Shannon’s Formula
\[ W \cdot \log(1 + \text{SNR}) : \]
A Historical Perspective

on the occasion of Shannon’s Centenary

Oct. 26th, 2016

Olivier Rioul
<olivier.rioul@telecom-paristech.fr>
Outline

Who is Claude Shannon?
Shannon’s Seminal Paper
Shannon’s Main Contributions
Shannon’s Capacity Formula

Hartley’s rule $C' = \log_2 \left( 1 + \frac{A}{\Delta} \right)$ is not Hartley’s

Many authors independently derived $C = \frac{1}{2} \log_2 \left( 1 + \frac{P}{N} \right)$ in 1948.

Hartley’s rule is exact: $C' = C$ (a coincidence?)

$C'$ is the capacity of the “uniform” channel

Shannon’s Conclusion
April 30, 1916  Claude Elwood Shannon was born in Petoskey, Michigan, USA
Claude Shannon (1916–2001)
100th birthday 2016

April 30, 1916  Claude Elwood Shannon was born in Petoskey, Michigan, USA

April 30, 2016  centennial day celebrated by Google:
April 30, 1916  Claude Elwood Shannon was born in Petoskey, Michigan, USA
April 30, 2016  centennial day celebrated by Google:

here Shannon is juggling with bits (1,0,0) in his communication scheme

“father of the information age”
Do you Know Claude Shannon?

Shannon’s Formula $W \log(1 + SNR)$: A Historical Perspective
Do you Know Claude Shannon?
Do you Know Claude Shannon?
Do you Know Claude Shannon?

“the most important man... you've never heard of”
Do you Know Claude Shannon?

Shannon's Formula $W \log(1 + SNR)$: A Historical Perspective
Do you Know Claude Shannon?

Shannon’s Formula $W \log(1 + SNR)$: A Historical Perspective
Do you Know Claude Shannon?

Shannon’s Formula $W \log(1 + SNR)$: A Historical Perspective
Do you Know Claude Shannon?

“the most important man... you've never heard of”
Do you Know Claude Shannon?

Shannon’s Formula $W \log(1 + \text{SNR})$: A Historical Perspective
Do you Know Claude Shannon?

Shannon’s Formula $W \log(1 + \text{SNR})$: A Historical Perspective
Do you Know Claude Shannon?

Shannon’s Formula $W \log(1 + SNR)$: A Historical Perspective
Do you Know Claude Shannon?

Shannon’s Formula $W \log(1 + SNR)$: A Historical Perspective
Do you Know Claude Shannon?
Do you Know Claude Shannon?

Shannon’s Formula \( W \log(1 + \text{SNR}) \): A Historical Perspective
Do you Know Claude Shannon?

Shannon’s Formula $W \log(1 + SNR)$: A Historical Perspective
Do you Know Claude Shannon?

"the most important man... you've never heard of"
Do you Know Claude Shannon?
Do you Know Claude Shannon?

Shannon’s Formula $W \log(1 + \text{SNR})$: A Historical Perspective
Do you Know Claude Shannon?

Shannon’s Formula $W \log(1 + SNR)$: A Historical Perspective
Do you Know Claude Shannon?
Do you Know Claude Shannon?

Shannon’s Formula $W \log(1 + \text{SNR})$: A Historical Perspective
Do you Know Claude Shannon?

Shannon’s Formula $W \log(1 + SNR)$: A Historical Perspective
Do you Know Claude Shannon?

Shannon’s Formula $W \log(1 + SNR)$: A Historical Perspective
Do you Know Claude Shannon?

Shannon’s Formula $W \log(1 + SNR)$: A Historical Perspective
Do you Know Claude Shannon?

“the most important man... you’ve never heard of”
Well-Known Scientific Heroes

Alan Turing (1912–1954)
Well-Known Scientific Heroes

Alan Turing (1912–1954)
Well-Known Scientific Heroes

John Nash (1928–2015)

Shannon’s Formula \( W \log(1 + SNR) \): A Historical Perspective
Well-Known Scientific Heroes

John Nash (1928–2015)
The Quiet and Modest Life of Shannon
Shannon with Juggling Props
Shannon is known for riding through the halls of Bell Labs on a unicycle while simultaneously juggling four balls.
Crazy Machines

Theseus (labyrinth mouse)
Crazy Machines

Shannon’s Formula $W \log(1 + \text{SNR})$: A Historical Perspective
Crazy Machines

calculator in Roman numerals
Crazy Machines

“Hex” switching game machine
Crazy Machines

Rubik’s cube solver
Crazy Machines

3-ball juggling machine
Wearable computer to predict roulette in casinos
(with Edward Thorp)
ultimate useless machine
“Serious” Work

At the same time, Shannon made decisive theoretical advances in ...

- logic & circuits
- cryptography
- artifical intelligence
- stock investment
- wearable computing

... and information theory!
“Serious” Work

At the same time, Shannon made decisive theoretical advances in ...

- logic & circuits
- cryptography
- artificial intelligence
- stock investment
- wearable computing
- ...and **information theory**!
Outline

Who is Claude Shannon?

Shannon’s Seminal Paper

Shannon’s Main Contributions

Shannon’s Capacity Formula

Hartley’s rule $C' = \log_2 \left(1 + \frac{A}{A}\right)$ is not Hartley’s

Many authors independently derived $C = \frac{1}{2} \log_2 \left(1 + \frac{P}{N}\right)$ in 1948.

Hartley’s rule is exact: $C' = C$ (a coincidence?)

$C'$ is the capacity of the “uniform” channel

Shannon’s Conclusion
The Mathematical Theory of Communication (BSTJ, 1948)

Shannon’s Formula $W \log(1 + SNR)$: A Historical Perspective
The Mathematical Theory of Communication (BSTJ, 1948)

One article (written 1940–48): A REVOLUTION !!!!!!
Shannon’s Theorems

Yes it’s Maths!!

1. Source Coding Theorem
   (Compression of Information)

2. Channel Coding Theorem
   (Transmission of Information)
Outline

Who is Claude Shannon?

Shannon’s Seminal Paper

**Shannon’s Main Contributions**

Shannon’s Capacity Formula

Hartley’s rule $C’ = \log_2 \left( 1 + \frac{A}{\Delta} \right)$ is not Hartley’s

Many authors independently derived $C = \frac{1}{2} \log_2 \left( 1 + \frac{P}{N} \right)$ in 1948.

Hartley’s rule is exact: $C’ = C$ (a coincidence?)

$C’$ is the capacity of the “uniform” channel

Shannon’s Conclusion
A tremendous impact!
Shannon’s Paradigm... in Communication

Example: Broadcast following crisis

A television station → B TV broadcasting equipment → C Circuitry, waves → D TV sets; → E viewing public

Signal: A storm!

Received Signal: A storm!

Noise Source:
Storm damages TV equipment; static from storm in reception
Shannon’s Paradigm... in Linguistics

A SPEECH EVENT

ADDRESSER
emotive function
(expression of attitude)

MESSAGE
poetic function
(focus on msg for its own sake)

CONTEXT
referential function
(locates msg in experience)

CONTACT
phatic function
(aspect of msg linking addr + addee)

CODE
metalingual function
(shared aspects of msg)

ADDRESSEE
conative function

Roman Jakobson’s 1960 model of communication
drawn by jjs
Shannon’s Paradigm... in Biology

Shannon’s Formula $W \log(1 + SNR)$: A Historical Perspective
Shannon’s Paradigm... in Psychology

Shannon’s Formula $W \log(1 + SNR)$: A Historical Perspective
Shannon’s Paradigm... in Social Sciences

Shannon’s Formula $W \log(1 + SNR)$: A Historical Perspective
Shannon’s Paradigm... in Human-Computer Interaction

Shannon’s Formula $W \log(1 + SNR)$: A Historical Perspective

Diagram:
- Source
  - User intention
- Encoder
  - Movement mapping
- Channel
- Decoder
  - Target recognition
- Destination
  - Target hit

Noise:
- Neural noise, tremor,...
The Bandwagon

CLAUDE E. SHANNON

INFORMATION theory has, in the last few years, become something of a scientific bandwagon. Starting as a technical tool for the communication engineer, it has received an extraordinary amount of publicity in the popular as well as the scientific press. In part, this has been due to connections with such fashionable fields as computing machines, cybernetics, and automation; and in part, to the novelty of its subject matter. As a consequence, it has perhaps been ballooned to an importance beyond its actual accomplishments. Our fellow scientific subjects are aimed in a very specific direction, a direction that is not necessarily relevant to such fields as psychology, economics, and other social sciences. Indeed, the hard core of information theory is, essentially, a branch of mathematics, a strictly deductive system. A thorough understanding of the mathematical foundation and its communication application is surely a prerequisite to other applications. 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
Shannon’s Viewpoint

“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; [...] 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 [...] unknown at the time of design.”

$X$: a message symbol modeled as a random variable

$p(x)$: the probability that $X = x$
Andreï Kolmogorov (1903–1987)

- founded modern probability theory in 1933
- a strong early supporter of information theory!

"Information theory must precede probability theory and not be based on it. [...] The concepts of information theory as applied to infinite sequences [...] can acquire a certain value in the investigation of the algorithmic side of mathematics as a whole."
A Logarithmic Measure

- 1 digit represents 10 numbers 0,1,2,3,4,5,6,7,8,9;
- 2 digits represents 100 numbers 00, 01, ... , 99;
- 3 digits represents 1000 numbers 000, ... , 999;
- \( \log_{10} M \) digits represents \( M \) possible outcomes

Ralph Hartley (1888–1970)

“[...] take as our practical measure of information the logarithm of the number of possible symbol sequences”

Transmission of Information, BSTJ, 1928
The Bit

- $\log_{10} M$ digits represents $M$ possible outcomes
- or...
- $\log_2 M$ bits represents $M$ possible outcomes

John Tukey (1915–2000) coined the term “bit” (contraction of “binary digit”) which was first used by Shannon in his 1948 paper
The Bit

- \( \log_{10} M \) digits represents \( M \) possible outcomes
- or...
- \( \log_2 M \) bits represents \( M \) possible outcomes

John Tukey (1915–2000) coined the term “bit” (contraction of “binary digit”) which was first used by Shannon in his 1948 paper:

any information can be represented by a sequence of 0’s and 1’s — the Digital Revolution!
The Unit of Information

bit (binary digit, unit of storage) ≠ bit (binary unit of information)

- less-likely messages are more informative than more-likely ones
- 1 bit is the information content of one equiprobable bit ($\frac{1}{2}, \frac{1}{2}$)

otherwise the information content is $< 1$ bit:
The Unit of Information

bit (binary digit, unit of storage) \neq \text{bit (binary unit of information)}

- less-likely messages are more informative than more-likely ones
- 1 bit is the information content of one equiprobable bit \( \left( \frac{1}{2}, \frac{1}{2} \right) \)

otherwise the information content is < 1 bit:

The official name (International standard ISO/IEC 80000-13) for the information unit:
The Unit of Information

- less-likely messages are more informative than more-likely ones
- 1 bit is the information content of one equiprobable bit ($\frac{1}{2}, \frac{1}{2}$)

otherwise the information content is $< 1$ bit:

The official name (International standard ISO/IEC 80000-13) for the information unit:

...the Shannon (symbol Sh)
Fundamental Limit of Performance

- Shannon does not really give *practical* solutions but solves a *theoretical* problem:

  - *No matter what you do,*
    
    (as long as you have a given amount of resources)
    
    you *cannot* go beyond than a certain bit rate limit to achieve reliable communication
Fundamental Limit of Performance

- before Shannon: communication technologies did not have a landmark
- the limit can be calculated: we know how far we are from it and you can be (in theory) arbitrarily close to the limit!
- the challenge becomes: how can we build practical solutions that are close to the limit?
Fundamental Limit of Performance

- before Shannon: communication technologies did not have a landmark
- the limit can be calculated: we know how far we are from it and you can be (in theory) arbitrarily close to the limit!
- the challenge becomes: how can we build practical solutions that are close to the limit?
Asymptotic Results

- to find the limits of performance, Shannon’s results are necessarily asymptotic
- a source is modeled as a sequence of random variables $X_1, X_2, \ldots, X_n$

where the dimension $n \to +\infty$.

- this allows to exploit dependences and obtain a geometric “gain” using the law of large numbers
- where limits are expressed as expectations $\mathbb{E}\{\cdot\}$
Asymptotic Results: Example

Consider the source $X_1, X_2, \ldots, X_n$ where each $X$ can take a finite number of possible values, independently of the other symbols.

The probability of message $\underline{x} = (x_1, x_2, \ldots, x_n)$ is the product of the individual probabilities:

$$p(\underline{x}) = p(x_1) \cdot p(x_2) \cdots \cdots \cdot p(x_n).$$

Re-arrange according to the value $x$ taken by each argument:

$$p(x) = \prod_x p(x)^{n(x)}$$

where $n(x) = \text{number of symbols equal to } x$. 
Asymptotic Results: Example (Cont’d)

By the *law of large numbers*, the empirical probability (frequency)

$$\frac{n(x)}{n} \rightarrow p(x) \quad \text{as } n \rightarrow +\infty$$

Therefore, a “typical” message $\underline{x} = (x_1, x_2, \ldots, x_n)$ satisfies

$$p(\underline{x}) = \prod_x p(x)^{n(x)} \approx \prod_x p(x)^{np(x)}$$
By the law of large numbers, the empirical probability (frequency)

\[ \frac{n(x)}{n} \to p(x) \quad \text{as} \quad n \to +\infty \]

Therefore, a “typical” message \( x = (x_1, x_2, \ldots, x_n) \) satisfies

\[ p(x) = \prod_x p(x)^{n(x)} \approx \prod_x p(x)^{np(x)} = 2^{-n \cdot H} \]

where

\[ H = \sum_x p(x) \log_2 \frac{1}{p(x)} = \mathbb{E} \left\{ \log_2 \frac{1}{p(X)} \right\} \]

is a positive quantity called entropy.
Shannon’s entropy

\[ H = \sum_{x} p(x) \log_2 \frac{1}{p(x)} \]

analogy with statistical mechanics

Ludwig Boltzmann (1844–1906)
Shannon’s entropy

\[ H = \sum_x p(x) \log_2 \frac{1}{p(x)} \]

- analogy with statistical mechanics

Ludwig Boltzmann (1844–1906)

suggested by

John von Neumann (1903–1957)
Shannon’s entropy

\[ H = \sum_x p(x) \log_2 \frac{1}{p(x)} \]

- analogy with statistical mechanics

Ludwig Boltzmann (1844–1906)

- suggested by

“You should call it entropy [...] no one really knows what entropy really is, so in a debate you will always have the advantage.”

John von Neumann (1903–1957)
Shannon’s entropy

\[ H = \sum_x p(x) \log_2 \frac{1}{p(x)} \]

- analogy with statistical mechanics

Ludwig Boltzmann (1844–1906)

suggested by

“You should call it entropy [...] no one really knows what entropy really is, so in a debate you will always have the advantage.”

John von Neumann (1903–1957)

studied in physics by

Léon Brillouin (1889–1969)
The Source Coding Theorem

Compression problem: noiseless channel, minimize bit rate

A “typical” sequence $x = (x_1, x_2, \ldots, x_n)$ satisfies $p(x) \approx 2^{-nH}$.

Summing over the $N$ typical sequences:

$$1 \approx N 2^{-nH}$$

since the probability of $x$ being typical is $\approx 1$. So $N \approx 2^{nH}$.

It is sufficient to encode only the $N$ typical sequences:

$$\frac{\log_2 N}{n} \approx H \quad \text{bits per symbol}$$
The Source Coding Theorem

Theorem (Shannon’s First Theorem)

*Only* $H$ bits per symbol suffice to reliably encode an information source.

The entropy $H$ is the bit rate lower bound for reliable compression.
The Source Coding Theorem

Theorem (Shannon’s First Theorem)

Only $H$ bits per symbol suffice to reliably encode an information source.

The entropy $H$ is the bit rate lower bound for reliable compression.

- This is an asymptotic theorem ($n \to +\infty$) not a practical solution.
- Variable length coding solution by Shannon and Robert Fano (1917–2016)
  - Optimal code (1952) by David Huffman (1925-1999)
  - Elias, Golomb, Lempel-Ziv, ...
Relative Entropy (or Divergence)

\[ D(p, q) = \sum_x p(x) \log_2 \frac{p(x)}{q(x)} \geq 0 \text{ with } D(p, q) = 0 \text{ iff } p \equiv q. \]
Relative Entropy (or Divergence)

\[ D(p, q) = \sum_x p(x) \log_2 \frac{p(x)}{q(x)} \geq 0 \text{ with } D(p, q) = 0 \text{ iff } p \equiv q. \]

Bounds of the type \(2^{-n \cdot D(p, q)}\) useful in statistics:
- large deviations theory
- asymptotic behavior in hypothesis testing

Chernoff information to classify empirical data

Herman Chernoff (1923–)

Fisher information for parameter estimation

Ronald Fisher (1890–1962)
Shannon’s Mutual Information

Shannon’s entropy of a random variable $X$:

$$H(X) = \sum_x p(x) \log_2 \frac{1}{p(x)} = \mathbb{E} \left\{ \log_2 \frac{1}{p(X)} \right\}$$

Shannon’s (mutual) information between two random variables $X, Y$:

$$I(X; Y) = \sum_{x,y} p(x, y) \log_2 \frac{p(x, y)}{p(x)p(y)} = \mathbb{E} \left\{ \log_2 \frac{p(X, Y)}{p(X)p(Y)} \right\}$$

This exactly $D(p, q)$ where:

- $p(x, y)$ is the (true) joint distribution;
- $q(x, y) = p(x)p(y)$ is what would have been in the case of independence.

Therefore $I(X; Y) \geq 0$ with $I(X; Y) = 0$ iff $X$ and $Y$ are independent.
Shannon’s Mutual Information

Shannon writes

\[ I(X; Y) = \mathbb{E} \left\{ \log_2 \frac{p(X|Y)}{p(X)} \right\} = H(X) - H(X|Y) \]

where \( H(X|Y) \) is the conditional entropy of \( X \) given \( Y \).

- \( H(X|Y) \leq H(X) \): knowledge decreases uncertainty by a quantity equal to the information gain \( I(X; Y) \).
- Intuitive and rigorous!
The Channel Coding Theorem

Transmission problem: noisy channel, maximize bit rate for reliable communication

It is sufficient to decode only sequences $x$ jointly typical with $y$. 
The Channel Coding Theorem (Cont’d)

*But* another code is also jointly typical with $y$ with probability bounded by

$$2^{-n \cdot I(X;Y)}.$$

Summing over the $N$ code sequences, the total probability of decoding error is bounded by

$$N \cdot 2^{-n \cdot I(X;Y)}$$

which tends to zero only if the bit rate

$$\frac{\log_2 N}{n} < I(X;Y)$$

**Definition (Channel Capacity)**

$$C = \max_{p(x)} I(X;Y)$$
The Channel Coding Theorem (Cont’d)

If the bit rate is \(< C\), then the error probability, \textit{averaged over all possible codes}, can be made as small as desired.

Therefore \textit{there exists at least one code} with arbitrarily small probability of error.

Theorem (Shannon’s Second Theorem)

*Information can be transmitted reliably provided that the bit rate does not exceed the channel capacity C.*

The capacity \(C\) is the bit rate upper bound for reliable transmission.
If the bit rate is $< C$, then the error probability, averaged over all possible codes, can be made as small as desired.

Therefore there exists at least one code with arbitrarily small probability of error.

**Theorem (Shannon’s Second Theorem)**

*Information can be transmitted reliably provided that the bit rate does not exceed the channel capacity $C$.***

The capacity $C$ is the bit rate upper bound for reliable transmission.

**Revolutionary!** Transmission noise does not affect quality—it only impacts the bit rate. This is the theorem that led to the digital revolution!
Shannon’s Result is Paradoxical!

- Shannon theorems show that good codes exist, but give no clue on how to build them in practice
- but choosing a code at random would be almost optimal!
- however random coding is impractical ($n$ is large)...
- only 50 years later were found *turbo-codes* (by Claude Berrou & Alain Glavieux) that imitate random coding to approach capacity
Outline

Who is Claude Shannon?
Shannon’s Seminal Paper
Shannon’s Main Contributions
Shannon’s Capacity Formula

Hartley’s rule $C' = \log_2 \left( 1 + \frac{A}{A} \right)$ is not Hartley’s

Many authors independently derived $C = \frac{1}{2} \log_2 \left( 1 + \frac{P}{N} \right)$ in 1948.

Hartley’s rule is exact: $C' = C$ (a coincidence?)

$C'$ is the capacity of the “uniform” channel

Shannon’s Conclusion
Claude Shannon

Claude Shannon

anagram
A sound channel

Shannon’s Formula $W \log(1 + \text{SNR})$: A Historical Perspective
Shannon’s formula:

\[ C = W \log_2 \left( 1 + \frac{P}{N} \right) \text{ bits/second} \]
Shannon’s formula:

\[ C = \frac{1}{2} \log_2 \left( 1 + \frac{P}{N} \right) \ \text{bits/symbol} \]
Additive White Gaussian Noise Channel

A very common model: \( Y = X + Z \) where \( Z \) is Gaussian \( \mathcal{N}(0, \sigma^2) \).

Shannon finds the exact expression:

\[
C = W \cdot \log_2 \left( 1 + \frac{P}{N} \right) \text{ bit/s}
\]

where \( W \) is the bandwidth and \( P/N \) is the signal-to-noise ratio.

- a “concrete” finding of information theory – the most celebrated formula of Shannon!
- to derive this formula, Shannon popularized the Whittaker-Nyquist sampling theorem — “Shannon’s Theorem”!
Claude Shannon

Shannon’s formula:

\[ C = W \log_2 \left( \frac{P + N}{N} \right) \]


Shannon’s formula:

\[ C = W \log_2 \left( \frac{P + N}{N} \right) \]


Note on the Theoretical Efficiency of Information Reception with PPM*

For small \( P/N \) ratios, the \textbf{now classical} expression for the information reception capacity of a channel

\[ C = W \log_2 (1 + P/N) \]

can be written, substituting \( kT W \) for \( N \),

\[ CT_0 = WT \frac{P}{N} \log_2 e = \frac{PT_0}{kT} \log_2 e = \frac{E}{kT} \log_2 e \]

* Received by the Institute, February 23, 1949.
Ralph Hartley

20 years before... in the same journal...

Hartley’s rule:

\[ C' = \log_2 \left( 1 + \frac{A}{\Delta} \right) \text{ bits/symbol} \]

Ralph Hartley

Hartley’s rule:

\[ C' = \log_2 \left( 1 + \frac{A}{\Delta} \right) \]

(Wozencraft-Jacobs textbook, 1965)

Figure 1.1 Distinguishable receiver amplitudes. Hartley considered received pulse amplitudes to be distinguishable only if they lie in different zones of width \(2\Delta\). Thus pulses \(a\) and \(c\) are distinguishable but \(a\) and \(b\) are not. For the case shown, \(A/\Delta = 4\) and there are five distinguishable zones.
Hartley’s rule:

\[ C' = \log_2 \left( 1 + \frac{A}{\Delta} \right) \]

- amplitude “SNR” \( A/\Delta \) (factor 1/2 is missing)
- no coding involved (except quantization)
- zero error

Hartley’s formulation exhibits a simple but somewhat inexact inter-relation among the time interval \( T \), the channel bandwidth \( W \), the maximum signal magnitude \( A \), the receiver accuracy \( \Delta \), and the allowable number \( M \) of message alternatives. Communication theory is intimately concerned with the determination of more precise interrelations of this sort.

(Wozencraft-Jacobs textbook, 1965)
Outline

Hartley’s $C' = \log_2\left(1 + \frac{A}{\Delta}\right)$ came 20 years before Shannon
Hartley’s $C' = \log_2 \left(1 + \frac{A}{\Delta}\right)$ came 20 years before Shannon

Shannon’s $C = \frac{1}{2} \log_2 \left(1 + \frac{P}{N}\right)$ came unexpected in 1948
Outline

Hartley’s $C' = \log_2 \left(1 + \frac{A}{\Delta}\right)$ came 20 years before Shannon

Shannon’s $C = \frac{1}{2} \log_2 \left(1 + \frac{P}{N}\right)$ came unexpected in 1948

Hartley’s rule is inexact: $C' \neq C$
Outline

Hartley’s $C' = \log_2\left(1 + \frac{A}{\Delta}\right)$ came 20 years before Shannon

Shannon’s $C = \frac{1}{2} \log_2\left(1 + \frac{P}{N}\right)$ came unexpected in 1948

Hartley’s rule is inexact: $C' \neq C$

Besides, $C'$ is not the capacity of a noisy channel
Wrong!
Outline

Hartley’s rule \( C' = \log_2 \left( 1 + \frac{A}{\Delta} \right) \) is not Hartley’s
Hartley’s rule $C' = \log_2 \left(1 + \frac{A}{\Delta}\right)$ is not Hartley’s.

Shannon’s $C = \frac{1}{2} \log_2 \left(1 + \frac{P}{N}\right)$ was derived by 7 other authors in 1948!
Outline

Hartley’s rule $C' = \log_2 \left(1 + \frac{A}{\Delta}\right)$ is not Hartley’s

Shannon’s $C = \frac{1}{2} \log_2 \left(1 + \frac{P}{N}\right)$ was derived by 7 other authors in 1948!

Hartley’s rule is exact: $C' = C$ (a coincidence?)
Hartley’s rule $C' = \log_2 \left(1 + \frac{A}{\Delta}\right)$ is not Hartley’s

Shannon’s $C = \frac{1}{2} \log_2 \left(1 + \frac{P}{N}\right)$ was derived by 7 other authors in 1948!

Hartley’s rule is exact: $C' = C$ (a coincidence?)

$C'$ is the capacity of a noisy “uniform” channel
Hartley’s rule $C' = \log_2 \left(1 + \frac{A}{\Delta}\right)$ is not Hartley’s

Shannon’s $C = \frac{1}{2} \log_2 \left(1 + \frac{P}{N}\right)$ was derived by 7 other authors in 1948!

Hartley’s rule is exact: $C' = C$ (a coincidence?)

$C'$ is the capacity of a noisy “uniform” channel
Outline

Who is Claude Shannon?
Shannon’s Seminal Paper
Shannon’s Main Contributions
Shannon’s Capacity Formula

Hartley’s rule $C' = \log_2 \left(1 + \frac{A}{\Delta}\right)$ is not Hartley’s

Many authors independently derived $C = \frac{1}{2} \log_2 \left(1 + \frac{P}{N}\right)$ in 1948.

Hartley’s rule is exact: $C' = C$ (a coincidence?)

$C'$ is the capacity of the “uniform” channel

Shannon’s Conclusion
Hartley or not Hartley

Quote from Shannon, 1984:

"I was working on typology and then I started on the general aspects of information theory. I started with information theory, inspired by Hartley's paper, which was a good paper, but it did not take account of things like noise and best encoding and probabilistic aspects.³

R.P.: You have said to other people that these were closely

- In Hartley's paper, no mention of signal vs. noise or $A$ vs. $\Delta$
- Why was $C' = \log_2\left(1 + \frac{A}{\Delta}\right)$ mistakenly attributed to Hartley?
A HISTORY OF THE THEORY OF INFORMATION

By E. COLIN CHERRY, M.Sc., Associate Member.

(The paper was first received 7th February, and in revised form 28th May, 1951.)

increased. Although not explicitly stated in this form in his paper, Hartley\textsuperscript{12} has implied that the quantity of information which can be transmitted in a frequency band of width $B$ and necessary 256 data in a time $t$, and the vertical the "smallest distinguishable" amplitude change; in practice this smallest step may be taken to equal the noise level, $n$. Then the quantity of information transmitted may be shown to be proportional to

$$Bt \log \left(1 + \frac{a}{n}\right)$$
Outline

Who is Claude Shannon?
Shannon’s Seminal Paper
Shannon’s Main Contributions
Shannon’s Capacity Formula

Hartley’s rule $C' = \log_2 \left( 1 + \frac{A}{\Delta} \right)$ is not Hartley’s

Many authors independently derived $C = \frac{1}{2} \log_2 \left( 1 + \frac{P}{N} \right)$ in 1948.

Hartley’s rule is exact: $C' = C$ (a coincidence?)

$C'$ is the capacity of the “uniform” channel

Shannon’s Conclusion
And then there were eight

Quote from Shannon, 1948:

Formulas similar to $C = W \log \frac{P + N}{N}$ for the white noise case have been developed independently by several other writers, although with somewhat different interpretations. We may mention the work of N. Wiener, W. G. Tuller, and H. Sullivan in this connection.

And then there were eight

Quote from Shannon, 1948:

Formulas similar to $C = W \log \frac{P + N}{N}$ for the white noise case have been developed independently by several other writers, although with somewhat different interpretations. We may mention the work of Norbert Wiener, William G. Tuller, and Herbert Sullivan in this connection.

And then there were eight

Quote from Shannon, 1948:

Formulas similar to $C = W \log \frac{P + N}{N}$ for the white noise case have been developed independently by several other writers, although with somewhat different interpretations. We may mention the work of N. Wiener, W. G. Tuller, and H. Sullivan in this connection.

3. Herbert Sullivan (unpublished)
And then there were eight

Quote from Shannon, 1948:

Formulas similar to $C = W \log \frac{P + N}{N}$ for the white noise case have been developed independently by several other writers, although with somewhat different interpretations. We may mention the work of N. Wiener, W. G. Tuller, and H. Sullivan in this connection.

3. Herbert Sullivan (unpublished)
4. Jacques Laplume, April 1948
And then there were eight

Quote from Shannon, 1948:

Formulas similar to \( C = W \log \frac{P + N}{N} \) for the white noise case have been developed independently by several other writers, although with somewhat different interpretations. We may mention the work of Norbert Wiener, William G. Tuller, Herbert Sullivan (unpublished), Jacques Laplume, and Charles W. Earp in this connection.

3. Herbert Sullivan (unpublished)
4. Jacques Laplume, April 1948
5. Charles W. Earp, June 1948
And then there were eight

Quote from Shannon, 1948:

Formula similar to $C = W \log \frac{P + N}{N}$ for the white noise case have been developed independently by several other writers, although with somewhat different interpretations. We may mention the work of Norbert Wiener, William G. Tuller, Herbert Sullivan (unpublished), Jacques Laplume, Charles W. Earp, André G. Clavier in this connection.

3. Herbert Sullivan (unpublished)
4. Jacques Laplume, April 1948
5. Charles W. Earp, June 1948
6. André G. Clavier, December 1948
And then there were eight

Quote from Shannon, 1948:

Formulas similar to \( C = W \log \frac{P + N}{N} \) for the white noise case have been developed independently by several other writers, although with somewhat different interpretations. We may mention the work of Norbert Wiener, William G. Tuller, Herbert Sullivan (unpublished), Jacques Laplume, Charles W. Earp, André G. Clavier, and Stanford Goldman in this connection.

3. Herbert Sullivan (unpublished)
4. Jacques Laplume, April 1948
5. Charles W. Earp, June 1948
6. André G. Clavier, December 1948
7. Stanford Goldman, May 1948
And then there were eight

Quote from Shannon, 1948:

Formulas similar to $C = W \log \frac{P + N}{N}$ for the white noise case have been developed independently by several other writers, although with somewhat different interpretations. We may mention the work of Norbert Wiener, William G. Tuller, and Herbert Sullivan in this connection.

3. Herbert Sullivan (unpublished)
4. Jacques Laplume, April 1948
5. Charles W. Earp, June 1948
6. André G. Clavier, December 1948
7. Stanford Goldman, May 1948
There is a large class of phenomena in which what is observed is a numerical quantity, or a sequence of numerical quantities, dis-

An interesting problem is that of determining the information gained by fixing one or more variables in a problem. For example, let us suppose that a variable $u$ lies between $x$ and $x + dx$ with the probability $\exp \left(-\frac{x^2}{2a}\right) \frac{dx}{\sqrt{2\pi a}}$, while a variable $v$ lies between the same two limits with a probability $\exp \left(-\frac{x^2}{2b}\right) \frac{dx}{\sqrt{2\pi b}}$. How much information do we gain concerning $u$ if we know that $u + v = w$? In this case, it is clear that $u = w - v$, where $w$ is
The excess of information concerning \( x \) when we know \( w \) to be that which we have in advance is

\[
\frac{1}{\sqrt{2\pi ab/(a+b)}} \int_{-\infty}^{\infty} \left\{ \exp \left[ -(x - c_2)^2 \left( \frac{a + b}{2ab} \right) \right] \right\} \\
\times \left[ -\frac{1}{2} \log_2 2\pi \left( \frac{ab}{a+b} \right) \right] - (x - c_2)^2 \left( \frac{a + b}{2ab} \right) \log_2 e \right] dx \\
- \frac{1}{\sqrt{2\pi a}} \int_{-\infty}^{\infty} \exp \left( -\frac{x^2}{2a} \right) \left( -\frac{1}{2} \log_2 2\pi a - \frac{x^2}{2a} \log_2 e \right) dx \\
= \frac{1}{2} \log_2 \left( \frac{a + b}{b} \right)
\]  

(3.091)
Norbert Wiener

The excess of information concerning $x$ when we know $w$ to be that which we have in advance is

$$\frac{1}{\sqrt{2\pi[ab/(a+b)]}} \int_{-\infty}^{\infty} \left\{ \exp \left[ -(x - c_2)^2 \left( \frac{a + b}{2ab} \right) \right] \right\} \times \left[ -\frac{1}{2} \log_2 2\pi \left( \frac{ab}{a+b} \right) - (x - c_2)^2 \left[ \left( \frac{a + b}{2ab} \right) \log_2 e \right] \right] dx$$

$$= \frac{1}{2} \log_2 \left( \frac{a + b}{b} \right)$$

(3.091)

Later... in 1956:

What is Information Theory?

NORBERT WIENER

INFORMATION THEORY has been identified in the public mind to denote the theory of information by bits, as developed by Claude E. Shannon and myself. This notion is certainly impor-
Relationship Between Rate of Transmission of Information, Frequency Bandwidth, and Signal-to-Noise Ratio

By C. W. Earp


channels, channel maximum signal to root-mean-square noise ratio = $S_{SSB}/\sqrt{n}$ and maximum signal-to-peak-noise ratio = $S_{SSB}/(p\sqrt{n})$.

In each channel, the available power may be used to provide $N$ instantaneous values, this being achieved without ambiguity provided that

$$N < \left( \frac{S_{SSB}}{p\sqrt{n}} - 1 \right).$$
Charles W. Earp

Relationship Between Rate of Transmission of Information, Frequency Bandwidth, and Signal-to-Noise Ratio*

By C. W. EARP


In each channel, the available power may be used to provide \( N \) instantaneous values, this being achieved without ambiguity provided that

\[
N < \left( \frac{S_{SSB}}{P \sqrt{n}} \right) - 1.
\]

* The present paper was written in original form in October, 1946, when the author had no knowledge of any practical development of pulse-code modulation, as the
The number of significant amplitude levels is usually determined by the noise in the system. If the system is of a linear nature, and the maximum signal amplitude is $S$, while the noise amplitude is $N$, then the number of significant amplitude levels is essentially

$$L = (S/N) + 1$$  \hspace{1cm} (2)$$

where the “1” is due to the fact that the zero signal level can be used.

The value of $L$ can be expressed in a

---

*Some Fundamental Considerations Concerning Noise Reduction and Range in Radar and Communication*

STANFORD GOLDMAN†, SENIOR MEMBER, I.R.E.

---

Shannon’s Formula $W \log(1 + SNR)$: A Historical Perspective
Some Fundamental Considerations Concerning Noise Reduction and Range in Radar and Communication

STANFORD GOLDMAN†, SENIOR MEMBER, I.R.E.

The number of significant amplitude levels is usually determined by the noise in the system. If the system is of a linear nature, and the maximum signal amplitude is $S$, while the noise amplitude is $N$, then the number of significant amplitude levels is essentially

$$L = (S/N) + 1 \quad (2)$$

where the “1” is due to the fact that the zero signal level can be used.

Equation (5) has been derived independently by many people, among them W. G. Tuller, from whom the writer first learned about it.
Theoretical Limitations on the Rate of Transmission of Information*

WILLIAM G. TULLER†, SENIOR MEMBER, IEEE

Then, if $N$ is the rms amplitude of the noise mixed with the signal, there are $1 + S/N$ significant values of signal that may be determined. This sets $s$ in

$$H = kn \log s = k2f_s T \log (1 + S/N).$$

This is an important expression, to be sure, but gives
William G. Tuller

Theoretical Limitations on the Rate of Transmission of Information*

WILLIAM G. TULLER†, SENIOR MEMBER, IRE

Recognizable.¹⁴ Then, if $N$ is the rms amplitude of the noise mixed with the signal, there are $1 + S/N$ significant values of signal that may be determined. This sets $s$ in

$$H = kn \log s = k2f_cT \log (1 + S/N).$$

(2)

This is an important expression, to be sure, but gives


¹¹ The existence of this work was learned by the author in the spring of 1946, when the basic work underlying this paper had just been completed. Details were not known by the author until the summer of 1948, at which time the work reported here had been complete for about eight months.
Communication in the Presence of Noise*  

CLAUSE E. SHANNON†, MEMBER, IRE  

Theorem 2: Let P be the average transmitter power, and suppose the noise is white thermal noise of power N in the band W. By sufficiently complicated encoding systems it is possible to transmit binary digits at a rate

\[ C = W \log_2 \frac{P + N}{N} \]  

(19)

with as small a frequency of errors as desired. It is not pos.
Communication in the Presence of Noise*  

CLAUSE E. SHANNON†, MEMBER, IRE  

Communication in the Presence of Noise*

CLAUDE E. SHANNON†, MEMBER, IRE

[10] A. Hodges, *Alan Turing: The Enigma*, New York: Simon and Schuster, 1983. [The following information was obtained from C. E. Shannon on March 3, 1984: “On p. 552, Hodges cites a Shannon manuscript date of 1940, which is, in fact, a typographical error. While results for coding statistical sources into noiseless channels using the plog(p) measure were obtained in 1940–1941 (at the Institute for Advanced Study in Princeton), first submission of this work for formal publication occurred soon after World War II.”]
What about the French?

Deux ingénieurs français ont publié la même « formule de Shannon » en 1948:
What about the French?

Deux ingénieurs français ont publié la même « formule de Shannon » en 1948:

Clavier & Laplume
Evaluation of Transmission Efficiency According to Hartley’s Expression of Information Content*

By A. G. CLAVIER

Federal Telecommunication Laboratories, Incorporated, Nutley, New Jersey

small percentage of error due to noise. The total number of distinguishable levels on the ideal line is thus given by

\[ \frac{S + N\sqrt{2}}{N\sqrt{2}} = 1 + \frac{S}{N\sqrt{2}} , \]

with a reasonable approximation. It follows that the amount of information transmittible on the ideal line is measured by

\[ H_{lm} = k_0 \cdot 2f_t \cdot t \cdot \log \left( 1 + \frac{S_t}{N_t N^2} \right) . \]
André G. Clavier

Evaluation of Transmission Efficiency According to Hartley’s Expression of Information Content*

By A. G. CLAVIER

Federal Telecommunication Laboratories, Incorporated, Nutley, New Jersey

* A symposium on “Recent Advances in the Theory of Communication” was presented at the November 12, 1947, meeting of the New York Section of the Institute of Radio Engineers. Four papers were presented by A. G. Clavier, Federal Telecommunication Laboratories; B. D. Loughlin, Hazeltine Electronics Corporation; and J. R. Pierce and C. E. Shannon, both of Bell Telephone Laboratories. The
Jacques Laplume

Meanwhile (1948), far away...

**Physique Mathématique. — Sur le nombre de signaux discernables en présence du bruit erratique dans un système de transmission à bande passante limitée.**
Note de M. Jacques Laplume.

Si $N$ et $n$ sont suffisamment grands, on peut former une expression approchée de $\log M$ en utilisant la formule de Stirling limitée aux termes prépondérants. On trouve ainsi

\[ \log M \approx N \log \frac{N + n}{N} + n \log \frac{N + n}{n}. \]

Si, de plus, $N \gg n$,

\[ \log M \approx n \log \frac{N}{n} = TW \log \frac{P}{b}. \]
Cette note vise à faire sortir de l'oubli un travail original de 1948 de l'ingénieur français Jacques Laplume, relatif au calcul de la capacité d'un canal bruité de bande passante donnée. La publication de sa Note dans les Comptes Rendus de l'Académie des sciences a précédé de peu celle de l'article du mathématicien américain Claude E. Shannon, fondateur de la théorie de l'information, ainsi que celles de plusieurs chercheurs aux U.S.A. qui avaient proposé la même année 1948 des formules de capacité analogues.

La singularité de Jacques Laplume réside dans le fait qu'il travaillait indépendamment et isolément en France, et que son approche est (contrairement à Shannon) plus physique que mathématique bien qu'exploitant explicitement (comme Shannon) le caractère aléatoire du bruit. La Note est semble-t-il tombée rapidement dans l'oubli et n'a été exhumée qu'en 1998, à l'occasion du cinquantenaire de la théorie de l'information. Même au début de cette année 2016 qui marque le centième anniversaire de la naissance de Shannon, nous ne connaissions que très peu de choses sur l'œuvre et la vie de Jacques Laplume. Il nous a fallu un véritable travail de détective pour remonter jusqu'à sa famille et découvrir que cette Note de Jacques Laplume n'était qu'une parmi de nombreux travaux extrêmement divers d'une personnalité forte aux multiples talents qui s'inscrit dans une véritable saga familiale.
Who’s formula?

The “Shannon-Hartley” formula

\[ C = \frac{1}{2} \log_2 \left( 1 + \frac{P}{N} \right) \]
Who’s formula?

The “Shannon-Hartley” formula

\[ C = \frac{1}{2} \log_2 \left( 1 + \frac{P}{N} \right) \]

would actually be the

Shannon-Laplume-Tuller-Wiener-Clavier-Earp-Goldman-Sullivan formula
Outline

Who is Claude Shannon?
Shannon’s Seminal Paper
Shannon’s Main Contributions
Shannon’s Capacity Formula
Hartley’s rule $C' = \log_2 \left(1 + \frac{A}{\Delta} \right)$ is not Hartley’s
Many authors independently derived $C = \frac{1}{2} \log_2 \left(1 + \frac{P}{N} \right)$ in 1948.
Hartley’s rule is exact: $C' = C$ (a coincidence?)
$C'$ is the capacity of the “uniform” channel
Shannon’s Conclusion
“Hartley”’s argument

The channel input $X$ is taking $M = 1 + A/\Delta$ equiprobable values in the set $\{-A, -A + 2\Delta, \ldots, A - 2\Delta, A\}$:

$$P = \mathbb{E}(X^2) = \frac{1}{M} \sum_{k=0}^{n} (M - 1 - 2k)^2 = \Delta^2 \frac{M^2 - 1}{3}.$$

The input is mixed with additive noise $Z$ with accuracy $\pm \Delta$, i.e. having uniform distribution in $[-\Delta, \Delta]$:

$$N = \mathbb{E}(Z^2) = \frac{1}{2\Delta} \int_{-\Delta}^{\Delta} z^2 \, dz = \frac{\Delta^2}{3}.$$
“Hartley”’s argument

The channel input $X$ is taking $M = 1 + A/\Delta$ equiprobable values in the set $\{-A, -A + 2\Delta, \ldots, A - 2\Delta, A\}$:

$$P = \mathbb{E}(X^2) = \frac{1}{M} \sum_{k=0}^{n} (M - 1 - 2k)^2 = \Delta^2 \frac{M^2 - 1}{3}.$$

The input is mixed with additive noise $Z$ with accuracy $\pm \Delta$, i.e. having uniform distribution in $[-\Delta, \Delta]$:

$$N = \mathbb{E}(Z^2) = \frac{1}{2\Delta} \int_{-\Delta}^{\Delta} z^2 dz = \frac{\Delta^2}{3}.$$

Hence

$$\log_2 \left(1 + \frac{A}{\Delta}\right) = \frac{1}{2} \log_2 (1 + M^2 - 1) = \frac{1}{2} \log_2 \left(1 + \frac{3P}{\Delta^2}\right) = \frac{1}{2} \log_2 \left(1 + \frac{P}{N}\right)$$

i.e., $C' = C$. A mathematical coincidence?
Outline

Who is Claude Shannon?
Shannon’s Seminal Paper
Shannon’s Main Contributions
Shannon’s Capacity Formula
Hartley’s rule $C' = \log_2 \left( 1 + \frac{A}{\Delta} \right)$ is not Hartley’s
Many authors independently derived $C = \frac{1}{2} \log_2 \left( 1 + \frac{P}{N} \right)$ in 1948.
Hartley’s rule is exact: $C' = C$ (a coincidence?)
$C'$ is the capacity of the “uniform” channel
Shannon’s Conclusion
The uniform channel

The capacity of $Y = X + Z$ with additive uniform noise $Z$ is

$$\max_{X \text{ s.t. } |X| \leq A} I(X; Y) = \max_X h(Y) - h(Y|X)$$

$$= \max_X h(Y) - h(Z)$$

$$= \max_{X \text{ s.t. } |Y| \leq A + \Delta} h(Y) - \log_2(2\Delta)$$

Choose $X^*$ to be discrete uniform in $\{-A, -A + 2\Delta, \ldots, A\}$, then $Y = X^* + Z$ has uniform density over $[-A - \Delta, A + \Delta]$, which maximizes differential entropy:

$$= \log_2(2(A + \Delta)) - \log_2(2\Delta)$$

$$= \log_2 \left( 1 + \frac{A}{\Delta} \right)$$
What is the worst noise?

Thus $C' = \log_2 \left( 1 + \frac{A}{\Delta} \right)$ is correct as a capacity! But:

- the noise is *not* Gaussian, but uniform;
- signal limitation is *not* on the power, but on the amplitude.
What is the worst noise?

Thus $C' = \log_2 \left( 1 + \frac{A}{\Delta} \right)$ is correct as a capacity! But:

- the noise is not Gaussian, but uniform;
- signal limitation is not on the power, but on the amplitude.

Further analogy:

- Shannon used the entropy power inequality to show that under limited power, Gaussian is the worst possible noise in the channel:

$$\frac{1}{2} \log_2 \left( 1 + \alpha \frac{P}{N} \right) \leq C \leq \frac{1}{2} \log_2 \left( 1 + \frac{P}{N} \right) + \frac{1}{2} \log_2 \alpha,$$

where $\alpha = N/\bar{N} \geq 1$

- We can show: under limited amplitude, uniform noise is the worst possible noise one can inflict in the channel:

$$\log_2 \left( 1 + \frac{A}{\Delta} \right) \leq C' \leq \log_2 \left( 1 + \frac{A}{\Delta} \right) + \log_2 \alpha,$$

where $\Delta/\bar{\Delta} \geq 1$. 
Conclusion

Why is Shannon’s formula ubiquitous?

\[
C = \frac{1}{2} \log_2 \left(1 + \frac{P}{N}\right).
\]

The uniform (Tuller) and Gaussian (Shannon) channels are not the only examples. Using B-splines, we can construct a sequence of such additive noise channels:

\[\text{uniform channel} \rightarrow \text{Gaussian channel}\]
Conclusion

Why is Shannon’s formula ubiquitous?

we can explain the coincidence by deriving necessary and sufficient conditions s.t. $C = \frac{1}{2} \log_2\left(1 + \frac{P}{N}\right)$. 


Conclusion

Why is Shannon’s formula ubiquitous?

- we can explain the coincidence by deriving necessary and sufficient conditions s.t. $C = \frac{1}{2} \log_2 \left(1 + \frac{P}{N}\right)$.
- the uniform (Tuller) and Gaussian (Shannon) channels are not the only examples.
Conclusion

Why is Shannon’s formula ubiquitous?

- we can explain the coincidence by deriving necessary and sufficient conditions s.t. \( C = \frac{1}{2} \log_2 \left( 1 + \frac{P}{N} \right) \).
- the uniform (Tuller) and Gaussian (Shannon) channels are not the only examples.
- using B-splines, we can construct a sequence of such additive noise channels s.t.

  \[
  \text{uniform channel} \quad \longrightarrow \quad \text{Gaussian channel}
  \]
Conclusion

Why is Shannon’s formula ubiquitous?

- we can explain the coincidence by deriving necessary and sufficient conditions s.t. \( C = \frac{1}{2} \log_2 \left( 1 + \frac{P}{N} \right) \).
- the uniform (Tuller) and Gaussian (Shannon) channels are not the only examples.
- using B-splines, we can construct a sequence of such additive noise channels s.t.

\[
\text{uniform channel} \quad \longrightarrow \quad \text{Gaussian channel}
\]


http://www.mdpi.com/1099-4300/16/9/4892/
Proof. Since $p_{Z^d}(z) = \frac{1}{2} \cdot d \cdot z^2$ is the $(d+1)$th convolution power of the rectangle function of the interval $[0,1]$, the corresponding characteristic function is a $(d+1)$th power of a cardinal sine:

$$Z^d(z) = \text{sinc}_{d+1}(z \cdot M \cdot !).$$

Let $M > 0$ be an integer. From Example 4, we have

$$Z^d(M!) = \text{sinc}_{d+1}(M \cdot !) = \left[ \sin \left( M \cdot \frac{\pi}{2} \right) \right] = e^{i \left( M \cdot \frac{\pi}{2} \right)}.$$ 

This is the characteristic function of the random variable $X^d = \sum_{i=1}^{M} X_i$, where the $X_i$ are i.i.d. and take $M$ equiprobable values in the set $\{ (\frac{M}{1}), (\frac{M}{3}), \ldots, (\frac{M}{3}) \}$. Hence, Theorem 7 applies with $\varpi = M$ and cost function (7).

Again for $d = 0$ one recovers the case of the uniform channel with input $X^d = \sum_{i=1}^{M} X_i$, taking $M$ equiprobable values in the set $\{ (\frac{M}{1}), (\frac{M}{3}), \ldots, (\frac{M}{3}) \}$. In general, the probability distribution of $X^d$ is the $(d+1)$th discrete convolution power of the uniform distribution. For $d = 1$, the pdf of the noise has a triangular shape and the distribution of $X^d$ is also triangular (Figure 1b). As $d$ increases, it becomes closer to a Gaussian shape (Figure 1c,d).

Figure 1. Discrete plots of input probability distributions (of $X^d$) that attain capacity for $M = 15$ and different values of $d$.

(a) $d = 0$ (rectangular)  (b) $d = 1$ (triangular)

(c) $d = 2$  (d) $d = 3$
Outline

Who is Claude Shannon?
Shannon’s Seminal Paper
Shannon’s Main Contributions
Shannon’s Capacity Formula
Hartley’s rule $C' = \log_2 \left( 1 + \frac{A}{\Delta} \right)$ is not Hartley’s

Many authors independently derived $C = \frac{1}{2} \log_2 \left( 1 + \frac{P}{N} \right)$ in 1948.

Hartley’s rule is exact: $C' = C$ (a coincidence?)

$C'$ is the capacity of the “uniform” channel

Shannon’s Conclusion
Shannon on Information Theory

“I didn’t think at the first stages that it was going to have a great deal of impact. I enjoyed working on this kind of a problem, as I have enjoyed working on many other problems, without any notion of either financial or gain in the sense of being famous; and I think indeed that most scientists are oriented that way, that they are working because they like the game.”
Thank you!
Shannon’s Formula

\[ W \cdot \log(1 + \text{SNR}) : \]

A Historical Perspective

on the occasion of Shannon’s Centenary

Oct. 26th, 2016

Olivier Rioul

<olivier.rioul@telecom-paristech.fr>