Shannon's Formula and Hartley's Rule: A Mathematical Coincidence?

Olivier Rioul[*] and José Carlos Magossi[†]

[*]*Télécom ParisTech - Institut Mines-Télécom - CNRS LTCI, Paris, France*
[†]*School of Technology (FT) - Unicamp, Campinas, Brazil*

Abstract. *Shannon's formula* $C = \frac{1}{2}\log(1+P/N)$ is the emblematic expression for the information capacity of a communication channel. Hartley's name is often associated with it, owing to *Hartley's rule:* counting the highest possible number of distinguishable values for a given amplitude A and precision $\pm\Delta$ yields a similar expression $C' = \log(1 + A/\Delta)$. In the information theory community, the following "historical" statements are generally well accepted: (1) Hartley put forth his rule twenty years before Shannon; (2) Shannon's formula as a fundamental tradeoff between transmission rate, bandwidth, and signal-to-noise ratio came unexpected in 1948; (3) Hartley's rule is an imprecise relation while Shannon's formula is exact; (4) Hartley's expression is not an appropriate formula for the capacity of a communication channel.

We show that all these four statements are questionable, if not wrong.

INTRODUCTION

As researchers in information theory we all know that the milestone event that founded our field is Shannon's publication of his seminal 1948 paper [1]. What has rapidly become the emblematic classical expression of the theory is *Shannon's formula* [1, 2][1]

$$C = \tfrac{1}{2}\log_2\left(1 + \frac{P}{N}\right) \tag{1}$$

for the information capacity of a communication channel with signal-to-noise ratio P/N. The classical derivation of (1) was done in [1] as an application of Shannon's coding theorem for a memoryless channel, which states that the best coding procedure for reliable transmission achieves a maximal rate of $C = \max_X I(X;Y)$ bits per sample, where X is the channel input with average power $P = \mathbb{E}(X^2)$ and $Y = X + Z$ is the channel output. Here Z denotes the additive Gaussian random variable (independent of X) that models the communication noise with power $N = \mathbb{E}(Z^2)$.

---

[1] Hereafter we shall always express information capacity in binary units (bits) per *sample*. Shannon's well-known original formulation was in bits per second: $C = W\log_2(1 + P/N)$ bits/s. The difference between this formula and (1) is essentially the content of the *sampling theorem*, that the number of independent samples that can be put through a channel of bandwidth W hertz is $2W$ samples per second. We shall not discuss here whether this theorem should be attributed to Shannon or to other authors that precede him; see e.g., [3] for a recent account on this subject.

Formula (1) is also known as the *Shannon-Hartley formula*, giving the maximum rate at which information can be transmitted reliably over a noisy communication channel (*Shannon-Hartley theorem*) [4]. The reason for which Hartley's name is associated to it is commonly justified by *Hartley's law* (quote from Wikipedia [4]):

> *During 1928, Hartley formulated a way to quantify information and its line rate (also known as data signalling rate R bits per second) [5]. This method, later known as Hartley's law, became an important precursor for Shannon's more sophisticated notion of channel capacity. (...)*
>
> *Hartley argued that (...) if the amplitude of the transmitted signal is restricted to the range of $[-A, +A]$ volts, and the precision of the receiver is $\pm \Delta$ volts, then the maximum number of distinct pulses M is given by $M = 1 + \frac{A}{\Delta}$. By taking information per pulse in bit/pulse to be the base-2-logarithm of the number of distinct messages M that could be sent, Hartley [5] constructed a measure of the line rate R as $R = \log_2(M)$ [bits per symbol].*

In other words, within a noise amplitude limited by Δ, by taking regularly spaced input symbol values $-A, -A + 2\Delta, \ldots, A - 2\Delta, A$ in the range $[-A, A]$ with step 2Δ, one can achieve a maximum total number of $M = A/\Delta + 1$ possible distinguishable values[2]. Therefore, error-free communication is achieved with at most

$$C' = \log_2 \left(1 + \frac{A}{\Delta} \right) \tag{2}$$

bits per sample. This equation strikingly resembles (1). Of course, the "signal-to-noise ratio" A/Δ is a ratio of amplitudes, not of powers, hence should not be confused with the usual definition P/N; accordingly, the factor $1/2$ in (1) is missing in (2). Also, (2) is only considered as an approximation of (1) since it views the communication channel as an errorless M-ary channel, which is an idealization [4].

In the information theory community, the following "historical" statements are generally well accepted:

1. *Hartley put forth his rule (2) twenty years before Shannon.*
2. *The fundamental tradeoff (1) between transmission rate, bandwidth, and signal-to-noise ratio came unexpected in 1948: the times were not ripe for this breakthrough.*
3. *Hartley's rule is inexact while Shannon's formula is characteristic of the additive white Gaussian noise (AWGN) channel ($C' \neq C$).*
4. *Hartley's rule is an imprecise relation between signal magnitude, receiver accuracy and transmission rate that is not an appropriate formula for the capacity of a communication channel.*

In this article, we show that all these four statements are questionable, if not wrong. The organisation is as follows. For $i = 1$ to 4, Section i will defend the opposite view of statement i. The last section concludes.

---

[2] This holds in the most favorable case where A/Δ is an integer, where the "+1" is due to the sample values at the boundaries. Otherwise, M would be the integer part of $A/\Delta + 1$.

106

1. HARTLEY'S RULE IS NOT HARTLEY'S

Hartley [5] was the first researcher to try to formulate a theory of the transmission of information. Apart from stating explicitly that the amount of transmitted information is proportional to the transmission bandwidth, he showed that the number M of possible alternatives from a message source over a given time interval grows exponentially with the duration, suggesting a definition of information as the logarithm $\log M$. However, as Shannon recalled in 1984 [6]:

> *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.*

Indeed, no mention of signal vs. noise, or of amplitude limitation A or Δ was ever made in Hartley's paper [5]. One may then wonder how (2) was coined as Hartley's law. The oldest reference we could find which explicitly attributes (2) to Hartley seems to be the classical 1965 textbook of Wozencraft and Jacobs [7, p. 2–5]:

> *(...) Hartley's argument may be summarized as follows. If we assume that (1) the amplitude of a transmitted pulse is confined to the voltage range $[-A, A]$ and (2) the receiver can estimate a transmitted amplitude reliably only to an accuracy of $\pm\Delta$ volts, then, as illustrated in [the] Figure (...), the maximum number of pulse amplitudes distinguishable at the receiver is $(1 + A/\Delta)$. (...)*
>
> *Hartley's formulation exhibits a simple but somewhat inexact interrelation among (...) maximum signal magnitude A, the receiver accuracy Δ, and the allowable number of message alternatives. Communication theory is intimately concerned with the determination of more precise interrelations of this sort.*

The textbook was highly regarded and is still widely used today. This would explain why (2) is now widely known as Hartley's capacity law.

One may then wonder whether Wozencraft and Jacobs have found such a result themselves while attributing it to Hartley or whether it was inspired from other researchers. We found that the answer is probably in the first tutorial article in information theory that was ever published by E. C. Cherry in 1951 [8]:

> *Although not explicitly stated in this form in his paper, Hartley [5] has implied that the quantity of information which can be transmitted in a frequency band of width B and time T is proportional to the product: $2BT \log M$, where M is the number of "distinguishable amplitude levels." [...] He approximates the waveform by a series of steps, each one representing a selection of an amplitude level. [...] 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 $BT \log(1 + a/n)$ where a is the maximum signal amplitude, an expression given by Tuller [9], being based upon Hartley's definition of information.*

Here Cherry attributes (2) to an *implicit* derivation of Hartley but cites the explicit derivation of Tuller [9]. The next section investigates the contribution of Tuller and others.

2. INDEPENDENT 1948 DERIVATIONS OF THE FORMULA

In the introduction to his classic textbook, Robert McEliece [10] wrote:

> With many profound scientific discoveries (for example Einstein's discovery in 1905 of the special theory of relativity) it is possible with the aid of hindsight to see that the times were ripe for a breakthrough. Not so with information theory. (...) [Shannon's] results were so breathtakingly original that even the communication specialists of the day were at a loss to understand their significance.

One can hardly disagree with this statement when one sees the power and generality of Shannon's results. Being so deep and profound, [1] did not have an immediate impact. As Robert Gallager recalls [11]:

> The first subsequent paper was [12], whose coauthors were B. R. Oliver and J. R. Pierce. This is a very simple paper compared to [1], but it had a tremendous impact by clarifying a major advantage of digital communication. (...) It is probable that this paper had a greater impact on actual communication practice at the time than [1]. The second major paper written at about the same time as [1] is [2]. This is a more tutorial amplification of the AWGN channel results of [1]. (...) This was the paper that introduced many communication researchers to the ideas of information theory.

In [12], Shannon's formula (1) is used without explicit reference to the Gaussian nature of the added white noise, as the capacity of an "ideal system". On the other hand, [2] is devoted to a geometric proof of (1). It appears, therefore, that Shannon's formula (1) was *the* emblematic result that impacted communication specialists at the time, as expressing the correct tradeoff between transmission rate, bandwidth, and signal-to-noise ratio. It is one of Shannon's best known and understood results among communication engineers, if not the most.

As far as (1) is concerned, Shannon, after the completion of [1], acknowledges other works:

> Formulas similar to (1) 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 [13], W. G. Tuller [9], and H. Sullivan in this connection.

Unfortunately, Shannon gives no specific reference to H. Sullivan. S. Verdú [14] cites many more contributions during the same year of 1948:

> By 1948 the need for a theory of communication encompassing the fundamental tradeoffs of transmission rate, reliability, bandwidth, and signal-to-noise ratio was recognized by various researchers. Several theories and principles were put forth in the space of a few months by A. Clavier [15], C. Earp [16], S. Goldman [17], J. Laplume [18], C. Shannon [1], W. Tuller [9], and N. Wiener [13]. One of those theories would prove to be everlasting.

Lundheim [19] reviewed some of these independent discoveries and concludes:

> *(...) this result [Shannon's formula] was discovered independently by several researchers, and serves as an illustration of a scientific concept whose time had come.*

This can be contrasted to the above citation of R. McEliece.

Wiener's independent derivation [13] of Shannon's formula is certainly the one that is closest to Shannon's. He also used probabilistic arguments, logarithmic measures (in base 2) and differential entropy, the latter choice being done "mak[ing] use of a personal communication of J. von Neumann". Unlike Shannon, his definition of information is not based on any precise communication problem. There is also no relation to Hartley's argument leading to (2).

All other independent discoveries that year of 1948 were in fact essentially what is now referred to Hartley's rule leading to (2). Among these, the first published work in April 1948 was by the French engineer Jacques Laplume [18] from Thompson-Houston. He essentially gives the usual derivation that gives (2) for a signal amplitude range $[0, A]$. C. Earp's publication [16] in June 1948 also makes a similar derivation of (2) where the signal-to-noise amplitude ratio is expressed as a "root-mean-square ratio" for the "step modulation" which is essentially pulse-code modulation. In a footnote, Earp claims that his paper "was written in original form in October, 1946". He also mentions that

> *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 Insitute of Radio Engineers. Four papers were presented by A. G. Clavier (...); B.D. Loughlin (...); and J. R. Pierce and C. E. Shannon, both of Bell Telephone Laboratories.*

André Clavier is another French engineer from Le Matériel Téléphonique (LMT) laboratories (subsidiary of International Telephone & Telegraph Corporation), who published [15] in December 1948. He again makes a similar derivation of (2) as Earp's, expressed with root-mean-square values. As Lundheim notes [19, footnote 5], "it is, perhaps, strange that neither Shannon nor Clavier have mutual references in their works, since both [2] and [15] were orally presented at the same meeting (...) and printed more than a year afterwards."

In May 1948, Stanford Goldman again rederives (2), acknowledging that the equation "has been derived independently by many people, among them W. G. Tuller, from whom the writer first learned about it" [17, footnote 4]. William G. Tuller's thesis was defended in June 1948 and printed as an article in May 1949 [9]. His derivation uses again root-mean-square (rms) ratios:

> *Let S be the rms amplitude of the maximum signal that may delivered by the communication system. Let us assume, a fact very close to the truth, that a signal amplitude change less than noise amplitude cannot be recognized, but a signal amplitude change equal to noise is instantly recognizable. 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. (...) the quantity of information available at the output of the system [is $= \log(1 + S/N)$].*

In the 1949 article [9, footnote 11] he explains that

> *The existence of [Shannon's] 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.*

Considering that Tuller's work is—apart from Wiener's—the only work referenced by Shannon in [1], and that the oldest reference known (1946) is Tuller's, it should be perhaps appropriate to refer to (2) as *Tuller's formula* or to (1) as the *Tuller-Shannon formula*.

Interestingly, Shannon's 1949 article [2] explicitly mentions (and criticizes) Hartley's Law and proposes his own interpretation of (2) making the link with his formula (1):

> *How many different signals can be distinguished at the receiving point in spite of the perturbations due to noise? A crude estimate can be obtained as follows. If the signal has a power P, then the perturbed signal will have a power $P + N$. The number of amplitudes that can be reasonably well distinguished is $K\sqrt{\frac{P+N}{N}}$ where K is a small constant in the neighborhood of unity (...) The number of bits that can be sent in this time is $\log_2 M$ $[= \frac{1}{2} \log_2 K^2 \left(1 + \frac{P}{N}\right)]$.*

It may be puzzling to notice, as Hodges did in his historical book on A. Turing [20, p. 552], that Shannon's article [2] mentions a manuscript received date of 23 July, 1940! But this was later corrected by Shannon himself in 1984 (cited in [6, reference 10]):

> *(...) Hodges cites a Shannon manuscript date 1940, which is, in fact, a typographical error. (...) First submission of this work for formal publication occurred soon after World War ll.*

This would mean in particular that Shannon's work leading to his formula was completed in 1946, at about the same time as Tuller's.

3. HARTLEY'S RULE YIELDS SHANNON'S FORMULA: $C' = C$

Let us consider again the argument leading to (2). The channel input X is taking $M = 1 + A/\Delta$ values in the set $\{-A, -A + 2\Delta, \ldots, A - 2\Delta, A\}$, that is the set of values $(M - 1 - 2k)\Delta$ for $k = 0, \ldots, M - 1$. A maximum amount of information will be conveyed through the channel if the input values are equiprobable. Then, using the well-known formula for the sum of squares of consecutive integers, one finds :

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

The input is mixed with additive noise Z with accuracy $\pm\Delta$. The least favorable case would be that Z follows a uniform distribution in $[-\Delta, \Delta]$. Then its average power is

110

$N = \mathbb{E}(Z^2) = \frac{1}{2\Delta} \int_{-\Delta}^{\Delta} z^2 \, \mathrm{d}z = \frac{\Delta^2}{3}$. It follows that (2) becomes

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

A mathematical coïncidence?

In any case, such an identification of (1) and (2) calls for verification that Hartley's rule would in fact be "mathematically correct" as a capacity formula.

4. HARTLEY'S RULE AS A CAPACITY FORMULA

Consider the *uniform channel*, a memoryless channel with additive white noise Z with uniform density in the interval $[-\Delta, \Delta]$. If X is the channel input, the output will be $Y = X + Z$, where X and Z are independent. We assume that the input has the amplitude constraint $|X| \le A$ and that A/Δ is integral [3]. Then

Theorem. *The uniform channel has capacity C' given by* (2).

A similar calculation was proposed as a homework exercice in the excellent textbook by Cover and Thomas [21, Chapter 9, Problem 4]. The proof is omitted here due to lack of space (see [22]).

Thus there is a sense in which the "Tuller-Shannon formula" (2) is *correct* as the capacity of a communication channel, except that the communication noise is *not* Gaussian, but uniform, and that signal limitation is *not* on the power, but on the amplitude.

The analogy between the Gaussian and uniform channels can be pushed further. Both channels are memoryless and additive where the noise Z maximizes the differential entropy $h(Z)$ under the corresponding constraint. Shannon used these properties to show that, under limited *power*, Gaussian noise is the worst possible noise one can impose in the channel (in terms of its capacity) [1]. With our mathematical analysis it can be easily shown [22] that the uniform channel enjoys a similar property: under limited *amplitude*, *uniform* noise is the worst possible noise one can impose in the channel.

CONCLUSION

In this paper, we have criticized the four "historical" statements in the introduction:

1. Hartley's article contains no mention of signal amplitude vs. noise precision—the earliest reference to such "Hartley's rule" seems to be the classical 1965 textbook of Wozencraft and Jacobs;
2. at least seven authors have independently derived formulas very similar to Shannon's, most of them coinciding with "Hartley's rule", in the same year 1948—the earliest contribution seems to be Tuller's;

---

[3] If A/Δ is not integral, then the proof of the theorem can be used to show that $C' \le \log_2(1 + A/\Delta)$, yet C' cannot be obtained in closed form.

3. a careful calculation shows that "Hartley's rule" in fact coincides with Shannon's formula: $C = C'$;

4. "Hartley's rule" is in fact mathematically correct as the capacity of a communication channel, where the communication noise is not Gaussian but uniform, and the signal limitation is not on the power but on the amplitude.

As a further perspective, a detailed mathematical analysis can be carried out. We can explain the mathematical coincidence $C = C'$ by deriving necessary and sufficient conditions on an additive noise channel such that its capacity is given by Shannon's formula. The uniform (Hartley) and Gaussian (Shannon) channels are not the only examples. We can construct a sequence of such additive noise channels, starting with the uniform channel and converging to the Gaussian channel [22].

ACKNOWLEDGMENT

REFERENCES

1. C. E. Shannon, *Bell System Technical Journal* **27**, 379–423, 623–656 (1948), reprinted in C. E. Shannon and W. Weaver, *The Mathematical Theory of Communication*, Univ. Illinois Press, 1949.
2. C. E. Shannon, *Proceedings of the Institute of Radio Engineers* **37**, 10–21 (1949).
3. P. Butzer, M. Dodson, P. Ferreira, J. Higgins, O. Lange, P. Seidler, and R. Stens, *Applicable Analysis* **90**, 643–688 (2011).
4. Wikipedia, Shannon-Hartley theorem — Wikipedia, the free encyclopedia (2014), URL `http://en.wikipedia.org/wiki/Shannon-Hartley_theorem`.
5. R. V. L. Hartley, *Bell System Technical Journal* **7**, 535–563 (1928).
6. F. W. Ellersick, *IEEE Communications Magazine* **22**, 123–126 (1984).
7. J. M. Wozencraft, and I. M. Jacobs, *Principles of Communication Engineering*, John Wiley & Sons, 1965.
8. E. C. Cherry, *Proceedings of the Institute of Electrical Engineers* **98**, 383–393 (1951).
9. W. G. Tuller, *Theoretical limitations on the rate of transmission of information*, Ph.D. thesis, Massachusetts Institute of Technology (1948), published in *Proceedings of the Institute of Radio Engineers*, Vol. 37, No. 5, pp. 468–478, May 1949.
10. R. J. McEliece, *The theory of information and coding*, Cambridge University Press, 2002.
11. R. Gallager, *IEEE Transactions on Information Theory* **47**, 2681–2695 (2001).
12. B. Oliver, J. Pierce, and C. E. Shannon, *Proceedings of the Institute of Radio Engineers* **36**, 1324–1331 (1948).
13. N. Wiener, *Cybernetics, chapter III: Time series, information and communication*, Wiley, 1948.
14. S. Verdú, *IEEE Transactions on Information Theory* **44**, 2057–2078 (1998), ISSN 0018-9448.
15. A. G. Clavier, *Electronic Communication: ITT Technical Journal* **25**, 414–420 (1948).
16. C. W. Earp, *Electronic Communication: ITT Technical Journal* **25**, 178–195 (1948).
17. S. Goldman, *Proceedings of the Institute of Radio Engineers* **36**, 584–594 (1948).
18. J. Laplume, *Comptes rendus de l'Académie des Sciences de Paris* **226**, 1348–1349 (1948).
19. L. Lundheim, *Telektronikk, Special issue on Information theory* **98**, 20–29 (2002).
20. A. Hodges, *Alan Turing: The Enigma*, Simon and Schuster, 1983.
21. T. M. Cover, and J. A. Thomas, *Elements of information theory*, John Wiley & Sons, 2006.
22. O. Rioul, and J. C. Magossi, *Entropy, Special Issue on Information, Entropy and their Geometric Structures* **16**, 4892–4910 (2014), URL `http://www.mdpi.com/1099-4300/16/9/4892`.