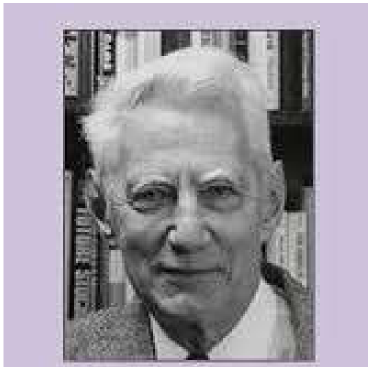
Shannon and my research

Where did I use the ideas of Claude Elwood Shannon?

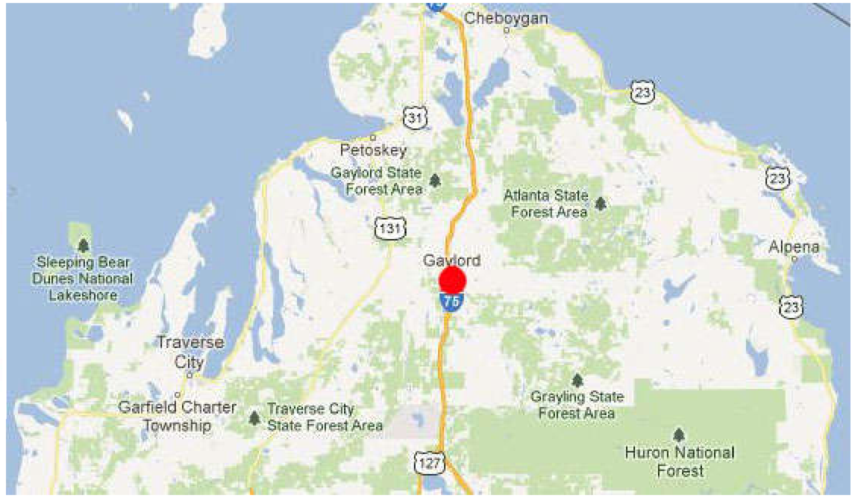
At the occasion of his 100th birthday, 2016

Han Vinck



Claude Shannon
"Father of the Information Age"

The historical perspective



- Born 1916, Shannon was almost a Canadian (Vijay Bhargava)
- Master thesis: 1937 (age 21!) - *A Symbolic Analysis of Relay and Switching Circuits*,
- „Work“ for PhD: 1940 - *An Algebra for Theoretical Genetics*,
- Visit: 40/41 the Institute for Advanced Study, in Princeton
- Work at Bell labs: 1941 – 1958
  - 1948 publiced hij *A Mathematical Theory of Communication*
  - 1949 *Communication Theory of Secrecy Systems*
- „Work“ at MIT: 1959 - 1978

Claude Shannon Gaylord

statue

Gaylord High School

This Certifies That

Claude E. Shannon

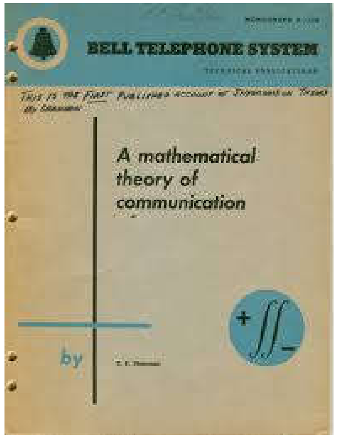
Academic

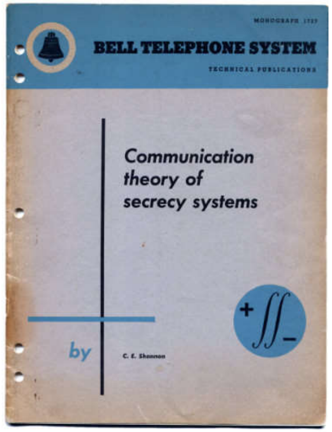
DIPLOMA

In Witness Whereof

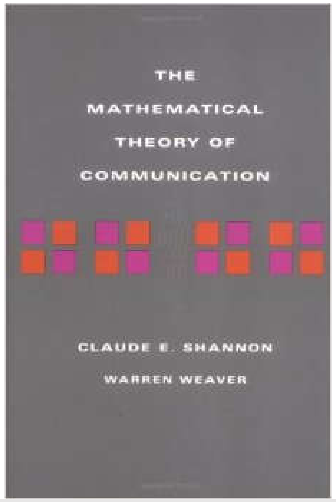
Later: flame-throwing trumpet.

Some pictures



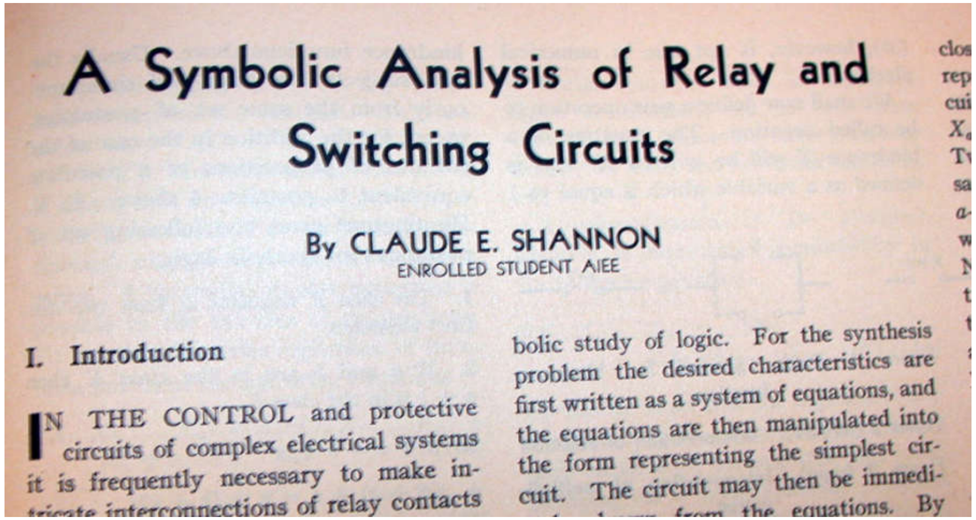


"transformed cryptography from an art to a science."



The book co-authored with Warren Weaver, *The Mathematical Theory of Communication*, reprints Shannon's 1948 article and Weaver's popularization of it, which is accessible to the non-specialist.[5] In short, Weaver reprinted Shannon's two-part paper, wrote a 28 page introduction for a 144 pages bool d changed the title from "A mathematical theory..." to "The mathematical theory..."

It all started with a master thesis!(1936)



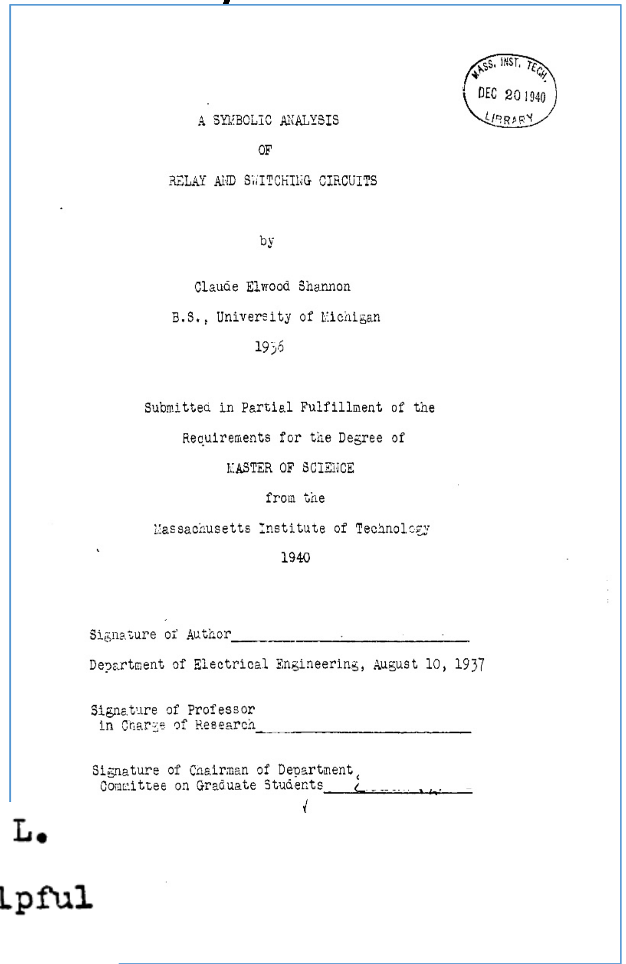
A Symbolic Analysis of Relay and Switching Circuits

By CLAUDE E. SHANNON
ENROLLED STUDENT AIEE

I. Introduction

IN THE CONTROL and protective circuits of complex electrical systems it is frequently necessary to make intricate interconnections of relay contacts

bolic study of logic. For the synthesis problem the desired characteristics are first written as a system of equations, and the equations are then manipulated into the form representing the simplest circuit. The circuit may then be immediately drawn from the equations. By



A SYMBOLIC ANALYSIS
OF
RELAY AND SWITCHING CIRCUITS

by

Claude Elwood Shannon
B.S., University of Michigan
1936

Submitted in Partial Fulfillment of the
Requirements for the Degree of
MASTER OF SCIENCE
from the
Massachusetts Institute of Technology
1940

Signature of Author
Department of Electrical Engineering, August 10, 1937

Signature of Professor
in Charge of Research

Signature of Chairman of Department
Committee on Graduate Students

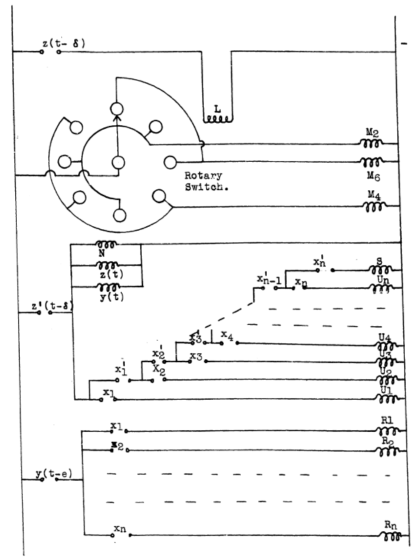
It’s the diagrams used in the final chapter of the thesis, which showed different types of circuits, that contained the central circuit that is still used in digital computers. The circuit is the 4-bit full adder.

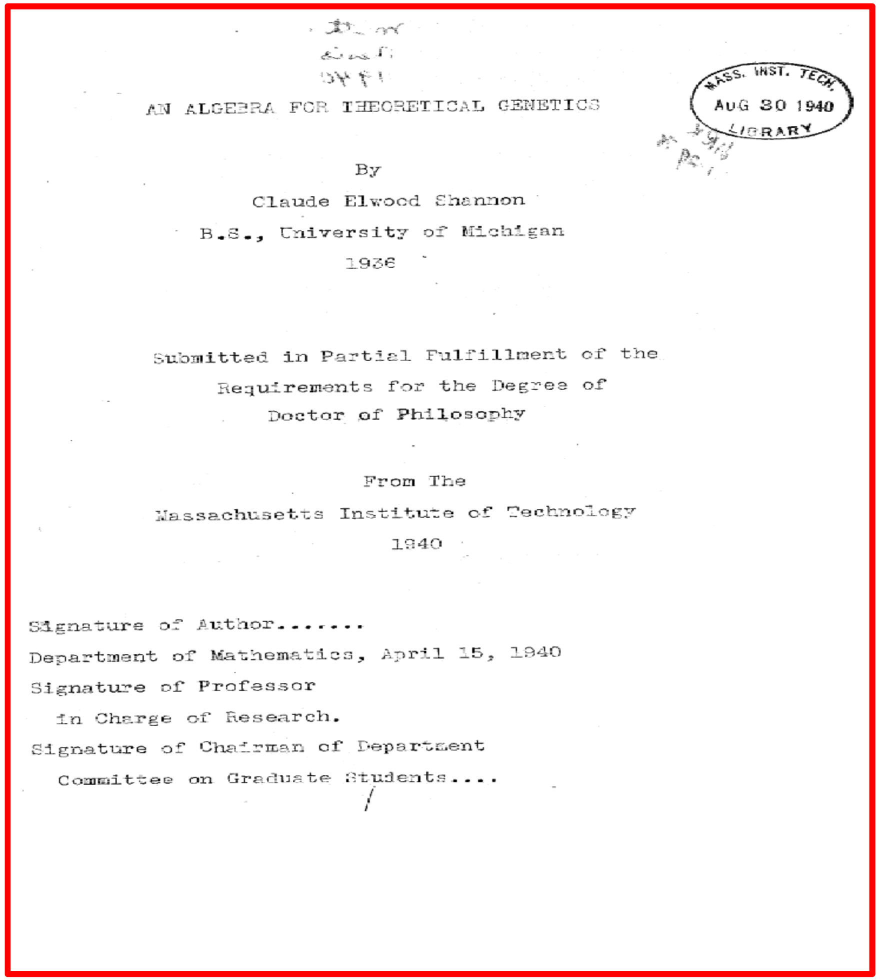
The author is indebted to Professor F. L. Hitchcock, who supervised the thesis, for helpful criticism and advice.

For „some“ reason, signatures were remove

Master thesis described a machine to find prime numbers

As to the practicability of such a device, it might be said that J.P. Kulik spent 20 years in constructing a table of primes up to 100,000,000 and when finished it was found to contain so many errors that it was not worth publishing. The machine described here could probably be made to handle 5 numbers per second so that the table would require only about 2 months to construct.





AN ALGEBRA FOR THEORETICAL GENETICS

By

Claude Elwood Shannon

B.S., University of Michigan

1936

Submitted in Partial Fulfillment of the
Requirements for the Degree of
Doctor of Philosophy

From The
Massachusetts Institute of Technology
1940

Signature of Author.......
Department of Mathematics, April 15, 1940
Signature of Professor
in Charge of Research.
Signature of Chairman of Department
Committee on Graduate Students....

Apparently, Shannon spent only a few months on the thesis.

With his creativity, if Shannon had stayed in population genetics, he would surely have made some important contributions. Nevertheless, I think it is fair to say that the world is far better off for his having concentrated on communication theory, where his work was revolutionary.

Because the thesis was unpublished, it had no impact on the genetics community.

Shannon also wrote a PhD thesis: who knows about it?

1956 **RELIABLE CIRCUITS USING LESS RELIABLE RELAYS**

BY

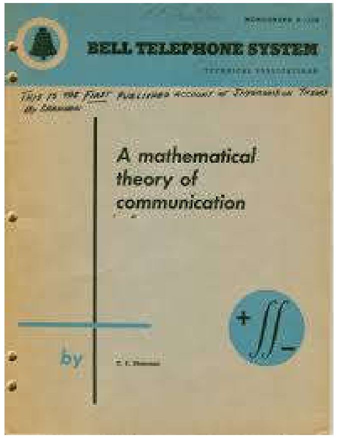
E. F. MOORE[1] **AND C. E. SHANNON**[1]

Shannon Married (2) E(lizabeth) Moore

1130	Bode H W, Mathematics	MH-236
1910	Wiggins Miss M F, Secretary	MH-201
1130	Darlington S	MH-978
	Blackman R B	MH-205
	Lakatos E	MH-437
	Ling D P	MH-234
	Zobel O J	MH-214
	Dietzold R L	MH-204
	Angell Miss D T	MH-202
	Hamming R W	MH-209
	McMillan B	MH-407
	Shannon C E	MH-209
	Schelkunoff S A	636
	Gray Miss M C	467
	MacColl L A	MH-215
	Shewhart W A	MH-239
	Harold Miss M S	MH-202
	Packer Miss M C	MH-427
	Tukey J W	MH-566
	Froelich Miss C L	MH-213
	Asbury Miss J G	MH-213
	Cooper Mrs H L	1464
	Moore Miss M E	1972 MH- 357
	Pecon Miss P A	MH-357
	Sumoska Miss H	1167
	Weiss Miss R A	MH-207
1140	MacNair W A, Military Research	MH-722
1910	Nimmo Miss P E, Secretary	MH-581
1140	Burger M J	MH-305
	Buntenbach R W	MH-219
	Clement G F	MH-219
	Knapp W A	MH-265
	Oestreicher J J	MH-792

Transmission problem

The fundamental problem of communication is that of reproducing at one point either exactly or approximately a message selected at another point. Frequently the messages have *meaning*; that is they refer to or are correlated according to some system with certain physical or conceptual entities. These semantic aspects of communication are irrelevant to the engineering problem. The significant aspect is that the actual message is one *selected from a set* of possible messages. The system must be designed to operate for each possible selection, not just the one which will actually be chosen since this is unknown at the time of design.



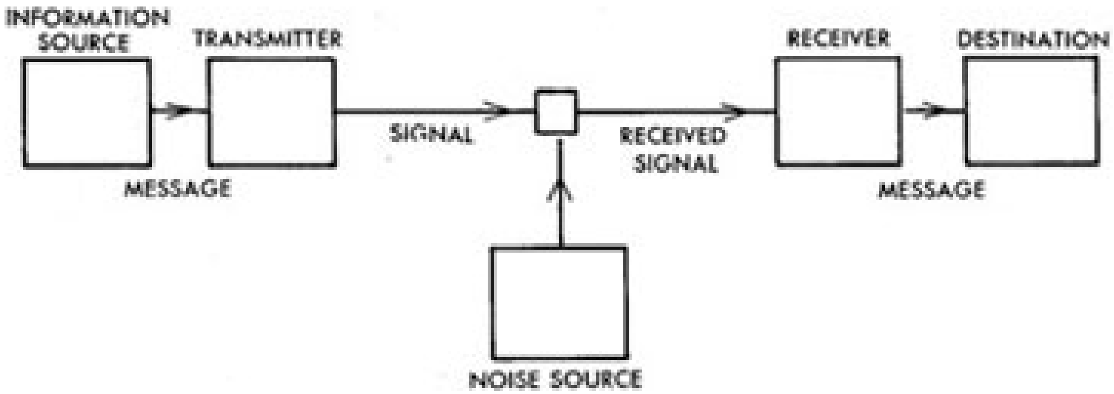


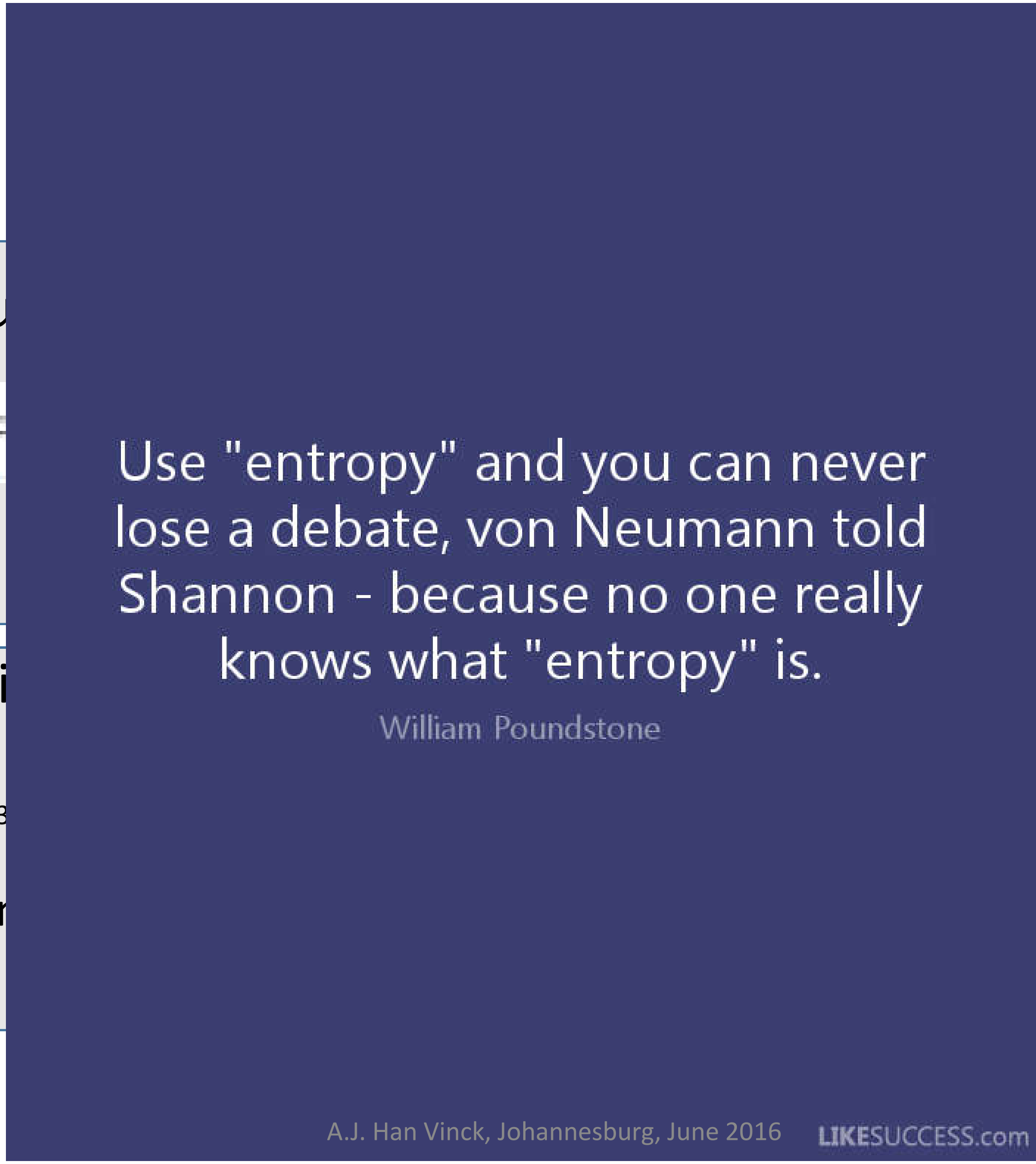
FIGURE 1

ENTROPY … Y

- Entropy:= minimu… of a source

We shall call H = … babilities $p_1, \ldots, p_n$.

  - Shannon 1948

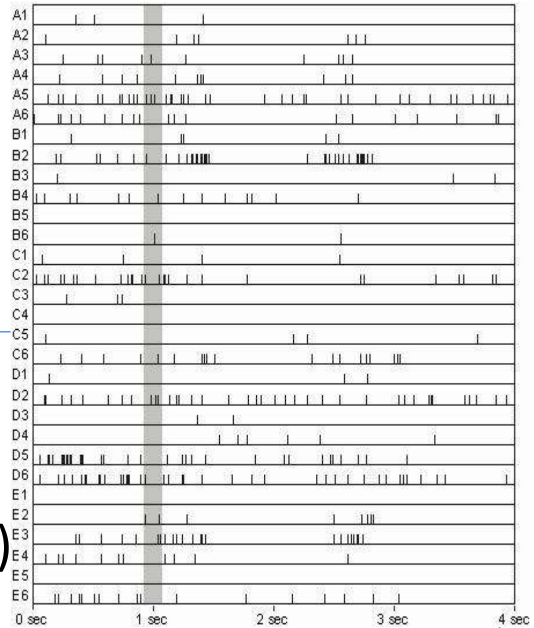


- Source coding: mi… o Coding)

Example: p_1, p_2, p_3 … sentation length H = 3/2

Applications: Intern…

Problem entropy estimation



- How to estimate the entropy for:
  - Limited number of samples (spike trains in neuro science)

$$\hat{H}_{\text{plugin}}(n) = -\sum_{m=1}^{M} \frac{n_m}{n} \ln\left(\frac{n_m}{n}\right), \quad \mathrm{E}\{-\frac{n_m}{n}\ln(\frac{n_m}{n})\} \le -\mathrm{E}\{\frac{n_m}{n}\}\ln(\mathrm{E}\{\frac{n_m}{n}\})$$

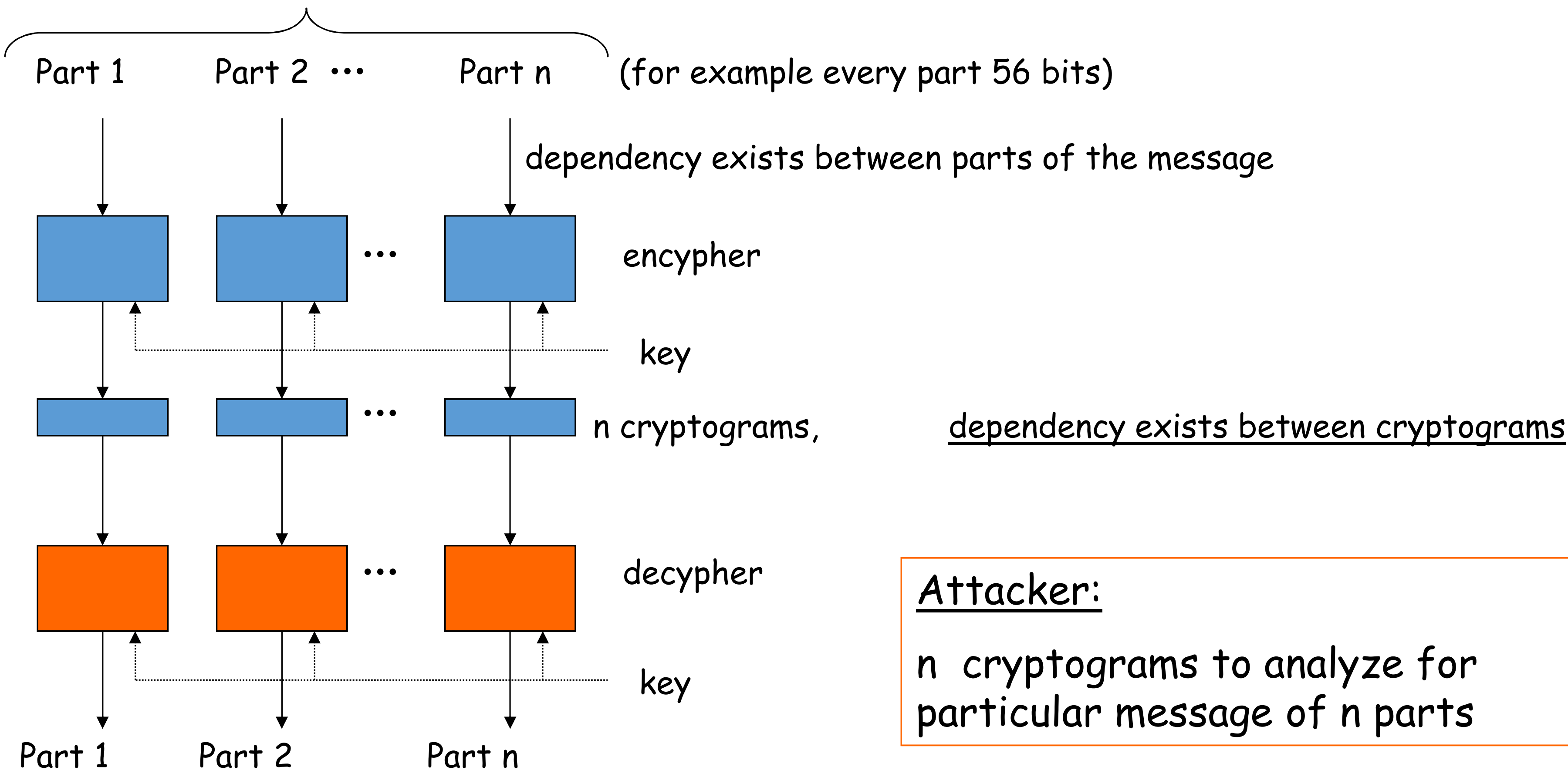
  - Sources with unknown memory

$$H_n = -\frac{1}{n}\sum_{i,\, j,\ldots,s} p(i,\, j,\cdots,s)\log_2 p(i,\, j,\cdots,s) \qquad (42)$$

$$H = \lim_{n\to\infty} H_n .$$

Estimation of the entropy based on its polynomial representation, Phys. Rev. E 85, 051139 (2012) [9 pages], Martin Vinck, Francesco P. Battaglia, Vladimir B. Balakirsky, A. J. Han Vinck, and Cyriel M. A. Pennartz

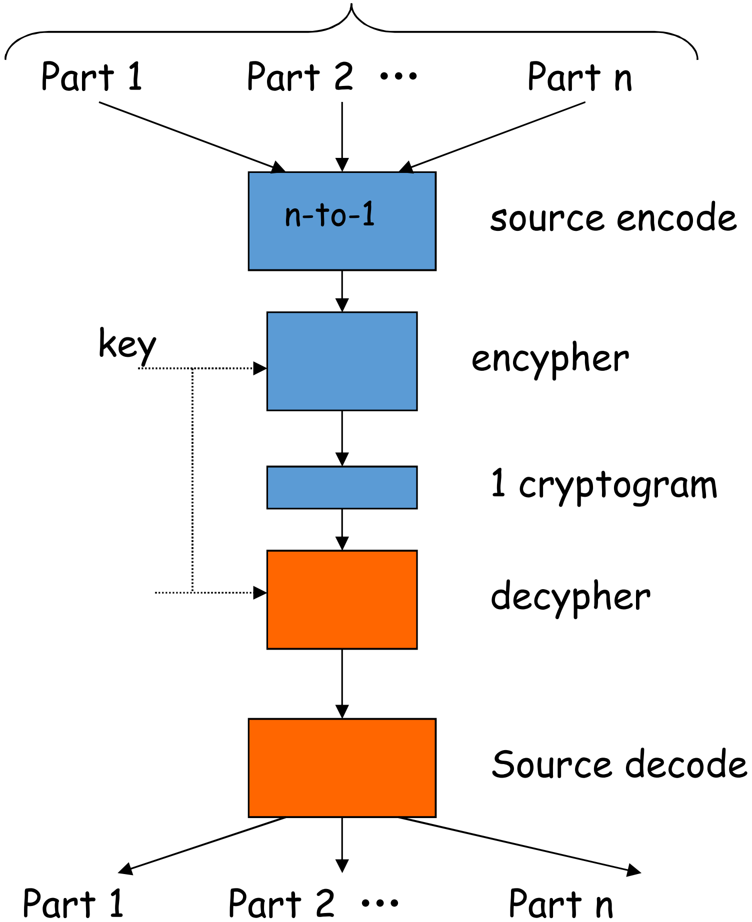
Message encryption without source coding



Attacker:

n cryptograms to analyze for particular message of n parts

Message encryption with source coding



(for example every part 56 bits)

Attacker:

- 1 cryptogram to analyze for particular message of n parts
- assume data compression factor n-to-1

Hence, less material for the same message!

Channel capacity: = maximum <u>reduction</u> in representation length with $P_e \leq \varepsilon$!

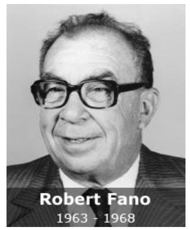
<u>Information</u>: = Entropy before transmission H(X) – Entropy after transmission $H_y(X)$

The capacity C of a noisy channel has been defined as

$$C = \mathrm{Max}\big(H(x) - H_y(x)\big) \qquad = H(Y) - H_X(Y)$$

where x is the input and y the output. The maximization is over all sources which might be used as input to the channel.

I(X;Y), Fano



Theorem 17: The capacity of a channel of band W perturbed by white thermal noise power N when the average transmitter power is limited to P is given by

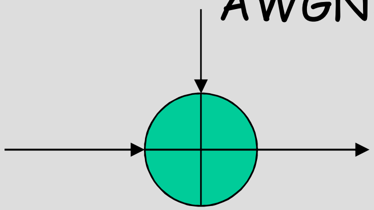
$$C = W \log \frac{P+N}{N}. \qquad = H(Y) - H_X(Y)$$

This means that by sufficiently involved encoding systems we can transmit binary digits at the rate $W \log_2 \frac{P+N}{N}$ bits per second, with arbitrarily small frequency of errors. It is not possible to transmit at a higher rate by any encoding system without a definite positive frequency of errors.

Capacity for AWGN

AWGN: σ_G^2

Gaussian input X



Output Y (also Gaussian)

$\sigma_X^2 \leq S/2W$

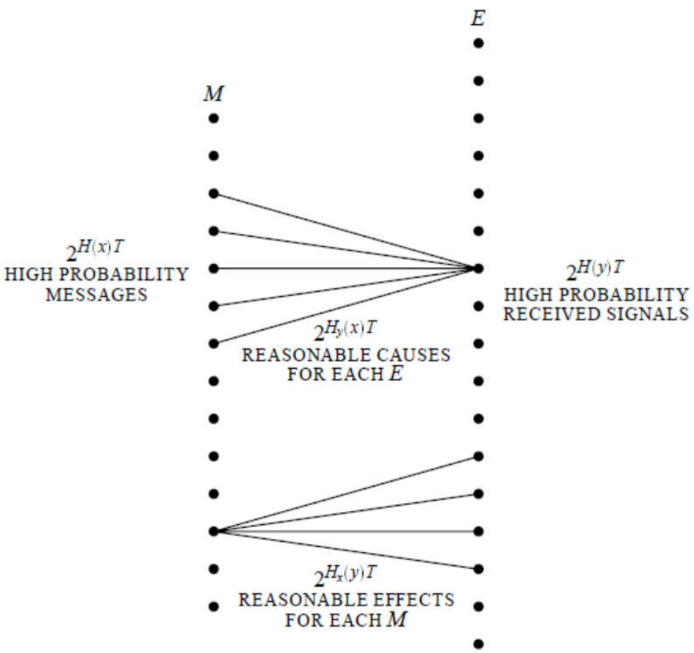
W is (single sided) bandwidth
S is average power

$\sigma_Y^2 = \sigma_X^2 + \sigma_G^2$

$$\text{Capacity} = W\log_2\left(1+\frac{\sigma_X^2}{\sigma_G^2}\right) \text{ bits/sec.} \quad \leq W\log_2\left(1+\frac{S/2W}{\sigma_G^2}\right)$$

Claude Shannon

Capacity achieving Codes exist (Shannon, 1948)



. 10—Schematic representation of the relations between inputs and outputs in a channel.

Sketch of proof:

Encoding: use a **random** code

Decoding:

1. look for a „closest“ code word (P_{ERROR} => 0, law of large numbers)

2. Probability that another codeword is in the decoding region => 0 (random code word selection)

The first rigorous proof for the discrete case is due to Amiel Feinstein in 1954.

Mathematicians did not like this (engineering) approach!

Capacity powerlimited channel (PLC channel)

Theorem 20: The channel capacity C for a band W perturbed by white thermal noise of power N is bounded by

$$C \geq W \log \frac{2}{\pi e^3} \frac{S}{N},$$

where S is the peak allowed transmitter power. For sufficiently large $\frac{S}{N}$

$$C \leq W \log \frac{\frac{2}{\pi e} S + N}{N} (1 + \epsilon)$$

Example CENELEC)

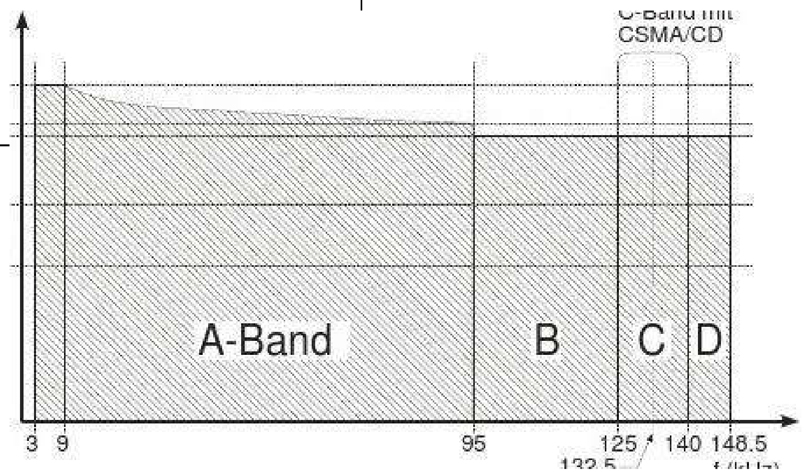
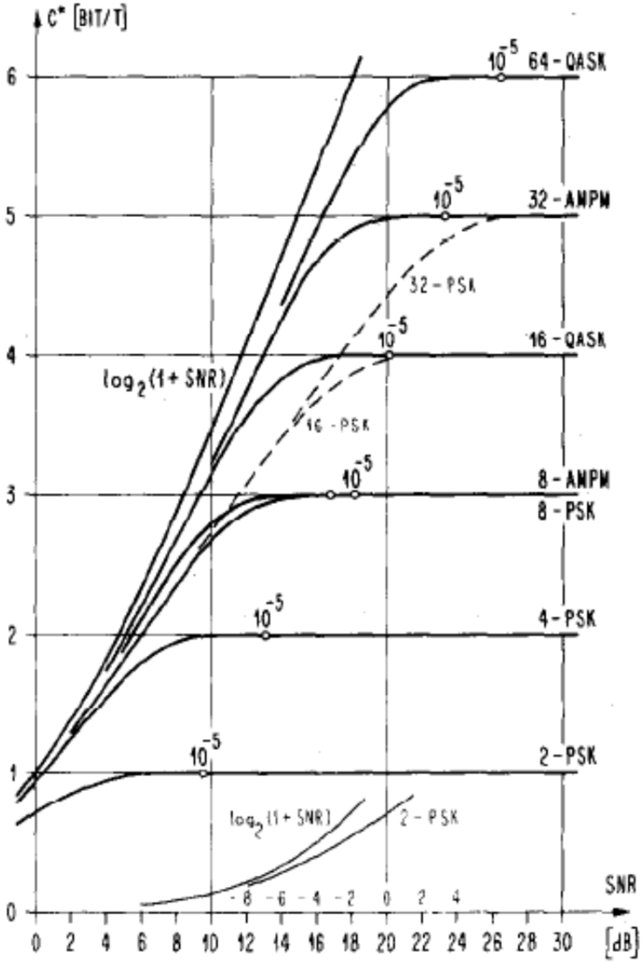


Figure 1: Maximum output level in the frequency range 3 kHz to 148.5 kHz in dB (µV)

Capacity over Gaussian inputs?

Channel Coding with Multilevel / Phase Signals

GOTTFRIED UNGERBOECK, MEMBER, IEEE



(b)

Fig. 2. Channel capacity C^* of bandlimited AWGN channels

Rate distortion theory

Coding Theorems for a Discrete Source With a Fidelity Criterion*

Claude E. Shannon**

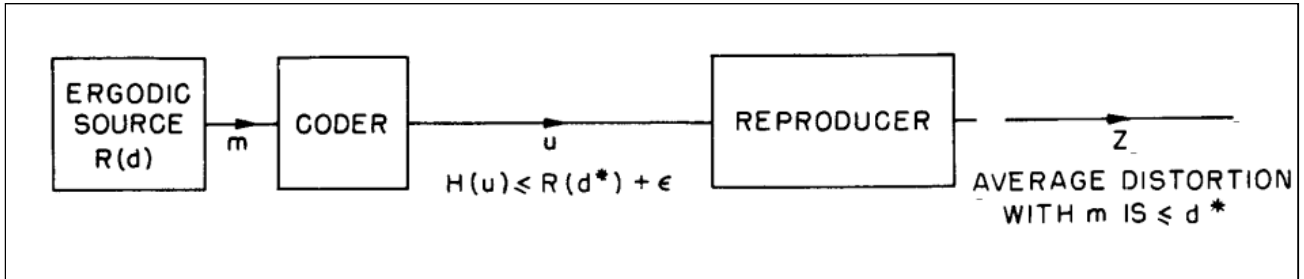


Fig. 4.

* Institute of Radio Engineers, *International Convention Record*, vol. 7, 1959.

** This work was supported in part by the U.S. Army (Signal Corps), the U.S. Air Force (Office of Scientific Research, Air Reserve and Development Command), and the U.S. Navy (Office of Naval Research).

Rate distortion theory (supposed to be difficult, covering problem)

Replace a source output X by another X' with average distortion $\leq D$: task is to minimize H(X')

example

since X is given, choose the quantizer (mapping of X to X')

Solution: the 16 Hamming codewords cover all sequences 128 of length 7 with a difference $\leq$ 1.

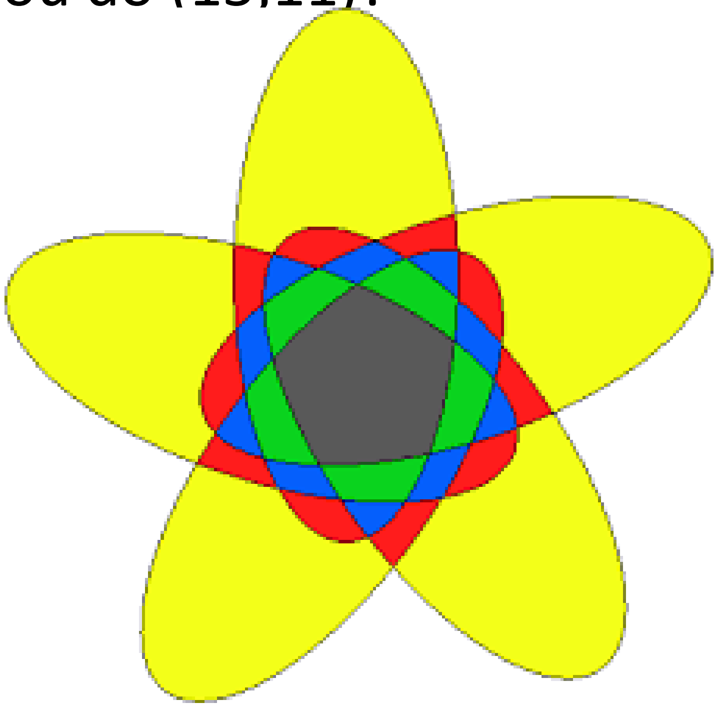
Hence, the efficiency is 4/7. How does it look like for general Hamming codes?

The Hamming codewords are linear combinations of the vectors: (1 0 0 0 1 1 1; 0 1 0 0 1 1 0; 0 0 1 0 1 0 1; 0 0 0 1 0 1 1)

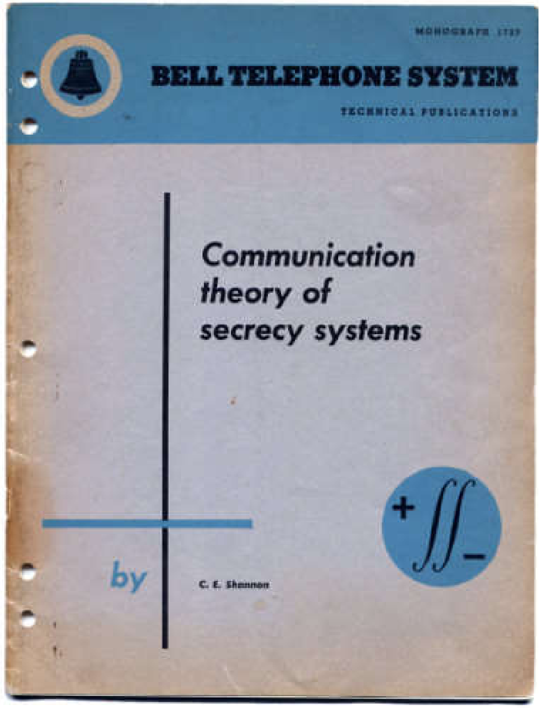
A Hamming code can be represented as Venn diagramm (why?)

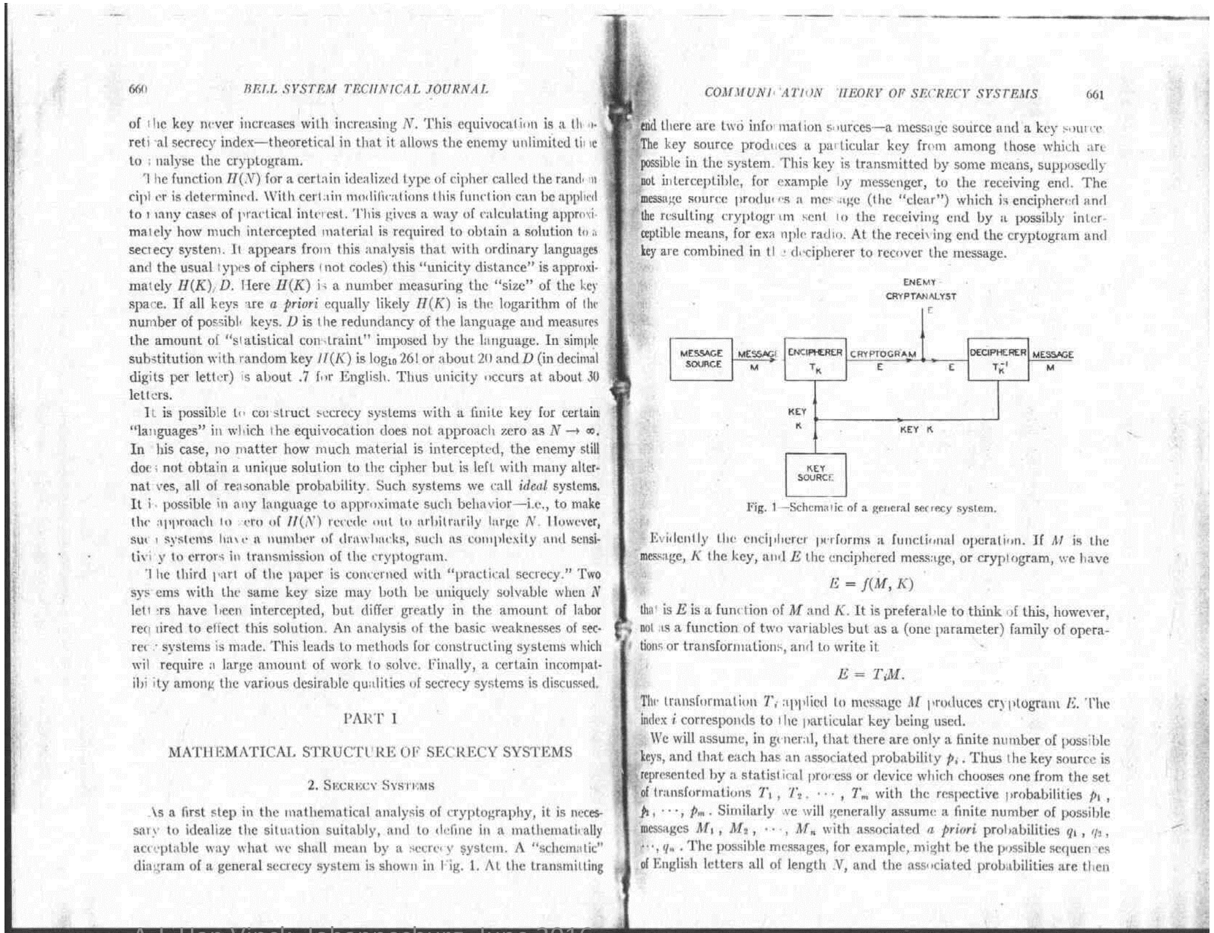
6th *Asia-Europe Workshop* on Information Theory, *Ishigaki* Island, Okinawa

- Example (31,26). Can you do (15,11)?



Shannon's original crypto paper, 1949





of the key never increases with increasing N. This equivocation is a theoretical secrecy index—theoretical in that it allows the enemy unlimited time to analyse the cryptogram.

The function $H(N)$ for a certain idealized type of cipher called the random cipher is determined. With certain modifications this function can be applied to many cases of practical interest. This gives a way of calculating approximately how much intercepted material is required to obtain a solution to a secrecy system. It appears from this analysis that with ordinary languages and the usual types of ciphers (not codes) this "unicity distance" is approximately $H(K)/D$. Here $H(K)$ is a number measuring the "size" of the key space. If all keys are *a priori* equally likely $H(K)$ is the logarithm of the number of possible keys. D is the redundancy of the language and measures the amount of "statistical constraint" imposed by the language. In simple substitution with random key $H(K)$ is $\log_{10} 26!$ or about 20 and D (in decimal digits per letter) is about .7 for English. Thus unicity occurs at about 30 letters.

It is possible to construct secrecy systems with a finite key for certain "languages" in which the equivocation does not approach zero as $N \to \infty$. In this case, no matter how much material is intercepted, the enemy still does not obtain a unique solution to the cipher but is left with many alternatives, all of reasonable probability. Such systems we call *ideal* systems. It is possible in any language to approximate such behavior—i.e., to make the approach to zero of $H(N)$ recede out to arbitrarily large N. However, such systems have a number of drawbacks, such as complexity and sensitivity to errors in transmission of the cryptogram.

The third part of the paper is concerned with "practical secrecy." Two systems with the same key size may both be uniquely solvable when N letters have been intercepted, but differ greatly in the amount of labor required to effect this solution. An analysis of the basic weaknesses of secrecy systems is made. This leads to methods for constructing systems which will require a large amount of work to solve. Finally, a certain incompatibility among the various desirable qualities of secrecy systems is discussed.

PART I

MATHEMATICAL STRUCTURE OF SECRECY SYSTEMS

2. Secrecy Systems

As a first step in the mathematical analysis of cryptography, it is necessary to idealize the situation suitably, and to define in a mathematically acceptable way what we shall mean by a secrecy system. A "schematic" diagram of a general secrecy system is shown in Fig. 1. At the transmitting

end there are two information sources—a message source and a key source. The key source produces a particular key from among those which are possible in the system. This key is transmitted by some means, supposedly not interceptible, for example by messenger, to the receiving end. The message source produces a message (the "clear") which is enciphered and the resulting cryptogram sent to the receiving end by a possibly interceptible means, for example radio. At the receiving end the cryptogram and key are combined in the decipherer to recover the message.

Fig. 1—Schematic of a general secrecy system.

Evidently the encipherer performs a functional operation. If M is the message, K the key, and E the enciphered message, or cryptogram, we have

$$E = f(M, K)$$

that is E is a function of M and K. It is preferable to think of this, however, not as a function of two variables but as a (one parameter) family of operations or transformations, and to write it

$$E = T_i M.$$

The transformation T_i applied to message M produces cryptogram E. The index i corresponds to the particular key being used.

We will assume, in general, that there are only a finite number of possible keys, and that each has an associated probability p_i. Thus the key source is represented by a statistical process or device which chooses one from the set of transformations $T_1, T_2, \cdots, T_m$ with the respective probabilities $p_1, p_2, \cdots, p_m$. Similarly we will generally assume a finite number of possible messages $M_1, M_2, \cdots, M_n$ with associated *a priori* probabilities $q_1, q_2, \cdots, q_n$. The possible messages, for example, might be the possible sequences of English letters all of length N, and the associated probabilities are then

Contribution still used in DES and AES

Confusion and Diffusion

Confusion

The relationship between the key and the ciphertext as complex and as involved as possible.

e.g. Enigma & complex substitution (S-boxes)

011011

S_5		Middle 4 bits of input															
		0000	0001	0010	0011	0100	0101	0110	0111	1000	1001	1010	1011	1100	1101	1110	1111
Outer bits	00	0010	1100	0100	0001	0111	1010	1011	0110	1000	0101	0011	1111	1101	0000	1110	1001
	01	1110	1011	0010	1100	0100	0111	1101	0001	0101	0000	1111	1010	0011	1001	1000	0110
	10	0100	0010	0001	1011	1010	1101	0111	1000	1111	1001	1100	0101	0110	0011	0000	1110
	11	1011	1000	1100	0111	0001	1110	0010	1101	0110	1111	0000	1001	1010	0100	0101	0011

Claude Shannon

Diffusion

Statistics of the plaintext is "dissipated" in the statistics of the ciphertext. If we change a character of the plaintext, then several characters of the ciphertext should change.

http://en.wikipedia.org/wiki/Permutation_box

14	17	11	24	1	5	3	28
15	6	21	10	23	19	12	4
26	8	16	7	27	20	13	2
41	52	31	37	47	55	30	40
51	45	33	47	44	49	39	56
34	53	46	42	50	36	29	32

P-Box

PERSPECTIVE OF SHANNON'S SECRECY SYSTEM

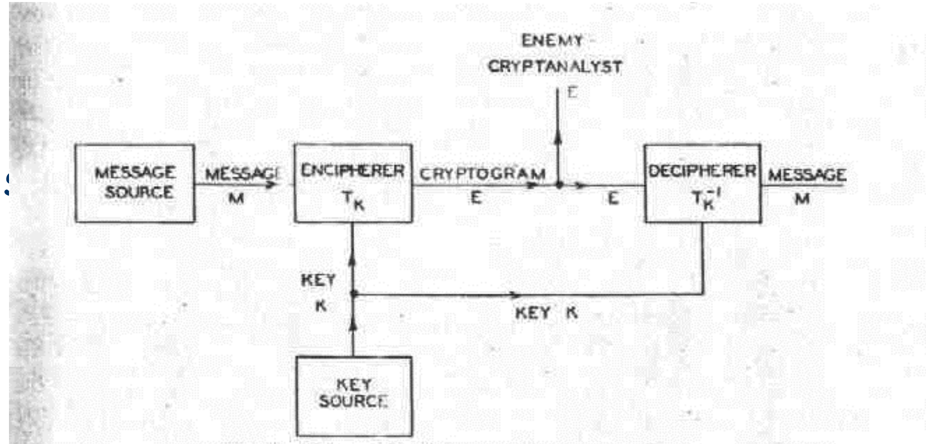


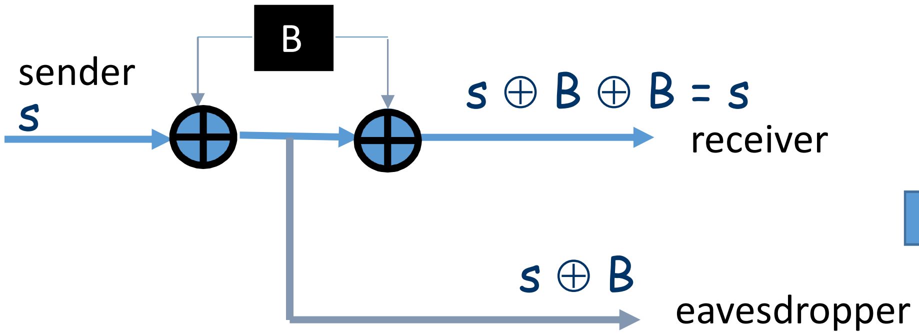
Fig. 1—Schematic of a general secrecy system.

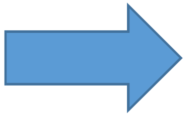
For Perfect secrecy we have a necessary condition:

$$H(S|X) = H(S)$$

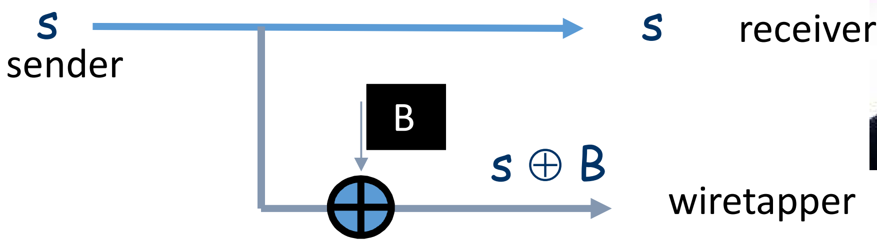
$$\Rightarrow H(S) \leq H(B)$$

i.e. # of messages ≤ # of keys





Wiretap channel model




Aaron Wyner

Secrecy rate: C_s = H(B) = amount of secret bits/tr

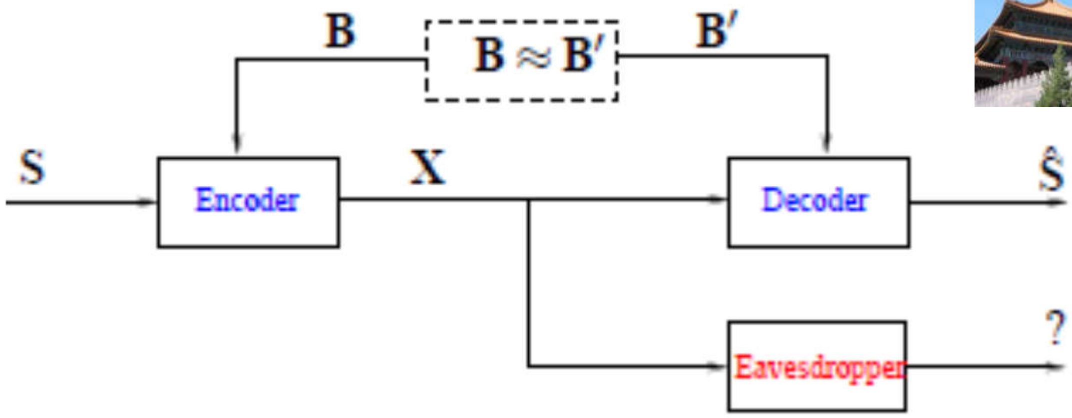
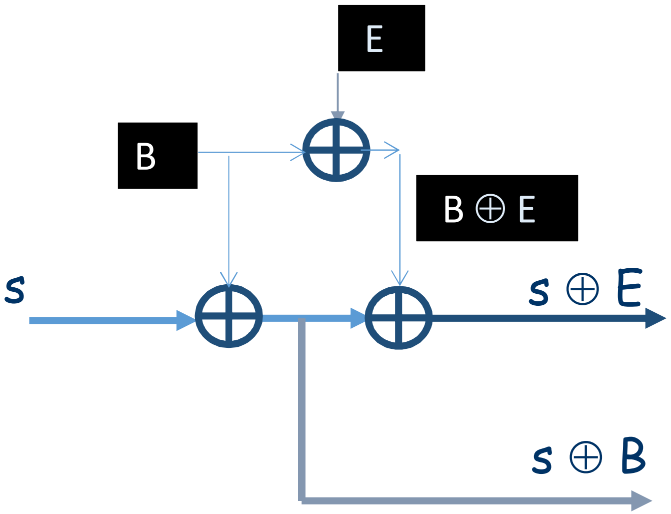


Figure 2: An extension of Shannon's secrecy system.

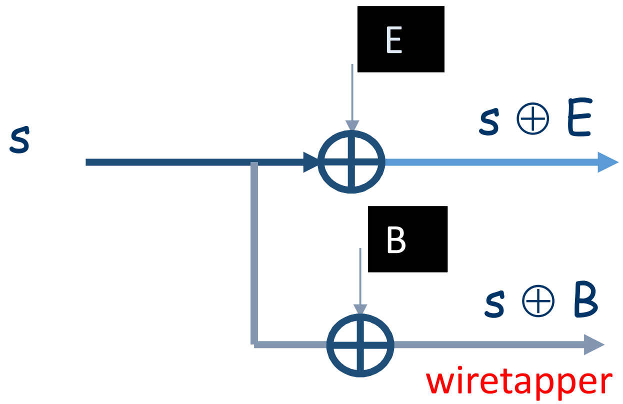
For Perfect secrecy H(S|X) = H(S)

H(S) ≤ H(B) – H(E)

i.e. we pay a price for the noise!



Wiretap channel model



Aaron Wyner

Secrecy rate C_s = H(B) - H(E) = # secret bits/transmission

Two-Way Channel, Shannon (1961)

TWO-WAY COMMUNICATION CHANNELS

CLAUDE E. SHANNON
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
CAMBRIDGE, MASSACHUSETTS

1. Introduction

A two-way communication channel is shown schematically in figure 1. Here x_1 is an input letter to the channel at terminal 1 and y_1 an output while x_2 is an

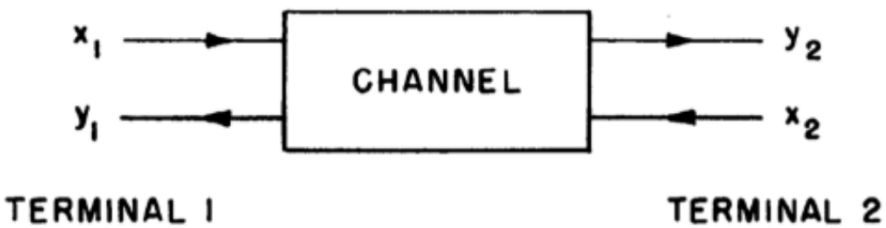
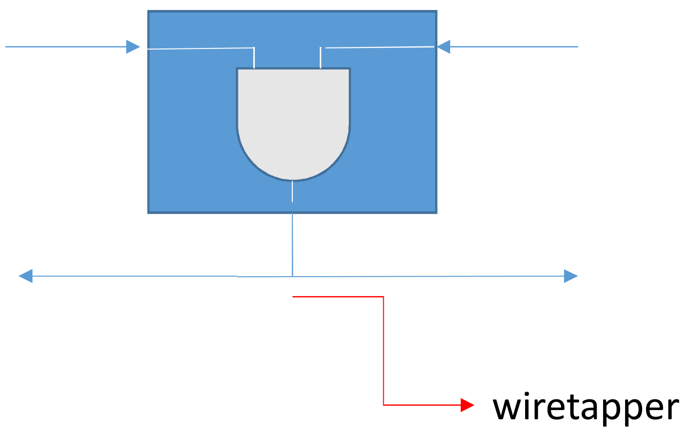


FIGURE 1

input at terminal 2 and y_2 the corresponding output. Once each second, say, new inputs x_1 and x_2 may be chosen from corresponding input alphabets and

Examples: binary input

	0	1
0	0	0
1	0	1



$x_2 \backslash x_1$	0	1
0	0	0
1	0	1

(a)

$x_2 \backslash x_1$	0	1
0	0	1
1	1	0

(b)

$x_2 \backslash x_1$	0	1
0	0	1
1	1	2

(c)

Only interesting cases!

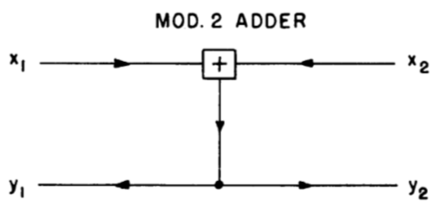
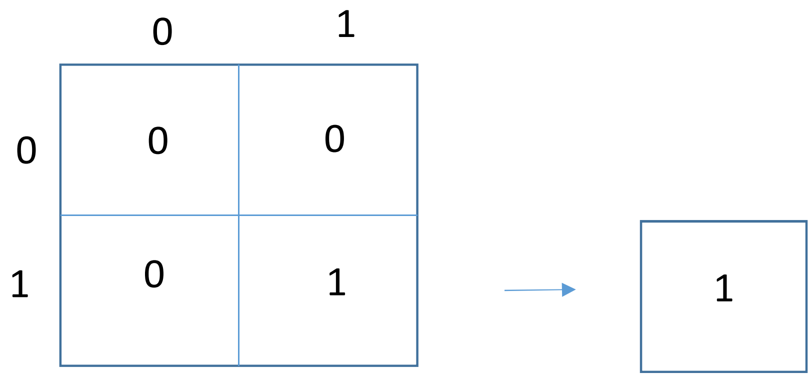
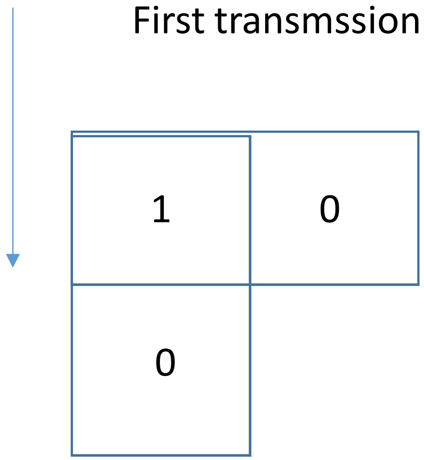


FIGURE 4

A simple code: Hagelbarger code for the and



First transmssion



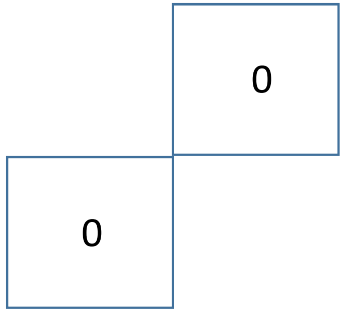
Second transmssion

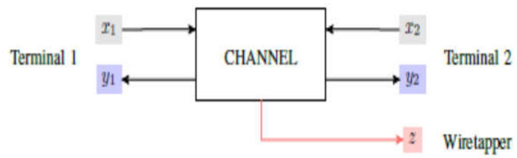
Code word length (av) : ¼ x 1 + ¾ x 2 = 7/4

Rate/user = 4/7 bit/tr

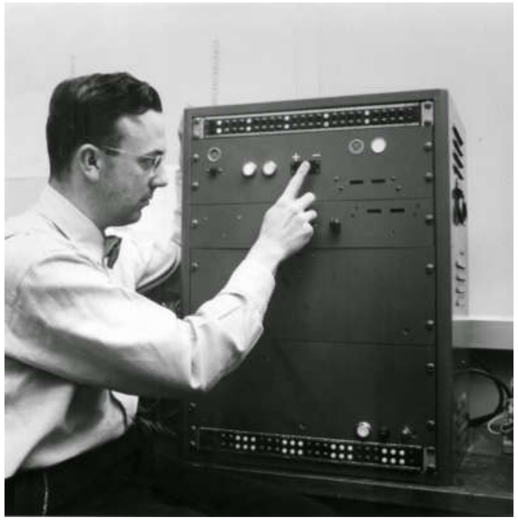
Wiretapper ambiguity: ½ x 1 bit/square

Hence: joint ambiguity = 2/7 bit/tr

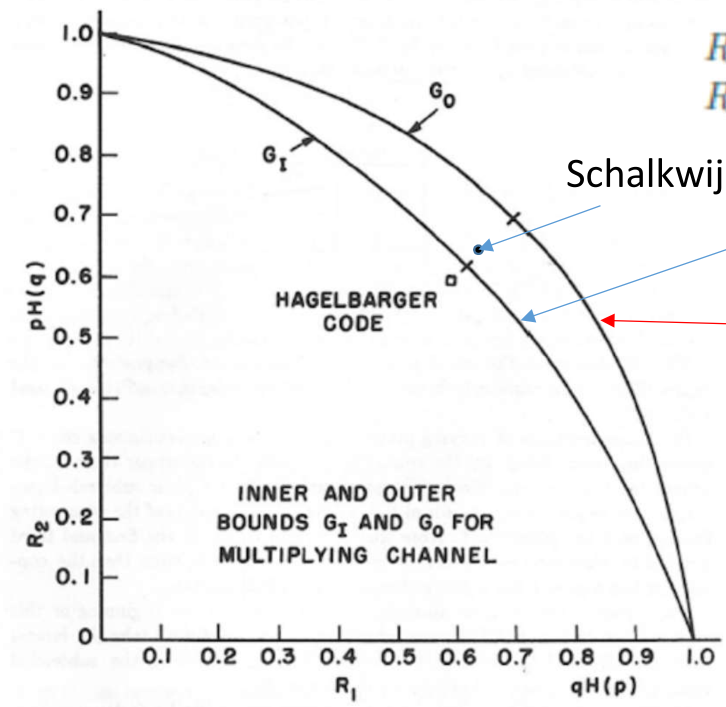




Only inner and outer bound are known (open)



David Hagelbarger at Bell labs



$$R_1 \le I(X_1;Y|X_2),$$
$$R_2 \le I(X_2;Y|X_1),$$

Schalkwijk

Inner bound: independent input

Outer bound: dependent inputs

Achievable security:

rate

$$R_1 \le I(X_1; Y|X_2),$$
$$R_2 \le I(X_2; Y|X_1),$$
$$d \le 1,$$
$$(R_1 + R_2)d \le I(X_1; Y|X_2) + I(X_2; Y|X_1) - I(X_1, X_2; Z),$$

Secrecy

Hagelbarger

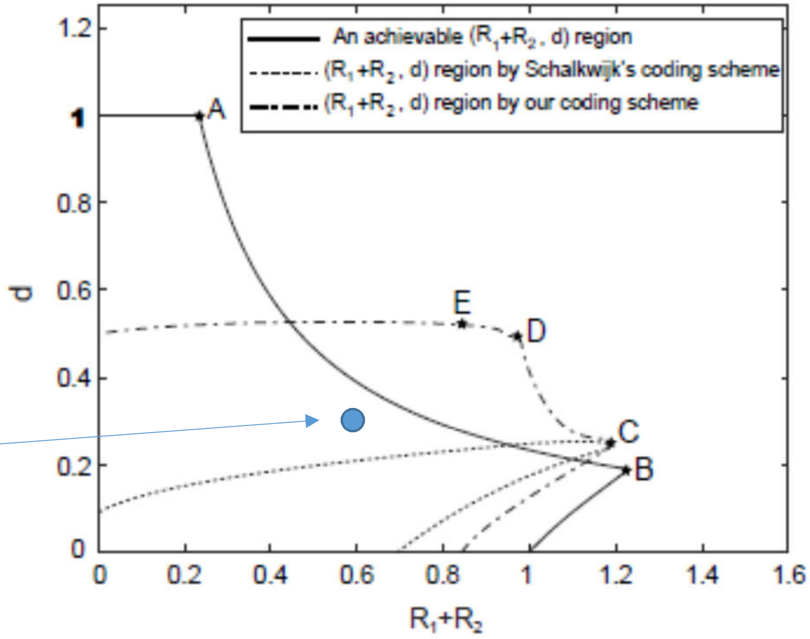


Figure 6: Comparison on the sum-rate equivocation $(R_1 + R_2, d)$ regions.

An interesting problem (smart grid)

- Suppose that - a question is public (give me your consumption,
  - but the answer is secret (I used xx KWh)

What does it mean for the security?

One-sided secrecy over the two-way wiretap channel

Chao Qi*, Yanling Chen†, A. J. Han.Vinck† and Xiaohu Tang*

One-sided secrecy over the two-way wiretap channel

Chao Qi*, Yanling Chen†, A. J. Han.Vinck† and Xiaohu Tang*

Corollary 1. *For the two-way channel with an external eavesdropper such that $Y_1 = Y_2 = Z$, an achievable one-sided secrecy rate region is given by the union of non-negative rate pairs (R_{1e}, R_2) satisfying*

$$R_2 \le I(X_2; Z|X_1),$$
$$R_{1e} \le R_2 - I(X_2; Z),$$

over all $p(x_1)p(x_2)$.

$$P_{e,i} \le \epsilon_n, \quad \textit{for} \quad i = 1, 2$$
$$\frac{1}{n} I(W_{1e}; Z^n) \le \tau_n,$$
$$\lim_{n\to\infty} \epsilon_n = 0 \quad \textit{and} \quad \lim_{n\to\infty} \tau_n = 0.$$

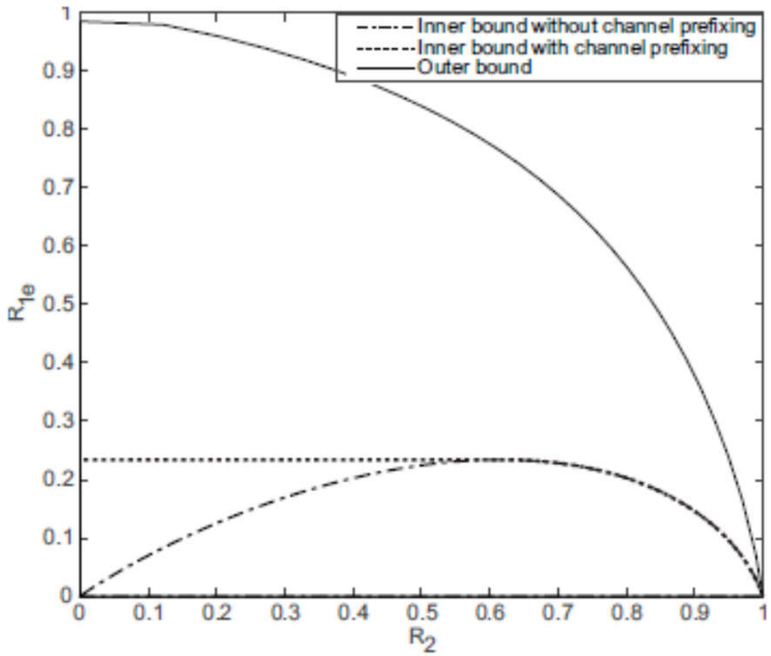
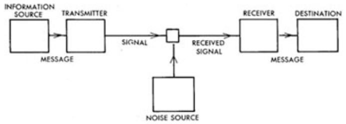
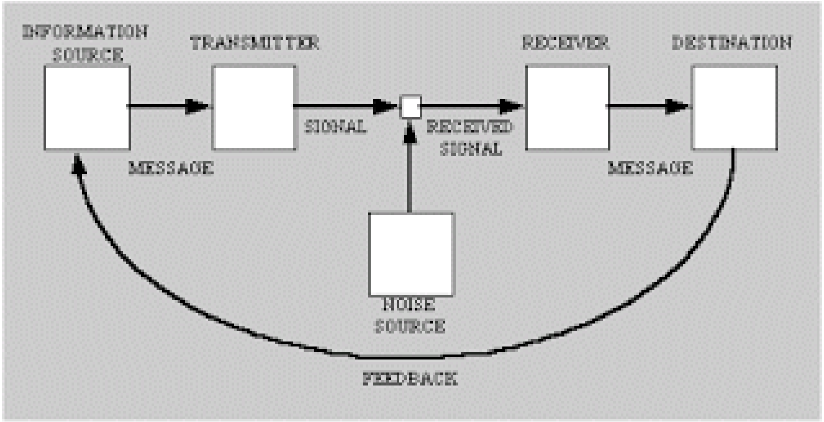


Fig. 3. The secrecy rate region for the BMC.



Shannon and feedback



One of the most surprising results in information theory was proven by Claude Shannon in 1956 [1]: instantaneous and noiseless feedback of the output of a discrete memoryless channel does not increase capacity. This

[1] C.E. Shannon, "The zero error capacity of a noisy channel," *IRE Trans. Inform. Theory*, vol. IT-2, pp. 8–19, September 1956.

But what about:

- Channels with memory
- Multi user channels like MAC?

Interesting observation from the two terminal paper from Shannon (1961)

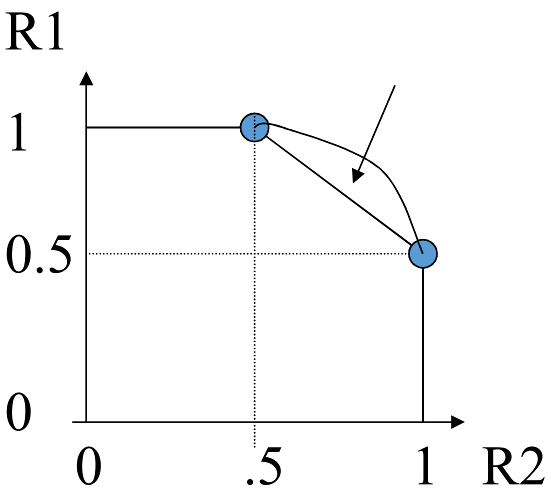
17. Generalization to T-terminal channels

Many of the tricks and techniques used above may be generalized to channels with three or more terminals. However, some definitely new phenomena appear in these more complex cases. In another paper we will discuss the case of a channel with two or more terminals having inputs only and one terminal with an output only, a case for which a complete and simple solution of the capacity region has been found.

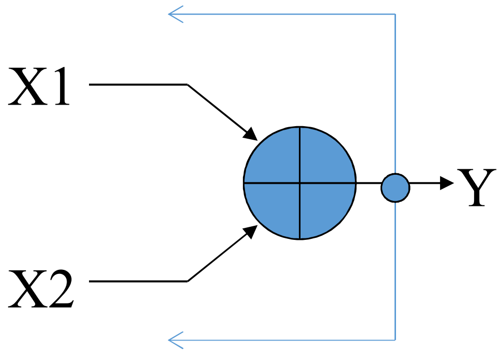
MAC

First results appear in: R. Ahlswede, “Multi-way communication channels,” in Proceedings of 2nd International Symposium on Information Theory (Thakadsor, Armenian SSR, Sept. 1971), Publishing House of the Hungarian Academy of Science, Budapest, 1973, pp. 23–52

Two-adder with feedback improves over non-feedback!



Capacity region



model

Improvements?

X1

X2 \ X1	1	0
0	1	0
1	2	1

- Solution: users know each others input due to the feedback
- They solve the problem for the receiver in total cooperation (log 3 bits/transmission)

- Coding Techniques and the Two-Access Channel,, ***In Multiple Access Channels: Theory and Practice***
  Eds. E. Biglieri, L. Györfi, *pp. 273-286, IOS Press, ISBN, 978-1-58603-728-4, 2007*

C. E. Shannon

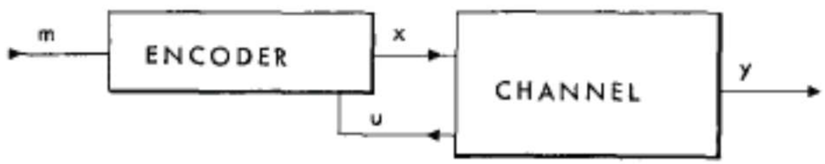
Channels with Side Information at the Transmitter

Abstract: In certain communication systems where information is to be transmitted from one point to another, additional side information is available at the transmitting point. This side information relates to the state of the transmission channel and can be used to aid in the coding and transmission of information. In this paper a type of channel with side information is studied and its capacity determined.

Introduction

Channels with feedback[1] from the receiving to the transmitting point are a special case of a situation in which there is additional information available at the transmitter which may be used as an aid in the forward transmission system. In Fig. 1 the channel has an input x and an output y.

Figure 1



Memory systems: defects known to writer, not to the reader

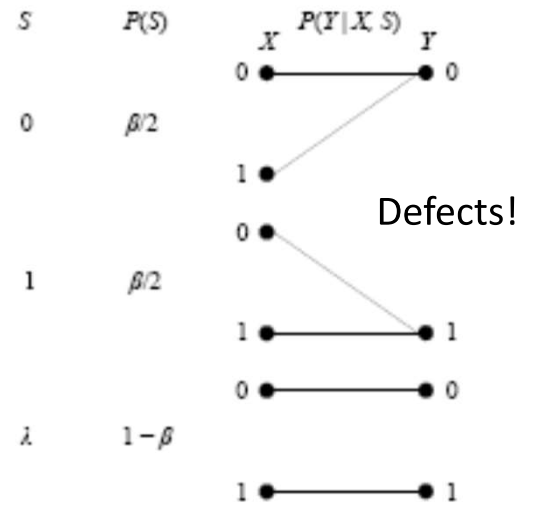
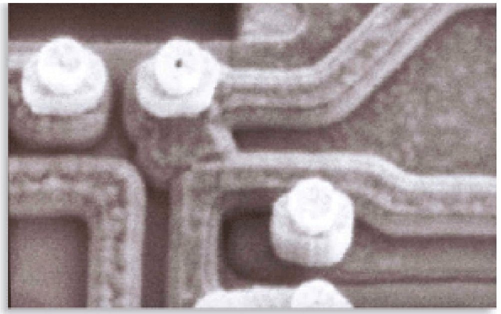


Fig. 1. Binary defect channel (BDC).

Capacity is 1-p bits/cell



RELIABLE CIRCUITS USING LESS RELIABLE RELAYS

BY

E. F. MOORE [1] AND C. E. SHANNON [1]

The first kind of failure allowed is the failure of a relay contact to close, which in actual relays is often due to a particle of dust preventing electrical closure.

The second type of failure is the failure of a contact to open, which in actual relays is usually due to the welding action of the current passing through the contacts. We shall consider relay circuits in which

On the Influence of Coding on the Mean Time to Failure for Degrading Memories with Defects

HAN VINCK AND KAREL POST, MEMBER, IEEE

For the simplest, and perhaps most practical, situation where $d_{\min} = 3$, we get

$$\eta \cong \frac{k}{n}\sqrt{N}$$

Q: how does coding influence the MTTF?

N = number of words

IEEE TRANSACTIONS ON INFORMATION THEORY, VOL. 40, NO. 6, NOVEMBER 1994

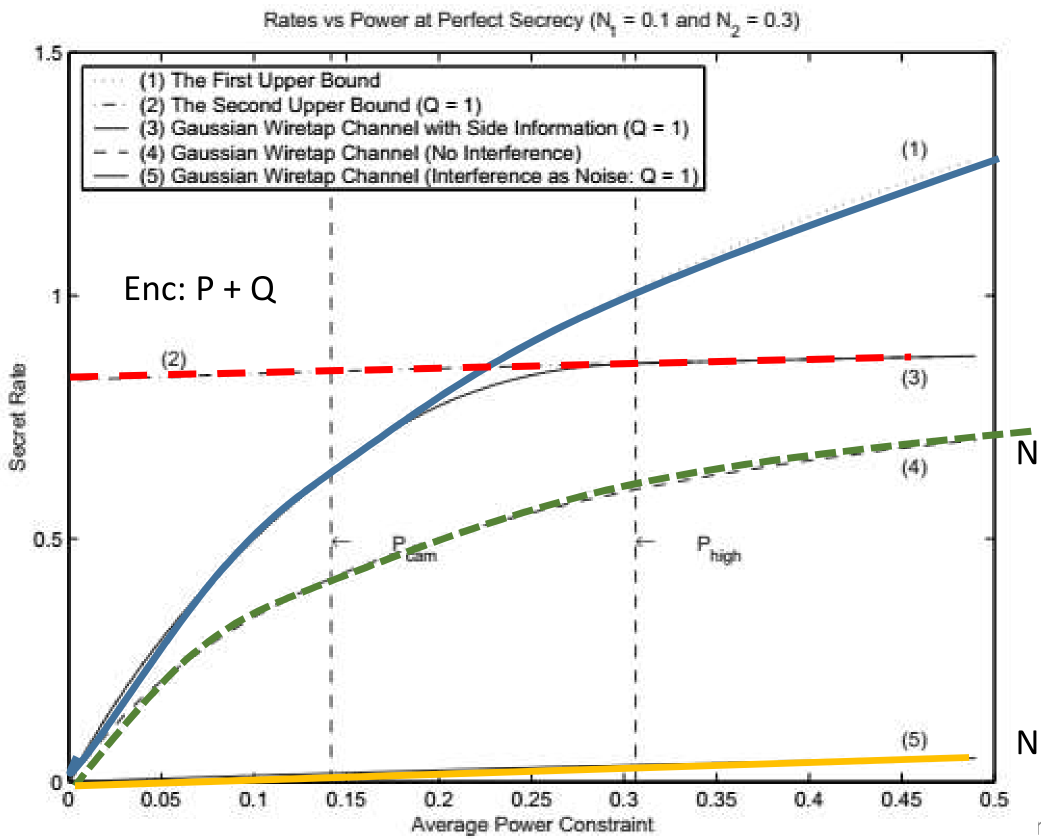
On the General Defective Channel with Informed Encoder and Capacities of Some Constrained Memories

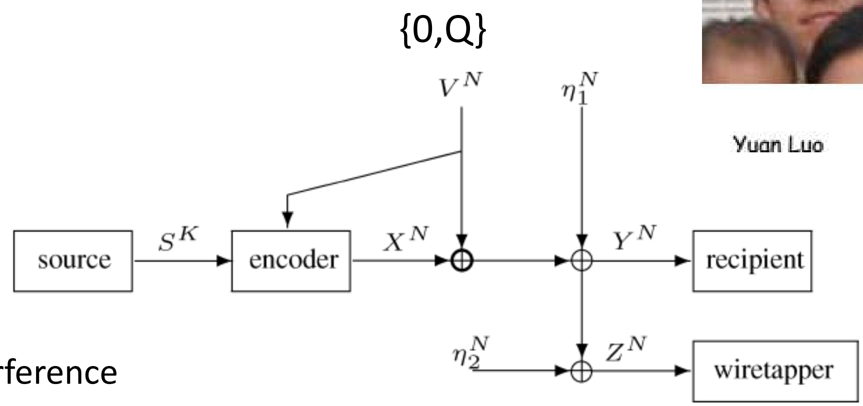
Alexander V. Kuznetsov and A. J. Han Vinck

Sacha Kuznetsov

[An Achievable Region for the Gaussian Wiretap Channel with Side Information,](#)

IEEE Transactions on Information Theory, May 2006, C. Mitrpant, A.J. Han Vinck and Yuan Luo, iSSN 0018-9448

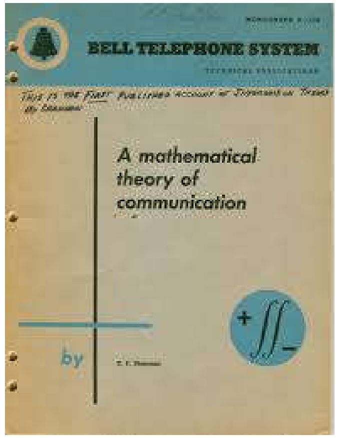




Yuan Luo

***Index Terms*—Dirty-paper channel, Gaussian wiretap channel, Gaussian wiretap channel with side information, perfect secrecy.**

CONSTRAINED SEQUENCES from the 1948 paper



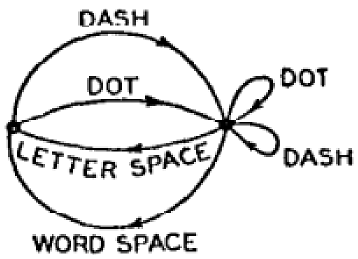


Fig. 2—Graphical representation of the constraints on telegraph symbols.

FIGURE 2!

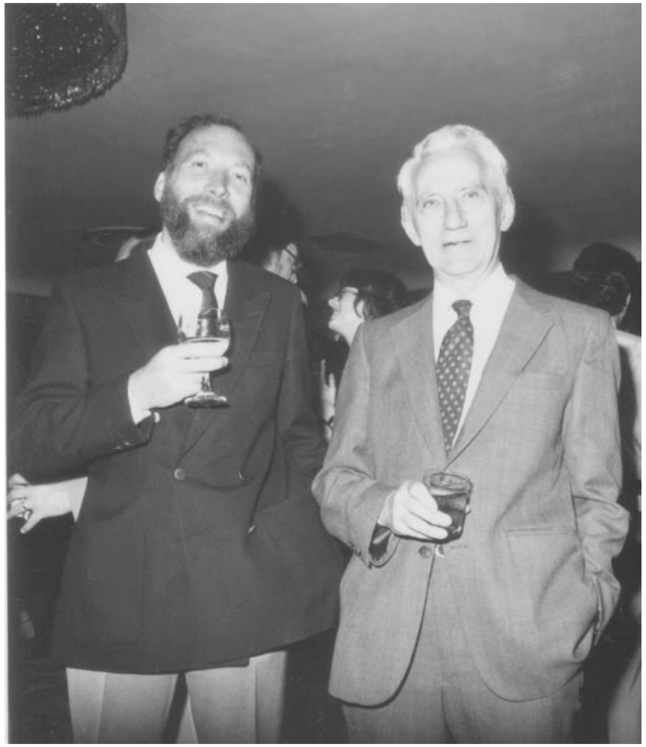
In the more general case with different lengths of symbols and constraints on the allowed sequences, we make the following definition:
Definition: The capacity C of a discrete channel is given by

$$C = \lim_{T \to \infty} \frac{\log N(T)}{T}$$

where $N(T)$ is the number of allowed signals of duration T.

„Our“ Kees Immink got famous for using constrained sequences for CD!

Johannesburg, 2014



- **AES Convention, New York, 1985**
- Claude Shannon, and Kees Immink

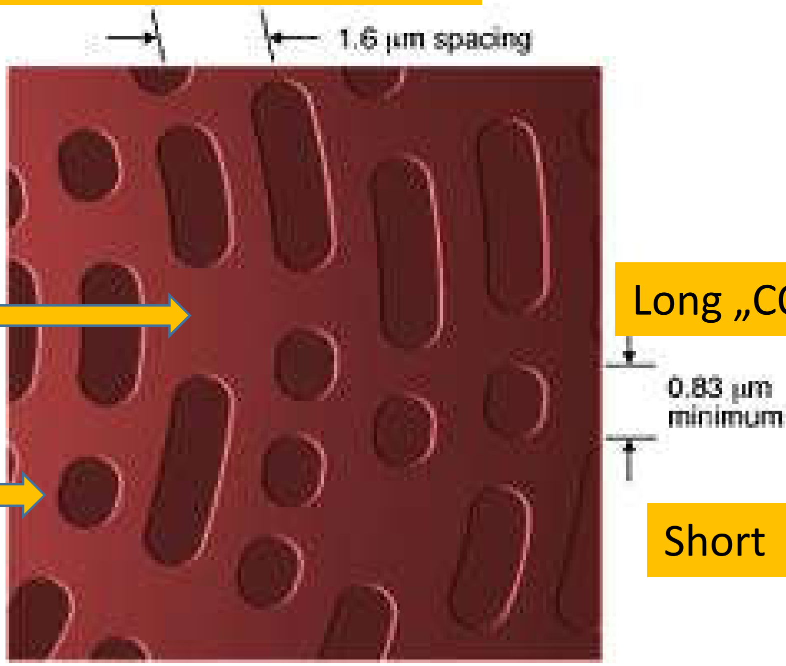
Johannesburg, 1994

What are the symbol constraints for writing on a CD ?

Symbol length has discrete values!

Not too long

Not too short



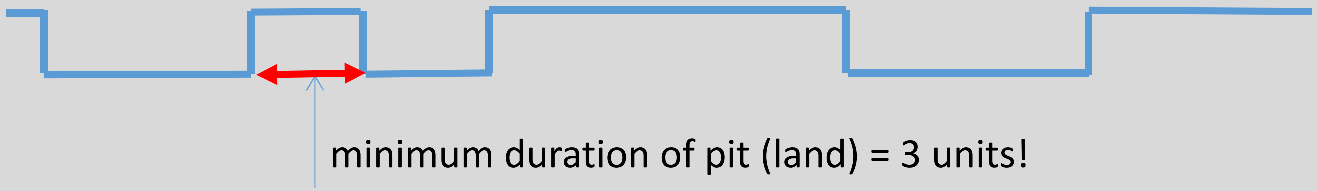
Long „CONSTANT“ sequences give synchronization problems

Short symbol duration gives detection problems

a remarkable observation can be made

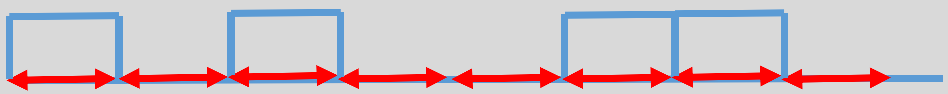
Constrained code: 0 1 0 0 0 0 1 0 0 1 0 0 1 0 0 0 0 0 0 1 0 0 0 0 0 1 0 0 0 0

8 information bits
in 17(16) positions



Traditional coding: the length of pits and lands is a multple of the minimum duration

8 information bits
in 24 positions



DENSITY GAIN
$\approx$ 40%

Coded modulation with a constraint on the minimum channel symbol duration, Mengi, A.J. Han Vinck, **Conference Proceeding:** 08/2009; DOI: 10.1109/ISIT.2009.5205832In proceeding of: IEEE International, Symposium on Information Theory, 2009. ISIT 2009.

Optical rewritable disk(Sony): writing only in 1 direction

Example: 6 messages, word length n = 5 => R = 0.51 > 0.5 !

6 messages

code book 0 $\rightarrow$ 1	
0 0 0 0 0	
1 0 0 0 0	0 1 0 0 1
0 1 0 0 0	1 0 1 0 0
0 0 1 0 0	0 1 0 1 0
0 0 0 1 0	0 0 1 0 1
0 0 0 0 1	1 0 0 1 0

code book 1 $\rightarrow$ 0		
1 1 1 1 1		0
0 1 1 1 1	1 0 1 1 0	1
1 0 1 1 1	0 1 0 1 1	2
1 1 0 1 1	1 0 1 0 1	3
1 1 1 0 1	1 1 0 1 0	4
1 1 1 1 0	0 1 1 0 1	5

Property: from any code word in code book 0 $\rightarrow$ 1 to any word in code book 1 $\rightarrow$ 0 and back

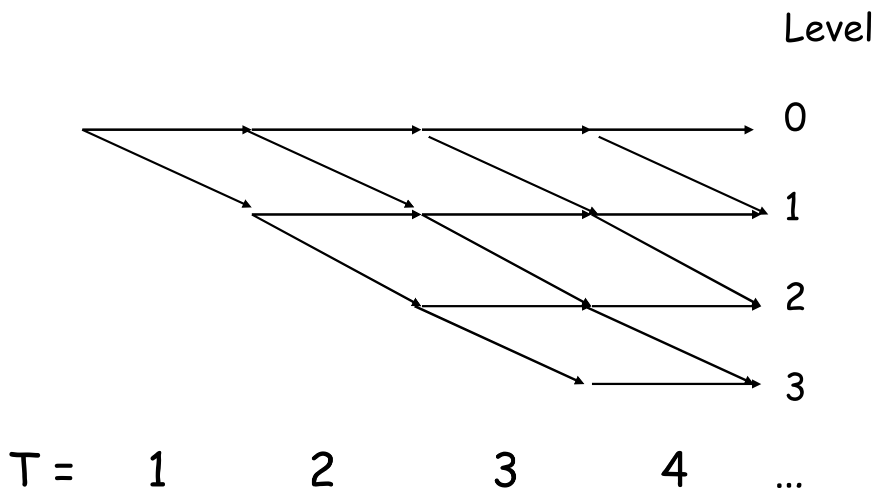
Example: 0 0 0 1 0 $\rightarrow$ 0 1 0 1 1 $\rightarrow$ 0 0 0 1 0 $\rightarrow$ 1 0 1 1 0 ...
0 $\rightarrow$ 1 1 $\rightarrow$ 0 0 $\rightarrow$ 1

Still work to do!

F. M. J. Willems and A. J. H. Vinck, "Repeated recording for an optical disk", *Proc. 7th Symp. Information Theory in the Benelux*, pp.49 -53 1986

Generalized WOM (flash), model

Example: $q = 4$
1 step increment

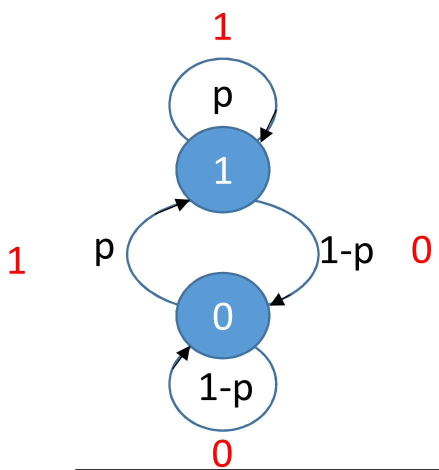


Log(# sequences) ≈ ?

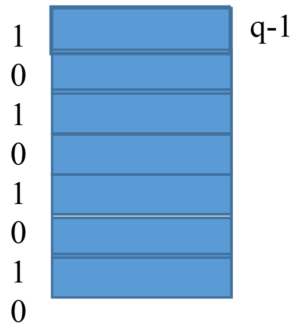
Log(# sequences) ≈ $(q-1) \log_2 (T+1)$ a factor of $(q-1)$!! more than the binary WOM

On the Capacity of Generalized Write-Once Memory with State Transitions Described by an Arbitrary Directed Acyclic Graph, Fang-Wei Fu and A. J. Han Vinck,
IEEE TRANSACTIONS ON INFORMATION THEORY, VOL. 45, NO. 1, JANUARY 1999

Performance 1 step-up for q-ary flash memory where P(1) = p



average increase
$2p(1-p)$ per writing



level flash

average number of writes
to reach top level (erase)

w = (q-1)/2p(1-p)

Average amount of information stored w x h(p)

$\approx$ ½ (q-1) $\log_2$(T+1) for p = 1/(T+1)

<u>Conclusion</u>: storage capacity improved
average time before erasure w $\approx$ Tq/2

Communication in the Presence of Noise

CLAUDE E. SHANNON, MEMBER, IRE

This paper is reprinted from the PROCEEDINGS OF THE IRE, vol. 37, no. 1, pp. 10–21, Jan. 1949.

X. THE CHANNEL CAPACITY WITH AN ARBITRARY TYPE OF NOISE

Of course, there are many kinds of noise which are not Gaussian; for example, impulse noise, or white noise that has passed through a nonlinear device. If the signal is perturbed by one of these types of noise, there will still be a definite channel capacity C, the maximum rate of transmission of binary digits. We will merely outline the general theory here.[10]

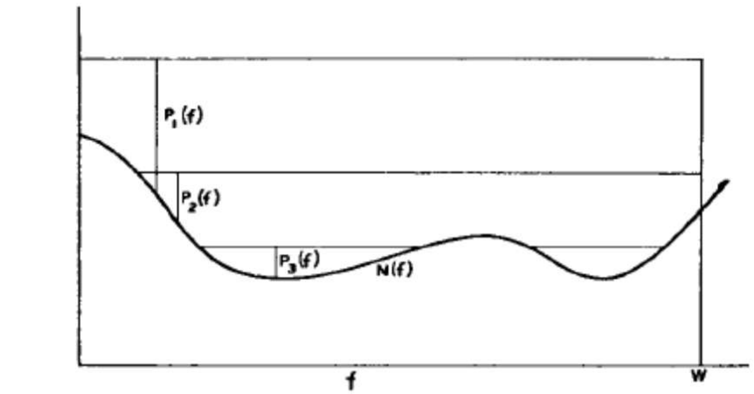
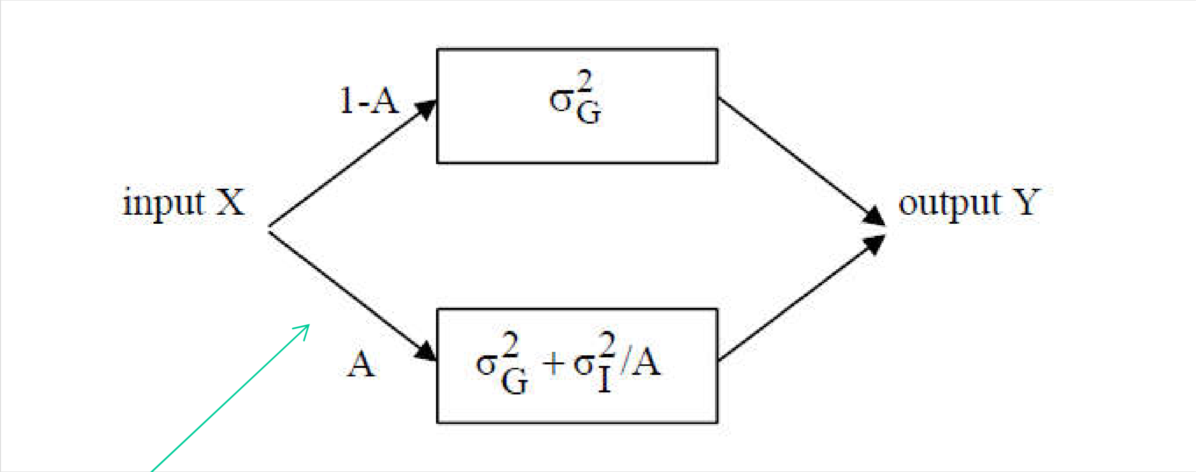


Fig. 8. Best distribution of transmitter power.

The waterfilling argument

A simple two states impulse model



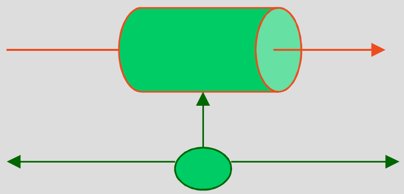
A influences the frequency of occurence of impulse noise

for $\sigma_I^2 = \sigma_G^2/T$

Example: average frequency of impulse $A = 0.1; T = 0.01$

$$\overline{\sigma^2} = \sigma_G^2 + \sigma_I^2 = 101\,\sigma_G^2; \qquad \sigma_G^2 + \sigma_I^2 \frac{1}{A} = 1001\,\sigma_G^2$$

Middleton class-A noise model: what is the channel capacity?

Input X  Output Y

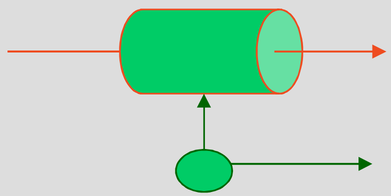
(un)known at transmitter channel state (un)known at receiver

$\sigma_G^2 + \sigma_I^2 \frac{1}{A}$ σ_G^2

Q1: channel capacity ?

It is not realistic to assume that the transmitter knows the state

Thus, Q2:what happens if only the receiver knows the state?



What can we gain by using the channel state? (memory of the noise)

- Using waterfilling argument (high P)

$$\text{(lowpower)capacity}(++) = (1-A)B\log(1+\frac{P/2B(1-A)}{\sigma_G^2})$$

$$\text{capacity}(+\ +) = (1-A)B\log_2(1+\frac{\sigma_I^2 + P/2B}{\sigma_G^2}) + AB\log_2(\frac{\sigma_G^2+\sigma_I^2+P/2B}{\sigma_G^2+\sigma_I^2/A})$$

- Using Gaussian input with average power $\leq$ P

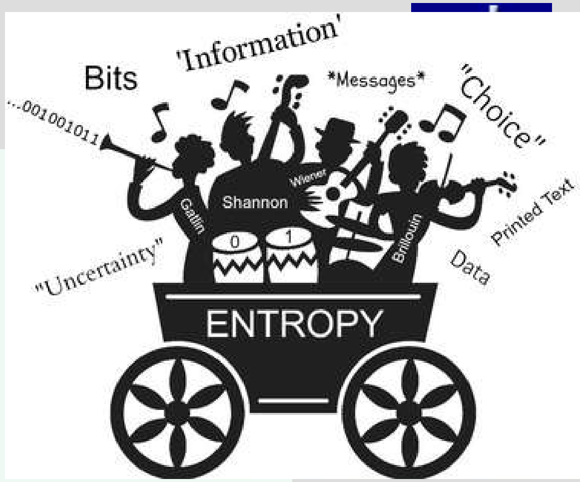
$$\text{capacity}(-\ +) = (1-A)B\log_2(1+\frac{P/2B}{\sigma_G^2}) + AB\log_2(\frac{\sigma_G^2+\sigma_I^2/A+P/2B}{\sigma_G^2+\sigma_I^2/A})$$

- The randomized (Gaussian) channel

$$\text{capacity}(-,-) = B\log_2(1+\frac{P/2B}{\sigma_G^2+\sigma_I^2})$$

$$\text{gain} \approx 10\log_{10}(1+\frac{\sigma_I^2}{\sigma_G^2})\ \text{dB}$$

1956, Shannon and the „BANDWAGON“

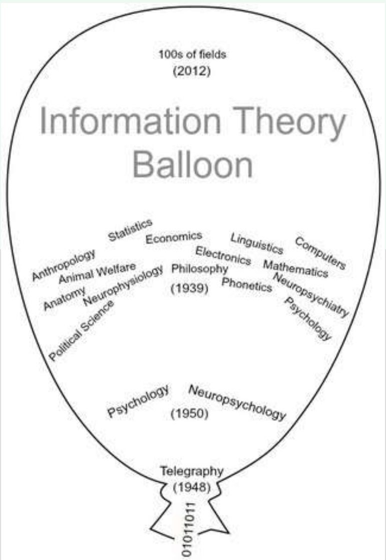


- Shannon was critical about „his information theory“

tions. I personally believe that many of the concepts of information theory will prove useful in these other fields—and, indeed, some results are already quite promising—but the estab
is not a trivial matter of
domain, but rather the
hypothesis and experim

Secondly, we must keep our own house in first class order. The subject of information theory has certainly been sold, if not oversold. We should now turn our attention to the business of research and development at the highest sc
tain. Research rather tha

tain. Research rather than exposition is the keynote, and our critical thresholds should be raised. Authors should submit only their best efforts, and these only after careful criticism by themselves and their colleagues. A few first rate research papers are preferable to a large number that are poorly conceived or half-finished. The latter are no credit to their writers and a waste of time to their readers. Only by maintaining



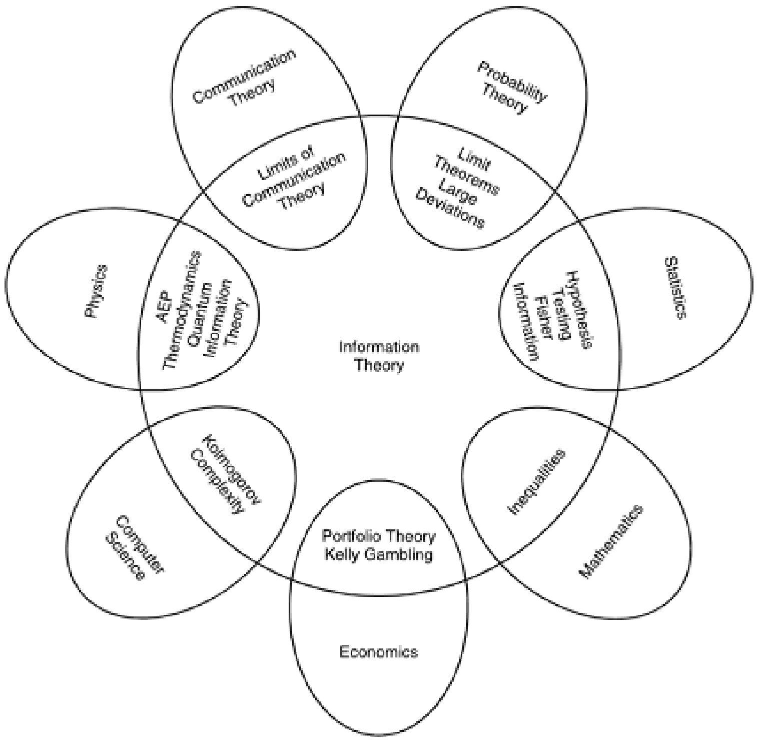
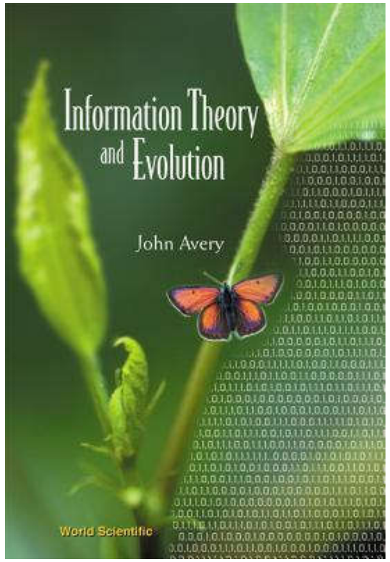
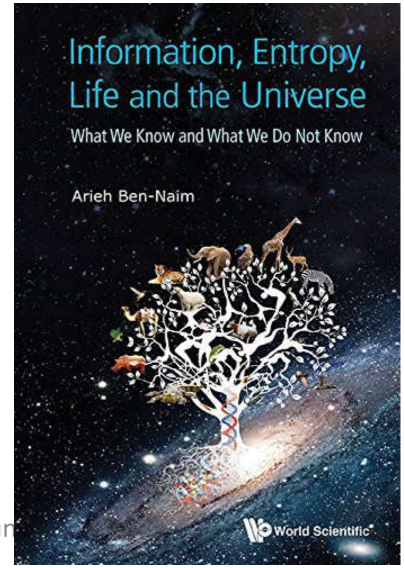
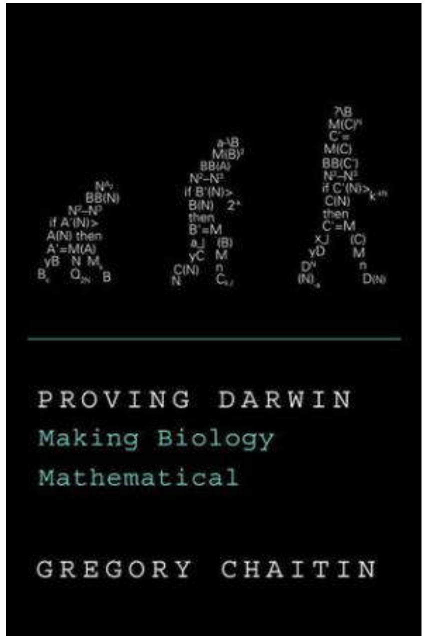


FIGURE 1.1. Relationship of information theory to other fields.

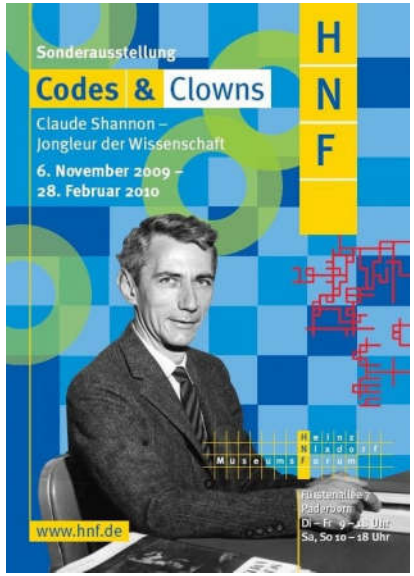


There are also other books than we are used to!



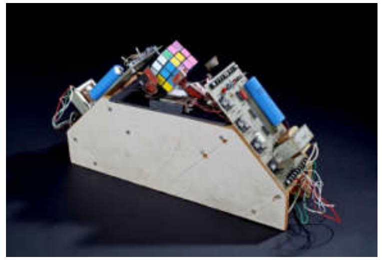


PLAY is the only way



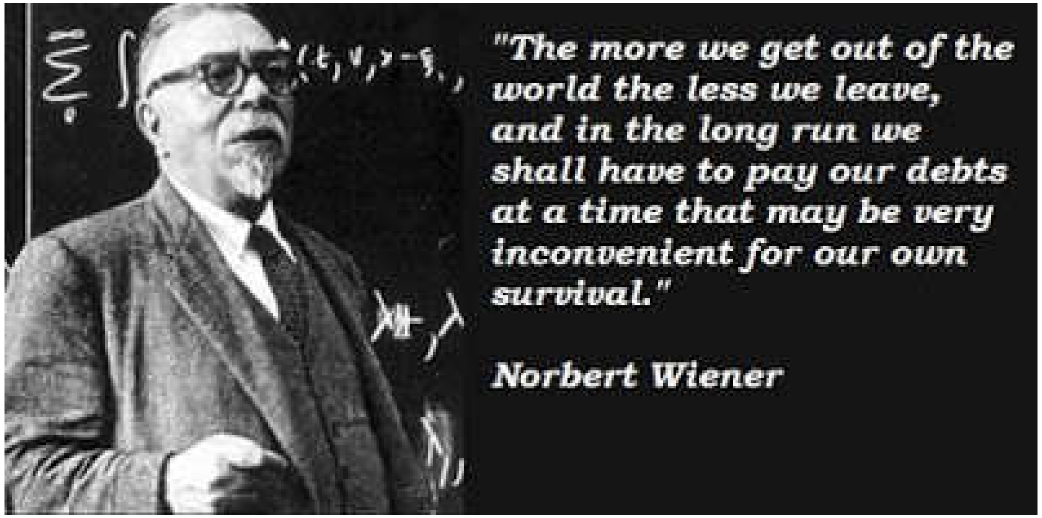
the highest intelligence of humankind

can unfold



J.C. Pearce.

Wiener influenced Shannon at MIT

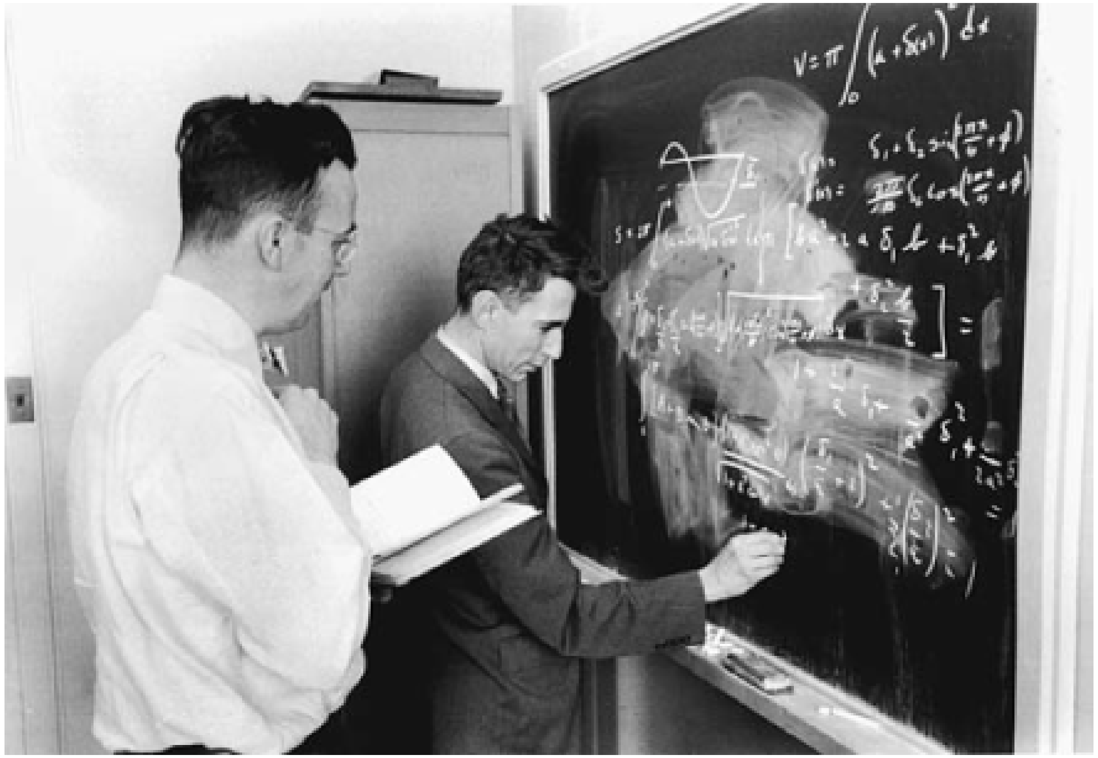


Norbert Wiener defined cybernetics in 1948 as

"the scientific study of control and communication in the animal and the machine."[2]

The word *cybernetics* comes from Greek κυβερνητική (*kybernetike*),

Dave Hagelbarger and Claude Shannon

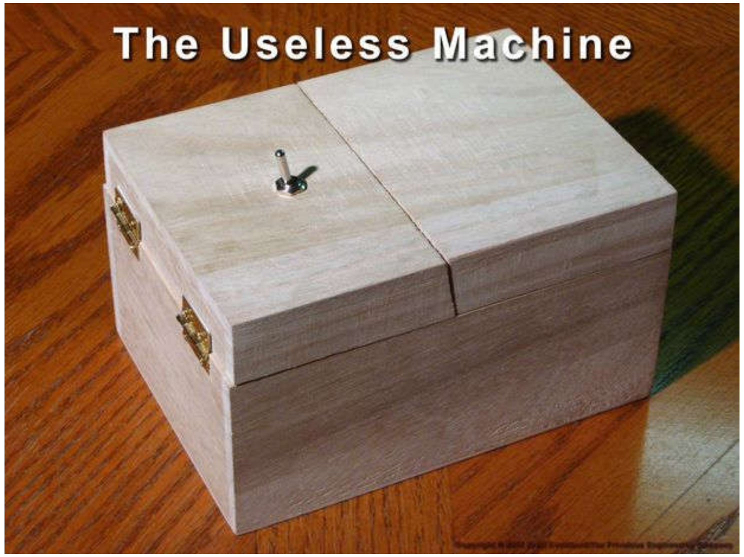


Claude Shannon's 1953
Outguessing Machine, at the MIT Museum.

(both Hagelbarger and Shannon produced a guessing machine)

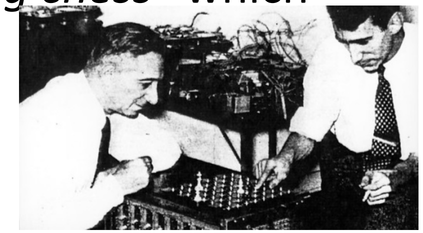
Shannon and the useless (ultimate) machine

- Many intelligent machines were produced (see Wikipedia), but also …
- https://youtu.be/urgL4Br2rqI



Summary of some other contributions of Shannon

- artificial intelligence, or AI.
- In 1950 he wrote a paper called "*Programming a digital computer for playing chess*" which essentially invented the whole subject of computer game playing.

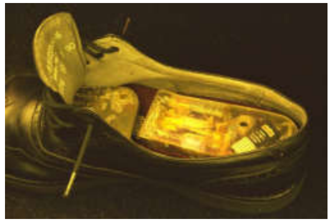


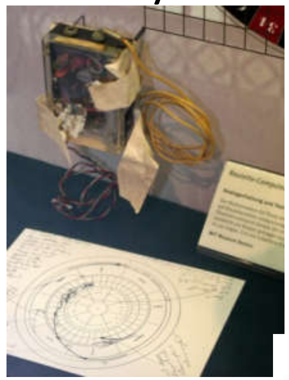
- JUGGLING (THEOREM)

Scientific American

[14] C.E. Shannon. Scientific aspects of juggling. In N. Sloane and A. Wyner, editors, *Claude Elwood Shannon – Collected Papers*, pages 850–864. IEEE Press, 1993.

- apply mathematics to beat the game of roulette. Thorp and Shannon build what is widely regarded to be the first wearable computer.



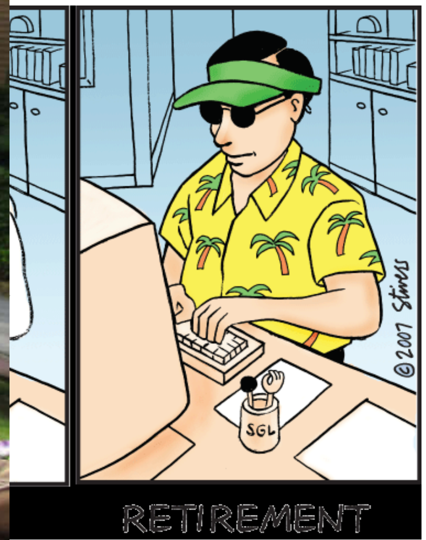


- Stock market/gambling

retirement is a transition from whatever you were doing to whatever you want to do, at whatever rate you want to make the transition."

• W

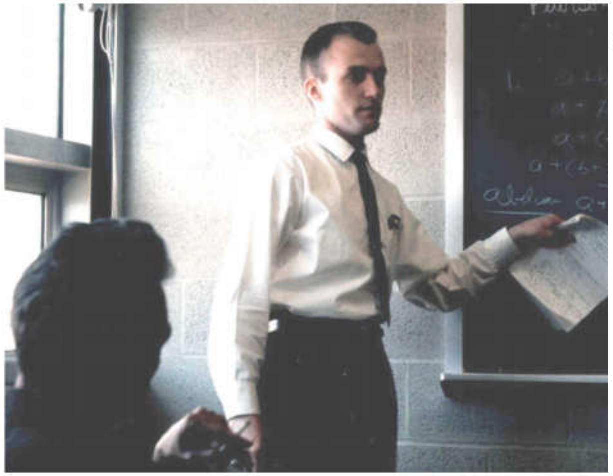
"Look here Wat



UNIVERSITÄT DUISBURG ESSEN Campus Essen

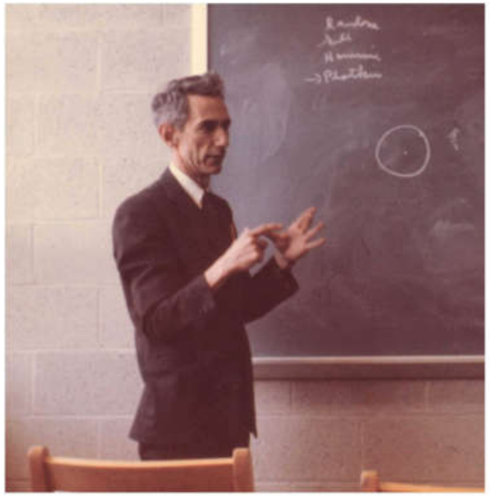
In Germany
Very famous
Gasthaus Petersberg,
Bonn

1961 **Claude Elwood Shannon**



Claude Elwood Shannon

17 April 1961



Left photo: Shannon, 1939, in a Piper Cub,