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

21/09/2014

Auteurs:
Publication MaxEnt 2014
OAI: oai:www.see.asso.fr:9603:11326
DOI:

MaxEnt 2014

Abstract

Collection

Authors

Olivier Rioul

Optimal Transport to Rényi Entropies
Shannon’s Formula and Hartley’s Rule: A Mathematical Coincidence?
Publication externe des contenus du portail des publications

Tout extrait des contenus du portail destiné à être utilisé dans des publicités, des communiqués de presse ou du matériel de promotion nécessite un accord préalable écrit de la SEE. Une version préliminaire du document proposé contenant ces extraits doit accompagner chacune de ces demandes. SEE se réserve le droit de refuser un tel usage externe pour quelque raison que ce soit.

Responsabilités

La SEE apporte tout le soin possible à la préparation des informations délivrées dans les contenus produits. Cependant elle ne peut être tenue pour responsable d'aucune perte ou frais qui pourrait résulter d'imprécisions, d'inexactitudes, d'erreurs ou de possibles omissions portant sur des informations publiées, ni des résultats obtenus par l'utilisation et la pratique des informations délivrées.

Utilisation des informations recueillies lors du téléchargement de contenu

Le portail des publications est susceptible d'utiliser des « cookies » afin notamment de permettre l'utilisation de paniers d'achat et de personnaliser les parcours sur le site. SEE se réserve la possibilité d'utiliser les informations recueillies lors des téléchargements pour ses besoins internes et notamment pour l'amélioration de ses services, sans qu'elles puissent être cédées à des partenaires commerciaux. Conformément à la loi "informatique et libertés" du 6 janvier 1978, chaque utilisateur du portail dispose d'un droit d'accès et de rectification aux informations qui le concernent. Pour exercer ce droit, les utilisateurs doivent s'adresser à SEE – 17 rue de l'amiral Hamelin – 75783 Paris Cedex 16, par simple lettre ou en utilisant le formulaire de contact disponible sur son site.

Paris, le 28 avril 2013

Sponsors

Scientific sponsors

Logistic sponsors
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 to 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 did 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 out 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. Keywords: Shannon's formula; Hartley's rule; additive noise channel; differential entropy; channel capacity; signal-to-noise ratio; additive white Gaussian noise (AWGN); uniform noise channel. PACS: 01.65.+g, 02.50.Cw

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 $[1, 2] C = 1 2 \log(1 + P/N)$ 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 $P = E(X^2)$ 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 $N = E(Z^2)$. 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 predate him in this discovery; 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 the so-called 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 distinguishable pulses $M$ is given by $M = 1 + A \Delta$. 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 $\Delta$, by taking regularly spaced input symbol values $-A,-A + 2\Delta,\ldots,A - 2\Delta,A$ in the range $[-A,A]$ with step $2\Delta$, one can achieve a maximum total number of $M = A/\Delta + 1$ distinguishable values$^2$. Therefore, error-free communication is achieved with at most $C = \log_2 1 + A)$ (2) bits per sample. This equation strikingly resembles (1). Of course, the “signal-to-noise ratio” $A/\Delta$ 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 did put forth his rule (2) twenty years before Shannon. 2. The fundamental tradeoff (1) between transmission rate, bandwidth, and signal-to-noise ratio came out unexpected in 1948: times were not even ripe for this breakthrough. 3. Hartley’s rule is inexact while Shannon’s formula is characteristic of the additive white Gaussian noise (AWGN) channel ($C = C$). 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 $A/\Delta$ is an integer, where the “+1” is due to the sample values at the boundaries. Otherwise, $M$ would be the integer part of $A/\Delta + 1$. 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 given a 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 $\Delta$ or $\Delta$ 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 mentions (2)—and is incidentally cited in the Wikipedia page [4]—seems to be the classical 1965 textbook of Wozencraft and Jacobs, most notably its introduction chapter [7, p. 2–5]; (...) in 1928, Hartley [5] reasoned that Nyquist’s result, when coupled with a limitation on the accuracy of signal reception, implied a restriction on the amount of data that can be communicated reliably over a physical channel. 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 (…) the maximum signal magnitude $A$, the receiver accuracy $\Delta$, and the allowable number of message alternatives. Communication theory is inter- mately concerned with the determination of more precise interrelations of this sort. The textbook was highly regarded and still widely used today. Its introductory text have become famous to many researchers in the field of communication theory and has had a tremendous impact. 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 them- selves while attributing it to Hartley or whether it was inspired from other researchers. The next section proposes an answer. 2. INDEPENDENT 1948 DERIVATIONS OF THE FORMULA In the introduction to his classic textbook, Robert McEliece [8] 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.3. Being so deep and profound, [1] did not have an immediate impact. As Robert Gallager recalls [9]: The first subsequent paper was [10], 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 [10], 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 Shannon’s result that is the best known and understood among communications engineers. 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 [11], W. G. Tuller [12], and H. Sullivan in this connection. S. Verdù [13] 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 [14], C. Earp [15], S. Goldman [16], J. Laplume [17], C. Shannon [1], W. Tuller [12], and N. S. To mention but a few: The introduction of the formal architecture of communication systems (Shannon’s paradigm) with explicit distinction between source, channel and destination; the emphasis on digital representation to make the chance of error as small as desired; the consideration of codes in high dimensions; the use of probabilistic models for the signal as well as for the noise, via information theoretic tools like entropy and mutual information. 4 Unfortunately, Shannon gives no specific reference to H. Sullivan. Wiener [11]. One of those theories would prove to be everlasting. Lundheim [18] 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 [11] 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 “making use of a personal communication of J. von Neumann”. Unlike Shannon, however, 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 [17] from Thompson-Houston. He essentially gives the usual derivation that gives (2) for a signal amplitude range [0, A]. C. Earp’s publication [15] 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 PCM. 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 Institute of Radio Engineers. Four papers were presented by A. G. Clavier (…); B. D. Loughlin (…); and R. R. Pierce and C. E. Shannon, both of Bell Telephone Laboratories. André Clavier is another French engineer from LMT laboratories (subsidiary of ITT Corporation), who published [14] in December 1948. He again makes a similar derivation (of (2) as Earp’s, expressed with root-mean-square values. As Lundheim notes [18, footnote 5]: “It is, perhaps, strange that neither Shannon nor Clavier have mutual refer- ences in their works, since both [2] and [14] 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” [16, footnote 4]. William G. Tuller’s thesis was defended in June 1948 and printed as an article in May 1949 [12]. His derivation uses again root-mean-square (rms) ratios: Let S be the rms amplitude of the maximum signal that may be 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 [12, 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 √ 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 log2 M = log2 K2 1+ P N. It may be puzzling to notice, as Hodges did in his historical book on A. Turing [19, 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 II. This would mean in particular that Shannon’s work was submitted to a journal 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/Δ values in the set {−A, −A+2Δ, −A−2Δ, A}, that is the set of values (M−1−2k)Δ for k = 0,…, 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 finds5 : P = E(X2) = 1 M ∑n k=0(M −1−2k)2 = Δ2 M2−1 3 . The input is mixed with additive noise Z with accuracy ±Δ. The least favorable case would be that Z follows a uniform distribution in [−Δ,Δ]. Then its average power is 5 Interestingly, this is the classical formula for the average power of a M-state PCM or PAM signal, as was derived by Oliver, Pierce and Shannon in [10]. N = E(Z2) = 1 2M Δ−Δ z=2 dz = Δ2 3 . It follows that (2) becomes C = log2 M = 1 2 log2(1+M2 −1) = 1 2 log2 1+ 3P Δ2 = 1 2 log2 1+ P N = C. A mathematical coincidence? 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 \( Z \) 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 \( |X| \leq A \) and that \( A/\Delta \) is integral. Then Theorem. The uniform channel has capacity \( C \) given by (2). A similar calculation was proposed as a homework exercise in the excellent textbook by Cover and Thomas [20, Chapter 9, Problem 4]. The proof is omitted here due to lack of space (see [21]). 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. Also in both cases, the noise \( Z = Z^* \) maximizes 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 inflict in the channel (in terms of its capacity) [1]. With our mathematical analysis it can be easily shown [21] that the uniform channel enjoys a similar property: under limited amplitude, uniform noise is the worst possible noise one can inflict 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 to Hartley's rule; 6. If \( A/\Delta \) is not integral, then the proof of the theorem can be used to show that \( C \log(1+A/\Delta) \), 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. It is possible to construct a sequence of such additive noise channels, starting with the uniform channel and converging to the Gaussian channel [21].


https://www.see.asso.fr/en/node/11326/landing