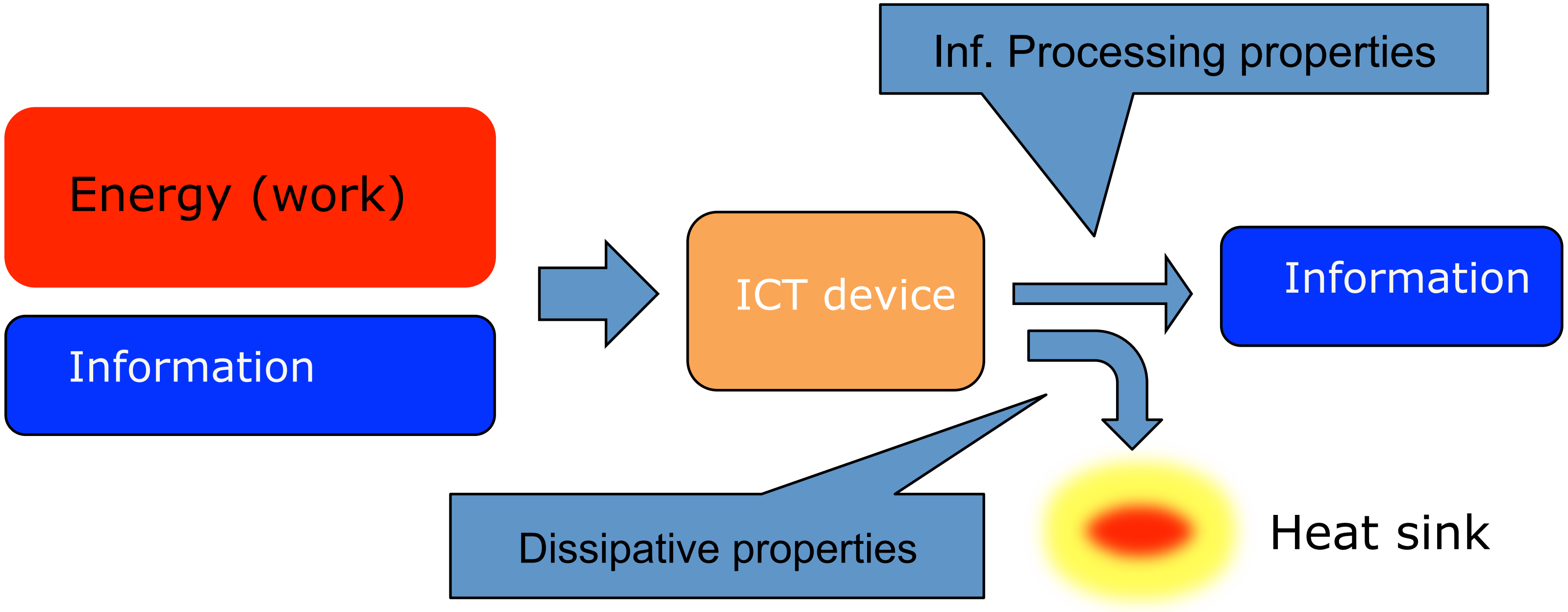
Introduction to
Energy, **Entropy** and **Information**
in small scale physical systems

Prof. Luca Gammaitoni

NiPS Laboratory, University of Perugia

Key-point in our approach:
A general approach to ICT devices as micro/nano thermal machines

An **ICT device** is a machine that inputs information and energy (under the form of work), processes both and outputs information and energy (mostly under the form of heat).



So... the main characters are: Energy (work, heat), Information.

Energy, Entropy, Information

Energy

Capability of doing WORK…
WORK = FORCE x Displacement

Energy is conserved (First Principle)

Entropy

Capability of exchanging WORK and HEAT …

Entropy increases in spontaneous transf. (Second Principle)

What about **Information** ?

From latin/italian: INFORMARE

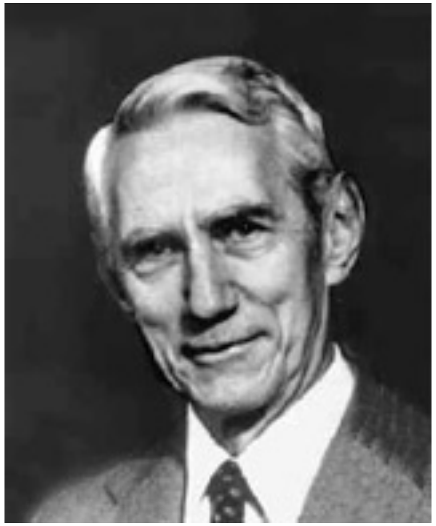
informo, informas, informavi, informatum, informāre

Meaning: “to **give shape** to something”

extended meaning“to instruct somebody (give shape to the mind)”

Relation between information and communication

Claude Elwood Shannon
(Gaylord, Michigan 1916 -
Medford, Massachussets 2001),



Reprinted with corrections from *The Bell System Technical Journal*,
Vol. 27, pp. 379–423, 623–656, July, October, 1948.

A Mathematical Theory of Communication

By C. E. SHANNON

INTRODUCTION

THE recent development of various methods of modulation such as PCM and PPM which exchange bandwidth for signal-to-noise ratio has intensified the interest in a general theory of communication. A basis for such a theory is contained in the important papers of Nyquist[1] and Hartley[2] on this subject. In the present paper we will extend the theory to include a number of new factors, in particular the effect of noise in the channel, and the savings possible due to the statistical structure of the original message and due to the nature of the final destination of the information.

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.

If the number of messages in the set is finite then this number or any monotonic function of this number can be regarded as a measure of the information produced when one message is chosen from the set, all choices being equally likely. As was pointed out by Hartley the most natural choice is the logarithmic function. Although this definition must be generalized considerably when we consider the influence of the statistics of the message and when we have a continuous range of messages, we will in all cases use an essentially logarithmic measure.

The logarithmic measure is more convenient for various reasons:

1. It is practically more useful. Parameters of engineering importance such as time, bandwidth, number of relays, etc., tend to vary linearly with the logarithm of the number of possibilities. For example, adding one relay to a group doubles the number of possible states of the relays. It adds 1 to the base 2 logarithm of this number. Doubling the time roughly squares the number of possible messages, or doubles the logarithm, etc.
2. It is nearer to our intuitive feeling as to the proper measure. This is closely related to (1) since we intuitively measures entities by linear comparison with common standards. One feels, for example, that two punched cards should have twice the capacity of one for information storage, and two identical channels twice the capacity of one for transmitting information.
3. It is mathematically more suitable. Many of the limiting operations are simple in terms of the logarithm but would require clumsy restatement in terms of the number of possibilities.

The choice of a logarithmic base corresponds to the choice of a unit for measuring information. If the base 2 is used the resulting units may be called binary digits, or more briefly *bits*, a word suggested by J. W. Tukey. A device with two stable positions, such as a relay or a flip-flop circuit, can store one bit of information. N such devices can store N bits, since the total number of possible states is 2^N and $\log_2 2^N = N$. If the base 10 is used the units may be called decimal digits. Since

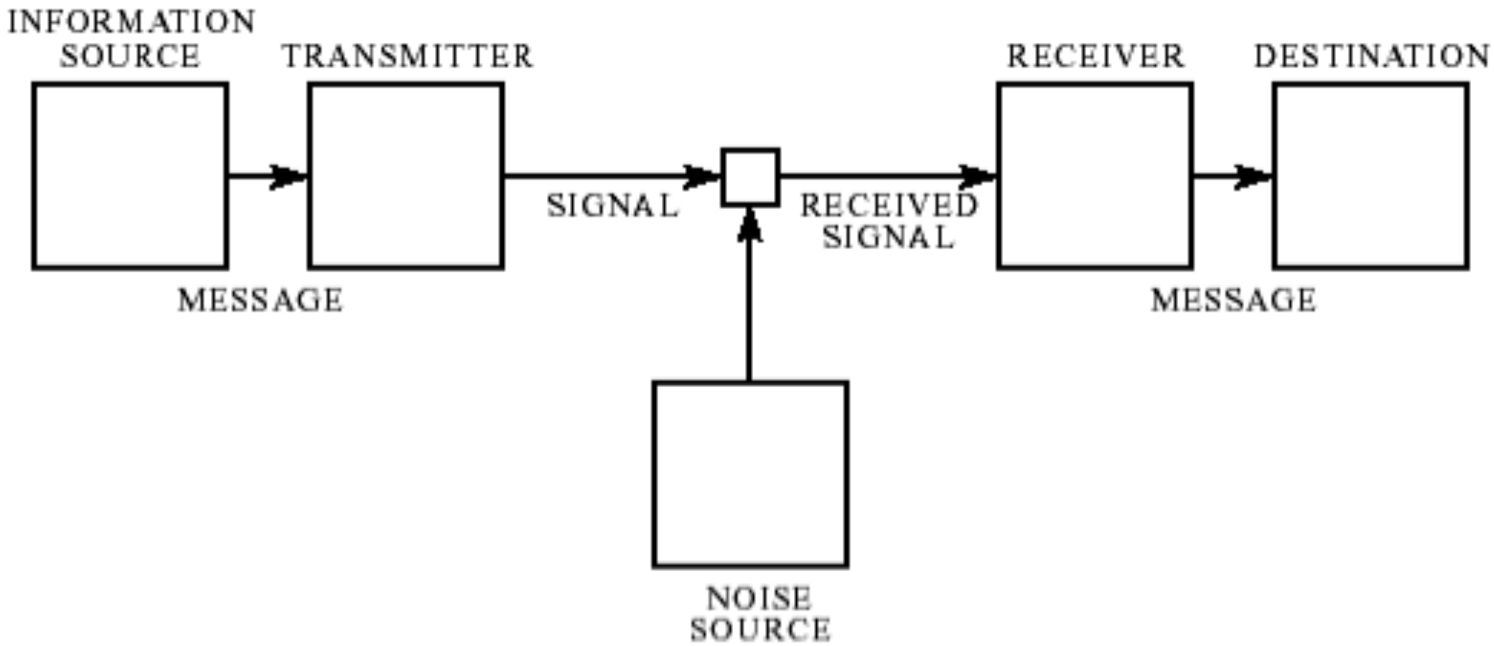
$$\log_2 M = \log_{10} M / \log_{10} 2 = 3.32 \log_{10} M.$$

[1] Nyquist, H., "Certain Factors Affecting Telegraph Speed," *Bell System Technical Journal*, April 1924, p. 324; "Certain Topics in Telegraph Transmission Theory," *A.I.E.E. Trans.*, v. 47, April 1928, p. 617.
[2] Hartley, R. V. L., "Transmission of Information," *Bell System Technical Journal*, July 1928, p. 535.

1

C. Shannon, 1948 - ***A Mathematical Theory of Communication***

Information and communication



C. Shannon, 1948 *A Mathematical Theory of Communication*

Available at: http://www.fisica.unipg.it/~gammaitoni/info1fis/documenti/shannon1948.pdf

Information: what is it?

It is a property of a message.
A message made for communicating something.

We say that the information content of a message is greater the greater is its *casualty*.

In practice the less probable is the content od the message the more is the information content of that message.

Let's see examples…

Information: what is it?

Let's suppose we are waiting for an answer to a question.
The answer is the message.

Case 1:	answer yes	(probability 50%)
	answer no	(probability 50%)

The two messages have the same information content.

Information: what is it?

Let’s suppose we are waiting for an answer to a question.
The answer is the message.

Case 2:	answer yes	(probability 75%)
	answer no	(probability 25%)

The two messages have the different information content.

Information: how do we measure it?

Let's suppose we want to transmit a text message:

My dear friend….

We have a number of symbols to transmit… 25 lower case letters + 25 upper case letters + punctuation + …

Too large a number of different symbols… it is unpractical.

We can use a coding that assign letters to numbers.
E.g. the ASCII code: A=65, B=66, C=67, … a = 97, b=98, c= 99 …

The advantage is that we have a small number of different symbols:

0,1,2,3,4,5,6,7,8,9

But the message becomes longer…
Example: caro amico --------------> 67 97 114 111 97 109 105 99 111

Information: how do we measure it?

We send the message: 67 97 114 111 97 109 105 99 111

How much information are we sending?

We assume that information is an additive quantity, thus the information of the message is the sum of the information of the single components of the message, i.e. the symbols.

Now: if I send the symbol “4” how much information is in it?

Answer: it depends on the probability of that symbol, meanaing the probability that the specific symbol “4” happens to be in my message.

Information: how do we measure it?

We send the message: 67 97 114 111 97 109 105 99 111

If we call p_4 the probability of having "4"and generically
p_x the probability of having the symbol "x" (a given number) we have:

$$H = - K\ p_x \log p_x$$

Amount of information associated with symbol "x".
This is technically known also as "Entropy".

Information: how do we measure it?

We send the message: 67 97 114 111 97 109 105 99 111

If we have a message with n_x symbol "x"; n_y symbol "y" and so on.. :

$$H = - K (n_x \, p_x \log p_x + n_y \, p_y \log p_y + \ldots)$$

Information: binary is better

In order to reduce the error probability during transmission is more convenient codify the numbers in base 2, with only two symbols: “0”, “1”

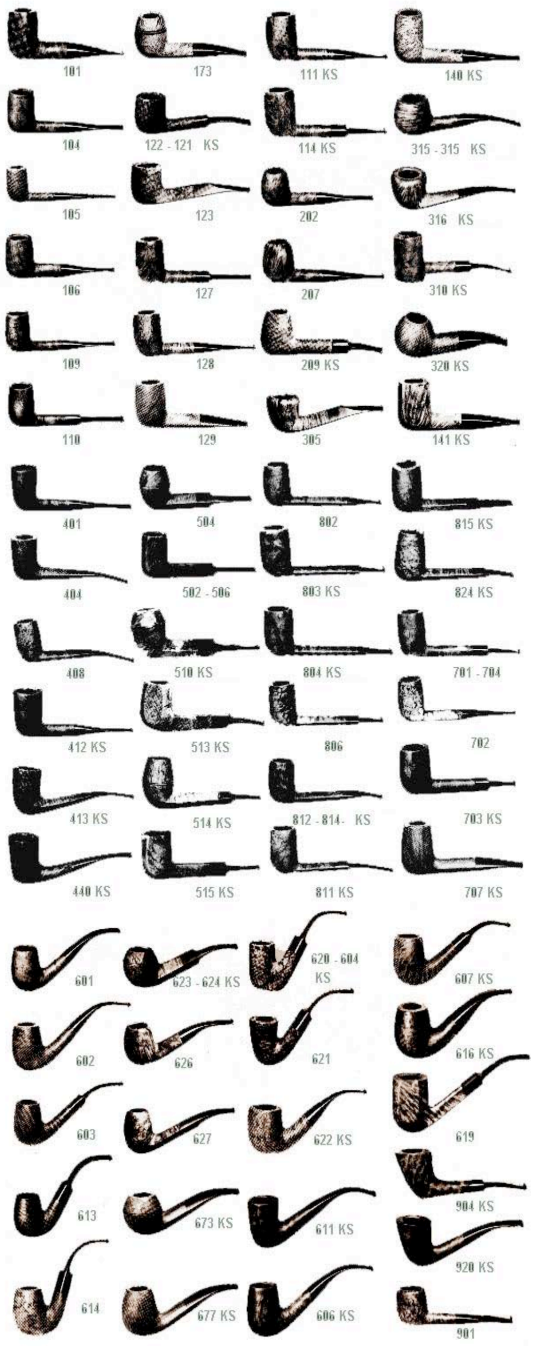
Now our message appears like: 0110110000101000111

If it is long m characters (with m large), the probability $p_1 = p_2 = 1/2$

$$H = - K\, m\, 1/n \log (1/n)$$
$$= - K\, m/n\, (- \log n) = K\, m/n \log n$$

$$H = K\, m/n \log n = 2\, m/2 \log_2 2 = m$$

Thus $H = m$ = number of bits



information

In-forma = in shape

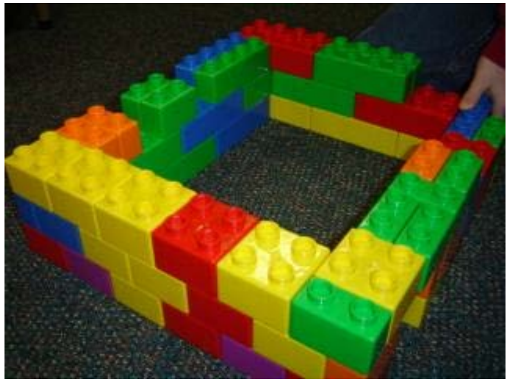
Information =
to put something in shape

Forma = shape

The shape of an object is
a visual maifestation of the amount of
Information encoded in that object…

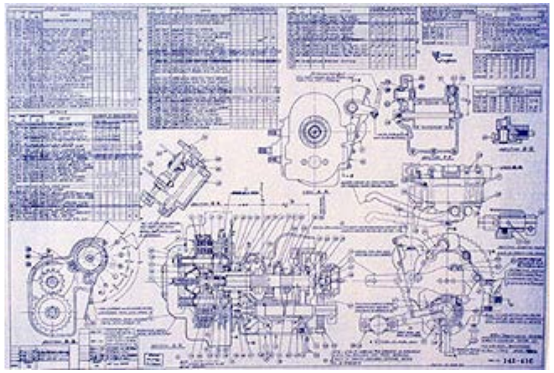
Example with LEGO bricks

Shape = Pattern = Configuration… FORMA



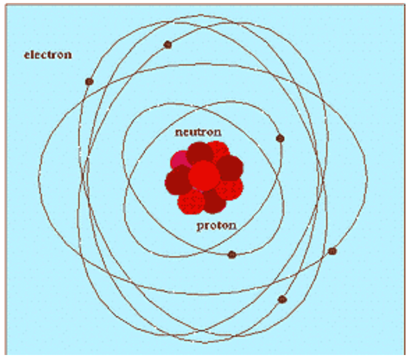
Object = bricks elements + information

+

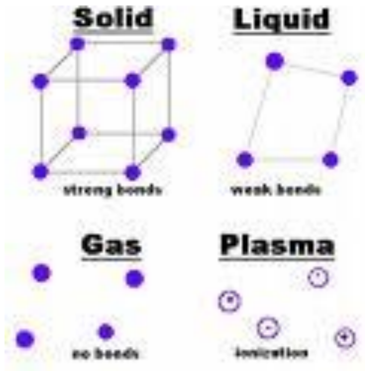


=

Atoms + information = matter



+

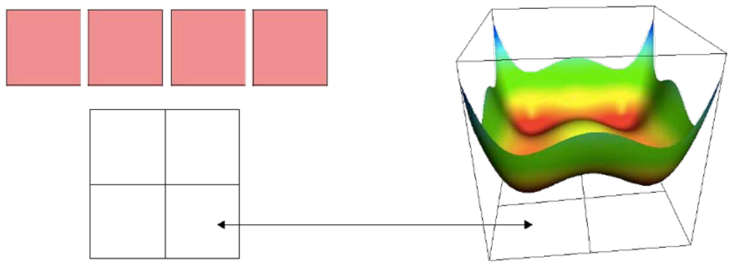


=

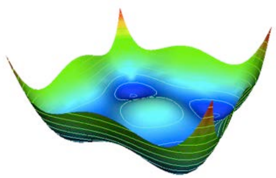
How much information ?

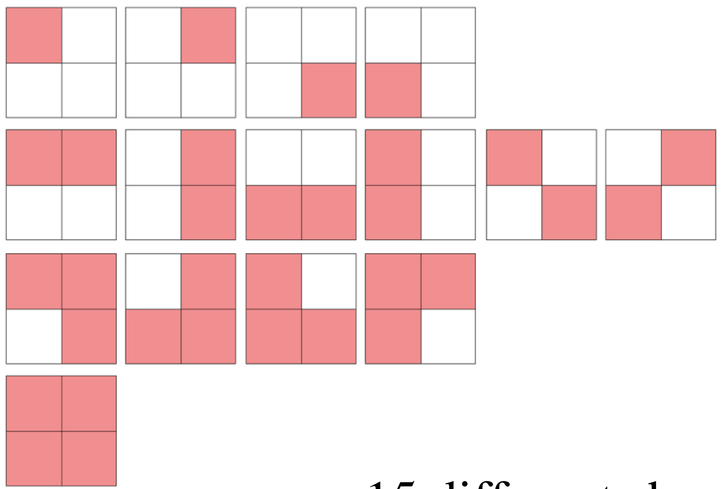
It depends on the coding mechanism.

Let's consider a simple example…



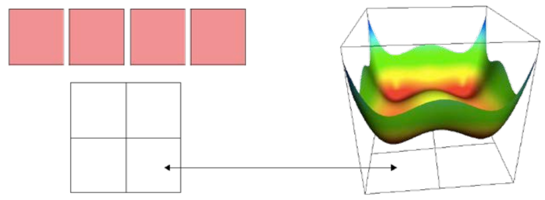
1) Define shape

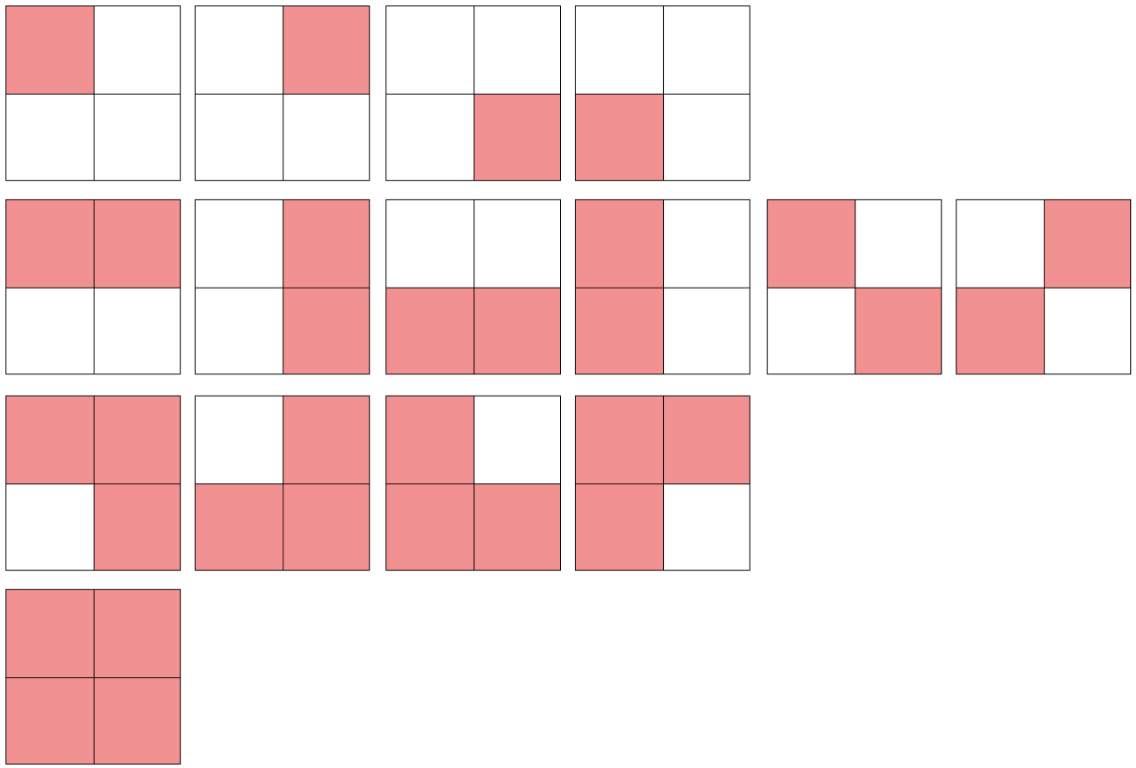




15 different shapes

2) Count shapes

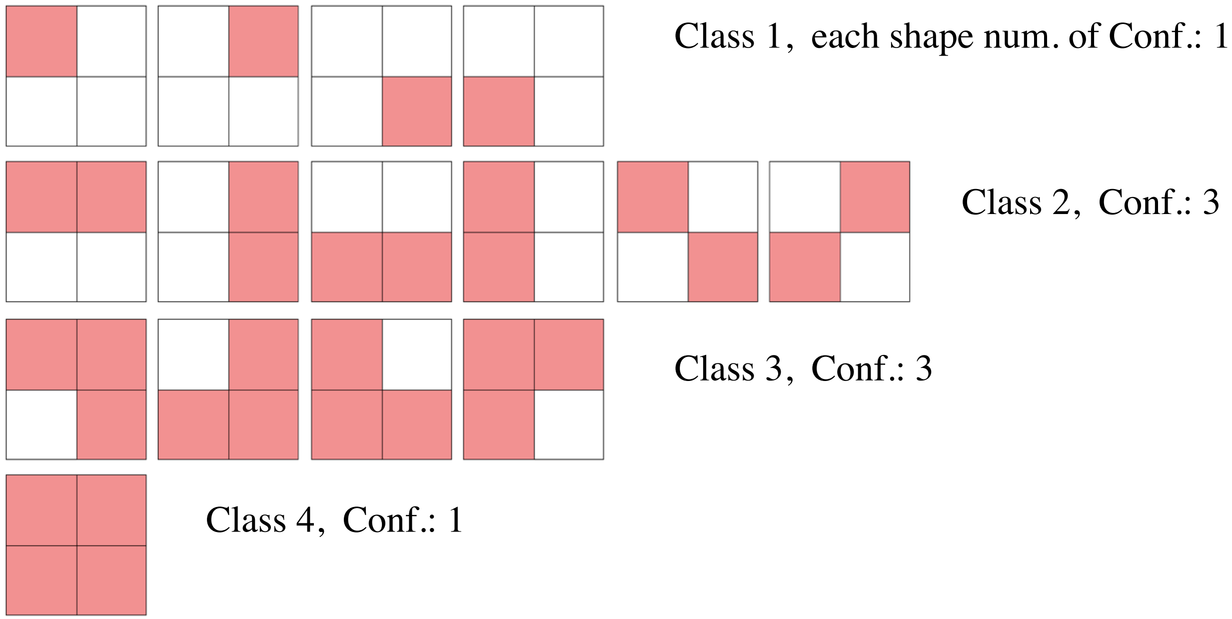




15 different shapes

3) Count configurations

Indistinguishable particles. Each shape can be realized with a different arrangement of the marbles



In general…

In general if we have q indistinguishable particles that can be distributed in r distinguishable sites, a single shape s_{ij} is characterized by two indexes: the class index $i = 1, 2, ..r$ and, within a single class, the shape index $j = 1, 2, ..C(r, i)$ where $C(r, i)$ is the binomial coefficient. The total number of different shapes is given by

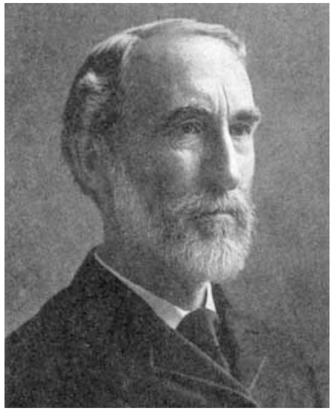
$$N_S = \sum_{i=1}^{r} C(r, i) \tag{1}$$

The number of configurations for each given shape s_{ij}, N_{ij}, depends only on the shape class, i.e. $N_{ij} = N_i$ and this is given by:

$$N_i = C(q-1, i-1) = \frac{(q-1)!}{(i-1)!(q-i)!} \tag{2}$$

The total number of possible configurations is given by $N = C(q + r - 1, r - 1)$.

In our example with $q = 4$ and $r = 4$ we have $N_S = 15$ and $N = 35$ while $N_1 = 1$, $N_2 = 3$, $N_3 = 3$ and $N_4 = 1$.



4) Entropy

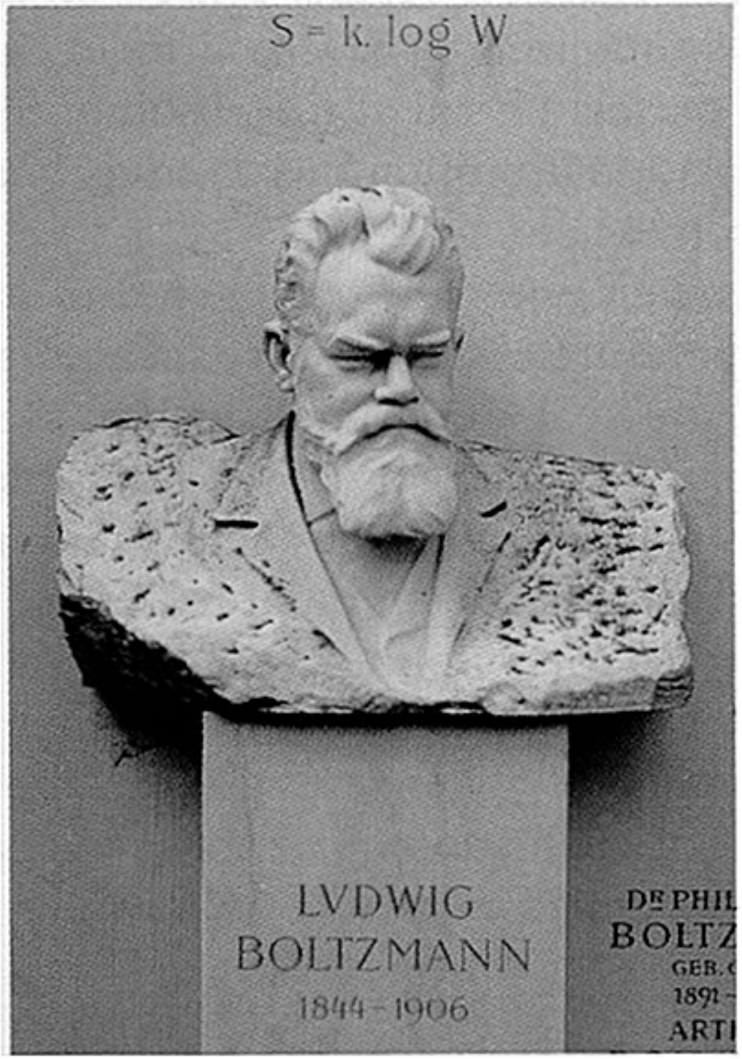
We define *shape entropy* the quantity

$$S_i = K \ln N_i$$

where K is an arbitrary constant. This quantity coincides with the microscopic form given by Boltzmann and Gibbs of the thermodynamic entropy initially introduced by Clausius, if we interpret the number of configurations N_i for a given shape as the number of accessible microstates for a given state of the thermodynamical system. Specifically, Gibbs entropy is given by

$$S_G = -K \sum_l p_l \ln p_l$$

p_l is the probability of the microstate of index l and the sum is taken over all the microstates.

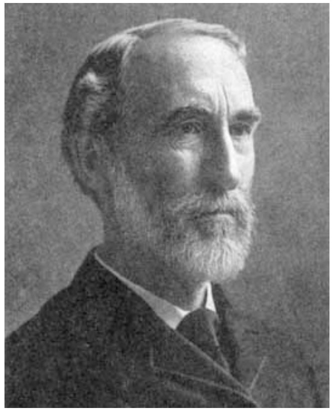


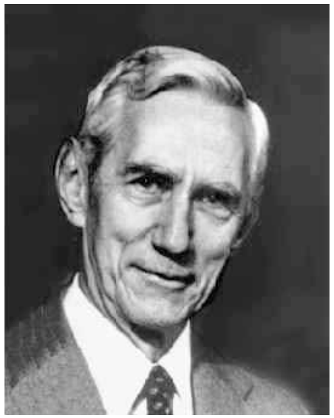
If the probability of the microstates are all the same, then the Gibbs entropy reduces to the Boltzmann entropy.

Thus if we identify the microstate of a physical object with a configuration that realizes one shape we have that the shape entropy IS the Boltzmann entropy of our object.

Up to this point we have shown that the shape of a physical object can be associated with a physical observable called “shape entropy” and that the shape entropy IS the physical entropy defined by Boltzmann.

What about information ?

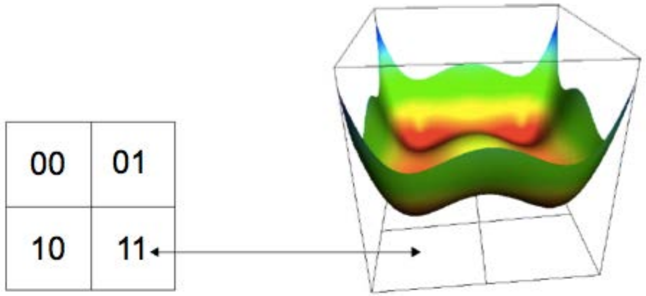




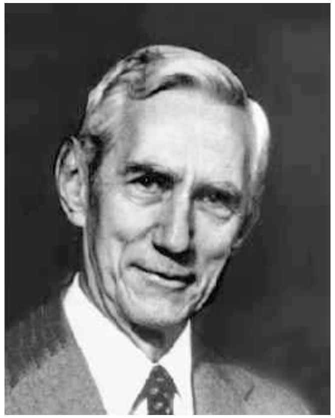
5) Shape and information

To associate an information content with a shape we select the following coding system: we use 2 bits per site identifying the occupation of a site as follows.

a particle on the:
- upper left 00,
- upper right 01,
- lower right 10,
- lower left 11.

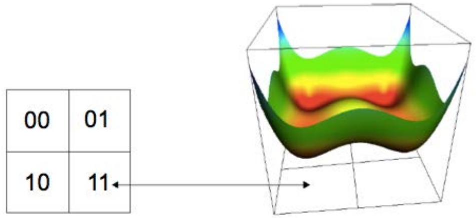


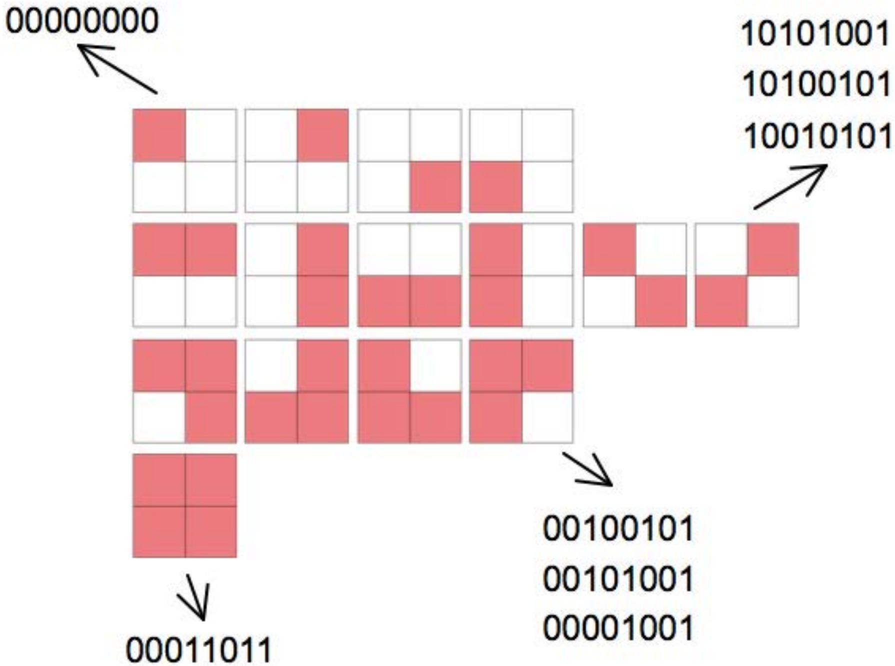
One configuration is represented by the occupation of the fours sites and thus requires 8 bits (whose order is immaterial due to the undistinguishable character of the particles).

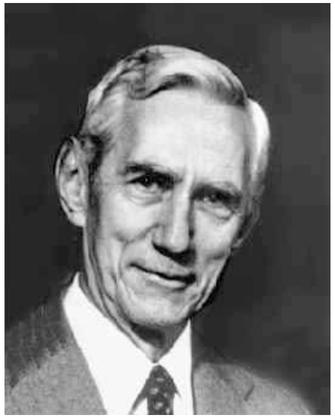


5) Shape and information

Each configuration corresponds to a different set of 8 bits

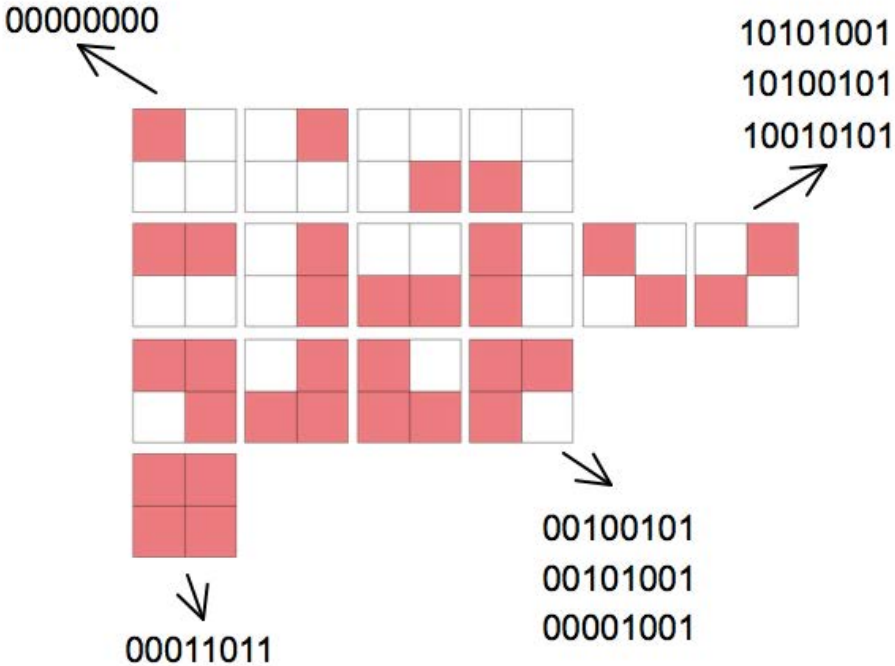


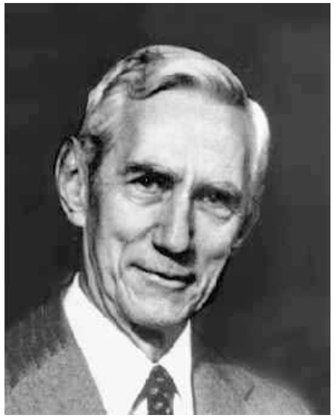




5) Shape and information

How much information is there in each set of 8 bits?
(i.e. how much information is there in each configuration and thus in each shape?)





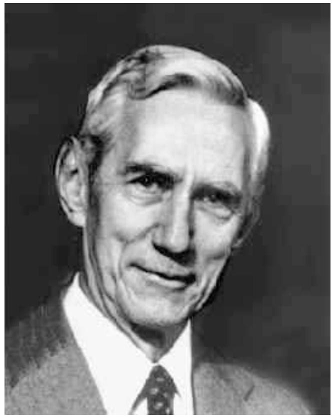
5) Shape and information

How much information is there in each set of 8 bits?
(i.e. how much information is there in each configuration and thus in each shape?)

As we have seen, a given shape can be realized by $\mathbf{N_i}$ different configurations. The probability of a single configuration (represented by a given set of 8 bits) is given by $\mathbf{p_i} = 1/\mathbf{N_i}$ thus the shape information is computed according to Shannon by:

$$S_i = -K \sum_{l=1}^{N_i} p_l \ln p_l = -K\, N_i \frac{1}{N_i} \ln \frac{1}{N_i} = K \ln N_i$$

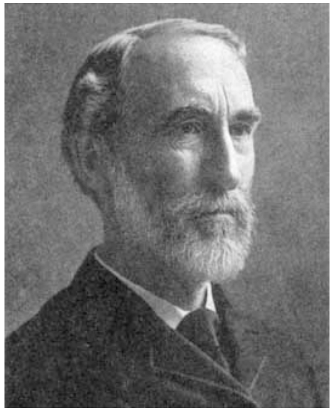
This is same quantity that we have called shape entropy and thus we can interpret the shape entropy as a measure of the information content of a given shape.

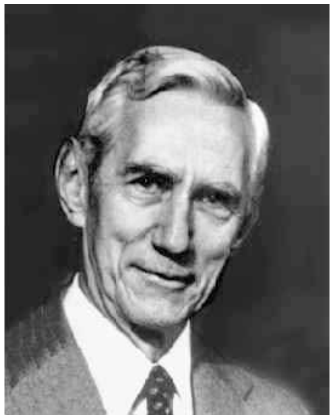


5) Shape and information

It is interesting to note that the amount of information differs from class to class. The maximum information is embedded into a random shape, meaning with this a shape whose *class* is populated by the whole configuration space: $S_{max} = K \ln N$. In our four particle example, if we assume $K = \log_2 e$ and the base 2 for the log function, we obtain $S_{max} = 7.40\ bits$. Accordingly, the information in any shape belonging to class $i = 2$ and $i = 3$ is $S_i = K \log_2(3) = 2.29\ bits$, while the shape information in any shape belonging to class $i = 1$ and $i = 4$ is just $S_i = K \log_2(1) = 0\ bits$.

6) The shape of things changes spontaneously with time





Thus at thermal equilibrium the shape of things changes spontaneously with time according to the second principle of thermodynamics.

According to the laws of thermodynamics, an object in thermal equilibrium with the environment will spontaneously change its shape according to the **maximization of the shape entropy**.

By randomly shaking our marble cartoon we will produce a shape change according to a maximization of the shape entropy (information) associated with each shape.

6) The shape of things changes spontaneously with time

before

after

6) The shape of things changes spontaneously with time

7) The shape of things changes in a preferred direction

Sometimes this is called irreversibility of spontaneous transformations but is simply a manifestation of the tendency of a system to evolve toward the **most probable shape** (that has the largest number of configurations). This is the content of the second law of thermodynamics according to Boltzmann.

We go from order to disorder

before

after

Question: can we change the shape of things the other way around?

If so,… is there a cost to pay?


Shape is physical

In fact, once we have the information associated with each shape, following Landauer and Bennet we deduce that any shape dynamics that involves a change in the information content must play a physical role as well. Specifically, for an isolated system a shape change that implies an increase of shape entropy should come at expenses of a corresponding decrease of the free energy.

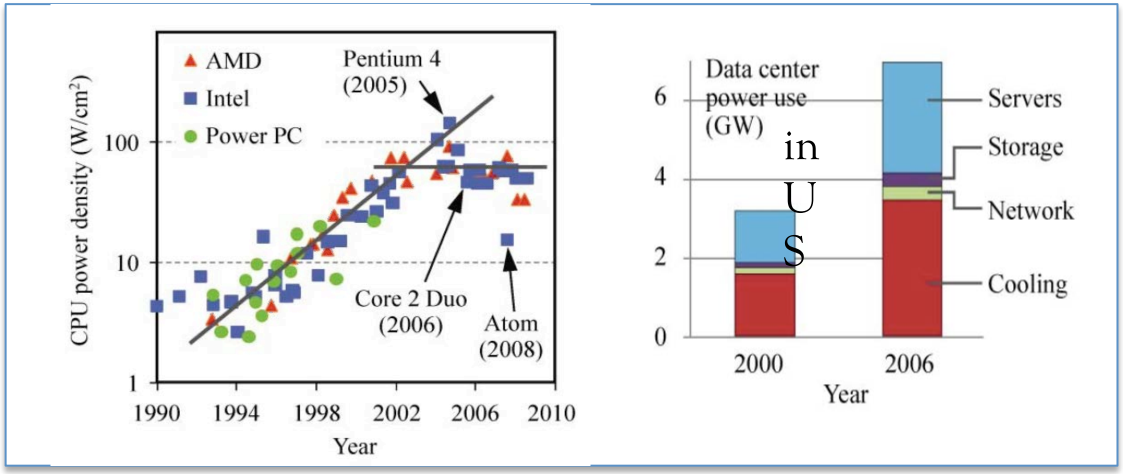

Answer to question

On the other hand, if we want to perform a shape modification that implies a net shape entropy change of $\Delta S < 0$, this requires a minimum of energy to be dissipated during the transformation equal to $Q = -k_B T \Delta S$.

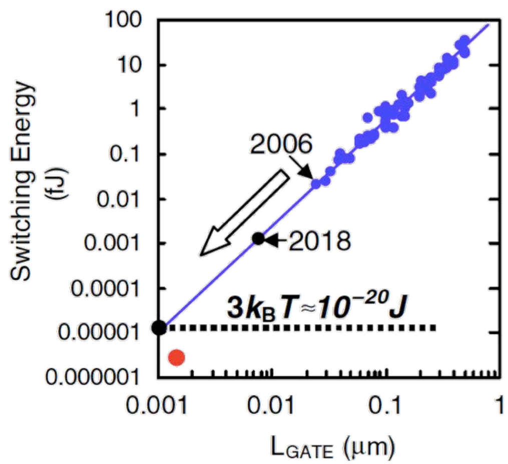
Consequences for ICT…

ICT - Energy

Energy efficiency in computing systems has become a major issue for the future of ICT



E. Pop, *Energy Dissipation and Transport in Nanoscale Devices*, Nano Res (2010) 3: 147–169



Research directions and challenges in nanoelectronics
R. K. Cavin1, V. V. Zhirnov, D. J. C. Herr1, Alba Avila and J. Hutchby, 2006

Present status

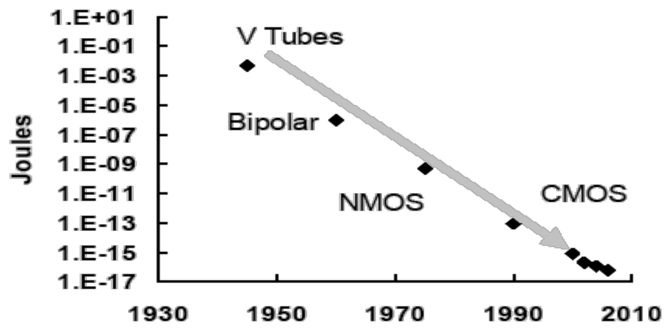


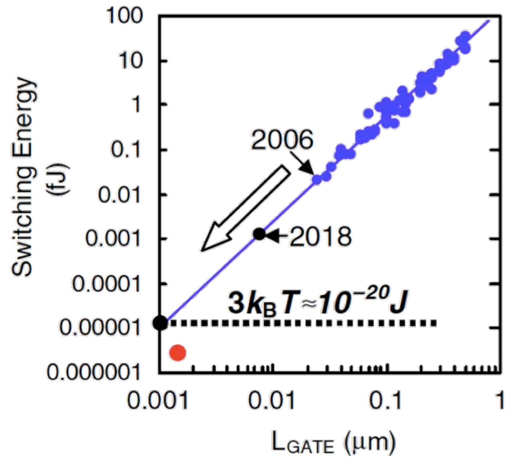
Figure 3: Energy per logic operation

Electronics Beyond Nano-scale CMOS, Shekhar Borkar

ICT - Energy

…the resulting power density for these switches at maximum packing density would be on the order of 1MW/cm^2 – orders of magnitude higher than the practical air-cooling limit..

Jeffrey J. Welser
The Quest for the Next Information Processing Technology , 2008



The search for alternative switches is presently very active*, however, even if a new information vector will be found that overcomes the limitation of charge driven FET switches a more fundamental energy limit stands in the path toward zero-power dissipation: the Landauer's limit.

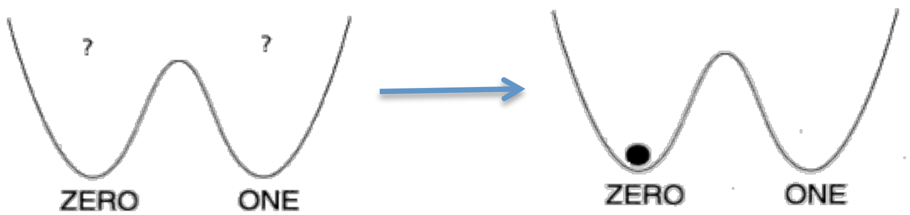
*To take on this grand challenge, the Nanoelectronics Research Initiative (NRI) (nri.src.org) was formed in 2004 as a consortium of Semiconductor Industry Association (SIA) (www.sia-online.org) companies to manage a university-based research program as part of the Semiconductor Research Corporation (SRC) (www.src.org). The NRI was founded by six U.S. semiconductor companies (AMD, Freescale, IBM, Intel, Micron, and TI), and partners with the National Science Foundation (NSF), the National Institute of Standards and Technology (NIST), and state governments, sponsoring research currently at 35 U.S. universities in 20 states.

THE LANDAUER'S LIMIT

The Landauer's principle (1) states that erasing one bit of information (like in a resetting operation) comes unavoidably with a decrease in physical entropy and thus is accompanied by a minimal dissipation of energy equal to

$$Q = k_B T \ln 2$$

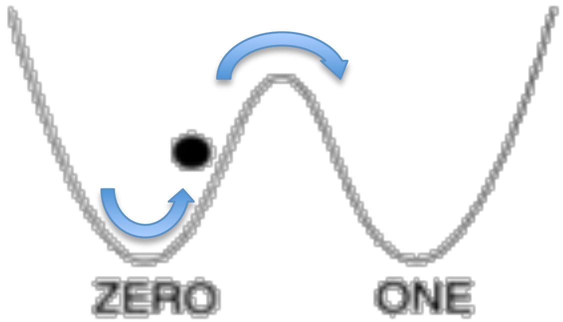
More technically this is the result of a change in entropy due to a change from a random state to a defined state

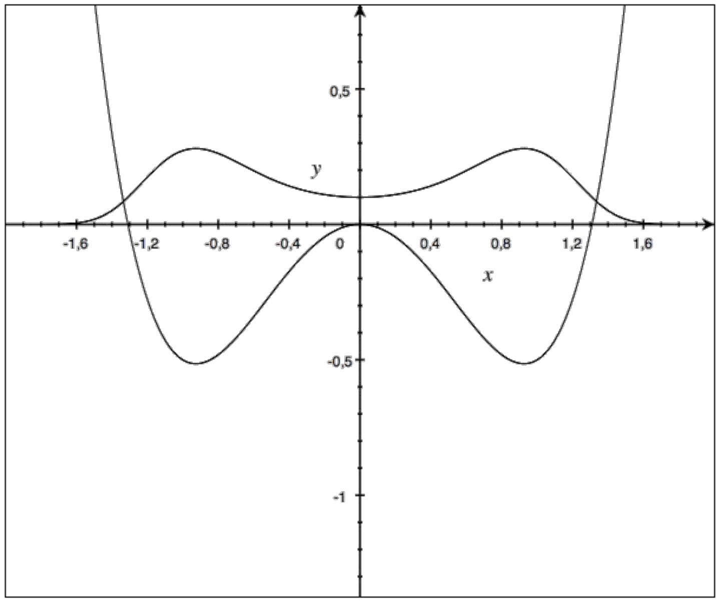


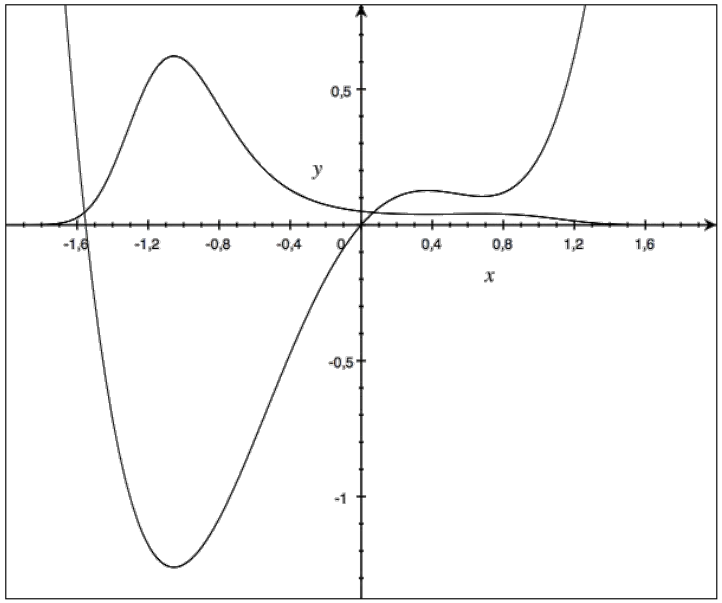
(1) R. Landauer, "Dissipation and Heat Generation in the Computing Process" *IBM J. Research and Develop. 5*, 183-191 (1961),

Real world scenario

The bistable switch is in contact with a thermal reservoir at temperature T. The point particle perform random motion.



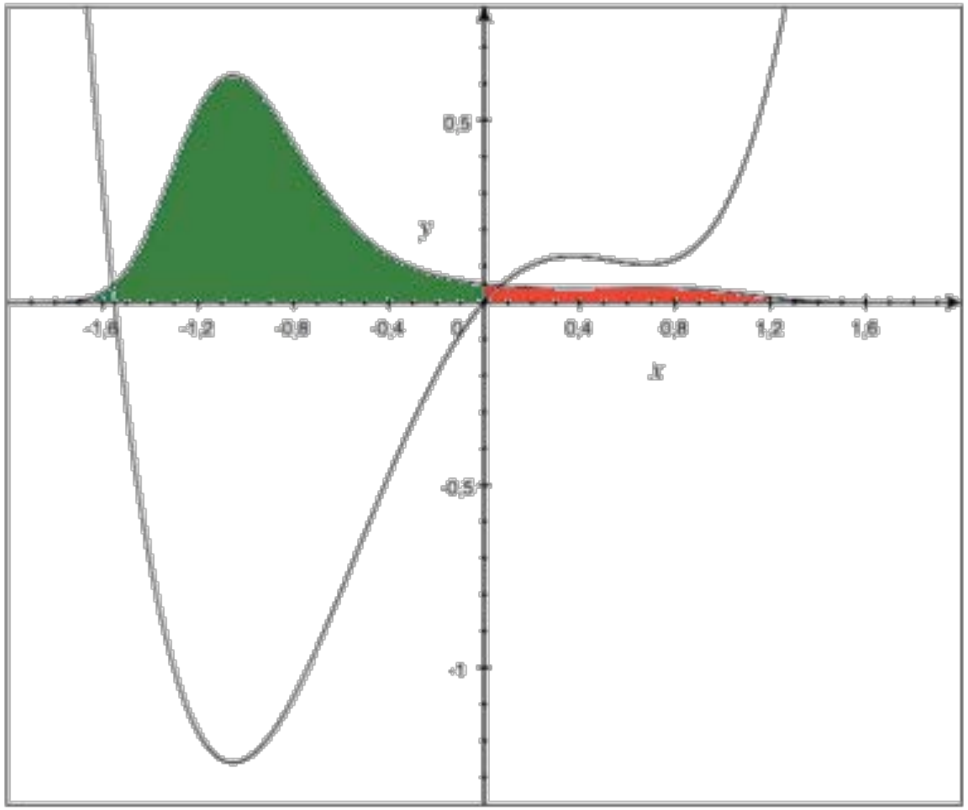




Reset operation

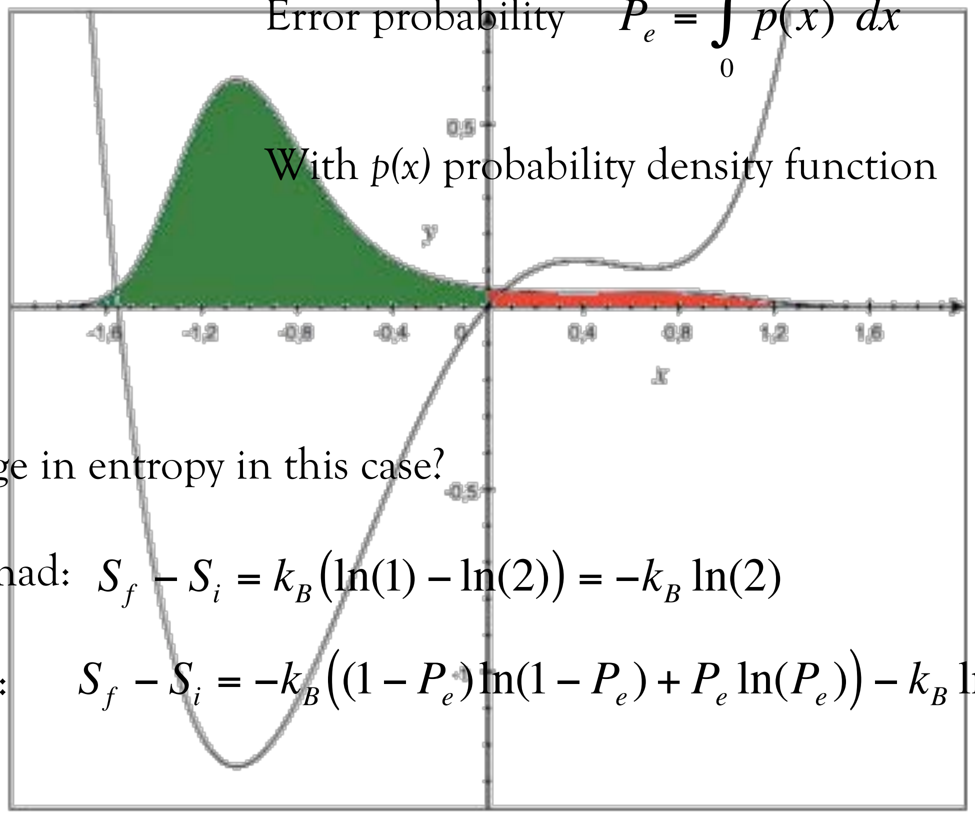
Probabilities

In a real world switch there is a finite probability that the reset operation generate errors



Probabilities

In a real world switch there is a finite probability that the reset operation generate errors



Error probability $P_e = \int_0^{\infty} p(x)\ dx$

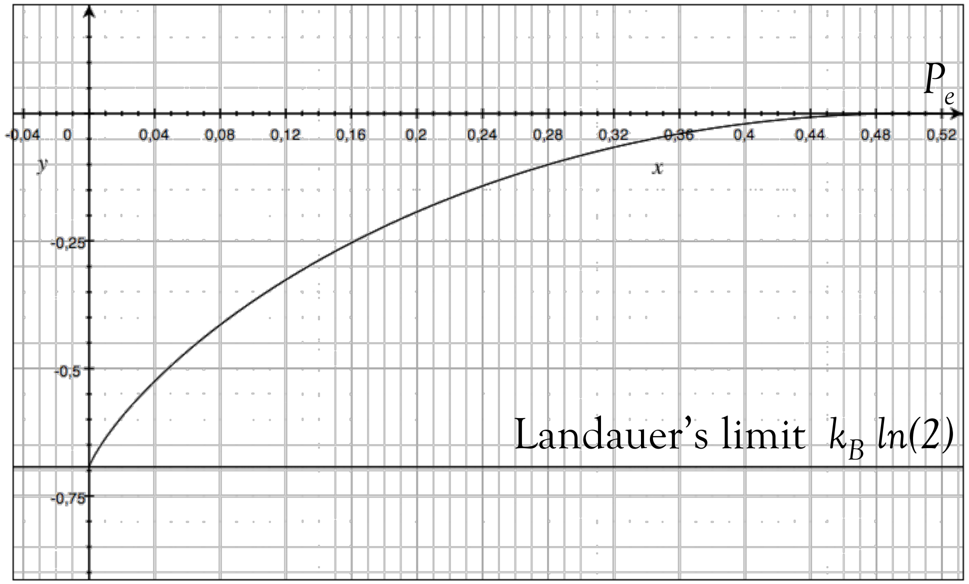
With $p(x)$ probability density function

What is it the change in entropy in this case?

With Landauer we had: $S_f - S_i = k_B\left(\ln(1) - \ln(2)\right) = -k_B \ln(2)$

In this case we have: $S_f - S_i = -k_B\left((1 - P_e)\ln(1 - P_e) + P_e \ln(P_e)\right) - k_B \ln(2)$

Entropy difference as a function of $\boldsymbol{P_e}$



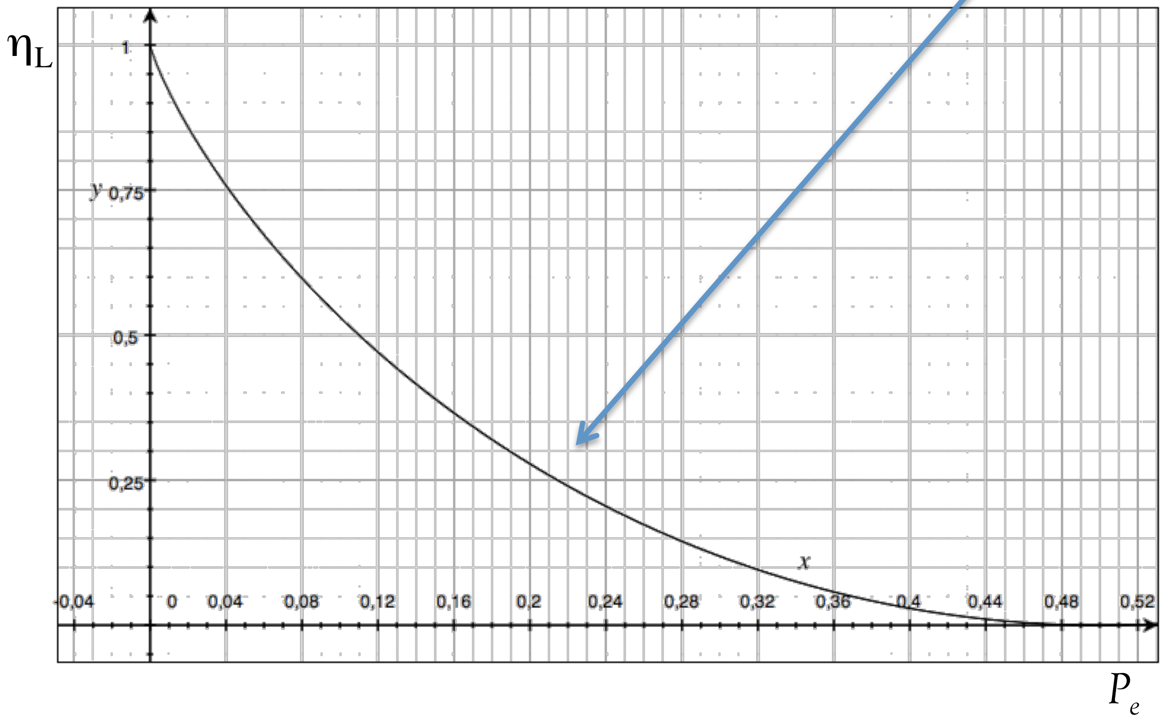
The more increases the probability of error
The more decreases the entropy difference
…and thus the dissipated energy.

When the probability of error is zero we recover the Landauer's limit.

Minimum energy ratio for reset operation*

$$\eta_L(P_e) = \frac{Q(P_e)}{Q} = 1 + \frac{(1-P_e)\ln(1-P_e) + P_e\ln(P_e)}{\ln(2)}$$

With 20% probability of error
The minimum energy is about ¼ of Landauer's limit.



* L. Gammaitoni, *Beating the Landauer's limit by trading energy with uncertainity*, 2011, Arxiv

Conclusions

1) Energy, entropy and Information are connected
2) Information is a manifestation of shape entropy
3) Changing shape may take energy
4) Computing is altering information and thus may take energy
5) A physical switch is a changing shape system and thus…