ENR-325/325L Principles of Digital Electronics and Laboratory

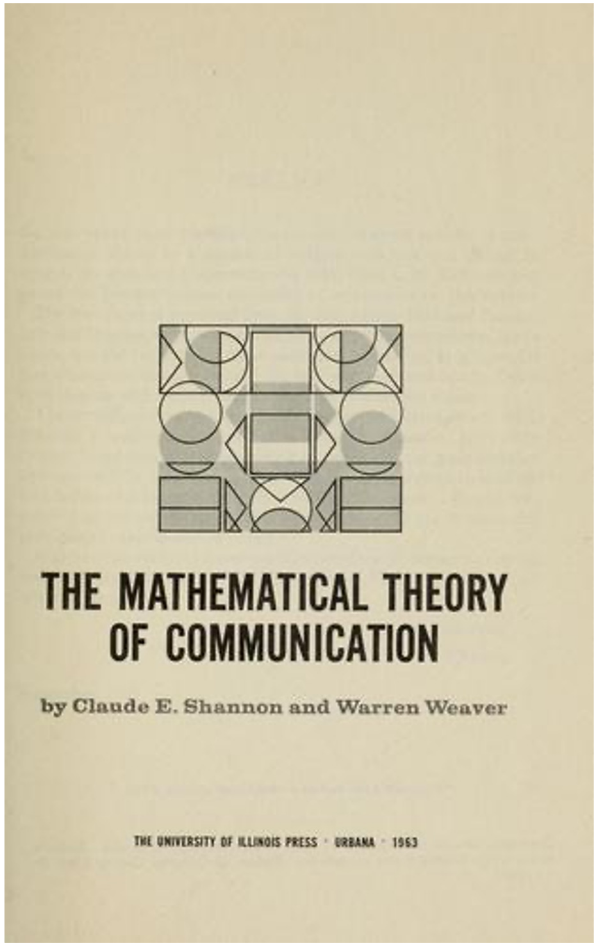
Xiang Li

Fall 2025

Coe College

Why go digital? – the EE version

Now let's get back to Shannon's information theory


THE MATHEMATICAL THEORY OF COMMUNICATION

by Claude E. Shannon and Warren Weaver

THE UNIVERSITY OF ILLINOIS PRESS · URBANA · 1963

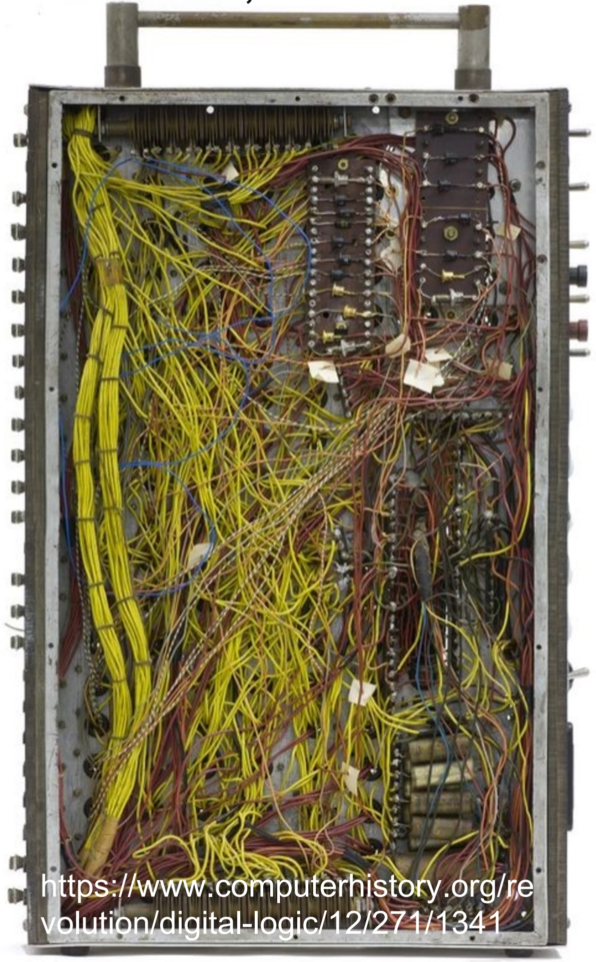
It doubtless seems queer, when one first meets it, that information is defined as the *logarithm* of the number of choices. But in the unfolding of the theory, it becomes more and more obvious that logarithmic measures are in fact the natural ones.

The system at Shannon’s time

Guardian Electric Series 200 Relay

Available on ebay by seller

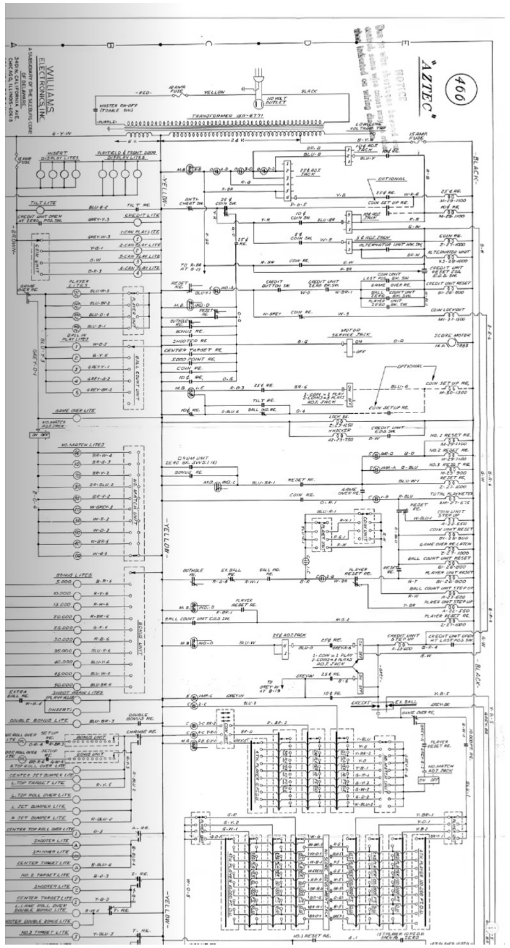
Simon 1 relay logic machine, 1950



https://www.computerhistory.org/revolution/digital-logic/12/271/1341

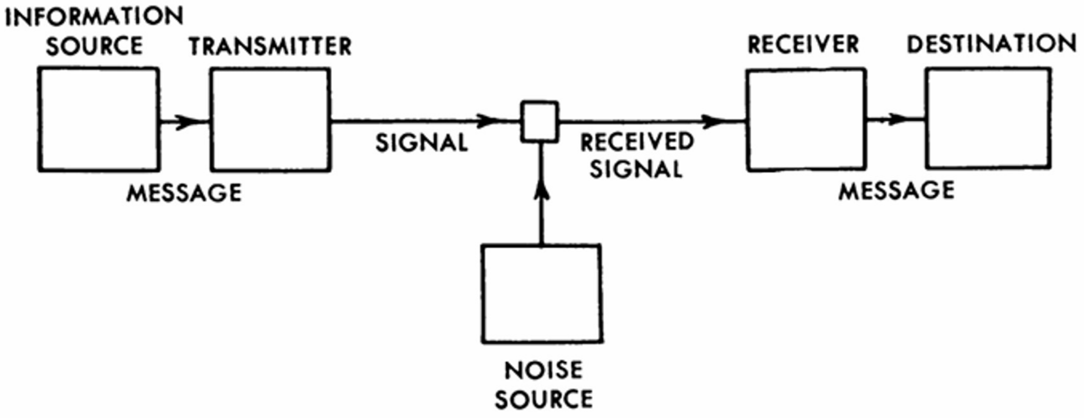
Aztec pinball machine, 1970s

https://www.youtube.com/watch?v=ue-1JoJQaEg&t=2423s



The information theory (aka the communication theory)

Some Recent Contributions: Weaver 7



To be sure, this word information in communication theory relates not so much to what you *do* say, as to what you *could* say.

The significant aspect is that the actual message is one *selected from a set* of possible messages.

The most basic definition of information, should still meet the need of the most basic form of communication: the switch on/off of relays!

Shannon's communication is NOT communication with any specific languages, it's more like sending 1s and 0s through long distance.

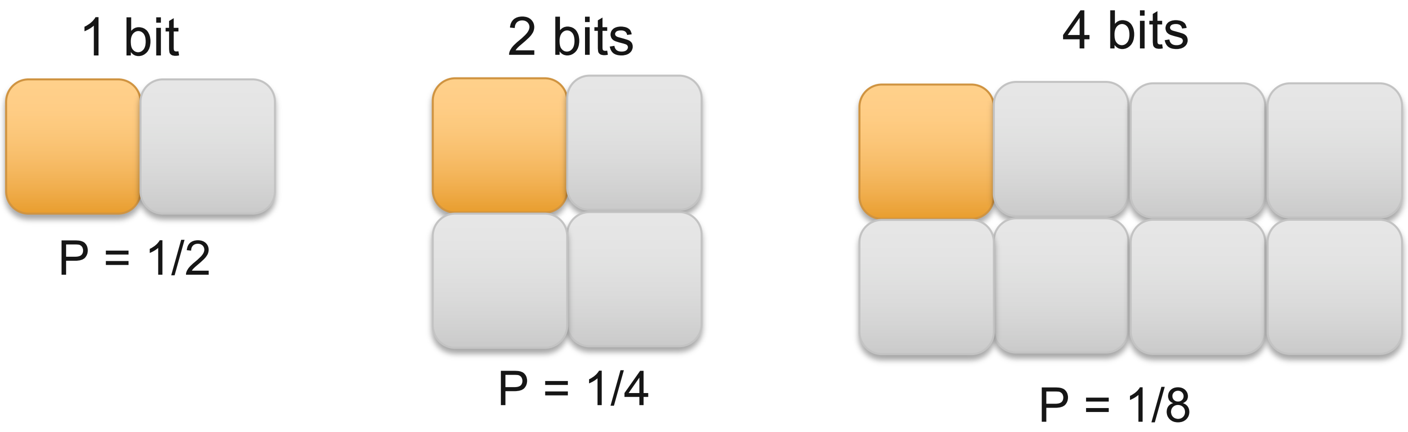
COE COLLEGE

The information theory (aka the communication theory)

Shannon's information unit: 1 bit = on/off of one relay (where the possibility of on/off is p=½)

And such information increase in the **power of 2**:

1 on/off relay (2^1): 0 1
2 on/off relays(2^2): 00 01 10 11
3 on/off relays(2^3): 000 001 010 011 100 101 110 111



To Shannon, the opposite of communication is:
RNG (random number generation).

The information theory (aka the communication theory)

Number of switches

$$\left(\frac{1}{2}\right)^{I} = p$$

The "odds", the probability

The "information"

$$\mathrm{I} = \log_2\left(\frac{1}{p}\right)$$

"Bit" (when use base 2) becomes the information's unit of measure.

Shannon's entropy (H)

The goal: $$H = -\sum_{i=1}^{n} p_i \log_2 p_i$$ - to quantify how accurately can the symbols of communication be transmitted.

The form is very like something from thermo dynamics

Specifically, entropy is a logarithmic measure for the system with a number of states, each with a probability p_i of being occupied (usually given by the Boltzmann distribution):

$$S = -k_B \sum_i p_i \ln p_i$$

where k_B is the Boltzmann constant and the summation is performed over all possible microstates of the system.[27]

https://en.wikipedia.org/wiki/Entropy

Shannon’s entropy (H)

- To quantify the degree of uncertainty, “surprise”, randomness or.. entropy of the communication system.

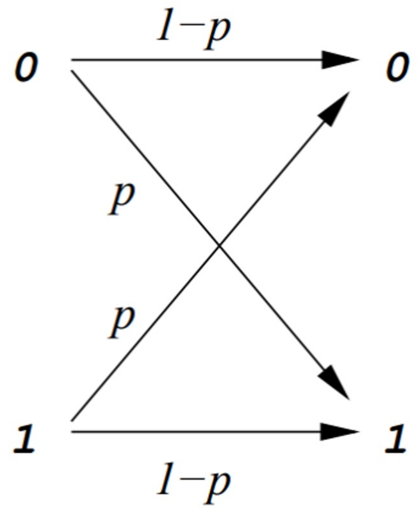
Shannon's main problem was what to label his new legislation. I considered calling it "information," but that term is overused, so I chose "uncertainty." John von Neumann had a better idea after we spoke. Von Neumann said, "Calculate entropy for two reasons." First, your uncertainty function already has a name in statistical mechanics. No one understands entropy, and thus you will always win a discussion" [19].

DOI: 10.36785/jaes.131550

H in application: Shannon's channel capacity

For Binary Symmetric Channel (BSC) $C = 1 - H(p)$

BSC



- 1 bit input (0 or 1).
- The probability of error is p.
- The probability of error is the same for both 0 and 1.

H in application: Shannon's channel capacity

For Binary Symmetric Channel (BSC) ***C* = 1 – *H*(p)**

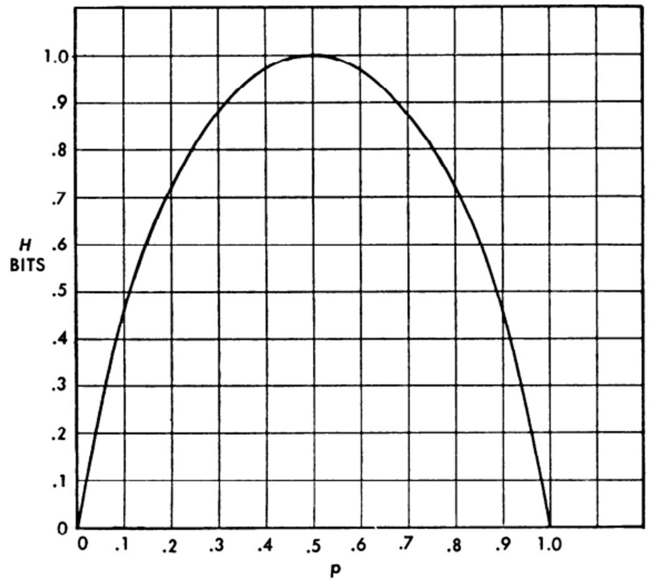
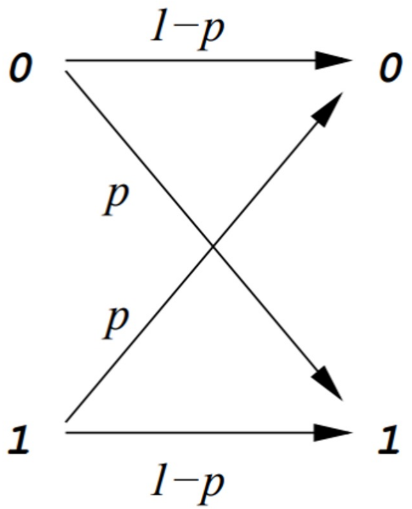


Fig. 7. — Entropy in the case of two possibilities with probabilities p and (1—p).

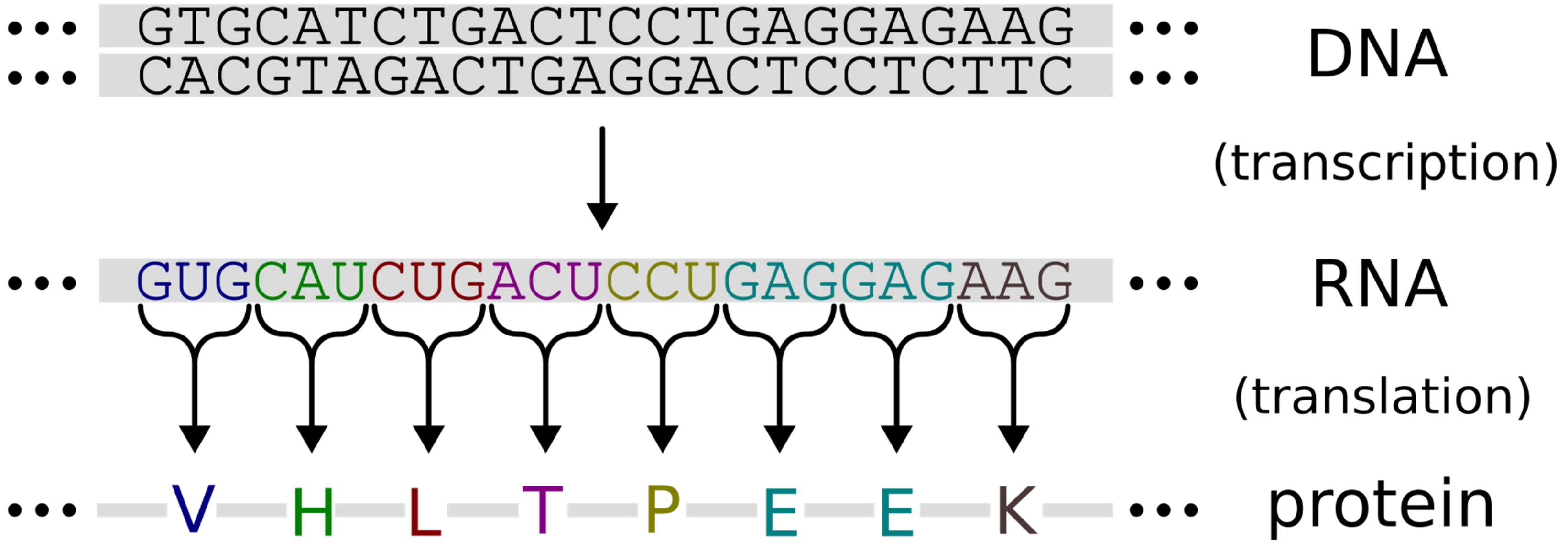
BSC



When p = 1/1000
H(p) = 0.011 bits, C=?

When p = 1/2
H(p) = 1 bit, C=?

H in application: genetics



https://en.wikipedia.org/wiki/Genetics

So our genes runs on a 2 bit system?

$$p_{[A,C,G,T]} = [0.29\ 0.21\ 0.21\ 0.29]$$

How many bits for the function (protein)?

Similarly, the entropy of the codon distribution in humans is $H(p_{codons}) = 5.7936 < 3 \times H(p_{[A,C,G,T]}) = 5.9445$ using the frequencies reported in Nei and Kumar (2000).

https://doi.org/10.1016/j.jtbi.2017.01.046

H in application: genetics

RESEARCH ARTICLE | GENETICS

Defining diversity, specialization, and gene specificity in transcriptomes through information theory

Octavio Martínez and M. Humberto Reyes-Valdés Authors Info & Affiliations

July 15, 2008 | 105 (28) 9709-9714 | https://doi.org/10.1073/pnas.0803479105

Table 1.

Examples of genes with distinct value of specificity (S_i)

OPEN IN VIEWER

Completely specific (S_i = 5), highly expressed genes (tissue)

S_i	HUGO	Description (tissue where expressed)
5.00	*LIPF*	Lipase (stomach)
5.00	*ELA3B*	Enastase 3B (pancreas)
5.00	*RHO*	Rhodopsin; opsin 2, rod pigment (retina)
5.00	*AZU1*	Azurocidin 1 (bone marrow)
5.00	*MYL2*	Myosin, light polypeptide 2 (heart)

Consider the division of an organism in tissues; the transcriptomes of each tissue can then be simply described as the set of relative frequencies, p_{ij}, for the *i*th gene (i = 1, 2, ..., g) in the *j*th tissue (j = 1, 2, ..., t). Then the diversity of the transcriptome of each tissue can be quantified by an adaptation of Shannon's entropy formula,

$$H_j = -\sum_{i=1}^{g} p_{ij}\log_2(p_{ij}). \qquad [1]$$

H_j will vary from zero when only one gene is transcribed up to $\log_2(g)$, where all g genes are transcribed at the same frequency: $1/g$. If we consider the average frequency of the *i*th gene among tissues, say,

$$p_i = \frac{1}{t}\sum_{j=1}^{t} p_{ij}, \qquad [2]$$

and define gene specificity as the information that its expression provides about the identity of the source tissue as

$$S_i = \frac{1}{t}\left(\sum_{j=1}^{t} \frac{p_{ij}}{p_i}\log_2\frac{p_{ij}}{p_i}\right). \qquad [3]$$

H in application: cross-entropy

The true H of a system:

$$H(P) = \sum p \log_2 \frac{1}{p}$$

The probability

The “information”

What if we don’t have data of the true system?

COE COLLEGE

H in application: cross-entropy

The true H of a system:

$$H(P) = \sum p \log_2 \frac{1}{p}$$

The probability

The "information"

What if we don't have data of the true system?

$$H(Q) = \sum q \log_2 \frac{1}{q}$$

We can make educated guess.

COE COLLEGE

H in application: cross-entropy

The true H of a system:

$$H(P) = \sum p \log_2 \frac{1}{p}$$

The probability

The "information"

What if we don't have data of the true system?

$$H(Q) = \sum q \log_2 \frac{1}{q}$$

We can make educated guess.

Hence, the cross-entropy:

$$H(P,Q) = \sum p \log_2 \frac{1}{q}$$

$$H(Q,P) = \sum q \log_2 \frac{1}{p}$$

COE COLLEGE

H in application: cross-entropy

The true H of a system:

$$H(P) = \sum p \log_2 \frac{1}{p}$$

The probability

The "information"

What if we don't have data of the true system?

$$H(Q) = \sum q \log_2 \frac{1}{q}$$

We can make educated guess.

Hence, the cross-entropy:

$$H(P,Q) = \sum p \log_2 \frac{1}{q}$$

$$H(Q,P) = \sum q \log_2 \frac{1}{p}$$

And a handy tool called K-L divergence*:

$$D_{KL}(P,Q) = H(P,Q) - H(P) = \sum p \log_2 \frac{p}{q}$$

COE COLLEGE

*: Kullback–Leibler divergence

H and C, more explanations

- The following slides are backups to help better understand the information theory, and Shannon’s original deduction.
-

Shannon’s entropy (H)

What properties (H) have?

- *H* is dealing with probabilities and be continuous. $H(p_1, p_2, \dots p_n)$.
- When no freedom of choice (certain outcome), $H = 0$.
- When $p_i \uparrow$, $H \downarrow$.
- *H* is largest when all outcomes are equally possible.
- The information from **independent sets** is linearly additive $H(p_1)+H(p_2)=H(p_1 * p_2)$.

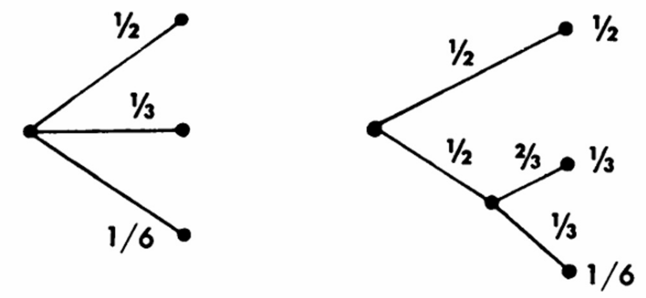


Fig. 6. — Decomposition of a choice from three possibilities.

$$H\left(\tfrac{1}{2}, \tfrac{1}{3}, \tfrac{1}{6}\right) = H\left(\tfrac{1}{2}, \tfrac{1}{2}\right) + \tfrac{1}{2} H\left(\tfrac{2}{3}, \tfrac{1}{3}\right).$$

COE COLLEGE

Shannon's entropy (H)

Theorem 2: The only H satisfying the three above assumptions is of the form:

$$H = -K\sum_{i=1}^{n} p_i \log p_i$$

where K is a positive constant. When $K = 1$,

$$H = -[p_1 \log p_1 + p_2 \log p_2 + \cdot \cdot \cdot + p_n \log p_n],$$

or

$$H = -\Sigma\, p_i \log p_i.$$

Shannon's entropy (H)

Appendix 2. **Derivation of $H = -\Sigma p_i \log p_i$**

Let $H\left(\frac{1}{n}, \frac{1}{n}, \cdots, \frac{1}{n}\right) = A(n)$. From condition (3) we can decompose a choice from s^m equally likely possibilities into a series of m choices each from s equally likely possibilities and obtain

$$A(s^m) = m\,A(s).$$

Similarly

$$A(t^n) = n\,A(t).$$

We can choose n arbitrarily large and find an m to satisfy

$$s^m \le t^n < s^{(m+1)}.$$

Thus, taking logarithms and dividing by $n \log s$,

$$\frac{m}{n} \le \frac{\log t}{\log s} \le \frac{m}{n} + \frac{1}{n} \quad \text{or} \quad \left| \frac{m}{n} - \frac{\log t}{\log s} \right| < \epsilon$$

where ϵ is arbitrarily small. Now from the monotonic property of $A(n)$,

$$A(s^m) \le A(t^n) \le A(s^{m+1})$$
$$m\,A(s) \le nA(t) \le (m+1)\,A(s).$$

Hence, dividing by $nA(s)$,

$$\frac{m}{n} \le \frac{A(t)}{A(s)} \le \frac{m}{n} + \frac{1}{n} \quad \text{or} \quad \left| \frac{m}{n} - \frac{A(t)}{A(s)} \right| < \epsilon$$

$$\left| \frac{A(t)}{A(s)} - \frac{\log t}{\log s} \right| \le 2\epsilon \qquad A(t) = -K \log t$$

where K must be positive to satisfy (2).

Now suppose we have a choice from n possibilities with commeasurable probabilities $p_i = \frac{n_i}{\Sigma n_i}$ where the n_i are integers. We can break down a choice from Σn_i possibilities into a choice from n possibilities with probabilities $p_1, \cdots, p_n$ and then, if the ith was chosen, a choice from n_i with equal probabilities. Using condition 3 again, we equate the total choice from Σn_i as computed by two methods

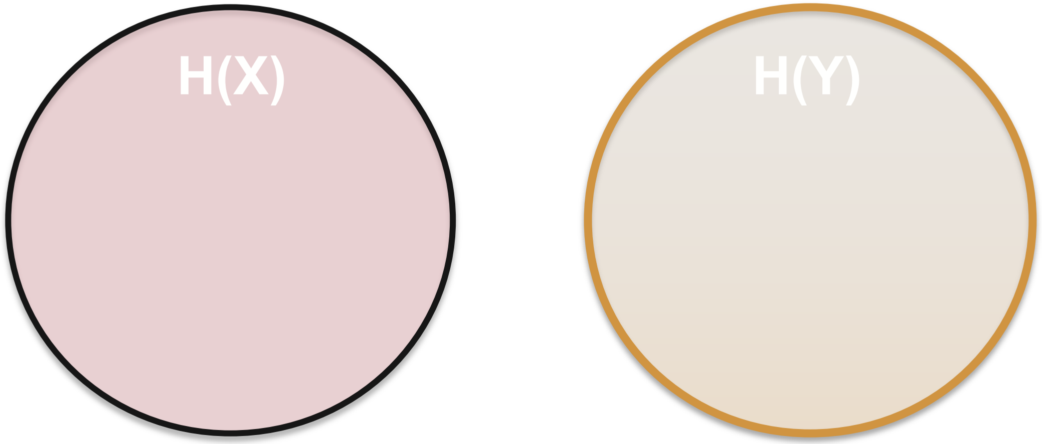
$$K \log \Sigma n_i = H(p_1, \cdots, p_n) + K\Sigma\, p_i \log n_i.$$

Hence

$$\begin{aligned} H &= K[\Sigma p_i \log \Sigma n_i - \Sigma p_i \log n_i] \\ &= -K\Sigma p_i \log \frac{n_i}{\Sigma n_i} = -K\Sigma p_i \log p_i. \end{aligned}$$

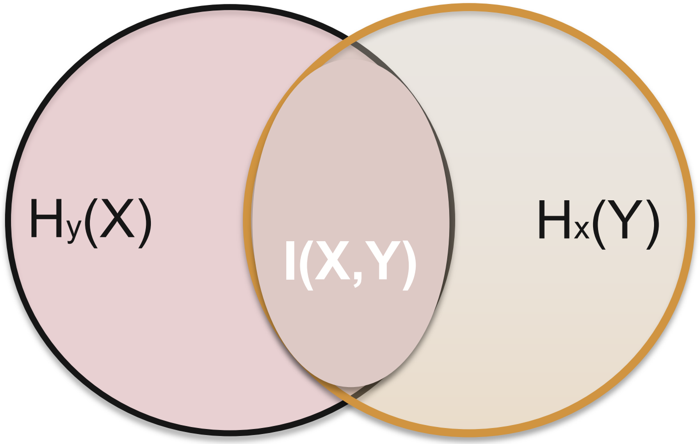
COE COLLEGE

Hy(X) starts with conditional entropy, basically



H(X,Y) = H(X) + H(Y)

$H_y(X)$ starts with conditional entropy, basically

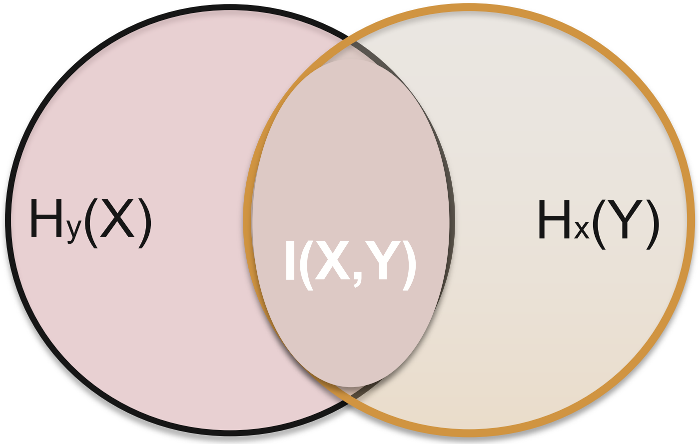


$$H(X,Y) = H(X) + H_x(Y) = H_y(X) + H(Y)$$
$$H(X,Y) = H_x(Y) + H_y(X) + I(X,Y)$$

?

COE COLLEGE

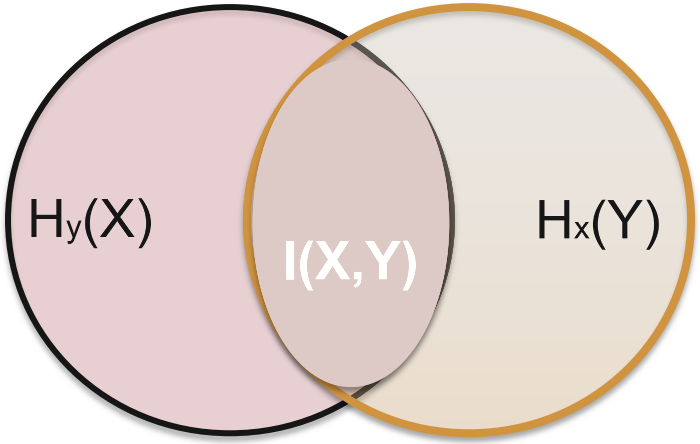
$H_y(X)$ starts with conditional entropy, basically



$$H(X,Y) = H(X) + H_x(Y) = H_y(X) + H(Y)$$
$$H(X,Y) = H_x(Y) + H_y(X) + I(X,Y)$$

Mutual info.

$H_y(X)$ starts with conditional entropy, basically



$$H(X,Y) = H(X) + H_x(Y) = H_y(X) + H(Y)$$
$$H(X,Y) = H_x(Y) + H_y(X) + I(X,Y)$$

Mutual info.

COE COLLEGE

H in application: Shannon's channel capacity

"the worst pipe misalignment"

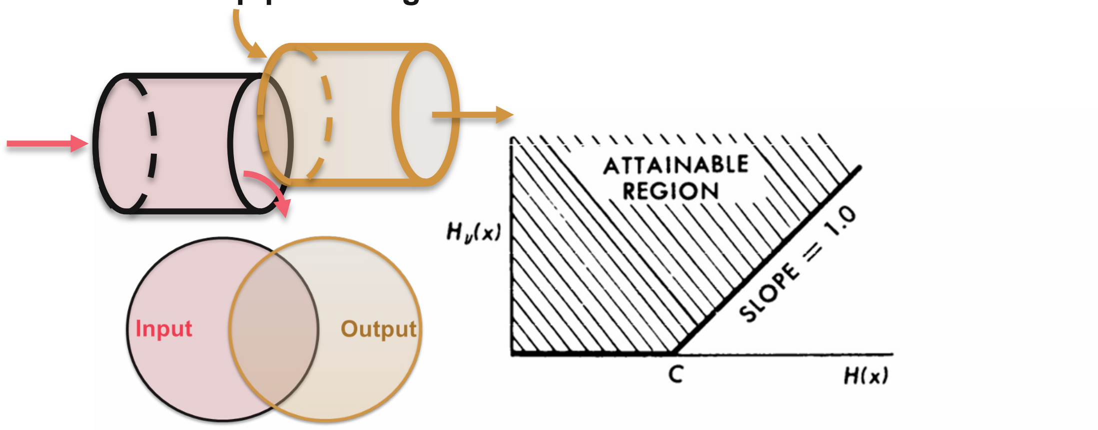
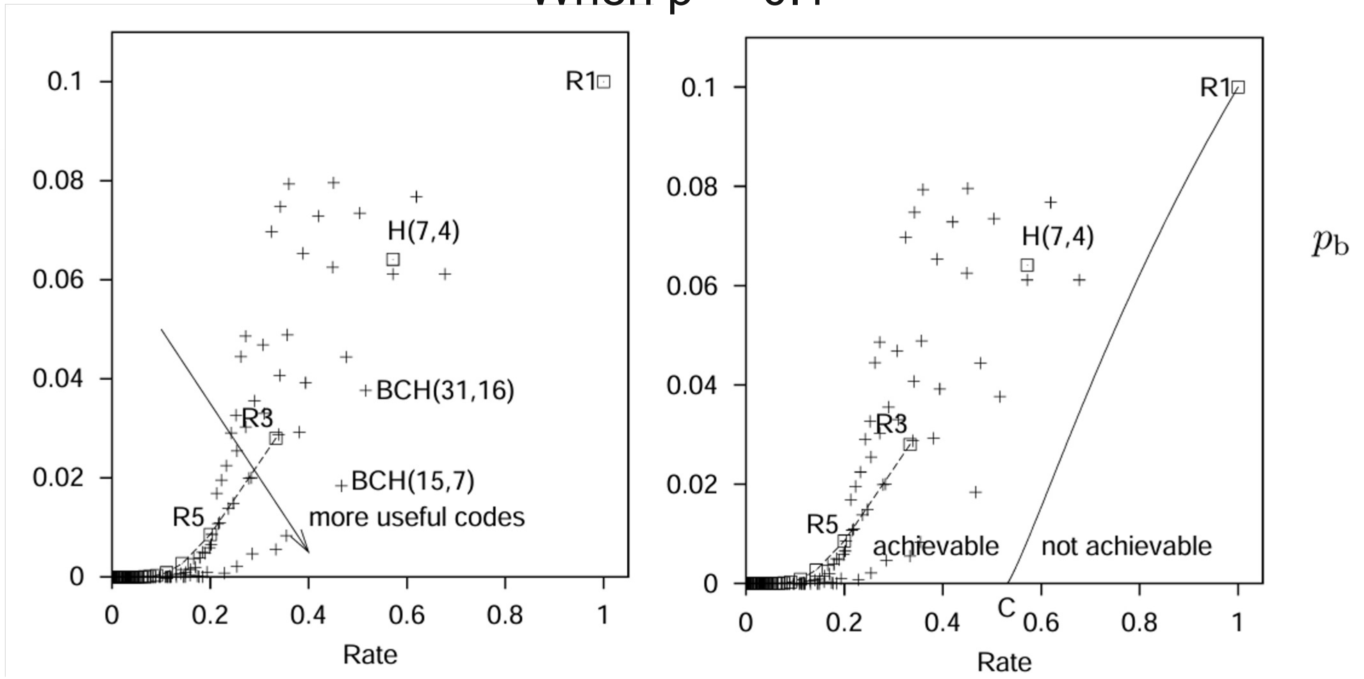


Fig. 9. — The equivocation possible for a given input entropy to a channel.

$$C = \text{Max}\ (H(X) - H_y(X)) = \text{Max}\ I(X,Y)$$

When p = 0.1



MacKay, David JC. *Information theory, inference and learning algorithms*. Cambridge university press, 2003.

But how? Shannon didn't give the answer.

Additional read: https://www.quantamagazine.org/how-claude-shannons-information-theory-invented-the-future-20201222/

H in application: genetics

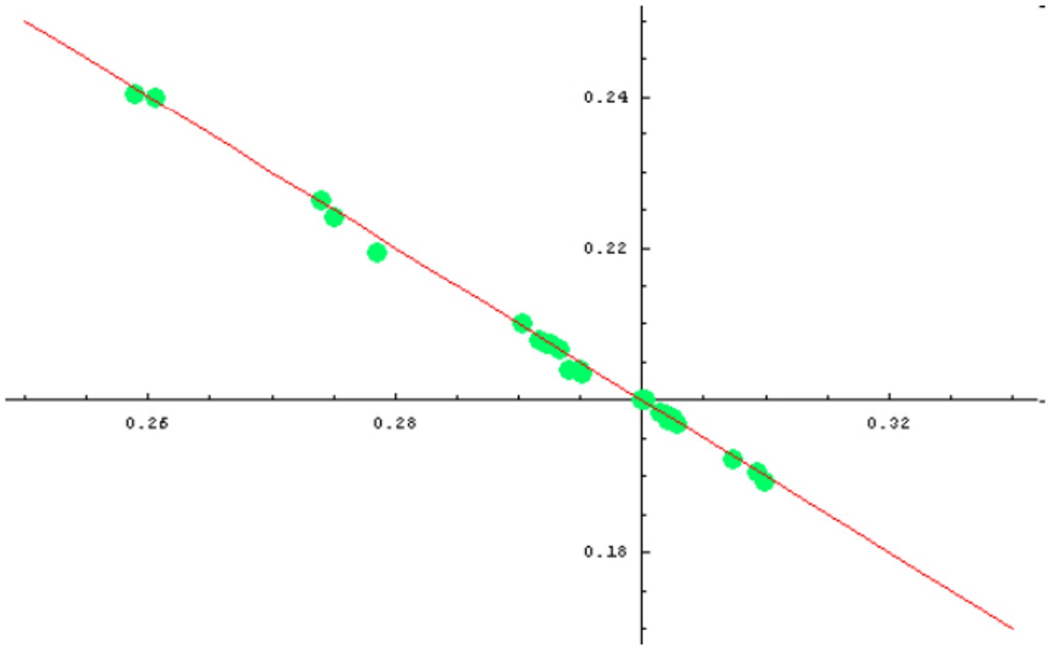


Figure 1: In red the line $P_A + P_C = \frac{1}{2}$, and in green the observed points (P_A, P_C) for each human chromosome

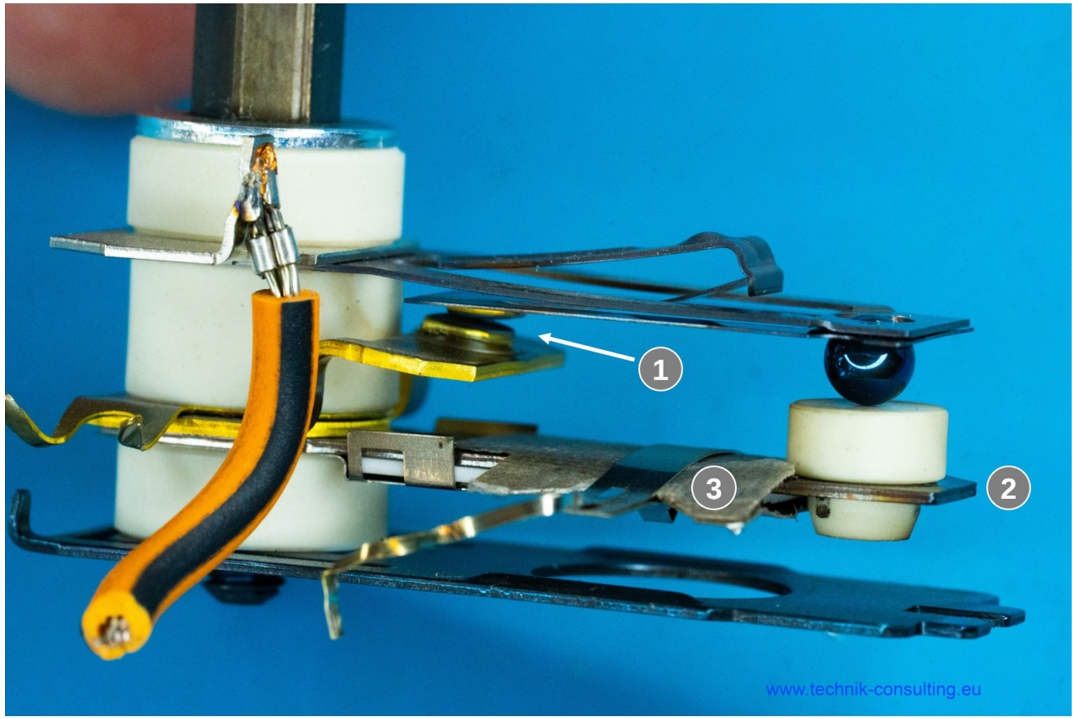
Chromosome	P_A	P_C	P_G	P_T	k
Chrom 1	0.2916	0.2080	0.2080	0.2922	3
Chrom 2	0.3000	0.2003	0.2005	0.2997	2
Chrom 3	0.3019	0.1980	0.1980	0.3020	2
Chrom 4	0.3093	0.1905	0.1906	0.3094	1
Chrom 5	0.3020	0.1974	0.1975	0.3011	2
Chrom 6	0.3024	0.1975	0.1976	0.3023	2
Chrom 7	0.2950	0.2040	0.2040	0.2951	3
Chrom 8	0.3002	0.2001	0.2000	0.2999	2
Chrom 9	0.2933	0.2067	0.2067	0.2931	3
Chrom 10	0.2922	0.2074	0.2074	0.2928	3
Chrom 11	0.2925	0.2072	0.2075	0.2926	3
Chrom 12	0.2950	0.2040	0.2033	0.2956	3
Chrom 13	0.3072	0.1922	0.1922	0.3080	1
Chrom 14	0.2951	0.2034	0.2039	0.2974	3
Chrom 15	0.2903	0.2101	0.2099	0.2895	3
Chrom 16	0.2750	0.2040	0.2040	0.2750	4
Chrom 17	0.2740	0.2261	0.2258	0.2713	5
Chrom 18	0.3014	0.1982	0.1985	0.3017	2
Chrom 19	0.2588	0.2403	0.2409	0.2598	7
Chrom 20	0.2785	0.2194	0.2202	0.2817	5
Chrom 21	0.2940	0.2040	0.2039	0.2952	3
Chrom 22	0.2605	0.2398	0.2397	0.2598	6
Chrom X	0.3027	0.1968	0.1967	0.3033	2
Chrom Y	0.3098	0.1893	0.1889	0.3118	1

Table 2: Nucleotide Frequencies for all human chromosomes

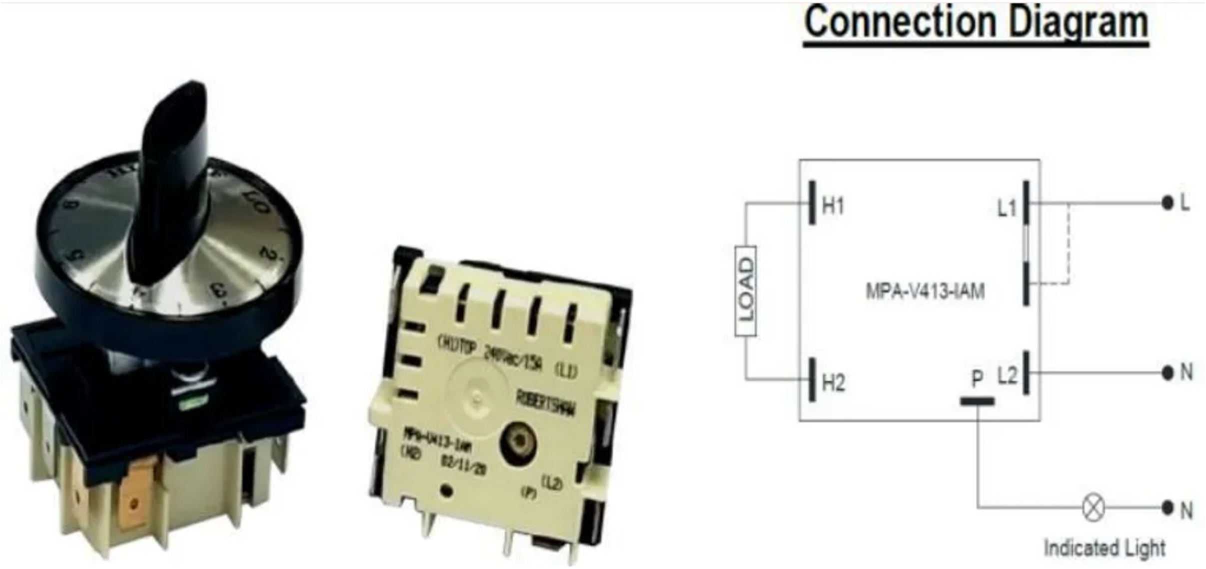
Yamagishi, Michel E. Beleza, and Alex Itiro Shimabukuro. "Nucleotide frequencies in human genome and Fibonacci numbers." *Bulletin of mathematical biology* 70.3 (2008): 643-

Why NOT go digital? – the EE version

Or Simmerstat, a >100 year old tech but still going strong



https://en.wikipedia.org/wiki/Infinite_switch



Robertshaw Simmerstat (MPA-V413-IAM)

https://gii.easy.co/products/robertshaw-simmerstat-mpa-v413-iam-