

Towards a General Theory of Information Transfer

Rudolf Ahlswede

More than restoring
strings of symbols transmitted
means transfer today

- A. Probabilistic Models
- B. Combinatorial Models
- C. Further Perspectives

A. Probabilistic Models

I. **Transmission** via DMC (Shannon Theory)

II. **Identification** via DMC (including Feedback)

III. Discovery of
Mystery Numbers = Common Randomness Capacity

“Principle”

1. Order Common Randomness Capacity C_R
=
2. Order Identification Capacity C_{ID}

IV. “Consequences” for **Secrecy Systems**

V. More General **Transfer** Models

VI. Extensions to Classical/**Quantum Channels**

VII. **Source Coding for Identification**
Discovery of Identification Entropy

B. Combinatorial Models

VIII. Updating Memories with cost constraints - Optimal **Anticodes**

Ahlsvede/Khachatrian
Complete Intersection Theorem
Problem of Erdős/Ko/Rado 1938

IX. Network Coding for **Information Flows**

Shannon's Missed Theory

X. **Localized** Errors

Ahlsvede/Bassalygo/Pinsker
Almost Made it

XI. **Search**

Rényi/Berlekamp/Ulam Liar Problem
(or Error Correcting Codes with feedback)

Berlekamp's Thesis II

Rényi's Missed Theorem

XII. **Combi-Probabilistic** Models

Coloring Hypergraphs did a problem by Gallager

C. Further Perspectives

a. Protocol Information ?

b. Beyond Information Theory:

Identification as a **New Concept of Solution**
for Probabilistic Algorithms

c. A New Connection between Information Inequalities
and Combinatorial Number Theory (Tao)

d. A Question for Shannon's Attorneys

e. Could we ask Shannon's advise !

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?”

\mathcal{X} = input alphabet, \mathcal{Y} = output alphabet $W^n(y^n|x^n) = \prod_{t=1}^n W(y_t|x_t)$ channel

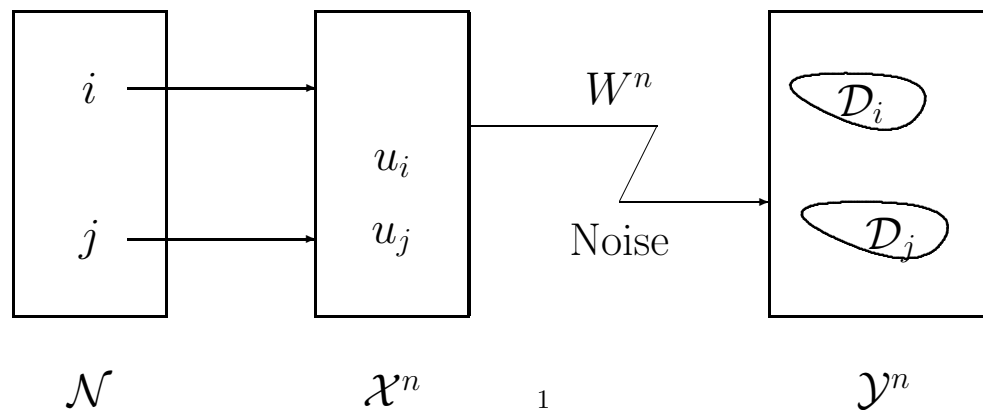
W = stochastic matrix $x^n = (x_1, x_1, \dots, x_n) \in \mathcal{X}, y^n \in \mathcal{Y}^n$.

Definition: (n, N, ε) Code: $\{(u_i, D_i) : 1 \leq i \leq N\}$ with $u_i \in \mathcal{X}^n, D_i \subset \mathcal{Y}^n, D_i \cap D_j = \emptyset (i \neq j), W^n(D_i|u_i) \geq 1 - \varepsilon$.

Definition: $N(n, \varepsilon) = \max N$

Shannon 48: $\lim_{n \rightarrow \infty} \frac{1}{n} \log M(n, \varepsilon) = C$

$$\underbrace{C}_{\text{capacity}} = \max \underbrace{H(X)}_{\text{entropy}} - \underbrace{H(X|Y)}_{\text{cond. entropy}} = I(X \wedge Y)_{\text{mutual information}}$$



II. Identification via DMC (including Feedback)

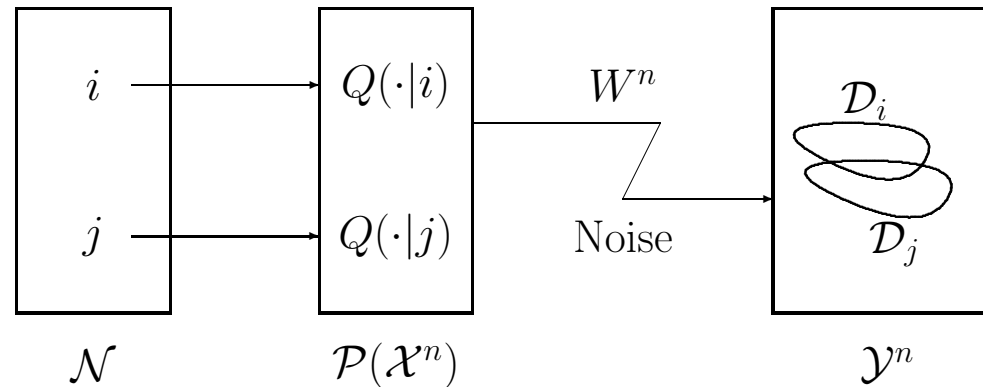
How many possible messages can the receiver of a noisy channel identify?

Identification means there is an answer to the question

“Is the actual message i ?”

Here i can be any member of the set of possible messages $\{1, 2, \dots, N\}$.

Here randomisation helps!!!



Definition $(n, N, \varepsilon_1, \varepsilon_2)$ ID-code

$$\{(Q(\cdot|i), D_i) : 1 \leq i \leq N\}$$

with $Q(\cdot|i) \in \mathcal{P}(\mathcal{X}^n) =$ 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 1. 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 2. kind: i accepted, but some $j \neq i$ present)

Definition $N(n, \varepsilon) = \max N$ for which $\exists(n, N, \varepsilon, \varepsilon)$

ID-code

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

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

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

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

C = second order identification capacity

= Shannon's (first order) transmission capacity.

Theorem AD₂: In case of feedback the 2-order ID-capacities are, if $C > 0$

without randomisation: $C_f(W) = \max_{x \in \mathcal{X}} H(W(\cdot|x))$

with randomisation: $C_f(W) = \max_P H(P \cdot W) \geq C$

Phenomena:

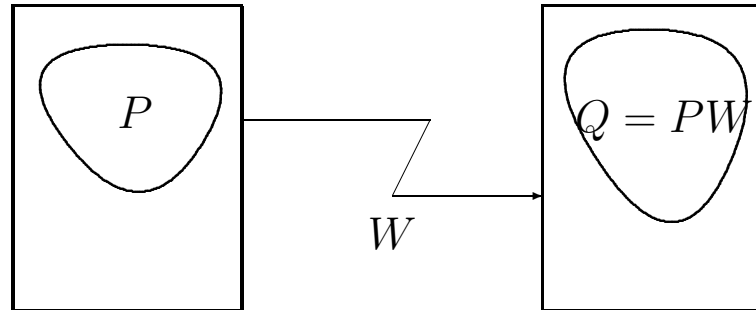
1. Feedback increases the optimal rate for identification.
2. Noise can increase the identification capacity of a DMC in case of feedback (think about probabilistic algorithms, here noise creates the randomisation, not the case for Shannon's theory of transmission)
3. **Idea:** Produce “big” (large entropy) random experiment with a result known to sender and receiver.
 \sqrt{n} -trick, random keys)

“**Principle**”: Entropy of a large common random experiment = ID-capacity of 2. order (region).

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

Approximation of output distributions.

Converse for Theorem AD_1

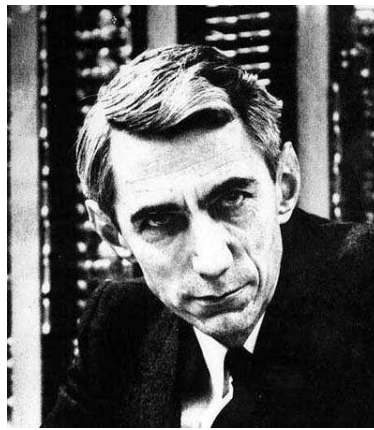


How can we count?

Find $\mathcal{U} \subset \mathcal{X}^n$ with uniform distribution $P_{\mathcal{U}}$:

$$P_{\mathcal{U}} \cdot W \sim Q$$

Minimize $|\mathcal{U}|$, then $N \lesssim \binom{|\mathcal{X}^n|}{|\mathcal{U}|}$.



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

What is information?

Cn bits in Shannon’s fundamental theorem

or

$\log Cn$ bits in our Theory of Identification

III. Discovery of Mystery Numbers = Common Randomness Capacity

Mystery Number

In dealing with different kinds of feedback strategies it is convenient to have the following concept. Let \mathcal{F}_n ($n = 1, 2, \dots$) 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 $n = n_1 + n_2$

$$\mathcal{F}_n \supset \mathcal{F}_{n_1} \times \mathcal{F}_{n_2} \quad (*)$$

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

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

By (*) and the memoryless character of the channel

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

and therefore

$$\boxed{\mu = \mu((\mathcal{F}_n)_{n=1}^\infty = \lim_{n \rightarrow \infty} 1/n \mu(\mathcal{F}_n) \text{ exists.}}$$

It is called **mystery number** to attract attention.

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 for length n on a set \mathcal{K} with $|\mathcal{K}| < e^{cn}$ with

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

From common randomness to identification: (now also called shared randomness)

The \sqrt{n} -trick

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

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

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

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

Lemma: Given $M, M' = \exp\{\sqrt{\log M}\}$ and $\epsilon > 0$ there exists a family $\mathcal{T} = \mathcal{T}(\epsilon, M)$ such that

- $|\mathcal{T}| = N \geq \exp\{M - c(\epsilon)\sqrt{n}\}$
- $Q_i(K_i) = 1$ for $i = 1, \dots, N$
- $Q_i(K_j) \leq \epsilon \forall i \neq j$

Note In typical applications the common random experiment has range $M = \exp\{C_R n\}$ and uses for its realisation

blocklength n

while the extension by the T_i uses

blocklength \sqrt{n}

Capacity Regions

	Transmission	Identification
DMC	X	X
MAC	X	X
BC	?	X
TWC	?	?
With Feedback		
DMC	X	X
MAC	?	XD
MAC 1.)	?	?R
BC	?	?D
BC 2.)	?	XR
TWC	?	XD

Amazing dualities transmission versus 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.

and concerning MAC and BC:

- 1.) Mystery numbers for MAC?
- 2.) Mystery number region for BC known!

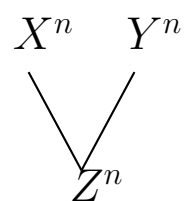
For transmission capacity regions the situation is reversed.

Further cases of validity of the “principle”

- Ahlswede/Zhang 1993: Cryptography (Wire tape channel)
- Ahlswede/Balakirsky (Correlated source helps)
- Ahlswede/Cai (More on AVC, correlated source helps)
- Ahlswede/Csiszár: AVC See Blackwell/Breiman/Thomasian 1960
- Ahlswede, Steinberg: MAC
- Ahlswede: BC
- Bassalygo/Burnashev (relation to cryptography)
- Ahlswede/Csiszár: Common Randomness: its role in Information Theory and Cryptography, Part II, 1998
- This work was continued in several papers by Csiszár/Narayan
- Venkatesan/Anantharam: The common randomness capacity of a pair of independent discrete memoryless channels, 1998

Entropy characterisation to calculate entropy of common random experiment

Example MAC:



–common

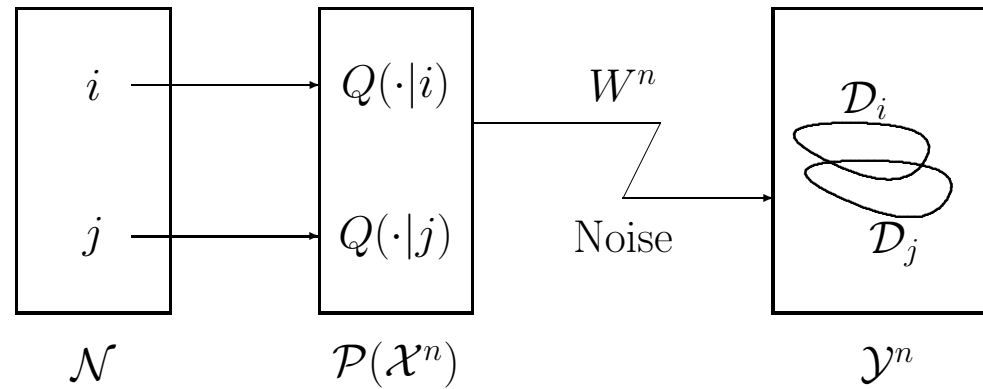
(identification–transmission)

$$X_{n+1} = f_{n+1}(Z^n)$$

$$Y_{n+1} = g_{n+1}(Z^n)$$

$$\boxed{\text{Max } H(Z^n)=?}$$

DMC



How much does a correlated source

U^n
sender side

V^n
receiver side

help for IDENTIFICATION over the DMC?

The “Principle” suggests the question:

How much does DMC help for Common Randomness in (U^n, V^n) ?

What is

COMMON INFORMATION

?

Gács/Körner

Common Information is far less than mutual information 1972

Let $(U_t, V_t)_{t=1}^n$ be pairwise i.i.d.

Maximize cardinality of ranges of f or g with

$$\text{Prob}(f(U^n) = g(V^n)) \sim 1$$

Their result: $\max \text{rate}(f) = 0$

if for instance $P_{UV}(u, v) > 0$ for all $(u, v) \in \mathcal{U} \times \mathcal{V}$

Now we see that their quantity is a common randomness capacity in case the DMC above has capacity 0.

Its significance is now understood.

Following the idea of having a concept of common information

Wyner: Common Information

C_{Wyner} depends on **probabilities** not only on **positivities**

Claimed to have **the** notion of common information.

Note: $C_{\text{Wyner}}(U, V) \geq I(U \wedge V)$

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 \sqrt{n} -trick always

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

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

(Kleinewächter found also an example involving feedback with $0 < C_I < C_R$)

We use a Gilbert type construction of error correcting codes with constant weight words. This was done for certain parameters. 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, \dots, A_N of subsets of \mathcal{Z} exists with the properties*

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

and

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

Notice that $\lambda \log \left(\frac{1}{3} - 1\right) > 2$ and that for ℓ with $2^{-\ell} = \varepsilon$ necessarily $\ell > \frac{2}{\lambda}$.

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

Consider now a discrete channel $(W^n)^\infty$, where the input alphabets $\mathcal{X}_t = \{1, 2, \dots, N(t, \lambda)\}$ 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

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

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

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

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 $\frac{1}{n} \log M(n, \lambda) \leq \frac{\ell+1}{n} \leq \frac{3/\lambda+1}{n} \rightarrow 0$ ($n \rightarrow \infty$).

The transmission capacity is 0. Consequently also $C_R = 0$.

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.

(Cover, van der Meulen; Körner and Marton)

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

Achievable are (with maximal errors)

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

resp.

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

This is our surprising result.

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

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

where

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

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

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

R. Ahlswede and N. Cai

Problems on General Theory of Information Transfer, particularly, **transmission, identification and common randomness, via a wire-tap channel with secure feedback**

are studied in this work.

Wire-tap channels were introduced by A. D. Wyner and were generalized by I. Csiszár and J. Körner.

Its identification capacity was determined by R. Ahlswede and Z. Zhang.

Here 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. More General Transfer Models

Our work on identification has led us to reconsider the basic assumptions of Shannon's Theory. It deals with "messages", which are elements of a *prescribed set of objects*, known to the communicators. The receiver wants to know the true message. This basic model occurring in all engineering work on communication channels and networks addresses a very special communication situation. More generally they are characterized by

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

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

One sender answering several questions of receivers

A general communication model for one sender

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

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

$$D_i \cap D_{i'} = \emptyset \text{ for } i = 1, \dots, M,$$

$$W(D_i^c | u_i) \leq \lambda \text{ for } i = 1, \dots, M.$$

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 occurred 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: $\mathcal{M} = \{0, 1\}^\ell$, $\Pi_C = \{\pi_t : 1 \leq t \leq \ell\}$ 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_ℓ) .

In all these models we can consider the first (or second) order capacities. They are known for models 1, 2. It is shown 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 $\{(Q(\cdot|i), D_\pi) : 1 \leq i \leq N, \pi \in \Pi_K\}$, where the $Q(\cdot|i)$'s are PD's on \mathcal{X}^n , $D_\pi \subset \mathcal{Y}^n$, and where for all $\pi = \{S, \mathcal{M} \setminus S\}$ ($S \in \binom{\mathcal{M}}{K}$)

$$\sum_{x^n} Q(x^n|i)W(D_\pi^c|x^n) \leq \lambda \quad \text{for all } i \in S,$$

$$\sum_{x^n} Q(x^n|i)W(D_\pi|x^n) \leq \lambda \quad \text{for all } i \notin S.$$

We also write D_S instead of D_π . A coding theorem is established.

Remark 1: Most models fall into the following category of regular transfer models. By this we mean that the set of partitions Π of \mathcal{M} is invariant under all permutations $\sigma : \mathcal{M} \rightarrow \mathcal{M}$:

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

Remark 2: Many of the models introduced concern bivariate partitions. More generally they are described by a hypergraph $\mathcal{H} = (\mathcal{M}, \mathcal{E})$, where decoder E wants to know whether the m occurred is in E or not.

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

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

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

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

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

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

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

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.

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 $A_0 \not\subset A_1 \cup A_2 \cup \dots \cup A_r$ holds for all distinct $A_0, A_1, \dots, A_r \in \mathcal{F}$. 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

Great progress in recent years with fruitful exchanges between Information Theory and Physics.

Note: Common Randomness — Entanglement

Since I cannot expect that many listeners understand this just **give classical methods which extend or have analoga.**

The elimination technique

This method was introduced by us in 1978.

It is a general method to obtain from a **correlated code for AVC** with probability

$$\lambda_n \leq e^{-\epsilon n}, \quad \epsilon > 0,$$

an **ordinary code** of essentially the same rate and average error probability

$$\lambda_n = o(1).$$

Since for correlated codes the capacity is known, we thus obtain the ordinary capacity.

We obtain this result, if **the ordinary capacity is known to be positive**.

Capacity of Quantum Arbitrarily Varying Channels

Rudolf Ahlswede and Vladimir Blinovsky

We prove 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 $C_q > 0$.

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 $E_1, \dots, E_L \in \mathcal{E}$ such that with

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

the following holds:

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

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

Remark: Applies also to identification for (classical) quantum channels
(Ahlsvede/Winter)

The blowing up technique

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

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

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

Blowing up Lemma (Ahlsvede/Gács/Körner, 1976)

For any DMC W there is a constant $c(W)$:

$$\forall x^n \in \mathcal{X}^n, B \subset \mathcal{Y}^n$$

$$W^n(\Gamma^k B | x^n) \geq \Phi(\Phi^{-1}(W^n(B | x^n))) + n^{-1/2}(k - 1)c$$

$$\text{if } \Phi(t) = \int_{-\infty}^t (2\pi)^{-1/2} e^{-u^2/2} du.$$

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

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

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

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

and

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

Remark: Presently only method to prove strong converse for transmission for (classical) **quantum** multiple-access channel (Ahlsvede/Cai).

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 $L(P_N)$ is approximately 2.

Actually it has been shown that

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

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

$$H_I(P) = 2 \left(1 - \sum_{u=1}^N P_u^2 \right) \quad (1.1)$$

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 $\mathcal{C} = \{c_1, \dots, c_N\}$ 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 $L_{\mathcal{C}}(P, u)$ of checkings?

Related quantities are

$$L_{\mathcal{C}} = \max_{1 \leq u \leq N} L_{\mathcal{C}}(P, u), \quad (1.2)$$

that is, the expected number of checkings for a person in the **worst case**, if code \mathcal{C} is used,

$$L(P) = \min_{\mathcal{C}} L_{\mathcal{C}}(P), \quad (1.3)$$

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

$$L_{\mathcal{C}}(P, Q) = \sum_{v \in \mathcal{U}} Q_v L_{\mathcal{C}}(P, v) \quad (1.4)$$

the average number of expected checkings, if code \mathcal{C} is used, and also

$$L(P, Q) = \min_{\mathcal{C}} L_{\mathcal{C}}(P, Q) \quad (1.5)$$

the average number of expected checkings for a best code.

A natural special case is the mean number of expected checkings

$$\bar{L}_{\mathcal{C}}(P) = \sum_{u=1}^N \frac{1}{N} L_{\mathcal{C}}(P, u), \quad (1.6)$$

which equals $L_{\mathcal{C}}(P, Q)$ for $Q = (\frac{1}{N}, \dots, \frac{1}{N})$, and

$$\bar{L}(P) = \min_{\mathcal{C}} \bar{L}_{\mathcal{C}}(P). \quad (1.7)$$

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

$$L(P, P) = \min_{\mathcal{C}} L_{\mathcal{C}}(P, P). \quad (1.8)$$

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 (see Example 3 below). It was noticed already in [4], [5] that a construction of code trees balancing probabilities like in the Shannon–Fano code is often better. In fact the Theorem of [5] establishes that $L(P) < 3$ for every $P = (P_1, \dots, P_N)$!

Still it is also interesting to see how well Huffman codes do with respect to identification, because of their classical optimality property. This can be put into the following

Examples for Huffman codes

We start with the uniform distribution

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

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

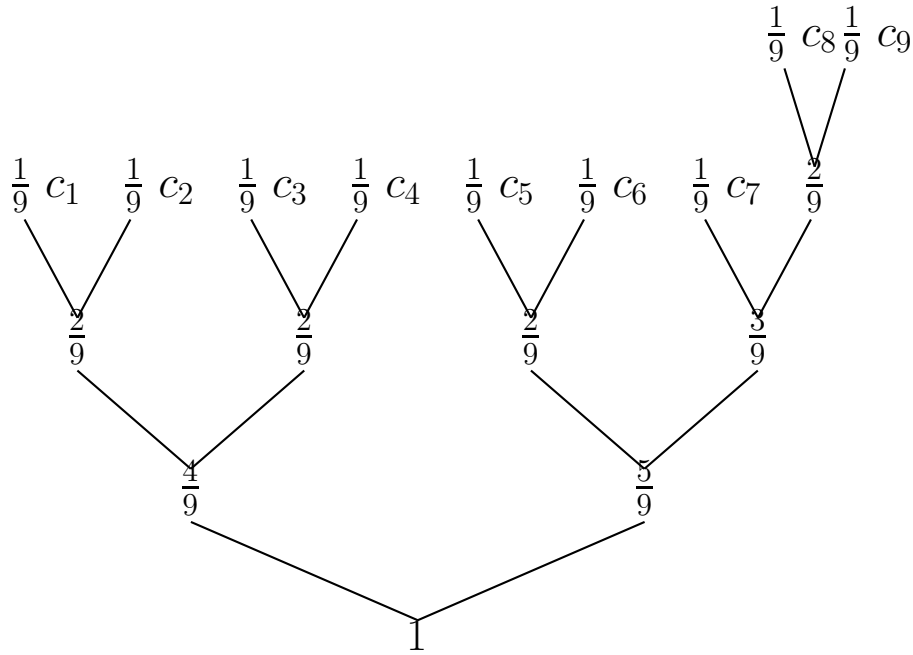
Then $2^{n+1} - N$ codewords have the length n and the other $2N - 2^{n+1}$ other codewords have the length $n + 1$ in any Huffman code. We call the $N - 2^n$ nodes of length n of the code tree, which are extended up to the length $n + 1$ **extended nodes**.

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 1: $N = 9$, $\mathcal{U} = \{1, 2, \dots, 9\}$, $P_1 = \dots = P_9 = \frac{1}{9}$.



Here $L_{\mathcal{C}}(P) \approx 2.111$, $L_{\mathcal{C}}(P, P) \approx 1.815$ because

$$L_{\mathcal{C}}(P) = L_{\mathcal{C}}(c_8) = \frac{4}{9} \cdot 1 + \frac{2}{9} \cdot 2 + \frac{1}{9} \cdot 3 + \frac{2}{9} \cdot 4 = 2\frac{1}{9}$$

$$L_{\mathcal{C}}(c_9) = L_{\mathcal{C}}(c_8), L_{\mathcal{C}}(c_7) = 1\frac{8}{9}, L_{\mathcal{C}}(c_5) = L_{\mathcal{C}}(c_6) = 1\frac{7}{9},$$

$$L_{\mathcal{C}}(c_1) = L_{\mathcal{C}}(c_2) = L_{\mathcal{C}}(c_3) = L_{\mathcal{C}}(c_4) = 1\frac{6}{9}$$

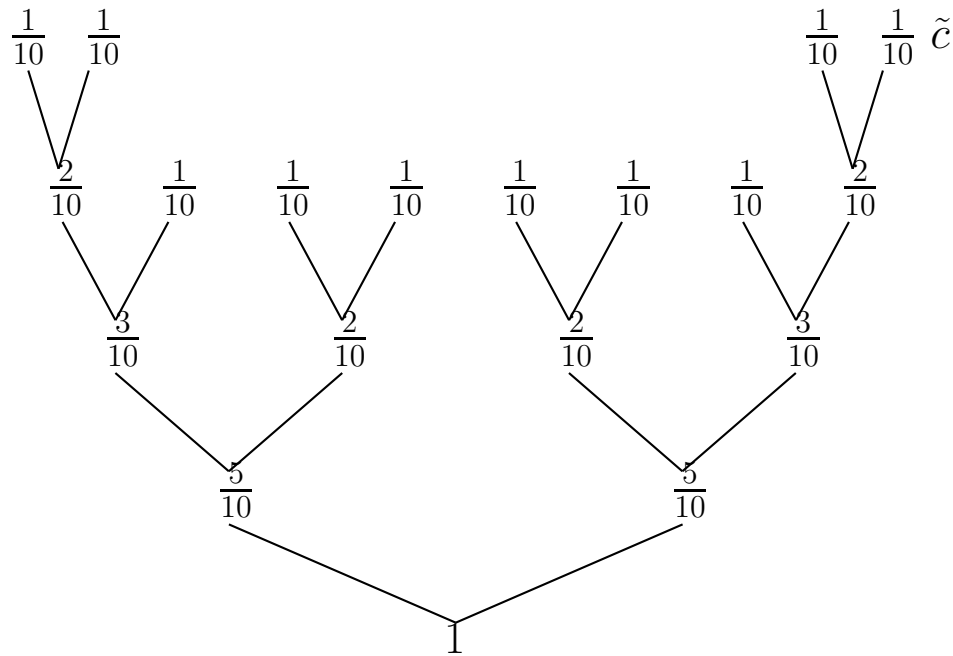
and therefore

Here $L_C(P) = 2.2$ and $L_C(P, P) = 1.880$, because

$$L_C(P) = 1 + 0.6 + 0.4 + 0.2 = 2.2$$

$$L_C(P, P) = \frac{1}{10}[1.6 \cdot 4 + 1.8 \cdot 2 + 2.2 \cdot 4] = 1.880.$$

One of the 16 best Huffman codes



Here $L_C(P) = 2.0$ and $L_C(P, P) = 1.840$ because

$$L_C(P) = L_C(\tilde{c}) = 1 + 0.5 + 0.3 + 0.2 = 2.000$$

$$L_C(P, P) = \frac{1}{5}(1.7 \cdot 2 + 1.8 \cdot 1 + 2.0 \cdot 2) = 1.840$$

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

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

Theorem: For $P^N = (P_1, \dots, P_N)$

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

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

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

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 $2 \left(1 - \frac{1}{2N} \right)$, which is better than the bound $2 \left(1 - \frac{1}{N^2} \right)$ for uniform distributions.

Finally we derive

Corollary.

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

It shows the lower bound of $L(P^n, P^N)$ by $H_I(P^N)$ 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

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

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 $\lim_{\alpha \rightarrow 1} S_\alpha(P) = H(P) = S_1(P)$, 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 **superextensivity**, **extensivity**, and **subextensivity**).

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 a forthcoming paper “An interpretation of identification entropy” the author and Ning Cai have discussed the coding theoretical meanings of the factors $\frac{q}{q-1}$ and $\left(1 - \sum_{u=1}^N P_u^2\right)$.

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 $P_U = P$, $P_V = Q$

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

B. Combinatorial Models

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

VIII. Updating Memories with cost constraint.

Optimal anticodes, binary constant weight equivalent to Erdős/Ko/Rado, 1938.

N	O	T	S	O	C	L	E	A	R
---	---	---	---	---	---	---	---	---	---

N	O	W	C	L	E	A	R	E	R
---	---	---	---	---	---	---	---	---	---

$$d = 7$$

Cost letterwise of transitions given by Hamming distance.

How many messages can be updated into each other, if cost $\leq c$?

The diametric theorem in Hamming Spaces

— optimal anticodes

R. Ahlswede and L. H. Khachatrian

For a Hamming space (\mathcal{X}_q^n, d_H) , the set of n -length words over the alphabet $\mathcal{X}_q = \{0, 1, \dots, q-1\}$ 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 $\{i, i+1, \dots, j\}$ 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

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

and $I(n, t)$ denotes the set of all such systems. Moreover, we define $I(n, k, t) = \left\{ \mathcal{A} \in I(n, t) : \mathcal{A} \subset \binom{[n]}{k} \right\}$.

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 Erdős, Ko, and Rado.

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

Theorem EKR *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

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

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 for $t \geq 15$ and subsequently by Wilson 1984 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 AK

Define $\mathcal{F}_i = \left\{ F \in \binom{[n]}{k} : |F \cap [1, t + 2i]| \geq t + i \right\}$ for $0 \leq i \leq \frac{n-t}{2}$.

For $1 \leq t \leq k \leq n$ with

(i) $(k - t + 1) \left(2 + \frac{t-1}{r+1}\right) < n < (k - t + 1) \left(2 + \frac{t-1}{r}\right)$ 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) $(k - t + 1) \left(2 + \frac{t-1}{r+1}\right) = n$ 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* (Erdős, Ko, Rado 1938)

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

Most recent result

- on intersecting families:

Ahlswede/Aydinian/Khachatrian, 2006.

with many further references

- and the new shifting technique:

Ahlswede/Aydinian/Khachatrian, More about shifting techniques, 2003.

- Local-global principle of Ahlswede/Cai:

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

Diametric Theorem

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

$$\max\{|\mathcal{A}| : \mathcal{A} \subset \mathcal{X}_q^n, \text{diam}(\mathcal{A}) \leq d\} =$$

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

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

Another diametric theorem in Hamming spaces: optimal group anticomodes

R. Ahlswede

In the last century together with Levon Khachatrian we established a diametric theorem in Hamming space $\mathcal{H}^n = (\mathcal{X}^n, d_H)$.

Now we contribute a diametric theorem for such spaces, if they are endowed with the group structure $\mathcal{G}^n = \sum_1^n \mathcal{G}$, the direct sum of group \mathcal{G} on $\mathcal{X} = \{0, 1, \dots, q-1\}$, and as candidates are considered which form a **subgroup** of \mathcal{G}^n .

For all finite groups \mathcal{G} , every permitted distance d , and all $n \geq d$ subgroups of \mathcal{G}^n with diameter d have maximal cardinality q^d .

Other extremal problems can also be studied in this setting.

A report on **Extremal Problems in Number Theory** and especially also in **Combinatorics**, which arose in Information Theory can be found in

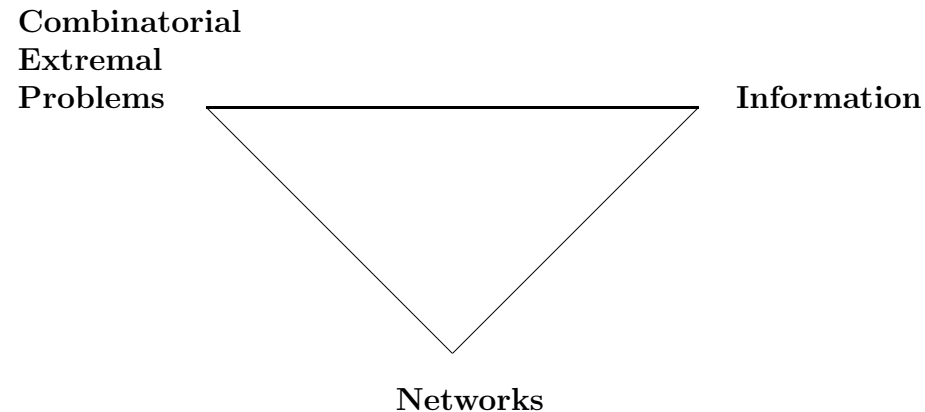
R. Ahlswede, Advances on extremal problems in Number Theory and Combinatorics, European Congress of Mathematicians, Barcelona 2000, Vol. I, 147.175, Progress in Mathematics, Vol. 201.

R. Ahlswede and V. Blinovsky, Modern Problems in Combinatorics, forthcoming book.

K. Engel, Sperner Theory, Cambridge University Press, Cambridge, 1997.

General Theory of Information Transfer and Combinatorics, Report on a Research Project at the ZIF (Center of interdisciplinary studies) in Bielefeld Oct. 1, 2001 – August 31, 2004, edit R. Ahlswede with the assistance of L. Bäumer and N. Cai, Lecture Notes in Computer Science, No. 4123.

IX Network Coding for Information Flows



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** - at a time where another giant John von Neumann was still fighting unreliability of systems by repetitions -

Shannon also wrote together with Peter Elias and Amiel Feinstein a basic paper on networks containing the L.R. Ford/D.R. Fulkerson - Min Cut - Max Flow Theorem, 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, \dots, s_n to their destinations, sets of node sinks D_1, \dots, 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 Multi-Casting.

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.

The **Min-Max-Theorem was discovered and proved for Information Flows** by Ahlswede, Cai, Li, and Yeung (2000).

Its statement can be simply explained. For one source only, that is $n = 1$, in the notation above, and $D_1 = \{d_{11}, d_{12}, \dots, d_{1t}\}$ 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 $\min_{1 \leq j \leq t} F_{1j}$ bits can go **simultaneously** to d_{11}, d_{12}, \dots and d_{1t} . Obviously, this is best possible. It has been referred to as ACLY-Min-Max-Theorem (**It also could be called Shannon's Missed 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 Multicasting allowing only copies.

Network Flows with **more than one source** are much harder to analyze and lead to a wealth of old and new Combinatorial Extremal problems. **This is one of the most striking examples of an interplay between Information Transfer and Combinatorics.**

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 our characterization of optimal **Anticodes** obtained with the late Levon Khachatrian arises in such a role!

On the classical side for instance orthogonal **Latin Squares** - on which Euler went so wrong - arise.

The Min-Max-Theorem has been made practically more feasible by a polynomial algorithm by Peter Sanders, Sebastian Egner and Ludo Tolhuizen as well as by his competitors (or groups of competitors) in other parts of the world, leading to the joint publications.

With NetCod 2005 - the first workshop on Network Coding Theory and Applications, April 7, 2005, Riva, Italy the **New Subject Network Coding** was put to start.

Research into network coding is growing fast, and Microsoft, IBM and other companies have research teams who are researching this new field.

A few American universities (Princeton, MIT, Caltec and Berkeley) have also established research groups in network coding.

The holy grail in network coding is to plan and organize (in an automated fashion) network flow (that is to allowed to utilize network coding) in a feasible manner. Most current research does not yet address this difficult problem.

There may be a great challenge not only coming to **Combinatorics** but also to **Algebraic Geometry** and its present foundations.

An Introduction to the area of Network Coding is given in the book of R. Yeung.

The case $|D_i \cap D_j| = \emptyset$ for $i \neq j$ and $|D_i| = 1$ for $i = 1, \dots, n$, that is, **each source sends its message to its sink** has an obvious symmetry and appeal. Soren Riis established the equivalence of this flow problem to a **guessing game**, which is **cooperative**.

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 $Q = \{0, 1, \dots, q-1\}$ 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 at most 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.

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

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 $\{u_i(E) : 1 \leq i \leq M, E \in \mathcal{E}_t\}$ 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 $e' \in V(E')$

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

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 for the binary case to equal $H_2(\tau)$. For general q the best known result is

Theorem Ahlswede/Bassalygo/Pinsker

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

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

Competing Ideas:

Ahlsweide: With increase of q the Hamming space should become more flexible for packing.

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

Input alphabet $\mathcal{X} = \mathcal{Q}$ and output alphabet $\mathcal{Y} = \mathcal{Q}$.

$M_f(n, t, q)$ 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 ($j = 1, 2, \dots, n$).

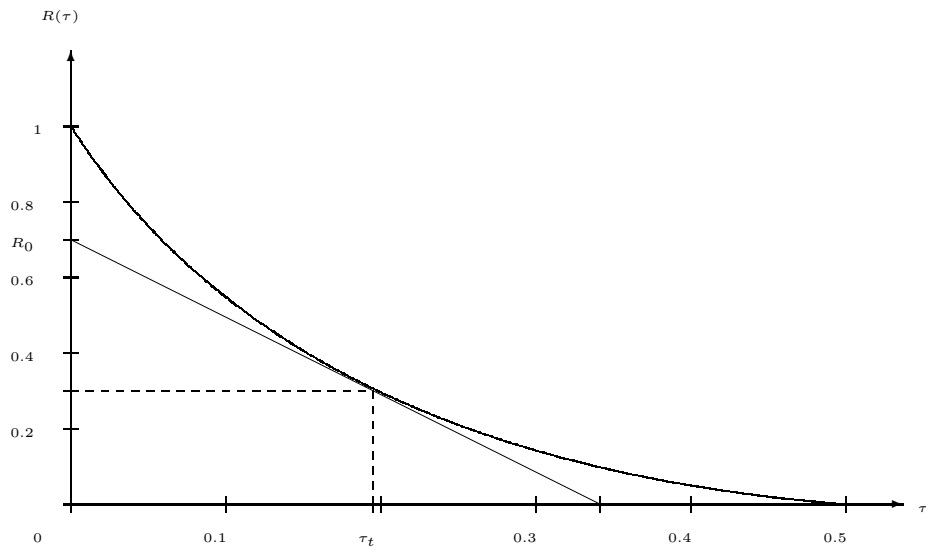
Relative error $\tau = t/n$ and rate $R = n^{-1} \log M$.

$C_q^f(\tau)$ the supremum of the rates achievable for τ and all large n (capacity error function).

Theorem[Berlekamp 64, Zigangirov 76]

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

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



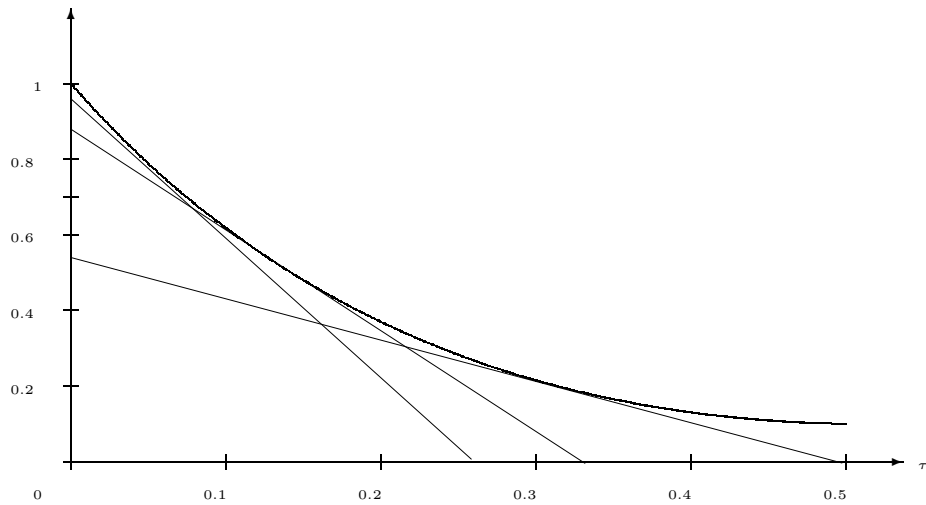
Theorem[Ahlsvede, Deppe, Lebedev, Annals of the EAS, 2006]

Let $q \geq 3$

(i)

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

(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

$$b : \mathcal{M} \rightarrow \{1, 2, \dots, q - 1\}^{n-2t}$$

bijection between messages and of 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 $i \leq n$ not needed dummy $x_i = 1$ 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 ($x_p \rightarrow y_p \in \{1, 2, \dots, q - 1\}$), 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 ($x_p \rightarrow y_p = 0$), 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.**

Example: $n = 5, t = 2, q = 3$

$b(1) = 1, b(2) = 2$

Let $i = 1$:

sent: **10111** **10101** **10001**

received: **20111** **20201** **22001**

r -rubber method:

skeleton: $\{x^{n-(r+1)t} \in \{0, 1, \dots, q-1\}^{n-(r+1)t} : \text{the sequence contains } \leq r-1 \text{ consecutive zeros} \}$

protocol string: 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 $(\frac{1}{r+1}, 0)$.

Ratewise-optimal non-sequential search strategies under constraints on the tests

R. Ahlswede

Already in his Lectures on Search **Rényi** suggested to consider a search problem, where an unknown $x \in \mathcal{X} = \{1, 2, \dots, n\}$ is to be found by asking for containment in a minimal number $m(n, k)$ of subsets A_1, \dots, A_m with the restrictions $|A_i| \leq k < \frac{n}{2}$ for $i = 1, 2, \dots, m$.

Katona gave in 1966 the lower bound $m(n, k) \geq \frac{\log n}{h(\frac{k}{n})}$ in terms of binary entropy and the upper bound $m(n, k) \leq \left\lceil \frac{\log n+1}{\log n/k} \right\rceil \cdot \frac{n}{k}$, which was improved by Wegener in 1979 to $m(n, k) \leq \left\lceil \frac{\log n}{\log n/k} \right\rceil (\lceil \frac{n}{k} \rceil - 1)$.

We prove here for $k = pn$ that $m(n, k) = \frac{\log n + o(\log n)}{h(p)}$, **that is, ratewise optimality** of the entropy bound.

Quite surprisingly, eventhough this result is known for several decades nobody proved that – or even seems to have wondered whether – it is essentially best possible for instance for k -set tests “carrying $h(\frac{k}{n})$ bit of information”.

However, when we looked for a proof we realized an obstacle, which blocked the development even for people who may have believed in the entropy bound. The known proofs for upper bounds (Theorem KW in Section 3) are constructive and apparently hard to improve. In such a situation often a probabilistic argument helps. However, a **standard approach by random choice is suboptimal** even for the simple case of unrestricted tests as was noticed already by Renyi 1965. Using the uniform distribution for choosing a separating system (see Section 2) requires

$$m \geq 2 \log n + 6$$

sets, **where** $\lceil \log n \rceil$ **is optimal**. So we are by a factor of 2 away from the optimum!

XII. Combi-Probabilistic Models

Coloring Hypergraphs did a problem by Gallager

Slepian/Wolf Model 1973

for DMCS $((X^n, Y^n))_{n=1}^\infty$.

$$Y^n \longrightarrow X^n$$

Encoding $f : \mathcal{Y}^n \rightarrow \mathbb{N}$

Decoding $g : \mathcal{X}^n \times \mathbb{N} \rightarrow \mathcal{X}^n \times \mathcal{Y}^n$

$$Prob(g(X^n, f(Y^n)) = (X^n, Y^n)) \sim 1 \quad (1)$$

Opt. Rate(f) = $H(Y|X)$

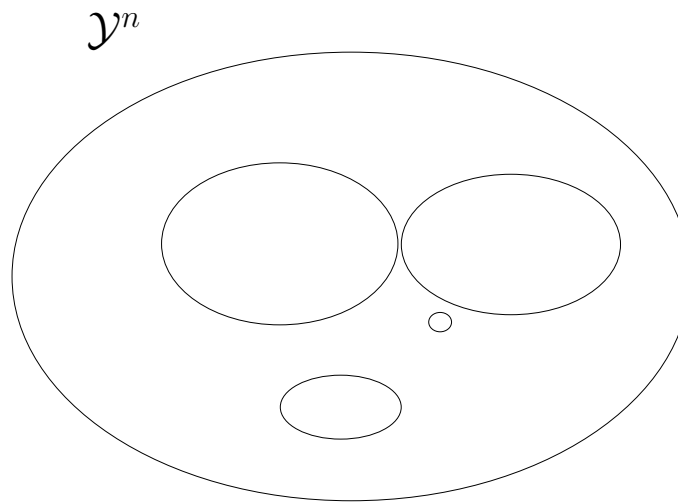
Gallager Model 1976

for discrete, memoryless conditional distribution

$(\{Y^n(x^n) : x^n \in \mathcal{X}^n\})_{n=1}^\infty$ (Generic $P_{Y|X}$)

$$Prob(g(x^n, f(Y^n)) = (x^n, Y^n)) \sim 1 \quad \forall x^n \in \mathcal{X}^n \quad (2)$$

$$\text{Opt. Rate}(f) = \max_x H(Y|X = x)$$



Hypergraph $(\mathcal{Y}^n, \{Carrier(x^n) : x^n \in \mathcal{X}^n\})$

must be close to 1:1 on all edges

(called binning or coloring).

They have sizes between exponential in n and constant numbers.

Therefore RANDOM SELECTION fails.

Our solution already in

Channel Capacities for List Codes **1973**

by a counting argument and in

Coloring hypergraphs: A new approach to multi-user source I,II 78/79

by **combined greedy/random selection.**

C. Further Perspectives

a. Protocol Information (Gallager?)

“Protocol” information ought to be investigated more deeply. We encountered it in the Theory of Localized Errors and in the Rubber Method.

b. Beyond Information Theory:

Identification as a New Concept of Solution for Probabilistic Algorithms

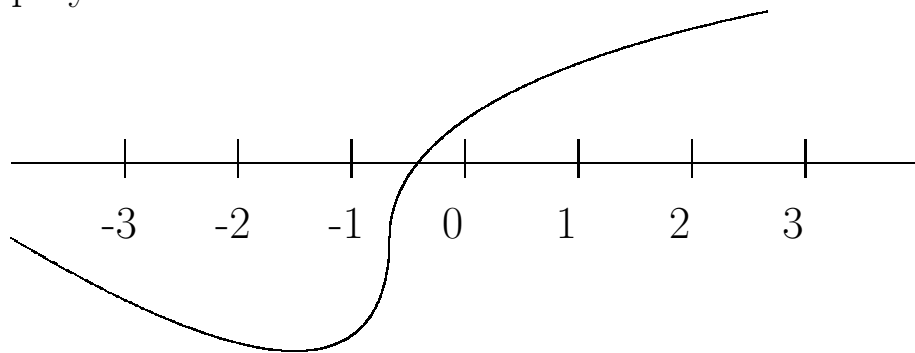
We mention as perhaps one of the most promising directions the study of probabilistic algorithms with identification as *concept of solution*.

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

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



c. A new connection between information inequalities and Combinatorial Number Theory (Tao)

The final form of Tao's inequality relating conditional expectation and conditional mutual information

R. Ahlswede

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 , and X, X' be discrete 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 $(2\ell n 2)^{\frac{1}{2}}$ and that this is the best possible constant.

d. A 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 my 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 "A channel with two senders and two receivers", 1974 I showed that the region obtained with independent sender's distributions is generally smaller than the capacity region.**

e. Could we ask Shannon's advise !!!

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.

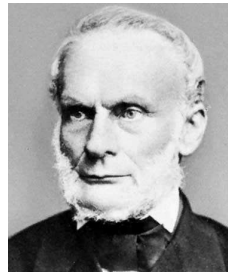
"In a somewhat dual way, evaluating the rate-distortion function $R(D)$ for a source amounts, mathematically, to *minimizing* a mutual Information under variant of the $q_i(j)$, again with a linear inequality constraint. The solution leads to a function $R(d)$ which is *convex* downward. Solving this 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 now our ideas about creating order involving knowledge of past and future and wonder what Shannon would think about them. **They are motivated by Clausius' second law of thermodynamics. He also introduced entropy.**

“Clausius, Rudolf”, Encyclopaedia Britannica. 2006.

Clausius, Rudolf (Julius Emanuel) born January 2, 1822, Koflin, Prussia died August 24, 1888, Bonn



German mathematical physicist **who formulated the second law of thermodynamics and is credited with making thermodynamics a science.**

Clausius was appointed professor of physics at the Artillery and Engineering School at Berlin in 1850, the same year in which he presented a paper stating the second law of thermodynamics in the well-known form: **“Heat cannot of itself pass from a colder to a hotter body.”** He applied his results to an exhaustive development of the theory of the steam engine, **stressing the concept of entropy** (dissipation of available energy). He became professor of physics at Zurich Polytechnikum in 1855, and, two years later, contributed to the theory of electrolysis (the breaking down of a compound by electricity) by **suggesting that molecules are made up of continually interchanging atoms and that electric force does not cause but simply directs the interchange.** This view later was used as the basis of the **theory of electrolytic dissociation** (breakdown of molecules into charged atoms or ions).

He became professor of physics at the University of Wurzburg in 1867 and at the University of Bonn in 1869. In molecular physics, Clausius restated the French physicist Sadi Carnot’s principle concerning efficiency of heat engines and thus provided a much sounder basis for the theory of heat.

Entropy

Rudolf Clausius

$$\Delta S = \frac{\Delta Q}{T}$$

Einstein: Deepest law in physics

Boltzmann

$$- \sum p_i \log p_i$$

(empirical distribution = composition = type = complexion)

Shannon

$$I(x \wedge y)$$

individual information

Identification Entropy

$$\frac{q}{q-1} \left(1 - \sum_{i=1}^N p_i^2 \right)$$

“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 uncertainty). 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.”

A. Renyi, Probability Theory, North Holland, Amsterdam, p. 554, 1970.

Creating order in sequence spaces

People spend a large amount of time creating order in various circumstances. Our aim is to start or to 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”.

The non-probabilistic model

We have a box that contains β objects at time t labeled with numbers from $\mathcal{X} = \{1, \dots, \alpha\}$. 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 behaviour 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 $f_n : \mathcal{X}^n \times \mathcal{S} \rightarrow \mathcal{X}^n$ 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**

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

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

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

Furthermore the following relations hold.

$$\begin{aligned} \omega_2(\infty, \beta, \varphi) &= \nu_2(\infty, \beta, \varphi), & \omega_2(\pi, \beta, \infty) &= \nu_2(\pi, \beta, \infty), \\ \lim_{\beta \rightarrow \infty} \nu_3(0, \beta, 0) &= 1, & \tau_2(\pi, \beta, \varphi) &= \nu_2(\infty, \beta, \varphi) \text{ for } \pi \geq 1, \\ \tau_2(0, 2, 0) &= \log((\sqrt{5} + 1)/2). \end{aligned}$$

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 $\lambda^\beta - \lambda^{\beta-1} - 1 = 0$.

Conjecture 1. $\lim_{\varphi \rightarrow \infty} \nu_2(\pi, \beta, \varphi) \neq \nu_2(\pi, \beta, \infty)$.

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

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

A probabilistic model

The initial content of the box and the input $(X_t)_{t=1}^n$ are produced by i.i.d. random variables with generic distribution P_X . This gives rise to an output process $(Y_t)_{t=1}^n$. The constraints on matter and mind are again meaningful. The performance of a strategy f is now measured by the entropy $H(Y^n)$ and the mean entropy $\overline{H}_f = \limsup_{n \rightarrow \infty} (1/n)H(Y^n)$ to be minimized.

Define

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

For this model we obtained the following result. Consider the events $E_k = \{Y^k = 01010 \dots\}$ and $D_k = E_k \setminus E_{k+1}$. Denote $q(k) = \text{Prob}(D_k)$.

Then for $P_X(0) = P_X(1) = 1/2$ we have

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

Towards a theory of creating order

1. Directions of developments of our basic model for sequences
 - (a) Multiple in- and outputs: s inputs and s outputs, varying number of outputs, merging, splitting, correlation
 - (b) Objects with special features: Varying-length objects, death–birth, idle objects, box with exclusion rule
 - (c) Compound objects: Box with reaction rules, representatives, objects with many properties, exchanging parts of objects
 - (d) Errors: Probabilistic, confusion rule, frequency rule, receiver can distinguish only certain objects
2. Examples: Production of goods, arrival of goods and documents, garbage collection
3. Ordering and source coding
4. Ordering, sorting and Maxwell’s demon
5. A calculus of machines: Comparisons of machines, commutativity
6. Why do we want to create order?

Ahlsvede/Zhang, Contributions to a theory of ordering for sequence spaces, 1989

Ahlsvede/Ye/Zhang, Creating order in sequence spaces with simple machines, 1990