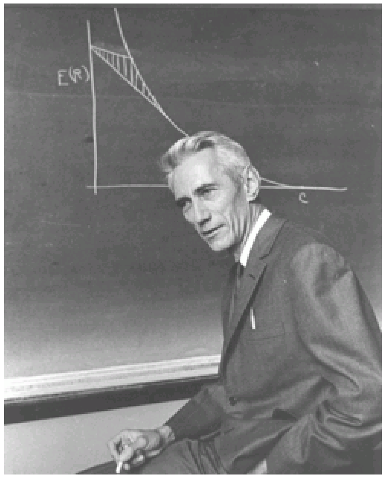
TC10 / **0. Introduction**

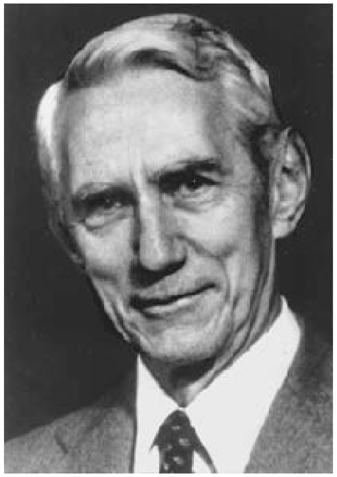
8·2·10 SX

The digital era

A Mathematical Theory of Communication (1948)

- Mathematical foundations of communication systems
- Definition of information, of channel capacity and of error-correcting codes.
- Source and channel coding theorems

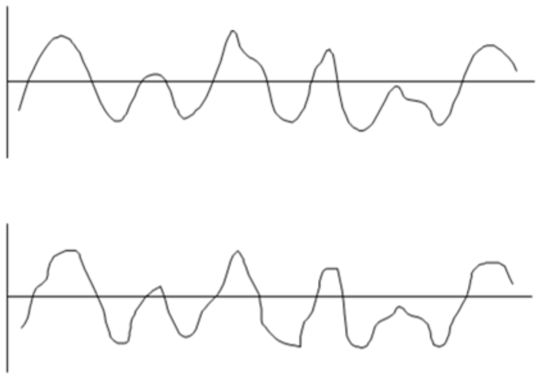




Claude Shannon (1916-2001)

- Master’s thesis (1937): logic circuits
- II WWII: Cryptography

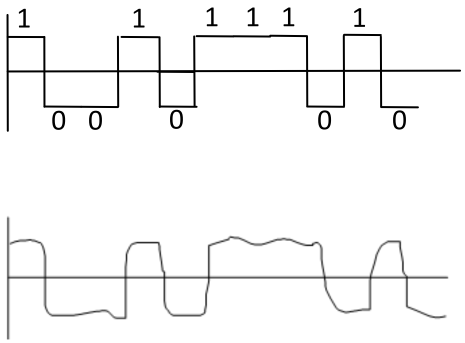
Analogical signal



Binary alphabet

$B = \{0,1\}$

Digital signal



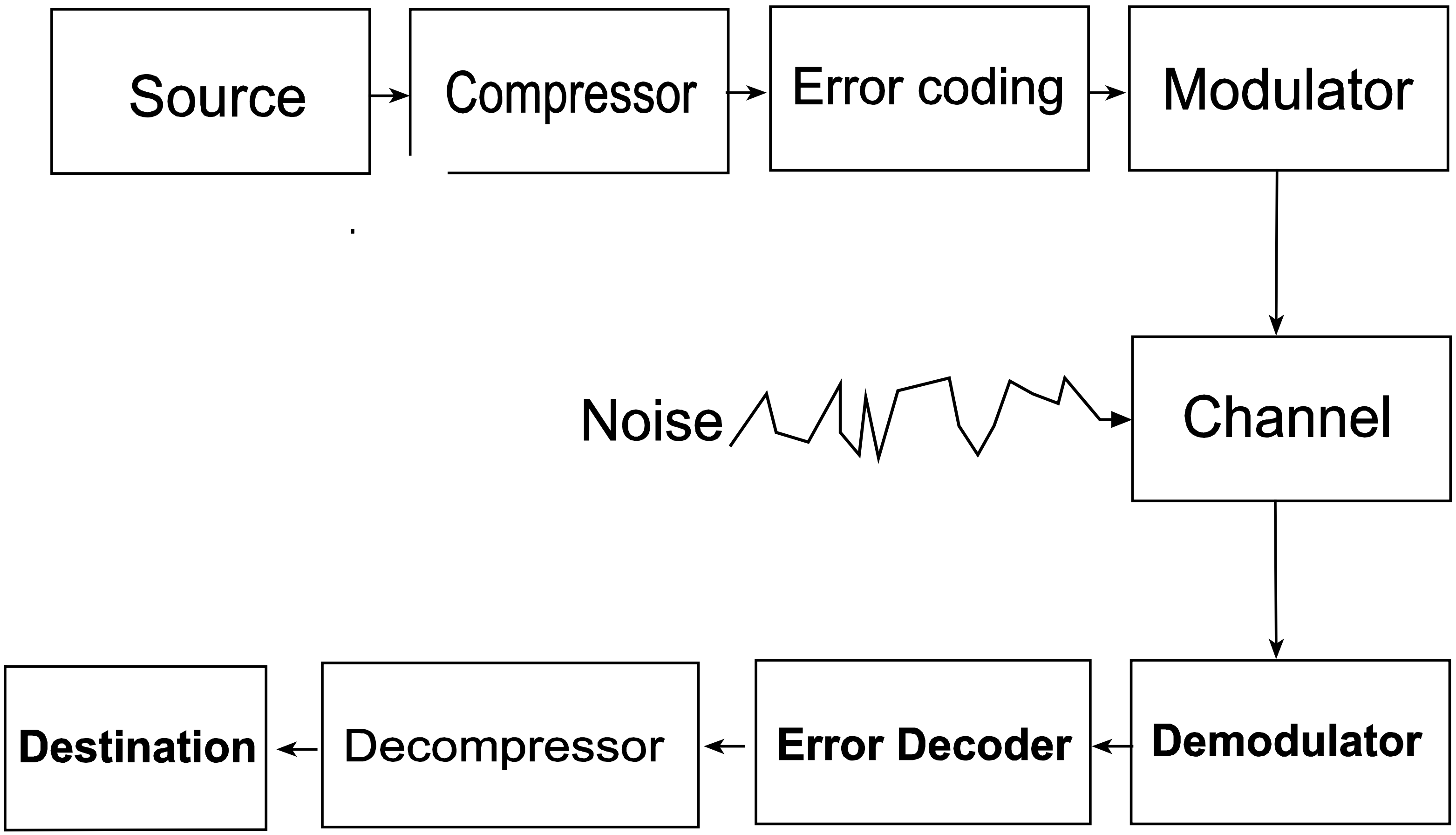
…and sound

Image

Exercise

- 1 s of digital audio: $\simeq 1$ Mb
- Jackson, *Classical electrodynamics*: $\simeq 5$ Mb
- Large encyclopedia: $\simeq 0.5$ Gb
- 1 s of digital video: $\simeq 1.5$ Gb

Model of a communication system



Source

Compressor

Error coding

Modulator

Noise

Channel

Demodulator

Error Decoder

Decompressor

Destination

Binary symmetric channel (BSC): $\begin{array}{ccc} & 0 & 1 \\ 0 & 1-p & p \\ 1 & p & 1-p \end{array}$

$$\begin{array}{c|cc} & 0 & 1 \\ \hline 0 & 1-p & p \\ 1 & p & 1-p \end{array}$$

Example: *The repetition code* [3,1,3]

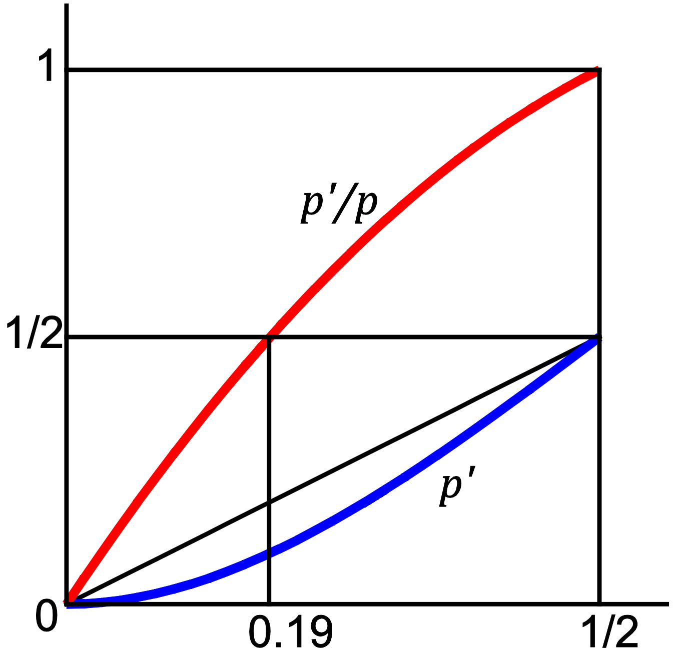
$$f(u) := uuu$$

$$g(x) \coloneqq \textbf{if } \mathrm{weight}(x) > 1$$

$$\textbf{then } 1$$

$$\textbf{else } 0$$

(weight(x)) is the number of 1's in x)



What is the probability p' of error per bit after decoding?

$$p' = 3p^2(1-p) + p^3 \simeq 3p^2$$ (aproximation valid if p is small).

Thus $p'/p \simeq 3p$. For example, if $p = 0.01$, $p'/p \simeq 0.03$, which means that for every 100 channel only 3 remain (on average) after decoding.

Example (the Hamming code [7,4,3]). Let $B = \mathbb{Z}_2$ be the field of binary digits. Consider the B-matrix

$$R = \begin{pmatrix} 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 \end{pmatrix}.$$

Note that the columns of R are the binary vectors of length 3 whose weight is at least 2 (the *weight* of a binary vector is the number of 1's it contains). Writing I_r to denote the identity matrix of order r, let

$$G = I_4|R^T \text{ and } H = R|I_3.$$

Note that the columns of H are precisely the seven non-zero binary vectors of length 3.

Define the block encoding

$$f: B^4 \to B^7,\ u \mapsto uG = u(I_4|R^T) = u|uR^T.$$

The image of this function is $C = \langle G \rangle$, the B-linear subspace spanned by the rows of G (» $\boldsymbol{A}$). We say that C is a [7, 4] code. Since

$$GH^T = (I_4|R^T)\binom{R^T}{I_3} = R^T + R^T = 0$$

(the latter because arithmetic is mod 2), we see that the rows of G, and hence the elements of C, are in the kernel of H^T. In fact,

$$C = \{y \in B^7 | yH^T = 0\},$$

as the right-hand side contains C and both expressions are B-linear subspaces of dimension 4. From the fact that all columns of H are distinct, it is easy to conclude that any two elements of C differ in at least 3 positions, and in some cases exactly in 3, and we express this by saying that C has type [7, 4, 3].

As a decoding function we take the map $g: B^7 \to B^4$ defined by the following recipe:

1. Let $s = yH^T$ (this length 3 binary vector is said to be the *syndrome* of the vector y with respect to H).
2. If $s = 0$, return y (as we said above, $s = 0$ is equivalent to say that $y \in C$). So we can assume that $s \neq 0$.
3. Let j be the index of s as a row of H^T.
4. Negate the bit y_j of y (and still cal y the resulting vector).
5. Return the first four components of y.

Let us show that this decoder corrects 1 error. Indeed, assume that $x \in C$ is the vector that has been sent. If there are no errors, then $y = x$, $s = 0$, and the decoder returns x. Now assume that the j-th bit of $x \in C$ is changed during the transmission, and that the received vector is y. We can write $y = x + \varepsilon_j$, where ε_j is the vector with 1 on the j-th entry and 0 on all the others. Then

$$s = yH^T = xH^T + \varepsilon_j H^T = \varepsilon_j H^T,$$

which clearly is the j-th row of H^T. Thus the result of negating the bit y_j is $y + \varepsilon_j = x = u|uR^T$ and therefore $g(y) = u$.

We notice that the Hamming code has the same correcting capacity as the repetition code [3,1,3], but on the other hand its information rate is 4/7, which is considerably better than 1/3.

Conclusions

Error correction is possible, at the price of increasing the channel time and of incorporating coding and decoding schemes, which also imply higher operation time.

Error correction can be improved, as the Hamming code [7,4,3] does with respect to the repetition code [3,1,3].

This leads to some of the key questions that have driven error-correcting coding techniques in the last six decades.

Basic problems

What limits can be achieved, by means of coding schemes, with regard to:

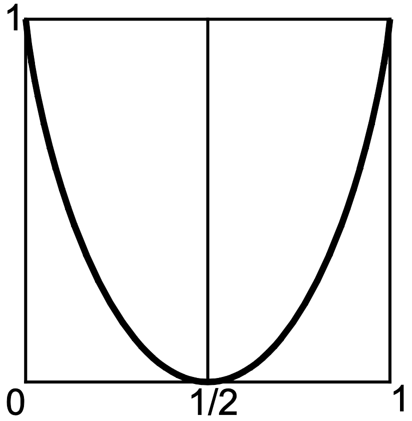
1. *Decreasing errors?*
2. *Increasing the information transmission rate?*
3. *Bounding the computational costs?*

Channel capacity

The *capacity* of a channel is a number c in the interval $[0,1]$ that measures the maximum fraction of the information sent through the channel that is available at the receiving end. In the case of a binary symmetric channel it turns out that

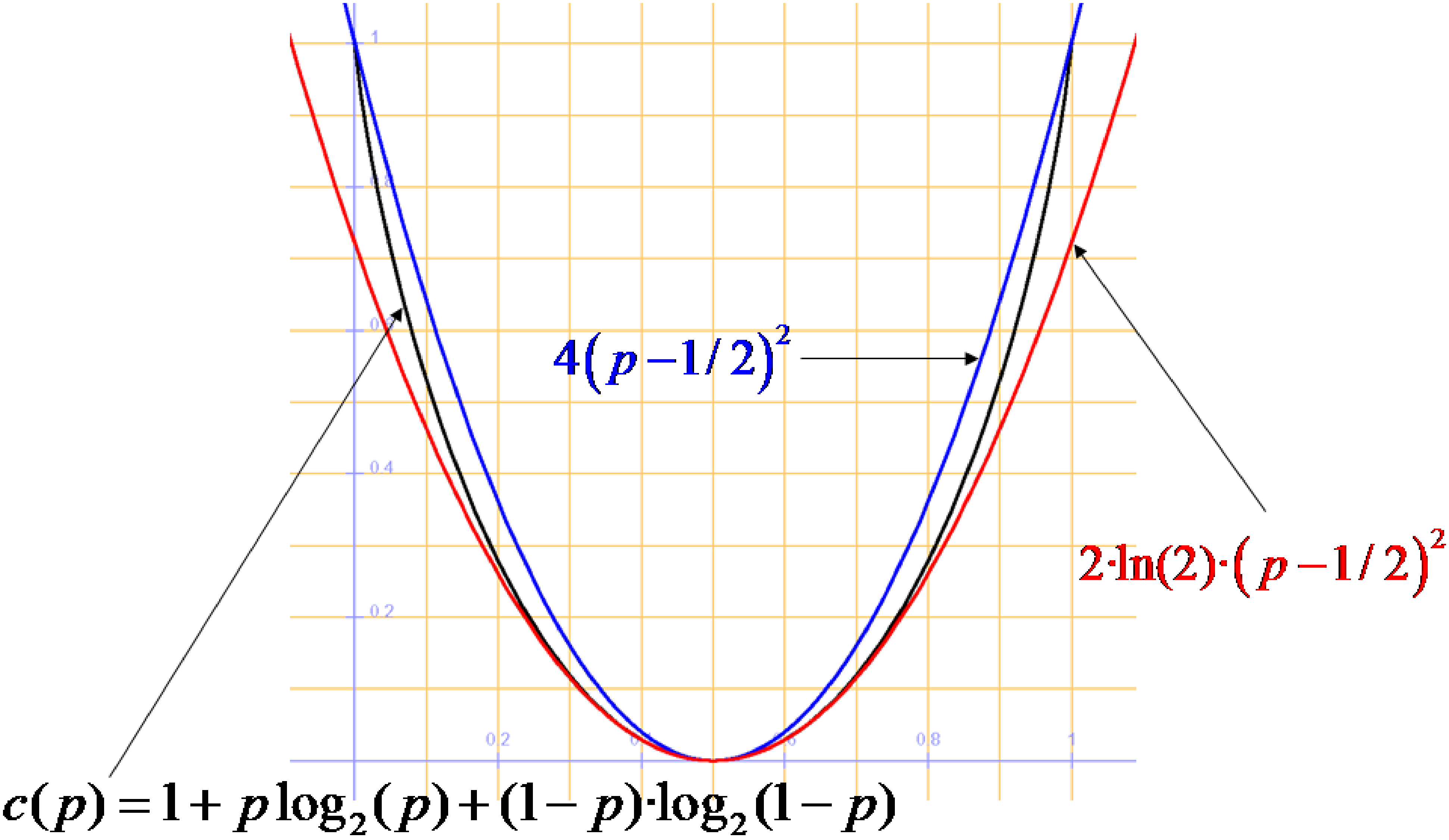
$$C = 1 + p\log_2(p) + (1-p)\ \log_2(1-p),$$

where $\log_2$ is the base 2 logarithm function. Notice that $c = c(p)$ is a strictly decreasing function in the interval $[0,1/2]$, with $c(0) = 1$ and $c(1/2) = 0$, and that $c(p) = c(1-p)$ (see Figure).



$$c\left(10^{-k}\right), \qquad k = 1..6$$

1	2	3	4	5	6
0.531	0.919	0.989	0.9995	0.9998	0.9999



$4\left(p-1/2\right)^2$

$2\cdot\ln(2)\cdot\left(p-1/2\right)^2$

$c(p)=1+p\log_2(p)+(1-p)\cdot\log_2(1-p)$

Channel coding theorem

Shannon's *channel coding theorem* states that if R is a positive real number less than the *capacity* c of a binary symmetric channel ($0 < R < c$), and ε is any positive real number, then there are 'codes' with *rate* at least R and with a probability of code-error less than ε.

Shannon's theorem shows that in theory it is possible to transmit information with sufficient confidence and with a transmission time increase by a factor that can be as close to $1/c$ as desired.

Unfortunately, his methods only show the existence of such codes, but do not produce them, nor their coding and decoding, in an *effective* way.

It can be said that the main motivation of the theory of error-correcting codes in the last sixty years has been, to a great extent, to find

- explicit codes with good rates,
- small code-error probabilities, and
- with fast coding and decoding procedures.

$\boldsymbol{A}$. In general, the product uG of a vector $u = (u_1, \dots, u_k)$ and a $k \times n$ matrix G is the linear combination $u_1 g^1 + \cdots + u_k g^k$ of the k rows $g^1, \dots, g^k$ of G with coefficients the components of u. It follows that the set $\{uG \mid \text{any}\ u\ \}$ is the (linear) span of the rows of G, often denoted $\langle G \rangle$.