An Elementary Proof of the Strong Converse Theorem for the Multiple–Access Channel

Rudolf Ahlswede

1 Introduction

In [1] we determined the capacity region of the multiple–access channel (MAC) by proving a coding theorem and its weak converse. Recently Dueck [2] proved a strong converse theorem in the sense of Wolfowitz [3]. His proof uses the Ahlswede–Gács–Körner [4] method of "blowing up decoding sets" in conjunction with a new "*wringing technique*". This technique makes it now possible to prove strong converses, if the average error probability criterion is genuinely used (as in the case in the results for the MAC mentioned above, c.f. [5]).

In this paper we prove Dueck's result without using the method of "blowing up decoding sets", which is based on non–elementary combinatorial work of Margulis [6].

Our proof follows our old approach of [7] to derive upper bounds on the length of *maximal* error codes. In [7] we considered the TWC, the MAC can be treated in essentially the same way. In conjunction with a suitable "wringing technique" (Lemma 3) this approach becomes applicable also to *average* error codes. The heart of the matter is the fact that codes for the MAC have subcodes with a certain independence structure. Actually even this fact can be understood from a more basic simple principle concerning the comparison of two probability distributions on a product space (Lemma 4). This general principle makes the combinatorial or probabilistic nature of Dueck's technique and our improvement thereof (Lemma 3) fully transparent. It also leads to a somewhat sharper result on coding: Strong converse with $\sqrt{n}\log n$ deviation.

The paper is self–contained and all ideas are explained in detail.

2 The Strong Converse Theorem for the MAC

$\mathcal{X}, \mathcal{Y}$ are the (finite) input alphabets and $\mathcal{Z}$ is the (finite) output alphabet of a MAC with transmission matrix w. For words of length n the transmission probabilities are

$$W(z^n|x^n y^n) = \prod_{t=i}^{n} w(z_t|x_t y_t) \text{ for } x^n = (x_1, \ldots, x_n) \in \mathcal{X}^n = \prod_{1}^{n} \mathcal{X}, y^n \in \mathcal{Y}^n, z^n \in \mathcal{Z}^n. \tag{2.1}$$

A code $(n,M,N,\overline{\lambda})$ for the MAC is a system $\big\{(u_i,v_j,D_{ij}):1\le i\le M, 1\le j\le N\big\}$ with

(a) $u_i\in\mathcal{X}^n, v_j\in\mathcal{Y}^n, D_{ij}\subset\mathcal{Z}^n$ for $1\le i\le M, 1\le j\le N$

(b) $D_{ij}\cap D_{i'j'}=\varnothing$ for $(i,j)\neq(i',j')$

(c) $\frac{1}{MN}\sum\limits_{i=1}^{N}\sum\limits_{j=1}^{M}W(D_{ij}|u_iv_j)\ge 1-\overline{\lambda}$

A pair of non–negative reals (R_1,R_2) is an achievable pair of rates for $\overline{\lambda}\in(0,1)$, if for all sufficiently large n there exist codes $\big(n,\lceil\exp R_1 n\rceil,\lceil\exp R_2 n\rceil,\overline{\lambda}\big)$. $\mathcal{R}(\overline{\lambda})$ denotes the set of those pairs and $\mathcal{R}=\bigcap\limits_{\overline{\lambda}\in(0,1)}\mathcal{R}(\overline{\lambda})$ is called the capacity region. The characterization found in [9], which is somewhat different from the original one in [1], is

$$\begin{aligned}\mathcal{R}=\operatorname{conv}\big\{&(R_1,R_2)\in\mathbb{R}_+^n: R_1\le I(X\wedge Z|Y), R_2\le I(Y\wedge Z|X),\\ &R_1+R_2\le I(XY\wedge Z)\text{ for some indep. } X,Y\big\},\end{aligned}\tag{2.2}$$

where X,Y are input variables, Z is the corresponding output variable, $I(X\wedge Z), I(X\wedge Z|Y)$ etc. denote mutual resp. conditional mutual information, and "conv" stands for the convex hull operation.

Dueck's strong converse theorem states

$$\mathcal{R}(\overline{\lambda})\subset\mathcal{R}\text{ (and hence }\mathcal{R}=\mathcal{R}(\overline{\lambda}))\text{ for }\overline{\lambda}\in(0,1).\tag{2.3}$$

We prove the

Theorem. *For every n and every $(n,M,N,\overline{\lambda})$ code:*

$$(\log M,\log N)\in\big(n+0(\sqrt{n}\log n)\big)\mathcal{R}.$$

The approach of [7] makes use of Augustin's [11] strong converse estimate for one–way channels. Wolfowitz gave in [12] a general lemma for proving strong converses, which he credited as follows: "It is a formalization and slight generalization of methods used by Kemperman, Yoshihara, and the author". We formulate and prove here a slight extension thereof, called packing lemma, which yields also the result of [11]. This way one has one key tool for proving strong converses and also, the paper becomes self–contained.

3 The Packing Lemma and a Bound on Codes for the MAC

Let $\mathcal{K}$ and $\mathcal{L}$ be finite sets and let P be a $|\mathcal{K}|\times|\mathcal{L}|$–stochastic matrix.

An (M,λ)–code is a system $\big\{(u_i,D_i):1\le i\le M\big\}$ with

(a) $u_i \subset \mathcal{K}, D_i \subset \mathcal{L}$ for $1 \leq i \leq M$

(b) $D_i \cap D_{i'} = \varnothing$ for $i \neq i'$

(c) $P(D_i|u_i) \geq 1 - \lambda$ for $1 \leq i \leq M$

For a probability distribution (PD) r on $\mathcal{L}$ and a number $\theta > 0$ define

$$B_k(\theta, r) = \left\{ l \in \mathcal{L} : \frac{P(l|k)}{r(l)} \geq e^\theta \right\} \text{ for } k \in \mathcal{K}. \tag{3.1}$$

Lemma 1. *Suppose that for an* (M, λ)*–code* $\big\{(u_i, D_i) : 1 \leq i \leq M\big\}$ *there exists a PD* r *on* $\mathcal{L}$ *and positive numbers* $\theta_1, \dots, \theta_M$ *such that*

$$\max_{1 \leq i \leq M} \sum_{l \in B_{u_i}(\theta_i, r)} P(l|u_i) < \kappa, \tag{3.2}$$

then

$$M < (1 - \lambda - \kappa)^{-1} \exp\left(\frac{1}{M} \sum_{i=1}^M \theta_i \right), \tag{3.3}$$

provided that $\lambda + \kappa < 1$.

(The case $\theta_i = \theta$ *for* $1 \leq i \leq M$ *is the result of [12]).*

Proof. Consider the code $\{u_i, D_i) : 1 \leq i \leq M\}$ and define for $1 \leq i \leq M$

$$A_i = \left\{ l : l \in D_i, \frac{P(l|u_i)}{r(l)} < e^{\theta_i} \right\}.$$

Then for $l \in A_i$ $e^{\theta_i} r(l) > P(l|u_i)$ and hence

$$e^{\theta_i} r(A_i) > P(A_i|u_i) = P(D_i|u_i) - P(D_i - A_i|u_i) \geq 1 - \lambda - \kappa.$$

It follows that

$$\theta_i \geq \log \frac{1 - \lambda - \kappa}{r(A_i)} \geq \log \frac{1 - \lambda - \kappa}{r(D_i)}$$

and also that

$$\frac{1}{M} \sum_{i=1}^M \theta_l \geq \frac{1}{M} \sum_{i=1}^M \log \frac{1 - \lambda - \kappa}{r(D_i)} \geq \frac{1}{M} \sum_{i=1}^M - \log \frac{1}{M} + \log(1 - \lambda - \kappa) = \log M + \log(1 - \lambda - \kappa).$$

This implies

$$M \leq (1 - \lambda - \kappa)^{-1} \exp\left(\frac{1}{M} \sum_{i=1}^M \theta_i \right).$$

Q.E.D.

Remark 1. The lemma can be further generalized to average error codes. We did not present this more general form, because we have no genuine applications for it.

Since it is necessary to take the convex hull in (2.2) a proof of the Theorem naturally has to involve non–stationary DMC's, which are defined by a sequence $(w_t)_{t=1}^{\infty}$ of $|\mathcal{X}| \times |\mathcal{Z}|$–stochastic matrices and

$$W(z^n|x^n) = \prod_{t=1}^{n} w_t(z_t|x_t) \text{ for every } n = 1, 2, \ldots; \text{ every } x^n \in \mathcal{X}^n; \text{ and every } z^n \in \mathcal{Z}^n \tag{3.4}$$

as transmission probabilities for words. We show next how to prove the familiar strong converse for non–stationary DMC's via Lemma 1. In applying this lemma one has some freedom in the choice of r. Kemperman [10] used $r^{*n} = r_1^* x \ldots x r_n^*$, where r_t^* is the maximizing output distribution for w_t, that is,

$$R(p_t^*, w_t) = \sum_{x,z} p_t^*(x) w_t(z|x) \log \frac{w_t(z|x)}{r_t^*(z)} = \max_{X_t} I(X_t \wedge Z_t) = C_t.$$

For a given (n, M, λ)–code $\big\{(u_i, D_i) : 1 \leq i \leq M\big\}$ Augustin [11] used $r^n = r_1 x \ldots x r_n$, where

$$r_t(z) = \frac{1}{M} \sum_{i=1}^{M} w_t(z|u_{it}) \text{ for } u_i = (u_{i1}, \ldots, u_{in}). \tag{3.5}$$

In order to understood this choice let us choose first r as

$$r(z^n) = \frac{1}{M} \sum_{i=1}^{M} W(z^n|u_i),$$

that is the output distribution corresponding to the "Fano–distribution": $\frac{1}{M}$ probability on each code word u_i.

With $\theta_i = c \sum_{z^n} W(z^n|u_i) \log \frac{W(z^n|u_i)}{r(z^n)}$, c a constant, we get that $\theta = \frac{1}{M} \sum_{i=1}^{M} \theta_i$ is a mutual information up to a constant c. By a suitable choice of c one can derive the *weak* converse by using Lemma 1. One does not get the strong converse, because $\log \frac{W(\cdot|u_i)}{r(\cdot)}$ is not a sum of independent RV's and therefore the variance is too big. r^n is the output distribution obtained by choosing as input distribution

$$p^n = \prod_{t=1}^{n} p_t, p_t(x) = \sum_{i=1}^{M} \frac{1}{M} \delta(u_{it}, x), x \in \mathcal{X}, 1 \leq t \leq n, \tag{3.6}$$

that is the product of 1–dimensional marginal distributions of the "Fano–distribution" and may therefore be called Fano*–distribution. This way one achieves both, the independence property and the "matching" of an information quantity. r^n reflects structural properties of the set of code words, which r^{*n} doesn't.

Now with the choices $\mathcal{K} = \mathcal{X}^n, \mathcal{L} = Z^n, r = r^n, P = W, y = \frac{1-\lambda}{2}$, and for $i = 1, \dots, M$

$$\theta_i = E_{W(\cdot|u_i)} \log \frac{W(\cdot|u_i)}{r^n(\cdot)} + \left(\frac{2}{1-\lambda} \mathrm{Var}_{W(\cdot|u_i)} \log \frac{W(\cdot|u_i)}{r^n(\cdot)} \right)^{1/2}.$$

By Chebyshev's inequality

$$W\big(B_{u_i}(\theta_i, r^n)|u_i\big) \leq \frac{1-\lambda}{2} \text{ for } 1 \leq i \leq M \tag{3.7}$$

and hence Lemma 1 yields

$$M < \frac{2}{1-\lambda} \exp\left\{ \frac{1}{M} \sum_{i=1}^{M} \theta_i \right\}. \tag{3.8}$$

In order to bound the right–side expression set

$$T_1 = \frac{1}{M} \sum_{i=1}^{M} E_{W(\cdot|u_i)} \log \frac{W(\cdot|u_i)}{r^n(\cdot)},$$

$$T_2 = \frac{1}{M} \sum_{i=1}^{M} \left(\mathrm{Var}_{W(\cdot|u_i)} \log \frac{W(\cdot|u_i)}{r^n(\cdot)} \right)^{1/2}$$

Clearly,

$$\begin{aligned} T_1 &= \sum_{i=1}^{M} \sum_{z^n} \frac{1}{M} W(z^n|u_i) \log \frac{W(z^n|u_i)}{r^n(z^n)} \\ &= \sum_{i=1}^{n} \sum_{z} \sum_{x} \frac{1}{M} \delta(u_{it}, x) w_t(z|x) \log \frac{w_t(z|x)}{r_t(z)} \\ &= \sum_{i=1}^{n} I(X_t \wedge Z_t), \text{ where } \Pr(X_t = x) = p_t(x) \end{aligned} \tag{3.9}$$

and Z_t is the corresponding output distribution.

T_2 was bounded in [11] as follows:

By the convexity of the square root function

$$T_2 \leq \left(\sum_{i=1}^{M} \frac{1}{M} \mathrm{Var}_{W(\cdot|u_t)} \log \frac{W(\cdot|u_i)}{r^n(\cdot)} \right)^{1/2}$$

and

$$\begin{aligned} \sum_{i=1}^{M} \frac{1}{M} \mathrm{Var}_{W(\cdot|u_t)} \log \frac{w(\cdot|u_i)}{r^n(\cdot)} &= \sum_{t=1}^{n} \sum_{i=1}^{M} \frac{1}{M} \mathrm{Var}_{w_t(\cdot|u_{it})} \log \frac{w_t(\cdot|u_{it})}{r^t(\cdot)} \\ &= \sum_{t=1}^{n} \sum_{x} \sum_{z} p_t(x) w_t(z|x) \left(\log \frac{w_t(z|x)}{r_t(z)} - E_{w_t(\cdot|x)} \log \frac{w_t(\cdot|x)}{r_t(\cdot)} \right)^2. \end{aligned}$$

Since for any RV F and any constant a $\mathrm{Var}F \leq E(F+a)^2$, the last quantity can be upperbounded by

$$\sum_{t=1}^{n}\sum_{x}\sum_{z} p_t(x)w_t(z|x)\left(\log\frac{w_t(z|x)}{r_t(z)} + \log p_t(x)\right)^2$$
$$= \sum_{t=1}^{n}\sum_{x}\sum_{z} r_t(z)\frac{P_t(x)w_t(z|x)}{r_t(z)}\left(\log\frac{w_t(z|x)p_t(x)}{r_t(z)}\right)^2.$$

Since for a probability vector $(a_1,\dots,a_c)$

$$\sum_{i=1}^{c} a_i\log^2 a_i \leq \max(\log^2 3, \log^2 c),$$

also

$$\sum_x \frac{p_t(x)w_t(z|x)}{r_t(z)}\left(\log\frac{p_t(x)w_t(z|x)}{r_t(z)}\right)^2 \leq \max(\log^2 3, \log^2|\mathcal{X}|) \leq 3|\mathcal{X}|.$$

Thus

$$T_2 \leq (3|\mathcal{X}|n)^{1/2}. \tag{3.10}$$

Thus, (3.9) and (3.8) yield

$$\log M \leq \sum_{t=1}^{n} I(X_t\wedge Z_t) + \left(\frac{2}{1-\lambda}3|\mathcal{X}|n\right)^{1/2} + \log\frac{2}{1-\lambda}$$

and hence the

Corollary 1 (Augustin [11]): *For an (n,M,λ)–code $\{(u_i,D_i): 1\leq i\leq M\}$ for the non–stationary DMC $(w_t)_{t=1}^{\infty}$*

$$\log M \leq \sum_{t=1}^{n} I(X_t\wedge Z_t) + \frac{3}{1-\lambda}|\mathcal{X}|n^{1/2}, 0<\lambda<1, \tag{3.11}$$

where the distributions of the RV's are (as usual) determined by the Fano–distribution on the code words.

Already in [7] we showed how to use Fano–distributions to derive upper bounds on the lengths of codes for the restricted TWC in case of *maximal* errors. We apply this approach now to (n,M,N) codes $\{(u_i,v_j,D_{ij}): 1\leq i\leq M, 1\leq j\leq N\}$ for the MAC with *average* error $\overline{\lambda}$, that is,

$$\frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N} W(D_{ij}|u_i,v_j) = 1-\overline{\lambda}. \tag{3.12}$$

$$\mathcal{A} = \left\{ (i,j) : W(D_{ij}|u_i,v_j) \geq \frac{1-\overline{\lambda}}{2} \triangleq 1-\lambda, 1 \leq i \leq M, 1 \leq j \leq N \right\}. \tag{3.13}$$

$$\begin{aligned} \mathcal{C}(i) &= \big\{(i,j) : (i,j) \in \mathcal{A}, 1 \leq j \leq N\big\}, \\ \mathcal{B}(j) &= \big\{(i,j) : (i,j) \in \mathcal{A}, 1 \leq i \leq M\big\}. \end{aligned} \tag{3.14}$$

Consider the subcode $\big\{(u_i,v_j,D_{ij}) : (i,j) \in \mathcal{A}\big\}$ and define with its Fano–distribution RV's X^n, Y^n

$$\Pr\big((X^n,Y^n) = (u_i,v_j)\big) = |\mathcal{A}|^{-1}, \text{ if } (i,j) \in \mathcal{A}. \tag{3.15}$$

It follows from Corollary 1 that

$$\log|\mathcal{B}(j)| \leq \sum_{t=1}^n I(X_t \wedge Z_t|Y_t = v_{jt}) + \frac{3}{1-\lambda}|\mathcal{X}|n^{1/2}, \tag{3.16}$$

$$\log|\mathcal{C}(i)| \leq \sum_{t=1}^n I(Y_t \wedge Z_t|X_t = u_{it}) + \frac{3}{1-\lambda}|\mathcal{X}|n^{1/2}, \tag{3.17}$$

and

$$\log|\mathcal{A}| \leq \sum_{t=1}^n I(X_tY_t \wedge Z_t) + \frac{3}{1-\lambda}|X|n^{1/2}. \tag{3.18}$$

Since $\mathrm{Prob}(Y_t = y) = |\mathcal{A}|^{-1}\sum_{(i,j)\in\mathcal{A}} \delta(v_{jt},y)$, it follows from (3.16) that

$$\begin{aligned} &|\mathcal{A}|^{-1} \sum_{(i,j)\in\mathcal{A}} \log|\mathcal{B}(j)| \\ &\leq |\mathcal{A}|^{-1} \sum_{(i,j)\in\mathcal{A}} \sum_{t=1}^n I(X_t \wedge Z_t|Y_t = v_{jt}) \sum_y \delta(v_{jt},y) + \frac{3}{1-\lambda}|\mathcal{X}|n^{1/2} \\ &= \sum_{t=1}^n I(X_t \wedge Z_t|Y_t) + \frac{3}{1-\lambda}|\mathcal{X}|n^{1/2}. \end{aligned} \tag{3.19}$$

Since $|\mathcal{A}| + \frac{1-\overline{\lambda}}{2}(MN - |\mathcal{A}|) \geq (1-\overline{\lambda})MN$, we get

$$|\mathcal{A}| \geq \frac{1-\overline{\lambda}}{1+\overline{\lambda}}MN = \left(1 - \frac{2\overline{\lambda}}{1+\overline{\lambda}}\right)MN \triangleq (1-\lambda^*)MN. \tag{3.20}$$

Furthermore,

$$\begin{aligned}
|\mathcal{A}|^{-1} \sum_{(i,j)\in\mathcal{A}} \log|\mathcal{B}(j)| &= |\mathcal{A}|^{-1} \sum_{j=1} |\mathcal{B}(j)| \log|\mathcal{B}(j)| \\
&\geq |\mathcal{A}|^{-1} \sum_{j:|\mathcal{B}(j)|\geq \frac{1-\lambda^*}{n}M} |\mathcal{B}(j)| \log|\mathcal{B}(j)| \\
&\geq |\mathcal{A}|^{-1} \left(|\mathcal{A}| - \frac{1}{n}|\mathcal{A}|\right) \log \frac{1-\lambda^*}{n} M \\
&= \left(1 - \frac{1}{n}\right) \log \frac{1-\lambda^*}{n} M,
\end{aligned}$$

and therefore by (3.19)

$$\begin{aligned}
\log M &\leq \left(1 + \frac{2}{n}\right) \left(\sum_{t=1}^{n} I(X_t \wedge Z_t|Y_t) + \frac{3}{1-\lambda}|\mathcal{X}|n^{1/2}\right) - \log(1-\lambda^*) + \log n \\
&\leq \sum_{t=1}^{n} I(X_t \wedge Z_t|Y_t) + c_1(\overline{\lambda})n^{1/2}.
\end{aligned}$$

Analogously,

$$\log N \leq \sum_{t=1}^{n} I(Y_t \wedge Z_t|X_t) + c_2(\overline{\lambda})n^{1/2}$$

and by (3.18), (3.20) also

$$\log MN \leq \sum_{t=1}^{n} I(Y_t \wedge Z_t|X_t) + c_3(\overline{\lambda})n^{1/2}.$$

Thus we have proved

Lemma 2. *An $(n, M, N, \overline{\lambda})$–code $\{(u_i, v_j, D_{ij}) : 1 \leq i \leq M, 1 \leq j \leq N\}$ for the MAC satisfies for $0 \leq \overline{\lambda} < 1$ and $c(\overline{\lambda})$ suitable*

$$\log M \leq \sum_{t=1}^{n} I(X_t \wedge Z_t|Y_t) + c(\overline{\lambda})n^{1/2},$$

$$\log N \leq \sum_{t=1}^{n} I(Y_t \wedge Z_t|X_t) + c(\overline{\lambda})n^{1/2},$$

$$\log NM \leq \sum_{t=1}^{n} I(X_tY_t \wedge Z_t) + c(\overline{\lambda})n^{1/2},$$

where the distributions of the RV's are determined by the Fano–distribution on the code words $\{(u_i, v_j) : (i,j) \in \mathcal{A}\}$. $\mathcal{A}$ is defined in (3.13).

Remark 2. This does not yet prove the Theorem, because X_t and Y_t are not necessarily independent.

4 Wringing Techniques

To fix some ideas let us quickly recall the attempt of [7], which may be considered as the first "wringing idea". In order to gain the independence of X^n, Y^n mentioned in Remark 2 it would suffice to find for an $(n, M, N, \overline{\lambda})$–code a maximal error subcode of essentially the same rates, that is, a set $\mathcal{A}^* = \mathcal{B}^* \times \mathcal{C}^*$ with $\mathcal{B}^* \subset \{1, \dots, M\}$, $\mathcal{C}^* \subset \{1, \dots, N\}$ such that

$$W(D_{ij}|u_i, v_j) > \varepsilon \text{ for } (i,j) \in \mathcal{A}^* \tag{4.1}$$

and

$$|\mathcal{B}^*| \geq M \exp\big\{-o(n)\big\}, |\mathcal{C}^*| \geq N \exp\big\{-o(n)\big\}. \tag{4.2}$$

Abstractly the problem can be stated as follows:

Given $\mathcal{A} \subset \{1, \dots, M\} \times \{1. \dots, N\}, |\mathcal{A}| \geq \delta MN$, $M = \exp\{R_1 n\}$, $N = \exp\{R_2 n\}$, does there exist an $\mathcal{A}^* = \mathcal{B}^* \times \mathcal{C}^* \subset \mathcal{A}$ satisfying (4.2)?

This is exactly the *problem of Zarankiewics* [13] for certain values of the parameters (there exists an extensive literature on this problem for $|\mathcal{B}^*|, |\mathcal{C}^*|$ small). In [17] we showed that the question has in general a negative answer and Dueck [5] proved that also the reduction to a maximal error subcode is in general impossible, because average and maximal error capacity regions can be different.

Next observe that the existence of subcodes with weaker properties suffices. It is enough that X^n and Y^n are *almost* independent. As a possible approach one might try to achieve this by considering a Quasi–Zarankiewics problem in which the condition $\mathcal{A}^* = \mathcal{B}^* \times \mathcal{C}^* \subset \mathcal{A}$ is replaced by

$$|\mathcal{A}^* \cap \mathcal{B}(j)| \geq (1-\eta)|\mathcal{B}^*|, |\mathcal{A}^* \cap \mathcal{C}(j)| \geq (1-\eta)|\mathcal{C}^*|$$

for $j \in \mathcal{C}^*, i \in \mathcal{B}^*$ and η close to 1.

Selecting $\mathcal{A}$ at random it is readily verified that this is in general again not possible for the parameters specified above.

However, in order to prove the strong converse via Lemma 2 it suffices to find subcodes, whose associated *component* variables X_t, Y_t are almost independent for $t = 1, 2, \dots, n$. The answer is given by Lemma 3 below.

Dueck's original solution is based on a wringing technique, which is slightly weaker (see Remark 3). He doesn't need to produce a sub–code, because he uses instead of Lemma 2 the method of blowing up decoding sets [4] in conjunction with Fano's Lemma.

Lemma 3. *Let X^n, Y^n be RV's with values in $\mathcal{X}^n, \mathcal{Y}^n$ resp. and assume that*

$$I(X^n \wedge Y^n) \leq \sigma.$$

Then for any $0 < \delta < \sigma$ there exist $t_1, \dots, t_k \in \{1, \dots, n\}$, where $0 \leq k < \frac{2\delta}{\sigma}$, such that for some $\overline{x}_{t_1}, \overline{y}_{t_1}, \overline{x}_{t_2}, \overline{y}_{t_2}, \dots, \overline{x}_{t_k}, \overline{y}_{t_k}$

$$I(X_t \wedge Y_t | X_{t_1} = \overline{x}_{t_1}, Y_{t_1} = \overline{y}_{t_1}, \dots, X_{t_k} = \overline{x}_{t_k}, Y_{t_k} = \overline{y}_{t_k}) \leq \delta \tag{4.3}$$

for $t = 1, 2, \ldots, n$, *and*

$$Pr(X_{t_1} = \overline{x}_{t_1}, Y_{t_1} = \overline{y}_{t_1}, \ldots, X_{t_k} = \overline{x}_{t_k}, Y_{t_k} = \overline{y}_{t_k}) \geq \left(\frac{\delta}{|\mathcal{X}||\mathcal{Y}|(2\sigma - \delta)} \right)^k. \tag{4.4}$$

Proof. If (4.3) does not hold already for $k = 0$, then for some t_I $I(X_{t_1} \wedge Y_{t_1}) > \delta$. Since

$$\sigma \geq I(X^n \wedge Y^n) \geq I(X^n \wedge Y^n | X_{t_1} Y_{t_1}) + I(X_{t_1} \wedge Y_{t_1}),$$

we obtain

$$I(X^n \wedge Y^n | X_{t_1} Y_{t_1}) < \sigma - \delta.$$

Set

$$\sigma_1 = \sigma, \varepsilon_1 = \frac{\delta}{2\sigma_1 - \delta}$$

and

$$A_{t_1}(\varepsilon_1) = \left\{ (x_t, y_t) : \Pr(X_t = x_t, Y_t = y_t) \geq \frac{\varepsilon_1}{|\mathcal{X}||\mathcal{Y}|} \right\}.$$

Then

$$\sigma_1 - \delta \geq \sum_{(x_{t_1}, y_{t_1}) \in A_{t_1}(\varepsilon_1)} I(X^n \wedge Y^n | X_{t_1} = x_{t_1}, Y_{t_1} = y_{t_1}) \Pr(X_{t_1} = x_{t_1}, Y_{t_1} = y_{t_1})$$

and since $\Pr\big((X_{t_1}, Y_{t_1}) \notin A_{t_1}(\varepsilon_1)\big) \leq \varepsilon$ there exists an $(\overline{x}_{t_1}, \overline{y}_{t_1}) \in A_{t_1}(\varepsilon_1)$ such that

$$\sigma_1 - \delta \geq I(X^n \wedge Y^n | X_{t_1} = \overline{x}_{t_1}, Y_{t_1} = \overline{y}_{t_1})(1 - \varepsilon_1).$$

Using $(\sigma_1 - \delta)(1 - \varepsilon_1)^{-1} = \sigma_1 - \frac{\delta}{2}$ we get therefore

$$\sigma_1 - \frac{\delta}{2} \geq I(X^n \wedge Y^n | X_{t_1} = \overline{x}_{t_1}, Y_{t_1} = \overline{y}_{t_1}) \tag{4.5}$$

and

$$\Pr(X_{t_1} = \overline{x}_{t_1}, Y_{t_1} = \overline{y}_{t_1}) \geq \frac{\varepsilon_1}{|\mathcal{X}||\mathcal{Y}|}. \tag{4.6}$$

We repeat now the argument with the choices $\sigma_2 = \sigma_1 - \frac{\delta}{2}, \varepsilon_2 = \frac{\delta}{2\sigma_2 - \delta}$.
We are either done or there exists a t_2 with

$$I(X_{t_2} \wedge Y_{t_2} | X_{t_1} = \overline{x}_{t_1}, Y_{t_1} = \overline{y}_{t_1}) > \delta.$$

Then

$$\begin{aligned}
\sigma_2 &\geq I(X^n \wedge Y^n | X_{t_1} = \overline{x}_{t_1}, Y_{t_1} = \overline{y}_{t_1}) \\
&\geq I(X^n \wedge Y^n | X_{t_1} = \overline{x}_{t_1}, Y_{t_1} = \overline{y}_{t_1}, X_{t_2}, Y_{t_2}) + I(X_{t_2} \wedge Y_{t_2} | X_{t_1} = \overline{x}_{t_1}, Y_{t_1} = \overline{y}_{t_1})
\end{aligned}$$

and there exists a pair $(\overline{x}_{t_2}, \overline{y}_{t_2})$ with

$$\sigma_2 - \frac{\delta}{2} \geq I(X^n \wedge Y^n | X_{t_1} = \overline{x}_{t_1}, Y_{t_1} = \overline{y}_{t_1}, X_{t_2} = \overline{x}_{t_2}, Y_{t_2} = \overline{y}_{t_2}) \tag{4.7}$$

and with

$$\Pr(X_{t_2} = \overline{x}_{t_2}, Y_{t_2} = \overline{y}_{t_2} | X_{t_1} = \overline{x}_{t_1}, Y_{t_1} = \overline{y}_{t_1}) \geq \frac{\varepsilon_2}{|\mathcal{X}||\mathcal{Y}|}. \tag{4.8}$$

Iterating the argument with the choices $\sigma_i = \sigma_{i-1} - \frac{\delta}{2}, \varepsilon_i = \frac{\delta}{2\sigma_i - \delta} (i = 3, 4, \dots)$ we obtain either for some $i = k < \frac{2\sigma - \delta}{\delta}$,

$$I(X_t \wedge Y_t | X_{t_1} = \overline{x}_{t_1}, Y_{t_1} = \overline{y}_{t_1}, \dots, X_{t_k} = \overline{x}_{t_k}, Y_{t_k} = \overline{y}_{t_k}) \leq \delta$$

or for $k = \frac{2\sigma}{\lfloor \delta \rfloor}, \sigma_k = \sigma \left(\frac{2\sigma}{\lfloor \delta \rfloor} - 1 \right) \frac{\delta}{2} \leq \delta$, and hence again

$$\begin{aligned} \delta \geq \sigma_k &\geq I(X^n \wedge Y^n | X_{t_1} = \overline{x}_{t_1}, Y_{t_1} = \overline{y}_{t_1}, \dots, X_{t_k} = \overline{x}_{t_k}, Y_{t_k} = \overline{y}_{t_k}) \\ &\geq I(X_t \wedge Y_t | X_{t_1} = \overline{x}_{t_1}, Y_{t_1} = \overline{y}_{t_1}, \dots, X_{t_k} = \overline{x}_{t_k}, Y_{t_k} = \overline{y}_{t_k}) \end{aligned}$$

for $t = 1, \dots, n$.

In any case also

$$\begin{aligned} &\Pr(X_{t_1} = \overline{x}_{t_1}, Y_{t_1} = \overline{y}_{t_1}, \dots, X_{t_k} = \overline{x}_{t_k}, Y_{t_k} = \overline{y}_{t_k}) \\ &\geq \prod_{i=1}^{k} \frac{\varepsilon_1}{|\mathcal{X}||\mathcal{Y}|} = \prod_{i=1}^{k} \frac{\delta}{|\mathcal{X}||\mathcal{Y}|(2\sigma_i - \delta)} \geq \left(\frac{\delta}{|\mathcal{X}||\mathcal{Y}|(2\sigma - \delta)} \right)^k. \end{aligned}$$

Q.E.D.

Remark 3. Dueck's result is that under the assumption of the Lemma

$$I(X_t \wedge Y_t | X_{t_1}, Y_{t_1}, \dots, X_{t_k}, Y_{t_k}) \leq \delta \text{ for } t = 1, 2, \dots, n$$

and some $t_1, \dots, t_k; k < \frac{\sigma}{\delta}$.

In the following it is convenient to adopt the notation:

For a RV $X^n = (X_1, \dots, X_n)$ with values in $\mathcal{X}^n$ and distribution P we define

$$P(x^n) = \Pr(X^n = x^n)$$

and

$$P(x_{s_1}, \dots, x_{s_l} | x'_{t_1}, \dots, x'_{t_m}) = \Pr(X_{s_1} = x_{s_1}, \dots, X_{s_l} = x_{s_l} | X_{t_1} = x'_{t_1}, \dots, X_{t_m} = x'_{t_m})$$

for any not necessarily distinct

$$s_1, \dots, s_l, t_1, \dots, t_m \in \{1, \dots, n\}.$$

Lemma 4. *Let P and Q be probability distributions on $\mathcal{X}^n$ such that for a positive constant c*

$$P(x^n) \leq (1+c)Q(x^n) \text{ for all } x^n \in \mathcal{X}^n, \tag{4.9}$$

then for any $0 < \gamma < c, 0 \leq \varepsilon < 1$ there exist $t_1, \dots, t_k \in \{1, \dots, n\}$, where $0 \leq k \leq \frac{c}{\gamma}$, such that for some $\overline{x}_{t_1}, \dots, \overline{x}_{t_k}$

$$P(x_{t_1}|\overline{x}_{t_1}, \dots, \overline{x}_{t_k}) \leq \max\big((1+\gamma)Q(x_t|\overline{x}_{t_1}, \dots, \overline{x}_{t_k}), \varepsilon\big) \text{ for all } x_t \in \mathcal{X} \text{ and all } t = 1, 2, \dots, n \tag{4.10}$$

and

$$P(\overline{x}_{t_1}, \dots, \overline{x}_{t_k}) \geq \varepsilon^k. \tag{4.11}$$

Proof. If (4.10) does not hold already for $k = 0$, then for some t_1 and some $\overline{x}_{t_1}$

$$P(\overline{x}_{t_1}) > \max\big((1+\gamma)Q(\overline{x}_{t_1}), \varepsilon\big)$$

and we derive from (4.9)

$$(1+c)Q(\overline{x}_{t_1}) \geq P(\overline{x}_{t_1}) > \max\big((1+\gamma)Q(\overline{x}_{t_1}), \varepsilon\big).$$

This insures (4.11) for $k = 1$ and $P(\overline{x}_{t_1}) > (1+\gamma)Q(\overline{x}_{t_1}) > 0$. From (4.9) we can derive therefore

$$P(x^n|\overline{x}_{t_1}) \leq \frac{1+c}{1+\gamma} Q(x^m|\overline{x}_{t_1}) \text{ for all } x^n \in \mathcal{X}^n. \tag{4.12}$$

Repeating the argument we get either $P(x_1|\overline{x}_{t_1}) \leq \max\big((1+\gamma)Q(x_t|\overline{x}_{t_1}), \varepsilon\big)$ for $x_1 \in \mathcal{X}, 1 \leq t \leq n$ (and we are done) or there exists a t_2 and an $\overline{x}_{t_2}$ with

$$\frac{1+c}{1+\gamma} Q(\overline{x}_{t_2}|\overline{x}_{t_1}) \geq P(\overline{x}_{t_2}|\overline{x}_{t_1}) > \max\big((1+\gamma)Q(\overline{x}_{t_2}|\overline{x}_{t_1}), \varepsilon\big).$$

This yields (4.11) for $k = 2$ and implies with (4.12)

$$P(x^n|\overline{x}_{t_1}, \overline{x}_{t_2}) \leq \frac{1+c}{(1+\gamma)^2} Q(x^n|\overline{x}_{t_1}, \overline{x}_{t_2}).$$

Clearly, after k steps (without the procedure having ended before) (4.11) holds and

$$P(x^n|\overline{x}_{t_1}, \overline{x}_{t_2}, \dots, \overline{x}_{t_k}) \leq \frac{1+c}{(1+\gamma)^k} Q(x^n|\overline{x}_{t_1}, \dots, \overline{x}_{t_k}),$$

which implies

$$P(x_t|\overline{x}_{t_1}, \overline{x}_{t_2}, \dots, \overline{x}_{t_k}) \leq \frac{1+c}{(1+\gamma)^k} Q(x_t|\overline{x}_{t_1}, \dots, \overline{x}_{t_k})$$

for all $x_t \in \mathcal{X}, 1 \leq t \leq n$.

Now for $k+1 \geq \frac{c}{\gamma} \geq \frac{\log(1+c)}{\log(1+\gamma)} : \frac{1+c}{(1+\gamma)^k} \leq 1+\gamma$. Q.E.D.

Corollary 2. *Let $\mathcal{A} \subset \{1,\dots,M\} \times \{1,\dots,M\}$, $|\mathcal{A}| \geq (1-\lambda^*)MN$, and let $\{(u_i, v_j, D_{ij}) : (i,j) \in \mathcal{A}\}$ be a code for the MAC with maximal error probability λ.*

Then for any $0 < \gamma < c \triangleq \frac{\lambda^*}{1-\lambda^*}, 0 \leq \varepsilon < 1$ there exist $t_1, \dots, t_k \in \{1,\dots,n\}$, where $k \leq \frac{\lambda^*}{\gamma(1-\lambda^*)}$, and some $(\overline{x}_{t_1}, \overline{y}_{t_1}), \dots, (\overline{x}_{t_k}, \overline{y}_{t_k})$ such that

$$\{(u_i, v_j, D_{ij}) : (i,j) \in \mathcal{A}\} \triangleq \{(u_{it_l} = \overline{x}_{t_l}, v_{jt_l} = \overline{y}_{t_l} \text{ for } 1 \leq l \leq k\} \tag{4.13}$$

is a subcode with maximal error λ and

(a) $|\overline{\mathcal{A}}| \geq \varepsilon^k |\mathcal{A}|, \overline{M} = |\{u_i : (i,j) \in \mathcal{A}\}| \geq \varepsilon^k M,\ \overline{N} = |\{v_j : (i,j) \in \overline{\mathcal{A}}\}| \geq \varepsilon^k N$

(b) $\big((1+\gamma) Pr(\overline{X}_t = x) Pr(\overline{Y}_t = y) - \gamma - |\mathcal{X}||\mathcal{Y}|\varepsilon\big) \leq Pr(\overline{X}_t = x, \overline{Y}_t = y)$
$\leq \max\big((1+\gamma) Pr(\overline{X}_t = x) Pr(\overline{Y}_t = y), \varepsilon\big)$ *for all $x \in \mathcal{X}, y \in \mathcal{Y}, 1 \leq t \leq n$.*

$\overline{X}^n = (\overline{X}_1, \dots, \overline{X}_n), \overline{Y}^n = (\overline{Y}_1, \dots, \overline{Y}_n)$ are distributed according to the Fano–distribution of the subcode.

Proof. Apply Lemma 4 with P as Fano–distribution of the code, that is,

$$P(x^n, y^n) = \Pr(X^n = x^n, Y^n = y^n) = \frac{1}{|\mathcal{A}|}, \text{ if } (x^n, y^n) = (u_i, v_j) \text{ for } (i,j) \in \mathcal{A}$$

and Q defined by

$$Q(x^n, y^n) = \Pr(X^n = x^n) \Pr(Y^n = y^n), (x^n, y^n) \in \mathcal{X}^n \times \mathcal{Y}^n.$$

$\mathcal{X}^n \times \mathcal{Y}^n$ takes the role of $\mathcal{X}^n$ in the Lemma.

Now $Q(x^n, y^n) = 0$ implies $P(x^n, y^n) = 0$, $Q(x^n, y^n) = \frac{1}{|\mathcal{A}|}$ implies $P(x^n, y^n) = \frac{1}{MN}$, and by our assumption on $\mathcal{A}$, $\frac{1}{|\mathcal{A}|} \leq \frac{1}{1-\lambda^*} MN$.

Therefore (4.9) holds with $c = \frac{1}{1-\lambda^*} - 1 = \frac{\lambda^*}{1-\lambda^*}$ and the Lemma yields immediately (a) and the right side inequality in (b). This inequality implies

$$\begin{aligned}
\Pr(\overline{X}_t = x, \overline{Y}_t = y) &= 1 - \sum_{(x',y') \neq (x,y)} \Pr(\overline{X}_t = x', \overline{Y}_t = y') \\
&\geq 1 - \sum_{(x',y') \neq (x,y)} \max\big((1+\gamma)\Pr(\overline{X}_t = x', \overline{Y}_t = y'), \varepsilon\big) \\
&\geq 1 - |\mathcal{X}||\mathcal{Y}|\varepsilon - (1+\gamma)\big(1 - \Pr(\overline{X}_t = x, \overline{Y}_t = y)\big) \\
&= (1+\gamma)\Pr(\overline{X}_t = x, \overline{Y}_t = y) - \gamma - |\mathcal{X}||\mathcal{Y}|\varepsilon.
\end{aligned}$$

Q.E.D.

5 Proof of the Theorem

We simply have to combine Lemma 2 and Corollary 2.

For an $(n, M, N, \overline{\lambda})$ code $\big\{(u_i, v_j, D_{ij}) : 1 \leq i \leq M, 1 \leq j \leq N\big\}$ define $\mathcal{A}$ as in (3.13). Then $|\mathcal{A}| \geq (1-\lambda^*)MN$ for $\lambda^* = \frac{2\overline{\lambda}}{1+\overline{\lambda}}$. Apply Corollary 2 with the parameters

$$\gamma = n^{-1/2}, \varepsilon = n^{-1}. \tag{5.1}$$

Thus for some $k \leq \frac{\lambda^*}{1-\lambda^*} n^{1/2}$

$$|\overline{\mathcal{A}}| \geq \varepsilon^k |\mathcal{A}| \geq n^{-\lambda^* n^{1/2}/(1-\lambda^*)}(1-\lambda^*)M, \overline{N} \geq n^{-\lambda^* n^{1/2}/(1-\lambda^*)}. \tag{5.2}$$

Application of Lemma 2 to this subcode yields

$$\begin{aligned}
\log M &\leq \frac{\lambda^*}{1-\lambda^*} n^{1/2} \log n + \log \overline{M} \\
&\leq \sum_{t=1}^n I(\overline{X}_t \wedge \overline{Z}_t | \overline{Y}_t) + C(\overline{\lambda}) n^{1/2} \log n \\
\log N &\leq \sum_{t=1}^n I(\overline{Y}_t \wedge \overline{Z}_t | \overline{X}_t) + C(\overline{\lambda}) n^{1/2} \log n \\
\log MN &\leq \sum_{t=1}^n I(\overline{X}_t \overline{Y}_t \wedge \overline{Z}_t) + C(\overline{\lambda}) n^{1/2} \log n
\end{aligned}$$

with

$$C(\overline{\lambda}) = c(\overline{\lambda}) + \frac{\lambda^*}{1-\lambda^*} - \log(1-\lambda^*).$$

Since

$$\begin{aligned}
I(\overline{X}_t \overline{Y}_t \wedge \overline{Z}_t) &= H(\overline{X}_t \overline{Y}_t) + H(\overline{Z}_t) - H(\overline{X}_t \overline{Y}_t \overline{Z}_t), \\
I(\overline{X}_t \wedge \overline{Z}_t | \overline{Y}_t) &= I(\overline{X}_t \overline{Y}_t \wedge \overline{Z}_t) - I(\overline{X}_t \wedge \overline{Z}_t) \\
&= H(\overline{X}_t, \overline{Y}_t) - H(\overline{X}_t \overline{Y}_t \overline{Z}_t) - H(\overline{X}_t) + H(\overline{X}_t \overline{Z}_t)
\end{aligned}$$

etc., using (b) we complete the proof by showing that for $n^{-1/2} \geq |\mathcal{X}||\mathcal{Y}| n^{-1}$

$$|H(\overline{X}_t, \overline{Y}_t) - H(\overline{\overline{X}}_t, \overline{\overline{Y}}_t)| \leq \text{const. } n^{-1/2} \log n \text{ etc.}, \tag{5.3}$$

where $\Pr(\overline{\overline{X}}_t = x, \overline{\overline{Y}}_t = y) = \Pr(\overline{X}_t = x, \overline{Y}_t = y)$.

Clearly,

$$(1 + n^{-1/2}) \Pr(\overline{X}_t = x) \Pr(\overline{Y}_t = y) - 2n^{-1/2} \leq \Pr(\overline{X}_t = x, \overline{Y}_t = y)$$

$$\leq (1+n^{-1/2})\Pr(\overline{X}_t = x)\Pr(\overline{Y}_t = y) + n^{-1}$$

and hence

$$|\Pr(\overline{X}_t = x)\Pr(\overline{Y}_t = y) - \Pr(\overline{X}_t = x, \overline{Y}_t = y)| \leq 2n^{-1/2}. \tag{5.4}$$

This implies with

$$\Pr(\overline{\overline{Z}}_t = z | \overline{\overline{X}}_t = x, \overline{\overline{Y}}_t = y) = w(z|xy) = \Pr(\overline{Z}_t = z | \overline{X}_t = x, \overline{Y}_t = y)$$

$$|\Pr(\overline{X}_t = x, \overline{Y}_t = y, \overline{Z}_t = z) - \Pr(\overline{\overline{X}}_t = x, \overline{\overline{Y}}_t = y, \overline{\overline{Z}}_t = z)| \leq 2 \cdot n^{-1/2}$$
$$\text{for } x \in \mathcal{X}, y \in \mathcal{Y}, z \in \mathcal{Z}. \tag{5.5}$$

For $0 \leq a \leq b \leq a + \text{const. } n^{-1/2} \leq 1$ obviously

$$|a \log a - b \log b| \leq \text{const. } n^{-1/2} \log n. \tag{5.6}$$

This and (5.5) imply (5.3). Q.E.D.

Remark 4. Using Lemma 3 instead of Lemma 4, one can proceed as follows:

1. One shows that for X^n, Y^n associated with the code $I(X^n|Y^n) \leq \sigma = f(\overline{\lambda})$.
2. Application of Lemma 3 and the analogue of Corollary 3 gives a subcode with the usual desired properties and $I(\overline{X}_t \wedge \overline{Y}_t) \leq \delta$ for $1 \leq t \leq n$. Since $I(\overline{X}_t \wedge \overline{Y}_t)$ is an I–divergence Pinsker's inequality implies

   $$\sum_{x,y} |\Pr(\overline{X}_t = x, \overline{Y}_t = y) - \Pr(\overline{X}_t = x)\Pr(\overline{Y}_t = y)| \leq 2\delta^{1/2}.$$

   For $\delta = n^{-1/2}$ this approach yields a strong converse with the weaker $n^{3/4} \log n$–deviation.

Remark 5. The fact that our question concerning the Quasi–Zarankiewicz problem has a negative answer has also the consequence that the conclusion in Lemma 4 cannot be replaced by

$$P(x^n | \overline{x}_{t_1}, \dots, \overline{x}_{t_k}) \leq \max\big((1+\gamma)Q(x^n | \overline{x}_{t_1}, \dots, \overline{x}_{t_k}|, \varepsilon\big) \tag{4.10*}$$

for all $x^n \in \mathcal{X}^n$ and $\overline{x}_{t_1}, \dots, \overline{x}_{t_k}$ suitable and (4.11) if for instance $\varepsilon \geq 1/n$.

6 References

[1] R. Ahlswede, “Multi–way communication channels”, Second Intern. Sympos. on Inf. Theory, Thakadsor, 1971, Publ. House of the Hungarian Acad. of Sciences, 23–52, 1973.

[2] G. Dueck, “The strong converse to the coding theorem for the multiple–access channel”, Preprint.

[3] J. Wolfowitz, “The coding of messages subject to chance errors”, Illinois J. Math., 4, 591–606, 1957.

[4] R. Ahlswede, P. Gács, and J. Körner, “Bounds on conditional probabilities with applications in multi–user communication”, Z. Wahrscheinlichkeitstheorie u. verw. Gebiete, 34, 157–177, 1976.

[5] G. Dueck, “Maximal error capacity regions are smaller than average error capacity regions for multi–user channels”, Problems of Control and Inf. Theory, 7(1), 11-19, 1978.

[6] G.A. Margulis, “Probabilistic characteristics of graphs with large connectivity”, Problemi Perdachi Informatsii, 10, 101–108, 1974 (Russian).

[7] R. Ahlswede, “On two–way communication channels and a problem by Zarankiewics”, Trans. Sixth Prague Conf. on Inf. Theory, Stat. Dec. Fct’s, Rand. Proc., Sept. 1971, Publ. Hourse of the Czechoslovakian Acad. of Sc., 23–37, 1974.

[8] J. Wolfowitz, *Coding Theorems of Information Theory*, Springer, Berlin–Heidelberg–New York, 3rd edition, 1978.

[9] R. Ahlswede, “The capacity region of a channel with two senders and two receivers”, Ann. of Probability, 2(5), 805–814, 1974.

[10] J.H.B. Kemperman, “Studies in coding theory”, Mimeogr. Notes.

[11] U. Augustin, “Gedächtnisfreie Kanäle für diskrete Zeit”, Z. Wahrscheinlichkeitstheorie u. verw. Gebiete, 6, 10–61, 1966.

[12] J. Wolfowitz, “Note on a general strong converse”, 12, 1–4, 1968.

[13] K. Zarankiewicz, “Problem P 101”, Golloq. Math., 2, 301, 1951.