P • A • R • T 1

PRINCIPLES AND TECHNIQUES

On the CD-ROM

Basic Phenomena
Mathematics, Formulas, Definitions, and Theorems
Circuit Principles

SECTION 1

INFORMATION, COMMUNICATION, NOISE, AND INTERFERENCE

The telephone profoundly changed our methods of communication, thanks to Alexander Graham Bell and other pioneers (Bell, incidentally, declined to have a telephone in his home!). Communication has been at the heart of the information age. Electronic communication deals with transmitters and receivers of electromagnetic waves. Even digital communications systems rely on this phenomenon. This section of the handbook covers information sources, codes and coding, communication channels, error correction, continuous and band-limited channels, digital data transmission and pulse modulation, and noise and interference. C.A.

In This Section:

Section Bibliography:

Of Historical Significance

Davenport, W. B., Jr., and W. L. Root, "An Introduction to the Theory of Random Signals and Noise," McGraw-Hill, 1958. (Reprint edition published by IEEE Press, 1987.)

Middleton, D., "Introduction to Statistical Communication Theory," McGraw-Hill, 1960. (Reprint edition published by IEEE Press, 1996.)

Sloane, N. J. A., and A. D. Wyner (eds.), "Claude Elwood Shannon: Collected Papers," IEEE Press, 1993.

General

Carlson, A. B., et al., "Communications Systems," 4th ed., McGraw-Hill, 2001.

Gibson, J. D., "Principles of Digital and Analog Communications," 2nd ed., Macmillan, 1993.

Haykin, S., "Communication Systems," 4th ed., Wiley, 2000.

Papoulis, A., and S. U. Pillai, "Probability, Random Variables, and Stochastic Processes," 4th ed., McGraw-Hill, 2002.

Thomas, J. B., "An Introduction to Communication Theory and Systems," Springer-Verlag, 1987.

Ziemer, R. E., and W. H. Tranter, "Principles of Communications: Systems, Modulation, and Noise," 5th ed., Wiley, 2001.

Information Theory

Blahut, R. E., "Principles and Practice of Information Theory," Addison-Wesley, 1987.

Cover, T. M., and J. A. Thomas, "Elements of Information Theory," Wiley, 1991.

Gallagher, R., "Information Theory and Reliable Communication," Wiley, 1968.

Coding Theory

Blahut, R. E., "Theory and Practice of Error Control Codes," Addison-Wesley, 1983.

Clark, G. C., Jr., and J. B. Cain, "Error-correction Coding for Digital Communications," Plenum Press, 1981.

Lin, S., and D. J. Costello, "Error Control Coding," Prentice-Hall, 1983.

Digital Data Transmission

Barry, J. R., D. G. Messerschmitt, and E. A. Lee, "Digital Communications," 3rd ed., Kluwer, 2003.

Proakis, J. G., "Digital Communications," 4th ed., McGraw-Hill, 2000.

CHAPTER 1.1
COMMUNICATION SYSTEMS

Geoffrey C. Orsak, H. Vincent Poor, John B. Thomas

CONCEPTS

The principal problem in most communication systems is the transmission of information in the form of messages or data from an originating *information source S* to a *destination* or *receiver D*. The method of transmission is frequently by means of electric signals under the control of the sender. These signals are transmitted via a channel C, as shown in Fig. 1.1.1. The set of messages sent by the source will be denoted by $\{U\}$. If the channel were such that each member of U were received exactly, there would be no communication problem. However, because of channel limitations and noise, a corrupted version $\{U^*\}$ of $\{U\}$ is received at the information destination. It is generally desired that the distorting effects of channel imperfections and noise be minimized and that the number of messages sent over the channel in a given time be maximized.

These two requirements are interacting, since, in general, increasing the rate of message transmission increases the distortion or error. However, some forms of message are better suited for transmission over a given channel than others, in that they can be transmitted faster or with less error. Thus it may be desirable to modify the message set $\{U\}$ by a suitable *encoder E* to produce a new message set $\{A\}$ more suitable for a given channel. Then a decoder E^{-1} will be required at the destination to recover $\{U^*\}$ from the distorted set $\{A^*\}$. A typical block diagram of the resulting system is shown in Fig. 1.1.2.

SELF-INFORMATION AND ENTROPY

Information theory is concerned with the quantification of the communications process. It is based on probabilistic modeling of the objects involved. In the model communication system given in Fig. 1.1.1, we assume that each member of the message set $\{U\}$ is expressible by means of some combination of a finite set of symbols called an *alphabet*. Let this source alphabet be denoted by the set $\{X\}$ with elements $x_1, x_2, \ldots, x_M$, where M is the size of the alphabet. The notation $p(x_i)$, $i = 1, 2, \ldots, M$, will be used for the probability of occurrence of the ith symbol x_i. In general the set of numbers $\{p(x_i)\}$ can be assigned arbitrarily provided that

$$p(x_i) \geq 0 \qquad i = 1, 2, \ldots, M \tag{1}$$

and

$$\sum_{i=1}^{M} p(x_i) = 1 \tag{2}$$

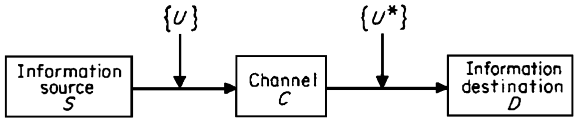

FIGURE 1.1.1 Basic communication system.

A measure of the amount of information contained in the ith symbol x_i can be defined based solely on the probability $p(x_i)$. In particular, the *self-information* $I(x_i)$ of the ith symbol x_i is defined as

$$I(x_i) = \log 1/p(x_i) = -\log p(x_i) \tag{3}$$

This quantity is a decreasing function of $p(x_i)$ with the endpoint values of infinity for the impossible event and zero for the certain event.

It follows directly from Eq. (3) that $I(x_i)$ is a discrete random variable, i.e., a real-valued function defined on the elements x_i of a probability space. Of the various statistical properties of this random variable $I(x_i)$, the most important is the expected value, or mean, given by

$$E\{I(x_i)\} = H(X) = \sum_{i=1}^{M} p(x_i)I(x_i) = -\sum_{i=1}^{M} p(x_i)\log p(x_i) \tag{4}$$

This quantity $H(X)$ is called the *entropy* of the distribution $p(x_i)$. If $p(x_i)$ is interpreted as the probability of the ith state of a system in phase space, then this expression is identical to the entropy of statistical mechanics and thermodynamics. Furthermore, the relationship is more than a mathematical similarity. In statistical mechanics, entropy is a measure of the disorder of a system; in information theory, it is a measure of the uncertainty associated with a message source.

In the definitions of self-information and entropy, the choice of the base for the logarithm is arbitrary, but of course each choice results in a different system of units for the information measures. The most common bases used are base 2, base e (the natural logarithm), and base 10. When base 2 is used, the unit of $I(\cdot)$ is called the *binary digit* or *bit*, which is a very familiar unit of information content. When base e is used, the unit is the *nat*; this base is often used because of its convenient analytical properties in integration, differentiation, and the like. The base 10 is encountered only rarely; the unit is the *Hartley*.

ENTROPY OF DISCRETE RANDOM VARIABLES

The more elementary properties of the entropy of a discrete random variable can be illustrated with a simple example. Consider the binary case, where $M = 2$, so that the alphabet consists of the symbols 0 and 1 with probabilities p and $1 - p$, respectively. It follows from Eq. (4) that

$$H_1(X) = -[p \log_2 p + (1 - p) \log_2 (1 - p)] \text{ (bits)} \tag{5}$$

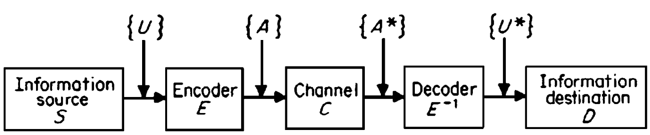

FIGURE 1.1.2 Communication system with encoding and decoding.

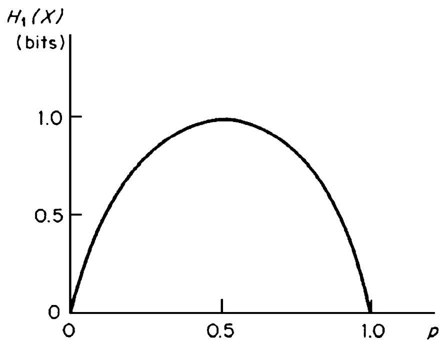

FIGURE 1.1.3 Entropy in the binary case.

Equation (5) can be plotted as a function of p, as shown in Fig. 1.1.3, and has the following interesting properties:

1. $H_1(X) \geq 0$.
2. $H_1(X)$ is zero only for $p = 0$ and $p = 1$.
3. $H_1(X)$ is a maximum at $p = 1 - p = 1/2$.

More generally, it can be shown that the entropy $H(X)$ has the following properties for the general case of an alphabet of size M:

1. $H(X) \geq 0$. (6)
2. $H(X) = 0$ if and only if all of the probabilities are zero except for one, which must be unity. (7)
3. $H(X) \leq \log_b M$. (8)
4. $H(X) = \log_b M$ if and only if all the probabilities are equal so that $p(x_i) = 1/M$ for all i. (9)

MUTUAL INFORMATION AND JOINT ENTROPY

The usual communication problem concerns the transfer of information from a source S through a channel C to a destination D, as shown in Fig. 1.1.1. The source has available for forming messages an alphabet X of size M. A particular symbol x_1 is selected from the M possible symbols and is sent over the channel C. It is the limitations of the channel that produce the need for a study of information theory.

The information destination has available an alphabet Y of size N. For each symbol x_i sent from the source, a symbol y_j is selected at the destination. Two probabilities serve to describe the "state of knowledge" at the destination. Prior to the reception of a communication, the state of knowledge of the destination about the symbol x_j is the a priori probability $p(x_i)$ that x_i would be selected for transmission. After reception and selection of the symbol y_j, the state of knowledge concerning x_i is the conditional probability $p(x_i|y_j)$, which will be called the a posteriori probability of x_i. It is the probability that x_i was sent given that y_j was received. Ideally this a posteriori probability for each given y_j should be unity for one x_i and zero for all other x_i. In this case an observer at the destination is able to determine exactly which symbol x_i has been sent after the reception of each symbol y_j. Thus the uncertainty that existed previously and which was expressed by the a priori probability distribution of x_i has been removed completely by reception. In the general case it is not possible to remove all the uncertainty, and the best that can be hoped for is that it has been decreased. Thus the a posteriori probability $p(x_i|y_j)$ is distributed over a number of x_i but should be different from $p(x_i)$. If the two probabilities are the same, then no uncertainty has been removed by transmission or no information has been transferred.

Based on this discussion and on other considerations that will become clearer later, the quantity $I(x_i; y_j)$ is defined as the information gained about x_i by the reception of y_j, where

$$I(x_i; y_j) = \log_b [p(x_i|y_j)/p(x_i)] \tag{10}$$

This measure has a number of reasonable and desirable properties.

Property 1. The information measure $I(x_i; y_j)$ *is symmetric in* x_i and y_j; that is,

$$I(x_i; y_j) = I(y_j; x_i) \tag{11}$$

Property 2. The mutual information $I(x_i; y_j)$ is a maximum when $p(x_i|y_j) = 1$, that is, when the reception of y_j completely removes the uncertainty concerning x_i:

$$I(x_i; y_j) \leq -\log p(x_i) = (x_i) \tag{12}$$

Property 3. If two communications y_j and z_k concerning the same message x_i are received successively, and if the observer at the destination takes the a posteriori probability of the first as the a priori probability of the second, then the total information gained about x_i is the sum of the gains from both communications:

$$I(x_i; y_j, z_k) = I(x_i; y_j) + I(x_i; z_k | y_j) \tag{13}$$

Property 4. If two communications y_j and y_k concerning two *independent* messages x_i and x_m are received, the total information gain is the sum of the two information gains considered separately:

$$I(x_i, x_m; y_j, y_k) = I(x_i; y_j) + I(x_m; y_k) \tag{14}$$

These four properties of mutual information are intuitively satisfying and desirable. Moreover, if one begins by requiring these properties, it is easily shown that the logarithmic definition of Eq. (10) is the simplest form that can be obtained.

The definition of mutual information given by Eq. (10) suffers from one major disadvantage. When errors are present, an observer will not be able to calculate the information gain even after the reception of all the symbols relating to a given source symbol, since the same series of received symbols may represent several different source symbols. Thus, the observer is unable to say which source symbol has been sent and at best can only compute the information gain with respect to each possible source symbol. In many cases it would be more desirable to have a quantity that is independent of the particular symbols. A number of quantities of this nature will be obtained in the remainder of this section.

The mutual information $I(x_i; y_j)$ is a random variable just as was the self-information $I(x_i)$; however, two probability spaces X and Y are involved now, and several ensemble averages are possible. The *average mutual information* $I(X; Y)$ is defined as a statistical average of $I(x_i; y_j)$ with respect to the joint probability $p(x_i; y_j)$; that is,

$$I(X;Y) = E_{XY}\{I(x_i;\, y_j)\} = \sum_i \sum_j p(x_i,\, y_j) \log[p(x_i|y_j)/p(x_i)] \tag{15}$$

This new function $I(X; Y)$ is the first information measure defined that does not depend on the individual symbols x_i or y_j. Thus, it is a property of the whole communication system and will turn out to be only the first in a series of similar quantities used as a basis for the characterization of communication systems. This quantity $I(X; Y)$ has a number of useful properties. It is nonnegative; it is zero if and only if the ensembles X and Y are *statistically independent*; and it is symmetric in X and Y so that $I(X; Y) = I(Y; X)$.

A source entropy $H(X)$ was given by Eq. (4). It is obvious that a similar quantity, the destination entropy $H(Y)$, can be defined analogously by

$$H(Y) = -\sum_{j=1}^{N} p(y_j) \log p(y_j) \tag{16}$$

This quantity will, of course, have all the properties developed for $H(X)$. In the same way the *joint or system entropy* $H(X, Y)$ can be defined by

$$H(X,\, Y) = -\sum_{i=1}^{M} \sum_{j=1}^{N} p(x_i,\, y_j) \log p(x_i,\, y_j) \tag{17}$$

If X and Y are *statistically independent* so that $p(x_i, y_j) = p(x_i)p(y_j)$ for all i and j, then Eq. (17) can be written as

$$H(X, Y) = H(X) + H(Y) \tag{18}$$

On the other hand, if X and Y are not independent, Eq. (17) becomes

$$H(X, Y) = H(X) + H(Y|X) = H(Y) + H(X|Y) \tag{19}$$

where $H(Y|X)$ and $H(X|Y)$ are *conditional entropies* given by

$$H(Y|X) = -\sum_{i=1}^{M}\sum_{j=1}^{N} p(x_i, y_j)\log p(y_j|x_i) \tag{20}$$

and by

$$H(X|Y) = -\sum_{i=1}^{M}\sum_{j=1}^{N} p(x_i, y_j)\log p(x_i|y_j) \tag{21}$$

These conditional entropies each satisfies an important inequality

$$0 \le H(Y|H) \le H(Y) \tag{22}$$

and

$$0 \le H(X|Y) \le H(X) \tag{23}$$

It follows from these last two expressions that Eq. (15) can be expanded to yield

$$I(X; Y) = -H(X, Y) + H(X) + H(Y) \ge 0 \tag{24}$$

This equation can be rewritten in the two equivalent forms

$$I(X; Y) = H(Y) - H(Y|X) \ge 0 \tag{25}$$

or

$$I(X|Y) = H(X) - H(X|Y) \ge 0 \tag{26}$$

It is also clear, say from Eq. (24), that $H(X, Y)$ satisfies the inequality

$$H(X, Y) \le H(X) + H(Y) \tag{27}$$

Thus, the joint entropy of two ensembles X and Y is a maximum when the ensembles are independent.

At this point it may be appropriate to comment on the meaning of the two conditional entropies $H(Y|X)$ and $H(X|Y)$. Let us refer first to Eq. (26). This equation expresses the fact that the average information gained about a message, when a communication is completed, is equal to the average source information less the average uncertainty that still remains about the message. From another point of view, the quantity $H(X|Y)$ is the average additional information needed at the destination after reception to completely specify the message sent. Thus, $H(X|Y)$ represents the information lost in the channel. It is frequently called the *equivocation.* Let us now consider Eq. (25). This equation indicates that the information transmitted consists of the difference between the destination entropy and that part of the destination entropy that is not information about the source; thus the term $H(Y|X)$ can be considered a *noise entropy* added in the channel.