Universität Bielefeld/IMW

Working Papers
Institute of Mathematical Economics

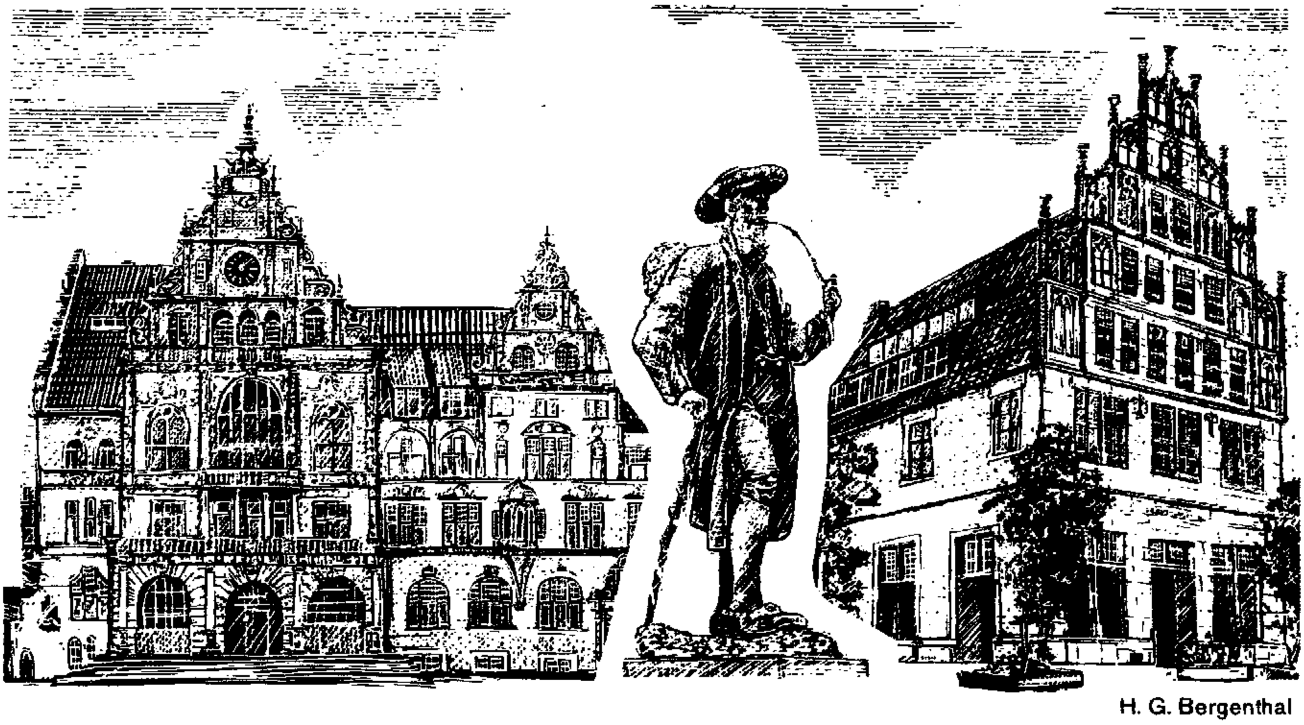
Arbeiten aus dem
Institut für Mathematische Wirtschaftsforschung

Nr. 142

Discrimination of equally-sized subsets for discrete, memoryless channels

Hans-Martin Wallmeier

November 1985



H. G. Bergenthal

Institut für Mathematische Wirtschaftsforschung
an der
Universität Bielefeld
Adresse/Address:
Universitätsstraße
4800 Bielefeld 1
Bundesrepublik Deutschland
Federal Republic of Germany

Abstract

A subset-discriminating code is defined by a set of codewords and a partition consisting of equally-sized subsets. These subsets are to be discriminated at the output of a discrete, memoryless channel. The rate-region given in terms of the pair: (rate of codewords, rate of elements of the partition) is determined in single-letter form.

I. Introduction

This paper is devoted to the determination of the rate-region for a subset-discriminating code for discrete, memoryless channels. Let us be given a DMC W and a set of messages to be transmitted via the channel. The rate of this set is assumed to exceed the capacity of the channel. It is well known that the error-probability arising from such a scheme converges to unity exponentially, that is, the use of a decoder consisting of a surjective function from the channel output sequences onto the set of messages becomes completely useless. Instead of decoding onto the set of messages we shall discriminate between subsets of messages, all of them sharing a given size. Our task now will be to give a computable formula for the maximum number of subsets which can be distinguished when the average probability of error is prescribed not to exceed an arbitrarily small $\lambda \in (0,1)$.

II. Notation and elementary known facts

Script capitals $\mathcal{U}, \mathcal{X}$, ... denote finite, nonempty sets. The letters P and Q stand for probability distributions, U, X ... for random variables. We shall write $U \ominus X \ominus Z$ if the random variables U, X and Z form a markov-chain in this order. The cardinality of a set and the range of a random-variable are denoted by $|\mathcal{U}|$ and $\|U\|$ respectively. $\mathcal{U}^c$ is the set-theoretical complement of $\mathcal{U}$. The logical symbols $\bigwedge_i F$ and $\bigvee_i F$ mean: "for all i proposition F holds" and "there exists i such that proposition F holds", respectively, $\hat{=}$ means "identity by definition" and $|t|^+ \hat{=} \max\{0,t\}$.

In the following we give notations and basic statements which are needed later, in their simplest form.

For any alphabet $\mathcal{X}$ $\mathcal{P}(\mathcal{X})$ denotes the set of all probability distributions on $\mathcal{X}$. The type of an n-sequence $x^n = (x_1,\dots,x_n) \in \mathcal{X}^n$ is a probability distribution P_{x^n} on $\mathcal{X}$ defined by

$$P_{x^n}(x) = n^{-1} N(x|x^n), \ x \in \mathcal{X},$$

where $N(x|x^n)$ is the number of occurences of x in the sequence x^n. $\mathcal{P}^n(\mathcal{X})$ is the set of all types in $\mathcal{X}^n$ and $T_P = \{x^n \in \mathcal{X}^n | P_{x^n} = P\}$ is the set of all n-sequences of type P. If $\mathcal{Z}$ is also an alphabet the joint type of a pair $(x^n,z^n) \in \mathcal{X}^n \times \mathcal{Z}^n$ is the probability distribution $P_{x^n z^n}$ on $\mathcal{X} \times \mathcal{Z}$ defined by $P_{x^n z^n}(x,z) = n^{-1} N(x,z|x^n,z^n)$, $(x,z) \in \mathcal{X} \times \mathcal{Z}$, and the conditional type of z^n given x^n is the conditional probability distribution

$$P_{z^n|x^n}(z|x) = \frac{N(x,z \mid x^n,z^n)}{N(x|x^n)}, \ (x,z) \in \mathcal{X} \times \mathcal{Z}.$$

For a conditional probability distribution V on $\mathcal{Z}$ given $\mathcal{X}$, denoted as

$V \mid \mathcal{X} \Rightarrow \mathcal{Z}$, and $x^n \in \mathcal{X}^n$ the V-shell around x^n is

$$T_V(x^n) = \{z^n \in \mathcal{Z}^n / \; P_{z^n|x^n} = V\}$$

For the entropy $H(X)$, conditional entropy $H(Z|X)$ and mutual information $I(X \wedge Z)$ we also write $H(P)$, $H(V|P)$ and $I(P;V)$, respectively.

For $P,Q \in \mathcal{P}(\mathcal{X})$

$$D(P \| Q) = \sum_{x \in \mathcal{X}} P(x) \log \frac{P(x)}{Q(x)}$$

denotes the Kullback-Leibler I-divergence and for $V,W \mid \mathcal{X} \Rightarrow \mathcal{Z}$

$$D(V \| W|P) = \sum_{x \in \mathcal{X}} P(x)\; D(V(\cdot|x) \,\|\, W(\cdot|x))$$

denotes the conditional I-divergence.

For $P \in \mathcal{P}(\mathcal{X})$, $V \mid \mathcal{X} \Rightarrow \mathcal{Z}$ the distribution $P(x,z) = P(x) \cdot V(z|x)$ on $\mathcal{X} \times \mathcal{Z}$ is denoted by (P,V), the distribution $P(z) = \sum_x P(x) \cdot V(z|x)$ on $\mathcal{Z}$ is denoted by $V \circ P$ in reminiscence to the composition of functions.

The following inequalities are folklore.

For any $\delta > 0$ there exists n sufficiently large such that

$$|\mathcal{P}^n(\mathcal{X})| \le (n+1)^{|\mathcal{X}|} \le \exp\{n \cdot \delta\},$$

$$\bigwedge_{P \in \mathcal{P}^n(\mathcal{X})} \exp\{n \cdot (H(P)-\delta)\} \le |T_P| \le \exp\{n \cdot H(P)\} ,$$

$$\bigwedge_{P \in \mathcal{P}^n(\mathcal{X})} P^n(x^n) = \prod_{t=1}^{n} P(x_t) = \exp\{- n \cdot (D(P_{x^n} \| P) + H(P_{x^n}))\},$$

$$\le \exp\{- n \cdot (H(P)-\delta)\}$$

$$\bigwedge_{(P,V) \in \mathcal{P}^n(\mathcal{X} \times \mathcal{Z})} \exp\{n \cdot (H(V|P)-\delta)\} \le |T_V(x^n)| \le \exp\{n \cdot H(V|P)\}$$

$$\bigwedge_{(P,V) \in \mathcal{P}^n(\mathcal{X} \times \mathcal{Z})} \; \bigwedge_{y^n \in T_{V \circ P}} \exp\{n \cdot (H(P)-I(P;V)-\delta)\} \le |\{x^n \in T_P | y^n \in T_V(x^n)\}|$$

$$\le \exp\{n \cdot (H(P)-I(P,V))\}$$

and

$$\bigwedge_{W|\mathcal{X}\Rightarrow\mathcal{Z}} W^n(z^n|x^n) = \prod_{t=1}^{n} W(z_t|x_t) = \exp\{-n\cdot(D(P_{z^n|x^n}||W|P_{x^n})+H(P_{z^n|x^n}\ P_{x^n}))\}$$

We shall use the elementary properties of typical and generated sequences without reference and refer the reader to the book [3] for further details.

III. The model and results

Let us be given channel-input and -output alphabets $\mathcal{X}$ and $\mathcal{Z}$ and denote the channel by $W \,|\mathcal{X} \Rightarrow \mathcal{Z}$. The transmission of sequences is assumed to be memoryless, that is

$$W^n(z^n|x^n) = \prod_{t=1}^{n} W(z_t|x_t)$$

For convenience the messages to be sent over the channel consist of channel input sequences $x_1,\dots,x_N$ of length n. They are independently chosen according to the uniform distribution on $\{1,\dots,N\}$.

We are interested in discriminating elements F_i, $i=1,\dots,N_0$ of a partition of $\{x_1,\dots,x_N\}$ consisting of nearly equally-sized subsets. Thus, for given $\delta > 0$ we define a <u>subset-discriminating (n,N_0,N,δ)-code</u> to consist of N distinct sequences $x_1,\dots,x_N$, a partition of $\{x_1,\dots,x_N\}$ for which

$$|F_i| \leq \frac{N}{N_0} \exp\{n \cdot \delta\} \quad , \; i = 1,\dots,N_0$$

and disjoint decoding sets $D_1,\dots,D_{N_0} \subset \mathcal{Z}^n$.

These codes will be denoted by $C = \{(F_i,D_i) \mid i=1,\dots,N_0, \sum_k |F_k| = N\}$, by omission of δ indicating that we are interested in codes with smaller and smaller $\delta > 0$. The pair $(n^{-1} \log N_0, u^{-1} \log N)$ is called <u>rate-pair</u> of the code. As mostly in information theory also in discriminating between subsets we allow small error probability: A pair (R_0,R) is called <u>achievable</u>, if for any $\varepsilon, \delta > 0$, $0 < \lambda < 1$ and sufficiently large block-length $n \in \mathbb{N}$ there exists an $(n, \exp\{n\cdot(R_0-\varepsilon)\}, \exp\{n\cdot(R-\varepsilon)\}, \delta)$ -code such that

$$\lambda(C,W) = N^{-1} \sum_{i=1}^{N_0} \sum_{x^n \in F_i} W^n(D_i^c \mid x^n) \leq \lambda \qquad \text{holds.}$$

For the discrete, memoryless channel W we denote the set of achievable rates by $\mathcal{R} \triangleq \mathcal{R}(W)$. The set $\mathcal{R}$ is a convex subset of the positive orthant of $\mathbb{R}^2$ since time-sharing is available. Our main result gives $\mathcal{R}$ in a computable manner. Define

$$\mathcal{R}^{*} = \{(R_0,R) \mid 0 \leq R_0 \leq \sup_{\substack{U \ominus X \ominus Z,\ P_{Z|X} = W\} \\ H(X|U) + I(U \wedge Z) \geq R \geq R_0}} \{I(U \wedge Z)\} \ \}$$

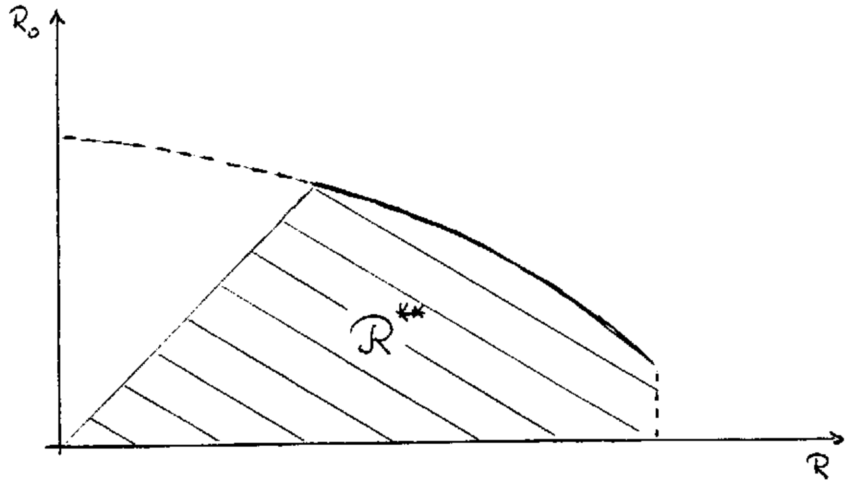
and

$$\mathcal{R}^{**} = \{(R_0,R) \mid 0 \leq R_0 \leq \max_{\substack{U \ominus X \ominus Z,\ P_{Z|X} = W\} \\ H(X|U) + I(U \wedge Z) \geq R \geq R_0 \\ \|U\| \leq |\mathcal{X}| + 1}} \{I(U \wedge Z)\} \ \}$$

Observing

$$\mathcal{R}^{**} = \{(R_0,R) \mid 0 \leq R_0 \leq \max_{\substack{U \theta X \theta Z,\ P_{Z|X} = W \\ H(X|U) + I(U \wedge Z) \geq R \\ \|U\| \leq |\mathcal{X}| + 1}} \{I(U \wedge Z)\},\ R_0 \leq R\}$$

the rate-region may be visualized as



The set $\mathcal{R}^{**}$ is well defined since the information-theoretical quantities involved are continuous with respect to the defining distributions and since the restriction on the range of U implies the set of admissible markov-chains to be compact. The set $\mathcal{R}^{**}$ accordingly may be calculated by a computer and thus our main

Theorem

$$\mathcal{R} = \mathcal{R}^{**}$$

gives a meaningful characterization of the rate-region for our model.

Our proceeding to derive this result will be as follows:

At first a coding theorem will be proved showing that each code out of some class of codes may be prolonged by adjunction of a suitable set of messages, a new element of a partition. Since, as is shown, in using this prolongation argument we may start from scratch and continue as long as $n^{-1} \log N_0 < I(U \wedge Z)$ at least the inclusion $\mathcal{R}^{**} \subset \mathcal{R}$ is shown. The additional subsets found in each step have a common rate of about $H(X\ U)$. Using the identity $\mathcal{R}^{*} = \mathcal{R}^{**}$ as shown in appendix A,

next the converse is proved showing the inclusion $\mathcal{R} \subset \mathcal{R}^{*}$. Here we prove that for smaller and smaller $\delta > 0$ the rates $(n^{-1} \log N_0$, $u^{1} \log N)$ of any subset-discriminating (n, N_0, N, δ)-code are limited by a function $B(\cdot)$ describing the border of $\mathcal{R}^{*}$.

IV. The coding result

Within this section the inclusion $\mathcal{R}^{**} \subset \mathcal{R}$ will be proved. Let $\mathcal{U}$ be any finite set and let U, X and Z be random variables with values in $\mathcal{U}$, $\mathcal{X}$ and $\mathcal{Z}$ respectively. Further assume U, X and Z to form a makrov-chain with $P_{Z|X} = W$. Let numbers $R_0 < I(U \wedge Z)$ and R such that $R_0 \leq R < I(U \wedge Z) + H(X|U)$ be given.

The coding theorem to be fomulated and proved below says that codes for which the set of messages has rate $R_0 + H(X|U)$ (basically) may be built up iteratively provided the partition given by the sets to be distinguished consists of at most $\exp\{n \cdot I(U \wedge Z)\}$ objects. More precisely, given any set of subsets $\{F_i \mid i = 1,\dots,N_0\}$, if the F_i resemble the V-shell around some centers $u_i \in \mathcal{U}^n$ close enough, then an $N_0 + 1^{st}$ center U_{N_0+1} may be chosen such that the average probability of decoding errors increases only exponentially. Our procedure to obtain this result follows [1] as far as the construction of codes is concerned.

By a random selection argument an appropriate center U_{N_0+1} is shown to exist, the new set of messages used for transmission is basically defined to consist of the V-shell around U_{N_0+1}. However, the V-shell around the new center may contain some old messages, thus we have to be careful in bounding the number of newly created words. The new center has to be chosen such that the mutual positions of the centers are such that the probability of decoding errors may be bounded successfully. It should be remarked that our method of prolonging a code turns out to be applicable only in case that the sets F_i to be distinguished from one another are basically V-shells, the centers, however may be arbitrary. Of course an exponential error bound only arises if those centers have a great enough distance.

Let $Q \in \mathcal{P}(\mathcal{U})$, $V \mid \mathcal{U} \Rightarrow \mathcal{X}$ be such that $(Q,V) \in \mathcal{P}^n(\mathcal{U} \times \mathcal{X})$. Then the following is true:

4.1 Lemma

For $\delta > 0$, n sufficiently large and any $u_1,\dots,u_{N_o} \in T_Q$ there exists u_{N_o+1} such that

$$\text{(i)} \qquad \left| T_V(u_{N_o+1}) \cap \bigcup_{i=1}^{N_o} T_V(u_i) \right|$$

$$\leq \exp\{n\,(H(V|Q) - |I(Q,V) - R_o|^+ + 3\delta)\}$$

$$\text{(ii)} \qquad \bigwedge_{\substack{P_{U\tilde{U}X} \\ P_{X|U}=V,\, P_U=P_{\tilde{U}}=Q}} \; \sum_{x^n \in T_V(u_{N_o+1})} \; \sum_{i=1}^{N_o} 1_{T_{P_{U\tilde{U}X}}} (u_{N_o+1}, u_i, x^n)$$

$$\leq \exp\{n(R_o + H(X|U\tilde{U}) - I(U\wedge\tilde{U}) + 3\delta)\}$$

$$\text{(iii)} \qquad \bigwedge_{\substack{P_{U\tilde{U}X} \\ P_{X|U}=V,\, P_U=P_{\tilde{U}}=Q}} \; \sum_{i=1}^{N_o} \; \sum_{x^n \in T_V(U_i)} 1_{T_{P_{U\tilde{U}X}}} (u_i, u_{N_o+1}, x^n)$$

$$\leq \exp\{n(R_o + H(X|U) - I(\tilde{U}\wedge UX) + 3\delta)\}$$

Proof:

It will be shown that the expectations of the left side will be smaller than the expressions on the right such that the application of the Markov-inequality $\Pr\{X \geq \alpha \cdot E(X)\} \leq \frac{1}{\alpha}$ - valid for non-negative random variables - will yield the desired result.

(i) is a special case of lemma 1 in [1]. For the sake of completeness we repeat its proof. Let here (and in the following steps) U_o denote a random-variable uniformly distributed on T_Q. Then

$$
\begin{aligned}
& E\,[\,|T_V(U_o) \cap \bigcup_{i=1}^{N_o} T_V(u_i)|\,] \\
\leq\ & \sum_{i=1}^{N_o} E\,[\,|T_V(U_o) \cap T_V(u_i)|\,] \\
=\ & N_o \cdot E\,[\,|T_V(U_o) \cap T_V(u_1)|\,] \\
\leq\ & N_o \cdot \sum_{u \in T_V(u_1)} \sum_{u_o \in T_Q} \exp\{-\, n \cdot (H(Q)-\delta)\} \cdot 1_{T_V(u_o)}(u) \\
\leq\ & \exp\{n \cdot (R_o + H(V|Q) - H(Q) + \delta + H(Q) - I(Q,V))\}
\end{aligned}
$$

On the other hand

$$
\begin{aligned}
& E\,[\,|T_V(U_o) \cap \bigcup_{i=1}^{N_o} T_V(u_i)|\,] \\
\leq\ & E\,[\,|T_V(U_o)|\,] \\
\leq\ & \exp\{n \cdot H(V|Q)\}
\end{aligned}
$$

whence

$$
\begin{aligned}
& E\,[\,|T_V(U_o) \cap \bigcup_{i=1}^{N_o} T_V(u_i)|\,] \\
\leq\ & \exp\{n \cdot (H(V|Q) - |I(Q;V) - R|^+ + \delta)\}
\end{aligned}
$$

Next we shall compute the expectation occuring in (ii):

$$
\begin{aligned}
& E\,[\sum_{X^n \in T_V(U_o)} \sum_{i=1}^{N_o} 1_{T_{P_{U\tilde{U}X}}}(U_o, u_i, x^n)] \\
\leq\ & \sum_{u_o \in T_Q} \exp\{-n \cdot (H(Q)-\delta)\} \cdot \sum_{x^n \in T_V(u_o)} \sum_{i=1}^{N_o} 1_{T_{P_{U\tilde{u}x}}}(u_o, u_i, x^n) \\
=\ & \exp\{n \cdot (R_o - H(U) + \delta)\} \cdot \sum_{u_o \in T_Q} \sum_{x^n \in T_V(u_o)} 1_{T_{P_{U\tilde{U}X}}}(u_o, u_1, x^n) \\
\leq\ & \exp\{n \cdot (R_o - H(U) + H(U,X|\tilde{U}) + \delta)\} \\
=\ & \exp\{n \cdot (R_o - I(U \wedge U) + H(X|U,\tilde{U}) + \delta)\}\ .
\end{aligned}
$$

Concerning case (iii) we have

$$E\ \Big[\sum_{i=1}^{N_o} \sum_{x^n\in T_V(u_i)} 1_{T_{P_{U\tilde{U}X}}} (u_i,U_o,x^n)\Big]$$

$$= \sum_{i=1}^{N_o} \sum_{x^n\in T_V(u_i)} E\ [1_{T_{P_{U\tilde{U}X}}} (u_i,\ U_o,x^n)]$$

$$\leq \exp\{n\cdot(R_o+H(V|Q))\} \sum_{u_o\in T_Q} \exp\{-n\ (H(Q)-\delta)\}\ 1_{T_{P_{U\tilde{U}X}}} (u_1,u_o,x^n)$$

$$\leq \exp\{n\cdot(R_o+H(X|U) - H(\tilde{U}) + H(\breve{U}|U,X) +\delta)\}$$

Now

$$\Pr\{|T_V(U_o) \cap \bigcup_{i=1}^{N_o} T_V(u_i)|$$

$$\geq \exp\{n\cdot(H(V|Q) - |I(Q,V)-R_o|^+ + 3\delta)\}$$

or, for some P_{UUX} with $P_{X|U} = V$, $P_U = P_U = Q$:

$$\sum_{x^n\in T_V(U_o)} \sum_{i=1}^{N_o} 1_{T_{P_{U\tilde{U}X}}} (U_o,u_i,x^n)$$

$$\geq \exp\{n\cdot(R_o+H(X|U,\tilde{U}) - I(U\wedge\tilde{U}) + 3\delta)\}$$

or

$$\sum_{i=1}^{N_o} \sum_{x^n\in T_V(u_i)} 1_{T_{P_{U\tilde{U}X}}} (u_i,U_o,x^n)$$

$$\geq \exp\{n\cdot(R_o+H(X|U) - I(\tilde{U}\wedge UX) + 3\delta)\}\}$$

$$\leq \sum_{\substack{P_{U\tilde{U}X} \in \mathcal{P}^n(\mathcal{U}^2 x \mathcal{X}): \\ P_{X|U}=V, P_U=P_{\tilde{U}}=Q}} \Pr\{|T_V(U_o) \cap \bigcup_{i=1}^{N_o} T_V(u_i)|$$

$$\geq \exp\{n2\delta\} E[|T_V(U_o) \cap \bigcup_{i=1}^{N_o} T_V(u_i)|]$$

or

$$\sum_{x^n \in T_V(U_o)} \sum_{i=1}^{N_o} 1_{T_{P_{U\tilde{U}X}}}(U_o,u_i,x^n)$$

$$\geq \exp\{n\cdot 2\delta\}\cdot E[\sum_{x^n \in T_V(U_o)} \sum_{i=1}^{N_o} 1_{T_{P_{U\tilde{U}X}}}(U_o,u_i,x^n)]$$

or

$$\sum_{i=1}^{N_o} \sum_{x^n \in T_V(u_i)} 1_{T_{P_{U\tilde{U}X}}}(u_i,U_o,x^n)$$

$$\geq \exp\{n\cdot 2\delta\} E[\sum_{i=1}^{N_o} \sum_{x^n \in T_V(u_i)} 1_{T_{P_{U\tilde{U}X}}}(u_i,U_o,x^n)]\}$$

$$\leq 3 \quad \exp\{- n \cdot \delta\}$$

$$< 1$$

whence the existence of a suitable U_{N_o+1} follows.

Property (i) shows that a center U_{N_o+1} may be chosen such that the V-shells do not intersect too much. Indeed, let us consider a code with a pair of rates lying in $\mathcal{R}^{**}$. Then $R_o < I(U \wedge Z) \leq I(U \wedge X) = I(Q;V)$, where the inequality follows from markovity of U, X and Z. Now assume the sets F_i to be contained in a V-shell of some $u_i \in T_Q$, each. Then the set of words to be deleted from the V-shell of a suitable U_{N_o+1} has a rate strictly smaller than $H(V|Q)$ since $|I(Q,V) - R_o|^+ > 0$. Thus the rate of the remaining words is still nearly $H(V|Q)$. This will ensure thereafter that the probability of decoding errors may be computed as if the entire shell is used as set of codewords.

We may now formulate our coding result concerning the iterative code construction and show the achievability of an exponential error bound.

4.2 Theorem

For any finite set $\mathcal{U}$, numbers $R_0 > 0$ and $\delta > 0$, $Q \in \mathcal{P}(\mathcal{U})$, $V|\mathcal{U} \Rightarrow \mathcal{X}$ such that $(Q,V) \in \mathcal{P}^n(\mathcal{U} \times \mathcal{X})$ and n sufficiently large the following is true:

Let $C = \{(F_i,D_i) \mid i = 1,\dots,N_0, \sum_{i=1}^{N_0} |F_i| = N\}$ be a set discriminating (n, N_0, N, δ)-code such that

$$n^{-1} \log N_0 \leq R_0$$

$$F_i \subset T_V(u_i) \quad \text{for suitable} \quad u_i \in T_Q$$

and

$$|F_i| \geq \exp\{n \cdot (H(V|Q) - \delta)\}.$$

Then there exists $u_{N_0+1} \in T_Q$ such that for suitable chosen decoding sets D_i^*, $i=1,\dots,N_0$ the enlarged code

$$C^* = \{(F_i,D_i^*) \mid i = 1,\dots,N+1\}$$

with

$$F_{N_0+1} = T_V(u_{N_0+1}) - \bigcup_{i=1}^{N_0} F_i$$

satisfies

$$|F_{N_0+1}| \geq \exp\{n \cdot (H(V|Q) - \delta)\}$$

and

$$\lambda(C^*,W) \leq \Big(\sum_{i=1}^{N_0+1} |F_i|\Big)^{-1} N \cdot \lambda(C,W) + 2 \cdot \exp\{-n \cdot (E_r(R_0,R,Q,V,W) - 4\delta)\},$$

where

$$E_r(R_0,R,Q,V,W)$$
$$= \min_{P_{UXZ}=(Q,V,W)} \{D(P_{Z|V\circ Q} \| W \mid V \circ Q) + |I(U \wedge Z) - R_0|^+\}$$

$$H(X|U) + I(U \wedge Z) \geq R.$$

<u>Proof:</u> Let u_{N_o+1} be chosen consistent with the inequalities of the lemma. We define the decoding sets for the prolonged code to be

$$D_i^* = D_i - \{z^n \mid I(u_{N_o} \wedge z^n) > I(u_i \wedge z^n)\},\ i = 1,\dots,N_o$$

and

$$D_{N_o+1}^* = \{z^n \mid \ _{i=1\dots N_o} \ I(u_{N_o} \wedge z^n) > I(u_i \wedge z^n)\}.$$

Obviously, the sets D_i^* , $i = 1,\dots,N_o+1$ are disjoint. We may now estimate the average probability of decoding errors:

$$\lambda\,(C^*, W)$$

$$= \Big(\sum_{i=1}^{N_o+1} |F_i|\Big)^{-1} \sum_{i=1}^{N_o} \sum_{x^n \in F_i} W^n(D_i^{*\,c} \mid x^n)$$

$$= \Big(\sum_{i=1}^{N_o+1} |F_i|\Big)^{-1} \Big(\sum_{i=1}^{N_o} \sum_{x^n \in F_i} [W^n(D_i^c \mid x^n) + W^n\,(D_i - D_i^* \mid x^n)]$$

$$+ \sum_{x^n \in F_{N_o+1}} W^n\,(D_{N_o+1}^{*c} \mid x^n)\Big)$$

$$= \Big(\sum_{i=1}^{N_o+1} |F_i|\Big)^{-1} \Big(N \cdot \lambda(C,W) + \sum_{i=1}^{N_o} \sum_{x^n \in F_i} W^n\,(D_i - D_i^* \mid x^n)$$

$$+ \sum_{x^n \in F_{N_o+1}} W^n\,(D_{N_o+1}^{*c} \mid x^n)\Big)$$

Due to the markovity of U, X and Z we may replace the channel W by $W^* \mid \mathfrak{U} \times \mathfrak{X} \Rightarrow \mathfrak{Z}$ defined through $W^*(z \mid u,x) = W\,(z \mid x)$.

First we estimate the error-probability resulting from the transmission of the newly chosen center.

$$\Big(\sum_{i=1}^{N_o+1} |F_i|\Big)^{-1} \sum_{x^n \in F_{N_o+1}} W^n(D^{*c}_{N_o+1} \mid x^n)$$

$$\leq N_o^{-1} \exp\{-n \cdot (H(V|Q) - \delta)\} \sum_{x^n \in T_V(u_{N_o+1})} W^n(D^{*c}_{N_o+1} \mid x^n)$$

$$= N_o^{-1} \exp\{-n \cdot (H(V|Q) - \delta)\} \sum_{x^n \in T_V(u_{N_o+1})} W^{*n}(D^{*c}_{N_o+1} \mid u_{N_o+1}, x^n),$$

where the inequality results from $F_{N_o+1} \subset T_V(u_{N_o+1})$,

$|F_{N_o+1}| \geq \exp\{n \cdot (H(V|Q) - \delta)\}$ and our assumption

$|F_i| \geq \exp\{n \cdot (H(V|Q) - \delta)\}$, $i = 1, \dots, N_o$.

Now

$$\sum_{x^n \in T_V(u_{N_o+1})} W^{*n}(D^{*c}_{N_o+1} \mid x^n)$$

$$= \sum_{x^n \in T_V(u_{N_o+1})} W^{*n}(\{z^n \mid \bigvee_i I(u_{N_o+1} \wedge z^n) \leq I(u_i \wedge z^n)\} \mid u_{N_o+1}, x^n)$$

$$= \overline{\sum_{\substack{P_{U\tilde{U}XZ} \\ P_{X|U} = V, \\ I(U\wedge Z) \leq I(\tilde{U}\wedge Z)}}} \exp\{-n(D(P_{Z|UX} \| W | P_{UX}) + H(Z|UX))\} \cdot \Big[\sum_{x^n \in T_V(u_{N_o+1})} |\{z^n \mid \bigvee_i P_{u_{N_o}, u_i, x^n, zn} = P_{U\tilde{U}XZ}\}|\Big]$$

On one hand the expression in brackets may be upperbounded by

$$\exp\{n \cdot (H(X|U) + H(Z|UX))\}$$

yielding the upper bound

$$\overline{\sum_{\substack{P_{U\tilde{U}XZ} \\ P_{X|U} = V}}} \exp\{-n \cdot (D(P_{Z|UX} \| W^* \mid P_{UX}) - H(X|U))\}$$

for the error probability resulting from the use of the newly chosen codewords.

On the other hand the previous lemma gives the bound

$$\sum_{x^n\in T_V(u_{N_0+1})} \left| \{z^n \mid \bigvee_i P_{u_{N_0+1},\, u_i, x^n, z^n} = P_{UUXZ}\}\right|$$

$$\leq \exp\{n\cdot(H(Z|U\tilde{U}X)\}\cdot\exp\{n\cdot(R_0+H(X|U\tilde{U}) - I(U\wedge\tilde{U}) + 3\delta)\}$$

such that the error-probability may be upperbounded by

$$\sum_{\substack{P_{U\tilde{U}XZ}\\ P_{X|U}=V\\ I(U\wedge Z)\leq I(\tilde{U}\wedge Z)}} \exp\{-n\cdot(D(P_{Z|UX}\| W^*|P_{UX})+I(\tilde{U}\wedge Z|UX)-H(X|U\tilde{U})+I(U\wedge\tilde{U})-R_0-3\delta)\}$$

$$\leq \sum_{\substack{P_{U\tilde{U}XZ}\\ P_{X|U}=V\\ I(U\wedge Z)\leq I(\tilde{U}\wedge Z)}} \exp\{-n\cdot(D(P_{Z|UX}\| W^* \mid P_{UX})-H(X|U)+I(\tilde{U}\wedge UXZ)-R_0-3\delta)\}$$

$$\leq \sum_{\substack{P_{U\tilde{U}XZ}\\ P_{X|U}=V\\ I(U\wedge Z)\leq I(\tilde{U}\wedge Z)}} \exp\{-n\cdot(D(P_{Z|X}\| W|V\circ Q)-H(X|U)+I(U\wedge Z)-R_0-3\delta)\}$$

whence, additionally using the bound derived above,the inequality

$$\left(\sum_{i=1}^{N_0+1}|F_i|\right)^{-1}\underset{x^n\in F_{N_0+1}}{} W^n(D^{*c}_{N_0+1}\mid x^n)$$

$$\leq N_0^{-1}\exp\{-n\,(\min_{\substack{P_{UXZ}\\ P_{X|U}=V,P_U=Q}}\{D(P_{Z|X}\| W|P_X)+|I(U\wedge Z)-R_0|^+\} - 4\delta)\}$$

follows.

Our next point will be estimating the error-probability resulting from the use of the old words through the modified decoding sets. Here, as was seen above only the probability to reach the deleted part of the old decoding set has to be investigated. The inequality

$$\Big(\sum_{i=1}^{N_0+1}|F_i|\Big)^{-1}\sum_{i=1}^{N_0}\sum_{x^n\in F_i}W^n(D_i-D_i^*\mid x^n)$$

$$\le N_0^{-1}\exp\{-n\,(H(V|Q)-\delta)\}\cdot\sum_{i=1}^{N_0}\sum_{x^n\in F_i}W^{*n}(D_i-D_i^*\mid u_i,x^n)$$

is proved analogously to the proceeding above. Then

$$\sum_{i=1}^{N_0}\sum_{x^n\in F_i}W^{*n}\,(D_i-D_i^*\mid u_i,x^n)$$

$$\le\sum_{i=1}^{N_0}\sum_{x^n\in T_V(u_i)}W^{*n}(\{z^n\in D_i\mid I(u_{N_0+1}\wedge z^n)>I(u_i\wedge z^n)\}\mid u_i,x^n)$$

$$\le\sum_{\substack{P_{U\tilde{U}XZ}\\ P_{X|U}=V\\ I(U\wedge Z)\le I(\tilde{U}\wedge Z)}}\exp\{-n\,(D(P_{Z|UX}\|W^*\mid P_{UX})+H(Z|UX))\}\cdot\Big[\sum_{i=1}^{N_0}\sum_{x^n\in T_V(u_i)}|\{z^n\mid P_{u_iu_{N_0+1}x^nz^n}=P_{U\tilde{U}XZ}\}|\Big]$$

We obtain as before the upper bounds

$$\sum_{\substack{P_{U\tilde{U}XZ}:\\ P_{X|U}=V\\ I(U\wedge Z)\le I(\tilde{U}\wedge Z)}}\exp\{-n\,(D(P_{Z|UX}\|W^*\mid P_{UX})-H(X|U))\}$$

and

$$\sum_{\substack{P_{U\tilde{U}XZ}:\\ P_{X|U}=V\\ I(U\wedge Z)\le I(\tilde{U}\wedge Z)}}\exp\{-n\,(D(P_{Z|UX}\|W^*\mid P_{UX})-H(X|U)+I(U\wedge Z)-R_0-3\delta)\}$$

for the term in brackets. Here instead of (ii) of the preceding lemma the inequality (iii) has to be used. Thus

$$\Big(\sum_{i=1}^{N_0+1} F_i\Big)^{-1} \sum_{i=1}^{N_0} \sum_{x^n \in F_i} W^n(D_i - D_i^* \mid x^n)$$

$$\leq N_0^{-1} \exp\{-n \cdot (\min_{\substack{P_{UXZ} \\ P_{X|U}=V, P_U=Q}} \{D(P_{Z|X} \| W|P_X) + |I(U \wedge Z) - R_0|\}^+ - 4\delta)\}$$

The coding-theorem as formulated in section III is now obtained as a corollary. In fact we prove even more since codes may be built up iteratively and given an universably obtainable error-bound.

4.3 Corollary

For any $\mathfrak{U}$, $Q \in \mathcal{P}(\mathfrak{U})$ and $V|\mathfrak{U} \Rightarrow \mathcal{X}$ such that $(Q,V) \in \mathcal{P}^n (\mathfrak{U} \times \mathcal{X})$, any $\delta > 0$, and n sufficiently large there exists a subset-discriminating (n, N_0, N, δ) -code C such that

$$n^{-1} \log N_0 \geq R_0 > I(U \wedge Z) - \delta$$

$$n^{-1} \log N \geq R > R_0 + H(X|U) - 2\delta$$

for which

$$\lambda(C,W) \leq 2 \cdot \exp\{-n \cdot (E_r(R_0,R,Q,V,W) - 4\delta)\},$$

$E_r(R_0,R,Q,V,W) > 0$ and the bound on the error-probability holds simultaneously for all channels $W|\mathcal{X} \Rightarrow \mathcal{Z}$ for which $(R_0,R) \in \mathcal{R} = \mathcal{R}(W)$.

Proof: Given (R_0,R) as above and a markov chain $U \ominus X \ominus Z$ with probability distribution $P_{UXZ}(u,x,z) = Q(u) \cdot V(x|u) \cdot W(Z|X)$, we may iteratively choose centers $u_1,\dots,u_{N_0}$ and sets $F_1,\dots,F_{N_0}$ satisfying the conditions of lemma 4.1. In particular $|F_i| \geq \exp\{n \cdot (H(V|Q) - \delta)\}$ such that $N = \sum_{i=1}^{N_0} |F_i| \geq \exp\{n \cdot (I(U \wedge Z) + H(X|U) - \delta)\}$.

The exponent $E_r(R_0,R,Q,V,W)$ is positive (and can be assumed to exceed 4δ) for $R_0 < I(U \wedge Z)$. The universality of the code follows from the construction of the code. The centers u_i are chosen according to their mutual positions and the error-estimation,as provided in the proof of the theorem,is solely based on these positions.

Given $U \ominus X \ominus Z$ we proved until now the achievability of rate-pairs from $\{(R_0,R) \mid 0 \leq R_0 \leq I(U \wedge Z),\ R = R_0 + H(X|U)\}$. To see the achievability of the larger set $\{(R_0,R) \mid 0 \leq R_0 \leq I(U \wedge Z),\ R_0 \leq R \leq R_0 + H(X|U)\}$ we observe that the rate-pair (R_0,R_0) corresponds just to a subset-discriminating code in which the elements F_i of the partitions are singletons. Such a code may be identified with a common code for the DMC W possessing rate R_0. The achievability of rates (R_0,R) in the interval $[(R_0,R_0),(R_0,R_0 + H(X|U))]$ is now obtained by using time-sharing. As far as the error-exponent is concerned it is easily seen that the one corresponding to the transmission of the entire V-shell is worse (smaller) than the one corresponding to sets F_i being singletons. $E_r(R_0,R,Q,V,W)$ is therefore attainable even when time-sharing is used.

The set of achievable rates as obtained by the coding theorem and the preceding argument has to be augmented a second time to give the region as defined in the main theorem. We observe that according to theorem 4.2 the rate-pair $(I(U \wedge Z),\ I(U \wedge Z) + H(X|U))$ is achievable. Forming unions of an exponential number of elements of the partition, i.e.

$E_j = \bigcup_{i \in T_j} F_i,\ |T_j| = \exp\{n \cdot c\},\ T_j \cap T_{j'} = \emptyset$ for $j \neq j'$ and

considering the sets E_j as new partition of the messages at hand, rate pairs $(R_0,\ I(U \wedge Z) + H(X|U)),\ 0 \leq R_0 \leq I(U \wedge Z)$ are achievable. This procedure of course does not enlarge the error-probability since less objects are to be discriminated. Previously it was argumented that the pair $(R_0,R_0 + H(X|U)$ is achievable. Again by time-sharing between codes of rate-pairs $(R_0,R_0 + H(X|U))$ and $(R_0,\ I(U \wedge Z) + H(X|U))$ we get achievability of pairs (R_0,R) such that $R_0 + H(X|U) \leq R \leq I(U \wedge Z)+H(X|U)$

with an error-exponent being at least the smallest of all codes used intermediately, i.e. $E_r(R_0,R,Q,V,W)$ as defined above. Together with our first enlargening of the achievable rate-region $E_r(R_0,R,Q,V,W)$ is discerned to be an exponent bounding the probability of decoding errors.

At last comment concerning the coding theorem is in order. The finiteness assumption on $\mathcal{U}$, $|\mathcal{U}| \leq |\mathcal{X}| + 1$ used in the definition of R^{**} yields that any distribution $P_{UX} = (Q,V)$ may be approximated arbitrarily well by types. Remembering the continuity of the information-theoretical functions as depending on varying distributions this shows the achievability of the rate-region as given in our main theorem, which in contrast to the coding theorem and subsequent remarks of section IV does not take care of types. In the same manner the error-exponent may be defined by additionally allowing variation on probability distributions which are not types, of course.

V. The Converse

The converse of the coding-theorem - formalized by the inclusion $\mathcal{R} \subset \mathcal{R}^*$ - consists of proving that codes only may exist with rate-pairs up to a certain size. We shall give a description of the rate-region by means of some antitonic and concave (∩) function such that the shape of the region as visualized in section IV is obtained. We shall prove that this boundary of $\mathcal{R}^*$ is given by the function

$$B^* : [0,\infty) \longrightarrow [0,\infty) \quad \text{defined through}$$

$$B^*(c) = \sup_{\substack{U \ominus X \ominus Z,\; P_{Z|X}=W \\ H(X|U)+I(U \wedge Z) \geq c}} \{I(U \wedge Z)] \;;$$

$B^*(c)$ is defined to be zero if the constraint yields a void set. An alternative description may be given by

$$B^{**}(c) = \max_{\substack{U \ominus X \ominus Z,\; P_{Z|X}= W, \|U\| \leq |\mathfrak{X}| + 1 , \\ H(X|U)+I(U \wedge Z) \geq c}} \{I(U \wedge Z)\}$$

(see appendix A) such that for deriving properties of $B = B^* = B^{**}$ we may make use of either description.

To prove the inclusion $\mathcal{R} \subset \mathcal{R}^*$ we have to provide some technical tools. At first we shall investigate the continuity of $B(\cdot)$. Since the proof of the corresponding lemma needs some digression we refer it to appendix B.

Define $S = \{c \mid c \geq 0$, there exists $U \ominus X \ominus Z$ such that $H(X|U) + I(U \wedge Z) \geq c\}$

(Here as in the definition of B^* the restriction on the range of may be omitted, see appendix A).

5.1 Lemma:

$B(\cdot) = B^{**}(\cdot)$ is continuous on S .

A second property of B needed in the sequel is given in

5.2 Lemma:

$B(\cdot) = B^*(\cdot)$ is concave (∩) on S.

Proof: For c_τ, $\tau = 1,2$ let $U_\tau \ominus X_\tau \ominus Z_\tau$ denote markov-chains such that $H(X_\tau|U_\tau)+I(U_\tau \wedge Z_\tau) \geq c$, $\tau = 1,2$. Define $c = \alpha\cdot c_1 + (1-\alpha)\cdot c_2$, $\alpha \in (0,1)$ and (U,X,Z) such that $\Pr\{(\tilde{U},\tilde{X},\tilde{Z}) = (U_1,X_1,Z_1)\}$ $= \alpha = 1 - \Pr\{(\tilde{U},\tilde{X},\tilde{Z}) = (U_2,X_2,Z_2)\}$. Let T be a random variable independent of all other random variables such that $\Pr\{T = 1\}$ $= \alpha = 1 - \Pr\{T = 2\}$. Since

$$\begin{aligned} I((T,\tilde{U}) \wedge \tilde{Z}|\tilde{X}) &= I(T\wedge \tilde{Z}|\tilde{X}) + I(\tilde{U}\wedge\tilde{Z}|\tilde{X},T) \\ &= \alpha\; I(U_1\wedge Z_1|X_1) + (1-\alpha)\;\; I(U_2\wedge Z_2|X_2) \\ &= 0 \end{aligned}$$

we have $(T,\tilde{U}) \ominus \tilde{X} \ominus \tilde{Z}$ and

$$\begin{aligned} & H(\tilde{X}|T,\tilde{U}) + I(T,\tilde{U}\wedge\tilde{Z}) \\ \geq\; & H(\tilde{X}|T,\tilde{U}) + I(\tilde{U}\wedge\tilde{Z}|T) \\ =\; & \alpha\cdot(H(X_1|U_1) + I(U_1\wedge Z_1)) + (1-\alpha)\;(H(X_2|U_2) + I(U_2\wedge Z_2)) \\ \geq\; & c\;. \end{aligned}$$

Thus

$$\begin{aligned} B^*(c) &\geq I(T,\tilde{U}\wedge\tilde{Z}) \\ &\geq I(\tilde{U}\wedge\tilde{Z}|T) \\ &\geq \alpha\cdot I(U_1\wedge Z_1) + (1-\alpha)\; I(U_2\wedge Z_2), \end{aligned}$$

whence $B^*(c) \geq \alpha(B^*(c_1)-\varepsilon) + (1-\alpha)\;(B^*(c_2)-\varepsilon)$ for any $\varepsilon > 0$. The claim follows.

5.3 Lemma:

For $U \ominus X^n \ominus Z^n$ with memoryless transition from $\mathfrak{X}^n$ to $\mathfrak{Z}^n$ following holds:

(i) $U_t \ominus X_t \ominus Z_t$, $t = 1,\dots,n$

(ii) $Z^{t-1} \ominus U_t \ominus Z_t$, $t = 1,\dots,n$

(iii) $I(U \wedge Z^n) \leq \sum_{t=1}^{n} I(U_t \wedge Z_t)$

(iv) $H(X^n|U)+I(U \wedge Z^n) \leq \sum_{t=1}^{n} H(X_t|U_t)+I(U_t \wedge Z_t)$,

where

$$Z^{t-1} = (Z_1,\dots,Z_{t-1})$$
$$X^{t-1} = (X_1,\dots,X_{t-1})$$

and
$$U_t = (U,X^{t-1}), \quad t = 1,\dots,n.$$

Proof: To prove (i) we observe that due to the inequality $H(Z_t|X_t,U_t) \leq H(Z_t|X_t)$, $t=1,\dots,n$ it is sufficient to show

$$\sum_t H(Z_t|X_t,U_t) \geq \sum_t H(Z_t|X_t).$$

But

$$\begin{aligned}\sum_t H(Z_t|X_t,U_t) &\geq \sum_t H(Z_t|X_t,U_t,Z^{t-1}, X_{t+1},\dots,X_n)\\ &= H(Z^n|X^n,U)\\ &= H(Z^n|X^n)\\ &= \sum_t H(Z_t|X_t) .\end{aligned}$$

In case of property (ii) it is sufficient to show

$$\sum_t H(Z_t|U_t,Z^{t-1}) \geq \sum_t H(Z_t|U_t)$$

We have

$$\begin{aligned}&\sum_t H(Z_t|U,X^{t-1}Z^{t-1})\\ = \ &\sum_t (H(Z_t,X_t|U,X^{t-1},Z^{t-1}) - H(X_t|Z_t,U,X^{t-1},Z^{t-1}))\\ = \ &H(Z^n,X^n|U) - \sum_t H(X_t|Z_t,U,X^{t-1},Z^{t-1})\\ = \ &H(X^n|U) + H(Z^n|X^n,U) - \sum_t H(X_t|Z_t,U,X^{t-1},Z^{t-1})\end{aligned}$$

$$\begin{aligned}
&= H(X^n|U) + H(Z^n|X^n) - \sum_t H(X_t|Z_t,U,X^{t-1},Z^{t-1}) \\
&= H(Z^n|X^n) + \sum_t H(X_t|U,X^{t-1}) - H(X_t|U,X^{t-1},Z_t,Z^{t-1}) \\
&= \sum_t H(Z_t|X_t) + \sum_t I(X_t \wedge Z_t,Z^{t-1}|U,X^{t-1}) \\
&\geq \sum_t (H(Z_t|X_t) + \quad I(X_t \wedge Z_t|U,X^{t-1})) \\
&= \sum_t (H(Z_t|X_t) + H(Z_t|U,X^{t-1}) - H(Z_t|U,X^{t-1},X_t)) \\
&= \sum_t (H(Z_t|U,X^{t-1}) \ ,
\end{aligned}$$

where the last equality follows from application of (i).

To see (iii) we proceed

$$\begin{aligned}
I(U \wedge Z^n) &= H(Z^n) - H(Z^n|U) \\
&\leq \sum_t (H(Z_t) - H(Z_t|U,Z^{t-1})) \\
&\leq \sum_t (H(Z_t) - H(Z_t|U,X^{t-1},Z^{t-1})) \\
&= \sum_t (H(Z_t) - H(Z_t|U,X^{t-1}))
\end{aligned}$$

using (ii) to derive the last equality.
The chain follows considering the definition of U_t.

(iv) is a trivial consequence of (iii) since we observe the equality $H(X^n|U) = \sum_t H(X_t|U_t)$ holding due to the very definition of U_t.

The preceding lemma will be applied to derive a connection between the quantities $H(X^n|U) + I(U \wedge Z^n)$ and $I(U \wedge Z^n)$, where $U \ominus X^n \ominus Z^n$. Giving a lower bound to the first expression we shall obtain a single-letterized upper bound to the second.

5.4 Lemma: For $U \ominus X^n \ominus Z^n$ let X^n and Z^n be connected via a memoryless transition system. Then

$$n^{-1} \left(H(X^n|U) + I(U \wedge Z^n)\right) \geq c$$

implies

$$n^{-1} I(U \wedge Z^n) \leq B(c).$$

Proof: Let $U \ominus X^n \ominus Z^n$ satisfy the assumption.
Using lemma 5.3 , (i) and (ii) we get

$$U_t \ominus X_t \ominus Z_t$$

and $$I(U \wedge Z^n) \leq \sum_t I(U_t \wedge Z_t).$$

This yields

$$\begin{aligned} & I(U \wedge Z^n) \\ \leq & \sum_t I(U_t \wedge Z_t) \\ \leq & \sum_t \sup \{I(U \wedge Z) \mid H(X|U)+I(U \wedge Z) \geq H(X_t|U_t)+I(U \wedge Z)\} \\ = & \sum_t B \left(H(X_t|U_t) + I(U_t \wedge Z_t)\right). \end{aligned}$$

Now concavity of $B(\cdot)$ gives

$$\begin{aligned} & n^{-1} I(U \wedge Z^n) \\ \leq & B(n^{-1} \sum_t H(X_t|U_t) + I(U_t \wedge Z_t)). \end{aligned}$$

Since (iv), lemma 5.3 implies

$$n^{-1} \sum_t H(X_t|U_t) + I(U_t \wedge Z_t) \geq c \text{ , the antitonicity of } B(\cdot)$$

which follows trivially from the definition, shows our statement.

We have now provided all material to formulate and prove the converse. It will be proved by showing that for a subset - discriminating code the number of messages contained gives an upper bound to the subsets which may be distinguished, formally:

5.5 Theorem: Given a subset-discriminating (n,N_0,N,δ)-code with average error probability not exceeding $\lambda \in (0,1)$, then $n^{-1} \log N \hat{=} R$ implies $R_0 \hat{=} n^{-1} \log N_0 \le (1-\lambda)^{-1} B((1-\lambda)\, R - n^{-1} h(\lambda)) + n^{-1} h(\lambda) + \delta$

Proof: Let X^n denote a random-variable uniformly distributed on the set of messages $\{x_1,\dots,x_N\}$. Define a random variable U with values in $\{1,\dots,N_0\}$ by

$$U = i \quad \text{iff} \quad X^n \in F_i$$

and let Z^n be connected with X^n via the DMC W. We firstly observe that U, X^n and Z^n form a markov-chain in this order.

Since

$$\Pr\{U = 1\} = \frac{|F_i|}{N} \le \frac{\frac{N}{N_0} \exp\{n\cdot\delta\}}{N} = \exp\{-n \cdot (R_0 - \delta)\},$$

using the trivial proposition:

$$\bigwedge_{a \in \mathfrak{A}} \Pr\{A = a\} \le \alpha$$

implies

$$H(A) \ge \log \frac{1}{\alpha} ,$$

we conclude

$$H(U) \ge n \cdot (R_0 - \delta) .$$

Now

$$\begin{aligned} n \cdot R_0 &= \log N_0 \\ &\le H(U) + n \cdot \delta \\ &= I(U \wedge Z^n) + H(U|Z^n) + n \cdot \delta \end{aligned}$$

Using Fano's inequality we may continue upperbounding

$$\log N_0 \le I(U \wedge Z^n) + \lambda \cdot \log N_0 + h(\lambda) + n\delta$$

where $h(\cdot)$ denotes the binary entropy function.

Thus

$$(1-\lambda) \log N_0 \leq I(U \wedge Z^n) + h(\lambda) + n \cdot \delta$$

whence

$$(1-\lambda) R_0 - n^{-1} h(\lambda) - \delta \leq n^{-1} I(U \wedge Z^n).$$

On the other hand

$$\begin{aligned} n \cdot R &= \log N \\ &= H(X^n) \\ &= H(X^n, U) \\ &= I(U \wedge Z^n) + H(U | Z^n) + H(X^n | U), \end{aligned}$$

where the third equality holds since U is a function of X^n. Again using Fano's inequality we obtain

$$(1-\lambda) R \leq n^{-1} [I(U \wedge Z^n) + H(X^n | U) + h(\lambda)] .$$

The application of lemma 5.4 with

$c = (1-\lambda) R - n^{-1} h(\lambda)$ results in

$$\begin{aligned} &(1-\lambda) R_0 - n^{-1} h(\lambda) - \delta \\ &\leq n^{-1} I(U \wedge Z^n) \\ &\leq B((1-\lambda) R - n^{-1} h(\lambda)) \end{aligned}$$

thereby yielding

$$R_0 \leq (1-\lambda)^{-1} B((1-\lambda) R - n^{-1} h(\lambda)) + n^{-1} h(\lambda) + \delta.$$

Since this upperbound holds for any $\lambda \in (0,1)$ and $\delta > 0$ we have by the continuity of $B(\cdot)$ as proved in the appendix B the result $R_0 \leq B(R)$.

Appendix A

Until now we gave an inner bound $\mathcal{R}^*$ to the rate-region $\mathcal{R}$ for the discrimination-problem of equally sized subsets by proving a coding theorem. The proof of the converse, the outer bound to the rate-region hinged upon the coincidence of $B^*(\cdot)$ and $B^{**}(\cdot)$.

This supposition has to be verified now, in fact we shall prove a little more, namely the identity of $\mathcal{R}^*$ and $\mathcal{R}^{**}$.

A.1 Lemma

$$\mathcal{R}^* = \mathcal{R}^{**}$$

Proof:

Obviously, the result would be implied by the identity

$$\{(I(U\wedge Z),\ I(U\wedge Z) + H(X|U)) \mid U \ominus X \ominus Z\}$$

$$= \{(I(U\wedge Z),\ I(U\wedge Z) + H(X|U)) \mid U \ominus X \ominus Z,\ \|U\| \leq |\mathcal{X}| + 1\}$$

or, equivalently by

$$\{(H(X|U),\ H(Z|U),\ H(Z)) \mid U \ominus X \ominus Z\}$$

$$= \{(H(X|U),\ H(Z|U),\ H(Z)) \mid U \ominus X \ominus Z, \|U\| \leq |\mathcal{X}| + 1\}$$

The latter identity is obtained basically as a particular case of lemma 3.5 of [3]. It is hardly worth mentioning that going through the proof of lemma 3.5 we observe the distribution of the random variable X remaining fixed when the original random variable U is replaced by a new one with range bounded by $|\mathcal{X}| + 1$. Consequently, since Z is connected with X via the channel W, the distribution of Z also does not alter whence of course $H(Z)$ is fixed too.

As a trivial consequence we obtain

A.2 Corollary:

$$B(\cdot) = B^*(\cdot) = B^{**}(\cdot)$$

Appendix B

For the sake of brevity we assume some familiarity of the reader with the notion of a "correspondence", sometimes also denoted as "multivalued function".

From [2], p. 123 we cite the "maximum theorem" which represents the basis to our proof of continuity of $B(\cdot)$.

Theorem: Let $f : M \to \mathbb{R}$ be a continuous function and $\varphi : T \to M$ a continuous correspondence such that $\varphi(t) \neq \emptyset$ for all $t \in T$, $T \subset \mathbb{R}$. Then

$$h(t) = \max_{m \in \varphi(t)} \{f(m)\}$$

is continuous.

Let M be a compact set and $g : M \to T$ denote a continuous function. Then $\varphi(t) \mathrel{\hat{=}} m \{g(m) \geq t\}$ is closed and as a subset of a compact set compact itself. Restricting the domain of $\varphi(\cdot)$ to the image T of g yields

$$\bigwedge_t \varphi(t) \neq \emptyset$$

The continuity of a correspondence may in case of compact-valuedness of φ equivalently be derived by a criterion reverting to sequencees (see [4], pp. 24 and 27).

It will be seen that from g assumed to be concave (∩) the continuity of φ as defined above is easily derived.

φ is to be shown to be upper- and lower- hemi continuous. Upper-hemi continuity rules out implosions. To verify upper-hemi-continuity at each point t we assume to be given a sequence (t_n) converging to t and for each n an element $m_n \in \varphi(t_n)$ such that $m_n \to m$. By definition of φ we have

$$g(m_n) \geq t_n \quad \text{and}$$

thus by continuity of g

$$g(m) \geq \lim_{n\to\infty} t_n = t ,$$

yielding $m \in \varphi(t)$ as was to be shown. Lower-hemi-continuity excludes explosion of the values of φ. Formally we have to assume $(t_n) \to t$ and $m \in \varphi(t)$ to be given and show the existence of a sequence (m_n) converging to m such that $m_n \in \varphi(t_n)$.

Since $\varphi(t_1) \neq \emptyset$ there is m_1 such that $g(m_1) \geq c_1$. For any $n \in \mathbb{N}$ choose $\lambda_n \in [0,1]$ such that

$$\lambda_n \cdot g(m_1) + (1-\lambda_n)\, g(m) = c_n$$

Since $(c_n) \to c = g(m)$ we may assume $(\lambda_n) \to 0$. Obviously, for

$$m_n \triangleq \lambda_n \cdot m_1 + (1-\lambda_n) \cdot m$$

the sequence (m_n) coverges to m. Further, due to the concavity $(\cap)$ of $g(\cdot)$ we have $g(m_n) \geq c_n$ representing the lower-hemi-continuity of $\varphi(\cdot)$.

Proof of lemma 5.1

Observe that the set P of all probability distributions on $\mathcal{U} \times \mathcal{X} \times \mathcal{Z}$ with $\mathcal{U}$ such that $|\mathcal{U}| \leq |\mathcal{X}| + 1$ and such that the corresponding random variables U, and Z form a markov chain with $P_{Z|X} = W$ is compact. Define $T = \{c \mid c \geq 0$ there exists $U \ominus X \ominus Z$ such that $H(X|U) + I(U\ Z) \geq c\} \subset \mathbb{R}$, for $p = P_{UXZ}$ define $f(p) = I(P_U;P_{Z|U}) = I(U \wedge Z)$ and $g(p) = H(P_{X|U} \mid P_U) + I(P_U;P_{Z|U}) = H(X|U) + I(U \wedge Z)$. Then f and g are continuous and g is concave $(\cap)$ in addition, the latter property was obtained as a by-product in the proof of lemma 5.2. Thus all postulates concerning the "maximum theorem" are met showing the continuity of $B(\cdot) = B^{**}(\cdot)$.

List of References

[1] R. Ahlswede/G.Dueck: "Good codes can be produced by a few permutations", IEEE Trans.Inform. Theory, Vol II-28, pp. 430-443, 1982.

[2] C. Berge: "Espaces topologiques-fonctions multivoques", 2nd edition, Dunod, Paris, 1966.

[3] I. Csiszar/J.Körner: "Information Theory: Coding theorems for discrete memoryless systems", Akademiai Kiado, Budapest, 1982.

[4] W. Hildenbrand: "Core and equilibria of a large economy", Princeton University Press, Princeton, 1974.