s = 1,\dots,d\}$$

Define

$$B^* = \{X_s \leq d(\bar{\lambda} + \epsilon) \quad \text{for } s = 1,\dots,d\}$$

and

$$B_s = \{X_s > d(\bar{\lambda} + \epsilon)\}, \quad s = 1,..,d \ .$$

Then

$$P^*(B_s) \leq \frac{E(X_s)}{d(\bar{\lambda}+\epsilon)} \leq \frac{\bar{\lambda}}{d(\bar{\lambda}+\epsilon)} \ .$$

Hence

$$P^*(\bigcup_{s=1}^{d} B_s) \leq \frac{\bar{\lambda}}{\bar{\lambda} + \epsilon}$$

and therefore

$$P^*(B^*) \geq 1 - \frac{\bar{\lambda}}{\bar{\lambda} + \epsilon} = \frac{\epsilon}{\bar{\lambda} + \epsilon} \ .$$

By the definition of P^*

$$|B^*| \geq N \cdot \frac{\epsilon}{\bar{\lambda} + \epsilon} \geq N \cdot \frac{\epsilon}{1 + \epsilon} \geq [N \frac{\epsilon}{1 + \epsilon}] \ .$$

The elements of B^* are the desired sequences. This proves Lemma 1.

In Lemmas 2 and 3 only we let $|S| = 1$ and $(P_n(\cdot|\cdot))$, $n = 1,2,..$ be the only element of $\mathcal{C}^*$. We then have:

<u>Lemma 2</u>: (Shannon's) Lemma 3.1.1 in [6])

Let $\{(u_i,A_i)|i = 1,\ldots,N\}$ be a code for $P_n(\cdot|\cdot)$ with average error $\bar{\lambda}$, then there exists a subcode of length $N_1 = \frac{N\epsilon}{\bar{\lambda} + \epsilon}$ with maximal error $\bar{\lambda} + \epsilon$.

<u>Proof</u>:

Denote $|\{u_i|P_n(A_i|u_i) < 1 - \bar{\lambda} - \epsilon\}|$ by Z, then $Z(1-\bar{\lambda}-\epsilon) + (N-Z) \geq N(1 - \bar{\lambda})$ and therefore $N_1 = N - Z \geq \frac{\epsilon}{\bar{\lambda} + \epsilon} N$.

Lemma 3:

Given a random code $\{(p^i, A_i) | i = 1,..,N\}$ for $P_n(\cdot|\cdot)$ with average error $\bar{\lambda}$, we can construct a nonrandom code of the same length N with average error $\leq \bar{\lambda}$.

(As a consequence of Lemma 3, for given length N the average error is minimized by a non-random code. Obviously the maximal length of a code of average error $\bar{\lambda}$ increases with increasing $\bar{\lambda}$. Hence, for given average error, a nonrandom code is at least as long as any random code.)

Proof of Lemma 3:

Let $\{(p^i, A_i) | i = 1,...,N\}$ be a random code with $\frac{1}{N} \sum_{i=1}^{N} \sum_{x_n \in X_n} p^i(x_n) P_n(A_i | x_n) = 1 - \bar{\lambda}$. The contribution of message i to $N(1 - \bar{\lambda})$ is clearly $\sum_{x_n \in X_n} p^i(x_n) P_n(A_i | x_n)$. Suppose now that $P_n(A_i | x_n^{(1)}) \geq P_n(A_i | x_n^{(2)}) \geq \dots \geq P_n(A_i | x_n^{(a^n)})$. Instead of using $\{x_n^{(1)}, \dots, x_n^{(a^n)}\}$ with the probabilities $\{p^i(x_n^{(1)}), \dots, p^i(x_n^{(a^n)})\}$ for message i, now use $x_n^{(1)}$ with probability 1, and keep A_i as the decoding set which corresponds to message i. The contribution of message i to $N(1 - \bar{\lambda})$ is now replaced by the larger quantity $P_n(A_i | x_n^{(1)})$. Using the same procedure for all i one achieves a nonrandom code $\{(u_i, A_i) | i = 1,...,N\}$ with average error $\leq \bar{\lambda}$.

(One can improve on the code even more by keeping the u_i of the new code, and replacing the A_i by the maximum-likelihood sets B_i.)

2) Averaged channels:

Let $S = \{1,\dots,d\}$, and let $g = (g,\dots,g_d)$ be a probability vector on S. The sequence

$$(P_n(\cdot|\cdot)) = (\sum_{s=1}^{d} g_s P_n(\cdot|\cdot|s)), \quad n = 1,2, \dots$$

is called an <u>averaged</u> channel. Let $N_a(n,\lambda)$ be the maximal length of any code (n,N,λ) for this channel. Denote $\lim_{n\to\infty} \frac{1}{n} \log N_a(n,\lambda)$ by $C_a(\lambda)$ for those λ for which the limit exists.

Theorem 1 and remark 2 of [1] imply that

$$C_a(\lambda) = \max_{\{S'|S'\subset S, g(S') > 1-\lambda\}} \max_{\pi} \inf_{s\in S} R(\pi,s)$$

at least for $\lambda \notin \{\Sigma_{i\in S'} g_i | S'\subset S\}$. Furthermore, as a consequence of Lemma 2 we have

$$C_a(\lambda) = C_a(\overline{\lambda}) \quad \text{for} \quad \lambda = \overline{\lambda} \notin \{\sum_{i\in S'} g_i | S' \subset S\}.$$

Also, as a consequence of Lemma 3 we have

$$C_a(\overline{\lambda}_R) = C_a(\overline{\lambda}) .$$

Obviously, $C_a(\overline{\lambda}_R) \geq C_a(\lambda_R) \geq C_a(\lambda)$ and therefore

$$C_a(\overline{\lambda}_R) = C_a(\lambda_R) = C_a(\overline{\lambda}) = C_a(\lambda) \quad \text{for} \quad \lambda = \overline{\lambda} \notin \{\sum_{i\in S'} g_i | S' \subset S\}.$$

3) Compound channels with side information were introduced in [4]. If the sender knows the c.p.f. in $\mathcal{C}$ which governs the transmission of a message to be sent, an (n,N,λ) code is defined as a system

$$\{(u_i(s), A_i) \mid u_i(s) \in X_n, \ A_i \subset Y_n, \ A_i \text{ disjoint}, \ i=1,\dots,N; \ s \in S\}$$

which satisfies $P_n(A_i | u_i(s) | s) \geq 1-\lambda$ for $i = 1,\dots,N$; $s \in S$.

The capacity is then given by $\inf_{s \in S} \max_{\pi} R(\pi, s)$ (Theorem 2 of [4]).

We will need a slightly more general theorem. In the situation just described the sender knows precisely the channel which actually governs the transmission of any word; in other words, he has complete knowledge. We shall say that the sender has the partial knowledge

$$K = \{(S_1,\dots,S_h) \mid S_i \subset S, \ i = 1,\dots,h\},$$

if the sender knows only that the governing channel has an index which belongs to a set of K, the set itself being known to him.

Lemma 4:

The capacity of the compound channel $\mathcal{C}^*$ with the sender's partial knowledge $K = (S_1,\dots,S_h)$ equals

$$\inf_{i=1\dots h} \ \max_{\pi} \ \inf_{s \in S_i} R(\pi, s) .$$

The proof follows the lines of the proof of Theorem 2 of [4] and will therefore be omitted.

3. The structure of $C(\overline{\lambda})$.

The determination of $C(\overline{\lambda})$ at its points of discontinuity seems to be difficult, and it is even undecided whether $\lim_{n\to\infty} \frac{1}{n} \log N(n,\overline{\lambda})$ exists at these points. (Compare also [5] and [1]. The determination of $C(\overline{\lambda})$ becomes more and more complicated as $|S|$ increases, and it seems to us that a simple recursion formula does not exist. However, the following results help clarify the structure of $C(\overline{\lambda})$.

Theorem 1.

Given $\mathcal{C} = \{w(\cdot|\cdot|s)|s = 1,\ldots,k\}$, then $C(\overline{\lambda})$ is well defined except perhaps for finitely many points $\lambda_1,\ldots,\lambda_{K*(k)}$, and for every $\overline{\lambda} \neq \lambda_i (i = 1,\ldots,K^*(k))$ $C(\overline{\lambda})$ equals an expression

$$(3.1) \qquad C_{\ell r \ldots} = \max_{\pi} \inf_{s=\ell, r, \ldots} R(\pi,s)$$

The points λ_i belong to a finite set D^* which is characterized in Theorem 2 below.

Since $0 \leq \log N(n,\overline{\lambda}) \leq n \log a$, $C^+(\overline{\lambda}) = \overline{\lim_{n\to\infty}} \frac{1}{n} \log N(n,\overline{\lambda})$ and $C^-(\overline{\lambda}) = \underline{\lim}_{n\to\infty} \frac{1}{n} \log N(n,\lambda)$ are well defined for all $\overline{\lambda}$. Let $\{(u_i,A_i)|i = 1,\ldots,N\}$ be a $(n,N,\overline{\lambda})$-code for $\mathcal{C}_n^*$ of maximal length. For every $\epsilon > 0$ define

$$(3.2) \qquad G_{\ell r \ldots}(\epsilon) = \{u_i | P_n(A_i|u_i|s) > \epsilon \text{ for } s = \ell, r, \ldots \text{ and for no other index}\}$$

and

$$G_o(\epsilon) = \{u_i | P_n(A_i | u_i | s) \leq \epsilon \quad \text{for all} \quad s \in S\}.$$

The G's form a partition of the code into disjoint subcodes. Applying Lemma 2 with ϵ sufficiently small for any one value of s, say $s = 1$, we obtain that $|G_o(\epsilon)|$ is bounded by a fixed multiple of $N(n,\overline{\lambda})$. Since $N(n,\overline{\lambda})$ grows exponentially, we can, and do, omit $G_o(\epsilon)$ from our code without any essential loss, provided ϵ is sufficiently small.

Define $\alpha_{\ell r \ldots}(n,\epsilon) = \dfrac{|G_{\ell r \ldots}(\epsilon)|}{N(n,\overline{\lambda})}$. Let $n_1, n_2, \ldots,$ be a subsequence of the integers such that

$$\overline{\lim_{t\to\infty}} \frac{1}{n_t} \log N(n_t,\overline{\lambda}) = C^+(\overline{\lambda}) \quad . \tag{3.3}$$

We can now define

$$\alpha_{\ell r \ldots}(\epsilon) = \overline{\lim_{t\to\infty}} \alpha_{\ell r \ldots}(n_t,\epsilon) \quad . \tag{3.4}$$

Let

$$L(\epsilon) = \{(\ell,r,\ldots,) | \alpha_{\ell r \ldots}(\epsilon) > 0\}.$$

If $(\ell,r,\ldots,) \in L(\epsilon)$ then, as a consequence of the strong converse for compound channels (Theorem 4.4.1 of [6]), $C^+(\overline{\lambda}) \leq C_{\ell r \ldots}$, and therefore

$$C^+(\overline{\lambda}) \leq \inf \{C_{\ell r \ldots} | (\ell,r,\ldots,) \in L(\epsilon)\}. \tag{3.5}$$

Since ϵ was arbitrary,

$$(3.6) \qquad C^{+}(\bar{\lambda}) \leq \liminf_{\epsilon \to 0} \{C_{\ell r \ldots} \mid (\ell, r, \ldots) \in L(\epsilon)\} .$$

Define

$$f_t(s) = |\{u_i \mid P_{n_t}(A_i \mid u_i \mid s) > \epsilon\}|$$

for $s = 1, \ldots, k$. Hence $f_t(s) + (N - f_t(s))\,\epsilon \geq N(1-\bar{\lambda})$ and consequently

$$f_t(s) \geq N\left(\frac{1-\bar{\lambda}-\epsilon}{1-\epsilon}\right) \qquad (s = 1, \ldots, k)$$

On the other hand,

$$(3.7) \qquad \frac{f_t(s)}{N} = \sum_{\substack{(\ell, r, \ldots) \\ s \in \{\ell, r \ldots\}}} \alpha_{\ell r \ldots}(n_t, \epsilon) \geq \frac{1-\bar{\lambda}-\epsilon}{1-\epsilon}, \qquad s = 1, \ldots, k$$

Clearly, for $\eta > 0$ there exists a $n_o(\eta)$ such that, for $n_t \geq n_o(\eta)$, $\alpha_{\ell r \ldots}(n_t, \epsilon) \leq \eta$ for $(\ell, r \ldots) \notin L(\epsilon)$, because there are only finitely many sets of indices. From (3.7) it follows that, for $s = 1, \ldots, k$,

$$(3.8) \qquad \sum_{\substack{(\ell, r, \ldots) \in L(\epsilon) \\ s \in (\ell, r \ldots)}} \alpha_{\ell, r \ldots}(n_t, \epsilon) \geq \frac{1-\bar{\lambda}-\epsilon}{1-\epsilon} - \eta \cdot 2^k$$

Consider a code (n_t, N', δ) of maximal length for the compound channel with the sender's partial knowledge

$$K = \{(\ell,r,\ldots) | (\ell,r,\ldots) \in L(\epsilon)\}.$$

For each $(\ell,r,\ldots) \in L(\epsilon)$ choose $N' \cdot \alpha_{\ell,r,\ldots}(\epsilon)$ indices from $1, \ldots, N'$ (the choice is arbitrary, but different complexes which are in $L(\epsilon)$ must correspond to disjoint sets of indices), and for these indices use as message sequences (i.e., $u_i's$) only those message sequences which would have been used if the sender knew that the governing channel was in $(\ell,r,\ldots)$. By (3.8) and Lemma 4 this leads to a code $(n_t,N,\overline{\lambda}')$ for $\mathcal{C}^*_{n_t}$ of length

(3.9) $$N(n_t,\overline{\lambda}') \geq \exp\,[n_t \cdot \inf\,\{C_{\ell r\ldots} | (\ell,r,\ldots) \in L(\epsilon)\} - \text{const.}\ \sqrt{n}$$

where $1-\overline{\lambda}' = (\frac{1-\overline{\lambda}-\epsilon}{1-\epsilon} - \eta \cdot 2^k)\,(1-\delta)$. Using the same α's for all n sufficiently large, we get

$$N(n,\overline{\lambda}') \geq \exp\,[n \cdot \inf\,\{C_{\ell r\ldots} | (\ell,r,\ldots) \in L(\epsilon)\} - \text{const.}\ \sqrt{n}\ ]$$

and consequently

$$C^-\ (\overline{\lambda}') \geq \inf\,\{C_{\ell r\ldots} | (\ell,r,\ldots) \in L(\epsilon)\}$$

Furthermore, $\overline{\lambda} = \lim\limits_{\epsilon,\eta,\delta\to 0} \overline{\lambda}'$, and therefore

$$C^-(\overline{\lambda}) \geq \lim_{\epsilon\to 0} \inf\ \{C_{\ell,r\ldots} | (\ell,r,\ldots) \in L(\epsilon)\}$$

for every $\overline{\lambda}$ which is a continuity point of $C^-(\overline{\lambda})$. Using (3.6) we get

(3.10) $$C^+(\overline{\lambda}) = C^-(\overline{\lambda}) = C(\overline{\lambda}) = \lim_{\epsilon\to 0} \inf\,\{C_{\ell r\ldots} | (\ell,r,\ldots) \in L(\epsilon)\}$$

for all $\bar{\lambda}$ which are continuity points of $C^-(\bar{\lambda})$. However, $C^-(\bar{\lambda})$ is a monotonic function on $[0,1]$ and can therefore have only countably many discontinuities. It follows from (3.10) that $C^-(\bar{\lambda})$ takes only finitely many values on the set of its continuity points. Hence $C^-(\bar{\lambda})$, and therefore also $C(\bar{\lambda})$, have only finitely many discontinuities. This proves the theorem.

From the definition of $C(\bar{\lambda})$, every point of continuity of $C(\bar{\lambda})$ is a point of continuity of $C^-(\bar{\lambda})$. From (3.10) and the fact that $C^-(\bar{\lambda})$ is a step function it follows that every point of continuity of $C^-(\bar{\lambda})$ is a point of continuity of $C(\bar{\lambda})$. Therefore $C(\bar{\lambda})$ and $C^-(\bar{\lambda})$ have the same points of continuity.

Theorem 1 says that, except perhaps for at most finitely many points, $C(\bar{\lambda})$ is given by an expression

$$C_{\ell r \ldots} = \max_{\pi} \inf_{s=\ell, r, \ldots} R(\pi, s)$$

For different channels $C(\bar{\lambda})$ may be given by different expressions. We now seek a formula for $C(\bar{\lambda})$ which does not depend on the channel. (The actual values taken by this formula will, of course, depend on the channel.)

We introduce the class of formulas

$$\text{(3.11)} \quad \tilde{f} = \{I \mid I \text{ is given by maxima and minima of expressions}$$

$$C_{\ell r \ldots} = \max_{\pi} \inf_{s=\ell, r, \ldots} R(\pi, s)\}.$$

The value of a formula I for $\mathcal{C}$ will be denoted by $I(\mathcal{C})$. A partial ordering is defined in $\tilde{f}$ by

(3.12) $I_1 \leq I_2$ if and only if $I_1(\mathcal{C}) \leq I_2(\mathcal{C})$ for all $\mathcal{C}$ with

$$|\mathcal{C}| = k.$$

$\tilde{f}$ need not be totally ordered. It can happen that, for $I_1, I_2 \in \tilde{f}$ and two channels $\mathcal{C}_1$, $\mathcal{C}_2$, $I_1(\mathcal{C}_1) > I_2(\mathcal{C}_1)$ and $I_1(\mathcal{C}_2) < I_2(\mathcal{C}_2)$.

We start our considerations for a fixed $\mathcal{C}$ which has k elements and develop an algorithm for the computation of $C(\overline{\lambda})$. For any real numbers z_1 and z_2 define $z_1 \cap z_2 = \min(z_1, z_2)$, $z_1 \cup z_2 = \max(z_1, z_2)$. Obviously

$$(3.13) \quad C_{12\ldots k} \leq \underline{\lim_{n\to\infty}} \frac{1}{n} \log N(n,\overline{\lambda}) \leq \overline{\lim_{n\to\infty}} \frac{1}{n} \log N(n,\overline{\lambda}) \leq \bigwedge_{s=1,\ldots,k} C_s$$

Every term $C_{\ell r\ldots}$ which is a possible value of $C(\overline{\lambda})$ for some value of $\overline{\lambda}$ therefore has to satisfy

$$(3.14) \quad C_{\ell r\ldots} = C_{\ell r\ldots} \bigwedge_{s\notin\{\ell, r, \ldots\}} C_s$$

Every index $1,\ldots,k$ appears in the right member of (3.14). We now write $C_{\ell r\ldots}$ as

$$(3.15) \quad C_{\ell r\ldots} = C_{\ell_1 r_1\ldots} \wedge C_{\ell_2 r_2\ldots} \wedge\ldots\wedge C_{\ell_t r_t},$$ where

a) no index can be added to any set $\{\ell_i, r_i, \ldots\}$ without violating (3.15),

b) no additional term can be added on the right without violating (3.15) or condition a).

The representation (3.15) is therefore unique. Let the number of terms on the right of (3.15) be t. For $s = 1,\ldots,k$ and $i = 1,\ldots,t$ define

$$\delta(s,i) = 1 \quad \text{if } s \in (\ell_i, r_i, \ldots)$$

$$\delta(s,i) = 0 \quad \text{if } s \notin (\ell_i, r_i, \ldots)$$

Let $\alpha = (\alpha_1,\ldots,\alpha_t)$ be a probability t-vector. We define

$$(3.16) \quad \bar{\lambda}(\ell, r, \ldots) = 1 - \max_{\alpha} \min_{s} \sum_{i=1}^{t} \alpha_i \, \delta(s,i).$$

We will now prove that, for $\bar{\lambda} > \bar{\lambda}(\ell, r \ldots)$,

$$(3.17) \quad \underline{\lim}_{n\to\infty} \frac{1}{n} \log N(n, \bar{\lambda}) \geq C_{\ell r \ldots} \; .$$

Let α^* be the maximizing value of α in (3.16). Let $\epsilon > 0$ be small enough. For suitable $m(\epsilon) > 0$ we construct a code

$$(n, N = \exp_2 \{n\, C_{\ell r \ldots} - \sqrt{n}\, m(\epsilon)\}, \epsilon)$$

for the compound channel with the sender's partial knowledge

$$K = \{(\ell_1, r_1, \ldots), \; \ldots \; , (\ell_t, r_t, \ldots)\}.$$

Let the code be written as

$$(u_i^{(1)},\ldots,u_i^{(t)}, A_i), \quad i = 1,\ldots,N.$$

Consider the new code

$$(u_i^{(1)}, A_i),\ i = 1,\dots,N\cdot\alpha_1^*$$

$$(u_i^{(2)}, A_i),\ i = (N\cdot\alpha_1^* + 1),\dots,N\cdot(\alpha_1^* + \alpha_2^*)$$

$$\cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot$$

$$(u_i^{(t)}, A_i),\ i = N\cdot(\alpha_1^* + \dots + \alpha_{t-1}^*) + 1,\dots,N.$$

For $s = 1,\dots,k$ the average error of this code is not greater than

$$1 - (1-\epsilon) \min_s \sum_{i=1}^{t} \alpha_i^* \delta(s,i).$$

When ϵ is small enough we obtain (3.17).

Now define

$$(3.18)\quad V_{\ell r\dots}(\bar{\lambda}) = \begin{cases} C_{\ell r\dots} & \text{for } \bar{\lambda} > \bar{\lambda}(\ell, r, \dots) \\ 0 & \text{otherwise} \end{cases}$$

and

$$(3.19)\quad V(\bar{\lambda}) = \max_{S'} \{V_{\ell,r\dots}(\bar{\lambda}) \mid S' = \{\ell, r, \dots\} \subset S\}$$

$V(\bar{\lambda})$ is a step-function with at most finitely many jumps. It follows from (3.17) that

$$(3.20)\quad \underline{\lim}_{n\to\infty} \frac{1}{n} \log N(n,\bar{\lambda}) \geq V(\bar{\lambda})$$

at every point of continuity of $V(\bar{\lambda})$.

Let $\bar{\lambda}$ be a point of continuity of $C(\bar{\lambda})$ and $V(\bar{\lambda})$. Let $\epsilon_0 > 0$ be so small that $L(\epsilon_0) = L(\epsilon)$ for $0 < \epsilon < \epsilon_0$. From (3.6) we know that $C(\lambda^-)$ is the smallest, say $C_{\ell r\dots}$, of a finite number of expressions of this type whose index sets belong to $L(\epsilon_0)$. Passing to the limit in (3.8) we have, for $s = 1,\dots,k$,

$$(3.21) \quad \Sigma\, \alpha_{\mu}(\epsilon_0) \geq \frac{1-\overline{\lambda}-\epsilon_0}{1-\epsilon_0} - \eta \cdot 2^k,$$

where the summation is over all index sets μ which contain s and belong to $L(\epsilon_0)$. Write $C_{\ell r \ldots}$ in the form (3.15) and suppose, without loss of generality, that (3.15) is the actual representation. Assign each element of $L(\epsilon_0)$ to some one of the sets in the right member of (3.15) which contains this element, and define $\alpha^*(\epsilon_0)$ of the latter set as the sum of the $\alpha(\epsilon_0)$ of the sets assigned to it; $\alpha^*(\epsilon_0)$ will be zero for a set to which no sets have been assigned. A fortiori, for $s = 1,\ldots,k$,

$$(3.22) \quad \sum_{i=1}^{t} \delta(s,i)\alpha^*_{\ell_i r_i \ldots}(\epsilon_0) \geq \frac{1-\overline{\lambda}-\epsilon_0}{1-\epsilon_0} - \eta \cdot 2^k.$$

Letting η and ϵ_0 approach zero we obtain from (3.16) and (3.22) that

$$(3.23) \quad \overline{\lim_{n\to\infty}} \frac{1}{n} \log N(n,\overline{\lambda}) \leq V(\overline{\lambda}).$$

From (3.20) and (3.23) we obtain that

$$(3.24) \quad C(\overline{\lambda}) = V(\overline{\lambda})$$

at the points of continuity of both functions. $C(\overline{\lambda})$ is defined and continuous at all but a finite number of points, and monotonic. $V(\overline{\lambda})$ is defined everywhere and monotonic. Both are step-functions. Hence the two functions are identical at every point of continuity of $C(\overline{\lambda})$.

We now have that (3.15), (3.16), (3.18), and (3.19) determine an algorithm for the computation of $C(\overline{\lambda})$. (See Section 5 for applications.)

It follows from (3.18) and (3.19) that any point of discontinuity λ_i of $C(\overline{\lambda})$ must be one of the set

(3.25) $\{\overline{\lambda}(\ell, r, \ldots) | (\ell, r, \ldots) \subset S\}$.

Now $\overline{\lambda}(\ell, r, \ldots)$ depends upon the representation (3.15). However, it does not depend on the actual values C which enter into that representation, but only upon the indices which enter into the right member of (3.15). All possible sets of such indices are finite in number. Moreover, for any given $\mathcal{C}$ with $|S| = k$, the set of indices in the right member of (3.15) depends <u>only on the ordering according to size of the various C's</u> of $\mathcal{C}$, and not at all on the actual values taken by them. When $|S| = k$ there are a fixed (finite) number of expressions of the form $C_{\ell r \ldots}$. A finite number of channels with $|S| = k$ and alphabets of sufficient length will produce all the possible orderings of these expressions. Call one such set of channels

(3.26) $Q = \{T_1, \ldots, T_q\}$.

We have therefore proved:

(3.27) For <u>any</u> channel $\mathcal{C}$ with $|S| = k$, the set of points of discontinuity of its function $C(\overline{\lambda})$ coincides with the set of points of discontinuity of the function $C(\overline{\lambda})$ of $T(\mathcal{C})$, where $T(\mathcal{C})$ is that member of Q whose C's have the same ordering according to size as those of $\mathcal{C}$, and

(3.28) The set D^* of all possible points of discontinuity of $C(\overline{\lambda})$ for all $\mathcal{C}$ with $|S| = k$ consists of all points of the form

(3.25), and can be evaluated by the algorithm implied by (3.16) and (3.15), and

(3.29) Two channels, $\mathcal{C}_1$ and $\mathcal{C}_2$, say, both with $|S| = k$, have the same points of discontinuity for their respective functions $C(\bar{\lambda})$ if the set

$$\{c_{\ell r \ldots} | (\ell, r, \ldots) \subset S\}$$

has the same ordering according to size for both $\mathcal{C}_1$ and $\mathcal{C}_2$.

The representation (3.15) is defined for a fixed $\mathcal{C}$. To indicate the dependence on $\mathcal{C}$ we write

$$C(\bar{\lambda}, \mathcal{C}),\ c_{\ell_1 r_1 \ldots}(\mathcal{C}), \ldots, c_{\ell_t r_t}(\mathcal{C}).$$

Suppose now that, for a fixed $\bar{\lambda}$ not in D^*,

$$(3.30)\quad c_{\ell r \ldots}(\mathcal{C}) = c_{\ell_1 r_1 \ldots}(\mathcal{C}) \wedge c_{\ell_2 r_2 \ldots}(\mathcal{C}) \wedge \cdots \wedge c_{\ell_t r_t \ldots}(\mathcal{C})$$

$$= C(\bar{\lambda}, \mathcal{C})$$

and for channel T_1

$$(3.31)\quad c_{\ell^* r^* \ldots}(T_1) = c_{\ell_1^{(1)} r_1^{(1)} \ldots}(T_1) \wedge \cdots \wedge c_{\ell_t^{(1)} r_t^{(1)} \ldots}(T_1)$$

$$= C(\bar{\lambda}, T_1)$$

In (3.22) let α^* correspond to channel $\mathcal{C}$ and α^{**} correspond to channel T_1. Both $\{\alpha^*\}$ and $\{\alpha^{**}\}$ satisfy (3.22). Hence, by the argument which follows (3.8) we have

$$(3.32)\quad C(\bar{\lambda}, \mathcal{C}) \geq c_{\ell_1^{(1)} r_1^{(1)} \ldots}(\mathcal{C}) \wedge \cdots \wedge c_{\ell_t^{(1)} r_t^{(1)} \ldots}(\mathcal{C}).$$

Hence, from (3.30) and (3.32),

$$(3.33)\quad C(\bar{\lambda},\mathcal{C}) = [C_{\ell_1 r_1 \ldots}(\mathcal{C}) \wedge \ldots\ ]$$

$$\vee\ [C_{\ell_1^{(1)} r_1^{(1)} \ldots}(\mathcal{C}) \wedge \ldots\ ]$$

Repeating this argument we obtain

$$(3.34)\quad C(\bar{\lambda},\mathcal{C}) =$$

$$[C_{\ell_1 r_1 \ldots}(\mathcal{C}) \wedge \ldots\ ] \vee$$

$$\bigvee_{i=1}^{q} [C_{\ell_1^{(i)} r_1^{(i)} \ldots}(\mathcal{C}) \wedge \ldots\ ]$$

where, for $i = 1,\ldots,q$,

$$C_{\ell_1^{(i)} r_1^{(i)} \ldots} \wedge \ldots$$

is the representation (3.15) of $C(\bar{\lambda}, T_i)$ in terms of the C's of channel T_i.

Assume temporarily that we can show that

$$(3.35)\quad C(\bar{\lambda},\mathcal{C}) = \bigvee_{i=1}^{q} [C_{\ell_1^{(i)} r_1^{(i)} \ldots}(\mathcal{C}) \wedge \ldots\ ].$$

We could then regard (3.35) as an identity in the "free variable" (argument) $\mathcal{C}$ (with $|S| = k$) if we could show that the system of subscripts of the C's which occurs in the right member of (3.30) does not depend on $\mathcal{C}$. (It may, and actually does, depend on the fixed $\bar{\lambda}$.) To prove this it is sufficient to see that the system of subscripts is determined by

(3.36) $C(\overline{\lambda},T_1),\ldots,C(\overline{\lambda},T_q)$.

Write the points of D^* as

(3.37) $a_1 < a_2 < \ldots < a_{Z(k)-1}$.

Also write $a_0 = 1$, $a_{Z(k)} = 1$. Suppose $a_z < \overline{\lambda} < a_{z+1}$. Then clearly (3.35) is valid for all points in the interval (a_z,a_{z+1}), because both members are constant in the interval.

The formula (3.35) depends upon the interval (a_z,a_{z+1}); there may be a different formula for a different interval. However, since $C(\overline{\lambda},\mathcal{C})$ is monotonic in $\overline{\lambda}$ for <u>any</u> $\mathcal{C}$, the different right members of (3.35) for different intervals are monotonic for <u>any</u> $\mathcal{C}$, and thus are totally ordered.

It remains to prove that we can omit the first bracket on the right of (3.34). The subscripts in it are determined by the representation (3.15) of

$$C_{\ell,r\ldots}(\mathcal{C}) = C(\overline{\lambda},\mathcal{C})$$

in terms of the C's of $\mathcal{C}$. We have already seen, in (3.27), that this representation is the same as that in terms of the C's of $T(\mathcal{C})$. Hence the first bracket on the right of (3.34) is already included among the square brackets in $\bigvee_{i=1}^{q}[\quad]$ in the right member of (3.34). This proves (3.35).

We sum up our results in

<u>Theorem 2</u>. For any integer k there is a finite set D^*, described in

(3.28). The points of discontinuity of $C(\bar{\lambda})$ for any $\mathcal{C}$ with $|S| = k$ belong to D^*. The right member of (3.35) is constant in any $\bar{\lambda}$ - interval between two consecutive points of D^*, and is determined by this interval. (Different such intervals in general determine different right members of (3.35).) $C(\bar{\lambda})$ is given by (3.35).

Remarks

1.) It is not possible to use only formulas of $\tilde{f}$ which are built up only by minima. In Example 2 of Section 5, for instance, we have

$$C(\bar{\lambda}) = (c_{12} \vee c_{13} \vee c_{23}) \wedge c_1 \wedge c_2 \wedge c_3$$

$$= (c_{12} \wedge c_3) \vee (c_{13} \wedge c_2) \vee (c_{23} \wedge c_1)$$

$$\text{for } \bar{\lambda} \in (\tfrac{1}{2}, \tfrac{2}{3})$$

Suppose $c_{12} \wedge c_3 > c_{13} \wedge c_2,\ c_{23} \wedge c_1$ then $C(\bar{\lambda}) = c_{12} \wedge c_3$. Permuting the indices we would get $C(\bar{\lambda}) \neq c_{12} \wedge c_3$.

2.) It is not true that any two terms in square brackets on the right of (3.35) can be transformed into each other by permutation of indices, as can be seen from Example 3 in Section 5 for $\bar{\lambda} \in (\frac{3}{5}, \frac{2}{3})$.

4. The relationships of $C(\lambda_R)$, $C(\bar{\lambda}_R)$, and $C(\bar{\lambda})$

Theorem 3:

$$C(\lambda_R) = C(\bar{\lambda}) = C(\bar{\lambda}_R) \quad \text{for} \quad \lambda_R = \bar{\lambda} = \bar{\lambda}_R ,$$

at the points of continuity of $C(\bar{\lambda})$. [$C(\bar{\lambda})$ has only finitely many points of discontinuity.] The proof will be given in several steps.

For any positive integer n there exists a random code for $\mathcal{C}_n^*$

$$\{(p^i, A_i) \mid i = 1,\ldots,N\} \tag{4.1}$$

which satisfies, for any $s \in S$,

$$\frac{1}{N} \sum_{i=1}^{N} \sum_{x_n \in X_n} p^i(x_n) \, P_n(A_i | x_n | s) \geq 1 - \bar{\lambda}_R , \tag{4.2}$$

and which is of maximal length $N(n, \bar{\lambda}_R)$. Define, for $i = 1,\ldots,N$,

$$B^i_{\ell r \ldots}(\epsilon) = \{x_n | P_n(A_i | x_n | s) > \epsilon \quad \text{for } s = \ell, r, .., \text{ and no other index}\} \tag{4.3}$$

and also

$$B^i_o(\epsilon) = \{x_n | P_n(A_i | x_n | s) \leq \epsilon \quad \text{for every index } s \in S\} \tag{4.4}$$

There are 2^k possible index sets $\{\ell, r, \ldots,\}$. Denote these sets in some order by $\rho_1, \ldots, \rho_{2^k}$. For every $i (i = 1,\ldots,N)$

$\{B^i_{\rho_j}(\epsilon) \mid j = 1,\ldots,2^k\}$ is a disjoint partition of X_n. Define the column vector

$$B_{\rho_j}(\epsilon) = \begin{pmatrix} B^1_{\rho_j}(\epsilon) \\ \vdots \\ B^N_{\rho_j}(\epsilon) \end{pmatrix} \tag{4.5}$$

and the matrix

$$B(\epsilon) = \left(B^i_{\rho_j}(\epsilon) \right) \quad \begin{array}{l} i = 1,\ldots,N \\ j = 1,\ldots,2^k \end{array} \tag{4.6}$$

Henceforth we operate only on the matrix $B(\epsilon)$.

Define $C^+(\lambda_R) = \overline{\lim_{n\to\infty}} \frac{1}{n} \log N(n,\lambda_R)$

$$\begin{aligned} C^-(\lambda_R) &= \underline{\lim_{n\to\infty}} \frac{1}{n} \log N(n,\lambda_R) \\ C^+(\overline{\lambda}_R) &= \overline{\lim_{n\to\infty}} \frac{1}{n} \log N(n,\overline{\lambda}_R) \\ C^-(\overline{\lambda}_R) &= \underline{\lim_{n\to\infty}} \frac{1}{n} \log N(n,\overline{\lambda}_R) \quad . \end{aligned} \tag{4.7}$$

Let n_1, n_2, be a sequence such that

$$\lim_{t\to\infty} \frac{1}{n_t} \log N(n_t, \overline{\lambda}_R) = C^+(\overline{\lambda}_R)$$

Assume now that for every $n (n = 1,2,\ldots)$ a random code $(n, N, \overline{\lambda}_R)$ with maximal length $N(n,\overline{\lambda}_R)$ is given. To indicate the dependence on n we now write $B^i_{\rho_j}(\epsilon,n)$. Denote by $\beta_{\rho_j}(\epsilon,n)$ the number of components (rows) of $\beta_{\rho_j}(\epsilon,n)$ which are non-empty sets. We say that the index set ρ_j is ϵ-essential if

$$(4.8) \qquad \overline{\lim_{t\to\infty}} \{[N(n_t, \overline{\lambda}_R)]^{-1} \beta_{\rho_j}(\epsilon,n)\} = \beta_{\rho_j}(\epsilon) > 0.$$

Let $M(\epsilon)$ be the set of ϵ-essential index sets ρ_j. It follows from the definitions (4.7) and (4.8) and from the strong converse for compound channels (Theorem 4.4.1 of [6]) that

$$C^+(\overline{\lambda}_R) \leq C_{\rho_j}, \; \rho_j \text{ in } M(\epsilon).$$

Hence

$$C^+(\overline{\lambda}_R) \leq \inf \{C_{\rho_j} | \rho_j \text{ in } M(\epsilon)\}$$

This is true for every $\epsilon > 0$. Hence, when $\lambda_R = \overline{\lambda}_R$,

$$(4.9) \qquad C^+(\lambda_R) \leq C^+(\overline{\lambda}_R) \leq \inf_{\epsilon>0} \inf \{C_{\rho_j} | \rho_j \text{ in } M(\epsilon)\},$$

the first inequality being obvious.

We now prove the converse. Since there are only finitely many indices ρ_j we can conclude the following for any $\eta > 0$: There exists an $n_o(\eta)$ such that, for $n_t \geq n_o(\eta)$,

$$(4.10) \qquad \beta_{\rho_j}(n_t, \epsilon) \leq \eta, \quad \rho_j \text{ not in } M(\epsilon).$$

Then, for n sufficiently large, in the matrix (4.6) for a code $(n_t, N, \overline{\lambda}_R)$, we delete column $B_o(\epsilon)$ and all <u>columns</u> $B_{\rho_j}(\epsilon)$ for which ρ_j is not in $M(\epsilon)$. As a result of this the average error of the resulting code is less than

$$\bar{\lambda}_R + 2^k \cdot \eta + \epsilon \tag{4.11}$$

Now take an (n_t, N', δ) code

$$\{(u_i(\rho_j), A_i^*) \mid i = 1,\ldots,N';\ \rho_j \text{ in } M(\epsilon)\}$$

of length

$$N' \geq \exp\left[\inf\{C_{\rho_j} \mid \rho_j \in M(\epsilon)\} \cdot n_t - K'(\delta)\sqrt{n}\right] \tag{4.12}$$

for the compound channel with the sender's partial knowledge

$$K = \{\rho_j \mid \rho_j \text{ in } M(\epsilon)\}$$

For any $\ell \in \{1,2,\ldots,N'\}$ define

$$p^i(u_\ell(\rho_j)) = p^i(B^i_{\rho_j}) \quad \text{for } i = 1,\ldots,N;\ \rho_j \in M(\epsilon). \tag{4.13}$$

Also define $\delta_{\rho_j s} = 1$ when $s \in \rho_j$ and 0 when $s \notin \rho_j$. Then we can conclude that

$$\sum_{\rho_j \in M(\epsilon)} p^i(u_\ell(\rho_j))\, \delta_{\rho_j s}\, P(A_\ell^* \mid u_\ell(\rho_j) \mid s) \tag{4.14}$$

$$\geq [1-\delta] \sum_{\rho_j \in M(\epsilon)} \sum_{x_n \in B^i_{\rho_j}} p^i(x_n) P(A_i \mid x_n \mid s) - \epsilon$$

$$\text{for } i = 1,\ldots,N;\ s \in S;\ \ell = 1,\ldots,N'$$

It follows from (4.11) and (4.14) that

$$(4.15)\quad \frac{1}{N}\sum_{i=1}^{N}\sum_{\rho_j\in M(\epsilon)} p^i(u_\ell(\rho_j))\delta_{\rho_j s}\, P(A_\ell^*|u_\ell(\rho_j)|s)$$

$$\geq [1-\delta]\,\frac{1}{N}\sum_{i=1}^{N}\sum_{\rho_j\in M(\epsilon)}\sum_{x_n\in B^i_{\rho_j}} p^i(x_n)\, P(A_i|x_n|s)-\epsilon$$

$$\geq [1-\bar{\lambda}_R-2^k\cdot\eta-\epsilon][1-\delta]-\epsilon \quad \text{for} \quad s\in S \quad \text{and}\ \ell=1,\dots,N'\,.$$

Defining now

$$(4.16)\qquad P(\rho_j) = \frac{1}{N}\sum_{i=1}^{N} p^i(B^i_{\rho_j})$$

for $\rho_j \in M(\epsilon)$, we conclude, using (4.15), that

$$(4.17)\qquad \sum_{\rho_j\in M(\epsilon)} p(\rho_j)\ \delta_{\rho_j s}\ P(A_\ell^*|u_\ell(\rho_j)|s)$$

$$\geq [1-\bar{\lambda}_R-2^k\eta-\epsilon][1-\delta]-\epsilon \quad \text{for} \quad \ell = 1,\dots,N';\ s\in S.$$

Thus we now have a random code with <u>maximal</u> error λ' defined by

$$1-\lambda' = (1-\bar{\lambda}_R-2^k\eta-\epsilon)(1-\delta)-\epsilon$$

and length given by (4.12).

Now define

$$\alpha_j = [p(\rho_j)\cdot N'] \quad \text{for}\ \rho_j \in M(\epsilon).$$

If necessary we renumber the elements of $M(\epsilon)$ so that

$$M(\epsilon) = \{\rho_j | j = 1,\dots,k^*(\epsilon)\}\ .$$

Consider the non-random code

$$(4.18) \qquad (u_1(\rho_1),\ A_1^*), \ldots, (u_{\alpha_1}(\rho_1),\ A_{\alpha_1}^*)\ ,$$

$$(u_{\alpha_1+1}(\rho_2),\ A_{\alpha_1+1}^*), \ldots, (u_{N'}(\rho_{k^*(\epsilon)}),\ A_{N'}^*)$$

It is a consequence of (4.17) that this code has an average error less than λ'. Hence, passing to the limit with ϵ, η, and δ we obtain, just as in the argument which led to (3.10), that

$$(4.19) \qquad C^-(\bar{\lambda}) \geq \inf_{\epsilon > 0} \inf \{ C_{\rho_j} | \rho_j \ \text{ in } \ M(\epsilon)\}$$

at the continuity points of $C^-(\bar{\lambda})$, and

$$(4.20) \qquad C^-(\lambda_R) \geq \inf_{\epsilon > 0} \inf \{ C_{\rho_j} | \rho_j \ \text{ in } \ M(\epsilon)\}$$

at the continuity points of $C^-(\lambda_R)$. From (4.9) and (4.20) we obtain that $C(\lambda_R)$ exists at the points of continuity of $C^-(\lambda_R)$ and that there

$$(4.21) \qquad C(\lambda_R) = C^+(\bar{\lambda}_R),\ \lambda_R = \bar{\lambda}_R$$

From (4.9) and (4.19) we obtain that at the points of continuity of $C^-(\bar{\lambda})$,

$$(4.22) \qquad C^-(\bar{\lambda}_R) \geq C^-(\bar{\lambda}) \geq C^+(\bar{\lambda}_R)\ , \quad \bar{\lambda} = \bar{\lambda}_R\ ,$$

the first inequality being obvious.

Finally, from (3.10), (4.21), and (4.22) we obtain that, at the points of continuity of $C^{-}(\overline{\lambda})$ and of $C^{-}(\lambda_R)$ we have

$$(4.23) \qquad C(\overline{\lambda}) = C(\overline{\lambda}_R) = C(\lambda_R), \; \overline{\lambda} = \lambda_R = \overline{\lambda}_R$$

Since $C(\overline{\lambda})$ and $C^{-}(\overline{\lambda})$ have the same points of continuity, we have that

$$(4.24) \qquad C(\overline{\lambda}_R) = C(\lambda_R) = C(\overline{\lambda}), \; \overline{\lambda}_R = \lambda_R = \overline{\lambda}$$

at the points of continuity of $C(\overline{\lambda})$ and $C^{-}(\lambda_R)$.

Earlier we proved that $C(\overline{\lambda})$ has only finitely many points of discontinuity, takes on the set of continuity points only finitely many values, and is monotonic. The function $C^{-}(\lambda_R)$ is monotonic, and hence has at most denumerably many points of discontinuity. If it had a point of discontinuity which is not a point of discontinuity of $C(\overline{\lambda})$ this would result in a contradiction of (4.24). Hence every point of continuity of $C(\overline{\lambda})$ is a point of continuity of $C^{-}(\lambda_R)$. Theorem 3 follows from this and (4.24).

5. Evaluation of $C(\bar{\lambda})$ in several examples.

Example 1. $S = \{1,2\}$.

We shall show that then

$$C(\bar{\lambda}) = \begin{cases} \max_{\pi} \inf_{s=1,2} R(\pi,s) & \text{for } 0 < \bar{\lambda} < \frac{1}{2} \\ \inf_{s=1,2} \max_{\pi} R(\pi,s) & \text{for } \frac{1}{2} < \bar{\lambda} < 1 \end{cases}$$

Proof:

That $C(\bar{\lambda}) \geq \max_{\pi} \inf_{s=1,2} R(\pi,s)$ for $0 < \bar{\lambda} < \frac{1}{2}$ follows from Theorem 4.3.1 of [6] (coding theorem for compound channels). On the other hand, given a $(n,N,\bar{\lambda})$ code for a $\bar{\lambda} < \frac{1}{2}$, we choose $\epsilon > 0$ such that $2(\bar{\lambda} + \epsilon) < 1$. Application of Lemma 1 with $d = 2$ guarantees the existence of a code with length $[\frac{\epsilon}{1+\epsilon} N]$ and maximal error $2(\bar{\lambda} + \epsilon)$. Hence, from Theorem 4.4.1 of [6] (strong converse for compound channels) it follows that

$$C(\bar{\lambda}) \leq \max_{\pi} \inf_{s=1,2} R(\pi,s) \text{ for } 0 < \bar{\lambda} < \frac{1}{2}.$$

Case: $\frac{1}{2} < \bar{\lambda} < 1$

Choose $\epsilon < \bar{\lambda} - \frac{1}{2}$. $\{(u_i(1), u_i(2), A_i) \mid i = 1,\dots, N\}$ be a code with maxim error ϵ for the compound channel with complete knowledge by the sender. Then

$$\{(u_j(1),A_j) \mid j = 1,\dots,[N|2]\} \cup \{(u_j(2),A_j) \mid j = [N|2] + 1,\dots,N\}$$

is a code for $\mathcal{C}^*$ with average error less than $\bar{\lambda}$. It follows from

Theorem 4.5.3 of [6] that $C(\bar{\lambda}) \geq \inf_{s=1,2} \max_{\pi} R(\pi,s)$ for $\frac{1}{2} < \bar{\lambda} < 1$, and from Lemma 2 that

$$C(\bar{\lambda}) \leq \inf_{s=1,2} \max_{\pi} R(\pi,s).$$

Example 2. $S = \{1,2,3\}$

We shall show that

$$C(\bar{\lambda}) = \begin{cases} C_{123} & \text{for } 0 < \bar{\lambda} < \frac{1}{3} \\ C_{12} \wedge C_{13} \wedge C_{23} & \text{for } \frac{1}{3} < \bar{\lambda} < \frac{1}{2} \\ (C_{12} \vee C_{13} \vee C_{23}) \wedge C_1 \wedge C_2 \wedge C_3 & \text{for } \frac{1}{2} < \bar{\lambda} < \frac{2}{3} \\ C_1 \wedge C_2 \wedge C_3 & \text{for } \frac{2}{3} < \bar{\lambda} < 1 . \end{cases}$$

Proof:

Case: $0 < \bar{\lambda} < \frac{1}{3}$

Use the coding theorem for compound channels with maximal error (Theorem 4.3.1 of [6]) for proving $C(\bar{\lambda}) \geq C_{123}$, and Lemma 1 and the strong converse for compound channels (Theorem 4.4.1 of [6]) for proving $C(\bar{\lambda}) \leq C_{123}$.

Case: $\frac{1}{3} < \bar{\lambda} < \frac{1}{2}$

1 Choose $\epsilon < \bar{\lambda} - \frac{1}{3}$. Let $\{u_i(12), u_i(13), u_i(23), A_i \mid i = 1,\dots,N\}$ be a (n,N,ϵ) code for $\mathcal{C}^*$, where the sender has the partial knowledge $K = (\{12\}, \{13\}, \{23\})$. Then $u_1(12), \dots, u_{[\frac{N}{3}]}(12), u_{([\frac{N}{3}]+1)}(13), \dots u_{[2\cdot\frac{N}{3}]}(13), u_{([2\cdot\frac{N}{3}]+1)}(23), \dots u_N(23)$; $A_1, \dots A_N$ is a $(n,N,\bar{\lambda})$

code for $\mathcal{C}^*$. Application of the coding theorem for compound channels (Theorem 4.3.1 of [6]) gives $C(\bar{\lambda}) \geq C_{12} \wedge C_{13} \wedge C_{23}$. Suppose now, without loss of generality (w.l.o.g.) $C_{12} = C_{12} \wedge C_{13} \wedge C_{23}$, then $C(\bar{\lambda}) \leq C_{12}$ by example 1.

<u>Case</u>: $\frac{1}{2} < \bar{\lambda} < \frac{2}{3}$

Choose $\epsilon < \bar{\lambda} - \frac{1}{2}$ and assume, w.l.o.g., that $(C_{12} \vee C_{13} \vee C_{23}) \cap C_1 \wedge C_2 \wedge C_3 = C_{12} \wedge C_3$. Then define $K = (\{12\}, \{3\})$. Apply Theorem 4.3.1 of [1] and select $u_1(12), \ldots u_{[\frac{N}{2}]}(12)$, $u_{[\frac{N}{2}]+1}(3) \ldots$, $u_N(3)$. By the usual procedure we finally get $C(\bar{\lambda}) \geq C_{12} \wedge C_3$. For proving the converse part we use the result for averaged channels. If $C_3 = C_{12} \wedge C_3$, obviously $C(\bar{\lambda}) \leq C_3$. Assume therefore that $C_{12} = C_{12} \wedge C_3$ $[\geq C_{23}, C_{13}]$. An $(n,N,\bar{\lambda})$ code for $\mathcal{C}^*$ is a $(n,N,\bar{\lambda})$ code for the averaged channel

$$(P_n(\cdot|\cdot)\ n = 1,2,\ldots) = (\sum_{s=1}^{3} \frac{1}{3} P_n(\cdot|\cdot|s),\ n = 1,2,\ldots)$$

Therefore $C_a(\bar{\lambda}) = C_a(\lambda) \geq C(\bar{\lambda})$, if $\lambda = \bar{\lambda}$ and not equal to 0, $\frac{1}{3}$, $\frac{2}{3}$, or 1. We get for $\frac{1}{2} < \bar{\lambda} < \frac{2}{3}$, that $C_a(\lambda) = C_{12}$, since $C_{12} \geq C_{23}, C_{13}$. Hence $C(\bar{\lambda}) \leq C_{12}$. This proves the desired result.

<u>Case</u>: $\frac{2}{3} < \bar{\lambda} < 1$.

Choose $\epsilon < \bar{\lambda} - \frac{2}{3}$ and define $K = (\{1\}, \{2\}, \{3\})$. Apply Theorem 4.3.1 of [6] and select

$$u_1(1), \ldots, u_{[\frac{N}{3}]}(1),\ u_{[\frac{N}{3}]+1}(2), \ldots, u_{[\frac{2N}{3}]}(2),\ u_{[\frac{2N}{3}]+1}(3), \ldots, u_N(3$$

Prove $C(\bar{\lambda}) \geq C_1 \wedge C_2 \wedge C_3$ as usual. $C(\bar{\lambda}) \leq C_1 \wedge C_2 \wedge C_3$ is obvious.

The converse parts could have been proved in all four cases by using suitable averaged channels. This will be illustrated in

Example 3. $S = \{1,2,3,4\}$

$$
C(\bar{\lambda}) = \begin{cases}
C_{1234} & \text{for } \bar{\lambda} \in (0,\frac{1}{4}) \\
C_{123} \wedge C_{124} \wedge C_{134} \wedge C_{234} & \text{for } \bar{\lambda} \in (\frac{1}{4},\frac{1}{3}) \\
\bigvee_{i \neq j \neq h \neq \ell} (C_{ijh} \wedge C_{ij\ell} \wedge C_{h\ell}) & \text{for } \bar{\lambda} \in (\frac{1}{3},\frac{2}{5}) \\
(C_{123} \vee C_{124} \vee C_{134} \vee C_{234}) \wedge C_{12} \wedge C_{13} \wedge & \\
\wedge C_{14} \wedge C_{23} \wedge C_{24} \wedge C_{34} & \text{for } \bar{\lambda} \in (\frac{2}{5},\frac{1}{2})
\end{cases}
$$

$$
\begin{cases}
(C_{12} \wedge C_{34}) \vee (C_{13} \wedge C_{24}) \vee (C_{14} \wedge C_{23}) \vee & \\
\vee (C_{123} \wedge C_4) \vee (C_{124} \wedge C_3) \vee (C_{134} \wedge C_2) & \\
\vee (C_{234} \wedge C_1) & \text{for } \bar{\lambda} \in (\frac{1}{2},\frac{3}{5}) \\
(C_{12} \wedge C_{13} \wedge C_{23} \wedge C_4) \vee (C_{12} \wedge C_{14} \wedge C_{24} \wedge C_3) \ \ldots & \\
\vee (C_{12} \wedge C_{34}) \vee \ldots & \text{for } \bar{\lambda} \in (\frac{3}{5},\frac{2}{3}) \\
(C_{12} \vee C_{13} \vee C_{14} \vee C_{23} \vee C_{24} \vee C_{34}) & \\
\wedge C_1 \wedge C_2 \wedge C_3 \wedge C_4 & \text{for } \bar{\lambda} \in (\frac{2}{3},\frac{3}{4}) \\
C_1 \wedge C_2 \wedge C_3 \wedge C_4 & \text{for } \bar{\lambda} \in (\frac{3}{4},1)
\end{cases}
$$

Proof: Case $(0, \frac{1}{4})$:

Obviously $C(\bar{\lambda}) \geq C_{1234}$. Use the averaged channel $P_n(\cdot|\cdot) = \sum_{s=1}^{4} \frac{1}{4} P_n(\cdot|\cdot|s)$ for proving $C(\bar{\lambda}) \leq C_{1234}$

Case: $(\frac{1}{4}, \frac{1}{3})$

Choose $\epsilon < \bar{\lambda} - \frac{1}{4}$. Let $\{u_i(123), u_i(124), u_i(134), u_i(234), A_i | i = 1,\dots,N\}$ a (n,N,ϵ) code for $\mathcal{C}^*$, where the sender has partial knowledge

$$K = (\{123\}, \{124\}, \{134\}, \{234\}).$$

Then $\{u_1(123),\dots,u_{[\frac{N}{4}]}(123), u_{[\frac{N}{4}]+1}(124), \dots u_N(234),$

$A_1, \dots A_N\}$ is a $(n,N,\bar{\lambda})$ code for $\mathcal{C}^*$.

Application of Theorem 4.3.1 of [6] gives $C(\bar{\lambda}) \geq C_{123} \wedge C_{124} \wedge C_{134} \wedge C_{234}$

We want to prove the converse in $(\frac{1}{4}, x_0)$. Assume the infimum is taken for C_{123}.

We introduce an averaged channel

$$P_n(\cdot|\cdot) = \sum_{s=1}^{4} p_s \, P_n(\cdot|\cdot|s)$$

for which

(a) $p_1 + p_2 + p_3 \geq 1 - \frac{1}{4}$

(b) $p_1 + p_2 + p_4, \; p_1 + p_3 + p_4, \; p_2 + p_3 + p_4 \leq 1 - x$

and x_0 is the maximal value of x for which a solution of (a), (b) exists.

Univ. Bibliothek Bielefeld

We use the solution

$$p_1 = p_2 = p_3 = \frac{1}{3},\ p_4 = 0$$

$$x_o = \frac{1}{3}$$

It follows that

$$C_a(\overline{\lambda}) = C_{123} \text{ for } \overline{\lambda} \in (\tfrac{1}{4},\tfrac{1}{3}), \text{ and therefore } C(\overline{\lambda}) \leq C_{123}.$$

<u>Case</u>: $(\frac{1}{3}, \frac{2}{5})$

Assume that the maximum is taken for $C_{123} \wedge C_{124} \wedge C_{34}$. Then $C(\overline{\lambda}) \geq C_{123} \wedge C_{124} \wedge C_{34}$ follows as usual by taking

$\frac{1}{3}$ of the $\{u_i(123) \mid i = 1,\ldots,N\}$,

$\frac{1}{3}$ of the $\{u_i(124) \mid i = 1,\ldots,N\}$,

and $\frac{1}{3}$ of the $\{u_i(34) \mid i = 1,\ldots,N\}$.

In the future we shall say shortly that we use a $(\frac{1}{3},\frac{1}{3},\frac{1}{3})$ - <u>fraction</u> (or in general a $(\alpha_1,\ldots,\alpha_n)$ - fraction).

If now $C_{34} = C_{123} \wedge C_{124} \wedge C_{34}$, then we use an average $p = (p_1,\ldots,p_4) = (0,0,\frac{1}{2},\frac{1}{2})$ and obtain the desired result.

We can therefore assume w.l.o.g. that

$$C_{123} \leq C_{124} \wedge C_{34}$$

$$C_{123} \geq C_{134} \wedge C_{24}$$

$$C_{123} \geq C_{234} \wedge C_{14}$$

If $C_{134} \wedge C_{24} = C_{24}$ [or $C_{234} \wedge C_{14} = C_{14}$], we immediately get $C(\overline{\lambda}) \leq C_{24}$ by using an average $p = (0, \frac{1}{2}, 0, \frac{1}{2})$ [or $p = (\frac{1}{2}, 0, 0, \frac{1}{2})$.]

It remains to consider

$$C_{123} \leq C_{124}, \; C_{34}$$

$$C_{123} \geq C_{234}, \; C_{134}$$

In order to get an averaged channel with $C_a(\overline{\lambda}) = C_{123}$ in $(\frac{1}{3}, x_0)$, $p = (p_1, \ldots, p_4)$ must satisfy $p_1 + p_2 + p_3 \geq 1 - \frac{1}{3}$

$$p_1 + p_2 + p_4 \leq 1 - x$$

$$p_1 + p_3 \leq 1 - x$$

$$p_2 + p_3 \leq 1 - x$$

$$p_4 + p_3 \leq 1 - x$$

Let x_0 be the maximal x for which a solution exists. We get $x_0 = \frac{2}{5}$, $p_1 = p_2 = p_4 = \frac{1}{5}$; $p_3 = \frac{2}{5}$ as a solution.

<u>Case</u>: $(\frac{2}{5}, \frac{1}{2})$

We can assume the infimum $= C_{123} \wedge C_{14} \wedge C_{24} \wedge C_{34}$
Use the fraction $(\frac{2}{5}, \frac{1}{5}, \frac{1}{5}, \frac{1}{5})$ for $K = \{(123), (14), (24), (341)\}$ to prove

$$C(\overline{\lambda}) \geq C_{123} \wedge C_{14} \wedge C_{24} \wedge C_{34}$$

If the infimum is taken for C_{123} use $p = (\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4})$ and if the infimum is taken for C_{14}, for instance, use $p = (\frac{1}{2}, 0, 0, \frac{1}{2})$. In either case we get $C(\overline{\lambda}) \leq C_{123} \wedge C_{14} \wedge C_{24} \wedge C_{34}$.

Case: $(\frac{1}{2}, \frac{3}{5})$

That the expression given above in Example 3 is a lower bound, is trivial; take the fraction $(\frac{1}{2},\frac{1}{2})$. To prove that the expression given is an upper bound, we consider first the case

1.) The maximum is taken by $C_{123} \wedge C_4$.

Subcase a.)

$$C_{123} \wedge C_4 = C_{123}$$

Thus $C_{123} \geq C_{12} \wedge C_{34},\ C_{13} \wedge C_{24},\ C_{23} \wedge C_{14}$

$C_{12}, C_{13}, C_{23} \geq C_{123}$ implies $C_{123} \geq C_{34}, C_{24}, C_{14}$.

We can assume that $C_{123} \geq C_{jkl}$, because if for instance $C_{134} > C_{123}$, then $C_2 = C_{123}$ and we can use the average (0,1,0,0).

We have therefore finally $C_4 \geq C_{123} \geq C_{34}, C_{24}, C_{14}, C_{jkl}$.

Now define $p = (\frac{1}{5}, \frac{1}{5}, \frac{1}{5}, \frac{2}{5})$. Then $C(\overline{\lambda}) \leq C_a(\overline{\lambda}) = C_{123}$.

Subcase b.)

$$C_4 \leq C_{123}$$

Use p = (0,0,0,1).

2.) The maximum is taken by $C_{12} \wedge C_{34}$. W.l.o.g. $C_{12} = C_{12} \wedge C_{34}$.

W.l.o.g. $C_{13} \leq C_{12}$

Case a.) $C_{23} \leq C_{12}$

Use $p = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3}, 0)$

Case b.) $C_{14} \leq C_{12}$

Use $(\frac{2}{5}, \frac{1}{5}, \frac{1}{5}, \frac{1}{5})$. $C_{\ell h} > C_{12}$ implies $\ell, h \neq 1$, but then

$p(\ell) + p(h) = \frac{2}{5} < 1 - \bar{\lambda}$. $C_{\ell hn} > C_{12}$ implies $\{\ell, h, n\} = \{2,3,4\}$. But $C_{234} \wedge C_1 \leq C_{12}$ implies that $C_1 = C_{12}$. Use $p = (1,0,0,0)$.

Case: $(\frac{3}{5}, \frac{2}{3})$

1.) The maximum is attained by $(*) = C_{12} \wedge C_{13} \wedge C_{23} \wedge C_4$ and by no term $(C_{lh} \wedge C_{nu})$.

Use fraction $(\frac{1}{5}, \frac{1}{5}, \frac{1}{5}, \frac{2}{5})$ to prove $C(\bar{\lambda}) \geq (*)$.

If $C_4 = (*)$, then the converse is obvious: $p = (0,0,0,1)$. Assume therefore, w.l.o.g., that $C_{12} = (*)$. It follows that $C_{34} < C_{12}$ and also C_{24}, $C_{14} < C_{12}$. Use $p = (\frac{1}{3}, \frac{1}{3}, 0, \frac{1}{3})$ to prove $C(\bar{\lambda}) \leq C_a(\bar{\lambda}) \leq (*)$.

2.) W.l.o.g. assume $C_{12} \wedge C_{34} = (*)$. Use the fraction $(\frac{1}{2}, \frac{1}{2})$ to prove $C(\bar{\lambda}) \geq (*)$.

Assume $C_{12} \leq C_{34}$ W.l.o.g. $C_{13} \leq C_{12}$

Case a.) $C_{23} \leq C_{12}$

Use $p = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3}, 0)$

Case b.) $C_{14} \leq C_{12}$

therefore C_{14}, $C_{13} \leq C_{12} \leq C_{34}$. Again, two cases:

α.) $C_{34} > C_{12}$

$(C_{34} \wedge C_{23} \wedge C_{24} \wedge C_1) \leq C_{12} \wedge C_{34}$ implies either $C_1 = C_{12}$, and we are finished, or $C_{23} \wedge C_{24} \leq C_{12}$, and therefore w.l.o.g. $C_{23} \leq C_{12}$. We have C_{13}, C_{14}, $C_{23} \leq C_{12} \leq C_{34}$. Use $p = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3}, 0)$.

β.) $C_{34} = C_{12}$

Therefore $C_{34} = C_{12} \geq C_{13}, C_{14}$. Use $p = (\frac{1}{3}, 0, \frac{1}{3}, \frac{1}{3})$.
$C_a(\overline{\lambda}) = C_{34} \geq C(\overline{\lambda})$.

<u>Case</u>: $(\frac{2}{3}, \frac{3}{4})$

W.l.o.g. let the value I of the formula be $C_{12} \wedge C_3 \wedge C_4$.
Use the fraction $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$ to get $C(\overline{\lambda}) \geq C_{12} \wedge C_3 \wedge C_4$.
Suppose $I = C_{12}$, use $p = (\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4})$.
Suppose $I = C_3$ for instance, then use $p = (0,0,1,0)$.

<u>Case</u>: $(\frac{3}{4}, 1)$ is obvious.

References

[1] Ahlswede, R., "Certain results in coding theory for compound channels" (to appear in Colloquium on Information Theory, Debrecen 1967).

[2] Ahlswede, R., and Wolfowitz, J., "Correlated decoding" to appear.

[3] Blackwell, D., Breiman, L., and Thomasian, A.J., "The capacity of a class of channels" Ann. Math. Stat. 30, No. 4 (1959), 1229-1241.

[4] Wolfowitz, J., "Simultaneous channels" (Arch. Rat. Mech. Analysis, 4, No. 4 (1960), 371-386).

[5] Wolfowitz, J., "Channels without capacity" Inf. and Control 6, No. 1 (1963), 49-54.

[6] Wolfowitz, J., "Coding theorems of information theory", Springer Verlag, Berlin-Heidelberg - New York. First edition, 1961; Second edition, 1964.

[7] Shannon, C.E., "Two-way communication channels", Proc. Fourth Berkeley Symp. on Math. Stat. and Prob., 611-644, University of California Press, Berkeley and Los Angeles, 1961.