Information Channel Capacity

Kirk T. McDonald
Joseph Henry Laboratories, Princeton University, Princeton, NJ 08544

(July 23, 2011; updated Nov. 22, 2016)

1 Problem

Estimate the maximum number C of bits of information that can be transmitted per second in a communication channel of bandwidth B, meaning that this channel acts as a low-pass filter with cutoff frequency B.[1]

2 Solution

The solution proceeds by taking the channel capacity C to be the product of the maximum number of pulses per second that can be transmitted times the maximum amount of information that can be encoded onto a single pulse.

2.1 Maximum Number of Pulses per Second

A simple model of a pulse is that it is one-half period of a sine wave. The shortest period wave that can be transmitted down a channel of bandwidth B is $1/B$. Hence, the shortest pulse that can be transmitted has period $1/2B$, and the maximum number of (distinct) pulses per second that can be transmitted is

$$N_{\rm pulse} = 2B. \tag{1}$$

The estimate (1) is sometimes called Nyquist's theorem [1].

2.2 Maximum Information per Pulse

If the pulse can have n distinguishable amplitudes (often called levels), this amplitude can be interpreted as a binary number consisting of $\log_2 n$ bits.

Hence, the maximum number of bits per second that can be transmitted by encoding pulses is[2]

$$C = N_{\rm pulse} \log_2 n = 2B \log_2 n. \tag{2}$$

[1] A communication channel consists of a transmitter, a receiver, and some medium between these two (possibly vacuum) that supports transmission of signals. In most cases, the transmitter and/or receiver have limited bandwidth, while the medium could transmit pulses of arbitrarily small time width. Hence, "channel capacity" is more a property of the transmitter and receiver, than of the signal-transmission medium.

[2] In simple pulse telegraphy, the presence of a pulse corresponds to a 1 and its absence to a 0. The number of levels is $n = 2$, in which case eq. (2) has the same form as eq. (1). If it is desired to transmit a digitized analog signal with n analog levels via simple pulse telegraphy, the bandwidth of the analog signal is limited to B/n, since a block of n two-level pulses is used to represent each n-level analog sample. Equivalently, if the analog sampling is performed to m-bit accuracy, then the analog bandwidth is limited to $B/2^m$.

If the number of levels per pulse can be arbitrarily high, the amount of information that can be transmitted can also be arbitrarily large. However, in practice there will be some maximum transmittable pulse amplitude, $A_{\rm max}$, and there will always be some kind of noise in the transmission. A simple model of the noise is that its average amplitude is zero, with a Gaussian distribution of variance $A_{\rm noise}$.

We estimate that the number of distinguishable levels in this case is $n = A_{\rm max}/A_{\rm noise}$ (although it is doubtful that a pulse of amplitude $A_{\rm noise}$ can be clearly distinguished from one of amplitude $2A_{\rm noise}$ in the presence of the noise). From this we estimate that the number of bits per pulse that can be transmitted down a noisy channel is $n = \log_2 A_{\rm max}/A_{\rm noise} = \log_2 \sqrt{S/N} = \frac{1}{2}\log_2 S/N$, where the signal power S and the noise power N are proportional to the squares of the respective amplitudes.

While this estimate makes sense if S/N is large, the logarithm is negative when $S/N < 1$. To keep the number of bits per pulse positive when the signal-to-noise ratio is low, we revise our estimate to be

$$\frac{1}{2}\log_2(1 + S/N). \tag{3}$$

Combining this with the estimate (1), we obtain the final result that the information capacity of a noisy channel is

$$C = B \log_2(1 + S/N). \tag{4}$$

This result is often called Shannon's theorem [2, 3, 4].[3]

A simple model is that the noise power per unit bandwidth is the constant η. If so, the channel capacity for a fixed maximum signal power S but variable bandwidth B is

$$C(B) = B \log_2(1 + S/\eta B). \tag{5}$$

If we can increase the bandwidth B indefinitely then eventually $S/\eta B$ becomes small, and

$$C \to \frac{1}{\ln 2}\frac{S}{\eta} = 1.44\frac{S}{\eta} \tag{6}$$

is the maximum channel capacity for infinite bandwidth but fixed signal power (in which case the signal-to-noise ratio is very poor, and we should consider the processing of weak signals).

2.3 Weak Signals

When the signal-to-noise ratio is low it is better to speak of the number of pulses needed to successfully transmit one bit of information, rather than the number of bits per pulse.

When $S/N \ll 1$ the number of bits per pulse according to eq. (3) is $(1/2\ln 2)S/N = 0.7S/N$. That is, it takes approximately $1.4N/S$ pulses to successfully transmit one bit. To

[3]We have arrived at the results (1) and (4) without requiring the use of Fourier analysis, but only that there exists a minimum time width of a recognizable pulse. It is claimed in [5] that these results depend on Fourier analysis, and that the information capacity of "nonperiodic" channels can exceed these bounds. Note that messages are generally nonperiodic in that a periodic message, such as 01010101..., has very little information content. Hence, all practical information channels are "nonperiodic", and the claims of [5] seem ill founded.

see that this is reasonable, note that the standard deviation of the mean of a sample of m pulses of mean amplitude A is $\sigma_A/\sqrt{m}$, where σ_A is the variance of the amplitudes of the pulses. For the case of $A = \sqrt{S}$ the variance is $\sigma_A = \sqrt{N}$, so with $m = 1.4N/S$ the standard deviation of the mean is $\sigma_A/\sqrt{m} = \sqrt{N}/\sqrt{1.4N/S} = \sqrt{S/1.4} = A/\sqrt{1.4}$. Thus, the quality of the measurement of the $1.4N/S$ pulses is $\sqrt{m}A/\sigma_A = \sqrt{1.4} = 1.2$ standard deviations, which is barely statistically significant.

2.4 Information and Entropy

So far, we have considered the number of bits is a message as the measure of its information content. But, if all the bits in a long message are the same, a one-bit version would suffice to carry this message. This led people, starting with Szilard [6], and more famously Shannon [2], to define a statistical measure H of the information content[4] of a set $\{i\}$ of possible messages, of which message i is transmitted with probability p_i, to be

$$H = -\sum_i p_i \log_2 p_i. \tag{7}$$

This measure is often called Shannon entropy. For a review, see [7]. For discussion of the relation between entropy and Maxwell's demon, see, for example, [8].

A way to apply this measure to a single message of m bits is to suppose that there are only two submessages, a 0 and a 1. So, if $0 \le n \le m$ of the bits are 0's, and $m - n$ are 1's we obtain,

$$H(m,n) = -\frac{n}{m}\log_2\frac{n}{m} - \frac{m-n}{m}\log_2\frac{m-n}{m} = \log_2\frac{m}{m-n} + \frac{n}{m}\log_2\left(\frac{m}{n} - 1\right) \le 1. \tag{8}$$

$H(m,n)$ is zero when $n = 0$ or m and 1 when $n = m/2$. This measure successfully identifies messages in which all bits are the same as having low information content, but it does not consider the order of the bits, and rates the message 01010101 as having high information content.[5]

A more practical question is: how much shorter could a message be made without losing any of its information content? This is the challenge of data compression. The measure H gives a clue as to how much a message could be compressed, but it does not suggest the best way to do that compression. One of the most popular data-compression algorithms (with zero loss of information) was developed by Huffman [9], and is the basis for zip files.

---

[4]The symbol H honors Boltzmann who first wrote expressions of the form (7).

[5]A surprising debate persists as to whether H or $-H$ should be considered the measure of information content of a set of messages. The definition (7) implies that messages consisting of random bits have the highest information content, whereas many people feel that if the bits are random we should say that the message has no content.

A lesson I have learned from quantum information theory, which applies also to the classical case, is that we can say that we have knowledge of a state/message only if we can copy it. In this view, very little knowledge is required to make a copy of a message in which all the bits are the same, but a lot of knowledge is required to make a copy if the bits are random. Hence, I am of the persuasion that a random message actually has high information content, and that definition (7) is appropriate.

Furthermore, as discussed in [8], the task of erasing a message leads to a net increase of entropy of the Universe, which I find consistent with the notion that a nontrivial message is in some way equivalent to a high, not low, state of entropy.

2.5 Increased Channel Capacity via Waves with Field Angular Momentum

Historically, the earliest physical realization of an information channel was two-wire transmission lines (telegraph lines), and followed by use of (two-conductor) coaxial transmission lines. These lines support transverse (TEM) waves of any wavelength, with field patterns that are static solutions, multiplied by an axial wavefunction $e^{i(kz-\omega t)}$ (for lines parallel to the z-axis). For a given wavelength, there is only a single possible wave, unless the wavelength is smaller than the separation between the two conductors of the line. In the latter case, there also exist waves for coaxial lines with axial wavefunctions of the form $e^{i(kz-\omega t\pm m\phi)}$ for integer m. Waves with nonzero m carry field angular momentum, and lines of the Poynting vector (rays) form helices.

Waves of each index m constitute a separate information channel, for which Shannon's results hold separately. Hence, use of short (but finite) wavelengths on two-conductor transmission lines permits, in principle, infinite total channel capacity if the waves with field angular momentum could be exploited. In practice, two-conductor transmission lines are so lossy for such short wavelengths that only long wavelengths and the TEM mode is used, and the channel capacity is finite.

Hollow, conducting waveguides, and more particularly dielectric waveguides (optical fibers) do not support TEM waves, but only waves of small wavelength, which can carry field angular momentum.[6] The two modes, often called TE and TM, with index $m = 0$ are the most common, but there exists an infinite set of modes with nonzero m, called EH or HE depending on whether E_z is larger or smaller than H_z.[7] If the source and receiver on such a waveguide can generate and detect waves of n different values of index m, for which two distinct modes exist with each index, the channel capacity would be $2n$ times larger than that if only a single mode were used.

Likewise, electromagnetic waves in free space, such as laser beams, have an infinite set of modes at each wavelength. Such modes can also be characterized by an index m related to field angular momentum, with two modes (polarizations) for each index m.[8] Again, in principle, use of modes with n different values of index m would permit information channel capacity $2n$ times that for a single mode.

At present, such mode separation seems not to be practical for large m, but is an active area of research. See, for example, [13].

References

[1] H. Nyquist, *Certain Factors Affecting Telegraph Speed*, Bell System Tech. J. **3**, 324 (1924), http://physics.princeton.edu/~mcdonald/examples/statmech/nyquist_bstj_3_324_24.pdf

[6]So-called weakly guiding dielectric waveguides have two modes, often both called HE_{11} that propagates at any wavelength. However, practical optical fibers are not weakly guiding, and have diameter much larger than an optical wavelength, so this mode does not play as special a role as does the TEM mode of two-conductor lines.

[7]For an introduction to such modes, see sec. 8.11B of [10]. For a discussion of hollow, conducting, circular waveguides that emphasizes angular momentum, see [11].

[8]For discussion by the author of Gaussian laser beams with angular momentum, see [12].

Certain Topics in Telegraph Transmission Theory, Trans. AIEE **47**, 617 (1928),
http://physics.princeton.edu/~mcdonald/examples/statmech/nyquist_taiee_47_617_28.pdf

[2] C.E. Shannon, *A Mathematical Theory of Communication*, Bell System Tech. J. **20**, 379, 623 (1948),
http://physics.princeton.edu/~mcdonald/examples/statmech/shannon_bstj_20_27_379_48.pdf
http://physics.princeton.edu/~mcdonald/examples/statmech/shannon_bstj_20_27_623_48.pdf

[3] C.E. Shannon, *Communication in the Presence of Noise*, Proc. IRE **37**, 10 (1949),
http://physics.princeton.edu/~mcdonald/examples/statmech/shannon_pire_37_10_49.pdf

[4] C.E. Shannon and W. Weaver, *The Mathematical Theory of Communication* (U. Illinois Press, 1949).

[5] J. Prothero, *The Shannon Law for Non-Periodic Channels* (May 2, 2012),
http://www.astrapi-corp.com/wp-content/uploads/2014/02/R-2012-1-The-Shannon-Law-for-Non-Periodic-Channels-v2.pdf
http://physics.princeton.edu/~mcdonald/examples/statmech/prothero_120502.pdf

[6] L. Szilard, *On the Decrease of Entropy in a Thermodynamic System by the Intervention of Intelligent Beings*, Z. Phys. **53**, 840 (1929), English translation:
http://physics.princeton.edu/~mcdonald/examples/statmech/szliard_zp_53_840_29_english.pdf

[7] J. Machta, *Entropy, information, and computation*, Am. J. Phys. **67**, 1075 (1999),
http://physics.princeton.edu/~mcdonald/examples/statmech/machta_ajp_67_1075_99.pdf

[8] K.T. McDonald, *Maxwell's Demon* (Oct. 3, 2004),
http://physics.princeton.edu/~mcdonald/examples/demon.pdf

[9] D.A. Huffman, *A Method for the Construction of Minimum-Redundancy Codes*, Proc. IRE **40**, 1098 (1952),
http://physics.princeton.edu/~mcdonald/examples/statmech/huffman_pire_40_1098_52.pdf

[10] J.D. Jackson, *Classical Electrodynamics*, 3rd ed. (Wiley, 1999).

[11] K.T. McDonald, *Angular Momentum in Circular Waveguides* (June 29, 2013),
http://physics.princeton.edu/~mcdonald/examples/circular.pdf

[12] K.T. McDonald, *Gaussian Laser Beams via Oblate Spheroidal Waves* (Oct. 19, 2002),
http://physics.princeton.edu/~mcdonald/examples/circular.pdf

[13] M. Krenn *et al.*, *Twisted light transmission over 143 km*, Proc. Nat. Aca. Sci. **113**, xxx (2016), http://physics.princeton.edu/~mcdonald/examples/QM/krenn_pnas_16.pdf