Shannon Lecture at ISIT in Seattle 13th July 2006: Towards a General Theory of Information Transfer

R. Ahlswede,
Faculty of Mathematics,
University of Bielefeld,
Bielefeld, Germany
http://www.math.uni-bieleteld.de/ahlswede/it.html

"More than restoring
strings of symbols transmitted
means transfer today."

Introduction

We all know that C.E. Shannon in his paper [97] presented a theory of Transmission over Noisy Channels based on the concept of codes. He considered crucial performance criteria like rates and error probabilities, predicted there connections and outlined proofs for them. (For a unification in terminology we refer to [11]). Subsequently he was involved in refining estimates and in inspiring others to do so. As highlights we point at the two papers [100] and [101]. Another aspect, complexity of coding, gave a strong impetus to several theoretical and practical inventions of concepts and methods, which kept a large community of scientists busy for more than fifty years now. It is not our aim to describe or comment on these developments. That has been done in many books and articles. Recently two of them, [56] and [71], came in our hands and we can recommend them. A similar situation can be encountered in Data Compression—lossless and lossy meeting fidelity criteria—Shannon's other favorite research area within Information Theory. Also here most optimality results (Source Coding Theorems) are—very similar to the situation in Statistics—of an asymptotic nature. A reminder for going to more practicability came from Ziv's lecture [117].

After these preliminary remarks we come to start to justify our (ambitious) title. Perhaps the most direct and easy way is to draw attention to two statements of Shannon from his paper [97] concerning communication and filtering (also called denoising), respectively.

> *"The fundamental problem of communication is that of reproducing at one point either exactly or approximately a message selected at another point."*

> *"If the source already has a certain redundancy and no attempt is made to eliminate it … a sizable fraction of the letters can be received incorrectly and still reconstructed by the context."*

Both times the goal of reproducing transmitted data is clearly expressed.

However **transmission** is not the only goal of communication for human beings (and animals?!). A step beyond this goal of Shannon's celebrated theory of communication was made with our creation of a theory of identification in the presence of noise. The mathematical foundations were laid together with G. Dueck and carried on by Verboven, van der Meulen, Zhang, Cai, Csiszár, Han, Verdu, Steinberg, Anantharam, Venkataram, Wei, Csibi, Yeong, Yang, Shamai, Merhav, Burnashev, Bassalygo, Narayan and many others.

To fix ideas, *transmission* (classical) 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 $\mathcal{M} = \{1, \ldots, M\}$?"

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 $\mathcal{N} = \{1, 2, \ldots, N\}$.

This theory initiated other research areas like Common Randomness, Authentication in Cryptology, Alarm Systems. It also led to the discovery of new methods which become fruitful also for the classical theory of transmission, for instance in studies of robustness like arbitrarily varying channels, optimal coding procedures in case of complete feedback, novel approximation problems for output statistics and generation of common randomness, the key issue in Cryptology.

Moreover 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. It has been emphasized with the two citations from Shannon 1948 above! However, 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.

It seems that the whole body of present day Information Theory will undergo serious revisions and some dramatic expansions. A general theory of information transfer abbreviated as GTIT, was developed and to some degree analyzed in [13]. It extends 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 [13]. *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 based on randomized encoding*. Quite surprisingly whereas Shannon's coding theorem for transmission shows that sizes of maximal message sets grow *exponentially* in the block length of optimal codes used for fixed guaranteed error probability, now, for instance already for identification typically there is a *double exponentially* growth and we determine again the best exponent, called now *second order capacities*, exactly (see [16], [34], [35], [41], [44], [47], [77], [78], [85], [107]).

On the source coding side we introduced data compression for identification and discovered the identification entropy.

Finally we mention as a new and perhaps most promising direction the study of probabilistic algorithms with identification as *concept of solution* in mathematics. (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?!

A General Communication Model

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

$$\begin{aligned} D_i \cap D_{i'} &= \varnothing \quad \text{for } i',\ i \neq 1, \dots, M, \\ W^n(D_i^c | u_i) &\le \lambda \quad \text{for } i = 1, \dots, M. \end{aligned}$$

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**.

One can conceive of many situations in which the receiver has (or many receivers have) **different goals**.

A nice class of such situations can, abstractly, be described by **a family $\Pi(\mathcal{M})$ of partitions of $\mathcal{M}$**. Decoder $\pi \in \Pi(\mathcal{M})$ 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: $\Pi_S = \{\pi_{sh}\}, \pi_{sh} = \{\{m\} : m \in \mathcal{M}\}$. This describes Shannon's classical transmission problem stated above.

Model 2: $\Pi_I = \{\pi_m : m \in \mathcal{M}\}$ with $\pi_m = \{\{m\}, \mathcal{M} \setminus \{m\}\}$. Here decoder π_m wants to know whether m occured or not. This is the identification problem.

Model 3: $\Pi_K = \{\pi_S : |S| = K, S \subset \mathcal{M}\}$ with $\pi_S = \{S, \mathcal{M} \setminus S\}$. This is an interesting generalisation of the identification problem. We call it **K-identification** (relation to superimposed codes, Kautz/Singleton Codes).

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

Model 5: $\Pi_B = \{\{A, \mathcal{M} \setminus A\} : A \subset \mathcal{M}\}$. Here $\pi_A = \{A, \mathcal{M} \setminus A\}$ wants to know the answer to the binary question "Is m in A?".

Model 6: $\mathcal{M} = \{0,1\}^\ell$, $\Pi_C = \{\pi_t : 1 \le t \le \ell\}$ with $\pi_t = \{\{(x_1, \dots, x_\ell) \in \mathcal{M} : x_t = 1\}, \{(x_1, \dots, x_\ell) \in \mathcal{M} : x_t = 0\}\}$. 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. They are known for models 1, 2. It is shown in [13] 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 $\{(Q(\cdot|i), D_\pi) : 1 \le i \le N, \pi \in \Pi_K\}$, where the $Q(\cdot|i)$'s are PD's on $\mathcal{X}^n$, $D_\pi \subset \mathcal{Y}^n$, and where for all $\pi = \{S, \mathcal{M} \setminus S\}$ $(S \in \binom{\mathcal{M}}{K})$

$$\begin{aligned} \sum_{x^n} Q(x^n|i) W^n\left(D_\pi^c | x^n\right) &\le \lambda \quad \text{for all } i \in S, \\ \sum_{x^n} Q(x^n|i) W^n\left(D_\pi | x^n\right) &\le \lambda \quad \text{for all } i \notin S. \end{aligned}$$

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 choosen 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

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

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 a subsequence of events, say,

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

The ℓ persons are to be informed with high probability correctly about the correct sequence of events.

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

There are of course several other situations described by a family $\Pi(\mathcal{M})$ of partitions of $\mathcal{M}$. There are others, which do not fall exactly in this setting. One of them is that of L-identification introduced by Christian Heup in [82] for source coding. For a one-way channel, which we assume to be a discrete memoryless channel, abbreviated as DMC, L-identification refers to the situation where an L-subset of $\mathcal{M}$ is given to the encoder. For example, the encoder knows L persons $m_1,\dots,m_L\in\mathcal{M}$, who have won a lottery. On the receiver's side, a member of $\mathcal{M}$, wants to know whether or not he or she is among the winners. However, the information in which a participant is interested can no longer be represented by a partition of $\mathcal{M}$. We have to partition $\binom{\mathcal{M}}{L}$ and get

$$\Pi_{L,\text{set}}=\{\pi_m\ :\ m\in\mathcal{M}\},$$

where $\pi_m=\{S_m,\binom{\mathcal{M}}{L}\backslash S_m\}$ and $S_m=\{S\in\binom{\mathcal{M}}{L}\ :\ m\in S\}$. We call this model *L-identification for sets*.

One could also think of situations where the L objects, which are known to the encoder, need not be pairwise different. We call this *L-identification for vectors*. The model for this is

$$\Pi_L=\{\pi_m\ :\ m\in\mathcal{M}\},$$

where $\pi_m=\{A_m,\mathcal{M}^L\backslash A_m\}$ and

$$A_m=\{A\in\mathcal{M}^L : A\text{ has at least one component equal to }m\}.$$

Encoding and decoding have to be devised so that every participant, a member of $\mathcal{M}$ can make his decision with small probability of error.

The theory of identification led us to the discovery of the concept of common randomness. The interplay between second order identification capacity and first order common randomness capacity is discussed in the Introduction of the book [19] on pages 6–16 and in our Shannon Lecture (http://media.itsoc.org/isit2006/ahlswede/).
For further important work on common randomness we refer to the papers [31], [32], [65], [84], [91], [92], [95], and [112].

Whereas Shannon is usually credited as the founder of Information Theory in a probabilistic setting, Hamming is often mentioned as the originator of combinatorial models for communication.

In our report of results we follow this devision, also comment on combi-probabilistic models, and finally give further perspectives.

It has been reported that Shannon devoted a great part of his lecture at the 1973 ISIT in Ashkelon to feedback. Still it is recorded that in 1998 at a workshop of the Netherlands Academy of Sciences a well-known member of the Information Theory Society articulated his position that feedback problems don't belong to channel coding theory, especially for AVC-channels. In his lecture "Information Theory after Shannon" in Bielefeld, August 12, 2003 Massey mentioned that "Shannon never treated feedback with the same sweeping generality that he did with almost everything else he founded." and added that in his opinion "Shannon was trying to encourage his "followers" to take up this subject." In his 1990 paper [90] he gave his support to **Marko's** ([89]) concept of directional information for channels with feedback. Those are views, which we share, and this is documented in chapter 2 of [42].

We emphasize that GTIT can also be studied in the presence of feedback.

In the late 70-ties we wrote with I. Wegener the first book on search [42]. Its title "Search Problems" indicates the explorative state of the subject at that time. Contributions (often the same several times) came from Statistics, Economy, Computer Science or even entertainment games. We started viewing the subject as part of Information Theory.

In general a search problem can be formulated as a coding problem for channels with passive noiseless feedback. Furthermore noiseless source coding can be described as a search problem with probabilities on the search space. It is remarkable that very different kinds of problems could be classified as search problems and that researchers from various fields often know little about results achieved in areas in which they don't work.

The significance of the feedback scheme in [2] (see also [3] and [25] for robust versions), which is based on iterative list reductions, has been recognized after many years and the construction has been made the basis of the book [96].

Again there are probabilistic and combinatorial models.

In order to convey the flavor of the later subject to a broader scientific community we present below the solution of a seemingly basic classical problem.

The Rényi-Berlekamp-Ulam **search game with t lies** has an equivalent formulation in terms of transmission of messages over a noisy binary channel with t errors in the presence of feedback. For block length n and error fraction $\tau = \frac{t}{n}$ the optimal rate for all large n we call capacity-error function $C_2(\tau)$.

In his 1964 MIT thesis Berlekamp found a coding strategy achieving equality of C_2 with the Hamming bound H_2 for infinitely many τ. He also showed that C_2 coincides with the tangent at H_2 through the point $(\frac{1}{3}, 0)$.

In joint work with C. Deppe and V. Lebedev we discovered a coding scheme which gives such results for all alphabet sizes q. Surprisingly the corresponding tangent at H_q, the q-ary Hamming bound, starts for every $q \geq 3$ at $(\frac{1}{2}, 0)$!

There is a great variety of search problems caused by different error concepts, types of questions, and the structure of objects searched for (the most prominent example being group testing). The flavor can be gained from [66].

A. Probabilistic Models

I Transmission via DMC (Shannon Theory)

How many possible messages can we transmit over a noisy channel? Transmission means there is an answer to the question: **"What is the actual message?"**

Define $M(n, \lambda) = \max\{M : \exists(n, M, \lambda)\text{-code}\}$.

Shannon 48: $\lim_{n\to\infty} \frac{1}{n} \log M(n, \lambda) = C$, where the capacity $C = \max_X I(X \wedge Y)$ and the mutual information $I(X \wedge Y)$ equals $H(X) - H(X|Y)$, that is, the difference of the entropy $H(X)$ and the conditional entropy $H(X|Y)$.

II Identification via DMC (Including Feedback)

How many possible messages can the receiver of a noisy channel identify? Identification means there is an answer to the question "**Is the actual message i?**", where i can be any member of the set of possible messages $\{1, 2, \ldots, N\}$. Here randomisation helps!!!

$\{(Q(\cdot|i), D_i) : 1 \leq i \leq N\}$ is an $(n, N, \varepsilon_1, \varepsilon_2)$ ID–code of $Q(\cdot|i) \in \mathcal{P}(\mathcal{X}^n) =$ set of all PD on $\mathcal{X}^n$, $D_i \subset \mathcal{Y}^n$, and

(1) $\sum_{x^n \in \mathcal{X}^n} Q(x^n|i) W^n(D_i^c|x^n) \leq \varepsilon_1 (1 \leq i \leq N)$ (Error of 1st kind: i rejected, but present)

(2) $\sum_{x^n \in \mathcal{X}^n} Q(x^n|j) W^n(D_i|x^n) \leq \varepsilon_2 \forall i \neq j$ (Error of 2nd kind: i accepted, but some $j \neq i$ present).

Define $N(n, \varepsilon) = \max\{N : \exists(n, N, \varepsilon, \varepsilon)$ ID–code$\}$.

Theorem AD$_1$ (Double exponent.–Coding Theorem and soft converse)

(1) $\underline{\lim}_{n\to\infty} \frac{1}{n} \log\log N(n, \varepsilon) \geq C \ \forall \varepsilon \in [0, 1]$

(2) $\overline{\lim}_{n\to\infty} \frac{1}{n} \log\log N(n, 2^{-\delta n}) \leq C \ \forall \delta > 0$.

(Han/Verdú $\lim_{n\to\infty} \frac{1}{n} \log\log N(n, \varepsilon) = C \ \forall \varepsilon \in (0, \frac{1}{2})$)

$C =$ second order identification capacity $=$ Shannon's (first order) transmission capacity.

Theorem AD$_2$ *In case of **feedback** the 2–order ID–capacities are, if* $C > 0$,
without randomisation $C_f(W) = \max_{x\in\mathcal{X}} H(W(\cdot|x))$ *and*
with randomisation $C_f(W) = \max_P H(PW) \geq C$.

Phenomena:

1. **Feedback** increases the optimal rate for identification.

2. **Noise** can increase the identification capacity of a DMC in case of feedback (think about probabilistic algorithms, here noise creates the randomisation, which is not the case for Shannon's theory of transmission)

3. **Idea:** Produce a "big" (large entropy) random experiment with a result known to sender and receiver and use the Transformer Lemma below.

"Principle": Entropy of a large common random experiment $=$ ID–capacity of 2-order (region).

Remark ID–theory led to the foundation of new areas and stimulated further research.

Approximation of output distributions

It originated from converse proofs in Theorem AD$_1$. How can we count? For $P \in \mathcal{P}(\mathcal{X}^n)$ find minimal $\mathcal{U} \subset \mathcal{X}^n$ with uniform distribution $P_{\mathcal{U}}$ such that $P_{\mathcal{U}} W \sim PW$. Then $N \lesssim \binom{|\mathcal{X}^n|}{|\mathcal{U}|}$.

Information measures

How do we measure information, by the Cn bits in Shannon's fundamental theorem or by the $\log Cn$ bits in our Theory of Identification?

III Discovery of Mystery Numbers = Common Randomness Capacity

It was observed in [35] that in Identification the second order rate is essentially determined by the first order rate of a random experiment set up by the communicators and whose outcome is known to both, sender and receiver, with high probability. In other words instead of the requirement for the receiver to recover the message sent by the sender with high probability it is required for the communicators to know the value of the same random variable with high probability. Thus a new concept, different from

both, transmission and identification, but with interesting connections to them was introduced. It is now called **common randomness**. A systematic presentation can be found in [31], [32]. Many interesting and important results and applications of common randomness have been obtained so far. **When we speak of GTIT today we mean it to include at its core the theory of information transmission, common randomness, identification and its generalizations and applications, but it goes far beyond it even outside communication theory when we think about probabilistic algorithms with identification (or more general tasks) as concepts of solution!**

Actually, the origin of the concepts **common randomness and common randomness capacity** took a fascinating path. Immediately after [34] the results of [35] were discovered—the papers appeared face by face in the same volume. An output process $Y_1, \ldots, Y_n$ produced by a DMC from an input process $X_1, \ldots, X_n$ is not only known to the receiver of the channel W, but also to its sender, if there is a noiseless (passive) feedback channel. This common knowledge of the random process was used in [35] for the randomization in a randomized identification procedure, which devotes a blocklength n to creating $Y_1, \ldots, Y_n$ and does then the identification in blocklength $\sqrt{n}$ (also called $\sqrt{n}$-trick). The size of the identification code obtained is of order $e^{e^{H(Y)n}}$! Making a best choice of X one gets the second order rate $C_F = \max_X H(Y)$, and the identification works if Shannon's transmission capacity $C_{Sh} = \max_X (H(Y) - H(Y|X))$ is positive.

Now the second idea was to wonder whether there is also or can be constructed also a random experiment (or process) in the original case of no feedback in [34], where the second order identification capacity equals C_{Sh}. Well, just choose a transmission λ-code $\{(u_i, D_i) : 1 \leq i \leq exp\{(C_{Sh} - \delta)n\}\}$ and define X^n as the RV taking codewords as values with equal probabilities.

Thus of course the sender knows X^n, but the receiver knows it almost, namely with an error probability not exceeding λ, if he uses the decoding sets D_i. This slight deviation from exact knowledge was not essential, the described experiment in conjunction with the Transformator Lemma below gave a second proof of the direct part of the coding theorem in [34].

This discovery was followed up by [41] and led to solutions of identification problems for multi-way channels with noiseless feedback. The paper contains a novel method to prove weak converses by exploiting Schur concavity of the entropy function. In addition it has two new features, firstly it settles a rather rich class of channel models unheard of in multi-user theory for transmission, where it can be said—"cum grano salis"—that after struggles of more than 30 years the frontiers could not be moved very far beyond [4], secondly the identification schemes are all **constructive** modulo the production of rich random experiments. This richness is measured by what was called **Mystery Numbers** or **Regions of k-tuples of Mystery Numbers** in [41].

The constructions are based on Freivald's Lemma for hashing. As byproduct it gives also a constructive scheme for deterministic channels because they automatically have feedback. Shortly thereafter another construction was given for these special channels in [108].

In dealing with different kinds of feedback strategies it is convenient to have the following concept. Let $\mathcal{F}_n (n = 1, 2, \ldots)$ be a subset of the set of all randomized feedback strategies $\mathcal{F}_n^r$ of a DMC W with blocklength n and let it contain the set $\mathcal{F}_n^d$ of all deterministic strategies.

We call $(\mathcal{F}_n)_{n=1}^{\infty}$ a smooth class of strategies if for all $n_1, n_2 \in \mathbb{N}$ and $n = n_1 + n_2$

$$\mathcal{F}_n \supset \mathcal{F}_{n_1} \times \mathcal{F}_{n_2} \tag{1}$$

where the product means concatenation of strategies.

For $f^n \in \mathcal{F}_n$ the channel induces an output sequence $Y^n(f^n)$. For any smooth class we define numbers

$$\mu(\mathcal{F}_n) = \max_{f^n \in \mathcal{F}_n} H(Y^n(f^n))$$

By (1) and the memoryless character of the channel

$$\mu(\mathcal{F}_n) \geq \mu(\mathcal{F}_{n_1}) + \mu(\mathcal{F}_{n_2}),$$

and therefore

$$\mu = \mu((\mathcal{F}_n)_{n=1}^{\infty}) = \lim_{n \to \infty} \frac{1}{n} \mu(\mathcal{F}_n) \textbf{ exists.}$$

It was called **mystery number** in [41] and has subsequently been called by us in lectures and papers, in particular also in [13], **common randomness capacity.**

The common randomness capacity C_{CR} 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 of length n on a set $\mathcal{K}$ with $|\mathcal{K}| < e^{cn}$ with

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

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

Let $[M] = \{1, 2, \ldots, M\}$, $[M'] = \{1, 2, \ldots, M'\}$ and let $\mathcal{T} = \{T_i : i = 1, \ldots, N\}$ be a family of maps $T_i : [M] \to [M']$ and consider for $i = 1, 2, \ldots, N$ 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* $|\mathcal{T}| = N \geq \exp\{M - c(\epsilon)\sqrt{\log M}\}$, $Q_i(K_i) = 1$ *for* $i = 1, \ldots, N$, *and* $Q_i(K_j) \leq \epsilon \ \forall i \neq j$.

Note In typical applications the common random experiment has range $M = \exp\{C_R n\}$ and uses for its realization the blocklength n, while for the extension to the T_i the blocklength $\sqrt{n}$ suffices.

A further enlightening development concerned what we formulated as a **PRINCIPLE:**

Second order identification capacity equals (first order) common randomness capacity

After [34], [35], and [41] a lot spoke for it and it became a driving dream leading to many results like [47], where the remarkable fact, that a wire-tapper cannot reduce identification capacity, if he cannot prohibit identification for 2 alternatives, and otherwise the identification capacity equals zero, was discovered and proved by arguments, which are by no means simple.

The same paper also started the investigation of identification in the presence of noisy (passive) feedback channels. This is discussed in [13].

Continuing the line of children of the principle there are [30] and striking work on the AVC in [22] and on the arbitrarily varying MAC in [23], [24], and above all for the maximal error concept for the AVC with complete feedback in [25] a determination of the capacity formula, which has a **trichotomy**.

Let's recall that the Ahlswede-dichotomy was for average error and no feedback [6].

What was called "correlation in random codes", originally introduced in the pioneering paper [57], can now be **understood as common randomness**.

Also its elimination in [6] is an early version of what now Computer Scientists call **derandomization**.

Finally, we report on the removal of another heavy stone. Having understood how correlation in random codes, a form of common randomness, helps the communicators for AVC a next question is how a Slepian/Wolf type correlated source (U^n, V^n) [102] helps the identification for a DMC W, when the sender knows U^n and the receiver knows V^n. Well, the principle says that it should be equivalent to asking how much common randomness can the communicators extract from (U^n, V^n), if they are assisted by the DMC W with capacity $C_{Sh}(W)$.

Now just notice that the case $C_{Sh}(W) = 0$ leads to the problem of finding what I. Csiszar asked for, and according to [115] also D. Slepian, and named **Common Information**. It was determined by P. Gács and J. Körner [73]. As expressed in their title the question was to know how this common information relates to Shannon's **mutual information**, in particular whether they are equal.

As we know the quantities are far apart, and under natural conditions, $C_{GK}(U,V)$ equals zero and it only depends on the positions of positivity of the joint distribution P_{UV}.

This got A. Wyner started, who believed that the quantity $C_W(U,V)$ he introduced was the right notion of common information. For one thing it does depend on the actual values of P_{XY}. On the other hand it satisfies $C_W(U,V) \geq I(U \wedge V)$ and is therefore rather big. The authors of [40] gave a critical analysis about the problems at hand.

By the foregoing it is clear that the common randomness capacity of R. Ahlswede and V. Balakirsky, say $C^W_{AB}(U,V)$, equals $C_{GK}(U,V)$, if $C_{Sh}(W) = 0$. However, if $C_{Sh}(W) > 0$ $C^W_{AB}(U,V)$ nicely depends on the actual value of P_{UV}. Furthermore, $C_{GK}(U,V)$, which was always considered to be somewhat outside Information Theory proper, turns out to be a common randomness capacity. The proof of the characterization of $C^W_{AB}(U,V)$ is a natural extension of the one in case $C_{Sh}(W) = 0$ given in [40].

More importantly we feel that the analysis and discussion in [40] are still of interest today.

The first systematic investigation of common randomness started in [31] and was continued after ideas had matured with [32], in particular, with a revival of another old friend: balanced coloring for hypergraphs ([7], [8]).

Very remarkable work has been done since then by Csiszár and Narayan ([64], [65]), and we are particular intrigued by the work of Venkatesan and Anantharam [105], [106].

In conclusion of the subject, we mention that common randomness and entanglement go into the center of Quantum Information Theory. But there according to [111] already for simple channels identification and common randomness can be far apart.

The exploration of new concepts, ideas and models does not end at the discovery of identification. It actually was a starting point for them. We mentioned already that in [13] more general communication systems were introduced and studied.

We have explained the role of common randomness for identification (The Principle!).

In the absence of feedback, one possibility to achieve the maximal possible rate of such a common random experiment is that the sender performs a uniform random experiment and transmits the result to the receiver using an ordinary transmission code. If noiseless feedback is available, the sender sends letters in such a way, that the entropy of the channel output (which he gets to know by the feedback channel) is maximized, where he can either use a deterministic or randomized input strategy, depending on the kind of code he may use. This interpretation proved to be the right one also for other kind of channels like the multiple access channel (see [41]).

State of knowledge about Capacity Regions

For identification the letter D indicates a deterministic encoding and its absence refers to randomized encoding. (For a refined analysis with maximal versus average error probability we refer to [6].)

	Transmission	Identification
DMC	X	X [34]
MAC	X	X [13], [103]
BC	?	X [13]
TWC	?	?
	With Feedback Transmission	Identification
DMC	X	X [35]
MAC	?	X D [41]
MAC	?	?
BC	?	? D
BC	?	X [41]
TWC	?	X D [41]

There are amazing dualities between transmission and identification. For instance concerning feedback there is **a rather unified theory of Multi-user identification with feedback—with constructive solutions,** whereas for transmission with feedback most capacity regions are unknown. Furthermore using randomness mystery numbers regions are known for the BC, but not for the MAC, whereas, as is well known, for transmission capacity regions the situation is reversed.

Actually common randomness corresponds to the key space in cyptography (see [53]). There are many open entropy characterisation problems: for instance to **calculate** the maximal output entropy $\sup_n \max \frac{1}{n} H(Z^n)$ of a MAC, if $X_{n+1} = f_{n+1}(Z^n)$, $Y_{n+1} = g_{n+1}(Z^n)$ define the encoding processes.

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

One of the observations was that random experiments, to whom the communicators have access, essentially influence the value of the identification capacity C_I. Actually, if sender and receiver have a common random capacity C_R then by the Transformator Lemma always

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

For many channels, in particular for channels with feedback, 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 4 $C_I = 1, C_R = 0$. (Fundamental)

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

Proposition. *Let $\mathcal{Z}$ be a finite set and let $\lambda \in (0, 1/2)$ be given. For $\varepsilon < (2^{2/\lambda} + 1)^{-1}$ a family $A_1, \ldots, A_N$ 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 $\lambda \log(\frac{1}{3} - 1) > 2$ 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 $(W^n)^\infty$, where the input alphabets $\mathcal{X}_t = \{1, 2, \ldots, N(t, \lambda)\}$ are increasing, $\mathcal{X}^n = \prod_{t=1}^n \mathcal{X}_t$ 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

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

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

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

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

and

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

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 $C_R = 0$.

In [86] Kleinewächter presents a counterexample for the other direction. For given real numbers C_{ID} and C_{CR} with $0 < C_{ID} < C_{CR}$, he constructed a discrete channel with memory and noiseless passive feedback with identification capacity C_{ID} and common randomness capacity C_{CR}. This channel is constructed in such a way that it can be used in two ways. In one respect, the channel is good for the generation of common randomness, in the other it is suitable for identification.

IV "Consequences" for Secrecy Systems

Characterisation of the capacity region for the BC for identification

We need the direct part of the ABC Coding Theorem for transmission ([59], [109], [87]).

Here, there are separate messages for decoder $\mathcal{Y}$ (resp. $\mathcal{Z}$) and common messages for both decoders.

Achievable are (with maximal errors)

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

resp.

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

This is our surprising result.

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

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

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

Remark $\mathcal{B}$ gives also the achievable pairs of first order rates for common randomness. Proof goes via identification!

Remark The theorem 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.

Furthermore, 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.

Transmission, identification and common randomness capacities for wire-tape channels with secure feedback from the decoder

Recall that wire-tap channels were introduced by A. D. Wyner [114] and were generalized by I. Csiszár and J. Körner [62]. Its identification capacity was determined by R. Ahlswede and Z. Zhang in [47].

Now 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!

V Analysis of a specific model: K–identification

A relation to standard identification

For reasons, which become apparent soon, we assume K to grow

exponentially in the blocklength n, that is,

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

where κ is a first order rate.

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

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

where R is a second order rate.

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

Theorem *For every DMC the set $\mathcal{K}$ of all achievable rate pairs satisfies*

(i) $\{(R, \kappa) : 0 \leq R, \kappa, R + 2\kappa \leq C_{sh}\} \subset \mathcal{K}$

(ii) $\{(R, \kappa) : 0 \leq R, \kappa, R + \kappa \leq C_{sh}\} \supset \mathcal{K}$

(iii) *For a noiseless DMC there is equality in* (i).

In general?

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 by Kautz/Singleton.

VI Extensions to Classical/Quantum Channels

There has been great progress in recent years with fruitful exchanges between Information Theory and Physics.

Since most readers are not familiar with this we just **give classical methods which extend or have analoga.**

We prove in [20] that the average error capacity C_q of a quantum arbitrarily varying channel (QAVC) equals 0 or else the random code capacity $\bar{C}$ (Ahlswede's dichotomy). We also establish a necessary and sufficient condition for $C_q > 0$.

It is interesting to note, that in our proof of this we essentially use the **elimination technique** (an early candidate of what is now called derandomization in Computer Sciences) from [18]. There a necessary and sufficient condition for positivity of the capacity was given, if the set of transmission matrices is row-convex closed—that is under a practically satisfactory assumption of robustness. The mathematical problem of characterizing positivity without this assumption in terms of symmetrizability was started in [69] and completely solved in [63] with a non-standard decoding rule and without use of the elimination technique.

On the other hand in the present quantum case we have not found a suitable decoding rule and follow the elemination technique. Analogously the positivity problem for the QAVC can be settled by reducing it to a **related classical AVC** to which then the result of [63] can be applied.

We emphasize that the very hard maximal error capacity problem for AVC (including Shannon's zero error capacity problem as special case) is based on a more realistic communication model. It was solved for a nice class of channels in [9], where for the first time in the area of AVC a non-standard decoding rule was used. Extension to QAVC constitute a challenging problem!

A hypergraph covering lemma useful for deriving capacity results

— in the theory of identification

— in the theory of common randomness

Lemma *Let $\Gamma = (\mathcal{V}, \mathcal{E})$ be a hypergraph, with a measure Q_E on each edge E, such that $Q_E(v) \leq \eta$ for all $E, v \in E$. For a probability distribution P on $\mathcal{E}$ define*

$$Q = \sum_{E \in \mathcal{E}} P(E) Q_E,$$

and fix $\epsilon, \tau > 0$. Then there exist vertices $\mathcal{V}_0 \subset \mathcal{V}$ and edges $E_1, \ldots E_L \in \mathcal{E}$ such that with

$$\bar{Q} = \frac{1}{L} \sum_{i=1}^{L} Q_{E_i}$$

the following holds:

$$Q(\mathcal{V}_0) \leq \tau, \ \forall v \in \mathcal{V} \backslash \mathcal{V}_0 \ (1 - \epsilon) Q(v) \leq \bar{Q}(v) \leq (1 + \epsilon) Q(v),$$

$$L \leq 1 + \eta |\mathcal{V}| \frac{2 \ln 2 \log(2|\mathcal{V}|)}{\epsilon^2 \tau}.$$

Remark The lemma applies also to identification for (classical) quantum channels (Ahlswede/Winter [43]).

The blowing up technique

We define the k–Hamming–neighbourhood $\Gamma^k B$ of a set $B \subset \mathcal{Y}^n$ as

$$\Gamma^k B \triangleq \{y^n \in \mathcal{Y}^n : d(y^n, y'^n) \leq k \text{ for some } y'^n \in B\}$$

where $d(y^n, y'^n) \triangleq (\{t : 1 \leq t \leq n, y'_t \neq y_t\})$.

Blowing up Lemma (Ahlswede/Gács/Körner, 1976, [36])

For any DMC W there is a constant $c(W)$: $\forall x^n \in \mathcal{X}^n, B \subset \mathcal{Y}^n$ $W^n(\Gamma^k B | x^n) \geq \Phi(\Phi^{-1}(W^n(B|x^n))) + n^{-1/2}(k-1)c$ if $\Phi(t) = \int_{-\infty}^{t} (2\pi)^{-1/2} e^{-u^2/2} du$.

Remark We have no quantum version!

A wringing technique

useful for

— strong converse for multi-user channels

— converses for multiple-descriptions in rate-distortion theory

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

(1) $P(x^n) \le (1+c)Q(x^n)$ *for all* $x^n \in \mathcal{X}^n$,

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

(2) $P(x_{t_1}|\bar{x}_{t_1}, \dots, \bar{x}_{t_k}) \le \max((1+\gamma)Q(x_i|\bar{x}_{t_1}, \dots, \bar{x}_{t_k}), \epsilon)$ *for all* $x_t \in \mathcal{X}$ *and all* $t = 1, 2, \dots, n$ *and*

(3) $P(\bar{x}_{t_1}, \dots, \bar{x}_{t_k}) \ge \epsilon^k$

Remark Presently only method to prove strong converse for transmission for (classical) **quantum** multiple-access channel (Ahlswede/Cai [28]).

VII Source Coding for Identification: a Discovery of Identification Entropy

Shannon's Channel Coding Theorem for Transmission is paralleled by a Channel Coding Theorem for Identification. We introduced noiseless source coding for identification and suggested the study of several performance measures.

Interesting observations were made already for uniform sources $P_N = (\frac{1}{N}, \dots, \frac{1}{N})$, for which the worst case expected number of checkings $L(P_N)$ (defined in the next column) is approximately 2. Actually it has been shown that

$$\lim_{N\to\infty} L(P_N) = 2.$$

Recall that in channel coding going from transmission to identification leads from an **exponentially** growing number of manageable messages to **double exponentially** many.

Now in source coding roughly speaking the range of average code lengths for data compression is the interval $[0, \infty)$ and it is $[0, 2)$ for an average expected length of optimal identification procedures.

Note that no randomization has to be used here.

A discovery is an identification entropy, namely the functional

$$H_I(P) = 2\left(1 - \sum_{u=1}^{N} P_u^2\right)$$

for the source $(\mathcal{U}, P)$, where $\mathcal{U} = \{1, 2, \dots, N\}$ and $P = (P_1, \dots, P_N)$ is a probability distribution.

Its operational significance in identification source coding is similar to that of classical entropy $H(P)$ in noiseless coding of data: it serves as a good bound.

Noiseless identification for sources and basic concept of performance

For the source $(\mathcal{U}, P)$ let $C = \{c_1, \dots, c_N\}$ be a binary prefix code (PC) with $\|c_u\|$ as length of c_u.

Introduce the RV U with $\text{Prob}(U = u) = P_u$ for $u \in \mathcal{U}$ and the RV C with $C = c_u = (c_{u1}, c_{u2}, \dots, c_{u\|u\|})$ 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 letters occur or when $C = c_u$. What is the expected number $L_C(P, u)$ of checkings?

Related quantities are

$$L_C = \max_{1\le u\le N} L_C(P, u),$$

that is, the expected number of ckeckings for a person in the **worst case**, if code C is used,

$$L(P) = \min_C L_C(P),$$

the expected number of checkings in the worst case for the best code, and finally, if **users are chosen by a RV** V independent of U and defined by $\text{Prob}(V = v) = Q_v$ for $v \in \mathcal{V} = \mathcal{U}$, we consider

$$L_C(P, Q) = \sum_{v\in\mathcal{U}} Q_v L_C(P, v)$$

the average number of expected checkings, if code C is used, and also

$$L(P, Q) = \min_C L_C(P, Q)$$

the average number of expected checkings for a best code.

A natural special case is the mean number of expected checkings

$$\bar{L}_C(P) = \sum_{u=1}^{N} \frac{1}{N} L_C(P,u),$$

which equals $L_C(P,Q)$ for $Q = (\frac{1}{N},\ldots,\frac{1}{N})$, and

$$\bar{L}(P) = \min_C \bar{L}_C(P)$$

Another special case of some "intensive appeal" is the case $Q = P$. Here we write

$$L(P,P) = \min_C L_C(P,P).$$

It is known that Huffman codes minimize the expected code length for PC.

This is not the case for $L(P)$ and the other quantities in identification. It was noticed already in [13], [17] that a construction of code trees balancing probabilities like in the Shannon–Fano code is often better. In fact the Theorem of [17] establishes that $L(P) < 3$ for every $P = (P_1,\ldots,P_N)$!

Still it is also interesting to see how well Huffman codes do with respect to identification, because of their classical optimality property.

Examples for Huffman codes

We start with the uniform distribution

$$P^N = (P_1,\ldots,P_N) = \left(\frac{1}{N},\ldots,\frac{1}{N}\right),$$

$2^n \le N < 2^{n+1}$.

Then $2^{n+1} - N$ codewords have the length n and the other $2N - 2^{n+1}$ other codewords have the length $n+1$ in any Huffman code. We call the $N - 2^n$ nodes of length n of the code tree, which are extended up to the length $n+1$ **extended nodes**.

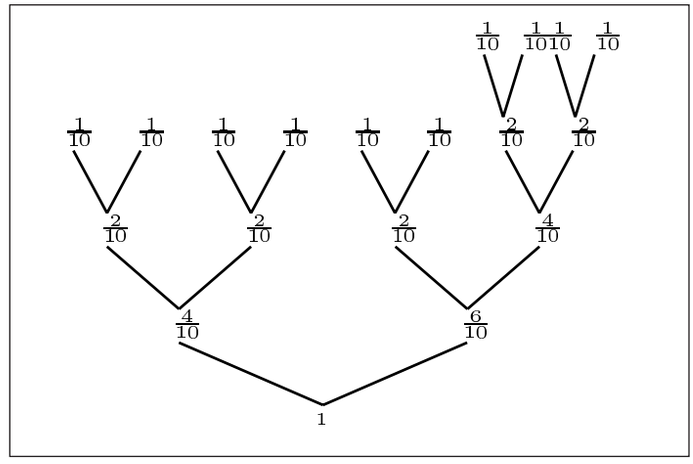
All Huffman codes for this uniform distribution differ only by the positions of the $N - 2^n$ extended nodes in the set of 2^n nodes of length n.

The average codeword length (for transmission) does not depend on the choice of the extended nodes.

However, the choice influences the performance criteria for identification!

Example 5 $N = 10$. There are $\binom{2^3}{10-2^3} = 28$ Huffman codes.

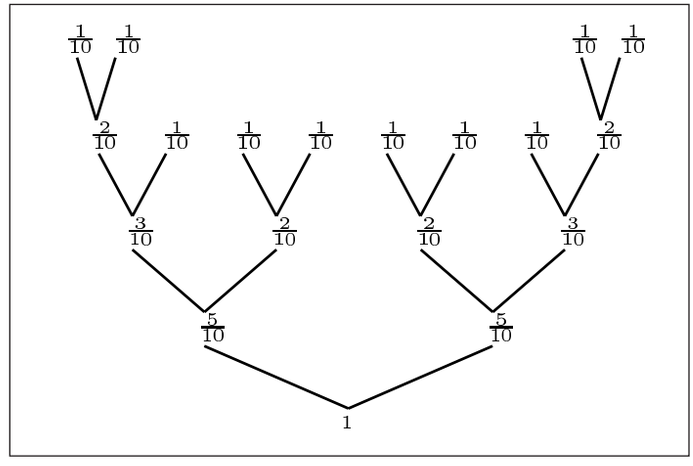
The 4 worst Huffman codes are maximally unbalanced.



Here

$$L_{\mathcal{C}}(P) = 1 + 0.6 + 0.4 + 0.2 = 2.2$$
$$L_{\mathcal{C}}(P,P) = \frac{1}{10}[1.6\cdot 4 + 1.8\cdot 2 + 2.2\cdot 4] = 1.880.$$

One of the 16 best Huffman codes



Here

$$L_{\mathcal{C}}(P) = L_{\mathcal{C}}(\tilde{c}) = 1 + 0.5 + 0.3 + 0.2 = 2.000$$
$$L_{\mathcal{C}}(P,P) = \frac{1}{5}(1.7\cdot 2 + 1.8\cdot 1 + 2.0\cdot 2) = 1.840$$

Remark Notice that Shannon's data compression gives $H(P) + 1 = \log 10 + 1 > \sum_{n=1}^{10} P_n \|C_n\| = \frac{1}{10}\cdot 3\cdot 6 + \frac{1}{10}\cdot 4\cdot 4 = 3\frac{2}{5} \ge \log 10$

Theorem *For every source* $(\mathcal{U}, P^N)$

$$L(P^N) \geq L(P^N, P^N) \geq H_l(P^N).$$

Theorem *For* $P^N = (P_1, \ldots, P_N)$

$$\bar{L}(P^N) \leq 2\left(1 - \frac{1}{N^2}\right).$$

Theorem *For* $P^N = (2^{-\ell_1}, \ldots, 2^{-\ell_N})$ *with 2-powers as probabilities*

$$L(P^N, P^N) = H_l(P^N).$$

Theorem

$$L(P^N, P^N) \leq 2\left(1 - \sum_u \left(\sum_{s=1}^{\alpha(u)} P_{us}^2\right)\right) \leq 2\left(1 - \frac{1}{2}\sum_u P_u^2\right).$$

For $P_u = \frac{1}{N}(u \in \mathcal{U})$ this gives the upper bound $2(1 - \frac{1}{2}N)$, which is better than the bound $2(1 - \frac{1}{N^2})$ for uniform distributions.

Finally we derive

Corollary

$$L(P^N, P^N) \leq H_l(P^N) + \max_{1 \leq u \leq N} P_u.$$

It shows the lower bound of $L(P^n, P^N)$ by $H_l(P^N)$ and this upper bound are close.

Further Remarks

1. Our results can be extended to q-ary alphabets, for which then identification entropy has the form

$$H_{l,q}(P) = \frac{q}{q-1}\left(1 - \sum_{u=1}^{N} P_u^2\right).$$

2. Tsallis generalized Boltzmann's entropy

$$H(P) = -k\sum P_u \ln P_u$$

to

$$S_\alpha(P) = k\frac{1}{\alpha - 1}\left(1 - \sum_{u=1}^{N} P_u^\alpha\right)$$

for any real $\alpha \neq 1$.

Clearly $\lim_{\alpha \to 1} S_\alpha(P) = H(P) = S_1(P)$, say.

One readily verifies that for product-distributions $P \times Q$ for independent random variables

$$S_\alpha(P \times Q) = S_\alpha(P) + S_\alpha(Q) - \frac{(\alpha - 1)}{k} S_\alpha(P) S_\alpha(Q).$$

Since in all cases $S_\alpha \geq 0, \alpha < 1, \alpha = 1$ and $\alpha > 1$ respectively correspond to **superadditivity, additivity and subadditivity** (also called for the purposes in statistical physics **superextensitivity**, **extensitivity**, and **subextensitivity**).

We have been told by several experts in physics that the operational significance of the quantities S_α (for $\alpha \neq 1$) in statistical physics seems not to be undisputed.

In contrast we **have demonstrated the significance of identification entropy**, which is formally close, but essentially different for two reasons: always $\alpha = 2$ and $k = \frac{q}{q-1}$ is uniquely determined and **depends on the alphabet size** q!

3. In [26] we have discussed the coding theoretical meanings of the factors $\frac{q}{q-1}$ and $(1 - \sum_{u=1}^{N} P_u^2)$.

In particular we have the

Theorem *For a DMS* $(U^n, V^n)_{n=1}^\infty$ *with generic distribution* $P_{UV} = PQ$, *i.e. the generic random variables* U *and* V *are independent and* $P_U = P$, $P_V = Q$

$$\lim_{n \to \infty} L(P^n, Q^n) = \begin{cases} 1 & \text{if } P \neq Q \\ \frac{q}{q-1} & \text{if } P = Q. \end{cases}$$

B. Combinatorial Models

That Combinatorics and Information Sciences often come together is no surprise, because they were born as twins (Leibniz in Ars Combinatoria gives credit to Raimundus Lullus from Catalania, who wanted to create a formal language).

VIII Updating Memories with Cost Constraints: Optimal Anticodes

In the example

N	O	T	S	O	C	L	E	A	R

N	O	W	C	L	E	A	R	E	R

$d = 7$ letters have to be changed for an updating, where d is the Hamming distance, measuring the cost.

How many messages can be updated into each other, if cost $\leq c$? This is equivalent to the diametric problem in Hamming spaces. It was solved in [39].

For a Hamming space $(\mathcal{X}_q^n, d_H)$, the set of n–length words over the alphabet $\mathcal{X}_q = \{0, 1, \ldots, q-1\}$ endowed with the distance d_H, we determine the maximal cardinality of subsets with a prescribed diameter d or, in another language, anticodes with distance d. We refer to the result as Diametric Theorem.

In a sense anticodes are dual to codes, which have a prescribed

lower bound on the pairwise distance. It is a hopeless task to determine their maximal sizes exactly.

We find it remarkable that the Diametric Theorem (for arbitrary q) can be derived from the Complete Intersection Theorem, which can be viewed as a Diametric Theorem (for $q = 2$) in the constant weight case, where all n–length words considered have exactly k ones.

$\mathbb{N}$ denotes the set of positive integers and for $i, j \in \mathbb{N}$, $i < j$, the set $\{i, i+1, \ldots, j\}$ is abbreviated as $[i, j]$. Moreover, for $[1, j]$ we also write $[j]$. For $k, n \in \mathbb{N}$, $k \leq n$, we set

$$2^{[n]} = \{F : F \subset [1, n]\} \quad \text{and} \quad \binom{[n]}{k} = \{F \in 2^{[n]} : |F| = k\}.$$

A system of sets $\mathcal{A} \subset 2^{[n]}$ is called t–intersecting, if

$$|A_1 \cap A_2| \geq t \quad \text{for all} \quad A_1, A_2 \in \mathcal{A},$$

and $I(n, t)$ denotes the set of all such systems.

Moreover, we define $I(n, k, t) = \{\mathcal{A} \in I(n, t) : \mathcal{A} \subset \binom{[n]}{k}\}$.

The investigation of the function $M(n, k, t) = \max_{\mathcal{A} \in I(n,k,t)} |\mathcal{A}|$, $1 \leq t \leq k \leq n$, and the structure of maximal systems was one of the oldest problems in combinatorial extremal theory and was initiated by Erdös, Ko, and Rado.

They proved already in the year 1938 the following theorem, which was published only in 1961 [68].

Theorem *For $1 \leq t \leq k$ and $n \geq n_0(k, t)$ (suitable)*

$$M(n, k, t) = \binom{n-t}{k-t}.$$

Clearly, the system

$$A(n, k, t) = \left\{A \in \binom{[n]}{k} : [1, t] \subset A\right\}$$

is t–intersecting, has cardinality $\binom{n-t}{k-t}$, and is therefore optimal for $n \geq n_0(k, t)$.

The smallest $n_0(k, t)$, for which this is the case, has been determined by Frankl 1978 in [72] for $t \geq 15$ and subsequently 1984 in [110] for all t:

$$n_0(k, t) = (k - t + 1)(t + 1).$$

We have settled all the remaining cases: $n < (k - t + 1)(t + 1)$.

Complete Intersection Theorem [38] *Define $\mathcal{F}_i = \{F \in \binom{[n]}{k} : |F \cap [1, t + 2i]| \geq t + i\}$ for $0 \leq i \leq \frac{n-t}{2}$. For $1 \leq t \leq k \leq n$ with*

(i) $(k - t + 1)(2 + \frac{t-1}{r+1}) < n < (k - t + 1)(2 + \frac{t-1}{r})$ *for some* $r \in \mathbb{N} \cup \{0\}$

we have

$$M(n, k, t) = |\mathcal{F}_r|$$

and $\mathcal{F}_r$ is—up to permutations—the unique optimum. By convention $\frac{t-1}{r} = \infty$ for $r = 0$.

(ii) $(k - t + 1)(2 + \frac{t-1}{r+1}) = n$ *for* $r \in \mathbb{N} \cup \{0\}$

we have

$$M(n, k, t) = |\mathcal{F}_r| = |\mathcal{F}_{r+1}|$$

and an optimal system equals—up to permutations—either $\mathcal{F}_r$ or $\mathcal{F}_{r+1}$.

Remark In particular this proves the so called $4m - Conjecture$ (Erdös, Ko, Rado 1938, [68])

$$M(4m, 2m, 2) = \left|\left\{F \in \binom{[4m]}{2m} : F \cap [1, 2m] \geq m + 1\right\}\right|.$$

Remarks Our most recent results on intersecting families can be found in [15], which contains many further references, and our most advanced method is the shifting technique of [14]. We also draw attention to the local-global principle [21] which plays a key role in the recent book [79].

For non-constant weight anticodes the complete solution is this.

Diametric Theorem [39] *For $q \geq 2$ let $r \in \{0\} \cup \mathbb{N}$ be the largest integer such that*

$$n - d + 2r < \min\left\{n + 1, n - d + 2\frac{n - d - 1}{q - 2}\right\},$$

then

$\max\{|\mathcal{A}| : \mathcal{A} \subset \mathcal{X}_q^n, diam(\mathcal{A}) \leq d\} = |\{a^n \in \mathcal{X}_\alpha^n : \sum_{s=1}^{n-d+2r} w_H(a_s) \leq r\}|$.

(By convention $\frac{n-d-1}{q-2} = \infty$ for $q = 2$.)

Another diametric theorem in Hamming spaces concerns optimal group anticodes [12].

A report on **Extremal Problems in Number Theory** and especially also in **Combinatorics**, which arose in Information Theory, can be found in [10], [19] and [67].

IX Information Flows in Networks

We continue now with the subject whose origin is generally attributed to [29]. The founder of Information Theory Claude E. Shannon, who set the standards for efficient transmission of channels with noise by introducing the idea of coding also wrote together with Peter Elias and Amiel Feinstein a basic paper on networks [99] discussing algorithmic aspects of the Min Cut—Max Flow Theorem [70], saying that for flows of physical commodities like electric currents or water, satisfying Kirchhoff's laws, the maximal flow equals the minimal cut.

With the stormy development of Computer Science there is an ever increasing demand for designing and optimizing information flows over networks—for instance in the internet.

Data, that is strings of symbols, are to be send from sources $s_1, \ldots, s_n$ to their destinations, sets of node sinks $D_1, \ldots, D_n$.

Computer scientist quickly realized that it is beneficial to copy incoming strings at processors sitting at nodes of the network and to forward copies to adjacent nodes. This task is called multicasting.

However, quite surprisingly **they did not consider coding**, which means here to produce not only copies, but, more generally, new output strings as deterministic functions of incoming strings.

A **Min-Max-Theorem was discovered and proved for information flows** by Ahlswede, Cai, Li, and Yeung in [29].

Its statement can be simply explained. For one source only, that is $n = 1$, in the notation above, and $D_1 = \{d_{11}, d_{12}, \ldots, d_{1t}\}$ let F_{1j} denote the max-flow value, which can go for any commodity like water in case of Ford/Fulkerson from s_i to d_{1i}. The same water cannot go to several sinks. However, the amount of $\min_{1 \le j \le t} F_{1j}$ bits can go **simultaneously** to $d_{11}, d_{12}, \ldots$ and d_{1t}. Obviously, this is best possible. It has been referred to as ACLY-Min-Max-Theorem. To the individual F_{1j} Ford/Fulkerson's Min-Cut-Max-Flow Theorem applies.

It is very important that in the starting model there is no noise and it is amazing for how long Computer Scientists did the inferior multi-casting allowing only copies. It is perhaps surprising that Shannon seems not to have realized the consequences of the basic difference between classical and information flows. We substantiate this by citing from his Kyoto lecture [98].

> *"A basic idea in information theory is that information can be treated very much like a physical quantity, such as mass or energy. For example, an information source is like a lumber mill producing lumber at a certain point. The channel might correspond to a conveyor system for transporting the lumber to a second point. In such a situation there are two important quantities: the rate R (in cubic feet per second) at which lumber is produced at the mill and the capacity C (in cubic feet per second) of the conveyor. These two quantities determine whether or not the conveyor system will be adequate for the lumber mill. If the rate of production R is greater than the conveyor capacity C, it will certainly be impossible to transport the full output of the mill; there will not be sufficient space available. If R is less than or equal to C, it may or may not be possible, depending on whether the lumber can be packed efficiently in the conveyor. Suppose, however, that there is a sawmill at the source. This correspond in the analogy to the encoder or transmitter. Then the lumber can be cut into small pieces in such a way as to fill out the available capacity of the conveyor with 100 percent efficiency. Naturally, in this case a carpenter would be provided at the receiving point to fasten the pieces back together in their original form before passing them on to the consumer.*
>
> *If this analogy is sound, it should be possible to set up a measure R, in suitable units, giving the rate at which information is produced by a given information source, and a second measure C that determines the capacity of a channel for transmitting information. Furthermore, the analogy would suggest that by a suitable coding or modulation system, the information can be transmitted over the channel if and only if the rate of production R is not greater than the capacity C. A key result of information theory is that it is indeed possible to set up measures R and C having this property."*

Network flows with **more than one source** are much harder to analyze and lead to a wealth of old and new combinatorial extremal problems.

Even nicely characterized classes of **error correcting codes** come up as being isomorphic to a complete set of solutions of flow problems **without errors**!

Also **optimal anticodes** (see theorem above) arise in such a role!

On the classical side for instance orthogonal **Latin Squares** arise.

It is known that classical network flows have many connections to combinatorial extremal problems like Baranyai's factorization theorem [52] or especially for matching problems. Information flows promise more such connections as for example in [113]. There may be a great challenge not only coming to **Combinatorics** but also to **Algebraic Geometry** and its present foundations.

We draw attention to the chapter on Network Coding in [19], pages 858–897.

X Localized Errors

A famous problem in coding theory consists in finding good bounds for the maximal size, say $N(n, t, q)$, of a t-error correcting code over a q-ary alphabet $Q = \{0, 1, \ldots, q-1\}$ with blocklength n.

This code concept is suited for communication over a q-ary channel with input and output alphabets Q, where a word of length n sent by the encoder is changed by the channel in atmost t letters. Here neither the encoder nor the decoder knows in advance where the errors, that is changes of letters, occur.

It is convenient to use the notation relative error $\tau = t/n$ and rate $R = n^{-1} \log M$.

The Hamming bound is an upper bound on it.

$$H_q(\tau) = \begin{cases} 1 - h_q(\tau) - \tau \log_q(q-1) & \text{if } 0 \le \tau \le \frac{q-1}{q} \\ 0 & \text{if } \frac{q-1}{q} < \tau \le 1. \end{cases}$$

We turn now to another model. Suppose that the **encoder**, who wants to encode message $i \in \mathcal{M} = \{1, 2, \dots, M\}$, knows the t-element set $E \subset [n] = \{1, \dots, n\}$ of positions, in which only errors may occur. He then can make the codeword presenting i dependent on $E \in \mathcal{E}_t = \binom{[n]}{t}$, the family of t-element subsets of $[n]$. We call them "a priori error pattern". A family $\{u_i(E) : 1 \le i \le M,\ E \in \mathcal{E}_t\}$ of q-ary vectors with n components is an $(M, n, t, q)_l$ code (for localized errors), if for all $E, E' \in \mathcal{E}_t$ and all q-ary vectors $e \in V(E) = \{e = (e_1, \dots, e_n) : e_j = 0 \text{ for } j \notin E\}$ and $e' \in V(E')$

$$u_i(E) \oplus e \neq u_{i'}(E') \oplus e' \quad \text{for } i \neq i',$$

where $\oplus$ is the addition modulo q.

We denote the capacity error function, that is the supremum of the rates achievable for τ and all large n, by C_q^l. It was determined by Bassalygo/Gelfand/Pinsker [54] for the binary case to equal $H_2(\tau)$. For general q the best known result is

Theorem

(i) $C_q^l(\tau) \le H_q(\tau)$, for $0 \le \tau \le \frac{1}{2}$.

(ii) $C_q^l(\tau) = H_q(\tau)$, for $0 \le \tau < \frac{1}{2} - \frac{q-2}{2q(2q-3)}$.

Competing Ideas:

Ahlswede: With increase of q the Hamming space should become more flexible for packing and the Hamming bound should be tight for $0 \le \tau \le \frac{1}{2}$.

Pinsker: Knowing the a-priori error pattern E gives less **protocol** information if q increases.

Who wins?

XI Search

After we wrote with I. Wegener one of the first books on search in 1978, the subject has grown terrifically. **Still progress is possible on basic questions**.

For input alphabet $\mathcal{X} = Q$ and output alphabet $\mathcal{Y} = Q$ let $M_f(n, t, q)$ be the maximal size of a t-error correcting code over a q-ary alphabet with block length n in the presence of noiseless feedback, that means having sent letters $x_1, \dots, x_{j-1} \in \mathcal{X}$ the encoder knows the letters $y_1, \dots, y_{j-1} \in \mathcal{Y}$ received before he sends the next letter $x_j (j = 1, 2, \dots, n)$. Define the relative error $\tau = t/n$, the rate $R = n^{-1} \log M$, and the capacity error function $C_q^f(\tau)$ as the supremum of the rates achievable for τ and all large n.

Theorem ([55], [116])

$$C_2^f(\tau) = \begin{cases} h_2(\tau) & \text{if } 0 \le \tau \le \tau_t \\ (-3R_0\tau) + R_0 & \text{if } \tau_t \le \tau \le \frac{1}{3}, \end{cases}$$

where $R_0 = \log_2(\frac{1+\sqrt{5}}{2})$ *and* $\tau_t = (3+\sqrt{5})^{-1}$.

Theorem ([33]) *Let* $q \ge 3$

(i)

$$C_q^f(\tau) \begin{cases} \le H_q(\tau) & \text{if } 0 \le \tau \le \frac{1}{q} \\ = (1-2\tau)\log_q(q-1) & \text{if } \frac{1}{q} \le \tau \le \frac{1}{2} \\ = 0 & \text{if } \frac{1}{2} \le \tau \le 1\,. \end{cases}$$

(ii) *The rate function obtained by the* r*-rubber method is a tangent to* $H_q(\tau)$ *going through* $(\frac{1}{r+1}, 0)$.

The rubber method

Let $b : \mathcal{M} \to \{1, 2, \dots, q-1\}^{n-2t}$ be a bijection between the messages and the used sequences.

The "0" is used for error correction only.

Given $i \in \mathcal{M}$ the sender chooses $b(i) = (x_1, x_2, \dots, x_{n-2t}) \in \{1, 2, \dots, q-1\}^{n-2t}$ as a **skeleton for encoding**, which finally will be known to the receiver.

For all positions $i \le n$ not needed dummies $x_i = 1$ are defined to fill the block length n.

Transmission algorithm: The sender sends x_1, x_2 until the first error occurs, say in position p with x_p sent.

If a **standard error** occurs $(x_p \to y_p \in \{1, 2, \dots, q-1\})$, the sender transmits, with smallest l possible, $2l+1$ times 0 until the decoder received $l+1$ zeros. Then he transmits at the next step x_p, again, and continues the algorithm.

If a **towards zero error** occurs $(x_p \to y_p = 0)$, the sender decreases p by one (if it is bigger than 1) and continues (transmits at the next step x_p).

Decoding algorithm: The receiver just regards the "0" as a protocol symbol—he erases it by a rubber, **who in addition erases the previous symbol.**

r-rubber method: Let the skeleton defined by $\{x^{n-(r+1)t} \in \{0, 1, \dots, q-1\}^{n-(r+1)t}$: the sequence contains $\le r-1$ consecutive zeros $\}$ and the protocol string defined as r consecutive zeros.

Relation between Berlekamp's strategies and r-rubber method

— For $q = 2$ and $r > 1$ the r-rubber strategies have the same rate as Berlekamp's strategies (tangents to the Hamming bound

going through $(\frac{1}{r+1}, 0)$).

— Especially for $q = 2$ and $r = 2$ we get Berlekamp's tangent bound.

— More general we get for $q > 2$ and $r \geq 1$ tangents to the Hamming bound going through $(\frac{1}{r+1}, 0)$.

XII Combi-probabilistic Models: Coloring Hypergraphs did a Problem by Gallager

Slepian/Wolf Model 1973 ([102])

For a DMCS $((X^n, Y^n))_{n=1}^{\infty}$ with alphabets $\mathcal{X}$ and $\mathcal{Y}$ for encoding $f : \mathcal{Y}^n \to \mathbb{N}$ and decoding $g : \mathcal{X}^n \times \mathbb{N} \to \mathcal{X}^n \times \mathcal{Y}^n$ with $Prob(g(X^n, f(Y^n)) = (X^n, Y^n)) \sim 1$ it is true that the optimal rate (f) equals $H(Y|X)$.

Gallager Model 1976 ([75])

For a discrete, memoryless conditional distribution $(\{Y^n(x^n) : x^n \in \mathcal{X}^n\})_{n=1}^{\infty}$ (Generic $P_{Y|X}$) with alphabets $\mathcal{X}$ and $\mathcal{Y}$ for encoding $f : \mathcal{Y}^n \to \mathbb{N}$ and decoding $g : \mathcal{X}^n \times \mathbb{N} \to \mathcal{X}^n \times \mathcal{Y}^n$ with $Prob(g(x^n, f(Y^n)) = (x^n, Y^n)) \sim 1 \quad \forall x^n \in \mathcal{X}^n$, we proved that the optimal rate (f) equals $\max_x H(Y|X = x)$.

Here RANDOM SELECTION fails.

Our solution is given already in [1] by a counting argument and in [7] it proceeds by a **combined greedy/random selection**.

C. Further Perspectives

Protocol Information

"Protocol" information we encountered in the Theory of Localized Errors and in the Rubber Method. The subject was started by R.G. Gallager [74] and deserves further investigations.

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

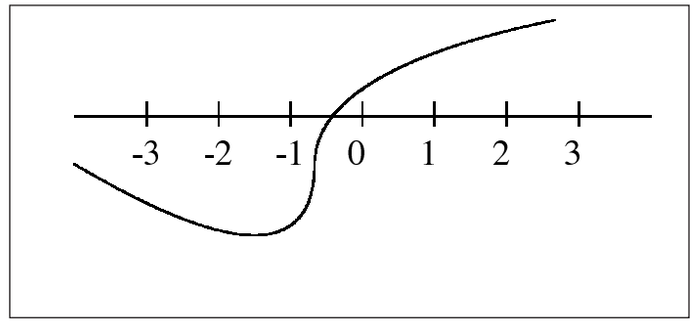
We mention as perhaps one of the most promising directions the study of probabilistic algorithms with identification as *concept of solution and underline its importance by repeating:*

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?!

Example 6

Develop probabilistic algorithms which answer very quickly with high probaility correctly whether a polynomial $P : \mathbb{R} \to \mathbb{R}$ has a root in the interval $[i, i+1]$ or not, for any $i \in \mathbb{N}$.



A new connection between information inequalities and Combinatorial Number Theory: The final form of Tao's inequality relating conditional expectation and conditional mutual information

Recently Terence Tao approached Szemerédi's Regularity Lemma from the perspectives of Probability Theory and of **Information Theory** instead of Graph Theory and found a stronger variant of this lemma, which involves a new parameter.

To pass from an entropy formulation to an expectation formulation he found the following

Lemma. *Let Y, X, and X' be random variables taking values in $\mathcal{Y}$ and $\mathcal{X}$, respectively, where $\mathcal{Y} \subset [-1, 1]$, and with $X' = f(X)$ for a (deterministic) function f. Then we have*

$$\mathbb{E}(|\mathbb{E}(Y|X') - \mathbb{E}(Y|X)|) \leq 2I(X \wedge Y|X')^{\frac{1}{2}}.$$

We show that the constant 2 can be improved to $(2\ln 2)^{\frac{1}{2}}$ and that this is the best possible constant.

Could we ask Shannon's advice !!!

The following last paragraph on page 350 is taken from "Coding theorems for a discrete source with a fidelity criterion", C. Shannon Collected Papers, 325–350.

> *"In a somewhat dual way, evaluating the rate-distortion function $R(D)$ for a source amounts, mathematically, to minimizing a mutual Information under variant of the $q_i(j)$, again with a linear inequality constraint. The solution leads to a function $R(d)$ which is convex downward. Solving this problem corresponds to finding a channel that is just right for the source and allowed distortion level. This duality can be pursued further and is related to the duality between past and future and the notions of control and knowledge.* ***Thus we may have knowledge of the past***

but cannot control it; we may control the future but have no knowledge of it."

The often cited last sentence, which we put here in boldface, has made several thinkers curious.

We sketch below our ideas about creating order involving knowledge of past and future and wonder what Shannon, whom we never met, would think about them. **They are motivated by Clausius' second law of thermodynamics**

"Heat cannot by itself pass from a colder to a hotter body."

He also introduced entropy, for which Boltzmann gave a famous formula.

We quote A. Rényi, Probability Theory, North Holland, Amsterdam, p. 554, 1970, for his opinion about this.

"The quantity $\sum_{k=1}^{n} p_k \log_2 \frac{1}{p_k}$ is frequently called the entropy of the distribution $\mathcal{P} = (p_1, \dots, p_k)$. Indeed, there is a strong connection between the notion of entropy in thermodynamics and the notion of information (or uncertainly). L. Boltzmann was the first to emphasize the probabilistic meaning of the thermodynamical entropy and thus he may be considered as a pioneer of information theory. It would even be proper to call the formula the Boltzmann-Shannon formula. Boltzmann proved that the entropy of a physical system can be considered as a measure of the disorder in the system. In case of a physical system having many degrees of freedom (e.g. perfect gas) the number measuring the disorder of the system measures also the uncertainty concerning the states of the individual particles."

Creating order with simple machines

In [45] and [46] a new field of research, creating order in sequence spaces with simple machines, was introduced. People spend a large amount of time creating order in various circumstances. We contribute to a theory of ordering. In particular we try to understand how much "order" can be created in a "system" under constraints on our "knowledge about the system" and on the "actions we can perform in the system".

We have a box that contains β objects at time t labeled with numbers from $\mathcal{X} = \{0, \dots, \alpha - 1\}$. The state of the box is $s_t = (s_t(1), \dots, s_t(\alpha))$, where $s_t(i)$ denotes the number of balls at time t labeled by i.

Assume now that an arbitrary sequence $x^n = (x_1, \dots, x_n) \in \mathcal{X}^n$ enters the box iteratively. At time t an organizer $\mathcal{O}$ outputs an object y_t and then x_t enters the box. $x^n = (x_1, \dots, x_n)$ is called an input and $y^n = (y_1, \dots, y_n)$ an output sequence. The organizer's behavior must obey the following rules.

Constraints on matter. The organizer can output only objects from the box. At each time t he must output exactly one object.

Constraints on mind. The organizer's strategy depends on

(a) his knowledge about the time t. The cases where $\mathcal{O}$ has a timer and has no timer are denoted by T^+ and T^-, respectively.

(b) his knowledge about the content of the box. O^- indicates that the organizer knows at time t only the state s_t of the box. If he also knows the order of entrance times of the objects, we write O^+.

(c) the passive memory (π, β, φ). At time t the organizer remembers the output letters $y_{t-\pi}, \dots, y_{t-1}$ and can see the incoming letters $x_{t+1}, \dots, x_{t+\varphi}$.

Let $\mathcal{F}_n(\pi, \beta, \varphi, T^-, O^-)$ be the set of all strategies for (T^-, O^-), length n and a given memory (π, β, φ) and $\mathcal{S}$ be the set of all states. A strategy $f_n : \mathcal{X}^n \times \mathcal{S} \to \mathcal{X}^n$ assigns to each pair (x^n, s_1) an output y^n. Denote $\mathcal{Y}(f_n)$ the image of $\mathcal{X}^n \times \mathcal{S}$ under f_n. Also denote $||\mathcal{Y}(f_n)||$ the cardinality of $\mathcal{Y}(f_n)$.

Now we define the **size**

$$N_\alpha^n(\pi, \beta, \varphi) = \min\{||\mathcal{Y}(f_n)|| : f_n \in \mathcal{F}_n(\pi, \beta, \varphi, T^-, O^-)\}$$

and the **rate**

$$\nu_\alpha(\pi, \beta, \varphi) = \lim_{n \to \infty} \frac{1}{n} \log N_\alpha^n(\pi, \beta, \varphi).$$

Analogously, we define in the case (T^-, O^+) the quantities $O_\alpha^n(\pi, \beta, \varphi)$, $\omega_\alpha(\pi, \beta, \varphi)$, in the case (T^+, O^-) the quantities $T_\alpha^n(\pi, \beta, \varphi)$, $\tau_\alpha(\pi, \beta, \varphi)$ and in the case (T^+, O^+) the quantities $G_\alpha^n(\pi, \beta, \varphi)$, $\gamma_\alpha(\pi, \beta, \varphi)$.

(d) the active memory. Now the organizer has additional memory of size m, where he is free to delete or store any relevant information at any time. Here we are led to study the quantities $N_\alpha^n(\pi, \beta, \varphi, m)$, $\nu_\alpha(\pi, \beta, \varphi, m)$, etc.

Survey of the results

π	φ	$\nu_2(\pi, \beta, \varphi)$
0	0	1
0	1	1
1	0	$\sup_\delta (1 - (\beta - 1)\delta) h\left(\frac{\delta}{1-(\beta-1)\delta}\right)$
π	∞	$1/\beta$
∞	$\leq \beta - 1$	$\log \lambda^*$, where λ^* is the largest root of $\lambda^{\beta+1+\varphi} = \lambda^{\lceil(\beta+1+\varphi)/2\rceil} + \lambda^{\lfloor(\beta+1+\varphi)/2\rfloor}$
∞	$\geq \beta - 1$	$1/\beta$

Furthermore the following relations hold. $\omega_2(\infty,\beta,\varphi) = \nu_2(\infty,\beta,\varphi)$, $\omega_2(\pi,\beta,\infty) = \nu_2(\pi,\beta,\infty)$, $\lim_{\beta\to\infty}\nu_3(0,\beta,0) = 1$, $\tau_2(\pi,\beta,\varphi) = \nu_2(\infty,\beta,\varphi)$ for $\pi \geq 1$,

$$\tau_2(0,2,0) = \log((\sqrt{5}+1)/2).$$

In the model of active memory we have for the memory size $m = 2$ that $\nu_2(0,\beta,0,2) = \nu_2(1,\beta,0) = \log\lambda_\beta$, where λ_β is the positive root of $\lambda^\beta - \lambda^{\beta-1} - 1 = 0$.

The general case, where the size α of the set $\mathcal{X}$, the size β of the box, and the memory parameters π, φ and m are arbitrary, has not been solved yet. This is the cardinal goal for our research to aim at within this field. We have the following **conjectures**.

1. $\lim_{\varphi\to\infty}\nu_2(\pi,\beta,\varphi) \neq \nu_2(\pi,\beta,\infty)$ (in the analogous case for $\pi \to \infty$ equality holds)

2. $\lim_{\beta\to\infty}\nu_\alpha(0,\beta,0) = \log_2\lceil(\alpha+1)/2\rceil$ (for $\alpha = 2$ and $\alpha = 3$ this is true)

3. $\omega_2(0,\beta,0) = \nu_2(1,\beta-1,0)$

In a probabilistic model the objects or letters are produced by a stochastic process, which in the simplest case is a sequence $(X_t)_{t=1}^\infty$ of i.i.d. RV's with values in $\mathcal{X} = \{0,1,\ldots,\alpha-1\}$ and generic distribution P_X. In Information Theory this is also called a discrete, memoryless source. For a strategy f_n, which depends on the triple (π,β,φ), let $Y^n = Y_1\ldots Y_n$ be the output sequence corresponding to $X^n = X_1\ldots X_n$. Let $F_\alpha^n(\pi,\beta,\varphi,P_X)$ be the set of strategies restricted to block length n.

We use the "per letter" entropy $\frac{1}{n}H(Y^n)$ as performance criterion and define

$$\eta_\alpha(\pi,\beta,\varphi,P_X) = \lim_{n\to\infty}\min_{f_n\in F_\alpha^n(\pi,\beta,\varphi,P_X)}\frac{1}{n}H(Y^n).$$

This is the smallest mean entropy of the output process, which can be achieved by $\mathcal{O}$ with strategies based on his knowledge. It corresponds to the optimal rate $\nu_\alpha(\pi,\beta,\varphi)$ in the non-probabilistic model. Our new quantity is much harder to analyze.

In the first non-trivial case $\beta = 2$ and $\pi = \infty, \varphi = 0$ only the simplest non-trivial source, namely the binary symmetric source defined by $P_X(0) = P_X(1) = 1/2$, could be analyzed.

Theorem *The strategy which is locally optimal for every $t = 1,2,\ldots$ is optimal. Moreover for the disjoint events $D_k = E_k \setminus E_{k+1}$, where $E_k = \{Y^k = 01010\ldots\}$, $q(k) = Prob(D_k)$ satisfies $\sum_{n=1}^\infty q(k) = 1$ and*

$$\eta_2(\infty,2,0,P_X) = \frac{H(q)}{\sum_{k=1}^\infty kq(k)} = 0,5989\ldots \qquad (2)$$

Conjecture *The formula (2) has a nice structure. It suggests a general principle for arbitrary sources. However, already the binary non-symmetric source is difficult to solve.*

Finally we mention the survey of Vanroose, pages 603–613 in [48].

Directions of developments of our basic model for sequences

Multiple in- and outputs: s inputs and s outputs, varying number of outputs, merging, splitting, correlation

Objects with special features: Varying-length objects, death–birth, idle objects, box with exclusion rule

Compound objects: Box with reaction rules, representatives, objects with many properties, exchanging parts of objects

Errors: Probabilistic, confusion rule, frequency rule, receiver can distinguish only certain objects

Applications

Production of goods, arrival of goods and documents, garbage collection

Extensions of the basic model

A combined theory of ordering and source coding

Ordering, sorting and Maxwell's demon

A calculus of machines: comparisons of machines, commutativity

Other topics

When after an interruption of a decade we attended the ISIT again, namely in Seattle 2006, we learned from the outside world about seemingly important topics: oblivious transfer capacity, denoising, fountain capacity, and timing channels with jamming. All these can be studied also in the context of GTIT.

One can conceive of Information Theory in the broad sense as covering the theory of Gaining, Transferring, and Storing Information, where the first is usually called Statistics. For a somewhat different view the reader is advised to look at [61]. A broad class of statistical problems arises in the framework of hypothesis testing in the spirit of identification for different kinds of sources, with complete or partial side information or without it. Paper [37] is a start.

Information concepts play an important role in Game Theory. Information theorists usually think about choosing portfolios (see [60]), a direction started by J. Kelly ([83]), but there are many more connections which ought to be studied. We hint at them by listing basic papers [49], [50], [80], [104] and by pointing at the titles of the following survey articles [76], [88], [93] and [94] in the

Handbook of Game Theory. They contain the terms common knowledge, communication, correlated equilibria, search, and signalling.

We feel that animal communication ([58], [81]), psychology, and also neurology ought to be studied experimentally in the light of GTIT, with and without feedback.

A final question to Shannon's attorneys

The following last paragraph on page 376 is taken from "Two way communication channels", C. Shannon Collected Papers, 351–384.

> *"The inner bound also has an interesting interpretation. If we artificially limit the codes to those where the transmitted sequence at each terminal depends only on the message and not on the received sequences at that terminal, then the inner bound is indeed the capacity region. This results since in this case we have at each stage of the transmission (that is, given the index of the letter being transmitted) independence between the two next transmitted letters. It follows that the total vector change in equivocation is bounded by the sum of n vectors, each corresponding to an independent probability assignment. Details of this proofs are left to the reader.* ***The independence required would also occur if the transmission and repetition points at each end were at different places with no direct cross communication.***"

According to our understanding the last sentence in this quote (which is put here in boldface) implies the solution of the capacity region problem for what is now called Interference Channel. **Already in [5] we showed that the region obtained with independent sender's distributions is generally smaller than the capacity region.**

References

[1] R. Ahlswede, Channel capacities for list codes, J. Appl. Probability, 10, pp. 824–836, 1973.

[2] 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, Publ. House Czechosl. Academy of Sc., 1–22, 1973.

[3] R. Ahlswede, Channels with arbitrarily varying channel probability functions in the presence of noiseless feedback, Z. Wahrscheinlichkeitstheorie und verw. Geb. 25, pp. 239–252, 1973.

[4] R. Ahlswede, Multi-way communication channels, Proceedings of 2nd International Symposium on Information Theory, Thakadsor, Armenian SSR, Sept. 1971, Akademiai Kiado, Budapest, pp. 23–52, 1973.

[5] R. Ahlswede, The capacity region of a channel with two senders and two receivers, Ann. Probability, vol. 2, no. 5, pp. 805–814, 1973.

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

[7] R. Ahlswede, Coloring hypergraphs: A new approach to multi-user source coding I, Journ. of Combinatorics, Information and System Sciences, vol. 4, no. 1, pp. 76–115, 1979.

[8] R. Ahlswede, Coloring hypergraphs: A new approach to multi-user source coding II, Journ. of Combinatorics, Information and System Sciences, vol. 5, no. 3, pp. 220–268, 1980.

[9] R. Ahlswede, A method of coding and its application to arbitrarily varying channels, J. Combinatorics, Information and System Sciences, vol. 5, no. 1, pp. 10–35, 1980.

[10] R. Ahlswede, Advances on extremal problems in Number Theory and Combinatorics, European Congress of Mathematicians, Barcelona 2000, vol. I, 147.175, Progress in Mathematics, vol. 201, 2001.

[11] 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, pp. 639–663, 2006.

[12] R. Ahlswede, Another diametric theorem in Hamming spaces: optimal group anticodes, Proc. IEEE Information Theory Workshop, Punta del Este, Uruguay, March 13–17, pp. 212–216, 2006.

[13] R. Ahlswede, General theory of information transfer: updated, General Theory of Information Transfer and Combinatorics, Special Issue of Discrete Applied Mathematics, to appear.

[14] R. Ahlswede, H. Aydinian, and L.H. Khachatrian, More about shifting techniques,, European Journal of Combinatorics 24, pp. 551–556, 2003.

[15] R. Ahlswede, H. Aydinian, and L.H. Khachatrian, Intersection theorems under dimension constraints part I: the restricted case and part II: the unrestricted case, J. Comb. Theory, Series A 113, pp. 483–519, 2006.

[16] R. Ahlswede and V. Balakirsky, Identification under random processes, Problems of Information Transmission, vol. 32, no. 1, pp. 123–138, 1996.

[17] 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, pp. 51–61, 2006.

[18] R. Ahlswede, L.A. Bassalygo, and M.S. Pinsker, Nonbinary codes correcting localized errors, IEEE Trans. Inf. Theory, vol. 39, no. 4, pp. 1413–1416, 1993.

[19] R. Ahlswede, L. Bäumer, N. Cai, H. Aydinian, V. Blinovsky, C. Deppe, and H. Mashurian (Eds.), General Theory of Information Transfer and Combinatorics, Lecture Notes in Computer Science,

vol. 4123, Springer Verlag, 2006.

[20] R. Ahlswede and V. Blinovsky, Classical capacity of classical-quantum arbitrarily varying channels IEEE Trans. Inf. Theory, vol. 53, no. 2, pp. 526–533, 2007.

[21] R. Ahlswede and N. Cai, General edge-isoperimetric inequalities, Part 2: A local-global principle for lexicographical solutions, European J. of Combinatorics 18, pp. 479–489, 1997.

[22] R. Ahlswede and N. Cai, Correlated sources help the transmission over AVC, IEEE Trans. Inf. Theory, vol. 43, no. 4, pp. 1254–1255, 1997.

[23] R. Ahlswede and N. Cai, Arbitrarily varying multiple-access channels, Part I. Ericson's symmetrizability is adequate, Gubner's conjecture is true, IEEE Trans. Inf. Theory, vol. 45, no. 2, pp. 742–749, 1999.

[24] R. Ahlswede and N. Cai, Arbitrarily varying multiple-access channels, Part II. Correlated sender's side information, correlated messages, and ambiguous transmission, IEEE Trans. Inf. Theory, vol. 45, no. 2, pp. 749–756, 1999.

[25] 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, pp. 151–176, 2000.

[26] R. Ahlswede and N. Cai, An interpretation of identification entropy, IEEE Trans. Inf. Theory, vol. 52, no. 9, pp. 4198–4207, 2006.

[27] R. Ahlswede and N. Cai, Transmission, identification and common randomness capacities for wire-tape channels with secure feedback from the decoder, General Theory of Information Transfer and Combinatorics, Lecture Notes in Computer Science, vol. 4123, Springer Verlag, pp. 258–275, 2006.

[28] R. Ahlswede and N. Cai, A strong converse theorem for quantum multiple access channels, General Theory of Information Transfer and Combinatorics, Lecture Notes in Computer Science, vol. 4123, Springer Verlag, pp. 460–485, 2006.

[29] R. Ahlswede, N. Cai, S.Y.R. Li, and R.W. Yeung, Preprint 98-033, SFB 343 "Diskrete Strukturen in der Mathematik", Universitt Bielefeld, IEEE Trans. Inf. Theory, vol. 46, no. 4, pp. 1204–1216, 2000.

[30] R. Ahlswede, N. Cai, and Z. Zhang, Erasure, list, and detection zero-error capacities for low noise and a relation to identification, IEEE Trans. Inf. Theory, vol. 42, no. 1, pp. 55–62, 1996.

[31] R. Ahlswede and I. Csiszár, Common randomness in information theory and cryptography, part I: secret sharing, IEEE Trans. Inf. Theory, vol. 39, no. 4, pp. 1121–1132, 1993.

[32] R. Ahlswede and I. Csiszár, Common randomness in information theory and cryptography, part II: CR capacity, IEEE Trans. Inf. Theory, vol. 44, no. 1, pp. 225–240, 1998.

[33] R. Ahlswede, C. Deppe, and V. Lebedev, Nonbinary error correcting codes with noiseless feedback, localized errors or both, Annals of European Academy of Sciences, no. 1, pp. 285–309, 2005.

[34] R. Ahlswede and G. Dueck, Identification via channels, IEEE Trans. Inf. Theory, vol. 35, no. 1, pp. 15–29, 1989.

[35] R. Ahlswede and G. Dueck, Identification in the presence of feedback—a discovery of new capacity formulas, IEEE Trans. Inf. Theory, vol. 35, no. 1, pp. 30–39, 1989.

[36] R. Ahlswede, P. Gács, and J. Körner, Bounds on conditional probabilities with applications in multiuser communication, Z. Wahrscheinlichkeitstheorie und verw. Geb. 34, pp. 157–177, 1976.

[37] R. Ahlswede and E. Haroutunian, On logarithmically asymptotically optimal testing of hypothesis and identification, General Theory of Information Transfer and Combinatorics, Lecture Notes in Computer Science, vol. 4123, Springer Verlag, pp. 553–571, 2006.

[38] R. Ahlswede and L.H. Khachatrian, The complete intersection theorem for systems of finite sets, European J. Combinatorics, 18, pp. 125–136, 1997.

[39] R. Ahlswede and L.H. Khachatrian, The diametric theorem in Hamming spaces—optimal anticodes, Proceedings First INTAS International Seminar on Coding Theory and Combinatorics 1996, Thahkadzor, Armenia, 1–19, 6–11 October 1996; Advances in Applied Mathematics 20, pp. 429–449, 1998.

[40] R. Ahlswede and J. Körner, On common information and related characteristics of correlated information sources, General Theory of Information Transfer and Combinatorics, Lecture Notes in Computer Science, vol. 4123, Springer Verlag, pp. 664–677, 2006.

[41] R. Ahlswede and B. Verboven, On identification via multiway channels with feedback, IEEE Trans. Inf. Theory, vol. 37, no. 5, pp. 1519–1526, 1991.

[42] R. Ahlswede and I. Wegener, Suchprobleme, Teubner Verlag, Stuttgart, 1979. Russian Edition with Appendix by Maljutov 1981. Search problems, Wiley-Interscience Series in Discrete Mathematics and Optimization, John Wiley & Sons., 1987.

[43] R. Ahlswede and A. Winter, Strong converse for identification via quantum channels, IEEE Trans. Inf. Theory, vol. 48, no. 3, pp. 569–579, 2002.

[44] R. Ahlswede, E. Yang, and Z. Zhang, Identification via compressed data, IEEE Trans. Inf. Theory, vol. 43, no. 1, pp. 48–70, 1997.

[45] R. Ahlswede, J.-P. Ye, and Z. Zhang, Creating order in sequence spaces with simple machines, Information and Computation, vol. 89, no. 1, pp. 47–94, 1990.

[46] R. Ahlswede and Z. Zhang, Contributions to a theory of ordering for sequence spaces, Problems of Control and Information Theory, vol. 18, no. 4, pp. 197–221, 1989.

[47] R. Ahlswede and Z. Zhang, New directions in the theory of identification via channels, IEEE Trans. Inf. Theory, vol. 41, no. 4, pp. 1040–1050, 1995.

[48] I. Althöfer, N. Cai, G. Dueck, L.H. Khachatrian, M. Pinsker, A. Sárközy, I. Wegener, and Z. Zhang (Eds.), Numbers, Information and Complexity, Special volume in honour of R. Ahlswede on occasion of his 60th birthday, Kluwer Acad. Publ., Boston, Dordrecht, London, 2000.

[49] R.J. Aumann, Agreeing to disagree, The Annals of Statistics, vol. 4, pp. 1236–1239, 1976.

[50] R.J. Aumann, Subjectivity and correlation in randomized strategies, J. of Mathematical Economics, vol. 1, pp. 67–96, 1974.

[51] R.J. Aumann and S. Hart (Eds.) Handbook of Game Theory with Economic Applications, vol. 2, Elsevier, 1994.

[52] Z. Baranyai, On the factorization of the complete uniform hypergraph. Infinite and finite sets, vol. 1, Proceedings of a Colloquium held at Keszthely, June 25–July 1, 1973, Dedicated to Paul Erdős on his 60th Birthday, Eds. A. Hajnal, R. Rado, and V. T. Sós, pp. 91–108, 1975.

[53] L.A. Bassalygo and M.V. Burnashev, Authentication, identification, and pairwise-separated measures, Problems Inform. Transmission 32, no. 1, pp. 33–39, 1996.

[54] L.A. Bassalygo, S.I. Gelfand, and M.S. Pinsker, Coding for channels with localized errors, Proc. Fourth Soviet–Swedish Workshop in Information Theory, Gotland, Sweden, pp. 95–99, 1989.

[55] E.R. Berlekamp, Block Coding with Noiseless Feedback, Doctoral Dissertation, MIT, 1964.

[56] E.R. Berlekamp, The Performance of Block Codes, Notices of the AMS, pp. 17–22, 2002.

[57] D. Blackwell, L. Breiman, and A.J. Thomasian, The capacity of certain channel classes under random coding, Ann. Math. Statist. 31, pp. 558–567, 1960.

[58] J.W. Bradbury and S.L. Vehrencamp, Principles of Animal Communication, Sinauer Assoc., Inc., Publishers Sunderland, Mass., USA, 1998.

[59] T.M. Cover, Broadcast channels, IEEE Trans. Inf. Theory, vol. 18, no. 1, pp. 2–14, 1972.

[60] T.M. Cover, Shannon and investment, IEEE Inform. Theory Society Newsletter, special Golden Jubilee issue, pp. 10–11, 1998.

[61] I. Csiszár, Information theoretic methods in probability and statistics (Shannon Lecture 1996), IEEE Inform. Theory Society Newsletter, vol. 48, no. 1, p. 1 and pp. 30–33, 1996.

[62] I. Csiszár and J. Körner, Broadcast channel with confidential messages, IEEE Trans. Inf. Theory, vol. 24, pp. 339–348. 1978.

[63] I. Csiszár and P. Narayan, The capacity of the arbitrarily varying channel revisited: positivity constraints, IEEE Transactions of IT, vol. 34, no, 2, pp. 181–193, 1988.

[64] I. Csiszár and P. Narayan, Common randomness and secret key generation with a helper, IEEE Trans. Inf. Theory, vol. 46, no. 2, pp. 344–366, 2000.

[65] I. Csiszár and P. Narayan, Secrecy capacities for multiterminal channel models, submitted to IEEE Trans. Inf. Theory, spec. issue on information theoretic secrecy.

[66] C. Deppe, Searching with lies and coding with feedback, Entropy, Search, Complexity, Bolyai Society Mathematical Studies, Katona (Ed.), vol. 16, pp. 27–70, 2007.

[67] K. Engel, Sperner Theory, Cambridge University Press, Cambridge, 1997.

[68] P. Erdös, C. Ko, and R. Rado, Intersection theorems for systems of finite sets, Quart. J. Math. Oxford, vol. 12, pp. 313–320, 1961.

[69] T. Ericson, Exponential error bounds for random codes in the arbitrarily varying channel, IEEE Transactions on IT, vol. 31, no. 1, pp. 42–48, 1985.

[70] L. R. Ford and D. R. Fulkerson, Maximal flow through a network, Canad. J. Math. 8, pp. 399–404, 1956.

[71] G.D. Forney, Jr., Performance and complexity (Shannon Lecture 1995), IEEE Inform. Theory Society Newsletter, vol. 46, no. 1, pp. 3–5 and 23–25, 1996.

[72] P. Frankl, On intersecting families of finite sets, J. Combinatorial Theory Ser. A 24, no. 2, pp. 146–161, 1978.

[73] P. Gács and J. Körner, Common information is far less than mutual information, Problems of Control and Information Theory/Problemy Upravlenija i Teorii Informacii 2, no. 2, pp. 149–162, 1973.

[74] R.G. Gallager, Basic limits on protocol information in data communication networks, IEEE Trans. Inf. Theory, vol. 22, no. 4, pp. 385–398, 1976.

[75] R.G. Gallager, Source coding with side information and universal coding, IEEE Int. Syposium on Inf. Theo., Renneby, Sweden, 1976.

[76] J. Geanakoplos, Common knowledge, Handbook of Game Th. with Econ. Appl., vol. 2, pp. 1437–1496, 1994.

[77] T.S. Han and S. Verdú, New results in the theory and application of identification via channels, IEEE Trans. Inf. Theory, vol. 38, no. 1, pp. 14–25, 1992.

[78] T.S. Han and S. Verdú, Approximation theory of output statistics, IEEE Trans. Inf. Theory, vol. 39, no. 3, 1993.

[79] L.H. Harper, Global Methods for Combinatorial Isoperimetric Problems, Cambridge University Press, Cambridge, 2004.

[80] J.C. Harsanyi, Games with incomplete information played by "Bayesian" players Part I-III, Management Sci., vol. 14, pp. 159–182, 320–334, 486–502, 1967–1968.

[81] M.D. Hauser, The Evolution of Communication, MIT Press, Cambridge, Mass., London, England, 1997.

[82] C. Heup, *L*-identification for Sources, PhD-thesis, University of Bielefeld, 2006.

[83] J. Kelly, A new interpretation of information rate, Bell Syst. Tech. J., pp. 917–926, 1956.

[84] J. Kilian, Founding cryptography on oblivious transfer, STOC 1998, pp. 20–31, 1988.

[85] C. Kleinewächter, Identification of messages and identifying zeroes of a function, Diploma Thesis, University of Bielefeld, 1996.

[86] C. Kleinewächter, On identification, General Theory of Information Transfer and Combinatorics, Lecture Notes in Computer Science, vol. 4123, Springer Verlag, pp. 62–83, 2006.

[87] J. Körner and K. Marton, General broadcast channels with degraded message sets, IEEE Trans. Inf. Theory, vol. 23, no. 1, pp. 60–64, 1977.

[88] D.M. Kreps and J. Sobel, Signalling, Handbook of Game Th. with Econ. Appl., vol. 2, pp. 849–867, 1994.

[89] H. Marko, Die Theorie der bidirektionalen Kommunikation und ihre Anwendung auf die Nachrichtenübermittlung zwischen Menschen (Subjektive Information), Kybernetik, vol. 3, pp. 128–136, 1966.

[90] J.L. Massey, Causality, feedback and directed information, Proceedings of the International Symposium on Information Theory and its Applications, Honolulu, 1990.

[91] U. Maurer, Secret key agreement by public discussion, IEEE Trans. Inf. Theory, vol. 39, no. 3, pp. 733–742, 1993.

[92] U. Maurer, The strong secret key rate of discrete random triples, Communications and Cryptography: Two sides of One Tapestry, R.E. Blahut et.al. Eds., Kluwer, pp. 271–285, 1994.

[93] J. McMillan and M. Rothschild, Search, Handbook of Game Th. with Econ. Appl., vol. 2, pp. 905–927, 1994.

[94] R.B. Myerson, Communication, correlated equilibria and incentive compatibility. Handbook of Game Th. with Econ. Appl., vol. 2, pp. 827–847, 1994.

[95] A. Nascimento and A. Winter, On the oblivious transfer capacity of noisy correlations, Proc. ISIT 2006, Seattle, pp. 1871–1875, 2006.

[96] J.M. Ooi, Coding for Channels with Feedback, Kluwer Academic Publishers, 1998.

[97] C.E. Shannon, A mathematical theory of communication, Bell Syst. Tech. J. 27, pp. 379–423, 623–656, 1948.

[98] C.E. Shannon, Development of Communication and Computing, and my Hobby, Commemorative Lectures of the 1985 (1st) Kyoto Prize Laureates (Advanced Technology: Rudolf E. Kalman, Basic Sciences: Claude E. Shannon, Creative Arts and Moral Sciences: Olivier Messiaen), pp. 135–153, 1985.

[99] C. Shannon, P. Elias, and A. Feinstein, A note on the maximum flow through a network, IEEE Trans. Inf. Theory, vol. 2, no. 4, pp. 117–119, 1956.

[100] C.E. Shannon, R.G. Gallager, and E.R. Berlekamp, Lower bounds to error probability for coding on discrete memoryless channels I, Information and Control 10, pp. 65–103, 1967.

[101] C.E. Shannon, R.G. Gallager, and E.R. Berlekamp, Lower bounds to error probability for coding on discrete memoryless channels II, Information and Control 10, pp. 522–552, 1967.

[102] D. Slepian and J.K. Wolf, Noiseless coding of correlated information sources, IEEE Trans. Inf. Theory, vol. 19, no. 4, pp. 471–480, 1973.

[103] Y. Steinberg, New converses in the theory of identification via channels, IEEE Trans. Inf. Theory, vol. 44, no. 3, 1998.

[104] G.J. Stigler, The economics of information, J. of Political Economy, vol. 69, no. 3, pp. 213–225, 1961.

[105] S. Venkatesan and V. Anantharam, The common randomness capacity of a pair of independent discrete memoryless channels, IEEE Trans. Inf. Theory, vol. 44, no. 1, pp. 215–224, 1998.

[106] S. Venkatesan and V. Anantharam, The common randomness capacity of a network of discrete memoryless channels, IEEE Trans. Inf. Theory, vol. 46, no. 2, pp. 367–387, 2000.

[107] 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. Inf. Theory, vol. 36, no. 6, pp. 1197–1205, 1990.

[108] S. Verdú and V. Wei, Explicit construction of constant-weight codes for identification via channels, IEEE Trans. Inf. Theory, vol. 39, no. 1, pp. 30–36, 1993.

[109] E.C. van der Meulen, Random coding theorems for the general discrete memoryless broadcast channel, IEEE Trans. Inf. Theory, vol. 21, no. 2, pp. 180–190, 1975.

[110] R.M. Wilson, The exact bound on the Erdős-Ko-Rado theorem, Combinatorica, 4, pp. 247–257, 1984.

[111] A. Winter, Identification via quantum channels in the presence of prior correlation and feedback, General Theory of Information Transfer and Combinatorics, Lecture Notes in Computer Science, vol. 4123, Springer Verlag, pp. 486–504, 2006.

[112] A. Winter, A. Nascimento, and H. Imai, Commitment capacity of discrete memoryless channels, Cryptography and Coding 2003, LNCS 2898, pp. 35–51, Springer 2003.

[113] Y. Wu, K. Jain, and S.-Y. Kung, A unification of Edmonds' routing theorem and Ahlswede et al's network coding theorem, the joint special issue of the IEEE Trans. Inf. Theory and the IEEE/ACM Trans. on Networking and Information Theory, 2006.

[114] A.D. Wyner, The wire-tap channel, Bell Sys. Tech. J., vol. 54, pp. 1355–1387, 1975.

[115] A.D. Wyner, A theorem on the entropy of certain binary sequences and applications II, IEEE Trans. Inf. Theory, vol. 19, no. 6, pp. 772–777, 1973.

[116] K.Sh. Zigangirov, Number of correctable errors for transmission over a binary symmetrical channel with Feedback, Problems Inform. Transmission, vol. 12, pp. 85–97, 1976.

[117] J. Ziv, Back from infinity: a constrained resources approach to information theory (Shannon Lecture 1997), IEEE Inform. Theory Society Newsletter, vol. 48, no. 1, p. 1 and pp. 21–30, 1996.

Award Announcements

Marc Fossorier

2008 Claude E. Shannon Award

Prof. Bob Gray from Stanford was awarded the "2008 Claude E. Shannon Award." Prof. Gray will held his Shannon lecture at ISIT 2008 in Toronto, Canada.

2007 Information Theory Society Paper Award

The 2007 Information Theory Society Paper Award recognizes an exceptional publication in information theory, appearing in the period January 1, 2005 through December 31, 2006. At ISIT 2007 in Nice, it was announced that the award goes to:

"The Capacity Region of the Gaussian Multiple-Input Multiple-Output Broadcast Channel," by H. Weingarten, Y. Steinberg and S. Shamai (Shitz), which appeared in the IEEE Transactions on Information Theory, vol. 52, No. 9, pp. 3936-3964, September 2006.

This paper establishes the capacity region of one of the most important class of broadcast channels. In the process, new concepts and analytical tools are introduced. These results already impacted many other works in information theory.

A special mention should be given to the runner-up paper which was recognized by the award subcommittee to be an extremely strong contender:

"Mutual Information and Minimal Mean-Squared Error in Gaussian Channels", by D. Guo, S. Shamai (Shitz) and S. Verdu, IEEE Transactions on Information Theory , vol. 51, pp. 1261-1282, April 2005.

2007 Information Theory Society Aaron D. Wyner Distinguished Service Award

Dr. Jack Wolf, Stephen O. Rice Professor of Magnetics at the University of California at San Diego, has been awarded the 2007 Aaron D. Wyner Distinguished Service Award. The award honors individuals who have shown outstanding leadership in, and provided long standing exceptional service to, the Information Theory community.

2007 Information Theory Society Chapter of the Year Award

The award goes to the Seoul Chapter. The award recognizes the most active chapter during the previous year.

2007 Information Theory Student Paper Award

The first Information Theory Student Paper Award has been awarded to:

"Minimum Expected Distortion in Gaussian Layered Broadcast Coding with Successive Refinement," by Chris T.K. Ng, Deniz Gunduz, Andrea Goldsmith, and Elza Erkip

and

"Uplink Macro Diversity with Limited Backhaul Capacity," by Amichai Sanderovich, Oren Somekh, and Shlomo Shamai

Papers with a student author as the major contributor and presenter were eligible for this award. In total, 193 of the submissions to ISIT were self-marked as eligible for the student paper award. Of these, 106 were accepted for publication in ISIT and considered in the selection of the finalists. The complete list of the finalists is available at http://www.isit2007.org/index.php.