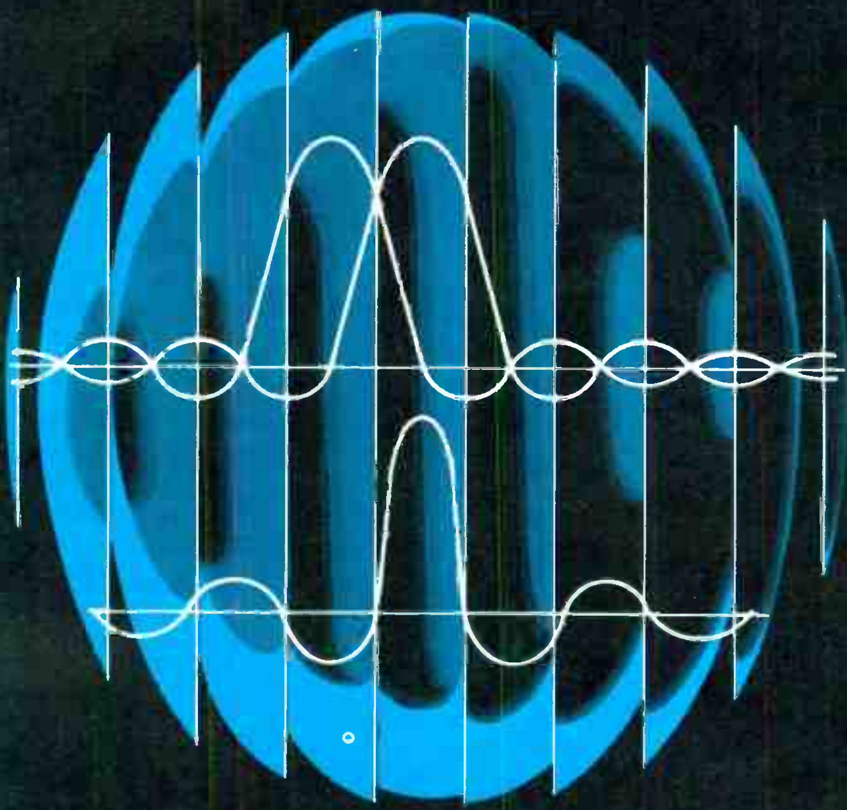
GTE LENKURT

DEMODULATOR

AUGUST 1974



data transmission: principles and problems

"The fundamental problem of communication is that of reproducing at one point either exactly or approximately a message selected at another point." C.E. Shannon

In moving from primitive to modern, man has consistently sought for ways to make tasks less difficult or tedious: this search has almost invariably resulted in the invention of machines. Communications "machines" have run the gamut from papyrus rolls and rune staffs to printed books and computer print-outs.

Data Language

Communication implies a language or symbology: almost without exception, the language employed in communication between machines consists of electrical signals, with the information conveyed in a digital form.

Digital symbology is based on the premise that all information can be conveyed in a two-state code. Samuel Morse applied this principle to the development of his code, which uses long and short transmission times as coding to represent the letters of the alphabet, numbers, and various symbols.

The two-state code concept is derived by considering information to be present only where there is doubt: that is, where a choice, selection, or discrimination is required. In other words, a message is considered to have the potential to communicate something, with the information communicated dependent upon the selection of one message from a set of messages. The simplest choice is between two equally possible messages, which may be yes-or-no, on-or-off, A-or-B, 0-or-1, or any other two-state condition. As an example, figure 1 presents a group of eight messages, 1 through 8. Assuming that one has been selected by a hypothetical source and signaled to a receiver, and that all of the messages are equally likely to have been selected, the question is, how much information must be conveyed for the receiver to identify the proper message. Choice 1 asks the question, "Is it in the first half of the group, yes or no?" (yes = 1, no = 0). Having thus eliminated half of the possibilities, choice 2 asks, "Is it in the first half of this half?" Similarly, choice 3 determines in which half of that half the selected message lies. Should message 6 be selected, it could be located by the following responses: no (0) to choice 1, yes (1) to choice 2, and no (0) to choice 3. In this way, three simple yes-or-no (1-or-0) choices have served to identify one unique message out of a set of eight.

Since the two possible messages correspond to the two symbols in the binary number system, the information content of a message based on two-state coding is measured in units called binary digits, or bits. Expressed in bits, information content can be determined by the formula:

$$H = \log_2 m$$

where

H = the number of bits of information

m = the number of likely choices

If m is 8 — as it is in the example of figure 1, where eight possible messages are presented — the formula shows

CHOICE 1	CHOICE 2	CHOICE 3	MESSAGE
1	1	1	1
1	1	0	2
1	0	1	3
1	0	0	4
0	1	1	5
0	1	0	6
0	0	1	7
0	0	0	8

Figure 1. Amount of information required to identify a unique message from a set of eight equal possibilities.

that $\log_2 8$, or 3, bits of information are required to isolate a given message, which is what was found in the example. Finding one word in a dictionary of a quarter of a million words would require at most 18 bits of information, since 2^{18} is greater than 250,000.

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 select the correct message from a set of two. Mechanical and electrical devices, which can be made to change states very rapidly — as from conduction to non-conduction — are ideally suited to application of this principle.

Communication Systems

A general communication system consists of five elements:

1. An information source, which selects a message from a set of possible messages.
2. A transmitter, which transforms the selected message into a signal that can be sent over the communication channel.
3. A communication channel. This is the physical medium over which the signal is transmitted, and may be air, as in oral speech, a pair of wires, a coaxial cable, a beam of light, or simply space with the signal in the form of an electromagnetic wave.
4. A receiver, which performs the inverse operation of the transmitter to reproduce the message from the signal.
5. A destination, which is the person or machine for which the message is intended.

In digital communication systems, the information source and destination are most commonly computers, tape terminals, or business machines. Data from the source are typically represented by binary signals (messages) at the input to the transmitter, which is generally a modulation or encoding device, whose purpose is to transform the binary signal into another digital signal that requires less channel bandwidth. In data applications, it is normal to refer to two channels: the modulation channel and the coding channel. While both are physically part of the same communication system, the modulation element is considered to be analog and to consist of the transmission medium and the physical transducers needed to couple the signal to the medium. The coding element is concerned with the digital data; it is a digital link which not only includes the modulation channel, but is also extended to incorporate the encoding and decoding devices.

Bandwidth Conservation

The need to use less channel bandwidth is due to the band limitations placed on transmission media. If there were no limits, rectangular binary pulses could be sent directly through a communication channel; in reality, however, bandwidth is limited and its conservation — through maximum utilization — is of extreme importance.

It was recognized as early as 1924 that each pulse in a data stream can be made to convey additional information if one or more of its characteristics is varied in a prescribed manner. In this way, not only the presence or absence of a pulse, but its phase, frequency, and amplitude levels as well, can be used to transmit information. The use of multilevel systems represents a significant improvement in bandwidth utilization over binary transmission since, for a given bandwidth, more information can be conveyed by making each pulse represent a group of binary digits, rather than a single digit as in binary systems.

However, increasing the number of levels in a signal while maintaining constant power makes that signal more sensitive to noise, and more inclined to produce errors; in addition, complex equipment is required to generate the multilevel pulses. Because of this, alternatives to multilevel systems have been sought which use fewer levels for the same bandwidth and bit speed. One method, developed by GTE Lenkurt, provides a two-to-one bandwidth compression relative to binary signaling; that is, the duobinary technique, as it is called, affords twice the transmission speed in bits per second as a binary system of the same fixed bandwidth. To achieve this with multilevel techniques, four-level coding would be required; a duobinary system does it with three levels, and the equipment needed is comparable in complexity to that of a straight binary system.

In the duobinary system, the encoding and decoding of a signal involves dealing with the algebraic sum, at the sample instant, of the value of the present digit and of interference from other digits. To accomplish this, a binary input signal is passed through a shaping filter having a transfer function designated $H_1(\omega)$, which delays the waveform one interval and adds it to itself. As a result, the original two-level signal is transformed into a 3-level signal formed by an algebraic addition of the present and previous pulses. For example, if three unipolar pulses are sent in succession at time intervals of T seconds as shown in figure 2, sampling the signal at odd multiples of T/2 seconds (T/2, 3T/2, etc.) allows the generation of pulses using interference from the adjacent pulse.

When pulse No. 1 is sampled at $t = T/2$, the value of pulse No. 2 at the same instant in time is not zero but is exactly the same as that of pulse No. 1; the same can be said for pulses No. 2 and No. 3 at $t = 3T/2$. Thus, overlapping of pulses is deliberately introduced at the sampling instants; at the receiving end, the algebraic sum is observed and interpreted. When both pulses are present, as at $t = T/2$ and $t = 3T/2$, the value of a sampling instant is $4/\pi$, or twice that of a single pulse. At $t = 5T/2$, only one pulse is present, so the value is $2/\pi$. Finally, when no pulse is present at $t = 7T/2$, the value of the sampling instant is zero. Consequently, the signal has three distinguishable levels at the sampling instants.

Limitations on signal bandwidth are not the only bonds on data transmission. As a rule, data communication utilizes existing common carrier telephone plants which were designed primarily for voice communication. The channels provided by these facilities, while adequate for their intended purposes, are afflicted with noise, linear distortion, non-linearities (such as harmonic distortion and spurious signals caused by system compander action), and the abrupt changes in channel characteristics known as phase and amplitude hits. These impairments modify the transmitted data signals, resulting in errors when the receiver reconstructs the message.

Noise in a communication channel is of two basic types: white and impulse — the latter appears as sharp bursts of energy arising from such

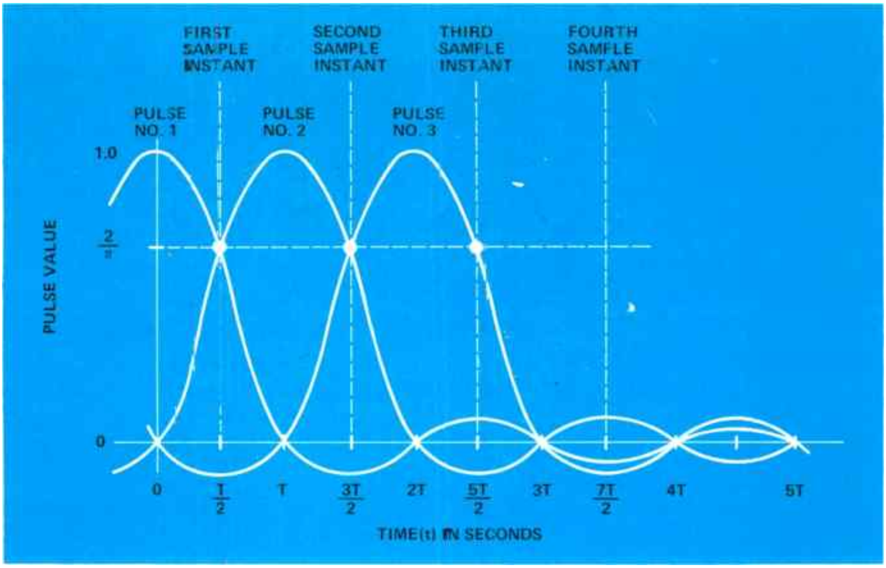


Figure 2. Superposition of pulses in the duobinary coding process results in the creation of a three-level signal with values of $4/\pi$ at sample instants $t = T/2$ and $3T/2$, $2/\pi$ at $t = 5T/2$, and zero at $t = 7T/2$.

sources as electrical storms and switching in telephone plants. White noise interferes more with speech than it does with binary data, because the amplitude of the data pulses can be set above the noise level; however, as the number of signal levels is increased for a fixed power, the noise margin decreases and white noise can become a channel capacity limiting factor. Impulse noise is a much greater problem in data transmission than white noise, because the short duration and high instantaneous amplitude of an impulse spike may be easily interpreted by the receiver as a data signal, resulting in an error in message reproduction.

Equalization

Of most concern in high-speed data transmission is a form of distortion called intersymbol interference, which is defined as the distortion caused by the tails of preceding and succeeding pulses intruding into the time slot of the pulse currently being transmitted.

Intersymbol interference results from the non-uniform delay and amplitude characteristics of communication channels. An equalizer – basically an adjustable filter – is used to introduce a controlled amount of delay at certain frequencies to achieve uniform delay and amplitude over the entire bandwidth. At lower transmission speeds (up to about 2400 bits per second), it is possible to take a statistical sampling of signals sent over the network and determine the average performance of the channel; it is then possible to design compromise equalizers that will compensate for average distortion characteristics. At higher speeds, however, more efficient utilization of bandwidth is required, as is equalization that is more precisely tailored to the particular channel in use during any one random connection. This is beyond the capability of fixed equalizers. To meet these requirements, equalizers have been developed that change their characteris-

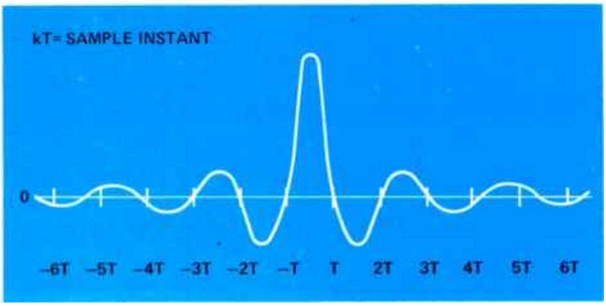


Figure 3. Output of a 6-tap zero-forcing equalizer. Zero-forcing equalization tries to make the pulse cross zero at sample instants corresponding to tap placement (±T, ±2T, ±3T).

ties automatically to fit those of the communications channel.

Automatic equalizers make use of a transversal filter – a tapped delay line whose taps are connected to gain control circuits. By varying the gains of the taps, the controlled amplitude and delay distortion introduced into the signal components can be set to compensate for the effects of the channel. In general, two types of automatic equalization can be distinguished: preset and adaptive. In preset equalization, test signals are used to adjust the tap gains prior to, or during breaks in, data transmission; a major shortcoming of this technique is that it requires the test pulses to be transmitted during a separate training time whenever the channel characteristics change, thus interrupting data flow. Adaptive equalization provides for constant monitoring of the data signal and adjustment of the tap gain settings to give optimum equalization at all times.

In practical adaptive equalization, data transmission is preceded by a short training period, during which test pulses allow determination of an initial equalization level and the setting of the tap gains accordingly. As data are transmitted, samples are taken and compared to the equalization level; any difference between channel characteristic – as revealed by the data sample – and equalized sample results in the generation of an error sample and the setting of the tap gains to cancel the difference.

Phase modulation has proved to be an effective and reliable technique for high-speed data transmission, and adaptive equalization techniques suitable for use with phase-modulated systems are becoming increasingly important in commercial applications. Implementation of these processes depends upon logic circuitry in the equalizer solving simultaneous equations. One procedure, based on a zero-forcing algorithm (an algorithm is a procedure for solving a mathematical problem in a finite number of steps), involves looking at the equalizer output pulse shape; by changing the tap

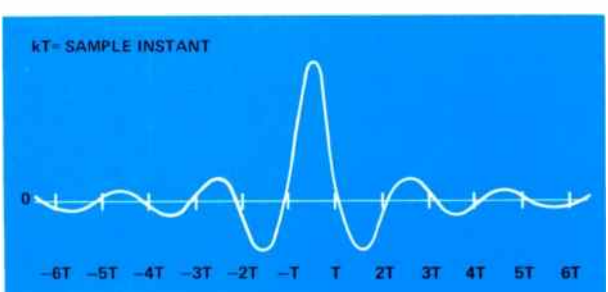


Figure 4. Output of a 6-tap mean-square equalizer. The pulse is not constrained to pass through zero; instead, the equalizer minimizes the sum of the square of the errors, resulting in less distortion over the entire bandwidth.

gain settings, the equalizer attempts to get the pulse to cross zero at the sample times (see figure 3). If, for example, three taps are placed on each side of the delay line center tap, the equalizer will try to force the output pulses to cross zero at three sampling instants on either side of the main pulse. Under heavy distortion conditions, however, there is the likelihood that, while forcing a certain number of pulse samples to cross zero, the samples not being looked at will be distorted.

Another procedure for equalizing phase-modulated systems was recently developed by GTE Lenkurt, utilizing an incoherent, mean-square algorithm. This technique, which is called incoherent because it does not require carrier phase or amplitude reference information, looks at all of the output pulse samples and attempts to minimize the mean square (rms) error, rather than forcing certain samples to cross zero (see figure 4). Each received signal sample is used to generate a reference to which the next succeeding sample is compared. Due to the use of the previous received samples for both phase and amplitude reference, operation of the equalizer is almost totally independent of frequency offset and phase jitter, and avoids the problem of establishing a reasonable initial phase reference in fast start-up systems. In actual use, this algorithm has allowed the analog components to be replaced by A/D converters and a digital processor.

The common carrier communications systems constitute a vast network of transmission facilities which, while designed primarily to handle speech signals, can be made to transmit digital signals. In transmitting data over such communication systems, accuracy is of utmost importance; the redundancy which allows significant interference in speech is missing in data, and channel characteristics that are acceptable for analog signals may alter a digital signal and produce errors in message reproduction. Solving the problems presented by channel distortion is therefore essential if the most efficient use of transmission facilities is to be realized.

BIBLIOGRAPHY

Cherry, C. *On Human Communication.* Cambridge, Mass.: The M.I.T. Press, 1966.

"Data Communications," *The Lenkurt Demodulator.* September and October 1966.

"Information Theory and Coding," *The Lenkurt Demodulator,* April and May 1965.

Lender, A. *Correlative Signal Processing.* Stanford, Ca.: Stanford University Center for Systems Research, 1972.

Lucky, R.W. "Techniques for Adaptive Equalization of Digital Communication Systems," *The Bell System Technical Journal.* Vol. 45 (February 1966), 255-286.

Lucky, R.W., et al. *Principles of Data Communication.* New York: McGraw-Hill Book Company, 1968.

Shannon, C.E. and Weaver, W. *The Mathematical Theory of Communication.* Urbana, Ill.: University of Illinois Press, 1949.

Tracey, R.J. "An Incoherent Adaptive Mean Square Equalizer," *International Conference on Communications,* (June 1974), 9F-1 – 9F-5.

The GTE Lenkurt Demodulator is circulated monthly to selected technicians, engineers and managers employed by companies or government agencies who use and operate communications systems, and to educational institutions. Permission to reprint granted on request.