Communication in the Presence of Noise

CLAUDE E. SHANNON, MEMBER, IRE

Classic Paper

A method is developed for representing any communication system geometrically. Messages and the corresponding signals are points in two "function spaces," and the modulation process is a mapping of one space into the other. Using this representation, a number of results in communication theory are deduced concerning expansion and compression of bandwidth and the threshold effect. Formulas are found for the maximum rate of transmission of binary digits over a system when the signal is perturbed by various types of noise. Some of the properties of "ideal" systems which transmit at this maximum rate are discussed. The equivalent number of binary digits per second for certain information sources is calculated.

I. INTRODUCTION

A general communications system is shown schematically in Fig. 1. It consists essentially of five elements.

1) An Information Source: The source selects one message from a set of possible messages to be transmitted to the receiving terminal. The message may be of various types; for example, a sequence of letters or numbers, as in telegraphy or teletype, or a continuous function of time $f(t)$, as in radio or telephony.

2) The Transmitter: This operates on the message in some way and produces a signal suitable for transmission to the receiving point over the channel. In telephony, this operation consists of merely changing sound pressure into a proportional electrical current. In telegraphy, we have a encoding operation which produces a sequence of dots, dashes, and spaces corresponding to the letters of the message. To take a more complex example, in the case of multiplex PCM telephony the different speech functions must be sampled, compressed, quantized and encoded, and finally interleaved properly to construct the signal.

3) The Channel: This is merely the medium used to transmit the signal from the transmitting to the receiving point. It may be a pair of wires, a coaxial cable, a band of radio frequencies, etc. During transmission, or at the receiving terminal, the signal may be perturbed by noise or distortion. Noise and distortion may be differentiated on the basis that distortion is a fixed operation applied to the signal, while noise involves statistical and unpredictable perturbations. Distortion can, in principle, be corrected by applying the inverse operation, while a perturbation due to noise cannot always be removed, since the signal does not always undergo the same change during transmission.

This paper is reprinted from the PROCEEDINGS OF THE IRE, vol. 37, no. 1, pp. 10–21, Jan. 1949.
Publisher Item Identifier S 0018-9219(98)01299-7.

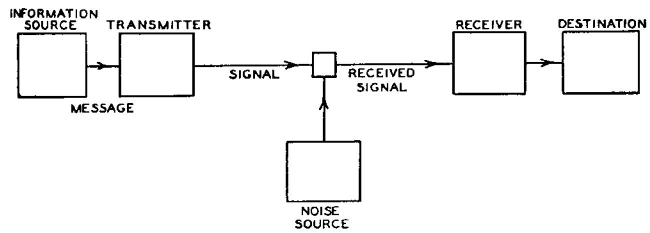

Fig. 1. General communications system.

4) The Receiver: This operates on the received signal and attempts to reproduce, from it, the original message. Ordinarily it will perform approximately the mathematical inverse of the operations of the transmitter, although they may differ somewhat with best design in order to combat noise.

5) The Destination: This is the person or thing for whom the message is intended.

Following Nyquist[1] and Hartley,[2] it is convenient to use a logarithmic measure of information. If a device has n possible positions it can, by definition, store $\log_b n$ units of information. The choice of the base b amounts to a choice of unit, since $\log_b n = \log_b c \log_c n$. We will use the base 2 and call the resulting units binary digits or bits. A group of m relays or flip-flop circuits has 2^m possible sets of positions, and can therefore store $\log_2 2^m = m$ bits.

If it is possible to distinguish reliably M different signal functions of duration T on a channel, we can say that the channel can transmit $\log_2 M$ bits in time T. The *rate* of transmission is then $\log_2 M/T$. More precisely, the *channel capacity* may be defined as

$$C = \lim_{T\to\infty} \frac{\log_2 M}{T}. \tag{1}$$

[1] H. Nyquist, "Certain factors affecting telegraph speed," *Bell Syst. Tech. J.*, vol. 3, p. 324, Apr. 1924.

[2] R. V. L. Hartley, "The transmission of information," *Bell Syst. Tech. J.*, vol. 3, p. 535–564, July 1928.

A precise meaning will be given later to the requirement of reliable resolution of the M signals.

II. The Sampling Theorem

Let us suppose that the channel has a certain bandwidth W in cps starting at zero frequency, and that we are allowed to use this channel for a certain period of time T. Without any further restrictions this would mean that we can use as signal functions any functions of time whose spectra lie entirely within the band W, and whose time functions lie within the interval T. Although it is not possible to fulfill both of these conditions exactly, it is possible to keep the spectrum within the band W, and to have the time function very small outside the interval T. Can we describe in a more useful way the functions which satisfy these conditions? One answer is the following.

Theorem 1: If a function $f(t)$ contains no frequencies higher than W cps, it is completely determined by giving its ordinates at a series of points spaced 1/2 W seconds apart.

This is a fact which is common knowledge in the communication art. The intuitive justification is that, if $f(t)$ contains no frequencies higher than W, it cannot change to a substantially new value in a time less than one-half cycle of the highest frequency, that is, $1/2W$. A mathematical proof showing that this is not only approximately, but exactly, true can be given as follows. Let $F(\omega)$ be the spectrum of $f(t)$. Then

$$f(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} F(\omega)e^{i\omega t}\,d\omega \qquad (2)$$

$$= \frac{1}{2\pi}\int_{-2\pi W}^{+2\pi W} F(\omega)e^{i\omega t}\,d\omega \qquad (3)$$

since $F(\omega)$ is assumed zero outside the band W. If we let

$$t = \frac{n}{2W} \qquad (4)$$

where n is any positive or negative integer, we obtain

$$f\left(\frac{n}{2W}\right) = \frac{1}{2\pi}\int_{-2\pi W}^{+2\pi W} F(\omega)e^{i\omega\frac{n}{2W}}\,d\omega. \qquad (5)$$

On the left are the values of $f(t)$ at the sampling points. The integral on the right will be recognized as essentially the nth coefficient in a Fourier-series expansion of the function $F(\omega)$, taking the interval $-W$ to $+W$ as a fundamental period. This means that the values of the samples $f(n/2W)$ determine the Fourier coefficients in the series expansion of $F(\omega)$. Thus they determine $F(\omega)$, since $F(\omega)$ is zero for frequencies greater than W, and for lower frequencies $F(\omega)$ is determined if its Fourier coefficients are determined. But $F(\omega)$ determines the original function $f(t)$ completely, since a function is determined if its spectrum is known. Therefore the original samples determine the function $f(t)$ completely. There is one and only one function whose spectrum is limited to a band W, and which passes through given values at sampling points separated 1.2 W seconds apart. The function can be simply reconstructed from the samples by using a pulse of the type

$$\frac{\sin 2\pi W t}{2\pi W t}. \qquad (6)$$

This function is unity at $t = 0$ and zero at $t = n/2W$, i.e., at all other sample points. Furthermore, its spectrum is constant in the band W and zero outside. At each sample point a pulse of this type is placed whose amplitude is adjusted to equal that of the sample. The sum of these pulses is the required function, since it satisfies the conditions on the spectrum and passes through the sampled values.

Mathematically, this process can be described as follows. Let x_n be the nth sample. Then the function $f(t)$ is represented by

$$f(t) = \sum_{n=-\infty}^{\infty} x_n \frac{\sin \pi(2Wt - n)}{\pi(2Wt - n)}. \qquad (7)$$

A similar result is true if the band W does not start at zero frequency but at some higher value, and can be proved by a linear translation (corresponding physically to single-sideband modulation) of the zero-frequency case. In this case the elementary pulse is obtained from $\sin x/x$ by single-side-band modulation.

If the function is limited to the time interval T and the samples are spaced $1/2W$ seconds apart, there will be a total of $2TW$ samples in the interval. All samples outside will be substantially zero. To be more precise, we can define a function to be limited to the time interval T if, and only if, all the samples outside this interval are exactly zero. Then we can say that any function limited to the bandwidth W and the time interval T can be specified by giving $2TW$ numbers.

Theorem 1 has been given previously in other forms by mathematicians[3] but in spite of its evident importance seems not to have appeared explicitly in the literature of communication theory. Nyquist,[4,5] however, and more recently Gabor,[6] have pointed out that approximately $2TW$ numbers are sufficient, basing their arguments on a Fourier series expansion of the function over the time interval T. This given TW and $(TW + 1)$ cosine terms up to frequency W. The slight discrepancy is due to the fact that the functions obtained in this way will not be strictly limited to the band W but, because of the sudden starting and stopping of the sine and cosine components, contain some frequency content outside the band. Nyquist pointed out the fundamental importance of the time interval $1/2W$ seconds in connection with telegraphy, and we will call this the Nyquist interval corresponding to the band W.

[3] J. M. Whittaker, *Interpolatory Function Theory*, Cambridge Tracts in Mathematics and Mathematical Physics, no. 33. Cambridge, U.K.: Cambridge Univ. Press, ch. IV, 1935.

[4] H. Nyquist, "Certain topics in telegraph transmission theory," *AIEE Trans.*, p. 617, Apr. 1928.

[5] W. R. Bennett, "Time division multiplex systems," *Bell Syst. Tech. J.*, vol. 20, p. 199, Apr. 1941, where a result similar to Theorem 1 is established, but on a steady-state basis.

[6] D. Gabor, "Theory of communication," *J. Inst. Elect. Eng.* (London), vol. 93, pt. 3, no. 26, p. 429, 1946.

The $2TW$ numbers used to specify the function need not be the equally spaced samples used above. For example, the samples can be unevenly spaced, although, if there is considerable bunching, the samples must be known very accurately to give a good reconstruction of the function. The reconstruction process is also more involved with unequal spacing. One can further show that the value of the function and its derivative at every other sample point are sufficient. The value and first and second derivatives at every third sample point give a still different set of parameters which uniquely determine the function. Generally speaking, any set of $2TW$ independent numbers associated with the function can be used to describe it.

III. Geometrical Representation of the Signals

A set of three numbers x_1, x_2, x_3, regardless of their source, can always be thought of as coordinates of a point in three-dimensional space. Similarly, the $2TW$ evenly spaced samples of a signal can be thought of as coordinates of a point in a space of $2TW$ dimensions. Each particular selection of these numbers corresponds to a particular point in this space. Thus there is exactly one point corresponding to each signal in the band W and with duration T.

The number of dimensions $2TW$ will be, in general, very high. A 5-Mc television signal lasting for an hour would be represented by a point in a space with $2 \times 5 \times 10^6 \times 60^2 = 3.6 \times 10^{10}$ dimensions. Needless to say, such a space cannot be visualized. It is possible, however, to study analytically the properties of n-dimensional space. To a considerable extent, these properties are a simple generalization of the properties of two- and three-dimensional space, and can often be arrived at by inductive reasoning from these cases. The advantage of this geometrical representation of the signals is that we can use the vocabulary and the results of geometry in the communication problem. Essentially, we have replaced a complex entity (say, a television signal) in a simple environment [the signal requires only a plane for its representation as $f(t)$] by a simple entity (a point) in a complex environment ($2TW$ dimensional space).

If we imagine the $2TW$ coordinate axes to be at right angles to each other, then distances in the space have a simple interpretation. The distance from the origin to a point is analogous to the two- and three-dimensional cases

$$d = \sqrt{\sum_{n=1}^{2TW} x_n^2} \tag{8}$$

where x_n is the nth sample. Now, since

$$f(t) = \sum_{n=1}^{2TW} x_n \frac{\sin \pi(2Wt - n)}{\pi(2Wt - n)} \tag{9}$$

we have

$$\int_{-\infty}^{\infty} f(t)^2 dt = \frac{1}{2W} \sum x_n^2 \tag{10}$$

using the fact that

$$\int_{-\infty}^{\infty} \frac{\sin \pi(2Wt - m)}{\pi(2Wt - m)} \frac{\sin \pi(2Wt - n)}{\pi(Wt - n)} dt = \begin{cases} 0 & m \neq n \\ \frac{1}{2W} & m = n. \end{cases} \tag{11}$$

Hence, the square of the distance to a point is $2W$ times the energy (more precisely, the energy into a unit resistance) of the corresponding signal

$$d^2 = 2WE = 2WTP \tag{12}$$

where P is the average power over the time T. Similarly, the distance between two points is $\sqrt{2WT}$ times the rms discrepancy between the two corresponding signals.

If we consider only signals whose average power is less than P, these will correspond to points within a sphere of radius

$$r = \sqrt{2WTP}. \tag{13}$$

If noise is added to the signal in transmission, it means that the point corresponding to the signal has been moved a certain distance in the space proportional to the rms value of the noise. Thus noise produces a small region of uncertainty about each point in the space. A fixed distortion in the channel corresponds to a warping of the space, so that each point is moved, but in a definite fixed way.

In ordinary three-dimensional space it is possible to set up many different coordinate systems. This is also possible in the signal space of $2TW$ dimensions that we are considering. A different coordinate system corresponds to a different way of describing the same signal function. The various ways of specifying a function given above are special cases of this. One other way of particular importance in communication is in terms of frequency components. The function $f(t)$ can be expanded as a sum of sines and cosines of frequencies $1/T$ apart, and the coefficients used as a different set of coordinates. It can be shown that these coordinates are all perpendicular to each other and are obtained by what is essentially a rotation of the original coordinate system.

Passing a signal through an ideal filter corresponds to projecting the corresponding point onto a certain region in the space. In fact, in the frequency-coordinate system those components lying in the pass band of the filter are retained and those outside are eliminated, so that the projection is on one of the coordinate lines, planes, or hyperplanes. Any filter performs a linear operation on the vectors of the space, producing a new vector linearly related to the old one.

IV. Geometrical Representation of Messages

We have associated a space of $2TW$ dimensions with the set of possible signals. In a similar way one can associate a space with the set of possible messages. Suppose we are considering a speech system and that the messages consist

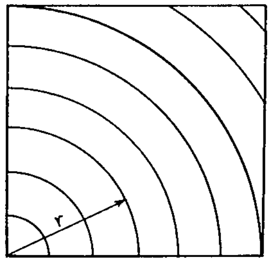

Fig. 2. Reduction of dimensionality through equivalence classes.

of all possible sounds which contain no frequencies over a certain limit W_1 and last for a time T_1.

Just as for the case of the signals, these messages can be represented in a one-to-one way in a space of $2T_1W_1$ dimensions. There are several points to be noted, however. In the first place, various different points may represent the same message, insofar as the final destination is concerned. For example, in the case of speech, the ear is insensitive to a certain amount of phase distortion. Messages differing only in the phases of their components (to a limited extent) sound the same. This may have the effect of reducing the number of essential dimensions in the message space. All the points which are equivalent for the destination can be grouped together and treated as one point. It may then require fewer numbers to specify one of these "equivalence classes" than to specify an arbitrary point. For example, in Fig. 2 we have a two-dimensional space, the set of points in a square. If all points on a circle are regarded as equivalent, it reduces to a one-dimensional space—a point can now be specified by one number, the radius of the circle. In the case of sounds, if the ear were completely insensitive to phase, then the number of dimensions would be reduced by one-half due to this cause alone. The sine and cosine components a_n and b_n for a given frequency would not need to be specified independently, but only $\sqrt{a_n^2 + b_n^2}$; that is, the total amplitude for this frequency. The reduction in frequency discrimination of the ear as frequency increases indicates that a further reduction in dimensionality occurs. The vocoder makes use to a considerable extent of these equivalences among speech sounds, in the first place by eliminating, to a large degree, phase information, and in the second place by lumping groups of frequencies together, particularly at the higher frequencies.

In other types of communication there may not be any equivalence classes of this type. The final destination is sensitive to any change in the message within the full message space of $2T_1W_1$ dimensions. This appears to be the case in television transmission.

A second point to be noted is that the information source may put certain restrictions on the actual messages. The space of $2T_1W_1$ dimensions contains a point for *every* function of time $f(t)$ limited to the band W_1 and of duration T_1. The class of messages we wish to transmit may be only a small subset of these functions. For example, speech sounds must be produced by the human vocal system. If we are willing to forego the transmission of any other sounds, the effective dimensionality may be considerably decreased. A similar effect can occur through probability considerations. Certain messages may be possible, but so improbable relative to the others that we can, in a certain sense, neglect them. In a television image, for example, successive frames are likely to be very nearly identical. There is a fair probability of a particular picture element having the same light intensity in successive frames. If this is analyzed mathematically, it results in an effective reduction of dimensionality of the message space when T_1 is large.

We will not go further into these two effects at present, but let us suppose that, when they are taken into account, the resulting message space has a dimensionality D, which will, of course, be less than or equal to $2T_1W_1$. In many cases, even though the effects are present, their utilization involves too much complication in the way of equipment. The system is then designed on the basis that all functions are different and that there are no limitations on the information source. In this case, the message space is considered to have the full $2T_1W_1$ dimensions.

V. Geometrical Representation of the Transmitter and Receiver

We now consider the function of the transmitter from this geometrical standpoint. The input to the transmitter is a message; that is, one point in the message space. Its output is a signal—one point in the signal space. Whatever form of encoding or modulation is performed, the transmitter must establish some correspondence between the points in the two spaces. Every point in the message space must correspond to a point in the signal space, and no two messages can correspond to the same signal. If they did, there would be no way to determine at the receiver which of the two messages was intended. The geometrical name for such a correspondence is a mapping. The transmitter maps the message space into the signal space.

In a similar way, the receiver maps the signal space back into the message space. Here, however, it is possible to have more than one point mapped into the same point. This means that several different signals are demodulated or decoded into the same message. In AM, for example, the phase of the carrier is lost in demodulation. Different signals which differ only in the phase of the carrier are demodulated into the same message. In FM the shape of the signal wave above the limiting value of the limiter does not affect the recovered message. In PCM considerable distortion of the received pulses is possible, with no effect on the output of the receiver.

We have so far established a correspondence between a communication system and certain geometrical ideas. The correspondence is summarized in Table 1.

VI. Mapping Considerations

It is possible to draw certain conclusions of a general nature regarding modulation methods from the geometrical picture alone. Mathematically, the simplest types of mappings are those in which the two spaces have the same

Table 1

Communication System	*Geometrical Entity*
The set of possible signals	A space of $2TW$ dimensions
A particular signal	A point in the space
Distortion in the channel	A warping of the space
Noise in the channel	A region of uncertainty about each point
The average power of the signal	$(2TW)^{-1}$ times the square of the distance from the origin to the point
The set of signals of power P	The set of points in a sphere of radius $\sqrt{2TW\,P}$
The set of possible messages	A space of $2T_1W_1$ dimensions
The set of actual messages distinguishable by the destination	A space of D dimensions obtained by regarding all equivalent messages as one point, and deleting messages which the source could not produce
A message	A point in this space
The transmitter	A mapping of the message space into the signal space
The receiver	A mapping of the signal space into the message space

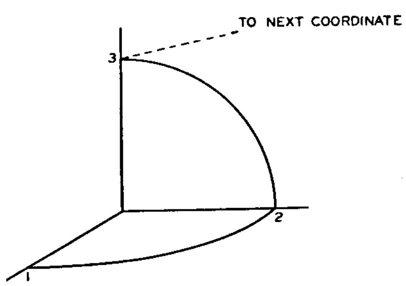

Fig. 3. Mapping similar to frequency modulation.

number of dimensions. Single-sideband amplitude modulation is an example of this type and an especially simple one, since the coordinates in the signal space are proportional to the corresponding coordinates in the message space. In double-sideband transmission the signal space has twice the number of coordinates, but they occur in pairs with equal values. If there were only one dimension in the message space and two in the signal space, it would correspond to mapping a line onto a square so that the point x on the line is represented by (x, x) in the square. Thus no significant use is made of the extra dimensions. All the messages go into a subspace having only $2T_1W_1$ dimensions.

In frequency modulation the mapping is more involved. The signal space has a much larger dimensionality than the message space. The type of mapping can be suggested by Fig. 3, where a line is mapped into a three-dimensional space. The line starts at unit distance from the origin on the first coordinate axis, stays at this distance from the origin on a circle to the next coordinate axis, and then goes to the third. It can be seen that the line is lengthened in this mapping in proportion to the total number of coordinates. It is not, however, nearly as long as it could be if it wound back and forth through the space, filling up the internal volume of the sphere it traverses.

This expansion of the line is related to the improved signal-to-noise ratio obtainable with increased bandwidth. Since the noise produces a small region of uncertainty about each point, the effect of this on the recovered message will be less if the map is in a large scale. To obtain as large a scale as possible requires that the line wander back and

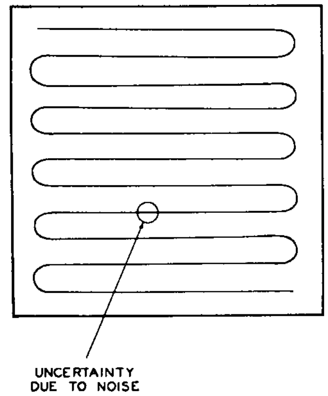

Fig. 4. Efficient mapping of a line into a square.

forth through the higher dimensional region as indicated in Fig. 4, where we have mapped a line into a square. It will be noticed that when this is done the effect of noise is small relative to the length of the line, provided the noise is less than a certain critical value. At this value it becomes uncertain at the receiver as to which portion of the line contains the message. This holds generally, and it shows that any system which attempts to use the capacities of a wider band to the full extent possible will suffer from a threshold effect when there is noise. If the noise is small, very little distortion will occur, but at some critical noise amplitude the message will become very badly distorted. This effect is well known in PCM.

Suppose, on the other hand, we wish to reduce dimensionality, i.e., to compress bandwidth or time or both. That is, we wish to send messages of band W_1 and duration T_1 over a channel with $TW < T_1W_1$. It has already been indicated that the effective dimensionality D of the message space may be less than $2T_1W_1$ due to the properties of the source and of the destination. Hence we certainly need no more than D dimension in the signal space for a good mapping. To make this saving it is necessary, of course, to isolate the effective coordinates in the message space, and to send these only. The reduced bandwidth transmission of speech by the vocoder is a case of this kind.

The question arises, however, as to whether further reduction is possible. In our geometrical analogy, is it possible to map a space of high dimensionality onto one of lower dimensionality? The answer is that it is possible, with certain reservations. For example, the points of a square can be described by their two coordinates which could be written in decimal notation

$$x = .a_1a_2a_3\cdots$$
$$y = .b_1b_2b_3\cdots . \quad (14)$$

From these two numbers we can construct one number by taking digits alternately from x and y

$$z = .a_1b_1a_2b_2a_3b_3\cdots . \quad (15)$$

A knowledge of x and y determines z, and z determines both x and y. Thus there is a one-to-one correspondence between the points of a square and the points of a line.

This type of mapping, due to the mathematician Cantor, can easily be extended as far as we wish in the direction of reducing dimensionality. A space of n dimensions can be mapped in a one-to-one way into a space of one dimension. Physically, this means that the frequency-time product can be reduced as far as we wish when there is no noise, with exact recovery of the original messages.

In a less exact sense, a mapping of the type shown in Fig. 4 maps a square into a line, provided we are not too particular about recovering exactly the starting point, but are satisfied with a nearby one. The sensitivity we noticed before when increasing dimensionality now takes a different form. In such a mapping, to reduce TW, there will be a certain threshold effect when we perturb the message. As we change the message a small amount, the corresponding signal will change a small amount, until some critical value is reached. At this point the signal will undergo a considerable change. In topology it is shown[7] that it is not possible to map a region of higher dimension into a region of lower dimension *continuously*. It is the necessary discontinuity which produces the threshold effects we have been describing for communication systems.

This discussion is relevant to the well-known "Hartley law," which states that "an upper limit to the amount of information which may be transmitted is set by the sum for the various available lines of the product of the line-frequency range of each by the time during which it is available for use."[2] There is a sense in which this statement is true, and another sense in which it is false. It is not possible to map the message space into the signal space in a one-to-one, continuous manner (this is known mathematically as a *topological* mapping) unless the two spaces have the same dimensionality; i.e., unless $D = 2TW$. Hence, if we limit the transmitter and receiver to continuous one-to-one operations, there is a lower bound to the product TW in the channel. This lower bound is determined, not by the product $W_1 T_1$ of message bandwidth and time, but by the number of *essential* dimension D, as indicated in Section IV. There is, however, no good reason for limiting the transmitter and receiver to topological mappings. In fact, PCM and similar modulation systems are highly discontinuous and come very close to the type of mapping given by (14) and (15). It is desirable, then, to find limits for what can be done with no restrictions on the type of transmitter and receiver operations. These limits, which will be derived in the following sections, depend on the amount and nature of the noise in the channel, and on the transmitter power, as well as on the bandwidth-time product.

It is evident that any system, either to compress TW, or to expand it and make full use of the additional volume, must be highly nonlinear in character and fairly complex because of the peculiar nature of the mappings involved.

[7] W. Hurewitz and H. Wallman, *Dimension Theory*. Princeton, NJ: Princeton Univ. Press, 1941.

VII. The Capacity of a Channel in the Presence of White Thermal Noise

It is not difficult to set up certain quantitative relations that must hold when we change the product TW. Let us assume, for the present, that the noise in the system is a white thermal-noise band limited to the band W, and that it is added to the transmitted signal to produce the received signal. A white thermal noise has the property that each sample is perturbed independently of all the others, and the distribution of each amplitude is Gaussian with standard deviation $\sigma = \sqrt{N}$ where N is the average noise power. 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

$$K\sqrt{\frac{P+N}{N}} \tag{16}$$

where K is a small constant in the neighborhood of unity depending on how the phrase "reasonably well" is interpreted. If we require very good separation, K will be small, while toleration of occasional errors allows K to be larger. Since in time T there are $2TW$ independent amplitudes, the total number of reasonably distinct signals is

$$M = \left[K\sqrt{\frac{P+N}{N}}\right]^{2TW}. \tag{17}$$

The number of bits that can be sent in this time is $\log_2 M$, and the rate of transmission is

$$\frac{\log_2 M}{T} = W \log_2 K^2 \frac{P+N}{N} \text{ (bits per second).} \tag{18}$$

The difficulty with this argument, apart from its general approximate character, lies in the tacit assumption that for two signals to be distinguishable they must differ at some sampling point by more than the expected noise. The argument presupposes that PCM, or something very similar to PCM, is the best method of encoding binary digits into signals. Actually, two signals can be reliably distinguished if they differ by only a small amount, provided this difference is sustained over a long period of time. Each sample of the received signal then gives a small amount of statistical information concerning the transmitted signal; in combination, these statistical indications result in near certainty. This possibility allows an improvement of about 8 dB in power over (18) with a reasonable definition of reliable resolution of signals, as will appear later. We will now make use of the geometrical representation to determine the exact capacity of a noisy channel.

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

$$C = W \log_2 \frac{P+N}{N} \tag{19}$$

with as small a frequency of errors as desired. It is not possible by any encoding method to send at a higher rate and have an arbitrarily low frequency of errors.

This shows that the rate $W \log (P+N)/N$ measures in a sharply defined way the capacity of the channel for transmitting information. It is a rather surprising result, since one would expect that reducing the frequency of errors would require reducing the rate of transmission, and that the rate must approach zero as the error frequency does. Actually, we can send at the rate C but reduce errors by using more involved encoding and longer delays at the transmitter and receiver. The transmitter will take long sequences of binary digits and represent this entire sequence by a particular signal function of long duration. The delay is required because the transmitter must wait for the full sequence before the signal is determined. Similarly, the receiver must wait for the full signal function before decoding into binary digits.

We now prove Theorem 2. In the geometrical representation each signal point is surrounded by a small region of uncertainty due to noise. With white thermal noise, the perturbations of the different samples (or coordinates) are all Gaussian and independent. Thus the probability of a perturbation having coordinates $x_1, x_2, \cdots, x_n$ (these are the differences between the original and received signal coordinates) is the product of the individual probabilities for the different coordinates

$$\prod_{n=1}^{2TW} \frac{1}{\sqrt{2\pi 2TWN}} \exp - \frac{x_n^2}{2TWN} = \frac{1}{(2\pi 2TWN)^{TW}} \exp \frac{-1}{2TW} \sum_1^{2TW} x_n^2.$$

Since this depends only on

$$\sum_1^{2TW} x_n^2$$

the probability of a given perturbation depends only on the *distance* from the original signal and not on the direction. In other words, the region of uncertainty is spherical in nature. Although the limits of this region are not sharply defined for a small number of dimensions $(2TW)$, the limits become more and more definite as the dimensionality increases. This is because the square of the distance a signal is perturbed is equal to $2TW$ times the average noise power during the time T. As T increases, this average noise power must approach N. Thus, for large T, the perturbation will almost certainly be to some point near the surface of a sphere of radius $\sqrt{2TWN}$ centered at the original signal point. More precisely, by taking T sufficiently large we can insure (with probability as near to one as we wish) that the perturbation will lie within a sphere of radius $\sqrt{2TW(N+\epsilon)}$ where ϵ is arbitrarily small. The noise regions can therefore be thought of roughly as sharply defined billiard balls, when $2TW$ is very large. The received signals have an average power $P+N$, and in the same sense must almost all lie on the surface of a sphere of radius $\sqrt{2TW(P+N)}$. How many different transmitted signals can be found which will be distinguishable? Certainly not more than the volume of the sphere of radius $\sqrt{2TW(P+N)}$ divided by the volume of a sphere of radius $\sqrt{2TWN}$, since overlap of the noise spheres results in confusion as to the message at the receiving point. The volume of an n-dimensional sphere[8] of radius r is

$$V = \frac{\pi^{n/2}}{\Gamma\left(\frac{n}{2}+1\right)} r^n. \tag{20}$$

Hence, an upper limit for the number M of distinguishable signals is

$$M \leq \left(\sqrt{\frac{P+N}{N}}\right)^{2TW}. \tag{21}$$

Consequently, the channel capacity is bounded by

$$C = \frac{\log_2 M}{T} \leqq W \log_2 \frac{P+N}{N}. \tag{22}$$

This proves the last statement in the theorem.

To prove the first part of the theorem, we must show that there exists a system of encoding which transmits $W \log_2 (P+N)/N$ binary digits per second with a frequency of errors less than ϵ when ϵ is arbitrarily small. The system to be considered operates as follows. A long sequence of, say, m binary digits is taken in at the transmitter. There are 2^m such sequences, and each corresponds to a particular signal function of duration T. Thus there are $M = 2^m$ different signal functions. When the sequence of m is completed, the transmitter starts sending the corresponding signal. At the receiver a perturbed signal is received. The receiver compares this signal with each of the M possible transmitted signals and selects the one which is nearest the perturbed signal (in the sense of rms error) as the one actually sent. The receiver then constructs, as its output, the corresponding sequence of binary digits. There will be, therefore, an overall delay of $2T$ seconds.

To insure a frequency of errors less than ϵ, the M signal functions must be reasonably well separated from each other. In fact, we must choose them in such a way that, when a perturbed signal is received, the nearest signal point (in the geometrical representation) is, with probability greater than $1-\epsilon$, the actual original signal.

It turns out, rather surprisingly, that it is possible to choose our M signal functions at random from the points inside the sphere of radius $\sqrt{2TWP}$, and achieve the most that is possible. Physically, this corresponds very nearly to using M different samples of band-limited white noise with power P as signal functions.

A particular selection of M points in the sphere corresponds to a particular encoding system. The general scheme of the proof is to consider all such selections, and to show that the frequency of errors averaged over all the particular selections is less than ϵ. This will show that there are

[8] D. M. Y. Sommerville, *An Introduction to the Geometry of N Dimensions.* New York: Dutton, 1929, p. 135.

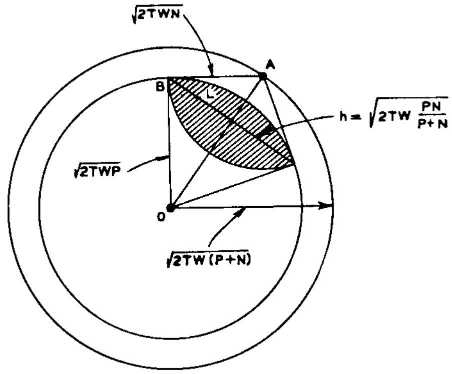

Fig. 5. The geometry involved in Theorem 2.

particular selections in the set with frequency of errors less than ϵ. Of course, there will be other particular selections with a high frequency of errors.

The geometry is shown in Fig. 5. This is a plane cross section through the high-dimensional sphere defined by a typical transmitted signal B, received signal A, and the origin 0. The transmitted signal will lie very close to the surface of the sphere of radius $\sqrt{2TWP}$, since in a high-dimensional sphere nearly all the volume is very close to the surface. The received signal similarly will lie on the surface of the sphere of radius $\sqrt{2TW}(P+N)$. The high-dimensional lens-shaped region L is the region of possible signals that might have caused A, since the distance between the transmitted and received signal is almost certainly very close to $\sqrt{2TWN}$. L is of smaller volume than a sphere of radius h. We can determine h by equating the area of the triangle OAB, calculated two different ways

$$\frac{1}{2}h\sqrt{2TW(P+N)} = \frac{1}{2}\sqrt{2TWP}\sqrt{2TWN}$$

$$h = \sqrt{2TW\frac{PN}{P+N}}.$$

The probability of any particular signal point (other than the actual cause of A) lying in L is, therefore, less than the ratio of the volumes of spheres of radii $\sqrt{2TW\,PN/P+N}$ and $\sqrt{2TWP}$, since in our ensemble of coding systems we chose the signal points at random from the points in the sphere of radius $\sqrt{2TWP}$. This ratio is

$$\left(\frac{\sqrt{2TW\frac{PN}{P+N}}}{\sqrt{2TWP}}\right)^{2TW} = \left(\frac{N}{P+N}\right)^{TW}. \qquad (23)$$

We have M signal points. Hence the probability p that all except the actual cause of A are *outside* L is greater than

$$\left[1-\left(\frac{N}{P+N}\right)^{TW}\right]^{M-1}. \qquad (24)$$

When these points are outside L, the signal is interpreted correctly. Therefore, if we make P greater than $1-\epsilon$, the frequency of errors will be less than ϵ. This will be true if

$$\left[1-\left(\frac{N}{P+N}\right)^{TW}\right]^{(M-1)} > 1-\epsilon. \qquad (25)$$

Now $(1-x)^n$ is always greater than $1-nx$ when n is positive. Consequently, (25) will be true if

$$1-(M-1)\left(\frac{N}{P+N}\right)^{TW} > 1-\epsilon \qquad (26)$$

or if

$$(M-1) < \epsilon\left(\frac{P+N}{N}\right)^{TW} \qquad (27)$$

or

$$\frac{\log(M-1)}{T} < W\log\frac{P+N}{N} + \frac{\log\epsilon}{T}. \qquad (28)$$

For any fixed ϵ, we can satisfy this by taking T sufficiently large, and also have $\log(M-1)/T$ or $\log M/T$ as close as desired to $W\log P+N/N$. This shows that, with a random selection of points for signals, we can obtain an arbitrarily small frequency of errors and transmit at a rate arbitrarily close to the rate C. We can also send *at* the rate C with arbitrarily small ϵ, since the extra binary digits need not be sent at all, but can be filled in at random at the receiver. This only adds another arbitrarily small quantity to ϵ. This completes the proof.

VIII. Discussion

We will call a system that transmits without errors at the rate C an ideal system. Such a system cannot be achieved with any finite encoding process but can be approximated as closely as desired. As we approximate more closely to the ideal, the following effects occur.

1) The rate of transmission of binary digits approaches $C = W\log_2(1+P/N)$.
2) The frequency of errors approaches zero.
3) The transmitted signal approaches a white noise in statistical properties. This is true, roughly speaking, because the various signal functions used must be distributed at random in the sphere of radius $\sqrt{2TWP}$.
4) The threshold effect becomes very sharp. If the noise is increased over the value for which the system was designed, the frequency of errors increases very rapidly.
5) The required delays at transmitter and receiver increase indefinitely. Of course, in a wide-band system a millisecond may be substantially an infinite delay.

In Fig. 6 the function $C/W = \log(1+P/N)$ is plotted with P/N in dB horizontal and C/W the number of bits per cycle of band vertical. The circles represent PCM systems of the binary, ternary, etc., types, using positive and negative pulses and adjusted to give one error in about 10^5 binary digits. The dots are for a PPM system with two,

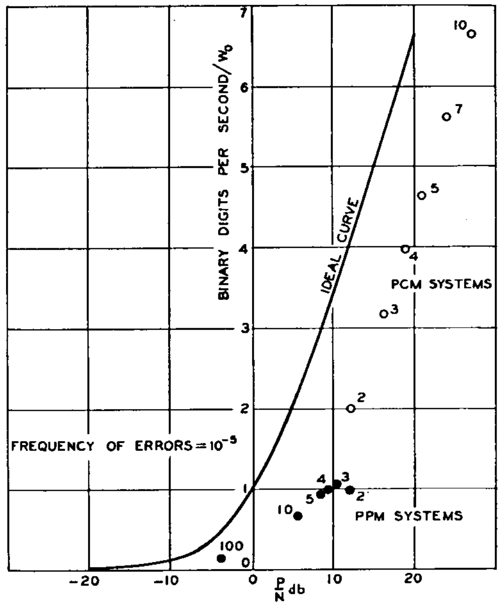

Fig. 6. Comparison of PCM and PPM with ideal performance.

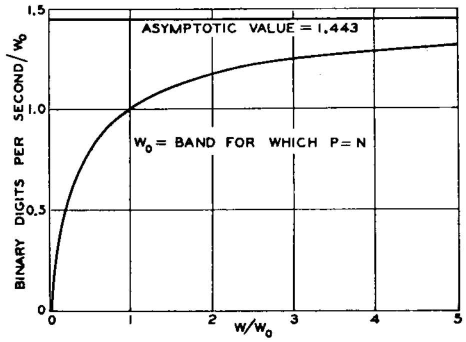

Fig. 7. Channel capacity as a function of bandwidth.

three, etc., discrete positions for the pulse.[9] The difference between the series of points and the ideal curve corresponds to the gain that could be obtained by more involved coding systems. It amounts to about 8 dB in power over most of the practical range. The series of points and circles is about the best that can be done without delay. Whether it is worth while to use more complex types of modulation to obtain some of this possible saving is, of course, a question of relative costs and valuations.

The quantity $TW \log(1 + P/N)$ is, for large T, the number of bits that can be transmitted in time T. It can be regarded as an exchange relation between the different parameters. The individual quantities T, W, P, and N can be altered at will without changing the amount of information we can transmit, provided $TW \log(1+P/N)$ is held constant. If TW is reduced, P/N must be increased, etc.

Ordinarily, as we increase W, the noise power N in the band will increase proportionally; $N = N_0 W$ where N_0 is the noise power per cycle. In this case, we have

$$C = W \log\left(1 + \frac{P}{N_0 W}\right). \tag{29}$$

If we let $W_0 = P/N_0$, i.e., W_0 is the band for which the noise power is equal to the signal power, this can be written

$$\frac{C}{W_0} = \frac{W}{W_0} \log\left(1 + \frac{W_0}{W}\right). \tag{30}$$

In Fig. 7, C/W_0 is plotted as a function of W/W_0. As we increase the band, the capacity increases rapidly until the total noise power accepted is about equal to the signal power; after this, the increase is low, and it approaches an asymptotic value $\log_2 e$ times the capacity for $W = W_0$.

[9] The PCM points are calculated from formulas given in B. M. Oliver, J. R. Pierce, and C. E. Shannon, "The philosophy of PCM," *Proc. IRE*, vol. 36, pp. 1324–1332, Nov. 1948. The PPM points are from unpublished calculations of B. McMillan, who points out that, for very small P/N, the points approach to within 3 dB of the ideal curve.

IX. Arbitrary Gaussian Noise

If a white thermal noise is passed through a filter whose transfer function is $Y(f)$, the resulting noise has a power spectrum $N(f) = K|Y(f)|^2$ and is known as Gaussian noise. We can calculate the capacity of a channel perturbed by any Gaussian noise from the white-noise result. Suppose our total transmitter power is P and it is distributed among the various frequencies according to $P(f)$. Then

$$\int_0^W P(f) df = P. \tag{31}$$

We can divide the band into a large number of small bands, with $N(f)$ approximately constant in each. The total capacity for a given distribution $P(f)$ will then be given by

$$C_1 = \int_0^W \log\left(1 + \frac{P(f)}{N(f)}\right) df \tag{32}$$

since, for each elementary band, the white-noise result applies. The maximum rate of transmission will be found by maximizing C_1 subject to condition (31). This requires that we maximize

$$\int_0^W \left[\log\left(1 + \frac{P(f)}{N(f)}\right) + \lambda P(f)\right] df. \tag{33}$$

The condition for this is, by the calculus of variations, or merely from the convex nature of the curve $\log(1 + x)$

$$\frac{1}{N(f) + P(f)} + \lambda = 0 \tag{34}$$

or $N(f) + P(f)$ must be constant. The constant is adjusted to make the total signal power equal to P. For frequencies where the noise power is low, the signal power should be high, and vice versa, as we would expect.

The situation is shown graphically in Fig. 8. The curve is the assumed noise spectrum, and the three lines correspond to different choices of P. If P is small, we cannot make $P(f) + N(f)$ constant, since this would require negative power at some frequencies. It is easily shown, however, that in this case the best $P(f)$ is obtained by making $P(f) + N(f)$ constant whenever possible, and making $P(f)$

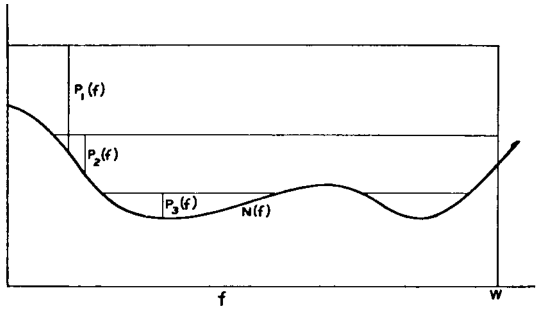

Fig. 8. Best distribution of transmitter power.

zero at other frequencies. With low values of P, some of the frequencies will not be used at all.

If we now vary the noise spectrum $N(f)$, keeping the total noise power constant and always adjusting the signal spectrum $P(f)$ to give the maximum transmission, we can determine the worst spectrum for the noise. This turns out to be the white-noise case. Although this only shows it to be worst among the Gaussian noises, it will be shown later to be the worst among all possible noises with the given power N in the band.

X. The Channel Capacity with an Arbitrary Type of Noise

Of course, there are many kinds of noise which are not Gaussian; for example, impulse noise, or white noise that has passed through a nonlinear device. If the signal is perturbed by one of these types of noise, there will still be a definite channel capacity C, the maximum rate of transmission of binary digits. We will merely outline the general theory here.[10]

Let $x_1, x_2, \cdots, x_n$ be the amplitudes of the noise at successive sample points, and let

$$p(x_1, x_2, \cdots, x_n)dx_1 \cdots dx_n \tag{35}$$

be the probability that these amplitudes lie between x_1 and $x_1 + dx_1$, x_2 and $x_2 + dx_2$, etc. Then the function p describes the statistical structure of the noise, insofar as n successive samples are concerned. The *entropy* H of the noise is defined as follows. Let

$$H_n = \frac{1}{n}\int \cdots \int p(x_1, \cdots, x_n) \cdot \log_e p(x_1, \cdots, x_n)\, dx_1, \cdots, dx_n. \tag{36}$$

Then

$$H = \lim_{n \to \infty} H_n. \tag{37}$$

This limit exists in all cases of practical interest, and can be determined in many of them. H is a measure of the randomness of the noise. In the case of white Gaussian noise of power N, the entropy is

$$H = \log_e \sqrt{2\pi e N}. \tag{38}$$

[10] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, pp. 379–424, July 1948; pp. 623–657, Oct. 1948.

It is convenient to measure the randomness of an arbitrary type of noise not directly by its entropy, but by comparison with white Gaussian noise. We can calculate the power in a white noise having the same entropy as the given noise. This power, namely

$$\overline{N} = \frac{1}{2\pi e} \exp 2H \tag{39}$$

where H is the entropy of the given noise, will be called the *entropy power* of the noise.

A noise of entropy power $\overline{N}$ acts very much like a white noise of power $\overline{N}$, insofar as perturbing the message is concerned. It can be shown that the region of uncertainty about each signal point will have the same volume as the region associated with the white noise. Of course, it will no longer be a spherical region. In proving Theorem 1 this volume of uncertainty was the chief property of the noise used. Essentially the same argument may be applied for any kind of noise with minor modifications. The result is summarized in the following.

Theorem 3: Let a noise limited to the band W have power N and entropy power N_1. The capacity C is then bounded by

$$W \log_2 \frac{P + N_1}{N_1} \leqq C \leqq W \log_2 \frac{P + N}{N_1} \tag{40}$$

where P is the average signal power and W the bandwidth.

If the noise is a white Gaussian noise, $N_1 = N$, and the two limits are equal. The result then reduces to the theorem in Section VII.

For any noise, $N_1 < N$. This is why white Gaussian noise is the worst among all possible noises. If the noise is Gaussian with spectrum $N(f)$, then

$$N_1 = W \exp \frac{1}{w} \int_0^W \log N(f)\, df. \tag{41}$$

The upper limit in Theorem 3 is then reached when we are above the highest noise power in Fig. 8. This is easily verified by substitution.

In the cases of most interest, P/N is fairly large. The two limits are then nearly the same, and we can use $W \log(P + N)/N_1$ as the capacity. The upper limit is the best choice, since it can be shown that as P/N increases, C approaches the upper limit.

XI. Discrete Sources of Information

Up to now we have been chiefly concerned with the channel. The capacity C measures the maximum rate at which a random series of binary digits can be transmitted when they are encoded in the best possible way. In general, the information to be transmitted will not be in this form. It may, for example, be a sequence of letters as in telegraphy, a speech wave, or a television signal. Can we find an equivalent number of bits per second for information sources of this type? Consider first the discrete case; i.e., the message consists of a sequence of discrete symbols. In general, there may be correlation of various sorts between the different symbols. If the message is English text, the letter E is the most frequent, T is often followed by H, etc. These correlations allow a certain compression of the text by proper encoding. We may define the entropy of a discrete source in a way analogous to that for a noise; namely, let

$$H_n = -\frac{1}{n} \sum_{i,\, j, \cdots, s} p(i,\, j, \cdots, s) \log_2 p(i,\, j, \cdots, s) \tag{42}$$

where $p(i,\, j, \cdots, s)$ is the probability of the sequence of symbols $i,\, j, \cdots, s$, and the sum is over all sequences of n symbols. Then the entropy is

$$H = \lim_{n \to \infty} H_n. \tag{43}$$

It turns out that H is the number of bits produced by the source for each symbol of message. In fact, the following result is proved in the appendix.

Theorem 4: It is possible to encode all sequences of n message symbols into sequences of binary digits in such a way that the average number of binary digits per message symbol is approximately H, the approximation approaching equality as n increases.

It follows that, if we have a channel of capacity C and a discrete source of entropy H, it is possible to encode the messages via binary digits into signals and transmit at the rate C/H of the original message symbols per second.

For example, if the source produces a sequence of letters A, B, or C with probabilities $p_A = 0.6$, $p_b = 0.3$, $p_c = 0.1$, and successive letters are chosen independently, then $H_n = H_1 = -[0.6 \log_2 0.6 + 0.3 \log_2 0.3 + 0.1 \log_2 0.1] = 1.294$ and the information produced is equivalent to 1.294 bits for each letter of the message. A channel with a capacity of 100 bits per second could transmit with best encoding $100/1.294 = 77.3$ message letters per second.

XII. Continuous Sources

If the source is producing a continuous function of time, then without further data we must ascribe it an infinite rate of generating information. In fact, merely to specify exactly one quantity which has a continuous range of possibilities requires an infinite number of binary digits. We cannot send continuous information *exactly* over a channel of finite capacity.

Fortunately, we do not need to send continuous messages exactly. A certain amount of discrepancy between the original and the recovered messages can always be tolerated. If a certain tolerance is allowed, then a definite finite rate in binary digits per second can be assigned to a continuous source. It must be remembered that this rate depends on the nature and magnitude of the allowed error between original and final messages. The rate may be described as the rate of generating information *relative to the criterion of fidelity*.

Suppose the criterion of fidelity is the rms discrepancy between the original and recovered signals, and that we can tolerate a value $\sqrt{N_1}$. Then each point in the message space is surrounded by a small sphere of radius $\sqrt{2 I_1 W_1 N_1}$. If the system is such that the recovered message lies within this sphere, the transmission will be satisfactory. Hence, the number of different messages which must be capable of distinct transmission is of the order of the volume V_1 of the region of possible messages divided by the volume of the small spheres. Carrying out this argument in detail along lines similar to those used in Sections VII and IX leads to the following result.

Theorem 5: If the message source has power Q, entropy power $\overline{Q}$, and bandwidth W_1, the rate R of generating information in bits per second is bounded by

$$W_1 \log_2 \frac{\overline{Q}}{N_1} \leqq R \leqq W_1 \log_2 \frac{Q}{N_1} \tag{44}$$

where N_1 is the maximum tolerable mean square error in reproduction. If we have a channel with capacity C and a source whose rate of generating information R is less than or equal to C, it is possible to encode the source in such a way as to transmit over this channel with the fidelity measured by N_1. If $R > C$, this is impossible.

In the case where the message source is producing white thermal noise, $\overline{Q} = Q$. Hence the two bounds are equal and $R = W_1 \log Q/N_1$. We can, therefore, transmit white noise of power Q and band W_1 over a channel of band W perturbed by a white noise of power N and recover the original message with mean square error N_1 if, and only if

$$W_1 \log \frac{Q}{N_1} \leqq W \log \frac{P+N}{N}. \tag{45}$$

Appendix

Consider the possible sequences of n symbols. Let them be arranged in order of decreasing probability, $p_1 \geqq p_2 \geqq p_3 \cdots \geqq p_s$. Let $P_i = \sum_1^{i-1} p_j$. The ith message is encoded by expanding P_j as a binary fraction and using only the first t_i places where t_i is determined from

$$\log_2 \frac{1}{p_i} \leqq t_i < 1 + \log_2 \frac{1}{p_i}. \tag{46}$$

Probable sequences have short codes and improbable ones long codes. We have

$$\frac{1}{2^{t_i}} \leqq p_i \leqq \frac{1}{2^{t_i - 1}}. \tag{47}$$

The codes for different sequences will all be different. P_{i+1}, for example, differs by p_i from P_i, and therefore its binary expansion will differ in one or more of the first t_i places, and similarly for all others. The average length of the encoded message will be $\sum p_i t_i$. Using (46)

$$-\sum p_i \log p_i \leqq \sum p_i t_i < \sum p_i (1 - \log p_i) \tag{48}$$

or

$$n H_n \leqq \sum p_i t_i < 1 + n H_n. \tag{49}$$

The average number of binary digits used per message symbol is $1/n \sum p_i t_i$ and

$$H_n \leqq \frac{1}{n} \sum p_i t_i < \frac{1}{n} + H_n. \tag{50}$$

As $n \to \infty$, $H_n \to H$, and $1/n \to 0$, so the average number of bits per message symbol approaches H.