The Lenkurt® Demodulator

VOL. 14, NO. 4 APRIL, 1965

INFORMATION THEORY and CODING

Part 1

Information theory began as an academic mathematical inquiry into factors which limit communication. The fantastic growth of man's ability to communicate in the past several decades has confirmed the importance of the inquiry and added to pressures of many important practical problems. Billions of dollars have been spent on communications in recent years and will be spent again to keep up with public demand. Improving the efficiency of communication facilities could channel these funds into providing even greater advances. This article is the first of two which discuss some of the highlights of information theory and how it is used to improve communications.

Transmitting information by various communications techniques is an important part of everyday life. Certainly everyone has used the phrase "a lot of information," but few people regard the fact that it is possible to measure information quantitatively. However, information has been given a numerical value that is very useful in the study of communications.

What is the importance of information theory to a communications engineer? It seems quite reasonable that an engineer who is responsible for selecting solid-state radio equipment for a communications system should have a knowledge of the theory of transistors. It is very unlikely that such an engineer will have anything to do with the design of transistors, but a knowledge of basic transistor theory is nevertheless required for the engineer to be adequately qualified to do his job. In other words, a great deal of fundamental background knowledge is necessary for the broad understanding required by a competent engineer. For the communications engineers, information theory is becoming an increasingly important part of this background knowledge.

The accelerating growth of data transmission creates an obvious requirement for a broader understanding of information theory. In addition to other uses, information theory provides the fundamental principle for analyzing and comparing existing and future data transmission systems.

Unfortunately, many of the improvements in communications systems suggested by information theory are rather complicated and expensive, thus preventing them from being readily put into use. Nevertheless, as technology progresses, cheaper means of performing the complicated operations suggested by the theory will undoubtedly be found and information theory can be expected to play a more important role in practical communication systems of the future.

Meaning of Information

Information can perhaps be explained as *choice* or *uncertainty*. The effect of the information in a message is to change the probability concerning a situation, as far as the receiver of the message is concerned, from its value before the message is received to what is usually a larger value after the message is received. If an event is certain to occur, the mathematical probability of its occurrence is, by definition, one. If the event is certain not to occur, the mathematical probability is zero. The mathematical probability of the occurrence of any event whose occurrence or nonoccurrence cannot be predicted with certainty lies somewhere between zero and one.

One of the first steps in determining the exact nature of information was the selection of a unit, or yardstick, by which information could be measured. This unit had to be such that it could easily be determined and did not depend upon the *importance* of the message, since a message's importance is difficult to evaluate mathematically.

It turned out that the simplest and most basic unit was the amount of information necessary for a receiver (person or machine) to make the correct choice between two equally possible messages. This choice may be between the messages yes-or-no, on-or-off, A-or-B, 0-or-1, black-or-white, and so on. Since the two possible messages correspond to the two symbols in the binary number system, a unit of information based on two symbols (messages) came to be called a *binary digit* and was abbreviated bit.

A message consisting of one simple electrical pulse has the informational value of one bit because the presence or absence of the pulse permits the receiver to choose the correct message from a set of two. As shown in Figure 1, transmitting two pulses permits the receiver to select the correct message from four equally possible messages. Three pulses, or bits, will enable the correct selection from a set of eight. This selection process gives the average amount of information which must be transmitted to specify a message from a set of equal possibilities.

For instance, suppose a message is to be sent indicating a choice in an election among eight candidates. In the form of communication, the sender has a certain definite limited choice as to what message to send. The receiver of the message will have some uncertainty as to what it will say but he knows that there are only eight possible choices. In order for the sender to indicate to the receiver which candidate he has chosen, he must use some sort of signal.

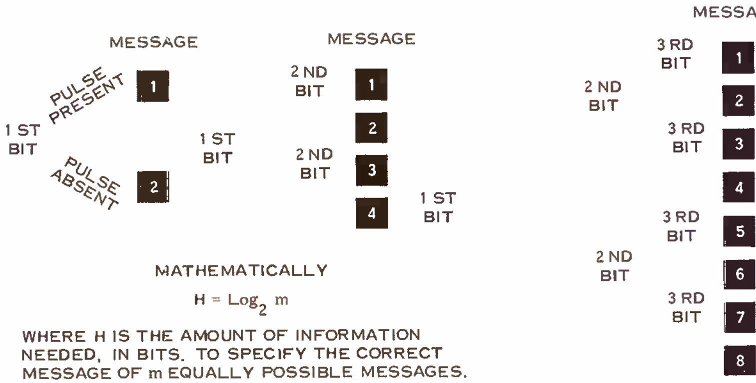


Figure 1. Amount of information required to specify a message from a set of equal possibilities.

One form of signal might be a series of pulses or absence of pulses (for convenience, a *one* can be used to represent a pulse and a *zero* to represent the absence of a pulse). Since there are eight unique series of three pulses or no-pulses possible, any one of the messages can be designated by one series of three pulses or no-pulses. For the receiver, three elementary decisions decide which message among eight was intended.

The information content of a message, expressed in bits, is determined by the formula:

$$H = \log_2 m$$

where

H = number of bits of information
m = number of equally likely choices

If m is 8, the first bit corresponds to a choice of which half of the 8 possibilities is chosen, the second bit to a choice between the first and second pair of the selected half, and the last bit to a choice between the first or second member of the chosen pair. Thus 3 (or $\log_2 8$) bits of information determine the selection, and this is the amount of information acquired by the receiver.

In the example cited, each pulse provides one bit of information. In binary code each code element may be either of two distinct kinds of values: for example, the presence or absence of a pulse. In a ternary or three level code, each code element may be any of three distinct kinds or values; in an N'ary code, each code element may be any one of N distinct kinds or values.

With a simple N'ary code, if all values are equally probable and the probability of any code element is independent of preceding code elements, the amount of information (H) is proportional to the number of code elements (n) comprising the message multiplied by the logarithm of N.

$$H = n \log_2 N$$

The average information per symbol in any of the commonly used sets of symbols or in any language is always considerably less than its maximum possible value. In effect, this means that parts of messages usually tell things which are already partly known. Thus the intersymbol influences (including interword influences) can predict the nearby succeeding parts of a message to a considerable extent. The actual reception of the message then gives partly a verification or correction of the prediction in addition to completely new ideas. This partial or complete repetition of message content which occurs in languages is called *redundancy*. However, despite the fact that it causes a loss in the rate of transmitting information, redundancy is a very useful property of languages, for it allows individual errors in the transmission of messages to be recognized easily and corrected.

Message Sources

In information theory, message sources are classified as either *discrete* or *non-discrete*. A speaker is an example of a non-discrete message source since the values of a speech wave are drawn from a continuum of possibilities. Such non-discrete sources are very difficult to evaluate mathematically. For this reason, the application of information theory to communications systems is concerned mainly with so-called discrete sources. A discrete source produces messages which are sequences of symbols, the symbols being drawn from some finite list. The most familiar example of such messages is printed English text. The sequence of telegrams passed to the telegrapher for transmission can be thought of as such a source of English text. Other examples of discrete information sources are the input tape to a large computer or the string of symbols printed on a stock exchange ticker-tape.

In analogy with English, the different symbols of a message from any discrete source can be called *letters,* and the finite list of letters from which messages are composed are called the source *alphabet*. The number of letters in the source alphabet represent the *size* of the alphabet. The occurrence of a letter in a message is called a *character* regardless of what letter it is. For example, the word *Mississippi* contains eleven characters but only four different letters.

The first step in describing a given information source is to list its alphabet. This is far from a complete description of the source, however, for the messages produced by most sources have an elaborate statistical structure. The character being printed now is not independent of characters just produced by the source, but depends upon them in a complicated way. What the next character produced by a source will be is not certain. A good guess as to the next character, however, depends strongly on how much of the past message has already been received. If the source is a telegraph and the message "General Eisenho" is observed, then it is quite certain what letter the next character will be. If, however, only the last letter of the message, "o", is observed, what will follow is not so clear.

The statistical structure of the messages produced by a given source can be described mathematically by associating with the source a long list of probabilities. The first probabilities on this list are quantities p_i—the probability that the source will produce the ith letter of the source alphabet. These

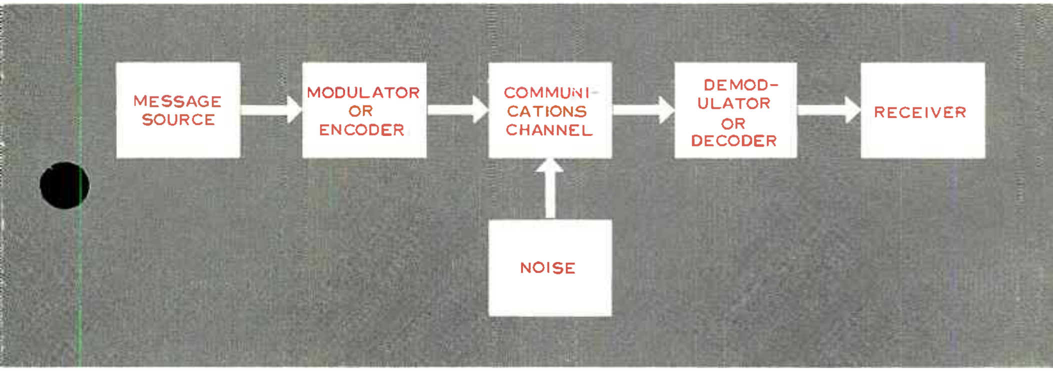


Figure 2. Generalized communications system used in the study of information theory.

quantities reflect the best guess as to what letter the source will produce when the text already produced has not been seen. The next set of characterizing numbers for the source are quantities p_{ij}—the conditional probability that the source will next produce the jth letter of the alphabet when it is known to have just produced the ith letter. Next are the quantities p_{ijk}—the probability that the source will next produce the kth letter of the alphabet when it is known that the source has just produced the ith letter followed by the jth letter. Listing such probabilities in this manner can continue indefinitely, each set giving more information about the long range structure of the messages. In the mathematical model of a message source used in *information theory*, this list of probabilities, along with the source alphabet, characterizes a particular information source.

In actual sources the infinitely remote past of a message certainly exerts no influence on characters being printed now. In fact, in many sources correlation between characters does not extend very far into the past at all. Sources may be classified then according to the number of past characters that exert an influence on the character being produced by the source at the present moment. A source in which the past characters exert no influence on the present character is called a *monogram* source; one in which only the last character produced influences the choice of the present character is called a *digram* source, and so on.

Information Content of English Text

As previously stated, information can be measured in terms of bits or informative yes's or no's, and an element of a binary code contains one bit of information. For the more complex or non-binary codes the information content of each code element increases. If there is an equal possibility of any code element appearing in a sequence, the information value of each code element is equal to the logarithm to the base two of the possible number of code elements.

In written English, there is a finite set composed of 27 symbols, the 26 letters of the alphabet and a word space. If the next letter in a written message appeared at random with a probability of 1/27 for each letter, then the information content of a message expressed in bits would merely be equal to the number of symbols multiplied by $\log_2 27$, using the formula previously given.

However, when the information in a message is in the form of a language, occurrence of the various symbols which comprise the language's alphabet is never completely random. Thus, the appearance of a given letter or a given word in English is subject to "constraints" which act to modify an otherwise completely random probability of occurrence.

In setting up his dot-dash code, Morse made one of the first applications of statistics to a communication problem. On the basis of type counts made in a printing shop, Morse assigned a short code to the most frequently used letters and longer codes to the less frequent. Thus, he could transmit E, the most frequent, by simply sending a dot, but for V, one of the least frequent, he had to send dot dot dot dash.

Thus, Morse would have expected to use more time transmitting letters of gibberish, which might use V as frequently as E, than sensible English in which letters appeared with their familiar frequency.

To get the most possible combinations from its alphabet, a language should allow its letters to fall with uniform probability. Constraints on where the letters fall serve to introduce substantial redundancy in the transmission of information.

As an example, consider the number of bits of information that are con-

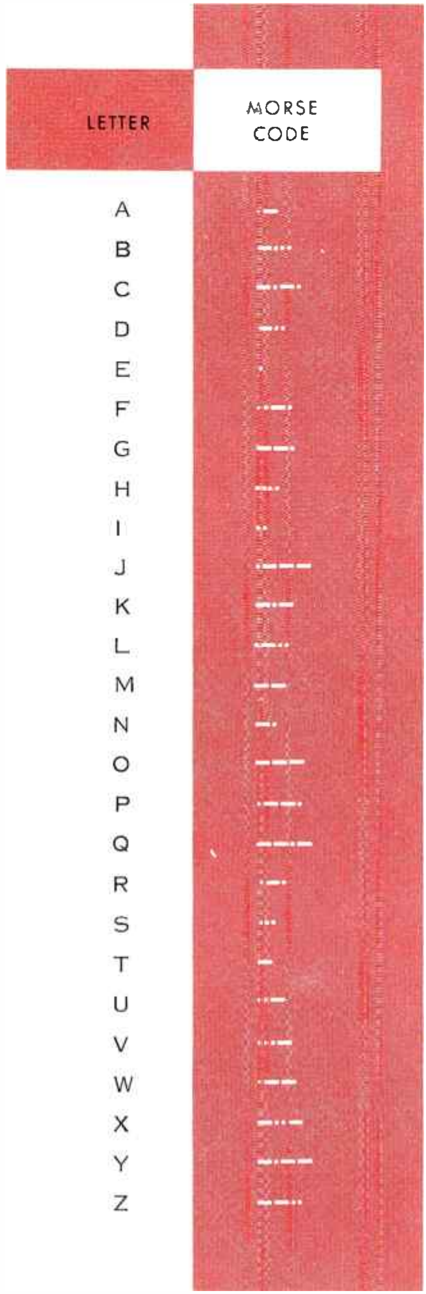


Figure 3. The familiar Morse code was one of the first applications of statistics to a communications problem.

tained in a letter of the English alphabet. If any letter of a 27 letter alphabet were equally probable, the information in one letter would be the logarithm to the base of 2 of 27 or 4.76 bits. Actually, since all letters are not equally probable, when the known probabilities are applied for each letter, it develops that, on the average, each letter probably contains less than two bits. Putting it another way, an 18 letter alphabet of uniform probability could do the same job as our less efficient 27 letter alphabet. However, the redundancy of the English language permits great liberties in transmitting written messages. For example, the telegraph message

PLLESE SXND MONEZ

can easily be interpreted as

PLEASE SEND MONEY

Transmitting Information

If there were no noise to degrade transmission, there would be no limit to information transmission. By transmitting a perfectly measured voltage to represent information, for example, any desired rate of communication could be achieved. In reality, noise masks signals transmitted over communication circuits and introduces uncertainty as to their exact value. Signals tend to be converted into noise by a process of degradation and distortion. In transmitting many channels of information over a multiplex system, each channel requires a certain bandwidth in order to distinguish the signal from random noise. Thus, the greater the number of channels, the greater the bandwidth required. However, as the number of channels increases, each channel signal represents a smaller and smaller portion of the total band. As this occurs, it becomes increasingly difficult to distinguish the signals from background noise unless transmission power is increased.

Information theory studies have revealed the exact relationship between information capacity, signal power, noise, and bandwidth. While these studies have generally confirmed knowledge acquired on an experimental basis, a number of possibilities were revealed that had not been self-evident. It was well known that a smaller signal-to-noise ratio would be acceptable in communications if greater bandwidth were employed, as in FM. It was surprising, however, to discover that in principle, bandwidth could be reduced by increasing signal-to-noise ratio. Heretofore, it was firmly believed that channel bandwidth could never be less than the bandwidth of the original message.

H. Nyquist, a mathematician at the Bell Telephone Laboratories, proved mathematically that the required bandwidth for a communications channel is directly proportional to signaling speed, and that the minimum bandwidth required for transmission of a signal is essentially equal to half the number of binary pulses per second.

Nyquist showed that although there was a limit to the number of pulses per second that could be transmitted over a given communications channel, each pulse might have several distinguishable states or conditions, each of which could carry information. Thus, if amplitude were the variable conveying the information, and each pulse had four possible amplitudes, twice as much information could be transmitted as in a system where pulses had only two possible values.

Nyquist showed that the limit to the number of information-carrying states was related to the noise in the circuit. As stated previously, without noise, there would be no limit to the rate at which information could be transmitted. In the presence of noise, however, the difference in value between two levels or states must be at least twice the value of peak noise. Otherwise, there will be uncertainty as to the value of the pulse.

The same limitation applies to continuous waveforms as well as pulse signals. Actually, there is no real difference between the two. Although a continuous wave may contain an infinite number of points which define its shape, it does not contain an infinite number of information-carrying values. In fact, periodic samples of the waveform can be used to reconstruct or define the waveform perfectly if they are taken often enough. The waveform doesn't have to be sampled very often to make a perfect reconstruction—sampling at twice the highest useful frequency in the signal will do. Thus, if 3000 cps is the highest useful frequency in a telephone channel, a series of brief samples taken at the rate of 6000 per second will precisely and exactly duplicate the telephone conversation! The samples can be as brief as desired, in fact, the shorter the better. Thus, a series of pulses can serve in lieu of a continuous waveform, with no loss whatsoever.

The 3000 cps telephone circuit is a universal communications channel, available almost anywhere in the world. Almost all general purpose communications facilities are designed to accommodate voice signals. Accordingly, this bandwidth has been taken into consideration in designing equipment used to transmit telegraph and data signals and other forms of information.

According to Nyquist's formula for maximum signaling speed, a 3000 cycle channel should be capable of carrying 6000 binary pulses per second. Translated to words per minute, and using the standard Baudot or teletypewriter code, this is approximately 8000 words per minute. Furthermore, the information capacity is considerably higher if codes other than binary are used. The relationship between bandwidth, signal power, and noise is complex and depends upon many factors such as the kind of noise present in the channel, the nature of the power limitation, the type of modulation used, and the method of encoding the information. In 1948, C. E. Shannon, also of the Bell Telephone Laboratories, devised a mathematical formula which defined the *capacity* of a communications channel or the maximum transmission rate. This formula, which relates information rate to the bandwidth and the amount of interfering noise in the system, is shown graphically in Figure 4.

Using Shannon's formula, a channel of 3000 cycles bandwidth and a signal-to-noise ratio (signal power/noise power) of 30 db has a capacity (C) of about 30,000 bits per second:

$$\begin{aligned} C &= W \log_2 (1 + \frac{S}{N}) \\ &= 3000 \log_2 1001 \\ &= 3000\ (9.96) \\ &= 29{,}880 \text{ bits per second} \end{aligned}$$

Of course, this is ideal, non-surpassable performance, and achievable only by the most elaborate coding. Practical communications systems cannot begin to approach that rate of transmission. To achieve such a rate, three conditions would have to be met. First, the transmission medium must be distortionless. Second, the noise power in the channel must be equal throughout the frequency band. Third, the method used to encode the signals must be so complex that no possible combination of noise impulses will ever cause errors to occur during transmission. None of these conditions can be met or even closely achieved with present-day techniques.

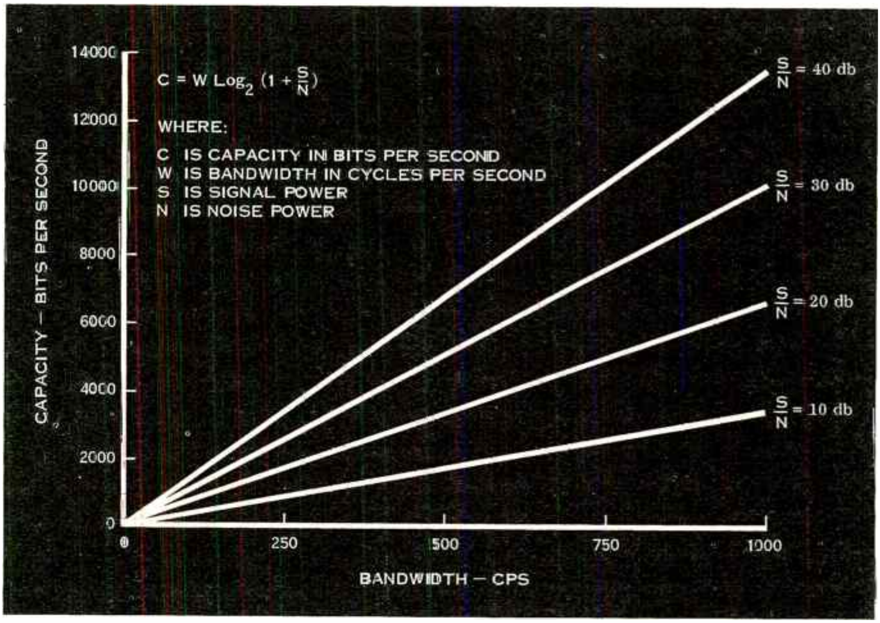


Figure 4. Graph of Shannon's formula which relates information rate to the bandwidth and the interfering noise in a communications system.

Such performance would require whole buildings-full of equipment to encode and decode the message. In addition, the time required for encoding and decoding messages would be far too great for practical needs.

Information theory studies have indicated the existence of ideal codes for transmission over the noisiest channel at rates up to the theoretical limit, and permitting as low a probability of transmission error as required. Redundancy or repetition reduces error probability. By introducing a controlled amount of redundancy in proportion to the channel noise, the desired transmission reliability can be maintained under the worst conditions, at the cost of reduced transmission rate. If reduced transmission rate cannot be tolerated, more complicated coding can theoretically be employed to keep down the probability of error. The subject of coding and error control is discussed in Part 2.

GLOSSARY

Some of the terms often encountered in the fields of communications, information theory, and computer engineering are listed below with their specialized meanings.

BINARY—Anything that is composed of two parts or elements, or which has only two states or conditions; for example, a switch may be either on or off.

BINARY CODE—Any *code* employing only two distinguishable *code elements* or states; mark-space and on-off are examples of *binary codes.*

BINARY DIGIT—A unit of information content; one element or *bit* of a binary or two-element code; mark and space are examples of *binary digits* used in communications codes.

BIT—Commonly used short form of *bi*nary digi*t*: see above.

CHANNEL CAPACITY — The maximum possible information rate through a channel. *Channel capacity* is often stated as the maximum number of *bits* per second that may be transmitted through the channel.

CLOCK—The primary source of synchronizing signals for a computer data system or data transmission system. In high-speed systems, the *clock* may be an oscillator, the output of which is used as a reference or timing frequency by the system.

CODE—A system for representing that which is to be transmitted. For example, Morse *code* may be used to represent letters in telegraphy. Words and language may be considered a way of *coding* ideas.

CODE CHARACTER—One of the elements which make up a *code* and which represents a specific symbol or value to be encoded. Dot-dot-dot-dash is the Morse *code character* for the letter *v*.

CODE ELEMENT—One of a set of parts of which *code characters* may be composed. Mark or space, dash or dot, are examples of *code elements.*

COMPUTER—A machine for carrying out calculations, or for performing specified changes or transformations of *information.* Some types of decoders operate on *computer* principles.

DATA—Information, usually originating as numbers, values, or digital symbols. *Data* usually excludes speech, music, or other continuous-wave information, even when converted to digital form for transmission.

ENTROPY—Disorder; reduction from an easily distinguishable condition to a less easily distinguishable condition; equal and opposite to "amount of information." An example of entropy is the gradual distortion and loss of a signal pulse as it travels down a long wire or cable.

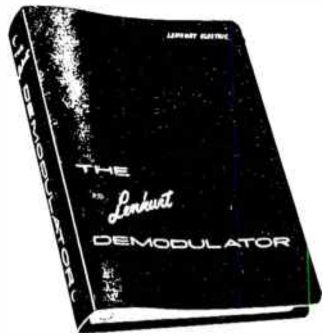
ERROR—A false transformation of *information;* a mistake in transmission; improper alteration of *information;* an incorrect step, process, or result in transmitting *information.*

INFORMATION—News, intelligence; that which improves or adds to a representation. *Information* is an unpredictable event or item. If the transmission specifies something already known to the receiver, it is not *information.*

REDUNDANCY — Added or repeated *information* employed to reduce ambiguity or *error* in a transmission of *information.* As signal-to-noise ratio decreases, *redundancy* may be employed to prevent an increase in transmission *error.*

QUANTIZE—To convert a continuous variable, such as a waveform, into a series of levels or steps. There are no "in between" values in such a *quantized* waveform. All values of signal are represented by the nearest standard value.

Binder



A special binder is available for storing issues of **The Lenkurt Demodulator.** These binders are covered in durable, embossed Lexide; each binder holds 24 issues. The cost is 60 cents for one binder, or 50 cents each for two or more, postpaid.

Reprint Book

The 32 most-requested articles from the first seven years of **The Lenkurt Demodulator** have been compiled into book form.

The attractive, cloth-bound book is titled **Carrier and Radio Articles Selected from The Lenkurt Demodulator**, and costs $2.50, postpaid.

Lenkurt Electric Co., Inc.
San Carlos, California

Bulk Rate
U.S. Postage
Paid
San Carlos, Calif.
Permit No. 37

MR. E. A. GARCIA
CHIEF OPERATOR
% MILWAUKEE ROAD DEPOT
SAVANNA, ILL.

R-24948 6B

RETURN REQUESTED

Improved Data Set

Lenkurt's improved 26C Data Set offers a highly accurate means of transmitting serialized digital data over standard 3-kc voice channels at speeds of 1200 or 2400 bits per second (bps). Transmission at 2400 bps is achieved through the use of Lenkurt's unique Duobinary coding technique. Automatic error detection—plus an extremely low error rate—are bonus features of the Duobinary technique. More information about this small-size, flexible data set is contained in Form 26C-P4, available on request from

THE EDITOR
The Lenkurt Demodulator

LENKURT ELECTRIC San Carlos, California, U.S.A.
SUBSIDIARY OF
GENERAL TELEPHONE & ELECTRONICS GT&E

Lenkurt Offices

San Carlos | New York City
Atlanta | Washington, D. C.
Dallas | Chicago

The Lenkurt Demodulator is a monthly publication circulated free to individuals interested in multi-channel carrier, microwave radio systems, and allied electronic products. Permission to reproduce material from the *Demodulator* will be granted upon request. Please address all correspondence to the Editor.

LITHO IN USA