General theory of information transfer: updated

Rudolf Ahlswede

Universität Bielefeld
Fakultät für Mathematik
Postfach 100131
33501 Bielefeld
Germany

Abstract

We report on ideas, problems and results, which occupied us during the past decade and which seem to extend the frontiers of information theory in several directions. The main contributions concern information transfer by channels. There are also new questions and some answers in new models of source coding. While many of our investigations are in an explorative state, there are also hard cores of mathematical theories. In particular we present a unified theory of information transfer, which naturally incorporates Shannon's theory of information transmission and the theory of identification in the presence of noise as extremal cases. It provides several novel coding theorems. On the source coding side we introduce data compression for identification. Finally we are led beyond information theory to new concepts of solutions for probabilistic algorithms.

The original paper [43] gave to and received from the ZIF-project essential stimulations which resulted in contributions added as GTIT-Supplements "Search and channels with feedback" and "Noiseless coding for multiple purposes: a combinatorial model".

Other contributions - also to areas initiated - are published in the recent book [57].

The readers are advised to study always the pioneering papers in a field - in this case the papers [16], [17] on identification. It is not only the most rewarding way to come to new ideas, but it also helps to more quickly grasp the more advanced formalisms without going through too many technicalities. Perhaps also the recent Shannon Lecture [58], aiming at an even wider scope, gives further impetus.

Footnote: Some of the ideas and results have been presented at the Tenth Symp. on Inf. Theory in the Benelux, Houthalen, Belgium May 25–26, 1989, at the Inf. Theory Meeting in Oberwolfach, April 1992, at an IEEE Workshop on Inf. Theory in San Salvador, Brazil, June 1992, at the ISIT San Antonio, Texas, Jan. 1993, the IEEE Workshop on Inf. Theory in Rydzyna, Poland, June 1995, and at the World Congress of the Bernoulli Society, Vienna, August 1996.

0. Contents

Part I: One sender answering several questions of receivers

Part II: Models with prior knowledge of the receiver

Part III: Models with prior knowledge at the sender

Part IV: Identification and transmission with multi–way channels

Part V: Data Compression

Part VI: Perspectives

References

1 Introduction

We have included in the references several articles and books ([19], [20], [21], [23], [24]), which deal with information not just in a more or less technical engineering sense. They are meant to enlarge our horizon, stimulate our awareness of what is unknown about "information", and to bring us into the spirit for new adventures. Some questions from [23] give indications of the kind of thoughts which took us into their chains.

In the Appendix of [23] one finds the following definition or explication of the concept "communication":

"The establishment of a social unit from individuals, by the shared usage of language or signs. The sharing of common sets of rules, for various goal–seeking activities. (There are many *shades of opinions.*)"

Again in [23] on page 41 we read:

"Perhaps the most important technical development which has assisted in the birth of communication theory is that of telegraphy. With its introduction the speed of transmission of "intelligence" arose. When its economic value was fully realized, the problems of compressing signals exercised many minds, leading eventually to the concept of "quantity of information" and to theories of times and speed of signalling."

and on page 43:

"Hartley went further and defined information as the successive selection of signs or words from a given list, rejecting all "meaning" as a more subjective factor (it is the signs we transmit, or physical signs; we do not transmit their "meaning"). He showed that a message of N signs chosen from an "alphabet" or code book of S signs has S^N possibilities and that the "quantity of information" is most reasonably defined as the logarithm, that is, $H = N \log S$."

This concept of information is closely related to the idea of selection, or discrimination and therefore sometimes called selective–information. It is also at the very basis of Shannon's celebrated statistical theory of communication [1].

This theory has by now been developed into a sophisticated mathematical discipline with many branches and facets. Sometimes more concrete engineering problems led to or gave the incentive to new directions of research and in other cases new discoveries were made by exploring inherent properties of the mathematical structures. Some of our views on the state of this theory, to which we also shall refer as the "Shannon Island", are expressed in [8].

The price for every good theory is simplification and its permanent challenge is reality.

“We live in a world vibrating with information” and in most cases we don’t know how the information is processed or even what it is at the semantic and pragmatic levels. How does our brain deal with information? It is still worthwhile to read von Neumann’s ideas about this [21].

Cherry writes on page of [23]:

“It is remarkable that human communication works at all, for so much seems to be against it; yet it does. The fact that it does depends principally upon the vast store of habits which one of us possess, the *imprints of all our past experiences*. With this, we can hear snatches of speech, the vague gestures and grimaces, and from this shreds of evidence we are able to make a continual series of inferences, guesses, with extra ordinary effectiveness.”

We shall discuss the issue of “prior knowledge” later and we shall show that some aspects are accessible to a rigorous mathematical treatment.

There are various stimuli concerning the concepts of communication and information from the sciences, for instance from quantum theory in physics, the theory of learning in psychology [19], theories in linguistics [40], etc.

These hints give an idea of the size of the ocean around the Shannon Island.

We don’t have the intention to drown in this ocean. However, since the ocean is large there ought to be some other islands. In fact there are.

Among those, which are fairly close to the Shannon Island we can see

1.) Mathematical Statistics
2.) Communication Networks
3.) Computer Storage and Distributive Computing
4.) Memory Cells

Since those islands are close there is hope that they can be connected by dams.

A first attempt to explore connections between Multi–user source coding and hypothesis testing was made in [10]. For interesting ideas about relations between Multiple–access channels and communication networks see Gallager [24]. A multitude of challenges to Information Theory comes from Computer Science. A proper frame for storage in memory cells is our abstract coding theory [8]. Our work on identification has led us to reconsider the basic assumptions of Shannon’s Theory. It deals with “messages”, which are elements of a *prescribed set of objects*, known to the communicators. The receiver wants to know the true message. This basic model occurring in all engineering work on communication channels and networks addresses a very special communication situation. More generally they are characterized by

(I) The questions of the receivers concerning the given "ensemble", to be answered by the sender(s)
(II) The prior knowledge of the receivers
(III) The senders prior knowledge.

Accordingly the paper starts with three parts.

It seems that the whole body of present day Information Theory will undergo serious revisions and some dramatic expansions. We open several directions of future research and start the mathematical description of communication models in great generality. For some specific problems we provide solutions or ideas for their solutions.

We continue in Part IV with (promised) capacity theorems for identification via multi–way channels. We also study identification in conjunction with transmission.

The proof of the "polynomial" weak converse is new even for the discrete memoryless channel (DMC).

In Part V we discuss a new direction of research on sources, which goes back to a problem of [15]: noiseless coding for multiple purposes. It stimulated to go for a new concept: identification for sources.

Part VI concludes with striking results on the relation of identification and common randomness and a general discussion.

Part I: One sender answering several questions of receivers

2 A general communication model for one sender

To simplify matters we assume first that the noise is modelled by a DMC with finite input (resp. output) alphabet $\mathcal{X}$ (resp. $\mathcal{Y}$) and transmission matrix W.

The goal in the classical Shannon communication theory is to transmit many messages reliably over this channel. This is done by coding. An (n, M, λ)–code is a system of pairs $\Big\{(u_i, \mathcal{D}_i) : 1 \leq i \leq M\Big\}$ with $u_i \in \mathcal{X}^n, \mathcal{D}_i \subset \mathcal{Y}^n$ and

$$\mathcal{D}_i \cap \mathcal{D}_{i'} = \varnothing \text{ for } i \neq i', \tag{2.1}$$

$$W^n(\mathcal{D}_i^c | u_i) \leq \lambda \text{ for } i = 1, \dots, M. \tag{2.2}$$

Given a set of messages $\mathcal{M} = \{1, \dots, M\}$, by assigning i to codeword u_i we can transmit a message from $\mathcal{M}$ in blocklength n over the channel with a maximal error probability less than λ. Notice that the underlying assumption in this classical transmission problem is that both, sender and receiver, know that the message is from a specified set $\mathcal{M}$. They also know the code. The receiver's goal is to get to know the message sent. Having received an element in decoding set $\mathcal{D}_i$ he decides for codeword u_i and then for message i. By the assumptions his (maximal) error probability is bounded by λ.

An (n, M, λ) transmission code **with randomization** assigns to message i a probability distribution P_i on $\mathcal{X}^n$, for which

$$\sum_{x^n \in \mathcal{X}^n} W^n(\mathcal{D}_i^c | x^n) P_i(x^n) \leq \lambda.$$

Observe that for some $v_i \in \mathcal{X}^n$

$$W^n(\mathcal{D}_i^c | v_i) \leq \sum_{x^n \in \mathcal{X}^n} W^n(\mathcal{D}_i^c | x^n) P_i(x^n) \leq \lambda$$

and that therefore the code $\{(P_i, \mathcal{D}_i) : 1 \leq i \leq M\}$ with randomization in the encoding can be replaced by the (deterministic) code $\{(v_i, \mathcal{D}_i) : 1 \leq i \leq M\}$ satisfying also the bound λ on the error probability. Obviously the same reduction holds for channels without time structure.

This implies that randomization is of no advantage for transmission over one-way channels like the DMC. However, **it has a dramatic effect on performance for identification**. To fix ideas, transmission concerns the question "How many messages can we transmit over a noisy channel?" One tries to give an answer to the question "What is the actual message from $\mathcal{M} = \{1, \dots, M\}$?"

On the other hand in identification it is asked "How many possible messages can the receiver of a noisy channel identify?" One tries to give an answer to the question "Is the actual message i?" Here i can be any member of the set of possible messages $\mathcal{N} = \{1, 2, \dots, N\}$.

Certain error probabilities are again permitted. From the Theory of Transmission one cannot derive answers for these questions in the Theory of Identification, which therefore goes beyond Shannon's Theory.

An (n, N, λ) identification code for the DMC with transmission probability matrix W is a system of pairs $\Big\{(P_i, \mathcal{D}_i) : 1 \leq i \leq N\Big\}$ with $P_i \in \mathcal{P}(\mathcal{X}^n)$ and $\mathcal{D}_i \subset \mathcal{X}^n$ with error probability of misacceptance and also misrejection less than λ, that is,

$$\sum_{x^n} P_i(x^n) W^n(\mathcal{D}_i|x^n) > 1 - \lambda \text{ for all } i \text{ and } \sum_{x^n} P_i(x^n) W^n(\mathcal{D}_j|x^n) < \lambda \text{ for } i \neq j.$$

We know from [16] that any (second order) rate $R < C_{Sh} = C$ is achievable for any $\lambda > 0$ and all large n, that is, there are (n, N, λ) codes with $R \leq \frac{1}{n} \log \log N$.

It is convenient to introduce the maximal code size

$$N(n, \lambda) = \max\Big\{N : (n, N, \lambda) \ \text{ code exists } \Big\}.$$

Already in [16] it was shown that for any exponentially small sequence of error probabilities $\lambda_n = e^{-\varepsilon n}$ $(\varepsilon > 0)$

$$\overline{\lim_{n \to \infty}} \frac{1}{n} \log \log N(n, \lambda_n) \leq C.$$

This converse was named soft converse in [16]. We use here the more instructive name "exponential weak converse".

The (classical) weak converse states that

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

As a statement between these two we introduce now a *polynomial weak converse*:

For some $\alpha > 0$

$$\overline{\lim_{n \to \infty}} \frac{1}{n} \log \log N\left(n, \frac{1}{n^\alpha}\right) \leq C.$$

Such a statement was derived for $\alpha = 1$ in [35].

Again already in [16] a version of the strong converse was conjectured:

$$\overline{\lim_{n \to \infty}} \frac{1}{n} \log \log N(n, \lambda) \leq C \quad \text{for} \quad 0 \leq \lambda < 1/2.$$

In case of feedback this was proved in [17] and the conjecture of [16] was established by Han/Verdú [28] and with a simpler proof in [29].

Remark 1: The **capacity concept** used in [16], [17] is often called **pessimistic capacity**, that is, the maximal rate achievable with arbitrary small **constant** error probability λ. Sometimes in the literature also the optimistic capacity $\bar{C}$ is used. Actually for many channels (like for instance non-stationary memoryless channels) other performance criteria like **capacity functions** say more about them. This is discussed in great detail in [62]. In this paper we discuss only pessimistic capacities C, C_{pol}, and C_{exp} where the latter are defined as optimal rates achievable for all polynomial error probabilities $\lambda_n = n^{-\alpha}$, $\alpha > 0$, resp. exponential error probabilities $\lambda_n = 2^{-\epsilon n}$ with some small $\epsilon > 0$. It is important to notice that in order to establish a number as the (pessimistic) capacity neither strong nor weak converses are necessary. Furthermore, $C \geq C_{pol} \geq C_{exp}$ and for instance for the DMC it is easy to prove that $C_{exp} \geq C$ and these capacities are equal. The same holds for regions of the multiple access channel (MAC) and can also be shown for **regions for identification** following the direct proofs of [16], [17] which are based on transmission codes and for maximal errors can be improved also by the Ahlswede/Dueck local converse [63]. It is essential that one stays near to memoryless channels; in general the concepts go apart.

One can conceive of many situations in which the receiver has (or many receivers have) different goals. They lead to decoding rules with not necessarily disjoint decoding sets.

A nice class of such situations can, abstractly, be described by a family $\Pi(\mathcal{M})$ of partitions of $\mathcal{M}$. Each $\pi \in \Pi(\mathcal{M})$ is associated with a receiver, who wants to know only which member of the partition $\pi = (A_1, \dots, A_r)$ contains m, the true message, which is known to the encoder.

We describe now some seemingly natural families of partitions.

Model 1: $\Pi_S = \{\pi_{Sh}\}$, $\pi_{Sh} = \Big\{\{m\} : m \in \mathcal{M}\Big\}$. This describes Shannon's classical transmission problem stated above.

Model 2: $\Pi_I = \{\pi_m : m \in \mathcal{M}\}$ with $\pi_m = \Big\{\{m\}, \mathcal{M} \smallsetminus \{m\}\Big\}$. Here decoder π_m wants to know whether m occurred or not. This is the identification problem introduced in [16].

Model 3: $\Pi_K = \{\pi_\mathcal{S} : |\mathcal{S}| = K, \mathcal{S} \subset \mathcal{M}\}$ with $\pi_\mathcal{S} = \{\mathcal{S}, \mathcal{M} \smallsetminus \mathcal{S}\}$. This is an interesting generalisation of the identification problem. We call it K–identification.

This case also arises in several situations. For instance every person π_S may have a set S of K closest friends and the sender knows that one person $m \in \mathcal{M}$ is sick. All persons π_S want to know whether one of their friends is sick.

Model 4: $\Pi_R = \Big\{\pi_r : \pi_r = \Big\{\{1,\dots,r\},\{r+1,\dots,M\}\Big\}, 1 \leq r \leq M-1\Big\}$. Here decoder π_r wants to know whether the true message exceeds r or not. We speak of the ranking problem.

Model 5: $\Pi_B = \{\pi_{\mathcal{A}} : \mathcal{A} \subset \mathcal{M}\}$. A receiver associated with $\pi_{\mathcal{A}} = \{\mathcal{A}, \mathcal{M} \smallsetminus \mathcal{A}\}$ wants to know the answer to the binary question "Is m in $\mathcal{A}$?" (Here, of course, $\pi_{\mathcal{A}}$ and $\pi_{\mathcal{M}\smallsetminus\mathcal{A}}$ can be viewed as the same questions).

Model 6: $\mathcal{M} = \{0,1\}^\ell$, $\Pi_C = \{\pi_t : 1 \leq t \leq \ell\}$ with $\pi_t = \Big\{\{(x_1,\dots,x_\ell) \in \mathcal{M} : x_t = 1\}, \{(x_1,\dots,x_\ell) \in \mathcal{M} : x_t = 0\}\Big\}$. Decoder π_t wants to know the t–th component of the vector valued message $(x_1,\dots,x_\ell)$.

In all these models we can consider the first (or second) order capacities, defined analogously to those in models 1,2, where they are known from [16] and [17]. It is shown in Section 4 that for models 4 and 5 the capacities equal Shannon's transmission capacity.

The most challenging problem is the general K–identification problem of model 3. Here an (n,N,K,λ)–code is a family of pairs $\Big\{(P(\cdot|i), \mathcal{D}_\pi) : 1 \leq i \leq N, \pi \in \Pi_K\Big\}$, where the $P(\cdot|i)$'s are PD's on $\mathcal{X}^n$, $\mathcal{D}_\pi \subset \mathcal{Y}^n$, and where for all $\pi = \{S, \mathcal{M} \smallsetminus S\}$ $\Big(S \in \binom{\mathcal{M}}{K}\Big)$

$$\begin{aligned}\sum_{x^n} P(x^n|i) W^n(\mathcal{D}_\pi^c|x^n) &\leq \lambda \ \text{ for all } \ i \in S,\\ \sum_{x^n} P(x^n|i) W^n(\mathcal{D}_\pi|x^n) &\leq \lambda \ \text{ for all } \ i \notin S.\end{aligned} \tag{2.3}$$

We also write $\mathcal{D}_S$ instead of $\mathcal{D}_\pi$. A coding theorem is established in Section 3.

Remark 2: K-identification applies whenever persons want to know whether a winner is among their favourite teams or lottery numbers or friends.

Remark 3: Most models fall into the following category of regular transfer models. By this we mean that the set of partitions Π of $\mathcal{M}$ is invariant under all permutations $\sigma : \mathcal{M} \to \mathcal{M}$:

$\pi = (A_1,\dots,A_r) \in \Pi$ implies $\sigma\pi = \Big(\sigma(A_1),\dots,\sigma(A_r)\Big) \in \Pi$.

Remark 4: Many of the models introduced concern bivariate partitions. More generally they are described by a hypergraph $\mathcal{H} = (\mathcal{M}, \mathcal{E})$, where decoder $E, E \in \mathcal{E}$, wants to know whether the m occurred is in E or not.

Example 1: In a certain lottery a player can choose ℓ of the numbers $1,\dots,L$, say, $\{a_1,\dots,a_\ell\}$. A set $\{b_1,\dots,b_\ell\}$ of ℓ numbers is chosen at random.

Suppose that T players have chosen $\{a_1^1, \dots, a_\ell^1\}, \dots, \{a_1^T, \dots, a_\ell^T\}$, resp. Every player wants to know whether he won, that shall mean, whether he has at least $\ell - 1$ correct numbers: For the t–th player

$$|\{a_1^t, \dots, a_\ell^t\} \cap \{b_1, \dots, b_\ell\}| \geq \ell - 1.$$

How many bits have to be transmitted in a randomized encoding, so that every player knows with high probability, whether he won.

Example 2: Lets view the elements of $\{1, \dots, a\}^n$ as sequences of events. Historians (or observers of stockmarkets) have each their subsequence of events, say,

$$(t_1^1, \dots, t_{s_1}^1), \dots, (t_1^\ell, \dots, t_{s_\ell}^\ell).$$

The ℓ persons are to be informed with high probability correctly about the correct sequence of events. (Idea of binning, see [6], [7], [8]).

Example 3: In some countries 40% of the healthy men of an age–class are drafted by random selection. Every candidate wants to know with high probability correctly whether he is among them. This falls under model 6.

3 Analysis of a specific model: K–identification

3.1 A relation to standard identification

Recall the definition of an (n, N, K, λ)–code given in Section 2. For reasons, which become apparent soon, we assume K to grow exponentially in the blocklength n, that is,

$$K = 2^{\kappa \cdot n}, \tag{3.1}$$

where κ is called a first order rate.

As for the standard identification problem $(K = 1, \kappa = 0)$ N can grow double exponentially, that is,

$$N = 2^{2^{Rn}}, R > 0, \tag{3.2}$$

where R is called a second order rate.

The pair (R, κ) is achievable, if for any $\lambda > 0, \delta > 0$ and all sufficiently large n $\left(n, 2^{2^{(R-\delta)n}}, 2^{(\kappa-\delta)n}, \lambda\right)$ –codes exist.

Proposition 1. For every DMC the set $\mathcal{K}$ of all achievable rate pairs contains

$$\Big\{(R, \kappa) : 0 \leq R, \kappa; R + 2\kappa \leq C_{Sh}\Big\},$$

where C_{Sh} is Shannon's familiar capacity of the DMC.

Proof: In [16] the achievable triples (R, η_1, η_2) of second order rate R and error exponents η_1, η_2 have been investigated. Theorem 2 of [16] completely characterizes the set of achievable pairs (R, η_2) in the limit $\eta_1 \to 0$ as follows:

$$\lim_{\eta_1 \to 0} \Big\{ (R, \eta_2) : (R_1, \eta_1, \eta_2) \text{ is achievable} \Big\} = \Big\{ (R, \eta_2) : R \leq C_{Sh} - 2\eta_2 \Big\}. \quad (3.3)$$

Now, any identification code $\Big\{ (P_i, \mathcal{D}_i) : 1 \leq i \leq N \Big\}$ with parameters (R, η_1, η_2) has an associated K–identification code $\Big\{ \mathcal{P}_i, \mathcal{D}_S : 1 \leq i \leq N, S \in \binom{N}{K} \Big\}$, where

$$\mathcal{D}_S = \bigcup_{i \in S} \mathcal{D}_i, \quad (3.4)$$

meeting the parameters $(R, \kappa, \eta_1, \eta_2 - \kappa)$.

This means that

$$\begin{aligned} &\sum_{x^n} P_i(x^n) W^n(\mathcal{D}_S | x^n) \geq 1 - 2^{-n\eta_1} \text{ for all } i \in S \text{ and} \\ &\sum_{x^n} P_i(x^n) W^n(\mathcal{D}_S | x^n) \leq K 2^{-n\eta_2} = 2^{-n(\eta_2 - \kappa)} \text{ for all } i \notin S. \end{aligned}$$

These inequalities and (3.3) imply that for sufficiently small η_1 there exists for all pairs of rates (R, κ) with $R \leq C_{Sh} - 2\kappa - \delta$ an $\eta_2 > \kappa$ satisfying (3.3) such that for n large enough all error probabilities above fall below any $\lambda > 0$.

Remark 5: Especially, for $\kappa = 0$, Proposition 1 gives the standard Coding Theorem for Identification.

There is a very important connection to r–cover–free families.

A family of sets $\mathcal{F}$ is called *r–cover–free* if $A_0 \not\subset A_1 \cup A_2 \cup \cdots \cup A_r$ holds for all distinct $A_0, A_1, \ldots, A_r \in \mathcal{F}$. Let $M(n, r)$ denote the maximum cardinality of such an $\mathcal{F}$ over an n–element underlying set. This notion was introduced in terms of superimposed codes in [50], where for suitable constants c_1, c_2 the inequalities

$$\frac{c_1}{r^2} \leq \frac{\log M(n, r)}{n} \leq \frac{c_2}{r}$$

were proved. This result was rediscovered several times. In [51], with a rather complicated proof, the upper bound was improved to

$$\frac{\log M(n, r)}{n} \leq 2 \frac{\log r + O(1)}{r^2}.$$

After the purely combinatorial proof of [52] by a simpler argument (implicitly contained in [51]) the slightly weaker bound

$$\frac{\log M(n,r)}{n} \leq 4\frac{\log r + O(1)}{r^2}$$

was obtained in [53]. Let $a = |\mathcal{X}|$. With the replacements $r \longrightarrow a^{\kappa n}$, $n \longrightarrow a^n$ we obtain

$$\frac{\log M(a^n, a^{\kappa n})}{a^n} \leq c \cdot \frac{\log a^{\kappa n}}{a^{2\kappa n}}$$

and thus

$$R_n \triangleq \frac{\log\log M(a^n, a^{\kappa n})}{n} \leq (1-2\kappa)\log a + o(1). \tag{3.5}$$

In particular, for $a = 2$, $R \leq 1 - 2\kappa$.

This raises the question of optimality of the bound in Proposition 1. For its answer one needs a suitable bound for r–cover–free uniform families $\mathcal{F}$ of subsets, each of cardinality ℓ exponential in n. However, the existing bounds are too rough!

Technically very simple is the case of K–identification for noiseless channels, if we require the error of first kind to be 0, because thus $\mathcal{D}_S$ equals the union of the support sets $\mathcal{D}_i$ for the random strategies $P_i (i \in S)$ and to just obtain error probability of second kind to be *less than* 1, necessarily $\mathcal{D}_j \not\subset \mathcal{D}_S$ for $j \notin S$. Now the bound on $a^{\kappa n}$–cover–free families is applicable.

Proposition 2. In the noiseless case and for zero error probability of first kind the bound in Proposition 1 is tight.

Notice that in our definition of achievability of a pair (R, κ) we required the existence of (n, N, K, λ)–codes for all small $\lambda > 0$ and n large. It is very convenient to introduce the concept of $\lambda(n)$–achievable pairs (R, κ) by the property that for all large n $(n, N, K, \lambda(n))$–codes exist. Moreover (R, κ) shall be called **polynomially achievable**, if for $\lambda(n) = n^{-\alpha}$, with arbitrary $\alpha > 0$ and n large, $(n, N, K, \lambda(n))$–codes exist. Similarly (R, κ) is **exponentially achievable**, if for an $\varepsilon > 0$ it is $\lambda(n)$–achievable for $\lambda(n) = e^{-\varepsilon n}$.

Correspondingly we speak about $\mathcal{K}_{\lambda(n)}$, the region $\mathcal{K}_{\text{pol}}$ of polynomially achievable rate pairs and the region $\mathcal{K}_{\text{exp}}$ of exponentially achievable rate pairs.

This terminology is consistent with the terminology for converses, which we introduced in Section 2. Further qualifications for several kinds of probabilities are given when needed. Actually for many coding problems several regions coincide. However, as long as we don't know this it is convenient to have this flexible language.

3.2 An equivalence of two coding problems

Let us start with an (n,N,K,λ)–code $\Big\{P_i, \mathcal{D}_S : 1 \le i \le N, S \in \binom{\mathcal{N}}{K}\Big\}$.

We say that S is λ^**–decodable* for this code, if there is a partition $\mathcal{E}_S = \{E_s : s \in S\}$ of $\mathcal{D}_S$ such that

$$\sum_{x^n} W^n(E_s|x^n)P_s(x^n) \ge 1-\lambda^* \ \text{ for all } \ s \in S. \tag{3.6}$$

If for an (n,N,K,λ)–code every $S \in \binom{\mathcal{N}}{K}$ is λ^*–decodable, then we speak of an $(n,N,K,\lambda,\lambda^*)$–code. $\mathcal{K}^*$ denotes the set of pairs of rates for such codes, which are achievable for every $\lambda > 0, \lambda^* > 0$.

Equivalence Theorem 1 *For every DMC*

$$\mathcal{K}_{\text{pol}} \subset \mathcal{K}^* \subset \mathcal{K}.$$

Proof: Obviously, $\mathcal{K}^* \subset \mathcal{K}$. The rate pairs in $\mathcal{K}_{pol}$ are achievable for every $\lambda(n) = n^{-\alpha}$. We show now that an (n,N,K,λ)–code with $N = 2^{2^{Rn}}, K = 2^{\lceil \kappa n \rceil}, \lambda(n)$ can be transformed in an $\big(n,N,K,\lambda(n),\lambda^*(n)\big)$–code with

$$\lambda^*(n) \le \lceil \kappa n \rceil \lambda(n). \tag{3.7}$$

Fix any $S \in \binom{\mathcal{N}}{K}$ and label its elements by the mapping

$$\varphi : S \to \{0,1\}^{\lceil \kappa n \rceil}. \tag{3.8}$$

Then define for $j = 1, 2, \dots, \lceil \kappa n \rceil$

$$\underline{S}_j = \Big\{ s \in S : \varphi(s)_j = 1 \Big\} \tag{3.9}$$

and

$$S_j = \underline{S}_j \cup \overline{S} \ \text{ for } \ \overline{S} \subset \mathcal{N} \smallsetminus S, |\overline{S}| = \frac{1}{2}K. \tag{3.10}$$

The S_j's are elements of $\binom{\mathcal{N}}{K}$ and the S_j's (and also the $\underline{S}_j$'s) form a separating system on S : for every $s, s' \in S$, $s \neq s'$, we have for some j

$$s \in S_j \ \text{ and } \ s' \notin S_j. \tag{3.11}$$

Introduce now the function $\varepsilon_j : S \to \{0,1\}$ by

$$\varepsilon_j(s) = \begin{cases} 1 & \text{if } \ s \in S_j \\ 0 & \text{if } \ s \in S_j^c \end{cases}$$

and use the convention $A^1 = A$ and $A^0 = A^c$.

Then the sets

$$E_s \triangleq \bigcap_{j=1}^{\lceil \kappa n \rceil} (\mathcal{D}_{S_j})^{\varepsilon_j(s)}, s \in S, \tag{3.12}$$

are disjoint, because for $s \neq s'$ there is an S_j with $s \in S_j$ and $s' \notin S_j$ and so $\varepsilon_j(s) \neq \varepsilon_j(s')$.

Finally, we have by the properties of the original code

$$\sum_{x^n} W^n(E_s|x^n)P_s(x^n) \geq 1 - \lceil n\kappa \rceil \lambda(n), s \in S. \tag{3.13}$$

The choice $\lambda(n) = \frac{1}{n^2}$ is good enough. Every S is λ^*–decodable.

Furthermore, it becomes an exercise to show that the same argument also yields for a DMC a relation weaker than Proposition 1, namely

$$\mathcal{K} \supset \Big\{(R,\kappa) : R + 2\kappa \leq C_{er}\Big\},$$

where C_{er} is the erasure capacity (c.f. [59]).

Indeed, for an erasure code $\Big\{(u_i, \mathcal{D}_i) : 1 \leq i \leq M\Big\}$ with erasure probability ε we have

$$W^n(\mathcal{D}_j|u_i) = 0 \ \text{ for } \ i \neq j$$
$$W^n(\mathcal{D}_i|u_i) \geq 1 - \varepsilon; i = 1, \dots, M.$$

In the previous argument we can replace $\{0,1\}^n$ by $\mathcal{U} = \{u_1, \dots, u_M\}$. Subcodes of cardinalities $2^{\rho n}$ and intersecting in at most $2^{-\kappa n} 2^{\rho n}$ words give rise to identification codes (by averaging) of error probability of second kind $\lambda_2 \leq 2^{-\kappa n}$.

The erasure probability is only relevant for the error probability of first kind.

From here on we apply Gilbert's bound with 2^n replaced by $2^{\rho n}$; $\rho \geq \kappa$, $\rho \leq C_{er}$.

Remark 6: $\lambda - K$–identification, λ^*–decodable codes give rise to associated identification codes with error probabilities smaller than $\lambda + \lambda^*$ by assigning to every $i \in \mathcal{N}$ a K-element subset S_i containing i and the decoding set $\mathcal{D}_i = E_i \in \mathcal{E}_{S_i}$. Therefore $R < C_{Sh}$, and by Shannon's Coding Theorem also $\kappa \leq C_{Sh}$.

Remark 7: There is another instructive relation. Let us view $\binom{\mathcal{N}}{K}$ as set $\mathcal{M}$ of objects, one of which, say S, is given to the sender for encoding. The receiver wants to know whether it equals S' (any element of $\mathcal{M}$) or not. This is a standard identification problem with $|\mathcal{M}| = \binom{N}{K}$.

Since $\frac{1}{n}\log\log|\mathcal{M}|$ cannot exceed C_{Sh}, we see that for $K = 2^{\kappa n}$ and $N = 2^{2^{Rn}}$ $\binom{N}{K} \sim 2^{2^{(\kappa+R)n}} \lesssim 2^{2^{C_{Sh}\cdot n}}$, or $\kappa + R \leq C_{Sh}$. Thus κ cannot exceed C_{Sh}. Actually, this is true even if N grows exponentially only, say like $N = 2^{\varepsilon n}$, $\varepsilon > \kappa$, because then

$$2^{2^{C_{Sh}n}} \gtrsim \binom{N}{K} = \binom{2^{\varepsilon n}}{2^{\kappa n}} \geq 2^{(\varepsilon n - \kappa n)2^{\kappa n}} \geq 2^{2^{\kappa n}} \text{ gives } \kappa \leq C_{Sh}.$$

3.3 An outer bound on the capacity region $\mathcal{K}$

The simple idea here is to work with a "net" $\mathcal{S} \subset \binom{\mathcal{N}}{K}$ "almost" of cardinality N^K.

View a set S as 0–1–sequence of length $N = 2^{2^{Rn}}$ with exactly $K = 2^{\kappa n}$ 1's. By Gilbert's bound we can find $\mathcal{S} = \{S_1, S_2, \dots, S_{\tilde{N}}\}$ with the properties

$$|S_i \triangle S_j| \geq (1-\alpha)2K, 0 < \alpha < 1,$$

$$\tilde{N} \geq \binom{N}{K}\left[2^K (N-K)^{(1-\alpha)K}\right]^{-1}.$$

Therefore

$$\tilde{N} \gtrsim N^{\alpha K} = 2^{2^{Rn}\cdot\alpha 2^{\kappa n}} = 2^{\alpha 2^{(R+\kappa)n}}$$

and

$$\frac{1}{n}\log\log\tilde{N} \geq R + \kappa - \frac{1}{n}|\log\alpha|.$$

We summarize this.

Lemma 1. *For every $\alpha \in (0,1)$ there is a family $\mathcal{S} = \{S_1, \dots, S_{\tilde{N}}\} \subset \binom{\mathcal{N}}{K}$ with*

(i) $|S_i \triangle S_j| \geq (1-\alpha)2K$ *and* $|S_i \cap S_j| \leq \alpha K$.
(ii) $R + \kappa - \frac{1}{n}|\log\alpha| \leq \frac{1}{n}\log\log|\mathcal{S}| \leq \frac{1}{n}\log\log\binom{N}{K} \leq R + \kappa$.

We can therefore by (ii) upperbound $\binom{\mathcal{N}}{K}$ by upperbounding $|\mathcal{S}|$. For this we relate $\mathcal{S}$ to a standard identification problem. For $S \in \mathcal{S}$ define $P_S \in \mathcal{P}(\mathcal{X}^n)$ by

$$P_S(x^n) = \frac{1}{K} \sum_{i \in S} P(x^n|i), x^n \in \mathcal{X}^n, \tag{3.14}$$

if $P(\cdot|i)$ is the randomized encoding for i. Now by Lemma 1 (i) and the code definition in (2.1) and (2.2) we have for $S, S' \in \mathcal{S}, S \neq S'$,

$$\sum_{x^n} P_S(x^n) W^n(\mathcal{D}_S|x^n) \geq 1 - \lambda$$

and

$$\sum_{x^n} P_S(x^n) W^n(\mathcal{D}_{S'}|x^n) \leq \lambda + \alpha.$$

This is an $(n, |\mathcal{S}|, \lambda')$ identification code with

$$\lambda' = \lambda + \alpha \geq \lambda.$$

By the weak converse in Section 12 and Lemma 1 (ii) we get the desired bound for $\mathcal{K}$. The same proof works for the K–separating codes of Section 6, if we define $\mathcal{D}_E = \bigcup\limits_{i \in E} \mathcal{D}_{E,i}$.

So for this capacity region $\mathcal{K}^{++}$ we have the same bound.

Proposition 3. $\mathcal{K} \subset \Big\{(R, \kappa) : R + \kappa \leq C_{Sh}\Big\}$

Remark 8: There is a very simple proof for the noiseless BSC. Since the decoding sets $\mathcal{D}_S$ are distinct, it follows that

$$\left|\binom{\mathcal{N}}{K}\right| \leq 2^{2^n} \quad \text{and thus} \quad \frac{1}{n} \log\log \mathcal{N}^K = \frac{1}{n} \log\log \mathcal{N} + \frac{1}{n} \log K = R + \kappa \leq 1.$$

Remark 9: The two Propositions 1, 2 imply for $\kappa = 0$ the standard Identification Capacity Theorem.

Remark 10: Using also the Equivalence Theorem we see that for $R = 0$ we get the converse to Shannon's Coding Theorem and only the achievable rate $\frac{1}{2} C_{Sh}$!

3.4 On K–identification in case of noiseless feedback

As in [17] we assume the presence of a letter by letter noiseless feedback link. Again deterministic encoding functions for i are denoted by f_i^n and randomized encoding functions for i are denoted by F_i^n. The corresponding regions of achievable rate pairs are denoted by $\mathcal{K}_f$ and $\mathcal{K}_F$. Analogously, if all $S \in \binom{\mathcal{N}}{K}$ are λ–decodable we denote the regions by $\mathcal{K}_f^*$ and $\mathcal{K}_F^*$. We formulate now results, which are analog to those under 3.2 and 3.3. Notice that the argument leading to (3.13) applies also in cases of deterministic and randomized feedback strategies. The results in [27], including constructive coding strategies, go considerably beyond [17] and also, if necessary, [27] can be consulted for detailed definitions of all concepts used in this section, when they are not immediately clear.

Equivalence Theorem 2. *For every DMC*

(i) $\mathcal{K}_{f\ pol} \subset \mathcal{K}_f^* \subset \mathcal{K}_f$

(ii) $\mathcal{K}_{F\ pol} \subset \mathcal{K}_F^* \subset \mathcal{K}_F$.

Proposition 4. For every DMC W

$$\mathcal{K}_F \subset \Big\{(R,\kappa) : R+\kappa \leq \max_{P\in\mathcal{P}(\mathcal{X})} H(Q)\Big\},$$

where $Q = P \cdot W$.

We use our entropy property for all discrete distributions.

Lemma 2 (Included in [27]) *For* $P = (P_1, P_2, \dots) \in \mathcal{P}(\mathbb{N})$ *define*

$$\varepsilon(d,P) = \max\left\{\sum_{j\in J} P_j : J \subset \mathbb{N}, |J| = 2^{\lceil H(P)d\rceil+1}\right\},$$

and set

$$\varepsilon(d) = \min_{P\in\mathcal{P}(\mathbb{N})} \varepsilon(d,P).$$

Then

$$\varepsilon(d) = 1 - \frac{1}{d} \ \textit{ for all } \ d \geq 1.$$

Proof of Proposition 4: In any (n,N,K,λ)–code with feedback

$$\left\{(F_i,\mathcal{D}_S) : 1 \leq i \leq N; S \in \binom{\mathcal{N}}{K}\right\}$$

let Y_i^n be the output process generated by F_i via the channel. Furthermore define the process Y_S^n by the distribution

$$\text{Prob}(Y_S^n = y^n) = \frac{1}{K}\sum_{i\in S}\text{Prob}(Y_i^n = y^n).$$

By assumption

$$\text{Prob}[Y_i^n \in \mathcal{D}_S] \geq 1-\lambda, \quad \text{if} \;\; i \in S, \tag{3.15}$$

$$\text{Prob}[Y_i^n \in \mathcal{D}_{S'}] \leq \lambda, \quad \text{if} \;\; i \notin S'. \tag{3.16}$$

By Lemma 2 there are sets $\mathcal{E}_S \subset \mathcal{Y}^n$ $\left(S \in \binom{\mathcal{N}}{K}\right)$ with

$$\text{Prob}[Y_S^n \in \mathcal{E}_S] \geq 1 - \frac{1}{d}, \tag{3.17}$$

$$|\mathcal{E}_S| \leq 2^{\lceil d \; H(Y_S^n)\rceil + 1}. \tag{3.18}$$

We show later that the net $\mathcal{S} \subset \binom{\mathcal{N}}{K}$ with the properties (i), (ii) in Lemma 1 satisfies

$$\mathcal{D}_S \cap \mathcal{E}_S \neq \mathcal{D}_{S'} \cap \mathcal{E}_{S'} \;\; \text{for} \;\; S, S' \in \mathcal{S}; S \neq S', \tag{3.19}$$

provided that λ is sufficiently small.

We know from [17], that

$$H(Y_S^n) \leq n \max_{P:Q=P\cdot W} H(Q) = \overline{H} \;\; \text{(say)}. \tag{3.20}$$

Therefore by (3.18) and (3.19)

$$|\mathcal{S}| \leq 2^{2^{dn\overline{H}}} \tag{3.21}$$

and since d can be made arbitrarily close to 1 we conclude that

$$|\mathcal{S}| \simeq N^K \lesssim 2^{2^{g(\lambda)n\overline{H}}} \tag{3.22}$$

with $\lim\limits_{\lambda \to 0} g(\lambda) = 1$ (weak converse).

Therefore

$$\frac{1}{n} \log\log N^K = \frac{1}{n}(\log K + \log\log N) = \kappa + R \leq \overline{H} \; g(\lambda).$$

It remains to be seen that (3.19) holds.

Suppose that for $S, S' \in S$ $\mathcal{E}_S \cap \mathcal{D}_S = \mathcal{E}_{S'} \cap \mathcal{D}_{S'}$. Then by (3.15) and (3.17)

$$\text{Prob}(Y_{S'}^n \in \mathcal{E}_{S'} \cap \mathcal{D}_{S'}) = \text{Prob}(Y_{S'}^n \in \mathcal{E}_S \cap \mathcal{D}_S) \geq 1 - \frac{1}{d} - \lambda. \tag{3.23}$$

On the other hand, by (3.16)

$\text{Prob}(Y_i^n \in \mathcal{E}_S \cap \mathcal{D}_S) \leq \text{Prob}(Y_i^n \in \mathcal{D}_S) \leq \lambda$ for $i \in S' \smallsetminus S$ and by definition of $\mathcal{S}$ $|S' \smallsetminus S| \geq (1-\alpha)K$.

Therefore $\text{Prob}(Y_{S'}^n \in \mathcal{E}_S \cap \mathcal{D}_S) = \frac{1}{K} \sum\limits_{i \in S'} \text{Prob}(Y_i^n \in \mathcal{E}_S \cap \mathcal{D}_S) \leq \lambda + \alpha$.

This contradicts (3.23), if

$$\lambda + \alpha < 1 - \frac{1}{d} - \lambda. \tag{3.24}$$

This is equivalent with $\lambda < \frac{1}{2}\left(1 - \frac{1}{d}\right) - \frac{\alpha}{2}$.

So in order to show that for any $\varepsilon > 0$ $\kappa + R \leq \overline{H} + \varepsilon$, choose first d so that $d > 1$ and $d\overline{H} \leq \overline{H} + \varepsilon$, then choose λ smaller than $\frac{1}{4}\left(1 - \frac{1}{d}\right)$, and finally choose α smaller than $\frac{1}{2}\left(1 - \frac{1}{d}\right)$.

Remark 11: Notice that we have used that $\frac{1}{K} \sum\limits_{i \in S} f_i$ defines a *randomized* feedback strategy F_S. So this approach does not work for the case of deterministic feedback strategies!

Remark 12: We have upperbounded $\binom{N}{K}$ via upperbounding $|\mathcal{S}|$, for which we used our old idea of "distinct carriers". Instead we could also follow the approach under 3.2, in which we relate the modified K–identification problem with a standard identification problem. In case of feedback we get the upper bound for randomized strategies by the strong converse of [17].

Remark 13: For small K, say for constant K while n grows, K–identification reduces of course to K identifications and thus to identification.

K–identification means that any person E is interested in the question whether the edge E in the hypergraph $\left(\mathcal{N}, \binom{\mathcal{N}}{K}\right)$ occurred. Naturally, we can replace $\binom{\mathcal{N}}{K}$ by any set $\mathcal{E}$ of edges, if this describes the interests.

In order to motivate this model $\mathcal{H} = (\mathcal{V}, \mathcal{E})$ let us suppose that $\mathcal{V}$ is the set of roads in a region and $\mathcal{E}$ is the set of drivers. Driver E is primarily interested in the roads of his tour. In case there has been an accident on one road $v \in \mathcal{V}$ and this road is blocked, then all E's want to know whether $v \in E$ or not (and in the affirmative case secondarily also which road it is).

There are more efficient ways of transferring the information of interest than to broadcast the complete information, which specifies the road with the accident.

The converses in case of feedback show that

$$|\mathcal{E}| < 2^{2^{\overline{H}n}}. \tag{3.25}$$

Now, if we choose $\mathcal{E} = 2^{\mathcal{N}}$, the power set, $R_1 = \text{rate } (N) \leq \overline{H}$.

By Section 4 decoding all subsets, gives optimal rate C_{Sh}. So the bound in (3.25) is not achievable.

Problem 1: Does the Equivalence Theorem hold for general hypergraphs?

3.5 A combinatorial consequence

It is remarkable that a result for K–identification (Proposition 1) has an important consequence for r–cover–free families in relation to packings. We use a result of Kuzjurin [55].

A family $\mathcal{A}$ of k–subsets of $[m] = \{1, 2, \dots, m\}$ is called (m, k, ℓ)*–packing* iff each ℓ–subset of $[m]$ is contained in at most one member $A \in \mathcal{A}$. Therefore two members of $\mathcal{A}$ intersect in at most $\ell - 1$ elements. (In other words $\mathcal{A}$ can be viewed as a code with constant weight k, word length m and distance $d_H = 2(k - \ell) + 2$.)

The *density* $d(\mathcal{A})$ of a packing $\mathcal{A}$ is the average number of k–subsets of $\mathcal{A}$ containing an ℓ–subset, that is, $d(\mathcal{A}) = \frac{|\mathcal{A}|\binom{k}{\ell}}{\binom{m}{\ell}}$. Let $k = k(m)$ and let $\ell = \ell(m) \geq 2$.

A sequence of packings $(\mathcal{A}_m)_{m \geq k}$ is called *asymptotically good* if

$$\lim_{m \to \infty} d(\mathcal{A}_m) = 1.$$

Roughly speaking the result of [55] says that $k = \sqrt{m}$ is the threshold function for the existence of asymptotically good packings. Here is the precise result.

Theorem K. *Let α be the minimum constant such that for every $\varepsilon > 0$ and sufficiently large n every interval $[n, n + n^{\alpha+\varepsilon}]$ contains a prime number. It is known that $\alpha \leq \frac{23}{43}$. The following bounds hold:*

(i) *Let $c < 1$ and $k(m) < c\sqrt{m}$, where $\lim\limits_{n \to \infty} k(m) = \infty$. Further, let for some $\varepsilon > o$ $\ell(m) = o\Big(\sqrt{k(m)}\Big)$ and $\ell(m) = o\left(\binom{m}{k(m)}^{1-\alpha-\varepsilon}\right)$.*
Then asymptotically good (m, k, ℓ)–packings exist.

(ii) *Let $c > 1$, $k(m) > c\sqrt{m}$ and let $\ell(m) = o\big(k(m)\big)$. Then nontrivial asymptotically good (m, k, ℓ)–packings do not exist.*

Corollary 1.

(i) *Let* $m(n) = e^{\mu n}$*,* $k(n) = e^{\gamma n}$*, and* $\ell(n) = e^{\beta n}$*. For* $\frac{\mu}{2} > \gamma$*,* $\gamma/2 > \beta$ *and* $(\mu - \gamma)\frac{20}{43} > \beta$ *we have asymptotically good* (m, k, ℓ)*–packings.*

(ii) *Let* $m(n) = e^{\mu n}$*,* $k(m) = e^{\left(\frac{\mu}{2}+\varepsilon\right)n}$*, and let* $\ell(m) = e^{\beta n}$ *with* $\beta < \frac{\mu}{2} + \varepsilon$*, then asymptotically good* (m, k, ℓ)*–packings do not exist.*

We derive from the assumptions on μ, γ, β

$$\mu > 2\gamma, \gamma > 2\beta, \mu > \gamma + \frac{43}{20}\beta. \tag{3.26}$$

We apply this and (ii) to the set of codewords $\mathcal{U} \subset \mathcal{X}^n$ of a channel code with error probability λ, $|\mathcal{U}| \sim e^{In} = m$, and $\frac{1}{n}\log K(n) = \kappa$. Then $I = \mu$, $\kappa = \gamma - \beta$ and we get for the maximal packing cardinality

$$N^*(n, I, \kappa) \lesssim \frac{\binom{e^{In}}{e^{\beta n}}}{\binom{e^{\gamma n}}{e^{\beta n}}} = \frac{\binom{e^{In}}{e^{\beta n}}}{\binom{e^{(\beta+\kappa)n}}{e^{\beta n}}}, \tag{3.27}$$

$$\frac{1}{n}\log\log N^* \lesssim \beta, \tag{3.28}$$

and for $\gamma \sim \frac{I}{2}$ the lower bound $\beta = \gamma - \kappa \sim \frac{I}{2} - \kappa$. Moreover, $\beta_{\max} \leq \min\left(\frac{I}{4}, \frac{20I}{86}\right) = \frac{10}{43}I$, $\kappa_{\min} = \frac{I}{2} - \beta_{\max} = \frac{23}{86}I$, and $R = \frac{10}{43}I$.

However, our bound $R = I - 2\kappa = \frac{20}{43}I$ in Proposition 1 is much better!

It can be seen from its derivation in 3.1 that this bound can be interpreted as a lower bound on the size $N(n, I, \kappa)$ of optimal r–cover–free families, where r has rate κ. It is known and readily verified that always

$$N(n, I, \kappa) \geq N^*(n, I, \kappa).$$

We know now that the quantities can be very different!

4 Models with capacity equal to the ordinary capacity

Some of the cases considered here were first treated by Já Já [13] for non–randomized encoding on the BSC. If randomisation is permitted, the analysis is somewhat more complicated. In this section we describe the various codes and capacities by words.

4.1 The ordering problem

Suppose that one of the events $\{1, 2, \ldots, N\}$ occurred and is known to the sender. By proper coding he shall enable the receiver to answer the question "Is the true number less than or equal to j?" Here j is any element of $\{1, \ldots, N\}$. We can also use the ordering function

$$f_0(i,j) = \begin{cases} 1 & \text{for } \ i \leq j \\ 0 & \text{otherwise.} \end{cases}$$

A (randomized) ordering code $(n, N, \lambda_1, \lambda_2)$ is a family

$$\Big\{(P(\cdot|i), \mathcal{D}_i) : i = 1, 2, \ldots, N\Big\}$$

of pairs with

$$P(\cdot|i) \in \mathcal{P}(\mathcal{X}^n), \mathcal{D}_i \subset \mathcal{Y}^n \ \text{ for } \ i = 1, 2, \ldots, N \tag{4.1}$$

and with errors of the first (resp. second) kind satisfying for *every* j

$$\sum_{x^n \in \mathcal{X}^n} P(x^n|i) W^n(\mathcal{D}_j|x^n) \geq 1 - \lambda_1 \ \text{ for } \ i = 1, \ldots, j \tag{4.2}$$

and

$$\sum_{x^n \in \mathcal{X}^n} P(x^n|i) W^n(\mathcal{D}_j|x^n) \leq \lambda_2 \ \text{ for } \ i > j. \tag{4.3}$$

Of course, we can define this way deterministic ordering codes by letting $P(\cdot|i)$ denote point masses on points $u_i \in \mathcal{X}^n$.

Theorem 3. *Even for randomized encoding the polynomial ordering problem capacity does not exceed the transmission capacity. The same holds in case of noiseless feedback.*

Proof: Suppose first that $N \leq \Big(2|\mathcal{X}|\Big)^n$ and that $\lambda_1, \lambda_2 \leq \frac{1}{n^2}$.

The ordering problem code gives rise to a transmission code as follows:

Choose first $j_1 = \lceil \frac{M}{2} \rceil$. In case of a "yes" iterate the search for the "true message" in $\{1, \ldots, \lfloor \frac{M}{2} \rfloor\}$ and otherwise in $\{\lceil \frac{M}{2} \rceil, \ldots, M\}$ by choosing next j_2 in the middle of these sets, resp. After $\log N$ iterations we are done. The total error probability is bounded by

$$\frac{1}{n^2} \log N \leq \frac{2|\mathcal{X}|}{n}.$$

Next, if $N > \Big(2|\mathcal{X}|\Big)^n$, choose *any* subset of $\{1,2,\dots,N\}$ of a cardinality $\exp\Big\{(C+\delta)n\Big\}$ for some $\delta > 0$.

Apply to the subcode corresponding to this set the previous argument. This leads to a transmission code of a rate exceeding capacity and this contradiction proves that actually $N > \Big(2|\mathcal{X}|\Big)^n$ does not occur.

Finally, the same argument applies to the case of feedback.

Remark 14: We have shown that, generally speaking, whenever $\log N$ bits specify an event with the code concept used, its rate does not exceed C. Thus we have also the next result.

4.2 All binary questions

By proper coding the sender shall enable the receiver to answer all the questions "Is the true number in A?" Here A is any subset of $\{1,\dots,N\}$.

Theorem 4. *Even for randomized encoding the binary questions capacity does not exceed the transmission capacity. The same holds in case of noiseless feedback.*

4.3 Identification of a component

In model 6, the number of components is linear in the blocklength. For exponentially small error probability words can therefore be reproduced with small error probability. (For small, but constant error probabilities, rate–distortion theory is to be used).

Theorem 5. *Even for randomized encoding the component identification capacity does not exceed the transmission capacity. The same holds in case of feedback.*

Part II: Models with prior knowledge of the receiver

The a priori structure is a hypergraph $\mathcal{H} = (\mathcal{V},\mathcal{E})$. The encoder of channel W knows the message vertex $v \in \mathcal{V}$ and the decoder D_E $(E \in \mathcal{E})$ knows beforehand whether the message to be transmitted is in E or not. In case it is, he wants to know which element of E it is.

We consider first abstract hypergraphs.

5 Zero–error decodable hypergraphs

If the decoder wants to know $v \in E$, then any two vertices $x, y \in E$ must be separable for instance by different colors assigned to them.

Definition: The separability graph $\mathcal{G}(\mathcal{H}) = (\mathcal{V}, \mathcal{E}^*)$ is defined by

$$\{x, y\} \in \mathcal{E}^* \Leftrightarrow \exists\, F \in \mathcal{E} : \{x, y\} \subset F. \tag{5.1}$$

Let $\Psi(\mathcal{G})$ be the chromatic number of $\mathcal{G}$, then $\mathcal{H}$ is 0–error decodable iff $\Psi(\mathcal{G}) \leq 2^{C_0 n}$, where C_0 is the zero–error capacity of the channel W used for the transmission of this color. Now $\mathcal{H}$ is λ–identifiable iff $\Psi(\mathcal{G}) \lesssim 2^{2^{C(W)n}}$.

Remark 15: Also if 2–separable only within edges by the results of [16], [28] the answer is the same.

6 K–separating codes

Instead of zero–error decodability for hypergraphs one can consider λ-decodability, that is, an error probability not exceeding λ is permitted.

We call $\Big\{(P_i, \mathcal{D}_{E,i}) : E \in \mathcal{E}, i \in E\Big\}$ an $[n, N, \lambda]$–code for $\mathcal{H} = (\mathcal{V}, \mathcal{E})$ and W, if $P_i \in \mathcal{P}(\mathcal{X}^n)$ for $i \in \mathcal{V} = \{1, 2, \dots, N\}$, $\mathcal{D}_{E,i} \subset \mathcal{Y}^n$, and for all $E \in \mathcal{E}$

$$\mathcal{D}_{E,i} \cap \mathcal{D}_{E,i'} = \varnothing \;\text{ for }\; i, i' \in E, i \neq i' \tag{6.1}$$

$$\sum_{x^n} P_i(x^n) W^n(\mathcal{D}_{E,i} | x^n) \geq 1 - \lambda \;\text{ for }\; i \in E. \tag{6.2}$$

The issue is to minimize n for given $\mathcal{H}$ (and thus N) and λ for the channel W.

For abstract hypergraphs $\mathcal{H}$ not very much can be said. The subject becomes interesting under reasonable assumptions on $\mathcal{H}$.

Example 4: $\mathcal{E} = \{\mathcal{V}\}$ describes Shannon's theory of transmission.

Example 5: $\mathcal{E} = \binom{\mathcal{V}}{K}$, the family of all K–element subsets of $\mathcal{V}$, defines the complete K–uniform hypergraph. The codes defined above are denoted here by $[n, N, K, \lambda]$ and called *K–separating codes.*

Clearly, their capacity region $\mathcal{K}^{++}$ contains $\mathcal{K}^*$ and by the Equivalence Theorem 1 also $\mathcal{K}_{\text{pol}}$.

Moreover, the same proof as for Proposition 3 in Section 3 works for K-separating codes, if we define $\mathcal{D}_E = \bigcup\limits_{i \in E} \mathcal{D}_{E,i}$.

Corollary 2.

(i) $\mathcal{K}^{++} \supset \mathcal{K}^* \supset \mathcal{K}_{\text{pol}}$.
(ii) $\mathcal{K}^{++} \subset \{(R,\kappa) : R + \kappa \leq C_{Sh}\}$.

Problem 2: Determine $\mathcal{K}^{++}$!

6.1 Second order 2–separation capacity without and with feedback

Let us start with the first meaningful case $K = 2$.

For $E = \{i,j\}$ we can write

$$\mathcal{D}_{E,i} = \mathcal{D}_{ij} \text{ and } \mathcal{D}_{E,j} = \mathcal{D}_{ji}.$$

We also say that any two messages are λ–decodable.

Notice that an (n,N,λ)–ID–code $\Big\{(P_i,\mathcal{D}_i) : 1 \leq i \leq N\Big\}$ satisfies

$$\sum_{x^n} P_i(x^n) W^n(\mathcal{D}_i|x^n) \geq 1-\lambda \text{ and } \sum_{x^n} P_j(x^n) W^n(\mathcal{D}_i|x^n) \leq \lambda (i \neq j).$$

Therefore setting $\mathcal{D}_{ij} = \mathcal{D}_i \smallsetminus \mathcal{D}_j$ and $\mathcal{D}_{ji} = \mathcal{D}_j \smallsetminus \mathcal{D}_i$ we see that i and j are 2λ–separable. It immediately follows that the second order capacity for $K = 2$, say C_2, is not smaller than the ID–capacity C_{Sh}. Whereas in ID-codes the decoding sets carry one index, 2–separating codes carry two indices. The decoding sets for two messages are adapted for these two and no other message. Therefore *2–separation* is a weaker notion than identification (except, perhaps, for a small shift in error probability caused by the disjointness of the two decoding sets).

Theorem 6.

(i) *The 2–separation capacity of second order C_2 equals the second order identification capacity C_{Sh}.*
(ii) *The corresponding capacities for channel (deterministic and randomized) feedback strategies are also equal.*

Proof: The issues are the converses.

(i) Here we can be brief, because inspection of the strong converse proof for identification of Han/Verdú [28] shows that it is actually designed for 2–separation. The key fact, called resolvability in [29], is this:

For $P \in \mathcal{P}(\mathcal{X}^n)$ with $Q = PW^n$ and $\varepsilon > 0$ there is a $P^* \in \mathcal{P}(\mathcal{X}^n)$, which is an equidistribution over at most $\sim \exp\{nC_{Sh}\}$, not necessarily distinct, members of $\mathcal{X}^n$ and such that for $Q^* = P^*W^n$

$$\|Q - Q^*\| \leq \varepsilon \ \text{ for } \ n \geq n(\varepsilon). \tag{6.3}$$

(Here $\|\ \|$ denotes total–variation).

In this way to every encoding distribution $P_i (1 \leq i \leq N)$ we can find a distribution P_i^* such that the corresponding output distribution is close to that of P_i. By the code properties the Q_i's and also the Q_i^*'s are distinct. Therefore the P_i^*'s must be distinct and there number in second order rate does not exceed C_{Sh}.

(ii) Let us consider the deterministic case. For the randomized case we just have to replace $\underline{H} = \max\limits_x H(W(\cdot|x))$ by $\overline{H} = \max\limits_P H(PW)$.

We know from Lemma 2 in Section 3.4, that for encoding function f_i there exists an $\mathcal{E}_i \subset \mathcal{Y}^n$ such that for $Q_i = W^n(\cdot|f_i)$, $Q_i(\mathcal{E}_i) \geq 1 - \frac{1}{d}$, and $|\mathcal{E}_i| \leq 2^{\lceil d\underline{H}n \rceil + 1}$. Omit from $\mathcal{E}_i$ the elements with smallest probability until we get a set $\mathcal{E}_i^* \subset \mathcal{E}_i$ with $Q_i(\mathcal{E}_i^*) \geq 1 - \frac{1}{d}$ and which is minimal with this property.

Set $T = \max\limits_i |\mathcal{E}_i^*|$. The number of different such sets is

$$\left| \binom{y^n}{T} \right| \leq 2^{(n \log |\mathcal{Y}|) 2^{\lceil d\underline{H}n \rceil + 1}}. \tag{6.4}$$

This is the desired upper bound. However, not all $\mathcal{E}_i^*$'s are necessarily different. Therefore, we have to upperbound the multiplicity with which a set, say $\mathcal{F}$, occurs among the $\mathcal{E}_i^*$'s. W.l.o.g. we label them $\mathcal{E}_1^*, \dots, \mathcal{E}_M^*$. By our definitions

$$1 - \frac{1}{d} + \frac{Q_i(\mathcal{E}_i^*)}{|\mathcal{F}|} \geq Q_i(\mathcal{E}_i^*) \geq 1 - \frac{1}{d}. \tag{6.5}$$

For $i, j \in \{1, \dots, M\}$ we have for λ small

$$Q_i(\mathcal{F} \cap \mathcal{D}_{ij}) \geq 1 - \lambda - \frac{1}{d} > \lambda, Q_j(\mathcal{F} \cap \mathcal{D}_{ji}) \geq 1 - \lambda - \frac{1}{d} > \lambda,$$

$$\text{and} \ \ Q_i(\mathcal{D}_{ji}), Q_j(\mathcal{D}_{ij}) \leq \lambda.$$

If we now set $\mathcal{D}'_{\ell k} = \mathcal{F} \cap \mathcal{D}_{\ell k}$ and renormalize the measure Q_i on $\mathcal{F}$ from total measure $\sim 1 - \frac{1}{d}$ (see 6.5) to 1, then we have a 2–separating code of size M with output space $\mathcal{F}$.

To this situation we apply the idea of resolvability in the following setting: We want to know how many distributions can be 2–separated on a finite set $\mathcal{T}$ with T elements which we can view as subset of $\{0,1\}^n$, $T \le 2^m$. This is covered by Han/Verdú's result, when W is the noiseless BSC. We get the bound $M \le 2^{2^m}$ or

$$M \le 2^{2^{d\underline{H}n}}. \tag{6.6}$$

Together with (6.4) we get

$$N \le 2^{(n\log|\mathcal{Y}|)2^{d\underline{H}n}} \cdot 2^{2^{d\underline{H}n}} \le 2^{(1+n\log|\mathcal{Y}|)2^{d\underline{H}n}},$$

and thus the weak converse by choosing d close to 1, λ then small enough and $n \ge n(d,\lambda)$.

6.2 Strong converses by the method of [17] for 2–separation in case of feedback

We begin with Theorem 6, (ii) in 6.1. By Lemma 2 of [17] for any $\varepsilon \in (0,1)$ we can find sets $\mathcal{E}_i^*(i=1,\dots,N)$ of minimal size with

$$1 \ge W^n(\mathcal{E}_i^*|f_i) \ge 1-\varepsilon, \tag{6.7}$$

$$|\mathcal{E}_i^*| \le 2^{\left(\underline{H}+\frac{c(\varepsilon)}{\sqrt{n}}\right)n}. \tag{6.8}$$

How many can be equal to $\mathcal{F}$, say?

Now just repeat the previous proof in 6.1. Now (the sharper) (6.7) takes the role of (6.5). Instead of (6.6) we get now the stronger

$$M \le 2^{2^{\left(\underline{H}+\frac{c(\varepsilon)}{\sqrt{n}}\right)n}} \tag{6.9}$$

and finally

$$N \le 2^{(n\log|\mathcal{Y}|)2^{\left(\underline{H}+\frac{c(\varepsilon)}{\sqrt{n}}\right)n}} \cdot 2^{2^{\left(\underline{H}+\frac{c(\varepsilon)}{\sqrt{n}}\right)n}}$$

and thus

$$\frac{1}{n}\log\log N \le \underline{H} + \frac{c(\varepsilon)}{\sqrt{n}} \quad \text{(Strong Converse)}. \tag{6.10}$$

Replacing f_i by F_i and $\underline{H}$ by $\overline{H}$ the same proof applies otherwise literally and gives a strong converse for randomized encoding.

Remark 16: The results obviously generalize to any constant K.

Problem 3: Are the optimal rates for 2–separable codes and ID–codes equal if they satisfy $\lambda_2 \le e^{-\eta_2 n}$?

7 Analysis of a model with specific constraints: 2–separation and Rényi's entropy H_2

Let us assume that a set of persons $\mathcal{N} = \{1, 2, \dots, N\}$ are at a party. The persons move randomly between α rooms and the set of persons in room i at some time is A_i of cardinality

$$|A_i| = P_i\, N;\ i = 1, \dots, \alpha. \tag{7.1}$$

We say that the partition $\Pi = (A_1, \dots, A_\alpha)$ is of type $P = (P_1, P_2, \dots, P_\alpha) \in \mathcal{P}(\mathcal{N})$.

Let now $\Pi_1, \Pi_2, \dots, \Pi_m$ be a sequence of independent random partitions taking as values a partition of type P with equal probabilities. Equivalently we can say that a person $z \in \mathcal{N}$ belongs to the randomly chosen A_i with probability P_i independently of what happens to the other persons. (At discrete time points $1, 2, \dots$ the partition of the persons in several rooms is reported.)

Imagine now that somebody, the interrogator, has difficulties to distinguish any two persons in his interest at the party, but is reported the sequence of partitions described. So he knows at every time instance the set of persons in all rooms, but he cannot identify the persons in a set.

Let now $\lambda_{N,m}$ denote the probability that m such partitions separate any two persons in $\mathcal{N}$. Rényi [34] has shown that $m_2(N, \varepsilon)$, the smallest m with $\lambda_{N,m} \geq 1 - \varepsilon$, satisfies

$$m_2(N, \varepsilon) \sim \frac{2 \log_2 N + o(\varepsilon)}{H_2(P)}, \tag{7.2}$$

where H_2 is Rényi's entropy of order 2.

Now let us go a step further. The interrogator is at the receiver side of a noisy channel. For partition $\Pi_i = (A_{i1}, \dots, A_{i\alpha})$ let

$$F_i(z) = j, \quad \text{if} \ \ z \in A_{ij}. \tag{7.3}$$

For every $z \in \mathcal{N}$ $\Big(F_1(z), \dots, F_m(z)\Big)$ is known to the encoder. How fast can the interrogator decide his question with high probability correctly?

Answer: Match $(F_1, \dots, F_m)$ with a 2–*separation code.*

It would be stupid to use a transmission code. There are several variations of this model.

In many situations of information transfer reduction to transmission would be of poor performance.

8 Binning via channels

In Section 5 we considered vertex colorings with different colors in each edge. They have been called strict colorings in [6] and [8]. Other colorings discussed there are

(α) colorings, where in every edge no color occurs more than ℓ times (leading to list–knowledge)

(β) colorings, where in every edge a high percentage of colors occurs only one time

(γ) colorings, which are good, in the senses of (α) and/or (β) in average under given probability distributions on vertices and/or edges.

The present investigations have born still another coloring (or binning) concept.

Indeed, let us look at K–separation. We know from Proposition 1 that we can choose N with second order rate R and K with rate κ, $R + 2\kappa \leq C_{Sh}$, and achieve K–identification.

Further, by the Equivalence Theorem the hypergraph $\left(\mathcal{N}, \binom{\mathcal{N}}{K}\right)$ is in addition K–separable. *What does this mean*? Well, the "color" on vertex i is the randomized encoding P_i and within every edge $S \in \binom{\mathcal{N}}{K}$ containing i this i is decoded correctly with probability at least $1 - \lambda$!

Notice that for the price of a small error probability λ now — in contrast to the situation in (β) (or also (γ)) — *every* vertex can be decoded correctly.

Furthermore, the theory in [6], [8] works, if the number of vertices, the number of edges, and the edge sizes are roughly of the same growth, **namely exponential in n**.

Here the edge sizes are at most exponential in n, but the number of vertices and edges can grow double exponentially in n!

9 K–identifiability, K–separability and related notions

We discuss here connections between code concepts.

To fix ideas let us first compare 1–identification (the classical identification) and 2–separation. In both cases we have a fixed *encoding structure* (set of codewords, set of probability distributions or set of randomized or non–randomized–feedback functions). In any case they specify via the channel a set of *output distributions*

$$\mathcal{Q} = \{Q_i : i \in \mathcal{N}\}. \tag{9.1}$$

The various code concepts associate with such a set a *decoding structure.*

In case of identification the decoding structure is

$$\mathcal{D} = \{\mathcal{D}_i : i \in \mathcal{N}\}. \tag{9.2}$$

It is of *precision* λ, if

$$Q_i(\mathcal{D}_i) \geq 1 - \lambda(i \in \mathcal{N}) \text{ and } Q_i(\mathcal{D}_j) \leq \lambda(i \neq j). \tag{9.3}$$

The precision relates to the whole encoding structure $\mathcal{Q}$, however, in a pairwise fashion (as specified in (9.3)).

The concept 2–separation allows more freedom in the decoding structure. We say $\mathcal{Q}$ is *2–separable with precision* λ, if for any $S = \{i, j\} \in \binom{\mathcal{N}}{2}$ there are two sets $\mathcal{D}_{Si}$ and $\mathcal{D}_{Sj}$ with

$$\mathcal{D}_{Si} \cap \mathcal{D}_{Sj} = \varnothing, Q_i(\mathcal{D}_{Si}), Q_j(\mathcal{D}_{Sj}) \geq 1 - \lambda. \tag{9.4}$$

These sets relate only to i and j.

Lemma 3. *1–identifiable with precision λ implies 2–separable with precision 2λ.*

Proof: Define $\mathcal{D}_{Si} = \mathcal{D}_i \smallsetminus \mathcal{D}_j$ and $\mathcal{D}_{Sj} = \mathcal{D}_j \smallsetminus \mathcal{D}_i$, then $\mathcal{Q}_\ell(\mathcal{D}_{S\ell}) \geq 1 - 2\lambda$ for $\ell = i, j$.

There is also a general connection.

Lemma 4. *K–identifiable with precision $\lambda(n)$ implies K–separable with precision $\lambda'(n) = \lceil n\kappa \rceil \lambda(n)$, where*

$$\kappa = \textit{rate } (K) = \frac{1}{n} \log K.$$

Proof: See proof of Equivalence Theorem 1 in Section 3.

Problem 4: For $L \geq K$, how does K–identifiability relate to L–separability?

Finally we mention related concepts.

a. We say that a K–identification decoding is based on a 1–identification decoding $\{\mathcal{D}_i : i \in \mathcal{N}\}$ of precision λ, if

$$\mathcal{D}_S = \bigcup_{i \in S} \mathcal{D}_i, S \in \binom{\mathcal{N}}{K} \tag{9.5}$$

and

$$Q_i(\mathcal{D}_i) \geq 1 - \lambda \ \text{ for } \ i \in \mathcal{N}, Q_i(\mathcal{D}_S) \leq \lambda \ \text{ for } \ i \notin S. \tag{9.6}$$

For the disjoint sets

$$\mathcal{D}_{Si} = \mathcal{D}_i \smallsetminus \bigcup_{j \in S \smallsetminus \{i\}} \mathcal{D}_j \ \text{ for } \ i \in S$$

we have

$$Q_i(\mathcal{D}_{Si}) \geq 1 - 2\lambda \ \text{ for } \ i \in S, \tag{9.7}$$

a generalisation of Lemma 3.

b. As a weaker notion than K–separability we define for positive integers α, β with $\alpha + \beta = K$ that $\mathcal{Q}$ *is* (α, β)*–separable with precision* λ, if for every $S \in \binom{\mathcal{N}}{K}$ and every partition $\{S_0, S_1\}$ of S, where $|S_0| = \alpha$ and $|S_1| = \beta$, there are disjoint sets $\mathcal{D}_{S_0}$ and $\mathcal{D}_{S_1}$ with

$$Q_j(\mathcal{D}_{S_0}) \geq 1 - \lambda \ \text{ for } \ j \in S_0$$

and

$$Q_j(\mathcal{D}_{S_1}) \geq 1 - \lambda \ \text{ for } \ j \in S_1.$$

c. Analogously we say that $\mathcal{Q}$ is (α, β)–identifiable with precision λ, if there is a decoding structure $\left\{\mathcal{D}_{S'} : S' \in \binom{\mathcal{N}}{\alpha} \cup \binom{\mathcal{N}}{\beta}\right\}$ such that for $S = S_0 \dot{\cup} S_1$, $|S_0| = \alpha$, $|S_1| = \beta$

$$Q_i(\mathcal{D}_{S_\varepsilon}) \geq 1 - \lambda \ \text{ for } \ i \in S_\varepsilon$$

and

$$Q_i(\mathcal{D}_{S_\varepsilon}) \leq \lambda \ \text{ for } \ i \in S_{1-\varepsilon}$$

for $\varepsilon = 0, 1$.

K–identification concerns partitions $\{S, \mathcal{N} \smallsetminus S\}$, $S \in \binom{\mathcal{N}}{K}$. One can consider partitions π_ℓ, $\ell \in \mathcal{L}$, into more than 2 sets. Person ℓ wants to know the set in its partition, which contains the "message". There may be several channels. ("From which country is a sportsman?", "what is his age?" etc.)

This model includes compound channels, where the receiver knows the individual channel, broadcast channels (also with degraded message sets) etc.

Part III: Models with prior knowledge at the sender

10 Identification via group testing and a stronger form of the Rate–Distortion Theorem

Suppose that from the set $\mathcal{N} = \{1, 2, \ldots, N\}$ of persons any subset $\mathcal{S} \subset \mathcal{N}$ of persons may be the set of sick persons. Moreover it is known that with probability q a person is sick and that the RV S has the distribution

$$\text{Prob}(S = \mathcal{S}) = q^{|\mathcal{S}|}(1-q)^{N-|\mathcal{S}|}. \tag{10.1}$$

For each subset of the test subjects, ($B \subseteq \mathcal{N}$), the binary, error-free test, which determines whether at least one person in B is sick or not, is admissible. In the group testing model introduced in [65] the goal is to determine the expected number of tests $L(N, q)$ for an optimal sequential strategy to diagnose all sick persons (see also [S10], pp. 112-117).

Theorem ([65]) $Nh(q) \leq L(N, q) \leq N$

In our model the decoder (person) s wants to know whether he is sick. Any other information is of much less relevance to him. In particular he does not care who the other sick persons are. In terms of partitions

$$\pi_s = \Big\{\{\mathcal{S} \subset \mathcal{N} : s \in \mathcal{S}\}, \{\mathcal{S} \subset \mathcal{N} : s \notin \mathcal{S}\}\Big\} \tag{10.2}$$

he wants to know which member of π_s occurred.

We can reformulate this problem by identifying $\mathcal{S} \subset \mathcal{N}$ with a word $x_{\mathcal{S}} = (x_1, \ldots, x_N) \in \{0,1\}^N$, $x_s = 1$ iff $s \in \mathcal{S}$. Thus the distribution defined in (10.1) describes a discrete memoryless source (DMS) $\Big(\{0,1\}^N, Q^N, X^N\Big)$ with $Q^N(x^N) = \prod_{t=1}^N Q(x_t)$, where

$$Q(x_t) = \begin{cases} q & \text{for } \; x_t = 1 \\ 1-q & \text{for } \; x_t = 0, \end{cases} \tag{10.3}$$

and for $X^N = (X_1, \ldots, X_N)$

$$\text{Prob}(X^N = x^N) = Q^N(x^N). \tag{10.4}$$

For any encoding function $f_N : \{0,1\}^N \to \mathbb{N}$ and decoding function $g_t(1 \leq$

$t \leq N) : \mathbb{N} \to \{0,1\}$ we can set

$$\hat{X}_t = g_t\Big(f_N(X^N)\Big) \tag{10.5}$$

and consider the error probability

$$\lambda_t = \mathbb{E}\ d(X_t, \hat{X}_t),$$

where d is the Hamming distance.

Now the Rate–distortion Theorem tells us how small a rate $R(q,\lambda)$ we can achieve with rate$(f_N) = \frac{\log(\text{Number of values of } f_N)}{N}$ under the constraint

$$\sum_{t=1}^{N} \mathbb{E}\ d(X_t, \hat{X}_t) \leq \lambda\ N. \tag{10.6}$$

However, we are interested in the stronger condition

$$\mathbb{E}\ d(X_t, \hat{X}_t) \leq \lambda \ \text{ for } \ 1 \leq t \leq N \tag{10.7}$$

and the corresponding minimal rate $R^*(q,\lambda)$. We know that

$$\lim_{\lambda \to 0} R(q,\lambda) = h(q)$$

and therefore as $\lambda \to 0$ by the Source Coding Theorem also $\lim\limits_{\lambda \to 0} R^*(q,\lambda) = h(q)$.

When λ is kept at a prescribed level we have the following result.

Theorem 7. *The identification after group testing in a group of N persons, everyone being independently sick with probability q, can be performed at error probability λ with $R(q,\lambda)N$ bits. Here $R(q,\lambda)$ is the rate–distortion function for the Bernoulli source with generic distribution $(q, 1-q)$ evaluated at distortion level λ.*

Remark 17: Since space does not permit we leave the proof as an exercise using balanced hypergraph covering, which we started in [8]. The Lemma in Section VI of [58] can be used for q-typical N sequences as vertex set $\mathcal{V}$ and p-typical N sequences as edge set $\mathcal{E}$ for covering or approximation. The exceptional set $\mathcal{V}_0$ in that lemma can be kept empty (see Lemma 9 of [64]). Now in addition to hypergraph $(\mathcal{V},\mathcal{E})$ use also hypergraph $(\mathcal{V}_1,\mathcal{E})$, where $\mathcal{V}_1 = [N]$. There is a selection of edges $E_1, \dots, E_L \in \mathcal{E}$ which simultaneously covers $\mathcal{V}$ and $\mathcal{V}_1$ in balanced ways. The second means (10.7), of course after polynomially many pairs (q', p') with q' close to q have been used.

Instead of two properties (sick and not sick) there can be any finite number of properties k defining k classes and every person wants to know its class.

This leads to a **Rate–distortion theorem for a DMS stronger than Shannon's.**

In case the encoding of S is transmitted via a *noisy* channel an argument for the separation of source and channel coding is needed. To elaborate conditions under which the "separation principle" is valid is a major subject in Information Theory.

Part IV: Identification and transmission with multi–way channels

11 Simultaneous transfer: transmission and identification

The issue of simultaneity comes up frequently in life and in science. In information theory we encounter situations where the same code is used for several channels, where several users are served by the same channel, where one code serves several users etc.

A. Let us discuss now a specific example. Suppose that one DMC is used *simultaneously* for transmission and identification. Since both, the transmission capacity and the (second order) identification capacity, equal C_{Sh}, here is the best we can do: We use an (n, M) transmission code $\Big\{(u_i, \mathcal{D}_i) : 1 \leq i \leq M\Big\}$ with *average* error $\overline{\lambda} = \frac{1}{M} \sum\limits_{i=1}^{M} W^n(\mathcal{D}_i^c|u_i)$. The randomness in the messages produces via this code a common random experiment for sender and receiver. Adding a few, say, $\sqrt{n}$ letters, we can get the desired identification code $(n + \sqrt{n}, N, \lambda)$ as in [17] (see also [58]) by the following approach.

From common randomness (also called shared randomness in physics) to identification: The $\sqrt{n}$-trick

Let $[M] = \{1, 2, \dots, M\}$, $[M'] = \{1, 2, \dots, M'\}$ and let $\mathcal{T} = \{T_i : i = 1, \dots, N\}$ be a family of maps $T_i : [M] \to [M']$ and consider for $i = 1, 2, \dots, N$ the sets

$$K_i = \{(m, T_i(m)) : m \in [M]\}$$

and on $[M] \times [M']$ the PD's

$$Q_i((m, m')) = \frac{1}{M} \text{ for all } (m, m') \in K_i.$$

Transformator Lemma *Given* $M, M' = exp\{\sqrt{\log M}\}$ *and* $\epsilon > 0$ *there* *exists a family $\mathcal{T} = \mathcal{T}(\epsilon, M)$ such that $|\mathcal{T}| = N \geq exp\{M - c(\epsilon)\sqrt{n}\}$, $Q_i(K_i) = 1$ for $i = 1, \ldots, N$, and $Q_i(K_j) \leq \epsilon$ $\forall i \neq j$.*

Hence, (C_{Sh}, C_{Sh}) is achievable.

Next suppose that there is a noiseless feedback channel and we use the same code as before. This generates an input process $X^n = (X_1, \ldots, X_n)$ and an output process $Y^n = (Y_1, \ldots, Y_n)$, which is known also to the sender by the feedback. So we get a common random experiment of rate $\frac{1}{n}H(Y^n)$. Again by the identification trick of [17] now

$$\begin{aligned} R_{\text{transm.}} &\sim \frac{1}{n} I(X^n \wedge Y^n) \\ R_{\text{ident.}} &\sim \frac{1}{n} H(Y^n), \text{ second order.} \end{aligned}$$

It is now easy to show the direct part in

Theorem 8. *$\mathcal{R} = conv\Big\{\Big(I(X \wedge Y), H(Y)\Big) : P_X \in \mathcal{P}(\mathcal{X})\Big\}$ is the set of achievable pairs of rates for the simultaneous transmission and identification over the DMC with noiseless feedback.*

Proof of converse: Let the RV U take values in the set of codewords $\mathcal{U} = \{u_1, \ldots, u_M\}$ for transmission with equal probabilities. Further let $F_i(u)$ be the randomized encoding for i and $u \in \mathcal{U}$, making use of the feedback. Then for the transmission and disjoint decoding sets $\mathcal{D}_j$

$$\frac{1}{M} \sum_{j=1}^{M} W^n\Big(\mathcal{D}_j^c | F_i(u_j)\Big) \leq \overline{\lambda} \text{ for all } i \tag{11.1}$$

and for identification with decoding sets $\mathcal{D}_i^*$

$$\frac{1}{M} \sum_{j=1}^{M} W^n\Big(\mathcal{D}_i^* | F_i(u_j)\Big) \geq 1 - \lambda \text{ for } i = 1, \ldots, N \tag{11.2}$$

and

$$\frac{1}{M} \sum_{j=1}^{M} W^n\Big(\mathcal{D}_k^* | F_i(u_j)\Big) \leq \lambda \text{ for } i \neq k. \tag{11.3}$$

For every $i = 1, 2, \ldots, N$ we get input variables $X_i^n = (X_{i1}, \ldots, X_{in})$ and output variables $Y_i^n = (Y_{i1}, \ldots, Y_{in})$.

By Shannon's weak converse proof for the DMC with feedback

$$\log M \leq \frac{I(X_i^n, Y_i^n)}{1 - \overline{\lambda}} \text{ for all } i \tag{11.4}$$

and by the weak converse proof for identification on the DMC with feedback ([27])

$$\log \log N \leq \max_i H(Y_i^n). \tag{11.5}$$

Therefore for some i

$$\Big(\frac{1}{n}\log M, \frac{1}{n}\log\log N\Big) \leq \Big(\frac{1}{n}I(X_{i_0}^n, Y_{i_0}^n), \frac{1}{n}H(Y_{i_0}^n)\Big)\cdot\frac{1}{1-\overline{\overline{\lambda}}}$$
$$\leq \Big(\frac{1}{n}\sum_{t=1}^n I(X_{i_0 t}, Y_{i_0 t}), \frac{1}{n}\sum_{t=1}^n H(Y_{i_0 t})\Big)\frac{1}{1-\overline{\overline{\lambda}}}$$
$$\leq \Big(I(\overline{X},\overline{Y}), H(\overline{Y})\Big)\frac{1}{1-\overline{\overline{\lambda}}},$$

if we use the concavity of I and of H. This completes the weak converse proof.

We draw attention to the fact that it is a lucky coincidence that these two proofs are available and can be combined. The known strong converses for the separate problems cannot be combined!

Finally we propose as

Problem 5: This proof assumes a deterministic transmission code. Can randomized transmission codes give better overall performance?

B. More generally there is a theory of multiple purpose information transfer. Different goal seeking activities are optimized in combinations. The familiar compound and broadcast (also with degraded message sets) channels are included.

Not just transmission and identification, but any collection of the models in Section 2 can occur in various combinations. For example consider a MAC with three senders. For a given sportsman sender 1 says from which country he comes, sender 2 informs about the age groups, and sender 3 is concerned about the fields of activities.

C. Memory decreases the identification capacity of a discrete channel with alphabets $\mathcal{X}$ and $\mathcal{Y}$ in case of noiseless feedback.

(α) For non–random strategies this immediately follows from the inequality

$$\max_{x^n} H\Big(W^n(\cdot|x^n)\Big) \leq \sum_{t=1}^n \max_{x_t} H\Big(W(\cdot|x_t)\Big)$$

(β) For a randomized strategy F

$$H\Big(W^n(\cdot|F)\Big) = H(Y_1,\dots,Y_n) = H(Y_n|Y_1,\dots,Y_{n-1}) + H(Y_1,\dots,Y_{n-1})$$

and

$$\begin{aligned} H(Y_n|Y_1,\dots,Y_{n-1}) &= \sum_{y^{n-1}} \text{Prob}(Y^{n-1}=y^{n-1}) \cdot H(Y_n|y_1,\dots,y_{n-1}) \\ &= H\left(\sum_x W_n(\cdot|x) \sum_{y^{n-1}} \text{Prob}\Big(F_n(y_1\dots y_{n-1})=x\Big)\right) \\ &\le \max_{P_X} H(P_X W_n). \end{aligned}$$

12 A proof of the weak converse to the identification coding theorem for the DMC

We present here a new approach to polynomial converses for identification, which are explained in Section 2. We consider the proof being simpler than its predecessors. (Except for those in case of feedback [17], [27].)

Moreover, the approach is applicable to multi–way channels.

Furthermore, in contrast to the proofs in [28], [29] the approach works also for channels without a strong converse for transmission.

We begin our analysis with any channel $W : \mathcal{X} \to \mathcal{Y}$, that is, a time free situation and its (N,λ) codes $\Big\{(P_i,\mathcal{D}_i) : 1 \le i \le N\Big\}$ with $P_i \in \mathcal{P}(\mathcal{X}), \mathcal{D}_i \subset \mathcal{Y}$,

$$\sum_x P_i(x)W(\mathcal{D}_i|x) > 1-\lambda \ \text{ for all } \ i \ \text{ and } \ \sum_x P_i(x)W(\mathcal{D}_j|x) < \lambda \ \ (i \ne j).$$

For any distribution $P_X \in \mathcal{P}(\mathcal{X})$ we write P_{XY} for $P_X \times W$.

For any set $G \subset \mathcal{X} \times \mathcal{Y}$ we introduce

$$\rho(G) = \min_{(x,y)\in G} \frac{P_{XY}(x,y)}{P_X(x)P_Y(y)} \tag{12.1}$$

and

$$\sigma(G) = \max_{(x,y)\in G} \frac{P_{XY}(x,y)}{P_X(x)P_Y(y)}. \tag{12.2}$$

The ratio $\rho(G)\sigma(G)^{-1}$ measures how "informationally balanced" the set G is under P_{XY}. Clearly $0 \le \rho(G)\sigma(G)^{-1} \le 1$ and the closer to 1 the ratio is the more balanced G is.

We state now our key results.

Lemma 5. (Codes in informationally balanced sets)

For any $G \subset \mathcal{X} \times \mathcal{Y}$, $P_{XY} = P_X W$, and any $\delta' < P_{XY}(G)$ there exists a transmission code $\Big\{(u_i, \mathcal{E}_i) : 1 \leq i \leq M\Big\}$ with

(i) $\mathcal{E}_i \subset G_{u_i} = \Big\{y : (u_i, y) \in G\Big\}$
(ii) $W(\mathcal{E}_i|u_i) > \delta'$ *for* $i = 1, 2, \dots, M$
(iii) $M \geq \Big(P_{XY}(G) - \delta'\Big)\rho(G)$
(iv) $M < \frac{\sigma(G)}{\delta'}$ *(This holds for* any *code with (i) and (ii))*
(v) $P_Y\left(\bigcup\limits_{i=1}^{M} \mathcal{E}_i\right) \geq P_{XY}(G) - \delta'$.
(vi) *For* $Q(y) \triangleq \frac{1}{M}\sum\limits_{i=1}^{M} W(y|u_i) \quad Q(y) \geq \delta'\rho(G)\sigma(G)^{-1}P_Y(y)$, *if* $y \in \mathcal{E} = \bigcup\limits_{i=1}^{M} \mathcal{E}_i$.

Proof: Let $u_1 \in \mathcal{X}$ satisfy $W(G_{u_1}|u_1) > \delta'$. Its existence follows from $P_{XY}(G) > \delta'$. Set $\mathcal{E}_1 = G_{u_1}$, then define $(u_2, \mathcal{E}_2), \dots, (u_{j-1}, \mathcal{E}_{j-1})$ and add $u_j \in \mathcal{X}$ with $\mathcal{E}_j = G_{u_j} \smallsetminus \bigcup\limits_{i=1}^{j-1} \mathcal{E}_i$ and $W(\mathcal{E}_j|u_j) > \delta'$.

The procedure terminates at M, when no pair can be added subject to the constraints (i) and (ii). Consequently for all $x \in \mathcal{X}$

$$W\left(G_x \smallsetminus \bigcup_{i=1}^{M} \mathcal{E}_i \Big| x\right) \leq \delta'. \tag{12.3}$$

Since obviously for all $(x, y) \in G$

$$W(y|x) \geq \rho(G)P_Y(y) \tag{12.4}$$

and since $1 \geq W(\mathcal{E}_i|u_i)$, we have

$$P_Y(\mathcal{E}_i) \leq \rho(G)^{-1}. \tag{12.5}$$

It follows from (12.3) that

$$P_{XY}\left(G \smallsetminus \mathcal{X} \times \bigcup_{i=1}^{M} \mathcal{E}_i\right) \leq \delta'$$

and therefore also with (12.5)

$$P_{XY}(G) \leq \delta' + \sum_{i=1}^{M} P_Y(\mathcal{E}_i) \leq \delta' + M\rho(G)^{-1}.$$

This is (iii).

From the definition of σ for $(x,y) \in G$ $P_Y(y)\sigma(G) \geq W(y|x)$ and thus

$$P_Y(\mathcal{E}_i)\sigma(G) \geq W(\mathcal{E}_i|u_i) \text{ for } i = 1,2,\dots,M.$$

This gives (iv):

$$\sigma(G) \geq \sum_{i=1}^{M} W(\mathcal{E}_i|u_i) > M\delta'.$$

Further, (12.3) leads to $W(G_x|x) - W\left(\bigcup_{i=1}^{M} \mathcal{E}_i \middle| x\right) < \delta'$, which implies

$\sum\limits_x P_X(x)W(G_x|x) - \sum_x P_X(x)W\left(\bigcup_{i=1}^{M} \mathcal{E}_i \middle| x\right) = P_{XY}(G) - P_Y\left(\bigcup_{i=1}^{M} \mathcal{E}_i\right) < \delta'$ and hence (v).

Finally, by definition of ρ for $y \in \mathcal{E}_i \subset G_{u_i}$

$$W(y|u_i) \geq \rho(G)P_Y(y)$$

and by (iv)

$$\frac{1}{M}W(y|u_i) \geq \delta'\sigma(G)^{-1}\rho(G)P_Y(y).$$

Therefore

$$Q(y) \geq \delta'\sigma(G)^{-1}\rho(G)P_Y(y)$$

for all $y \in \bigcup\limits_{i=1}^{M} \mathcal{E}_i$.

The freedom in the choice of G or even several G's makes the power of this approach. We explain this in Sections 13, 14 and 15.

Obviously, we get good bounds, if $\rho(G)$ and $\sigma(G)$ are close to each other. We achieve this with our next idea to partition

$$G_{XY} = \Big\{(x,y) \in \mathcal{X} \times \mathcal{Y} : P_{XY}(x,y) > 0\Big\}$$

into informationally balanced sets and a set with big value of ρ, which we exclude.

Introduce

$$\begin{aligned} G(I+\beta) &= G_{XY}(I(X \wedge Y) + \beta) \\ &= \left\{(x,y) \in G_{XY} : \log \frac{P_{XY}(x,y)}{P_X(x)P_Y(y)} < I(X \wedge Y) + \beta\right\} \end{aligned}$$

and for suitable $\theta > 0$ and positive integer L, to be specified below, the partition

$$G(I+\beta) = \bigcup_{\ell=0}^{L-1} G_{XY}^{\ell}(I+\beta), \quad \text{where}$$

$$G_{XY}^{\ell}(I+\beta) = G_{XY}(I+\beta-\ell\theta) - G_{XY}\Big(I+\beta-(\ell+1)\theta\Big).$$

Its atoms are balanced, because

$$\frac{\sigma\Big(G_{XY}^{\ell}(I+\beta)\Big)}{\rho\Big(G_{XY}^{\ell}(I+\beta)\Big)} \le e^{\theta}.$$

For the further analysis we need a simple fact about I–divergences.

Lemma 6. *For any PD's p, q on $\mathcal{Z}$ and any $\mathcal{Z}' \subset \mathcal{Z}$*

$$\sum_{z \in \mathcal{Z}'} p(z) \log \frac{p(z)}{q(z)} \ge -e^{-1} \log_2 e = -c, \quad \textit{say.}$$

Proof:

$$\begin{aligned}
\sum_{z\in\mathcal{Z}'} p(z) \log \frac{p(z)}{q(z)} &= p(\mathcal{Z}') \sum_{z\in\mathcal{Z}'} \frac{p(z)}{p(\mathcal{Z}')} \log \frac{p(z)/p(\mathcal{Z}')}{q(z)/q(\mathcal{Z}')} + p(\mathcal{Z}') \log \frac{p(\mathcal{Z}')}{q(\mathcal{Z}')} \\
&\ge p(\mathcal{Z}') \log \frac{p(\mathcal{Z}')}{q(\mathcal{Z}')} \quad \text{(by nonnegativity of } I\text{–divergence)} \\
&\ge p(\mathcal{Z}') \log p(\mathcal{Z}') \left(\text{since} \quad \log \frac{1}{q(\mathcal{Z}')} \ge 1 \right) \\
&\ge \min_{0\le t\le 1} t \log t = -e^{-1} \log_2 e.
\end{aligned}$$

We apply this fact to the PD's P_{XY} and $P_X \times P_Y$ and $\mathcal{Z}' = G(I+\beta)$. Thus

$$\begin{aligned}
I &= \sum_{(x,y)\in\mathcal{Z}'} P_{XY}(x,y) \log \frac{P_{XY}(x,y)}{P_X(x)P_Y(y)} + \sum_{(x,y)\notin\mathcal{Z}'} \log \frac{P_{XY}(x,y)}{P_X(x)P_Y(y)} \\
&\ge -c + \Big(1 - P_{XY}(G(I+\beta))\Big)(I+\beta)
\end{aligned}$$

or

$$P_{XY}\Big(G(I+\beta)\Big) \ge \frac{\beta - c}{\beta + I}.$$

We can choose ℓ such that

$$P_{XY}\Big(G^{\ell}_{XY}(I+\beta)\Big) \geq \frac{\beta - c}{(\beta+I)L}. \tag{12.6}$$

The set $G^{\ell}_{XY}(I+\beta)$ serves as our representation for P_{XY}.

Lemma 7. *For any distribution P_{XY} and set $D \subset \mathcal{Y}$ with*

$$P_Y(D) = \sum_{x\in\mathcal{X}} P_X(x) W(D|x) \geq 1 - \lambda$$

consider for any $\beta > 0$ and positive integer L the representative $G^{\ell}_{XY}(I+\beta)$. Then we have for $G = G^{\ell}_{XY}(I+\beta) \cap \mathcal{X} \times D$

(i) $P_{XY}(G) \geq \frac{\beta-c}{(\beta+I)L} - \lambda = \delta$, *say.*

For any $\delta' < \delta$ there is a code

$$\Big\{(u_i, \mathcal{E}_i) : 1 \leq j \leq M\Big\} \ \textit{ with } \ \mathcal{E}_j \subset G_{u_j} \subset D \ \textit{ for } \ j = 1, \dots, M$$

and the properties

(ii) $M \leq \frac{1}{\delta'} e^{I+\beta-\ell\cdot\theta}$

(iii) $P_Y\left(\bigcup\limits_{i=1}^{M} \mathcal{E}_i\right) \geq \delta - \delta'$

(iv) $\frac{1}{M}\sum\limits_{j=1}^{M} W(y|u_j) \geq \delta' e^{-\theta} P_Y(y)$ *for* $y \in E = \bigcup\limits_{i=1}^{M} \mathcal{E}_i$

(v) $\frac{1}{M}\sum\limits_{j=1}^{M} W(E|u_j) \geq \delta' e^{-\theta}(\delta - \delta') = \delta^*$, *say.*

Proof: (i) is a consequence of (12.6) and the assumption on D. Inequality (ii) follows from (iv) in Lemma 5 and inequality (iii) follows from (v) in Lemma 5 (and (i) above). Finally, this and (vi) in Lemma 5 give (iv) and (v).

Theorem 9. *Let the discrete (***not necessarily memoryless***) channel $W^n : \prod\limits_1^n \mathcal{X} \to \prod\limits_1^n \mathcal{Y}$ have an (n, N, λ_n) identification code $\{(P_i, \mathcal{D}_i) : 1 \leq i \leq N\}$, then for pairs of RV's (X_i^n, Y_i^n) with distribution $P_i \times W^n$*

$$\log\log N \leq \max_i I(X_i^n \wedge Y_i^n) + o(n) \ \textit{ if } \lambda_n \leq n^{-7}.$$

Proof: Consider any pair $(P_i, \mathcal{D}_i)$ and apply Lemma 7 for $D = \mathcal{D}_i$, $P_X = P_i$. However, we write now P_{X^n} instead of P_X. Also, for $P_i \times W^n$ we write $P_{X^nY^n}$ (instead of P_{XY}) and thus we write the representation for $P_{X^nY^n}$ as $G = G^{\ell}_{X^nY^n}(I+\beta) \cap (\mathcal{X}^n \times D)$.

Our goal is to choose parameters so that M in (ii) of Lemma 7 becomes small and δ^* in (v) of Lemma 7 becomes large. The first property guarantees that $\binom{|\mathcal{X}^n|}{M}$ is so small that the number of representing encoding sets $\{u_j : 1 \leq j \leq M\}$ meets the desired double exponential bound.

The second property insures an appropriate bound on the multiplicity of representing encoding sets.

Accordingly the proof goes in two steps.

Step 1: We choose for $\varepsilon > 0$ $\beta = \varepsilon n$ and for convenience we choose $\delta' = \delta/2$. Clearly, for n large by Lemma 7, (i) since c is constant

$$P_{X^nY^n}(G) \geq \frac{\beta}{(\beta + I)2L} - \lambda_n = \delta_n^*. \tag{12.7}$$

We choose $\theta = \frac{\beta+I}{2L}$.

Using (12.7) and Lemma 7 (i), (v) we get now

$$\delta_n^* \geq \frac{1}{4}\left(\frac{\beta}{(\beta+I)2L} - \lambda_n\right)^2 e^{-(I+\beta)/2L}.$$

Since $I = I(X^n \wedge Y^n) \leq n \log |\mathcal{X}|$, we get

$$\delta_n^* \geq \frac{1}{4}\left(\frac{\varepsilon}{(\log|\mathcal{X}| + \varepsilon)2L} - \lambda_n\right)^2 e^{-(\log|\mathcal{X}|+\varepsilon)(2L)^{-1}n}.$$

Notice that for any function $f(n) \to \infty (n \to \infty)$ the choice $L = L_n = n\ f(n)$ yields $\lim\limits_{n\to\infty} e^{-(\log|\mathcal{X}|+\varepsilon)L_n^{-1}n} = 1$ and the choices $f(n) = n^{1/2}$, $L_n = n^{3/2}$, $\lambda_n = n^{-7}$ yield $\delta_n^* \geq n^{-4}$ for n large.

These are not optimal calculations, but only polynomial growth and the fact $\delta_n^* \gg \lambda_n$ are relevant here!

By our choices and Lemma 7 (ii) – (v), $\delta \geq \lambda_n$ and

$$M \leq 2n^3\ e^{I(X_i^n \wedge Y_i^n)+\varepsilon n}. \tag{12.8}$$

This is the first desired property. The others are

$$P_Y\left(\bigcup_{i=1}^{M} \mathcal{E}_i\right) \geq \frac{\delta}{2} \geq \frac{1}{4}n^{-3/2}. \tag{12.9}$$

For $\mathcal{U} = \{u_1, \dots, u_n\}$

$$Q_{\mathcal{U}}(y) = \frac{1}{M} \sum_{j=1}^{M} W^n(y|u_i) \geq \frac{1}{2} n^{-3} P_Y(y) \tag{12.10}$$

and so

$$Q_{\mathcal{U}} \left(\bigcup_{i=1}^{M} \mathcal{E}_i \right) \geq \frac{1}{8} n^{-9/2}, \tag{12.11}$$

which is much bigger than $\lambda_n = n^{-7}$.

Step 2: If now $\mathcal{U}$ serves $K' \geq K$ other times as representative for $(P_{Y^j}, \mathcal{D}_{Y^j})$ with decoding sets $\{\mathcal{E}_i^j : 1 \leq i \leq M\}$, $j = 1, \dots, K'$, then K' can be suitably bounded.

Indeed, set $\mathcal{E}^j = \bigcup\limits_{i=1}^{M} \mathcal{E}_i^j$ and define disjoint sets

$$\mathcal{E}'^j = \mathcal{E}^j - \bigcup_{j' \neq j} \mathcal{E}^{j'}; j = 1, 2, \dots, K. \tag{12.12}$$

Since $\mathcal{E}^j \subset \mathcal{D}_{Y^j}$ and the identification code has error probabilities less than λ_n, we get from (12.9)

$$P_{Y^j}(\mathcal{E}'^j) \geq \frac{1}{4} n^{-3/2} - K\lambda_n \tag{12.13}$$

and thus by (12.10)

$$Q_{\mathcal{U}} \left(\bigcup_{j=1}^{K} \mathcal{E}'^j \right) = \sum_{j=1}^{K} Q_{\mathcal{U}}(\mathcal{E}'^j) \geq K \left(\frac{1}{4} n^{-3/2} - K\lambda_n \right) \cdot \frac{1}{2} n^{-3}.$$

Now for $K = 16\, n^{9/2}$ and $\lambda_n < \frac{1}{128} n^{-6}$ we have $\frac{1}{4}\, n^{-3/2} - K\lambda_n > \frac{1}{8}\, n^{-3/2}$ and thus

$$Q_{\mathcal{U}} \left(\bigcup_{j=1}^{K} \mathcal{E}'^j \right) > 1, \text{ a contradiction.}$$

So $\mathcal{U}$ serves at most $16\, n^{9/2}$ times as representative and the result follows with (12.8).

Remark 18: When determining pessimistic capacities or capacity regions the observations in Remark 1 are relevant.

13 Two promised results: characterisation of the capacity regions for the MAC and the BC for identification

We know from [3], [47] that the transmission capacity region $\mathcal{R}$ of a (classical: memoryless, stationary) MAC $W : \mathcal{X} \times \mathcal{Y} \to \mathcal{Z}$ can be characterized as the convex hull of the set of pairs $(R_{\mathcal{X}}, R_{\mathcal{Y}})$ of non–negative numbers which satisfy for some input distribution $P_{XY} = P_X \times P_Y$

$$\begin{aligned} R_{\mathcal{X}} &\leq I(X \wedge Z|Y) \\ R_{\mathcal{Y}} &\leq I(Y \wedge Z|X) \\ R_{\mathcal{X}} + R_{\mathcal{Y}} &\leq I(XY \wedge Z). \end{aligned} \tag{13.1}$$

Also, in [3] there is a non–single letter characterisation.

$$\mathcal{R} = \{ \Big(\frac{1}{n} I(X^n \wedge Z^n), \frac{1}{n} I(Y^n \wedge Z^n) \Big) : n \in \mathbb{N}, P_{X^n Y^n} \in \mathcal{P}(\mathcal{X}^n \times \mathcal{Y}^n),$$

$$P_{X^n Y^n} = P_{X^n} \times P_{Y^n} \}. \tag{13.2}$$

Quite surprisingly we can use this characterisation for the proof of the polynomial weak converse for identification via the MAC.

Theorem 10. *The second order identification capacity region for the MAC* equals *the first order transmission capacity region $\mathcal{R}$.*

The broadcast channel is a stochastic map

$$W^n : \mathcal{X} \to \mathcal{Y} \times \mathcal{Z}$$

with components $W_1^n : \mathcal{X} \to \mathcal{Z}$ and $W_2^n : \mathcal{X} \to \mathcal{Z}$ and set of messages or the object space is

$$\mathcal{N} = \mathcal{N}_{\mathcal{Y}} \times \mathcal{N}_{\mathcal{Z}}, |\mathcal{N}_{\mathcal{Y}}| = N_{\mathcal{Y}} \text{ and } |\mathcal{N}_{\mathcal{Z}}| = N_{\mathcal{Z}}$$

An identification code (n, N_1, N_2, λ) for the BC is a family

$$\Big\{ (P_{ij}, \mathcal{D}_i, \mathcal{F}_j) : 1 \leq i \leq N_1; 1 \leq j \leq N_2 \Big\},$$

where the $\mathcal{D}_i$'s are sets in $\mathcal{Y}^n$, the $\mathcal{F}_j$'s are sets in $\mathcal{Z}^n$ and $P_{ij} \in \mathcal{P}(\mathcal{X}^n)$, and

$$\sum_{x^n} W_1^n(\mathcal{D}_i|x^n)P_{ij}(x^n) \geq 1-\lambda \text{ for all } i \text{ and } j \tag{13.3}$$

$$\sum_{x^n} W_1^n(\mathcal{D}_{i'}|x^n)P_{ij}(x^n) \leq \lambda \text{ for all } i \neq i' \text{ and all } j \tag{13.4}$$

$$\sum_{x^n} W_2^n(\mathcal{F}_j|x^n)P_{ij}(x^n) \geq 1-\lambda \text{ for all } j \text{ and } i \tag{13.5}$$

$$\sum_{x^n} W_2^n(\mathcal{F}_{j'}|x^n)P_{ij}(x^n) \leq \lambda \text{ for all } j \neq j' \text{ and all } i. \tag{13.6}$$

Let $\mathcal{B}$ be the set of all achievable pairs $(R_\mathcal{Y}, R_\mathcal{Z})$ of second order rates. For its analysis we need the cones

$\mathbb{R}^{2+}_\mathcal{Y} = \{(R_1, R_2) \in \mathbb{R}^2 : R_1 \geq R_2 \geq 0\}$ and $\mathbb{R}^{2+}_\mathcal{Z} = \{(R_1, R_2) \in \mathbb{R}^2 : R_2 \geq R_1 \geq 0\}$.

We can write $\mathcal{B}$ as a union $\mathcal{B} = \mathcal{B}^+_\mathcal{Y} \cup \mathcal{B}^+_\mathcal{Z}$, where

$$\mathcal{B}^+_\mathcal{Y} = \mathcal{B} \cap \mathbb{R}^{2+}_\mathcal{Y} \text{ and } \mathcal{B}^+_\mathcal{Z} = \mathcal{B} \cap \mathbb{R}^{2+}_\mathcal{Z}.$$

Our key observation is that for identification we can relate the capacity regions for identification of independent messages to the capacity regions for identification for degraded message sets, $\mathcal{A}_\mathcal{Y}$ and $\mathcal{A}_\mathcal{Z}$, where $\mathcal{A}_\mathcal{Y}$ (resp. $\mathcal{A}_\mathcal{Z}$) concerns the pairs of the rates of separate messages for $\mathcal{Y}$ (resp. $\mathcal{Z}$) and of common messages for $\mathcal{Y}$ and $\mathcal{Z}$. Since common messages can be interpreted as separated messages obviously

$$\mathcal{A}_\mathcal{Y}, \mathcal{A}_\mathcal{Z} \subset \mathcal{B}.$$

We can also write

$$\mathcal{A}^+_\mathcal{Y} = \mathcal{A}_\mathcal{Y} \cap \mathbb{R}^{2+}_\mathcal{Y} \text{ and } \mathcal{A}^+_\mathcal{Z} = \mathcal{A}_\mathcal{Z} \cap \mathbb{R}^{2+}_\mathcal{Z}$$

and notice that

$$\mathcal{A}^+_\mathcal{Y} \subset \mathcal{B}^+_\mathcal{Y}, \mathcal{A}^+_\mathcal{Z} \subset \mathcal{B}^+_\mathcal{Z}.$$

We come now to a key tool

Lemma 8 (Reduction).

(i) $\mathcal{B}^+_\mathcal{Y} \subset \mathcal{A}^+_\mathcal{Y}$ *and* $\mathcal{B}^+_\mathcal{Z} \subset \mathcal{A}^+_\mathcal{Z}$.
(ii) $\mathcal{B}^+_\mathcal{Y} = \mathcal{A}^+_\mathcal{Y}$ *and* $\mathcal{B}^+_\mathcal{Z} = \mathcal{A}^+_\mathcal{Z}$.
(iii) $\mathcal{B} = \mathcal{A}$.

Proof: By previous observations it remains to show (i) and by symmetry only its first part.

Let $\Big\{(P_{ij}, \mathcal{D}_i, \mathcal{E}_j) : 1 \le i \le N_\mathcal{Y}, 1 \le j \le N_\mathcal{Z}\Big\}$ be an identification code for the BC with error probabilities $\le \lambda$. Since $R_\mathcal{Z} \le R_\mathcal{Y}$ we can define for

$$\ell = 1, \dots, N_\mathcal{Z} \ \text{ and } \ m = 1, \dots, \frac{N_\mathcal{Y}}{N_\mathcal{Z}}$$

(where w.l.o.g. divisibility of $N_\mathcal{Y}$ by $N_\mathcal{Z}$ can be assumed)

$$Q_{\ell,m} = P_{\ell,(m-1)N_\mathcal{Y}+\ell}.$$

The $\mathcal{Z}$–decoder identifies ℓ and the $\mathcal{Y}$–decoder identifies $(m-1)N_\mathcal{Y} + \ell$ or equivalently ℓ **and** m, that is, the common part and a separate part.

If $\mathcal{R}_\mathcal{Y} > \mathcal{R}_\mathcal{Z}$, then with error probabilities $\le \lambda$

$$2^{2^{R_y n}} \cdot 2^{-2^{R_z n}} \sim 2^{2^{R_y n}}.$$

If $R_\mathcal{Y} = R_\mathcal{Z}$, then we can make the same construction with rates $R_\mathcal{Y}$ and $R_\mathcal{Z} - \varepsilon$.

We need the direct part of the ABC (asymmetric broadcastchannel) Coding Theorem for transmission ([5], [31], [30]). Here, there are separate messages for decoder $\mathcal{Y}$ (resp. $\mathcal{Z}$) and common messages for both decoders.

Achievable are (with maximal errors)

$$\begin{aligned}\mathcal{T}_\mathcal{Y} = \Big\{(R_\mathcal{Y}, R_0) : R_0 \le I(U \wedge Z), R_0 + R_\mathcal{Y} \le \min\Big[I(X \wedge Y), I(X \wedge Y|U) + I(U \wedge Z)\Big],\\ U \circleddash X \circleddash YZ, \ \ \|U\| \le |\mathcal{X}| + 2\Big\}\end{aligned}$$

resp.

$$\begin{aligned}\mathcal{T}_\mathcal{Z} = \Big\{(R_0, R_\mathcal{Z}) : R_0 \le I(U \wedge Y), R_0 + R_\mathcal{Z} \le \min\Big[I(X \wedge Z), I(X \wedge Z|U) + I(U \wedge Y)\Big],\\ U \circleddash X \circleddash YZ, \ \ \|U\| \le |\mathcal{X}| + 2\Big\}.\end{aligned}$$

This is our surprising result.

Theorem 11. *For the (general) BC the set of achievable pairs of second order rates is given by*

$$\mathcal{B} = \mathcal{T}'_\mathcal{Y} \cup \mathcal{T}'_\mathcal{Z},$$

where

$$\mathcal{T}'_\mathcal{Y} = \{(R'_\mathcal{Y}, R'_\mathcal{Z}) : \exists (R_\mathcal{Y}, R_0) \in \mathcal{T}_\mathcal{Y} \text{ with } R'_\mathcal{Y} = R_\mathcal{Y} + R_0, R'_\mathcal{Z} = R_0\}$$

and $\mathcal{T}'_{\mathcal{Z}} = \{(R'_{\mathcal{Y}}, R'_{\mathcal{Z}}) : \exists (R_0, R_{\mathcal{Z}}) \in \mathcal{T}_{\mathcal{Z}}$ with $R'_{\mathcal{Y}} = R_0, R'_{\mathcal{Z}} = R_0 + R_{\mathcal{Z}}\}$.

14 The proof for the MAC

The proof of achievability is straightforward by the second method of Ahlswede/ Dueck [17], that is, the Transformator Lemma. Indeed, use an *average* error transmission code in blocklength n

$$\Big\{(u_i, v_j, \mathcal{D}_{ij}) : 1 \le i \le M_{\mathcal{X}}, 1 \le j \le M_{\mathcal{Y}}\Big\}$$

with

$$\frac{1}{M_{\mathcal{X}}}\frac{1}{M_{\mathcal{Y}}}\sum_{i,j} W^n(\mathcal{D}^c_{ij}|u_i, v_j) \le \lambda. \tag{14.1}$$

Then of course also

$$\frac{1}{M_{\mathcal{X}}}\sum_i \left(\frac{1}{M_{\mathcal{Y}}}\sum_j W^n\left(\left(\bigcup_{j'} \mathcal{D}_{ij'}\right)^c |u_i, v_j\right)\right) \le \lambda \tag{14.2}$$

and we have a random experiment U with $\text{Prob}(U = u_i) = \frac{1}{M_{\mathcal{X}}}$, whose outcome is known to sender $S_{\mathcal{X}}$ and with probability at least $1 - \lambda$ also to the receiver.

Analogously, there is a random experiment V for the sender $S_{\mathcal{Y}}$ and the receiver. We have used blocklength n.

As in [17] by the Transformator Lemma with relatively few, say $\sqrt{n}$, letters (actually even $o(\log n)$) identification of *second* order rate $\sim \frac{1}{n}\log M_{\mathcal{X}}$ can be performed from $S_{\mathcal{X}}$ to the receiver. Finally, with other $\sqrt{n}$ letters the identification of *second* order rate $\sim \frac{1}{n}\log M_{\mathcal{Y}}$ can be done from $S_{\mathcal{Y}}$ to the receiver.

Remark 19: In our proof of the direct part the identification is done separately for both encoders. The encoding strategy pair (P_i, Q_j) and the decodings $\mathcal{D}_i, \mathcal{F}_j$ identify i and j separately. We can also choose $\mathcal{E}_{ij} = \mathcal{D}_i \cap \mathcal{F}_j$ and notice that

$$\sum_{x^n, y^n} W^n(\mathcal{E}_{ij}|x^n, y^n)P_i(x^n)Q_j(y^n) > 1 - 2\lambda$$

$$\sum_{x^n, y^n} W^n(\mathcal{E}^c_{i'j'}(x^n, y^n)P_i(x^n)Q_j(y^n) \le 2\lambda \ \text{ for } \ (i', j') \ne (i, j).$$

On the other hand, starting with the $\mathcal{E}_{ij}$'s we can define $\mathcal{D}_i = \bigcup\limits_j \mathcal{E}_{ij}$, $\mathcal{F}_j = \bigcup\limits_i \mathcal{E}_{ij}$.

Remark 20: The decomposition principle (see [8]) does not hold for identification on the MAC. If both encoders have independent messages, but can cooperate, then

$$R_{\mathcal{X}\mathcal{Y}} = \max_{P_X \times P_Y} I(XY \wedge Z)$$

and $2^{2^{n\, R_{\mathcal{X}\mathcal{Y}}}}$ is **much** bigger than

$$2^{2^{n\, R_{\mathcal{X}}}} \cdot 2^{2^{n\, R_{\mathcal{Y}}}} \sim 2^{2^{n\, \max(R_{\mathcal{X}}, R_{\mathcal{Y}})}}.$$

Remark 21 (Updating): Steinberg [60] did not use the Transformator Lemma, but followed the first approach in [16], which is based on a transmission code with small maximal errors. With deterministic maximal error transmission code the (average error) capacity region of a MAC cannot be achieved. However, it can be achieved if stochastic encoders are used (as shown in [**61**]) and for those coding the approach of [16] again applies.

Problems:

6. Develop a theory for identification of correlated data (see "correlated codes" in [12]).
7. Develop approximation of output statistics for the MAC to obtain a strong converse. Use random coding instead of maximal coding with rates

$$\begin{aligned} I(X \wedge Z) &\leq R_{\mathcal{X}} \leq I(X \wedge Z|Y) \\ I(Y \wedge Z) &< R_{\mathcal{Y}} \leq I(Y \wedge Z|X) \\ I(XY \wedge Z) &\leq R_{\mathcal{X}} + R_{\mathcal{Y}} \end{aligned}$$

   and code structure $\{u_1, \dots, u_{M_{\mathcal{X}}}\}$ and $\{v_{i1}, \dots, v_{iM_{\mathcal{Y}}}\}$ for $i = 1, \dots, M_{\mathcal{X}}$.

Converse proof: We follow closely the proof for a one–way channel. **Here it is essential that our approach treats general channels with memory.** Secondly we use **the characterisation (13.2) of the rate–region $\mathcal{R}$ for the MAC**.

In addition we partition our encoding pairs $(P_i \times Q_j)_{\substack{i=1,\dots,N_{\mathcal{X}} \\ j=1,\dots,N_{\mathcal{Z}}}}$ according to the values of their corresponding pairs of mutual informations $\Big(I(X_i^n \wedge Z_{ij}^n), I(Y_j^n \wedge Z_{ij}^n)\Big)$ where $P_{X_i^n} = P_i$, $P_{Y_j^n} = Q_j$, $P_{Z_{ij}^n} = (P_i \times Q_j)W^n$, as follows.

Endow $\mathbb{R}^2$ and, particularly,

$$S = \Big\{(R_1, R_2) : 0 \leq R_1 \leq \log|\mathcal{X}|, 0 \leq R_2 \leq \log|\mathcal{Y}|\Big\}$$

with a rectangular lattice with side lengths η. So we get $g(\eta) = g_1(\eta) \cdot g_2(\eta)$ rectangles, if $g_1(\eta) = \frac{\log|\mathcal{X}|}{\eta}$, $g_2(\eta) = \frac{\log|\mathcal{Y}|}{\eta}$.

Label them as $S_{a,b}(1 \le a \le g_1(\eta), 1 \le b \le g_2(\eta)$ and associate with $P_i \times Q_j$ the rectangle $S_{a(i,j),b(i,j)}$, where

$$\left(\frac{1}{n} I(X_i^n \wedge Z_{ij}^n), \frac{1}{n} I(Y_j^n \wedge Z_{ij}^n)\right) \in S_{a(i,j),b(i,j)}. \tag{14.3}$$

There is a rectangle S^* with which at least $\frac{N_{\mathcal{X}} \cdot N_{\mathcal{Y}}}{g(\eta)}$ encodings $P_i \times Q_j$ are associated. Denote them by $(P_i \times Q_j)_{(i,j) \in \mathcal{N}(\eta)}$.

Their corresponding pairs of (normalized) mutual informations differ componentwise by at most η.

Furthermore, there is a row index i^* and a column index j^* so that

$$\left|\left\{(i^*, j) : (i^*, j) \in \mathcal{N}(\eta)\right\}\right| \ge \frac{|\mathcal{N}(\eta)|}{N_{\mathcal{X}}} \ge \frac{N_{\mathcal{Y}}}{g(\eta)}, \tag{14.4}$$

$$\left|\left\{(i, j^*) : (i, j^*) \in \mathcal{N}(\eta)\right\}\right| \ge \frac{|\mathcal{N}(\eta)|}{N_{\mathcal{Y}}} \ge \frac{N_{\mathcal{X}}}{g(\eta)}. \tag{14.5}$$

Now our previous converse proof comes in. To every triple $(P_i, Q_j, \mathcal{D}_{ij})$ we assign two codes $(\mathcal{U}_i^j, \mathcal{E}_i^j), (\mathcal{V}_j^i, \mathcal{F}_j^i)$, where $\mathcal{U}_i^j \subset \mathcal{X}^n$, $\mathcal{E}_i^j = \{E_{i1}^j, \dots, E_{iM_{i\mathcal{X}}^j}^j\}$, (pairwise disjoint), $\mathcal{V}_j^i \subset \mathcal{Y}^n$, $\mathcal{F}_j^i = \{F_{j1}^i, \dots, F_{jM_{j\mathcal{Y}}^i}^i\}$ (pairwise disjoint), and all decoding sets are subsets from $\mathcal{D}_{ij}$. Here

$$M_{i\mathcal{X}}^j \le \exp\left\{I(X_i^n \wedge Z_{ij}^n) + o(n)\right\}$$

$$M_{j\mathcal{Y}}^i \le \exp\left\{I(Y_j^n \wedge Z_{ij}^n) + o(n)\right\}$$

and (14.3) holds.

Moreover, for all indices

$$\frac{1}{M_{i\mathcal{X}}^j} \sum_{u \in \mathcal{U}_i^j} \sum_{y^n} W^n(E_{iu}^j \cap \mathcal{D}_{ij} | u, y^n) Q_j(y^n) \ge n^{-4} \tag{14.6}$$

and analogous relations hold for $\mathcal{V}_j^i$.

Now observe that for all $(i,j) \in \mathcal{N}(\eta)$

1.) $\frac{1}{n} \log M_{i\mathcal{X}}^j \le R_{\mathcal{X}}^* + \eta$ and $\frac{1}{n} \log M_{j\mathcal{Y}}^i \le R_{\mathcal{Y}}^* + \eta$.
2.) By (14.4), (14.5) there are at most $\binom{|\mathcal{X}|^n}{2^{(R_{\mathcal{X}}^* + \eta)n}}$ different codes $\mathcal{U}_{i^*}^j$ in row i^* and at most $\binom{|\mathcal{Y}|^n}{2^{(R_{\mathcal{Y}}^* + \eta)n}}$ codes $\mathcal{V}_{j^*}^i$ in column j^*.

Furthermore the multiplicity K_{i^*} of codes in row i^* (resp. K_{j^*} for column j^*) does not exceed n^6 (as previously).

Finally, therefore

$$\frac{1}{n}\log\log N_{\mathcal{X}} \leq R^*_{\mathcal{X}} + 2\eta \text{ and } \frac{1}{n}\log\log \mathcal{N}_{\mathcal{Y}} \leq R^*_{\mathcal{Y}} + 2\eta.$$

Problem 8 (Updating): In [60] Steinberg strengthens our polynomial converse to a weaker converse. The main difference of his proof is a sharpening of the bound in Theorem 9, which is based on a generalization of Lemma 5 in [29]. We suggest as a further improvement to establish a strong converse by our hypergraph lemma, which is presented in Section VI of [58]. Otherwise in his proof the same ideas are used, namely facts (13.1) and (13.2) and a suitable subcode selection. The whole proof with all auxiliary results exceeds the present one in length roughly by a factor 3.

15 The proof for the BC

The direct part: We use the Reduction Lemma and the ABC Coding Theorem mentioned in Section 13. Eventhough that theorem holds for maximal errors we use average errors so that the transmission codes establish two common random experiments of the sender with both receivers, resp., with rates in $\mathcal{T}'_{\mathcal{Y}} \cup \mathcal{T}'_{\mathcal{Z}}$.

The converse part: Suppose w.l.o.g. that $R_{\mathcal{Z}} < R_{\mathcal{Y}} + \varepsilon$, ε arbitrarily small, and that the $\mathcal{Y}$–decoder has a separate part coded into row numbers and that the common part for both decoders is coded into column numbers with the encodings $(P_{uv})_{\substack{u=1,\dots,N_{\mathcal{Y}}\\ v=1,\dots,N_{\mathcal{Z}}}}$.

Note that we can start with a smaller common rate, so that $M_{\mathcal{Y}} \sim M_{\mathcal{Z}} \cdot M_{\mathcal{Y}}$ (If the common rate is bigger in the ABC model, we can convert this by the Reduction Lemma 8).

We associate RV's and information quantities as follows:

Let U, V be auxiliary RV's with $\text{Prob}((U,V) = (u,v)) = \frac{1}{N_{\mathcal{Y}} N_{\mathcal{Z}}}$ for $u = 1, \dots, N_{\mathcal{Y}}$ and $v = 1, \dots, N_{\mathcal{Z}}$. Furthermore let X^n take values in $\mathcal{X}^n$ with conditional PD $P_{X^n|U=u,V=v}(x^n) = P_{uv}(x^n)$, let Y^n take values in $\mathcal{Y}^n$ with conditional PD $P_{Y^n|U=u,V=v}(y^n) = \sum_{x^n} P_{uv}(x^n) W_1^n(y^n|x^n)$, and let Z^n take values in $\mathcal{Z}^n$ with conditional PD $P_{Z^n|U=u,V=v}(z^n) = \sum_{x^n} P_{uv}(x^n) W_2^n(z^n|x^n)$.

Thus we get information quantities

$$I(U \wedge Z^n | V = v), I(X^n \wedge Y^n | U, V = v), \quad \text{and} \quad I(X^n \wedge Y^n | V = v)$$

and the Markov condition $(U,V) \multimap X^n \multimap (Y^n, Z^n)$.

As in the proof of Theorem 10 we make η–approximations, first for all $\frac{1}{n}I(X^n \wedge Y^n|V=v)$ with biggest class of value I_{η_3}.

This gives as in the one–way channel coding theorem for identification

$$\frac{1}{n}\log\log N_{\mathcal{Y}} \leq I_{\eta_3}. \tag{15.1}$$

In the remaining matrix keep I_{η_2} for $I(X^n \wedge Y^n|U, V=v)$ and then all $I(U \wedge Z^n|V=v)$ approximately I_{η_1}.

We upper bound the number of columns by upper bounding the number of codes (via Lemma 7) representing triples $(P_{U|V=v}, P_{Z^n|U,V=v}, \mathcal{D}_v)$. Thus for $\lambda_n = n^{-6}$ (as usual)

$$\frac{1}{n}\log\log N_{\mathcal{Z}} \leq I_{\eta_1} + 2\eta. \tag{15.2}$$

Within column v^* a significant number of terms has

$$\frac{1}{n}I(X^n \wedge Y^n|U=u, V=v^*) \leq I_{\eta_2} + \beta^*.$$

This gives the desired row number estimate

$$\frac{1}{n}\log\log N_{\mathcal{Y}} \leq \min(I_{\eta_1} + I_{\eta_2}, I_{\eta_3}) + 2\eta + \beta^*.$$

$$= \min\Big(I(U \wedge Z^n|V=v^*) + I(X^n \wedge Y^n|U, V=v^*), I(X^n \wedge Y^n|V=v^*)\Big) + 2\eta + \beta^*$$

and thus $(R_{\mathcal{Y}}, R_{\mathcal{Z}}) \in \mathcal{T}'_{\mathcal{Y}}$ by the converse in the ABC Coding Theorem, which shows that the information quantities single-letterize.

Remark 22: Theorem 11 has an important consequence. Whereas for one-way channels the common randomness capacity equals the transmission capacity and the transmission capacity region is still unknown for general broadcast channels **we know now its common randomness capacity region**, where common random experiments for $\mathcal{X}$-encoder and $\mathcal{Y}$-decoder and, simultaneously, for $\mathcal{X}$-encoder and $\mathcal{Z}$-decoder are generated. **Indeed it equals the second order identification capacity region!**

That the latter includes the former is clear from our proof of the direct part. The reverse implication follows indirectly by the same argument.

Interesting here is that the outer bound for the common randomness capacity region is proved via identification.

The situation changes, if constraints like independency or security are imposed on the two common random experiments.

A transmission code with rates $(R_{\mathcal{Y}}, R_{\mathcal{Z}})$ can be used for independent common random experiments and thus the transmission capacity region for the general broadcast channel is contained in the identification capacity region.

Finally we mention that the identification capacity region $T'_{\mathcal{Y}} \cup T'_{\mathcal{Z}}$ is convex, because it equals the common randomness capacity region for which time sharing applies and thus convexity is given.

Part V: Data compression

16 Noiseless coding for identification

Let $(\mathcal{U}, P)$ be a source, where $\mathcal{U} = \{1, 2, \dots, N\}$, $P = (P_1, \dots, P_N)$, and let $\mathcal{C} = \{c_1, \dots, c_N\}$ be a binary prefix code (PC) for this source with $\|c_u\|$ as length of c_u. Introduce the RV U with $\text{Prob}\,(U = u) = p_u$ for $u = 1, 2, \dots, N$ and the RV C with $C = c_u = (c_{u_1}, c_{u_2}, \dots, c_{u\|c_u\|})$ if $U = u$.

We use the PC for noiseless identification, that is user u wants to know whether the source output equals u, that is, whether C equals c_u or not. He iteratively checks whether $C = (C_1, C_2, \dots)$ coincides with c_u in the first, second, etc. letter and stops when the first different letter occurs or when $C = c_u$.

What is the expected number $L_{\mathcal{C}}(P, u)$ of checkings?

In order to calculate this quantity we introduce for the binary tree $T_{\mathcal{C}}$, whose leaves are the codewords $c_1, \dots, c_N$, the sets of leaves $\mathcal{C}_{ik} (1 \le i \le N; 1 \le k)$, where $\mathcal{C}_{ik} = \{c \in \mathcal{C} : c$ coincides with c_i exactly until the k'th letter of $c_i\}$. If C takes a value in $\mathcal{C}_{uk}, 0 \le k \le \|c_u\| - 1$, the answers are k times "Yes" and 1 time "No". For $C = c_u$ the answers are $\|c_u\|$ times "Yes". Thus

$$L_{\mathcal{C}}(P, u) = \textstyle\sum_{k=0}^{\|c_u\|-1} P(C \in \mathcal{C}_{uk})(k+1) + \|c_u\| P_u.$$

For code $\mathcal{C}$ $\;L_{\mathcal{C}}(P) = \max\limits_{1 \le u \le N} L_{\mathcal{C}}(P, u)$ is the expected number of checkings in the worst case and $L(P) = \min\limits_{\mathcal{C}} L_{\mathcal{C}}(P)$ is this number for a best code.

Analogously, if $\tilde{\mathcal{C}}$ is a randomized coding, we introduce

$$L_{\tilde{\mathcal{C}}}(P, u), \; L_{\tilde{\mathcal{C}}}(P) \text{ and } \tilde{L}(P).$$

What are the properties of $L(P)$ and $\tilde{L}(P)$? We call for a kind of "identification entropies" serving as bounds like Boltzmann's entropy does in Shannon's source coding. Notice that every user comes with the same fixed code much faster to his goal to know "it's me – it's not me" than the one person in Shannon's model, who wants to use the outcome of the source always.

Moreover, as in [44] one can replace the lengths $||c_u||$ by $\varphi(||c_u||)$ where $\varphi : \mathbb{R}_+ \to \mathbb{R}_+$ is continuous and strictly monotone increasing.

Thus one gets functionals

$$L(P, \varphi) \text{ and } \tilde{L}(P, \varphi).$$

We shall analyze these quantities on another occasion and confine ourself here to deriving some simple facts.

Let us start with $P_N = \left(\frac{1}{N}, \dots, \frac{1}{N}\right)$ and set $f(N) = L(P_N)$. Clearly

$$f(2^k) \leq 1 + \frac{1}{2} f(2^{k-1}), f(2) = 1$$

and therefore

$$f(2^k) \leq 2 - 2^{-(k-1)}. \tag{16.1}$$

On the other hand it can be verified that

$$f(9) = 1 + \frac{10}{9} > 2 \text{ and more generally, } f(2^k + 1) > 2.$$

1. What is $\sup_N \big(f(N)\big)$?
2. Is $\tilde{L}(P) \leq 2$?
3. Suppose that encoder and decoder have access to a random experiment with unlimited capacity of common randomness (see [46]). Denote the best possible average codeword lengths by $L^*(P)$.

For $P = (P_1, \dots, P_N)$, $N \leq 2^k$ write $P' = (P_1, \dots, P_N, 0, \dots, 0)$ with 2^k components. Use a binary regular tree of depth k with leaves $1, 2, \dots, 2^k$ represented in binary expansions.

The common random experiment with 2^k outcomes can be used to use 2^k cyclic permutations of $1, 2, \dots, 2^k$ for 2^k deterministic codes. For each i we get equally often 0 and 1 in its representation and an expected word length $\leq 2 - \frac{1}{2^{k-1}}$. The error probability is 0. Therefore $L^*(P) \leq 2 - 2^{-(k-1)} \leq 2$ for all P.

17 Noiseless coding for multiple purposes

In the classical theory of data compression the main concern is to achieve a short average length coding. Here we address a problem of noiseless coding, where different persons are interested in different aspects of the data and their accessibility. We begin with a specified question.

17.1 Persons are interested in different components of a Bernoulli Source

Consider a discrete memoryless binary, symmetric source (BSS) producing the output $X^n = (X_1, \dots, X_n)$. Suppose that there are n persons and that person t is interested in the outcome of $X_t (1 \leq t \leq n)$. A multiple purpose encoding (or program) shall be a sequence $f = (f_i)_{i=1}^{\infty}$ of functions $f_i : \{0,1\}^n \to \{0,1\}$. Person t requests sequentially the values $f_1(X^n), f_2(X^n), \dots$ and stops as soon as he/she has identified the value of X_t. Let $\ell(f,t)$ denote the number of requests of person t for program f. We are interested in the quantity

$$L(n) = \min_f \max_{1 \leq t \leq n} \mathbb{E}\ell(f,t). \tag{17.1}$$

The choice $f_i(X^n) = X_i (1 \leq i \leq n)$ gives $\ell(f,t) = t$ and thus $\max\limits_{1 \leq t \leq n} \ell(f,t) = n$.

Since $\frac{1}{n} \sum\limits_{t=1}^{n} \ell(f,t) = \frac{n+1}{2}$, one should do better. In [15] we stated the problem to determine $L(n)$. Don Coppersmith [26] gave a rather precise bound.

Theorem 12. $\frac{n+1}{2} \leq L(n) \leq \frac{n+2}{2}$.

Proof: The lower bound is obvious, because

$$L(n) \geq \min_f \frac{1}{n} \sum_{t=1}^{n} \mathbb{E}\ell(f,t)$$

and

$$\mathbb{E}|\{t : 1 \leq t \leq n, \ell(f,t) \leq i\}| \leq i.$$

For the upper bound set $f_1(X^n) = X_1$ and for $2 \leq i \leq n$ set $f_i(X^n) = \begin{cases} X_i & \text{if } X_1 = 0 \\ X_{n+2-i} & \text{if } X_1 = 1. \end{cases}$

For $t > 1$ the stopping time is either t or $n+2-t$, each with probability $\frac{1}{2}$, so that the mean is $\mathbb{E}\ell(f,t) = \frac{n+2}{2}$, while obviously $\ell(f,1) = 1$. Thus $L(n) \leq \frac{n+2}{2}$.

Remark 23: A weaker upper bound, but more uniform distribution of the stopping times is obtained as follows: Let the first $\lceil \log_2 n \rceil$ bits be

$$\Big(f_1(X^n), f_2(X^n), \dots, f_{\lceil \log n \rceil}(X^n) \Big) = (X_1, X_2, \dots, X_{\lceil \log n \rceil})$$

and let these $\log n$ bits index a cyclic shift of the remaining $n - \log n$ bits so that the distribution of stopping times is approximately uniform between $\log n$ and n for $t > \lceil \log n \rceil$. This leads to the weaker upper bound

$$L(n) \leq (n + \log_2 n + c)/2.$$

Remark 24: Notice that both procedures are probabilistic algorithms. They exploit the randomness of the source.

17.2 Noiseless source coding problems of infinite order: Ordering and Identification

We consider here a source coding version of the ordering problem and also of the identification problem.

To simplify technicalities we assume that $N = 2^n$. We also assume that any element of $\{0,1\}^n$ is a source output with equal probabilities.

For any $u^n \in \{0,1\}^n$: Is the source output $x^n = (x_1, x_2, \dots, x_n)$ before u^n, that is, $x^n \leq u^n$ (lexicographically), or not? There is a canonical encoding function $f = (f_1, \dots, f_n)$ with $f_t(X_1, \dots, X_n) = X_t$. The person interested in u^n stops, when his/her question is answered. He/she stops at the smallest t with $f_t(u_t) \neq f_t(X_t)$.

The distributions of the stopping times don't depend on u^n. Let T_n denote the expected stopping time.

Lemma 9. $T_n = 1 + \frac{1}{2} T_{n-1} = \frac{2^n - 1}{2^{n-1}}$, $n \geq 1$.

This is a simple exercise. Notice that

$$\lim_{n \to \infty} T_n = 2. \tag{17.2}$$

So the compression rate exceeds any finite order.

Now let the question be "Does X^n equal u^n or not?" (Identification)

We use again a multi–purpose encoding function. Actually we can use the same function as before. There is also the same recursion for T_n. Notice that in case of identification for $X^n = u_n$ we have maximal running time, namely n.

17.3 Problems

9. It is interesting to study the previous problems for other distributions on $\{0,1\}^n$. In general the previous encoding function is not optimal (for instance if $\text{Prob}(X_1 = 0) = 1$).

   An instructive source is given by the distribution which assigns probability $\frac{1}{n}$ to the sequences starting with k 1's and continuing with 0's only. For $u^n = (1,1,\dots,1)$ the running time of the previous encoding function is always n. However, by choosing $f_1(X^n) = X_{\lceil \frac{n}{2} \rceil}$ etc. the worst case expected running time is still less than 2.

10. For any distribution P on $\{0,1\}^n$, is the worst case expected running time less than 2? In case the answer is negative, determine the best constant (independent of n) upper bound! An obvious algorithm: number probabilities in decreasing order; $P_1 \geq P_2 \geq \cdots \geq P_N$ and divide as equally as possible $P_1 + P_2 + \cdots + P_{N_1}$, $P_{N_1+1} + \cdots + P_N$. $f_1(X^n)$ says whether $i \in \{1,\dots,N_1\}$ or not, etc.

    We *conjecture* that the bound 2 is achievable, if randomisation in the encoding is permitted. Two simple examples illustrate the advantage of randomisation. Denote by $E_{P,i}(f)$ the expected running time for source distribution P, object i, and encoding function f.

    For $P = \left(\frac{1}{4},\frac{1}{4},\frac{1}{4},\frac{1}{4}\right)$ and f based on division into two equal parts gives

    $$T_{P,i}(f) = 1 + \frac{1}{2} \ (i = 1,2,3,4).$$

    For $Q = \left(\frac{1}{3},\frac{1}{3},\frac{1}{3}\right)$ and f based on the division $\left\{\left\{\frac{1}{3}\right\},\left\{\frac{1}{3},\frac{1}{3}\right\}\right\}$ gives

    $$T_{Q,1}(f) = 1,\ T_{Q,i}(f) = 1 + \frac{2}{3} \ (i = 2,3).$$

    Therefore $\max_i T_{P,i}(f) < \max_i T_{Q,i}(f)$, however, $\sum_i T_{P,i}(f) > \sum_i T_{Q,i}(f)$ and randomisation takes advantage of this fact, by smoothing out the differences between the individual running times.

    Let F choose with probabilities $\frac{1}{3}$ the partitions

    $$\Big\{\{1\},\{2,3\}\Big\},\Big\{\{1,3\},\{2\}\Big\},\Big\{\{3\},\{1,2\}\Big\}$$

    in the first step, the second step is canonical. Then

    $$T_{Q,i}(F) = \frac{1}{3}\left(1 + \left(1 + \frac{2}{3}\right) + \left(1 + \frac{2}{3}\right)\right) = 1 + \frac{4}{9} < 1 + \frac{1}{2}(!).$$

11. It is also reasonable to study alphabetical source codes for identification. For example for different intervals of a pipeline different repairman are responsible. They want to know whether a defect occurred in their interval or not.
12. Suppose that $N = 2^k$ numbers are stored in 0–1 bits in a machine. Upon request a further bit is revealed by the machine. What is the average number of requests so that person i knows whether i occurred or not?
13. One can study multiple purpose coding problems with noise (see [22], which gives a common generalisation of Shannon's noiseless coding theorem and coding theorems for noisy channels). What are the generalisations (there is one in [22]) of Kraft's inequality?
14. These source coding problems open a whole area of research. Are there coding problems of an order between first order (as in the component problem) and infinite order (as in the ordering problem)?
15. It is remarkable in this context also that the ordering problem *via channels* is not easier than transmission, if maximal errors are used. However, if for the second kind error probability the average is taken, then the ordering problem becomes of infinite order (similar as the identification problem does). Indeed just map the numbers $1, \dots, N$ onto codewords of a transmission code $\big\{(u_i, \mathcal{D}_i) : 1 \leq i \leq\big\}$ as follows:

    For any $K < N$ write $j \in \{1, \dots, N\}$ as $j = rK + s$, $0 \leq s < K$, and map j on u_r. Now just let N go to infinity and choose $K = \left\lceil \frac{N}{M} \right\rceil$.
16. It is also interesting that for maximal second kind error probabilities the identification problem via channels has second order behaviour whereas — as mentioned before — the ordering problem has first order behaviour.

    We therefore ask the following question:

    Is there a reasonable coding problem with average error of second kind as performance criterion which is neither of first order nor of infinite order behaviour? In the positive case, what is the hierarchy of all orders?
17. If $\kappa < \frac{1}{2} C_{Sh}$, then first order capacity R_1 equals infinity. However, if $\kappa > \frac{1}{2} C_{Sh}$, is then $R_1 > C_{Sh}$ possible?

Part VI: Perspectives

Our models go considerably beyond Shannon's transmission model and the model of identification. They will greatly enlarge the body of Information Theory. We substantiate here this belief by a brief discussion of how already the identification model alone had a significant impact.

Right now the most visible influences are new approximation problems (like approximation of output statistics [29] or entropy approximations based on Schur–convexity [27] etc.), a new emphasis on random number generation [47] and, above all, an understanding of the concept of common randomness [17], in identification ([27], [35], [48]), cryptography [46], and classical transmission problems of arbitrarily varying channels ([45], [41], [43]), and the paper [42], with a novel capacity formula, which could not be derived before.

It is also fascinating to discover how transmission problems and identification problems in multi–user theory show often some kind of duality. Often identification problems are mathematically more complex and in other cases we encounter the opposite: there is a rather *complete* capacity theory for identification via multi–way channels in case of complete feedback ([27]), whereas for transmission with feedback we don't even understand the multiple access channel.

We conclude with three more recently encountered directions of research.

18 Comparison of identification rate and common randomness capacity: Identification rate can exceed common randomness capacity and vice versa

One of the observations of [17] was that random experiments, to whom the communicators have access, essentially influence the value of the identification capacity C_I. We introduce now **common randomness capacity**, which was called mystery number in [27], and has subsequently been called by us in lectures and papers by its present name.

The common randomness capacity C_R is the maximal number ν such, that for a constant $c > 0$ and for all $\epsilon > 0$, $\delta > 0$ and for all n sufficiently large there exists a permissible pair (K, L) of random variables for length n on a set $\mathcal{K}$ with $|\mathcal{K}| < e^{cn}$ with

$$Pr\{K \neq L\} < \epsilon \text{ and } \frac{H(K)}{n} > \nu - \delta.$$

Actually, if sender and receiver have a common random capacity C_R then by the so called $\sqrt{n}$–trick of [17], that is, the Transformator Lemma (discussed in [58]), always

$$C_I \geq C_R \ \text{ if } \ C_I > 0. \tag{18.1}$$

For many channels (see [17], [46]), in particular for channels with feedback ([17], [27]), equality has been proved.

It seemed therefore plausible, that this is always the case, and that the theory of identification is basically understood, when common random capacities are known.

We report here a result, which shows that this expected unification is not valid in general — **there remain two theories**.

Example 6: $C_I = 1, C_R = 0$. (Fundamental)

(Actually, in [56] one can find also an example with $0 < C_I < C_R$)

We use a Gilbert type construction of error correcting codes with constant weight words. This was done for certain parameters in [16]. The same arguments give for parameters needed here the following auxiliary result.

Proposition 5. Let $\mathcal{Z}$ be a finite set and let $\lambda \in (0, 1/2)$ be given.
For $(2^{3/\lambda})^{-1} < \varepsilon < (2^{2/\lambda} + 1)^{-1}$ a family $A_1, \dots, A_N$ of subsets of $\mathcal{Z}$ exists with the properties

$$|A_i| = \varepsilon|\mathcal{Z}|, |A_i \cap A_j| < \lambda\varepsilon|\mathcal{Z}| \ (i \neq j)$$

and

$$N \geq |\mathcal{Z}|^{-1} 2^{\lfloor \varepsilon|\mathcal{Z}| \rfloor} - 1.$$

Notice that $\lambda \log\left(\frac{1}{\varepsilon} - 1\right) > 2$ and that for ℓ with $2^{-\ell} = \varepsilon$ necessarily $\ell > \frac{2}{\lambda}$.

Choose now $\mathcal{Z} = \{0,1\}^n$, $\varepsilon = 2^{-\ell}$ and A_i's as in the Proposition. Thus $|A_i| = 2^{n-\ell}$, $N(n, \lambda) = 2^{-n}\ 2^{2^{n-\ell}} - 1$ and $|A_i \cap A_j| < \lambda 2^{n-\ell}$.

Consider now a discrete channel $(W^n)_{n=1}^{\infty}$, where the input alphabets $\mathcal{X}_t = \{1, 2, \dots, N(t, \lambda)\}$ are increasing, $\mathcal{X}^n = \prod\limits_{t=1}^{n} \mathcal{X}_t$ are the input words of length n, $\mathcal{Y}^n = \{0,1\}^n$ are the output words and $W^n : \mathcal{X}^n \rightsquigarrow \mathcal{Y}^n$ is defined by

$$W^n(\cdot|i_1 i_2 \dots i_n) = W^n(\cdot|i_n)$$

and $W^n(\cdot|i)$ is the uniform distribution on A_i for $1 \leq i \leq N(n, \lambda)$.

By Proposition 5 and $3/\lambda > \ell > 2/\lambda$

$$N(n, \lambda) \geq 2^{-n} 2^{2^{n-3/\lambda}}$$

and

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

However, for transmission every decoding set is contained in some A_i and for error probability λ must have cardinality $(1-\lambda)|A_i| = (1-\lambda)2^{n-\ell}$.

Therefore $M(n, \lambda) \leq \frac{2^n}{(1-\lambda)2^{n-\ell}} \leq 2^{\ell+1}$, if $\lambda < 1/2$, and $\frac{1}{n} \log M(n, \lambda) \leq \frac{\ell+1}{n} \leq \frac{3/\lambda+1}{n} \to 0$ $(n \to \infty)$. **The transmission capacity is** 0. Consequently also $C_R = 0$.

Remarks:

25. The case of bounded input alphabets remains to be analysed. What are "natural" candidates for equality of C_I and C_R?
26. For infinite alphabets one should work out conditions for finiteness of the identification capacity.

19 Robustness, Common Randomness and Identification

It is understood now ([46], [42]) how the theory of AV–channels is *intimately* related to the concept of robust common randomness. A key tool is the balanced hypergraph coloring ([8]). We sketch now another direction concerning robustness and identification.

For more robust channel models, for instance in jamming situations, where the jammer knows the word to be sent (c.f. AV–channels with maximal error criterion), the communicators are forced to use the maximal error concept. In case of identification this makes the randomisation in the encoding (see [16]) superfluous. Now, for a DMC W it was mentioned in [16] that in the absence of randomisation the identification capacity, say $C_I^*(W)$, equals the logarithm of the number of different row–vectors in W. This is easy to show, however, a formidable problem arises if the DMC W is replaced by the AVC $\mathcal{W}$. In fact, for 0–1–matrices only in $\mathcal{W}$ we are — exactly as for transmission — led to the equivalent Shannon–zero–capacity problem. But for general $\mathcal{W}$ the identification problem is quite different from the transmission problem.

In so far there is a lower bound on $C_I^*(\mathcal{W})$, which implies for

$$\mathcal{W} = \left\{ \left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right), \left(\begin{smallmatrix} 1 & 0 \\ \delta & 1-\delta \end{smallmatrix}\right) \right\}, \delta \in (0,1)$$

that $C_I^*(\mathcal{W}) = 1$, which is obviously tight. It exceeds the known capacity for transmission. The capacity for $\mathcal{W} = \left\{ \left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right), \left(\begin{smallmatrix} 1-\delta & \delta \\ \delta & 1-\delta \end{smallmatrix}\right) \right\}$ is unknown.

20 Beyond Information Theory: Identification as a new Concept of Solution for Probabilistic Algorithms

Finally we mention as the perhaps most promising direction the study of probabilistic algorithms with identification as *concept of solution.* (For example: for any i, is there a root of a polynomial in interval i or not?)

The algorithm should be fast and have small error probabilities. Every algorithmic problem can be thus considered. This goes far beyond Information

Theory. Of course, like in general information transfer also here a more general set of questions can be considered. As usual in complexity theory one may try to classify problems.

What rich treasures do we have in the much wider areas of information transfer?!

References

[1] C.E. Shannon, A mathematical theory of communication, Bell Syst. Techn. J. 27, 339–425, 623–656, 1948.

[2] C.E. Shannon, Two–way communication channels, Proc. 4th Berkeley Symp. Math. Statist. and Prob., Univ. of California Press, Berkeley, Vol. 1, 611–644, 1961.

[3] R. Ahlswede, Multi–way communication channels, Proc. 2nd Int. Symposium on Inf. Theory, Tsahkadsor Armenian SSR, 1971. Akadémiai Kiadó, Budapest, 23–52, 1973.

[4] T.M. Cover, Broadcast channels, IEEE Trans. Inform. Theory, Vol. 18, 2–14, 1972.

[5] T.M. Cover, An achievable rate region for the broadcast channel, IEEE Trans. Inform. Theory, Vol. 21, 399–401, 1975.

[6] R. Ahlswede, Channel capacities for list codes, J. Appl. Prob. 10, 824–836, 1973.

[7] D. Slepian and J.K. Wolf, Noiseless coding of correlated information sources, IEEE Trans. Inform. Theory, Vol. 19, 471–480, 1973.

[8] R. Ahlswede, Coloring hypergraphs: A new approach to multi–user source coding, Part I, J. Comb., Inf. and System Sciences, Vol. 1, 76–115, 1979, Part II, Vol. 5, No. 3, 220–268, 1980.

[9] A.C. Yao, Some complexity questions related to distributive computing, Proc. 11th Annual Symposium on Theory of Computing, Atlanta, 209–213, 1979.

[10] R. Ahlswede and I. Csiszár, Hypothesis testing under communication constraints, IEEE Trans. Inform. Theory, Vol. 32, No. 4, 533–543, 1986.

[11] R. Ahlswede and G. Dueck, Good codes can be produced by a few permutations, IEEE Trans. Inform. Theory, Vol. 28, No. 3, 430–443, 1982.

[12] R. Ahlswede and T.S. Han, On source coding with side information via a multiple–access channel and related problems, IEEE Trans. Inform. Theory, Vol. IT 29, No. 3, 396–412, 1983.

[13] J. Já Já, Identification is easier than decoding, Preprint.

[14] I. Csiszár and J. Körner, Information Theory: Coding Theorems for Discrete Memoryless Systems, Academic Press, N.Y. 1981.

[15] R. Ahlswede, Eight problems in information theory in Open Problems in Communication and Computation, T.M. Cover and B. Gopinath Editors, Springer Verlag, 1987.

[16] R. Ahlswede and G. Dueck, Identification via channels, IEEE Trans. Inform. Theory, Vol. 35, 15–29, 1989.

[17] R. Ahlswede and G. Dueck, Identification in the presence of feedback — a discovery of new capacity formulas, IEEE Trans. Inform. Theory, Vol. 35, 30–39, 1989.

[18] M. Lothaire, Combinatorics on words, Encycl. of Math. and its Applications, Vol. 17, 1982.

[19] E. Mittenecker and E. Raab, Informationstheorie für Psychologen, Verlag für Psychologie, Dr. C. Hofgrefe, Göttingen 1973.

[20] P. Stucki, Advances in digital image processing, Theory, Application, Implentation, The IBM Res. Symp. Series, 256-302, 1984.

[21] J. Von Neumann, The Computer and the Brain, Yale University Press, 1958.

[22] R. Ahlswede and P. Gács, Two contributions to information theory, Colloquia Mathematica Societatis János Bolyai, 16. Topics in Information Theory, I. Csiszár and P. Elias Edit., Keszthely, Hungaria, 17–40, 1975.

[23] C. Cherry, On human communication, A Review, a survey and a criticism, MIT–Press 1957, 1966.

[24] R.G. Gallager, A perspective on multi–access channels, IEEE Trans. Inform. Theory, Vol. 31, No. 2, March 1985.

[25] T. Berger, Rate–distortion Theory, Prentice–Hall, Inc. Englewood Cliffs, N.J., 1971.

[26] D. Coppersmith, Private Communication in 1987.

[27] R. Ahlswede and B. Verboven, On identification via multi–way channels with feedback, IEEE Trans. Inform. Theory, Vol. 37, No. 5, 1519–1526, 1991.

[28] T.S. Han and S. Verdú, New results in the theory and application of identification via channels, IEEE Trans. Inform. Theory, Vol. 38, 14–25, 1992.

[29] T.S. Han and S. Verdú, Approximation theory of output statistics, IEEE Trans. Inform. Theory, Vol. 39, No. 3, 1993.

[30] J. Körner and K. Marton, General broadcast channels with degraded message sets, IEEE Trans. Inform. Theory, Vol. 23, 60–64, 1977.

[31] E.C. van der Meulen, Random coding theorems for the general discrete memoryless broadcast channel, IEEE Trans. Inform. Theory, IT 21, 180–190, 1975.

[32] M.S. Pinsker, Capacity region of noiseless broadcast channels, (in Russian), Problemi Peredachii Informatsii 14, No 2, 28–32, 1978.

[33] B. Verboven and E.C. van der Meulen, Capacity bounds for identification via broadcast channels that are optimal for the deterministic broadcast channel, IEEE Trans. Inform. Theory, Vol. 36, No 6, 1197–1205, 1990. IEEE Workshop on Inf. Theory, Salvador, Brazil, June 1992.

[34] A. Rényi, On the foundations of information theory, Rev. Inst. Internat. Stat. 33, 1–14, 1965.

[35] R. Ahlswede and Z. Zhang, New directions in the theory of identification via channels, IEEE Trans. Inform. Theory, Vol. 41, No. 4, 1040–1050, 1995.

[36] A.D. Wyner, The wire–tap channel, Bell Syst. Tech. J., Vol. 54, 1355–1387, 1975.

[37] R. Ahlswede and N. Cai, Information and control: the matching channel, IEEE Trans. Inform. Theory, Vol. 44, No. 2, 542-563, 1998.

[38] R. Ahlswede, J.P. Ye, and Z. Zhang, Creating order in sequence spaces with simple machines, Information and Computation, Vol. 89, No. 1, 47–94, 1990.

[39] R. Ahlswede, E. Yang, and Z. Zhang, Identification via compressed data, IEEE Trans. Inform. Theory, Vol. 43, No. 1, 48–70, 1997.

[40] J. Singh, Great Ideas in Information Theory, Language and Cybernetics, Dover Publication, Inc. New York, 1966.

[41] R. Ahlswede and N. Cai, Arbitrarily varying multiple–access channels, Part I: Ericson's symmetrizability is adequate, Gubner's conjecture is true, Part II: Correlated sender's side information, correlated messages and ambiguous transmission, IEEE Trans. Inform. Theory, Vol. 45, No. 2, 742-749, 749-756, 1999.

[42] R. Ahlswede and N. Cai, The AVC with noiseless feedback and maximal error probability: a capacity formula with a trichotomy, Numbers, Information and Complexity, Special volume in honour of R. Ahlswede on occasion of his 60th birthday, editors I. Althöfer, N. Cai, G. Dueck, L.H. Khachatrian, M. Pinsker, A. Sárközy, I. Wegener, and Z. Zhang, Kluwer Acad. Publ., Boston, Dordrecht, London, 151–176, 2000.

[43] R. Ahlswede, General theory of information transfer, Preprint 97-118, SFB 343 Diskrete Strukturen in der Mathematik, Universität Bielefeld 1997.

[44] C.C. Campbell, Definition of entropy by means of a coding problem, Z. Wahrscheinlichkeitstheorie u. verw. Geb., 113–119, 1966.

[45] R. Ahlswede, B. Balkenhol and C. Kleinewächter, Identification for sources, General Theory of Information Transfer and Combinatorics, Lecture Notes in Computer Science, Vol. 4123, Springer Verlag, 51-61, 2006.

[46] R. Ahlswede and I. Csiszár, Common randomness in information theory and cryptography, Part I: Secret sharing, IEEE Trans. Inform. Theory, Vol. 39, No. 4, 1121–1132; 1993; Part II: CR capacity, IEEE Trans. Inform. Theory, Vol. 44, No. 1, 55–62, 1998.

[47] R. Ahlswede, The capacity region of a channel with two senders and two receivers, Ann. Probability, Vol. 2, No. 5, 805–814, 1974.

[48] T.S. Han, Information-spectrum Methods in Information Theory, Applications of Mathematics (New York), 50, Stochastic Modelling and Applied Probability, Springer, 2003.

[49] R. Ahlswede and V. Balakirsky, Identification under random processes,

Preprint 95–098, SFB 343 Diskrete Strukturen in der Mathematik, Universität Bielefeld, Problemy peredachii informatsii (special issue devoted to M.S. Pinsker), vol. 32, no. 1, 144–160, Jan.–March 1996.

[50] W.H. Kautz and R.C. Singleton, Nonrandom binary superimposed codes, IEEE Trans. Inform. Theory, Vol. 10, 363–377, 1964.

[51] A.G. Dyachkov and V.V. Rykov, Bounds on the length of disjunctive codes, Problemy Peredachi Informatsii 18, No. 3, 7–13 (in Russian), 1982.

[52] P. Erdös, P. Frankl, and Z. Füredi, "Families of finite sets in which no set is covered by the union of r others", Israel J. Math. 51, 79–89, 1985.

[53] M. Ruszinko, On the upper bound of the size of the r–cover–free families, J. Combinatorial Theory, Ser. A 66, 302–310, 1994.

[54] Z. Füredi, "On r–cover–free families", J. Combin. Theory Ser. A 73, No. 1, 172–173, 1996.

[55] N.N. Kuzjurin, On the difference between asymptotically good packings and coverings, European J. Combin. 16, 35–40, 1995.

[56] C. Kleinewächter, On identification, General Theory of Information Transfer and Combinatorics, Lecture Notes in Computer Science, Vol. 4123, Springer Verlag, 62-83, 2006.

[57] General Theory of Information Transfer and Combinatorics, Lecture Notes in Computer Science, Vol. 4123, Springer Verlag, R. Ahlswede et al. (Eds.), 2006.

[58] R. Ahlswede, Towards a General Theory of Information Transfer, Shannon Lecture at ISIT in Seattle 13th July 2006, IEEE Information Theory Society Newsletter, 2007.

[59] R. Ahlswede, N. Cai, and Z. Zhang, Erasure, list, and detection zero-error capacities for low noise and a relation to identification, IEEE Trans. Inform. Theory, Vol. 42, No. 1, 55-62, 1996.

[60] Y. Steinberg, New converses in the theory of identification via channels, IEEE Trans. Inform. Theory, Vol. 44, No. 3, 984-998, 1998.

[61] R. Ahlswede, Elimination of correlation in random codes for arbitrarily varying channels, Z. Wahrscheinlichkeitstheorie und verw. Geb. 44, 159-175, 1978.

[62] R. Ahlswede, On concepts of performance parameters for channels, General Theory of Information Transfer and Combinatorics, Lecture Notes in Computer Science, Vol. 4123, Springer Verlag, 639-663, 2006.

[63] R. Ahlswede and G. Dueck, Every bad code has a good subcode: a local converse to the coding theorem, Z. Wahrscheinlichkeitstheorie und verw. Geb. 34, 179-182, 1976.

[64] R. Ahlswede, C. Mauduit, and A. Sárközy, Large families of pseudorandom sequences of k symbols and their complexity, Part II, General Theory of Information Transfer and Combinatorics, Lecture Notes in Computer Science, Vol. 4123, Springer Verlag, 308-325, 2006.

[65] P. Ungar, The cut-off point for group testing, Commun. Pure Appl. Math. 13, 49-54, 1960.

GTIT-Supplement

Search and channels with feedback

I. Basic classical concepts

We say that a channel has noiseless feedback if the transmitter knows the received letter before sending the next letter.

We cannot at this point go into the importance of this channel model for noisy transmitting systems in communication situations, which occur above all in concrete communication technology, but also in physics and psychology. We can only mention that channels with feedback play a role, for example, in data transmission from satellites to fixed earth stations since the latter (at least so far) have enough energy to pretty much arrange for error-free retransmission to the satellites.

In [S10] Chapter IX it is shown that every sequential search problem with answers having chance errors is equivalent to an information-theoretical coding problem for channels with feedback. A first systematic treatment was made by considering the BSC, more generally the DMC, Gaussian channels, and the AVC.

Furthermore for every sequential "combinatorial" search problem with given number of errors or errors proportional to the search duration there exists an equivalent coding problem for channels with feedback in the combinatorial setting. This was already shown by Berlekamp in 1964 in [S19].

We address here first the DMC with transmission matrix $W : \mathcal{X} \to \mathcal{Y}$, named in short the DMC W.

For a search space $\mathcal{U}$ considered have been non-adaptive (automatically fixed block length) search strategies or (equivalently) coding procedures without feedback, and adaptive ones, which are further subdivided into those of fixed block length and those of variable length (covering as a special case the fixed block length).

For instance Shannon [S18] considered adaptive procedures for channels with feedback of fixed block length for the two cases of zero error probabilities and error probability approaching zero, respectively.

On the other hand his Noiseless Coding Theorem concerns the case of varying lengths for a noiseless forward channel. Incorporation as a special case of sequential procedures for the DMC was done much later in [22]. This work can also be found in [S10].

In case of variable lengths one has the following performance criteria for search times. Associated with procedure $\mathcal{C}$ is a stopping time $\ell(\mathcal{Y}^\infty)$, which depends not only on the object u searched for, but also on chance coming from the DMC. We denote by $E(\ell|\mathcal{C},u)$ the expected search time for u and by $M(\ell|\mathcal{C},u)$ the maximal search time for u. This leads to four concepts of search time for $\mathcal{C}$:

(a) $M(\ell|\mathcal{C}) = \max\limits_{u\in\mathcal{U}} M(\ell|\mathcal{C},u)$, the maximal (maximal) search time

(b) $E(\ell|\mathcal{C}) = \max\limits_{u\in\mathcal{U}} E(\ell|\mathcal{C},u)$, the maximal expected search time

and in the presence of an a-priori PD P on $\mathcal{U}$

(c) $M(\ell|\mathcal{C},P) = \sum\limits_{u\in\mathcal{U}} P_u M(\ell|\mathcal{C},u)$, the average (maximal) search time

(d) $E(\ell|\mathcal{C},P) = \sum\limits_{u\in\mathcal{U}} P_u E(\ell|\mathcal{C},u)$, the average expected search time.

Remark 1: In noiseless coding one has $M(\ell|\mathcal{C},u) = E(\ell|\mathcal{C},u)$ and the concepts in (a) and (b) are equal and so are the concepts in (c) and (d).

The remaining two cases constitute fixed block length and the standard varying length coding average search time.

For general DMC (a) gives the case of block-coding and we adopt (d) in the following for the sequential case.

II. Search and Identification: performance criteria

Classical search for the realized (for instance also defective) object $u \in \mathcal{U}$ and its connection to the DMC with **passive** noiseless feedback (in short: DMC with feedback) has just been classified. (Active feedback carries us out of the domain of Search Theory). Let us just mention that the sequential case has not been fully analysed for instance in the zero error case! In the ε-error case there is a strong converse for block coding, but not for sequential coding, but the weak capacities are equal.

We come now to the classification of search models with the aim of identification.

Again we consider only adaptive procedures.

The work started in [17] for **block coding**. For the DMC (second order) strong identification capacities were determined in [17]. If $C_{Sh} > 0$ then

$$C_f = \max_{x \in \mathcal{X}} H(W(\cdot|x)), \tag{2.1}$$

and here randomized procedures give a bigger capacity

$$C_F = \max_{P:Q=PW} H(Q). \tag{2.2}$$

This work was extended to a large class of multi-way channels even with a constructive coding scheme [27], again always block coding is used. So everywhere here we are in case (a). The sequential case, called noiseless source coding for identification, started in [S5], was continued in [45] and led to the identification entropy in [S6]. We give first the performance criteria used. **They all involve a PD P on the search space $\mathcal{U}$.** $L_{\mathcal{C}}(P,u)$, $L_{\mathcal{C}}(P)$, and $L(P)$ are defined in Section 16.

We also consider, if users are chosen by a RV V independent of U and defined by $\text{Prob}(V = v) = Q_v$ for $v \in \mathcal{V} = \mathcal{U}$,

$$L_{\mathcal{C}}(P,Q) = \sum_{v \in \mathcal{U}} Q_v L_{\mathcal{C}}(P,v) \tag{2.3}$$

the average number of expected checkings, if code $\mathcal{C}$ is used, and also

$$L(P,Q) = \min_{\mathcal{C}} L_{\mathcal{C}}(P,Q) \tag{2.4}$$

the average number of expected checkings for a best code.

A natural special case is the mean number of expected checkings

$$\bar{L}_{\mathcal{C}}(P) = \sum_{u=1}^{N} \frac{1}{N} L_{\mathcal{C}}(P,u), \tag{2.5}$$

which equals $L_{\mathcal{C}}(P,Q)$ for $Q = \left(\frac{1}{N}, \dots, \frac{1}{N}\right)$, and

$$\bar{L}(P) = \min_{\mathcal{C}} \bar{L}_{\mathcal{C}}(P). \tag{2.6}$$

Already in [45] it was proved that $L(P) < 3$ for every $P = (P_1, \dots, P_N)$!

All other quantities defined are even smaller. For their analysis we refer to [S6] and the recent work [S8].

III. Search and Identification: completing the landscape

In contrast to the classical theory described above, where search time always goes with the order $\log |\mathcal{U}|$, in the two situations described in II, it goes in block coding with the order $\log \log |\mathcal{U}|$ and in noiseless identification source coding, which is sequential, with the even smaller constant order.

This difference in behaviour is to be understood. We shall gain clarity by answering two questions.

1. Does sequentiality cause a big difference?
   Notice that already classical sequentiality does help somewhat. Eventhough it does not increase capacity it makes the strong converse fail.
2. What is the effect of the presence of an a-priori distribution P on $\mathcal{U}$?

We begin with the first question.

Ad1 We reconsider the two identification capacities C_f and C_F in [16] for a DMC with feedback and fixed blocklength procedures. For blocklength n we introduced $N_f(n,\lambda)$, the maximal number of messages for which identification over the DMC W is possible in blocklength n with both type of error probilities bounded by λ. Correspondingly $N_F(n,\lambda)$ stands for the analogously defined quantity, if randomization in the coding procedure is permitted.

Then it was proved that for $0 < \lambda < \frac{1}{2}$

$$\lim_{n\to\infty} \frac{1}{n} \log\log N_f(n,\lambda) = \max_{x\in\mathcal{X}} H(W(\cdot|x))$$
$$\lim_{n\to\infty} \frac{1}{n} \log\log N_F(n,\lambda) = \max_{P:Q=PW} H(Q)$$

(Coding theorem and strong converse).

We define now for cases (a) - (d) the corresponding functions

$$N_f^a(n,\lambda), N_F^a(n,\lambda), \dots, N_F^d(n,\lambda).$$

Here $N_f^a(n,\lambda)$ means that the maximisation is over coding or search procedures $(\mathcal{C},\ell)$ with

$$M(\ell|\mathcal{C}) \leq n. \tag{3.1}$$

Thus obviously

$$N_f^a(n,\lambda) = N_f(n,\lambda), N_F^a(n,\lambda) = N_F(n,\lambda).$$

Presently we deal with $N_f^b(n,\lambda)$ and $N_F^b(n,\lambda)$ where (3.1) is replaced by

$$E(\ell|\mathcal{C}) \le n. \tag{3.2}$$

Fortunately we found a satisfactory answer to question 1.

Theorem 1 (Weak converses).
For a DMC W

(i) $\inf\limits_{\lambda} \overline{\lim\limits_{n\to\infty}} \frac{1}{n} \log\log N_f^b(n,\lambda) \le \max\limits_{x\in\mathcal{X}} H(W(\cdot|x)) = C_f(W)$

(ii) $\inf\limits_{\lambda} \overline{\lim\limits_{n\to\infty}} \frac{1}{n} \log\log N_F^b(n,\lambda) \le \max\limits_{P:Q=PW} H(Q) = C_F(W)$

(Since $N_f^b(n,\lambda) \ge N_f(n,\lambda)$ and $N_F^b(n,\lambda) \ge N_F(n,\lambda)$ the capacities don't change if sequentiality is permitted).

Proof: We use our general Entropy-Set Size Relation (stated as Lemma 2 in in Section 3.4).
For the set of PD's $\mathcal{P}(\mathbb{N})$ on the set of positive integers and all $d \ge 1$

$$\min_{P=(P_1,P_2,\dots,)\in\mathcal{P}(\mathbb{N})} \left\{ \max \sum_{j\in J} P_j : J \subset \mathbb{N}, |J| = \left\lceil 2^{H(P)d} \right\rceil + 1 \right\} = 1 - \frac{1}{d}. \tag{3.3}$$

(i) Consider an (n,N,λ) IDF-code $\{(f_i,\mathcal{D}_i,\ell_i) : 1 \le i \le N\}$.

f_i generates a RV $Y_i^{\ell_i}$ with distribution $\Pr(Y_i^{\ell_i} = y^t) = W^t(y^t|f_i,\ell_i = t)$. Now

$$H(Y_i^{\ell_i}) \le \max_{x\in\mathcal{X}} H(W(\cdot|x)) \cdot E\ell_i. \tag{3.4}$$

Application of (3.3) to the distribution of $Y_i^{\ell_i}$ gives a set $\mathcal{E}_i \subset \mathcal{Y}^* = \bigcup\limits_{t=1}^{\infty} \mathcal{Y}^t$ with

$$\Pr(Y_i^{\ell_i} \in \mathcal{E}_i) \ge 1 - \frac{1}{d}$$

and

$$|\mathcal{E}_i| \le 2^{d\,H(Y_i^{\ell_i})} \le 2^{d\,E\ell_i \max\limits_{x\in\mathcal{X}} H(W(\cdot|x)))} \triangleq K, \text{ say, and } K \le 2^{dC_f n}.$$

For any integer s, since $E\ell_i \le n$, by Chebyshev's inequality for the set $S_i = \{\ell_i \ge ns\}$

$$\Pr(S_i) \le \frac{1}{s}. \tag{3.5}$$

Define now

$$\mathcal{D}_i^* = (\mathcal{D}_i \cap \mathcal{E}_i) \smallsetminus S_i.$$

Then

$$\Pr(Y_i^{\ell_i} \in \mathcal{D}_i^*) \ge 1 - \lambda - \frac{1}{d} - \frac{1}{s}.$$

Under the assumption $\lambda < 1 - \lambda - \frac{1}{d} - \frac{1}{s}$ for d, s sufficiently large these $\mathcal{D}_i^*$ are necessarily **distinct**. We get therefore (in the wonderful world of double exponentiality)

$$N \leq \sum_{t=1}^{ns} \sum_{k=1}^{K} \binom{|\mathcal{Y}^t|}{k} \leq ns \cdot K |\mathcal{Y}|^{ns \cdot K} \leq nsK |\mathcal{Y}|^{ns \cdot 2^{dC_f n}}$$

for every $d > 1$ and (i) follows.

(ii) Just observe that in case of randomized strategies

$$H(Y_i^{\ell_i}) \leq \max_{P:Q=PW} H(Q) \cdot E\ell_i. \tag{3.6}$$

This replaces now (3.4) in the foregoing argument, which otherwise goes through literally.

Remark 2: Since

$$H(Y^\ell) \leq \max H(Q) \cdot E\ell \tag{3.7}$$

can we give a strong converse proof like in [17]?

If so, then we have an example where a strong converse holds for identification (seemingly also for common randomness), but not for transmission.

Now we present the surprising results, that even for sequential procedures the strong converses hold for identification over a DMC with noiseless feedback:

Recall that for transmission the strong converse does not hold in the sequential case and we now have the first case, where passing from transmission to identification leads to stronger statements (Han/Verdú [28] in proving a strong converse for identification assumed it to hold for transmission for a channel in question!).

Lemma 1. (Image size for a sequential deterministic feedback strategy)

For any feedback strategy (f, ℓ) of expected length $E\ell \leq n$, where $\ell : \mathcal{Y}^* \to \mathbb{N}$ is the stopping function for f,

$$\min_{\mathcal{E} \subset \mathcal{Y}^* : W(\mathcal{E}|f) \geq 1-\nu} |\mathcal{E}| \leq K = 2^{nH(W(\cdot|x^*)) + \alpha\sqrt{n}} \tag{3.8}$$

where

$$H(W(\cdot|x^*)) = \max_{x \in \mathcal{X}} H(W(\cdot|x)), \ \alpha = \sqrt{\frac{\beta}{\nu}}, \tag{3.9}$$

and $\beta = \max\{\log^2 3, \log^2 |\mathcal{Y}|\}$.

Proof of Lemma 1: The cardinality of the set

$$\mathcal{E}_f^* = \{y^n \in \mathcal{Y}^* : -\log W^n(y^n|f) \leq \log K\} \cap \mathcal{L}(f) = \{y^n \in \mathcal{Y}^* : \ell(y^n) = \text{stop}\}$$

$\left(W^n(y^n|f) \geq \frac{1}{K}\right)$ is clearly smaller than K, and it suffices to show that

$$W^n(\mathcal{E}_f^*|f) \geq 1 - \nu.$$

For this we first give another description of $W^n(\mathcal{E}_f^*|f)$.

Strategy f induces the RV's $Y^s = (Y_1, \dots, Y_s)$; $s = 1, 2, \dots, \ell$; with distribution

$$\Pr(Y^s = y^s) = W^s(y^s|f) \text{ for } y^s \in \mathcal{Y}^s \cap \{\kappa^s : \ell(\kappa^s) \geq s\}.$$

Defining $Z_t = -\log W(Y_t|f(Y^{t-1}))$ we can write

$$W(\mathcal{E}_f^*|f) = \Pr\left(\sum_{t=1}^{\ell} Z_t \leq \log K\right). \tag{3.10}$$

We now analyse this expression by considering the conditional expectations $\mathbb{E}(Z_t|Y^{t-1})$.

Since $\Pr(Y_t = y_t|Y^{t-1} = y^{t-1}) = W(y_t|f(y^{t-1}))$, we have for $y^{t-1} \in \mathcal{Y}^{t-1}$

$$\mathbb{E}(Z_t|y^{t-1}) = -\sum_{y_t \in \mathcal{Y}} W(y_t|f(y^{t-1}) \log W(y_t|f(y^{t-1})) \leq H(W(\cdot|x^*))$$

and therefore

$$\mathbb{E}(Z_t|y^{t-1}) \leq H(W(\cdot|x^*)). \tag{3.11}$$

Finally, we introduce the RV's

$$U_t = Z_t - \mathbb{E}(Z_t|Y^{t-1}) \tag{3.12}$$

which obviously satisfy

$$\mathbb{E}(U_t|Y^{t-1}) = 0,\ \mathbb{E}U_t = 0. \tag{3.13}$$

Moreover, since U_s is a function of $Y_1, \dots, Y_s$, this implies for $s < t$

$$\mathbb{E}(U_t|U_s) = 0.$$

Therefore, the RV's $U_1, \dots, U_\ell$ are uncorrelated, i.e.

$$\mathbb{E}\ U_s U_t = 0 \text{ for } s \neq t. \tag{3.14}$$

Notice that (3.10)-(3.13) and the definition of K imply

$$W^\ell(\mathcal{E}_f^*|f) \geq \Pr\left(\sum_{t=1}^{\ell} U_t \leq \alpha\sqrt{n}\right) \tag{3.15}$$

and by Chebyshev's inequality

$$\Pr\left(\sum_{t=1}^{\ell} U_t \le \alpha\sqrt{n}\right) \ge 1-\nu$$

provided that

$$\text{var } U_t \le \beta \ \text{ for } \ t = 1, 2, \dots, \ell. \tag{3.16}$$

Verification of (3.16) will complete the proof. Using (3.13) we can write

$$\text{var } U_t = \mathbb{E}\ U_t^2 = \mathbb{E}(U_t - \mathbb{E}(U_t|Y^{t-1}))^2$$

$$= \sum_{y^t} \Pr(Y^{t-1} = y^{t-1}) \cdot \mathbb{E}(U_t - \mathbb{E}(U_t|Y^{t-1})^2|Y^{t-1} = y^{t-1})$$

and by the well-known minimality property of the expected value this can be upper bounded by

$$\sum_{y^{t-1}} \Pr(Y^{t-1} = y^{t-1})\mathbb{E}(U_t - \mathbb{E}(Z_t|Y^{t-1}))^2|Y^{t-1} = y^{t-1})$$

$$= \sum_{y^{t-1}} \Pr(Y^{t-1} = y^{t-1})\mathbb{E}(Z_t^2|Y^{t-1} = y^{t-1}).$$

By the definition of Z_t

$$\mathbb{E}(Z_t^2|Y^{t-1} = y^{t-1}) = \sum_{y_t \in \mathcal{Y}} W(y_t|f(y^{t-1})) \cdot \log^2(W(y_t|f(y^{t-1})).$$

Since $x \log^2 x$ is bounded in $[0,1]$, this quantity is bounded by a function of $|\mathcal{Y}|$ uniformly in t and y^{t-1}.

A Lagrange multiplier argument gives the bound

$$\beta = \max\{\log^2 3, \log^2 |\mathcal{Y}|\}.$$

Thus,

$$\text{var}(U_t) \le \beta. \tag{3.17}$$

Lemma 2. (Image size for a sequential randomized feedback strategy)

For any randomized feedback strategy (F, ℓ) with $\mathbb{E}\ \ell \le n$ and any $\nu \in (0,1)$

$$\min_{\mathcal{E}' \subset \mathcal{Y}^* : W(\mathcal{E}'|F)|\mathcal{E}'| \ge 1-\nu} \le K' = 2^{nH(Q')+\alpha\sqrt{n}}$$

where $H(Q') = \max_P H(P \cdot W)$, $\alpha = \sqrt{\beta/\nu}$ and $\beta = \max\{\log^2 3, \log^2 |y|\}$.

Proof: The randomized strategy F can be viewed as a PD Q_F on the set $\mathcal{F}_\ell$ of ℓ-length deterministic feedback strategies. Therefore,

$$W^\ell(\mathcal{E}'|F) = \sum_{g\in\mathcal{F}_\ell} Q_F(g) W^\ell(\mathcal{E}'|g). \tag{3.18}$$

Q_F induces the RV Y^ℓ with distribution

$$\Pr(Y^\ell = y^\ell) = \sum_{g\in\mathcal{F}_\ell} Q_F(g) W^\ell(y^\ell|g).$$

We write $Q(y^\ell) = \Pr(Y^\ell = y^\ell)$. The cardinality of the set

$$\mathcal{E}'^* = \{y^\ell : -\log(y^\ell) \le \log K'\}$$

is clearly smaller than K', and it suffices to show now that $Q(\mathcal{E}'^*) \ge 1-\nu$. Defining $Z'_t = -\log Q(Y_t|y^{t-1})$, we can write

$$Q(\mathcal{E}'^*) = \Pr\left(\sum_{t=1}^{\ell} Z'_t \le \log K'\right). \tag{3.19}$$

For its analysis we consider now $\mathbb{E}(Z'_t|Y^{t-1})$. Notice that

$$\mathbb{E}(Z'_t|y^{t-1}) = -\sum_{y_t\in\mathcal{Y}} Q(y_t|y^{t-1}) \log(y_t|y^{t-1})$$

and that $Q(\cdot|y^{t-1})$ is a distribution of the form PW, because

$$Q(y_t|y^{t-1}) = \sum_{g\in\mathcal{F}_\ell} Q_F(g) \frac{\prod_{i=1}^{t-1} W(y_i|g(y^{i-1}))}{\sum_s Q_F(g) \prod_{i=1}^{t-1} W(y_t|g(y^{t-1}))} \cdot W(y_t|g(y^{t-1}))$$

Therefore we have

$$\mathbb{E}(Z'_t|y^{t-1}) \le H(Q'). \tag{3.20}$$

This is the substitute for (3.11). Otherwise we continue exactly as before. We define functions

$$U'_t = Z'_t - \mathbb{E}(Z'_t|Y^{t-1})$$

which again have the desired properties $\mathbb{E}\ Y'_t = 0$, $\mathbb{E}\ U'_t U'_s = 0$ for $s \ne t$ and var $U'_t \le \beta$. Application of Chebyshev's inequality again establishes the result.

Ad2 Having understood that sequentiality **alone** does not help much, the superperformance in source coding ($L(P) < 3$) must depend on the conjunction with source distributions. It enters already **the individual waiting times** $L_{\mathcal{C}}(P,u)$ – without it the source identification problem is not even defined! In the model just discussed the channel has noise and we allow error probability. But those two ingredients alone should not be responsible for different behaviour – as was demonstrated in the classical situation in [17]. We check this idea now by allowing a message distribution in identification (even without feedback). We thus come back to a simple model **already rejected** in [27].

Use the uniform distribution $\left(\frac{1}{N},\dots,\frac{1}{N}\right)$ on the message set and consider (for simplicity) $\mathcal{X}=\mathcal{Y}=\{0,1,\dots,q-1\}$ and as DMC the identity matrix, $W(x|x)=1$ for all $x\in\mathcal{X}$. Let $N=q^m$ and partition $\mathcal{U}=\{1,\dots,N\}$ into $\mathcal{U}_0,\dots,\mathcal{U}_{q-1}$, each of cardinality q^{m-1}, the messages in $\mathcal{U}_j$ are all encoded by j and "decoded" by $\mathcal{D}_i=\mathcal{U}_j$, if $i\in\mathcal{U}_j$.

So i is thus never rejected in identification, if it is there. The error probability of second type – accepted if not there – is

$$\frac{q^{m-1}-1}{q^m}\le\frac{1}{q}\ \text{ for all }\ m. \tag{3.21}$$

Thus $\frac{1}{q}$ can be achieved for arbitrarily large N! Partitioning $\mathcal{U}$ into q^r sets of equal cardinality one gets

$$\frac{q^{m-r}-1}{q^m}\le\frac{1}{q^r}\ \text{ for all }\ m\ge r. \tag{3.22}$$

This already smells like "$L(P)<3$". To complete the picture we have to use the feedback to make the procedure sequential.

Start like in (3.21). In case of a rejection stop. Otherwise iterate with the set $\mathcal{U}_i$ if the identification goes for i, etc. But this is exactly what we did in the proof of Theorem 1 in [S6].

The expected waiting time $T_{\mathcal{C}}(i)$ equals

$$1+\frac{1}{q}+\frac{1}{q^2}+\dots+\frac{1}{q^{m-1}}=\frac{\left(\frac{1}{q}\right)^m-1}{\frac{1}{q}-1}=\frac{1-\frac{1}{q^m}}{1-\frac{1}{q}}\le\frac{1}{1-\frac{1}{q}}=\frac{q}{q-1}$$

and $\lim\limits_{m\to\infty}\left(1+\frac{1}{q}+\dots+\frac{1}{q^{m-1}}\right)=\frac{q}{q-1}$.

Also

$$\left(1+\frac{1}{q}+\dots+\frac{1}{q^{m-1}}\right)=H_{I,q}(P^N),N=q^m. \tag{3.23}$$

By Theorem 4 of [S6] we know that this is best possible.

Remark 3: By Theorem 1 of [S8] the answer in (3.23) is obtained for all PD's $P^N \in \mathcal{P}([N])$, so it is universal in this sense, and $\max\limits_{P^N \in \mathcal{P}([N])} H_{I,q}(P^N) = \frac{q}{q-1}\left(1 - \frac{1}{N}\right)$.

Remark 4: Caution is in order with respect to the PD's involved. The P in the source $(\mathcal{U}, P)$ is responsible for getting small values like in $L_{\mathcal{C}}(P,u)$ – the competition of u **against the randomly chosen object**, whereas in [16], [27] the competition is for all pairs (i,j).

Another PD Q enters in $L_{\mathcal{C}}(P,Q)$ – it serves only on averaging. In the analysis above with the uniform distribution P Q was not relevant, because $L_{\mathcal{C}}(P,u)$ was independent of u.

Finally we propose the following seemingly essential remaining tasks.

Problem 1: Extend the $L(P) < 3$ result to **noisy channels**.

Problem 2: Analyse the quantity $L(P,Q)$ for arbitrary P,Q and find the relevant entropy measure.

Problem 3: For the "pentagon" channel

$$W = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & 0 & 0 & 0 & \frac{1}{2} \end{pmatrix}$$

m' steps are needed until $\left\lceil \lceil N \cdot \frac{2}{5} \rceil \cdot \frac{2}{5} \right\rceil \cdots \leq 1$. So $m' \sim \left(\log_5 \frac{5}{2}\right)^{-1} \cdot \log_5 N$, expected waiting time for uniform distributions $\sim \frac{\frac{5}{2}}{\frac{5}{2}-1}\left(1 - \frac{1}{N}\right)$ and $\log_5 \frac{5}{2}$ is $C_{o\ell}$, the zero-error list code capacity for W.

But if $C_{o\ell}$ is not of the form $\log_q \frac{q}{r}$, what is the answer?

This is to be worked out and may not be trivial!

For instance for $W_\epsilon = \begin{pmatrix} 1-\varepsilon & \varepsilon \\ \varepsilon & 1-\varepsilon \end{pmatrix}$ we have $C_{o\ell} = o$. Is there a jump at $C_{o,\ell}$?

Problem 4:

One can use an (n, M, ε) transmission code of rate $(M) \sim 1-h(\varepsilon)$ for the DMC with transmission matrix W_ϵ and put $\frac{N}{M}$ objects in every class corresponding to a codeword. So now M corresponds to q and

$$1 + \frac{1}{M} + \cdots + \frac{1}{M^{m-1}} \leq \frac{M}{M-1}.$$

Choosing $M^m \geq N$ or $m \geq \frac{\log N}{\log M} \sim \frac{\log N}{n(1-h(\varepsilon))}$ the error probability is bounded by $\varepsilon + \varepsilon + \cdots + \varepsilon = m\varepsilon$.

Use the blocklength

$$n\frac{M}{M-1}\left(1 - \frac{1}{M^m}\right) \sim n$$

and analyse this model for identification.

IV. New search models

We start or contribute to the following topics

— search under secrecy constraints
— noisy response channels with random or arbitrary varying parameters
— source statistics described by a class of PD
— the $(\mathcal{Y}, \mathcal{N}, \mathcal{A})$ model for liars, a common generalization of Rényi's search model with size constrained tests and a recent liar model of Katona/Tichler [S17].

1. Search under secrecy constraints

We address here first wire-tap channels introduced by A. Wyner and report on [S9]. There transmission, identification and common randomness via a wire-tap channel with **secure feedback** are studied. By secure feedback we mean that the feedback is noiseless and that the wire-tapper has no knowledge about the content of the feedback except via his own output. Lower and upper bounds to the transmission capacity are derived. The two bounds are shown to coincide for two families of degraded wire-tap channels, including Wyner's original version of the wire-tap channel. The identification and common randomness capacities for the channels are completely determined. Also here again identification capacity is much bigger than common randomness capacity, because the common randomness used for the (secured) identification needs not to be secured!

Finally we mention that the adaptive schemes use constant blocklength and uniform message statistics.

Remark 5: Extension to general message distribution P with sequential procedures **seems to be doable**.

2. Search with random parameters for the response channel

S. Gelfand and M. Pinsker mention on page 20 of [S13] that the determination of the capacity is closely related to the study of broadcast channels. This relation was used by the authors to determine this capacity region of noiseless broadcast channels. On the other hand their direct coding theorem is analogous to the direct coding theorem of K. Marton for broadcast channels.

Our observation here is that feedback does not increase the capacity, thus the search problem is solved.

The same applies to our paper [S2], which treats arbitrarily varying parameters known to the sender.

Proof of converse:

Going along the proof in [S2] we mention the necessary changes.
Instead of

$$(\alpha)\ \Pr(V=m, S^n=\underline{s}, x^n=\underline{x}) = \begin{cases} M^{-1}P_{S^n}(\underline{s}), & \text{if } \underline{x}=\varphi(m,s) \\ 0, & \text{otherwise} \end{cases}$$

at the bottom of page 26, let

$$(\alpha')\ \Pr(V=m, S^n=s^n) = M^{-1}P_{S^n}(s^n)\ \Big(= M^{-1}\prod_{i=1}^{n} P_S(s_i)\Big),$$

$x_i=\varphi_i(m,Y^{i-1})$ and Y^i is the output of $W^i(\cdot|x_i,s_i)$, when given input $X_i=x_i$ and state $S_i=s_i$ and $U(i)$'s are still defined by

$$U(i)=(V,Y_1,\dots,Y_{i-1},S_{i+1},\dots,S_n)$$

(but for new X_i's and Y_i's).

Now

$$(\beta)\ P_{VS^nX^nY^n} = P_{VS^nX^n}\cdot P_{Y^n|X^nS^n}$$

does **not** hold necessarily, but it is actually **not** needed either.

We only need

$$(\beta')\ \Pr\Big(\big(U(i),S_i,X_i,Y_i\big)=(u,s_i,x_i,y_i)\Big) = \Pr\Big(\big(U(i),S_i,X_i\big)=(u,s_i,x_i)\Big)\Pr(Y_i=y_i|X_i=x_i,S_i=s_i),$$

which obviously holds, because in case $x_i \neq \varphi_i(m, y^{i-1})$ (see the definition of $U(i)$) both sides vanish and otherwise they are

$$\Pr\Big((U(i), S_i) = (u, s_i)\Big) W(y|x_i, s_i).$$

Finally, (4.2) still holds by Fano's inequality and (4.3) (Lemma 4) still holds, because the proof uses properties of mutual information without special reference to feedback.

Therefore also the two inequalities at the bottom of page 28 hold and give the converse (Proposition 3).

These channels with random parameters were robustified in [S2] to arbitrarily varying parameters. The foregoing arguments also imply that our capacity formula also holds in case of noiseless feedback and thus has its part in the Theory of Search. For (standard) AVC this was treated in detail in [S10], ch. IX. Robustness is also the issue in the next Section 3 and in the last Section 4 we shall see that the $(\mathcal{Y}, \mathcal{N}, \mathcal{A})$ model discussed there opens a new direction in the theory of AVC and zero-error capacity problems.

3. Noiseless coding for a class of PD's

We returned to this problem in the context of identification for sources. For classical block coding compound sources and AVS have been considered – but not or **less for abstract sources**.

How do we encode 2 sources P, Q both on $\mathcal{U}$?

Using the Min-Max Theorem we can proceed as follows. We use a PD r on $\mathcal{U}$ for encoding, that is, c_u gets length $\lceil -\log r_u \rceil$ and we try to minimize the worst case "expected length":

$$L = \min_r \max \left(\sum_u P_u \log \frac{1}{r_u}, \sum_u Q_u \log \frac{1}{r_u} \right) \tag{3.1}$$

and get that $\max\limits_{0 \leq \lambda \leq 1} H\Big(\lambda P + (1-\lambda) Q\Big)$ is best: $H \leq L \leq H + 1$.

Now it is clear how to handle a general class $\mathcal{P}$ of PD's and not just two PD's.

Let $\overline{\text{conv}(\mathcal{P})}$ be the closed convex hull of $\mathcal{P}$. Clearly

$$\max_{P \in \overline{\text{conv}(\mathcal{P})}} H(P) \geq \sup_{P \in \mathcal{P}} H(P).$$

Recall that the smaller quantity rules for compound sources and the larger quantity rules for AV-sources

— however, the two concepts are the same for abstract sources and the closed convex hull comes in.

Note that

$$\sup_{P\in\mathcal{P}} \min_{r} -\sum_{\kappa} P(x)\log r(x) = \max_{P\in\overline{\text{conv}(\mathcal{P})}} \min_{r} -\sum_{x} P(x)\log r(x).$$

By the Min-Max Theorem there exists a pair (P^*, r^*) assuming the max value

$$-\sum_{x} P^*(x)\log r^*(x).$$

Since $\sum\limits_{\kappa} P^*(x)\log\frac{P^*(x)}{r^*(\kappa)} \geq 0$,

$$-\sum_{x} P^*(x)\log r^*(x) \geq -\sum_{x} P^*(x)\log P^*(x)$$

and necessarily $r^* = P^*$.

Therefore we have the

Noiseless Source Coding Theorem for class $\mathcal{P}$

Let $L_{\mathcal{C},P}$ be the expected length of prefix code $\mathcal{C}$ for source $(\mathcal{U}, P)$, then

$$\max_{P\in\overline{\text{conv}(\mathcal{P})}} H(P) \leq \min_{\mathcal{C}} \sup_{P\in\mathcal{P}} L_{\mathcal{C},P} \leq \max_{P\in\overline{\text{conv}(\mathcal{P})}} H(P) + 1.$$

4. The $(\mathcal{Y}, \mathcal{N}, \mathcal{A})$ model and AV channels

Recently Katona/Tichler considered a model with lies depending on targets in the following sense: every permitted question specifies a subset $\mathcal{B} \subset \mathcal{U}$ and an $\mathcal{A} \subset \mathcal{B}$ with $|\mathcal{A}| \geq a$. If the object $u \in \mathcal{U}$ searched for satisfies $u \in \mathcal{A}$, the answer is arbitrary YES or NO. We refer to it as KT model. Their abstract in [S17] just says "We give results on the shortest search both in the adaptive and the non-adaptive case". **Presently we still don't know their results on the non-adaptive case.** The adaptive case is readily settled. It is done analogously to the proof of Katona for separating systems with constrained test sizes, which we first recall.

Consider the source $\mathcal{U} = \{u_1, \ldots, u_N\}$ and let $\mathcal{A} = \{A_1, \ldots, A_m\}$ be a system of subsets of $\mathcal{U}$, which is separating. That means that for any distinct u_j and

u_ℓ there exists an A_i $(1 \le i \le m)$ such that

$$u_j \in A_i \text{ and } u_\ell \notin A_i \text{ or } u_j \notin A_i \text{ and } u_\ell \in A_i. \tag{4.1}$$

For any $k \le \frac{N}{2}$ we want to determine $m(N,k)$ the minimal value of m for which there exists a separating system $\mathcal{A}$ with $|A_i| \le k$ $(1 \le i \le m)$.

We view $\mathcal{A}$ as a matrix with column vectors A_i $(1 \le i \le m)$ and row vectors

$$u_j = (u_{j1}, \dots, u_{jm}) \quad (1 \le j \le N).$$

By assumption these row vectors are distinct. We can then view the matrix also as a code $\mathcal{U} = \begin{pmatrix} u_1 \\ \vdots \\ u_N \end{pmatrix}$ with codewords u_j $(1 \le j \le N)$.

The specialty here is that the number of ones in every column of $\mathcal{U}$ does not exceed k.

Let $m^*(N,k)$ be the minimal number of questions in adaptive search. It is determined in [S15] and stated in [S10] as

Theorem 2. *For $N \le 2k$ $\quad m^*(N,k) = \lceil \log N \rceil$ and for $N > 2k$ we have for $\ell \triangleq \left\lceil \frac{N}{k} \right\rceil - 2$*

$$m^*(N,k) = \ell + \lceil \log N - k\ell \rceil.$$

Let $t^*(N,a)$ be the minimal number of questions in adaptive search for the KT-model with parameter a. Then

Theorem 3.

(i) For $a \le \frac{1}{3}N$ $\quad t^*(N,a) \le 1 + \left\lceil \log \left\lfloor \frac{2}{3}N \right\rfloor \right\rceil$
(ii) For $\frac{1}{3}N < a \le \frac{1}{2}N$ $\quad t^*(N,a) \le 1 + \lceil \log N \rceil$
(iii) For $a = N$ or $a = N-1$ no successful search is possible
(iv) For $a = N-2$ we get $m^*(N, N-2) = N-1$

Proof: Ad(i) Choose $\mathcal{B}$ and $\mathcal{A}$ with $|\mathcal{A}| = a$ and $|\mathcal{B}| = \frac{N+a}{2}$. Set $\mathcal{C} = [N] \smallsetminus \mathcal{B}$ and so $|\mathcal{C}| = \frac{N-a}{2}$.

We know that the target is in $\mathcal{B}$, if the answer is YES and that it is in $\mathcal{C} \cup \mathcal{A}$, if the answer is NO. In both cases we are left with $\frac{N+a}{2}$ possibilities. This number does not exceed $\frac{N}{2} + \frac{N}{6} = \frac{2}{3}N$. Let $[N_1]$ denote the set of left possibilities.

Next choose always $\mathcal{B}$'s with $\mathcal{A}$'s included in $[N] \smallsetminus [N_1]$. We are done with a total of $1 + \left\lceil \log \frac{2}{3} N \right\rceil$ queries.

Ad(ii) Again we have the formula $N_1 = \frac{N+a}{2}$ and $N - N_1 = \frac{N}{2} - \frac{a}{2}$. $[N] - [N_1]$ can be used for $\mathcal{A}$ to reduce a to $a_1 = \big(a - (N - N_1)\big) = \left(\frac{3a}{2} - \frac{N}{2}\right)$ and since $\frac{8}{3} a \leq \frac{4}{3} N$ and equivalently $3a - N \leq \frac{N}{3} + \frac{a}{3}$,

$$a_1 \leq \frac{1}{2}\left(\frac{N}{3} + \frac{a}{3}\right) = \frac{1}{3} N_1. \tag{4.2}$$

Now we continue with case (i) and get done with a total of

$$1 + 1 + \left\lceil \log \left\lfloor \frac{2}{3} N_1 \right\rfloor \right\rceil = 2 + \left\lceil \log \left\lfloor \frac{N+a}{3} \right\rfloor \right\rceil \leq 1 + \lceil \log N \rceil.$$

Finally, the **case (iii) being obvious** we come to the last case

Ad(iv) Choose $|\mathcal{A}| = N-2$, $|\mathcal{B}| = N-1$. After one question we get $N_1 = N-1$ and $a_1 = a - 1 = N - 3 = N_1 - 2$ and inductively we get the result.

Remark 6: The cases $a > \frac{N}{2}$ left are tricky because $m^*(N, a)$ varies between $1 + \lceil \log N \rceil$ and $N - 1$.

We turn now to the $(\mathcal{Y}, \mathcal{N}, \mathcal{A})$ model. There a test (query) is a partition of the search space $\mathcal{U}$ into 3 sets $\mathcal{Y}, \mathcal{N}$, and $\mathcal{A}$. If the object $u \in \mathcal{U}$ searched for satisfies $u \in \mathcal{Y}$ the answer is YES, if it satisfies $u \in \mathcal{N}$ the answer is NO, and it is for $u \in \mathcal{A}$ arbitrary YES or NO. The tasks become more difficult the larger $\mathcal{A}$ is, for instance for $|\mathcal{A}| = |\mathcal{U}|$ or $|\mathcal{U}| - 1$ nothing can be done.

In the KT model $|\mathcal{A}| \geq a$ was assumed, which immediately reduces to $|\mathcal{A}| = a$. One can make further specifications about $|\mathcal{Y}|$ and $|\mathcal{N}|$, for instance

$$|\mathcal{Y}| \leq k \quad \text{and} \quad |\mathcal{N}| = N - a - |\mathcal{Y}|$$

gives a model $(\leq k, a)$, which for $a = 0$ covers Rényi's problem mentioned above.

In the adaptive case the function of interest, defined canonically, is $m^*(N, k, a)$.

$m(N, k, a)$ is for the non-adaptive case, which we now consider.

Important here is the observation that we can use as response channel an **arbitrarily varying channel**, in short AVC (the "arbitrary" answers make the link not only in words).

As class of channel matrices we choose $\mathcal{W} = \{W_1, W_2\}$, where $\mathcal{X} = \{0, 1, a\}$ is the input and $\mathcal{Z} = \{0, 1\}$ is the output alphabet and $W_1(0|0) = W_1(1|1) = 1$, $W_1(1|a) = 1$ and $W_2(0|0) = W_2(1|1) = 1$, $W_2(0|a) = 1$.

The **maximal** error capacity of binary output AVC were determined in [S13] as

$$C = \max_P \min_{W \in \overline{\overline{\mathcal{W}}}} I(P|W) = \min_{W \in \overline{\overline{\mathcal{W}}}} \max_P I(P|W), \tag{4.3}$$

where $\overline{\overline{\mathcal{W}}}$ is the row-convex hull of $\mathcal{W}$ and $I(P|W)$ is the mutual information for channel W and input distribution P.

It has value 1 for the channel just defined (as is readily seen by not using letter a). However, the **new ingredient** in this search problem is a **frequency constraint for the use of letters**. In the non-adaptive case a search strategy corresponds to a code $\{u_1, \dots, u_M\} \subset \mathcal{X}^n$ such that the matrix

$$\mathcal{U} = \begin{pmatrix} u_{11} & \dots & u_{1n} \\ \vdots & & \\ u_{M1} & \dots & u_{Mn} \end{pmatrix}$$ satisfies for the KT-model the **column constraint**

$$\text{for all } 1 \leq t \leq n \quad |\{i : 1 \leq i \leq M, u_{it} = a\}| \geq a. \tag{4.4}$$

This forces to use letters which are "bad" for communication and changes AVC-theory. Notice that for instance our $\mathcal{W}$ has 0-1-matrices only, and therefore error probability $\lambda_{\max} < 1$ implies zero-error probability.

Actually it was shown that AVC-theory for 0-1-matrices is equivalent to Shannon's zero-error capacity theory for DMC's.

The column constraint starts a new enlargement of that theory!

To gain more insight we first come to subsections and **return to the AVC-model in the last section**.

Search with a cardinality constraint on tests: revisited

In [S7] we proved that the minimum size $m(N, \alpha N)$, $\alpha \in \left(0, \frac{1}{2}\right)$, of a separating system $\mathcal{B}$ for the search space $\mathcal{U} = \{1, 2, \dots, N\}$ with

$$|B| \leq \alpha N \quad \text{for} \quad B \in \mathcal{B} \tag{4.5}$$

is ruled by the entropy $h(\alpha)$, namely,

$$m(N, \alpha N) = \frac{\log N + o(\log N)}{h(\alpha)}. \tag{4.6}$$

We raise the following additional questions.

A. For a search space $(\mathcal{U}, P)$ with a PD assigned to $\mathcal{U}$ adaptive search is generally superior with respect to the expected search time already in the standard case with no constraint (4.5). How does the expected search time behave under constraint (4.5)?
B. The proof of [S7] can be termed a large deviational existence proof. What can we do constructively? [1]
C. The search model of size constraint tests can be viewed as a special case of what we call $(\mathcal{Y}, \mathcal{N}, \mathcal{A})$ model with parameters $(\alpha N, (1-\alpha)N, 0)$. What are the implications?
D. How can we perform search for a class of PD's, in particular also under constraint (4.5)?

Remark 7: Parallel to Search Theory there is a spectrum of research problems addressed in the Theory of Questionnaires. However, where solutions could be offered mostly questionnaires reduce to error free tests. Therefore many developments started to escape attention of researchers in search. In particular we draw attention to Fig. 3.2 in [S10] and the topic explained there. It relates to models where lies depend on targets. A better exchange between search and questionnaires will be attempted in a forthcoming updated edition of [S10].

Remark 8: For the KT-model (as well as for the $(k, N-k-a, a)$-model it is essential that adaptiveness is used immediately for **every single step** (which is not so for AVC with feedback in the schemes [S3], [42]). Therefore for this kind of problems our coding schemes with feedback (started in [S1]) are not optimal if applied to the $(k, N-k-a, a)$ model. We begin with A. and improve Theorem 2. to a probabilistic search space $(\mathcal{U}, P)$ with minimal expected number of questions $m^*(P, N, k)$.

The previous strategy is only slightly modified by following the ordered probabilities $P_1 \geq P_2 \geq P_3 \geq \cdots \geq P_N$ and beginning with the set $\{1, 2, \dots, k\}$ etc. Thus short search times go parallel with high probabilities.

[1] Recently we found a simple construction. Let m be minimal with the properties $N' = \binom{m}{l} \geq N$ and $\binom{m}{l} l \leq \alpha N m$ for $l < m$. Let $M^{(\alpha N)}_{m,N'}$ be the matrix with set of column vectors equal to the set $\binom{[m]}{l}$ of 0-1-vectors with l 1's. They are distinct. By **symmetry** every row contains $\binom{m}{l}\frac{l}{m} = \binom{m-1}{l-1} \leq \alpha N$ 1's. Omitting $N'-N$ columns gives a matrix $M^{(\alpha N)}_{m,N}$ with suitable tests in the rows.
Clearly for $l = \alpha m + o(m)$ we get (4.6) or equivalently $m(N, \alpha N) \leq \frac{\log N}{h(\alpha)} + o(\log N)$. Actually, letting m be minimal with the properties $N' = \sum_{i=0}^{l} \binom{m}{i} \geq N$ and $\sum_{i=0}^{l} \binom{m}{i} \frac{i}{m} = \sum_{i=1}^{l} \binom{m-1}{i-1} \leq \alpha N$ for $l < m$, we get even a little bit better result.

It is convenient to define

$$Q_j = P_{kj+1} + \cdots + P_{kj+k} \text{ for } j = 1, \dots, \ell \text{ and } Q_{\ell+1} = 1 - \sum_{j=1}^{\ell} Q_j.$$

Then the expected time is between

$$T \triangleq Q_1 + H\left(\frac{P_1}{Q_1}, \dots, \frac{P_k}{Q_1}\right) + Q_2 + H\left(\frac{P_{k+1}}{Q_2}, \dots, \frac{P_{2k}}{Q_2}\right) + \cdots + Q_\ell$$
$$+ H\left(\frac{P_{k\ell+1}}{Q_\ell}, \dots, \frac{P_{k\ell+k}}{Q_\ell}\right) + H(P_{k(\ell+1)+1}, \dots, P_N) \text{ and } T + \ell + 1.$$

More precise bounds become very technical to derive. We address now B.

After the discovery of the "entropy principle" (equation 4.6) in [S7] G. Katona pointed out that it must be derivable from his work [S16], [S17]. This is the case. In fact for some parameters we can give the following very simple construction, which also indicates how the remaining parameters can be settled with additional approximations.

Indeed, consider parameters m, k, and $N = \binom{m}{\ell}$ satisfying

$$\binom{m}{\ell}\ell = mk. \tag{4.7}$$

Then choose the set of vectors $\binom{[m]}{\ell}$ as set of row vectors in $\mathcal{U}$ (in any labelling). **By symmetry** all column vectors in $\mathcal{U}$ have by (4.7) k ones.

Elementary Information Theory gives the bounds

$$\frac{1}{m+1}\exp\left\{mh\left(\frac{\ell}{m}\right)\right\} \le \binom{m}{\ell} \le \exp\left\{mh\left(\frac{\ell}{m}\right)\right\}. \tag{4.8}$$

Therefore

$$m \le \frac{\log\binom{m}{\ell} + 2\log(m+1)}{h\left(\frac{\ell}{m}\right)} = \frac{\log N + 2\log(m+1)}{h\left(\frac{k}{N}\right)}$$

and thus

$$\begin{aligned} m(N,k) &\leq \frac{\log N + O(\log\log N)}{h\left(\frac{k}{N}\right)} \\ &\leq \frac{\log N + o(\log N)}{h\left(\frac{k}{N}\right)}, \end{aligned} \tag{4.9}$$

the desired entropy bound. Now comes the approximation.

Given N, k define $\alpha = \frac{k}{N} \leq \frac{1}{2}$, the maximal "density" of ones in $\mathcal{U}$ to be tolerated.

Step 1 Define the column number m by the inequalities

$$\binom{m}{\lfloor \alpha m \rfloor} \leq N < \binom{m+1}{\lfloor \alpha(m+1) \rfloor} \tag{4.10}$$

and choose submatrix $\mathcal{U}_1$ with the $N_1 = \binom{m}{\lfloor \alpha m \rfloor}$ row vectors from the set $\binom{[m]}{\lfloor \alpha m \rfloor}$. Thus all its rows have exactly $\lfloor \alpha m \rfloor$ ones and by symmetry all its columns have exactly

$$k_1 \triangleq \frac{N_1 \cdot \lfloor \alpha m \rfloor}{m} \leq N_1 \alpha = N_1 \frac{k}{N} = \frac{N_1}{N} k \tag{4.11}$$

ones.

Note that

$$\frac{1}{(m+1)^2} \exp\left\{ mh\left(\frac{\lfloor \alpha m \rfloor}{m}\right)\right\} \leq N \qquad \text{and thus}$$

$$m \leq \frac{\log N + 2\log(m+1)}{h\left(\frac{\lfloor \alpha m \rfloor}{m}\right)}.$$

Since $\lim\limits_{m\to\infty} \frac{\lfloor \alpha m \rfloor}{m} = \alpha$, therefore

$$m \leq \frac{\log N + O(\log\log N)}{h(\alpha)}. \tag{4.12}$$

It remains to be shown how $\mathcal{U}_1$ can be expanded to $\mathcal{U}$ by adding suitable rows.

Step 2 Let now t_1 be the smallest positive integer with

$$\binom{m}{\lfloor \alpha m \rfloor} + \binom{m}{\lfloor \alpha m \rfloor - t} \leq N$$

and add to $\mathcal{U}_1$ the submatrix

$$\mathcal{U}_2 = \binom{[m]}{\lfloor \alpha m \rfloor - t}.$$

Again by symmetry

$\begin{pmatrix} \mathcal{U}_1 \\ \mathcal{U}_2 \end{pmatrix}$ has fewer than k ones in its columns as simple calculation shows.

Step 3 Iterate the above construction with the smallest $t_2 > t_1$ such that

$$\binom{m}{\lfloor \alpha m \rfloor} + \binom{m}{\lfloor \alpha m \rfloor - t_1} + \binom{m}{\lfloor \alpha m \rfloor - t_2} \leq N.$$

Alternate approach

$$\binom{m}{\lfloor \alpha m \rfloor} \leq N < \binom{m+1}{\lfloor \alpha (m+1) \rfloor}.$$

From the left inequality we get

$$m \leq \frac{\log N + 2\log(m+1)}{h\left(\frac{\lfloor \alpha m \rfloor}{m}\right)}$$

and the right one guarantees the density and a subset of N rows quasi-balanced in the columns: they differ at most by 1 in the number of ones. This is also clear from Baranyai's work [S12], if $\lfloor \alpha(m+1) \rfloor | m+1$ and otherwise from the "generalized Baranyai" or Katona's Step C on page 179 of [S14].

Note that the notation there differs from ours by exchanged roles of rows and columns.

Probabilistic sources $(\mathcal{U}, P)$

We begin with an extension of Theorem 1 and bound $t^*(N, P, a)$, the minimal expected number of questions in adaptive search for the KT-model.

Theorem 4. *For* $a \leq \frac{1}{3}N \quad t^*(N,P,a) \leq 2 + H(P)$ $\Big(t^*(N,P,a) \geq H(P)$ *being obvious*$\Big)$

Remark 9: Presently we don't know whether 2 can be replaced by 1 in the bound and suggest to decide this question.

Proof: We begin as in the proof of Theorem 1 with the partition $(\mathcal{Y}, \mathcal{N}, \mathcal{A})$.

For any behaviour of the jammer partitioning $\mathcal{A}$ into $\mathcal{A}^+$, where he answers YES, and $\mathcal{A}^-$, where he answers NO, define the probabilities

$$Q_{\mathcal{Y}} = \sum_{i \in \mathcal{Y}} P_i, \quad Q_{\mathcal{N}} = \sum_{i \in \mathcal{N}} P_i, \quad Q_{\mathcal{A}^+} = \sum_{i \in \mathcal{A}^+} P_i, \text{ and } Q_{\mathcal{A}^-} = \sum_{i \in \mathcal{A}^-} P_i. \tag{4.13}$$

Also denote by $L(P')$ for any PD P' the minimal expected noiseless coding length for the source $(\mathcal{U}', P')$.

We know from Shannon that

$$H(P') \leq L(P') \leq H(P') + 1. \tag{4.14}$$

Then

$$\begin{aligned} t^*(N, P, a) \leq 1 &+ (Q_{\mathcal{A}^+} + Q_{\mathcal{Y}}) \left[H\left(\left(\frac{P_i}{Q_{\mathcal{A}^+} + Q_{\mathcal{Y}}} \right)_{i \in \mathcal{A}^+ \cup \mathcal{Y}} \right) + 1 \right] \\ &+ (Q_{\mathcal{A}^-} + Q_{\mathcal{N}}) \left[H\left(\left(\frac{P_i}{Q_{\mathcal{A}^-} + Q_{\mathcal{N}}} \right)_{i \in \mathcal{A}^- \cup \mathcal{N}} \right) + 1 \right]. \end{aligned}$$

We relate now the entropy quantities to $H(P)$ using the grouping axiom. The quantity

$$\begin{aligned} \alpha &\triangleq (Q_{\mathcal{A}^+} + Q_{\mathcal{Y}}) H\left(\left(\frac{P_i}{Q_{\mathcal{A}^+} + Q_{\mathcal{Y}}} \right)_{i \in \mathcal{A}^+ \cup \mathcal{Y}} \right) + (Q_{\mathcal{A}^-} + Q_{\mathcal{N}}) H\left(\left(\frac{P_i}{Q_{\mathcal{A}^-} + Q_{\mathcal{N}}} \right)_{i \in \mathcal{A}^- \cup \mathcal{N}} \right) \\ &= (Q_{\mathcal{A}^+} + Q_{\mathcal{Y}}) \left[H\left(\frac{Q_{\mathcal{A}^+}}{Q_{\mathcal{A}^+} + Q_{\mathcal{Y}}}, \frac{Q_{\mathcal{Y}}}{Q_{\mathcal{A}^+} + Q_{\mathcal{Y}}} \right) \right. \\ &\quad \left. + \frac{Q_{\mathcal{A}^+}}{Q_{\mathcal{A}^+} + Q_{\mathcal{Y}}} H\left(\left(\frac{P_i}{Q_{\mathcal{A}^+}} \right)_{i \in \mathcal{A}^+} \right) + \frac{Q_{\mathcal{Y}}}{Q_{\mathcal{A}^+} + Q_{\mathcal{Y}}} H\left(\left(\frac{P_i}{Q_{\mathcal{Y}}} \right)_{i \in \mathcal{Y}} \right) \right] \\ &+ (Q_{\mathcal{A}^-} + Q_{\mathcal{N}}) \left[H\left(\frac{Q_{\mathcal{A}^-}}{Q_{\mathcal{A}^-} + Q_{\mathcal{N}}}, \frac{Q_{\mathcal{N}}}{Q_{\mathcal{A}^-} + Q_{\mathcal{N}}} \right) + \frac{Q_{\mathcal{A}^-}}{Q_{\mathcal{A}^-} + Q_{\mathcal{N}}} H\left(\left(\frac{P_i}{Q_{\mathcal{A}^-}} \right)_{i \in \mathcal{A}^-} \right) \right. \\ &\quad \left. + \frac{Q_{\mathcal{N}}}{Q_{\mathcal{A}^-} + Q_{\mathcal{N}}} H\left(\left(\frac{P_i}{Q_{\mathcal{N}}} \right)_{i \in \mathcal{N}} \right) \right]. \end{aligned}$$

Obviously $Q_{\mathcal{A}^+} + Q_{\mathcal{A}^-} = Q_{\mathcal{A}}$.

Now we decompose $H(P)$.

$$H(P) = H(Q_{\mathcal{A}^+}, Q_{\mathcal{A}^-}, Q_{\mathcal{Y}}, Q_{\mathcal{N}}) + Q_{\mathcal{A}^+} H\left(\left(\frac{P_i}{Q_{\mathcal{A}^+}}\right)_{i\in\mathcal{A}^+}\right) + Q_{\mathcal{A}^-} H\left(\left(\frac{P_i}{Q_{\mathcal{A}^-}}\right)_{i\in\mathcal{A}^-}\right)$$
$$+ Q_{\mathcal{Y}} H\left(\left(\frac{P_i}{Q_{\mathcal{Y}}}\right)_{i\in\mathcal{Y}}\right) + Q_{\mathcal{N}} H\left(\left(\frac{P_i}{Q_{\mathcal{N}}}\right)_{i\in\mathcal{N}}\right)$$

and verify that

$$\begin{aligned} H(P) - \alpha = {} & H(Q_{\mathcal{A}^+}, Q_{\mathcal{A}^-}, Q_{\mathcal{Y}}, Q_{\mathcal{N}}) - (Q_{\mathcal{A}^+} + Q_{\mathcal{Y}}) H\left(\frac{Q_{\mathcal{A}^+}}{Q_{\mathcal{A}^+} + Q_{\mathcal{Y}}}, \frac{Q_{\mathcal{Y}}}{Q_{\mathcal{A}^+} + Q_{\mathcal{Y}}}\right) \\ & - (Q_{\mathcal{A}^-} + Q_{\mathcal{N}}) H\left(\frac{Q_{\mathcal{A}^-}}{Q_{\mathcal{A}^-} + Q_{\mathcal{N}}}, \frac{Q_{\mathcal{N}}}{Q_{\mathcal{A}^-} + Q_{\mathcal{N}}}\right) \\ = {} & H(Q_{\mathcal{A}^+} + Q_{\mathcal{Y}}, Q_{\mathcal{A}^-} + Q_{\mathcal{N}}) \leq 1. \end{aligned}$$

Thus

$$t^*(N, P, a) \leq 2 + H(P) - H(Q_{\mathcal{A}^+} + Q_{\mathcal{Y}}, Q_{\mathcal{A}^-} + Q_{\mathcal{N}}) \leq 2 + H(P).$$

Remark 10: For the $(N-k-a, k, a) = (\alpha_0 N, \alpha_1 N, \alpha N)$ model we can follow the first kind of steps until we are left with N_ℓ candidates and a_ℓ arbitrary tests with

$$\frac{N_\ell - a_\ell}{2} \leq \min(N-k-a, k)$$

so that we can continue with the KT-model. We thus get the bound $\ell(\alpha_1, \alpha_0, \alpha) + H(P)$. The details are left to interested readers.

The $(\mathcal{Y}, \mathcal{N}, \mathcal{A})$-model with non-adaptive strategies viewed in terms of AV-response channels with input letter frequency constraints.

Recall the channel $\mathcal{W}$ and a code $\{u_1, \dots, u_M\} \subset \mathcal{X}^n$ corresponding to a non-adaptive strategy such that the matrix

$$\mathcal{U} = \begin{pmatrix} u_{11} & \dots & u_{1n} \\ \vdots & & \\ u_{M1} & \dots & u_{Mn} \end{pmatrix} \text{ satisfies the } \textbf{column constraints} \text{ for all } 1 \leq t \leq n$$

$$\begin{aligned} &|\{i : 1 \leq i \leq M \ u_{it} = 0\}| = N - k - a \\ &|\{i : 1 \leq i \leq M \ u_{it} = 1\}| = k \\ &|\{i : 1 \leq i \leq M \ u_{it} = a\}| = a. \end{aligned} \tag{4.15}$$

To simplify calculations let us assume that

$$a = \alpha N, k = \alpha_1 N, \text{ and } N - k - a = \alpha_0 N. \tag{4.16}$$

We mentioned that AVC-theory for frequency constraint is not far developed. A capacity formula suggesting itself is

$$C_{\alpha,\alpha_1,\alpha_0} = \min_{W \in \overset{=}{\mathcal{W}}} I \left(\begin{pmatrix} \alpha_1 \\ \alpha_0 \\ \alpha \end{pmatrix} \middle| W \right). \tag{4.17}$$

A basic idea here is that list codes allow a much more unified capacity theory than ordinary codes do (see [6]). $(\mathcal{U}, \mathcal{D})$ is a list code of list size L, if for $\mathcal{D} = (D_1, \dots, D_M)$, $D_i \subset \mathcal{Z}^n$

$$\sum_{i=1}^{M} 1_{D_i}(z^n) \leq L \text{ for all } z^n \in \mathcal{Z}^n.$$

It has error probability less than λ if $W(D_i | u_i, s^n) \geq 1 - \lambda$ for all i and $s^n \in \{0,1\}^n$ defining a channel sequence.

Since we have 0-1-matrices only in $\mathcal{W}$ $\lambda < 1$ implies 0-error.

It was shown in [6] that for list sizes $L(n) \to \infty$ as $n \to \infty$ the formula gives the capacity!

This gives hope also under our additional input constraints.

Lemma. (see [S3], [6] or [S12])

Let $\mathcal{U} = \binom{[n]}{\alpha_1 n, \alpha_0 n, \alpha n} = \mathcal{T}_P^n$ *for* $P = \left(\begin{smallmatrix} \alpha_1 \\ \alpha_0 \\ \alpha \end{smallmatrix}\right)$ *and* $D_i = \Big\{ z^n : z_t = u_{it}$ *for* $u_{it} \in \{0,1\} \Big\}$, *then*

$$L = \max_{z^n} \sum_{i=1}^{M} 1_{D_i}(z^n) = exp_2\{h(\alpha)n + o(n)\}.$$

Proof: Can be given directly or as special case of Lemma 7.1 of [S12].

Now notice that $M = \exp_2\{H(\alpha_1, \alpha_0, \alpha)n + o(n)\}$ and therefore

$$\frac{M}{L} = \exp_2\{H(\alpha_1, \alpha_0, \alpha)n - h(\alpha)n + o(n)\}.$$

For $\alpha_1 = \alpha_0$ this specializes to $\frac{M}{L} = \exp_2\{(1-\alpha)n + o(n)\}$. Notice that this rate equals

$$I\left(\begin{pmatrix}\frac{1-\alpha}{2}\\ \frac{1-\alpha}{2}\\ \alpha\end{pmatrix} \middle| \begin{pmatrix}1&0\\0&1\\ \frac{1}{2}&\frac{1}{2}\end{pmatrix}\right) = h\left(\frac{1}{2}\right) - \alpha h\left(\frac{1}{2}\right) = 1-\alpha.$$

Final comments

Example: $\alpha_1 = \alpha_0 = \alpha = \frac{1}{3}$.

There are three natural bounds:

1. Choose $u_i = v_i\ a^{\frac{1}{3}n}$, where $a^{\frac{1}{3}n} = a\dots a \in \{0,1\}^{\frac{1}{3}n}$ and $v_i \in \{0,1\}^{\frac{2}{3}n}$.
   We get $|\mathcal{U}| = 2^{\frac{2}{3}n}$, rate $= \frac{2}{3}$.
2. Gilbert's bound gives the rate $= \log\frac{3}{2}$, which is smaller.
3. List size of $\binom{[n]}{\frac{1}{3}n, \frac{1}{3}n, \frac{1}{3}n}$
   The sequence $1^{n/2}0^{n/2} \in \mathcal{Z}^n$ has maximal list size $\binom{\frac{1}{2}n}{\frac{1}{3}n}^2$ of rate $h\left(\frac{1}{3}\right)$ and thus we get $\frac{M}{L}$ of rate $\log 3 - h\left(\frac{1}{3}\right) = 2/3$, again like in 1.

But now we can iterate the list reduction as in [6] and observe that this can be done with the frequency constraints in the columns (using balanced colorings as in [8] and in [S4]).

For small list size (but **not** constant) we get

$$\frac{\log N}{2/3}.$$

More generally for $\left(\frac{1-\alpha}{2}, \frac{1-\alpha}{2}, \alpha\right)$ case 1) and case 3) give $1-\alpha$ instead of $2/3$. By 3) again with small (not constant) list size we achieve

$$\frac{\log N}{1-\alpha}.$$

This cannot be superceded, but what is best, if we go for list size 1?

All this gives nice zero-error capacity theory under frequency constraints.

Presently we can handle here also list versions for general parameters $(\alpha_1, \alpha_0, \alpha)$. Furthermore we can extend this to q-ary models with parameters

$$(\alpha_{q-1}, \dots, \alpha_0, \alpha).$$

In the binary case for targets in $\mathcal{A}$ two answers can happen, YES and NO or 1 and 0. Now for instance for $q = 3$ there are different $\mathcal{A}$'s, namely, $\mathcal{A}_{012}$, where three answers are possible, namely, 2, 1, and 0. Furthermore $\mathcal{A}_{01}$, $\mathcal{A}_{02}$, $\mathcal{A}_{12}$ where the specified two answers are possible.

In the setting of an AV $\mathcal{W}$ contains 24 matrices of the structure

$$W = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 01 \\ 02 \\ 12 \\ 012 \end{array} \left(\begin{array}{c} 0 \qquad 1\ 2 \\ V_0 \\ V_1 \\ V_2 \\ V_{01} \\ V_{02} \\ V_{12} \\ V_{012} \end{array} \right),$$

where

$V_0 = (100)$, $V_1 = (010)$, $V_2 = (001)$
$V_{01} \in \{(100), (010)\}$, $V_{02} \in \{(100), (001)\}$, $V_{12} = \{(010), (001)\}$,
$V_{012} \in \{(100), (010), (001)\}$ and $2 \cdot 2 \cdot 2 \cdot 3 = 24$.

All these models can be handled in a new subject of **list searching** for moderate list sizes. Problems are **very difficult** for **constant** list sizes. We end here with the field of non-adaptive strategies wide open!

References

[S1] R. Ahlswede, A constructive proof of the coding theorem for discrete memoryless channels in case of complete feedback, Sixth Prague Conf. on Inf. Th., Stat. Dec. Fct's and Rand. Proc., Sept. 1971, Publishing House Czechosl. Academy of Sc., 39-50, 1973.

[S2] R. Ahlswede, Arbitrarily varying channels with states sequence known to the sender, IEEE Trans. Inf. Theory, Vol. IT-32, No. 5, 621-629, 1986.

[S3] R. Ahlswede, Channels with arbitrarily varying channel probability functions in the presence of noiseless feedback, Z. Wahrscheinlichkeitstheorie u. verw. Geb. 25, 239-252, 1973.

[S4] R. Ahlswede, Concepts of performance parameters for channels, (Original version: Concepts of performance parameters for channels, Preprint 00–126, SFB 343 "Diskrete Strukturen in der Mathematik", Universität Bielefeld) General Theory of Information Transfer and Combinatorics, Lecture Notes in Computer Science, Vol. 4123, Springer Verlag, 639-663, 2006.

[S5] R. Ahlswede, General theory of information transfer, Preprint 97-118, SFB 343 "Diskrete Strukturen in der Mathematik", Universität Bielefeld, 1997.

[S6] R. Ahlswede, Identification entropy, General Theory of Information Transfer and Combinatorics, Lecture Notes in Computer Science, Vol. 4123, Springer Verlag, 595-613, 2006.

[S7] R. Ahlswede, Rate-wise optimal non-sequential search strategies under a cardinality constraint on the tests, General Theory of Information Transfer and Combinatorics, Special Issue of Discrete Applied Mathematics, this volume.

[S8] R. Ahlswede and N. Cai, An interpretation of identification entropy, IEEE Trans. Inf. Theory, Vol. 52, No. 9, 4198-4207, 2006.

[S9] R. Ahlswede and N. Cai, Transmission identification and common randomness capacities for wire-tap channels with secure feedback from the decoder, General Theory of Information Transfer and Combinatorics, Lecture Notes in Computer Science, Vol. 4123, Springer Verlag, 258-275, 2006.

[S10] R. Ahlswede and I. Wegener, Suchprobleme, Teubner Verlag, Stuttgart, 1979, Russian Edition with Appendix by Maljutov 1981, English Edition "Search Problems", Wiley-Interscience Series in Discrete Mathematics and Optimization, R.L. Graham, J.K. Leenstra, R.E. Tarjan, edit., 1987.

[S11] R. Ahlswede and J. Wolfowitz, The capacity of a channel with arbitrarily varying channel probability functions and binary output alphabet, Z. Wahrscheinlichkeitstheorie und verw. Geb. 15, 186-194, 1970.

[S12] Z. Baranyai, On the factorization of the complete uniform hypergraph, Infinite and finite sets, Colloq. Kesthely 1973, Vol. I, Colloq. Math. Soc. J. Bolyai 10, North-Holland, Amsterdam, 91-108, 1975.

[S13] S. Gelfand and M.S. Pinsker, Coding for channels with random parameters, Problems of Control and Inf. Theory 9, No. 1, 19-31, 1980.

[S14] G. Katona, On separating systems of a finite set, J. Comb. Theory 1, 174-194, 1966.

[S15] G. Katona, Combinatorial search problems, A survey of combinatorial theory, Ed., Srivastava et al, 285-308, 1973.

[S16] G. Katona, Search with small sets in presence of a liar, preprint 2004.

[S17] G. Katona and K. Tichler, When the lie depends on the target, lecture at a Workshop on Combinatorial Search, Budapest, Hungary, April 23-26, 2005.

[S18] C.E. Shannon, The zero-error capacity of a noisy channel, IRE Trans. Inf. Th., 3, 3-15, 1956.

[S19] E. Berlekamp, Block coding with noiseless feedback, PhD thesis, MIT, 1964.

GTIT-Supplement

Noiseless coding for multiple purposes: a combinatorial model

Important developments in Computer Science concerning communication aspects in parallel computing have been brought to our attention by Micah Adler at the Dagstuhl Seminar "Algorithmic Aspects of Large and Complex Networks", 16.9.2001-21.9.2001.

Important references can be found in [A1] and [A2]. A basic model presented is readily described. The model is different from the probabilistic model discussed in Section 17.

Given is an $n \times n$-matrix M with 0 and 1 as entries to a sender. There are n receivers, receiver j wants to know the j-th column of M (perfectly).

There are two kinds of data:

Common information: the sender can give m bits about M **to all receivers**.

Separate information: each receiver can get a window with k rows of M with its content. This counts kn bits for him and there is no price for specifying the k rows (with $\lceil \log \binom{n}{k} \rceil$ bits).

Obviously, choosing $k = n$ for all receivers (and $m = 0$) they all get their desired information for the price n^2. The same can be achieved by choosing $k = 0$ and $m = n^2$ bits of common information.

How can one do better?

The pointer idea

Concerning receiver j the sender proceeds as follows:

1. He looks at the first k rows in M and reads $c_j(k)$ in the upper j-th column of length k.
2. If $c_j(k)$ appears as **substring** in some column between rows $k+1$ and n thus he gives as separate information to receiver j the k rows together

with this column, which requires $\lceil \log n \rceil$ bits counting for the common information and counts kn bits separate information. Knowing $2k$ positions receiver j is done with $n - 2k$ last common bits.

3. If there is no such column-substring, then the sender just sends the first k rows to receiver j.

Notice that in this case no substring $c_j(k)$ occurs in a column of the lower part matrix, in particular not in column j.

This dramatically reduces the remaining possibilities for the lower part matrix. Indeed, the number of 0-1-sequences of length $n - k$ not containing $c_j(k)$ is actually known to be independent of the value $c_j(k)$ of an k-length 0-1-sequence and is bounded by

$$\sum_{t=0}^{k-1} \binom{n-k}{t} \leq k \; 2^{h\left(\frac{k}{n-k}\right)(n-k)}.$$

This is bitwise less than $h\left(\frac{k}{n-k}\right)(n-k)t \log k$.

The procedure gives kn separate bits for every receiver and the common number of bits is bounded by

$$n \max \left(n - 2k + \log n, h\left(\frac{k}{n-k}\right)(n-k) + \log k\right)$$

and for $k = n^\alpha$, for instance, this equals

$$n\Big(n - 2k + o(k)\Big).$$

Finally we have not tried to be very efficient, but we have achieved our goal

$$n\Big(n - 2k + o(k)\Big) + nk = \Big(n - k + o(k)\Big) < n^2.$$

References

[A1] M. Adler, New coding techniques for improved bandwidth utilization, 37th Annual Symposium on Foundations of Computer Science, IEEE Comput. Soc. Press, Los Alamitos, 173–182, 1996.

[A2] M. Adler and T. Leighton, Compression using efficient multi-casting, Proceedings of the Thirty-Second Annual ACM Symposium on Theory of Computing, New York, 153–162, 2000.