selected
articles
from

The

Lenkurt

Demodulator

Printed in the United States of America

PREFACE

This volume contains 46 of the best and most popular articles selected from **The Lenkurt Demodulator** during the five year period from 1966 through 1970. The articles are grouped topically into five sections – I. Multiplex Technology, II. Microwave Radio, III. Data Transmission, IV. General Communications, and V. Solid-State Design. An expanded table of contents is included for quick reference to the subjects covered in each article. A detailed index is included at the end of this volume for easy reference to any particular subject covered in Volume 1 or Volume 2.

These articles have been reprinted without change; therefore, occasional references may be found to articles not appearing in this volume. Some outdated articles have been included for their historical and tutorial value.

EDITOR

CONTENTS

SECTION I

MULTIPLEX TECHNOLOGY

SECTION II

MICROWAVE RADIO

SECTION III

DATA TRANSMISSION

SECTION IV

GENERAL COMMUNICATIONS

SECTION V

SOLID-STATE DESIGN

SECTION I

MULTIPLEX TECHNOLOGY

MAY, 1966
VOL. 15, NO. 5

the Lenkurt Demodulator

Operating Standards

For Frequency Division Multiplex Systems

Operating standards in the communications industry help to provide an efficient means of linking together hundreds of separate and independent networks throughout the world. They are necessary to achieve uniform performance and to insure high quality transmission. This article mentions the prominent communications agencies that have contributed to standardization, and discusses some of the important characteristics needed to interconnect different frequency division multiplex systems.

One of the most valuable assets of any society is freedom of communication. The unrestricted transfer of information and ideas is vital to promote education, commerce, business, and government operations, and to protect the welfare and security of a free nation.

The vast telecommunications networks that have been developed in the United States and the rest of the free world have indeed become great national resources. These networks carry voice and telegraph messages and a variety of other forms of communications such as data, facsimile, and television, to almost any place in the world. The services provided by these networks must be reliable, economical, and increasingly useful in order to advance user satisfaction.

The enormous success of the communications industry certainly can be attributed to continual improvements made in the quality of service and to the increasing efficiency of equipment and facilities. This, of course, has resulted in lower costs and has permitted almost everyone to fulfill his essential communications needs.

Perhaps one of the most significant factors that has contributed to the progress of communication systems has been the development of universal *operating standards.* To achieve total worldwide communications, thousands of separate networks have to be linked together. It is extremely desirable, therefore, that each of these networks be able to handle the same types of electrical signals. If it is relatively easy to transfer signals from one system to another, the communications services are apt to be more economical and efficient. In a growing and dynamic world, it would certainly be impractical to develop communications networks that, because of technical differences, could not transfer messages to adjacent systems without complicated and expensive conversion equipment. This would be tantamount to railroad systems having tracks of different gauges!

It is also very important that each network preserve the quality of transmission. This means that the performance characteristics of these systems must conform to a *set of rules* which specify standards of operation. In answer to this, many written standards and practices have been developed to cover not only operating problems, but almost every aspect of electrical communications. These standards provide the basis of comparing and evaluating the performance of communications systems. Although the use of such standards often is not obligatory, they are essential and are generally recognized and accepted by the communications industry. The particular standards adopted depend, of course, on the type of system, its intended use, and the performance requirements necessary to interconnect it with other systems.

Who Issues Standards?

In the United States, the most widely used standards or performance objectives are those of the Bell System and the Department of Defense. The Bell System has developed most of the standard practices that are used by the telephone industry to interconnect long-haul multiplex and carrier systems in North America. These standards are contained in publications known as *Bell System Practices (BSP's).*

For the huge worldwide Defense Communications System (DCS), a separate set of standards has been established. Operation of the DCS is controlled by the Defense Communications Agency (DCA) which issues DCS Engineering-Installation Standards to assure uniform high-quality performance of each segment of the system. Where appropriate, these military standards agree with those developed for use by the telephone industry.

There are other agencies and organizations that play a very active role in developing operating standards for carrier and multiplex systems. Prominent among these are the Communication and Signal Section of the American Association of Railroads (AAR) and the Rural Electrification Administration (REA) of the Department of Agriculture. Also, the Electronics Industries Association (EIA) has been very active in helping to standardize the characteristics of digital data signals that are to be transmitted over communications systems.

Another important set of standards used in the development of carrier telephone systems is produced by an organization known as the International Telegraph and Telephone Consultative Committee (CCITT). This body is a branch of the International Telecommunications Union (ITU), located in Geneva, Switzerland. The ITU is an agency of the United Nations.

The CCITT issues recommendations for standardizing international telephone and telegraph circuits. The need for such recommendations developed originally in Europe where many dif-

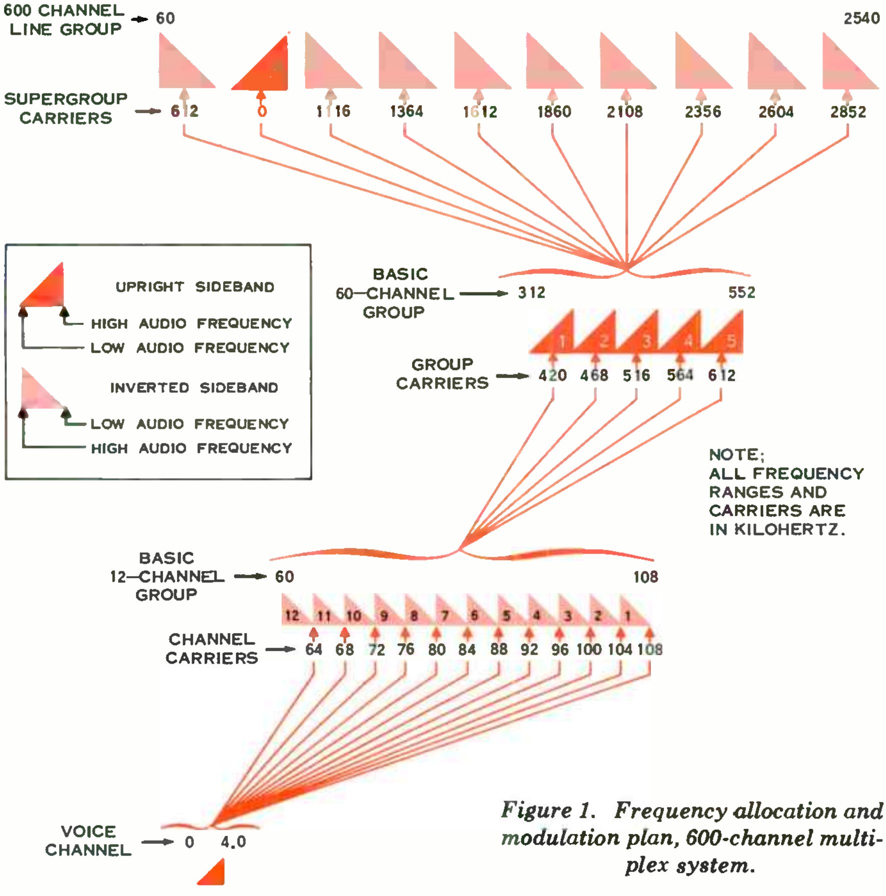


Figure 1. Frequency allocation and modulation plan, 600-channel multiplex system.

ferent telephone administrations had to interconnect at international borders. Unlike other parts of the world, Europe has many dense population centers concentrated in small political divisions. Because of the relatively short distances separating these populated areas, there is a great amount of telephone traffic between them. Therefore, it was necessary to establish an international cooperative organization where the nations involved could get together and agree on universal standards. Such agreements have been very effective in assuring that international circuits of various national telephone administrations and common carriers are compatible. Today, countries all over the world who are interested in promoting and developing international telecommunications networks, are represented in the CCITT.

Frequency Allocation and Modulation Plans

One of the most important aspects of interconnecting frequency-division multiplex and carrier systems is the assignment of frequencies. Each type of carrier and multiplex system employs some type of modulation scheme to shift the voice-frequency signals received from user equipment to some suitable line or baseband frequency range. These schemes are referred to as *frequency allocation and modulation plans.*

Whenever two carrier systems are connected in tandem, signals at the interface point must conform to the technical requirements of the receiving system. Of course, signals at line or baseband frequencies can be simply demodulated to the voice-frequency range and then transferred to the next system. This method, although acceptable, has proven to be rather inefficient in many cases. Extra equipment is needed to demodulate the signals and each additional modulation and demodulation step adds distortion to the signal. What was needed was a standard modulation plan which would allow different carrier and multiplex systems to be interconnected directly at line or baseband frequencies or at some intermediate stage of modulation. This would allow groups of channels to be transferred between systems without the need for extra equipment and unnecessary modulation steps.

When the Bell System began developing its wideband coaxial cable carrier system in the 1930's, considerable thought was given to standardizing single-sideband suppressed-carrier multiplex terminal equipment. One of the results of this effort was the establishment of a standard modulation plan for groups of channels. To accomplish this it was first necessary to standardize the spacing of channel carriers; the Bell System decided on a uniform spacing of 4 kHz. This would permit all channel carriers to be harmonically related to 4 kHz and would allow room to improve the quality of speech transmission with advances in filter design. The next step was to formulate a basic modulation plan that could be used in open-wire and multipair cable systems as well as wideband coaxial cable systems. With 4-kHz channel spacing, single-sideband suppressed-carrier open-wire systems operating with line frequencies above 30 kHz (this would place such systems above the Bell System's 3-channel type C carrier system) could only handle about 12 channels. So it was decided to establish a basic modulation plan for 12 channels. Coaxial cable systems (and later microwave radio systems) were, of course, not limited to 12 channels, but standard 12-channel groups could be used as building blocks to form systems with hundreds of channels by simply using additional stages of modulation. Since the practical lower frequency limit for coaxial cable was about 60 kHz, the standard 12-channel group was established with a frequency range of 60 to 108 kHz.

This standard 60 to 108 kHz 12-channel group has received wide acceptance as the basic building block for long-haul carrier and multiplex systems, and has been adopted by CCITT for use in international circuits and by the DCA for use in the Defense Communications System. Additionally, a standard 60-channel *supergroup,* formed from five 60 to 108 kHz channel groups, has been adopted for use in wideband systems. This supergroup has a frequency range of 312 to 552 kHz.

An example of a frequency allocation and modulation plan for a 600-channel multiplex system is shown in Figure 1. In the first modulation stage for this plan, each voice-frequency input signal modulates one of 12 *channel carriers* spaced 4 kHz apart. The lower sideband signals are selected to provide the standard 60 to 108 kHz 12-channel group. In the second modulation stage, five 12-channel groups each modulate

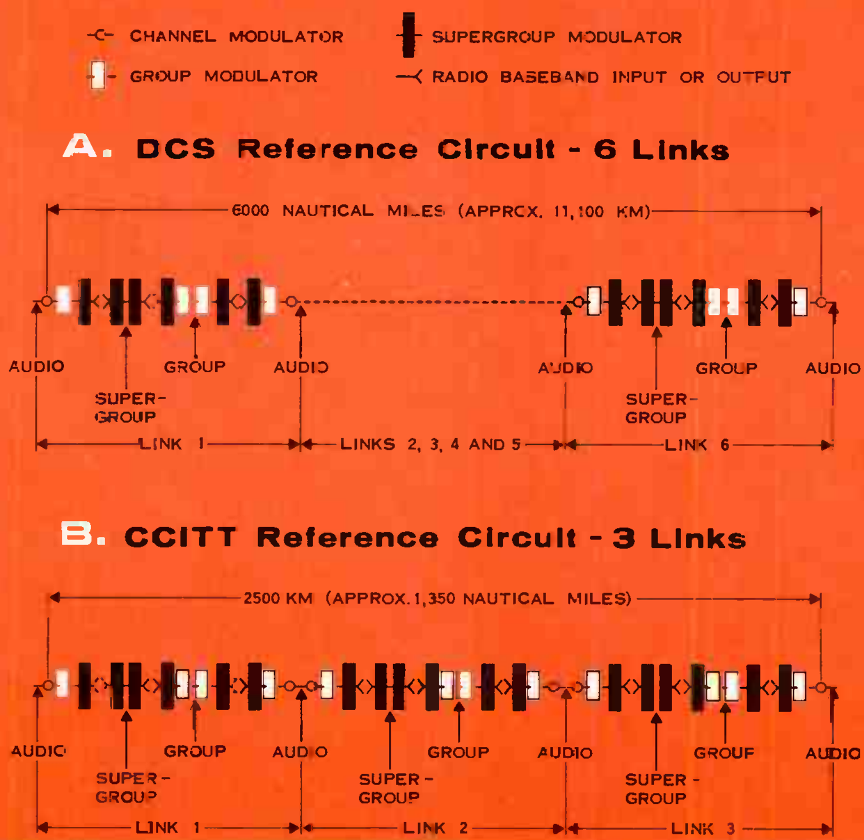


Figure 2. DCS and CCITT reference circuits.

a separate *group carrier* to produce a standard 60-channel supergroup with a frequency range of 312 to 552 kHz. Ten of these supergroups are needed to form the 600-channel system.

In the final stage of modulation, nine of the ten supergroups each modulate a separate *supergroup carrier*, resulting in line frequencies ranging from 60 to 2540 kHz. One of the supergroups (supergroup number 2) is applied directly to the line at the 312 to 552 kHz frequency level.

This particular 600-channel modulation plan is recommended by the CCITT and is standard for use in the Defense Communications System. The Bell System uses a slightly different line frequency range for their type L wideband carrier and multiplex systems. In the type L system, the modulation plan is the same as the one described, through supergroup 8. Supergroups 9 and 10, however, employ carriers of 1860 and 3100 kHz, respectively, resulting in an upper line frequency of 2788 kHz rather than 2540 kHz.

Modulation plans can be expanded to meet the future needs of higher density wideband multiplex systems requiring 2700 or more channels. These expanded plans are formed by additional modulation steps using higher order *master* and *supermaster* channel groups. By adhering to these standard frequency allocation and modulation plans, it is possible to directly interconnect 12-channel, 60-channel, and higher order channel groups of various carrier and multiplex systems, without having to first demodulate the signals down to the voice-frequency range.

For single-sideband suppressed-carrier open-wire carrier systems, the standard frequency allocation and modulation plan provides up to 12 channels. Since open-wire systems are typically 2-wire systems, the frequencies transmitted to the line must be different for each direction of transmission. This establishes what is known as an *equivalent 4-wire* system. The two directions are conveniently referred to as the *east-west* direction and the *west-east* direction.

After the 12 voice-frequency channels have been translated to the 60 to 108 kHz group level, they are then shifted to one of four staggered line frequency allocations. Staggered line frequency allocations are necessary to overcome unacceptable crosstalk where different systems share the same open-wire lead. The four staggered line groups are shown in Table 1.

This open-wire modulation plan is used in the Bell System type J carrier system and is specified standard by the DCA and CCITT.

The standard 60 to 108 kHz group modulation plan is also used in multi-pair cable carrier systems to provide 12 or 24 channels. The line frequencies for cable systems are also different for each direction of transmission. The DCA prescribes a 12-channel system with line frequencies of 6 to 54 kHz for one direction and 60 to 108 kHz for the other direction.

The standard 24-channel plan used by the telephone industry requires two basic 60 to 108 kHz 12-channel groups. Again the line frequencies are different for each direction of transmission. Typically, the channels in one direction are referred to as the *low line group,* and have a frequency range of 36 kHz to 132 kHz. The channels in the other direction, called the *high line group,* have a frequency range of 172 to 268 kHz.

TABLE 1.

Staggered line frequency allocations for 12-channel open-wire carrier system.

SYSTEM		WEST-EAST	EAST-WES
DCA	CCITT		
A	SOJ-A-12	36 to 84 kc	92 to 140 k
B	SOJ-B-12	36 to 84 kc	95 to 143 k
C	SOJ-C-12	36 to 84 kc	93 to 141 k
D	SOJ-D-12	36 to 84 kc	94 to 142 k

Performance Objectives

In order to define standard performance objectives of communications systems, the CCITT, the United States telephone industry, and the DCA have established hypothetical *reference circuits.* These circuits are of a specified length and are composed of a certain number of links. The amount of equipment in each link varies depending on whether the transmission path consists of open wire, cable, or radio. The reference circuits are complete transmission systems interconnecting two audio-frequency terminals. Each link consists of a number of 4-wire, nominally 4-kHz voice-frequency circuits derived from single-sideband suppressed-carrier frequency-division multiplex equipment using standard modulation plans. Such hypothetical reference circuits are very useful in establishing guidelines for the performance characteristics of a communications system.

The CCITT reference circuit, Figure 2A, is 2500 kilometers (1550 statute miles) long and consists of three tandem links with interconnections made at group and supergroup frequency levels. In the United States, the telephone industry uses a reference circuit of 4000 miles, made up of a maximum of seven links. However, from a performance standpoint the two circuits provide essentially the same results.

The DCS reference circuit for wideband systems is shown in Figure 2B. This hypothetical circuit is 6000 nautical miles long and consists of six tandem links each approximately 1000 miles long and interconnected on a 4-wire basis at the audio-frequency level. Each link is divided into three sections of equal length and consists of wire or radio facilities plus necessary repeaters and frequency division multiplex equipment.

Through the use of these hypothetical reference circuits, it is possible for various manufacturers to develop multiplex equipment with uniform performance capabilities. In addition to the type of multiplexing and associated frequency allocations and modulation plans, the reference circuits are used to define and standardize other circuit characteristics such as noise objectives, power levels, impedances, pilots, and signaling in order to interconnect groups of channels at carrier frequencies.

Power level is a very important factor which must be considered when establishing guidelines for interconnecting multiplex systems. The amount of power required at the voice-frequency input and output circuits of multiplex systems is determined by the needs of the subscriber or user equipment, or the switching center in the communications system. In the United States, the standard used to set the levels for speech transmission is a 1000-Hz test tone at a level of 0 dBm0. The CCITT specifies an 800-Hz tone for the same purpose.

Both the Bell System and the DCA have standardized the input level of speech signals at −16 dBm and the output level at +7 dBm with a balanced circuit impedance of 600 ohms. These levels result in a net gain of 23 dB from the input of the multiplex transmit channel to the output of the distant multiplex receive channel. The CCITT also recommends a voice-frequency circuit impedance of 600 ohms, but has not specified any standard voice-frequency power levels.

Conclusion

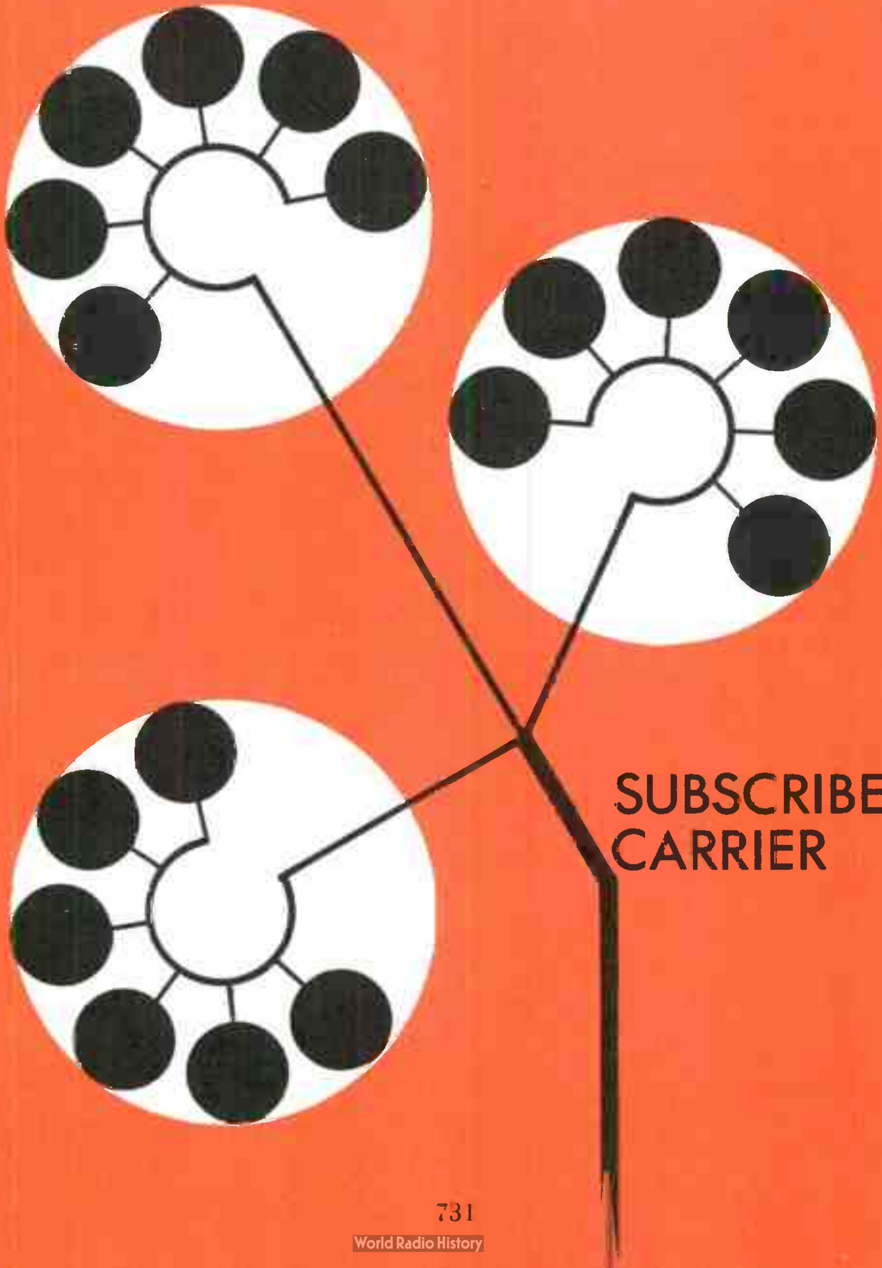
The development of universal operating standards for carrier and multiplex systems has certainly been a tremendous help in advancing worldwide international communications. Such standardization has made it possible to transfer groups of channels at carrier frequencies directly from one communications system to another. This has resulted in communication services with greater efficiency, better quality, and lower costs.

Direct Distance Dialing in the United States is an excellent example of what can be achieved through the use of operating standards for communications transmission equipment. With the advent of worldwide multiple-access communications satellites the need for universal standards for interconnecting carrier and multiplex systems will certainly become more and more significant.

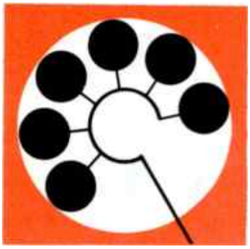
The Lenkurt

MARCH 1970

DEMODULATOR



SUBSCRIBER CARRIER



Subscriber carrier transmission systems economically expand existing cable capacity to meet the demand for additional voice and data circuits.

Man's unbounded desire to communicate has led to an ever increasing demand for new telephone circuits. A method of increasing telephone circuits without adding thousands of miles of wire is obviously desirable (Figure 1). While putting paired wires into cables serves to remove some of the wire from view, the problem of continually enlarging the number of physical circuits to satisfy the increasing demand remains.

Multiplexing permits simultaneous transmission of two or more signals over the same telephone circuit and has been one of the most valuable developments in the telephone industry since the early 1900's. Of course, the first multiplex carrier equipment was expensive and only practical for long distance circuits. As technology improved this condition changed.

The reduction in cost of multiplex equipment played an important role in the expansion of carrier for short haul transmission. One advance that made carrier practical for short haul transmission was in semiconductor technology, resulting in high performance, low cost transistors. The use of integrated circuits also improved performance and reliability, thereby decreasing cost still further.

Carrier transmission over open wire or cable between the central office and a subscriber is called subscriber carrier or station carrier. These systems provide less expensive transmission by using the medium more effectively. An ordinary cable pair carries one voice or data channel. The same cable pair, using a station carrier system can carry six carrier frequency derived circuits which can be used for voice or data.

Three Types

There are three basic types of subscriber carrier systems in use today. Each is designed to satisfy specific requirements, and each offers individual economic advantages. One class of subscriber carrier effectively extends the reach of the central office to outlying areas (as much as 30 miles away) and may carry 20 or more channels. A second type, providing six channels, serves to expand cable facilities closer to the central office. The third is a single-channel system specifically designed to add one additional subscriber to a cable pair easily and inexpensively at distances under 3.5 miles (5.4 kilometers).

Although the three types have different uses, they have some common advantages. All can be used with a combination of standard gauge cables. Minimum maintenance is possible because of the advantage of simply replacing defective equipment. Cable additions can, also, be deferred and planning flexibility provided using any of these carrier systems.

Multi-channel systems provide even greater service flexibility because each channel can be used for party-line service.

Multi-channel carrier equipment, for relief of existing facilities, is presently most practical at about 4 miles (6.4 kilometers) from the central office (Figure 2). Whereas presently, for

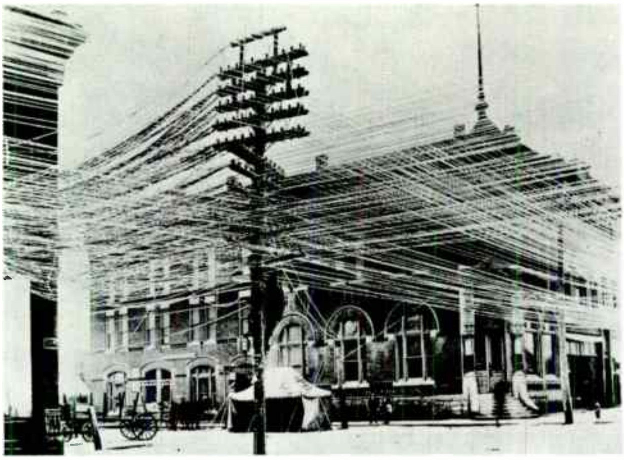


Figure 1. Maze of wires in this 1909 photograph epitomized telephone circuits before multiplex techniques were developed.

Bell Labs RECORD

initial installation, closer than 10 miles (16 kilometers) multi-channel equipment is not always practical. Cable must be installed initially, even when carrier is used; therefore, the economic trade-offs differ for initial and relief installations.

Upgrading and Expanding

Subscriber carrier first proved suitable for expanding cable and wire to rural and sparsely populated areas. Customers in these locations were accustomed to multi-party service and shared a circuit with as many as nine neighbors. As the economy of the established rural population changed, demands for urban quality service increased. The migration of urban workers — accustomed to single or two-party service — to rural areas has also added pressure for improved telephone service.

One-party service is being established under ambitious upgrading programs. Improving service by adding new physical circuits to remote areas, however, is not always economically practical. Factors, such as the need for automatic toll ticketing and increased copper and labor costs, also influence the decisions of planning engineers toward layout of new systems and planning the company's approach to growth areas.

The general need to upgrade service, relieve cable congestion, and provide growth margins in new developments has enhanced the addition of carrier equipment to exchange loop facilities. Carrier systems upgrade existing service and provide new service in areas without spare cable pairs. Additional benefits are realized in planning new cable installations around carrier systems.

Station carrier is also being used to provide temporary service for such functions as charitable and political campaigns, conventions, county fairs, home shows, sports events, etc. The largest application to date has been to provide immediate relief for a fraction of the cost of new cable installation. Carrier has gained recognition controlling new construction while meeting customer requests for additional and upgraded service in areas where facilities are already used to capacity.

Comparison

Carrier systems used to supplement existing physical circuits must provide a high degree of reliability, transmis-

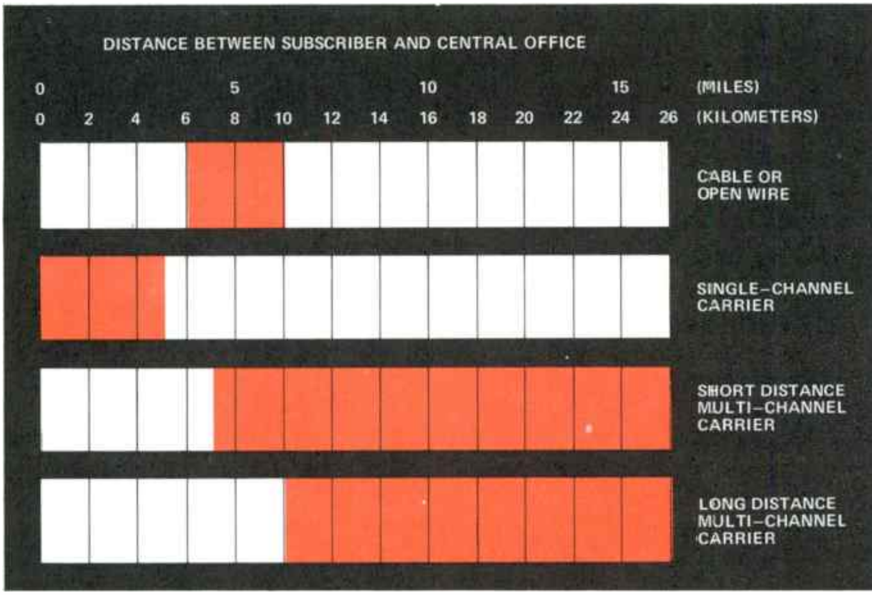


Figure 2. For relief of overloaded facilities, the green areas indicate, at the present level of technology, each system's recommended operating range.

sion performance, flexibility, and economy compared to the alternative use of cable.

Since carrier systems operate over physical circuits, total system reliability is the sum of the carrier reliability and the reliability of the physical circuit. Designing high reliability carrier equipment is thus a primary requirement. Although such tight quality control increases the product cost, frequent service complaints and high repair costs are not an acceptable compromise. Today's carrier systems have been able to maintain high reliability at reduced costs by using integrated circuits and semiconductor devices.

Transmission Performance

Transmission advantages obtained from carrier derived circuits include signal consistency and stability. The quality of a carrier signal is almost identical at any cable length. In addition, net losses can be carefully controlled, and environmental conditions will have little effect on the stability of the circuits.

Long physical circuits (approximately 8 - 10 miles) contribute to increased noise, delay distortion, and degraded frequency response. These are major considerations especially when data is being transmitted. These parameters are degraded with increasing cable length. Using carrier, long loop transmission performance is significantly improved.

With station carrier systems such as Lenkurt's 82A, all carrier frequency signals leaving the central office are fixed at the same level. In the individual subscriber terminal, the received carrier level is regulated and the voice frequency is then detected. The carrier signal from the subscriber back to the central office is automatically preadjusted to compensate for cable loss. Channel signal levels within a carrier system will be different because each channel is a different frequency and line loss varies with frequency. However, the level of a particular channel frequency in one system will be the same as the level of the same channel frequency in another system operating over cable pairs within a common cable sheath.

This automatic regulation reduces far-end crosstalk between systems by maintaining similar levels for all systems within the cable sheath, regardless of channel terminal location.

The automatic signal regulation eliminates the need for manual level adjustment and also makes it possible to install and maintain the equipment by personnel not familiar with electronic circuits.

Installation

Carrier systems can be added to almost any type physical circuit – open wire, cable (aerial and buried), or a combination of open wire and cable – to increase the number of circuits available for voice and data transmission, or to reach out distances where transmission limitations prevent the use of voice frequency circuits.

Once preliminary engineering is completed and a cable has been specified, carrier installation with systems such as Lenkurt's 82A and 83A is routine and much faster than installa-

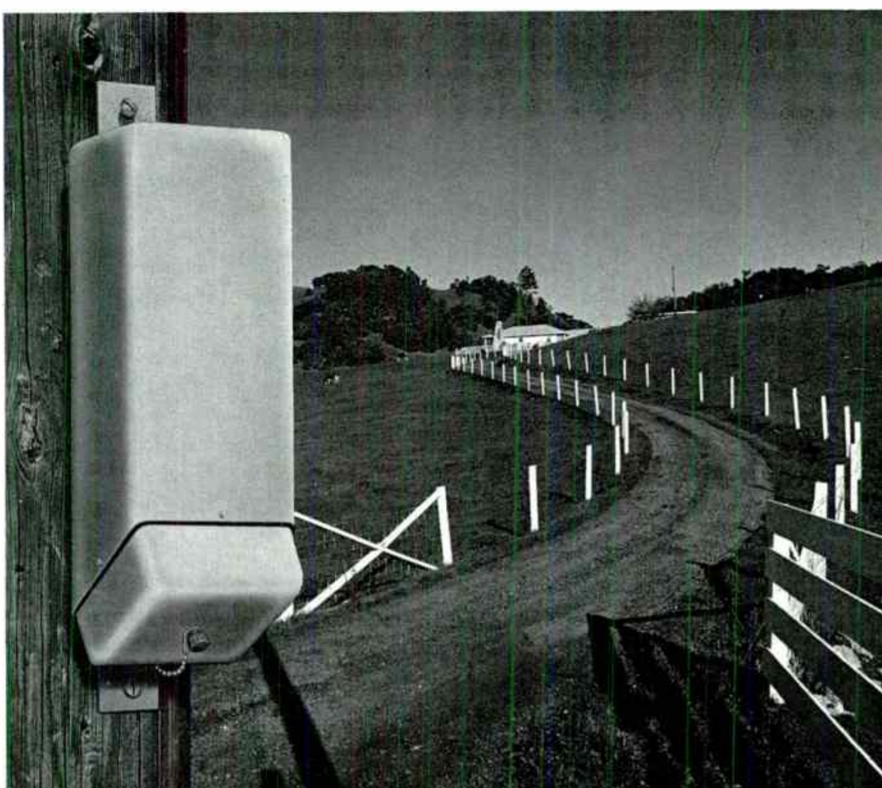


Figure 3. The subscriber terminal, shown here pole mounted, may be placed at any point along the cable.

Figure 4. Operations at the Charlotte Speedway in North Carolina, were able to start on time, with the necessary additional telephone service provided by Lenkurt's 82 A Station Carrier system.

tion of additional cable. Carrier equipment is installed at the same time as the home instrument, and can be placed at any point along the cable. No field adjustments are needed. The subscriber terminals and repeaters – identical in appearance – are designed for pole, crossarm, strand, or pedestal mounting (Figure 3).

Flexibility

Carrier equipment is readily available making it practical to work existing cables to a 100 percent voice frequency fill, instead of the 80 to 90 percent usually specified. Carrier equipment can be drawn from stock, or inventory, and installed as quickly as any non-carrier circuit – even when cable pairs are available. This is an important feature which is sometimes overlooked.

If carrier equipment is installed and the anticipated growth is not realized, the unnecessary carrier equipment can be removed and returned to stock with relative ease.

Short haul carrier equipment is practical for long term use in instances where the alternative is a long cable with a small circuit capacity.

If, however, carrier equipment is used as an interim measure to defer new cable provisions, carrier advantages are improved and shorter routes or larger growth rates are warranted. The degree of use which can be justified depends on the installed cost of a new cable in particular localities and on the expected return on investment.

After the addition of a new cable, station carrier systems used as a temporary means of providing additional circuits can be moved to a new location. There is a definite advantage in not having to rely on precise cable section forecasting. Instead, it is quite sufficient to use general circuit requirement forecasting for an area. Carrier equipment can be moved to meet changing demands; but, cable once laid, is fixed.

Economy

In a typical exercise to determine cable size for a new facility, the area's expected five year circuit requirements are reviewed, and possibly revised to reflect the personal experience of local engineers. Based on the results, a cable size is selected. Since cable is supplied in standard numbers of pairs, the usual practice is to choose the next larger size cable to ensure adequate facilities. This method will sometimes result in specifying cable with as many as 50 percent more pairs than are actually necessary. This overspecification can be costly when great distances are involved.

Using carrier systems in planning for maximum future requirements, the plant engineer realistically sizes his new cable to the nearest number of pairs instead of the next larger size. Extra circuits can be added with carrier equipment as required – reducing both initial and annual costs.

Practical Solution

When the Charlotte Motor Speedway in North Carolina, moved its administrative headquarters from downtown Charlotte to the race track near Harrisburg, N.C., it needed five additional circuits at the track immediately.

To install more cable facilities between the Speedway and the Concord Telephone's central office eight miles away would have been a time consuming project. Besides, a cable installation to provide five more circuits would be economically unsound.

In response to the problem, Concord Telephone turned to Lenkurt's 82A Station Carrier system. The 82A, designed to expand the capacity of a single subscriber cable pair to six private lines, proved to be an ideal solution.

The 82A, an extremely adaptable system, required no modification to the existing central office equipment or telephone subsets (Figure 4). All power for the repeaters and subscriber equipment was received from the central office supply, and no batteries were required for subscriber terminal or repeater operation.

A few days before the World 600 race, the Charlotte Motor Speedway officials requested an additional teletype circuit. Using 82A, another circuit was quickly and easily provided.

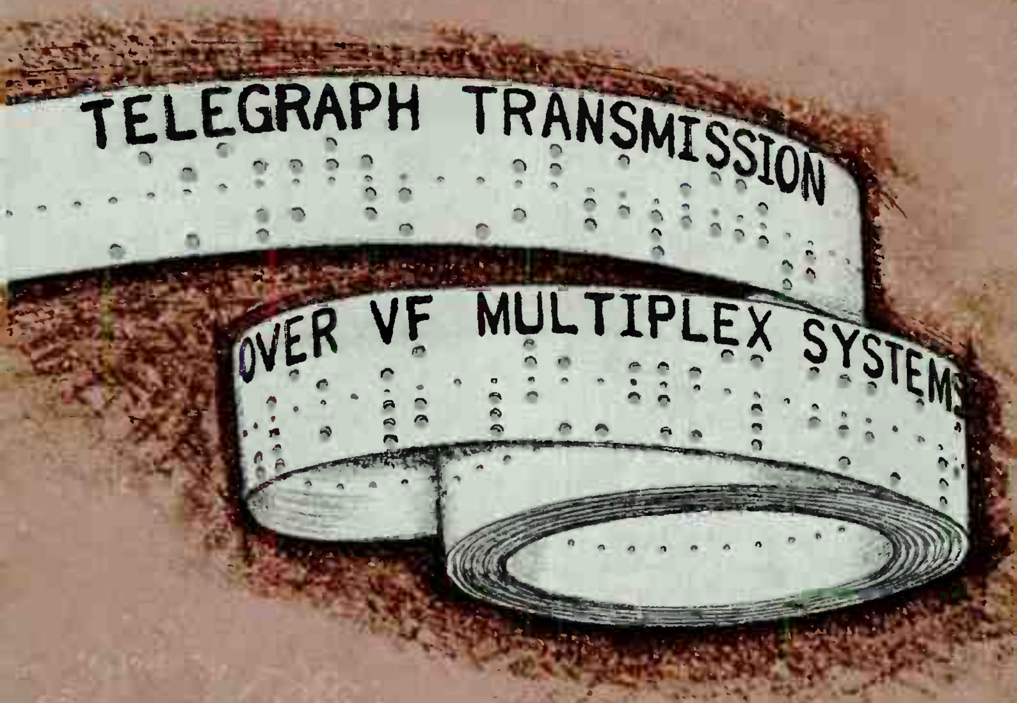
Even Greater Capacity

Subscriber carrier transmission can be an economical alternative to cable when the annual circuit requirements are small or where the circuit length is quite long. If future requirements are uncertain, carrier provides interim relief, allowing time to gather reliable data before making a major cable addition. Carrier systems are also being used where new cable additions would provide a great many extra pairs that would lie idle for several years.

Until such time as an even greater density, lower cost transmission system becomes practical, carrier will continue to replace cable transmission on shorter and shorter loops.

FEBRUARY, 1966
VOL. 15, NO. 2

The Lenkurt DEMODULATOR



The techniques of modulation and frequency division multiplexing play an important role in the transmission of telegraph messages over voice-frequency channels. This is especially significant in view of the present need for more and more channels in which to handle the increasing amount of business information now being transmitted over telephone networks.

This article reviews briefly the techniques used to convert telegraph pulses to voice-frequency tones and describes several direct-current telegraph loops in common use. Also, the voice channel loading effect of multiplexed telegraph signals is discussed.

Telegraph systems provide a means of transmitting information using electrical pulses which conform to a preestablished code. In earlier days, telegraph messages were transmitted by hand-operated *keys* using the familiar Morse code. Modern telegraph systems, however, use electromechanical machines, called teleprinters, page printers, or tape printers, that employ some type of machine code.

Conventional telegraph machines use the standard 5-level Baudot code and normally operate at transmission speeds of 60, 75, and 100 words per minute, at pulse rates of 45, 57, and 75 bits per second, respectively. A new 7-bit code called ASCII (American Standard Code for Information Interchange) was recently developed and is expected to find wide application in data processing systems as well as for message processing. Telegraph systems currently using the new ASCII code have added an eighth bit to provide a parity check, thus making it an 8-level code. The new 4-row keyboard teleprinters designed to handle the new code operate at a transmission speed of 100 words per minute with a pulse rate of 110 bits per second. These various telegraph machines provide a printed copy of the message or a punched tape which is then used to operate a printer.

Telegraph machine signals consist of a sequence of current and no-current pulses of equal length, known as *mark* and *space,* respectively. Using the 5-level Baudot code, for example, the letter A is indicated by a signal of mark-mark-space-space-space. Before these dc telegraph signals can be transmitted over standard voice frequency communication channels, they must be converted to ac tones. There are two basic methods used to convert the dc loop pulses to tones suitable for transmission over a voice-frequency multiplex channel. These are amplitude modulation, and frequency modulation.

In both AM and FM telegraph multiplex systems, a tone oscillator, in each transmitting channel, is used to provide the necessary voice-frequency carrier. Frequency division is the type of multiplexing ordinarily used and so the carrier frequency in each channel is different.

Amplitude Modulation

Amplitude modulation methods are historically related to direct-current telegraphy. In dc telegraph, a battery or other source of direct current is keyed on and off. At the receiving end, the signals are detected by some sort of magnetic device. In AM, the process is similar except that a tone oscillator is keyed on or off to indicate mark and space conditions and, for this reason, is sometimes referred to as on-off modulation.

This method has several disadvantages. It does not use bandwidth efficiently, since two sidebands of the carrier are produced and, unlike single-sideband voice communications methods, the carrier and one sideband cannot be completely eliminated and still do a satisfactory job.

Sidebands are produced when the modulating wave causes the carrier to change from one value or state to another. In voice communications, the modulating waveform is continuous, thus causing modulation products (sidebands) to be formed continuously. If the carrier and one sideband are eliminated, the other sideband remains to convey the modulating intelligence.

In telegraphy, where on-off pulses are the modulating signal, modulation products are formed only during the transition from "on" to "off," and from

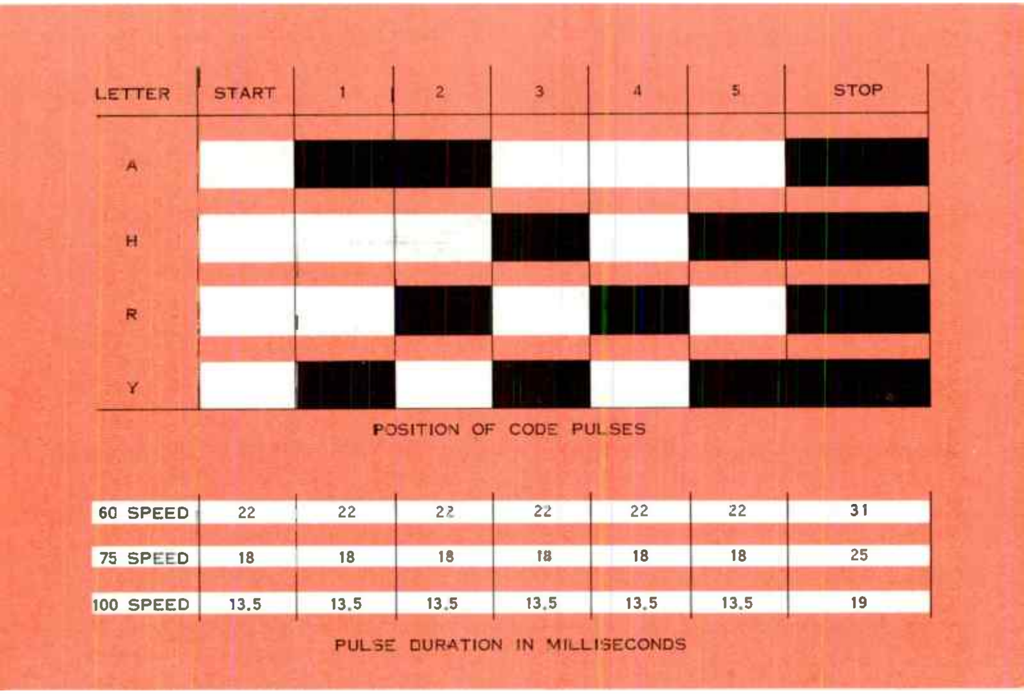


Figure 1. Examples of the 5-level Baudot code for the letters A, H, R, and Y, including pulse lengths for the three standard telegraph speeds. Mark pulses are shown in color while space pulses are shown in white.

"off" to "on." These modulation products are transients whose bandwidth is a function of the keying or switching rate. Except when a pulse is started or ended, no modulation products can appear in the transmission path. Thus, it would be impossible to continuously transmit a steady mark or space.

The information-carrying characteristic of an AM signal is its amplitude. For this reason, AM is particularly vulnerable to impulse noise and changes in transmission level. Impulse noise is particularly disturbing. Noise pulses caused by electrical storms, switching transients, and similar disturbances, may equal or exceed the information pulses in amplitude and duration. Under severe conditions, impulse noise may completely obliterate an AM information pulse.

Frequency Modulation

In FM systems, the carrier frequency is shifted in one direction for a mark condition and the opposite direction for a space condition. A *diode keyer* in the tuned circuit of the tone oscillator changes the circuit resonance so as to shift the tone back and forth between the two frequencies. Such frequency shifting does not occur instantaneously, however. The inherent resonance of the tuned circuit causes the resulting waveform to change smoothly from one frequency to the other. The amount of shift is the same for both directions and varies from about ± 30 to 42.5 Hz depending upon the operating speed of the telegraph equipment. This type of modulation is also referred to as frequency-shift keying (FSK).

Since the mark and space signals are represented by different frequencies of equal strength, amplitude variations have no effect on the signal unless the signal has the same or less amplitude than the noise. This contrasts strongly with amplitude modulation where a mark is represented by the presence of the carrier and a space is represented by the absence of the carrier. Level changes due to fading, noise, and other interference have a strong effect on AM signals. FM systems can tolerate level changes of about 40 to 50 dB, and are about 12 dB less sensitive to impulse noise than AM systems.

Bandwidth

The bandwidth required for a voice frequency multiplex telegraph channel depends on such things as the code pulse rate, noise, filter attenuation to adjacent channels, and whether or not both sidebands are transmitted (AM systems). A bandwidth of 120 Hz is usually satisfactory for 5-level code telegraph signals at speeds up to 100 words per minute for both FM and double-sideband AM systems. The usual bandwidth for 8-level coded telegraph signals is 170 Hz.

Since the required bandwidth is much smaller than that required for speech signals, a normal 3-kHz voice band can be divided by frequency division multiplexing into sub-bands or channels each capable of transmitting a telegraph signal. Approximately 18 channels can be obtained with 170 Hz spacing, while up to 26 channels can be obtained with 120 Hz spacing. This means that up to 18 or 26 voice-frequency multiplexed telegraph signals can be transmitted simultaneously over a single voice channel.

Telegraph Loops

The circuit between the telegraph machine and the multiplex terminal is called a *loop* circuit. Each telegraph loop is made up of two legs which are the conductors (full metallic or ground return) between the terminal points of the loop. In half-duplex operation, the same loop is used for sending and receiving. However, full-duplex operation, which permits simultaneous transmission in both directions, requires both a sending and a receiving loop.

Because of differences in applications and because of the variations in lengths, any one of a number of circuit arrangements may be employed in telegraph loops.

Neutral Loops

One of the simplest and most direct circuit arrangements is the *neutral* or *open-and-close* loop, illustrated in Figure 4(A). The neutral loop requires a battery only at the central office, and the difference between mark and space is determined by whether or not current is flowing in the loop.

When the printer is sending, closing of the printer contacts closes the loop circuit and the current flowing in the loop applies a potential to the multiplex-channel keying circuit. In the receiving direction, the carrier frequencies are applied to a discriminator. In the discriminator, the two frequencies that

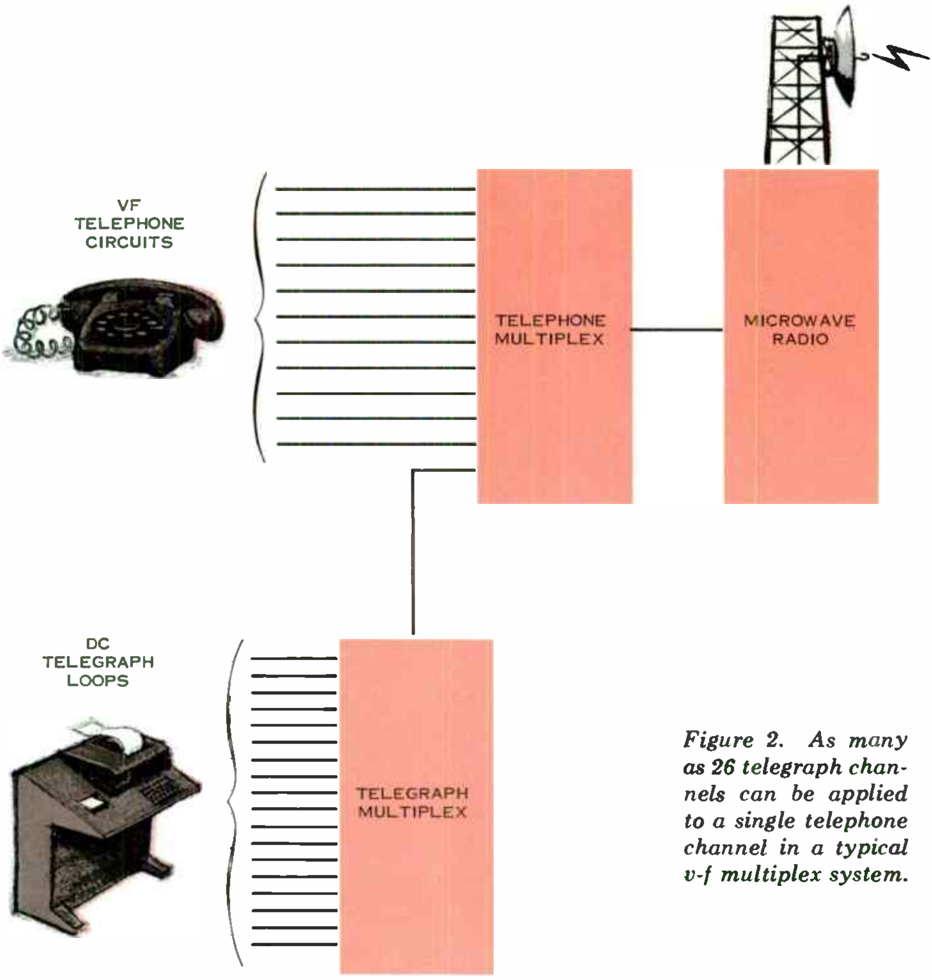


Figure 2. As many as 26 telegraph channels can be applied to a single telephone channel in a typical v-f multiplex system.

are used to transmit the marking and spacing conditions are separated and are rectified to obtain dc for operation of the polar receiving relay. The contacts of this relay open or close the receiving neutral loop to reproduce the transmitted character at the receiving printer.

Balanced Loop

While neutral loops offer the advantage of simplicity, they are restricted to the shorter loops in which either leakage or the distributed capacity of the path does not severely affect the signal. To reduce these problems a balanced loop (also called effective polar loop) may be used. An example is shown in Figure 4(B).

A balanced loop is similar to a neutral loop in that the difference between mark and space is determined by whether or not current flows in the loop. However, the balanced loop differs

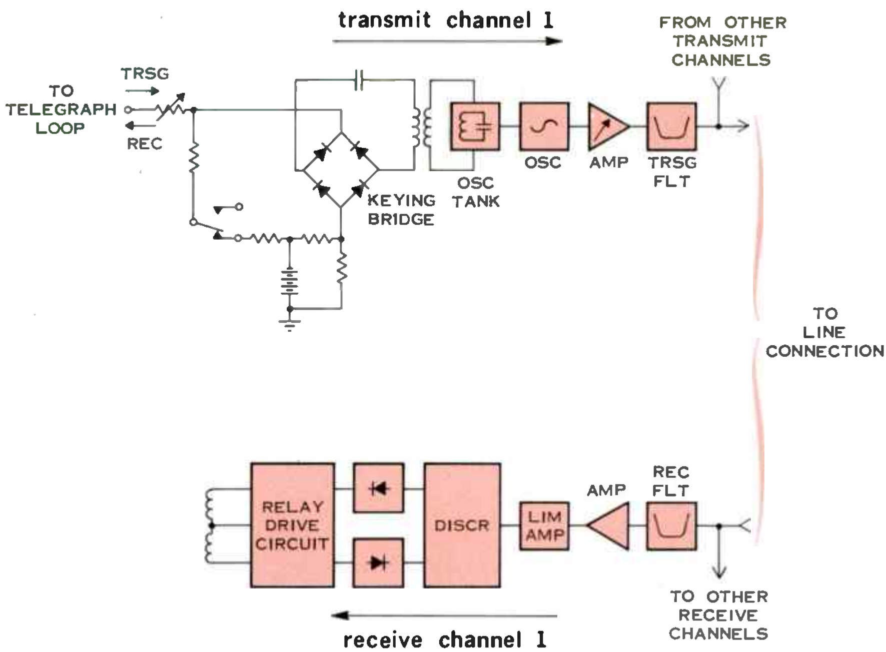


Figure 3. Simplified schematic diagram of telegraph multiplex terminal operating into a half-duplex neutral loop.

from the neutral loop in that a battery potential is applied at the printer location as well as at the central office. The printer battery, in conjunction with the battery potential applied to the marking contact, applies a higher potential to the loop. The increased potential improves the rise time of the marking pulse which tends to increase the length of the pulse. In addition, the increase in potential permits operation over longer loops.

When a spacing signal is received, application of equal potentials to both ends of the loop discharges the line more rapidly than simply opening the loop, resulting in an improvement of the pulse shape. Adjustments can be made in battery potentials to eliminate bias in the loop as required for changing conditions in the loop.

Polar Loop

The most effective transmission method commonly employed is called polar operation. In this case equal currents of opposite polarity are used for the marking and spacing conditions. In addition to the two voltages, this method requires the use of a polar relay in which the direction of current flow in a winding causes the relay to operate to either the marking or spacing position. Since printers normally oper-

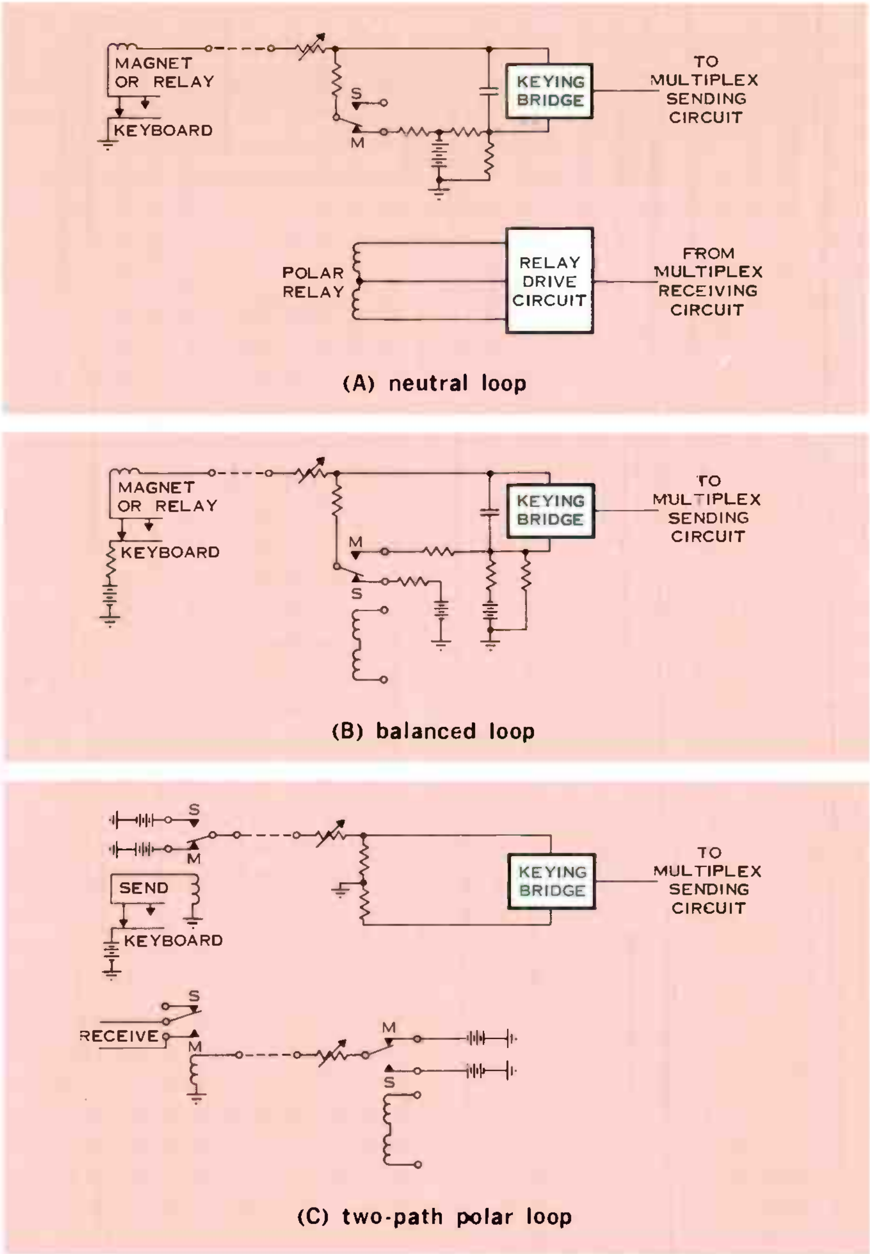


Figure 4. Three types of basic telegraph loop circuits. (A) neutral, (B) balanced, and (C) two-path polar.

ate only from on and off signals, a relay is usually required at the printer location. An example of a polar loop is shown in Figure 4(C). Where battery potentials are the same, the loop characteristics do not change between the sending of a mark or space signal, and if the relay is properly adjusted, the mark and space signals are equal and no bias is obtained.

However, because of the requirement for two batteries, the method is normally only used in transmission from the office to the subscriber, and either neutral or effective polar transmission is used in transmitting from the subscriber to the office.

Break Feature

In a half-duplex loop it is sometimes necessary for the operator at the receiving printer to interrupt the sending printer. This requirement led to the use of an additional relay in the telegraph loop, called the break relay, arranged to accomplish this purpose. The receiving operator may interrupt by opening his loop.

When the receiving loop is opened (effective spacing condition) signals received from the distant terminal are applied to the local-terminal keying circuit, but are inverted. The combination of the retransmitted signals with the original signal causes a continuous spacing signal condition at the sending terminal. When this occurs, the sending operator knows that the receiving operator wishes to interrupt.

Hub Operation

In some telegraph applications, it is occasionally desirable to connect a number of telegraph circuits together in such a way that telegraph signals originating in one circuit are transmitted to all other interconnecting circuits. A method of doing this is through a *hub* board. In this arrangement the dc sides of the multiplex channels are connected together on a high impedance basis. Thus, only a small amount of current is required.

Battery potentials of ± 130 volts are required in the hub equipment unit. The hub is supplied with a $+130$ volt potential through the hub potentiometer. The hub circuitry is such that in the normal marking condition the hub voltage is $+60$ volts.

The changes in current that result from one circuit sending a space signal into the hub changes the hub potential from $+60$ volts for marking to -30 volts for spacing. When applied to the sending portion of the remaining channels, these potentials effect simultaneous transmission of the desired signal condition. Three telegraph circuits are interconnected in the simplified diagram of a hub shown in Figure 5. Each circuit is connected to a multiplex channel through a hub-equipment unit.

Hubs may be operated either half or full duplex as with normal telegraph loops. Like the normal telegraph loop, it is sometimes necessary on half-duplex hubs for a receiving operator to break in.

Interruption is accomplished as in the normal loop by a receiving operator sending a spacing signal into the hub. The circuit is arranged so that the hub potential drops to -60 volts when two or more machines are sending spacing signals into the hub. This low potential causes all machines to go to spacing, including the original sending machine, and the sending operator then knows that someone wants to interrupt.

Channel Loading

When transmitting several telegraph tones over a voice frequency channel of a multiplex system, great care must be exercised in establishing the levels at which the signals are applied. Multiplex telegraph signals have greater average power than voice signals. If the power handling capability of the multiplex system amplifiers is exceeded, intermodulation products from the telegraph tones have far greater interfering effect on other channels than do voice signals.

For this reason, a standard signal level is usually specified for voice frequency telegraph signals transmitted over multiplex voice channels. This level is conservative, and is based on the loading effect produced by the maximum number of telegraph channels that can be handled by the voice channel. A common standard per-channel level is −21 dBm at the zero transmission level point. For most applications,

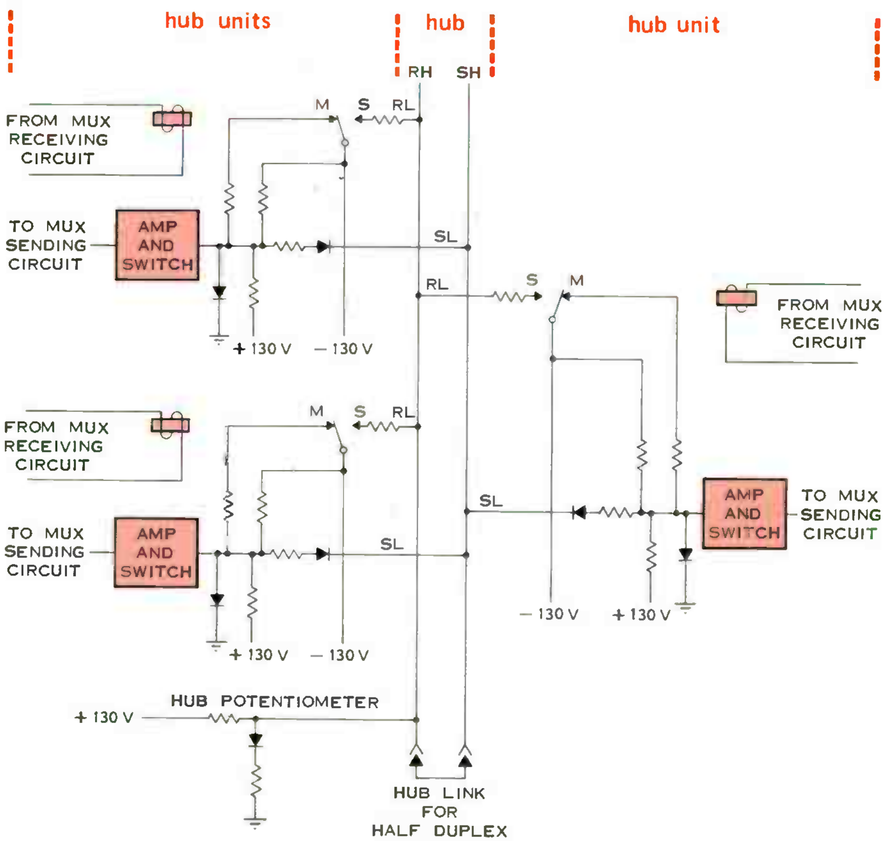
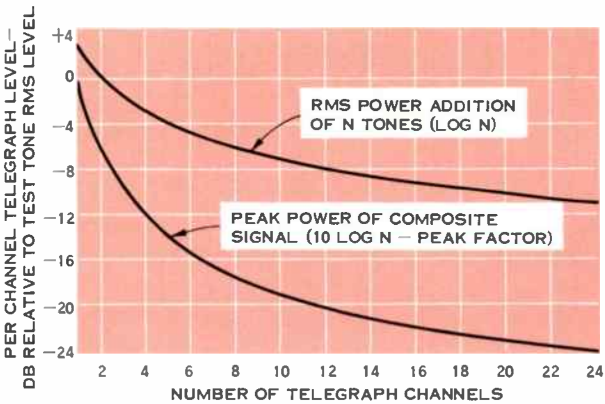


Figure 5. Hub operation, showing three multiplex telegraph channels interconnected on the d-c loop side.

Figure 6. Theoretical maximum transmission levels for various numbers of telegraph tones transmitted over a multiplex voice channel.



this level is high enough to provide good service over a voice frequency multiplex channel.

However, in applications where the maximum telegraph channel capacity is not used, it may be desirable to increase the level of each telegraph tone in order to improve the signal-to-noise ratio. The increased level is a function of the number of telegraph tones to be transmitted.

In calculating the loading effect, peak power must be used, since distortion will occur if the peak power exceeds the load handling capacity of the multiplex equipment. When telegraph signals are applied to a single voice frequency channel of a multiplex system, the permissible peak power is normally +3 dBm at the zero transmission level point. This value is assumed in the following discussion.

For a single telegraph channel, the calculation of peak power is straightforward. A sine wave is normally assumed. Peak voltage of a sine wave is 1.4 times the rms value of the wave, or 3 dB greater in power than the rms power value. When only one telegraph channel is involved, the level of the telegraph tone may be equal to the level of the normal test tone, since both signals are sine waves.

As the number of telegraph channels increases, the peak power that the composite waveform may reach also increases. Since there is a possibility that this value can become quite high for a large number of tones, a *peak factor* is used. This peak factor is based on the statistical probability that the peak power of a complex wave will almost never add up in such a way as to exceed the sum of the rms value of the wave and the peak factor. For a single tone, the peak factor is 3 dB. Peak factor increases as the number of channels is increased, reaching a maximum of 13 dB for approximately 20 channels.

As an example, assume that ten telegraph channels are to be applied to voice frequency multiplex channel normally adjusted to a −16 dBm test tone level. In this example, peak power should not exceed −13 dBm. Each telegraph channel transmitting level must be lower than −13 dBm by the sum of the combined power of the ten tones (rms power addition) and the peak factor.

First, the combined tone level is calculated by taking ten times the logarithm of the number of channels (10 log 10 = 10 dB). Adding a 12-dB peak factor to this 10-dB level gives a peak value 22 dB above a single channel peak. The per-channel transmitting power is then obtained by subtracting the 22-dB peak level from the maximum permissible level (—13 dBm minus 22 dB = —35 dBm). Similar calculations may be made for different numbers of telegraph channels. Figure 6 shows how the telegraph tone levels must be reduced as the number of channels increases. It is important to note that these calculations yield *theoretical* maximum levels for telegraph tones and pertain to the loading of a *single* voice channel in a multiplex system.

Conclusion

The transmission facilities provided by telephone communications systems constitute a vast network which is capable of interconnecting locations almost anywhere in the world. Although these facilities are made up in many forms and have different types of transmission media, they do have one very important thing in common — the standard voice frequency channel, which has a useful bandwidth of about 3 kHz.

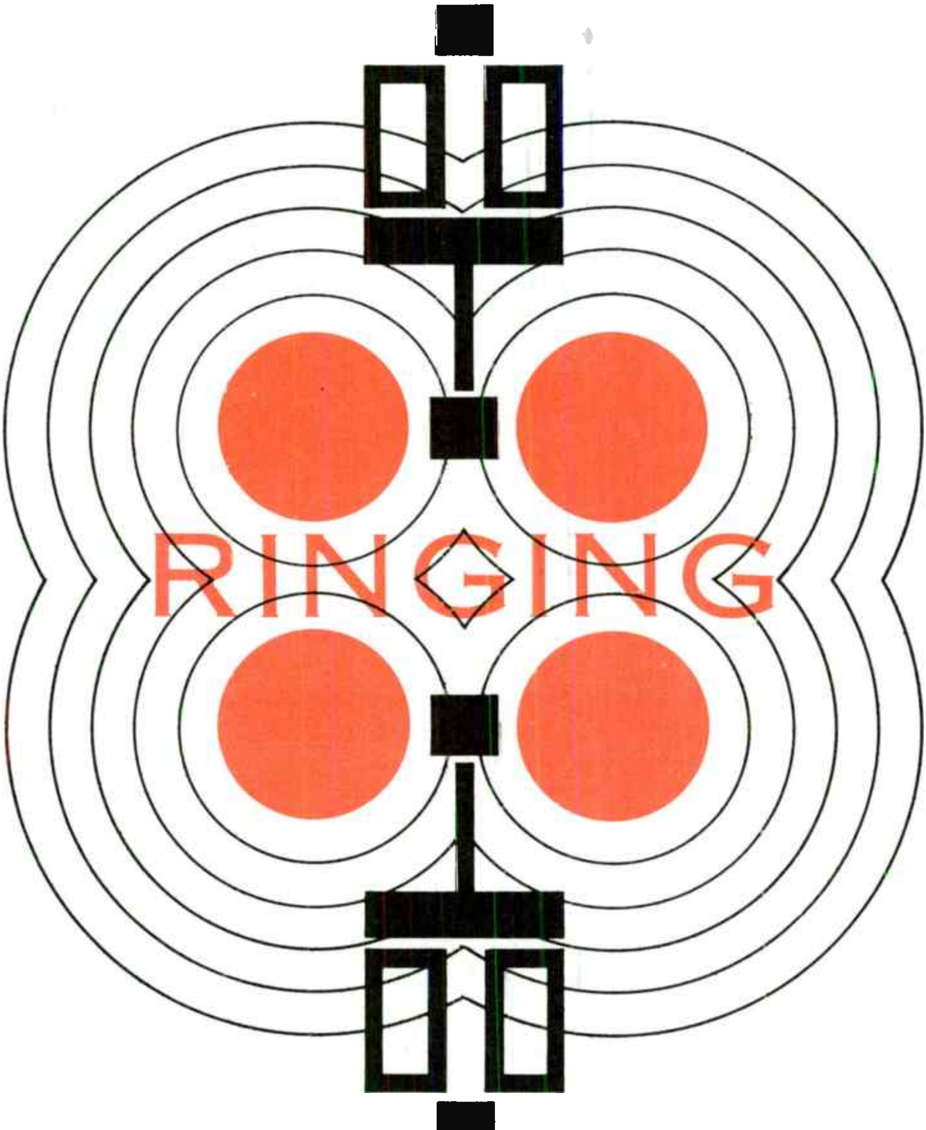
While this vast network of multiplexed telephone channels was designed primarily to handle speech signals, the circuits can be used to transmit other forms of information such as telegraph. The techniques of modulation and multiplexing provide a practical means of converting the dc telegraph signals to ac tones suitable for transmission over telephone circuits.

Through the use of frequency division multiplexing, as many as 26 narrow-band voice frequency telegraph channels can be derived within a single 3 kHz telephone channel.

Such efficient use of a single telephone channel is a tremendous asset in view of the present growth of machine communication to process business information.

The Lenkurt®
JUNE 1968

DEMODULATOR



Multiparty service through ringing.

The modern telephone system provides a pleasant and functional means of communication over literally any distance. But for the service to be efficient and convenient, a number of functions beyond just the conveyance of the voice must be performed.

By dialing ten digits, most subscribers in the United States can call any other telephone in the country. Soon direct dialing all over the world will be possible.

But before the caller gets his party, the telephone system must somehow alert the called party. This important information, called *ringing*, begins when the connection is made and remains until the called party answers or until the calling party hangs up.

The first telephones had no signaling device at all, and the lines between a number of phones were simply connected together. There was no central office or switching equipment, and service was strictly local.

Hallo!

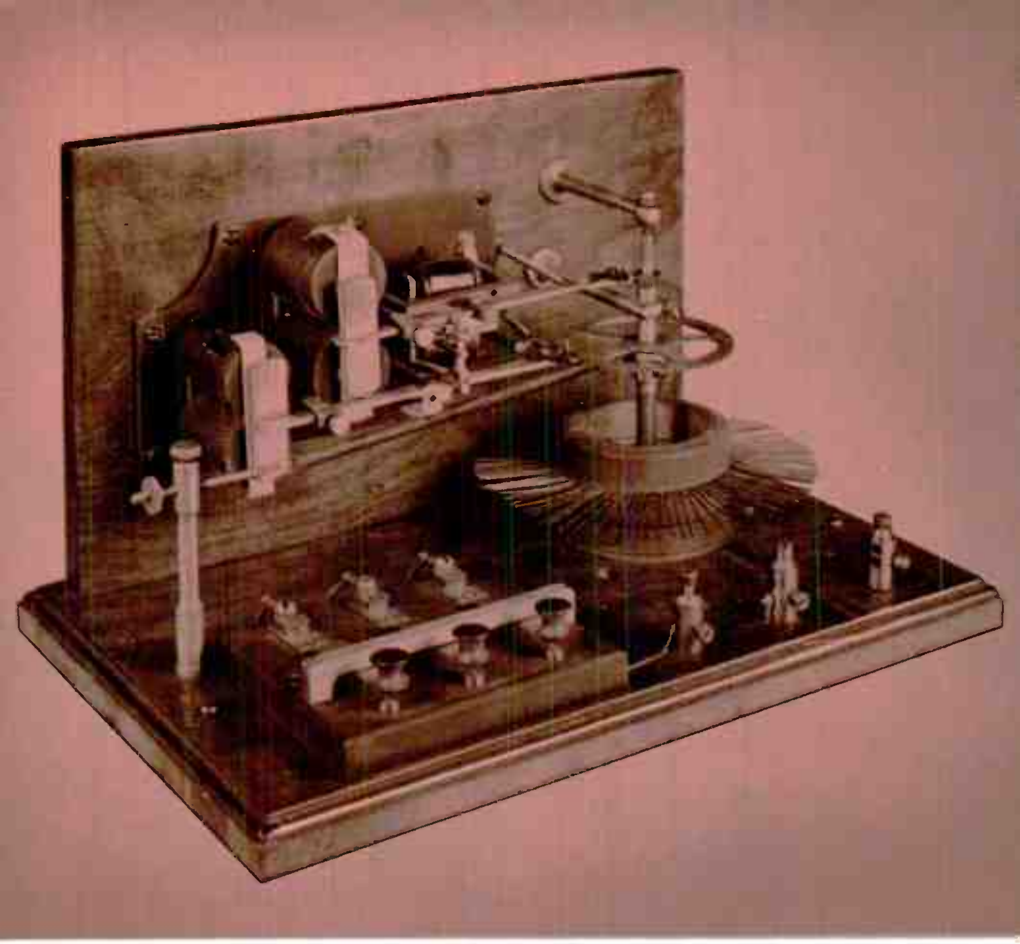
To attract attention the caller shouted into the mouthpiece. One of the common exhortations was "Hallo!", originally an exclamation to incite hunting dogs. With much usage, it became "Hello!", and one of the words contributed to our language by the telephone industry.

Callers soon learned to strike the mouthpiece diaphragm with a pencil to arouse attention. But this caused the diaphragm to become damaged, and a hammer-like device was designed to perform the same function.

A buzzer was added later, but its offensive sound was not popular with customers. This brought forth the two-gong bell, which still exists today as the most common form of signaling or ringing.

At first none of these signaling methods proved to be practical, mainly because they did not *selectively* identify the called party. Any number of parties could answer the phone or even listen in.

But along came "Central", a manual switchboard which helped the problem. This method of terminating lines at jacks and interconnecting them with patch cords still finds use in many small telephone companies.



Courtesy Automatic Electric Co.

Figure 1. Strowger's Automatic Switch, patented in 1891, provided automatic party selection without the aid of an operator. This first model had two ratchet wheels. The smaller wheel selected the tens numbers, and the larger, the ones, to give a total of 100 combinations or 100 lines.

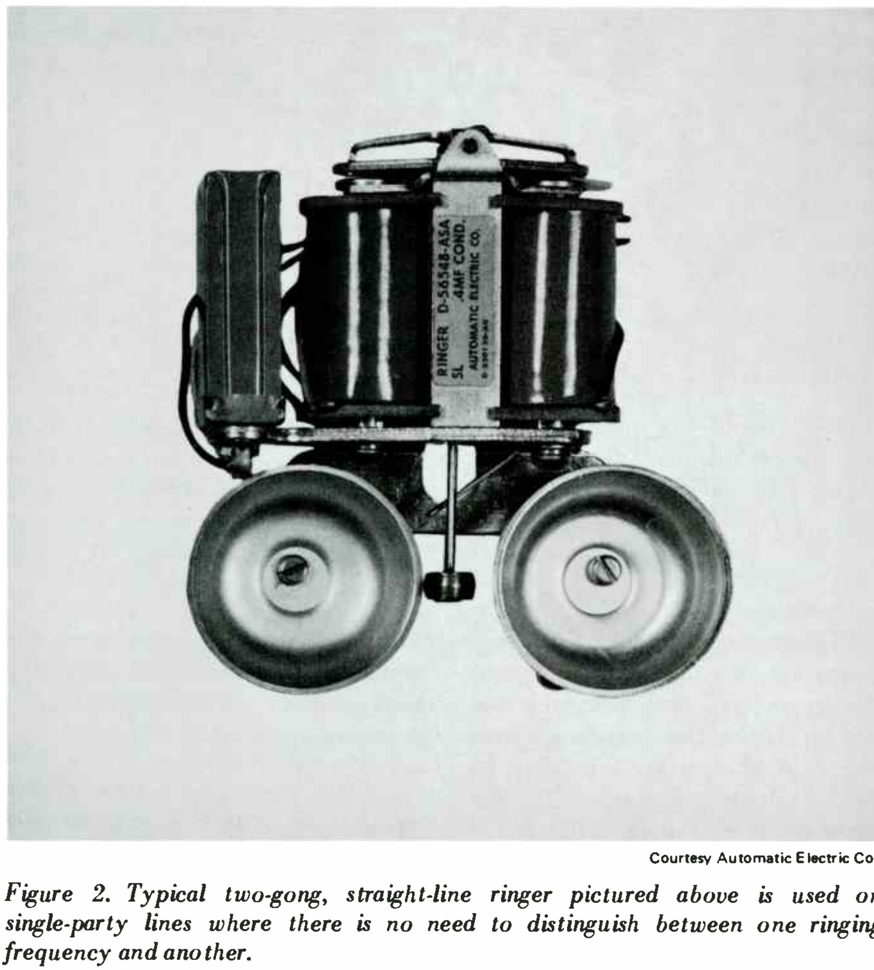
The magneto phone introduced about the same time fit well into the scheme and provided acceptable service even when the transmission path was a system of barbed wire fence. To make a call, the customer would crank the magneto to release a flag at the central office. When the operator answered, the customer would ask for the desired number – or in most cases, just the name of a neighbor down the road. The operator then patched the lines together, signaled the called party, and announced to both to "go ahead".

Strowger Switch

To Almon B. Strowger, a Kansas City undertaker, this intervention was suspect, for he was losing valuable business through the partiality of an operator. Because the operator was diverting calls to a competitor, he reacted by inventing the first automatic switch. The Strowger Automatic System was publicized as the "girl-less, cuss-less, out-of-order-less, wait-less telephone".

It accomplished automatic party selection by means of an electromechanical pawl-and-ratchet mechanism that moved a wiper over a bank of contacts, each connected to a different telephone (Figure 1). The calling telephone was permanently attached to the wiper, and by sending the proper number of pulses, the caller automatically guided the wiper to the correct contact, and as a result, the desired phone.

At first, pushbuttons were used for "dialing", but were followed by the



Courtesy Automatic Electric Co.

Figure 2. Typical two-gong, straight-line ringer pictured above is used on single-party lines where there is no need to distinguish between one ringing frequency and another.

rotary finger-wheel dial similar to those used today.

As the number of customers increased, party-line service was established with many phones sharing the same pair of wires. Now even more selectivity was required to satisfy customer needs. This was done by developing methods of identifying each customer by ringing.

In the beginning special codes were established in conjunction with the number of cranks of a magneto. But this did not provide full selectivity. Two or more phones would ring, and each customer would recognize the code assigned to him before answering his phone.

Later, more sophisticated methods provided fully selective signaling for each customer, and the job of ringing was transferred to a specialized section of the central office. Of the several different schemes developed through the years, only a few are still in general use.

The Ringing Circuit

All schemes share the same telephone ringing circuit, which is made up of a ringing generator and interrupter, a connector, and the customer's station ringing device. Their purpose is to direct a ringing signal to a desired party and to alert the party that he is being called.

The ringing voltage is either ac or a composite of ac and dc. The ac is supplied by the ringing generator, and the dc by the office battery. Types of generators include vibrating-reed, rotary, static-magnetic (sub-cycle), and electronic tube or transistor. Each fulfills a specific purpose in a central office depending on its frequency and capacity.

The first system rang the customer's bell continuously. But this was found to be irritating, and the interrupter was added. The interrupter is a mechanical device consisting of rotating cams whose peripheral lengths control the on/off timing of the ringing cycle. The standard interrupter ringing cycle for single-party service is 6 seconds – a 1.2-second ring followed by a 4.8-second period of silence. To equalize the load capacity of the ringing generator, the interrupter consists of five ringing groups sequentially connected to the generator for 1.2 seconds, or a total of 6 seconds.

In most exchanges ringing equipment is part of the station signaling rack. It usually operates continuously except in some small offices where the unit is on only when a call is made.

In response to digit dialing information, the central office connector or equivalent circuit used to complete a call checks whether the called line is idle. If so, a ringing voltage is applied to the called telephone. Simultaneously, a ring-back tone with the same on/off cycle as the distant end ringing lets the caller know that the phone is being rung. If the line is in use, a busy tone is sent back to the caller. Both the ring-back and busy tones are merely a subjective means to give the caller full control over the telephone connection.

Ringing current to the customer's phone uses the same physical wire pair or carrier-derived circuit as used for voice transmission. Some of the earlier systems, however, used a separate third wire for ringing only.

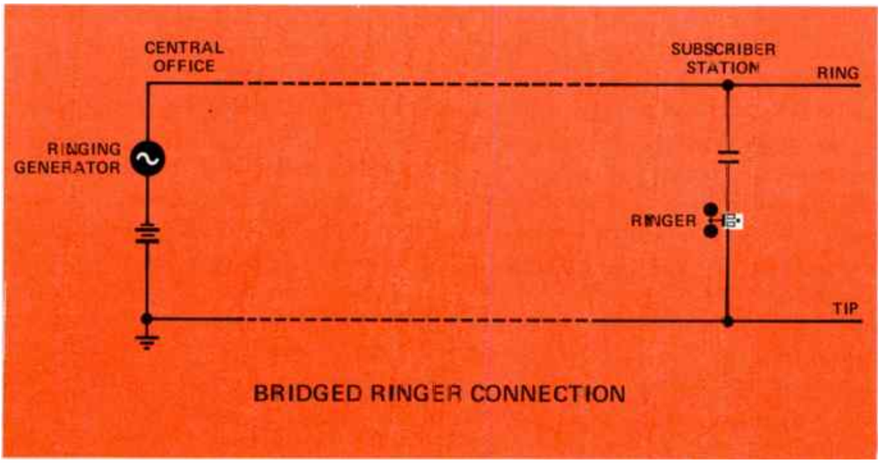


Figure 3. The connection for the bridged ringer is across the tip and ring transmission pair.

In order to signal a called party it is necessary to provide a ringing device at the customer's premises. It normally consists of a two-gong mechanical ringer.

The common two-gong mechanical ringer (Figure 2) is variously referred to as a polarized, biased, or straight-line ringer. All are the same device. Functioning parts include a two-coil electromagnet, an armature supporting a bell clapper, a bias spring, and two gongs. The armature is held to one side by spring tension to prevent bell tapping during dialing or accidental jarring.

One polarity of ac current causes the clapper to oppose the bias spring and strike one gong. The opposite polarity aids the clapper in returning to the other gong in the same direction as the spring.

In series with the ringer is a capacitor to prevent flow of direct current through the ringer coils. The capacitor coupled with ringer inductance resonates at about the ringing frequency (nominally 16-2/3 to 66-2/3 Hz) to increase ac current through the ringer coils and, thus, improve efficiency.

Ringer Connections

In general use are two types of connections for the ringer: *bridged* and *divided*. The bridged connection (Figure 3) has the ringer across the tip and ring transmission pair.

Divided ringing (Figure 4), also known as ground return ringing, uses either the tip or ring wire to ground. With one ringer connected tip to ground, and another connected ring to ground, two-party service in its simplest form is provided. The number of stations which may be served by divided ringing is double that for bridged ringing. And only one ringing frequency (usually 20 Hz) is needed to provide full-selective service to two customers over the same loop.

For multiparty service to four or more customers, there are a number of ringing schemes. Among the standard techniques are *frequency selective*, *superimposed* and *coded*. Frequency selective is commonly found in Independent telephone systems, whereas superimposed (or biased) ringing is used by the Bell System. Coded ringing, the simplest of all three, is employed by both.

Frequency Selective

Frequency selective, also called multifrequency ringing, makes use of five different frequencies to provide five-party service with a bridged connection, or ten-party service with a divided connection (Figure 5). Single-party and up to four-party frequency selective service almost always uses the bridged ringer connection.

Each station set is equipped with a mechanically tuned ringer whose reeds respond to a particular frequency. The three most common sets or groups of frequencies applied at the central office are *decimonic*, *harmonic*, and *synchromonic* (or anharmonic).

Decimonic frequencies are 20 Hz, 30 Hz, 40 Hz, 50 Hz, and 60 Hz (multiples of 10 Hz). Harmonic frequencies are multiples of 8-1/3 Hz: 16-2/3 Hz, 25 Hz, 33-1/3 Hz, 50 Hz, and 66-2/3 Hz. Non-multiple, synchromonic frequencies are 20 Hz, 30 Hz, 42 Hz, 54 Hz, and 66 Hz.

Decimonic and harmonic frequencies simplify the design of the ringing generator. However, "cross-ring" problems are sometimes encountered. For example, a 16-2/3-Hz ringing signal rich in third harmonics could give a weak ring or tinkle on a 50-Hz ringer. Additionally, power line interference could cause cross ringing

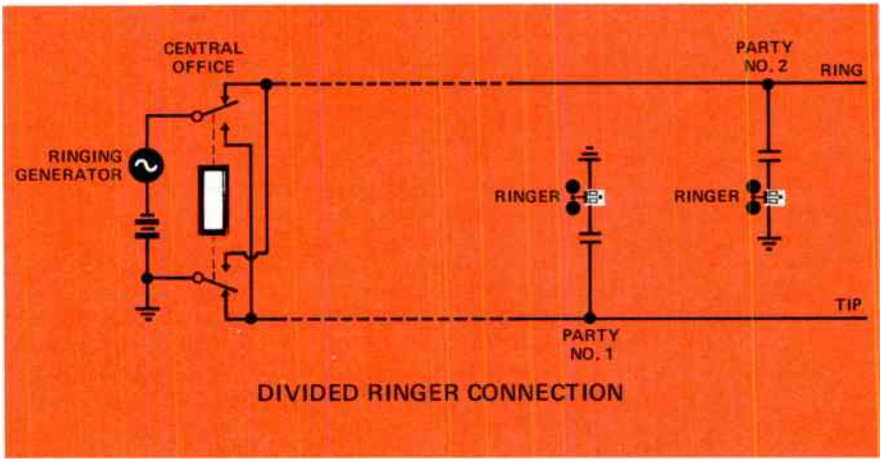


Figure 4. Divided ringing connects the ringer between either the tip or ring wire and ground.

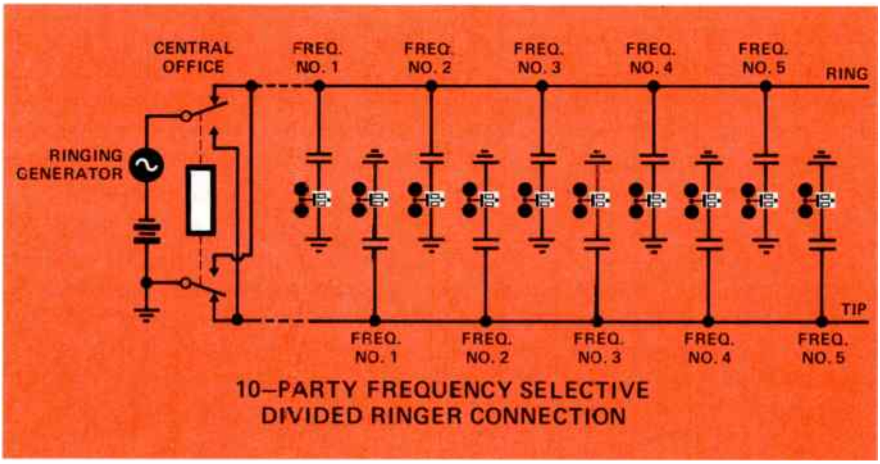


Figure 5. For ten-party frequency-selective service two sets of frequencies, either decimonic, harmonic or synchromonic, are used with divided ringer connections.

with the 60-Hz ringer of the decimonic frequency set.

Frequency selective ringing offers the greatest amount of fully-selective customer signaling. However, it does have one shortcoming – this is the requirement for five ringing generator frequencies and five station ringers.

Superimposed

Superimposed ringing uses both direct and alternating current to provide fully selective two-party bridged ringing or four-party divided ringing (Figure 6). The bridged connection is not generally used, however.

With superimposed ringing two sets of dc potentials of opposite polarities (± 38 to ± 48 Vdc) are applied to the tip and ring conductors for station selection. Telephones respond to only one of the two polarities.

Unlike frequency selective ringers, all superimposed telephone ringers are the same, but in series they have a three- or four-element, cold-cathode, gas-filled rectifier tube rather than the usual capacitor. For station selection the gas tubes are polarized for a particular polarity of ringing potential. The gas tube will pass the signal only if the 20-Hz ringing is superimposed on the proper dc voltage.

Superimposed ringing requires only a single-frequency ringing generator, which is a distinct advantage over frequency selective. However, it can only supply selective signaling to four customers.

Coded

Another form of multiparty signaling is coded ringing. It is nonselective since two or more phones on a party line are rung at the same time. Party identification is based on the number and duration of rings. Five ringing codes have been established consisting of combinations of shorts and longs.

Coded ringing requires only a single frequency ringing generator with the interrupter providing the codes. It can supply five-party service with the bridged ringer connection or ten-party service with the divided ringer connection. Combined with either frequency selective or superimposed schemes, it can provide semiselective service for up to ten parties.

Table A compares the different ringing schemes in general use, standard ringer connections for these schemes, the number of stations per line, and selectivity.

Revertive Ringing

Party lines do impose some special considerations when one customer desires to call another customer on the same line. A customer cannot make a call in the normal manner because once the call is initiated, the line is made busy. The term *reverting call* describes such a call on a party line, and there are several methods of revertive ringing depending on the ringer connection, ringing scheme, and number of parties on the line.

Two of the more common types are simultaneous revertive and alternate revertive ringing. Simultaneous revertive is normally used for coded ringing systems with either the bridged or divided ringer connection. The central office applies the called party's ringing code to the line and all ringers respond. When the called party answers his code, the ringing stops, and the calling party then picks up his receiver.

Alternate revertive finds use in bridged and divided frequency selec-

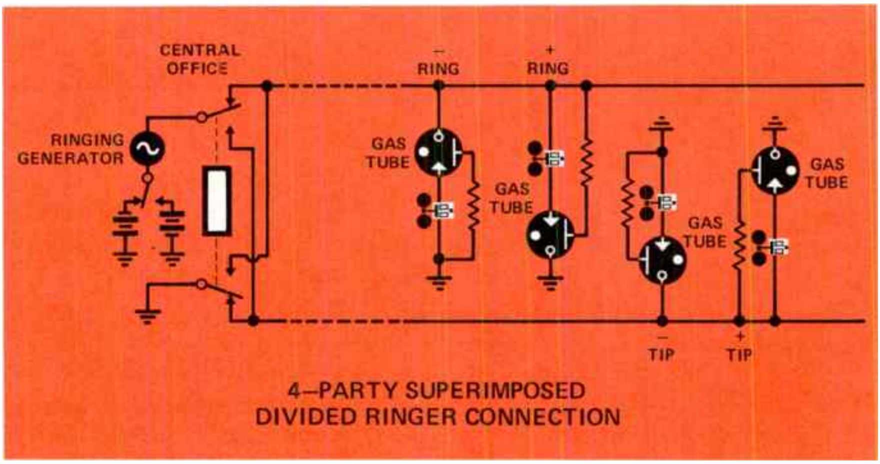


Figure 6. Four-party superimposed ringing uses a gas tube in series with each ringer, and a divided connection for the ringers. Each gas tube conducts the ringing current only if the applied superimposed dc voltage is of the correct polarity.

Table A.
Ringing schemes in general use.

TYPE OF RINGING	RINGER CONNECTION	STATIONS PER LINE MAXIMUM	RINGING SELECTIVITY
Single Party	Bridged	1	—
Single Party	Divided	2	Fully Selective
Frequency Selective	Bridged	5	Fully Selective
Frequency Selective	Divided	10	Fully Selective
Superimposed	Divided	4	Fully Selective
Coded	Bridged	5	Nonselective
Coded	Divided	10	Nonselective
Coded Frequency Selective	Bridged	10	Semiselective
Coded Superimposed	Divided	10	Semiselective

tive and superimposed ringing schemes. With this method the central office alternately applies ringing to the called party, then the calling party, etc. When the called party answers, the ringing stops.

TPL and TPS

For party-line systems, the central office linefinder is common to all circuits, but the connector has the option of being arranged on a terminal-per-line (TPL) basis or on a terminal-per-station (TPS) basis.

The TPL arrangement has one set of terminals for each party line. A final digit of the directory number identifies each party on the line.

The TPS arrangement, on the other hand, uses a separate set of terminals for each station on the party line, with unrelated directory numbers assigned to each customer. Generally used in expanding localities, it makes the most efficient use of office name codes and aids in providing full intercepting service. Also with TPS, a customer may be changed to a different transmission pair or moved from one party line to another without changing directory numbers.

Ringing on Carrier Systems

With all the combinations of ringer connections and schemes to signal customers on a party line, it is imperative that subscriber carrier systems such as Lenkurt's 82A, 83A, TFM and XU systems, accurately reproduce central office ringing. But with carrier systems, it is not possible to actually send the ringing frequency, the correct polarity, and separate the bridged or divided connection without special circuitry. And this circuitry must adhere to industry standards of reliability and ease of maintenance. Therefore, it must be straightforward and simple.

For example, the six-channel 82A Station Carrier System responds to ringing from the central office by turning the carrier on and off at the ringing frequency rate. A transistor switch at the subscriber unit detects the change in the carrier and applies about 80 Vac to the customer's ringer.

The ringing frequency applied to the 83A Single Channel Station Carrier System FM modulates the system's 64-kHz carrier. At the customer's station the signal is demodulated and applied as an accurate ringing signal.

Lenkurt's TFM Carrier System, like the 82A, turns the carrier on and off at the ringing frequency rate. Two in-band frequencies identify positive or negative superimposed ringing, and whether it is applied to the tip or ring wire.

The XU system operates similarly to the TFM but uses a 4-kHz out-of-band tone. All provide the widest flexibility without the need to make any changes to existing central office equipment.

Destiny

Specialized ringing schemes have provided important benefits by making possible the expansion of our nationwide telephone network and by aiding communications in general. The industry objective of one hundred percent single-party service in the next ten to fifteen years should eliminate the need for all of these special ringing schemes. The associated equipment will then perhaps become museum pieces along with the old magneto telephones and switchboards.

The Lenkurt DEMODULATOR

JULY 1968

DATA LOADING FOR MILITARY COMMUNICATIONS

Military communications must operate within unique requirements — including data loading different than CCIR and CCITT.

The impact of data and other forms of digital transmission on the modern communications system is increasing steadily. The demand for data circuits is, in fact, growing at a greater rate than for voice channels — especially in the military. This tug-of-war between data and voice has for the moment produced an interesting but solvable problem for the military system designer.

In the military communication networks, such as AUTOSEVOCOM, AUTODIN, and AUTOVON, where the signals are composed largely of digital forms, the problem becomes more complex for the system and design engineer. The critical point revolves around the FDM multiplexing/FM radio equipment, originally engineered for analog voice, but now called on to carry a higher percentage of digital signals at higher levels than originally designed. Because of this, a great deal of attention is being focused on the interrelation of data levels, loading, and noise performance in multichannel systems, especially those using microwave radio relay.

Loading

Loading, or load capacity, of a communications system may be defined as the volume of traffic that can be handled without exceeding the calculated distortion or noise originally designed into each link of the network. The actual physical makeup of the individual parts of a system determines the maximum load which the end-to-end system can handle. Load capacity, in terms of voice traffic, is traditionally (and realistically) measured on the basis of the probable load at the time of heaviest traffic — hence, the telephone term, busy hour.

Multichannel systems were typically designed to carry a specified number of voice channels during the busy hour, and at a load value derived from statistical evaluations. Loading has, in the past, been based on the equivalent load for a given number of voice channels. The formulas give the level of white noise that would be necessary to simulate the loading of a given number of channels. The equivalent load (P), expressed in dBm0 is:

$$P = -1 + 4 \log N \text{ (12-240 channels)}$$

$$P = -15 + 10 \log N \text{ (more than 240 channels)}$$

where

N is the number of channels.

Systems designed on these formulas are adequate if they are to carry only voice, or voice with a small percentage of data signals at a level higher than –15 dBm0 per channel. At –15 dBm0, voice and data can be mixed indiscriminately with no limitations.

The statistical properties of groups of tones used for data transmission are essentially identical to those of voice when the numbers are large. This is illustrated in Figure 1. If data is placed on the system in such a manner that the combined power of the data tones occupying a channel does not exceed –15 dBm0, the loading on the system will be essentially the same whether the channels are used for data or voice.

Unfortunately, –15 dBm0 is a rather low level for data transmission, especially in military applications. Restricting the data power to this level seriously affects the signal-to-noise ratio. It is for this reason that it is necessary to operate data at considerably higher levels, typically in the range of –8 dBm0 to –10 dBm0, and in some instances, as high as –5 dBm0. It is obvious that these are considerably higher than the –15 dBm0 average power level in a voice signal. Common data levels are shown in Figure 2.

Another factor to consider is that data is transmitted as a series of tones and presents a continuous load. Of specific interest are the levels at which these tones are presented and the relationship of the peak power signal to the power of the voice signal which it replaces. Consequently, as more and more channels of a system are shifted from voice to data, the total signal power – and, hence, the loading of the system – will increase.

System Noise

Noise in any form obscures the signal and causes transmission errors.

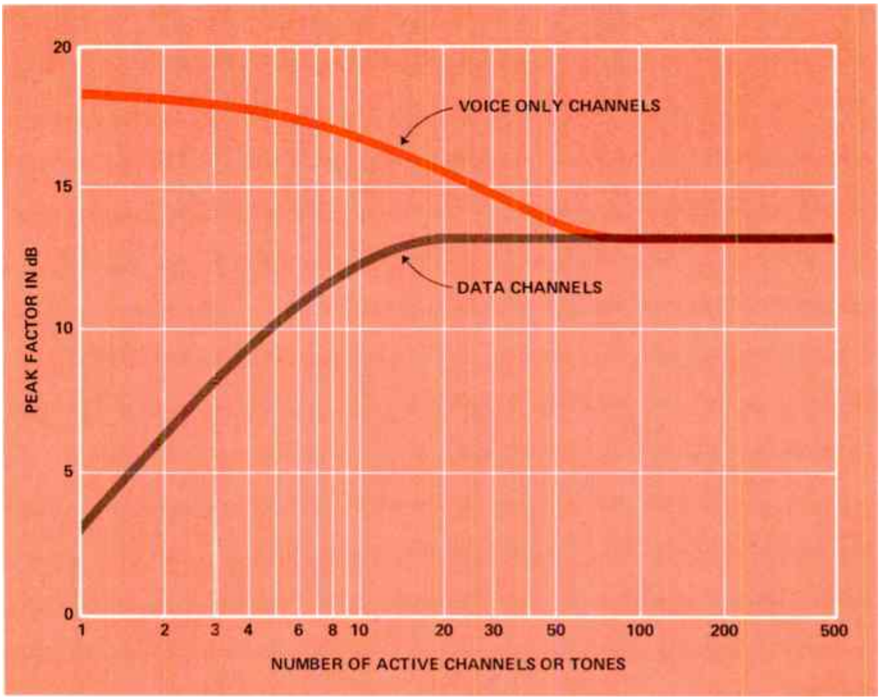


Figure 1. The patterns for the two peak factors – peak to rms ratios – of data and voice are essentially the same when the number of channels is large. However, restricting data to low levels affects signal-to-noise ratios.

	CCITT	U.S.
HIGH SPEED DATA	−10 dBm0 simplex −13 dBm0 duplex	−10 dBm0 switched − 8 dBm0 private line (− 5 dBm0 occasionally)
MEDIUM SPEED DATA		− 8 dBm0 total power
TELEGRAPH SPEED DATA		− 8 dBm0 total power
12 or less	−19.5 dBm0 each (−8.7 dBm0 total)	
18 or less	−21.25 dBm0 each (−8.7 dBm0 total)	
24 or less	−22.5 dBm0 each (−8.7 dBm0 total)	
PHOTOTELEGRAPH (FAX)	−10 dBm0 for FM 0 dBm0 absolute power for DSB—AM	− 8 dBm0

Figure 2. Commonly used levels for data applied to a voice channel.

Although noise is constantly introduced into the communications channel from the transmission medium and the equipment itself, it can be overcome by suitable design.

Actually, the amount of noise present is not as important as the relative strengths of the signal and the noise — the greater the signal-to-noise ratio, the better and clearer the transmission.

Within a communication system there are two basic types of noise: idle noise and intermodulation noise. Idle noise, present in the system at all times despite the absence of modulation, has accumulated a number of other names, among them are thermal noise and residual noise. This noise is basically in the electronics equipment itself and is of the random or "white noise" type. Idle noise varies inversely with receive input level, or signal level, while intermodulation noise varies with system loading.

Intermodulation noise is the result of nonlinearities in the equipment through which the signal must pass, and is a direct result of increasing the signal load. In an FM microwave radio, increased operating levels cause increased frequency deviation. More intermodulation noise is the result, although this is an effective method of decreasing some of the idle noise. While the highest level possible would be desirable in increasing the signal-to-noise ratio, a certain tradeoff must be made. As intermodulation noise increases, its effect is felt slowly at first. Then a "break point" is reached and it increases very rapidly. Near this level the optimum exists.

If the fixed amount of permissible deviation is shared by only a few channels, the signal-to-noise ratio in each channel will be quite good. As the number of channels is increased, intermodulation noise limitations demand that the per-channel frequency deviation be reduced, with a consequent increase in idle noise.

For a given system bandwidth, loading, and RF signal level, the per-channel deviation can be determined which will provide the best balance between idle noise and intermoduation noise. Beyond this, if the loading is increased, a different optimum deviation would apply, unless some other factor is also changed. Figure 3 illustrates the compromise necessary.

Effect on Equipment

Many segments of the communications network are affected by data loading, each in its own way. In the "hard wire" portion of the data systems – the path between the data modem and the multiplex equipment – the signal is at voice frequency as it travels through office cabling, line extensions and switches. Here, the signal is often subjected to high ambient noise levels, particularly the impulse type of noise which is very destructive to data though completely insignificant to voice operation.

There is also a rather strong threshold effect with data. Noise a few dB below the data determining level (i.e. mark or space) goes unnoticed. But, as it reaches this level, the noise will immediately result in data errors.

Most other noise, including intermodulation products in multichannel systems, is random and affects both voice and data. However, the effect of an increase or decrease in this kind of noise is far more dramatic on data than on voice. For example, an increase of only 1 dB in the signal-to-noise ratio provides a theoretical tenfold improvement in the data error rate. But, a 1-dB change is barely detectible in voice transmission.

To combat this type noise, the system designer may call for higher levels to improve the signal-to-noise ratio. There is nothing to overload in the "hard wire" region and crosstalk into other channels is about the only limitation on the level used for data.

Within the multiplexing process there are several stages involved. Channels are frequency translated to create groups, then supergroups, and then the line signal. The multiplex is the point where sharing of common equipment first comes into play. It is the vital stage at which relative values of data and voice power must be established.

It is important to note that once the relative values of voice and data are established at the input to the multiplex section, they cannot be changed throughout the rest of the system.

Voice vs. Data

Actually, the signal-to-noise ratio in good, modern communications systems is usually far in excess of what is needed for reliable voice communications. The subjective desire of the customer for added ease and convenience, rather than greater intelligibility, is good cause for commercial service to be offered with less noise.

In looking at how data loading affects the communications system, each segment needs to be treated separately. One portion is that part of the transmission system where the signal is in the baseband form. Included are the multiplex line equipment, the interconnecting cables, and the baseband portion of the microwave equipment. In this region the inputs, output and all intervening items are dedicated fixed circuitry. With proper design, no limitations on the relationship of voice and data are imposed.

In the hard wire and multiplex areas there are valid reasons for maintaining a relatively high data level with respect to voice. It is in this area that the bulk of impulse noise exists. Furthermore, it is not necessary to incur any technical penalty in these segments in order to allow data to be handled at a higher level.

In all of these areas, the data signal can be kept at relatively high levels, providing the multiplex equipment is designed for heavy power loading.

An important segment to be considered is the microwave radio system, where some natural limitations of a kind that do not exist in other parts of the system come into play. The signal is transmitted by FM, accomplishing a bandwidth-for-noise tradeoff. But while the medium itself does not impose any severe bandwidth restriction, a finite amount of spectrum is available – and this can be an obvious limitation.

The load carrying capabilities of a particular microwave system are a complex function of a number of factors, including system bandwidth, per channel deviation, RF signal level, and others.

Most of the complexities arise in this area, with two conditions pulling in opposite directions. It is desirable to keep the FM deviation high so that the signal level is well above that of idle noise. This is closely tied with the need to have adequate signal even during very deep fades. But, increasing the deviation causes intermodulation noise to increase.

Possible Solutions

There are several solutions available to the systems engineer when adding sizable percentages of data channels at

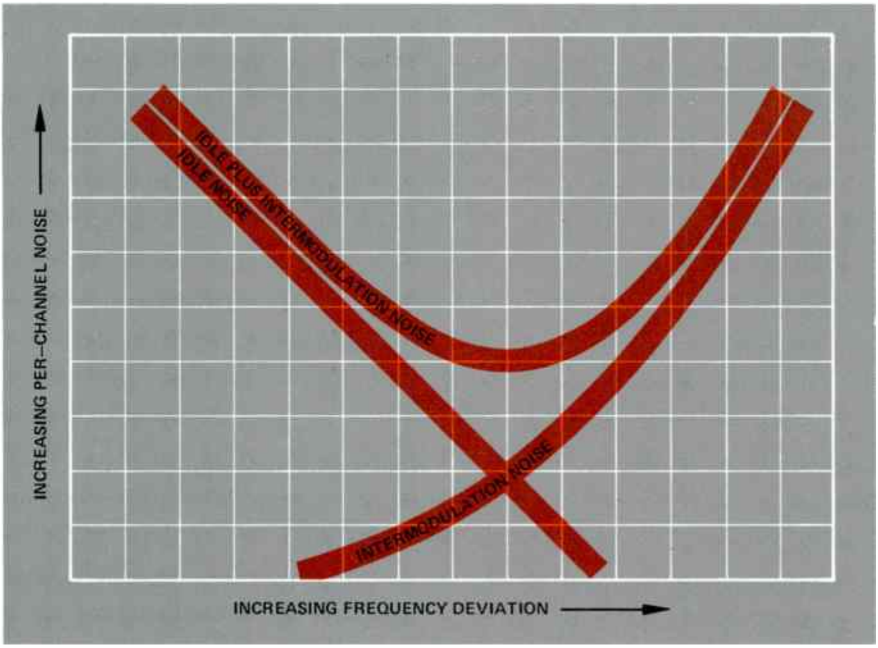


Figure 3. Idle noise is reduced in direct proportion to an increase in frequency deviation. But beyond a certain point peculiar to the equipment, the increasing intermodulation noise rapidly overcomes and reverses this advantage.

higher levels. The most obvious is to remove the equivalent number of voice channels for each data channel according to power load. For example, one –5 dBm0 data channel would require the removal of ten voice channels. While this is surely a workable answer, it becomes self-defeating – 30 data channels would use up the entire capacity of a 300-channel CCIR system.

A related and more widely accepted approach is to sharply limit the percentage of data channels and distribute them through the system so that no item of common equipment will be overloaded.

Another possibility is to drop data levels down nearer the equivalent voice channel load and juggle the total loading to provide an acceptable level of performance.

One solution centers on equipment design rather than operation procedure, and requires a multiplexer and radio capable of carrying higher loadings. This is the approach used in the AN/UCC–4, developed and manufactured by Lenkurt for the military. It will accept data levels as high as –5 dBm0 on all channels simultaneously. This capability allows complete freedom in assignment of channels that can be used for data, and almost no restrictions as to relative voice and data levels.

Military Loading

The Defense Communications Agency, realizing the effect of loading on the Defense Communications System, recently recommended certain revisions to the microwave standards of DCAC 330–175–1. DCA felt these changes were necessary to provide improved microwave equipments and subsystems of the DCS to support the DCA worldwide wideband transmission improvement program.

The recommendation for loading as suggested by DCA is:

$$P = -1 + 4 \log N \text{ (12-32 channels)}$$

$$P = -10 + 10 \log N \text{ (32-600 channels)}$$

The equivalent noise power for 600 channels under military loading is +17.8 dBm0. This is considerably above the recommended CCIR loading level – in fact, it corresponds more closely with the CCIR system designed for 1850 channels. Thus, it is obvious that the time-accepted CCIR loading formulas no longer apply to military systems.

Design Approach

Increasing the loading capability of the radio to achieve the acceptance of larger numbers of high level data signals is possible. This is the approach taken by Lenkurt in the design of the 75C microwave radio for military applications. Together with the AN/UCC–4 multiplexer the 75C offers a versatile communications package.

The impact of data on communications – especially on the DCS network – is forcing a critical look at operational standards and systems design. Loading formulas have changed to more accurately reflect the military communications environment, and improved transmission systems must meet today's needs while matching the design criteria of the future.

JUNE, 1967
VOL. 16 NO. 6

The Lenkurt DEMODULATOR

Coaxial Cable

Coaxial cable transmission systems have played an extremely important role in the phenomenal growth of the telephone industry in the U. S. and abroad. This article plots the history of the coaxial cable in communications, and describes some of the characteristics of the cable, the system, and its uses.

The heart of any communications system is the transmission medium over which the information signals pass. The makeup of the transmission medium places constraints on the design of the terminal communications equipment, such as multiplexing method, channel density and performance. These mediums include simple wire conductors, multipair cable, coaxial cable and microwave radio. Each medium has its own peculiar application advantages, and each plays an important role in our day-to-day communications.

The evolution of the coaxial cable in the 1920's—a significant structural innovation of a two-wire transmission line—has made possible the wideband, high-capacity communications of today.

Growth

In 1941, Bell System L1 coaxial cable routes were established between major metropolitan areas in the eastern United States. By 1948, a complete transcontinental coaxial cable facility was in operation. The L1 had a capacity of 600 message channels—an enormous amount compared to the few channels that could be transmitted over an open wire or multipair cable. Since microwave radio was not generally in use at that time, the coaxial cable medium was considered — and certainly was — the ultimate in multichannel communications.

As television became popular and network programming began to fill the airwaves, the coaxial cable seemed to be the ideal answer for conveying network broadcasts between stations. Although the L1 had only a 2.8-MHz bandwidth, performance was found entirely acceptable. The first TV application of the L1 was to transmit the Army-Navy game in 1945.

During the ensuing years the Bell System continued to develop multiplexing equipment and repeaters for more efficient use of the coaxial transmission line. In 1953, the L3 system went into service with an increased capacity of 1860 message channels, or 600 message channels and a 4.1-MHz TV signal on the same type of cable used for the L1.

More recently foreign systems have been developed with capacities up to 2700 message channels, and the Bell System L4 with 3600 channels is meeting the need for better utilization of existing and newly plowed-in cable routes in the U.S.

The coaxial cable has played an important part in long distance communications, accounting for about 25% of long distance services crisscrossing the country. Presently, about 13,000 miles of coaxial cable routes exist. An additional 10,000 miles is planned for the next five years.

The development of microwave in the late 1940's soon tended to stem expansion of coaxial cable systems. Microwave radio eliminated costly construction, right-of-way acquisition, maintenance, and other problems associated with establishing land lines. However, the relationship of microwave and coaxial cable proved to be valuable, mainly because the same basic multiplex equipment developed for coaxial cable could also be applied to the microwave baseband. Now, a second look is being given to coaxial cable in areas where allocations for microwave frequencies are not available.

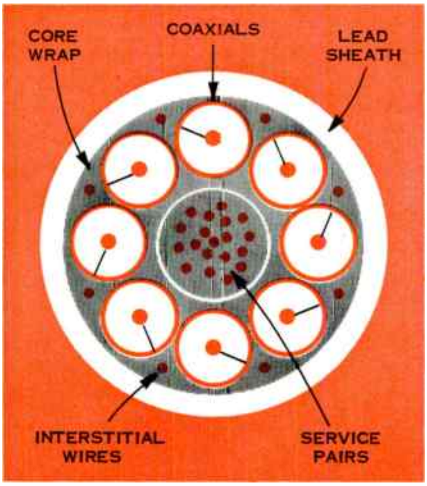


Figure 1. Typical communications coaxial cable consists of a number of "pipes" or "tubes" together with interstitial wires and service pairs inside a single sheath. Each pipe or tube provides one-way transmission for a large number of message channels or a TV signal.

Cable Construction

The communications coaxial cable consists basically of a single wire suspended in the center of a cylindrical conductor. The wire is held in the center of the tube by small disc-shaped dielectric or nonconducting insulators spaced closely together. Usually a number of these "pipes" or "tubes" (see Figure 1) are combined inside a single sheath.

Coaxial cable has a very low attenuation factor coupled with extremely good shielding from interference. In addition to its importance in the communications industry, other important uses of coaxial cable are associated with CATV and ETV, radar, navigation aids, aircraft, and test equipment.

Construction of the communications coaxial cable differs from the other types of cable which have the area between the inner and outer conductors separated by solid dielectric material. In addition, these types of cable normally have a braided copper outside conductor instead of the rigid copper tube, providing the needed flexibility for their particular use.

Disc insulated coaxial lines have much lower losses than the solid dielectric lines, but are more difficult to manufacture because of the mechanical problem of keeping the conductors concentric. The communications coaxial line with its spaced insulators approaches the ideal condition of having air as a dielectric, and is often referred to as air dielectric cable.

Included in the typical communications cable sheath with the coaxial tubes are interstitial wires and a cylindrical core containing service pairs. These added wires "round out" the cable. Around the cables is a layer of heavy insulation and a lead sheath. Interstitial wires may be used typically for v-f order wire between attended repeater stations, and for monitoring and control functions at unattended repeaters.

The service pairs, if provided, may be used as physical v-f circuits or with a cable multiplex facility.

Coaxial tube dimensions (see Figure 2) are usually described in terms of diameters of the inner and outer conductors. For example, when dimensions of 0.102/0.375 inches are given for a coaxial tube, this means that the outside diameter of the inner conductor is 0.102 inches, and the inside diameter of the outer conductor is 0.375 inches. More commonly, only the outer diameter is given; for example, 0.375 inches.

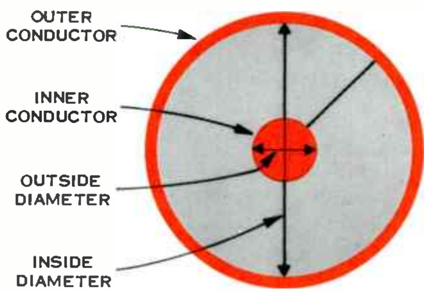


Figure 2. Coaxial tube dimensions are specified by the outside *diameter of the inner conductor, and the* inside *diameter of the outer conductor.*

Cable Sizes

The first cable installations in the U. S. were of coaxials having an outer diameter of 0.27 inches. Later installations were 0.375 inches, and this has become standard for long distance circuits. Other size cables in use include a 0.290 inch foam-filled dielectric cable used in Canada, and a "pencil gauge" cable of 0.174 inches.

The International Telegraph and Telephone Consultative Committee (CCITT) has established recommendations concerning the characteristics and performance of coaxial cable systems employing two standard size cables, the pencil gauge and 0.375 inch. According to their recommendations, the pencil gauge systems have a maximum bandwidth of about 6 MHz or 1260 voice-frequency channels, while the 0.375 inch systems have a capacity of 12 MHz or 2700 voice-frequency channels. Bell System technology has effectively exceeded this limit with the 3600 channels on 0.375 inch cable provided by the L4.

Characteristics

Electrically, what makes a coaxial cable attractive for communications is that it provides more conducting surface area than a two-wire transmission line and therefore suffers less resistance losses at higher frequencies. In addition, the electromagnetic energy propagation in a coaxial line is confined within the tube and isolated from outside interference or crosstalk because of its structure.

Generally, the effective bandwidth of a coaxial cable communications system is limited only by the required gain needed to maintain good signal quality. Spacing cable repeaters closer together makes it possible to increase the effective bandwidth by providing more amplification, and this approach has been used to increase the capacity of existing coaxial cable. However, economics and transmission reliability dictate certain limitations to such an approach.

Although a coaxial line will transmit signals down to zero frequency or dc, the lower practical frequency limit for communications is about 60 kHz. This is because the coaxial line does not provide good shielding at low frequency and because of equalization frequency limits.

The upper frequency limit for transmission in a given system is determined by cable dimensions, cable construction, and permissible attenuation. All three characteristics interact in such a manner that a compromise is usually made by giving appropriate attention to such factors as acceptable noise, repeater spacing, and amplification limits. The attenuation of a coaxial cable is given by the formula:

$$A = 40.1 \times 10^{-6} \frac{\sqrt{f}\left(\frac{1}{a} + \frac{1}{b}\right)}{\log \frac{b}{a}}$$

where:

A = attenuation in dB/1000 ft.

a = radius of inner conductor in centimeters

b = inner radius of outer conductor in centimeters

f = frequency in hertz

It can be seen from the equation that the attenuation of the cable varies directly with the square root of frequency and inversely with the size of the cable. Mathematically it has been proven that the minimum attenuation per unit length is accomplished with a ratio between the diameters of the inner and outer conductors of 3.6. With this particular ratio the impedance of a coaxial line, ignoring the losses of the dielectric, is obtained from the formula:

$$Z_0 = 138 \log \frac{b}{a} \text{ ohms}$$

Using $b/a = 3.6$, Z_0 is 77 ohms.

The insulating discs that support the center conductor of a coaxial cable represent shunt capacitive loading for the cable, and, therefore, lower the characteristic impedance and the velocity of propagation.

A coaxial cable having dimensions of 0.102 inches for the diameter of the inside conductor and 0.375 inches for the outside conductor has an attenuation of about 5.8 dB/mile at 2.5 MHz and a characteristic impedance of 75 ohms. A coaxial cable with dimensions of 0.047 inches and 0.174 inches has an attenuation of about 12.8 dB/mile at 2.5 MHz, and also a characteristic impedance of 75 ohms. See Figure 3 for a comparison of the attenuation versus frequency of the common types of communications coaxial cable.

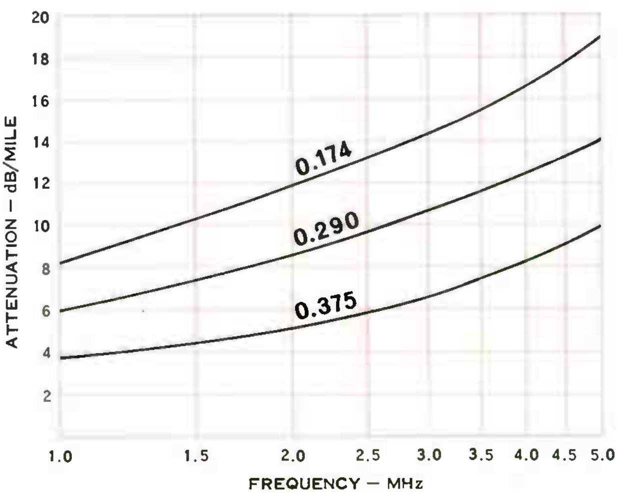


Figure 3. Response curves compare the attenuation versus frequency characteristics of the common types of communications coaxial cable.

The System

A communications coaxial cable system consists of a logical arrangement of repeater stations along the cable route. The basic requirement is that the cable network have uniform characteristic impedance, low losses and reflections, and proper protection from electric fields and disturbances such as lightning.

The original cable systems used vacuum tube repeaters. It was therefore mandatory that repeaters be spaced at wide intervals to increase reliability. With the invention of the transistor, more reliable repeaters were designed and power consumption was also substantially reduced. This made it possible to increase the number of repeaters and thereby increase the usable bandwidth of the coaxial cable.

A typical system usually contains widely spaced main repeater stations with several auxiliary stations situated between them. Customarily, the power feed for repeaters is through the center conductors of the coaxial pairs in a series loop from the main repeater stations. Hence, the maximum distance between main repeater stations is normally limited by the maximum voltage which can be efficiently applied to feed power from the main repeaters to the auxiliary repeaters. Intermediate repeater spacing depends on the loss of the cable and the problem of placing the repeater points at accessible locations.

Temperature variations are one of the most serious problems affecting the performance of a coaxial transmission system. Most of a coaxial system is buried underground, lessening the variations. Nevertheless, temperature-sensitive regulators must be employed to compensate for deviations in cable attenuation caused by temperature changes.

An example of a long-haul coaxial cable transmission system is the Bell System L4. This system employs three types of repeaters between main station repeaters: basic repeaters, regulating repeaters, and equalizing repeaters. Basic repeaters compensate for the normal loss of approximately 2 miles of cable. This repeater spacing is compared with 4 miles for the L3, and about 7.5 miles for the L1. Regulating repeaters spaced at up to 16-mile intervals provide additional compensation for changes in cable loss due to temperature variations. Equalizing repeaters at up to 54-mile intervals contain adjustable equalizers to compensate for random gain changes. These equalizers are remotely adjusted from the main station repeaters. Main station repeaters are spaced at up to 160-mile intervals. The main station repeaters contain all the functions of the other repeaters plus additional "mop-up" equalizers which compensate for un-

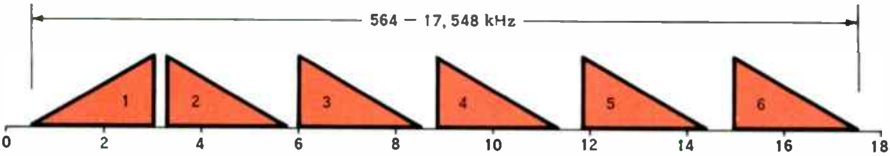


Figure 4. The modulation scheme used with the L4 combines six 600-channel mastergroups between 564 kHz and 17,548 kHz. Mastergroups 2 through 6 use the upper sideband while mastergroup 1 uses the lower sideband.

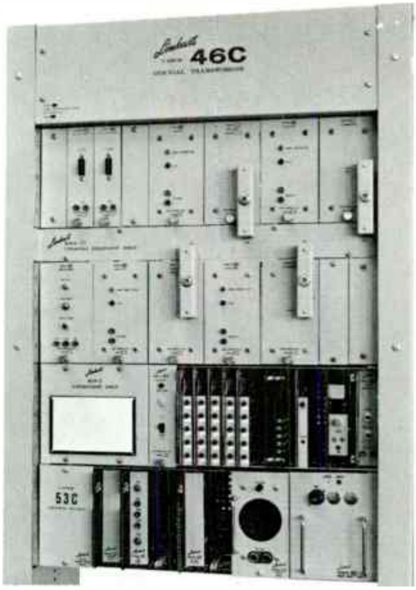


Figure 5. Up to 600 v-f channels may be processed by office arrangement of Lenkurt 46C Coaxial Transmission System shown above.

equal spacing of equalizing repeaters. These main station repeaters also supply direct current to the intermediate repeaters.

The multiplexing scheme for the L4 (Figure 4) is the combination of six 600-channel mastergroups with a frequency range of 564 kHz to 17,548 kHz. The mastergroups are separated by guard bands to permit dropping out any of the groups at a main station without demodulating the others.

Other Uses

The growth of the domestic telephone industry with each passing year decreases the number of frequency assignments available for microwave radio. For this reason increased emphasis is being given to the establishment of coaxial land lines as an alternative. Another use is in short haul cable extensions off backbone microwave radio routes. This particular application proves in some instances to be more economical than microwave.

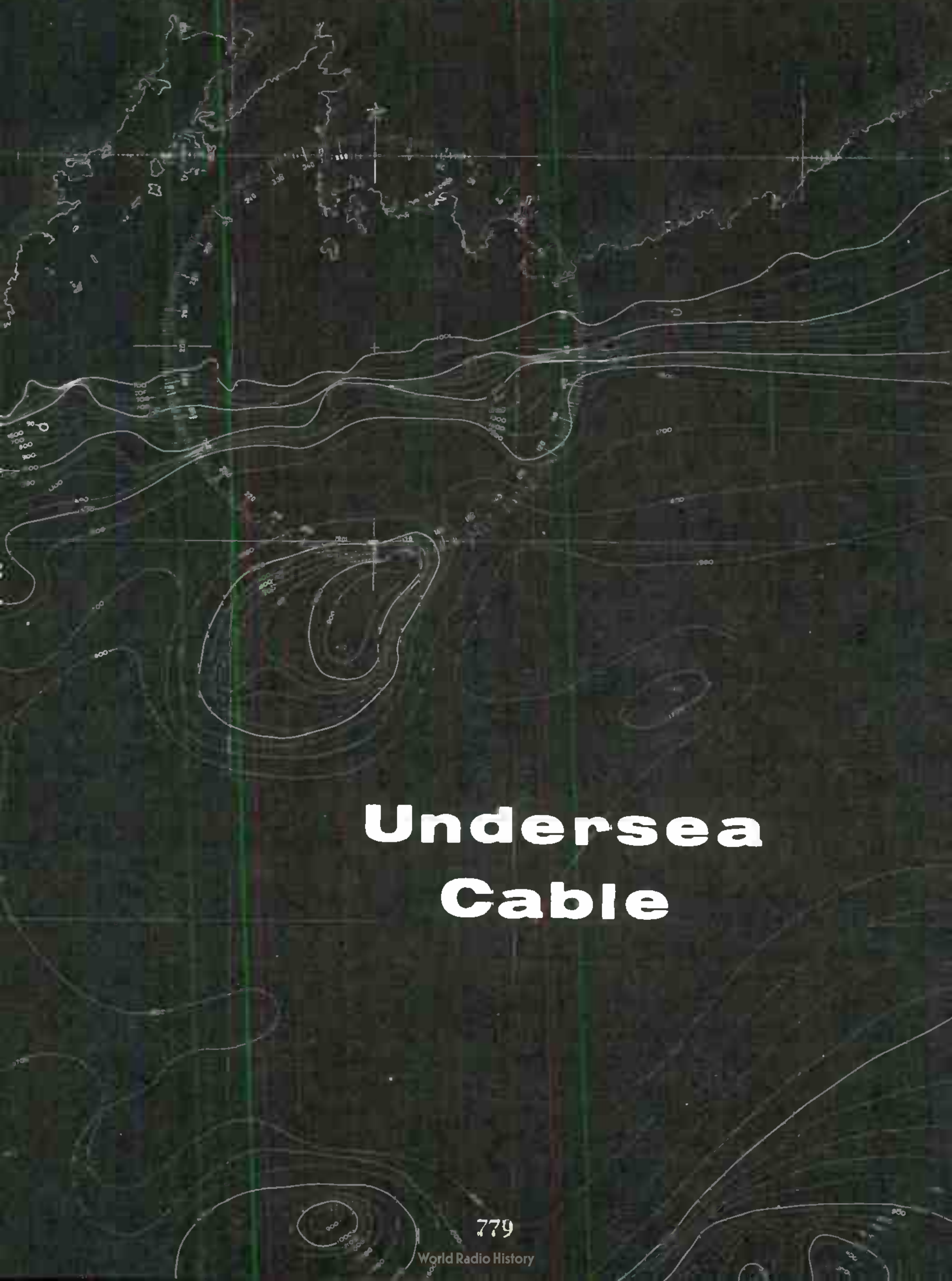
For example, the Lenkurt 46C Coaxial Transmission System—which complements the Lenkurt 46A Radio Multiplex System — provides for as many as 600 voice channels for interconnection between microwave radio installations and coaxial cable plant. The system permits transmission on 0.174, 0.290, and 0.375 inch cables. Repeaters along the buried cable are in watertight cabinets installed in manholes.

While the system capability is 600 channels, it can be proved-in for lower capacity systems by spacing repeaters at greater distances. For example, a typical system with an initial need of 60 channels on 0.174 inch cable would require repeaters at about 10-mile intervals. Expansion to 300 channels can be achieved by inserting intermediate repeaters at 5-mile intervals. Repeater spacing for 600 channels is 2½ miles. Selected repeater sites may be equipped for dropping and adding channels according to local needs.

The future holds growing applications for coaxial cable. For example, a pulse code modulation (PCM) system now under development at Bell Laboratories will carry 3600 to 4000 channels over coaxial cable. The digital transmission system, designated T4, will operate at 281 megabits per second and will be employed on long-haul toll circuits. Its commercial use is expected by the early 1970's.

JANUARY 1968

The Lenkurt® DEMODULATOR



Undersea Cable

Undersea telephone cable presents problems which have no parallel on dry land.

A dry land telephone cable system is relatively uncomplicated. It requires cable, of course, and uses repeaters energized by power sources available along the cable route. Whether above or below the ground, the system can be easily maintained because it is accessible.

When it comes to an undersea cable system, power and maintenance are not quite so simple. Power is not available four miles or even a few hundred feet below the ocean surface. Repair and maintenance services are hard to perform in a marine environment.

Yet power, maintenance, and the ocean environment are problems which designers of undersea cable have had to cope with and master.

Power for the water-isolated repeaters must come from land based sources. It must travel through the cable to distant repeaters. But the amount of power available is limited by the size of the cable which is itself limited by manufacturing and cable laying techniques.

Because power is fed from the shore end of the cable, it is a precious commodity. It must be carefully conserved. Each repeater – each power user—must be constructed to draw a minimum of power while producing maximum results.

In every phase of the design and construction of an undersea system reliability plays an important part. Again the isolation of an undersea system presents a unique challenge. For all practical purposes repeater maintenance is out of the question. Recovering a submerged cable takes hours and sometimes days.

As a result designers make every effort to use components which have low failure rates. Twenty years or more without a system failure is normal. To meet such a goal designers require components which have well documented use histories. This cautious approach has meant highly reliable undersea cable systems.

In addition the undersea cable itself must be strong enough to withstand pressures up to 12,000 pounds per square inch. The cable must also be light and pliable enough to withstand the high tensions of being lowered from an unsteady cable ship in the open sea.

The challenges of power, environment and reliability were met over a number of years. They are challenges peculiar to telephone cables—a johnny-come-lately to the world of undersea cable.

History

The first transatlantic cable, laid in 1858, was a telegraph link between the Old and New Worlds. It lasted approximately 20 days, carried 732 messages and allegedly saved the British government 50,000 pounds. It also took six-

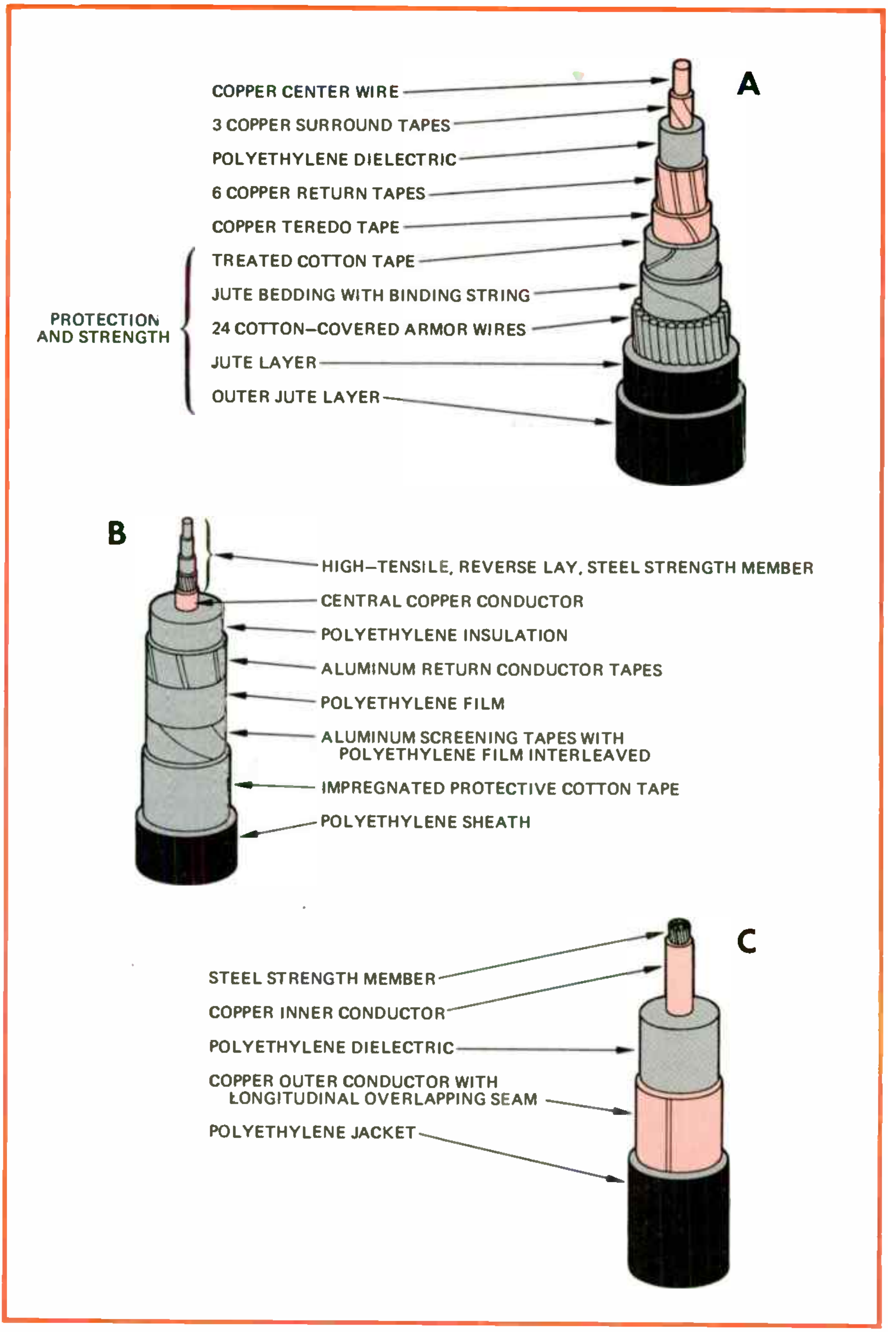


Figure 1. Armored cable shown in A was used in the SB system. B and C show armorless cable developed by the British Post Office (B) and American Telephone and Telegraph (C). All three cables measure 1-1/4 inches in diameter.

teen and a half hours to pass a 99 word message from Queen Victoria to President Buchanan—a long time for any woman, let alone a queen, to wait.

The next cable went down in 1866. This one lasted. In fact it lasted so well that the next major change to transatlantic communications did not come until 1927. In that year radio telephone bridged the Atlantic.

It was not until 1956 that the first transatlantic telephone cable connected Scotland with Newfoundland. Actually Alexander Graham Bell had tried without success to complete a transatlantic telephone call over existing telegraph circuits in 1879. At that time not enough was understood about bandwidth and attenuation to appreciate the reasons for Bell's failure.

Today the reasons are well known. Commercial telephone transmission requires much greater bandwidth than does telegraph. The greater bandwidth, in turn, requires higher frequencies which means more attenuation.

Only short undersea telephone systems were possible using telegraph cable technology. In the 1920's short systems to Havana and the Catalina Islands were laid from the United States mainland. Similar short systems connected the British Isles with the continent.

Breakthrough

But these systems could not have evolved into the three and four thousand mile undersea systems which exist today without the introduction of the submerged repeater. The British Post Office developed the first submerged repeater and put it into service in 1943. The American Telephone and Telegraph Company laid the first deep water repeater in 1950 between Key West and Havana.

The full story of undersea cable is not limited to the repeater—as significant as it is—or to modern communications technology. An undersea telephone cable system must also conform to recommendations and requirements of oceanographers and seamen.

A system, for instance, must be laid along a path as free as possible from deep trenches and jagged undersea mountains. It must be laid smoothly, steadily and at a reasonable speed. It must be resistant to the corrosive effects of water and boring of marine animals. Taken as a whole the system must be strong enough to support four

Figure 2. A comparison of three undersea cable systems developed by the American Telephone and Telegraph Company, showing growth of undersea cable capabilities.

	SB*	SD	SF
Capacity (3 kHz Channels)	48	128	720
Top Frequency On Cable	164 kHz	1.1 MHz	5.9 MHz
Cable	Two-0.620" Armored	One-1.00" Armorless	One-1.50" Armorless
Repeater Type	Flexible Vacuum Tube	Rigid Vacuum Tube	Rigid Transistor
Components Per Repeater	67	205	161
Repeater Spacing	44.5 Miles	23 Miles	11.5 Miles
Maximum System Length	2530 Miles	4025 Miles	4600 Miles

*Data for SB system based on operation after installation of new modulation scheme.

or five miles of its own weight in water.

The system which has evolved consists of coaxial cable, repeaters and equalizers. At the shore end are special terminals for multiplexing signals and supplying power, and fault location equipment. It is a fully complementary system, each part having been built specifically for undersea use.

The Cable

The first deep water telephone cable was similar to its telegraph counterpart. The only appreciable difference was a concentric return conductor added to form a coaxial structure.

The cable had a copper center wire surrounded by three thin copper tapes as its electrical member. A solid dielectric separated the center wire from a helix of six copper tapes. The solid dielectric—made of polyethylene—was necessary because of the high water pressure on the ocean bottom. Around these electrical members were several layers of protective and strengthening materials.

Telegraph cable did not use copper tapes but usually had strands of copper wire. Both cables were armored by wire rope and were further protected by tar, linseed oil and pitch.

The important difference between the telephone and telegraph systems was, of course, the repeater. Telegraph systems had operated for years without them, but telephone systems could not get along without periodic boosts from repeaters.

Even within the telephone community the undersea repeater made an important difference. In fact the use of different repeaters turned the first transatlantic system into two systems.

The first transatlantic telephone cable system — in spite of being two systems — was a triumph of international cooperation. It was the result of coordination between governments and private businesses in at least three countries.

The final venture included the active participation of the American Telephone and Telegraph Company, the British Post Office and the Canadian Overseas Telecommunications Corporation. The Americans and British were responsible for planning and laying the system.

Stiff or Soft

The project was divided into a deep water section—the American sphere—and a shallow water section which the British controlled. In both sections deep water repeaters were used, but in the interest of reliability and to avoid laying problems at sea, all concerned agreed that an American developed repeater should be used in the deep water section.

The American repeater had two advantages. It had a longer history of successful deep water operation, and it was a flexible repeater. To an extent the repeater behaved like a section of armored cable twisting with the tensions experienced during laying.

The British repeater was a rigid instrument which could not conform to a cable's twisting. At mid-ocean depths, where several miles of cable stretch under tension between ship and ocean bottom, the rigid repeater resisted the tensions placed on an armored cable. Such resistance causes damaging kinks and loops in the cable.

Both the British and Americans agreed that the risk of kinking was too great to try the British repeater in deep water laying operations. In addition the participants felt that a flexible repeater system could be handled and stowed aboard ship more easily and economically than could a rigid repeater system.

Irony

Ironically, the rigid repeater had a higher capacity than the flexible repeater. Its size did not impose the severe component limitations placed on the flexible repeater. As a result the rigid repeater system was able to accommodate 60 two-way voice channels on a single cable.

Even more ironic: the Americans were forced to lay two cables in the more difficult deep ocean section of the route because of the limited capacity of their repeater. The deep water system was a physical four-wire system using two cables of thirty-six 4-kHz voice channels each. Each cable carried voice transmissions in one direction through a string of 51 repeaters approximately 44.5 miles apart.

The cable itself was manufactured in lengths of about 200 miles, called blocks. During the laying of each block transmission measurements were made and analyzed aboard the cable ship. From the analysis the cable was equalized to correct for deviations in the cable arising from manufacture, temperature, depth and pressure.

Different types of cable were used in the system but the differences were physical rather than electrical. In shallow water up to 1300 feet, cables designated either type A or B were used. Both types had more protective and strength members than did the deep water, type D cable.

The heavier outer jacket in cable types A and B was necessary because of the frequent natural and man made disturbances which occur in shallow water. The type D cable did not need as much protective material because the deep ocean bottom is more serene.

Growth and Expansion

Since the first transatlantic system others have followed. There are now six coaxial cables spanning the Atlantic. Two cross the Pacific. Currently under construction is a system which will link Cape Town, South Africa and Lisbon, Portugal.

With this growth have come changes. The system used in the deep water section of the first transatlantic cable—dubbed the SB cable system—has been altered and has itself given way to radically different cable systems.

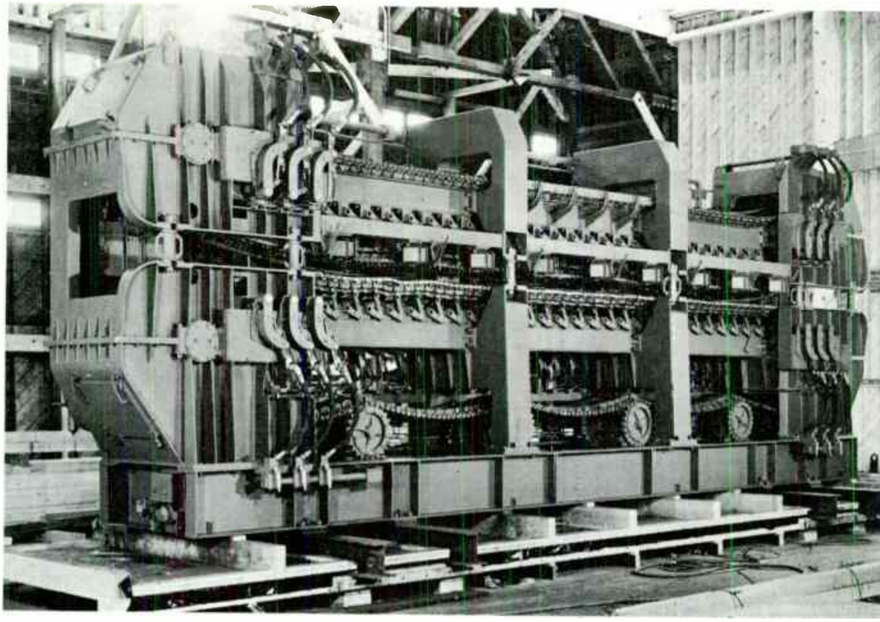
Originally the SB system had thirty-six 4-kHz voice channels. To optimize the use of these channels TASI (Time Assignment Speech Interpolation) was applied.

TASI made it possible to switch unused speech channels to a talker within milliseconds and switch away again to another user when the first talker stopped to listen. In effect TASI doubled the number of speech channels available.

In 1959 a new modulation scheme, called double modulation, was introduced to the SB system which reduced the 4-kHz voice channel to a 3-kHz channel. With double modulation it was possible to obtain 48 voice channels in the same frequency range that had carried 36 voice channels.

To cram the additional 12 voice channels into the system the band edges of adjacent channels had to be put 100 or 200 Hz apart, depending on the channel. This was much closer than the 800 Hz used for the 4-kHz channels. To support the closer channels much sharper cut-off filters were required. This made it possible to use 95 percent of the gross frequency band available.

Development of TASI and the slicing of frequencies are achievements which have no parallel in land cable systems. The developments do fit in with continuous efforts by undersea system designers to knock down formidable obstacles—inaccessibility, lack



COURTESY A T.&T

Figure 3. Cable engine used aboard C.S. Long Lines. *The engine is designed to run at payout speeds compatible with ship speeds of 8 knots.*

of power sources, the ocean environment, system capacity.

Armorless Cable

Another obstacle fell in 1961 when the English and Canadians teamed up to lay a second transatlantic telephone system. They used rigid repeaters in their entire system and added something new—armorless cable.

The armorless cable system brought its own circle of improvements starting with the cable core diameter. It increased from a 0.62 inch diameter to a full inch.

The overall diameter of the new cable was the same as the old, but the larger core gave the new cable 2/3 the attenuation of the old. Its expanded core made it possible to increase the line voltage from 2000 volts to 4000 volts which made it possible to put more repeaters on the system.

Finally, armorless cable made laying the rigid repeater easier. With the increased core size, the strength member could be put inside the central conductor. Placing the strength member there minimized torque tension coupling, thereby preventing the twisting and stretching characteristic of armored cable.

Reducing the tension had further advantages. With armorless cable it was possible to get a more consistent and predictable sea bottom performance. The risk of kinking caused by mechanical discontinuities at the rigid repeater decreased.

The cable itself used reverse lay strands—strands wrapped together in one direction enclosed by other strands wrapped in the opposite direction—for their center strength member. The reverse lay overcame internal torque.

The strength member was enclosed in a copper conductor surrounded by polyethylene. Around the polyethylene was a spiral of aluminum tape and around it a cover of overlapping turns of aluminum foil.

Armorless American Style

In 1963 a United States to England system went into operation also using armorless cable. It was American Telephone and Telegraph's SD cable system which used cable developed in the United States. It had a single lay strength member surrounded by a copper inner conductor. A polyethylene dielectric separated the inner and outer conductors.

Both the American and the English armorless cable were sealed in a thick polyethylene jacket. The resulting armorless cable was not as strong as armored cable, but because the newer cable was lighter, it retained the same strength-to-weight ratio.

The SD system employs rigid repeaters. These repeaters, like the flexible repeaters, contain a feedback amplifier which gives the system a wider frequency response with less distortion.

A common unit amplifies both directions of transmission by the use of directional filters. With the increased room in the rigid repeaters parallel amplifiers can be included which give added protection against failure.

Rigid Repeater

The SD system repeater is considerably more complex than the flexible repeater. It contains 205 components—about 3 times the number used in the earlier SB repeater. The new system carries 138 3-kHz channels in each direction. (It originally carried 128 voice channels.) The channels are derived by conventional frequency division multiplex.

Pilots in each group modulator are used for monitoring, equalization adjustments and automatic switching. Both the low band (108-504 kHz) and the high band (660-1052 kHz) have order wire channels. One of these channels is split so that it can be used for voice and teleprinter exchange.

The completed system has 182 repeaters, spaced every 23 miles. An equalizer follows every tenth repeater.

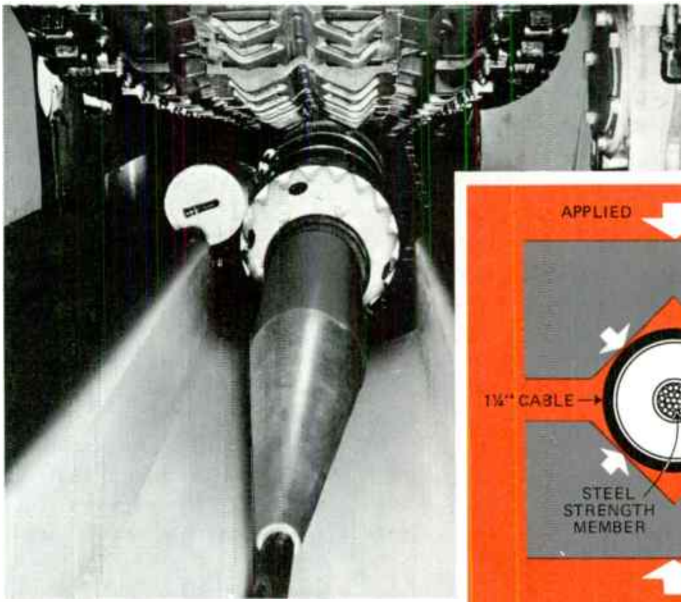
The repeater is made up of five sections with amplifiers and directional filters in the center three sections. Power separation filters which separate the power and information signals are at each end of the repeater.

Both the power separation filters and the directional filters create spurious feedback around the amplifier. This makes it necessary to use two transformer-regulated, symmetrical paths to cancel the unwanted signals.

The majority of the electrical components used in the SD repeater are similar to those in the earlier SB repeater. To fill new needs several new types of components were introduced but only after extensive testing. In all cases, each component and the whole repeater had to meet the reliability requirement of the earlier system—20 years of continuous operation.

The 500 pound repeater and housing are subjected to some 1700 tests. One test can find holes so tiny it would take 26 years to get a thimble of gas through them. The repeater is 50 inches long and 13 inches in diameter.

At the input and output, gas tubes protect the repeater against high voltage surges. The entire system – cable, repeaters and equalizers – requires 11,000 volts fed from 5500 volt power supplies at each end of the 4000 mile cable. The system draws 389 milliamperes of current.


COURTESY A.T.&T.

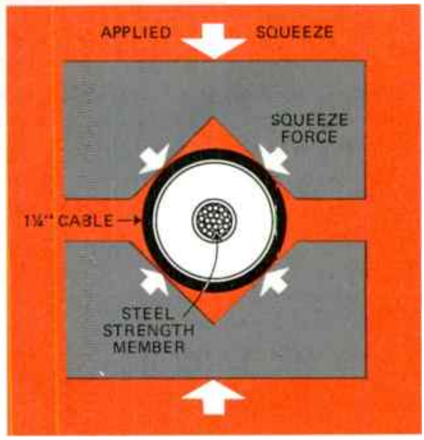


Figure 4. End view of cable engine as repeater enters engine. Insert shows how tracks grip a piece of cable. When repeater goes through, the flexible tracks separate to give the necessary clearance.

Special Ship

With the rigid repeaters and armorless cable came another development in undersea cable technology—a new cable laying ship, *C.S. Long Lines*, specially built for the American Telephone and Telegraph Company. The ship incorporated several innovations in cable handling.

Historically, cable laying has revolved around the circular drum. To use the drum, cable had to be bent around the drum's diameter. Sometimes this meant winding the cable around the drum several times.

While this was less a problem for armorless cable than for armored cable, it was a considerable problem for the rigid repeater. Complicating this was the requirement to lay cable and repeaters continuously at high speeds.

To avoid bending the repeaters a special cable engine with flexible, tractor-like tracks was developed and installed aboard *Long Lines*. The cable, repeaters or equalizers were fed between two of the tracks and pulled along by V-shaped blocks which gripped them at four points.

The engine was only one development. In place of a sheave with a diameter greater than 7 feet—required for rigid repeaters — a chute was molded into the stern of the ship's hull. The chute made it possible to pay out the cable and repeaters with a minimum of bending stresses.

At the bow of *Long Lines* a cable repair and recovery system was installed. The installation was a little more conventional, using large, wide drums which permit the passage of a rigid repeater at slow speeds.

On the Bottom

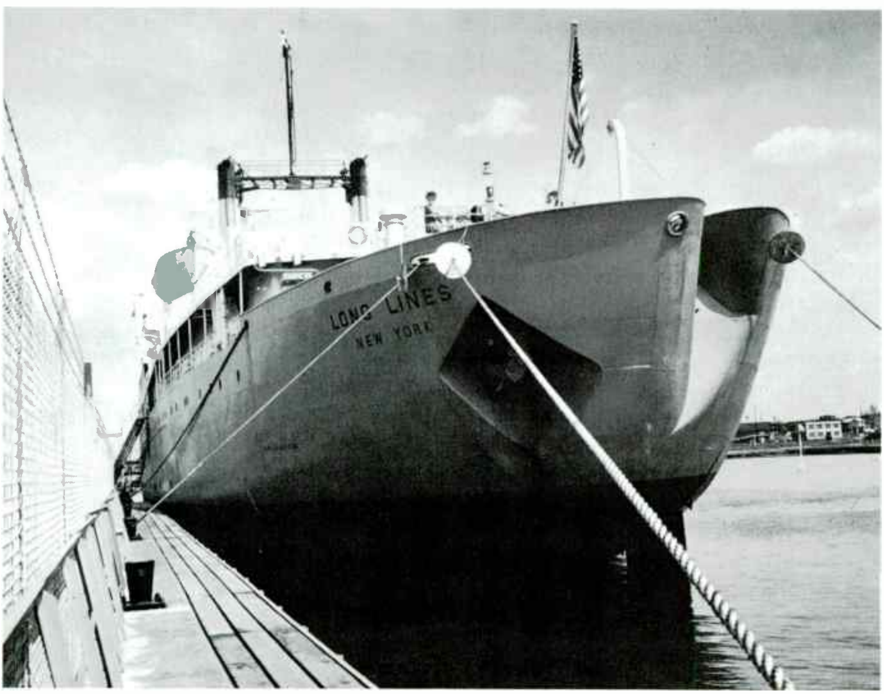
The first undersea cables were laid along existing shipping routes without much concern for the condition of the ocean bottom. Today a sizable amount of preliminary survey work is done to determine the best cable route. Ideally, such a route should avoid deep ocean trenches and steep grades, stay clear of centers of earthquake activity and the rough mountain ranges on the ocean bottom.

During the installation of one of the Pacific cable systems, oceanographers had to chart vast mountain ranges, deep trenches, thousands of volcanic seamounts and scores of live volcanoes in order to find an acceptable route for the cable.

While planning a section near Guam, it took six passes over the Magellan Seamount to find a safe passage. At the Marianas Trench (almost 6 miles deep) oceanographers searched for and found a natural bridge for the cable four miles down.

Remarkably it has been the forces of nature and man that have caused the most cable damage. Earthquakes and landslides are believed to have cut and washed away lengthy sections of cable. Other breaks have been caused by ship's anchors or trawlers dragging their nets.

In one study of recovered telegraph cable, it was found that 36 cables had suffered from trawler damage, 12 from corrosion, and 5 from



COURTESY A T & T

Figure 5. Stern of C.S. Long Lines *showing special chute to accommodate rigid repeaters. Most cable laying innovations incorporated in the ship's design were pioneered by the American Telephone and Telegraph Company.*

chafing. Four more had either deteriorated, been crushed or had telegraph repeater failures and three had armor pinches or tension breaks from a ship's anchor.

Most cable damage happens in shallow areas where the cable is not protected by several hundred feet of deep water. With this in mind, the American Telephone and Telegraph Company has begun to bury shallow water sections of cables.

Coming Up

The next generation of undersea cable will differ somewhat from the SD system. The new SF system will use diffused germanium transistors and a cable diameter of 1.75 inches.

Electrically the SF system is a direct successor of the SD. The cable construction remains the same. The repeater will be rigid but will contain only 161 components—44 less than the SD but 94 more than the SB system.

It does have definite advantages. Its 720 voice channels is one of them. Another is that a 4000 mile system will need only 3500 volts fed from each shore terminal and will draw 136 milliamperes of current.

The system will operate at higher frequencies than its predecessors—564 kHz to 5884 kHz—which will mean putting the repeaters closer together. In the new system there will be repeaters every 10 miles. Equalizers will still come every 200 miles.

Why Bother?

With the advent of satellites it might seem impertinent to talk about expanding undersea telephone capacity. Even in their infancy satellites can provide bandwidths which are just barely possible with the most advanced undersea systems.

But to look at the satellite as an immediate replacement for undersea cable systems is to overlook the virtues of each.

Satellites usually carry several repeaters in parallel, thereby avoiding complete system failures caused by the loss of a single repeater. In addition the terminal points in a satellite system can be changed, making it possible to re-route traffic when necessary. Being able to switch from one terminal to another gives the satellite system a flexibility which undersea cable systems do not have.

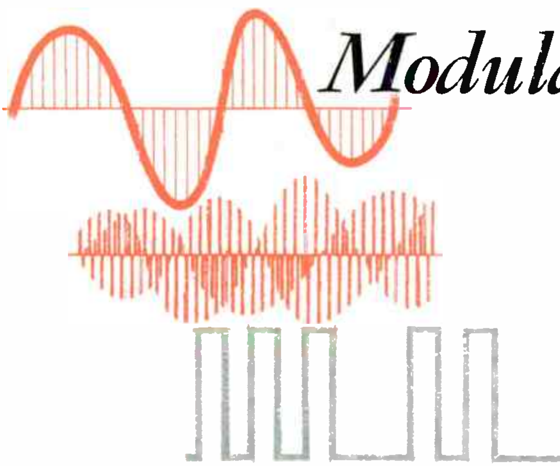
But undersea systems do not require the large, expensive terminals satellite systems do. In fact their fixed terminal points make undersea systems ideal for daily, well established international telephone service. Finally the ocean floor does not limit the number of undersea cables as much as does the area available for synchronous satellite orbits.

In the final analysis the two systems are complementary. With the growth of international communications both cable and satellites will have to share that growth.

NOVEMBER 1966
VOL. 15, NO. 11

the Lenkurt Demodulator

Pulse Code Modulation



The constant search for better communications at lower cost has led the telephone industry to a radically different method of transmitting speech information. This method, using binary digital pulses rather than conventional analog signals, provides high quality transmission and has proven to be very economical in short-haul carrier systems.

Up to the present time, telephone communications has been accomplished almost entirely with an analog electrical carrier wave which is varied continuously in proportion to the speech signals. Such systems have always been haunted by noise and crosstalk.

In recent years, the telephone industry has shown great interest in a method of coding speech information into digital electrical pulses. Unlike analog signals, these pulses, after becoming distorted by noise during transmission, can be completely regenerated at repeaters along the transmission path and at the receive station.

In the nineteenth century there were many attempts to code speech and music into digital electrical signals for transmission, using the techniques employed in telegraphy. Unfortunately, early experimenters did not have the mathematical tools provided by what is now termed information theory and were denied success because their coding schemes were too simple and did not convey enough information. Before they were able to advance their coding techniques, Alexander Graham Bell successfully transmitted speech using an analog electrical signal. The success of Bell's experiment was so immediately overwhelming that an immense telephone communications industry revolutionized around analog speech transmission.

Because of the outstanding success of analog transmission techniques, such as frequency division multiplexing (FDM), many years passed before serious attention was given to other methods of transmitting speech signals. However, with the ever-present problems of noise and crosstalk and the rising complexity and cost of electrical filters and other devices found in frequency division systems, it was certainly natural for engineers to search for more practical and efficient transmission methods. One of the most significant methods under investigation has been *time division multiplexing.*

It was demonstrated experimentally even before the development of FDM that time division techniques could be used to transmit many speech messages simultaneously over the same circuit. But such techniques could not be put into practical use at the time because of the limitations of mechanical devices for high-speed switching. The invention of the vacuum tube and the electric wave filter made frequency division multiplexing much more attractive for use in telephone transmission systems. However, researchers continued to investigate time division methods.

The first useful time division multiplex systems were developed in the early 1930's. In these systems a number of circuits share a common transmission path but at separate time intervals. Time division systems employ some type of pulse modulation, in contrast to the more familiar amplitude and frequency (AM and FM) techniques used in FDM.

The most popular type of pulse modulation has been pulse amplitude. In pulse amplitude modulation (PAM), a continuous signal, such as speech, is represented by a series of pulses called samples. The amplitude of each sample is directly proportional to the instantaneous amplitude of the continuous signal at the time of sampling. Since the amplitudes of the samples are continuously variable, the problems of cumulative noise and dis-

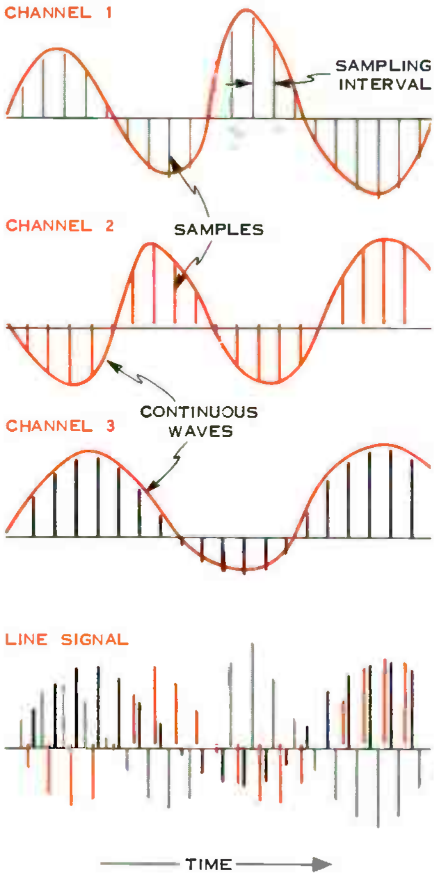


Figure 1. Example of time division multiplexing using pulse amplitude modulation. A pulse amplitude sample is placed on the line as each channel is sampled in turn.

tortion associated with analog signals are present in pulse amplitude modulation systems.

In 1937, Alec H. Reeves, then a member of the International Telephone and Telegraph Corporation Laboratory in Paris, resolved that the problems of cumulative noise and distortion could not be overcome in pulse modulation systems using pulses of varying amplitude. This prompted him to review the early idea of transmitting speech using coded pulses of constant amplitude, similar to those used in telegraphy. His investigation resulted in the invention of a radically different approach to transmitting speech signals. In 1938, Reeves patented his invention which became known as *pulse code modulation.* Unfortunately, the development of practical pulse code modulation systems had to await the arrival of high-speed solid-state switching devices, which occurred after World War II.

Pulse code modulation involves transforming continuously variable speech signals into a series of digitally coded pulses and then reversing the process to recover the original analog signals. This procedure can be carried out in three successive operations.

The first operation is to *sample* the speech signals at a suitable rate and to measure the amplitude of the signal at the time of sampling. This operation is equivalent to pulse amplitude modulation (PAM). Next, the voltage amplitude of each sample, which may assume *any* value within the speech range, is assigned to the nearest value of a set of discrete voltages. This process is known as *quantizing* and is equivalent in mathematics to rounding off to the nearest whole number or integer. The final step is to *code* each discrete amplitude value into binary digital form, similar to coding the letters of the alphabet for telegraphy. Now a series of binary coded digital pulses can be used to carry the message over a transmission line. These binary pulses are in fixed and predetermined time positions and only the presence or absence of a pulse determines the information content of the signal. Since the precise magnitude of the pulses is no longer critical, the problems of cumu-

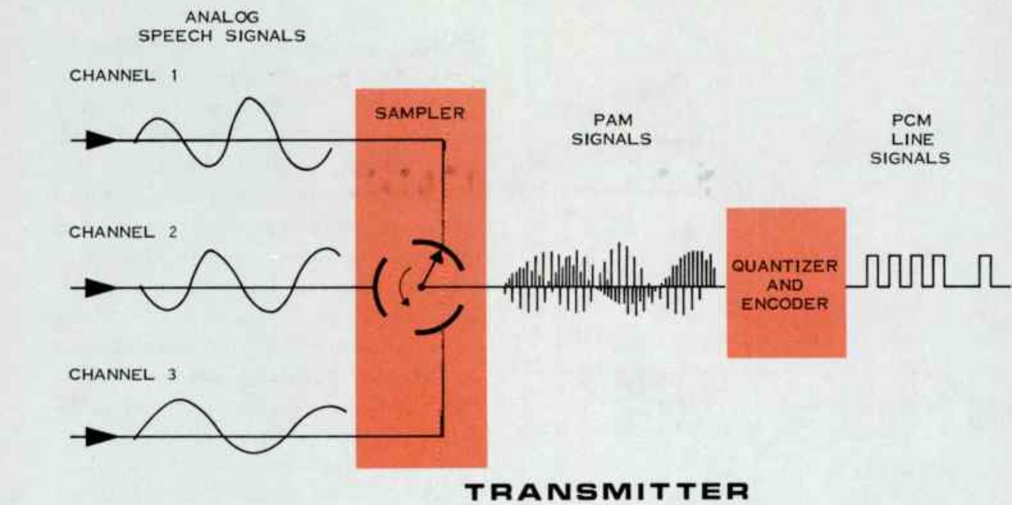


Figure 2. Simplified time division multiplex PCM system.

lative distortion and noise associated with pulses of varying amplitude are greatly reduced.

Sampling

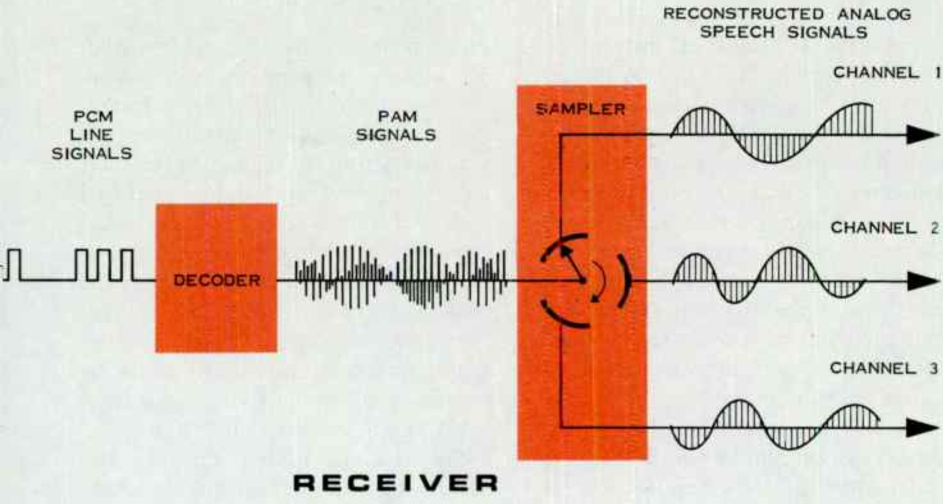
It has been proven mathematically that if a continuous electrical signal is sampled at regular intervals at a rate of at least twice the highest significant signal frequency, then the samples contain all of the information of the original signal. This principle is known as the *sampling theorem.* A continuous signal waveform, therefore, can be represented completely if at least two amplitude samples are transmitted for every cycle of the highest significant signal frequency.

In PCM systems designed for speech signals, a sampling rate of 8000 Hz, or one sample every 125 microseconds (1/8000 second), is ordinarily used. This sampling rate is sufficient since the bandpass of ordinary speech or telephone channels has an upper cutoff frequency below 4000 Hz. The 125 microsecond interval between samples of one voice channel can be allocated to other voice channels by means of time division multiplexing.

The number of channels that can be time division multiplexed using an 8000-Hz sampling rate depends, of course, on the duration of the time slot assigned to each sample — the shorter the duration, the greater the number of channels. In practice, the duration of the samples depends upon the operation and characteristics of a physical circuit. Thus, the number of time division channels is limited by the performance requirements and capabilities of a particular transmission system.

Quantizing

As previously stated, sampling a continuous speech signal at regular intervals results in a series of pulses whose voltage amplitudes are proportional to the level of the signal at the



time of sampling. The amplitudes might be any of an infinite number of values within the intensity range of speech. The usual intensity range encountered in telephone systems is about 60 dB, or a voltage ratio of 1000 to 1.

After sampling, the next step in the PCM process is to divide or quantize the 60-dB intensity range into increments or amplitude levels to permit binary digital coding. These discrete levels, known as quantum steps, are used to represent any level within the speech range. This is accomplished by using the quantum step nearest to the actual amplitude value of the pulse sample. For example, an actual amplitude sample with a value of say 8.24, would be represented by quantum step 8. A sample value of 8.61 would be represented by quantum step 9, and so on.

Since the quantum step only approximates the actual value, there is always some error. The maximum error is equal to one-half the size of the quantum step. In speech signals, such errors are random and cause what is usually referred to as *quantizing error or noise.* Quantizing noise is the major source of signal distortion in PCM systems. The degree of quantizing noise is mainly a function of the number of quantum steps used—the more quantum steps, the less the quantizing noise. However, increasing the number of quantum steps increases the bandwidth required to transmit the coded signals.

It is, of course, necessary that the quantizing process detect all of the positive and negative amplitude levels within the dynamic speech range. Experiments have shown that approximately 2048 *uniform-size* quantum steps are required to cover the speech range and to provide sufficient signal fidelity. An excessively large bandwidth is required to transmit the coded line signals representing such a large number of uniform quantum steps.

One way of reducing the number of quantum steps without sacrificing quality is to make the size of the quantum steps *non-uniform,* thereby taking advantage of the statistical distribution of speech amplitudes. Most of the information in speech signals is concentrated at low amplitude levels. If the quantum steps are all equal in size, then low level or weak signals suffer the greatest amount of quantizing error. Therefore, small quantum steps are needed more at the low amplitude levels than at the higher levels. If very small quantum steps are assigned where most of the speech information is concentrated, that is at low amplitude levels, and larger steps assigned to the rest of the amplitude range, then the total number of steps required can be greatly reduced. Varying the size of the quantum steps requires sophisticated coding techniques which are presently under development.

Another method is to *compress* the amplitude range of the pulse samples before uniform quantization and then to expand the range back to normal at the receiving end of the circuit. This technique, called instantaneous compression and expansion, or *companding*, achieves the same results as varying the size of the quantum steps. Instantaneous companding, which must be very fast-acting to respond to the short pulse samples, should not be confused with the slower-acting syllabic companding technique used in certain analog telephone circuits — although the principles are the same. The syllabic compandor responds to the envelope of analog speech signals directly while the instantaneous compandor responds to PAM samples of the analog signals.

Signal compression modifies the normal distribution of speech amplitudes by imparting more gain to weak signals than to strong signals. In typical applications, the technique reduces the amplitude ratio from 1000 to 1 to 63 to 1. Using a certain compression characteristic that reduces the speech range from about 60 dB to about 36 dB, and one that varies logarithmically with signal amplitude, the number of quantum steps can be reduced from

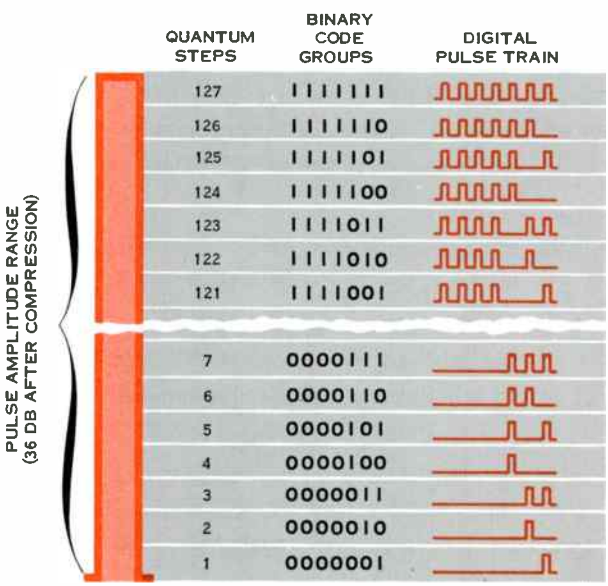


Figure 3. In PCM, amplitude samples of speech signals are compressed, quantized, coded into binary form, and placed on the line as digital pulses.

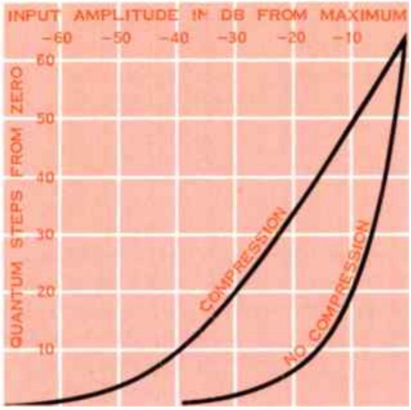


Figure 4. Signal compression using typical logarithmic compression characteristic.

2048 to 128 while maintaining the same quantizing noise performance. Signal compression using a typical logarithmic compression characteristic is shown in Figure 4.

Coding

The final step in the PCM process is to code the quantum steps into digital form. If each quantum step is numbered in decimal form, then some type of digital code can be developed to represent each of the numbers. Ordinarily, a binary code is used that consists of a combination or *code group* of binary 1's and 0's, each group representing a decimal number. Once the code is established, a series of on-off binary pulses, representing the code groups, can be used for transmission.

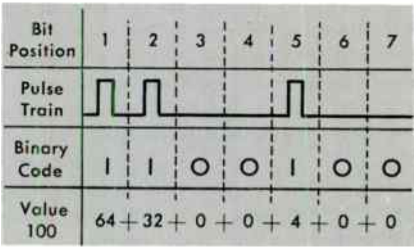
The number of quantum steps that can be represented with a binary code is 2^n, where n is the number of binary digits, or bits, required in each code group. Thus, a 5-bit code is required for 32 (or 2^5) quantum steps, while a 7-bit code is needed for 128 (2^7) steps. In systems using a 7-bit code, the speech amplitude range is normally divided into 127 quantum steps; step 64 is zero reference, with 63 steps positive and 63 steps negative.

The bandwidth required to transmit digital pulses is directly proportional to the number of bits in the code group. A code representing 2048 uniform quantum steps would have required 11 bits per code group ($2^{11} = 2048$). Compressing the amplitude range of the sample pulses before quantization, therefore, reduces the number of bits per code group required for quality speech transmission from 11 to 7.

With a 7-bit code, the first bit position has a value of $2^6 = 64$, the second has a value of $2^5 = 32$, and so on. The value of all seven positions is shown in the following table.

Bit Position	1	2	3	4	5	6	7
Value	2^6	2^5	2^4	2^3	2^2	2^1	2^0
	64	32	16	8	4	2	1

Typically, the coded line signal consists of a train of pulses in which binary 1's are represented by positive or negative pulses and binary 0's are represented by spaces (or no-pulses). A binary 1 in any of the bit positions means that the value of the position is to be summed. A binary 0 in any of the positions means that the value of the position is *not* to be summed. As an example, the pulse train and code group representing quantum decimal step number 100 would be:



Present Applications

One of the most outstanding features of PCM systems is that the coded line pulses can be regenerated at repeater stations. Since only the presence or absence of a pulse determines the message, the line signal can be completely renewed each time it passes through a repeater. This allows a high signal-to-noise ratio to be maintained through a long string of repeaters, thus overcoming most of the problems of cumulative noise which characterize analog transmission systems.

Unfortunately, the advantages of PCM are obtained at the expense of increased bandwidth. For example, the bandwidth of a voice channel in a PCM system using an 8000-hertz sample rate and a 7-bit code would be approximately 56 kHz compared to 4 kHz required for a single-sideband suppressed-carrier FDM system.

In typical long-haul high density transmission systems, especially microwave radio systems, the availability of bandwidth is usually very critical. Presently, PCM does not provide sufficient economical or technical improvements over analog techniques to justify its use in these long-haul systems. But the same is *not* true in short-haul cable transmission systems. There have been continuing efforts by the telephone industry to shorten the economical *prove-in* distance of multichannel carrier systems since they were introduced into the short-haul cable plant. The tremendous population growth around urban areas has greatly increased the need for low cost carrier systems in short-haul interoffice trunks. This need has stimulated

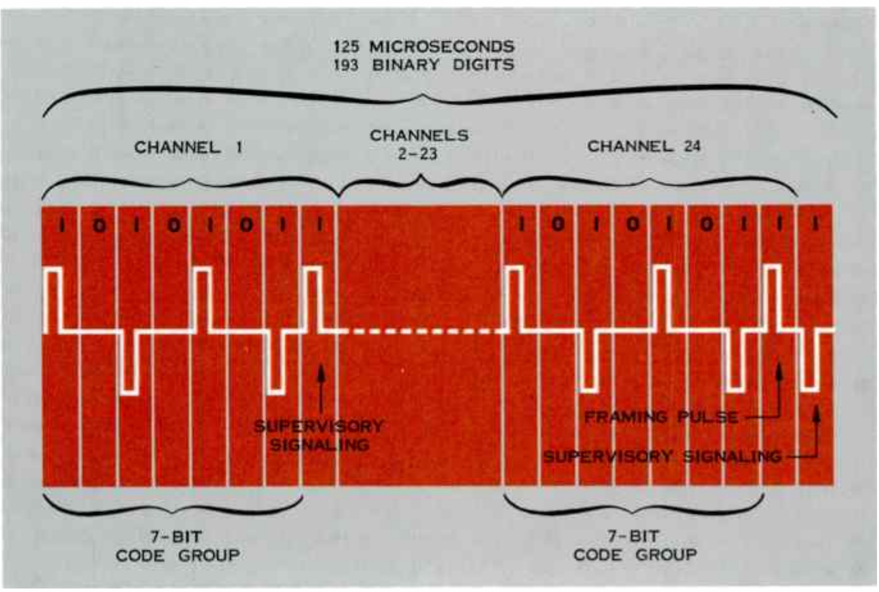


Figure 5. In the 24-channel T1 carrier system, 125 microsecond sampling interval or frame is divided into 193 time slots — 168 slots for speech, 24 slots for supervisory signaling, and 1 slot for synchronization.

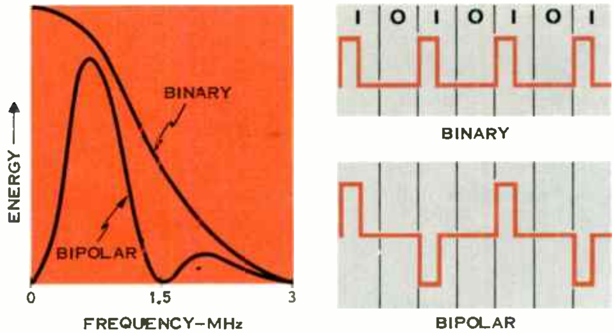


Figure 6. Energy distribution for binary and bipolar pulses with 50-percent duty cycle. In bipolar pulses most of the energy is concentrated at half the pulse repetition frequency and there is no dc component.

interest in PCM systems for use in telephone exchange cable trunks ranging in lengths from about 6 to 50 miles.

In 1962, the Bell System began production of a 24-channel PCM carrier system called the T1. Installation of T1 carrier systems in Bell's exchange plant marked the first large-scale use of time division multiplexing in commercial telephony.

The T1 carrier system was designed primarily for use with two non-loaded 22-gauge cable pairs in exchange area trunks. One cable pair is required for each direction of transmission. Regenerative repeaters, used with the T1 system, are spaced at intervals of about 6000 feet. This interval corresponds to the spacing of Western Electric's H-88 load coils on 22-gauge cable pairs. Since the load coils must be removed when the line is to be used for PCM operation, it is convenient to replace them with a regenerative repeater.

Each voice-frequency input channel in the T1 is sampled once every 125 microseconds or 8000 times per second. The variable amplitude pulses resulting from the sampling process are then compressed and quantized into one of 127 quantum steps coded into 7-digit binary code groups. An eighth digit or bit is added to the code group for each channel sample and is used to carry supervisory signaling information.

The time slots which make up one 125-microsecond period constitute what is called a *frame*. An additional bit time slot is added to each frame for use in synchronizing the two system terminals. This makes a total of 193 time slots per frame (24 channels x 8 bits per code group + 1 synchronizing slot). Multiplying the 193 time slots times the 8000 hertz sampling rate provides an output pulse train with a maximum bit rate of 1,544,000 bits per second.

The binary coded pulses transmitted to the cable pair have a fifty percent duty cycle, which means the width of the pulses is one-half the time slot allocated to each pulse. Bipolar transmission is used with successive pulses, representing binary 1's, alternating in polarity. Figure 5 illustrates a pulse train representing one frame.

There are several advantages of the bipolar pulse pattern over straight binary or unipolar transmission. As shown in Figure 6, most of the energy of bipolar signals is concentrated at frequencies of about half the pulse repetition frequency. Accordingly, there is much less energy coupled into other systems in the same transmission cable because of increased crosstalk coupling

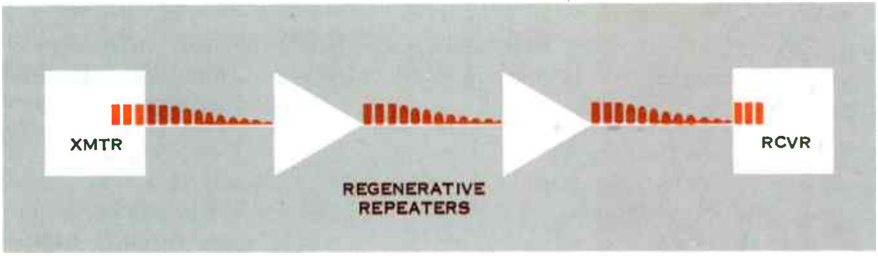


Figure 7. One of the most outstanding features of PCM systems is that the binary coded digital line pulses can be regenerated at repeaters and at the receiver station.

loss. Also, bipolar pulses do not have a dc component, thus permitting simple transformer coupling at repeaters. The unique alternating pulse pattern can also be used for error detection since errors tend to violate the pattern.

As the line signals travel along the cable pairs the pulses become distorted by the usual signal impairments such as noise and attenuation. When the pulses reach a repeater, they are re-timed and reshaped so that a new un-distorted pulse is produced for each pulse received. At the receive terminal the line pulses are again reconstructed before they are fed into the receiver detection and decoding equipment. The PCM coding process is reversed in the receiver in order to recover the original continuous speech signal. The continuous signal at the output of the PCM receiver should be a replica of the original signal, except for some distortion resulting from quantization.

However, noise troubles, like energy, seem to be conserved and only changed from one form to another. So it is with pulse code modulation. Although noise in PCM systems does not accumulate, it does prevent the perfect timing of re-generated pulses and shows up as jitter on the retransmitted pulse train. Successful practical solutions to this timing problem are the key to successful PCM cable carrier transmission.

Conclusions

New digital technology promises much more than just carrying out the tasks of transmission systems developed in the past. PCM systems will eventually handle all of the transmission functions of today's frequency division multiplex systems more efficiently and more economically. The development of digital techniques will enable different types of services such as voice and facsimile to be treated alike in transmission systems. Once the various types of analog signals are formed into digital signals they are all similar.

The new PCM cable carrier systems must carry on the tradition of the telephone industry. They must be used with the telephone plant that exists today, complete with its inheritance of old cables and switching systems produced by many different manufacturers.

In addition to operating over existing cable systems, there are already means to interconnect FDM systems with PCM networks using a device called an en*co*der-*dec*oder or codec. Network television in color will also be handled. Such high density systems will require line transmission rates close to 300 megabits per second. Also, the use of a technique called *pulse stuffing synchronization* will permit adding and dropping systems by digital means, and provide an easy method of interconnecting PCM systems.

High density PCM systems will use thousands of transistors in circuit configurations where switching times may be as fast as a fraction of a nanosecond. In the future, the use of integrated circuits instead of discrete components promises great economies as well as excellent reliability.

PCM systems are not without problems. Long chains of repeaters in tandem challenge the ingenuity of engineers to produce reliable repeaters at economical prices. The T1 system, for example, may use 50 repeaters in tandem for a 50 mile link. Further problems arise because PCM requires so much bandwidth. Bandwidth is readily available on cable but is not easily obtained from the available microwave spectrum. This fact ensures that single sideband frequency division multiplex will be around for a long time. PCM systems are also vulnerable to impulse noise which may prohibit their use in situations where cable plant and switching machines are not up to modern standards. In such situations present day FDM cable carrier systems, which do not exhibit the noise threshold characteristics of PCM, may do a better job. Also, cables already carrying FDM carrier systems will have to be filled out with the same type of systems since it is presently not possible to mix T1 and FDM systems in the same cable.

Although the T1 carrier system was developed primarily for the transmission of analog information in the form of processed voice signals, its repeatered line is a very fine high-speed digital transmission facility. Techniques have been developed to use these digital transmission systems to handle up to eight 50-kilobit data channels or two 250-kilobit data channels.

Pulse code modulation systems provide better handling of telephone supervisory signaling than the usual inband methods used with FDM systems. The systems employ time division signaling methods which are very economical and avoid the problems of speech simulation or *talkdown* inherent in inband signaling systems.

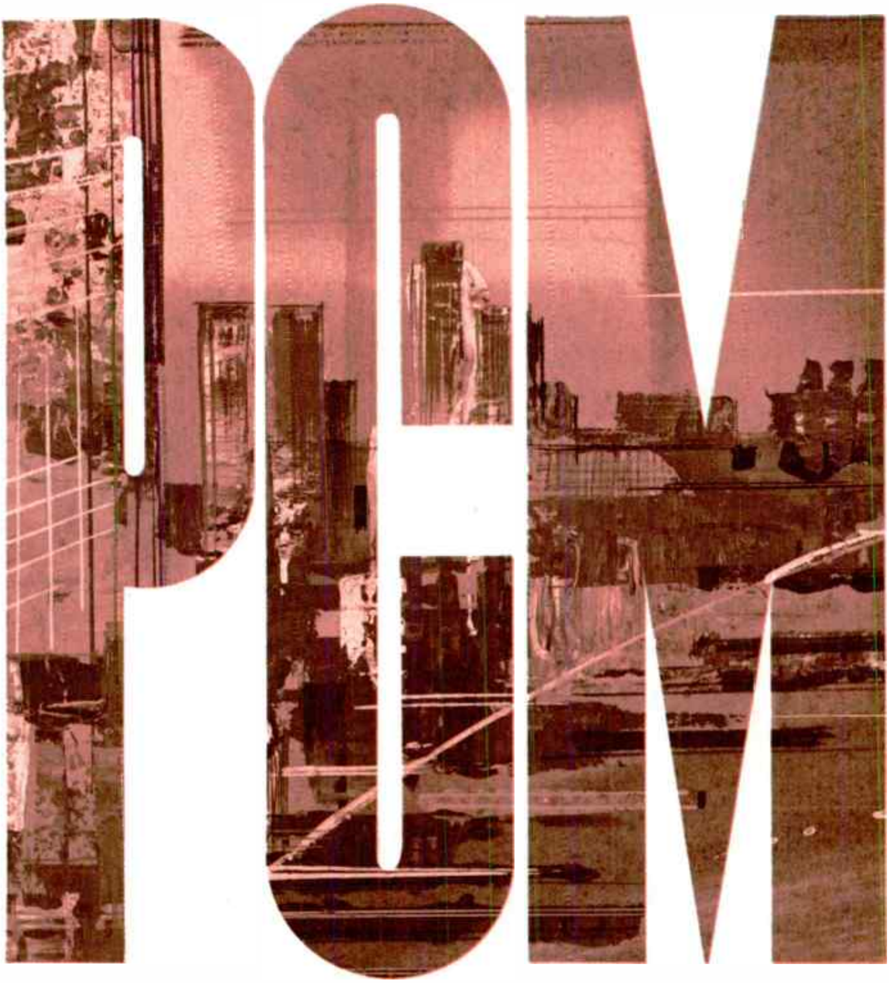
Some small switching machines employ time separation instead of the familiar space separation techniques used for so many years by electromechanical machines. Digital transmission is used with these time separation switching machines, and it is only a short technical step to join digital transmission and switching into an integrated communications system. It seems likely that in the 1970's integrated electronic switching and PCM transmission systems will be operating both in the United States and Europe.

BIBLIOGRAPHY

1. *Bell System Practices*, Plant Series, Section 365-010-100, Issue 1, AT&T Co., 1962.
2. *The Bell System Technical Journal*, Vol. 41, No. 1, AT&T Co., New York, Jan. 1962.
3. Deloraine, E. M. and Reeves, A. H., "The 25th Anniversary of Pulse-Code Modulation," *IEEE Spectrum*, The Institute of Electrical and Electronics Engineers, Inc., New York, May 1965.
4. Fultz, K. E. and Penick, D. B., "The T1 Carrier System," *The Bell System Technical Journal*, Vol. 44, No. 7, AT&T Co., New York, Sep. 1965.
5. Smith, B., "Instantaneous Companding of Quantized Signals," *The Bell System Technical Journal*, Vol. 36, No. 3, AT&T Co., New York, May 1957.

The Lenkurt®
MARCH 1968

DEMODULATOR



Part 1

PCM PCM PCM

. . . . a unique family of digital transmission systems for voice, video, or data.

The first serious interest in time division multiplexing came about 1930. Early experiments centered around pulse amplitude modulation, but noise and crosstalk were serious problems.

The invention of pulse code modulation (PCM) occurred in 1937. However, it was so basically different from the contemporary concept of telecommunications that its importance was not widely understood or appreciated. And even if it had been, the components necessary to accomplish the complex processes of sampling, quantizing, and coding were not available. PCM had to wait for transistors, integrated circuits, and a better knowledge of digital devices.

Frequency division multiplex (FDM) has served the telephone industry well for many years, but economic factors have kept engineers on the search for more practical and efficient transmission methods. As interest was renewed in digital communication, the advantages of PCM over FDM became apparent.

Solid State Economy

The economics of solid state technology point to PCM as the transmission method of the future. To the telephone company, this is another means of holding the line against constantly increasing costs; to the telephone user, it is another step toward more efficient and complete communications service.

With transistors as a working tool, and integrated circuits a promise for the future, Bell Telephone Laboratories readied their first PCM systems in the 1950's. The first working system in the United States was installed in 1962. Today there are about a half-million channels in service.

Bell will necessarily establish the system parameters for PCM performance, and as in commonly the case, independent manufacturers and users will find it convenient to refer to Bell nomenclature in describing their own systems.

(To avoid confusion, Bell terminology will be used here rather than Lenkurt equipment types.)

Time multiplexing PCM systems have been designated as T carrier in the Bell System. The first generation, now in service, is the T1. Future systems will expand the T family from T1 to T4. Along with these progressions will come refinements in channel banks and increased available service.

The T1 system can consist of either a D1 channel bank of 24 channels, a D2 channel bank of 96 channels, or several choices of data banks.

The D1 channel bank now in service provides 24 two-way channels on two exchange grade cable pairs, one for each direction of transmission. These may be in the same cable, or in separate cables.

Repeatered Line

Regenerative repeaters are spaced along the cable about every 6000 feet. Because information on a PCM system is transmitted in the form of binary pulses, the repeater need only recognize the presence or absence of a pulse to regenerate a clean, new pulse. Because of the lower signal-to-noise ratio required with regenerative sys-

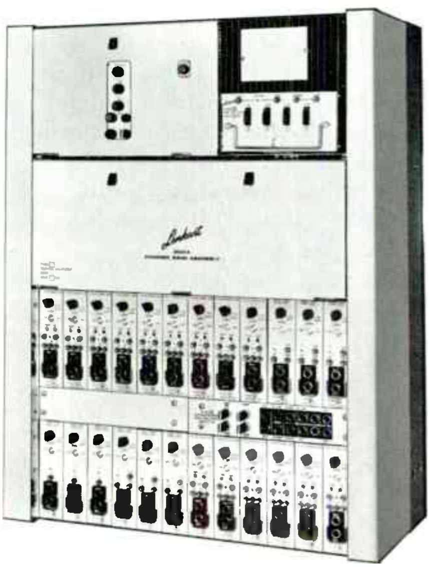


Figure 1. Lenkurt's 24 channel PCM system, the 91A, exhibits space saving design offered by TDM technology.

tems relative to FDM systems, large amounts of noise, interference, and distortion can be tolerated.

While noise and distortion do not accumulate along a PCM cable as they do with FDM repeaters, it's not a case of getting something for nothing. The noise appears in a different form, showing up as a jitter on the retransmitted pulse train. If allowed to accumulate, this jitter can prevent perfect retiming of pulses. Hence, special jitter-reducing circuits are needed for long haul PCM transmission systems.

Jitter is not a serious problem in the T1, which is intended for use in systems less than 100 miles in length. But the long haul, high density systems which will someday span the nation must be equipped to handle it.

Time Separation

The translation of an analog signal into PCM begins by switching sequentially from one channel to another at a rapid rate. Each channel occupies the transmission line for a fraction of the total time. Conversations are stacked in time rather than in frequency as in FDM systems, and the method is called time division multiplexing. By synchronizing the sampling rate at the receive end, each channel may be recreated in its original form.

If periodic samples of a waveform are taken often enough, the waveform can be perfectly reconstructed at the end terminal. The necessary sampling rate is just twice that of the highest frequency to be transmitted. Therefore, if 4000 Hz is the highest frequency on a telephone channel, samples taken at the rate of 8000 per second will precisely and exactly duplicate the telephone conversation. This is the sampling rate used in the T1, although in practice the channel bandwidth is 200-3400 Hz.

PCM Coding

Three successive operations are needed to transform the analog speech signal into the series of digitally coded pulses of PCM.

The first operation is to sample the speech signal at a suitable rate (8000

Hz) and measure the amplitude of the signal. This results in a train of pulses roughly analogous to the original waveform.

Next, the amplitude of each sample is compared to a scale of discrete values and assigned the closest value. This rounding off process is called quantizing. Each pulse, now with its discrete value, is then coded into binary form. These binary pulses are what appear on the transmission line.

A 7-digit binary code is used in telephone PCM. That is, each sampled pulse is coded into a combination of seven pulses representing one of 127 different discrete values. To each code group an eighth digit or bit is added for signaling.

Each of the 24 channels in a T1 system is sampled within a 125-microsecond period (1/8000 second), called a frame. To each frame an additional bit is added for synchronization of end terminals.

Eight bits per channel, times 24 channels, plus the synchronizing pulse, brings the total time slots needed per frame to 193. The resulting line bit rate, with 8000-Hz sampling, is 1,544,000 bits per second.

(A more complete analysis of the PCM coding process is found in the November 1966 *Demodulator.*)

The bandwidth required for the T1 system is about 1.5 MHz, or one cycle per bit. This is obviously much more than is needed to transmit 24 channels over an FDM system. On cable, the bandwidth is readily available, but the advisability of PCM on most microwave applications is clearly limited.

The most significant reason more bandwidth is not available to FDM systems on cable is the difficulty of designing wideband amplifiers with sufficiently flat response. In the digital PCM system, where each channel uses the entire system bandwidth and is separated from other channels by time,

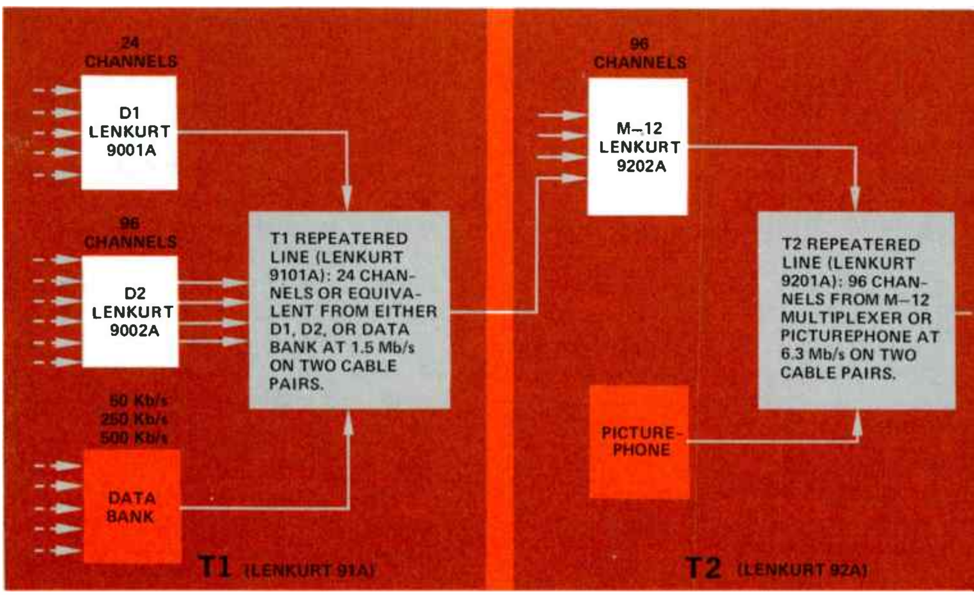


Figure 2. PCM systems, expanding from T1 to T4, will mix voice and oth

non-uniform attenuation of the transmission path is not a particular problem. The end terminal, and each repeater along the cable, need only register the presence or absence of pulses to successfully accomplish communication.

T Carrier Family

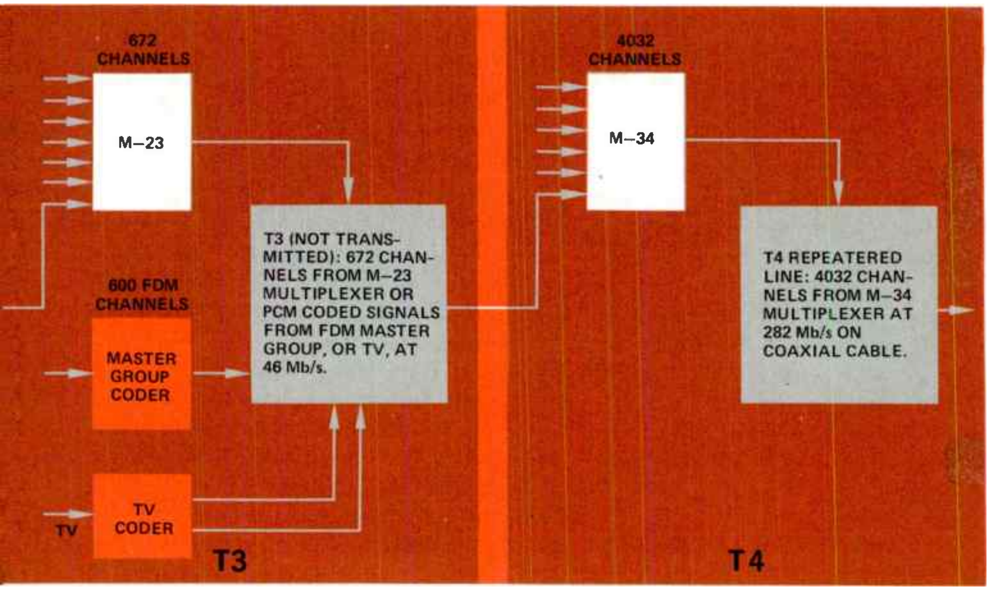
The overall concept of the T carrier family is illustrated in Figure 2. Beginning with the short haul exchange carrier T1 system now in service, the family expands into a transcontinental system capable of carrying voice, data, and television signals simultaneously.

As the network grows, four T1 signals will be combined into a 6.3 Mb/s stream called the T2. Seven T2 signals will be combined into a 46 Mb/s group, sometimes referred to as the T3 section although it is not intended for transmission. The final step in the planned T carrier system will be when six 46 Mb/s signals are multiplexed, forming the T4 signal at 281 Mb/s.

D2 Channel Bank

The first outgrowth of the T1 system will be contained in the channel bank equipment. By combining the common equipment of four 24-channel groups (D1 channel banks) into a 96-channel system, considerable savings in cost and size are promised, along with increased noise performance.

The result is the D2 channel bank. The transmission format is exactly the same as the D1, with each 24-channel group operating over a separate T1 repeatered line. However, some noise improvement is gained by making the eighth digit in each group do double duty. In the D1, this bit is committed to signaling information. But in the D2 the eighth digit will carry VF information most of the time – only occasionally will it be "borrowed" for signaling.



rvices using time division multiplex. Many T1 systems are now operating.

Current technology will not allow D1 and D2 channel banks to work end to end, but an interface between the two is theoretically possible.

The T1 system, although primarily developed for the transmission of voice, is ideally suited to the increasing data market. Various data banks now being developed will provide service at 50 kb/s, 250 kb/s, or 500 kb/s over the T1 repeatered line. When it is realized that 64 kb/s is the rate used for each voice channel, the hundreds-of-kilobit rates are not surprising.

T2 System

The intermediate speed T2 system will provide 96 channels at roughly four times the T1 speed, or 6.3 Mb/s. A multiplexer, called the M-12 by Bell, will bring together T1 pulse streams from either four D1 channel banks or a single D2 channel bank, retime them, and then transmit them to the T2 repeatered line. Two cable pairs are used, and present planning calls for a system extending out to 500 miles using a more complex repeater.

Bell's Picturephone will also operate at T2 bit rates and could be an alternate use of the system.

The next step in the T carrier hierarchy was originally named the T3 system. But this is no longer considered a necessary step in the logical growth of transmission systems and is now planned to be only a convenient way of gathering groups of signals together within one office.

At this point the M-23 multiplexer will combine seven T2 signals into a 46 Mb/s stream capable of carrying the equivalent of 672 voice channels.

T4 System

Finally, six M-23 signals will be accepted by the M-34 multiplexer to produce a total of 4032 channels in the T4 system at 281 Mb/s.

By its nature, a PCM system can accept almost any mixture of voice, data, or in the T4, television. Once the intelligence has been digitally coded, the pulses are exactly the same. The repeatered line equipment or the multiplexers themselves are no longer concerned with the type of information being transmitted. Likewise, there is virtually no interaction among the various signals.

Experimental systems operating at Bell Laboratories indicate that television signals with 10 Mb/s sampling and a 9-digit code group will produce acceptable picture quality. The total output of the TV encoder at 92 Mb/s will be divided into two 46 Mb/s streams for entrance to the M-34 multiplexer

Another encoder planned for the T4 system will convert an entire FDM master group of 600 voice channels to PCM. The master group will be sampled at 6 Mb/s, and will be transmitted with 9-bit quality. The output of this encoder will also be 46 Mb/s to match the standard M-34 input.

The planned T4 system is unique not only when compared to FDM systems, but to other members of its own PCM family. Because of its design function as a long haul carrier, the T4 must satisfy many special conditions. It should be capable of operating over any distance carrying several thousand telephone calls, several television signals, numerous wideband data signals, or mixtures of all.

To this end, coaxial tubes will be used to attain bandwidths sufficient for the 281 Mb/s signals. And at the toll office, T4 equipment will be called on to provide a number of more critical duties than were earlier systems.

A new code, for instance, has been devised to eliminate a number of limitations inherent in such a long system. A three-level code (minus, zero, or plus pulses) will replace the two-level bipolar code (pulse, no pulse) used in the T1. The new code, called paired selected ternary (PST), primarily conserves bandwidth by adding additional information capability with the same signal-to-noise ratio. PST provides a strong timing signal for repeaters by eliminating long series of zeros, and as in other bipolar PCM transmission, avoids a dc component in the signal and permits errors to be monitored on the repeatered line.

Pulse Stuffing

If the coded pulses of a single channel are to be extracted from over 4000 channels in a 281 Mb/s stream, some method of synchronizing T4 terminals must be available.

In the relatively short T1 system where all pulses originate at the same point and terminate together, the pulse stream itself is satisfactory to lock terminal timing clocks together. But with more complex systems having terminals thousands of miles apart, this is not practical.

Ideally, the various channel banks and coders feeding into the T4 network should be allowed to operate on their own clocks, which will vary slightly from the normal frequency.

One approach being studied at Bell Labs which will allow this independent action is called *pulse stuffing.* At certain intervals an extra time slot is "stuffed" into the digital stream. Because the presence of stuffed pulses is indicated by other coding on the line, these extra pulses are ignored at the receive end. But during transmission they have allowed the incoming signals to get onto the line even though their individual synchronization does not exactly agree. In a way, the stuffed pulses are a buffer between frames of different channel banks. The end terminal is capable of separating these frames and re-creating the synchronization necessary for that particular group of channels.

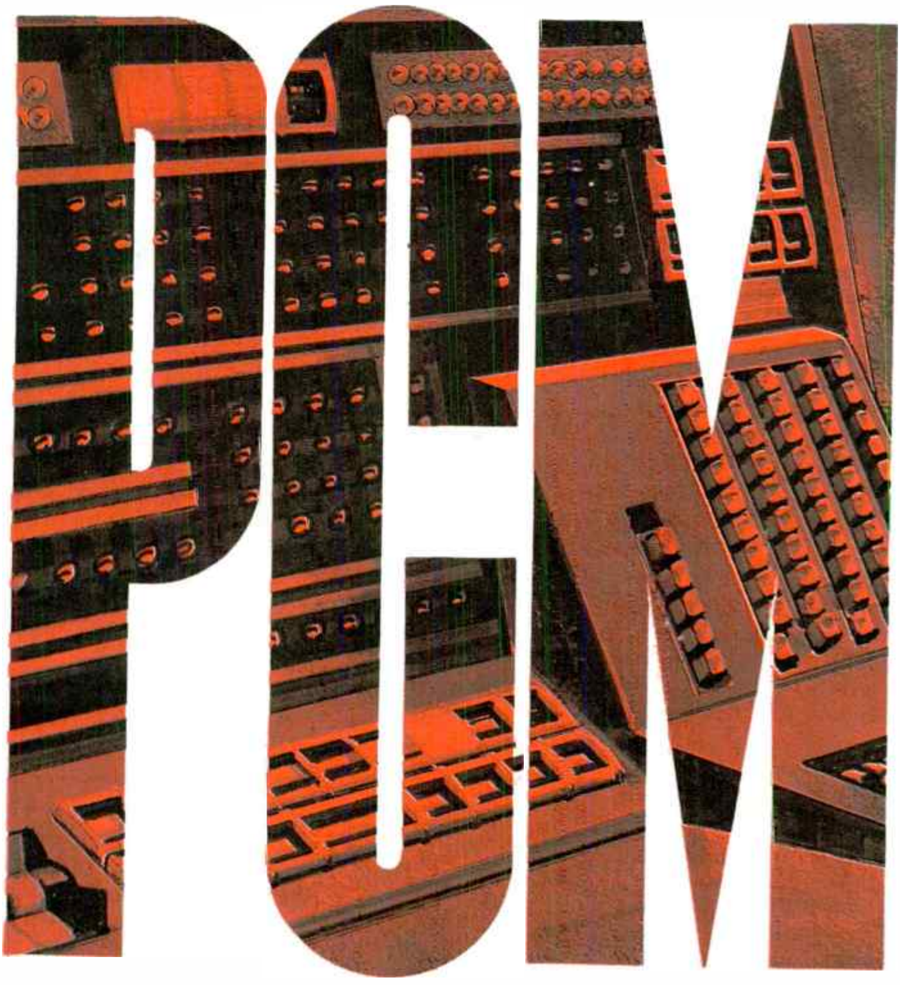
The basic elements for the implementation of systems beyond the T1 exist – some in operating systems, some in the laboratory. Introduction of the D2 channel bank is imminent, and the T2 transmission system probably is only several years away.

An experimental high-speed system, operating at 224 Mb/s, has been in operation at Bell Labs for over a year. But the T4 system as outlined must still be considered just a concept.

(The April issue of the *Demodulator* will discuss several additional facets of pulse code modulation, including data capabilities, the repeatered line, fault location at repeater sites, and PCM switching.)

The Lenkurt DEMODULATOR

APRIL 1968



Part 2

PCM PCM PCM

. . . . New concepts in data transmission, repeatered lines, and switching.

The entire concept of transmitting voice, television, or other analog information using a string of digital pulses – PCM – is unique. Beyond the initial treatment of the signal, so different in itself from frequency division multiplexing, there are many other significant and exciting factors which increase the telephone industry's responsiveness to this new technology.

The data carrying capabilities of PCM, for example, are startlingly high – for a reason. The span concepts of signal routing and repeatered lines themselves are dependent on digital transmission. And the possibility of direct PCM switching, without returning the signal to audio frequencies, is perhaps one of the strongest potentials of all.

Data Transmission

While PCM systems are designed especially for voice transmission, their digital format makes them particularly good carriers of data.

Relatively slow speed data, as would normally occupy no more than one voice channel, will be handled through an existing data modem such as the Lenkurt 26C. While this type of data set is designed to condition data signals for FDM transmission, and therefore has a tonal output, it can work into the T1 system. The data modem output is sampled by the PCM terminal the same as with a voice signal.

For higher speeds, special data modems must work directly into the repeatered line and in some cases be able to coordinate with D1 channel banks. Several data banks will be available to meet specific needs and will allow data speeds on the T1 line from 50 to 500 kb/s.

The economy of using the PCM line for data is clear. A 50 kb/s data signal displaces 12 voice channels on an FDM system, but takes only three channels on the T1 repeatered line. And even this efficiency is allowing for data that is not in synchronization with the line rates. If the incoming data rates precisely match the bit rates on the line – 1.5 Mb/s for the T1 – these synchronous signals can be transmitted on a bit-by-bit basis.

However, when asynchronous data must be handled, additional treatment of the information is necessary. Three T1 carrier bits are required for each data bit. The first transmitted bit indicates a data transition has occurred. The second bit carries information on the length of the data bit. The final bit relays the direction of the transition – that is, plus or minus.

Therefore, every time a data bit is received, three successive T1 bits are needed to transmit the information.

The mixture of voice and data signals on a T1 line is not without its own unique problems. The usual 8-digit "word" used to transmit voice is not a natural format for data. For example, if one-half the line bit rate is to be used for data, it seems logical that every other bit transmitted should be allocated to data. In the 8-digit grouping used for voice, this is impossible. But when data is the only service provided on the line, the most efficient method is to use evenly distributed data bits.

T1 LINE LOADING	NO. OF DATA CHANNELS	MAX. DATA RATE
1/8	8	64 kb/s
1/4	4	128 kb/s
1/2	2	256 kb/s
1/1	1	512 kb/s

Figure 1. Using alternate bits on T1 results in proportionate data loading.

The Western Electric T1WB–1 and T1WB–2 wideband data banks derive data channels in this way, using the appropriate number of alternating bits. Only simple timing changes are necessary to accomplish a variety of data speeds. If the data bank clock is arranged to put a data signal into every eighth bit on a T1 carrier, this channel will have a maximum capability of 64 kb/s. However, this will be standardized as a 50 kb/s channel to match other transmission requirements. Other possible data speeds arrived at in this way are listed in Figure 1.

The T1WB–1 has several possible arrangements depending on the data traffic needs between specific terminals. The T1WB–2 is basically a condensed version, more economical where only 250 kb/s data is transmitted. Other combinations are listed in Figure 2.

Voice With Data

In many cases it will be desirable to mix voice and data on the same line. Since their formats do not coordinate, some conversion will be necessary.

It is possible, for example, to retime the 8-bit word of the standard PCM voice signal to fit into alternate data bits. Or, the reverse can be accomplished by transforming the data stream into 8-bit words. This has been done in a third data bank, the T1WB–3. It will multiplex up to four data channels, using the remaining space for voice transmission. Possible combinations of data and voice are listed in Figure 3.

The T1WB–3 operates in synchronization with the D1 channel bank. Framing bits and the line bit stream for voice channels originate in the D1 and are received and retransmitted into the T1 line by the T1WB–3. In the data bank, a timing network clears pulses from selected voice channels and leaves them clear for data signals.

An additional task must be performed by the data bank in translating the data to the 8-bit format. This is accomplished by storing the data bits and rearranging them into a time sequence proper for the preempted voice channels. The data bits are combined with the retransmitted voice bits and the total digital stream is sent over the T1 line.

Although the data bits occupy the space of a voice channel in the T1WB–3, they are not directly affected by the 8-bit format: the organization of information in the data word bears no relationship to the channel word.

Complementing the three wideband data banks is a single channel modem for speeds up to 500 kb/s. The T1WM–1 will provide more economical service when there is reasonable assurance that additional wideband channels will not be required. The modem could be operated at the customer's location over a dedicated repeatered line.

WIDEBAND BANK	CHANNEL ARRANGEMENTS
T1WB–1	8 CHANNEL – 50 kb/s
	4 CHANNEL – 50 kb/s 1 CHANNEL – 250 kb/s
	2 CHANNEL – 250 kb/s
T1WB–2	2 CHANNEL – 250 kb/s

Figure 2. Wideband data banks offer a variety of channel options.

Span Concept

The very nature of a digital transmission system allows a new way of looking at the line and the signal it carries between offices. In PCM terminology, a series of regenerative repeaters from one office bay to another is called a "span line".

Using the span concept, it is possible to provide spare lines and fault location on an individual span line basis. And administration for assignment, maintenance, and powering becomes easier. Any span line is just like the next and direct substitution is possible.

In an analog FDM system, losses in the line, at terminals, and in repeater equipment are all cumulative. Therefore, the routing of calls and the design of alternate routes are restricted to some maximum attenuation. The span concept offers some relief from this historic problem. Taking advantage of the new technology, it becomes much less important for trunk routing to be along the shortest distance between terminals.

In practice, span lines between two central offices may be in different cable sheaths and may even follow different geographical routes. But they are indistinguishable in use and in design. The span line is a known quantity to be used as needed, but which remains an independent entity in the system of which it is a part.

With the PCM system that maintains its four-wire nature throughout, a standard loss of about 3 dB can be obtained irrespective of the number of span lines used to complete the route. In the T1 system, span lines may be connected in tandem to a limit of something less than 100 miles.

Fault Location

When trouble occurs in the repeatered line, another unique feature of the PCM system – fault location – makes it possible to identify the exact trouble spot from the central office. First the fault span is taken out of service by patching. A spare span is easily substituted. Then a fault locating test set is used to find the defective repeater.

WIDEBAND BANK	CHANNEL ARRANGEMENTS
T1WB–3	1 CHANNEL – 50 kb/s 21 CHANNELS VOICE
	2 CHANNELS – 50 kb/s 18 CHANNELS VOICE
	3 OR 4 CHANNELS – 50 kb/s 12 CHANNELS VOICE

Figure 3. Versatile T1WB–3 mixes voice and data channels.

The fault locating scheme of the T1 system uses twelve different audio frequencies and an equal number of matching single-frequency filters – one for each repeater site. The audio frequencies are generated in the test set and introduced to the repeatered line as a set of digital pulses. These pulses, which appear as errors to the repeater, actually have within them an audio component (Fig. 4).

The frequency selective filter bridged across the output of each repeater will pick off a specific tone intended to test that repeater. The audio component is looped back to a "fault locating pair" which returns the signal to the testing office. There it can be analyzed and the proper performance of the repeater determined.

In this way a technician working at the exchange can test each repeater in the span without leaving the office. The frequency used determines which repeater housing is being tested; the cable pair selected will determine which repeater within the housing is under scrutiny. The 12 frequencies available make it possible to test 12 different repeater locations, or a total of 300 repeaters if the housings each contain 25 repeaters. If more than this number must be tested from a single office, an additional fault locating pair may be used and the interrogation capacity doubled.

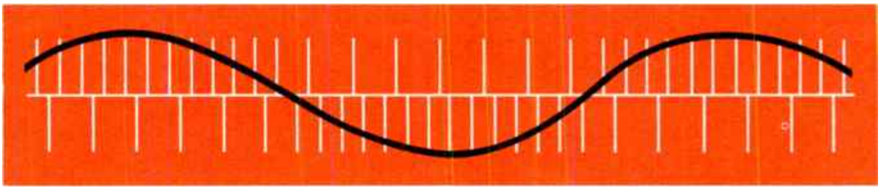


Figure 4. The audio component in the fault locating signal is caused by grouping "error" bits to create one of 12 frequencies.

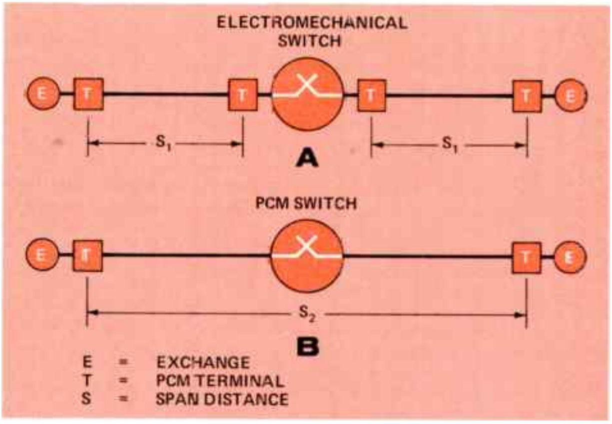
PCM Switching

In its normal operation the regenerative repeater looks at an incoming signal train and literally recreates new pulses in the same sequence as they were originally transmitted. If, instead, the repeater could store the pulses momentarily, and then regenerate them in a different order, a form of switching could be accomplished. For instance, pulses originally representing channel 4 might be regenerated in the time slot allocated to channel 7.

Taken one step further, the device might connect channel 4 pulses from one system to channel 7 slots in another system. This could be used as a trunk switch operating in the PCM mode, without changing information to audio frequency just for the purpose of switching. Such a method of switching could have a decided economic effect on the PCM systems of the future as well as offering some interesting possibilities in plant planning and design.

There are presently three basic types of exchange switching: manual, electromechanical, and electronic. The electromechanical switch operates in response to the dial pulsing of the

Figure 5. The distance between terminals, used in figuring prove-in economics, could be lengthened considerably with PCM switching.



subscriber's telephone and is the most common method. Operating at 10 digits per second, the system has been adequate for most applications.

But Direct Distance Dialing within the United States – and soon global dialing – places tighter restrictions on the signaling time available. Tone dialing will help, and the installation of advanced crossbar switching machines offers some relief. But a variation of electronic switching now in operation at several locations lends itself directly to application with PCM switching.

In PCM transmission, the signaling information is contained within the pulse train and therefore arrives at the exchange at the rapid microsecond rate of the PCM signal itself. If electronic switching devices were available to match this speed, all functions could be carried out in phenomenal time. It has been estimated that the entire process of switching might be accomplished in about 100 milliseconds.

Direct PCM switching could offer much more than speed. If signals did not have to be demodulated for switching, considerable economic gain as well as improved signal quality could be achieved. Terminal equipment contributes a large percentage of total system cost. And most degradation in the system takes place at the terminals where signals must be transformed from one form to another. The use of PCM switching could reduce the number of terminal units required.

Prove-in Distance

System planning dictates that the transmission engineer must consider prove-in distances when adding new equipment, especially carrier systems. The cost of carrier equipment versus physical cable, for example, must be compared for a given system. Generally it is less expensive to use physical cable for short distances and carrier equipment for longer routes.

Consider an exchange system (Fig. 5A) linked by PCM and using electromechanical switching at the hub. PCM terminals are required at both ends of the span between the switch and the exchange. Therefore, this span distance must be equal to or more than the prove-in distance if the system is to be economical.

By introducing a PCM switch at the tandem exchange, two terminals are eliminated and the longer span distance between the two far terminals becomes the criteria for prove-in judgment (Fig. 5B).

Two benefits to the operating company become obvious – half as many terminals must be supplied, and many more short-haul exchange routes become eligible for PCM.

In most telephone companies approximately 70 percent of all trunk lines are 10 miles long or less. If the prove-in distance is greater than 10 miles for a typical system, then 70 percent of the facilitites cannot be profitably used with PCM. But PCM switching could effectively double the trunk lengths and produce a sizable increase in the number of trunks available for PCM transmission.

Switch and Delay

The elementary switching function is to detect a new call, absorb signaling information, and set up a path through the exchange to the outgoing system. In PCM switching, routing would involve not only finding a clear circuit leaving the exchange in the proper direction, but would also necessitate matching in time of the two channels.

After the information has been switched out of an incoming pulse train, a finite amount of delay would be necessary to fit the signal into the proper time slot of the outgoing circuit. Preference would be given to switching situations where the incoming call could go out on the same channel, thus requiring no delay. Fixed delay lines would provide the proper delay for moving pulses into the time slots of any of the other 23 channels.

Fitting nicely into the span concept, PCM switching could increase the company's ability to give the customer substantially uniform quality communication regardless of the distance and the number of switching stages involved.

PCM Compatibility

The nationwide telephone network – giant machine that it is – must grow and change while it continues to operate. Any new technology, such as PCM, must work with existing facilities. This means that PCM transmission must be carried out over present cable and work through the exchange as it is today. The first PCM systems of the T1 class have been introduced in just that way.

If PCM switching is adopted, it will be added to the existing system. It is possible – perhaps even economically advisable – that the PCM switching will be called on to work with FDM systems. Local converters could change FDM audio signals to PCM and back again for the purpose of switching. These converter units would be relatively simple devices, free of complicated synchronization problems in that both ends of the conversion would be made in the same office.

Conceivably, PCM techniques could be extended to the subscriber level. This would not only open the way to increased wideband service available at the home and office, but have a strong effect on PCM switching, making possible a complete integrated telephone network.

Perhaps only with the eventual installation of the transcontinental T4 coaxial system will an entirely new PCM facility be established. But it too will have to coexist with other systems, accepting PCM-coded FDM supergroups, and pulse streams from other PCM systems, the T1 and T2.

SECTION II

MICROWAVE RADIO

The Lenkurt DEMODULATOR

AUGUST 1970

MTBF
MTTR

innage
outage

reliability
availability

99.9999%

some aspects of microwave system reliability

The following is the essential text of a talk given by Mr. R. F. White, a senior staff engineer at Lenkurt Electric Co., Inc., at the International Conference on Communications held in San Francisco in June, 1970. Because of the outstanding clarity of the discussion, and because the article is of general interest to communication systems users, it is being printed here for the benefit of all *Demodulator* readers.

In recent years the objectives for total reliability in microwave communications systems have become rather staggering. One example is the Bell System's stated objective of 99.98% *overall reliability* on a 4,000 mile system, which breaks down to an allowable *per hop outage* of about 25 *seconds* per year. Users of high reliability industrial systems are also talking about average per-hop reliabilities in the order of 99.9999%, or about 30 seconds per year, for their long-haul microwave systems.

This discussion is mainly concerned with the ways in which such microwave system reliabilities are being described, specified, and calculated, and with some apparent problems in some of the methods commonly used.

The microwave industry has long been accustomed to making estimates and calculations of outages due to propagation, using empirical or semi-empirical methods.

The results are usually stated either as a per hop annual outage, or as per hop reliability in percent. And it is interesting to note that calculations using these empirical methods indicate that by the use of suitable path engineering and diversity, it is possible to achieve propagation reliabilities in the above mentioned range.

Calculation methods for estimating the probable reliability of a microwave hop with respect to *equipment* outages have also come into the picture in recent years, using the principles and practices developed by reliability engineering experts in other fields.

It has also become fairly common practice to express the calculated equipment reliability results for microwave systems in terms similar to those used to describe propagation reliability in percent as the term is commonly used by microwave engineers, or as a per hop "availability." (The latter term as used in reliability engineering is the ratio, over the period of interest, of the innage time to the total time.)

A natural extension of this practice is to add the per hop annual outages for equipment and for propagation together to get an overall outage to be used as a reliability "figure of merit" for the hop.

This discussion will attempt to show that *none* of these parameters —per hop annual outage, per hop reliability in percent, or per hop availability— provides an adequate description of the *equipment* reliability performance in the case of ultra-reliable systems. It follows, of course, that if this is so, the "overall total reliability" concept and figures of merit are equally unsatisfactory.

Microwave equipment availability or outage calculations always rest in the end on two basic concepts: the "mean time between failures" (MTBF) and the "mean time to restore" (MTTR). The relationship between the two determines the outage ratio. (MTTR will be assumed to include notification time, travel time, diagnosis time, as well as the actual time to repair or replace the failed item. Thus, in this paper, it represents the actual average length of outage associated with a failure event.)

In high-reliability systems the relationships become quite simple, as shown in Figure 1.

The "innage ratio" is the term called "availability" by reliability engineers. Multiplied by 100 to convert it to percent, it is the "reliability" as used by microwave engineers.

Figure 2 shows how these parameters might look in a more or less typical *non-redundant* microwave hop.

The 5,000 hour figure in the denominator is an assumed value for the MTBF of all the equipment of a non-redundant microwave hop; it would correspond to an average of roughly two failures per year, and since the hop is non-redundant, each would be an actual outage.

A value of 5 hours is assumed for the MTTR, and as stated above, this is taken to mean all the time from the instant of failure until the equipment is restored and back in service.

These assumed values were chosen primarily for mathematical ease and are not intended to represent any actual system. This applies to any other values used.

What happens with a *fully redundant* configuration? Here, for simplicity, it is assumed that the non-redundant equipment is simply duplicated and that a perfect automatic sensing and switching system is provided. An MTBF of 5,000 hours for each side of the redundant configuration and an MTTR of five hours for any equipment failure are also assumed. Note, however, that in the redundant system, a single or one-side equipment failure will not cause an actual system outage. Only two simultaneously existing failures, one on each side, can do this.

One further assumption is made in Figure 3, that failures on the two sides are totally random and independent.

$$\text{OUTAGE RATIO } (U) = \frac{MTTR}{MTBF}$$

$$\text{INNAGE RATIO } (A) = 1 - U$$

$$\text{ANNUAL OUTAGE} = 8760 \times U \text{ hrs.}$$

Figure 1

NON-REDUNDANT

$$U = \frac{5}{5{,}000} = .001 \text{ or } 0.1\%$$

$$A = 1 - .001 = .999 \text{ or } 99.9\%$$

$$\text{ANNUAL OUTAGE} = .001 \times 8760 = 8.76 \text{ hours}$$

Figure 2

REDUNDANT
WITH ASSUMPTIONS

$$(MTBF)_{red} = \frac{(MTBF)^2}{MTTR}$$

Hence,

$$(MTBF)_{red} = \frac{(5{,}000)^2}{5} = 5{,}000{,}000 \text{ hours} = \text{about } 570 \text{ years}$$

Figure 3

These mathematics illustrate that, given these assumptions, the average time between outages (actual system failures) on this hop would be *570 years.*

Continuing with the same redundant example, Figure 4 gives the equation for calculating the outage ratio, U_{red}, for the redundant hop, and the actual calculation for this example.

This now represents a completed calculation which says, given all these assumptions, the equipment reliability characteristic for this hop can be described as 32 seconds of outage per year.

But, this figure of 32-seconds-per-year average outage is only a mathematical abstraction. Since an outage is by its very nature indivisible, there can only be, in any given year, either *no outage at all*, or an outage which, under the assumptions used, must be very much longer (5 hours per failure event in this model). Neither of these conditions —no outage or 5 hours outage— has any real relation to an annual outage of 32 seconds, and consequently the 32-second figure is a very inadequate way of describing this situation.

Figure 5 re-emphasizes the point that whenever the expected outage (MTTR) associated with a failure event is relatively large, the occurrence of such failure events must be extremely rare (MTBF very large)—if ultra-high reliability is to be achieved.

In real life microwave systems there are constraints imposed by the fact that the sytems (at least the long-haul ones where ultra-reliability is most urgently needed) involve unattended repeater stations spread over rather considerable geographic areas, and often in relatively inaccessible locations. This makes it rather unrealistic to assume that the average restoration time, even under favorable conditions, will be less than 1 or 2 hours. Travel time alone will often be greater than this, particularly for failures at isolated points occurring at night or on weekends. In fact, the mathematically convenient assumption of 5 hours may be overly optimistic.

A restoration time measured in hours must be accompanied by equivalent MTBF's measured in millions of hours (hundreds of years) in order to show calculated reliability in this range of 99.9999% per hop.

REDUNDANT – WITH ASSUMPTIONS

$$U_{red} = \frac{MTTR}{(MTBF)_{red}} = \left(\frac{(MTTR)}{(MTBF)}\right)^2 = \left(\frac{5}{5{,}000}\right)^2 = .000001 \text{ or } .0001\%$$

$$A_{red} = 1 - .000001 = .999999 \text{ or } 99.9999\%$$

$$\text{ANNUAL OUTAGE} = .000001 \times 8760 = .00876 \text{ hours}$$
$$= \text{about 32 seconds}$$

Figure 4

FOR 99.9999% RELIABILITY

MTBF MUST BE ONE MILLION TIMES THE MTTR!

e.g., If repair time is five hours, MTBF must be 5,000,000 hours (570 years).

If repair time is one hour, MTBF must be 1,000,000 hours (114 years).

Figure 5

This, coupled with the fact that it is impossible to have a fractional failure in a real system but only integral ones, is the real crux of the problem being discussed.

It has been shown how our example of a redundant hop could calculate out to an average per-hop annual outage of 32 seconds due to equipment. But it must be recognized that in a real system this is a meaningless value which cannot exist except by a wildly unlikely set of coincidences. Even if the analysis and the assumed parameters and conditions were precisely correct, the hop would have to be operated for at least 570 years in order to get even a minimum test, and in that time we get 569 years with zero outage and one year —which could be anywhere along the line— with 5-hours outage. Thus, "annual outage" is quite meaningless, and even the availability or reliability parameters would be meaningful only for the average performance over something like 10,000 years, or 10,000 hops.

The situation is quite different with respect to propagation outages and the kind of difference is shown in Figure 6. Here, a simple propagation situation has been made up which also leads to the same annual outage.

The propagation outages shown are based on a simple assumption of a diversity path with a 40-dB fade margin, Rayleigh fading on each side, and a diversity improvement factor of about 100. Under these assumptions, each side of the diversity would have a reliability of about 99.99% or about 53 minutes of outage per year, consisting of perhaps 1,000 individual hits averaging on the order of 3 seconds

	PROPAGATION	EQUIPMENT
#ONE-SIDE FAILURES PER YEAR	1,000	2
#OUTAGES PER YEAR	20	1/570
AVERAGE LENGTH OF EACH	1.5 sec.	18,000 sec.
TOTAL ANNUAL OUTAGE	30 sec.	about 30 sec.
RELIABILITY	99.9999%	99.9999%

Figure 6

each. The diversity improvement factor of 100 to 1 would lead to about 20 simultaneous hits per year, that is, 20 actual outages, each averaging about 1.5 seconds in length.

The 20 or so simultaneous hits, giving a total annual outage of about 30 seconds, constitute enough events to provide a reasonably adequate statistical population over a year, so that results expressed in this way are quite meaningful and can be related to real-life systems.

But the situation is quite different in the equipment column, in which there are about 2 one-side failures per year, and about 1/570th of an actual outage per year, so that the annual outage is 1/570th of 18,000 seconds, or about 30 seconds.

The difference in scale and sample size between the two situations is about 10,000 to 1, and it is clear that, despite the fact that in both cases there is a calculated annual outage of 30 seconds, the two types of outage are in fact totally and radically different in nature and cannot be usefully combined or treated in a similar fashion.

Twenty outages per year, each averaging less than two seconds, and one outage of several hours occurring only once every five or six *centuries* simply have nothing in common with each other.

The point is that in such ultra-reliable cases, the propagation reliability and equipment reliability of microwave hops must be treated and described separately.

Annual outage remains a good way to describe the propagation reliability. Availability, or reliability in percent, is equally good. It would be useful, however, to include information about the number of events and their average duration, the annual outage being the product of the two.

For equipment reliability, two alternative methods seem to have some merit, though neither is entirely satisfactory.

One is simply to state the equivalent system MTBF as a parameter. In the case of ultra-reliable systems this is usually the redundant MTBF. Preferably the MTBF in hours should be divided by 8760 and the result stated in years, since it is easier to relate to the real world. A statement that the MTBF of a microwave hop is 570 years is likely to arouse some skepticism on the part of engineers familiar with electronic equipment; whereas, a statement that it is 5,000,000 hours might not have the same impact.

A second possibility would be to use this equivalent redundant MTBF to calculate the probability that the hop will operate without failure for a period of a year, using the standard reliability formula as given in Figure 7.

The expression R(t) gives the reliability function in the nomenclature used by reliability engineers; that is, the probability that the device under consideration will operate without failure for a time t.

Summing up, the equipment reliability calculations, in situations of this type, are really saying that there is a very high probability that the outage due to equipment in any year will be zero, but if such an outage does occur, it will be very long (comparatively) and will probably use up the allocated outage time for hundreds, perhaps thousands of years.

This poses the very serious problem that if —as is very likely to happen— equipment reliability prognostications, showing average per-hop outages of seconds, or even a few minutes, per year, somehow get turned into specification requirements (rather than just calculations or estimates), the supplier is faced with the awesome realization

$R(t) = e^{-t/MTBF}$

Which for a t of 8760 hours and MTBF of 5,000,000 hours comes to 99.825%.

For MTBF of 1,000,000 hours, R comes to 99.124%.

Figure 7

that the only way he can meet such a specification at all, over any time period of interest —even the entire life of the equipment in some cases— is to have zero outages due to equipment.

Another serious —though perhaps less apparent— problem is that there is no evident way to make any realistic evaluation of the relative worth of simply changing the odds that there will or will not be an outage. For example, suppose one has a hop with a predicted probability of one outage every 100 years. How much would it be worth to reduce the outage probability to one every 600 years? In either case any outage in a year, or even over the life of the equipment, is highly unlikely, and in either case, if an outage does occur, its length will be the same— the X hours it takes to repair and restore the equipment.

The limitation discussed here is a basic one which does not depend at all on the validity of the assumptions or the calculations. It results simply from three things: microwave systems distributed over wide geographical areas; repeater stations (and often terminals as well) operated on an unattended basis; and outages due to equipment failure (unlike those due to propagation) requiring human intervention to restore and consequently, in general, requiring a rather large block of outage time associated with any outage event.

Regardless of the means used to describe it, there seems to be a parameter, with respect to equipment outages, which describes a situation that cannot exist in the real world, cannot be measured, and to which it is difficult to assign any economic or monetary value.

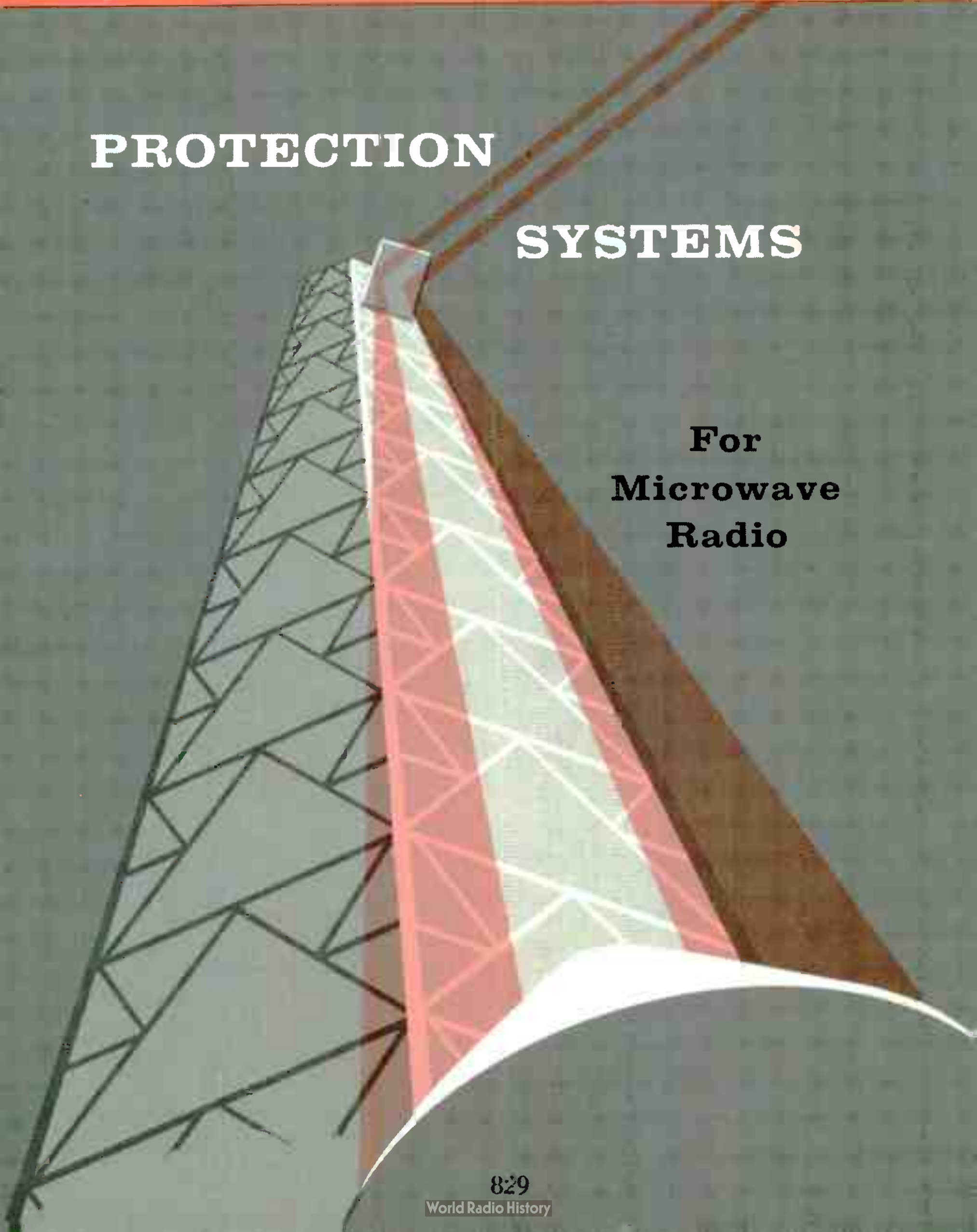
A further consideration is that the models customarily used in making equipment MTBF calculations consider only those outages or failures caused by chance, random failure of individual components for which no cause can be determined, and thus exclude most of the failures which occur in real systems— for example, failures due to human error in the design, the manufacture, the installation, the operation, and the maintenance areas; "early" or burn-in failures; wear-out failures; or unusual stress situations affecting both sides of a redundant system. Therefore, it is apparent that such *a priori* equipment reliability calculations should be treated with considerable caution.

FEBRUARY 1967
VOL. 16, NO. 2

The Lenkurt DEMODULATOR

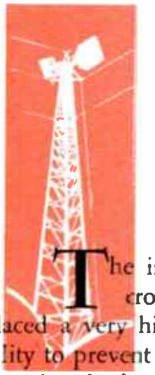
PROTECTION SYSTEMS

For Microwave Radio



829

Microwave radio systems are playing an extremely important role in the expansion and modernization of fixed communications facilities. The load handling capacity of microwave communications systems has been steadily increasing during the past ten years. This improved capability has made them increasingly useful for telephone traffic, broadcast audio and television, telephotos, and in meeting the rising demand for digital data communications.



The increased usage of microwave radio systems has placed a very high premium on reliability to prevent outages involving perhaps hundreds of communications circuits — especially those carrying vital messages and real-time data. The great importance placed on reliability has brought about several methods of protecting microwave systems against the possibility of interruptions in service.

Loss of primary power, equipment failure, and multiple path (multipath) fading are the three most serious events which can interrupt service on microwave radio communications systems. Primary power failures are counteracted by auxiliary sources of electric power standing by ready to assume the load. High quality components and complete "hot standby" systems that switch into service automatically if the primary system fails are used to protect against equipment failures.

The most troublesome event to guard against is multipath fading. This elusive and random phenomenon unpredictably *drops* the signal level at the receiver of microwave radio systems. Multipath fades occur mostly at night, are frequency selective, and pass very swiftly. They are caused by reflected or refracted energy arriving at the receiver out of phase with the direct beam.

Multipath Fading

Microwave radio beams exhibit many properties of light — they travel in relatively straight lines, are reflected by large flat surfaces, and are refracted by atmospheric conditions. During normal conditions, temperature, humidity and pressure in the lower atmosphere decrease almost linearly with increased altitude. This produces a corresponding linear decrease in the refractive index of the atmosphere. The velocity of microwaves traveling through the atmosphere increases as the refractive index decreases. As the wavefront passes through a normal atmosphere, the increased velocities at the top of the wavefront cause the microwave beam to bend slightly downward in a relatively uniform curve. This curve is about one-fourth the curvature of the earth.

Unfortunately, normal atmospheric conditions do not always prevail. Irregularities in the atmosphere cause energy components of a microwave beam to be reflected or refracted upwards or downwards instead of follow-

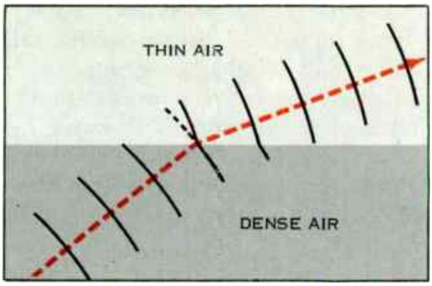


Figure 1. The velocity of microwaves traveling through the atmosphere increases as the refractive index of the air decreases. The increased velocities at the top of the wavefront cause the microwave beam to bend.

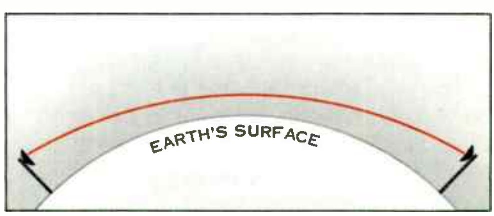
ing the normal slightly curved path to the receiving antenna. As a result, two or more separate wave components may travel to the receiver over slightly different paths. These components will be somewhat out of phase with each other because of the difference in the length of the path each has traveled. Also, at each point of reflection a 180° phase shift normally occurs.

If two equal signal components travel paths that are different by a wavelength, and one signal component has been reflected, they will arrive at the receiver 180° out of phase. Since the signal energy at the receiver is the vector sum of the wave components, the resultant strength of the two signal components will be zero. It is interesting to note that the wavelength of a 6000-MHz carrier is only about two inches. Thus, a one inch difference in the path traveled by two equal signal components (without a reflection) will cause total energy cancellation.

Normally, the information content of a microwave carrier is not disturbed by multipath fading. Differences in path length due to atmospheric refraction and reflection usually vary from a fraction of an inch to perhaps six or seven feet. Such differences cause many opportunities for severe phase separation of carrier wave components at microwave frequencies. However, such differences in path lengths are only a fraction of the wavelength of the information-bearing or modulating component of the microwave carrier. For example, the highest frequency in the baseband modulating signal of a typical 600-channel multiplex system is less than 3 MHz. A 3-MHz signal has a wavelength of about 328 feet. Path length variations of six feet represent a phase difference of only about 6.6 degrees between two 3-MHz components — not enough to cause significant signal distortion. The main problem is to ensure that enough microwave energy reaches the antenna so that the signal can be detected properly in the receiver.

Some multipath fades reduce the signal strength only a few dB, but deep fades may cause it to drop more than

Figure 2. In a normal atmosphere, microwaves beams bend slightly downward in a relatively uniform curve.



40 dB. As already mentioned, multipath fading occurs randomly. Lord Rayleigh showed that random phase cancellations occur in a predictable manner and follow the probability distribution shown in the curve of Figure 3. This curve indicates the probability of microwave fades of various depths. For example, fades of 35 dB or more from the median signal level are shown to occur 0.02% of the time. In a year this would amount to about an hour and 45 minutes during which transmission would be interrupted. This might seem like a relatively short period of time, but it could represent thousands of short term outages. Large numbers of short outages are extremely objectionable when transmititng high speed data.

It should be emphasized that the Rayleigh distribution applies only to the time when multipath fading is occuring. Also, this method of predicting the probability of fades is not precise, but is an expedient that can be used to *estimate* the propagation reliability of a particular system based upon a certain fade margin. The fade margin is the difference in dB between the practical noise threshold (or FM improvement threshold) and the median receiver input signal level selected for the system. Propagation reliability is simply the percentage of time that the receive signal is above the noise threshold point of the receiver. The fade margin required to meet propagation reliability objectives, determined by actual field tests and economic considerations, is a function of output power, receiver input power, antenna gain, and waveguide and free space losses. Typically, fade margins for line-of-sight microwave radio systems operating in the 6-GHz band range from about 30 to 40 dB, usually providing propagation reliabilities greater than 99.9 percent.

Engineering a microwave line-of-sight path with a fade margin that will assure almost 100 percent propagation reliability can be very costly. The expense of shorter hops, with larger antennas, reflectors, and towers needed to obtain extra power to offset deep fades places an economic limit on such an approach.

A more practical approach has been to use *diversity reception* — a means of reducing the effects of multipath fading. Three types of diversity reception commonly used in radio systems are *polarization, space,* and *frequency*. Polarization diversity has been useful in lower frequency radio systems that use sky waves, but it has not provided any significant advantage in line-of-sight microwave systems.

Space and Frequency Diversity

In space diversity, two or more receiving antennas intercept signals from the same transmitter. The receiving antennas are usually separated vertically on the same tower, providing separate direct paths from the transmitter. It is highly unlikely that signals traveling over vertically separated line-of-sight paths will incur deep fades simultaneously because of the vertical characteristic of the phenomenon. Thus, a sufficiently strong signal should usually be present at one of the receivers. The amount of vertical separation required for the receiving antennas may range from a few feet to over 80 feet.

A minimum of one transmitter and two receivers with separate antennas are required at the terminals of a space diversity system. However, a hot-standby transmitter may be required to protect against equipment failures. Since only one frequency is needed, this type of diversity reception is especially useful in conserving frequency allocations.

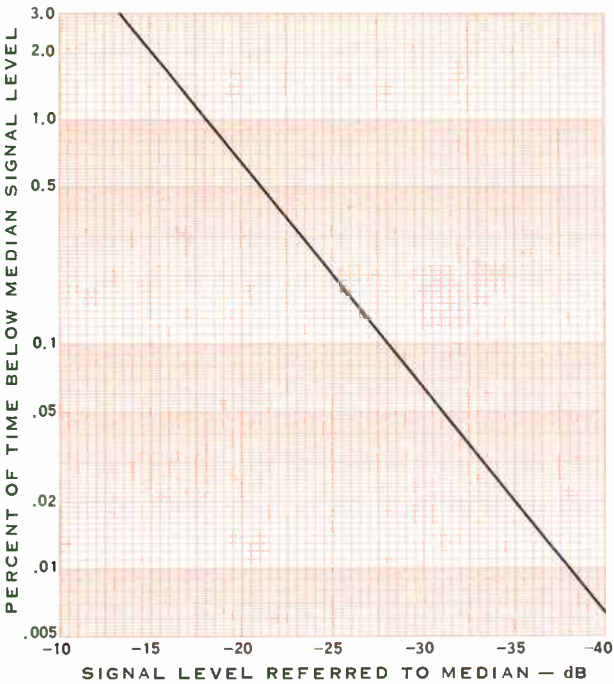


Figure 3. Rayleigh distribution curve shows probability of fades of various depths.

Frequency diversity has been the most popular type of diversity reception where the use of two or more frequencies in one system is permitted. Since multipath fades are frequency selective, microwave signals sufficiently different in frequency are not likely to fade simultaneously due to the different wavelengths of the signals.

This type of diversity reception requires two or more transmitters, each operating at a different frequency, and two or more receivers, generally using the same antenna. The completely redundant system also provides an improvement in overall equipment reliability. Like space diversity, a sufficiently strong signal should usually be present at one of the receivers. Normally, the frequencies selected for this type of diversity reception are within the same band allocation. For maximum propagation reliability, the frequencies should be separated by about 5 percent of the lower frequency. In practice, it is seldom possible to obtain more than about 2 or 3 percent separation because of limited frequency allocations within the band. An extension of frequency diversity, called cross-band diversity, is sometimes permitted using frequencies in different bands (e.g., 6 and 12 GHz).

In any type of space or frequency diversity system some sort of *combining* technique is used at the receive terminal to process the diversity signals. Three forms of combining in use are *variable gain, equal gain,* and *optimal switching*. Combining can be accomplished at the baseband level (post detection combining) or at the intermediate frequency level (predetection combining). Post detection combining is the most commonly used method.

Variable Gain Combining

In variable gain combining, the signals from two diversity receivers are amplified then added together to form a combined output. The amount of amplification each signal receives depends on the signal-to-noise ratio. The signal with the highest signal-to-noise ratio receives the largest gain and thus provides a larger portion of the baseband output. The two signals add together on a voltage basis (20 log) while noise in the two circuits is random and tends to add on a power basis (10 log). When the signal-to-noise ratios of the two signals are equal, the signal-to-noise ratio of the combined output, theoretically, will be 3 dB better than that of either receiver.

Figure 6 is a simplified diagram of a typical variable gain combiner circuit that combines the output signals from two diversity receivers. The signals from each receiver are at a fixed level and are adjusted so that they are equal both in magnitude and phase. The filtering and sensing unit bridged onto each receiver leg monitors the channel. A pilot monitor signal and a noise monitor signal (proportional to the noise sensed in the channel) are fed from each receiver leg into a control unit. The control unit continuously compares the noise voltages in the two legs and feeds a gain control signal to the two variable amplifiers. The control signal adjusts the gain of each amplifier (upward or downward) in accordance with the so-called *ratio-squared* principle. For this reason, this type of device is often called a ratio-squared combiner.

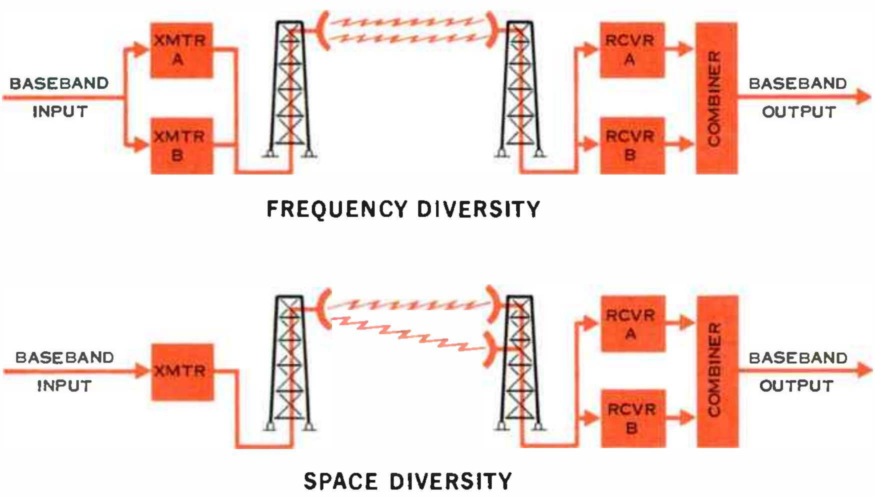


Figure 4. Two types of diversity reception commonly used in line-of-sight microwave systems are frequency diversity and space diversity.

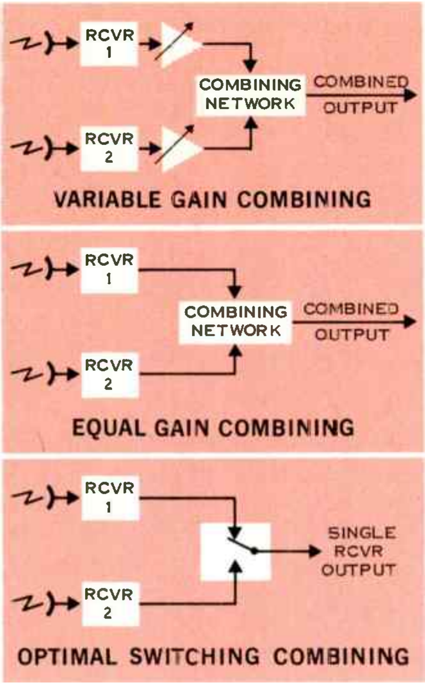


Figure 5. Some method of combining the multiple signals is required at the receive terminal of a diversity system. Three forms of combining in use are variable gain, equal gain, and optimal switching.

If the signal-to-noise ratio in both receiver legs is equal, the gains in the two amplifiers will be equal and each receiver will be contributing the same amount of signal to the baseband circuit. If the signal in one receiver fades, decreasing the signal-to-noise ratio, the gain of the variable amplifier in that leg will drop by a corresponding amount, while the gain in the other amplifier will be increased. Ideally, the amount of gain changes in the two amplifiers should provide an optimum signal-to-noise ratio for the combined output and maintain the signal at exactly the normal level.

If the continuity pilot in one leg disappears, the gain of the amplifier in that leg will be driven to zero while the gain of the other amplifier will be increased to keep the baseband signal at its normal level. The squelch or mute relays in each receiver leg are controlled by the noise monitor signal from each filtering and sensing device. When the noise in a leg reaches a set level, usually about 49 to 52 dBa0, the switch will open, disconnecting the noisy receiver from the combiner circuit. If the noise in both receiver legs reaches the set level at the same time, both circuits will be disconnected and there will be no output signal.

One of the disadvantages of the variable gain combiner is the gain variations it introduces into the system. Stabilizing a variable gain amplifier is far more difficult than stabilizing the gain of the fixed amplifiers in other parts of the system. Consequently, this type of combiner is inherently a source of unstable gains, a feature not found in the other types of combiners.

Also, variable amplifiers are a source of intermodulation noise. In wideband, high density microwave systems intermodulation noise can be a serious problem. To overcome this problem the combiner output signal level usually must be lower than the standard output levels established for microwave systems. In such cases, additional amplification has to be provided after the combiner. Redundant amplifiers with failure sensing devices may be required in this circuit to maintain a satisfactory degree of equipment reliability. All of the additional *active* devices used in this type of combiner tend to lower the overall equipment reliability of the microwave system.

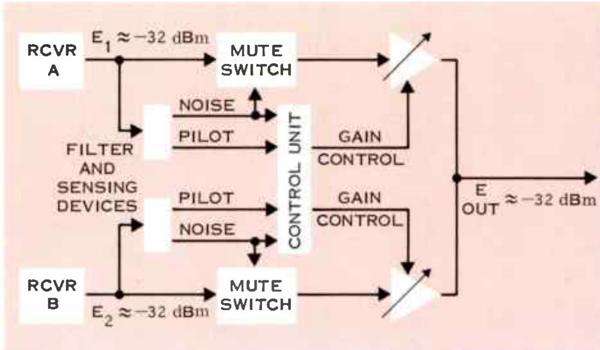


Figure 6. Simplified diagram of typical variable gain combiner circuit.

Equal Gain and Optimal Switching Combining

In equal gain combining, the two signal voltages and the noise power add together in the same manner as in the variable gain combiner. The signal applied to the baseband circuit, therefore, is the sum of the two receiver signals. When the signal-to-noise ratios of the two separate signals are equal, the signal-to-noise ratio of the combined output theoretically will be 3 dB better. As long as the two ratios remain within several decibels of each other, there will be a noise improvement in the combined output. However, the signal-to-noise ratio of the combined output can never be more than 6 dB better than the signal-to-noise ratio of the worst receiver channel. Consequently, this type of combining cannot overcome tremendous increases in noise from a receiver during a very deep fade. The device used to achieve equal gain combining is often called a *linear-adder combiner*.

Optimal switching combining is not actually a combining technique, although it is generally classified as such. In this type of system an automatic switching device monitors continuity pilots and noise levels from both receivers, determines which signal is best, and connects that receiver to the baseband circuit. The receiver with the weaker signal is disconnected from the baseband circuit.

A very practical and reliable method that has been successfully used is a combination of the optimal switching and equal gain techniques. In this type of system, when one of the diversity signals fades too far below normal, its receiver is disconnected from the combiner circuit and only the good signal from the other receiver is applied to the baseband circuit.

Figure 7 is a simplified diagram of a typical combiner circuit using the combination optimal-switching and equal-gain technique.

This circuit, like the ratio-squared combiner circuit, contains a filtering and sensing device bridged across each receiver leg. Also, the receiver output signals are at a fixed level and are adjusted to be equal in magnitude and phase. However, the fixed levels are typically set at about −15 dBm compared to a lower level of about −32 dBm used in ratio-squared systems.

Under normal conditions, half the signal voltage in each receiver leg is

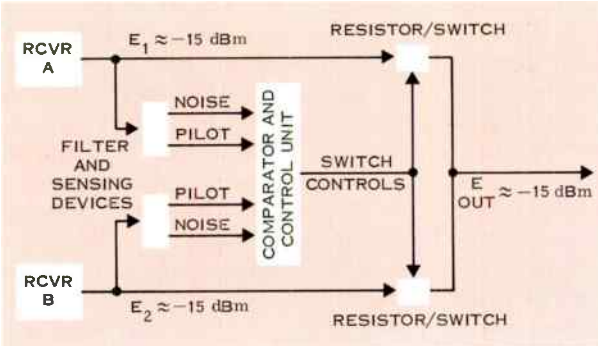


Figure 7. Simplified diagram of typical combination equal-gain optimal-switching combiner circuit.

dropped across a resistive network while the other two halves are combined in phase and applied to the baseband circuit. Because the two signal voltages are in phase, they add directly to produce an output signal voltage equal to that in each receiver leg. As stated previously, the two signals add together on a voltage basis (20 log) while the random noise tends to add on a power basis (10 log).

The comparator and control unit continuously monitors and compares the noise voltages in the two receiver legs. As long as the two voltages are within 5 to 6 dB of each other, the combiner is kept in the *combined* condition. If one signal fades to the point that the noise voltage becomes about 6 dB greater than that in the other leg, the receiver circuit will be automatically disconnected. At the same time, the full signal voltage from the other receiver will be connected to the baseband circuit. The switching time is usually very fast, about 1 to 2 microseconds, to prevent any level or phase changes. The circuit will remain in this *uncombined* condition as long as the noise differential persists. The same switching action will occur if the pilot in one of the receiver legs is lost.

The squelch or noise muting circuit used in the equal-gain optimal-switching type of combiner operates similarly to the one described for the ratio-squared combiner. However, there is no need for extra relays or switches since the function can be performed by the faster-acting solid-state noise differential switches.

Variable gain combining provides a slight advantage over the combination equal-gain optimal-switching type, with respect to receiver front end thermal noise, when the two diversity signals are unequal by a few dB. However, under usual signal conditions front end thermal noise is seldom the controlling noise problem in line-of-sight systems.

Standby Diversity

Frequency diversity systems require twice as much transmitting and receiving equipment as an unprotected system. In large systems, where many radio channels link the same locations, the cost of duplicate equipment for each channel may be prohibitive. Also, there may not be adequate space for all the extra equipment, or the signals might occupy such a large portion of the available frequency band that diversity reception is not feasible.

In such situations it has become practical to use a standby radio channel that is shared by perhaps three or more working channels. The standby channel and the working channels each require a separate frequency assignment. This standby arrangement reduces the cost of providing frequency diversity reception for each channel thereby conserving frequency allocations. Standby diversity, which cannot be used in space diversity systems, provides what is referred to as one-for-two protection, one-for-three protection, two-for-six protection, and so on. The first quantity refers to the number of standby channels and the second to the number of working channels. One-for-three protection, for example, means that one channel is providing standby protection for three operating channels. If one of the operating channels fails, or if its signal fades or contains excessive noise, a sensing mechanism at the receiver switches the signal from the failed channel to the standby channel with little or no interruption in service.

This type of diversity reception has become practical through the development of high-speed solid-state logic and switching circuits that automatically control the transfer of channels, usually on some priority basis. Channel switching can be accomplished at the microwave, intermediate frequency (IF), or baseband signal level, and many types of switching and priority arrangements have been developed to meet the needs of different applications.

Figure 8 illustrates a typical one-for-three protection arrangement. In this arrangement, channel B is operating in frequency diversity with channel A, while at the same time providing hot standby protection for channels C and D. Channel A would normally be carrying higher priority traffic and would have the controlling use of channel B. Thus, if channel A fails and is switched out of service, channels C and

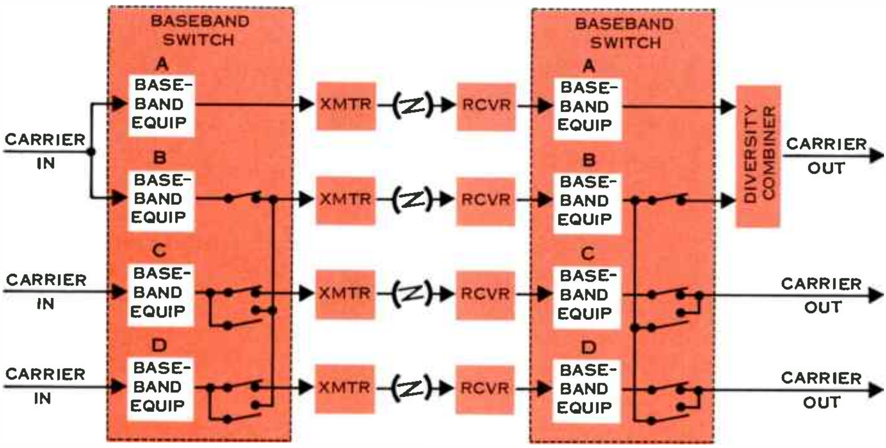


Figure 8. Block diagram of typical standby diversity protection system. Channel B operates in frequency diversity with channel A, while providing backup for channels C and D in the event of a failure.

D will not be protected by channel B until channel A is restored. If channel A is operating, and channel C or D fails, channel B will switch from its frequency diversity arrangement with channel A to support the failed channel. However, if channel A should also fail, channel B will drop channel C or D and restore service to higher priority channel A.

In another type of arrangement, the standby channel might be used to carry low-priority traffic, instead of operating in frequency diversity, while providing protection for perhaps three other channels carrying higher priority traffic. When an equipment failure or excessive fade occurs on one of the regular channels, the standby channel will drop its low-priority traffic and support the failed channel. While the standby channel is supporting a failed channel, the other regular channels are, of course, unprotected.

Conclusions

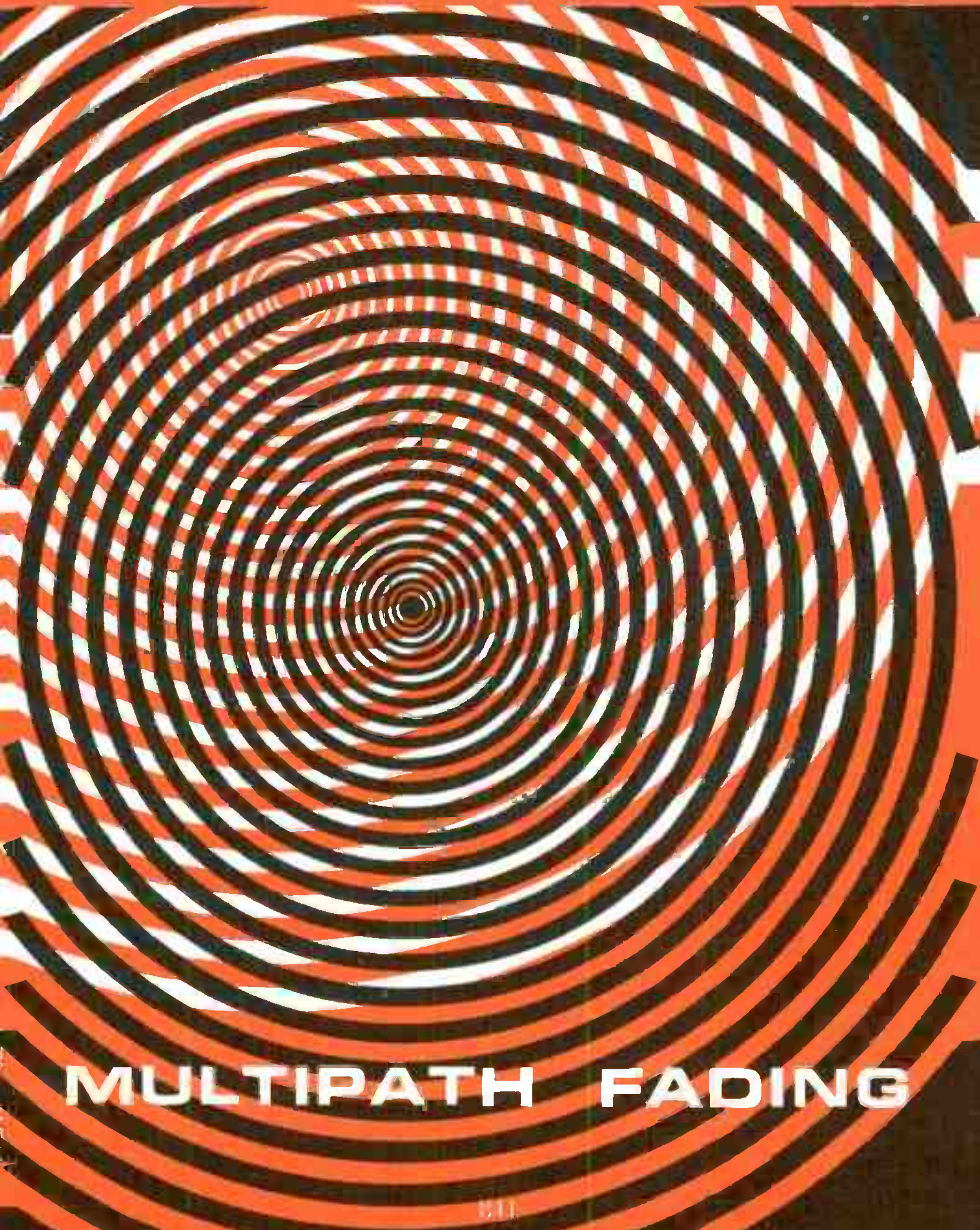
Frequency diversity systems have a significant advantage over hot standby or space diversity systems since any of the duplicate parallel channels can be completely tested while in service. Also, space diversity systems require additional waveguide runs and more complex antenna arrangements which, of course, increases their cost. However, frequency diversity operation is prohibited in many applications because of the amount of congestion within the allocated frequency band. In these situations the choice of protection systems is limited to hot standby or space diversity arrangements.

The choice of combining techniques used at the receive terminal in diversity systems depends, to a great extent, on the operating characteristics of the particular microwave radio system. Tropospheric scatter microwave systems, for example, are usually characterized by continuous fast fading, low signal levels, low fade margin, and relatively high receiver front end thermal noise. Under these conditions, the ratio-squared combiner generally provides the best noise improvement.

The same is not necessarily true, however, in line-of-sight microwave systems. These systems are characterized by relatively high signal levels, high fade margins, normally low receiver front end thermal noise, and only occasional periods of fading. Under average conditions, intermodulation and intrinsic noise usually are a greater problem in these systems than front end thermal noise. For this type of noise, the equal-optimal gain-switching combiner provides the best combiner noise improvement.

NOVEMBER, 1967
VOL. 16, NO. 11

The Lenkurt DEMODULATOR



MULTIPATH FADING

. . . . wave interference caused by atmospheric discontinuities means trouble for space waves.

Electromagnetic wave propagation depends on a number of varying factors. Each of these factors can cause radical changes in the reception of a radio signal.

The sun, the earth's terrain, the weather all work their peculiar influences. The sun, for instance, has a direct effect on the ion concentration of the ionosphere. The ionosphere in turn refracts radio waves.

How a particular signal is affected by sun, terrain or weather depends on its path through the atmosphere. Three such paths have been identified and designated as sky, ground and space waves.

Ground waves, as the name implies, hug the ground. Sky waves travel to the upper atmosphere where the ionosphere refracts them back to earth. Space waves propagate through the atmosphere just above the ground. They usually travel in straight lines.

At any given frequency only one or at most two of these waves are useful. The others are either attenuated or lost when they are not bent back to earth.

Point-to-point microwave systems rely on space wave propagation. As space waves they are supposed to follow straight lines. Occasionally, however, they do not because as electromagnetic waves they are susceptible to atmospheric bending from diffraction, reflection and refraction.

The Bends

A diffracted wave bends around corners because the edges of the wave tend to fill the areas masked by obstacles. A reflected wave bends because it encounters a reflecting medium. A refracted wave bends because its speed changes. Of the three, reflection and refraction pose the biggest problems in microwave transmission.

Radio waves can be reflected from a smooth surface just as light can. The amount of reflection depends on the angle of incidence or the approach angle and the reflective quality of the material doing the reflecting. At low angle reflection the wave undergoes a 180° phase shift.

Figure 1 shows an electromagnetic wave at the point of reflection. If there had not been an obstruction, the wave front would have continued to a'b. The reflecting surface caused a change of direction which resulted in the wave front acb. As the wave front continues

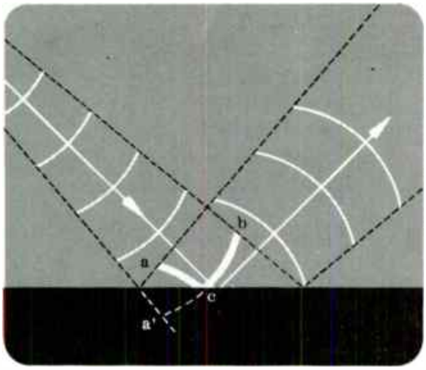


Figure 1. An electromagnetic wave undergoes a change of direction and usually a change of phase at reflection.

to arrive at the reflecting surface, it is redirected and its phase shifted 180°.

The third type of wave bending, refraction, is the least predictable. It is directly related to the condition of the atmosphere and as such also has some influence on the reflection of a radio wave.

Density changes which affect the speed of an electromagnetic wave cause refraction. To describe this effect an index of refraction—the ratio of the speed of light in a vacuum to the speed of light in another medium—was conceived.

The index was originally developed to analyze light rays, but because both light and radio waves are electromagnetic, the principle applies to radio waves as well. The speed of light, or for the purposes of this discussion the speed of a radio wave, decreases as the density of the medium of propagation increases. This relationship of density and speed is an important one.

As a radio wave moves obliquely between two differing densities (Figure 2), its change in speed alters its direction. If the wave enters a more dense area, the forward part of the wavefront slows, causing it to lag behind its upper portion. This uneven increase in speed across the wavefront forces the wave to pivot around its slower end just as a marching line of soldiers does when turning.

Atmospheric Alterations

If there were no atmosphere—therefore no density changes—a radiated signal would proceed in a straight line.

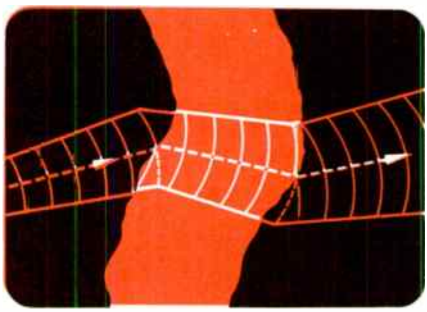


Figure 2. An electromagnetic wave is refracted when it encounters a medium of different density. The resulting speed change usually causes a change in direction.

But there is an atmosphere which can refract waves and therefore alter their relationship with the earth.

A single microwave beam, for instance, might follow a number of available paths. One might bring it down to the earth, another bend it away, or still another might lead it in a curve roughly equivalent to the curvature of the earth.

In a standard atmosphere where density decreases with height, the prevailing tendency of a space wave is to curve but at a slightly slower rate than the curvature of the earth. Unfortunately, the atmosphere does not always conform to a standard density pattern.

Figure 3 illustrates various atmospheric profiles each of which will change the propagation path of a space wave. The graphs use a modified refractive index, M, defined in units which relate the curvature of a microwave beam to the curvature of the earth. The value of M at a given altitude depends not only on the index of refraction at that altitude but also on the ratio of that altitude to the radius of the earth.

Figure 3A shows a slope which is standard for most of the earth. When the slope of the M curve is greater than normal but not negative, the tropospheric condition is known as superstandard (earth flattening), hence the radio horizon distance is increased. When the slope of the M curve is less than normal, the tropospheric condition is known as substandard (earth bulging).

Figures 3B and 3C show two profiles which indicate an atmospheric inversion. In both cases a space wave signal will be trapped at the elbow of the curve. This produces a condition known as ducting which can confine a signal to one height. When this happens, a signal caught in an unfavorably located duct can be at least partially blocked. An unfavorably located duct is one not at the same height as the receiving antenna.

K and the Earth

The need to correlate the conditions of the atmosphere and the curvature of a radio signal has led to the definition of an equivalent earth radius factor K. This factor compensates for apparent variations in the curvature of the earth caused when an electromagnetic wave bends. In effect the earth is flattened, bulged or depressed by the condition of the earth's atmosphere.

In a standard atmosphere K equals 4/3 of the curvature of the earth. With a standard atmosphere the earth does not fall away from a microwave beam in as short a distance as would be expected. The beam has a curvature less than that of the earth.

The earth appears to become increasingly flat as the value of K increases. When K equals infinity the earth ap-

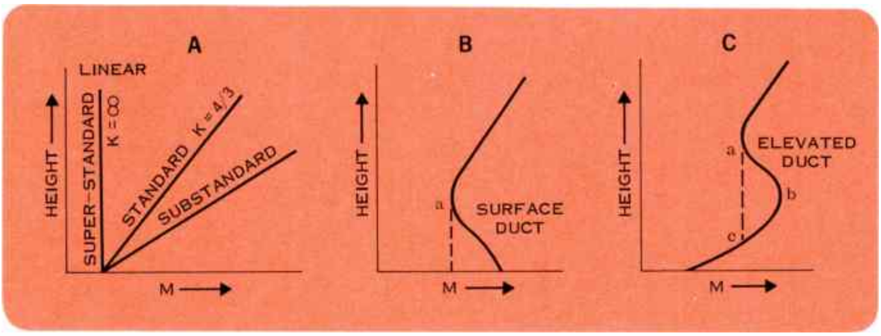


Figure 3. Typical M profiles showing atmospheric conditions. Profiles B and C indicate an inversion which can cause ducting. Each profile has associated with it different propagation characteristics.

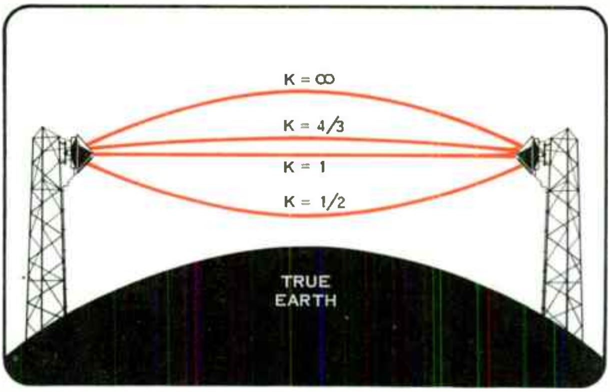


Figure 4. The equivalent earth radius factor K shows the path of a microwave beam relative to the surface of the earth. (For the purpose of illustrating this, the earth's curvature is exaggerated.)

pears to a microwave beam to be perfectly flat. In effect, when K is at infinity, the microwave beam curves at exactly the same rate as the earth.

If the value of K becomes less than one, the curvature of the beam becomes negative. The beam itself curves in a direction opposite to that of the earth.

To the beam the earth appears to bulge. The effect of bulging is to put obstacles in the way of the transmission path.

The earth's actual surface also has an influence on the atmosphere. Over certain kinds of terrain the influence is negligible but over others it is especially significant.

The atmosphere over flat lands or water, for instance, is subject to temperature inversions which can cause ducting. Atmospheric turbulence in mountainous areas causes mixing which aids space wave transmission.

The earth's terrain also affects the propagation of microwaves by causing or preventing reflections. Reflections from a rough surface are usually no problem because the incident and reflective angles are quite random. A relatively smooth surface, however, can reflect signals toward the receiving antenna.

Transmission Paths

Both reflection and refraction occasionally complicate transmission paths. In itself this is not a problem. The problem arises because a radio signal is not the neat little beam depicted in most diagrams. In reality radio signals spread as they advance, becoming not just one beam but, theoretically, an infinite number of them (Figure 5).

Each component of the wave, traveling its own unique path, is subject to different reflections and refractions. Some components do not reach the receiving antenna at all. Of those that do, there can be both principal and secondary wave components.

One convenient definition of these components hinges on the paths they follow. The principal component is the direct or unobstructed component and any reflected component of the wave. Secondary components are those which travel multiple paths through the atmosphere.

At any instant the signal strength at the receiving antenna is the vector sum of its components. Combinations of the principal and secondary components produce phase interference—one cause of multipath fading.

Two components of a wave can cause a 6 dB increase in signal strength or a complete null. The degree and character of interference depends on the amplitude and phase difference of the two components.

The exact nature of a cancellation or reinforcement varies with the circumstances. Components 180° out of phase experience a degree of cancellation directly related to amplitude. On the other hand components of the same amplitude experience a degree of interference dependent on phase differences.

There are examples of multipath fading which appear to be the result of more than two interfering waves. Investigations have led to the assumption that very deep multipath fading often results from the coincident arrival of a signal weakened by its reflected component and a secondary wave.

Solutions

Although multipath fading is quite random, there are ways to compensate for it. The least subtle and most obvious solution is to provide extra signal strength—increased by an amount known as the fade margin. This has the effect of increasing the amount of fading a signal can withstand before it becomes unusable.

A fade margin figure for a typical 6-GHz path is 30-40 dB. When the signal path is over a good reflecting surface such as water, additional fade margin must normally be provided.

Fade margin can be obtained in four ways. The first is to increase antenna gain by making the antenna larger. The second is to reduce the distance between antennas. The third and fourth are either to increase the transmitter power output or decrease the receiver noise figure. The effects of these adjustments vary widely and are often limited by expense or location.

Another solution to multipath fading is the use of diversity systems. Three methods have been tested and two are known to have been put into practical operation on microwave systems. Both frequency diversity and space diversity can reduce the amount of outage time due to deep multipath fading on most microwave systems.

The third method is polarization diversity. It requires two synchronized radios transmitting the same information on the same frequency but at different polarizations. The method has been useful in lower frequency radio systems using sky waves, but with space waves both polarizations have been found to fade simultaneously.

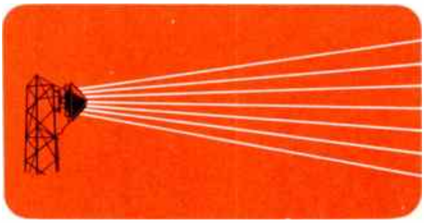


Figure 5. Radio signals spread as they travel through the atmosphere. In effect the signal becomes many beams each subject to the atmosphere's influence.

Frequency Diversity

Frequency diversity systems require at least two separate transmitters and two receivers operating on different frequencies. Normally it is not necessary to have a separate antenna for each transmitter and receiver. The receivers are connected to a diversity combiner which adds the two received signals to form a usable, combined output.

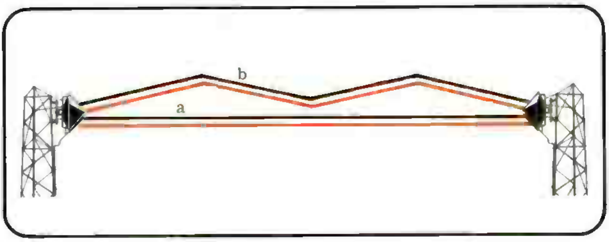


Figure 6. With frequency diversity two wavelengths travel the same refracted path (b) but will not have the same interfering effect on the direct wave (a).

Most frequency diversity systems have frequency separations of 2-5 percent of the lower frequency. This separation keeps the frequencies within the same band. Some systems use frequencies from two microwave bands, for example 6 and 12 GHz, thereby obtaining much greater separation. This latter method is called crossband diversity.

The effectiveness of frequency diversity depends on the wavelength differences of the frequencies in use. Fading occurs when the components of a signal interfere in such a way as to cause cancellation. Interference depends on the relationship of direct, reflected and secondary waves. With signals having different wavelengths but following the same paths, it is unlikely that they will cause simultaneous deep fades.

When considering any given path over which both frequencies must travel, it is easy to see why interference will not occur simultaneously on both frequencies. For each frequency there may be a number of different paths, but neither frequency can follow one path to the exclusion of the other frequency. When the wavelength of one frequency travels a distance which causes interference with the direct component of that frequency, the wavelength of the other frequency—traveling the same distance—will not have been delayed enough in its travels to interfere to the same extent with the direct component of its frequency.

As a solution to multipath fading, frequency diversity is simple and useful. The redundant arrangement of transmitters and receivers gives the system two complete electrical paths. This is a good hedge against equipment failures and an advantage when performing checks where service cannot be interrupted.

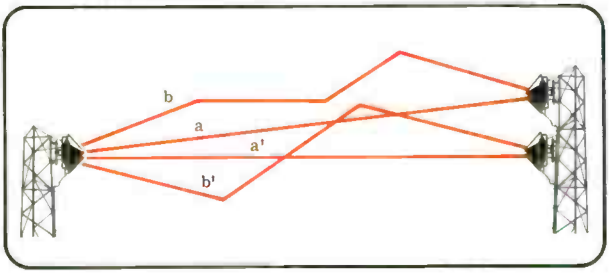


Figure 7. With space diversity the same wavelengths travel different refracted paths (b and b') but will not have the same interfering effect on the direct waves (a and a').

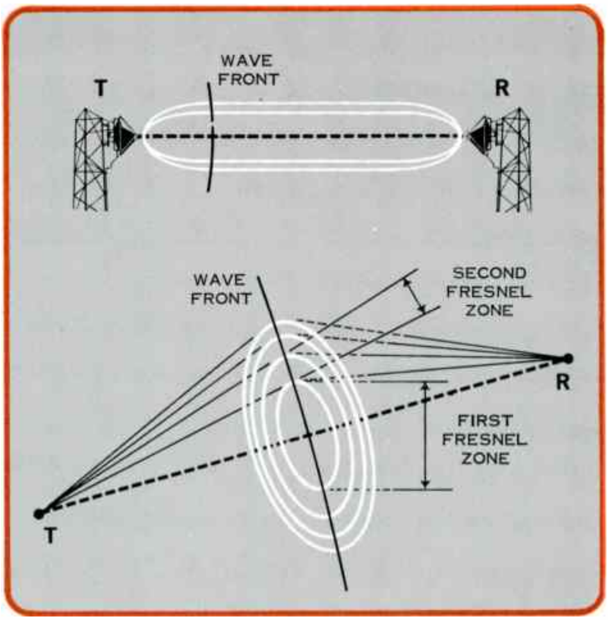


Figure 8. An unlimited number of Fresnel zones surround the direct beam. Here zones 1 through 4 are shown.

Space Diversity

In a typical space diversity system one signal is transmitted to two vertically spaced receiving antennas. At the receiving station the two signals are combined.

A space diversity system requires only one transmitter and two receivers, although most systems have a second transmitter in hot standby. In addition, a system must have at least two receiving antennas in order to provide the required vertical spacing. Each antenna must have its own waveguide.

Normally the additional antenna means a stronger if not separate antenna tower. In all probability the tower will also have to be taller in order to insure adequate vertical separation.

Unlike frequency diversity, which relies on wavelength differences, space diversity relies on path length differences. The working concept behind a space diversity system is that components of the same signal traveling different paths will not have the same interference points. The same wavelength is interfered with differently at two vertically separated points because it travels different length paths to the antenna.

Space diversity is the best protection against multipath fading if microwave frequencies are scarce. Although it does not have the advantage of two complete electrical paths, it avoids the problem of obtaining two frequencies for the same transmission path.

Interestingly, there is a growing controversy as to how the spacing between antennas should be determined. There are two schools of thought on the subject and, appropriately, they are diverse.

One approach is based on the assumption that multipath fading involves a complex interaction from more than one source of interference. This makes it difficult to calculate optimum antenna spacings on anything other than a statistical basis. As a result spacings are usually selected which are as wide as possible (based on the empirical conclusion that improvement would probably increase with separation) considering tower heights and other mechanical factors.

The other approach uses discrete, calculated spacings to combat simple two path, reflected component interference. It relies on a known vertical pattern of signal strength consisting of alternating nulls and maxima.

Fresnel Zones

Analysis of this pattern has shown that interference depends on the vertical distance between the direct component and a reflecting surface. The relationship is conveniently defined by Fresnel zones.

These zones form a series of concentric circles around the direct or shortest path between transmitter and receiver. The positions of the zones are wavelength dependent. Each zone contains wave components traveling paths no more than half a wavelength different in length. Two paths passing through corresponding points in adjacent zones will differ in length by half a wavelength.

Fresnel zones, of which there are an unlimited number surrounding a path, are numbered from the center out. Paths through the first Fresnel zone vary in length from the direct path by as much as half a wavelength. Paths through the second zone vary between half and one wavelength, those through the third by one to one and a half and so on. Each zone number represents an increase of half a wavelength in path length.

Figure 8 illustrates how the Fresnel zones surround the direct path. Each successive zone passes wave components which travel half wavelength differences. These zones can be used to determine where out-of-phase paths occur. Transmission engineers normally refer to paths passing through even zones as having components which cancel and those passing through odd zones as having components which reinforce the direct wave.

A knowledge of Fresnel zones is useful when planning a transmission path over reflecting surfaces. Because even Fresnel zones contain wave components which cancel the direct component, surfaces which reflect even zone components should be avoided. Logically the vertical distance between a direct component and a reflecting surface should

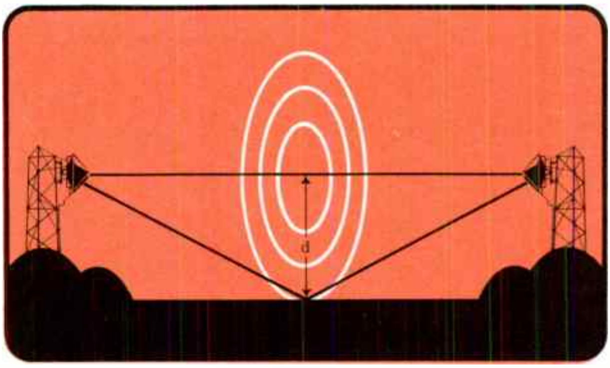


Figure 9. Reflected waves can interfere with direct waves. A distance, d, between reflecting point and direct wave equal to an odd Fresnel zone radius can cause reinforcement. If d equals an even zone radius, the two waves can cancel.

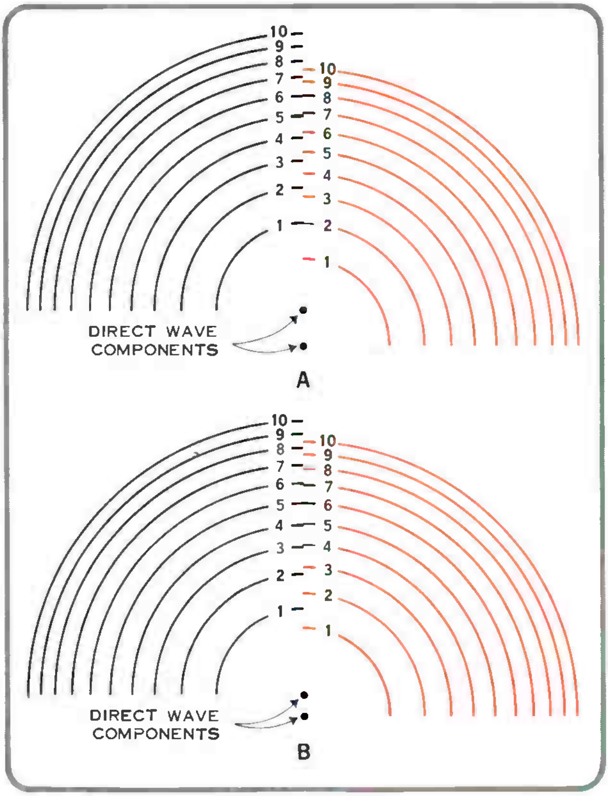


Figure 10. Fresnel zones around two direct waves can coincide. A reflection at zones 6 and 8 in A would cause cancellation. To overcome this other direct waves must be chosen to obtain the pattern in B. This is done by changing the position of at least one of the antennas.

be less than the radius of an even Fresnel zone.

Figure 9 shows the direct and reflected components of a wave. If the reflecting point is so located that components in an even zone are reflected, there will be a cancellation of the direct wave. It is for this reason that transmission paths are engineered to avoid reflections at even Fresnel zones.

On a transmission path receiving antennas are spaced to intercept two direct components. By calculating the reflecting points based on an expected value of K, it is possible to determine which direct components and which antenna heights are best suited to take advantage of the Fresnel zones.

Using the Zones

A rudimentary approach to calculated antenna spacings is to locate one antenna to receive waves reflected at an odd Fresnel zone for standard atmospheric conditions. The other antenna is then placed to receive those reflected at an adjacent even zone.

If the equivalent earth radius factor changes, the height of the direct wave changes. The positions of the Fresnel zones relative to the earth also change.

Unless there is a radical change in K, the antennas will continue to intercept reflections from adjacent Fresnel zones. As K increases, however, it is possible that reflection will occur at adjacent even or odd Fresnel zones.

This can happen because higher numbered Fresnel zones are closer together. This means that the Fresnel zones around two direct components creep up on each other (Figure 10). Eventually an even zone associated with one component can coincide with an even zone of another component. For instance, zone 8 might coincide with zone 6 on another direct component. When this happens at a reflecting point, there will be simultaneous nulls or maxima on both antennas.

To avoid this, antenna positions are sometimes fixed by determining or assuming the maximum possible value of K for the transmission path. With this method the distance between direct paths is chosen so that coincidence between even Fresnel zones will not occur at any reflecting points. There is some inefficiency in this system when K is at its normal value but this does not reduce the diversity effect.

The theory underlying the calculated approach is that reflections from locatable sources are the major contributers to multipath fading. On overland paths this is not always true. It is quite possible to have no reflections or to have two or more of them.

Studies of deep fading microwave signals have been made using paths with low coefficients of reflection. In spite of the low reflectivity, two, three, and sometimes more signal components were found. This discovery led to the belief that ground based reflection is not the only cause of multipath fading.

In Sum

Several years ago Lenkurt engineers, as well as other engineers, determined that diversity improvement would probably increase as vertical antenna separation increased. They concluded that in the 6-GHz band a spacing on the order of 30 to 40 feet offered a reasonable trade off between diversity improvement and tower height. Field experience with many systems has shown that space diversity engineered in this fashion provides extremely good protection against fading of the multipath type, whatever its source.

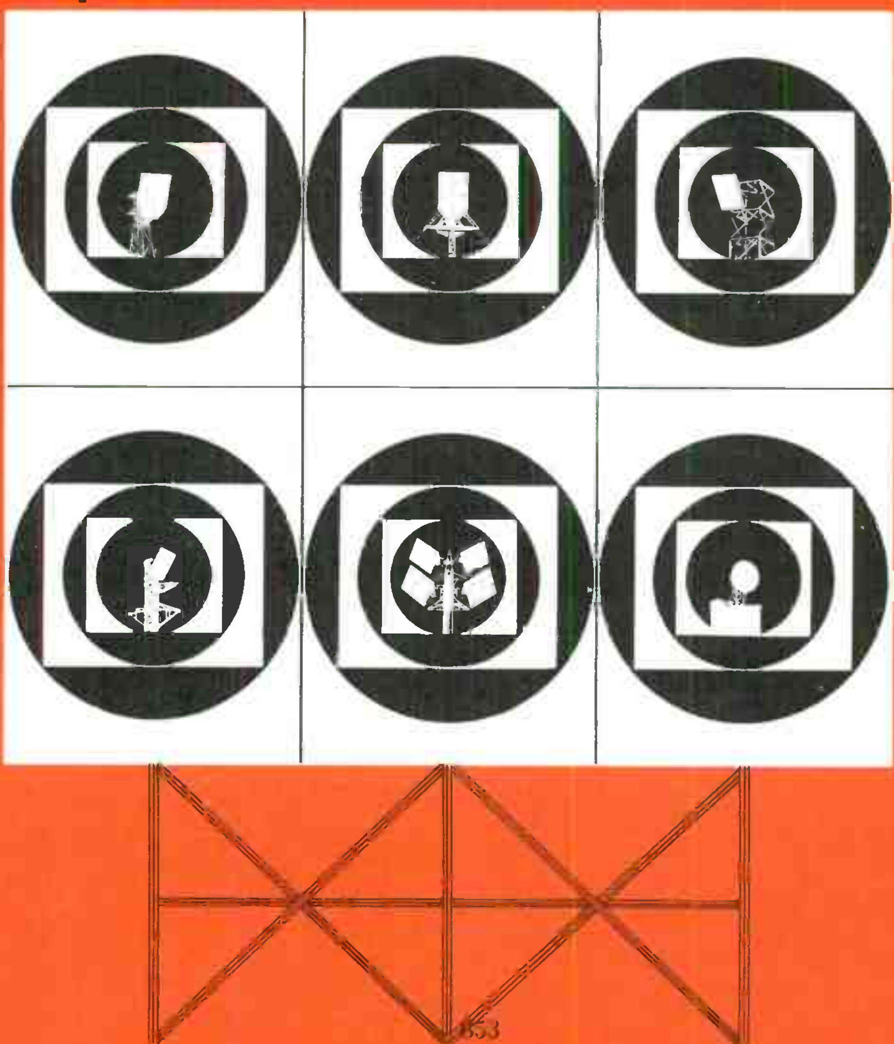
Based on these results space diversity appears to be an effective protection against multipath fading. It takes into account the disruptive influences of atmospheric reflection and refraction. In fact at times it capitalizes on these phenomena to obtain stronger signals than would be expected.

Whether the biggest fading damage is done by a reflecting surface or a refracting atmosphere depends on the specific path. It seems that over most microwave paths there is more multipath fading from atmospheric refraction than from reflection.

In either case space diversity or frequency diversity can protect against fading. Frequency diversity uses two frequencies and hence two wavelengths traveling over the same path. Space diversity uses two path lengths to send the same wavelength to the receiver. In both cases the different lengths prevent identical interference.

The Lenkurt
APRIL 1969
DEMODULATOR

Reflectors & Repeaters



853

Radio-reflective surfaces offer design engineers an efficient and relatively inexpensive alternative to some path redirection problems.

Because radio waves bounce off reflective surfaces in much the same way light is reflected by a mirror, radio reflectors can be thought of as radio mirrors.

In actual practice, the use of radio mirrors is dictated essentially by topographical conditions where the ruggedness of intervening terrain either makes a direct path impossible or requires that the antenna towers be extremely high.

Generally, speaking, radio mirrors fall into two categories – reflectors and passive repeaters. Those used in periscope antenna applications – rectangles, ellipticals, or "flyswatters" (Figure 1) are referred to as reflectors. The large "billboards", usually found on isolated hilltops, and certain "back-to-back" parabolic reflector arrangements are both classified as passive repeaters. (Figure 2).

In order to determine the relative advantages of one antenna reflector arrangement over another it is convenient to refer to some standard of measurement. In the case of microwave antennas, performance is measured in gain and expressed in decibels (dB). Using the isotropic antenna as a standard point of reference it is common practice to speak of an antenna's performance as the gain improvement (in dB) over what could be expected of an isotropic antenna.

An isotropic antenna would theoretically radiate or receive energy equally in all directions. (Figure 3). (A completely spherical radiation pattern is not really possible.) If an antenna could focus all its radiant energy into one-half a sphere, its gain (over isotropic) would be defined as 3 dB, since all the radiated energy would be concentrated in half the sphere and twice as much would appear on any given area of the half-sphere. Therefore, gain is $10 \log 2 = 3$ dB. Common beams run as small as 1 to 2 degrees and provide gains in the area of 40 dB.

The primary function of a good microwave antenna is to focus its radiant energy into the most concentrated and efficient beam possible.

Reflectors

All periscope antenna systems require a reflector of some kind to redirect the transmitted beam from

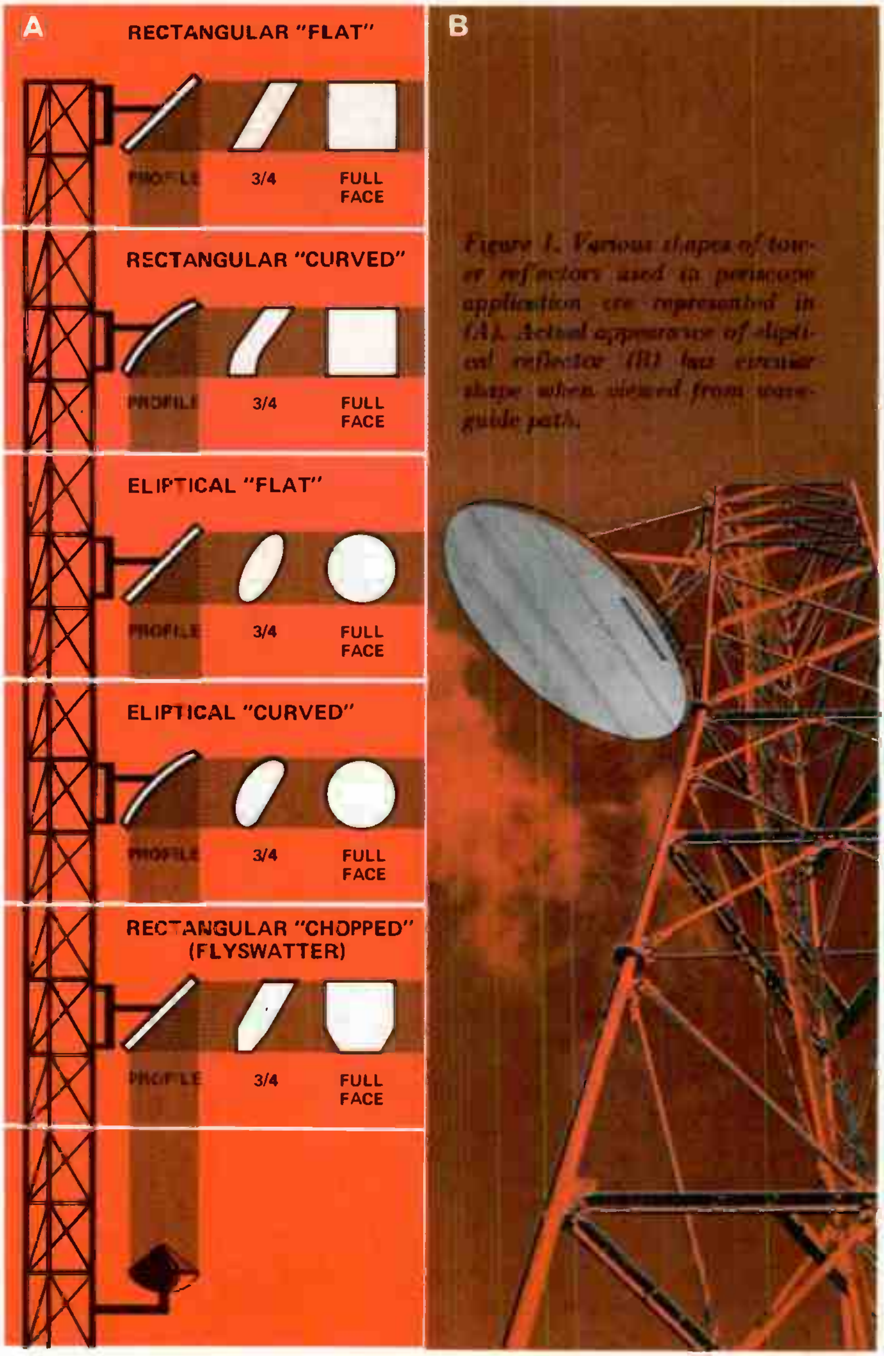


Figure 1. Various shapes of tower reflectors used in periscope application are represented in (A). Actual appearance of eliptical reflector (B) has circular shape when viewed from waveguide path.

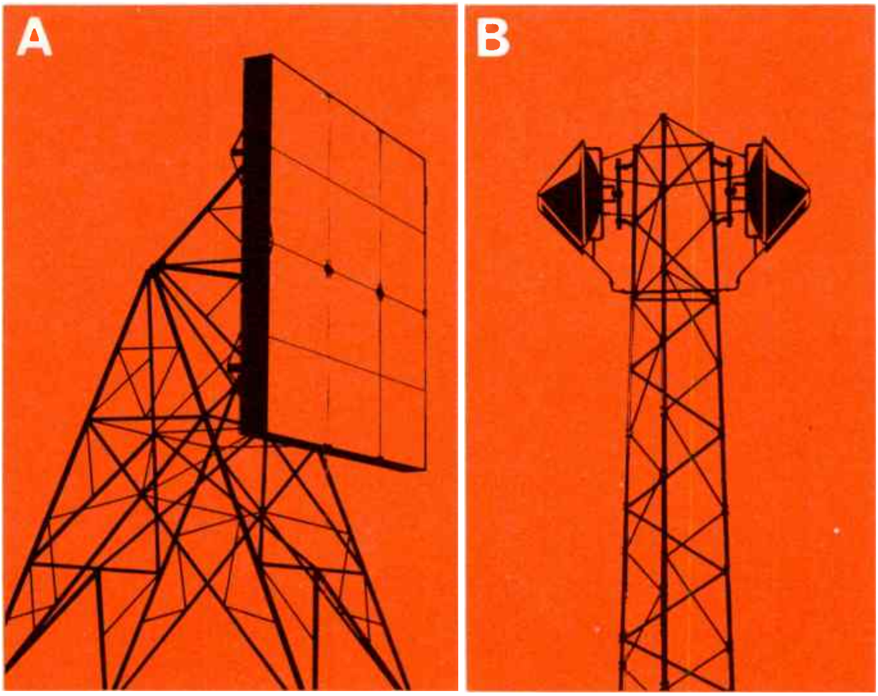


Figure 2. Passive repeaters are of two basic types: the "billboard" (A), so labelled because of its appearance, and the parabolic "back-to-back" passive (B) which uses two standard antenna dishes directly joined by a short length of waveguide.

the parabolic antenna to some distance receiver.

In most cases these reflectors are in close proximity to the transmitter-tower complex. There are some exceptions and because there is no absolute dividing line between what constitutes a periscope reflector and a passive repeater it is generally held that any system in which there is more than a few hundred feet of horizontal separation between reflector and illuminating dish is a passive repeater – not a periscope reflector.

The decision to use a periscope antenna arrangement is dependent on several considerations. Economic studies reveal that when the waveguide run to the parabolic antenna approaches distances of 150 feet and beyond, it is usually less expensive to use the periscope arrangement and beam the signal from the ground to the reflector atop the tower.

Tower height is not, however, the only consideration. Periscope antenna systems typically have somewhat higher side lobes and somewhat poorer

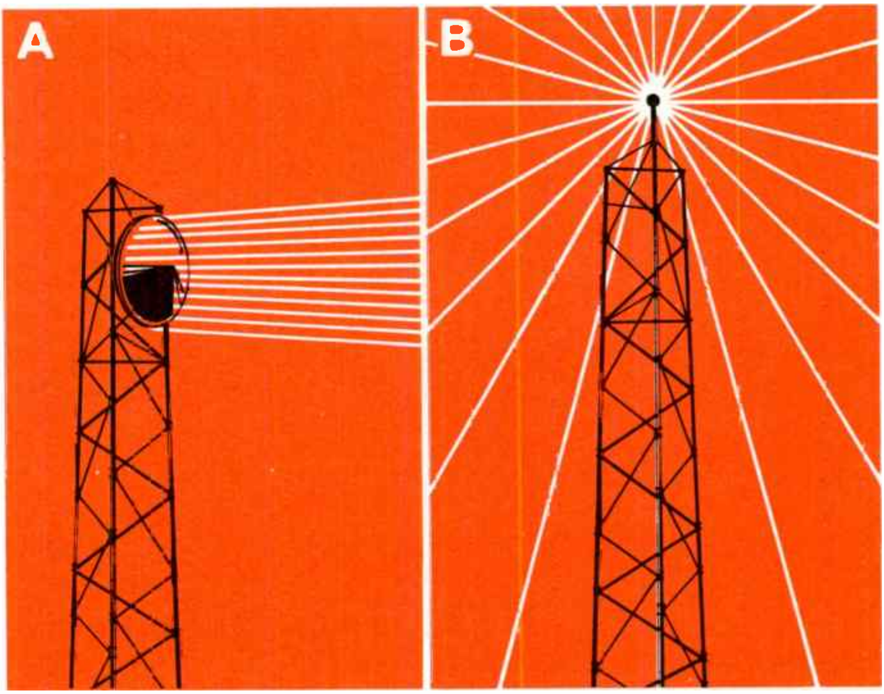


Figure 3. Energy radiation is theoretically assumed to radiate isotropically in all directions (B). Parabolic antennas are used to focus radiant energy into a directional beam (A) which has an obviously high amplitude gain over the hypothetical isotropic antenna.

discrimination patterns for radiation or reception at angles off the main beam than do direct-radiating antennas with comparable gains. They thus have a greater likelihood of creating interference to or receiving interference from other microwave systems operating in the same geographical area. This characteristic is probably the most negative aspect of periscope antennas. In areas of heavy microwave congestion it may be sufficiently important to preclude the use of periscope antenna systems, even though they might be advantageous from other points of view.

Another problem with the periscope setup is the "sneaking" of the signal from the illuminating dish to the distant receiver. (Figure 4). This bypassing of the reflector can occur when the direct path is not effectively blocked and allows a certain amount of the signal to reach the receiver ahead of the reflected main beam. This "sneaking" can produce troublesome noise levels requiring corrective engineering. In some instances, it has been

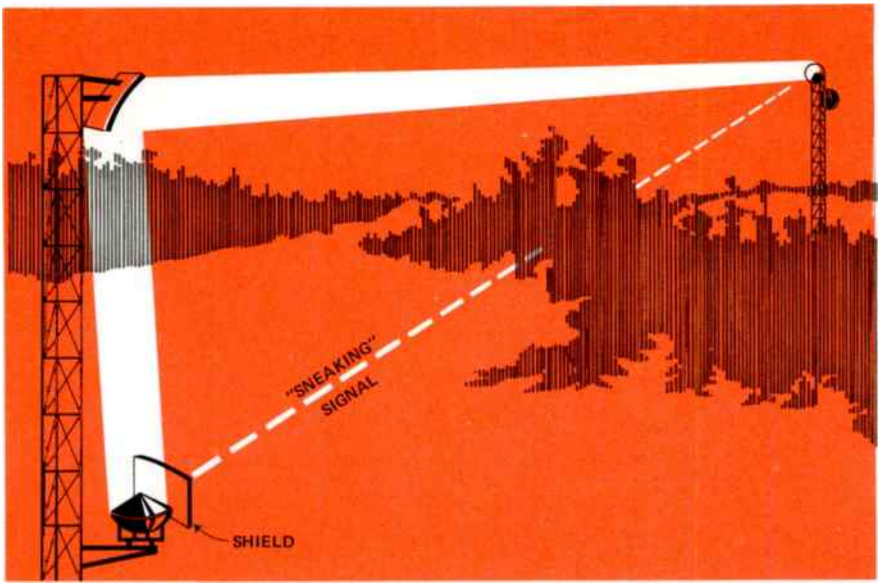


Figure 4. In situations where the path is not effectively obstructed, the signal may sneak from the illuminating dish, directly to the distant receiver. This bypassing of the reflector causes noise at the receiver. Shielding is used to control this problem.

necessary to place metal shields on the path sides of the illuminating dish in much the same manner blinders are used on race horses.

The periscope system has wide usage. Although the efficiency does not change with different frequencies, periscope application can nevertheless be more expensive at the lower frequencies (2 and 4 GHz) because the required dish sizes for these wave lengths are much larger.

One unique advantage of a properly laid-out periscope arrangement is the possible gain improvement over what can be expected from the parabolic antenna alone. This complex matter was clearly described in the July 1963 issue of the DEMODULATOR. By way of recapping, it is sufficient to say that the reflector's size, shape and its distance from the parabolic antenna can make it possible to reflect only first zone energy (Figure 6). When only first zone energy is reflected, the possibility of phase cancellation (caused by simultaneous reflection of the out-of-phase second zone energy) is almost completely eliminated. This arrangement produces sharper beams at distant points while giving net gains of from 2 to 3 dB for reflectors with flat faces.

Additional gain can be achieved by curving the face of the reflector to the approximate shape of a section of a paraboloid with the illuminating dish at its focus. (Figure 5) Actual practice has provided substantial evidence that a properly curved reflector can produce as much as 4 to 6 dB more gain than a flat uncurved reflector in the same application. This gain improvement results from the fact that the phase relationships of the various portions of a reflected beam are determined by the relative points at which the wavefront is intercepted by the reflective surface. It can be shown that because of this curving some of the second zone (out-of-phase) energy can be converted to in-phase energy thereby actually boosting the gain beyond anything possible with a flat reflector.

Passive Repeaters

Erecting an active radio relay station where inaccessibility and severe weather changes can inflate construction and maintenance costs beyond desirable limits is a situation every engineer tries to avoid. This is precisely the kind of problem the engineer can resolve by using a passive repeater.

The two general types of passive repeaters in common use are shown in Figure 2. One consists of two parabolic antennas connected back-to-back through a short length of waveguide. Because the size requirements and the associated cost, this type of passive is rarely used except for very short paths where small dishes are sufficient. The efficiency of this arrangement is approximately 30% compared to a 98% efficiency rating for the "billboard".

"Billboard" passives range in size from 4' x 6' single panels to 40' x 60' connected panels. The reflective surfaces are generally made of aluminum

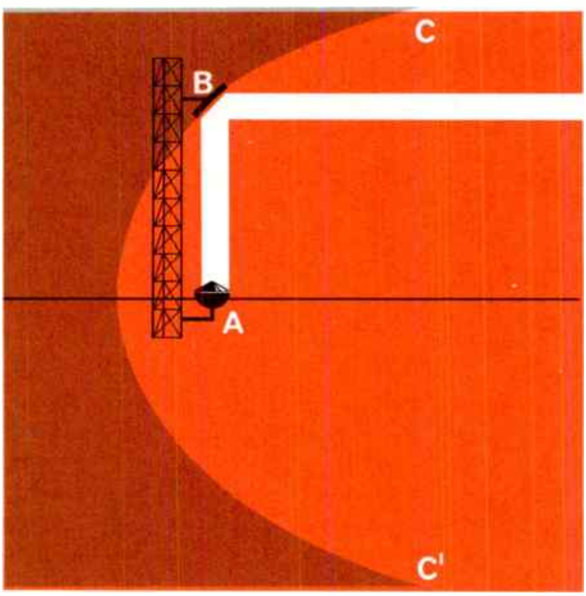


Figure 5. To calculate the curve of a periscope reflector (B), it is convenient to consider the illuminating dish (A) as the focal point of the parabola represented by CC¹. The reflector (B) may simply be considered a solid segment of the imaginary parabolic shell.

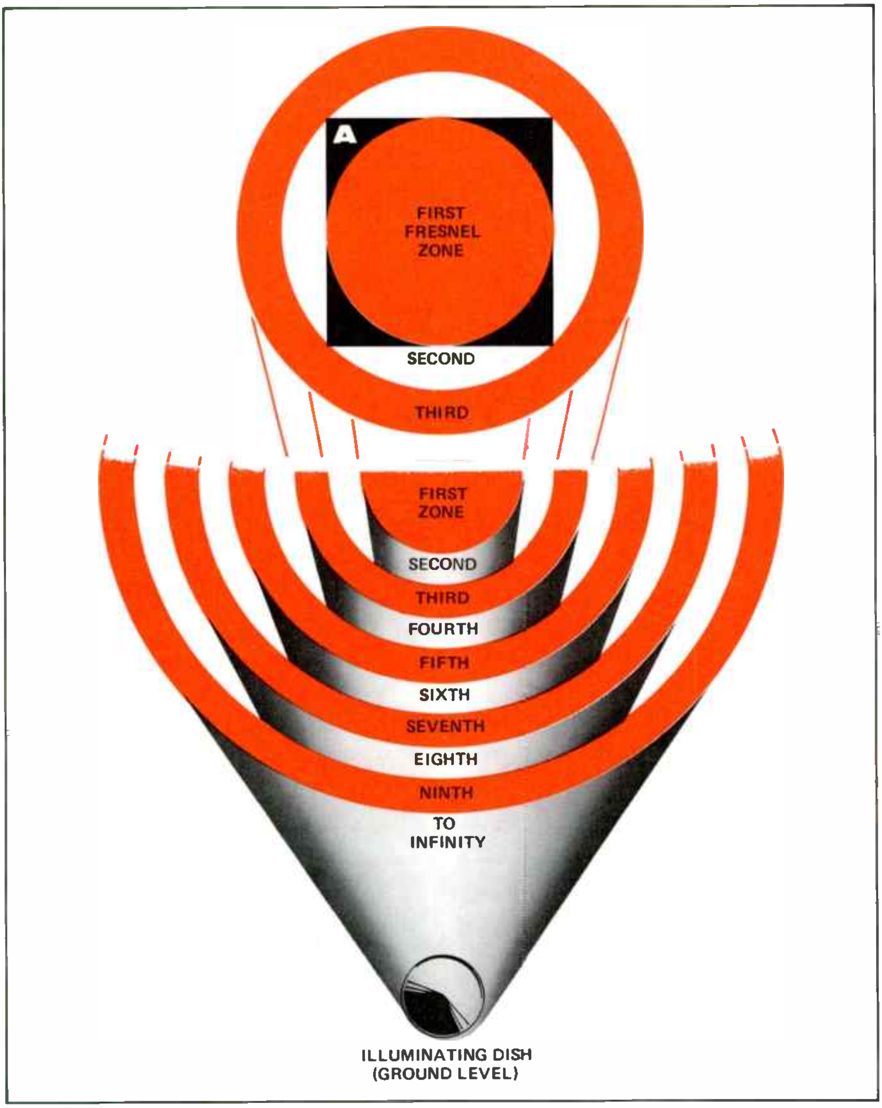


Figure 6. Microwave signals are transmitted in concentric bands of energy (fresnel zones). Each zone is 180° out-of-phase with its adjacent zone. All even numbered zones are in phase with each other and out-of-phase with all odd numbered zones. The 2nd zone energy which is picked up by a flat reflector (A) will tend to cancel it equivalent in first zone energy. By using a curved periscope reflector which extends into the 2nd energy zone it is possible to actually convert the out-of-phase 2nd zone energy to an in-phase relationship with the first zone improving the overall gain.

which has been treated to prevent corrosion. As a rule of thumb, face flatness should be within 1/8 the transmitted wavelength. It has been determined that the reflective surface of the passive must be flat to within 1/8" for 11-GHz transmissions, 1/4" for 6 GHz, and 3/4" for 2 GHz.

Once the existence of an obstruction makes it fairly certain that a passive must be used, it is then necessary to calculate the most efficient site available. The efficiency of any microwave path arrangement using a passive repeater has an inverse relationship to the product of the path distances.

Because of this it is obvious that any arrangement which reduces this product will improve the overall signal strength. It logically follows that those arrangements which place the passive nearest either of the path ends are therefore the most desirable. An additional benefit to this kind of site locating is the fact that the required surface area of the passive decreases as the distance to the path-end is shortened.

Some mention should be made of the fact that while certain topographical conditions appear to be well suited for passive repeater sites they may actually not be desirable at all. This has sometimes proved to be the case in heavily timbered areas immediately surrounding sites of small passive repeaters. Depending on their relationship to the passive and the signal beam, trees can produce serious interpath noise problems. This is also the situation which is occasionally created by unwittingly placing a small passive in front of a rock wall or bluff. For these reasons it is advisable to determine passive sites only after acquiring a thorough awareness of the particular terrain involved. Once the most realistic sites have been selected it is then possible to estimate their relative efficiencies.

Fields — Near and Far

One practical approach to determining antenna-reflector efficiency involves the calculated value of 1/K:

$$1/K = \frac{\pi \lambda}{4 a^2} d^1$$

Where:

λ = *the wavelength in feet*
d^1 = *the path length in feet*
a = *the effective area of the passive repeater*

When the value of 1/K is 2.5 or less, a near field condition exists. Once having determined a near-field condition it is then possible to decide the proper method to use in calculating the gain or loss of the proposed path arrangement.

If the passive is found to be in the far field, its gains and those of the end antennas are independent and the two-way gain of the passive repeater can be calculated by the following formula:

$$\text{Gain in dB} = 20 \log \frac{4 \pi A}{\lambda^2} \cos \alpha$$

Where:

α = *1/2 the horizontal included angle*
λ = *wavelength in feet*
A = *area in square feet*

To find the net loss between the two end points, it is only necessary to calculate the two path attenuations, add them together, then subtract from the result the two-way gain of the passive and the gains of the end antennas.

However, if the passive is found to be in the near field of either antenna then antenna and reflector gains are no longer independent but react with each other in such a way that the net gain would be reduced. In this case the above methods cannot be used, since they give overly optimistic results.

One way to evaluate gain where the passive is in the near-field is to consider the antenna and the nearby passive as a periscope antenna system. In this case a correction factor is calculated and applied to the gain of the antenna to obtain the net gain of the periscope combination. In these situations, the "path" is only taken to be the distance from the periscope reflector (passive in this case) to the far end – the distance between the antenna and reflector within this periscope arrangement is disregarded.

Double Passive Repeaters

If the passive repeater location is behind or off to one side of the near end path, so that the included angle between the two paths at the repeater does not exceed about 120°, a single billboard reflector is most efficient. One reason the angle of the passive to the path should not exceed 120° is that the surface dimension requirements increase unrealistically beyond this angle. If, however, the passive location is more or less along the line between the two end points, it is possible to use a double passive installation. (Figure 7).

Such an arrangement, in which two closely spaced billboards are so situa-

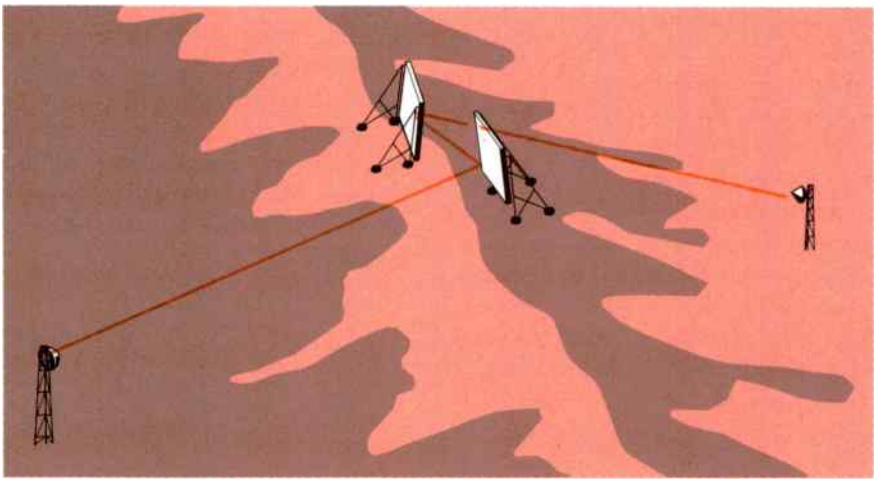


Figure 7. Double passive used to beam signal over a ridge.

Figure 8. An example of a "billboard" passive living up to its name.

ted, can provide the desired beam displacement with only slightly less gain than a single reflector – it is also true that twice as much billboard surface is required.

Reflections

The basic fact that microwave radio transmission is line-of-sight has imposed rather restrictive limitations on the methods microwave engineers can use in getting signals from one place to another. These limitations, like so many others, only serve to stimulate deeper investigations and more imaginative solutions to the problems which arise.

Although the use of radio mirrors in microwave path engineering is not a new development, it is indeed an excellent example of how imagination has provided a simple solution to a complex problem – both in terms of cost and performance. When viewed as simple components, these radio reflectors may be considered the only tool at the engineer's disposal whose efficiency approaches 100%.

It is important to point out that there is a wide distinction between reflective efficiency and overall path efficiency. Without exception, the reflective efficiency of periscope reflectors and passive repeaters is very close to 100%. Path efficiency is, however, a rather complex matter to determine and requires calculating whether or not a passive is in the near field or the far field. Additional figuring is then required to weigh the two-way antenna and passive repeater gains against the total path attenuation. To complicate matters further, in the case of periscope arrangements, it is customary to think of the reflector as simply an extension of the parabolic antenna and not a reflector as such.

These various approaches which experience has shown to be quite reliable, make it somewhat difficult to assess with a blanket statement the path efficiency of reflectors and passive repeaters, because it varies considerably with each application. It can be said without reservation, that the development of reflectors and passive repeaters has greatly increased the number of path engineering alternatives while measurably reducing installation and maintenance costs.

MARCH, 1966
VOL. 15, NO. 3

the Lenkurt. Demodulator

Community Antenna TV

Master antenna systems are serving an increasing number of American communities, providing television reception where before there was none. By extending the viewing area of both commercial and educational television stations, these small networks of coaxial cable and microwave radio have collectively established a new and sizeable industry.

This article discusses the operation of these systems, and related items of interest in the field of educational television.

The establishment of the television industry in the United States brought with it new concepts of entertainment, news, and education, adding immediacy and depth. At the moment there are over 700 television stations in this country (Figure 2), and over 100 construction permits have been granted by the Federal Communications Commission to establish new stations.

Simultaneously with the growth of television — and almost unnoticed during its early years — cable TV systems appeared, stretching TV coverage into otherwise poor signal areas. CATV, for Community Antenna Television, is now a sizeable industry of its own serving close to two million U.S. homes. Similar systems also are developing in other countries, including Canada, Mexico and Great Britain.

A Beginning

CATV is essentially a master antenna service for receiving television signals and distributing them to home receivers. When the first television stations went on the air in the late 1940's, it was found that signals in outlying areas were not always powerful enough for satisfactory reception. Potential TV viewers were either too far from the broadcasting station, or were in a shadow area behind a nearby mountain or other obstruction. Even the construction of costly roof-top antennas was not always successful.

The first meager steps toward the new CATV industry were made by local citizens joining forces to construct master antennas on nearby hilltops. The signal was carried down the hill by standard TV lead-in wire strung from tree top to fence post to pole, and interrupted regularly with unsophisticated booster amplifiers. The results were not always ideal. The first commercial in-

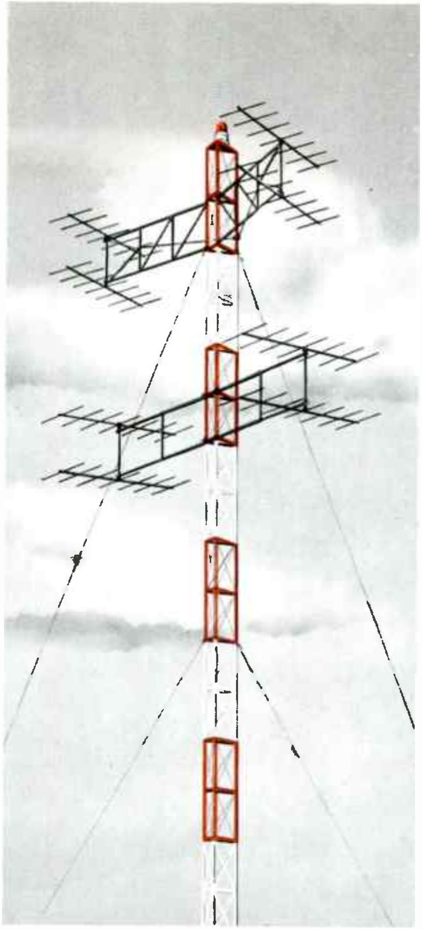


Figure 1. High gain Yagi antennas, one for each television station, receive off-the-air signals to be relayed through CATV system.

stallations, with better antennas and coaxial cable came in 1950. Soon equipment manufacturers were entering the field with specially designed CATV receivers, cable amplifiers and other components. When greater distances separated the TV station from the community, microwave radio replaced long coaxial cables as the most economical method of insuring good TV reception.

CATV Today

Today, approximately 1600 CATV systems are in operation, another thousand have the go-ahead from local franchising agencies, and over two thousand are pending approval. However, it should be understood that not all of these will be built immediately or even in the near future. Some predict about 100 new systems for 1966, increasing to a total of about 2600 by the end of 1970. Only one state does not have at least one operating system, but even there applications for service have been filed. Some larger states have hundreds of independent systems.

	VHF	UHF	TOTAL
COMMERCIAL	486	101	587
EDUCATIONAL	66	49	115
TOTAL	552	150	702

Figure 2. Numerical breakdown of TV stations now on the air.

CATV systems vary greatly in size and capability, from small operations carrying as few as two channels, to advanced and more elaborately equipped facilities bringing as many as 12 TV channels and a number of FM radio signals to the subscriber. The average subscriber receives five stations, while less than one percent get 10 or more channels. Four percent receive only two channels. Soon equipment advances may make it possible for a CATV system to carry 20 or more channels. Small operations may have only a hundred or so subscribers, while the largest in the United States serves nearly 20,000.

In its early years, CATV served the small population centers scattered some distance from television stations. More recently, however, CATV has been brought to the doorsteps and even into the parlors of major cities like New York, Los Angeles, and San Francisco. Tall buildings, natural obstructions, airplanes, and other such factors will degrade television signals from even nearby stations. The advent of color television also has increased the need for high-grade signals for satisfactory picture reproduction.

In addition to providing improved TV reception, CATV frequently includes bonus services placed on otherwise unused channels. An example is weather information from a camera continuously scanning temperature, wind, and other gauges. Another service allows home viewers to read the latest news as it is typed on news-wire machines.

At least one operator has gone further than that, setting up television-like studios to provide news, discussions, speeches, children's programs, and even live sports events. Equipment includes mobile units, video tape recorders, and professional studio consoles. Commercial background music also may be supplied by CATV, using already installed cables to carry recorded music to business concerns.

The CATV Signal

The first need of a CATV system — like the home receiver — is a good signal from the broadcast station. In some countries, broadcasters will provide a direct program feed from the station. But American CATV operators pick up television signals "off the air" with specialized receiving equipment. High gain antennas, typically of Yagi design, are situated on a mountain peak or other advantageous point in the terrain. These antennas are selective, narrow band devices, most efficient at only one frequency or channel. Therefore, a separate antenna is usually installed for each channel to be received. Ideally, TV signals at the antenna site should have a minimum strength of 50 microvolts per meter.

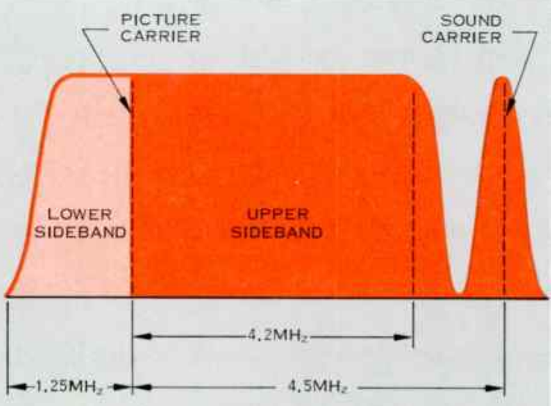
Special receivers detect the TV signals, convert UHF to VHF if necessary, and amplify them to suitable levels for transmission. This portion of the CATV system is known as the "head end" equipment. If the signals are to be fed directly into a cable trunk line, standard VHF television frequencies are used. However, if a local VHF station is carried on the cable, interference will usually result between the direct signal from the transmitter and the signal on the cable. In such cases it is necessary to translate this station's programs to a different channel prior to distribution.

When distant TV stations are to be carried, it is often more practical to use microwave radio links over sometimes as much as hundreds of miles to reach distribution trunk cables. For the head end equipment to feed a microwave system, incoming TV signals first must be demodulated to more usable frequencies. That is, the carrier frequency must be removed, leaving only pure video information in the range from 10 Hz to 4.2 MHz (Figure 3) with the audio on a subcarrier at 4.5 MHz. This is called a *composite signal,* and may be used to directly modulate the microwave radio. It is also possible to separate the video and audio signals at this point and place the sound signal on a higher frequency program channel.

The output of the microwave radio is transmitted by highly directive parabolic antennas to receiving stations 20 or 30 miles away. Then the signal may be retransmitted to another repeater, or fed into more head end equipment for cable distribution.

The Federal Communications Commission regulates all radio frequency allocations, and recently created a new Community Antenna Relay Service (CARS) for exclusive use by all CATV operations. The band is from 12700 to 12950 MHz.

Figure 3. Television spectrum. Upper sideband of video, plus audio carrier make up composite signal.



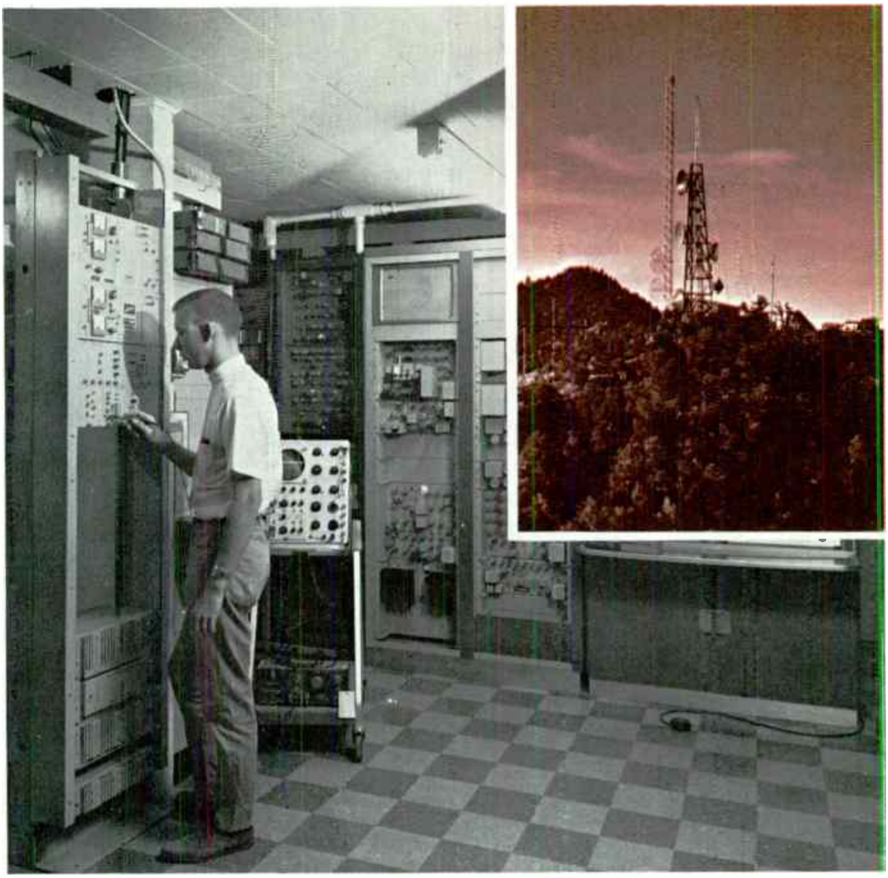


Figure 4. Lenkurt 76TV microwave radio in typical installation. Repeater stations (inset) receive and retransmit signals.

Approximately 25 percent of the CATV systems in the United States use microwave radio, the typical system requiring two or three hops to bring the signal to the cable distribution point.

Picture Distortion

The transmission of television, especially color television, by microwave includes a number of critical problems. Distortion, poor frequency response, and other transmission irregularities all tend to degrade the quality of the final picture image. Of particular importance is the extreme sensitivity of television signals to non-linear phase shift. Ideally, the entire system should be free of non-linear phase shift from almost zero frequency to at least 4.5 MHz (to cover the bandwidth of a video signal). In practice, this is difficult if not impossible to achieve. Components in the

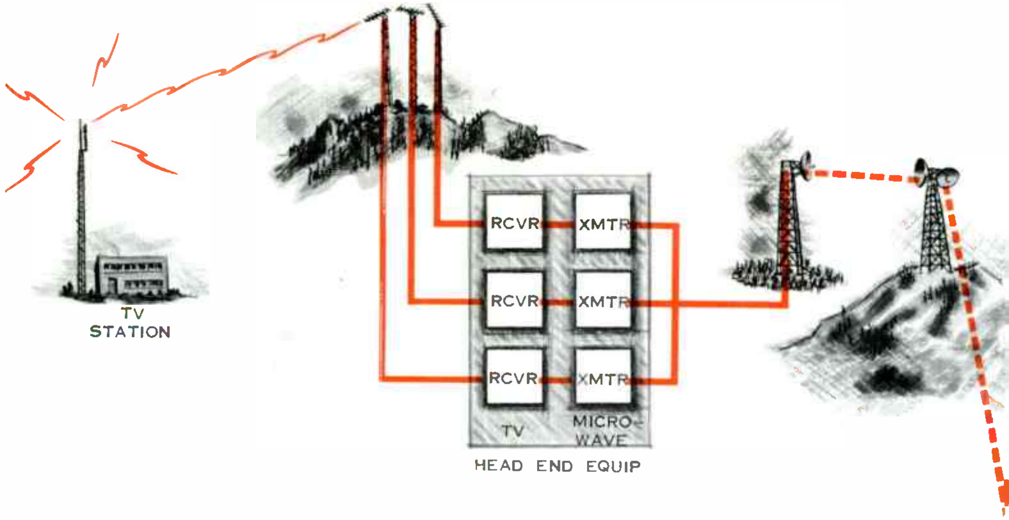


Figure 5. CATV system picks up signals outside the reach of home antennas; may use microwave radio to relay them to cable distribution point. Utility poles carry coaxial cables through residential areas.

system delay some frequencies more than others, distorting the waveform. Although delay distortion of speech or music is not readily detected by the ear, similar distortion of a television signal is very noticeable and grossly affects the quality of reproduction.

Color television is particularly vulnerable to *differential phase* and *differential gain*. The color appearing on the screen is determined by the exact phase relationship between two signals, the color burst and the color subcarrier. An unintentional shift in phase results in a different hue of color. Similarly, change in amplitude of the signal determines the saturation or *richness* of the color. (For additional discussions of these areas see the *Demodulator*, February, 1962; October, 1963; November, 1963; January, 1965).

Delay distortion is directly related to the bandpass characteristics of the entire transmission system, including head end equipment, microwave links, and cable facilities. System design must provide for a very wide bandwidth free from irregularities well beyond the actual frequency limits of the television signal itself. For example, the Lenkurt 76TV microwave system (Figure 4), designed specifically for television transmission, has a frequency response of ±0.5 dB from 20 Hz to 5.5 MHz.

Subjective testing has shown that phasing errors of 5° or more will be detected by the viewer as a change in hue. Likewise, he will find a 2 dB change in color saturation objectionable. In the 76TV, differential phase is less than 0.5° per terminal, while differential gain is held to 0.2 dB at up to 90 percent of applied picture loading.

Ultimately, the signals must be fed into the cable trunk line for distribution to home TV sets. In a system not using microwave, this occurs immediately after the signals are received by the master antenna. With microwave, more head end equipment is found at the final radio hop. Here, signals must be brought to proper levels, remodulated to VHF frequencies, combined, and fed into the main trunk lines (Figure 5).

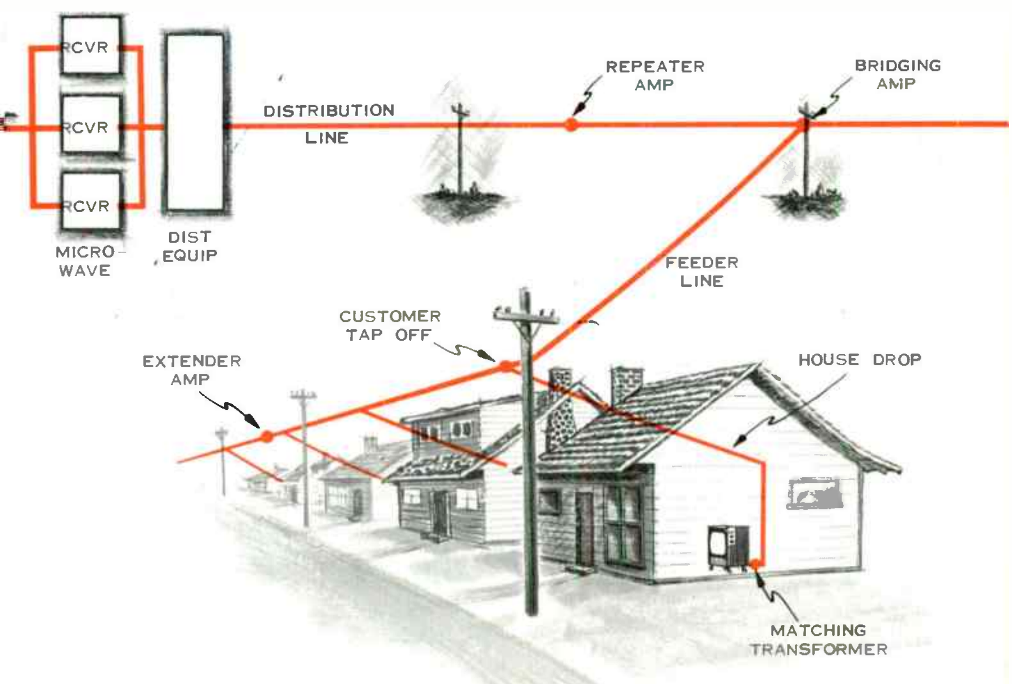
Cable System

The trunk line is the basic carrier of the CATV system and is never tapped to feed individual subscribers. At the intervals along the trunk line are a number of repeater amplifiers (Figure 6) to compensate for signal loss. These are usually less than a mile apart. Bridging amplifiers divert the signals onto feeder, or distribution lines.

Customer "tapoff" units (Figure 7) are placed along the feeder cables. These cause a slight disturbance on the line and therefore a limited number — usually 30 to 40 — are allowed on one line. Extender amplifiers, spaced about every 600 to 700 feet, are used to boost the signal along the feeder line. From the tapoffs come the house drops leading to the subscriber's TV set. However, before a connection can be made, the cable impedance of 75 ohms must be matched to the 300 ohm input impedance of the set through a matching transformer, placed on or near the back of the set.

Troposcatter

CATV operators in other countries have added their own variations to the methods of signal transmission. In Canada, for example, military-developed techniques of troposcatter are being used in some systems spanning rugged terrain. Dependent on the ability of the troposphere to diffuse or scatter a por-



tion of a high frequency signal well beyond the horizon, tropo systems send VHF television skipping over distances from 100 to 500 miles (Figure 8). Stationary tropo antennas may include a tower-supported wire mesh reflector 50 to 60 feet high, almost 300 feet wide, stretching in a parabolic curve around the antenna fixture mounted on the head end building. These antennas are highly directional, and have good ability to reject co-channel and adjacent channel interference.

British CATV

In Great Britain CATV is called "wired broadcast" or "communal aerial system", and uses two different methods of signal transmission. One system is essentially the same as that used in the United States, relaying signals at standard VHF TV frequencies directly to the receiver. Another popular technique is an outgrowth of the older "wired radio" system. This radio relay system was basically a public address system supplying audio directly to speakers in the home. The television version transmits unmodulated video signals (3-10 MHz) over twisted wire pairs to TV receivers built without the customary r-f front-end amplifiers. Economy is the prime justification for the technique — both the home receivers and the transmission wire are less expensive.

In the British system, head end equipment supplies approximately 40 watts of video power to the trunk lines, which may be up to 6000 yards long. Feeder lines may branch off the trunk lines for distances up to about 2000 yards. It has been found that four video signals with their accompanying audio, and four additional radio channels may be carried on two twisted pair (four wires) in a shielded cable.

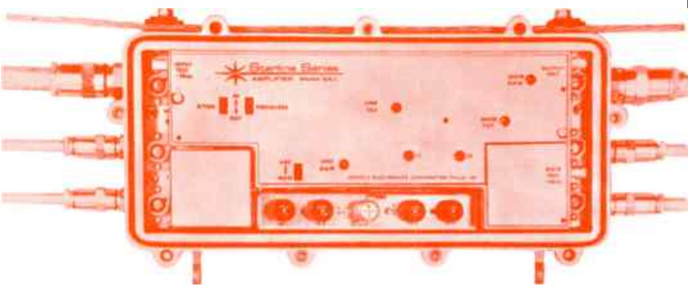


Figure 6. CATV signals pass through trunk-line and extender amplifiers before reaching the home receiver.

Educational TV

Sharing some of the problems, and related in many ways to CATV are the three overlapping areas of educational television (ETV), instructional television (ITV), and closed circuit television (CCTV).

ETV is generally meant to include non-commercial broadcast stations, both VHF and UHF. ITV refers to program content rather than facilities, and relates directly to formal education. CCTV describes the transmission of television by cable or microwave to a predetermined audience, as opposed to public broadcast. Additionally, our reference here is primarily to the use of CCTV in education.

There are four general types of licensees operating educational television stations: (1) universities, (2) public school systems, (3) statewide ETV commissions, and (4) non-profit "community" corporations. More than half the ETV stations in the country fall into the first two categories, being directly responsible to educational institutions. Likewise, a statewide commission's prime interest is usually with the school systems of the state. And while the community stations may have no direct connection with schools, they usually carry a regular schedule of instructional programs.

The average ETV station broadcasts 5 or 6 days a week, 10 to 11 hours a day. Programs are divided almost equally between classroom instruction and more general programming planned for home viewing by all age groups. Instructional material more likely will be seen during the normal school hours, with more general programs in the early evening, and informative discussions or entertainment features for adult viewing in the late evening.

At this time there are 115 ETV stations on the air, with another 65 under construction or with applications pending. Currently more than half of the ETV stations are on VHF frequencies (channels 2-13), but most reserved allocations for the future are in the UHF band (channels 14-83).

In the School

Closed circuit television is utilized by many schools to make more advantageous use of teachers and instructional material. There are about 800 CCTV installations in this country, split almost equally between elementary and secondary schools, and colleges and universities. These may operate within one school, delivering lectures or demonstrations to other buildings, or between various schools in a district. Within a single school, coaxial cable easily connects the cameras and studio equipment

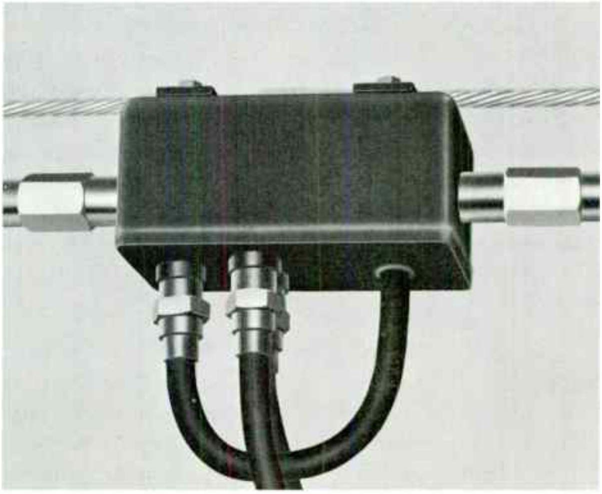


Figure 7. Tapoff units connect house drops to feeder lines, and are designed to prevent interfering signals from reentering cable.

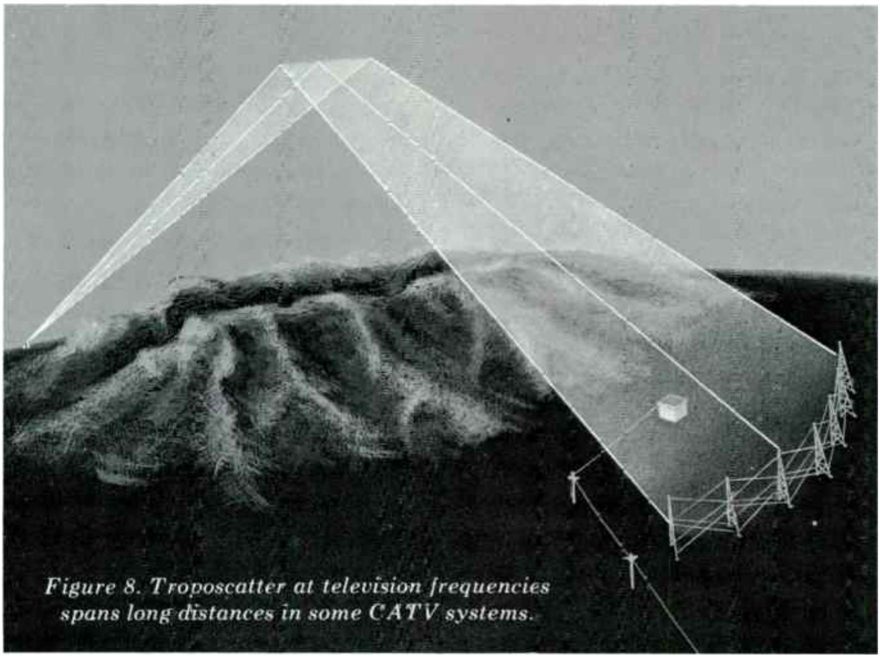


Figure 8. Troposcatter at television frequencies spans long distances in some CATV systems.

to other viewing locations, sometimes with three or four programs on a single cable. Longer runs between separate schools can also be practical, using telephone company cables on a leased basis.

An increasing trend in ITV is the use of microwave radio. The FCC has allocated 31 channels in the 2500 to 2690 MHz band for use by educational institutions. Some districts use these channels as direct links between two schools. Others operate a central transmitter beaming programs in several directions at once, much like a standard broadcast station, to be received off the air at various schools within the district.

ITV systems also operate point-to-point microwave relays on two higher frequency bands. The primary allocation is in the 12200 to 12700 MHz band. However, the FCC will consider applications on a case-by-case basis for the 6575 to 6875 MHz band when the operator can show that it is not technically feasible to use the higher frequency.

Frequently CATV systems will carry ETV programs, thereby greatly extending the range of the station. These may even be piped into the schools, hospitals, or other such facilities in distant towns for little or no charge. In many cases CATV operators also will allow two ETV stations to share programming over a spare microwave channel.

Networks

Many states have already installed widespread microwave networks connecting educational institutions hundreds of miles apart. Similarly, moves have been made to connect large numbers of ETV broadcast stations into educational networks. And many CATV systems are beginning to resemble small networks. It is possible that someday a combination of these efforts will bring to this country a "fourth" major television network joining the best educational and cultural programs in all parts of the nation. Moreover, a fifth network, with commercial UHF stations linked from one end of the country to the other, is being considered.

There are also other possibilities for bringing educational and cultural programs to larger audiences. One quite successful experiment has been undertaken by Purdue University, transmitting previously video taped programs from specially equipped airplanes circling 23,000 feet over Indiana. Daily, over a half-million students in six states (a total of 127,000 square miles) receive courses ranging from elementary to college-level subjects.

From the beginning, telephone companies have been involved with CATV systems, allowing cables to be strung on their utility poles. More recently, telephone companies have supplied cable transmission channels for CATV systems and instructional TV operations on a lease or tariff basis. And now many operating companies are expressing interest in becoming CATV operators themselves.

The Future

In the next few years both CATV and educational television undoubtedly will experience many changes. Advancing techniques will allow for greater numbers of channels to be carried over microwave and cable facilities, bringing even more programs into homes and schools across the nation. Satellite technology may add a new dimension with the possibility of broadcasting directly to schools — or even home receivers — anywhere in the nation.

In the United States close to 98 percent of the homes have at least one television set. Three percent of these homes are served by CATV. Educational programming is now available to an estimated 130 million people—another 10 million to be added this year with 14 new ETV stations. In addition, instructional television today reaches two out of three of the nation's 50 million students.

As new networks link one station or relay system to another and new operations spring up, a continually expanding measure of entertainment and education will be easily available to the television public.

The Lenkurt DEMODULATOR

MAY 1969

overloaded microwave systems — part 1

Complicated voice and data loading requires special equations to calculate actual capacity.

Microwave communication systems, like most systems built for use in a constantly expanding consumer market, seem to reach their maximum capacity before they should. Even when extra capacity has been painstakingly engineered into a system, it is not at all uncommon to find that even this additional capacity has been consumed earlier than anticipated.

As an FM–FDM (microwave-multiplex) system expands to its maximum capacity, problems arise as more circuits or services are required. In general, these systems consist of several microwave hops in tandem between the end points of the system, with spur or sideleg hops often branching from the intermediate points.

When is a System Overloaded?

In complex systems it is possible to have portions of the system operating at or near the overload point while the other portions are carrying much lighter loads. In determining an overload, it is only necessary to consider the single most heavily loaded microwave hop.

In an FM system there are several interrelated factors which limit maximum capacity. An overload exists when one or more of the following limits has been exceeded:

1) All of the available or usable baseband spectrum is in use.
2) The point at which total baseband signal power (system loading) if increased would cause unacceptable performance.
3) System usage is such that any increase in either the top baseband frequency or the system loading would cause emission bandwidth to exceed that legally allowed for the particular frequency band.

In FM systems, the first two of these limits often have some degree of elasticity. The third, however, is a legal limitation which cannot be exceeded without legal violation. Perhaps the best approach is to evaluate the nature of the emission, its limitations, a method by which it can be calculated, and how it is affected by various parameters of the microwave system.

Legal Limitations of Capacity

The allowable maximum bandwidth (necessary or occupied, whichever is greater) for microwave systems under the Industrial Radio Services is established in Paragraph 91.111 of the Federal Communications Commission rules. It is:

8 MHz in the 1850-1990 MHz band
800 kHz in the 2130-2150 and 2180-2200 MHz bands
10 MHz in the 6575-6875 MHz band
20 MHz in the 12,200-12,700 MHz band

Paragraph 2.202 of the FCC rules defines the various emission characteristics and provides formulas for calculating the "necessary bandwidth."

The type of service and the allowable bandwidth for a particular service is formalized in an "emission designator," which includes first the bandwidth in kHz, then a letter indicating the type of modulation (F for frequency modulated systems), then a code number indicating the type of transmission (usually "9" for composite transmission in case of FM systems with FDM multiplex). Thus the emission designators for the bands listed above would be 8000F9, 800F9, 10000F9 and 20000F9 respectively.

The formula given by FCC for calculating the necessary bandwidth of an F9 transmission is:

(A) $B_n = 2M + 2DK$

where:

B_n = *necessary bandwidth in kHz*

M = *maximum modulation frequency in kHz*

D = *peak deviation in kHz, defined as half the difference between the maximum and minimum values of instantaneous frequency.*

K = *a numerical factor depending upon the allowable distortion. A commonly used value for K in such systems is 0.9, though a value of 1.0 is sometimes used.*

The value of M for a particular system is easily established.[1] It is simply the frequency of the top modulating channel applied to the baseband. The value of D, however, is somewhat more elusive since the composite load applied to the baseband is a varying and complex quantity whose peak value can only be described statistically. The value of K is, as stated, very close to 1.0.

The multiplex used, except for systems of very low density, is almost exclusively of the single-sideband suppressed-carrier type (SSBSC).

Studies on operating systems have led to the following equations for calculating the rms (root mean square) value of white noise power, simulating the equivalent busy hour load of a given number of voice channels multiplexed into a baseband by SSBSC techniques (Fig. 1).

(B)

$P = (-15 + 10 \log N)$ dBm0

(N is 240 or more)

or:

(C)

$P = (-1 + 4 \log N)$ dBm0

(N is 60 to 240)

where:

P = *equivalent rms white noise power applied over the same baseband spectrum as occupied by the multiplex channels.*

N = *number of voice channels*

dBm0 = *dB with respect to the power of a single channel test tone at zero relative level.*

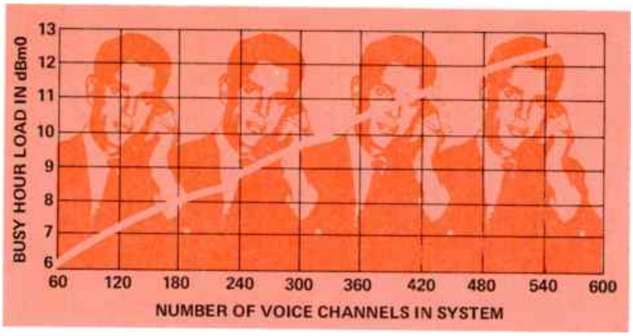
These equations, originated by CCITT and CCIR, are almost universally accepted as a basis for the design and testing of multi-channel microwave systems and provide a basis for calculating peak deviation (Factor D in equation (A)).

Calculating D for Voice Systems

The starting point for the calculation of D (peak deviation) is the

[1] Electronics Industries Association (EIA) has submitted to FCC a proposal that a peak factor of 11.5 dB be used instead of the 13 dB which has been customary, and that a value of 1.0 be used for the factor K for the present. The result of these changes would reduce the calculated values of 2 DK by approximately 8%. This would allow a slight increase in channel capacity for the same necessary bandwidth.
Industry's interpretation is that M should properly be taken as the frequency of the top information-bearing channel in the system, and that a sinusoidal continuity pilot located above the baseband should not be considered to be the "top modulation frequency" and should be excluded from the determination of M.

Figure 1. This graph, based on equations (A) and (B), shows the busy hour load (in dBm0) for the various number of voice channels used in a particular system.



known per-channel rms deviation and the known power of its signal. The per-channel deviation is a basic FM system parameter frequently chosen as 200 kHz rms. The baseband power of the test tone producing this deviation is 0 dBm0 rms.

The parenthetical expressions in (B) and (C), called the "noise loading ratio", express the dB ratio between the rms power of a white noise load whose peaks are equal to the peak values of the complex baseband signal during the busy hour, and the rms power of a test tone.

The peak value of white noise power is a statistical parameter with no specific value, but is commonly taken as 13 dB above the rms power. The use of two different equations for calculating the white noise load equivalent reflects the fact that the peak to rms factor of the complex signal from a number of voice channels is relatively constant at 13 dB for systems with more than 200 channels, but is variable and somewhat higher for systems with fewer channels (Fig. 2).

Deviation in an FM system has the dimension of voltage. Consequently, the effect of changes in deviation can be calculated as a 20 log function of changes in load power.

The following equations can be used to calculate the peak deviation for a multichannel SSBSC voice system:

(D)

$$D = 4.47d\left(\log^{-1} \frac{-15 + 10 \log N}{20}\right)$$

(N is 240 or more)

(E)

$$D = 4.47d\left(\log^{-1} \frac{-1 + 4 \log N}{20}\right)$$

(N is 60 to 240)

where:

D = *peak deviation in kHz*

d = *per-channel test tone deviation in kHz, rms*

N = *number of SSBSC voice channels in system*

$$\text{Peak factor} = \log^{-1} \frac{13}{20} = 4.47$$

Example A:

A 300 channel radio system could typically have a 200 kHz per channel rms deviation.

$$D = (4.47)(200)\left(\log^{-1} \frac{9.77}{20}\right)$$
$$= (4.47)(200)(\log^{-1} .4885)$$
$$= (4.47)(200)(3.08)$$
$$= 2753 \text{ kHz}$$

Equation (A) can be used to calculate B_n, noting that M = 1300 kHz (top channel of a 300 kHz system) and taking 0.9 for K. $B_n = 2 \times 1300 + 2 \times 2753 \times 0.9 = 7555$ kHz

For standard SSBSC multiplex configurations of 120 channels to about

960 channels, the frequency of the top channel in an N-channel system can be very closely approximated as (4.13 N + 60) kHz. By using this approximation for M, taking K as 0.9, and substituting the appropriate values of D from (D) and (E) respectively, the following equations for B_n in terms of N and d can be derived. (It should be emphasized that they apply only to systems used primarily for voice):

(F)

$$B_n = 120 + 8.26N + 1.43d\,N^{0.5}$$

(N is 240 or more)

(G)

$$B_n = 120 + 8.26N + 7.17d\,N^{0.2}$$

(N is 120 to 240)

These equations provide insight into the complicated way the necessary bandwidth varies as a function of the number of channels and per channel deviation in voice operation.

The equations permit calculation of any one of the three variables (B_n, N, and d) provided the other two are known, and can be used to determine what combinations of number of channels and per channel deviation can be used without exceeding a specific value of B_n.

Example B:

A typical microwave system in the 6 GHz industrial band has the limitation of 10000F9 emission. (From Example A, it is clear that there will be no problem with a 300 channel system using 200 kHz per channel deviation.)

But suppose 600 channels are desired in the same bandwidth.

What per channel deviation will allow staying within 10000F9?

By substituting 1000 for B_n and 600 for N in (F) it can be easily calculated that the deviation must be reduced to 140 kHz.

If d is left at 200 kHz per channel, it can be shown that N cannot exceed about 450 channels if B_n is not to exceed 10000 kHz.

Thus seven complete supergroups, or 420 channels, can be accommodated on a system using 200 kHz per channel deviation, within the 10000 kHz bandwidth limitation, but eight supergroups create an overload.

Calculating Voice and Data

Present day systems have a significant percentage of the derived SSBSC channels devoted to the transmission of systems of submultiplexed tones carrying data or telegraph. The number of tones of this type in an SSBSC channel can vary from one to 25 or more. Their power represents a relatively constant rms load to the baseband, since the tones are on continuously. When the total number of individual data signals on the system exceeds about 15, the peak to rms factor for their complex summation approaches that of white noise.

If the levels chosen for each data or telegraph circuit are such that the total rms power of their tones submultiplexed in any SSBSC channel is 15 dBm0, the data loading per SSBSC channel will be the same as if it had been used for voice. In this case these equations can be used to calculate deviation and bandwidth.

The common practice of putting data at a somewhat higher level means the loading due to the number of channels devoted to data will be much greater than if they had been devoted to voice. This also means greater overall loading and deviation.

The necessary calculations for a mixture of voice and data channels are simple in theory. They can become complicated in practice, however, because there are so many possible combinations of voice and non-voice circuits. The following equation is a generalized form of (D) and (E):

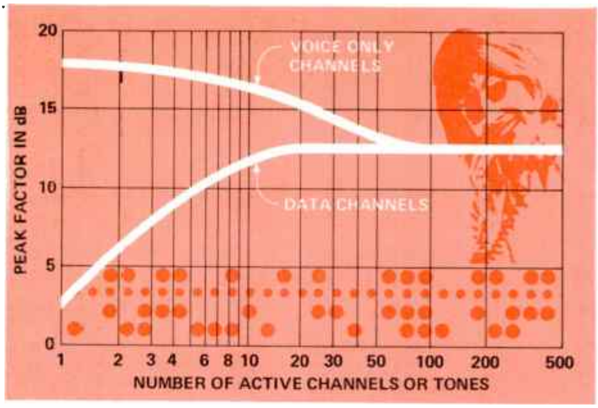


Figure 2. The patterns for the two peak factors – peak to rms ratios – of data and voice are essentially the same when the number of channels is large. However, restriction of data to low levels will affect the signal-to-noise ratios.

(H)

$$D = 4.47\, d\left(\log^{-1} \frac{NLR_{tot}}{20}\right) kHz$$

which leads to a generalized form of (F) and (G):

(I)

$$B_n = 120 + 8.26\, N + 8.05d \cdot \left(\log^{-1} \frac{NLR_{tot}}{20}\right)$$

where:

D, d, B_n, and N are all as before and NLR_{tot} is the Noise Loading Ratio corresponding to the total equivalent voice channel power plus the equivalent power of all non-voice groups.

(Note: N is the total number of SSBSC channels in the system, regardless of use. The function of N is only to establish the top modulating frequency M).

Before (H) and (I) can be used, a preliminary calculation must be made to determine the value of NLR_{tot}. The simplest way is to calculate separately the dBm0 equivalent noise power of the channels used for voice (using (B) or (C)), the equivalent dBm0 noise power of each non-voice group and then on a power summation basis, combine all the powers to obtain the equivalent total baseband load of white noise power. The NLR_{tot} in dB is then numerically equal to the dBm0 value of the equivalent white noise load. Once NLR_{tot} has been calculated, it can be used in (I) to obtain B_n, or with (H) and then (A) to determine both D and B_n.

The following example will illustrate the method.

Example C:

A 6 GHz system with 300 SSBSC channels, of which 200 channels are used for voice transmission, 40 channels are used for data at a power of –10 dBm0 per SSBSC channel, and 60 channels are used to carry submultiplex telegraph tones, each tone at a power level of –21 dBm0 and with each of the 60 SSBSC channels carrying 20 such tones. The per channel rms test tone deviation is 200 kHz. To calculate necessary bandwidth:

1. Calculate noise load power corresponding to 200 voice channels, using (C), as (–1 +4 log 200) = +8.2 dBm0.
2. Calculate noise power corresponding to 40 data channels at –10 dBm0 per channel as (–10 + 10 log 40) = +6.02 dBm0.
3. Calculate noise power corresponding to 20 tones in one SSBSC

channel as (–21 +10 log 20) = –8 dBm0 and the noise power corresponding to 60 such SSBSC channels as (–8 + 10 log 60) = m + 9.78 dBm0.

4. Sum the three noise powers, +8.2 dBm0, +6.02 dBm0, and +9.78 dBm0 on a power basis, by using appropriate curves or by converting each value to its equivalent in milliwatts, adding, and reconverting to dBm0. The power sum will be found to be very close to +13 dBm0 or about 3.2 dB higher than the equivalent noise loading of 300 voice channels.
5. As indicated above, the NLR corresponding to +13 dBm0 of noise power is 13 dB. Substitute this value for NLR_{tot} in (H), which gives the following:

$$\begin{aligned} D &= 4.47 \times 200 \times \left(\log^{-1} \frac{13}{20}\right) \\ &= 4.47 \times 200 \times 4.47 \\ &= 4000 \text{ kHz} \end{aligned}$$

(It is coincidental that the noise loading factor equals the peak factor. Generally, they will be different.)

6. With D known, use (A) to calculate the "necessary bandwidth". The value of M is still 1300 kHz, corresponding to the frequency of the top channel of a 300 channel SSBSC system, and the 0.9 value is still appropriate for K. This gives:

$$\begin{aligned} B_n &= 2 \times 1300 + 2 \times 4000 \times 0.9 \\ &= 9800 \text{ kHz} \end{aligned}$$

or:

Equation (I) could have been used to calculate B_n directly.

The methods used in Example C can be extended to cover other situations provided the basic principles are followed.

To avoid possible confusion, these calculations of peak deviation are based on systems which do not have emphasis and whose per-channel test tone deviations have the same value regardless of the position of the channel in the baseband. When emphasis and deemphasis networks are used, per channel test tone deviation is not a constant but is a function of channel baseband frequency. Higher channels deviate more than lower channels, but systems are generally so arranged that the total deviation remains the same and the equations are still valid.

Capacity Limitations

Microwave equipments are generally designed with some specific maximum capacity in mind, usually in some multiple of the standard 60-channel supergroup. In older systems, and in light route or spur legs, 120 channel and 240 channel systems were often used. Present usage tends toward systems with 300 channel capacity, even higher for backbone routes.

Selecting the proper equipment and employing the most effective field application necessarily requires some specific criterion of channel noise performance. Noise performance is an intricate function of the number of channels, the per channel deviation, the presence or absence of emphasis networks, the per channel loading, the receiver noise figure, the RF signal level, the fade margin needed to give the desired reliability, and the i-f bandwidth — to mention a few. There are many trade-offs and balances involved. The choices made when engineering a system for 300 channels would not be the same as those for 600 channels.

The Lenkurt DEMODULATOR

JUNE 1969

overloaded microwave systems — part 2

Actual capacity often can be extended by several methods short of costly updating

When a microwave system reaches its capacity limitation and additional service is required, there are several practical alternatives which should be explored.

Probably the most fundamental consideration is the nature of the equipment being evaluated; older systems with relatively small channel cross-sections often use older multiplex types such as transmitted carrier or double sideband multiplex – both of which use more bandwidth and load the system more heavily than single-sideband suppressed-carrier (SSBSC) multiplex. In systems using these older multiplex types, the obvious immediate solution is to update with the more efficient SSBSC system.

Other alternatives include sectionalizing the system by using baseband blocking or multiplex interconnects; evaluating the legal and technical possibilities of increasing the number of channels; evaluating the noise performance levels to see if modest relaxing might not permit an increase in capacity; and finally, considering whether it might not be more expedient to simply update the entire system.

In a simple microwave system consisting of only one hop or a few hops in tandem, with most of the required channels going end-to-end over the complete system, the problem is to increase the capacity of the individual hops.

In complex systems, on the other hand, the key to a more efficient use of capacity may lie in sectionalizing the system in such a way as to allow portions of the baseband to be blocked off at intervals and reused in different sections. This principle can be demonstrated by the following hypothetical example.

<u>Example A:</u>

An in-line microwave system (no spurs) connects the imaginary towns of (A), (B), (C), (D); channels are required only between these points, so intermediate repeater stations, if required, will not affect the situation. All microwave hops are designed for 300 channel capacity, and the full baseband is available at all four points. The system is operating at full capacity, with 60 channels in supergroup 1 between (A) and (B), 60 channels in SG 2 between (B) and (C), 60 channels in SG 3 between (C) and (D), and 120 channels in SG 4-5 end-to-end between (A) and (D).

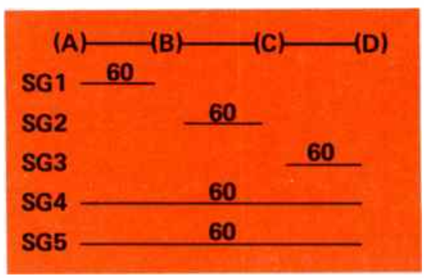


Figure 1.

In this system, a channel or group of channels once used anywhere in the system appear in the baseband throughout the whole system and cannot be used elsewhere. Consequently, this system is at full design capacity.

Sectionalizing With Filters

The system outlined in Figure 1 can be changed as follows: At (B) and at (C) the baseband can be sectionalized by putting in appropriate sets of high-pass/low-pass filters which divide the baseband so that supergroup 1-2 pass through the low-pass side and SG 3-4-5 through the high-pass side. At each of these stations there will be two sets of filters, one looking east and one looking west. The high-pass sides will be cross-connected through the station so that SG 3-4-5 pass through end-to-end. But insofar as SG 1-2 are concerned, the system is now broken into three independent sections, and the full 120 channels of SG 1-2 can be used in each of them creating the following situation:

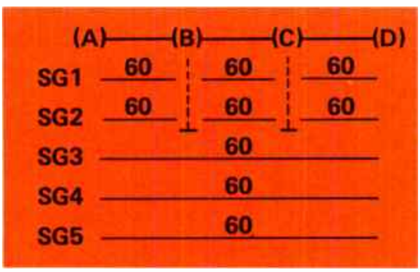


Figure 2.

This relatively simple change produces a microwave system capable of providing 120 channels from (A) to (B), 120 from (B) to (C), 120 from (C) to (D), and 180 from (A) to (D) for a total of 540 channels, yet no portion of the microwave system is carrying more than 300 channels. This has required additional multiplex equipment and some slight rearranging of original equipment. Figure 3 shows the filter arrangement used at the two intermediate stations. The cross-over point for the filters used in this particular application lies in the slot between SG 2 and SG 3, so that SG 1 and 2 pass through the LP side but are blocked from the HP side, while the reverse is true with SG 3 and higher supergroups.

Filters can be arranged either to pass the high groups through the station and drop the low groups, as shown in Figure 3, or the reverse. (The filters shown handle only one direction of transmission. Another identical set is needed for the opposite direction.)

There are many other possible filter arrangements. For example, filters are also available to split the baseband between SG 1 and SG 2, and others to split between SG 3 and 4, SG 3 and 5. Combinations of filters can be used to separate and drop an intermediate supergroup, while passing through the supergroups above and below it. Figure 4 is an example of such a complex filter arrangement. It could be used, for example, to pass SG 3, 4, 5 (and higher if necessary) straight through the station via the HP side of the upper pair of filters; SG 1 passes through via the LP (Low Pass) side of the upper pair and the LP side of the lower pair, while SG 2 is dropped and inserted via the HP (High Pass) side of the lower pair.

Figure 4 also shows another feature which may be needed – a pilot bypass arrangement. This might be required where a pilot must pass through the station which happens to be in the blocked section. The bypass equipment then is used to pick off this pilot and reinsert it on the other side.

Filter arrangements as shown in Figure 4 are relatively inexpensive and simple to apply, but they have some limitations. One is they are somewhat inflexible and difficult to modify or

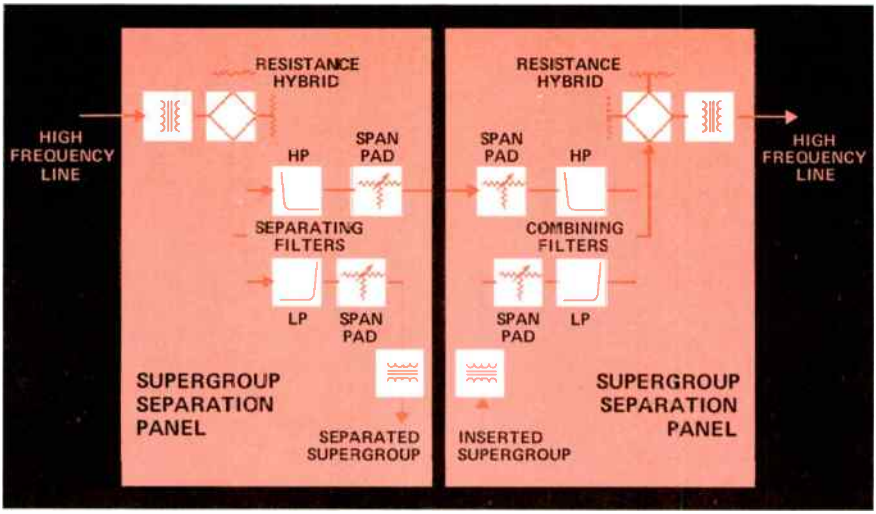


Figure 3. An example of two supergroup separation panels arranged for dropping and inserting a supergroup.

change without taking the system out of service. Another is the fact that separations between the higher supergroups (SG 4 and above) generally involve the loss of a few of the carrier channels in the vicinity of the crossover point. Further, if a number of such filters are used in tandem in a long system, there may be enough degradation in the response of the through paths to affect the end-to-end channels.

Sectionalizing With Carrier Interconnects

The filter type of sectionalizing is done in the microwave baseband. An alternative method of sectionalizing is available which provides maximum flexibility and does not have the limitations of the filter method, though it is somewhat more expensive.

When this alternative method is used, there are no through baseband connections at the sectionalizing station. Instead, each incoming microwave leg is completely terminated in a carrier terminal. Blocks of through channels are passed through the station by means of supergroup interconnects (60 channels) or group interconnects (12 channels) without demodulation. Those groups or supergroups destined for local drop are of course provided with channel modem equipment.

This carrier type of sectionalizing is almost universally used today in the telephone industry. In industrial systems of relatively high density there are often situations in which it is desirable to use carrier interconnect sectionalizing at some of the intermediate points. It is particularly advantageous, for example, at a junction station where several routes converge, with substantial numbers of channels terminating locally, but some blocks of channels needing to pass through the station in various ways. Setting up such a station on a carrier interconnect basis allows efficient use of the capacities of all the microwave branches, and perhaps even more important, allows great flexibility in any rearrangements which may develop as a result of changed requirements. Such rearrangements can be done without affecting service on anything except the particular blocks of channels being rerouted.

Figure 5 shows a simplified example of a three-way junction station arranged with carrier interconnects. Circuits passing through the station as shown are: SG 1 West to SG 2 North; SG 2 West to SG 2 East; SG 1 North to SG 3 East.

Thus this junction station could have 180 circuits to the west, 180 to the north, and 180 to the east, plus three supergroups passing through the station, giving a total of 720 channels.

The great flexibility available for rearrangements or for future additions is quite apparent. Although only supergroup interconnections are shown in Figure 5, it is also possible to make group interconnects in blocks of 12 channels, providing an added degree of flexibility.

Both types of sectionalizing have their advantages and disadvantages. In

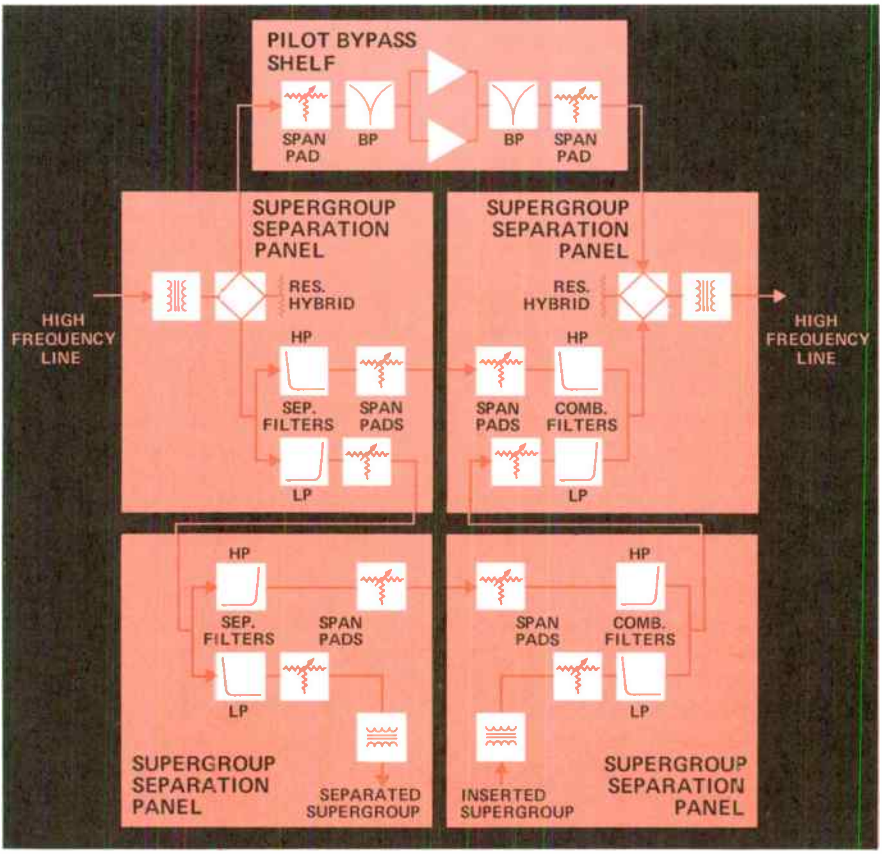


Figure 4. Four supergroup separation panels are shown with pilot bypass shelf arranged for dropping and inserting an intermediate supergroup.

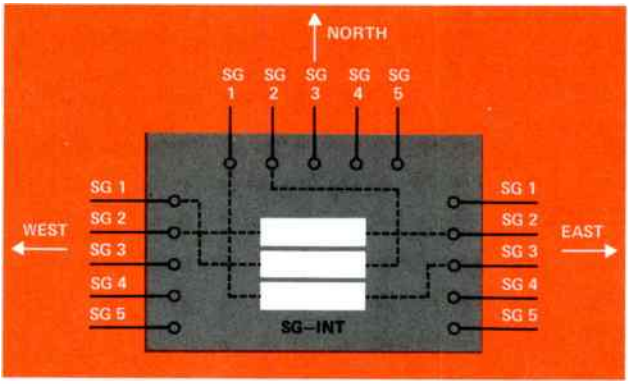


Figure 5. An example of a junction station with supergroup interconnects.

long microwave systems it is often desirable to use carrier interconnections to sectionalize at points of high channel density, and filter interconnects to sectionalize at points of lower channel density. Almost every situation has its own special characteristics, and only by studying the actual situation in light of the overall requirements can a decision be made as to the best method to use in a given station.

Noise Performance Criteria

A relaxation of a few dB in the requirements for channel noise performance can, under certain circumstances, permit a substantial increase in a microwave system's total channel capacity. This same circumstance can often make possible a more efficient use of the existing communications system.

The key lies in properly evaluating the true requirements for noise performance in relation to channel arrangements. Industrial microwave systems tend to perform essentially the same function as the public telephone networks. As a result they are designed to about the same standards of noise performance as long-haul telephone systems. For example, a common objective is 32 dBa0 or better in the worst channel for a system of 1,000 miles. There are very sound reasons for establishing such an objective for the initial system design. But in *using* the system, the user may very well recognize that a noise performance of 35 dBa0 would still provide an extremely good circuit which would be adequate to his needs. It should be remembered that the switching hierarchy in the public telephone network is such that up to eight or more trunks in tandem may occur on a given call, and the intertoll objectives thus must be very stringent indeed. Private microwave systems, even when they are very long, are much less complex and are more easily controlled by the user.

When it is not desirable to relax the end-to-end noise performance requirements in a system, there is still another useful possibility. This is to reserve the lower noise portions of the microwave baseband for the long-haul circuits, and use the higher-noise portions for short-haul circuits. This is in line with practices on the public networks, where considerably lower objectives are applied (on a per mile basis) for short-haul, toll-connecting and direct trunks than are applied to the intertoll trunks.

In an imaginary example, expansion of a 300 channel system to 420 channels could be accomplished by

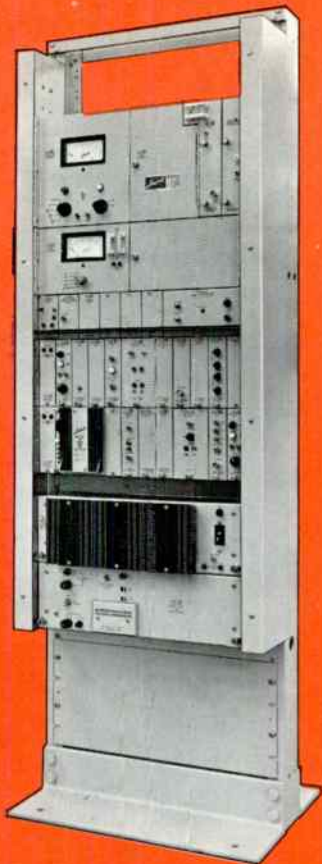


Figure 6. Lenkurt's 78 Microwave Radio System has been specifically designed with the wide flexibility needed to meet the rapid growth rates of modern message and data communications.

adding SG 6 and 7. If deviation were left unchanged the noise in the original 300 channels could be expected to increase by about 1.5 to 2 dB, but the noise in the top channel of SG 7 would be perhaps 5 dB poorer than the worst original channel. If the original system were a 1,000 mile system with end-to-end performance of 32 dBa0, and if the new SG 6 and 7 channels were used on relatively short sections of the system (up to 250 miles), the new channels would be around 31 dBa0 (6 dB improvement because they would traverse only 1/4 of the total system length, 5 dB degradation as indicated above) while the end-to-end channels would be no worse than about 34 dBa0. These values might well be thoroughly acceptable to the user.

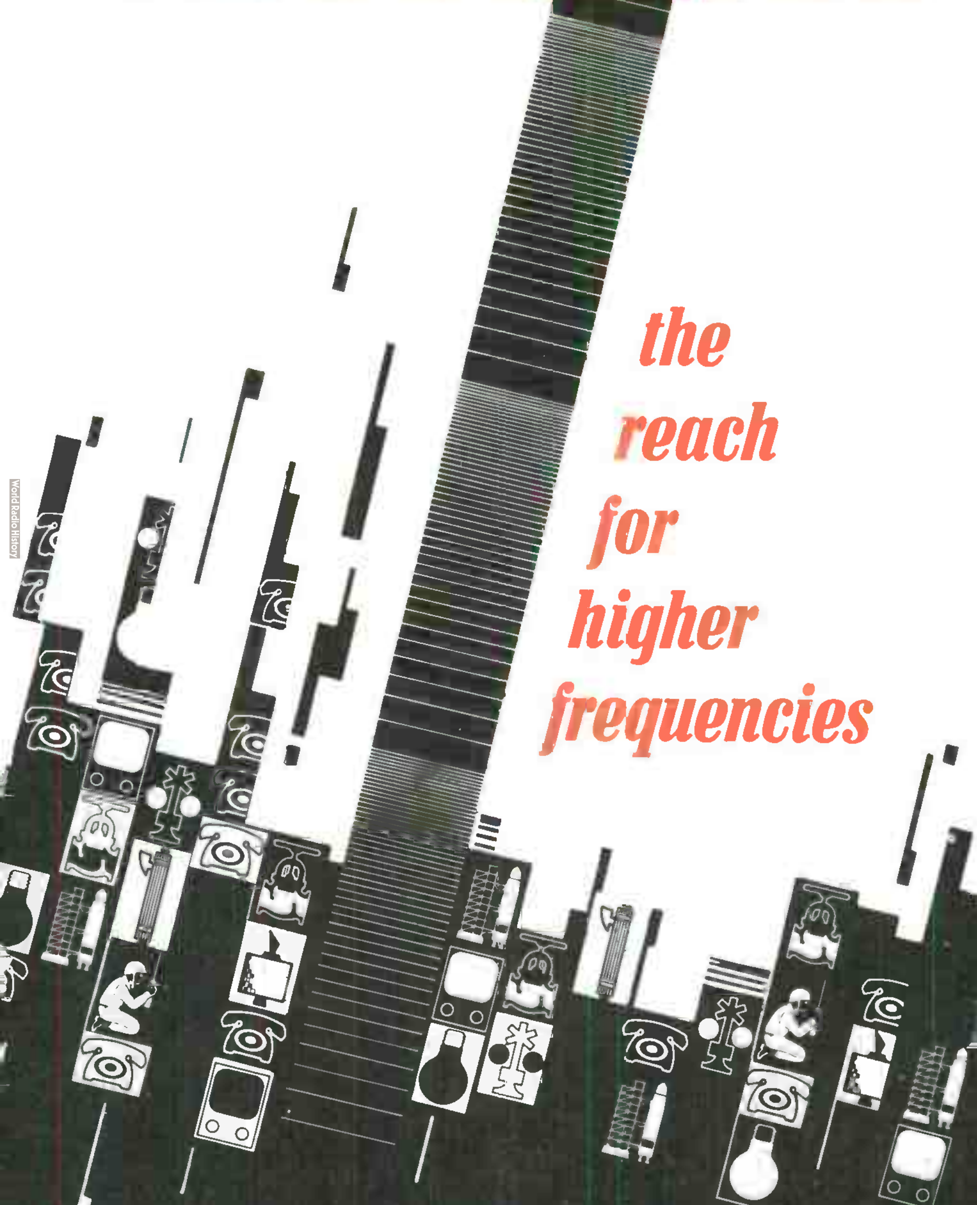
Obviously this approach would require careful evaluation and good judgment. It must be recognized that it would not be satisfactory if the noise performance on the original system were marginal. But when properly used in combination with the other methods of system expansion, it can provide very worthwhile results.

Update the System

If other measures prove inadequate the possibility of upgrading the microwave system itself should be considered at least in the most heavily loaded portions by such means as increasing antenna sizes or changing out i-f filters to provide wider bandwidth. In many situations these changes may be relatively simple and inexpensive when compared to the increased capacity obtained.

The Lenkurt DEMODULATOR

OCTOBER 1969



the reach for higher frequencies

PART 1

The trend toward higher radio frequencies creates new problems in system design.

Congestion of the radio-frequency spectrum has been foreseen almost from the beginning of commercial radio transmission. As radio technology has advanced, increasing demands for radio services have outpaced improvements in the efficiency of spectrum usage.

The result has been a continuing trend toward the use of higher and higher frequencies. There is no end in sight, with the upsurge in data transmission and other services that require the bandwidth of hundreds of voice channels.

Not too many years ago, "higher frequencies" referred to the VHF and UHF bands, with the then-new concept of line-of-sight propagation. Then came the microwave frequencies – 2, 4, and 6 GHz – where the signal behaved even more like a light beam. Each frequency "plateau" required new approaches to systems as well as hardware design.

Today, the 4 and 6 GHz bands form the heart of many communications systems. But new allocations in these bands are often difficult to get (impossible for some kinds of service). And in some major metropolitan areas, where the demand for service is the heaviest, no allocations are available at all. There is literally no place to go but up. For most types of service, the next available frequency bands are above 10.7 GHz.

The problems involved in building equipment for these higher frequencies were solved several years ago (Lenkurt introduced the highly reliable 76D Microwave Radio System in 1964). But the most reliable equipment is worth little if the system "goes down" because of propagation failure.

Path Considerations

In many ways, the higher microwave frequencies behave just like the lower ones. The path-attenuation calculations, for example, are identical. For a given path, they show that a 12 GHz signal has 6 dB more path attenuation than a 6 GHz signal. On the other hand, the gain of a parabolic antenna of given size is 6 dB higher at

12 GHz. Since both the transmitting and receiving antennas are involved, the 12 GHz signal would appear to have a 6 dB advantage. In practice, however, this advantage is essentially cancelled by higher receiver noise figures and higher waveguide losses at 12 GHz.

Selective fading at the higher frequencies can be effectively combatted by methods used at lower frequencies. Space diversity, for example, with its redundant transmission paths, substantially increases path reliability. The chance of both paths fading simultaneously is remote. In the types of service for which the allocations are available, in-band frequency diversity is an excellent defense against selective fading. Its chief disadvantage is the total amount of frequency spectrum it uses.

The effect of selective fading varies only slightly with frequency. For any given path reliability, the required fade margin in the 11 GHz band is at most a few dB greater than that needed at the lower frequencies.

Slightly less path clearance is required at the higher frequencies because the Fresnel-zone radii are smaller. Except in critical cases or on short hops, however, the difference is not likely to be very significant. For instance, on a 20 mile, 6.175 GHz hop, the first Fresnel-zone radius at 10 miles is 64.9 feet. If the frequency on the same hop is increased to 11.2 GHz, the radius decreases by only 16.8 feet.

Effects of Precipitation

In the design of most microwave systems, the path attenuation is assumed to be the same as that encountered in free space. This is a good approximation for frequencies up through the 6 GHz band. But as frequency increases, the signal becomes progressively more sensitive to precipitation.

Rain attenuates a microwave signal in two ways: the water absorbs energy, and the droplets scatter it. The severity of the attenuation is a function of the drop size, the temperature, the volume of water involved, and the signal frequency. The most significant part of this complex relationship can be summed up this way: the harder it rains, the bigger the drops, and the higher the frequency, the more severe the attenuation will be.

Of course, other forms of atmospheric moisture also affect signal attenuation, but rain is usually the dominant factor. Fog and mist are essentially light rain. Attenuation due to hail is only a small fraction of that caused by rain. The effect of snow varies widely, depending on the moisture content, the flake size, and the temperature; but snow generally carries a much lower volume of water than rain does.

At the higher frequencies, heavy rain can be a real problem. The theoretical curves of Figures 1 and 2 show how frequency would increase excess path loss if the rainfall rate were constant along the path. While actual rainfall rates are never uniform along the entire path, a hypothetical example based on Figure 1 gives some feel for the effect of extremely heavy rain (4 inches per hour) on signals of different frequencies.

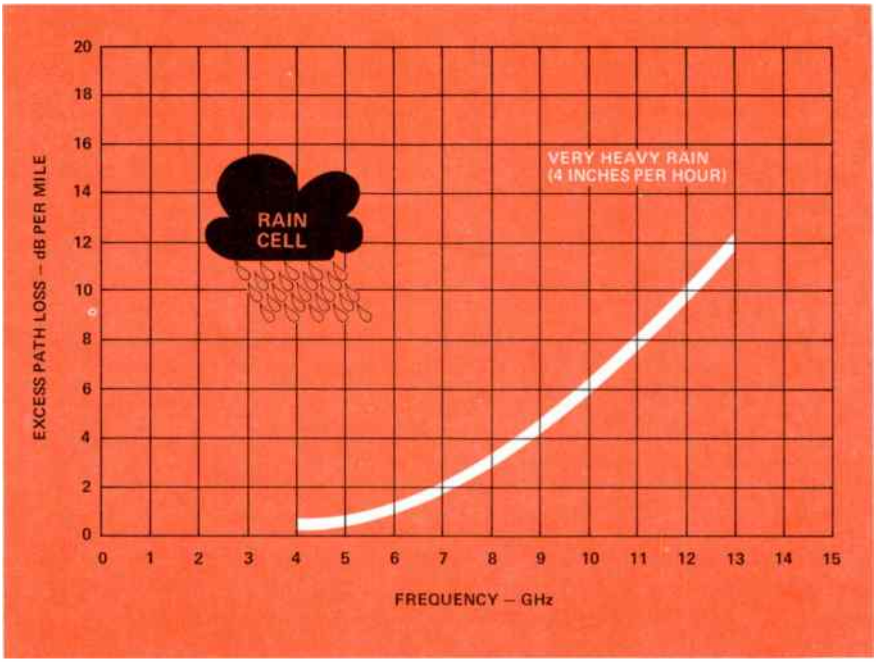


Figure 1. Excess path loss caused by heavy rain increases rapidly with increasing frequency.

A 6 GHz signal, with an excess path loss due to rain of only about 1.2 dB per mile, would be attenuated by about 24 dB over a 20 mile hop. That is centainly significant, but it is within the operating capability of a system engineered with, say, a 40 dB fade margin.

A 13 GHz signal, on the other hand, would suffer rainfall attenuation of about 240 dB.

As shown in Figure 2, a "heavy" rainfall of 0.6 inch per hour would cause excess path loss of only about 1.1 dB per mile at 13 GHz – roughly the same as the 6 GHz signal would suffer at the much higher rainfall rate.

The trouble with curves such as these is that they fail to take into account the changing nature of heavy rain.

Local Rainfall Distribution

It is relatively easy to measure the effect of rain on a microwave path. The difficulty arises in trying to measure the rainfall rates along the path for accurate correlation with the attenuation measurements. The harder it rains, the more likely it is that the rainfall rate will show wide and almost instantaneous variations. Furthermore, there may be very heavy rain at one point, and almost none a short dis-

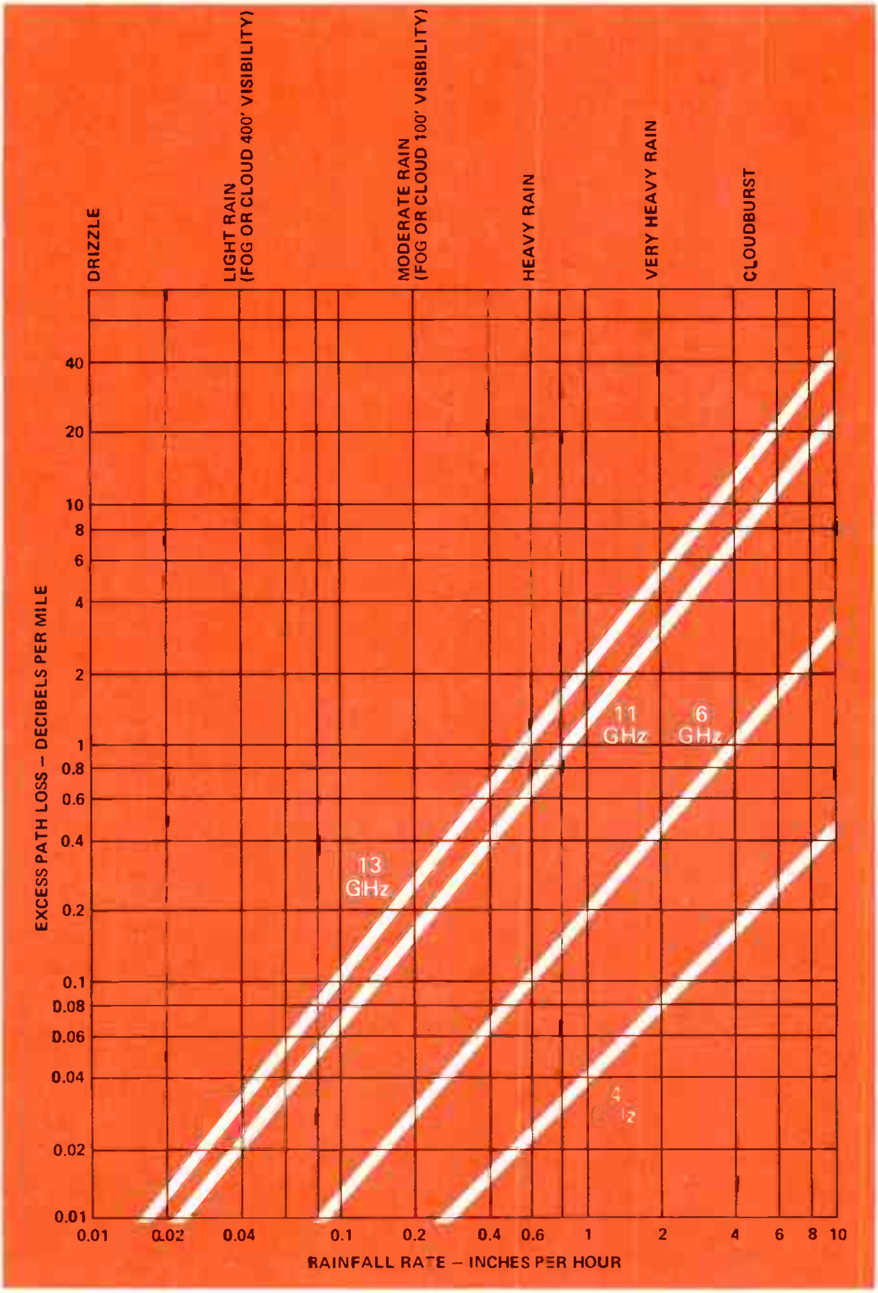


Figure 2. Theoretical curves show how attenuation increases with rainfall rate (based on calculations by Ryde and Ryde).

Figure 3. The contours of this map are for fixed transmission outage time and can generally be used in conjuction with the curves of Figure 4, to predict the effect of rainfall on outage time.

tance away. The cumulative figures for the total length of a microwave path are often irrelevant.

Much work has been done in recent years on the nature of rainfall patterns. The results are not conclusive, but they indicate that the most intense rain, the rain that significantly affects microwave propagation, occurs in relatively small cells. Available evidence indicates that these cells rarely exceed a few miles in diameter, and the rainfall rate varies even within the cell.

This variation means that even an intense cell may not block a microwave path for the entire time it takes to cross the path. A five mile wide cell, moving at 20 miles per hour, takes 15 minutes to cross a particular path at right angles. But regardless of its intensity, it may only cause some short outages. It is unlikely to block the path completely for 15 minutes.

Rainfall Distribution

Paradoxically, some geographical regions known for their large annual rainfall (such as the rain forests of Oregon and Washington) do not present as difficult a transmission problem as do other "drier" areas. The reason is, of course, that the total annual

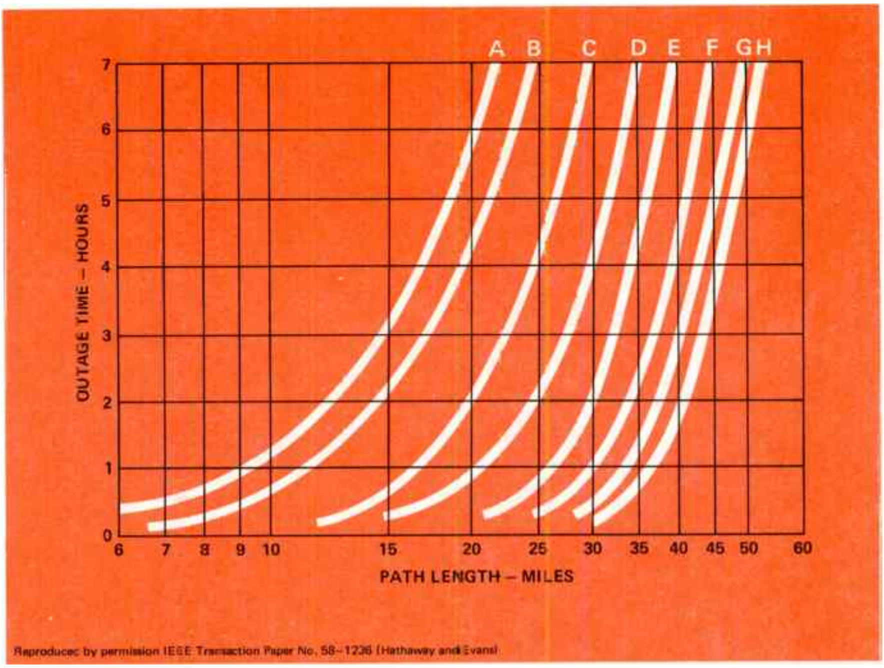


Figure 4. Expected outage time varies greatly with changing geographical rainfall distribution. These curves, for use with the contour map of Figure 3, are based on 11 GHz paths with 40 dB fade margins.

rainfall is of little consequence. *Concentrated* rain causes the trouble. The significant questions are: How heavy are the rainfall rates that can be expected? How often can such rates be expected?

The problem in any area is complicated by the fact that although much information is available on *annual* rainfall, very little is known about *instantaneous* rates. Gradually, however, the body of knowledge has built up so that it is now possible to generalize about many geographical areas.

The key factor, of course, is the amount of outage time a particular system is likely to suffer. Some types of service can tolerate substantial outages, while others cannot. Figures 3 and 4, the results of empirical studies, indicate generally how expected outage time varies with geography in the United States. For example, an 11 GHz path, 30 miles long and engineered for a 40 dB fade margin, would have an expected outage time of about 0.2 hour per year on Washington's Olympic Peninsula (contour H). This translates to a reliability of 99.998 percent.

If the same path were located on the coast of the Carolinas (contour C),

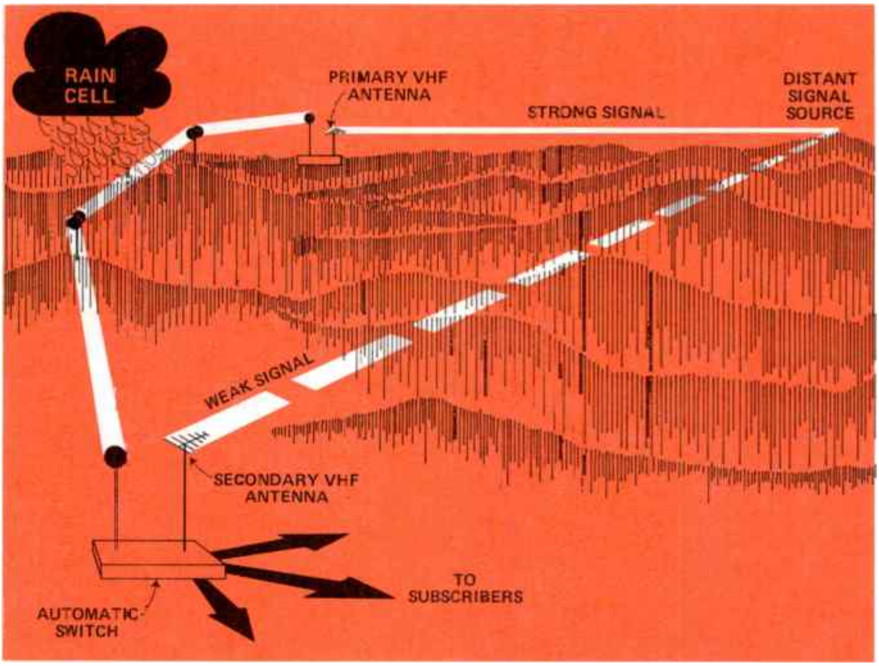


Figure 5. Many CATV systems can use a secondary off-the-air antenna as a back-up for a microwave link.

the predicted reliability would drop to 99.92 percent because the expected outage time would increase to 7 hours per year.

Now consider the same path on the Gulf Coast of Mississippi (contour A) and assume that the expected annual outage time must be held to the same 7 hours. The path would have to be shortened from 30 miles to 22 miles, or the fade margin would have to be increased substantially.

It must be remembered that these calculations are for a single hop. The outage time for the entire system can be expected to equal the sum of the single-hop outages. In terms of reliability, 10 hops with 99.99 percent reliability form a system that is only 99.9 percent reliable.

Living With the Problem

It may appear from what has been said that the frequencies about 10.7 GHz are a poor second choice, to be considered only when allocations at lower frequencies are not available. But this is only partly true. When the limitations of the higher bands are recognized, they give very good service. This really means learning to live with the rainfall problem.

First, of course, rainfall attenuation may not even be a problem. If the system is located in an area where rain rarely falls in cloudburst proportions, it can probably be engineered like any other system, with little worry about excess outage time.

Second, the type of service may be able to tolerate occasional outages of a few seconds to a few minutes. Some services, such as telephone common carrier, demand very high reliability. Others may be able to live with a few short outages.

A case in point is a CATV relay. Typically, the system operates only 18 hours a day. Thus, 25 percent of the heavy rains would be expected to occur during off-the-air hours. In some areas, heavy rainfall is consistently concentrated in these early morning hours.

Furthermore, many CATV systems can use a kind of "microwave/VHF diversity". If there is an outage in the microwave link, the signal is simply taken from the secondary antenna as shown in Figure 5. This provides an inferior signal, but it is usually watchable for the short periods of rain-induced outages.

The third factor in living with the rainfall problem is the length of the proposed system. Outages are cumulative. So a very long microwave system may have comparatively low reliability, even though the reliability of every hop is high. The longer the system, the greater the chance of a severe rain cell moving across the path somewhere. An obvious solution is to use the higher frequencies for short systems, and to reserve the lower frequencies for long, cross-country systems.

Diversity Arrangements

Heavy rain is not the only thing that will put a microwave system temporarily out of business. The mechanisms of selective fading are completely separate from those of rainfall attenuation. When the effects of rainfall attenuation cannot be completely controlled, one way to keep annual outage time down is to pay special attention to selective fading. This implies some sort of diversity arrangement. All diversity arrangements combat selective fading, and some provide protection against rainfall attenuation, as well.

Space diversity is no defense against rainfall attenuation because the two transmission paths are quite close, and subject to essentially the same rainfall pattern. In the typical case, where one path is directly above the other, the same rain will block both paths simultaneously.

While in-band frequency diversity also offers good protection against selective fading, it still provides no defense against rainfall attenuation. The two frequencies are so close together (typically separated by only two percent of the frequency) that the effect of rain is essentially the same on both.

On the other hand, if frequency diversity is extended to include both the 6 GHz and the 11 GHz bands (cross-band diversity), it can provide excellent protection against both selective fading and rain. It is true that when the 11 GHz path fails because of rain the entire load falls on the 6 GHz path. Fortunately, however, severe selective fading rarely occurs during per-

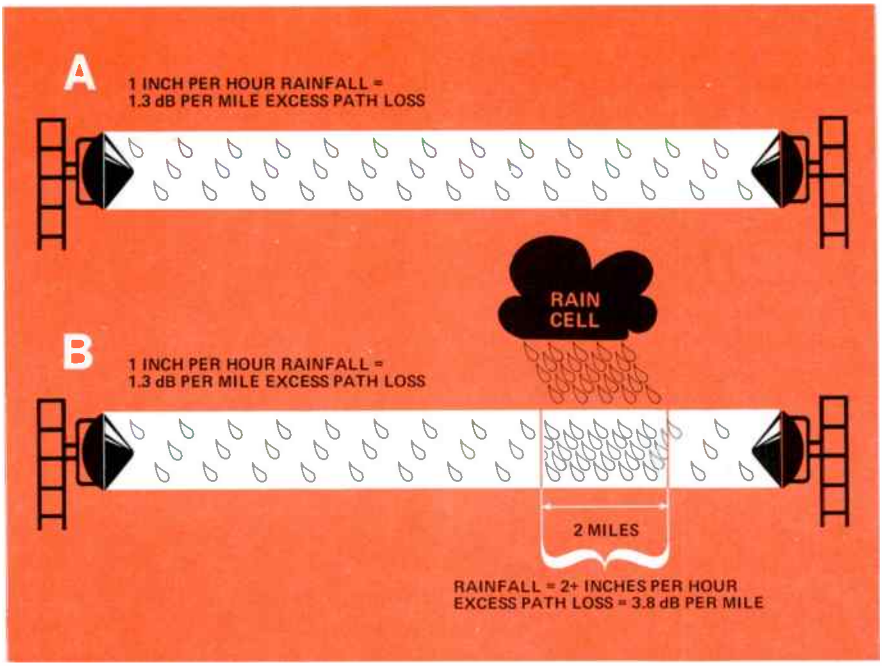


Figure 6. Raising the fade margin by 5 dB, (as in B above) can permit transmission through a rain cell of substantially greater intensity.

iods of heavy rain. This makes the 6 GHz path exceptionally reliable during such periods. Thus, cross-band diversity is probably the best solution – provided, of course, that frequency allocations and regulatory approval are available.

Perhaps the ideal solution would be route diversity – an extreme form of space diversity. The same signal would be sent over two paths separated by several miles. This would all but eliminate the possibility that rain would block both paths simultaneously. In practice, however, route diversity is not often used in present-day systems. The main reason is simple economics. Equipment and installation costs are quite high. Furthermore, the process of dropping and inserting channels is complicated – not to mention the difficulty encountered at the receiving end in trying to combine the signals from two paths of substantially different length.

Conservative Engineering

It is apparent that there is no easy, clear-cut way to avoid problems with rain. The most effective defense is a combination of techniques. And conservative engineering is the first one. A marginally engineered system is an invitation to excess outage time.

One thing that can be done, for example, is to increase the fade margin. This does not guarantee transmission through the heaviest rain, but it does effectively lower the expected outage time.

Because of variations in the instantaneous rainfall rate, it is not always possible to specify exactly how much effect a higher fade margin will have on rainfall attenuation. But some idea can be gained from a hypothetical example like this: Suppose a particular 11 GHz microwave hop can withstand an average rainfall along the path of 1 inch per hour – an excess path loss of about 1.3 dB per mile. If the fade margin is then raised by 5 dB, the hop can still withstand the 1 inch per hour rain along the path, except for a two mile segment where it passes through a rain cell. In that segment, it can withstand excess path loss of 3.8 dB per mile – equivalent to a rainfall rate of over 2 inches per hour (see Figure 6). In many areas, that much improvement will not eliminate rainfall outages. But it will reduce them.

Of course, increasing the fade margin may not always be desirable. If it means an increase in the number of hops, for instance, any gains in path reliability may be more than offset by the decrease in equipment reliability as more transmitters and receivers are added.

Equipment reliability is equally as important as path reliability. So is the reliability of the power source. And good maintenance is important, too. Improving the reliability of any one of these naturally improves the end product – total system reliability. Thus, economics is the common denominator in improving system reliability.

Where to Next?

The move to the 11 GHz microwave band is not the final rung on the ladder of ascending frequencies. This band will eventually become congested like all the others below it. What then? Still higher frequencies, with even more severe attenuation problems? Coaxial cable? Millimeter waveguides? Laser transmission?

The second part of this two-part article, to appear in next month's *Demodulator,* will discuss the future of transmission technology.

The Lenkurt DEMODULATOR

NOVEMBER 1969

the reach for higher frequencies

PART 2

The electromagnetic spectrum still has much untapped potential. Some frequencies are suitable for radio transmission through the atmosphere, while others require different transmission methods.

Part one of this two-part article discussed the transmission problems encountered as frequency congestion forces the shift to higher microwave frequencies. That discussion was limited to what is commonly called the 11 GHz band, because that is the highest band in general use today.

It is apparent to long-range planners, however, that this band will also become congested as the demand for more communications services accelerates. This trend shows no sign of tapering off and rapidly increasing services such as video and wideband data transmission continue to require enormous bandwidths.

The higher frequencies which can provide this bandwidth are unused, waiting to be tapped. They range in a continuous spectrum from the microwave frequencies through millimeter waves and infrared to the visible light region. Eventually, they may even include the ultraviolet range. The problem is finding a way to use them in practical communications systems.

The allocation of specific higher frequency bands is still under discussion. (It will be considered at the World Administrative Radio Conference, to be held in 1971 by the International Telecommunications Union.) Considerable work has already been done on microwave transmission in the 18 GHz region and above.

Atmospheric Considerations

Since rainfall attenuation is one of the most significant problems in the 11 GHz band, and the effect increases with frequency, the problem can be expected to be even more severe at higher frequencies. Figure 2 shows theoretical rainfall attenuation as a function of rainfall rate for selected frequencies up to 40 GHz. (Some empirical studies have indicated even higher attenuation than predicted.)

While rainfall attenuation is still the most significant atmospheric problem, fog and mist become increasingly important at the higher frequencies. The deciding factor is the volume of water in the air, which is perhaps easiest to understand in terms of visibility. At 30 GHz, for example, fog that cuts visibility to 150 feet attenuates the signal by about 0.5 dB per mile. It takes more than twice the moisture concentration to reduce the visibility to 100 feet at which point attenuation is about 1.6 dB per mile.

In this frequency region, another phenomenon – molecular absorption of the radio energy – also becomes a problem. Water vapor (not to be confused with water droplets) absorbs more energy as frequency increases, with the significant absorption occuring at resonant peaks. One such peak is at 22.4 GHz. At this frequency, a relative humidity of 60 percent produces absorption of about 0.4 dB per mile. At 18 GHz, the same humidity absorbs energy at the rate of only about 0.05 dB per mile.

Another minor effect is the molecular absorption of oxygen, which also increases with frequency. The loss only becomes significant, however at frequencies in the 50 GHz range.

Modulation Techniques

The frequency allocations for present-day microwave systems are intended primarily for equipment that uses low-deviation FM, with RF bandwidths of 20 MHz or less. This technique is well suited for voice traffic, which is usually multiplexed by frequency division. But, the nature of the traffic carried by microwave radio is being changed by two major factors. One is the tremendous increase in data communications, and the other is the increasing use of pulse-code modulation (PCM) for voice communications. The two are essentially the same from the microwave engineer's point of view. Either way, he is faced with the necessity to transmit pulses at a high rate (approximately 70,000 per second for each voice channel). One method is to use digital microwave transmission. Such a system becomes one more step in the time-division multiplex scheme.

An advantage of PCM is its relative immunity to noise. Because it is only necessary to detect the presence or absence of a pulse in a particular time slot (not its height, shape, or any other characteristic), a PCM system can operate at a very low signal-to-noise ratio. Consequently, it is quite tolerant of the severe atmospheric attenuation.

A binary system can use relatively simple repeaters. They need only produce new clean pulses to replace the old distorted and attenuated ones. A simple repeater is an inexpensive one. Since economics really dictate system performance, route diversity, with paths separated to avoid simultaneous heavy rainfall, may become economically feasible (Figure 3).

However, a more efficient use of bandwidth can be achieved by using multi-level transmission, rather than simple binary techniques. This, in turn, increases repeater complexity and cost. But, it still may be possible to build relatively inexpensive repeaters that are small enough for pole mounting.

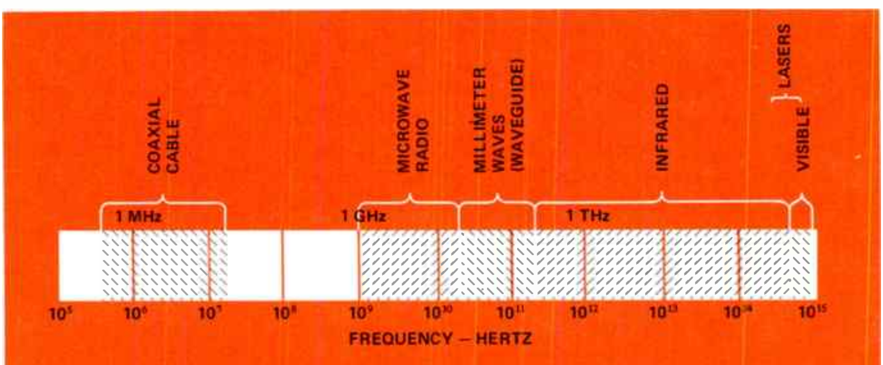


Figure 1. Examination of the electromagnetic spectrum shows large portions unused, particularly at frequencies above 10 GHz – where the largest information-carrying potential is.

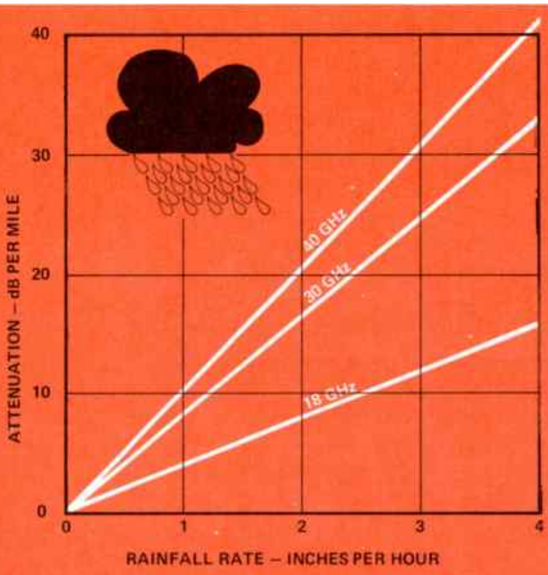


Figure 2. Attenuation caused by rain can be a formidable problem at the higher microwave frequencies. These theoretical curves (after Ryde) should be used only as approximations. Some measurements have indicated substantially higher attenuation.

A potential problem here is the public's increasing consciousness of aesthetic values. Current trends are toward underground installation of all utilities. And some communities might not be willing to accept pole-mounted microwave repeaters at one to two mile intervals.

Millimeter-Waves

The millimeter-wave region, from 30 to 300 GHz, is very attractive for wideband communications systems because of the tremendous bandwidth available. At these frequencies it is not at all unreasonable to think in terms of a 1 GHz baseband that could, in theory, accommodate over 200,000 voice channels – or the equivalent in other forms of communications.

Of course, the problems of atmospheric attenuation are exceptionally severe at these frequencies. In fact, transmission through the atmosphere may not be practical except for certain applications. One such case is satellite communications. Here, route diversity, in the form of widely separated earth stations can provide the necessary reliability. Futhermore, the signal path is primarily in free space rather than the atmosphere (Figure 3).

What about earthbound millimeter-wave communications? One answer is to shut out the atmosphere. A long roof is not practical, but a waveguide is.

The idea may sound strange to those used to thinking of waveguide in terms of the connecting link between a transmitter or receiver and a tower-mounted antenna. A significant loss can occur in 100 feet of this type of waveguide, and the loss increases as the frequency goes up. Losses would be prohibitive in a long system. For example, at only 4 GHz, one type of rectangular waveguide has a loss on the order of 50 dB per mile.

However, by using a circular electric wave in a round waveguide, the loss can be reduced dramatically (Figure 4). Also, the loss decreases as frequency increases. Since the physical

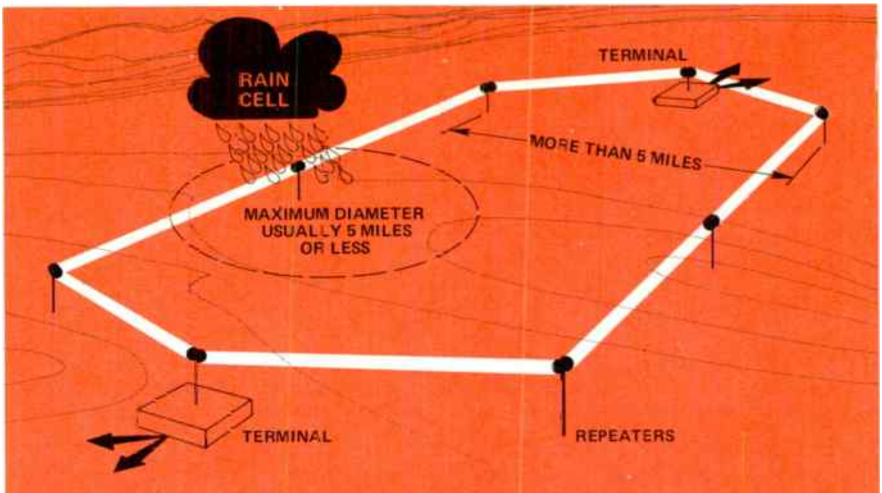


Figure 3. Route diversity provides protection against rainfall attenuation for microwave systems operating at high frequencies. If the paths are far enough apart, heavy rain is unlikely to block both at once.

size of the waveguide required also decreases with frequency, it is easy to visualize a small pipe carrying thousands of communications channels at millimeter wavelengths – with very low loss.

It happens that a 50 GHz signal loses only about 2 dB per mile in a waveguide with a 2 inch diameter – providing the mechanical tolerances are small enough. Theoretically, the loss would approach zero as the frequency approaches infinity. But, the mechanical requirements become so stringent that they limit the usable frequencies.

Any waveguide roughness or other imperfection causes mode conversion in the signal. Part of the energy gets "out of step" with the main signal. Not only is much of this converted energy lost, but the part that does get to the receiving end interferes with the desired signal.

Some mode conversion is inevitable, since a transmission line of any type cannot run indefinitely in a straight line – and any bend in the pipe causes mode conversion.

Consequently, the modulation method chosen must be resistant to interference. Once again, digital transmission becomes attractive. Not only is it interference resistant, but its adaptability to the increase in digital traffic is as important here as it is in atmospheric transmission.

Coaxial Cable

Another form of signal pipe is coaxial cable. It may seem strange to consider such an "old standby" in the same light as more exotic forms of transmission, such as millimeter waveguides. But coaxial transmission still has great unrealized potential. Equipment like Lenkurt's 46C Coaxial Transmission System carries 720 channels on routes of medium density. The Western Electric L4 system handles 3600 channels on high-density routes. Such systems may be only the beginning. Bigger systems are planned, with one intended to carry over 80,000 channels with multiple coaxial tubes. (Since the potential of coaxial transmission has more immediate impact than some of the other techniques discussed here, it will be the subject of a future DEMODULATOR article.)

Laser Transmission

Few people have been more excited over the useful possibilities of lasers than have communications engineers. The reason for their excitement is quite simple: the information-carrying potential of any communications channel is proportional to its operating frequency. Because lasers operate in a frequency range about 100,000 times higher than today's microwave radio systems, they have the potential to carry 100,000 times more information.

But, potential is sometimes far from reality. While laser beams have been used to burn through steel in industrial applications, their penetration range is limited. They are still light beams, and light beams do not penetrate very far through heavy clouds and other atmospheric obstructions. For this reason, unprotected laser transmission is practical only for short distances or in space communications. Long-range laser communications systems will have to follow an optically aligned tube. Here again, difficulties arise when the beam is bent – even enough to follow the curvature of the earth.

Therefore, any practical system will probably use a series of lenses to refocus the beam and change its direction slightly. In so doing, they will act somewhat as passive repeaters. Optical lenses can be used, but even the highest quality ones introduce substantial losses.

However, considerable promise is being shown by gas lenses. Such a lens can be formed by gas flowing through a heated tube. Because the gas is warmer near the tube wall and the cooler gas in the center is denser, it acts as a lens causing the beam to converge. The advantage of this type of lens is that it places no solid surface in the path of the light beam. Therefore the loss introduced by the lens is only that caused by the gas molecules scattering the light beam.

This principle sounds simple, but there are substantial obstacles to be overcome. A big problem is presented by the extremely critical mechanical tolerances required of a lens wave-

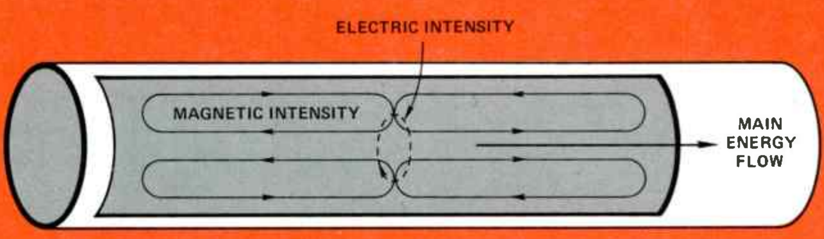


Figure 4. Electric field lines in a circular electric wave close on themselves, preventing significant charge accumulation on the waveguide walls, thus keeping wall currents low. Since charge accumulation is even less at higher frequencies, loss decreases with increasing frequency.

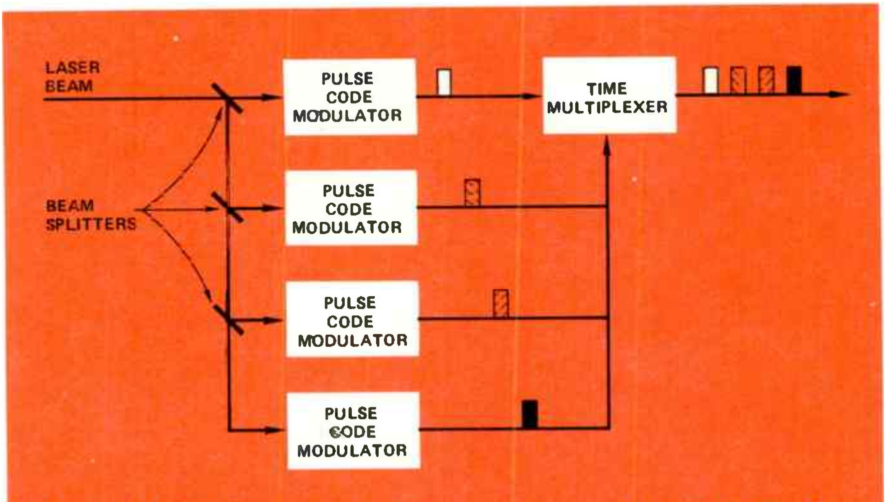


Figure 5. PCM shows considerable promise for modulating lasers. The beam-splitting arrangement shown here forms several high-speed channels from a single laser beam.

guide. The costs may make such an arrangement impractical.

Transmission is not the only area that presents problems for a laser communications system. Another hurdle is modulation and demodulation — and the associated area of multiplexing and demultiplexing.

One of the most promising modulation techniques is PCM — primarily because a laser can produce high pulse rates and very narrow pulses. If a laser beam is split as shown in Figure 5, parts of it can be sent to parallel modulators to form similar trains of narrow, relatively widely spaced, pulses. These pulse trains can then be interleaved for time-division multiplexing.

It is theoretically possible to add more multiplexing steps. If, say, 100 time-multiplexed signals were frequency multiplexed, the capacity would increase 100-fold. It is then conceivable that still another form of multiplexing, called spatial multiplexing, could be used. This means sending a number of beams simultaneously through a waveguide in different propagation modes.

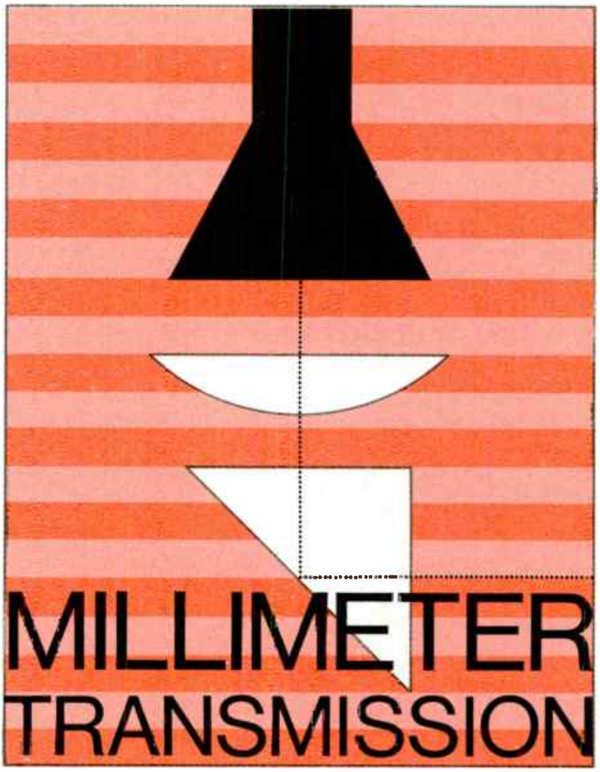
Such a system does not exist, and may never exist. However, a system has been suggested that would time-multiplex 32 channels in each of two polarization states, then frequency multiplex 100 of these "super channels," and finally use spatial multiplexing to combine 100 such beams.

The theoretical capacity of such a system staggers the imagination. The suggested bit rate would be about 2×10^{14} bits per second — the equivalent of 1,920,000 video signals.

The world has hardly begun to tap the potential of communications. It is not clear just what form the future uses of communication will take. But it is clear that man's capacity to devise communications systems has not been reached and the future is virtually unlimited.

OCTOBER, 1967
VOL. 16, NO. 10

The Lenkurt DEMODULATOR



more room for communicators in the frequency spectrum

The millimeter band in the electromagnetic spectrum lies in a gap between classical electric waves and optics. Traditionally this band, between the top microwave frequencies at 30 GHz and the bottom infrared at 300 GHz, has belonged to no one.

The physical properties of millimeter waves are such that until recently neither microwave nor optical techniques had been able to use them effectively. In theory the waves can serve several purposes. Their shorter wavelengths are well suited to radiometry, spectrometry, radar, and navigation.

To communicators the millimeter band is more than another source of transmission frequencies. By conventional standards it has about nine times the capacity of all the lower radio frequencies combined.

Broad bandwidth, highly directional antenna beam radiation, and small components are all available in millimeter communications systems. As might be expected these apparent advantages have some inherent difficulties. Physical size and atmospheric attenuation, aside from being valuable in certain applications, are two basic snags in the development of an operational system. Narrow beamwidth which requires precise pointing and, for a mobile system, accurate tracking are also drawbacks.

So far none of these problems has been insurmountable. Millimeter systems have been built and tested satisfactorily. All that remains is that the need for a millimeter system be great enough to support its cost.

From Optics

Because millimeter waves fall in the no man's land between electronics and optics, attempts to develop a satisfactory communications system have drawn on both fields. To date optical techniques have met with the least success.

Most optical approaches use atomic excitation, usually in a gas, to generate high frequencies. Known as the multiple quantum effect, this phenomenon is used to produce the maser or microwave version of the laser.

The phenomenon takes place in a resonant cavity where an external power source "pumps" atoms up to an excited state. When the atoms relax, they emit electromagnetic waves. In one experiment hydrogen cyanide produced a frequency of 105 GHz.

Unfortunately, the high frequency waves emitted have extremely low energy. The result is that effective generation by direct quantum mechanisms is not promising for communications applications.

Other efforts at generation include mixing coherent signals from two powerful lasers. Two lasers mixed in an element such as quartz or a potassium dihydrogen phosphate (KDP) crystal could theoretically produce millimeter waves.

Electronics

In spite of the ingenious approaches taken so far, adequate energy has yet to be developed using optical techniques. Electronic devices, in spite of obstacles, have brought more success.

Size is the major obstacle to electronic generation of millimeter waves. The usual method for generating power at radio wavelengths has been with tube type, free electron devices. But the free electron principle is directly linked to wavelength.

As the wavelength of a signal becomes smaller, a phenomenon known as the *transit time effect* seriously hampers the performance of a free electron tube. (Transit time refers to the time it takes an electron to travel from the tube's cathode to its plate.)

At short wavelengths the ac component of the voltage applied to the control grid reverses before an electron can transit the gap between plates. As a result, electrons cannot follow signal variations precisely. This causes losses in the oscillator which become excessive as the frequency increases and the wavelength shortens.

Fortunately the klystron tube was developed which overcame transit time limitations at microwave frequencies. The klystron produced continuous wave (cw) oscillations, but required small, resonant cavities at dimensions near the wavelength.

In the even smaller cavities required to produce millimeter waves a large amount of input power is lost as heat rather than converted to wave energy. This inefficiency could be accepted if the input power and the resulting output power were increased. But the small cavity does not dissipate heat fast enough to accommodate a larger input.

The klystron tubes which do operate in the low millimeter range are expensive. They have lifetimes limited to a few thousand hours and require high voltages.

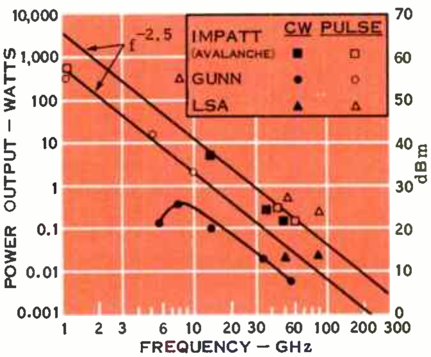


Figure 1. Maximum power produced by three solid-state devices. Note that the maximum pulsed power of the IMPATT avalanche type diode and Gunn diode decreases as frequency increases in the proportion $f^{-2.5}$.

Some electronic devices such as periodic beam devices have been able to produce relatively high outputs in the millimeter range. One such device has delivered 1 mW at 300 GHz. Similar devices have had outputs of one watt at frequencies ranging from 80 to 140 GHz but at efficiencies of 2 per cent.

Another Way

As an alternative, semiconductor devices have also been tried. They have the advantage of requiring less input power. But they still do not produce sufficient output power at high frequencies.

The frequency of most solid-state devices depends on the time it takes a space-charge (an electrical charge in space between the electrodes of a transistor or the plates of a tube) to travel through the device—the transit time. The smaller the transit time, the higher the frequency.

At General Telephone and Electronics Laboratories tests have been run using the thermoelectric effect of heated carriers in bulk semiconductors. Using this method General Telephone and Electronics Laboratories has been able to produce a pulsed output of 5 mW at 210 GHz. Generation took place in a frequency tripler. As might be expected the efficiency was quite low and, of course, the pulsed power output was not suitable to communications.

Until recently the two most promising solid-state devices were the avalanche-transit time diode and the Gunn effect diode. Both diodes have produced relatively high power and frequency outputs.

The avalanche diode uses an inductive cavity tuned to the diode capacity to build up oscillations. In the Gunn effect diode an electric field is applied across a crystal of gallium arsenide. The period of oscillation in the Gunn diode is roughly equal to the time it takes an electron to travel from one end of the crystal to the other at the voltages applied. Both diodes are transit time devices.

To attain higher frequencies in solid-state devices, therefore, it was necessary to shorten the transit time between electrodes. This meant shrinking the so called active region. The smaller active region, unfortunately, made transistors inefficient thereby limiting the output power.

The LSA

A new mode of oscillation, the "limited space-charge accumulation" (LSA), has been discovered which does not depend on transit times. The new mode makes possible high frequency, solid-state oscillators with useful power outputs. It is not susceptible to decreased powers at millimeter wave frequencies—the problem that had dogged every other attempt at generating short wavelengths.

LSA diodes, developed at the Bell Telephone Laboratories, have attained a continuous wave power output of 20 milliwatts at 88 GHz. This is reported to be the highest frequency recorded for a continuous wave, solid-state oscillator. LSA diodes operating at lower frequencies have produced correspondingly higher outputs.

In the LSA diode oscillator no space-charge is allowed to accumulate. As a result transit time does not play an important part in the oscillator's functioning, and physical size is not as critical as it is in other devices. The LSA diode, therefore, can be made thick enough to withstand relatively high applied voltages.

For LSA oscillation the diode functions as part of a resonant circuit which is tuned to the desired operating frequency. Both the diode and the resonant circuit must be properly designed and matched. Since the frequency is determined primarily by this circuit, the power is for all practical purposes independent of frequency.

LSA diodes have operated continuously for over four months in the solid-state repeater of an experimental millimeter communications system. The diodes are being tested as replacements for klystron tube oscillators which require high voltages from large regulated power supplies.

The operating life of LSA diodes is expected to be comparable to that of transistors and much greater than that of tubes which operate in the same frequency range.

Limited Successes

Most earlier millimeter transmission systems did not generate their waves directly. An experiment at TRW Systems used a 70 GHz klystron source and har-

Figure 2. Sylvania Electronics System's solid-state millimeter transceiver mounts on pedestal or tripod. At the bottom of the pedestal is the power supply.

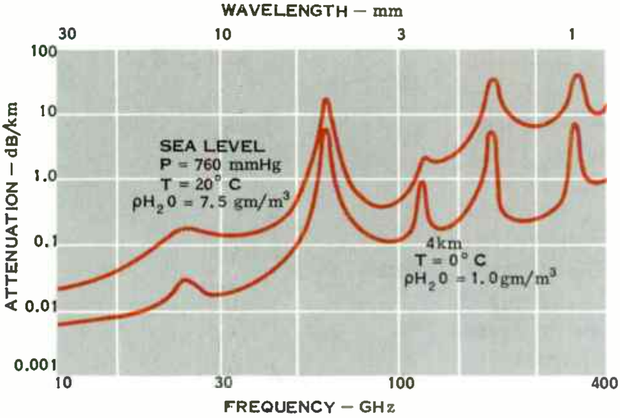


Figure 3. Attenuation of high frequencies at different altitudes is affected by oxygen and water vapor. P indicates barometric pressure, T atmospheric temperature and ρH_2O is water vapor density.

monic generation to reach 140 GHz. Another experiment at General Telephone and Electronics Laboratories operated a 90 GHz system across a distance of 1.2 miles with a high degree of reliability.

Sylvania Electronics Systems has developed a solid-state millimeter system which uses harmonic generation to reach 36-38 GHz. The transmitter uses a semiconductor amplifier and multiplier circuits to increase a low-frequency, crystal generated signal to a high frequency output. It is a frequency modulation system capable of 100 milliwatt output power.

The transmitter and superheterodyne receiver employ an identical horn and lens combination antenna. The surprising feature of the system is that it has a communications range of several miles through the atmosphere. Prior to the development of the Sylvania system, atmospheric propagation under anything less than ideal conditions was considered extremely tenuous.

In the Air

Changes in atmospheric temperature and pressure can have a critical effect on propagation. These two meteorological properties along with the oxygen and water vapor content of the atmosphere determine its dielectric constant and therefore its radio refractive index.

The refractive index indicates the speed at which a radio wave travels through a medium. In the atmosphere a change in the index of refraction can cause significant fluctuations in the angle of arrival of a signal at the receiver. In fact, a variation across the signal path can bend a signal enough to cause it to miss the receiver antenna. A change can also partially destroy a signal's coherence, making it unusable at the receiver. For a millimeter wave with its narrow beamwidth a variation in the index of refraction can be especially critical.

In general, atmospheric attenuation due to the molecular absorption of oxygen and water vapor easily disrupts millimeter waves. Fortunately, atmospheric attenuation is not constant for all wavelengths.

"Windows", as they are called, are spread across the millimeter band between oxygen and water vapor absorption lines. These windows are actually areas in the frequency spectrum which have relatively low attenuation. Fig. 3 shows where they occur in the millimeter band.

To complicate matters further the effects of scattering from water droplets are superimposed on atmospheric absorption. Here the ratio of particle size to wavelength is critical in determining the amount of energy scattered. If the scattering particles approach the size of the wavelength, near total extinction can result. Thus, it is important to consider the size of atmospheric particles in relation to the size of a wavelength.

Judged by typical particle sizes microwaves are relatively immune to rain, millimeter waves to fog, and infrared waves to haze. Obviously, atmospheric propagation in the millimeter band will be best in a hot, dry climate. Poor propagation would seem certain in wet weather, although Sylvania found that a moderate rainfall (2.5 mm/hr) had little effect on its millimeter system.

At rainfalls of 5 mm/hr and 12.5 mm/hr the range of the system did drop as low as 8.9 and 4.7 nautical miles respectively. Of course, dust and dirt could have the same effect as rain particles on these delicate waves.

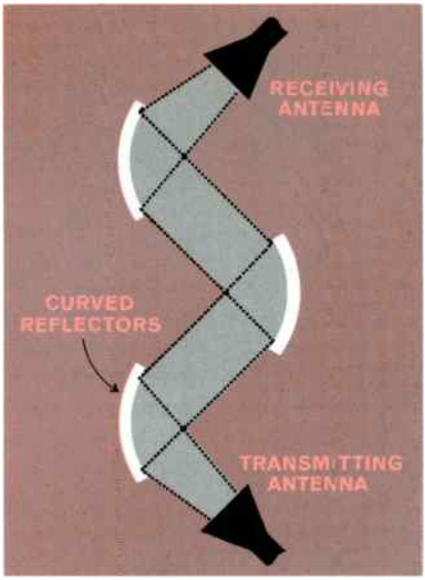


Figure 4. Reflecting beam waveguides use optical techniques to reflect millimeter wave energy along a path. Reflectors are usually curved to focus energy and reduce spreading losses.

Protection

Waveguides, on the other hand, are immune to the buffeting of the open atmosphere. They can provide a closed, controlled system which is well suited to millimeter transmission.

Waveguides do have drawbacks. They become less efficient as they are made smaller—a problem not unlike that encountered in generation.

Transmission losses increase at smaller wavelengths because currents crowd toward the surface of the guide. This phenomenon, known as the skin effect, increases resistance as the frequency becomes higher. The skin effect can be reduced by using circular waveguides with low transmission losses. Theoretically circular waveguides have an energy propagation mode with an electric field of zero along the inner wall of the waveguide. This zero electric field means low wall currents and low losses.

The circular waveguide transmits ever-increasing frequencies in the circular electric mode with ever-decreasing attenuation. Unfortunately, an extensive waveguide system is expensive. Its eventual use depends on the demand for broader communications channels.

More Optics

Other forms of guided transmission use such optical techniques as lenses, mirrors, and beam splitters. Curved mirrors form a reflective beam waveguide. In it a beam of millimeter wave energy travels from one mirror to another along a zig-zag path.

The mirrors are curved in order to focus the energy of the beam or otherwise collimate it to reduce spreading losses. Mirror size and materials must be carefully chosen to keep diffraction and surface losses low.

Beam waveguides using glass lenses to keep the beam tightly collimated suffer from losses on the order of 1-3 dB per lens. Gas lenses have much lower losses. They depend on a temperature gradient through the gas caused by heating the waveguide pipe.

The temperature difference across the gas causes a variation in the index of refraction. The gas, because of its index of refraction, forces the millimeter beam to remain in the tube's center (or just below the center where the coolest area of the gas settles because of convection). It actually focuses the beam.

Using heated gas in a long waveguide is unwieldy. In order to maintain the desired index of refraction, it is necessary to maintain the same temperature and pressure along the entire waveguide. This is not easily done over a long distance.

Detection

Detection does not run into as many problems as do transmission and propagation. Superheterodyne detection will work. The superheterodyne method mixes the incoming signal with one generated by a local oscillator to produce a usable signal. This new signal, the difference between the original two, is low enough in frequency to be applied and detected by available devices.

Any method of detection must take into consideration the angular tolerances of the narrow, millimeter wave beams. These tolerances are extremely small, imposing strict demands on the angle of arrival of a signal.

If a millimeter system is used on a moving vehicle such as a satellite or aircraft, precise position information is required. For heterodyne detection the vehicle's velocity must be known in order to compensate for Doppler frequency shifts. Moving vehicles must be tracked in angle and velocity—a function not normally associated with communications.

Useful

The broad bandwidth found in the millimeter region could accommodate high data rates easily. It appears to be ideally suited for machine-to-machine communications, especially for high speed computers. Photo transmission from space probes could be increased considerably if millimeter wave transmission were available.

In terms of equipment the millimeter band affords new possibilities for compactness. Because the size of component parts normally varies inversely with frequency, those in a millimeter system should be smaller and lighter than those used in a lower frequency system.

The beam emitted is highly directional. This, plus a high attenuation rate in the atmosphere, gives millimeter waves built-in security—of particular value to the military. On the other hand the negligible amount of attenuation in space and the expected size of equipment makes millimeter waves promising in this area.

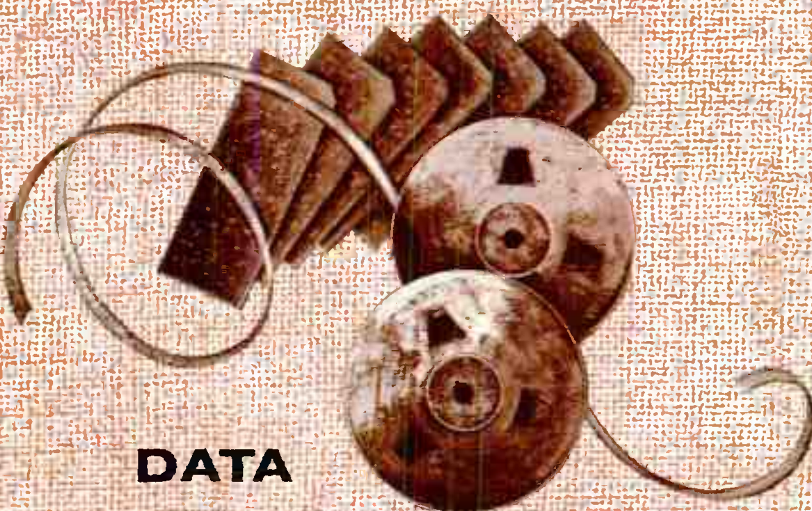
Commercially millimeter transmission systems will help keep pace with the expanding need for wideband communications.

At present the millimeter band is waiting to be used. Technology exists which can put at least part of the band to work. But economic considerations measured against need are the most influential factors slowing its extensive use. No doubt as demands on the frequency spectrum become greater, the millimeter band will come into its own.

SECTION III
DATA TRANSMISSION

SEPTEMBER, 1966
VOL. 15, NO. 9

the Lenkurt Demodulator



DATA COMMUNICATIONS

Part 1

Data communications is today a dynamic and rapidly expanding field stimulated by the increasing need to link electronic computers and other business machines across great distances. The resultant union of the data processor and the communicator has provided a vital service to the everyday operation of business, industry, and government.

This is the first of a two-part article which presents an introduction to data communications technology, and offers some insight into the future of this progressive field.

LESS than one percent of the computers in service today are interconnected through a communications network. However, it is estimated that ten years from now at least half of the computers in operation will be working together on a real-time basis. It is also anticipated that the volume of digital data transmitted over communications facilities will eventually equal and perhaps exceed the volume of voice traffic.

A little over a century ago the first data message was transmitted on wire lines by Morse Code. But this was not the beginning of the realization that the spoken word could be represented by some analogous language. Ancient records confirm that semaphore-type data or information transmission systems using the visual sense for perception existed even before the Greek and Roman empires.

But what is data? It might be described as factual information required as the basis for making decisions. Thus, statistical reports, engineering documents, and historical records all contain data. Data covers a broad range of information and plays an important part in the decision-making processes of our everyday lives. In this discussion, however, the meaning of data is limited to digital forms of information used in machine-to-machine communication.

The economics of computers and other types of business machines are based on moving information to achieve optimum use of what are usually expensive facilities. Some of the large-scale computers now in service are capable of input rates as high as 10,000,000 bits per second (b/s). Bit, a contraction of *bi*nary digi*t*, expresses a unit of information in a two-element binary code. The elements are called "mark" and "space", and indicate the choice between two equally possible events.

The requirement for data communications arises because modern business machines and computers can record and store information more efficiently, more accurately, and significantly faster than can humans. At present, thousands of data messages are transmitted over telephone networks at speeds many times faster than could be achieved by human speech.

Communicating with Data

Data signals are transmitted over various types of telephone circuits. They travel on wire from telephone pole to telephone pole, through underground cables, from mountain top to mountain top over microwave facilities, on the ocean floor in submarine cables, and via communications satellites from continent to continent. Some type of data conversion equipment is required to change the digital machine signals to a form suitable for transmission over these facilities.

The data machine which provides an input to the transmit section of the conversion equipment, or *modulator*, can be a keyboard, printer, card reader, paper tape terminal, computer, or magnetic tape terminal. The output from the receive section of the converter, or *demodulator*, can be applied to a tape punch, printer, card punch, magnetic tape unit, computer, or visual display terminal. Typically, both the modulator and demodulator sections of the converter are combined into a two-way data transmitter-receiver, commonly called a *data modem* or *data set*.

Figure 1 illustrates a typical full-duplex data transmission system including the originating data processing equipment and the interface assembly which consists of buffer and control units. The interface assembly at the transmitter accepts data at a rate determined by the operating speed of the data processor, stores the data temporarily, and regenerates it at a rate compatible with that of the data modem. At the receiving terminal the interface assembly accepts the received data, stores it, then feeds it to the data processor at the appropriate rate.

Timing signals from the interface assembly at the transmitter are applied to the data modem to synchronize the computer and the data set. At the receiver, synchronization pulses are derived from the data stream to synchronize the computer.

When more than one data set feeds into a computer, the capacity of the interface equipment is of major concern since it must determine the time slot allocation for each line. Various types of interface assemblies are employed, such as magnetic core memories, shift registers, and delay lines. Not all data communications terminals employ an interface between the data processor and the data modem. Without an interface the input, data transmission, and output functions proceed simultaneously and at the same rate of speed.

Since data signals are rarely in suitable form for transmission over the various types of transmission facilities, a signal coding process is normally performed. Ideally, the transmission medium should have linear attenuation and delay characteristics, but this is never so in practice, and transmission impairments are always present to disturb the data signals. As a comparison, in voice communications a high degree of transmission irregularities can be tolerated. If a voice circuit has a heavy loss or is noisy, the speakers compensate automatically by increasing the intensity of their voices. If words are missed because of transmission difficulties, they are often understood anyway because of the redundant nature of speech. In contrast, there is no inherent redundancy in data signals unless purposely inserted and, therefore, trans-

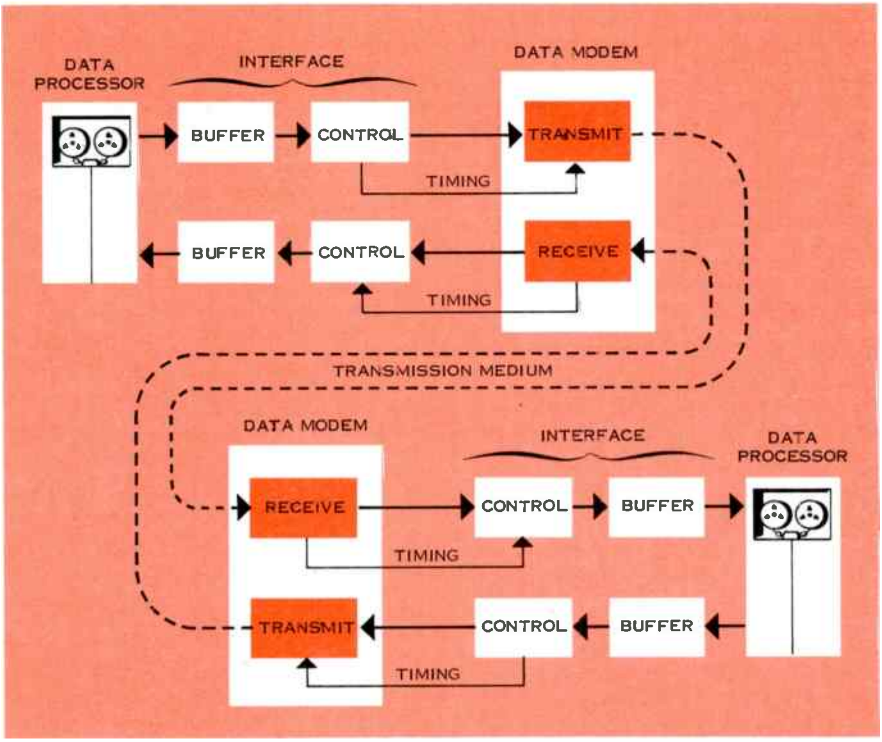


Figure 1. Typical full-duplex data transmission system arrangement. Timing signals from control unit synchronize data transmitter with data processor. At receiver, the element of time is important in reconstructing the original digital representation of the data. This is accomplished by deriving sync pulses from the data stream.

mission variations can only be compensated for over a very small range. In addition, data signals are sensitive to other transmission impairments which have little effect on speech.

Coding is undertaken to alleviate transmission irregularities, to increase the information capacity of the system, to enable error detection, and to provide message security. The coding process in the data transmitter (usually called *encoding*) simply rearranges the applied data machine signals into some other format. At the receiving end the reverse process *(decoding)* is performed to recover the original machine signals.

The diagrams in Figure 2 show the two types of information signals that are applied in digital form to a data modem. Shown in *A* is a binary *non-return to zero* (NRZ) signal. In *B* the same signal is shown in the *return to zero* (RZ) format. The difference between *A* and *B* is that in *A* successive marks or spaces follow one another, whereas in *B* there must be a return to the space level between successive marks. The voltage values of marks and spaces are arbitrary and may be positive, negative, or both.

Telephone Facilities

When data communications developed, a long established voice transmission facility already existed, and logically included service to those locations that would be the terminal ends of a data communications network. To provide economical data communications service, consideration must be given to using existing transmission facilities. It would be financially impractical to establish a completely new data communications network where existing voice facilities could satisfy the need.

An example of the use of voice facilities is the "time sharing" of a computer by several users for different purposes. Although the computer serves each user in sequence, it appears that all users are handled simultaneously because of the high-speed of the computer. A typical time sharing system uses a keyboard printer to connect to a remote computer via a data set. Eventually, data transmission over voice facilities might allow the automatic payment of bills, ordering of groceries, and a variety of other household tasks.

There are times when the nature of the data to be transmitted may prevent using normal voice facilities because of such factors as speed, quality, and compatibility. In this case, the use of a microwave wideband communications facility or a narrow band telegraph channel — but not a voice channel — might be required. At present, however, telegraph and public telephone line facilities are most commonly used for data transmission because of their wide availability and economy.

Data generated at such speeds that transmission requires part or all of a 3-kHz voice channel is normally referred to as *voice-band data.* Within this classification, data rates of 200 bits per second or less are called *low-speed* data. Data rates from 2000 to 2400 bits per second are referred to as *high-speed.* Between the two, data is called

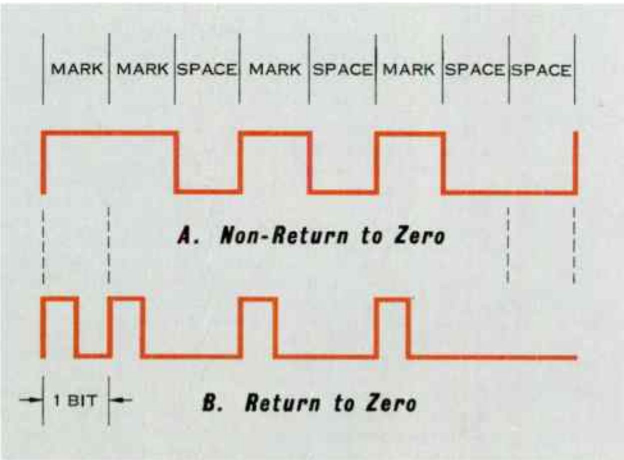
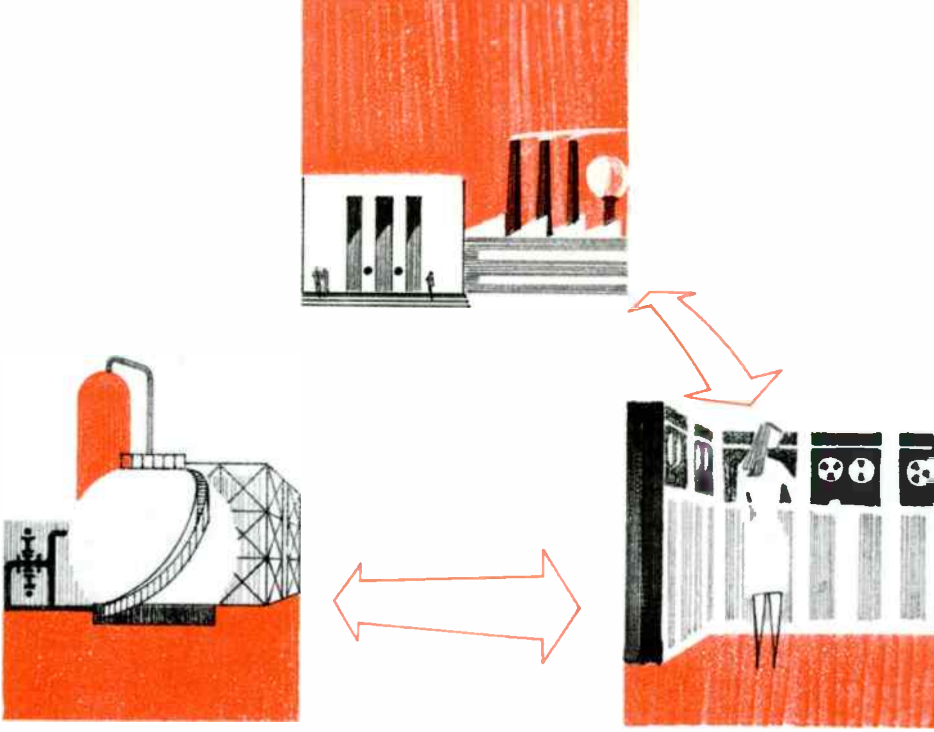


Figure 2. Digital representations of basic information signals.

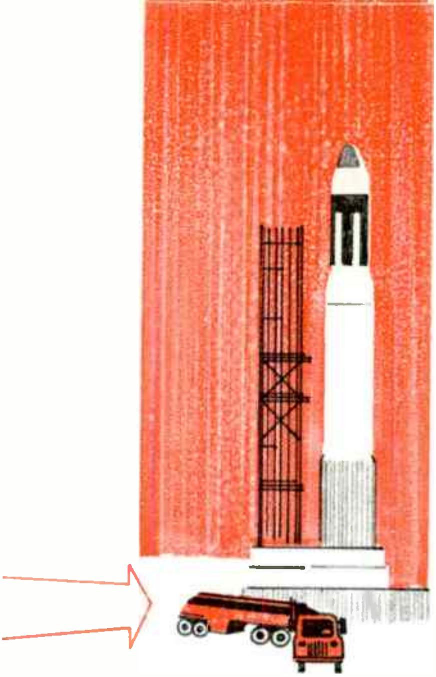


medium-speed. For typical speeds and uses of voice-band data, refer to Table A.

Data signals at speeds requiring more bandwidth than a single voice channel are called *wideband*. The most popular use of wideband data terminals is for remote access in real time to high-speed digital computers. Many of the present wideband data communications systems

TABLE A. Typical Speeds and Uses of Voice-Band Data

Speed	Classification	Use	Number of Circuits Per Voice Channel
75 b/s (120 Hz channel spacing)	Low-speed	5 level 100 wpm teletype 5 level 60 wpm teletype Variable frequency telemetering Pulse duration telemetering Alarm and control	25
110 b/s (170 Hz channel spacing)	Low-speed	8 level 100 wpm ASCII coded teletype All applications of 75 b/s speed	18
200 b/s (340 Hz channel spacing)	Low-speed	Data collection networks (remote to computer)	7
1200 b/s 2400 b/s	Medium-speed High-speed	Computer to computer Secure voice vocoders Pipeline telemetry and control	1



With the increased decentralization of business and industry, and with the need for a worldwide government digital network, data communications extends the services of data processing equipment far beyond the confines of a single office.

operate with 48- and 240-kHz bandwidths, which are the group and supergroup allocations of common multiplex systems. These multi-voice channel allocations may be used for regular voice traffic during busy periods, and during normally slack times used as a single wideband channel for data transmission.

Because cost and not speed ordinarily determines what type of system can be used efficiently, data rates within the voice-band and wideband classifications can vary to a broad extent. Why produce a highly complex and expensive wideband data set when an economical lower speed system will serve equally well? There are over 160 different types of data sets now being manufactured. These operate with approximately 16 different transmission codes, at least 12 different transmission speeds, and numerous methods of error detection and correction.

Transmitting Information

During the past fifty years several investigations have been made concerning the theoretical digital signal capacities of communications channels. In the late 1920's, H. Nyquist, a mathematician at Bell Telephone Laboratories, established a relationship between the bandwidth of an ideal *rectangular* distortionless communications channel and the speed of digital transmission. (Rectangular refers to the bandpass characteristic of a channel — linear throughout the band, with sharp attenuation at the ends.) Nyquist showed that the signaling rate in bits per second is equal to twice the bandwidth in hertz of a lowpass ideal rectangular channel. For example, using Nyquist's criterion, the normal 3000-Hz bandwidth telephone transmission channel could handle a maximum of 6000 bits per second. However, it was realized that the distortionless conditions laid down by Nyquist were ideal and could not be achieved in practice.

Later, C. E. Shannon, then at Bell Telephone Laboratories, examined how much information a channel of given bandwidth would pass in the presence of noise. Shannon's analysis yields a rate of nearly 30,000 b/s for an average telephone channel with a good signal-to-noise ratio. Shannon did not provide a practical means of achieving such transmission capacity, and Nyquist's rate has not been attained in modern data communications. In contrast to the idealized rectangular model of Nyquist, the actual physical channels are not rectangular but have gradual cutoff characteristics and, therefore, require about twice the Nyquist bandwidth, or approximately 1 cycle per bit for optimum binary transmission. (For more detailed information concerning Nyquist's and Shannon's formulas, refer to the April and May 1965 issues of *The Lenkurt Demodulator*.)

Bits and Bauds

The speed of signaling, measured in terms of the amount of information transmitted per unit time, depends on the transmission path and its associated apparatus. Bits per second expresses the total number of *information* pulses in one second and includes redundant bits used for checking errors. If the pulses are of varying length, or if start and

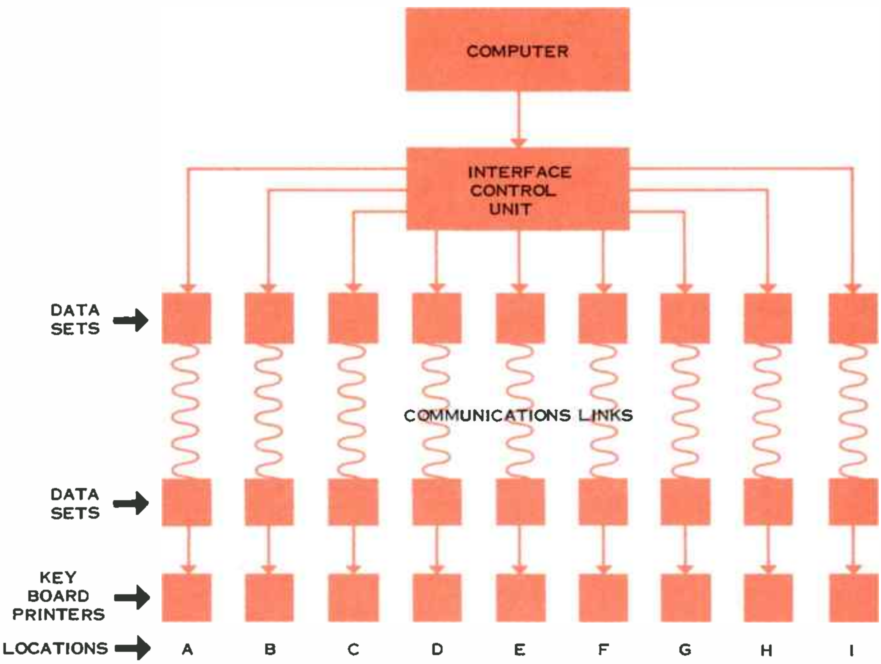


Figure 3. A typical time sharing computer employs an interface control unit to apportion its time among many users at different locations.

Figure 4. Data communications terminal (arrow) permits fast and reliable transmission at 2400 bits per second from a central computer.

stop pulses between each character are added that are not part of the message, the bit rate tells nothing of their number or duration. On the other hand, baud — from Jean Maurice Emile Baudot, an officer in the French Telegraph Service who contributed to early telegraph principles — is defined as the reciprocal of the time of the shortest signal element in a character. The term baud is often misinterpreted as a synonym for bits per second. However, the number of bauds equals the number of bits per second only when all time intervals are constant, and all signal pulses are information pulses, such as in binary transmission.

An example of the relationship between bits and bauds is in ordinary teletypewriter transmission, which makes use of a five-bit code, each bit being 13.5 milliseconds in length. The baud rate is therefore the reciprocal of 13.5 milliseconds, or approximately 74.2 bauds. A single character consists of a start pulse and the five information pulses or bits, each of 13.5 milliseconds duration, for a total of 81 milliseconds. A stop pulse of 19 milliseconds ends the character. The total time for a character is then 100 milliseconds.

Since the bit speed depends on the number of *information* pulses transmitted per unit of time, the equivalent rate for this type of transmission is 5/100 ms, or 50 bits per second.

Now, if a lapse period of 20 milliseconds is arbitrarily inserted between the stop pulse of this character and the start pulse of the next, the bit rate would be reduced to $\frac{5}{100\text{ ms} + 20\text{ ms}}$

or 41.7 bits per second. However, the teletype speed would remain at 74.2 bauds, because the baud rate depends only on the time length of the shortest pulse (13.5 milliseconds) in the character.

The number of words per minute can be determined using the ordinary telegraph definition of a word, which is 6 characters. The speed in bits per second is converted into bits per minute by multiplying by 60; hence, 50 bits per second equals 3000 bits per minute. Since there are 5 bits per character and 6 characters per word, there is a total of 30 bits per word. Dividing 3000 bits per minute by 30 bits per word equals 100 words per minute. Here again, the transmission rate in words per minute could be reduced by a slow teletypewriter operator, but the signal speed remains at 74.2 bauds.

It can be concluded that the baud rate is very important to the telephone engineer, since this rate establishes the type of telecommunications channel to be used. To a lesser degree the computer engineer is concerned with baud rate, but economics and speed of information flow are uppermost in this technology. Hence, to him the bit rate is the major concern, and is the more common expression used in dealing with binary data transmission.

Serial and Parallel Data

The terms *serial* and *parallel* are often used in descriptions of data transmission techniques. Both refer to the method by which information is processed. Serial indicates that the information is handled sequentially, similar to a group of soldiers marching in single file. In parallel transmission the information is divided into characters, words, or blocks which are transmitted simultaneously. This could be compared to a platoon of soldiers marching in ranks.

The output of a common type of business machine is on eight-level punched paper tape, or eight bits of data at a time on eight separate outputs. Each parallel set of eight bits comprises a character, and the output is referred to as *parallel by bit, serial by character*. The choice of either serial or parallel data transmission depends, of course, on the customer's data processing equipment and the transmission speed requirements.

Business machines with parallel outputs, however, can use either direct parallel data transmission or serial

TABLE B. Standards Organizations for Data Communications

Organization	Data Subdivision	Scope
International Telephone and Telegraph Consultative Committee (CCITT)	Study Group A	Data Communications including standards
Electronic Industries Association (EIA)	Committee TR30	Data transmission electrical standards
American Standards Association (ASA)	Subcommittee X3.3	Data transmission electrical standards
Institute of Electrical and Electronics Engineers (IEEE)	Data communications and telegraph systems committee	Electrical Information exchange

transmission, with the addition of a parallel-to-serial converter at the interface point of the business machine and the serial data transmitter. Similarly, another converter at the receiving terminal must change the serial data back to the parallel format.

Both serial and parallel data transmission systems have inherent advantages which are somewhat different. Parallel transmission requires that parts of the available bandwidth be used as guard bands for separating each of the parallel channels, whereas serial transmission systems can use the entire linear portion of the available band to transmit data. On the other hand, parallel systems are convenient to use because many business machines have parallel inputs and outputs. Though a serial data set has the added converters for parallel interface, the parallel transmitter requires several oscillators and filters to generate the frequencies for multiplexing each of the side-by-side channels and, hence, is more susceptible to frequency error.

Standards

Because of the wide variety of data communications and computer equipment available, industrial standards have been established to provide operating compatibility. These standards have evolved as a result of the coordination between manufacturers of communications equipment and the manufacturers of data processing equipment. Of course, it is to a manufacturer's advantage to provide equipment that is universally acceptable. It is also certainly apparent that without standardization intersystem compatibility would be almost impossible.

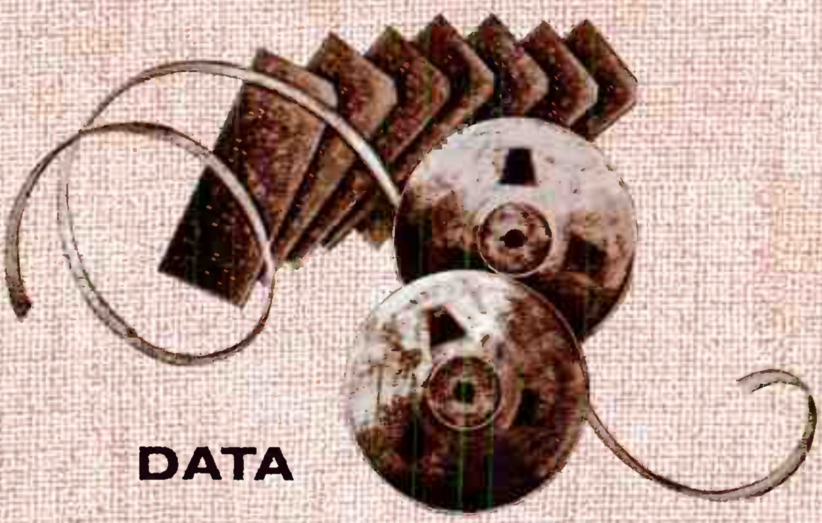
Organizations currently involved in uniting the data communications and computer fields are the CCITT, Electronic Industries Association (EIA), American Standards Association (ASA), and IEEE. (See Table B.)

A generally accepted standard issued by the EIA, RS-232-B, defines the characteristics of binary data signals, and provides a standard interface for control signals between data processing terminal equipment and data communications equipment. As more and more data communications systems are developed, and additional ways are found to use them, the importance of standards will become even more significant.

The second part of this two-part article will appear in the October issue of the Demodulator. Subjects to be covered include error detection and correction, transmission methods, and signal impairments.

OCTOBER, 1966
VOL. 15, NO. 10

the Lenkurt Demodulator



DATA COMMUNICATIONS

Part 2

Today, telephone companies are faced with the challenge of providing reliable data transmission at higher speeds over voice facilities. Because data signals do not possess the built-in redundancy of speech signals, data accuracy is very important. A single error could render a message incorrect. This article, the second of two parts, describes some of the major causes of errors and discusses some of the considerations involved in transmitting data over telephone lines.

ONE of the most important considerations in transmitting data over communications systems is accuracy. Data signals consist of a train of pulses arranged in some sort of code. In a typical binary system, for example, digits 1 and 0 are represented by two different pulse amplitudes. If the amplitude of a pulse changes beyond certain limits during transmission, the detector at the receiving end may produce the wrong digit, thus causing an error.

It is very difficult in most transmission systems to completely avoid such errors. This is especially true when transmitting digital signals over an analog transmission system designed for speech signals. Many of the inherent electrical characteristics of telephone circuits have an adverse effect on digital signals, often making the circuits unsatisfactory for data transmission—especially at high speeds. Frequently, these circuits must be specially treated before they can be used to handle data at speeds above 2000 bits per second.

Voice channels on the switched (dial-up) telephone network exhibit certain characteristics which tend to distort typical data signal waveforms. Since there is random selection of a particular route for the data signal with each dialed connection, transmission parameters will generally change, sometimes upsetting the effect of built-in compensation networks. Often data communications operations require speeds or types of signals which cannot be handled by the switched network. In addition, the switched network cannot be used for large multiple address data systems using time sharing. Because of these considerations, specially treated voice bandwidth circuits are made available for data use (see Table A). The characteristics and costs of these point-to-point *private lines* are published in documents called tariffs, which are merely regulatory agreements reached by the FCC, state public utilities commissions, and operating telephone companies regarding charges for particular types of telephone circuits. The main advantage of private or dedicated facilities is that transmission characteristics are fixed and remain so for all data communications operations.

Signal Impairments

Probably the most critical circuit qualities affecting data transmission are attenuation-frequency response, phase-frequency characteristics, and impulse noise. In addition, echoes and circuit net loss or over-all attenuation tend to degrade data pulses and usually have to be considered when selecting a voice circuit for data transmission.

The attenuation-frequency response limits the bandwidth for transmission. This characteristic causes the transmission loss of a circuit to vary with frequency. Ideally, the attenuation-frequency curve should be "flat" across the band required for data transmission. Most modern voice transmission circuits are sufficiently "flat" between 600 and 2700 hertz so that low-speed data is not seriously affected by this characteristic. However, for high-speed data transmission, it often is necessary to provide some sort of compensation in order to equalize the attenuation-frequency response.

The phase-frequency characteristics of a circuit cause what is known as *delay distortion*. (See *The Lenkurt Demodulator*, June, 1965.) Delay distortion results from the capacitive and inductive reactances common in all communications circuits, causing the propagation rate of signals to vary non-linearly with frequency over the desired bandwidth. This is not a problem in speech transmission because the human ear is not very sensitive to phase-frequency varia-

TABLE A. Specially Treated Voice Bandwidth Circuits for Data Use

	Use	Interstate Tariff FCC No.
Schedule 2 Telephoto	Alternate voice/facsimile (telephoto) or FAX only	140
Schedule 2 Telephoto with Special Conditioning	Alternate voice/facsimile (telephoto) or FAX only	140
Remote Operation & Control (FAA)	Simultaneous voice/remote operation and control or remote operation and control only	135
Schedule 4 Type 4 Data	Alternate voice/data or data only	237
Schedule 4 Type 4A Data	Alternate voice/data or data only	237
Schedule 4 Type 4B Data	Alternate voice/data or data only	237
Schedule 4 Type 4C Data	Alternate voice/data or data only	237
Schedule 5 Data	SAGE data circuits digital data	237

tions. But for data, such variations limit the speed of transmission and reduce the margin for error.

Delay distortion becomes more critical as data speeds increase. Typically, higher data speeds are achieved by shortening the width (duration) of each pulse. Because of the shorter pulse, slight shifts in relative phase between frequency components have a greater effect in distorting the signal.

Some type of compensating network or *delay equalizer* must be employed when the phase-frequency characteristics of a circuit are unsuitable for data transmission (see Figure 1). Such devices introduce a controlled amount of phase shift into the circuit at various frequencies.

The most difficult type of signal impairment to overcome is *impulse noise.* Such noise is extremely unpredictable and is commonly caused by electrical storms and the operation of switching equipment in the telephone plant. Quite often impulse noise will raise or lower the amplitude of a data pulse above or below a fixed detection or slicing level, causing the wrong binary symbol to be indicated at the receiver. One way of overcoming the problem of impulse noise would be to raise the signal power. However, communications systems have limited power handling capabilities which cannot be exceeded.

White noise, on the other hand, has a relatively uniform distribution of energy. Caused by the thermal agitation of electrons in resistances, white noise is always present in electrical circuits and cannot be eliminated. Since this type of noise is predictable, its effect can usually be overcome in the design of a data communications set.

Signal impairment also may be caused by variations in the carrier frequencies between the transmit and receive terminals of a single-sideband suppressed-carrier multiplex communications system. In such systems, carriers removed at the transmit terminal must be reinserted at the receive terminal. Any change in frequency between the transmit and receive carriers will shift the various frequency components of the data signals. If the frequency difference is great enough, the data signals can become so distorted that they cannot be correctly detected. This problem is

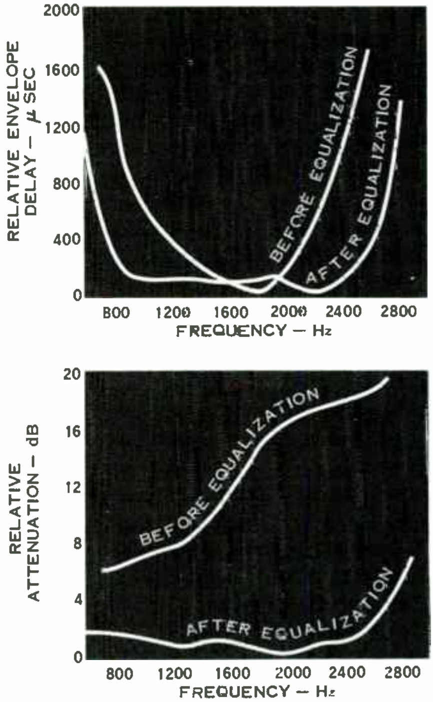


Figure 1. Both delay and amplitude distortion affect the quality of data transmission. Top panel shows the relative envelope delay of a typical voice channel both before and after equalization. The result of amplitude equalization is shown in bottom panel.

overcome by synchronizing the carrier generators of the two multiplex terminals, usually with a *pilot* signal which is transmitted through the system. Also, the use of a subcarrier in the modulation scheme of a data set helps to minimize the effect of frequency shift when data signals are fed through multiplex systems.

Error Detection and Correction

As data signals leave the business machine or computer and enter the data transmitter, they are essentially free from error. Since telephone systems were originally designed for voice transmission, various characteristics of the telephone circuit can impair the quality of a data message, possibly changing its context and rendering it unusable. Therefore, from the standpoint of service, the actual error rate or probability of error, which is determined statistically, is the most important factor in evaluating the overall performance of a data set.

Because of the significance of errors in data communications and since it is not practical to build error-free transmission circuits, the usual course is to equip data sets with some type of error control. Typically, error control systems in use today are capable of both error detection and correction, or error detection only.

Techniques which only detect errors are generally less complex than those which detect and correct errors. The simplest and most widely employed method of error detection is simple *parity check coding*. This technique uses redundant bits of information inserted into the digital message so that there is always an odd or even number of mark or space bits transmitted. Parity check coding, though vulnerable to

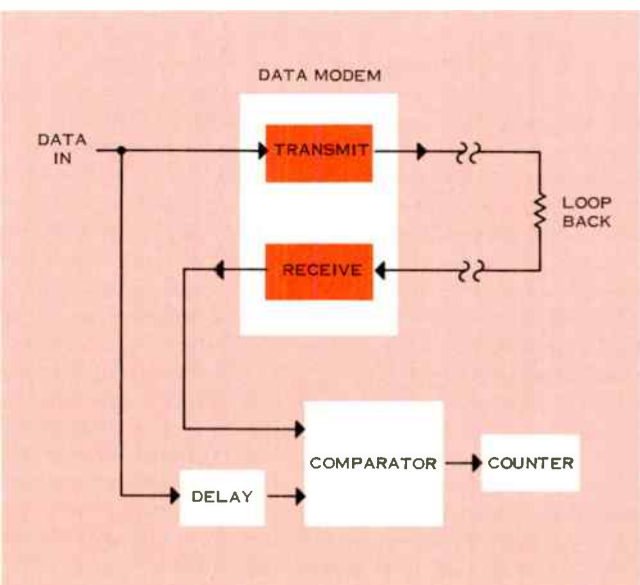


Figure 2. Typical method for measuring error performance of a data communications set. Data signals are looped back from transmitter to receiver. Delay corresponding to the amount in the path is added to the original data signals which are then compared bit-by-bit with the received data signals. An error counter keeps track of the number of errors.

many kinds of multiple errors which overcome the capacity of the coding method, has found wide use because the data receiver can be arranged to check parity without the need for complex circuits.

Error correction techniques use redundancy on a larger scale than the parity check method. The amount of redundancy determines the maximum number of errors that can be corrected. Error correction arrangements are extremely complicated, and the reduction in effective transmission speed necessary to accommodate redundant bits can become excessive. Furthermore, a substantial percentage of errors tend to come in bursts, so the utility of many correction schemes may be somewhat limited.

Data transmission systems using the Lenkurt - developed Duobinary technique (see *The Lenkurt Demodulator*, February, 1963) provide error detection without the need for adding redundant information. Instead of adding redundant bits to the message which, in effect, lowers the transmission rate, Duobinary coding follows a systematic pattern that provides a more powerful error detection method than the simple parity check.

For business purposes, data systems normally transmit information in sections or blocks. When errors in a block are detected, the system automatically keeps retransmitting the block until there are no errors. For real-time systems, such as high-speed telemetering, where data is obsolete almost immediately after it is received, duplicate or diversity transmission techniques are about the only practical solution to offset the problem of errors.

Error rate may be determined by comparing each and every bit of the received message with the transmitted message. This can be accomplished by either looping-back the received signal to the originating transmitter, or by transmitting a predetermined pattern of data signals known to the receiver. When the data signal is looped back (see Figure 2), a variable delay, corresponding to the transmission time of the data message in both directions, is inserted into the original message before it is applied to a bit-by-bit comparator. By

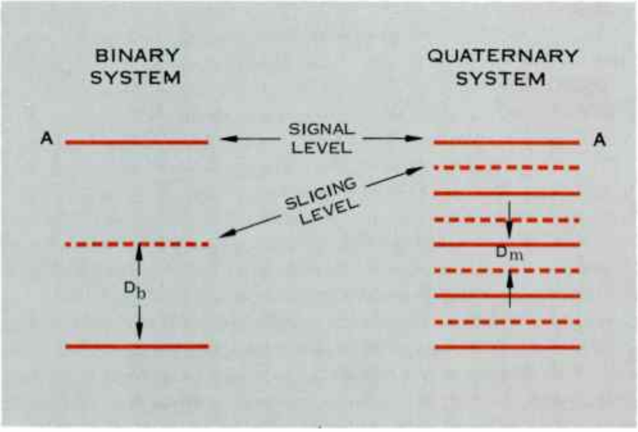


Figure 3. Determining noise penalty compared to binary. Because multilevel systems require more slicing points than binary systems, they are more susceptible to errors caused by noise.

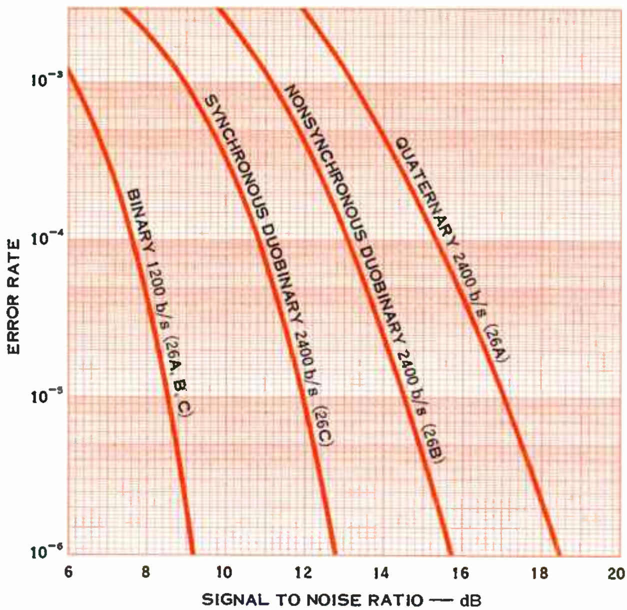


Figure 4. Comparison of error rate versus signal-to-noise ratio. Without normalization synchronous Duobinary has a noise penalty of about 3.4 dB compared to binary. This is a marked increase in performance over nonsynchronous Duobinary and quaternary.

analyzing the transmitted message and the received message together, the comparator signals an electronic counter when errors are present.

Modulation Methods

Selecting a suitable type of modulation is extremely important in the design of data transmission systems to provide simplicity and to achieve optimum performance. The three basic methods of modulation in data transmission are amplitude modulation (AM), frequency modulation (FM), and phase modulation (PM). Frequency modulation of a binary signal results in the shift of a two-state or binary signal about an FM carrier, and thus is usually referred to as frequency shift keying (FSK) rather than FM.

The most commonly used technique is binary FM (or FSK) because this type of signaling offers a good signal-to-noise ratio, and is not affected by changes in amplitude level. However, the signaling speeds achieved with binary modulation are limited and generally inadequate to meet present-day data requirements.

Vestigial sideband AM and techniques using more than two or *multilevel* signal states increase the amount of data that can be transmitted over communications facilities. Quaternary FSK and quaternary phase shift keying (PSK) are examples of multilevel systems which effectively double the data rate compared to binary that can be transmitted in a given bandwidth. But there is a greater sensitivity to noise with both vestigial sideband and multilevel systems, and error performance is usually poorer. Furthermore, distortion in the absence of noise, termed inter-

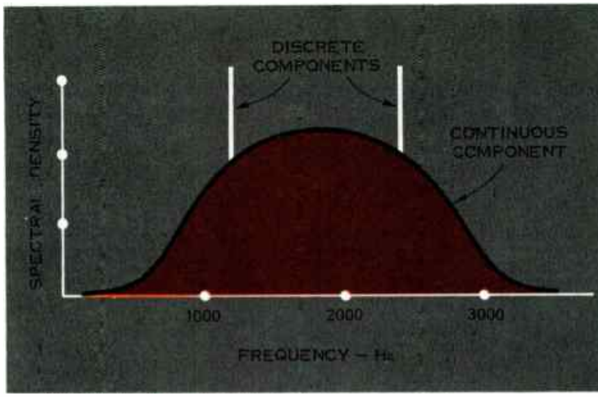


Figure 5. Spectral distribution of a binary FM signal. Discrete components present in the binary signal represent wasted power; these are not present in the synchronous Duobinary signal.

symbol interference, is increased with multilevel signals. This type of distortion is caused by the overlap of positive or negative overshoots of the past pulses into the time slots of other pulses. Experimentation on multilevel systems continues and new approaches have been investigated to improve their performance.

Correlative Technique

Correlative data transmission techniques, particularly the Duobinary principle, have aroused considerable interest because of the method of converting a binary signal into three equidistant levels. This correlative scheme is accomplished in such a manner that the predetermined level depends on past signal history, forming the signal so that it never goes from one level extreme to another in one bit interval.

The most significant property of the Duobinary process is that it affords a two-to-one bandwidth compression relative to binary signaling, or equivalently twice the speed capability in bits per second for a fixed bandwidth. The same speed capability for a multilevel code would normally require four levels, each of which would represent two binary digits.

Noise Penalty

Generating a signal with correlated levels permits overall spectrum shaping as well as individual pulse shaping, thus minimizing intersymbol interference. However, there is a noise penalty with respect to binary systems, and this applies to level-coded correlative systems as well as other multilevel systems. Though exact mathematical calculations of such penalties are usually complicated, there is a quick method of finding an approximate value. In Figure 3, both binary and multilevel representations are shown. Assuming an equal peak voltage of A volts for both cases, the noise penalty is the ratio of the distances between any signal level and the adjacent slicing level for each of the two systems.

The corresponding distances are $Dm = A/2(m - 1)$, where m is the number of levels

and for binary

$$Db = A/2$$

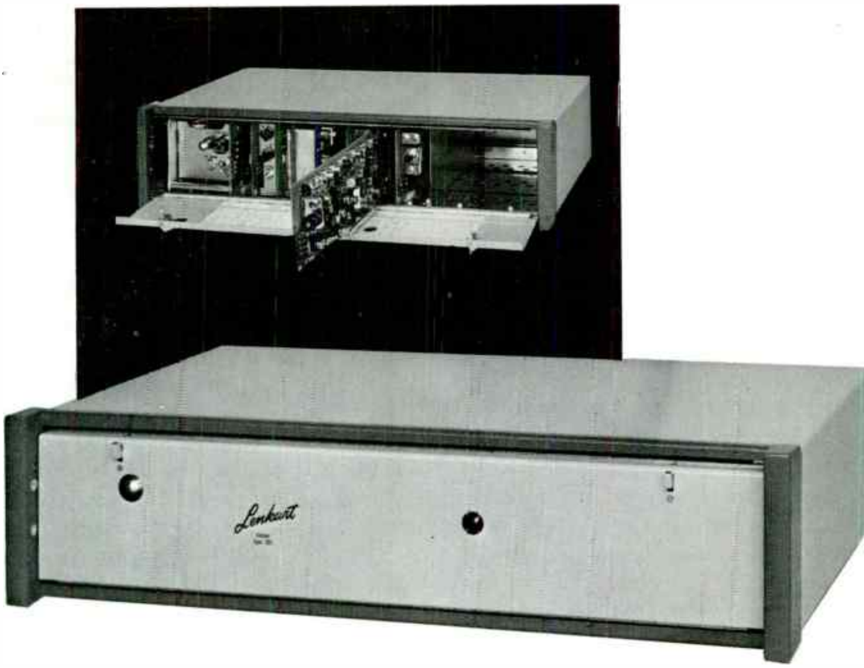


Figure 6. A typical data set shown here is Lenkurt's 26C which transmits serial digital data over standard 3-kHz voice channels at 1200 or 2400 bits per second. Other practical applications include telemetry, digitized voice or facsimile data transmission, and air-to-ground data communications over UHF or VHF radio links. Transmission at 2400 bits per second is achieved through the use of the Lenkurt synchronous Duobinary coding technique. Low error rate coupled with automatic error detection are added features of the Duobinary technique.

The approximate noise penalty (in decibels) for an m-level system relative to a binary system is then

$$20 \log Db/Dm$$

Substituting the values of Db and Dm gives

$$20 \log \frac{A/2}{A/2(m-1)} = 20 \log (m-1)$$

For a binary system, the noise penalty is 0 dB, since binary serves as a reference. The calculation for a quaternary system reveals that the noise penalty is approximately 9.4 dB relative to binary. The Duobinary signal has three levels ($m=3$). Calculating the noise penalty for a Duobinary system from the above formula results in a 6-dB value with respect to binary.

Synchronous Duobinary

A special situation occurs when data transitions are *synchronized* with the carrier phase in FM data transmission at 0°, 180°, or ±90°, as opposed to where the data transitions bear no relationship to the FM carrier phase. It would appear that because there is still the same number of slicing levels, the *synchronized* Duobinary signal has a 6-dB noise penalty. Yet, in practice, it is closer to 3 dB. This situation can be analyzed mathematically, but is more clearly demonstrated by comparative testing of actual working systems under identical conditions.

If error rate is measured as a function of signal-to-noise ratio, individual noise penalties can be determined by subtracting the resultant signal-to-noise ratio of nonsynchronous Duobinary, synchronous Duobinary, and quaternary from the signal-to-noise ratio of a binary system.

It is easy to establish identical conditions to determine the noise penalty by using three systems that have the same peak-to-peak FM deviation, namely, 1200 Hz, which has been proven to be optimum for binary transmission. Three such systems are the Lenkurt 26A, a quaternary system; the Lenkurt 26B, a nonsynchronous Duobinary system; and the Lenkurt 26C, a synchronous Duobinary system.

The conditions for plotting the curves shown in Figure 4 were established by supplying a random binary data signal to each system, introducing white noise with flat weighting over a 3.4-kHz bandwidth, and retiming and sampling the data with a clock derived from data transitions.

For clarity, the signal-to-noise ratio values shown are not *normalized* (put on the same speed basis), since the most significant factor is the performance of Duobinary and quaternary relative to binary rather than the absolute signal-to-noise ratio values. From Figure 4 and Table B, it can be seen that the experimental results and calculated values from the formula *20 log (m - 1)* are relatively similar for quaternary and nonsynchronous Duobinary when compared at an error rate of 10^{-5} (1 error in one hundred thousand bits), with a sufficiently long-time error averaging period.

For synchronous Duobinary, there is only a 3.4-dB noise penalty with respect to binary for random input data. This is because the synchronous system results in a power spectrum that does not have discrete spectral lines as does the nonsynchronous system. (See

TABLE B. Signal-to-noise ratio and noise penalty comparisons with respect to binary of synchronous Duobinary, nonsynchronous Duobinary, and quaternary systems for an error rate of 10^{-5}.

System	S/N Ratio (dB) from Figure 4	Noise Penalty (dB) relative to binary
Binary	8.5	
Synchronous Duobinary (Lenkurt 26C)	11.9	3.4
Nonsynchronous Duobinary (Lenkurt 26B)	14.5	6.0
Quaternary (Lenkurt 26A)	17.0	8.5

Figure 5.) Discrete components are two steady tones (sinusoids) at frequencies of 1200 and 2400 Hz, respectively, and which appear in binary FM. If the ratio of frequency difference in hertz between mark and space to the bit speed in bits per second (or deviation ratio) is unity, the total power is equally divided by the continuous component and the two discrete components. This phenomenon is inherent to binary FM transmission and there is nothing that can be done about it. Unlike this situation, the synchronous Duobinary signal contains only the continuous information carrying component of spectral density, but no discrete components. Consequently, the total power can be increased by 3 dB compared to nonsynchronous Duobinary FM which contains discrete components that do not carry information.

The Future

It is universally recognized that communications is essential at every level of organization. The United States Government utilizes vast communications networks for voice as well as data transmission. Likewise, businesses need communications to carry on their daily operations.

The communications industry has been hard at work to develop systems that will transmit data economically and reliably over both private-line and dial-up telephone circuits. The most ardent trend in data transmission today is toward higher speeds over voice-grade telephone channels. New transmission and equalization techniques now being investigated will soon permit transmitting digital data over telephone channels at speeds of 4800 bits per second or higher.

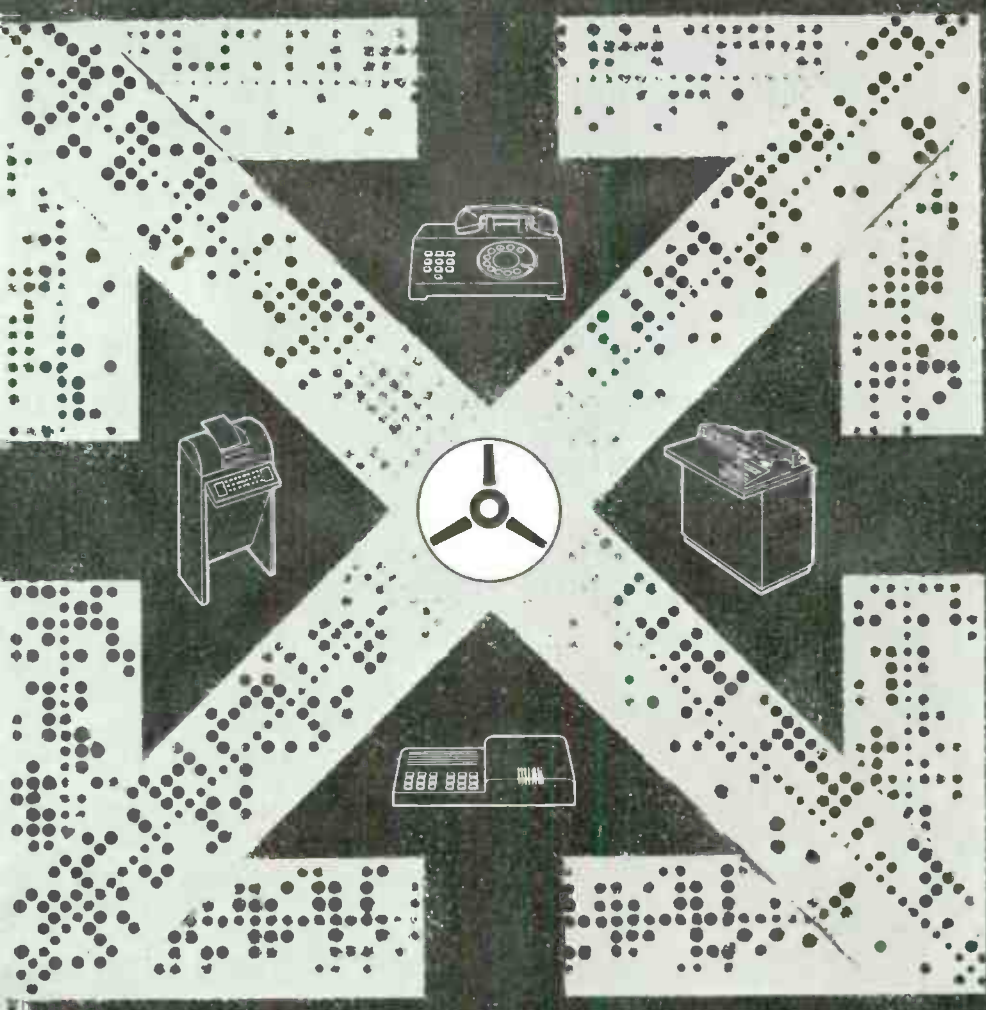
BIBLIOGRAPHY
(Parts 1 and 2)

1. Bennett, W. R., and J. R. Davey, *Data Transmission*, McGraw-Hill, 1965.
2. Lender, A., "Correlative Level Coding for Binary-Data Transmission," *IEEE Spectrum*, Vol. 3 (February, 1966), pp. 104-115.
3. Lender, A., "The Duobinary Technique for High-Speed Data Transmission," *IEEE Transactions on Communications and Electronics*, Vol. 82 (May, 1963), pp. 214-218.
4. Sunde, E. D., "Ideal Binary Pulse Transmission by AM and FM," *Bell System Technical Journal*, Vol. 38 (November, 1959), pp. 1357-1426.
5. "Data Communications," *Telephony*, First through Eleventh of a series (January 8 to July 23, 1966).

AUGUST, 1967
VOL. 15, NO. 8

The Lenkurt. DEMODULATOR

USES OF LOW SPEED DATA



With the growth of knowledge and the expansion of both business and government has come a need to provide timely and effective communications to widely separated but integrated organizations. The most obvious way to meet this need has been to increase communications speeds and expand the capacities of existing communications systems.

In this article, the Demodulator *explores some of the ways that low speed data transmission is expanding communications capacity.*

Explosions occupy an increasing amount of space in today's literature. Populations are exploding and information, aided and abetted by stories about itself, continues to explode.

At present rates, the volume of information is doubling every five years. A few years ago this took ten years and a century ago the body of knowledge available doubled at the leisurely pace of once every 50 years. Needless to say, the effective dissemination and use of all the facts and fancies coming into circulation today poses a staggering problem.

Perhaps the most promising solution to this problem is data communications—the use of machines and machine languages to transfer information from place to place by electrical means. Although the present tendency is to define telegraph and data communications separately, the two have similar transmission requirements and for the purposes of this discussion will be included under the general heading of data communications.

A typical data communications system consists of an input, modulator, transmission link, demodulator, and data sink. The input can be anything from a highly sophisticated computer to a simple business machine. The data sink might be nothing more than an output device. More often, the data sink is a depository, such as the memory bank of a computer, for the data being communicated.

The modulator and demodulator, often referred to as data sets, interface the input or output equipment with the transmission link. It is the function of the modulator to make input signals suitable for transmission. The demodulator converts the transmitted signals to their original form before sending them to an output device.

Two Categories

By using different combinations of input devices and data sinks, a remarkably comprehensive data communications system can be evolved. In general most systems fall into one of two cate-

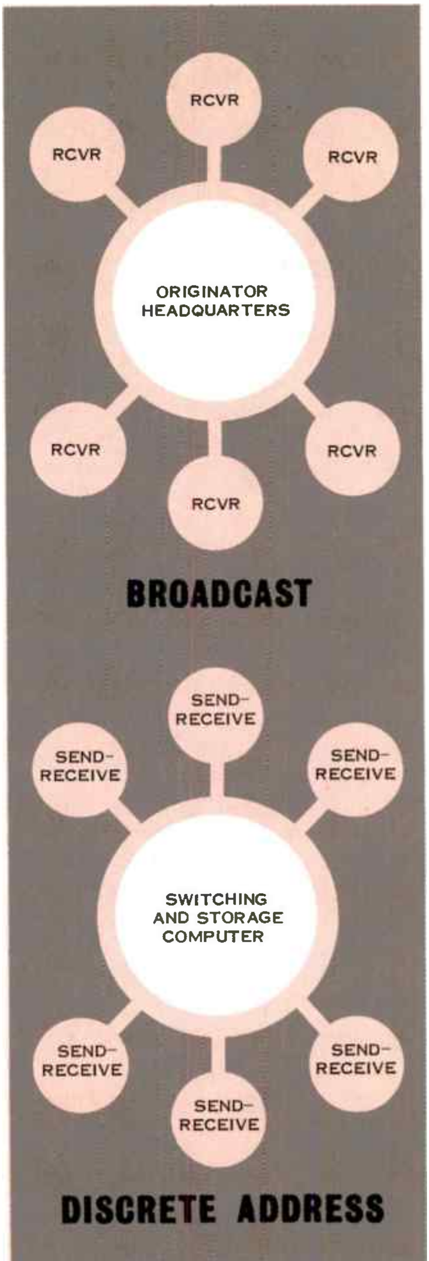


Figure 1. Data communications systems fall into one of two general categories. One is essentially a receive-only system while the other has stations with a send and receive capability.

gories. The simplest employs a number of receive-only terminals connected to a single data source. It is often compared to a radio broadcast. This procedure serves mainly to distribute information which is perishable—information intended to be used only once and then immediately.

Another procedure uses discrete addresses. The majority of the stations in this system have a send as well as a receive capability, although there are monitoring systems with send-only remote terminals which update computer information. Each remote terminal can be called by the use of addresses and any remote terminal can address other terminals in the system. More often than not this configuration uses a data sink to store information for future use.

As might be expected, data communications systems can be flexible enough to integrate both methods in a single system. For example, the broadcast method can be designed so that more than one station on the network can originate traffic to all or some of the subscribers. On the other hand the discrete address method can, by using preassigned address calls, broadcast information to several subscribers simultaneously.

New Economies

Recent changes in communications regulations have made low speed data communications more economical. Many systems use leased line circuits for transmission. These lines have been leased for specific purposes and, in the past, have had to terminate in telephone company interface equipment. Now a new tariff filed with the Federal Communications Commission makes it possible for a customer to install his own data set to interface on certain leased lines.

This change has given users the option of choosing interface equipment

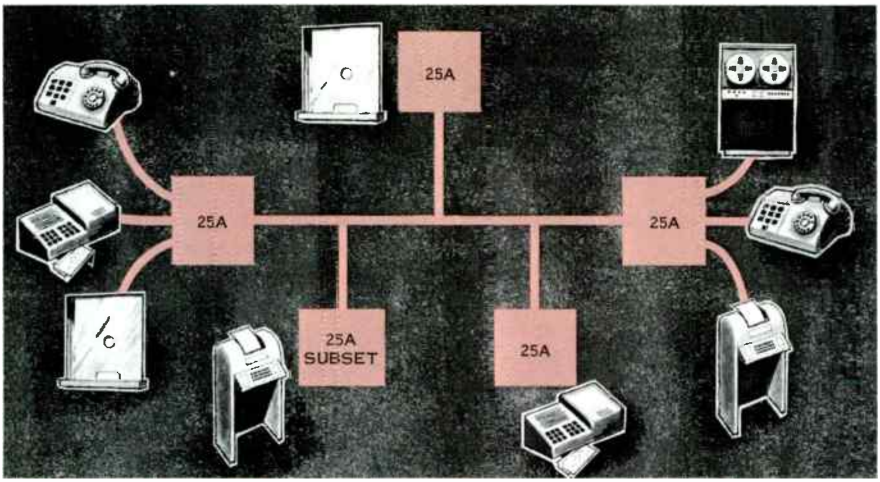


Figure 2. The Lenkurt 25A data transmission system makes it possible to use one voice circuit for a number of data channels. The 25A operates between 75 b/s and 200 b/s with a maximum of 25 channels per voice circuit.

which makes the most efficient use of a leased line. Individual users can now employ equipment which derives a number of channels from a single, leased voice-frequency circuit. From a user standpoint this enables him to satisfy several requirements—depending on the characteristics of his system—with only one voice circuit.

Speed of transmission is one system characteristic which influences the number of channels which can be derived from a single circuit. Direct computer-to-computer operations, for instance, run at extremely high speeds. Computers talking to each other under the most ideal circumstances can exchange bulk information in nanoseconds (a nanosecond is one billionth of a second). But high speed exchange requires wideband transmission paths which minimize impulse noise and delay distortion. Such considerations raise the cost of leased services.

Lower transmission speeds can use narrower bandwidths and less ideal circuit conditions. Data at rates up to 9600 bits per second has been carried on an equalized voice circuit. With even lower speed input equipment, sharp filtering can split a voice-frequency circuit into a number of separate channels. With the remote ends of each channel serving a different terminal, the over-all traffic volume can be increased without increasing circuit costs.

Increasing Efficiency

Making the most efficient use of data communications is not merely a matter of getting the maximum number of channels on a voice circuit. The computer must be protected against having to wait for all data from one channel before calling for inputs from another channel. With data coming at 200 bits per second or less, for instance, a computer capable of receiving information at 1.5+ million bits per second would be wasting time. On the other hand, allowing only one terminal at a time to have access to a computer would in effect turn a multi-channel voice circuit into a single channel circuit. To avoid such inefficiency each communications channel feeding a computer normally terminates in a buffer unit which uses store and forward techniques to accumulate data from many sources and send it to the computer at higher and more suitable speeds. The buffer acts as an intermediate storage device which compensates for the difference in the rate of flow between input and computer.

The buffer is not an essential component of a data transmission system. With it, however, a large number of remote terminals can communicate with a single computer. The buffer collects incoming data at relatively low bit rates and literally spits groups of bits into the computer in hunks called "bytes".

Accuracy

Of course, moving data at high speeds is of no avail if the material is not received accurately. This is especially critical in a system which passes large amounts of numerical or encoded data. With numbers and random letters an end user has a much more difficult task interpreting garbles than he would if plain language were being received. Obviously, data which goes directly into a computer or other business machine must be essentially error free.

Some sources of error, such as those introduced by malfunctioning equipment or operating personnel, are easily identified and can be corrected—at least temporarily. Others, which also affect the accuracy of a transmission, are not so easily dealt with. These depend on a number of variables including speed of transmission, noise and propagation characteristics.

Low speed data transmissions tend to minimize these sources of error because the signals exist longer and can more readily withstand disrupting influences. Although any given system will exhibit its own peculiarities regarding error rates, one test over leased lines found an error rate of from one to eight characters per 100,000 transmitted. At this rate an error occurred as often as every half hour and as infrequently as once every four hours. In terms of a 100 word per minute system, 144,000 characters could be sent between errors.

Nationwide

Both the Associated Press and United Press International are installing Lenkurt 25A data transmission systems to take advantage of the vast voice-frequency network already available in the United States. The 25A can derive as many as 25 channels from one voice circuit. This gives each wire service its own private broadcast network. To provide, say, 20 different news services a wire service only needs one voice circuit. Two channels might carry hard news, another feature stories, a fourth general sports news, and a fifth horse racing news. Other channels might have financial news, state news or special services such as news edited and processed prior to transmission so that a tape received along with the teletypewriter copy can be used to set type. Finally, other channels might be reserved for communications between bureau offices and wire service headquarters.

The ordinary news service subscriber, either with a data subset at his terminal or via local service from a data interface point at the wire service bureau office, receives only the channels contracted for. On the other hand the bureau office, in addition to terminating the desired news channels, can originate news traffic on certain channels. With a channel of its own, a bureau can provide the wire service with fast coverage of local events. For example, a flood story in the Midwest can be sent by the nearest bureau office to the news service head-

Figure 3. Lenkurt 25A data transmission subsets are being used by news service subscribers. Each subset is tuned and set to the channel contracted for by the subscriber.

quarters and to all subscribers on that channel.

Certain designated offices can block local delivery of channels on the backbone trunk. By blocking local delivery a bureau gains a free channel for broadcasting local news to regional subscribers. A bureau in the Southwest, then, can originate its own broadcast of news about the Southwest for local subscribers.

Wall St.

The nationwide brokerage firm of Paine, Webber, Jackson & Curtis is using Lenkurt 25A equipment to connect its 51 offices with a computer in New York City. The resulting data communications system represents a non-broadcast or discrete address network. As with the wire services, voice-frequency circuits serve as the backbone for a system which makes it possible for branch offices to feed data directly to a computer which acts as a message switching system.

Since installing its data communications system, Paine, Webber has more than doubled its communications capacity and has reduced the handling time of an execution order by as much as 20 or 25 minutes. With the new system an order for the New York or American stock exchange travels over one of the 22 two-way teletypewriter channels to the company's computer where it is directed to the correct booth on the exchange floor. The data communications system gives branch offices access to the firm's over-the-counter trading, institutional, bond, research, and underwriting departments. Copies of an order and its execution are also printed in the firm's bookkeeping office to speed billing and crediting customers' accounts. The entire transaction is also recorded on magnetic tape for record-keeping purposes.

Although the system is oriented toward handling orders, conceivably it could be put to such other uses as price quotations, reporting customer accounts or handling other information which is needed on a real time basis. Essentially it is an on-line, real time system. Each remote user is on-line with the computer continuously, guaranteeing an immediate or real time response from the computer. Where data must be processed outside the system, as would be the case on the stock exchange floor, response time will depend on the quickness of floor traders in handling an order.

Near and Far

Banks are making use of low speed, on-line data communications systems to link their branch offices to a central computer containing balance information on all customers' accounts. If a customer wishes to cash a check at the bank, a teller can query the computer on a push button intercom and receive a verbal response from the computer in about 20 seconds. By using a computer the bank has a single information retrieval system for all its branch offices.

A national radio-television network is using an information retrieval system to determine instantly the status of the air time it has for sale. By querying a computer any salesman can find out what air time is available. In addition, the network management has on call an up-to-the-minute report on all time sales. Of course, data communications also provides a means of keeping the computer updated on the status of sales.

Data communications is also helping create firms which offer computer services to a number of subscribers. This timesharing concept gives various unrelated organizations, many of which would not lease their own, a computer 24 hours a day. Each user has his own code, can store information, and can run his own programs. Several colleges offer timeshare services to other colleges and business firms as well as to their own students and faculty. Some companies and computer manufacturers are also sharing computer time with outside users.

Tymshare, Incorporated in California sells time on its computer to users as close as a local phone call and as far away as Colorado and New Mexico. At present all customers—remote and local—have access to the computer through data sets transmitting directly to the computer over ordinary dialed telephone circuits. In the future, however, remote users will be able to communicate with the computer by dialing a local number which interfaces with a leased line. The leased line, terminating in Lenkurt 25A equipment, will serve

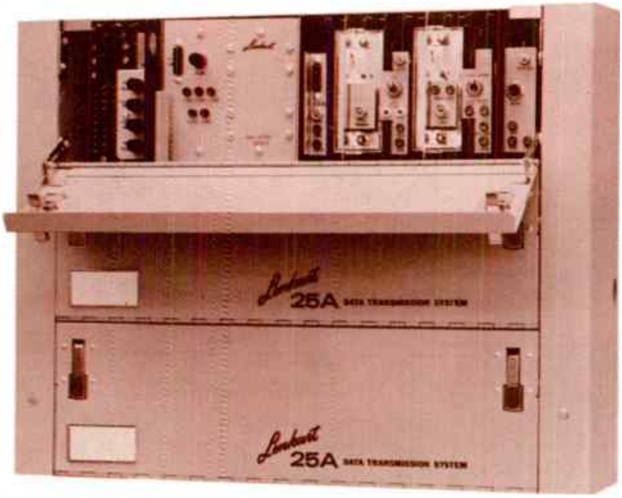


Figure 4. The Lenkurt 25A is an economical, highly reliable data transmission system which uses sharp filtering to transmit up to 50% more information than conventional systems.

as a trunk, able to handle several calls at once. This timesharing application allows users to run engineering and scientific computations or develop programs on a real time, on-line computer. It provides customers with simultaneous access to a computer with a maximum response delay of three seconds.

The Railroads

Railroads, many of which own their own communications networks, are one of the oldest users of data communications. By installing data transmission systems railroads have been able to increase their communications capacity at minimal cost. The increased capacity has in turn made it possible for the railroads to account for every car and train in their system on a real time basis.

Each car is represented by a data card and each train is represented by a stack of these cards—one for each car plus cards for the locomotive and caboose. As a car moves from yard to yard, its card, along with the other cards for cars in the train, is transmitted over data circuits to the next yard. This gives personnel at the next yard timely information for handling the incoming train.

At the same time this information is relayed to a computer to update its record of each shipment in progress. The computer, accessible from many sources by data communications, allows the railroad to make immediate replies to inquiries about a shipment.

Freight offices which book shipments are linked by teletypewriter to the communications system so that a shipping order can be processed immediately. A single message from the freight office provides the accounting department with billing information, the freight department with car requirements and the computer memory with a record of the transaction. From that point on the progress of the shipping order and the subsequent shipment can be recorded by relatively short inputs direct to the computer.

Railroads also use data for administrative traffic and train safety. One line has centralized its payroll accounting by sending timecard information over its communications system directly to a central office. At present the checks are drawn up and mailed back to the station submitting the timecards, although the processed information could be returned on a data circuit and printed directly onto check blanks.

For train safety, railroads are using a specially developed Lenkurt 960A journal data transmission system. This unit converts analog sensor signals received from heat detectors to tones which can be transmitted over standard communications facilities. Overheated journals have been cited as contributing to railroad accidents costing millions of dollars annually. The 960A, linked to a transmission system, is helping to reduce the number of these accidents.

Immediate Response

Airlines use on-line, low speed data communications to handle their passenger reservations, freight operations and aircraft maintenance schedules. During non-peak hours the airlines use the same communications to handle bulk jobs such as payroll and inventory. Pipeline and utility companies have joined their far flung operations with data communications also. The pipeline companies monitor such operational details as flow rate, pump pressure, suction, and fluid viscosity. Utility companies exercise load control by using data communications to connect substations and switching points. In both cases distant operations can be adjusted in response to variations as they occur.

Large business concerns which would not be immediately associated with a need for real time communications are using data communications to enhance

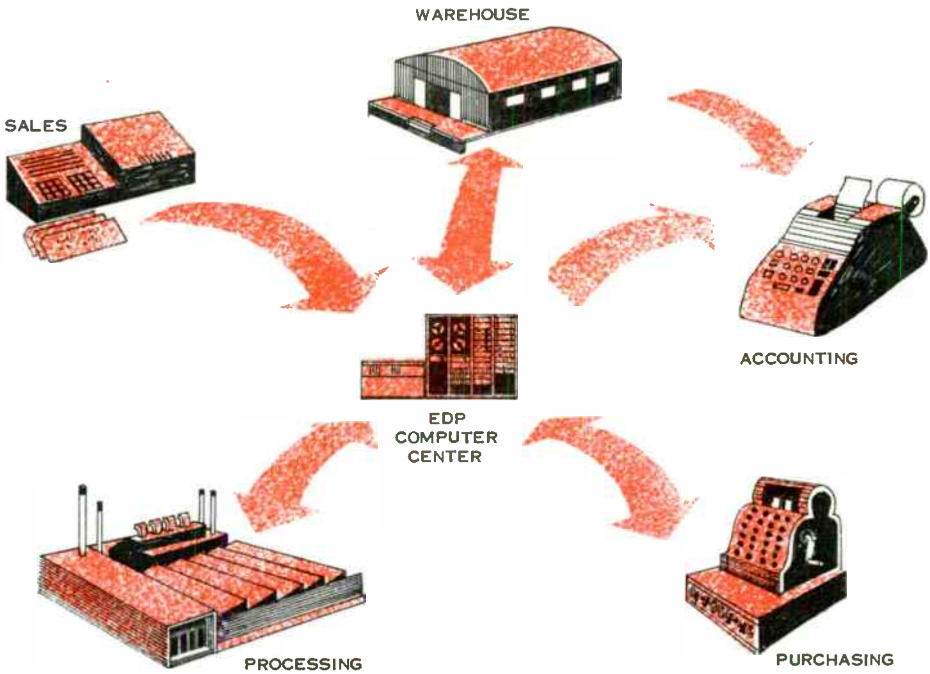


Figure 5. Management information systems can use data communications for their sales offices, manufacturing plants, and planning and support departments.

management information systems. Leased voice-frequency circuits connect central computers with warehouses, sales offices, processing plants, and accounting and purchasing departments.

With such an installation, a salesman booking an order at one of the sales offices sends that order via a data channel to the firm's computer. The computer determines which warehouse is best situated to fill the order and routes it there. At the same time the computer informs accounting that the customer's account should be billed for the item. If necessary and if programmed to do so, the computer will also inform production and purchasing of the order so that these departments can take action. If necessary, the warehouse can update the computer's inventory record.

Orders by data communications gain time—they can be filled within a few hours. Users claim that the time between writing and delivering an order can be less than twenty-four hours.

Supermarket chains with their own warehouses have used this method for restocking and inventory control. A large company which supplies consumer items to gas stations claims that orders placed by mid-afternoon can usually be on their way to the customer the next morning. Even at the slowest transmission rates information gets from place to place much faster than by conventional means. One company found that using data communications to process its daily sales spread the work load evenly over the week. Before switching to a data system, over 40% of its orders arrived on Monday and all orders spent at least two days in the mail.

Data communications can bring widely separated parts of an organization into closer coordination. Within seconds an order can be available at any number of user stations. If a computer is part of the system and is used for record keeping, the order is on file instantaneously. At the same time the computer can perform whatever accounting computations it is programmed to do.

Operating data from outlying points can be made available daily. Management then has timely and accurate information on which to base production schedules and sales promotion plans. Management also has for its immediate use information that can influence the purchase of basic commodities, the location of plants, and the regulation of cash flow.

Such interwoven systems reduce errors and paper work. A shipping invoice on a customer's order need be written only once. After that, all copies are either transmitted directly to users or are on call from a computer memory. In addition, records stored in the computer are updated automatically by direct inputs—no formal reports to be hand delivered by messenger.

More to Come

These examples only touch the more prosaic possibilities of data communications. What lies ahead depends more on man's imagination than on his technology. There has been speculation that houses might have a home learning center connected to a teacher-computer by data links.

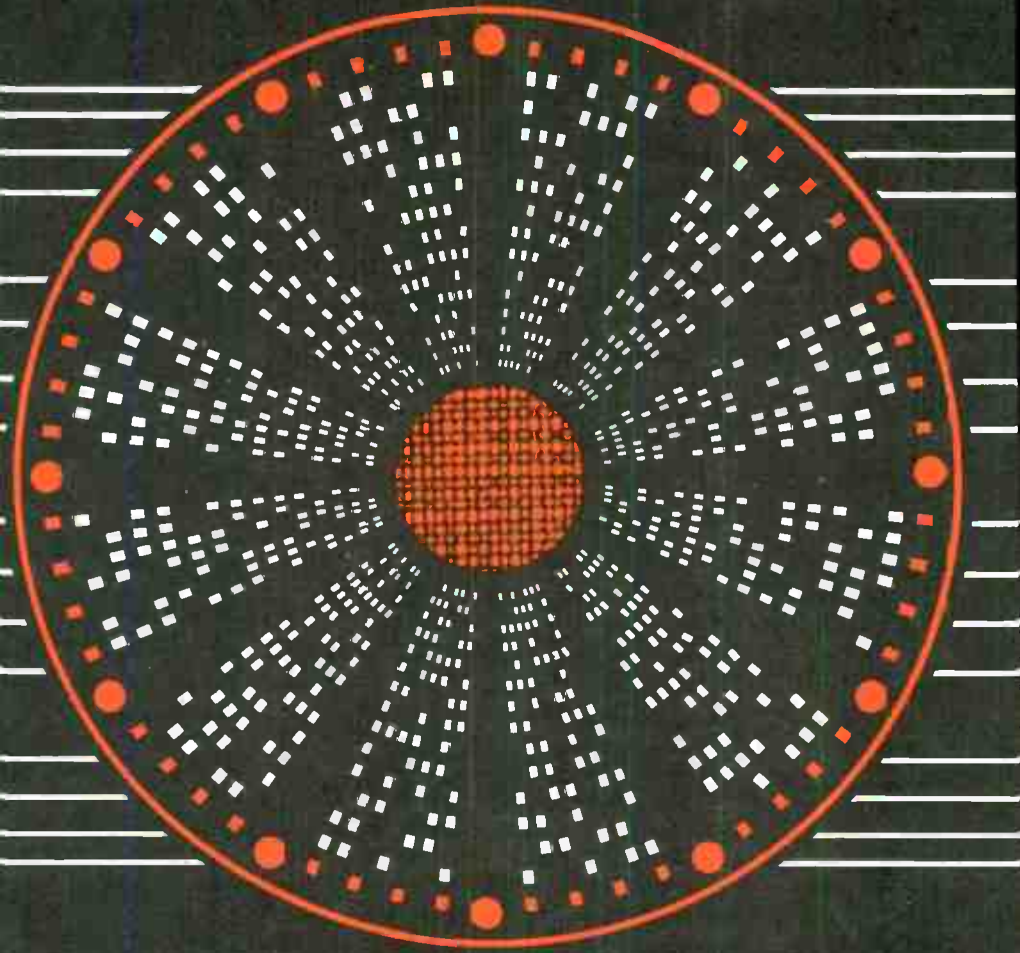
A few public utility companies are testing data systems which, by sending information directly to the billing office, could replace meter reading. At the same time these real time readouts could warn a control center of potential system overloads.

A data input device interconnecting retail stores with banks could eliminate money. On payday an employee's bank account would be credited by a deposit from his employer. The employee could draw on that deposit by inserting his personal money-card into a remote data terminal. The remote terminal would transmit information from the money-card, along with the amount required for the transaction, to the bank. The bank in turn would compare that amount with the balance of the account in question and, if there were sufficient funds, okay the transaction and debit the account.

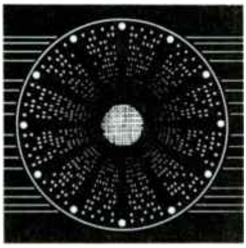
Coming even sooner than the money card scheme is a data communications system to give customers immediate access to hundreds of theater and sporting event tickets. By using a special terminal device in supermarkets and banks a customer can select tickets to a ballet in London, a football game in New York, and a play in Los Angeles within a few seconds. The customer's order is transmitted from the supermarket to a computer which searches for the tickets to the event and in the price range and area desired. When the seats are found, the computer causes the remote terminal to print the tickets for the customer who pays for them at the supermarket .

Data communications is bringing widely spread operations into more immediate contact and in so doing increasing the coordination between separate parts of large organizations. It is also keeping the right people informed in spite of ever increasing amounts of information. Hopefully, it will prove to be an effective solution to the problem of handling the exploding amounts of information becoming available.

The Lenkurt®
MAY 1968
DEMODULATOR



Computer Time-Sharing



Telecommunications and computers join forces to provide a new service

Until quite recently, only very large organizations — commercial, educational, or governmental — could afford the services of a high-speed computer. Now, however, thanks to the growth of time-sharing, even the smallest private concerns can have access to these multimillion dollar machines.

This revolutionary concept in the field of electronic data processing extends the use of large, general purpose, digital computers into hitherto unfamiliar territory. By serving 20, 40, or even 200 on-line users simultaneously, time-share computers are finding application in virtually every area of commerce.

Computer time-sharing was pioneered at Dartmouth University and the Massachusetts Institute of Technology in the early 1960's; since that time it has mushroomed into an international business. At latest count there were more than 70 time-share companies providing service to some 5000 on-line customers. New systems are entering the market at the rate of three or four a month.

Many of the new, commercial time-share systems began as in-house service functions for large computer manufacturers. Once the effectiveness of the concept was proven within their own organizations, the manufacturers saw the market potential of time-sharing and a new customer service was born.

Teleprocessing Essential

As a concept, time-sharing has existed for a long time. But the reality — the technology itself — is quite recent. Two major developments have contributed largely to making time-share a reality. The state of the computer art itself has advanced to and beyond the third generation. Integrated circuits have made possible speeds, capacities, and program sizes sufficient to allow a single computer to serve large numbers of users simultaneously.

The other major development is not in the computer industry proper, but in telecommunications technology; it is teleprocessing, the ability to transmit data over voice circuits. For time-sharing to have more than narrow geographical applications, there must be a means of communication between computer centers and subscriber terminals. Since they already exist, telephone circuits are the obvious media for the task.

Digital data signals, however, require some conditioning before they can be transmitted over voice circuits. This is done with data modems. Some are interfaced with regular telephone service, and others use dedicated circuits.

In cases where either traffic load, distance, or both make the use of regular switched service prohibitively expensive, many time-share subscribers will find it more feasible to use data modems and dedicated circuits. Lenkurt's 25A and 26C data sets are designed for just such applications. Depending upon bit rate requirements, the 25A can accommodate from 7 to 25 time-share terminals. The 26C, with its 2400 b/s rate, is designed for high speed transmission such as that required by graphic display units or in computer-to-computer links.

Time-Share Spectrum

Present time-sharing systems range in size and scope from relatively small in-house arrangements (12 to 16 users) to sprawling, commercial networks

```
USER NUMBER--P63000
SYSTEM--BASIC
NEW OR OLD--NEW
NEW PROBLEM NAME--STEM
READY.

100 LET X1=6*15
110 LET X2=2*10
120 LET X3=3*0.5
130 LET X4=12*(-0.02)
140 LET T=X1+X2+X3+X4
150 PRINT T
999 END

RUN
WAIT.


STEM        14:13

 111.26


TIME:    0 SECS.


BYE

*** OFF AT 14:14
```

Figure 1. Purely for purposes of illustration, the user's side of the conversation with the computer is printed in red. Note that the line numbers increase by tens rather than singly. This is so that the user may go back and make additions or corrections without disrupting the logic of the program.

serving up to 200 on-line terminals. Although there are many different "kinds" of time-share systems in the current market, most have similar features.

Externally, the computer systems used for time-sharing applications have little in common with the batch-processing systems familiar to most office workers. Rather than using stacks of punch cards or tapes, time-sharing operates on a real-time basis and is characterized by one-to-one interaction between man and machine.

Most commercial time-share computers are essentially problem-solving devices. Consequently, the majority of users – at the present state of the art – are scientists, engineers and other similarly oriented groups whose primary concern is the immediate solution of specific problems.

Among the myriad problems for which computational time-share systems might be used are mathematical problems, electronic circuit design, complex chemistry formulas, market analyses, banking and interest rates, and even precision tool design problems.

In addition to the problem-oriented, real-time computers, other time-share systems emphasize information storage and data retrieval; while still others are of the type designed for remote, on-line batch processing.

Probably the most striking feature of time-sharing is that regardless of the number of terminals, service to all users is immediate and simultaneous, although the simultaneity is actually only apparent. This is due to the disparity between computer speeds – measured in microseconds or even nanoseconds – and the relatively low speeds of the input/output hardware.

A computer is an electronic device, whereas terminals (usually teletypewriters or line printers) are electromechanical and operate much more slowly. Hence, a computer is able to switch from terminal to terminal and swap programs so quickly as to give each user the illusion that he alone is communicating with the machine. In a network using teletypewriter terminal devices, the computer is actually able to switch from one to another during the time an operator takes to move his finger from one key to another.

The physical arrangement of most time-share systems is similar to a star broadcasting network. This is illustrated by the diagram in Figure 2. The smaller systems are usually characterized by a small Central Processing Unit (CPU) with a core memory of 14 to 16 thousand characters. These computers may have as many as 16 terminal ports and find most of their application within a single organization.

For example, a large company might have several home-office departments as well as some in the field. In this case, even the in-house arrangement requires use of a transmission means to link together the various field offices and their headquarters. Using this smaller time-share system, all of a company's offices can have access to one computer – hence, to the same files and records.

At the other end of the time-share spectrum there are systems whose CPU's can accommodate 40 to 50 users (one newly developed system serves 200) spread over a broad geographic area. In these large arrangements, program sizes may go as high as 40 thousand characters with core storage rated at 264 thousand 36-bit words. Moreover, they incorporate huge multi-access libraries of stored data, have loop functions capable of virtually infinite repetitions of mathematical problems, as well as stored mathematical routines.

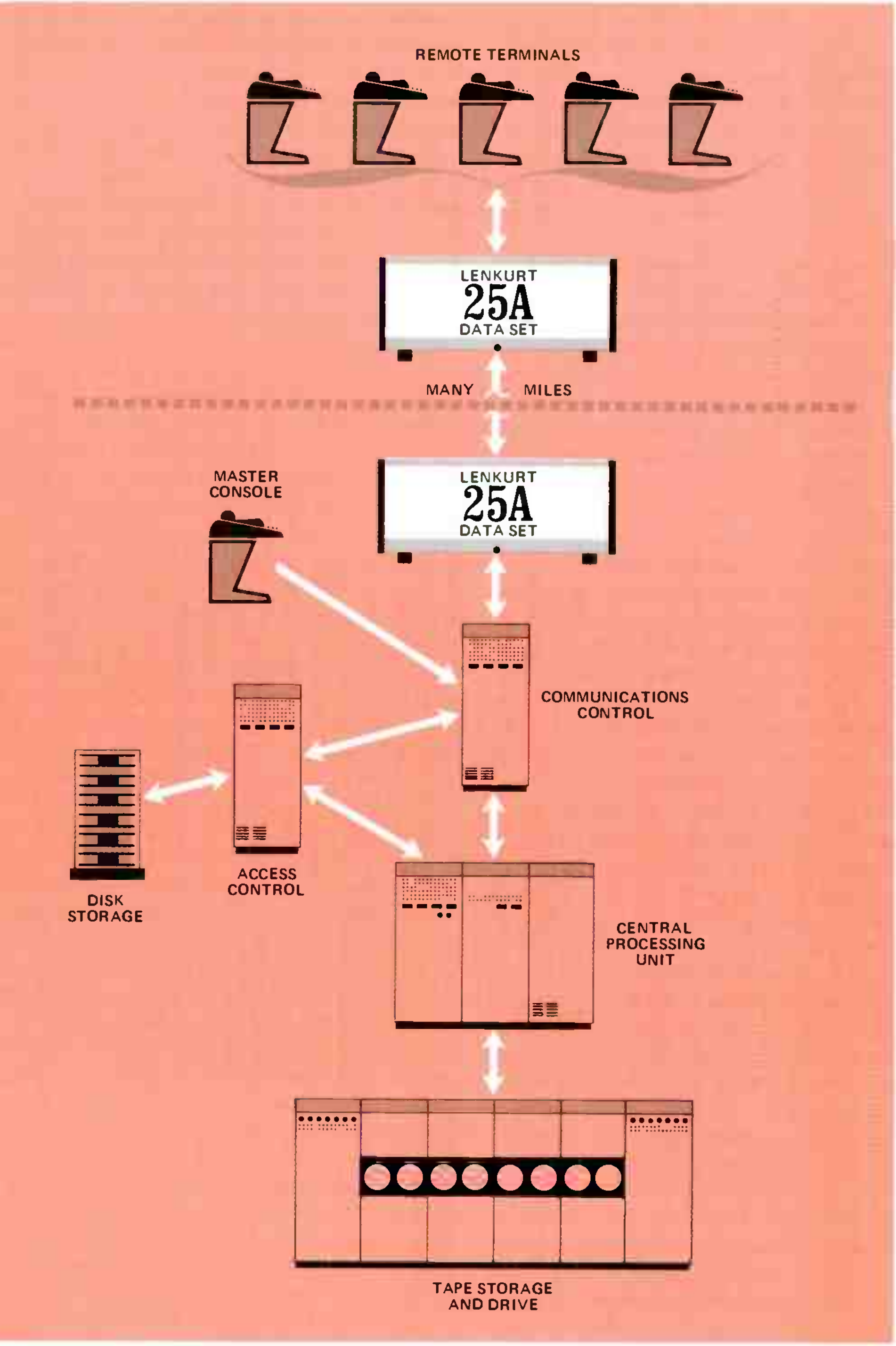
Figure 2. A data processing system such as this one is specifically designed for time-sharing service.

Through the use of Lenkurt 25A Data Transmission Sets or similar modems, the subscriber terminals are able to communicate with the central system. A switching system in the Communications Controller acts as a "traffic cop" and allows access to the system on some prearranged priority basis. This is the system's executive function.

A computer in itself, the Communications Controller also processes stored programs – tape or disk – and controls the remote terminals.

All the computational problems for which the system is programmed are done by the Central Processing Unit (CPU). This is the workhorse of the entire system. It compiles programs, does floating point arithmetic problems, works out matrices, and any number of other mathematical problems. Depending upon frequency of use, need, and immediacy, programs are stored in either the Disk Storage Unit or on magnetic tapes. Storage in the DSU is on oxide-coated, magnetic disks. Access to both storage areas is controlled by the Dual Access Controller. Usually access is by address code.

Each subscriber to a system is assigned a personal address code. This is for purposes of recognition by the computer, billing, and program security. Stored programs are identified by the same personal address code and are accessible only to the proper user.


REMOTE TERMINALS
LENKURT
25A
DATA SET
MANY
MILES
MASTER
CONSOLE
LENKURT
25A
DATA SET
COMMUNICATIONS
CONTROL
ACCESS
CONTROL
DISK
STORAGE
CENTRAL
PROCESSING
UNIT
TAPE STORAGE
AND DRIVE

Time

Obviously, the whole basis for the time-sharing concept is time. But there is a wide variety of "times" associated with computers – turnaround time, real time, swap time, switching time, and so on.

Real time is applied to a working situation where operator and computer work together concurrently. Swap time is that brief interval required for a computer to transfer data back and forth between primary (buffer) and secondary (core) storage. Switching time is the period during which a computer is switching from one user to another and back again. Often this is the time elapsing between separate key punches on a teletypewriter.

Basically, the time which is shared is the computer's actual operating time. This is, of course, different from terminal time. Typically, a subscriber might be operating his terminal for 30 or 40 minutes while using only 15 seconds of computer time.

Multilingual Machines

To the layman, one of the most intimidating aspects of computer communications is the confusing variety of computer languages currently in use. They bear such mystifying labels as FORTRAN, COBOL, SNOBOL, LISP, and even HELP – to name only a few of the more exotic. While these may be familiar to veteran data processors, they are alien to the average businessman or engineer. This situation has been alleviated in the time-share world by the introduction of much simpler computer languages. They are quickly learned, highly comprehensible, and easy to use. For general use, most time-share systems are programmed in BASIC (Beginners All-Purpose Symbolic Instruction Code). Scientists and engineers on the other hand usually require a more specialized vocabularly than found in BASIC. Consequently they may use ALGOL (Algorithmic Language), a simplified mathematical language specifically designed for solving scientific and engineering problems.

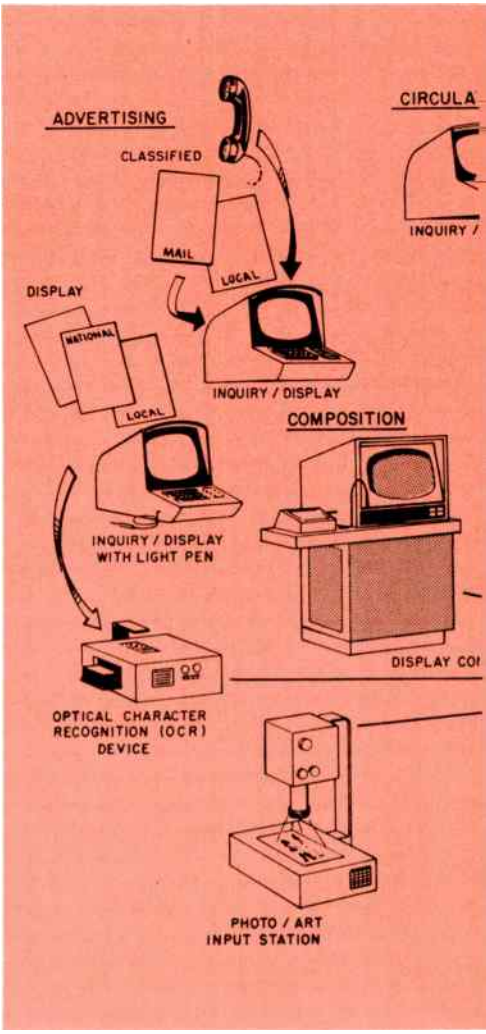
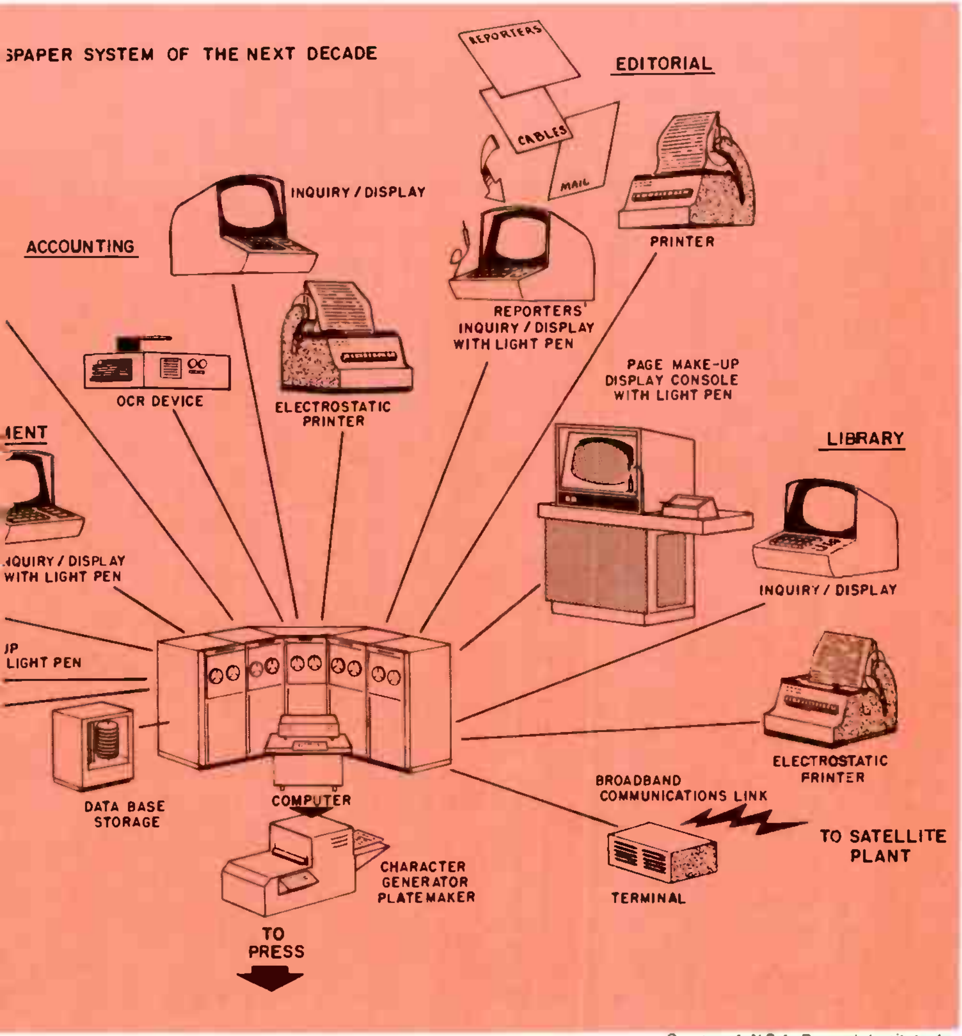


Figure 3. Though still in the future, time-share applications such as this computerized newspaper currently are being given serious consideration.

Using these languages, the time-share subscriber becomes his own programmer. This fact alone points up one of the advantages over batch processing systems. Obviously, if companies have their own computers, they are going to have to bear the cost of staffing the data processing section – clerks, programmers, analysts, and management. Often, this cost savings alone is enough to justify subscribing to time-share. Another advantage is the curtailment or elimination of turnaround time. In computer jargon, this is the elapsed time between input and output – that is, query and reply or problem and solution. In companies possessing their own data service, this time is virtually always measured in hours – and sometimes days or even weeks.

The person with the problem – say an engineer – must first translate it into terms acceptable to the computer.



Courtesy A.N.P.A. Research Institute, Inc.

He then takes it to a programmer who transposes the problem into the appropriate computer language and format for input to the computer. Once this is done there is then a period of waiting in line (called queuing time) if the computer is busy. Since many programs contain errors or "bugs", debugging time is also a factor.

Time-sharing eliminates all these steps. The subscriber simply dials up the computer and is immediately on-line. Programming is automatic, as is debugging; when an unacceptable program statement is fed into the computer, it automatically tells the operator so that he may correct the error immediately. Experience has proven it helpful for the time-share subscriber to rough out his programs before going on-line. A simplified sample program in BASIC is shown in Figure 1.

Some Limitations

While there are indeed many advantages in computer time-sharing as reflected by its growth, it is unrealistic at the present time to talk of a universal computer utility – that is, a utility in the sense that gas, power and water companies are public utilities. However, this concept may not be too far in the future.

At its present stage of development, time-sharing is not readily adaptable for multi-access batch processing of the type required for such things as payrolling, bookkeeping, inventory and the like. Nor are present generation time-share computers capable of the electronic "brain" functions often ascribed to computers. Obviously, long computer runs, as required by a payroll for instance, are not economically feasible for time share operation. Unless, of course, data modems are used in conjunction with flat-rate leased lines or dedicated circuits, and toll charges are no longer a factor.

Currently the bulk of time share subscribers falls into one of several categories – scientists, engineers, educators, and other researchers. Recently, though, more and more ordinary businessmen are finding applications for computers within their fields of interest.

Illustrative of this broadening trend in computer uses are some of the more highly specialized areas into which time-share has penetrated.

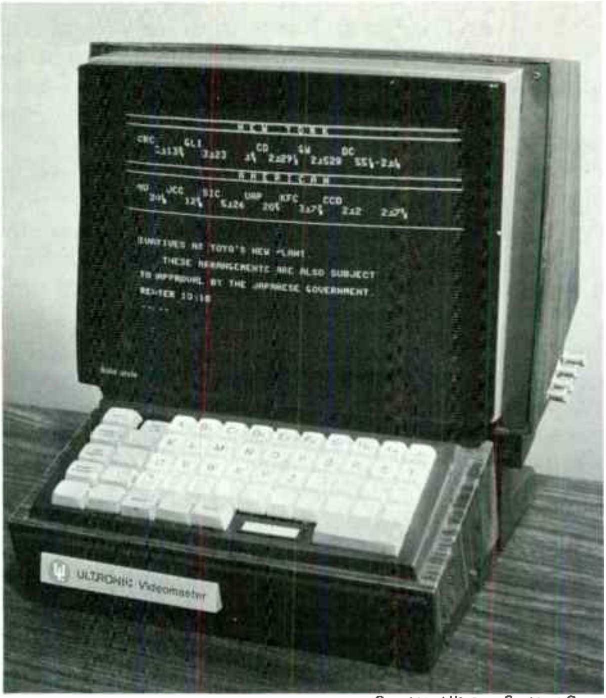
Unusual Applications

One of the most fascinating areas in which computers are finding new uses is in the publishing business – particularly newspapers. In the near future, the entire mechanical process of producing a daily paper may be aided by a computer. The chart in Figure 3 illustrates one way in which this might work as a new in-house function.

At present, only relatively simple tasks such as character generation and composition are being done electronically. However, some newspaper publishers are seriously studying the practicality of totally computerized publishing.

In such a system, a reporter would only have to transmit his story to the computer from his desk; the editor or rewrite man, in turn, edits the material on cathode ray tube (CRT) graphic display units making his changes with an electronic light pen or similar device. Other tasks such as accounting, classified ad placement, space sales, and layout also readily lend themselves to computerization.

Another highly specialized application of computer time-sharing is in the stock market. Ultronics Systems Corporation has developed and marketed a computerized stock quotation service. The system features a network of electronic graphic display units (Figure 4) connected to intermediate, regional



Courtesy Ultronic Systems Corp.

Figure 4. This stock quotation display is typical of the current trend toward graphic display units for time-share terminal devices.

computer systems; these in turn are tied to a central time-share system near the New York Stock Exchange via telephone lines. By simply querying the intermediate computer through his teletypewriter keyboard, a stock broker anywhere in the country can determine the status of as many as eighteen separate stock issues. Typical turnaround time is one second.

Just as time-sharing is an outgrowth of older concepts in the computer industry, it is itself already giving rise to other related kinds of computer enterprises.

Whereas most commercial time sharing firms are in the business of selling computer time, at least one organization, Data Services, Inc., sells computer core capacity on a kind of hybrid time-share basis. Access to a centrally located, multipurpose computer is sold to different businesses on a quota basis. That is, many firms engaged in the same business will cooperate in renting a certain portion of the computer's capacity. Typical of such firms are the multiple-listing realtors.

The computer is used to keep all real estate listings up to date for a given set of agencies. Access to the information is available to all subscribers – the result is that all the realtors in a geographic area might have access to every available piece of real estate in that area. Other subscribers to this kind of computer service are savings and loan associations, hospitals, ambulance services, banks, travel bureaus, insurance agencies, and ticket agencies, among others.

Generally, time-share is preferable to batch processing systems in applications where immediate access and response are essential and in any multifunctional requirements. In traditional areas — computing payrolls, inventories, accounts, etc. — where turnaround time is not a critical factor, batch processing is satisfactory. Even here though, general purpose time-share computers may have a role. With attachments, time-share computers can be altered for use in batch processing during off-peak hours of operation. This is particularly true of the smaller in-plant systems.

The primary value of time-sharing is not so much its advantages over batch-processing as is the fact that it has made the services of computers available to a much broader segment of the economy.

A Problem Becomes Progress

Since its inception, the computer industry has been saddled with two major problems — the developmental lag between hardware and software, and the input/output problem. The crux of this I/0 problem is the fact that terminal devices have yet to be designed which can make meaningful use of a high-speed computer's fantastic capabilities. Particularly since I/0 devices are of necessity geared to the relatively low operating speeds of human beings. But, as pointed out earlier, it is this disparity which makes time-sharing practicable — and time-sharing, in turn, is offering a tentative solution to the I/0 problem.

More recently, software developments have been growing apace with the hardware industry. Programs now exist which are of sufficient size and complexity to justify even higher computer speeds. Also, print-out devices are being marketed which have increasingly higher operating speeds; electronic typewriters now print 200-300 words per minute. This in itself is unimpressive compared to computer speeds, but with as many as 200 time-share subscribers working on-line simultaneously, the figure becomes more favorably comparable. Also, there are line-printers available with operating speeds of up to 3000 *lines* per minute.

However, the current interest in output devices is tending more toward electronic graphic display units such as cathode ray tubes. Among the most recent innovations in this area are some highly sophisticated machines — electron beam recorders, computer output microfilming machines, and

Courtesy General Electric Information Systems.

Figure 5. Computer time-sharing systems such as this one are springing up around the world at the rate of three or four a month.

even lasers; these devices provide instantaneous responses in the conversation between man and machine.

It is this aspect of time-sharing – conversation or interaction – which is potentially quite valuable. Not only does time-sharing serve more needs of more users more economically, it may pave the way for truly creative communications – hence, augmenting man's working intellect.

Time-Share Experiments

Scientists at Stanford Research Institute in Menlo Park, California, have been experimenting with computers in this area for the past two years. The system consists of a large, multi-access, general purpose computer with cathode ray tube terminal devices.

For better viewing, the CRT's are interfaced with closed-circuit television cameras; monitors are located in different rooms throughout the engineering department. Input is through an electronic typewriter keyboard. Usually, interaction with the computer is on a one-to-one basis; however, conference arrangements have been tried with considerable success. In such an arrangement, all the conferees have immediate access to the programs of all the others. And they are continually available for cross-reference or checking back. Perishability of data is not a factor; and conceivably then, a problem can be attacked simultaneously by several different men aided by the huge memory of a computer. The possibilities presented by such a concept are impressive, to say the least.

For example, each member of a group of scientists or engineers is given an identical problem. Working off-line and independently, each develops a possible solution; then the statement of the problem together with each of the proposed solutions is programmed into the computer. This done, each man sits down at his console and queries the computer for the problem as well as for each of his colleagues' work on the problem. With refinements, the participants will also be able to communicate directly with each other through the computer.

A Computer in Every Home?

Potentially, the applications for time-share computers are virtually unlimited. Computer hardware and software systems are advancing at an increasingly rapid rate.

Medium and large scale integrated circuitry, along with advanced construction techniques are pointing the way to the future.

Telecommunications engineers are developing ever faster methods of data transmission. Lenkurt's Duobinary technique is one example; PCM offers a broad range of possibilities for data communication. Working with PCM, transmission soon will be reckoned in the millions and even hundreds of millions of bits per second.

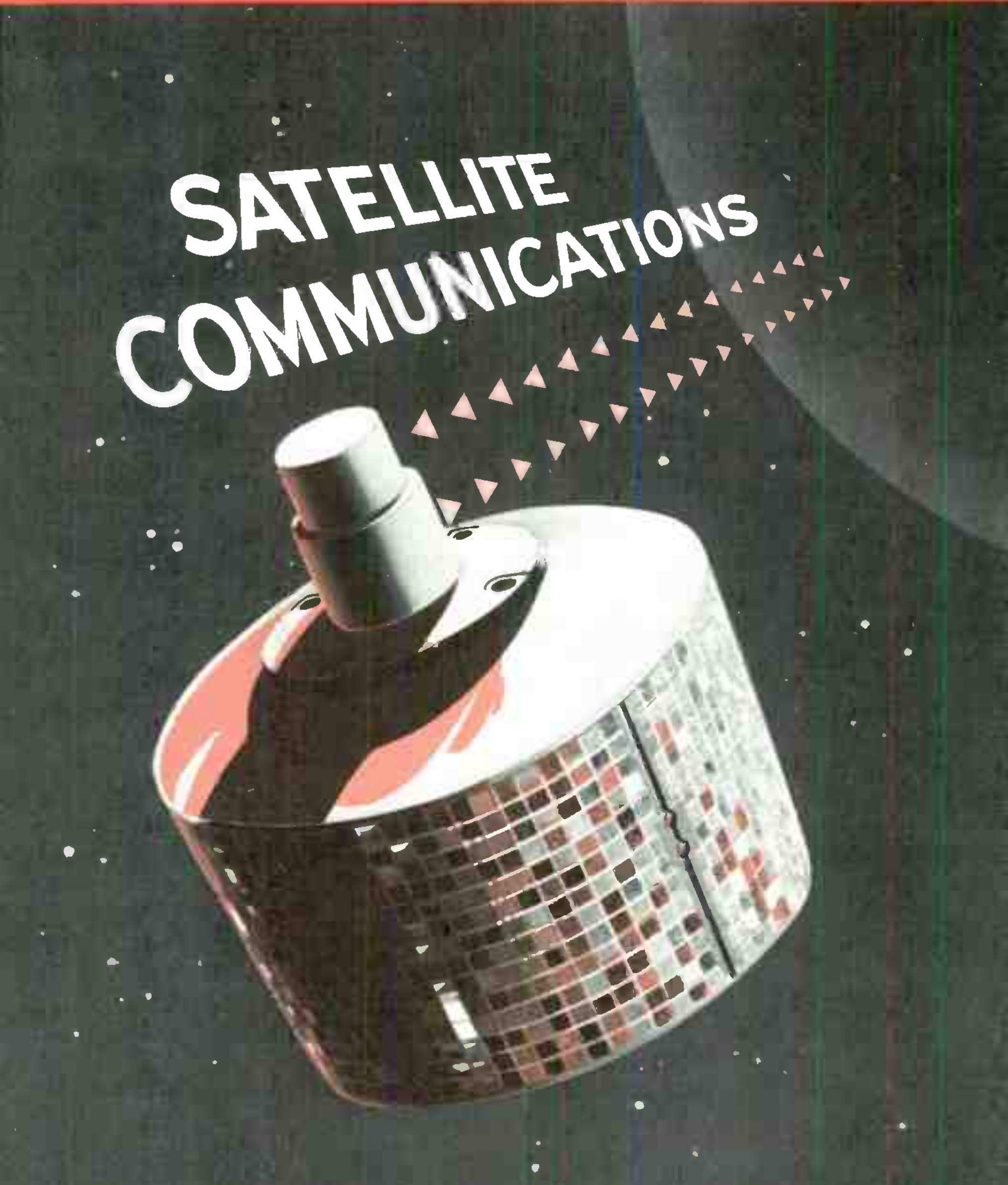
Far-sighted people in the computer world have envisioned such novel computer oriented services as automatic credit and banking where no cash changes hands, and networks of computers which enable businessmen to go to their "offices" without ever leaving home. Others foresee a time when movies, books, periodicals, and newspapers will simply be "dialed up" for viewing on the living room TV screen. These may be the first steps toward giving every home a computerized information center – a not unforeseeable situation in the light of current developments in computer time-sharing.

SECTION IV

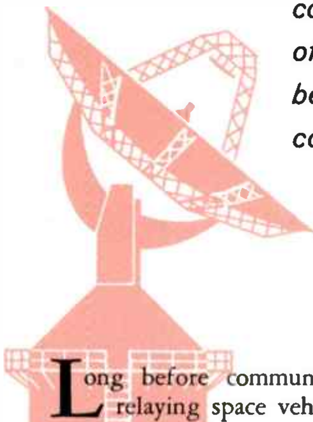
GENERAL COMMUNICATIONS

DECEMBER, 1966
VOL. 15, NO. 12

the Lenkurt. Demodulator



The first commercial use of man's ability to travel in space — not yet a decade old — has come in the field of communications. Soon, over 90 percent of the world's telephone facilities may be joined in a global network through communications satellites.



Long before communications-relaying space vehicles became a reality, scientists eyed the natural satellite of the earth—the moon. Early in 1946 Project Diana bounced the first radar signals off the moon. Twenty years later artificial satellites circle the globe, but the moon has not been forgotten. A ship-to-shore communications link using the moon as a reflector is expected to be operational sometime in 1967.

The first man-produced "beeps" directly from space were heard on October 4, 1957 from the Russian Sputnik I. The United States entered the space age on January 31, 1958 with Explorer I. Today, in addition to the spectacular manned exploration of space, commercial satellites stretch telephone circuits across the oceans, and telecasts from other continents are common.

American scientists placed the first communications satellite in orbit a year after Explorer I. Score—a short lived but highly successful "bird"—relayed messages up to 3000 miles and broadcast to the world a tape recorded Christmas greeting from President Eisenhower.

The 1960 flight of Echo I was witnessed around the world as the 100-foot balloon-like reflector satellite provided a "radio mirror" for powerful ground stations. Echo II went up in 1964 as experiments with passive reflectors continued.

An active repeater, Courier, extended the knowledge of space communications in 1960 with successful transmission of of high speed teletype, voice, and facsimile.

The commercial value of communications satellites was accentuated in 1962 with Telstar I, the joint project of NASA and AT&T. The first live telecasts between Europe and the United States added to the satellite's performance in transmitting high-quality voice, data, teleprint and other signals. Telstar II, and NASA's Relay I and II added more data.

By mid-1963, the first of three Syncom satellites was launched and communications milestones began to pile up. Syncom III brought the Tokyo Olympic games to the United States, and went on to demonstrate its value for all types of telecommunications. NASA has since concluded its planned tests, and both Syncom II and III are now working for the Defense Department, parked over the Pacific and Indian Oceans.

How High?

Most alternatives faced in the design of communications satellites are centered around the choice of orbits. Orbital mechanics govern precisely the height and period of a satellite. A satellite circles the earth in a period directly related to the satellite's altitude. (The mass of the satellite is negligible and can be ignored in most calculations.) A satellite 100 miles high circles the earth in about 87 minutes; at 1000 miles the period is 118 minutes. As the altitude increases, the orbital period becomes longer, until at an altitude of 22,300 miles a satellite orbits the earth in exactly the same time as one rotation of the earth—that is, every 24 hours. Placed in an easterly orbit over the equator, such a "synchronous" satellite appears stationary in the sky.

The first communications satellites—the Telstars and Relays—were in nonsynchronous orbits. The new breed, lead by Syncom and Early Bird, are synchronous and remain in precisely fixed positions. But depending on the application, each plan has advantages and disadvantages.

The lower the satellite, the shorter its period, and likewise the less time it will be in the simultaneous view of any two ground stations. For example, a 3000-mile-high satellite can be tracked for only 24 minutes by stations located 3000-miles apart—*if* the satellite passes directly over both stations.

Military Plan

A low-flying random orbit is especially appealing to the military, interested in the security of its communications system. The quasi- or nonsynchronous satellite does not require orbit-control commands from the ground, and therefore cannot be tampered with by an enemy. In the Initial Defense Communications Satellite Program (IDCSP), with up to 24 satellites placed in an 18,000-mile orbit, if a satellite fails for any reason another will soon move into view. The satellites drift around the earth at about 30° per day—any single satellite is in view for over four days at a time.

Commercial Advantage

Commercial systems planned through 1968 will be synchronous; the reasons are mostly economic. Fixed position synchronous satellites greatly simplify tracking, thus reducing the cost of ground stations. In developing a truly worldwide system, where each additional ground station may open communications to an entire region or country, the installation cost of these stations becomes increasingly important.

Tracking becomes a relatively simple function of the ground station of a synchronous system. As satellites are pushed by "solar wind" and pulled by gravity from the earth, moon, and sun, periodic adjustments in position are made by onboard thrusters. Between correction intervals typical 85-foot parabolic ground antennas track minor variations—measured in hundredths of a degree.

The synchronous satellite, at 22,300 miles above the earth, is visible to almost half of the globe at one time. Three satellites, spaced equally around the earth, would provide contact with any country served by an adequate ground station (Figure 1). An exception exists at the poles where signal strength is at a minimum. In practice, more than three satellites undoubtedly will be used to increase the number of circuits available in high density areas, and to ensure greater flexibility.

Long transmission delay time is the one serious disadvantage at synchronous altitudes. A one-way telephone path through such a satellite is about 50,000 miles—a delay of 265 milliseconds. Since delays of over 400 ms are unacceptable for voice communications, circuits probably will be limited to only one satellite hop in spanning the globe. The problem, however, does not concern the transmission of television or data.

The products of new technology are easily phased in with a synchronous system. Three or four satellites can replace an entire synchronous system, and older low-capacity units can be moved to areas where traffic is lighter.

The ability to relocate synchronous satellites is a needed feature should a failure occur. An extra satellite, parked in a low traffic area, could be moved in as a replacement in much less time than it would take to prepare and launch a new vehicle. Using onboard thrusters, a satellite can move about 10° per day — from a station over the Atlantic to the Pacific, for example, in about 15 days.

Intelsat

Global communications is being established through the International Telecommunications Satellite Consortium (Intelsat), made up of 54 participating countries. Congress has franchised the Communications Satellite Corporation (Comsat) to establish service for American common carriers, and to represent this country in dealings with Intelsat. Comsat is the major shareholder in Intelsat and serves as its manager.

Any country can join Intelsat, agreeing to share the financing of satellites and tracking equipment. Each country has the responsibility for its own ground stations, with at least 25 countries expected to have working stations by 1971.

The first commercial communications satellite, popularly known as Early Bird, began operation over the Atlantic in mid-1965. Two advanced Intelsat 2 satellites (Blue Birds) will establish the first service over the Pacific and add to the channel capacity over the Atlantic. The third generation Intelsat 3 satellites are scheduled for launch in early 1968 to further expand the worldwide system.

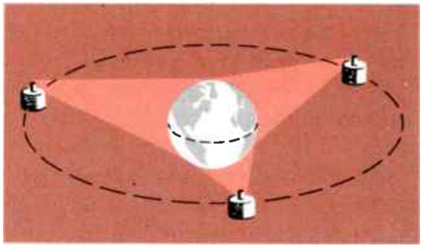


Figure 1. Three satellites in synchronous orbit could cover the entire earth.

Power Source

A sizeable tradeoff between rocket booster power and payload weight must be made in orbiting any object. Since the communications satellite must carry its own power source into space, energy for all electronics is necessarily limited. In turn, the less radiated power from the satellite's transmitter, the lower the signal-to-noise ratio and the fewer channels that can be relayed.

The solar cell remains the most practical power source in space, delivering about 6 watts per pound. Even at that, Intelsat 2 must make do with about 85 watts. Experts claim that up to 800 watts is possible using only skin-mounted solar cells, and deployable arrays might boost available power into the kilowatt area. But such arrays, like nuclear energy power for spacecraft, must remain in future plans.

Today's problem is doing the most effective job with the equipment available. With power output confined, attention has naturally turned to spacecraft antenna design—itself restricted by other physical considerations.

Satellites are prevented from tumbling uncontrollably through space by giving them a bullet-like spin of about 150 rpm. Spin stabilization eliminates some problems, but creates others, especially for antenna designers. With the satellite spinning, it is impossible to use a conventional directional antenna.

Present communications satellite antennas produce a toroidal or "doughnut shaped" pattern with about 9-dB gain (Figure 2). Better than a omnidirectional antenna, the method nevertheless "loses" considerable energy in the portion of the pattern not touching the earth.

Despun Antennas

Future spin-stabilized satellites will be equipped with devices for focusing this otherwise lost energy, increasing

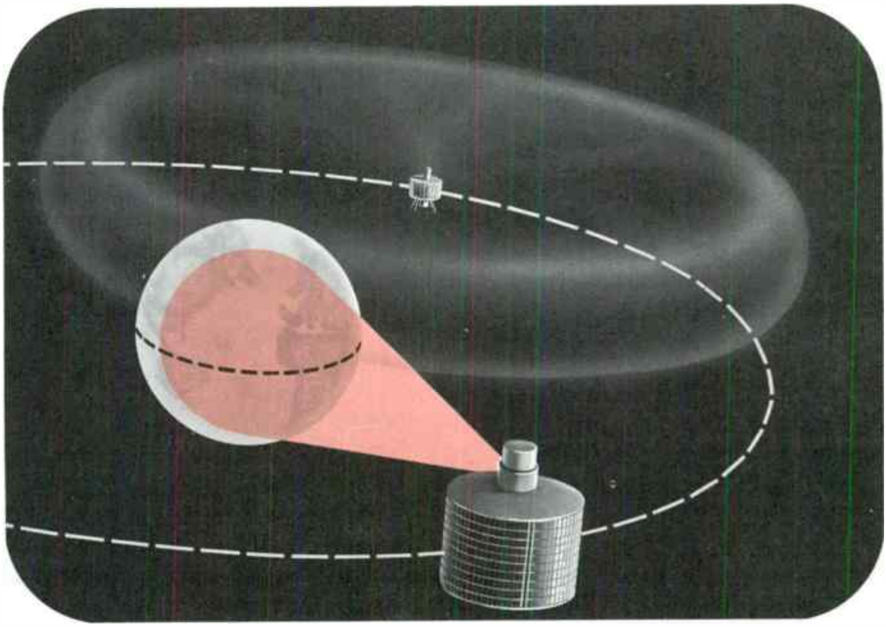


Figure 2. Toroidal pattern of first communications satellites (rear) loses much rf energy to space. Intelsat 3 will focus its communications beam toward earth, using new despun antennas, with up to 16 dB gain.

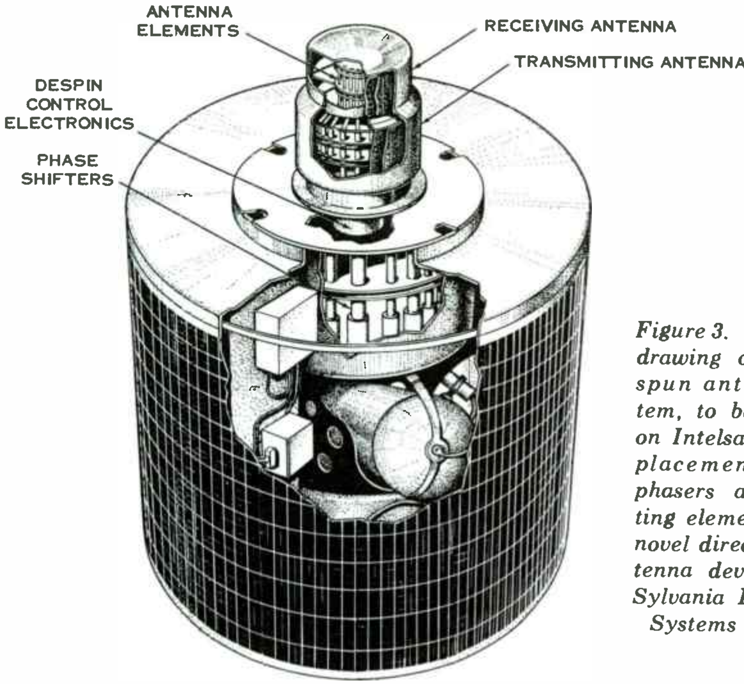


Figure 3. A cutaway drawing of the despun antenna system, to be initiated on Intelsat 3, shows placement of the phasers and radiating elements of the novel directional antenna developed by Sylvania Electronics Systems Division.

the gain considerably. Expected minimum gain will be 13 dB, with peak gain about 16 dB. The trick is to rotate the antenna in the opposite direction to the satellite, thereby keeping the "despun" beam pointed to earth.

Three types of despun antennas are possible: mechanical, electronically switched, and electronic. The mechanical method uses a directional antenna that is physically counter-rotated about the axis of the satellite. The greatest danger is mechanical failure. The electronically switched method systematically shifts rf power from antenna to antenna, keeping overlapping beams in the desired direction. The pure electronic approach—the one selected for the Intelsat 3 satellites—steers the beam by varying the phase of the signal as it feeds a series of radiating elements (Figure 3).

The electronically despun antenna has three major subsystems: an earth center reference system, control circuits, and the radiating assembly. Two redundant horizon sensors scan the earth as the satellite rotates. Control circuits regulate the action of phase shifters, which direct the rf energy to the radiating elements of the antenna. The result is a radio beam continually focused on the earth.

At synchronous altitudes, the earth's disk is just over 17° across. Allowing for satellite stabilization and antenna tracking errors, a beam approximately 19° x 19° would adequately cover most points on the globe. For specific purposes, the beam could be made more directional, thereby increasing the gain. For example, a satellite designed to relay traffic only from the United States to Europe might have a fan-shaped beam 19° x 10° (long to the east-west).

Minimum gain would be increased to 16 dB, with peak gain at 19 dB.

An alternative to spin stabilization is being tested, requiring no onboard thrusters or other control devices. Known as *gravity-gradient* stabilization, the method would maintain the same side of the satellite always facing the earth. A long object in space will tend to align itself vertically with the strongest source of gravity—in this case, the earth. Extendible arms could, in effect, make the satellite such a "long" object. Highly directional antennas could then be accurately pointed earthward.

Early Bird

Important advances are continually incorporated in new satellite designs. Our first communications satellite now seems small compared to vehicles being developed. Early Bird weighs 85 pounds, is 28 inches in diameter, has a solar power capacity of about 46 watts, and can relay 240 two-way voice channels, or two-way television. The communications system has two transponders (receiver-transmitters), one for each direction of traffic. The transmitter output comes from one 6-watt traveling wave tube; a second TWT is carried for redundancy. The bandwidth of the transponder is 25 MHz. Receiver frequencies are in the 6-GHz band, with transmission back to earth in the 4-GHz range. Telemetry and control signals to and from the satellite are at VHF frequencies in the 136 MHz range.

Intelsat 2

The latest addition to the global system, Intelsat 2, has five times the bandwidth (125 MHz) of the Early Bird, and three times the output power (18 watts). Increased power provides greater geographical coverage, while the wideband capability allows multiple access for the first time. Now a number of different ground stations can channel through the satellites at the same time. Capacity remains at 240 high-grade voice channels.

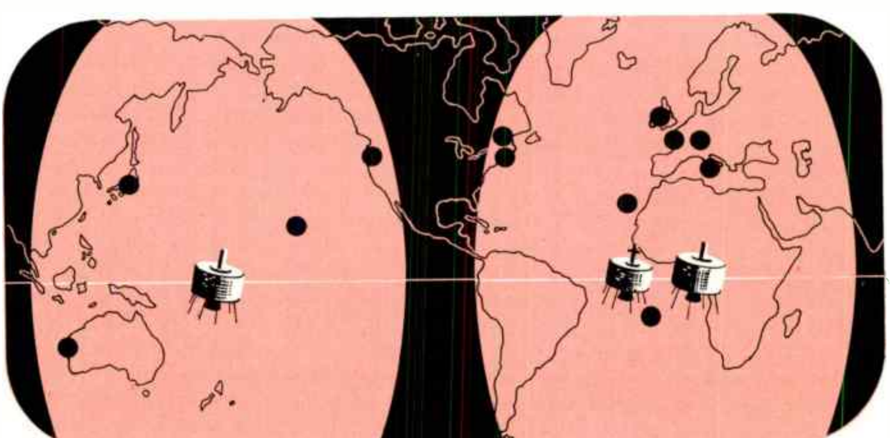


Figure 4. Ground stations on both sides of the Atlantic and Pacific will be connected through Early Bird and Intelsat 2 satellites.

So called "quasi-linear" transponders are a unique part of the multiple access design; transmitter output increases linearly with an increase in input power. There is no radiated power from the satellite until a signal is received from earth. If several signals are received simultaneously, transmitter power is divided among them proportionally according to the power of the received signal. The quasi-linear method reduces intermodulation products and crosstalk inherent in the Early Bird fixed-output transponder.

Four 6-watt TWT's are carried in the spacecraft. Three of them normally will work in parallel; the fourth is a spare. The Intelsat 2 satellite has an orbital weight of 165 pounds, is 56 inches in diameter, and produces 85 watts of power from solar cells. Communication and telemetry frequencies are virtually the same as in Early Bird.

Electronics

Within the Intelsat 2 transponder, the incoming 6-GHz signal passes through a low-noise tunnel diode rf amplifier to a directional coupler where command signals are extracted (see Figure 5). The communications signal is converted directly from 6 GHz to 4 GHz in a mixer section, then delivered to a driver TWT. The driver tube operates only on command, providing ground control selection of two redundant receivers. A beacon frequency (for tracking) is then added before the signal is supplied to the four output TWT's.

The satellite is capable of receiving a signal at any frequency between 6283

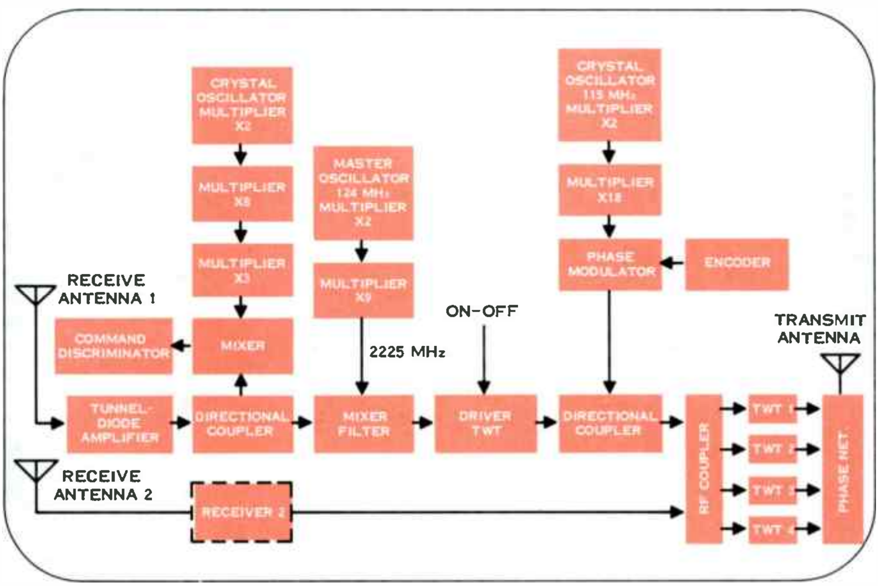


Figure 5. Communications system block diagram for Intelsat 2, built by Hughes Aircraft Co.

Figure 6. Seen inside Intelsat 3, to be launched in 1968, are many of the electronic and propulsion systems included in the third generation communications satellite being built by TRW Systems.

MHz and 6409 MHz. The incoming signal is translated at the mixer by 2225 MHz to the transmit band of 4058 MHz to 4184 MHz.

Telemetry and control signals are available on both VHF (136 MHz) and through the modulated beacon carried with the communications channels.

Together with adding satellite communications capability over the Pacific and increasing service in the Atlantic, Intelsat 2 will play a vital role in the Apollo space program. NASA will use a number of circuits for astronaut voice relay, spacecraft television, high-speed tracking data, and telemetry. Transmission path will be from the Apollo spacecraft to NASA surface stations (including special tracking ships at sea), and then via the communications satellites to mainland ground stations.

Intelsat 3

Even higher capability is being designed into the Intelsat 3 satellites, to be launched in 1968. Measuring 56 inches in diameter, weighing 250 pounds, and with solar power of 160 watts, the communications package will be able to handle at least 1200 two-way voice channels or four television channels (See Figure 6).

Each of the two transponders aboard Intelsat 3 will have a bandwidth of 225 MHz, with high-level 10-watt TWT output stages. Incorporating electronically despun antennas, the satellites will have an effective radiated power of about 22 dBw (decibels above one watt), compared to about 15 dBw for Early Bird. Like its predecessors, Intelsat 3 will use rf amplification, with translation from 6 GHz to 4 GHz.

Table A. Frequency allocations agreed on at the 1963 Geneva Extraordinary Administrative Radio Conference. Commercial bands, shared with terrestrial systems, are shaded; others are for special applications (including military).

Frequency Bands	Service	Frequency Bands	Service
1700-1710 MHz	Space Research (Telemetering & tracking) (shared)	5725-5850 MHz	Communication-Satellites (Earth-to-satellite) (shared)
1770-1790 MHz	Meteorological-Satellites (shared)	5850-5925 MHz	Communication-Satellites (Earth-to-satellite) (shared)
2290-2300 MHz	Space Research (Telemetering & tracking in deep space) (shared)	5925-6425 MHz	Communication-Satellites (Earth-to-satellite) (shared)
2690-2700 MHz	Radio Astronomy (exclusive)	7250-7300 MHz	Communication-Satellites (Satellite-to-Earth) (exclusive)
3400-4200 MHz	Communication-Satellites (Satellite-to-Earth) (shared)	7300-7750 MHz	Communication-Satellites (shared)
4400-4700 MHz	Communication-Satellites (Satellite-to-Earth) (shared)	7900-7975 MHz	Communication-Satellites (Earth-to satellite) (shared)
4990-5000 MHz	Radio Astronomy (shared in some areas)	7975-8025 MHz	Communication-Satellites (Earth-to-satellite) (exclusive)
5250-5255 MHz	Space Research (shared)	025-8400 MHz	Communication-Satellites (Earth-to-satellite) (shared)
5670-5725 MHz	Space Research (Deep space) (shared)	8400-8500 MHz	Space Research (shared) (exclusive in some areas)

Life expectancy of synchronous satellites is about five years, governed by the onboard fuel supply for positioning thrusters. When the fuel is expended, the satellite will begin to drift slowly westward. Its communications capability, however, could continue for some years. The life of the electronics is primarily dependent on the source of electrical power—solar cells. These cells deteriorate with exposure to radiation, a common hazard in space, especially near the Van Allen radiation belts. Intelsat 3, for example, will begin its service with 161 watts of available power. After five years only about 105 watts can be expected from the solar cells. But this is still enough to support at least limited communications.

Modulation

The satellite microwave repeater has much more bandwidth than its terrestrial cousin. Bandwidth in satellites is needed not only to increase channel capacity, but to allow multiple access from many ground stations. Each ground station will use a discrete carrier frequency, with a number of multiplexed channels. Frequency division multiplex with frequency modulation (FDM/FM) is used in current systems, but other modulation techniques are possible. Time division multiplex (TDM) tests have been completed with the Early Bird satellite. Pulse code modulation (PCM) was used successfully to carry voice and data signals between two North American terminals.

Russian Satellites

Even though Intelsat countries have 90 percent of the international communications potential, considerable work is being done by nonmember countries—especially the Soviet Union. Russia has lofted several Molniya-class communications satellites in 12-hour elliptical orbits. Successful transmissions of many types of signals, including color television, have been carried out in joint experiments with the French. While the Russian satellites apparently lack channel capacity, they do boast high-powered transmitters and other "weighty" equipment. The Molniya has a command receiver, 40-watt transmitter, two reserve transmitters, and two steerable parabolic antennas. In addition, the satellite has orbital adjustment and three-axis attitude control capabilities. Apparently with room to spare, the Russians are also including meteorological equipment on board, returning cloud photos to weathermen.

Expanded Uses

Great potential in distant communications exists in many areas beyond the telephone and television industries. The Federal Aviation Agency is interested in establishing service for transoceanic flights, so often out of reach of HF and VHF radio. This could come by mid-1967. The same advantage would be available to ships on the high seas.

The possibility of broadcasting television directly to the home via satellite has received considerable attention recently. Though the practicality of such a system may be questioned for some years, technologically it is not difficult to imagine.

Satellites vs. Cables

Of immediate interest to international common carriers is the commercial value of satellites. Coaxial submarine cables—only ten years old themselves—remain as the backbone for communications across the oceans. More high-capacity cables will be installed in the next five, ten, or even more years. From the present 3500 voice channels, international service could jump to over 7000 in the next five years, and triple in 10 years. Much of this growth, especially between dense traffic areas, can be handled by high-capacity cable. Of course, satellites will absorb their share of the market.

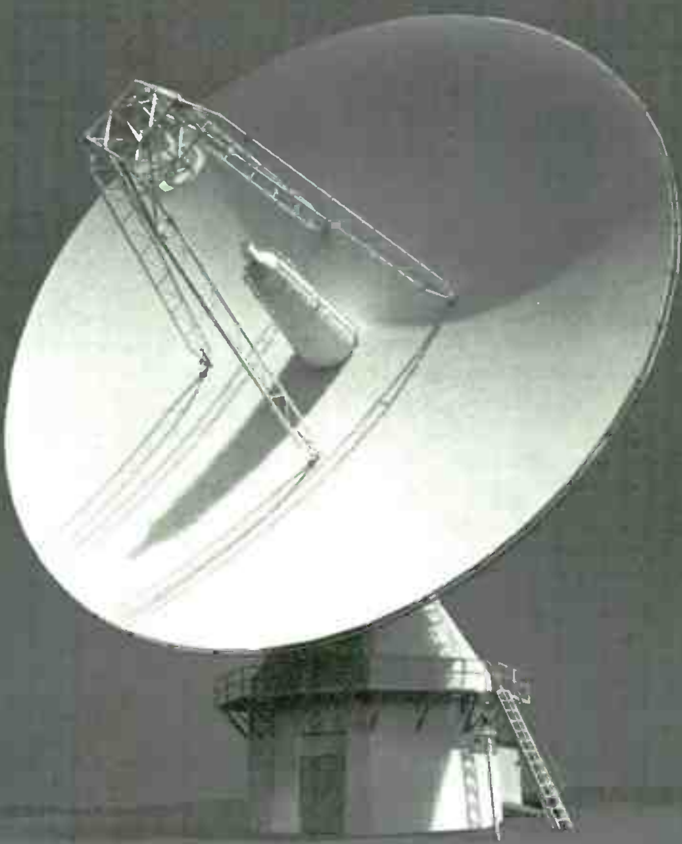
Probably more important are the new markets satellites will create for international communications. Continent-to-continent television, for example, had to wait for satellites—cables lacked the necessary bandwidth. Satellites will likewise provide a logical medium for high-speed data transmission between commercial centers around the world.

We have obviously crossed the threshold toward complete global telecommunications.

JANUARY, 1967
VOL. 16, NO. 1

The Lenkurt DEMODULATOR

GROUND STATIONS for Satellite Communications



Necessarily, most of a satellite communications system never leaves the ground. Highly directional antennas and elaborate electronics equipment must be stationed around the world to interconnect the orbiting microwave relay with its terrestrial users.

The techniques of communication through satellites and point-to-point microwave radio systems share conceptual similarities: both receive and transmit the same signals at the same frequencies through directional antennas. Differences are in the areas of power, bandwidth, and access to the system.

Point-to-point microwave transmitters span 30-mile distances with 1 to 5 watts, using dish antennas about 6 feet in diameter. Satellite ground stations must work with a repeater 22,300 miles distant, with 10-kilowatt transmitters driving 85-foot antennas. Moreover, the large antenna must be steerable and have the ability to accurately track a satellite in orbit.

At the same time, the satellite must offer access to many ground station "end terminals" — perhaps dozens of countries wanting to relay messages through the same repeater at the same time.

Bandwidths up to 500 MHz place additional demands on the earth station, which must "scoop in" meager satellite signals while maintaining relatively low noise figures. Large antennas, coupled with super-quiet receivers meet the requirements — but not without a great deal of engineering sophistication.

The commercial satellite system is being established through the International Telecommunications Satellite Consortium (Intelsat), with America's Communications Satellite Corporation (Comsat) as manager. Intelsat is providing the satellite relays for its 54 member nations, plus ground control equipment to keep the satellites properly positioned. The individual countries are responsible for establishing and operating their own ground stations — and will coordinate traffic and and tariffs with local carriers.

Since the summer of 1965, earth stations in Europe and North America have been gateways for international communications through the Early Bird satellite. The launching of improved satellites this year will open more channels across the Atlantic and establish service in the Pacific — spanning nearly two-thirds of the earth. For this service to be useful, many new ground stations must be built.

Locations

In the United States, three ground stations are now in operation for use with the new Intelsat 2 satellites. The

Figure 1. Brewster Flat ground station, with folded-horn and Cassagrain antenna (right, in distance). Microwave antennas are visible above building.

station at Andover, Maine, was built to work with Telstar in 1962, and is now a part of the Comsat system. Two new stations serving the Pacific were recently activated at Brewster Flat, Wash., and Paumalu, Hawaii. Comsat plans other stations at Moorefield, W. Va., and in the Caribbean. A station at Mill Village, Nova Scotia, will connect Canada to the global system.

In Europe, major stations are being operated in England, France, Germany and Italy. New stations are being built on Ascension Island by the United Kingdom; at Madrid and on the Grand Canary Island by Spain; and in Australia. Japan is joining the Pacific system with a station at Ibaraki.

Many other sites are being considered: Hong Kong, Bahrain on the Persian Gulf, Thailand, the Philippines, a number of countries in Latin America, the African Continent, and another location in Australia.

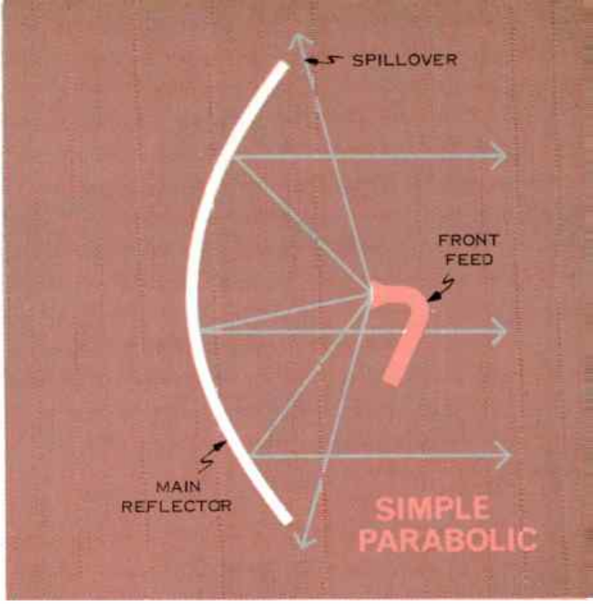
Earth stations must avoid all sources of electromagnetic noise — natural or man-made — if faint satellite signals are to be received satisfactorily. Geographic location, frequency allocations, and antenna design all must be considered if noise is to be reduced. Ideally, ground stations are situated in "radio quiet" locations, distant from sources of interference, yet close to power and telephone network interconnect.

Frequency Choice

The choice of frequencies to be used has a great influence on communications satellite technology. An antenna pointed skyward will "see" two types of noise: that from galactic sources, and that from the atmosphere itself. Galactic noise is relatively high at frequencies under 400 MHz, but falls off rapidly at the higher frequencies; galactic noise is negligible above 1 GHz. On the other hand, atmospheric noise increases above 8 GHz, and is considerable above 10 GHz. Logically, the spectrum between 1 and 10 GHz is considered most favorable for satellite communications.

Operating frequencies now in use were adopted at the 1963 Extraordinary Administrative Radio Conference in Geneva. Uplinks (ground to satellite) for commercial satellite systems are in the 6-GHz frequency range, while returning downlinks are in the 4-GHz

Figure 2. Three versions of the parabolic antenna, showing placement of rf feed and path of reflected energy. The Cassagrain is most practical type for satellite communications.



band. Other bands are used for special purposes, including military communications.

Since many of the satellite allocations are shared with other services, restrictions have been placed on the maximum flux density allowable at the earth's surface. While there are a number of variables, such as angle of arrival and method of modulation, the basic recommendation is for a maximum density of -130 dBw/m^2 in any 4-kHz bandwidth. This restriction limits the power a satellite may radiate, and likewise places greater emphasis on the design of the ground stations that must receive these signals. In many cases, the ground station will be joined to telephone networks by point-to-point microwave radio using the same 4 and 6 GHz bands.

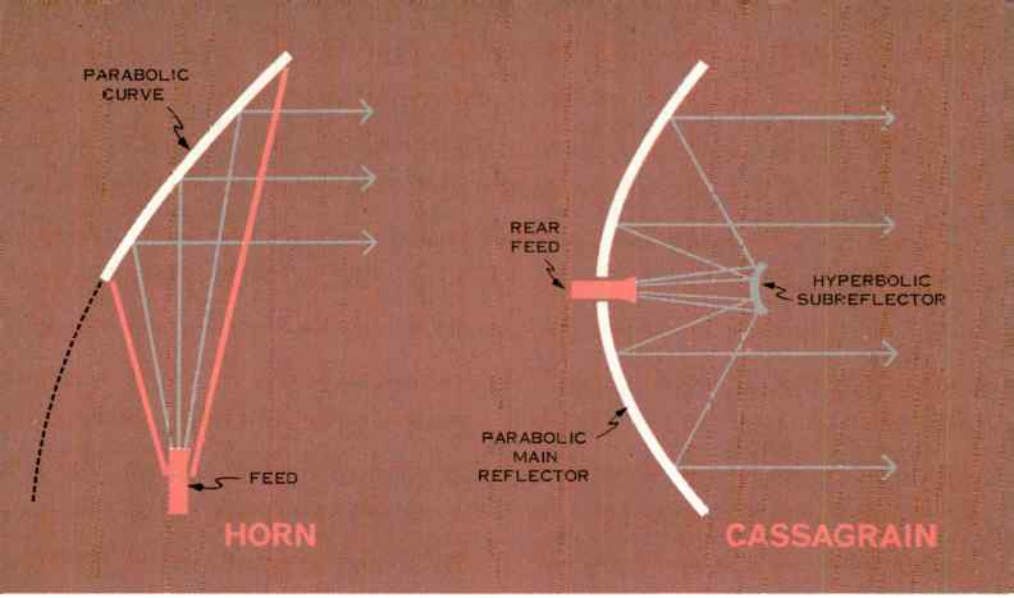
Antenna Gain

Probably the most critical single component in a ground station — and the most costly — is the antenna. It must concentrate transmitter power toward the satellite, and receive communications and tracking signals from the satellite. The desirability of a particular antenna design is directly proportional to the ratio of ántenna gain to receiving system noise. Gain is primarily dependent on the physical size of the antenna, and may be defined as the capability of the antenna to concentrate *transmitted* energy in the desired direction. Similarly, gain represents the *reception* of energy from a desired direction and the rejection of energy from other directions.

It is obvious that an effective ground station must have as narrow a beam as possible. While transmitting, the antenna must concentrate most of the radiated power in the main beam, known as the major lobe. Any energy in side lobes is wasted. Conversely, in a receiving antenna the side lobes represent sources of extraneous noise.

Noise Temperature

Noise in antenna-receiver systems is commonly expressed in degrees Kelvin. Since all objects radiate energy — the higher the temperature of the object, the more energy radiated — *noise temperature* is an appropriate unit of measure-



ment for receiving systems. For instance, a highly directional antenna pointed at the sun (surface temperature about 6000°K) will receive a noise power of 8.28×10^{-14} watts, or −101 dBm.

The concept of noise temperature is easily extended to other noise sources. Any device which produces random noise of 4.0×10^{-21} watt (−174 dBm) per cycle of bandwidth, may be said to have a noise temperature of 290°K — even though that may not be its physical temperature. In other words, the noise temperature of a device is the temperature at which a thermal noise source would have to be operated to produce the same noise power. Expressed as temperatures, noise units from antenna and receiver are conveniently added to arrive at a total system noise value.

Antenna Design

Three variations of the basic parabolic reflector can be considered for satellite communications ground stations — with noise performance the constant criterion.

The basic parabolic antenna (Figure 2), like those used in point-to-point microwave transmission, is a highly directional device. However, the center feed at the focal point of the antenna cannot be precisely controlled to illuminate only the reflector. A certain amount of "spillover" occurs in even the best system, adding undesirable lobes at the back and sides of the antenna. With the antenna reflector looking skyward, the feed is pointed back at the relatively noisy ground (approximately 290°K). Spillover from the simple parabolic adds to the antenna's noise temperature. Also, it is mechanically inconvenient to mount needed preamplifiers at the feed of the simple parabolic.

In an attempt to reduce minor lobes, the horn reflector antenna was developed. The antenna is basically a section of a parabola, with sides extended from the feed to the edges of the reflector. The horn is a highly efficient, low-noise antenna, but is large and costly. The first U.S. commercial satellite ground station at Andover uses a horn reflector weighing 380 tons.

The popular choice is the Cassagrain antenna. The double-reflector system incorporates a parabola main reflector, and a hyperbola subreflector. The feed is at the rear of the main reflector and looks, with the antenna, at the sky (noise temperature typically less than 30°K). In operation, rf energy from the feed strikes the subreflector and is bounced back to illuminate the main reflector as if it had come from the focal point of the parabolic.

The Cassagrain antenna (named for William Cassagrain who developed the subreflector method of improving optical telescopes) is superior to the conventional parabolic in a number of ways. The antenna has low spillover, shorter transmission lines, greater mechanical stability in the feed system, and more flexibility in design. Careful engineering eliminates the otherwise serious disadvantage of placing elements that could block radiation in front of the antenna.

The typical commercial ground station uses an 85-foot Cassagrain. Illustrated on the cover is the antenna recently installed by Sylvania Electric Products, Inc. at two Comsat stations.

Variations

Other variations exist, especially in mobile and transportable antennas. A 42-foot transportable folded-horn antenna is being used at some commercial ground stations while the more complex Cassagrains are readied for service.

The smaller the antenna, however, the higher the noise temperature of the system. Likewise, the lower the signal-to-noise ratio, the less bandwidth available for communications. The 42-foot folded horn is designed to provide a limited number of channels for telephone, telegraph, and high-speed data, but the 85-foot Cassagrain must be used for television transmission.

Willing to sacrifice bandwidth for mobility, the military is developing a number of smaller units for field and shipboard use. The smallest antenna system may service only one voice channel, but allow reliable long-range communications to be established at virtually any remote location in hours.

Pointing Accuracy

Large reflectors characteristically have high gain and narrow beamwidths; sophisticated control is required to keep these antennas properly aimed at the small point in the sky.

The accuracy imposed on earth stations is illustrated at Brewster Flat and Paumalu. The movable portion of the antennas at these new stations weighs more than 135 tons, but must be able to rotate 360 degrees in 120 seconds, and track satellites to within 1/500th of a degree!

It should be noted, however, that when commercial satellite systems are fully removed from the experimental stage and configurations are firmly established, little antenna adjustment should be required.

With any narrow-beam antenna, the transmitted signal must be pointed in a precisely known direction. Mechanical vs. electrical alignment is accomplished in much the same way an expert marksman corrects the sights on his rifle — by carefully aiming, firing a shot, and noting where it hits.

A 6-foot "boresight" antenna is the target for the ground station antenna. The boresight facility simulates the satellite's performance by translating a received signal to a different frequency and returning it to the ground station. By pointing the large antenna at the target and plotting returned signals, engineers can calibrate accurately the true direction of the narrow radiated beam.

Amplifiers

In operation, usable signals from the communications satellite are possible because of high-gain, low-noise antennas matched with ultralow-noise preamplifiers. The preamplifier. which might be a maser, parametric amplifier, or other similar device, is placed as near as possible to the antenna feed. A parametric amplifier operating as the initial microwave amplifier, may have a noise temperature of only 15°K. The entire receiving system, including antenna, could be rated at less than 50°K. Compared to the 1200°K temperature of a very good receiver only a few years ago, it is easy to recognize the great technological advances necessary for a practical satellite communications system.

Ground stations must transmit powerful signals to the satellite. Comsat stations use either traveling wave tube or klystron power amplifiers to produce 5 to 10 kilowatts of rf energy. The TWT is generally used for its greater bandwidth, while the more powerful klystron serves tunable narrow-band applications.

Telephone Interface

The national telephone network is connected to remotely located ground stations by point-to-point microwave radio relay. At Brewster Flat, for example, three separate microwave channels in the 11-GHz range — one for voice, one for television, and a protection channel — carry communications to and from the site. A parallel microwave system in the 2-GHz band carries order wire and control signals.

Multiplexed voice channels are interconnected to ground station equipment at the standard group frequency level (60-108 kHz). Comsat equipment further processes these voice signals, along with wideband television signals, to be fed to separate radio transmitters. The outputs of these transmitters, in the 5925 to 6425-MHz band, are combined through the power amplifier before they are supplied to the ground station antenna. Received signals are handled in essentially the reverse manner.

Signaling

Increasing global communication through satellites focuses attention on the technical necessity for agreement between many countries. One of the most difficult technical problems in international telecommunications is signaling. Successful interface with large numbers of unique national switching systems must be achieved.

Members of the International Telegraph and Telephone Consultative Committee (CCITT) have agreed on a system to send line or supervisory signals at two in-band frequencies in a link-by-link arrangement. The register or numerical signaling, also link-by-link, uses two out of six in-band frequencies.

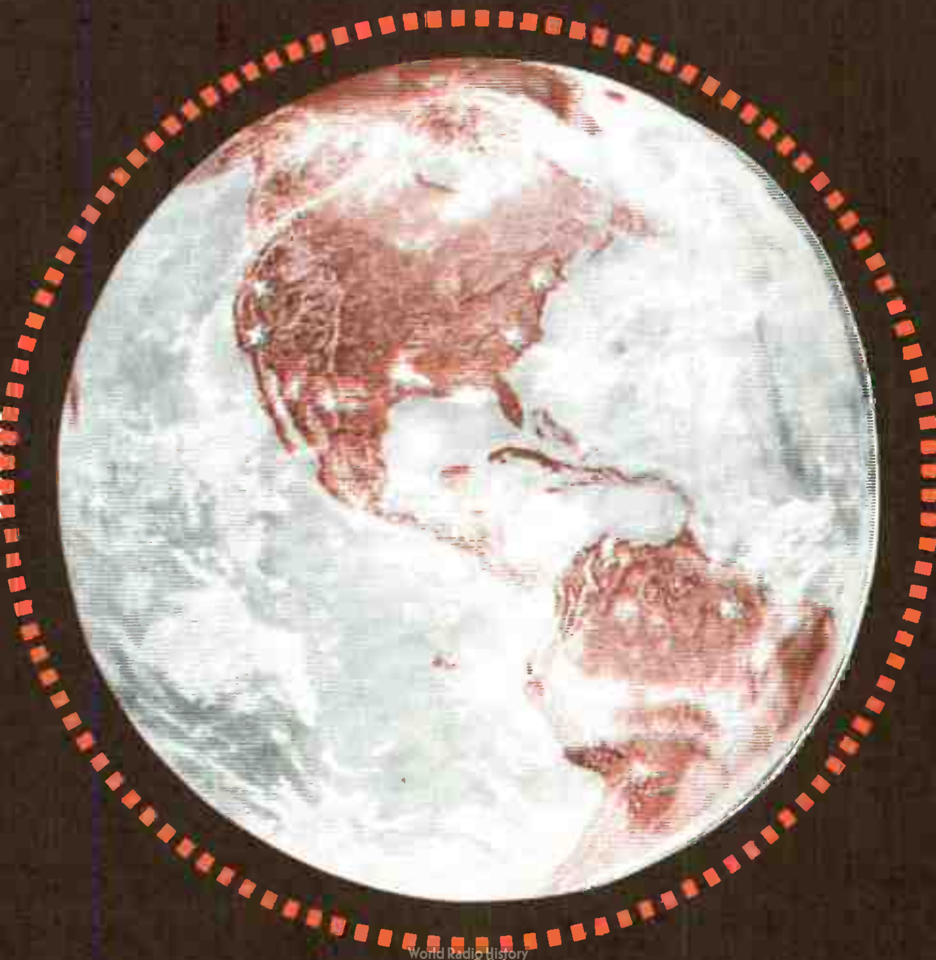
A proposed revision, if adopted, will mark the first time a completely new equipment design has been undertaken by international committee action. The new system would transmit signaling over a common channel separated from voice circuits.

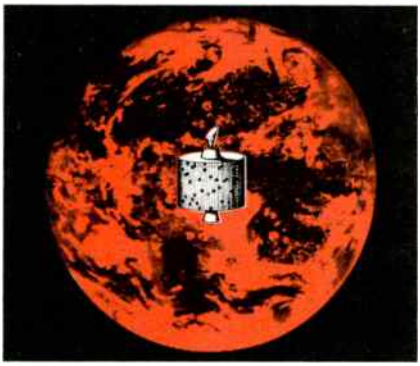
Discussion continues on future methods of handling carrier group and supergroup pilots, necessary for level regulation. CCITT recommendations place the pilots near the middle of the group frequency bands, but the Bell System has moved them to the edge to allow for wideband data service. The need for coordination between such varying systems has been accentuated by satellite communications.

The Lenkurt DEMODULATOR

AUGUST 1969

DEVELOPMENTS IN SATELLITE COMMUNICATIONS





In the four short years since the launching of Early Bird there have been dramatic advancements in both satellite and earth station technology.

The Communications Satellite Act of 1962, a unique piece of legislation recommended by the late President Kennedy, called for the establishment of a new and unprecedented private corporation to act as an instrument of the United States in establishing a world-wide communications satellite system as quickly and expeditiously as practicable.

This Act by Congress did not attempt to prescribe the nature or form of the international arrangements, but left this responsibility to the President, the Department of State, and a new corporation which was formed a few months later, the Communications Satellite Corporation (COMSAT).

After holding discussions with a number of countries around the globe, COMSAT drafted the Interim Agreements and opened them for signature in August of 1964. The original body of signatories representing 11 countries, now represents over 68 countries and is known as the International Telecommunications Satellite Consortium (INTELSAT).

In the brief time since its founding, INTELSAT has become the largest international joint venture ever undertaken.

Under the Interim Agreements, COMSAT represents the United States in INTELSAT, has 53 percent ownership in the system, and acts as manager for the consortium. The Agreements provide joint ownership of the space segment of the global system, with voting power being proportionate to ownership of participants as representatives of the Interim Communication Satellite Committee (ICSC). Membership of the interim committee is composed of space segment owners or combinations of owners having quotas of 1.5% or more. The committee is the executive organ of INTELSAT while COMSAT is the manager.

Geo-Stationary Satellites

The successful launching and orbiting of Telstar and Relay in 1962, with their low, non-synchronous orbits and brief 30 minute transmission periods assured the future of active satellite repeaters.

Under a contract from the National Aeronautics and Space Administration (NASA) the synchronous*, geo-

*A satellite in synchronous, geo-stationary orbit travels around the earth at 6,870 miles per hour in a circular, easterly path 22,300 statute miles above the equator. Because it travels at the same angular velocity around the earth's center, it appears to stand still – hence the reference to such satellites as being "fixed" or "stationary".

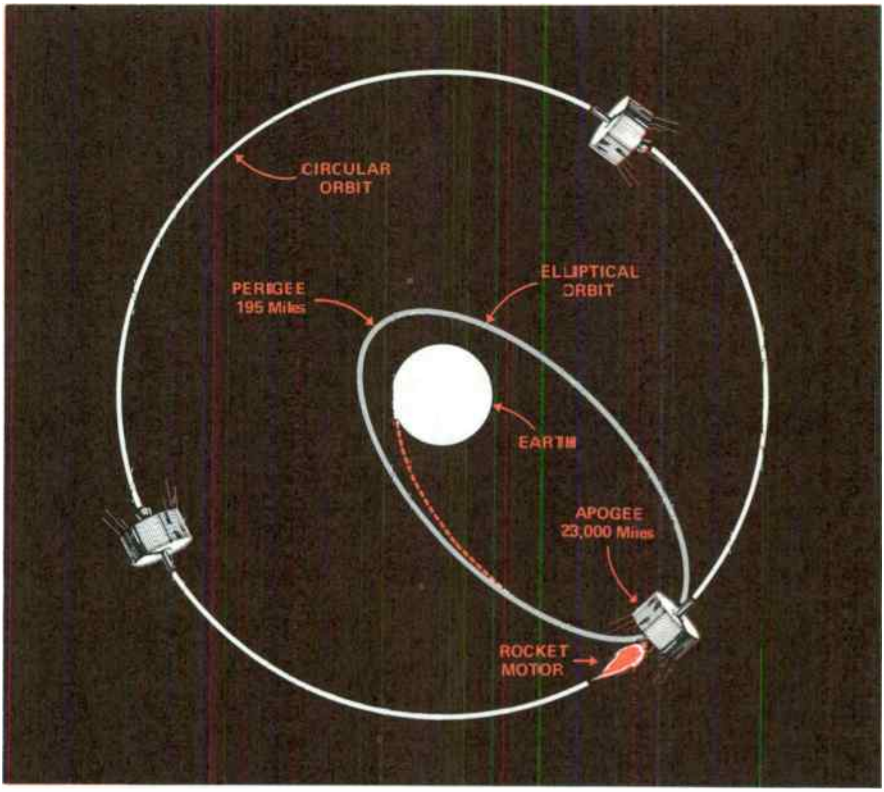


Figure 1. Accomplishing a synchronous, geo-stationary orbit requires the satellite be launched in an elliptical orbit, whose apogee aproximates 22,300 statute miles. When the satellite approaches this point an apogee rocket motor, on command from the ground, is fired emplacing the satellite in "fixed" orbit. Additional thrusters are employed from time to time to maintain proper attitude and earth alignment.

stationary satellite series termed SYNCOM was developed. SYNCOM I, intended for a synchronous equatorial orbit 22,300 miles above the earth early in 1963, failed to operate. SYNCOM II was launched later that year. SYNCOM III was placed in a stationary, equatorial orbit over the Pacific Ocean in mid-1964. The SYNCOM series dispelled any final doubts about the feasibility of a geo-stationary satellite communications network.

It is an interesting historical note that the concept of a geo-stationary satellite was first published in October, 1945 in an article entitled "Extra Terrestrial Relays" by British scientist, A.C. Clarke. Although Mr. Clarke's article seemed pure fiction to many at the time, it now stands as a prophecy to the events which have come to pass.

Clarke's idea of a satellite 22,300 miles above the earth in a stationary 24-hour orbit anticipated the actual event by some eighteen years.

The world's first commercial communication satellite was orbited April 6, 1965. This synchronous satellite, named Early Bird, was placed in a geo-stationary equatorial orbit over the Atlantic Ocean. From its vantage point 22,300 miles above the Atlantic, Early Bird linked North America and Europe with 240 high-quality voice circuits and made live television commercially available across the Atlantic for the first time. Early Bird, although the oldest communication satellite, has far exceeded its expected lifetime of 18 months and is still operating.

Since Early Bird

In the four short years since Early Bird was launched, six larger and more powerful satellites have successfully been placed in operation over the equatorial regions of the Atlantic, Pacific and Indian Oceans.

Early Bird, designated INTELSAT I, was the first of the INTELSAT series. By September 1967 four INTELSAT II satellites had been launched; the first of these failed to achieve orbit due to a malfunction of the apogee motor. The second was successfully parked in orbit over the Pacific Ocean, while the third was positioned, as planned, 6° W longitude over the Atlantic Ocean. The fourth of the INTELSAT II satellites was emplaced above the Pacific Ocean at 176° E longitude, only 2° W of the previous Pacific Ocean satellite.

INTELSAT I and II satellites were designed with 240 channel capacities, about one fifth the channel capacity of the INTELSAT III satellites which followed.

By May of 1969, three INTELSAT III satellites had been placed in orbit, one each over the Atlantic, Pacific and Indian Oceans. These three solar powered, 130 watt, giant satellites are capable of handling 1200 separate two-way telephone conversations, or four TV channels, or any combination of the two. In actual practice, part of their bandwidth has been assigned for TV use and the remainder for telephone, telegraph, data and facsimile communications.

There were seven commercial INTELSAT satellites over the equator in fixed, operational orbits by June of 1969 (See Figure 3). The objective, according to COMSAT's schedule, is to have an additional INTELSAT III satellite in orbit (over the Atlantic Ocean) by fall of 1969.

In contrast to the communications system of the INTELSAT III series, which includes two earth coverage transponders, the INTELSAT IV satellites will have a communications system that includes twelve transponders. Each of these receiver-transmitters is an independent frequency translator and amplifier. One of the unique features of the INTELSAT IV satellites will be the ability to switch the transponders in orbit to cover the precise earth area desired. This flexibility will make it posssible to adjust the satellite's capability to changing communications requirements. (Figure 2).

Using solar cells, the INTELSAT IV will generate more than 500 watts of power, about four times that of the INTELSAT III. This new model is expected to provide about 3600 circuits if all transponders are used for earth coverage, but as many as 9000 circuits if maxiınum use of spot coverage is made. It is expected that the mix of transponders will be such that each of these satellites will have a functioning capacity of at least 5000 circuits.

Earth Stations

The possibilities confirmed by the success of the synchronous satellite were revolutionary. With a minimum of three satellites, global coverage could be achieved. This would provide telephone, radio and television services, a vast communication and navigation service for aircraft and shipping, and many more applications.

The immediate implication of geostationary satellites greatly concerned earth station technology and mechanics. The huge tracking station antennas would require only very limited movement to follow a "fixed" satellite. This

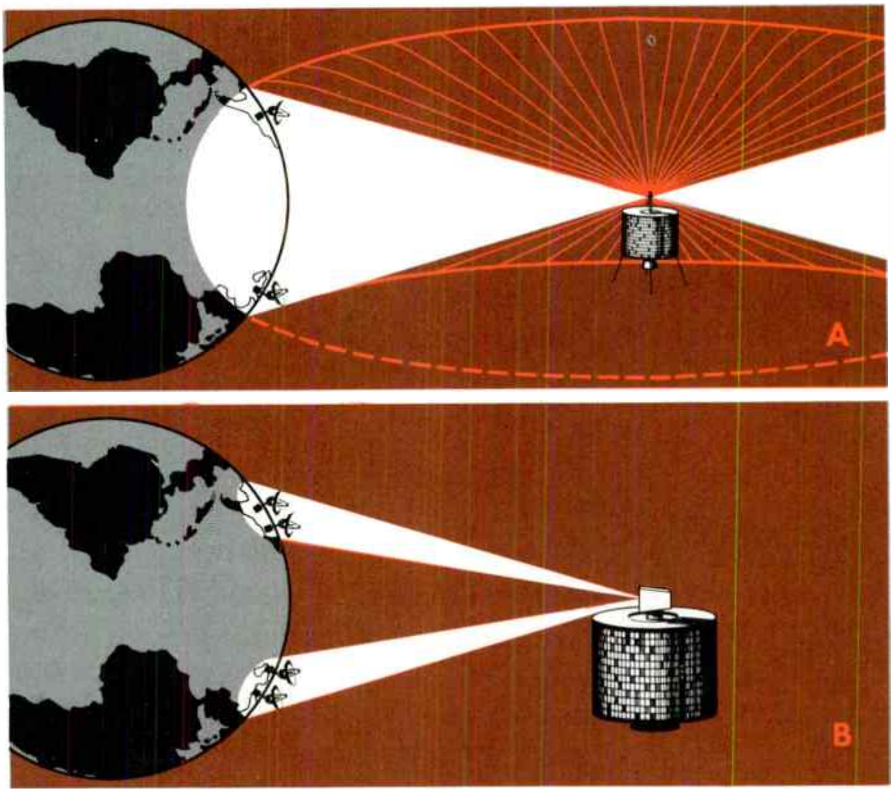


Figure 2. Early Bird employs a toroidal transmission pattern (A) which loses much rf energy to space. The proposed INTELSAT IV satellites will employ a more sophisicated and efficient transmission pattern (B).

would permit tremendous savings in the cost of material and maintenance. Further, because the tracking station would be operating with a synchronous satellite, the variable path-delay compensation required for operating a system of moving satellites as well as the complex hand-over problems created as a satellite passes from one earth station area to another, would become unnecessary.

Simplifying earth station technology, particularly the complex tracking requirements, not only allowed more attention to be focused on the development of new and better radio equipment, it realistically meant that many more earth stations could be erected in less time at less expense.

When Early Bird went into service in 1965, there were only a few experimental earth stations in the United, States, Japan, and Europe. By mid-1969, 25 earth stations were operating in 15 different countries. And 21 additional earth stations in 14 countries were either under construction or contract. This made a total of 46 stations in 29 countries (See Figure 3), around the world, with many more in the planning stage.

In spite of the fact that the INTELSAT group is not part of the space segment, it is understandable that they necessarily dictate the critical characteristics for each earth station interrelating with its satellites. A "standard" station specification has been

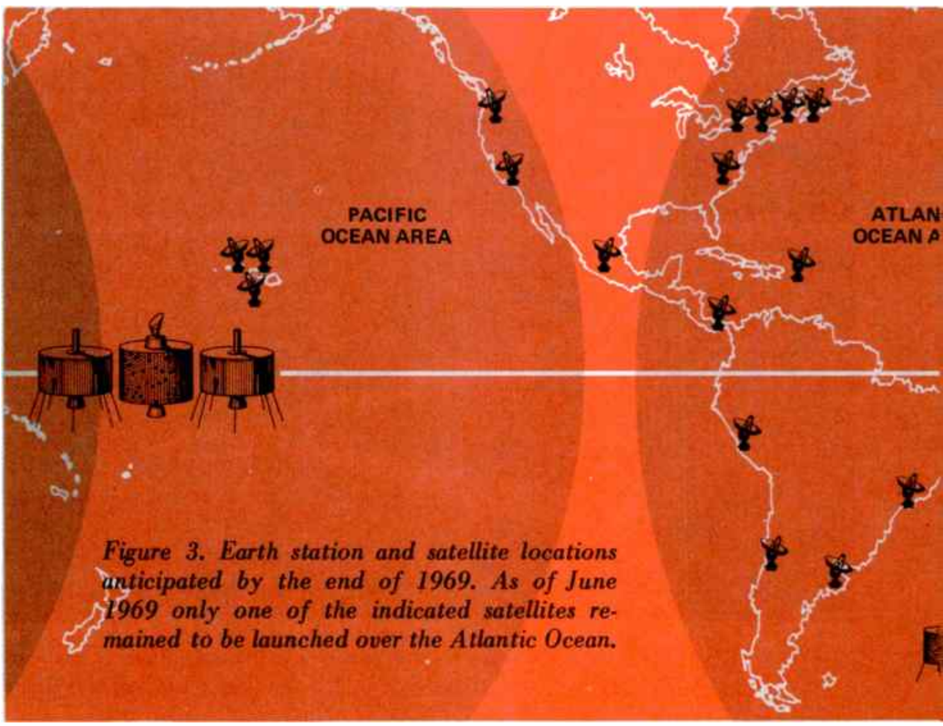


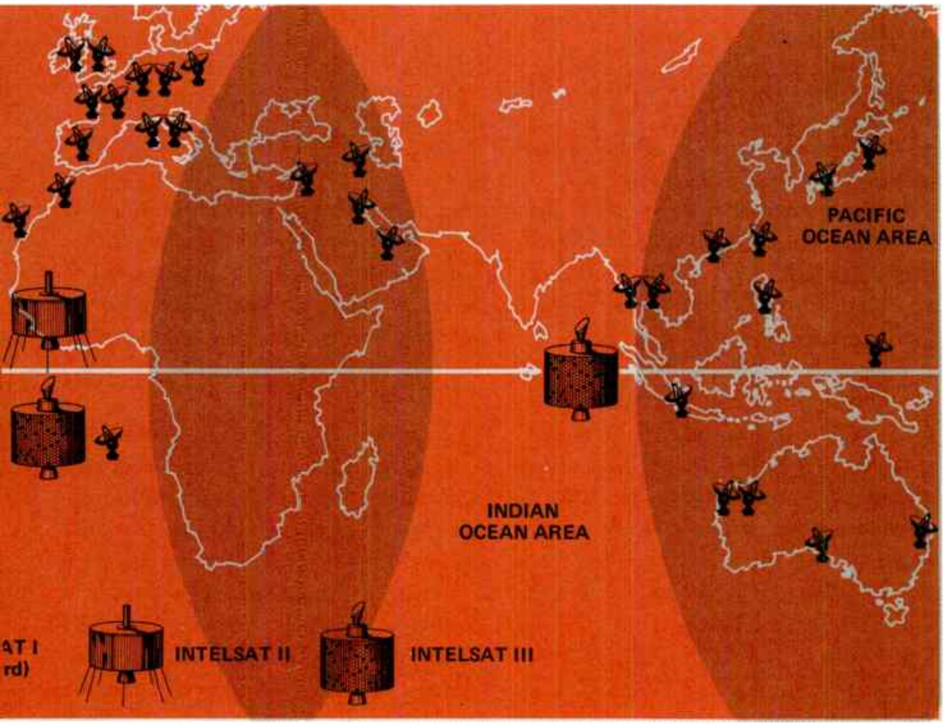
Figure 3. Earth station and satellite locations anticipated by the end of 1969. As of June 1969 only one of the indicated satellites remained to be launched over the Atlantic Ocean.

written to cover technical characteristics which may in any way affect the operation of the space segment or the use of it by any other earth station. Many features are specifically defined, such as the ratio of antenna gain to noise temperature, side lobe levels, maintenance of the transmit e.i.r.p. (effective isotropic radiated power) in the direction of the satellite to within ±0.5 dB of the nominal value, polarization, tracking modes, steering capabilities, RF out-of-band emission and amplitude frequency characteristics — to mention a few.

Improvements in earth station technology have not always received the same publicity as the launching and orbiting of the various satellites because of the more spectacular nature of the latter. There have been, however, some rather exciting developments in earth station systems. Some of these have been in specific response to unique problems accentuated by the satellite system. Lenkurt's 931C Echo Suppressor was especially designed for this type of service. The solid-state 931C suppresses the echos which are a problem partly due to the tremendous distances radio signals travel through space, to and from the extra-terrestrial satellites.

Molniya and Others

The Soviet Union, on October 4, 1957, launched Sputnik I. This event heralded man's first step into the field



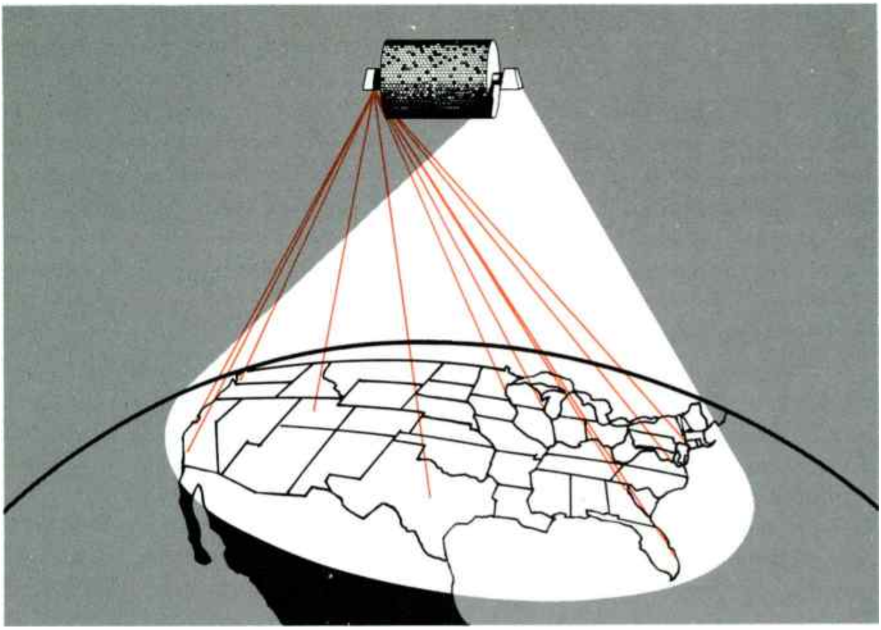


Figure 4. Advanced domestic satellites will use more efficient "direct beam" transmission.

of earth satellites with special attention going to the Soviet Union.

Only seventeen days after the launching of Early Bird, the Soviet Union successfully orbited their Molniya I on April 23, 1965. Molniya I was Russia's first experimental communications satellite. This active repeater satellite followed a highly elliptical orbit ranging from about 300 miles above the earth to more than 24,000 miles transmitting all forms of communication between Moscow and Vladivostok.

To date, Russia has launched several Molniya-class satellites in 12 hour elliptical orbits. Successful transmissions of many types of signals, including color television, have been carried out in joint experiments with the French. While the Russian satellites have an apparent lack of channel capacity, they do boast, among other things, high-powered transmitters. The Molniya has a command receiver, 40-watt transmitter, two reserve transmitters, and two steerable parabolic antennas.

Of the various nations which are non-members to the INTELSATgroup, the Soviet Union is the most noteworthy. Even though they do not have a synchronous, geo-stationary satellite, the Soviet Union has constructed a network of earth stations and, by using several Molniya-class satellites, have established a domestic satellite communication system.

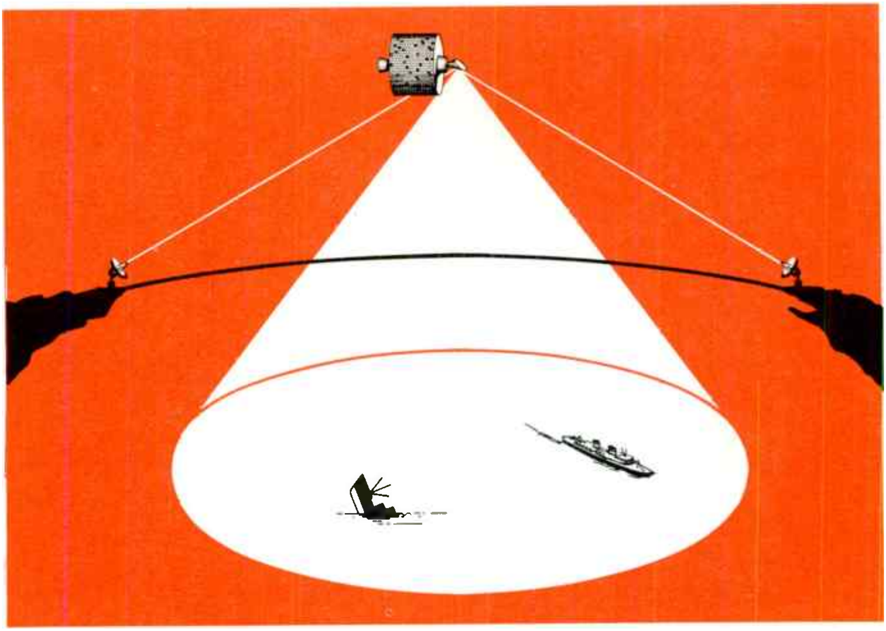


Figure 5. Unlike trans-oceanic cables, satellites can provide immediate communications at sea.

Frequency Utilization

All INTELSAT satellites presently employ radio frequencies within 500-MHz bands allocated for that purpose by the International Telecommunications Union (ITU) Radio Regulations. Earth station-to-satellite transmissions have been assigned frequencies in the 5925 to 6425 MHz band. Satellite-to-earth station transmissions use frequencies in the 3700 to 4200 MHz band. These frequency bands are shared on an equal right-of-use basis with terrestrial microwave systems.

The desirable range of frequencies for international communications lies within the range of 1 to 10 GHz. It is expected that the present band allocations will have been exploited within the next ten years.

A more efficient use of bandwidth has resulted from the increased power of the INTELSAT III and IV satellites. Even though this makes it possible to decrease the per channel bandwidth, it will still be necessary to look for new bands to allocate for exclusive satellite communications use.

Expansive Ideas

Modern satellites are capable of transmitting all forms of communications simultaneously – telephone, telegraph, television, data and facsimile. Satellites can operate competitively with cable networks and have the advantage of multipoint, multiple access. This means that synchronous satellites, in geo-stationary orbits can make possible direct communications between all countries having earth stations within a satellite's line of sight, eliminating much of the circuitous routing which has been so characteristic of international communications. A large number of innovations and expansions can be expected during the next decade.

A pilot domestic satellite program has been proposed to make available the benefits of satellite technology to the people of the United States. The development and use of "direct beam" transmission (See Figure 4) is now being planned.

In addition to people, computers in one country are now talking with increasing frequency to computers in other countries, and at speeds 20 times greater than anything possible by more conventional means.

High-quality telephone service is now becoming available from the United States to a growing number of countries that were previously difficult to reach by cable or short wave transmission.

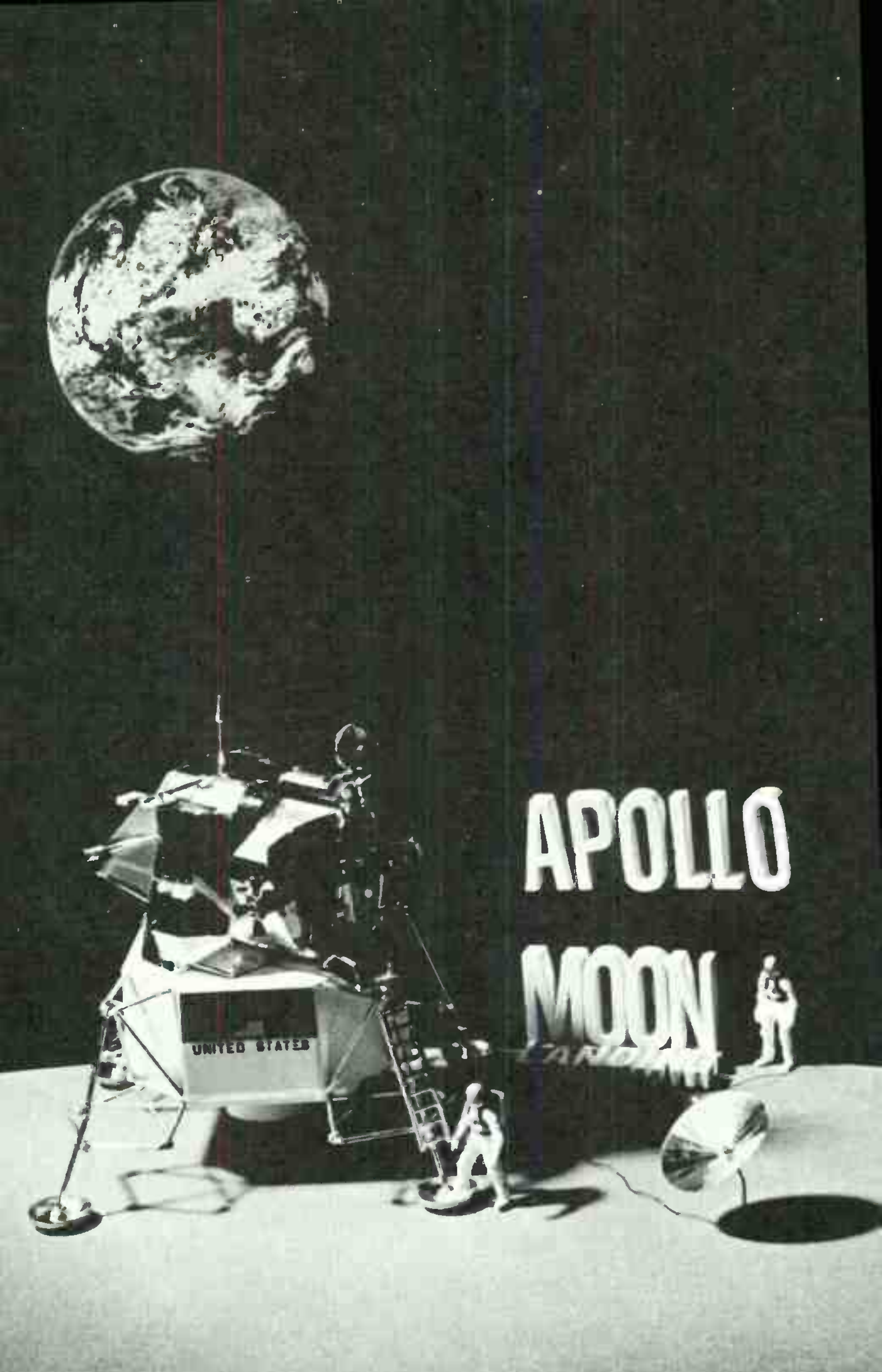
Entire news pages can be transmitted via facsimile across the ocean and reproduced a matter of hours after the original publication.

Weather maps are already being quickly transmitted via facsimilie from one country to another to assist intercontinental airline pilots in charting plans for flights.

Passenger and cargo data has also been transmitted across the Atlantic to customs officials in the United States while a plane was in flight. With all the necessary information on hand inspectors were able to speed passengers and cargo through customs.

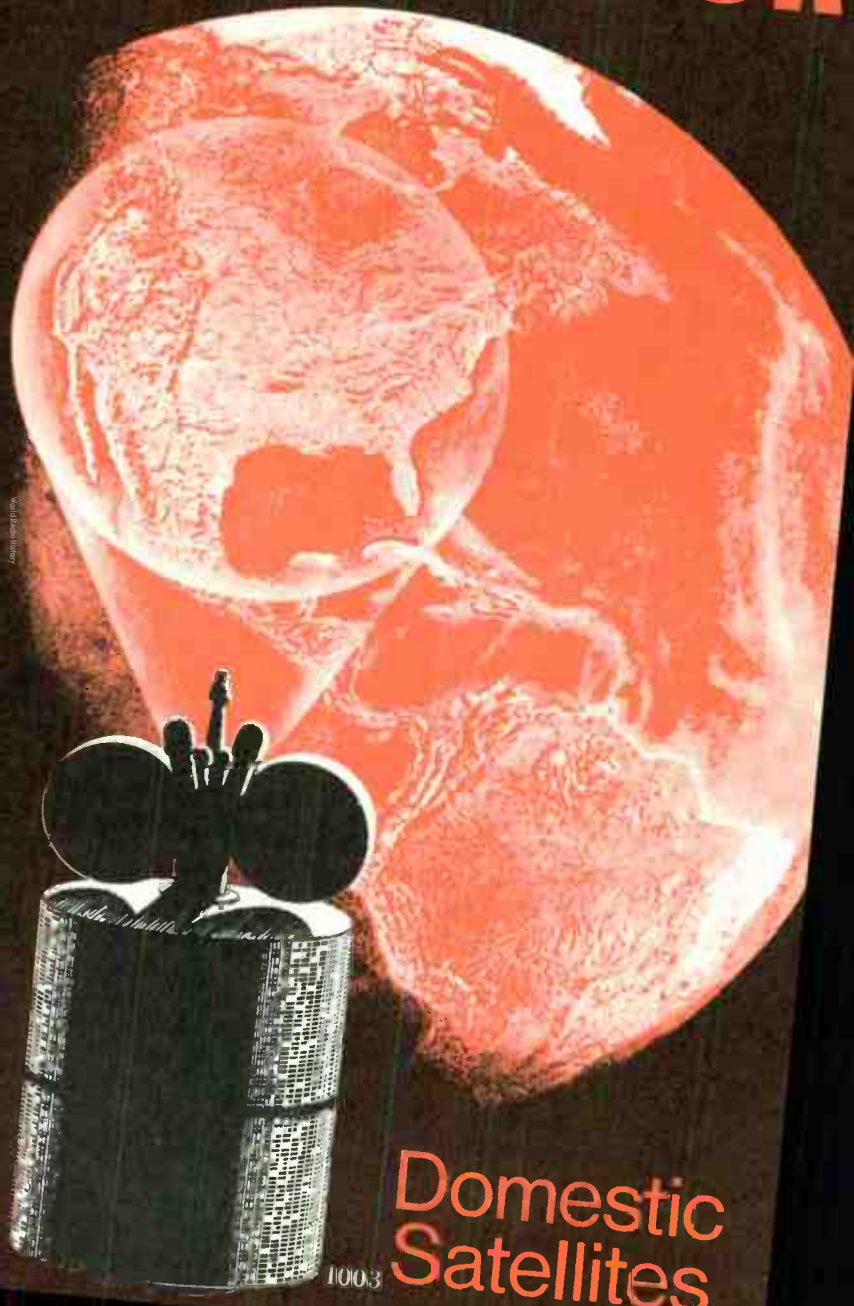
Satellites will soon make available maritime communications systems at sea which will compare, in quality and reliability with services now commonplace on land. This can mean the potential saving of hundreds of ships that are lost at sea each year, many without any communication.

The most dramatic role of the communications satellites is the part they play in the space program. Atlantic and Pacific satellites of the INTELSAT II series are a key part of the communications system developed for NASA's Project Apollo – the moonlanding program.


APOLLO
MOON
UNITED STATES

The Lenkurt® DEMODULATOR

JULY 1970



1003

Domestic Satellites

. . . . . . a versatile, reliable national communications network just over the horizon.

Synchronous, geostationary satellites capable of directing signals to a network of earth stations scattered throughout the United States is the image portrayed by a domestic satellite system (see Figure 1). Such a system could serve populous areas where the demand is greatest, or provide communication links to areas which are not now easily accessible with terrestrial methods.

The proposed system would provide communications for all areas within the country, and could someday be interconnected with trans-oceanic cable and international satellites to overseas points, making it possible for today's telephone user, even in remote areas, to reach 188 million telephones – 96 percent of the world's total.

Open Competition

The development of a domestic satellite system in the U. S. has been delayed pending the outcome of a government study. The results of this study have now been presented in the form of a Presidential memo suggesting the FCC give approval to any organization seeking to construct and operate a domestic satellite system, provided it meets certain guidelines.

The memo further suggests establishing a three- to five-year interim policy allowing competition to act within well defined limits to protect public interests.

Literally interpreted, the Presidential memo gives anyone with the funds and technology the opportunity to launch and operate a domestic satellite system, provided standards of compatibility are met and anti-trust laws are not violated. Specifically, the recommendations deal with financial ability, launching capability, room in space, and available frequencies.

When satellites are designed to fit a domestic network, they must compete, costwise, with existing service. Satellites can provide circuit performance and capacity equal to coaxial cable, digital systems and millimeter wave systems.

A satellite system able to compete economically with terrestrial facilities will have a limited number of drop-off points (earth stations). The earth stations should be limited for two reasons – cost and channel capacity. The channel capacity decreases with increasing station access; therefore, each earth station will have to gather traffic from a large area.

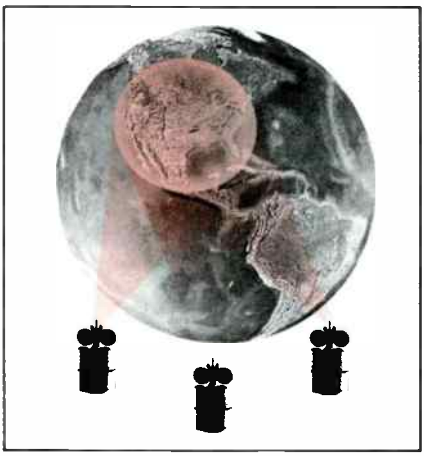


Figure 1. A domestic satellite system would use a series of synchronous, geostationary satellites for communication with all parts of a country.

The private enterprise approach to domestic satellite service may lead to a variety of special purpose systems. Three general plans are being considered, each in a different frequency range – 4 and 6 GHz, 10 to 40 GHz, and 30,000 GHz. The first system would operate in the 4- and 6-GHz range – presently used for most terrestrial radio transmission and for international satellite systems involving a few remotely located earth stations.

Radio Interference

Radio interference can be a serious problem if there is an extensive microwave network near the earth satellite station. This condition is typically found in the proximity of urban centers. In the United States, the proliferation of 4- and 6-GHz terrestrial systems makes these undesirable for satellite use. However, there are not as many 4- and 6-GHz terrestrial links in Canada, and their proposed system, using these frequencies, expects to avoid radio interference by placing the earth stations outside the metropolitan areas. Even with the interference shielding offered by hills, it will probably be necessary to place earth satellite stations 50 to 100 miles (80 to 160 kilometers) from urban centers.

Radio interference affecting a U.S. domestic satellite system operating in the 4- and 6-GHz region is shown in Figure 2. The greatest interference is between the 4-GHz radio relay transmitter and the highly sensitive earth station receiver, and between the high-gain earth station transmitter and the 6-GHz radio relay receiver.

International studies are in progress to find ways to avoid radio interference, with emphasis on the possible selection of preferred or segregated frequency assignments for satellite communication systems. The bands under study are above those generally used for terrestrial microwave systems. If exclusive assignments can be made for satellite service, earth stations can be placed near large centers where most circuits will be terminated.

Bell has studied a system that operates in the millimeter-wave frequency range between 10 and 40 GHz. Radio interference is no longer a hinderance with such a system. Atmospheric attenuation, however, is a much more serious problem, since electromagnetic waves in the frequency bands above 10 GHz are severely attenuated by rain and water vapor.

Atmospheric Attenuation

A domestic satellite system operating above 10 GHz must be designed to withstand a few dB of attenuation due to atmospheric conditions – sometimes for long periods – and must have a diversity earth station for the rare occasions when excessive rainfall causes large attenuation (See Figure 3). According to studies, the most intense rain occurs in limited

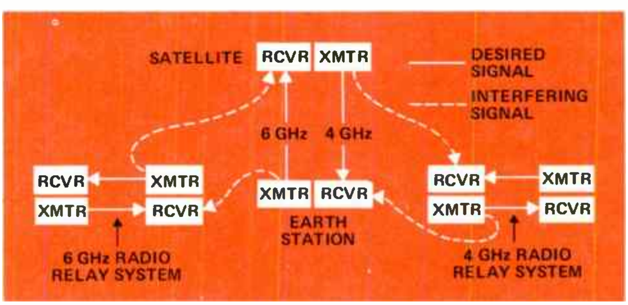


Figure 2. Radio interference is a serious problem when the earth station for a 4- and 6-GHz satellite system is located near an area with 4- and 6-GHz radio relay systems.

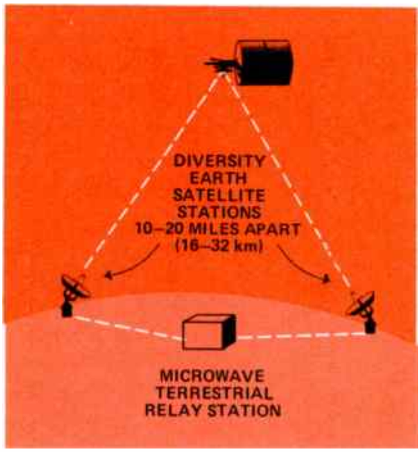


Figure 3. Diversity earth satellite stations avoid signal attenuation when heavy rainfall occurs.

cells, and rain covering large areas (several square miles) generally falls at the low rate of one inch per hour or less. Therefore, diversity earth stations separated by several miles have been proposed as a workable solution to atmospheric attenuation.

A third possible ground-to-satellite link would use a CO_2 laser. Such a system is not hampered by radio interference, and has a high tolerance against atmospheric attenuation. Although the frequency of a CO_2 laser (30,000 GHz) is higher than millimeter-waves, there is a transmission window, 40-GHz wide, centered at 30,000 GHz. This frequency, therefore, is less susceptible to attenuation than any in the visible or ultra-violet ranges. The CO_2 laser has also been suggested as a means of providing efficient inter-satellite communications – links between domestic and international systems

Time Delay and Echo

Long time delay and the associated echo became apparent with international geostationary communications satellites. The minimum distance between any two points via a geostationary satellite is 44,600 statute miles (72,000 kilometers). Consequently, a U.S. circuit via satellite will have a round trip delay of more than one-half second compared with about one-tenth second delay for terrestrial cross-country transmission in the U.S. This delay is due to the distances involved and the resulting transmission times.

The speaker's echo tolerance depends on the delay time and the loudness of the echo. There are two ways to suppress the echo within tolerable limits. One form of suppression is to attenuate the echo – making it barely noticeable compared with the speaker's voice. As time delay increases, the echo attenuation must also be increased. A voice-activated switching device can also be used in the return circuit to keep the echo from reaching the speaker.

In the early 1960's, Lenkurt Electric, Bell Laboratories, and the Dollis Hill Laboratories of the British Post Office were studying the effects of delay, echo, and echo suppression. These studies resulted in specifications for new suppressors designed for long-delay circuits. Lenkurt's 931C echo suppressor was designed to meet these specifications and is capable of compensating for delays of the magnitude encountered with satellite links.

Artificial delays were used in a simulated telephone test circuit carrying regular telephone traffic. These experiments began to show slight adverse public reactions to calls with a round trip delay of about 300 milliseconds, and a significant increase in adverse reactions with a round trip delay of about 500 milliseconds.

Three different situations arise with long delays which may be disturbing, but are tolerable. The first of these is called "simultaneous talking." If both parties start talking within one-quarter second of each other, both will continue talking until one party finally notices the other and ceases talking. When this happens, neither party will hear what the other has said.

"Hello calling" is the second condition encountered with long delays. When one party has been talking for some time, or has come to the end of what he wanted to say, he usually pauses and expects a response from the other party. This response may be delayed because the other party hesitates before answering. With the added satellite delay the talker may become impatient and start calling "Hello," indicating he is wondering if the other party is still on the line, or if the connection has been broken.

"Break-in difficulties" characterize the third delay situation. One of the parties may wish to start talking by taking advantage of a short pause in the other's speech. Therefore, he waits for a breathing pause by the other party. In a satellite call, it will take him about one-quarter second to note the pause. By the time his comments reach the original speaker, a minimum of another quarter second later, the latter may have resumed speaking. This condition is compounded if it leads to "simultaneous talking."

Although the distance to the moon is greater than it would be to a geostationary domestic satellite, most people were made aware of these long delays with the telephone conversation to the moon during the Apollo XI moon landing. All of these delay-related conditions may become more pronounced and lead to verbal communication difficulties under the pressure of time and argument.

Since subscriber dissatisfaction increases in proportion to delay time, the CCITT (International Telegraph and Telephone Consultative Committee) recommended the following limitations on mean, one-way propagation time with appropriate echo suppressors: 0 to 150 milliseconds, acceptable; 150 to 400 milliseconds, acceptable, provided increasing care is exercised on connections as the mean, one-way propagation time exceeds about 300 milliseconds; and unacceptable above 400 milliseconds.

A follow-up analysis on the simulated delay circuits showed that only a small proportion of the people had difficulty talking and hearing, and an even smaller proportion rated the connection "fair" or "poor" (see Figure 4). Therefore, it would seem that the delay-related problems are not as serious as first thought to be.

One possible way to minimize delay is to use the satellite path for transmission in only one direction, and use the shorter delay, terrestrial system for the return connection. In this way, the maximum one-way delay would be acceptable and the total delay would be significantly reduced.

Once a domestic satellite system is operating, it will be necessary to have automatic switching which would limit each call to only one satellite hop,

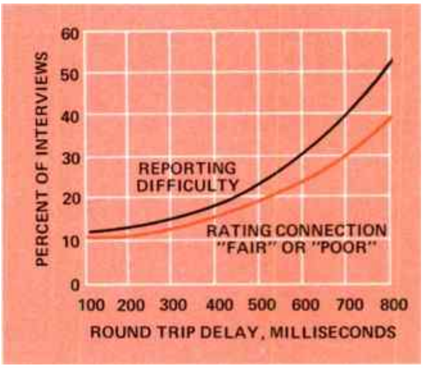


Figure 4. Only a small percentage reported having any difficulty hearing or talking during calls made on circuits with simulated delays. An even smaller portion rated the connection as "poor" or "fair".

keeping the delay within the acceptable range. One way to accomplish this is to make the domestic system available only for calls within the country. An alternative plan would develop an inter-satellite communication system requiring only one up-and-down link. The situation to be avoided is multiple up-and-down links. For example, on a call from England to Hawaii there could be as many as three up-and-down links, with a U.S. domestic system in operation (See Figure 5).

Long delays cause difficulties only when there is two-way communication. Data, television, and facsimile transmission would be unaffected by these delays and ideally suited to satellite communication systems.

Unique Outages

There are three types of outages affecting reliability on all satellite systems. The magnitude and exact occurrence of these outages depends upon the orbital placement of the satellite and the location of the earth station. The first is eclipse outage which occurs when the earth's shadow covers the satellite, causing the solar cells to become inoperative. These periods of eclipse last up to one hour, and occur each night for 43 consecutive nights in the spring and fall. However, the usual satellite design provides battery backup for most channels to insure their continuous operation. To conserve satellite weight, it would be possible to keep a channel without battery backup for television transmission – a service normally "off the air" during the eclipse periods.

Sun transit outage is caused by radiation of electromagnetic energy from the sun when it crosses directly behind the satellite. This radiation is proportional to temperature; therefore, the sun is an extremely powerful noise source which, when in direct line with a satellite, overrides the satellite signal. This condition, occurring on about five days, two times each year, causes an outage lasting approximately 10 minutes. Terrestrial protection channels can be provided to avoid losses, since these short outages can be predicted with reasonable accuracy.

If a satellite should fail, there is a distinct likelihood that it will be impossible to restore it to service, and a replacement satellite would have to be launched. A temporary means of restoring satellite circuits will have to be developed, to avoid the serious effects of this outage on a sophisticated telephone switching network.

Aircraft do not cause interference with terrestrial communication links. This is not, however, the case with satellite communication where aircraft corridors pass through satellite beams.

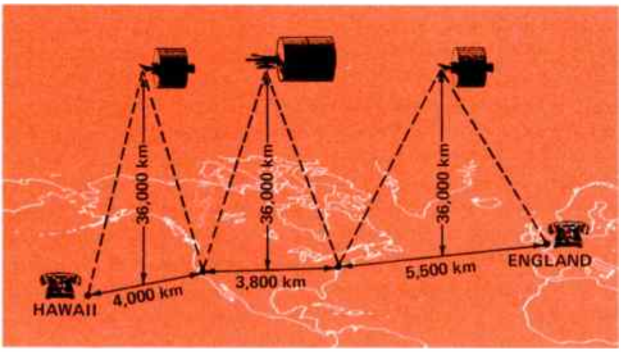


Figure 5. Limiting intercontinental communications to one up-and-down link, there are several possible routes for a call from England to Hawaii.

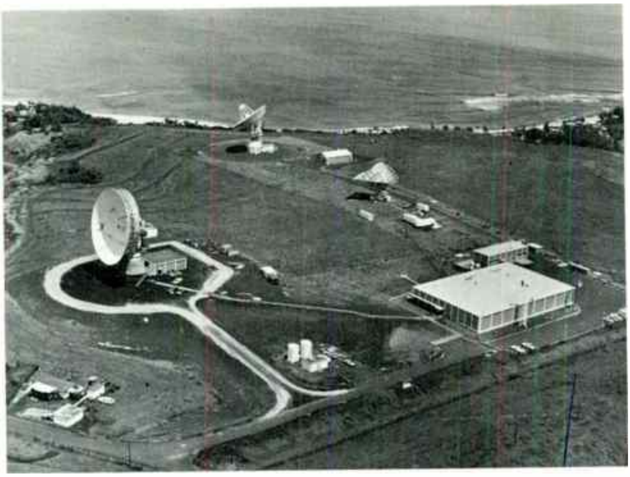


Figure 6. Domestic satellite earth stations will look very similar to this international satellite earth station.

This interference condition is still under investigation to determine its magnitude.

Weight vs Stabilization

Narrow antenna beams used for satellite communication require precise spacecraft stabilization. Accurate sensing for final attitude adjustments can be achieved by measuring the satellite's electromagnetic radiation. Attitude control appears to be primarily a question of the reliability of components to be used in a system designed for at least 10 years of operation. A basic aspect of attitude control is the amount of fuel (weight) required to stabilize the spacecraft.

In comparison with the amount of fuel required to keep a satellite station operating, the weight for attitude control is small – a few pounds per year per ton of satellite weight. Therefore, the stabilization weight penalty imposed on a satellite having a 10 year operational lifetime, while significant, is not prohibitive.

How Soon?

It is theorized that even if the FCC acts quickly, it would require about two years of planning and construction before a U.S. domestic satellite system could become operational. Experts predict the initial volume will be only great enough to support one satellite system, unless there is a significant increase in traffic over the next two years.

The Canadian domestic satellite system is scheduled for launching in late 1972. The specific requirements of the Canadian system are not the same as for a U.S. system; however, a great deal can be gained from their experience.

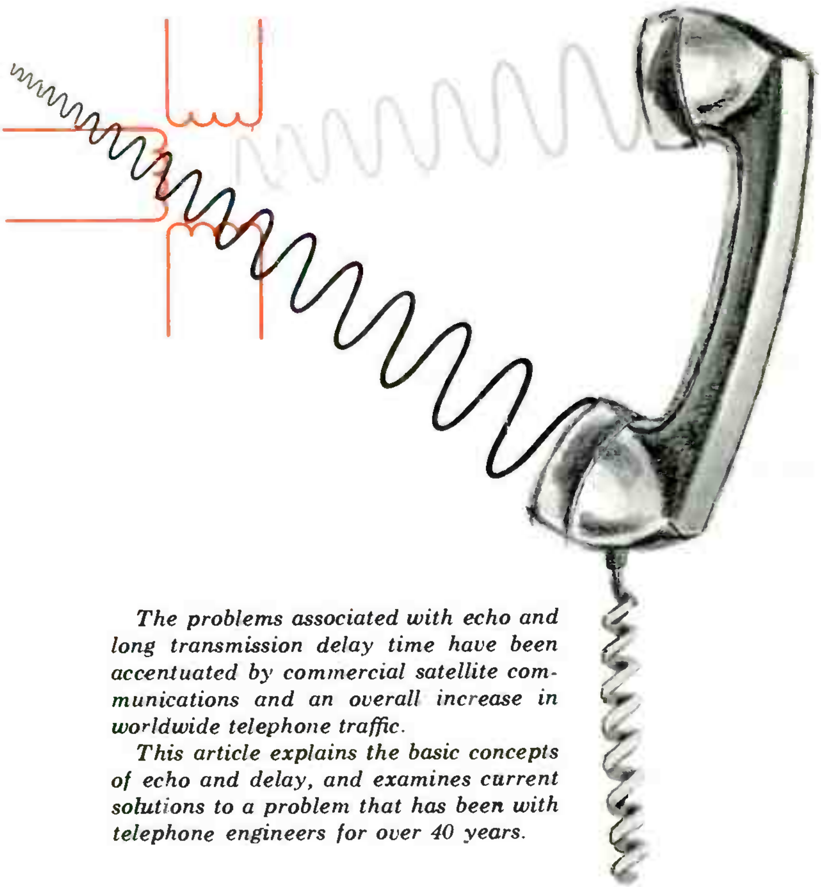
Although the cost of domestic satellite links is not expected to provide immediate economic advantages over terrestrial links, it is anticipated that its versatility and reliability combined with the present varied modes of terrestrial transmission will ultimately provide a more efficient total communications system.

Someday in the near future the United States will be covered with a network of earth stations similar to the one shown in Figure 6, and all phases of communication – voice, data, facsimile – will experience the advantages a domestic satellite system.

AUGUST, 1966
VOL. 15, NO. 8

the Lenkurt. Demodulator

Echo Suppression



The problems associated with echo and long transmission delay time have been accentuated by commercial satellite communications and an overall increase in worldwide telephone traffic.

This article explains the basic concepts of echo and delay, and examines current solutions to a problem that has been with telephone engineers for over 40 years.

Like the acoustic echo heard in a cave, or bounced from the side of a mountain, the echo in a long telephone circuit represents sound energy reflected back from some distant point. Unlike that found in nature, telephone echo is neither a pleasant nor desirable occurrence.

Telephone echo is created primarily at the far end of 4-wire transmission circuits where a junction is made with 2-wire subscriber loops. Because of unavoidable impedance mismatch at this point, energy transfer is not complete and some of the sound is reflected back to the talker. Thus, in the telephone receiver a talker hears his own voice, delayed proportionately to the length of the circuit.

Delay Time

Delay time, the basic factor causing echo to be objectionable, is a function of propagation rate and distance. The faster the propagation, the longer the distance that can be covered without serious degradation of the circuit. The upper limit, however, is the rate at which electromagnetic radiation travels in free space — 186,000 miles per second. And now communications satellites are proving that even this is not fast enough.

Delay is commonly measured in milliseconds (ms), or thousandths of a second. Long one-way delays of, say, 500 ms might be sensed by talkers only as a hesitancy in the response of the other party. But echo returned at a fraction of this time seriously degrades the circuit, and may cause the speaker to stammer, slur his words, or stop talking altogether. In fact, a round-trip echo delay of 45 ms is considered the maximum before some sort of echo suppression must be used. Of course, the speaker's tolerance to echo depends on both echo delay time and loudness. Echo with long delay and of sufficient magnitude is very noticeable.

Undoubtedly, the commercial satellite system of the future will be built around synchronous satellites whose orbital speed matches exactly the rotational speed of the earth. The satellite appears to "hang" in a stationary location over the equator at an altitude of 22,300 miles. A one-way telephone path through such a satellite is in the order of 50,000 miles, taking into account the geographical distance between ground stations. Delay between New York and London, for instance, is about 265 ms, or over a half-second for round-trip echo. This amount of delay has again accented the problem of echo suppressors.

Signal Path

The telephone path begins with a 2-wire loop from the subscriber's instrument to the local exchange office, and either 2- or 4-wire exchange trunks extending to the toll switching center. Long-haul 4-wire toll trunks eventually are returned to 2-wire subscriber loops at the receiver end of the circuit to complete the path to the called party's telephone. (See Figure 1.) In addition, signals are commonly returned to 2-wire circuits for more economical switching purposes.

Wherever it is necessary to match 2-wire to 4-wire circuits, hybrid transformers are used (Figure 2). Because of impedance irregularities at the hybrid, a certain amount of reflection occurs and echo is produced. Of primary concern is the echo resulting from the mismatch at the far end of the circuit.

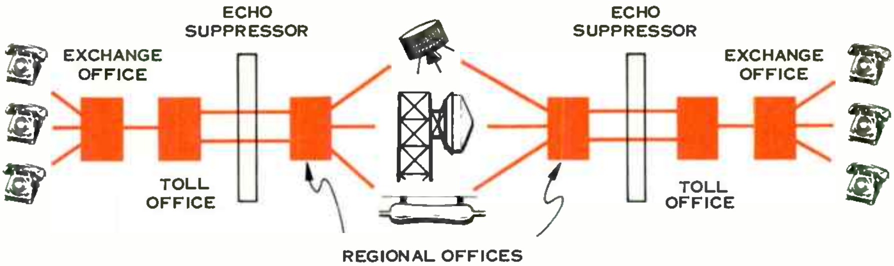


Figure 1. Telephone path through exchange, toll and regional offices. Echo suppressor is located at the beginning of 4-wire toll circuits.

Echo Return Loss

The terms transhybrid loss, return loss, and echo return loss are often used in discussing the operation of hybrid 4-wire terminations. Transhybrid loss is defined as the loss directly across the hybrid on the 4-wire side; i.e., the isolation between the transmit and receive branches of the 4-wire line. The return loss is the transhybrid loss *minus* the losses in the 4-wire to 2-wire paths of the hybrid and serves as a measure of energy returned to the talker. Echo return loss is measured with a known input of voice-weighted noise, between 500 and 2500 Hz. Echo return loss must be greater than 27 dB to meet toll switching requirements. The mean echo return loss at end offices is about 10 dB. (For further discussion, see the *Demodulator*, January 1964.)

It is possible to limit echo in short delay circuits by adding path attenuation. This loss will, of course, reduce the talker's signal as well, but echo returned through the same path will be attenuated twice as much. Until a few years ago circuits with up to 45 ms delay could be accommodated by the insertion of up to 14 dB total one way loss, known as terminal net loss (TNL). Beyond that point, further attenuation interferes seriously with the transmission of speech. Recent upgrading of nationwide service has reduced the amount of attenuation to be placed in high usage trunks. It is now policy to use echo suppressors in circuits having even 20 ms delay, or less.

First Suppressors

Original echo suppressor work began in the 1920's when the first transcontinental telephone systems were being planned. Carrier equipment was not commonly available and most transmission was at voice frequency, having the relatively low propagation rate of about 20,000 miles per second on loaded cable. At that speed a signal would travel a 900-mile circuit from New York to Chicago in 45 ms, or have an echo delay of 90 ms. As trunk lines were expanded across the country, the need for echo suppression became more and more obvious. But before echo equipment approached any degree of sophistication, carrier systems were developed, with their higher frequency increasing the propagation rate to over 110,000 miles per second. At this speed New York to Chicago is only 8 ms, or 16 ms echo delay.

The innovation of carrier transmission and its characteristically faster propagation speed greatly reduced the pressure on further echo suppressor development. Only years later as coast-to-coast trunks and longer undersea cables were established, did interest in echo suppression again take on new zeal. Even at 110,000 miles per second the round-trip echo delay from New York to London was about 70 ms—far too much to be handled by terminal net loss techniques.

Since it was impossible to reduce delay, impractical to eliminate echo at terminal points, and not feasible to introduce sufficient path loss to suppress echoes without reducing talker volumes to imperceivable levels, only one solution became immediately obvious. It was necessary to block the return path an echo must take. The first echo suppressors did just that, using amplified voice energy to activate a relay shorting the opposite path. More refined versions, such as the Bell 1A echo suppressor, detected relative speech energy between the two paths, and picked the stronger of the two to activate suppressor controls. In this case, rather than shorting the line, the suppressor introduced 40 to 50 dB loss in the return path. Because it was almost impossible for a second speaker to break in on a conversation, the circuit took on the qualities of a "push-to-talk" or simplex operation.

Split Suppressors

Echo suppressors were originally designed to be placed at the midway point of the voice-frequency circuit, but with carrier transmission and ocean cables, this was not feasible. First, suppressors were moved from the midpoint to one end of the circuit. Later, the "split" suppressor allowed one-half of the unit to be located at each end, providing some advantage. In all suppressors, attenuation is maintained after the party stops talking for the period of time necessary for the signal to make its round trip and return to the suppressor. This is called *hangover* time. By using split suppressors, hangover time can be held to a minimum, since one unit is always near the reflection point at the far end. Today, echo suppressors are commonly placed in toll or regional switching offices, and may also be found in future communications satellite ground stations.

One problem inherent in previous suppressors was *lockout*. If circuit switching resulted in two suppressors working in tandem—not at all unlikely, especially with Direct Distance Dialing—it was possible for each talker to take command of the suppressor nearest him, block the opposite

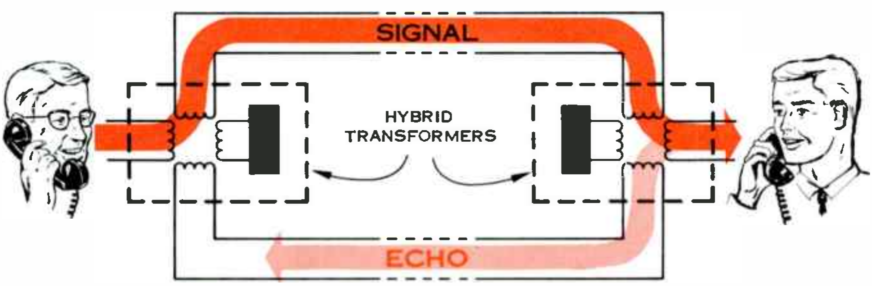


Figure 2. Hybrid transformers are needed at each end of a 4-wire circuit. Reflected energy at this point becomes echo.

path, and thereby prevent any further conversation until one of the parties stopped talking. Split suppressors have also reduced this possibility, but an extension of the problem still exists. A series of echo suppressors introducing individual path losses of about 12 dB could quickly add enough attenuation to reduce speech levels below acceptable values.

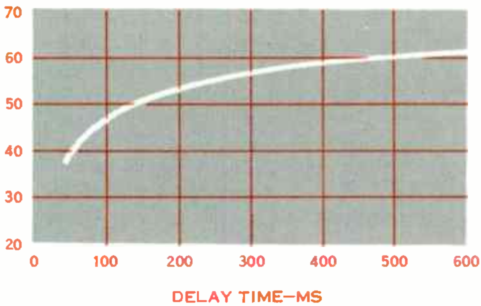


Figure 3. The attenuation needed to suppress echoes to tolerable levels is indicated for the average telephone talker.

Lenkurt 931B

In developing the 931B Echo Suppressor, Lenkurt design engineers were able to include the desirable feature of bi-directional operation — both parties talking at the same time—and minimize speech clipping at the beginning, and speech chopping in the middle of conversations. All modern echo suppressors have since adopted this bi-directional mode of operation.

The 931B suppressor has two modes of operation, allowing it to discriminate between single-party talking and two-party talking. With only one subscriber talking the suppressor is in Mode 1, and blocks echo by inserting 60 dB loss in the return path. Mode 2 provides for two-party talking by only partially suppressing echoes in both paths, on the assumption that echoes of short duration will be masked by the speech signals of the other party.

Mode 1

Block diagrams of the Lenkurt echo suppressor in both modes are shown in Figure 4. Identical units are placed at each end of the communications link, and contain two variable-gain amplifiers, an echo control switch, and associated control circuits. To analyze the set's operation, assume Station A is talking. A's voice operates the transmit control circuit closing both loss switches, producing a no loss, or unity gain condition in the transmit path. The echo control switch is normally closed.

At Station B, the signal operates the receive control circuit, opening the echo control switch. The loss switch remains closed and the signal arrives at the second talker unattenuated by the suppressor. The echo signal, reflected at the hybrid, is sensed by the transmit control circuit at Station B, but because of loss in the hybrid, lacks the energy to overcome the variable reference bias supplied from the receive path and is blocked by the echo control switch.

Mode 2

Mode 2 occurs when B attempts to interrupt A's conversation. B's speech energy is high enough to operate the transmit control switch on his end of the circuit. This results in closing of the echo switch, and the opening of loss switches in both paths at Station B. Similar action occurs as B's signal arrives at Station A. In Mode 2, with both parties talking, each speech signal is attenuated by two losses, and each echo signal reduced by four losses. These losses may be strapped in 1-dB

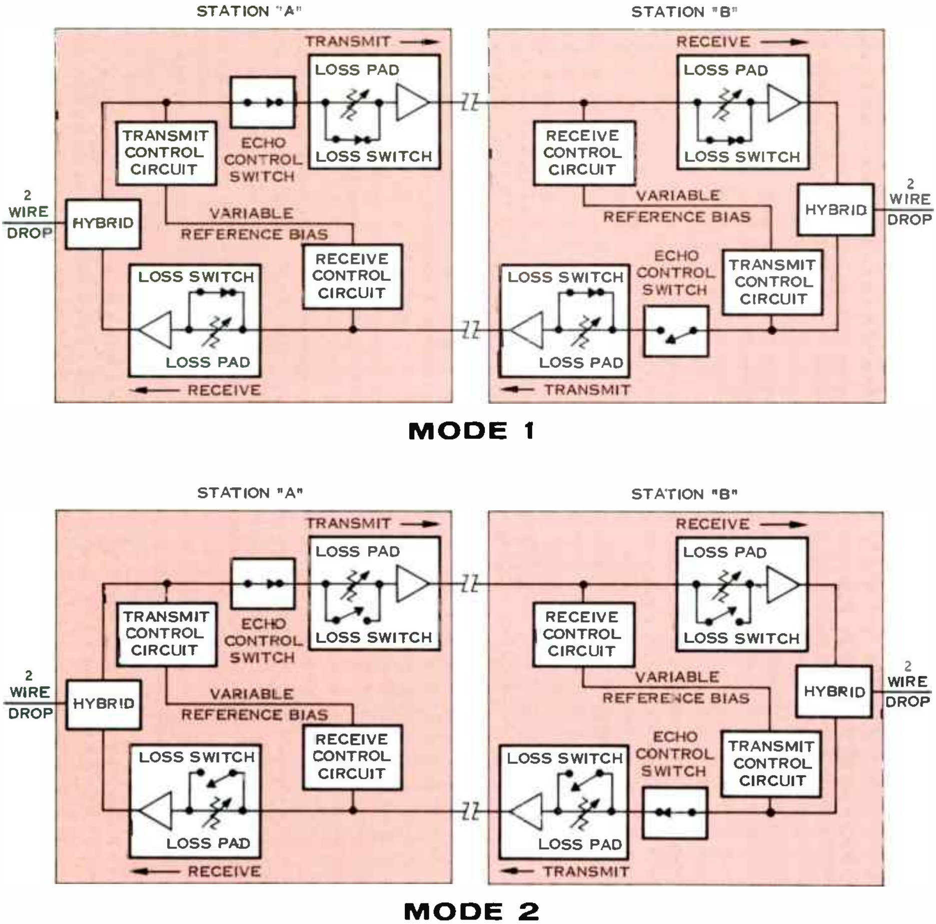


Figure 4. Simplified diagram of Mode 1 (uni-directional) and Mode 2 (bi-directional) operation of the Lenkurt 931B Echo Suppressor.

steps up to 6 dB, depending on the length of the circuit. If set for 6 dB, the talker's signal is reduced a total of 12 dB, while echo is suppressed 24 dB — this in addition to loss in the hybrid and other circuit losses.

The hangover time in the operation of the echo control switch is a compromise between echo and speech chopping. In Mode 1, the echo control switch at Station B is open as long as A is talking. When A stops talking the switch must be closed to allow B to talk. The time period before this happens is known as receive hangover and is set at 40 ms in the 931B.

In Mode 2, A is talking and B interrupts. The echo control switch at Station B closes to allow B's conversation to be transmitted. If B stops talking but A continues, the switch must again open to block any echo signal to A. If this time, known as transmit hangover, is too short chopping will result; if it is too long, A will hear a few moments of "trailing" echo. Transmit hangover is set at 85 ms in the Lenkurt echo suppressor as an adequate compromise.

In addition, it should be noted that since all operations consist of a *change* in attenuation rather than the actual opening and closing of the circuit, the recycling time between modes is not critical.

The Bell 3A echo suppressor operates similarly, using rectified speech energy through a differential detector to activate mechanical relays. During Mode 2 operation echo attenuation is introduced by a speech compressor, inserting loss in the receive path proportional to the level of the incoming signal.

Data Disabler

With the increasing use of voice circuits for data transmission, it is necessary to provide some means of removing the suppressor from the circuit. This is accomplished with a disabling circuit activated by the transmission of a continuous tone between 2000 and 2250 Hz for approximately 400 ms. Both paths are then held open for data until there is no signal for at least 100 ms. The suppressor then reverts to normal action. Note that echo delay presents no problem in one-way transmission, such as data or television.

Long delay and echo have been the subject of a number of recent tests performed by General Telephone and Electronics Laboratories, Bell Telephone Laboratories, Stanford Research Institute in Palo Alto, California, and others. Interestingly, talkers participating in the experiments became "sensitized" to the problems associated with exceptionally long delays, say of 1200 ms, and thereafter tended to be less tolerant of circuits with shorter delay. However, typical delays of 600 ms found in operating communications satellites apparently have not produced sensitizing to any noticeable degree and most customers find such service fully acceptable.

The Future

Authorities expect that round-trip delays longer than 600 ms must be avoided in future worldwide telephone links by limiting satellite systems to one hop. The practice in coming years may very well be to use satellites for only a portion of an around-the-world conversation, relying on conventional land circuits for the remainder of the path.

Generally, results of experimental testing and of actual performance in operating systems demonstrate that modern echo suppressors have satisfactorily eliminated undesirable qualities found in earlier models. Equipment such as this will be a necessary standard in all long telephone systems of the future, especially when distances are amplified by commercial and military communications satellites.

The Lenkurt DEMODULATOR

FEBRUARY 1968

SUPERVISORY CONTROL

1619

Telecommunications Provides Solutions for Industry's Problems of Remote Supervisory Control

The increasing complexity of modern society is perhaps best reflected in the areas of commerce and industry. It is here that the growing demands for goods and services – both quantitative and qualitative – by an exploding population is most strikingly illustrated. A major result of this rapidly increasing demand is a trend toward the monolithic among suppliers. Huge chains of supermarkets have replaced the neighborhood green grocer and butcher shop. The corner drugstore of a generation ago has given way to corporately-owned emporiums that merchandise everything from medicine to toys.

This same trend is also reflected in our utility industries. Where formerly fuel and power were supplied to consumers by locally-owned, co-operative concerns, they are now served by distributors whose operations encompass entire geographic regions. For example, most of the northeastern United States gets its electric power from a single source; also, most consumers in the American Southwest subscribe to the same monolithic natural gas supplier.

As these utility networks have grown, their transmission lines – pipe or wire – have extended more and more into remote areas – many of them virtually inaccessible. This extension has necessitated the establishment of relaying substations or pumping facilities in some of the remotest areas of the world. The obvious problem this situation presents is one of control.

Since it is neither humanly plausible nor economically feasible to operate these facilities manually, another means must be found. To meet this need the telecommunications industry has developed electronic status monitoring and remote control equipment, commonly referred to as supervisory systems.

Maintenance vs Control

The term status monitoring may be applied to a number of devices; some are quite familiar to us, others are highly unusual and specialized. Automotive fuel and pressure gauges, even the little red lights on the instrument panel which indicate high beams, low voltage, etc., are examples of simple status monitoring devices. There is a direct relationship between these simple applications of status monitoring and the highly sophisticated methods of reporting and control employed in power transmission and petroleum pipelines.

In discussing these systems, however, a distinction must be made between simple alarm systems and complete supervisory systems. In the case of the former, the main consideration is remote indication of "off normal" situations. This draws attention to potential trouble spots so that corrective action may be initiated. On the other hand, the most sophisticated supervisory systems report such quantitative information as flow, pressure, temperature, or voltage, and perform such functions as opening and closing circuit breakers, selecting basebands

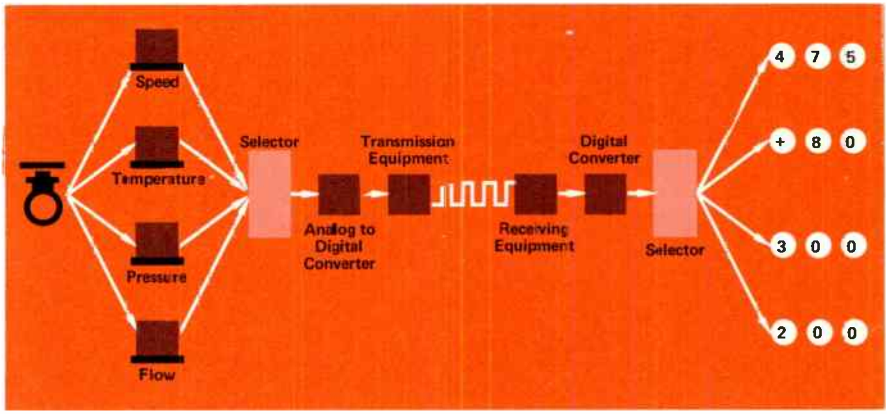


Figure 1. Simplified diagram of a typical supervisory control system.

for microwave radio, starting and stopping pumps and operating valves in the pipeline industry.

Microwave, A Typical User

A typical area in which simple maintenance-type monitoring is employed is in microwave radio transmission. Different functions of remote sites are scanned, and a simple stasis-change situation is continuously reported to a centrally located master. When change is indicated, the operater determines its nature and takes appropriate action. He may either attempt to correct the alarm situation electronically, or it may be necessary to dispatch maintenance personnel to the trouble spot.

In microwave radio, a common emergency situation might be a commercial power shortage or outage; in such cases one of the functions of monitoring systems is to switch to emergency dc power supply and begin monitoring that supply. If a predetermined critical output level is reached, the control station automatically shuts down the microwave system to prevent damage. While these maintenance-type systems do indeed come under the broad scope of status monitoring, the major concern here is with systems which perform control functions and how they do it, particularly in the electric power transmission and petroleum pipeline industries.

Typically, these systems are characterized by a centrally located master station and a number of remote stations. Ideally, the systems may employ cable, microwave radio, telegraph lines (open wire), or leased voice-plus-data lines. In most cases, the master station automatically and continuously interrogates the remotes. The remote stations in turn respond sequentially – usually indicating change or no change. Status of functions monitored by remotes is indicated at the master station on a lighted display panel or by means of audible alarms.

Different systems employ various electronic signaling techniques – some use frequency shift keying, others pulse code modulation, while still others use phase shift modulation. In all cases, stasis and change are translated as ON or OFF, or OPEN or CLOSED. Besides simple status surveillance, some systems also incorporate certain control functions, such as: activate emergency power supply, trip circuit breakers, open/close valves, or start/stop pumps. Still more complex systems involve telemetering analog data such as actual meter readings.

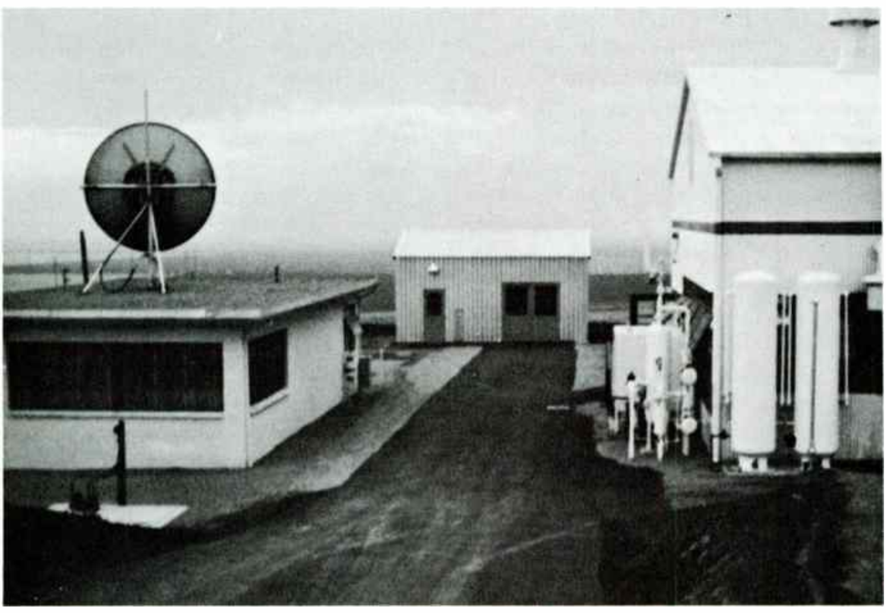
From India to the Alps

Petroleum pipeline networks crisscross some of the most remote and rugged terrain in the world – the Appalachians of western Pennsylvania for instance, or the entire northern section of the Indian subcontinent. There is a pipeline in Europe which originates at Trieste, Italy on the shores of the Adriatic, crosses the Austrian Alps and terminates at a tank farm in Ingolstadt, West Germany. In all three cases status monitoring and remote supervisory control are vital to the success of the entire operation. Volumetric output, flow rate, pressure, and temperature are monitored constantly; and valves are opened and closed, pumps activated or stopped automatically – hundreds of miles from the nearest civilization.

Although techniques and equipment vary considerably from one pipeline complex to another, the conditions monitored and the functions performed are substantially similar; this is equally true of either natural gas pipelines or crude oil carriers. Typically, they combine transducers, analog collectors, analog to digital (A/D) converters, digital collectors, digital transmission of status information, computerized storage, and visual/aural display facilities. (See Figure 1)

Quiescent-vs-Scanning

Early status monitoring systems were of the quiescent type – that is, they remained inert until an alarm situation occurred. And most had no control functions. Obviously, one



COURTESY OF UNION OIL CO.

Figure 2. Control room at a pipeline pumping station.

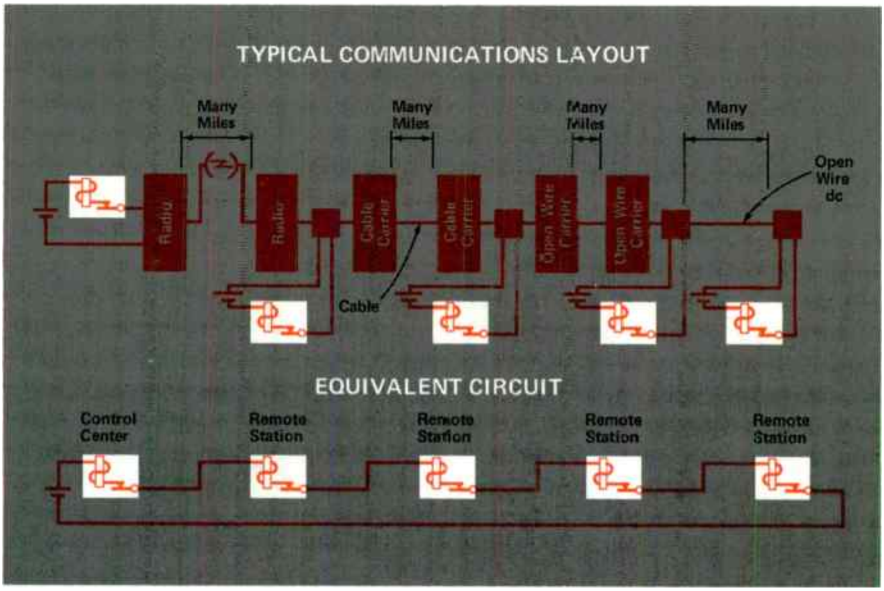


Figure 3. Typical communications layout for a supervisory system as might be used in pipeline networks.

problem, among others, was maintenance. It was virtually impossible to tell if such a system were operable unless it was reporting an alarm or malfunction.

Such systems have gradually been replaced by continuous reporting or sequential interrogation-type systems. These not only obviate the maintenance problems inherent in quiescent systems, but also provide more versatility and flexibility – particularly in the area of control functions. Furthermore, continuously scanning and reporting configurations more readily lend themselves to interfaced operation with digital transmission systems and computerized collection and storage of transmitted data.

As pointed out earlier, these supervisory control networks consist of one or more master stations and as many as 100 or more on-line stations – usually called slave stations.

In such scanning systems control of the scan is the responsibility of the control station or stations. The master station is able to alter the scan as required; that is, it may interrogate the outstations either continuously, sequentially, or randomly. Also, various functions of the different outstations – flow rate, metering, pressure, etc. – may be scanned.

The majority of high-speed supervisory systems use pulse position modulation (PPM). PPM is a type of pulse-time modulation in which instantaneous samples of a modulating wave are caused to modulate the position of a pulse in time. However, some also employ other techniques, such as pulse duration modulation, time phase modulation, or pulse code modulation. Using the binary digit system, PPM indicates 1 when a pulse is present and 0 when absent. The prime advantage of PPM is that all pulses are of the same

Figure 4. Electroluminescent map of Montreal-Hauterive power transmission network.

duration, hence data is transmitted in the minimum time.

A Model Supervisory System

A model supervisory control system as used by a pipeline, for example, should prove helpful in understanding the various processes involved and functions performed. Characteristically, status monitoring systems employ transducers, encoders, selectors, a communications loop, digital converters, and display facilities. More recently, there have been innovations in the application of computers.

Ideally, the master control station will be centrally located along the pipeline route; in some cases however, as in the Transalpine Pipeline in Europe, there will be two control stations, one at each terminus. Pumping and booster stations are spaced along the line according to terrain and path considerations. Slave stations of the supervisory system are collocated with these remote substations.

When the control station initiates interrogation, response begins with a sensory device (transducer) which meters and reports the requested variable – psi, degrees C, volumetric flow. This data is then fed through an analog selector to an A/D converter; the response is then transmitted through the communications loop to a digital converter and fed through a selector to the display facility. Display facilities are read-out dials, lighted panels, or, in more sophisticated systems, flat-fold print out.

Encoding for Telemetering

Although many types of encoding devices are used, a typical one is the digital voltmeter (DVM). This type is particularly effective since most transducers transform data into a current or voltage analog which is readily converted to digital form. The most frequently used operational techniques in DVM's for conversion purposes are digital servo, voltage-to-frequency, ramp comparison, and successive approximation. In connection with the model system, only the first of these will be discussed here. Digital servo techniques are also employed in supervisory control functions.

A servo element is graduated so as to follow a changing input voltage (or current). A feedback signal is then compared with the input, and when a null is reached the tracking element stops. Hence, the position of the tracking element can be read directly in digital form. These tracking elements are frequently in the form of stepping switches.

An earlier form of encoding device was the cyclic disc (see cover), with photo cell or carbon brush pick-off.

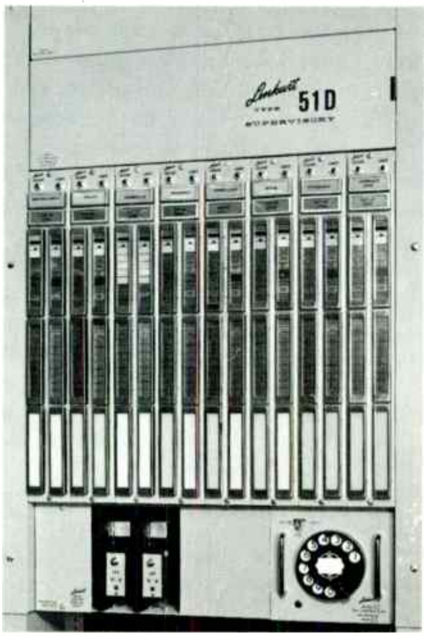


Figure 5. Alarm display panel of Lenkurt 51D supervisory system.

Using this digital disc, a remote station could scan each transducer and receive a digital reading for each variable reported. Although lately the cyclic disc has fallen into disuse, the trend toward computerization in master stations may bring it back into more widespread use. This is particularly likely since the use of discs obviates the requirement for A/D conversion found in present systems.

A Model Communications System

Once the information has undergone conversion from analog to digital form it must then be communicated to the control center from the slave stations – and vice versa. This function may be performed by various systems employing any ordinary 3-kHz voice channel. Ideally suited to such a system is the Lenkurt 25A Data Transmission System which has a 200 b/s capability for data transmission and is also adaptable for relaying of supervisory control commands. In cases where a higher bit rate is necessary or desirable – for example, computer-to-computer transmission – the Lenkurt 26C Duobinary data set is ideal.

Most communications systems used in status monitoring employ MARK (1) – SPACE (0) techniques with or without a return-to-zero function. Hence, the ordinary digital system is used wherein, say the number 4, or any decimal equivalent, is preprogrammed to elicit a desired response from the equipment under control. At the same time, but in the other direction, specific decimal equivalents indicate specific conditions, such as flow rate, temperature, or pressure. All such systems provide extremely high security from errors – most are virtually error free.

Readout

Once the reported data undergoes reconversion to digital form, it is

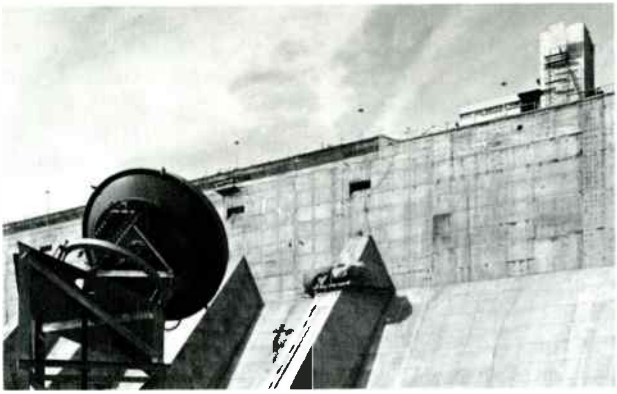


Figure 6. Microwave antenna and reflector as used by a hydro-electric dam.

picked up by a selector and relayed to the display facilities. The input is in digital form and any "off normal" data automatically triggers the alarm system. When this occurs, the operator at the control center initiates corrective action.

Also, all incoming data is automatically logged. The purpose of the alarm logger is to maintain a permanent record of the operation of the system. Intervening time and the sequence in which alarms are reported enable operators to evaluate operating techniques and reporting procedures. A typical log is shown in Table 2; others also indicate such information as station status on an hourly request basis and print out alarm status in red rather than black as in normal situations.

Function and Control

Pump engines, electronic valve controllers, booster pumps, etc., are all preprogrammed so as to respond to electronic commands from the control center. For example, to operate a bypass valve, the operator at master control in our model system must select the remote using rotating switches, pushbuttons, or thumb-wheel switches. He then selects the individual valve to be operated and the function to be performed. Finally, he cues the transmission system and transmits a command:

Start of Message (SOM) 01 2 10 (EOM) End of Message.

This might be translated as:

(SOM) Station 1 (01), Open (2) Bypass valve (10) (EOM).

The remote station then responds in kind. This is followed by a transmission that the requested function has been performed. The master station verifies the report and returns to status scan operation.

The prime functionary here is also a servo device. In response to a predetermined signal from the control center, the servo simply activates an electric motor which in turn performs the desired task, such as starting a booster pump or adjusting a valve. Another type, usually found in storage tanks, operates in conjunction with a float, and automatically stops the fluid input when a predetermined tank level is reached.

Conceivably then, one man at an adequately equipped control center is able to operate a pipeline network covering hundreds of miles. Furthermore, terrain, climate, and accessibility are no longer concerns — except

Controls:		
Engines	Stop	These controls are primarily for emergency occasions but the engine 'stop' controls may be used to adjust the throughput by stopping engines individually. Engines are started locally.
Inlet Valve	Open Close	
Discharge Valve	Open Close	
Bypass Valve	Open Close	
Repeater Mainline Valve	Open Close	
Intermediate Mainline Valve (Moran)	Open Close	
Security Sheet:		
Tank Level		Printout starts either, (i) On demand w/o reset (ii) At a predetermined time. (iii) On demand with reset
Flowmeter		
Total Throughput (per station)		
Flow Data Reset		
Monitoring: (Operational State)		
Booster Pumps (Running)		These functions are indicated by lamps lit on the dispatcher's graphic diagram – 'auto/manual' refers to the start of the balance tank stations where engines are controlled by the tank levels. During control 'off' state, telemetering from the site is maintained.
Main Pump (Running)		
Inlet Valve		
Discharge Valve		
Bypass Valve		
Auto/Manual		
Tank on Flow		
Repeater Station Valves		
Intermediate Mainline Valve (Moran)		
Dispatcher Control "Off"		
Duty Generator		
Alarms:		
Inlet Pressure (low)		The alarms listed are indicated by individual signals on the display panel. These alarms are accompanied by a flashing light and a bell. The 'multiple' alarm is a composite alarm to the dispatcher alarming functions not listed here, e.g. high water temperature, low lubricating oil pressure, etc.
Discharge Pressure (high)		
Balance Tank out of Band		
Tank Level		
Fire in Oil Area		
Equipment Room Temperature Rise		
Entry (Unauthorized)		
Duty Generator		
Emergency Generator		
Viscosity (High)		

Table 1. Some Functions of a Typical Pipeline Supervisory System.

of course to installers of the original equipment.

Power Companies, Too

In actual function, supervisory systems used by power transmission companies differ little from those employed in petroleum transport systems. The similarities between flow and current, pressure and voltage, etc., are easily seen. In general, however, reporting and scanning speed are considerably more important considerations in power transmission.

Sudden power outages – such as the one which darkened the American Northeast in 1965 – are illustrative of the need for viable supervisory control systems to power and light companies. Such a system, the Lenkurt 51D, is now in use in the Montreal-Hauterive complex of one of Canada's large power companies. This 750 Kv system utilizes both microwave radio and protective relaying systems as back-up for the power transmission lines. The microwave system features 13 repeaters spaced 15 to 50 miles apart.

The network also uses frequency diversity transmission and reception and provides more than 600 channels. Obviating the need for intermediate mechanical relays, the system employs the Lenkurt 937A Protective Relaying System to operate circuit breakers. This system has the advantage of reducing response time from 20 ms to 8 ms. Also, special circuits ensure that anything greater than a 5 dB increase

DAY	HOUR	MIN	LOG CODE	STN. NO.	PUMPS 1–5 1 – ON 0 – OFF	BOOSTERS 1 & 2	TANK MIXER 1 & 2	SUCTION PRESSURE	DIS. PRESSURE	BARREL COUNT	STN. NO.	PUMPS 1 & 2	SUCTION PRESSURE	DIS. PRESSURE	DIFF. PRESSURE	STN. PRESSURE	PUMPS 1–2	HEATER	SUCTION PRESSURE	DISCHARGE PRESSURE	DIFF. PRESSURE	TEMPERATURE IN	TEMPERATURE OUT	STN. NO.
215	10	00	7	1	11000	10	11	619	862	501632	2	10	660	873	214	3	11	10	603	902	299	39	48	4
215	10	06	8													3	11	10	601	898	293	39	50	
215	12	00	7	1	11000	10	11	621	860	502951	2	10	658	875	217	3	11	10	598	896	289	39	51	4
215	13	19	4																					

215 10 00 7 — This log produced by 2 hourly automatic request.

215 10 06 8 — This log produced because of change of pressure.

215 12 00 7 — This log produced by 2 hourly automatic request.

215 13 19 4 — This log produced by alarms.

2 — STN. LOCKOUT
01 — UNIT 1 LOCKOUT
08 — LOW LUB OIL
15 — HIGH PUMP 1 TEMP.
16

ALARM AND STATUS SCAN

Table 2.

in noise or carrier channel frequency will not cause false operation of the equipment.

The supervisory control system itself is capable of reporting up to 28 alarm situations from each of the slave stations along the line to either Montreal or Hauterive or both. Fully automatic operation is the main feature of the system. Fault conditions are automatically reported to control stations without the need of an operator; this is also true when conditions return to normal. Faults at mainline sites are displayed at both Montreal and Hauterive on a large electroluminescent map. (Figure 4)

Each site is scanned and reported every half second. An electronic clock and scanner at repeater sites sees an open circuit on each alarm lead; an indication of a fault opens the lead to the scanner which then shifts the frequency of the tone channel. This in turn transmits the fault data to the reporting center where clocks and decoders automatically synchronize to the remote stations and translate the alarm condition onto the illuminated display panel. Each station has 25 lamps to indicate different functions associated with that station. (Figure 5)

Interfaced with the supervisory system, a data logger notes all alarm conditions. These in turn are printed out on flat-fold paper by an electric typewriter capable of speeds up to 350 lines per minute. Information includes station name, fault, time and date.

A salient feature of the supervisory system is solid-state electronic equipment with module-type units for quick and easy replacement in the field. Further, all equipment is battery operated with a 24 hour reserve in conjunction with self-starting diesel generators which automatically come into play when commercial power fails. Continuous interrogation of slave stations with dual reporting in two directions simultaneously to two different control centers is another important feature. Buildings, towers, etc., are engineered so as to provide maximum immunity to the elements. The purpose of the system is to provide or maintain service during power outages.

Load Shedding

In dire emergency situations, a last ditch measure taken by supervisory control systems is to gradually shed the power load on the network. Of the many variables monitored, generator frequency is one of the most critical – a sudden drop in generator frequency is indicative of power overload. Preprogrammed servo devices automatically shed 25% of the load when frequency drops to 59.5 Hz. Should the frequency continue to decline, 50% of the load is shed at 59 Hz; the system completely shuts down if the generator frequency reaches 58.5 Hz.

However, since power companies are in the business of supplying – not reducing – power, these are to be considered as measures taken only when all else fails. In less drastic situations simple load transference or rerouting is usually sufficient.

Looking Ahead

As with other technologies, status monitoring is in a constant state of evolution – systems become increasingly complex and, paradoxically, more efficient. At the present time, probably the most striking innovation has been the transition from electromechanical circuitry to solid state devices. This change has resulted in reduced maintenance and increased reliability. Miniaturization has allowed more efficient use of shelf and rack space – hence, reduced operating costs resulting in greater profits to users. Further, the transition from analog to digital transmission of data has increased efficiency by eliminating one of the translation phases necessary in older systems.

In the near future, complex supervisory systems will undoubtedly be employed in such increasingly diverse areas as Rapid Transit Systems where instantaneous switching is a prime requirement; or small telephone companies characterized by unattended Community Dial Offices. Down lines may be immediately reported and corrective action initiated instantly through supervisory control facilities. Moreover, status monitoring is finding ever greater application in communication satellite programs.

Within the systems themselves we shall see a growing trend to solid state configurations and more widespread interfacing of existing equipment with high speed computers.

In the end it must be concluded that the potential growth of status monitoring and its applications is as limitless as that of technology itself.

JULY, 1966
VOL. 15, NO. 7

the Lenkurt Demodulator

SIGNALING
over telephone trunks

Signaling provides the means for operating and supervising a telephone communications system; it establishes connections, announces incoming calls, reports the fact that a line is busy. The functions of signaling are indeed most vital to the basic operation of the telephone plant.

Trunk signaling involves many considerations that are quite different from the basic techniques employed in signaling over a subscriber loop. This article reviews some of these considerations and describes the major techniques used to transmit signaling information over physical and carrier-derived trunk circuits.

Without signaling, a telephone system cannot operate. Even the simplest system, such as two local battery telephones connected by field wire, requires some means for the users to attract one another's attention when they want to talk. In early telephone systems, users simply cranked a hand magneto which caused a bell to ring at the subscriber station or a flag to drop at a switchboard. Over the years, signaling systems have had to keep pace with the advances made in telephone switching and transmission systems. The increasing complexity of the worldwide telephone plant has had a tremendous influence on the evolution of signaling techniques, from the simple hand cranked magneto to the many techniques employed today.

Many different signaling methods have evolved during the transition from one type of switching office or transmission system to another. Today's telephone plant includes various types of local exchange and toll switching offices, such as manual, step-by-step, panel, crossbar, and the modern electronic switching offices. In addition, there are many types of open-wire, cable and microwave radio transmission systems interconnecting the various switching offices.

Signaling Functions

There are a multitude of signaling functions that must be transmitted between the various manual and dial switching offices. These include functions whereby people must communicate with machines, machines must communicate with other machines, and machines must communicate with people.

The major functions can be somewhat arbitrarily classified as *ringing, supervisory,* and *address* (or dialing). Ringing signals are used to operate a visible or audible alarm to alert someone of an incoming call. Supervisory signals are used to convey information regarding switchhook conditions (on-hook or off-hook) at either end of a telephone circuit. Address signals convey dialing or digital information which is necessary to establish the desired connection.

In subscriber loops, supervisory and address signals are accomplished by means of direct current, while alternating current is used for ringing. Direct current signaling is also used on short-haul trunks between switching offices. However, such methods are not adequate for signaling on longer trunks, such as inter-toll, or on trunks derived from carrier or multiplex systems. As a result, various alternating current signaling systems have been developed for use over long-haul v-f and carrier-derived trunks.

Ringdown Trunks

In certain trunks, especially those interconnecting manual offices, it is necessary to transmit a ringing current to signal the switchboard operators. This type of signaling is known as *ringdown.* The ringing alternating current used in subscriber loops is at a frequency of 20 Hz. This same frequency is also used in certain short-haul trunks. On trunks equipped with composite telegraph, 20-Hz ringing cannot be used because of interference. In these circuits, a signaling frequency of 135 Hz is used. Neither of these frequencies, however, is suitable for long trunks because voice-frequency repeaters cannot pass them. Consequently, a 1000-Hz signaling tone, well within the v-f amplifier passband, has been adopted for use on longer circuits. To prevent voice signals from falsely operating the signaling equipment, the 1000-Hz tone is interrupted (modulated) at a 20-Hz rate.

Address Signals

Probably the most important and the most complicated signaling function is address or dialing. This function directs the operation of the switching equipment in the automatic offices. Consequently, the evolution of the various switching systems has brought about changes in address signaling techniques. Address signals originate at the telephone dial and consist of a train of dc pulses corresponding to the number dialed. Modern "touch calling" systems, which use keys or pushbuttons instead of a dial, employ tones at different frequencies rather than dc pulses.

In the step-by-step systems, the switching equipment responds directly to the dc pulses. However, in panel and crossbar systems, the switches cannot be controlled directly by the dial pulses. Consequently, these systems require a device known as a *sender* which stores the dial pulses and then controls the movement of the switches.

There are four basic methods commonly used to transmit address or dialing signals for use by the various switching offices. These are known as *dial pulsing, revertive pulsing, panel call indicator (PCI) pulsing,* and *multifrequency (MF) pulsing.*

Dial pulsing is the earliest and most commonly used method of transmitting address information — the numerical value of each digit is represented by the number of pulses in a train (ten pulses

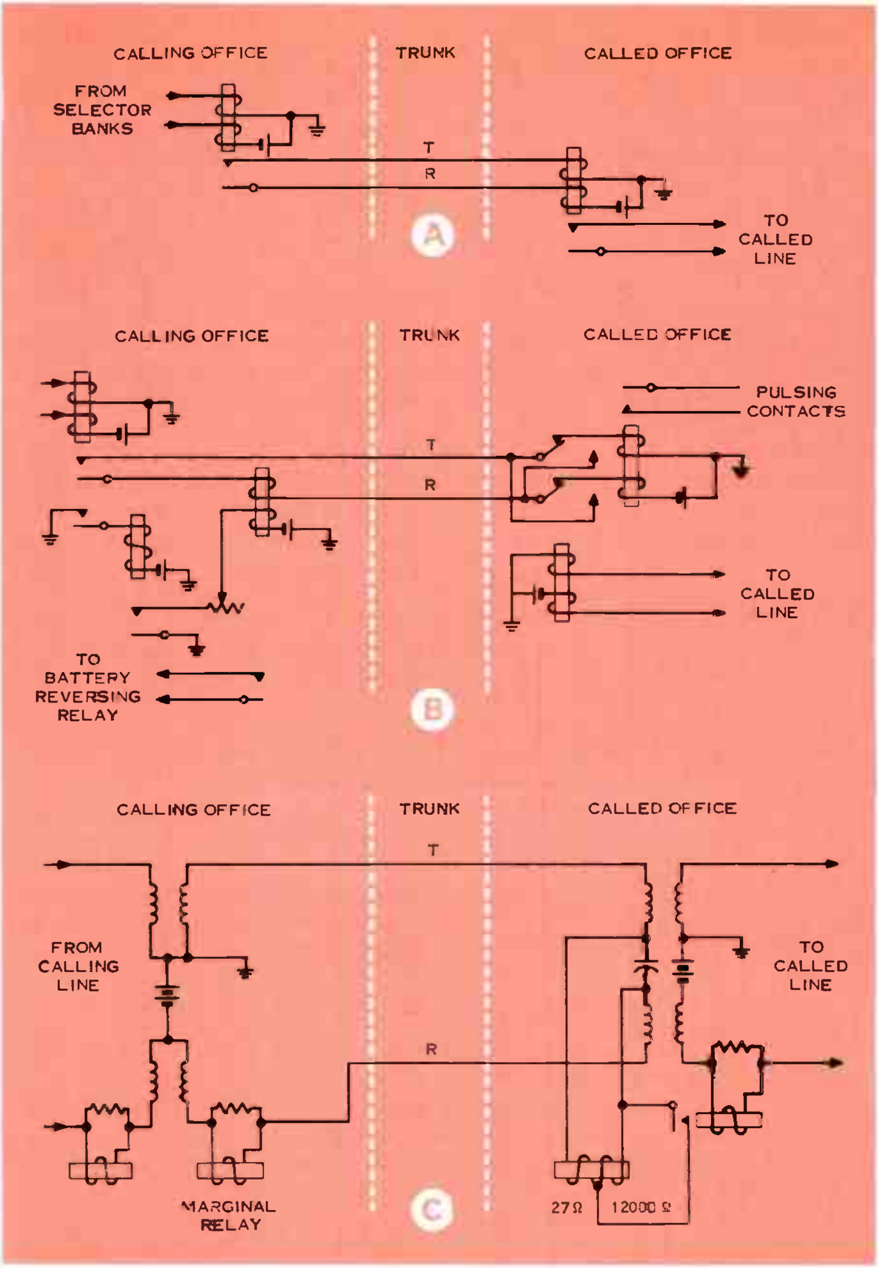


Figure 1. Loop signaling is accomplished by altering the flow of current in the trunk conductors. Three methods used to accomplish loop signaling are: (A) wet-dry, *(B)* reverse battery, *and (C)* high-low.

represents 0). Dial pulsing is used in all types of switching offices.

Revertive pulsing was originally developed for use in panel switching offices. In this type of pulsing, the address pulses are not transmitted by the originating office. When a call is made, a loop to the distant office is closed. This starts the movement of a panel selector switch at the distant office. As the selecting wipers pass each terminal, a commutator transmits pulses back to the sender at the originating office. When the proper number of these *revertive* pulses, corresponding to the called number, are received by the sender, a signal is sent back to the distant end to stop the movement of the selector. Revertive pulsing is used in certain crossbar offices as well as panel offices.

Panel call indicator (PCI) pulsing is a method of transmitting address signals between a dial office and a manual office. This technique converts pulses received from a dial office to lamp indications which appear on a switchboard. The switchboard operator then connects the incoming call to the called number and rings the subscriber.

Multifrequency (MF) pulsing is the newest method of transmitting address pulses between switching offices. Digital information is transmitted in the form of short tone bursts. Six signaling frequencies are used, each digit being represented by a combination of two of the six frequencies. The signaling frequencies fall within the speech band and are simply processed through the trunk in the same manner as speech signals. (A different form of multifrequency pulsing has recently been introduced to subscriber loop circuits through the use of telephones with pushbuttons instead of the conventional dial.)

Historically, signaling systems designed to transmit supervisory and address information have evolved from simple dc systems operating over 2-wire short-haul interoffice trunks, to complicated ac systems operating over multichannel carrier and microwave transmission systems. Today, there are essentially two fundamental techniques used to derive signaling paths on trunk circuits. The first of these is known as *loop signaling*. This technique requires a dc loop, and is the method used in all subscriber loops and in most short-haul 2-wire trunks. The second signaling technique, known as *E & M*, is used with both ac and dc signaling systems on 2-wire or 4-wire physical trunk circuits, and on carrier-derived trunk circuits. This type of signaling is standard for use in all intertoll trunks.

Loop Signaling

Loop signaling is the simplest of the two, and is used in certain exchange trunks, short-haul toll-connecting trunks, and one-way dialing toll trunks, where 2-wire voice-frequency circuits are employed. The dc signaling current flows over the same conductors used for voice transmission.

This type of signaling is accomplished by simply interrupting the condition of a dc voltage applied to the line to transmit both supervisory signals and dialing information. The range of loop signaling is usually limited to about 25 miles because of the dc resistance of the conductors.

There are three methods currently used to apply loop signaling to a 2-wire voice-frequency trunk: *wet-dry*, *reverse battery*, and *high-low*. (See Figure 1.)

In the wet-dry method, signaling information is indicated by the presence (wet) or absence (dry) of a battery and ground condition on the line at the called end of the trunk. Normally in the wet condition, the battery is placed on the ring conductor and ground on the tip conductor.

As its name implies, reverse-battery loop signaling is accomplished by reversing the polarity of the battery on the line to indicate supervisory conditions. For one condition, battery is on the ring conductor and ground on the tip conductor. The opposite supervisory condition is indicated by reversing the polarity of the battery, thus causing a polar relay to operate or release at the distant end of the trunk. This is the most prominent type of loop signaling used between exchange offices. To increase the operating range of reverse-battery loop signaling, batteries are sometimes placed at both ends of the circuit, in series. This variation of reverse-battery operation is called *battery and ground* signaling.

The third method, high-low, is used principally for supervisory signaling within a central office or from an automatic to a manual office. This type of signaling employs a marginal relay. During on-hook condition, a high resistance is placed in the loop. For off-hook, the resistance in the loop is reduced to a low value allowing more current to flow, and thereby causing the marginal relay to operate.

E & M Signaling

As mentioned previously, loop signaling is limited to trunks of about 25 miles in length. Also, such systems do not provide simultaneous signaling in both directions. In order to overcome these limitations, and especially to extend the dialing range of telephones, another type of signaling was developed.

This method of signaling employs two leads to connect the signaling equipment to the trunk circuit. These two leads are designated E and M, respectively. The name for the two leads was probably acquired from designations appearing in early drawings for this type of signaling circuit. The M lead transmits battery or ground signals to the distant end of the circuits, while incoming signals are received on the E lead as either a ground or open condition. Thus, the M lead reflects conditions at the near end of the circuit while the E lead reflects conditions at the far end.

There are several methods of deriving an E and M circuit to permit signaling between offices on a dc basis. These arrangements are known as simplex (SX), composite (CX), and duplex (DX). A simplex signaling circuit is obtained by means of a center-tap coil placed at both ends of the voice-frequency trunk circuit, as shown in Figure 2A. Signaling currents flow in both directions through the coils and, therefore, do not induce any interfering voltages into the voice channel. Conversely, voice currents do not flow through the simplex conductors (or legs) extending from the center tap of the coils. Since the two trunk conductors provide a parallel path for the signaling current, the dc resistance is approximately one-fourth of that presented to a loop-signaling arrangement over the same trunk. Thus, the dc signaling range is extended considerably. However, simplexing has certain disadvantages and has been largely superseded by the duplex arrangement.

In the composite method, a filter is used at each end of the trunk to separate the signaling current from the speech signals. The filter is called a *composite set*. Two composite signaling paths can be obtained from the two conductors of a v-f trunk and four can be obtained from a phantom circuit arrangement. This type of signaling, shown in Figure 2B, is used typically

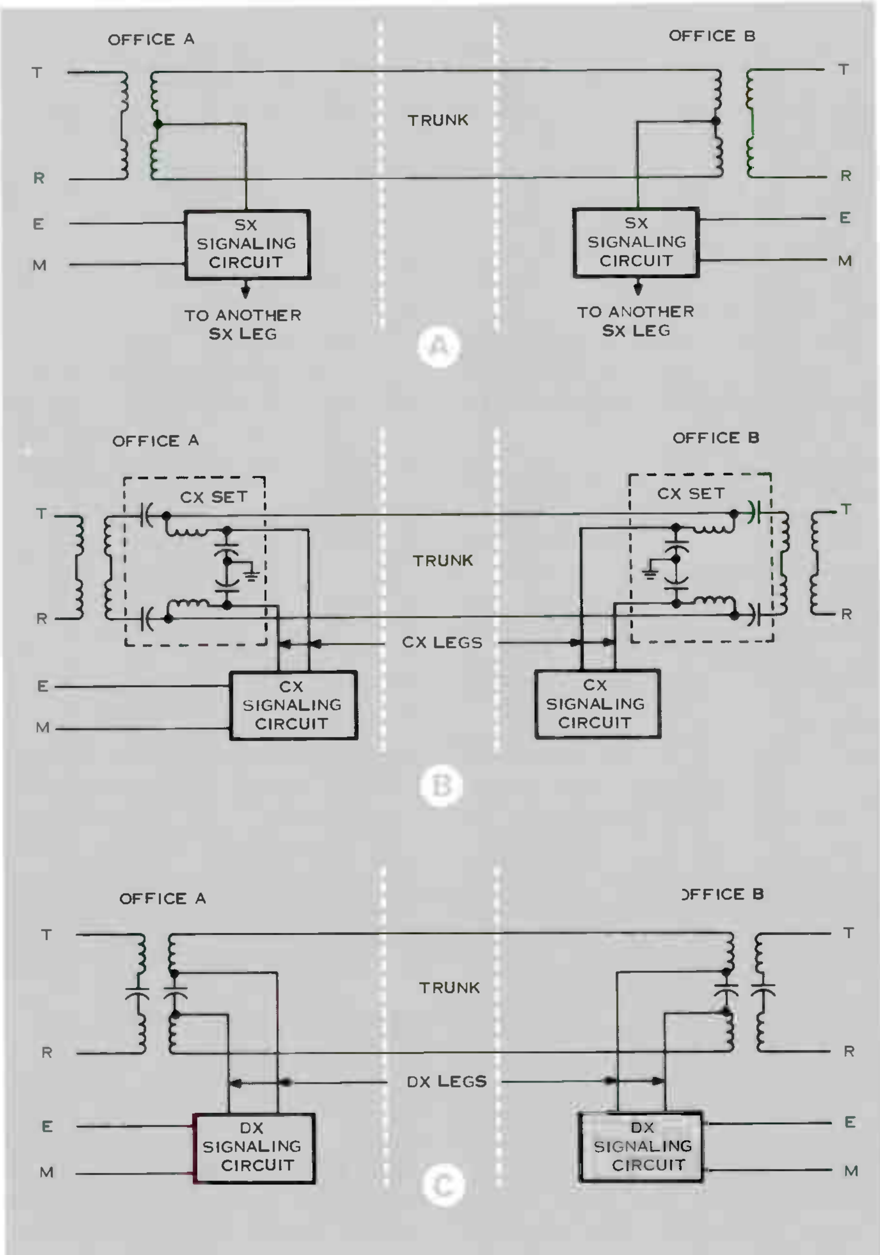


Figure 2. There are a number of different circuit arrangements designed for E & M signaling over telephone trunks. The three most prominent d-c arrangements are: (A) simplex, *(B)* composite, *and (C)* duplex.

on trunks derived from quadded cable where the conductors are arranged in phantom groups.

The duplex signaling arrangement, like the composite method, uses one conductor of the v-f circuit for signaling on a ground return basis, and the other conductor for ground potential compensation. Ground potential compensation is required because of the inherent instability of ground-return circuits. The composite set or filter, however, is not used with the duplex circuit. Instead, the signaling circuit is connected to the trunk pairs by means of a center-tap transformer and a capacitor, as shown in Figure 2C. This signaling arrangement is used primarily in paired cable trunks.

AC Signaling

The dc signaling systems described thus far are limited to relatively short v-f trunks containing a dc path. These systems are not suitable for use on long v-f trunks employing repeaters, or for carrier or multiplex trunk circuits because a dc path is not available. As a result, ac signaling systems had to be developed for use over the more modern exchange trunks and on the longer toll-connecting and intertoll trunks, especially where carrier is used.

The ac signaling systems use frequencies within the voice-frequency range so that the signals can be transmitted directly over the same path used for voice transmission. These ac systems usually employ E and M leads to connect the signaling circuit to the trunk. If the signaling frequency falls within the band used for speech transmission (typically 300 to 3400 Hz) the system is referred to as an *in-band* system. If the signaling frequency falls outside the speech band, the system is called an *out-of-band* system.

The ac systems must process the dc supervisory and address signals received from the switching office and convert them into ac signals for transmission over the trunk circuit. At the other end of the trunk, the ac signals must be converted back to dc signals before being applied to the switching equipment. Only one signaling frequency is required on 4-wire trunks. However, on 2-wire trunks two frequencies are required, one for each direction of transmission.

Early ac signaling systems used a frequency of 1600 hertz. On 2-wire trunks, 1600 hertz was used for one direction and 2000 hertz for the opposite direction. Later in-band systems used a frequency of 2600 hertz, with a second frequency of 2400 hertz for use with 2-wire trunks. The ac signaling frequencies easily pass through the same path used for voice transmission, and are amplified in repeaters in the same manner as speech signals. These so-called single-frequency (SF) signaling systems are used to transmit both supervisory signals and address or dial pulses. Multifrequency (MF) address pulsing, described previously, uses tones that are already in the voice band, so they do not require additional processing before being transmitted over a long-haul or carrier-derived trunk.

Signaling Over Carrier Channels

All trunk circuits equipped with carrier or multiplex equipment require some type of ac system for signaling. There are many different carrier signaling systems in use today employing either an in-band or out-of-band signaling frequency.

The most prevalent type of carrier signaling is accomplished with in-band frequencies. In-band signaling systems

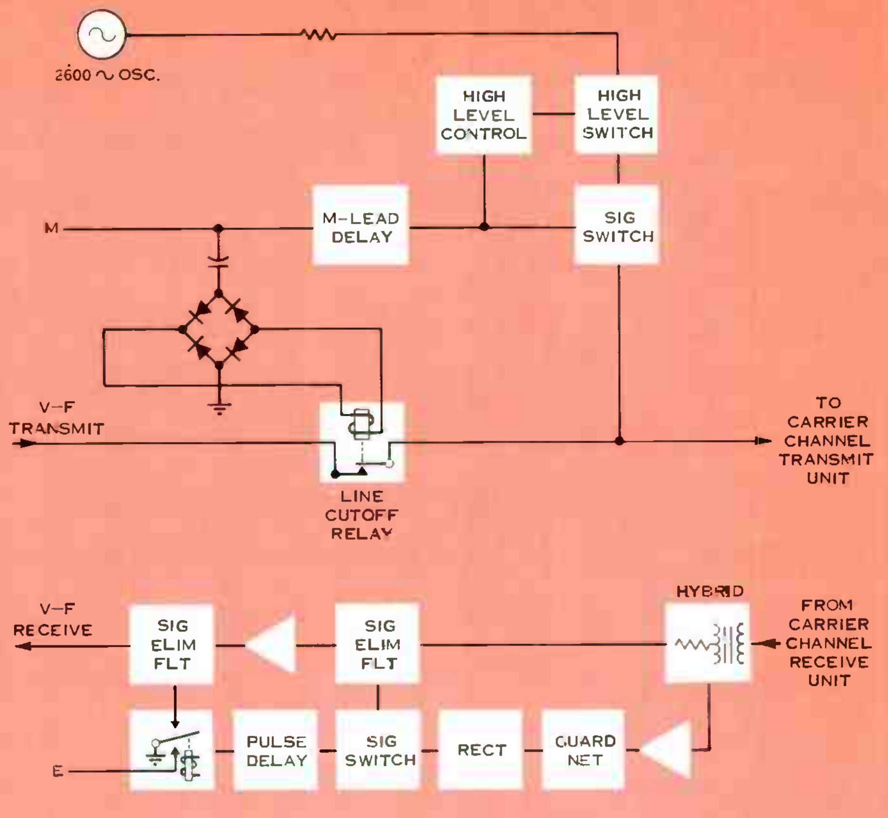


Figure 3. Simplified block diagram of a 2600-hz in-band signaling circuit applied to transmit channel and receive channel of a carrier system.

have an advantage over out-of-band systems in that they do not require extra bandwidth — the signals are passed directly through the voice channel. Another advantage is that signaling equipment is required only at the terminal stations of a trunk made up of several tandem links. Also, the in-band signaling system can be made a part of the office switching equipment rather than the particular carrier system, thus making it easier to patch trunk circuits to different carrier transmission systems.

The main disadvantage to in-band systems is that the signaling tones lie within the speech band. This leads to the possibility of speech energy at the signaling frequency "talking down" the signaling; that is, falsely operating the signaling equipment with speech energy. Protection against "talkdown" can be accomplished by using a time delay or guard circuit in the signaling system. By introducing a delay, the signaling circuit can be made insensitive to most voice energy or transient noise at the signaling frequency.

Additional protection is obtained by properly selecting the in-band signaling frequencies. Generally, it is desirable to use the highest possible frequency that will pass through the voice channel. Speech energy declines rapidly at the higher frequencies, thereby reducing the chances of "talkdown".

Most of the older voice-frequency telephone circuits use filters with an upper frequency cutoff of about 2800 hertz. For this reason, the most commonly used frequency for SF in-band supervisory and address signaling is 2600 hertz. In-band carrier signaling systems can be adapted for use with either loop signaling or E & M signaling arrangements.

Following the development of economical short-haul carrier systems, the need arose for inexpensive methods of signaling. This need resulted in the development of various out-of-band signaling systems. Out-of-band signaling equipment is generally less expensive than in-band equipment and also permits signaling during speech transmission, thus permitting extra functions such as regulation to be performed. Since the signaling frequency is outside of the speech band, there is no need for

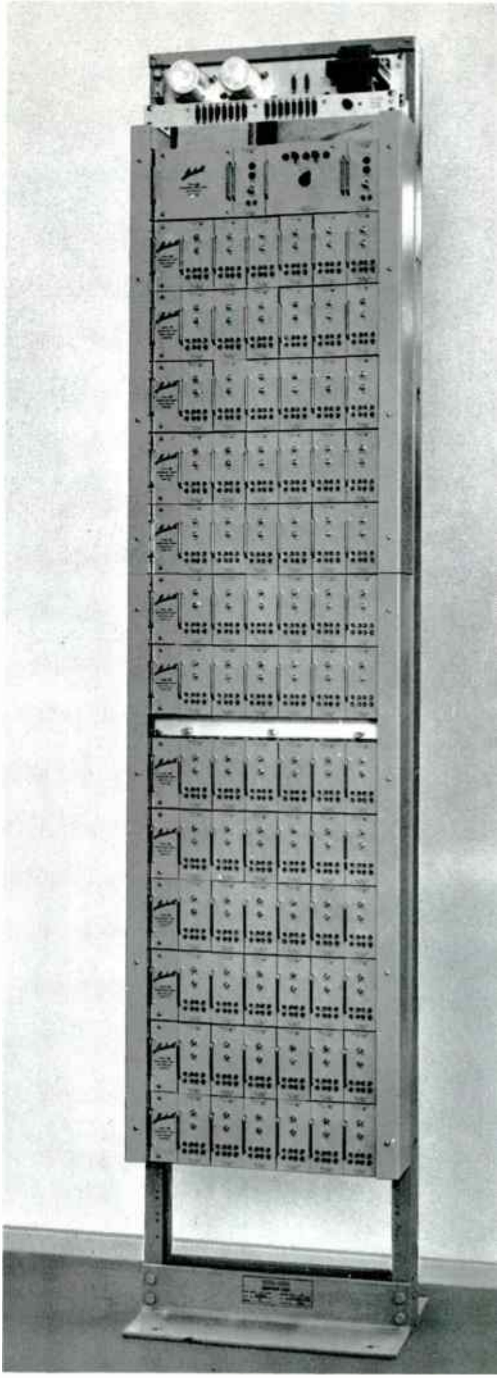


Figure 4. Photograph of typical in-band signaling units used with carrier systems. One signaling unit is required for each carrier channel. Many types of signaling units are required to accommodate the many different methods of signaling.

complicated guard circuits to prevent talkdown.

With out-of-band signaling, voice channel filters are designed with an upper cutoff frequency well below the top edge of the channel. This leaves the top portion of the channel passband available for transmitting out-of-band signaling tones. The most prevalent frequencies used for out-of-band signaling are 3700 hertz, which is standard throughout the Bell System, and 3825 hertz, which is recommended by the International Telegraph and Telephone Consultative Committee (CCITT) for use in international circuits.

Unfortunately, out-of-band signaling has certain disadvantages which tend to limit its use. Out-of-band signaling equipment has to be built-in to the carrier channel equipment and cannot be separated as in the case of in-band signaling equipment. This condition prevents randomly patching the circuit to other trunks.

Also, out-of-band signaling requires some sort of dc repeater at the end of each link of a multi-link trunk. As the signal passes from one link to another, the signal pulses must be detected and then made to operate a relay. The relay, in turn, keys the signaling equipment in the succeeding link. Thus, signaling equipment is required at both ends of each link in the trunk.

Another economical type of signaling, using time division multiplexing techniques, is used in Lenkurt's 81A exchange trunk carrier system. This unique method provides signaling for all 24 voice channels of the system using one common signaling channel. Each voice channel is assigned a specific time slot for signaling, and all 24 slots are scanned 500 times per second. The resulting signaling frequency modulates a pilot in the carrier system that is also used for slope regulation.

Although out-of-band and time division multiplex signaling techniques may be more economical for certain short-haul trunks, they lack the flexibility and other advantages offered by in-band signaling systems, especially when applied to long-haul trunks. As a result, single-frequency (SF) in-band signaling for supervisory functions and multifrequency (MF) pulsing for address functions have become the standard methods of signaling in modern interoffice, toll-connecting, and intertoll trunks.

JULY, 1967
VOL. 16, NO. 7

The Lenkurt DEMODULATOR

Remote Power from FUEL CELLS

Exotic power sources are one of the promising spin-offs from space age technology becoming available for commercial use. The fuel cell—popularized in the Gemini space flights—may become an economical source of power for remote communications sites.

Classically, a fuel is considered the storehouse of the sun's energy. In the "fossil fuels"— coal, petroleum, natural gas — this energy is usually liberated through burning. But conversions involving any intermediate process with high temperatures is not efficient. For example, the conversion of fuel to electricity in a steam turbine involves a number of steps. Fuel is first burned to produce heat. Pressurized steam then does work by turning large turbines. The turbines power an electrical generator. At each step of the conversion, energy is sacrificed.

Electrical energy is the most convenient form of energy to handle, and can be easily converted to mechanical power or heat. For this reason, investigators have been intensely interested in finding ways to produce electricity directly and with high efficiency. The fuel cell, which uses direct conversions for high efficiency, is one answer. In all direct conversion schemes, energy is extracted or transformed from one state to another without mechanical motion. With each step eliminated, greater efficiencies are gained.

The communicator is obviously interested in efficient and practical power sources to run remote relay stations and repeaters. These new devices are also being considered for large scale generation of electric power; to provide motive force for cars, trucks, boats, and even submarines; for use on space ships and satellites; for military communications systems; and almost anyplace where reliable, efficient, and quiet power generation is an advantage.

Energy Sources

Electrical power may be derived by direct conversion from a number of energy sources: thermal, nuclear, radiative, and chemical. Thermal, or heat energy was first used to produce electricity directly in a device discovered in the early 1800's. Known as a thermocouple, the thermoelectric generator consists of two dissimilar metals, such as copper and iron, joined together. When one end of the junction is heated and the other kept cool, a current is caused to flow through the thermocouple. If a number of thermocouples are stacked together forming what is called a thermopile, a usable amount of electricity can be obtained. Remotely located electronic equipment has been powered by such devices.

The heat source of the thermoelectric generator is commonly a gas flame. But a form of nuclear energy can also be tapped. Heat in this case is produced by the decay of radioactive isotopes. Workable units have been produced, but the initial cost is high. The advantage of isotopic power is reliability

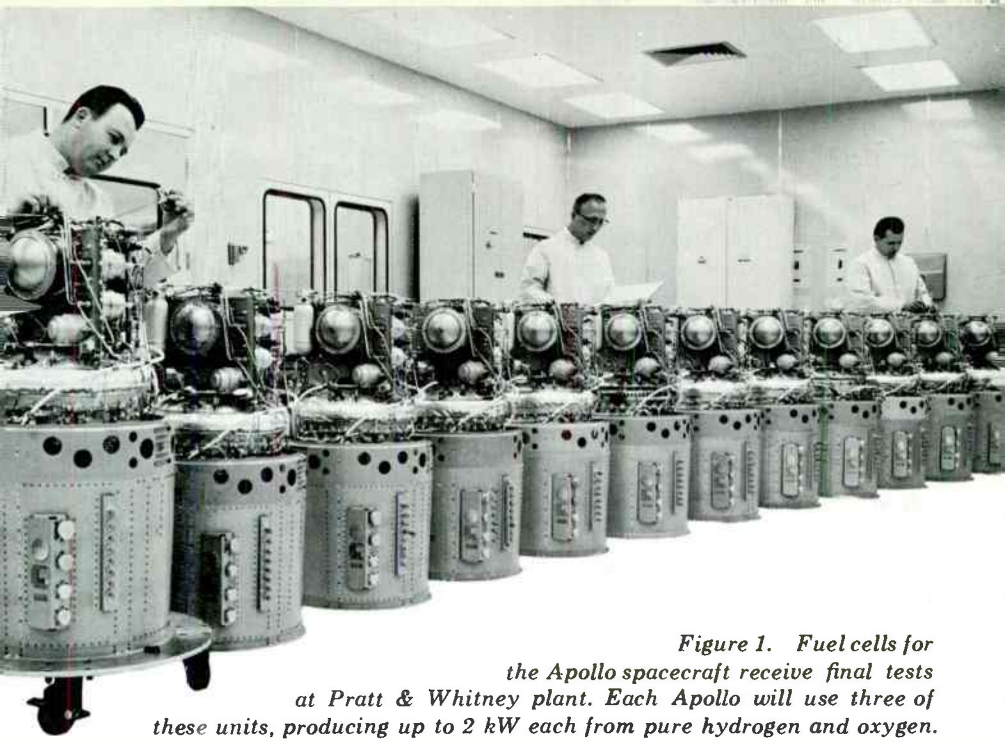


Figure 1. Fuel cells for the Apollo spacecraft receive final tests at Pratt & Whitney plant. Each Apollo will use three of these units, producing up to 2 kW each from pure hydrogen and oxygen.

and long life, even though economics makes the method prohibitive except in the most demanding cases.

Another thermal conversion device is the thermionic generator. Much like the vacuum tube, heat is used to "boil off" electrons from a cathode. Collected on an anode (plate), the electron flow becomes a usable force. Different from the electron tube, the generator derives its energy from a direct heat source — even the concentrated energy of the sun has been used.

Man's use of the radiant energy of the sun as a power source is probably better known through the space age use of the solar cell. Used almost exclusively as the power source for unmanned satellites and deep-space probes, the solar cell has proved to be a reliable power generator.

Not clearly associated with direct conversion devices, but worthy of mention as an exotic power source, is the magnetohydrodynamic (MHD) generator. Not unlike the conventional electromechanical generator, the MHD generator relies on the motion of a conductor through a magnetic field to produce electricity. In this case, however, the conductor is a plasma or highly ionized gas. The MHD generator is receiving considerable attention in the field of large scale power generation.

Chemical energy represents that energy stored in a substance and released in the form of heat, light, mechanical energy, or electrical energy during a chemical change. It is with electrical energy that we are at the moment concerned; this is the basis of the fuel cell.

Efficiencies

Most of the power used in the world comes from breaking the chemical bonds in the fossil fuels — coal, petroleum, and natural gas. But conversion efficiencies place definite limitations on the use of the resulting energy.

The internal combustion engine can approach 25 to 30 percent efficiency. However, when the mechanical linkage necessary to power an automobile is added, efficiency falls to about 15 percent for the total system. Steam generation of electricity can approach efficiencies of 25 to 30 percent, and high temperature gas turbines are rated as high as 40 percent. But whenever heat is a part of the energy conversion cycle, there is a definite limit to the efficiencies that may be obtained. And friction in any machine takes its toll.

In the fuel cell there are no moving parts, and the small amount of heat produced is not part of the conversion cycle. Theoretical efficiencies in the fuel cell approach 100 percent. In a practical device, 75 percent is more realistic, although laboratory models have exceeded this in special instances.

The economic desirability of such an efficient system is obvious. But it must be remembered that, as with any new technique, developmental costs are still high and initial investment in the machine still overshadows the advantages of laboratory-gained efficiencies.

Just the same, for applications where other more conventional forms of power are not readily available — in space, under water, and in remote land areas — efficiency may outweigh higher costs.

The Gemini 7 spacecraft, for example, carried a fuel cell system weighing about 575 pounds. To provide the same electrical capability (about 2 kW) for the two-week flight, it would have been necessary to burden the vehicle with 2000 pounds of conventional batteries.

Apollo will carry a fuel cell system rated at 6 kW which, including fuel and auxiliary equipment, weighs 1200 pounds — a tenth the weight of batteries with an equivalent output. In such applications cost is no limitation.

The communicator on earth is likewise interested in power efficiencies. At the installation of a remote microwave site, conventional power sources may represent as much as 30 percent of the total cost. The fuel cell (or other direct conversion device) may in the future reduce this cost and at the same time add convenience.

Fuel Cell Theory

The operation of the fuel cell is actually the reverse of the chemical process called electrolysis, known since the beginning of the 19th century. Electrolysis produces chemical change by passing current through an electrolyte, a solution capable of acting as a conductor. For example, if electrodes are suspended in water (H_2O) and a current passed between them, hydrogen gas (H_2) will form at the cathode (negative terminal) and oxygen gas (O_2) will appear at the anode (positive terminal).

While experimenting with electrolysis in 1839, Sir William Grove discovered that the reverse was also true: he brought hydrogen and oxygen together under controlled conditions and produced water and electric current. While Grove is credited with the discovery of the fuel cell, it was not until after World War II that any concentrated developmental effort was made. And even then it remained for NASA's space need to create the final incentive to develop a usable unit.

The first fuel cells used hydrogen and oxygen. Many other chemical reactions which produce electricity are the subject of much current research. But for a basic understanding of the fuel cell, it is simplest to examine the hydrogen-oxygen unit.

Chemical Reactions

The cell contains two electrodes separated by an electrolyte (Fig. 2). Hydrogen (the fuel) is available at the anode and oxygen (the oxidizer in the chemical reaction) is at the cathode. As the two gases are applied to the electrodes, separate reactions take place. When hydrogen is passed over the anode, electrons are generated and can be made to do electrical work through an external circuit. At the cathode the electrons join with the oxygen to produce what is called an hydroxl ion, with the chemical symbol OH . The ions travel through the electrolyte to complete the electrical circuit.

A closer look at the chemical reactions in the fuel cell will further explain the process. At the anode, hydrogen gas (H_2) is absorbed in the form of hydrogen atoms (H). In the electrolyte are hydroxl ions (OH^-) produced at the cathode. An ion is an atom or group of atoms that has either gained or lost an electron. In this case an extra electron is available and the OH grouping takes on a negative charge. The hydrogen atom and the hydroxl ion join to produce water (H_2O) and at the same time free the extra electron. This electron is now available to flow through the external circuit.

At the cathode, oxygen gas (O_2) is similarly absorbed through the electrode. Here the oxygen atom (O) reacts with both the water (H_2O) in the electrolyte and the incoming electron to form hydroxl ions.

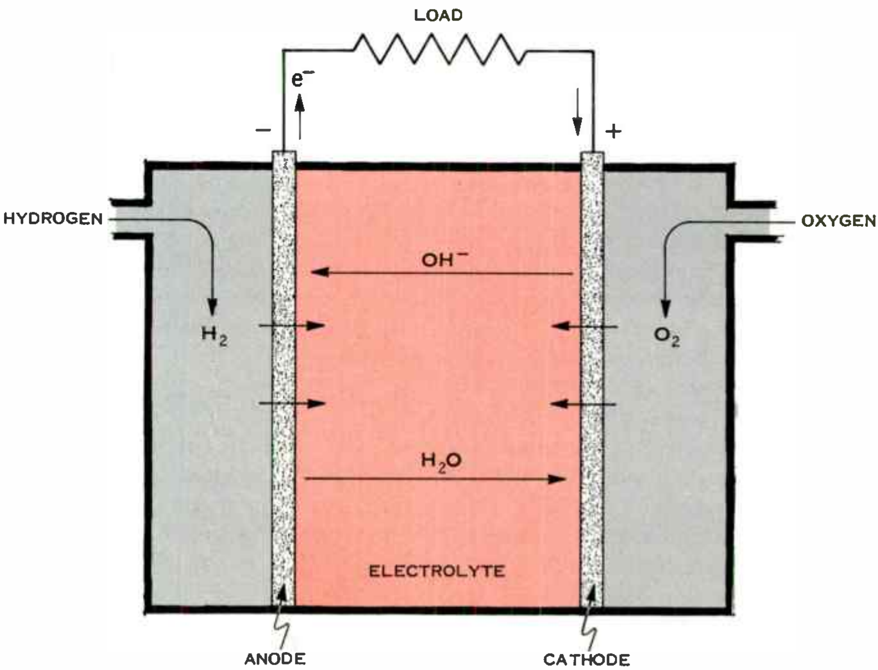


Figure 2. At the anode of the hydrogen-oxygen fuel cell, hydrogen joins with hydroxl ions to produce water, freeing electrons to do work. At the cathode, oxygen reacts with water in the electrolyte and electrons from the external circuit to form hydroxl ions.

These reactions may be summarized using chemical notation. At the cathode

$$1/2\ O_2 + H_2O + 2e^- \longrightarrow 2\ OH^-$$

That is, oxygen plus water plus electrons produce hydroxl ions.

At the anode:

$$H_2 + 2\ OH^- \longrightarrow 2\ H_2O + 2e^-$$

Hydrogen plus hydroxl ions produce water and electrons.

In summary, hydrogen and oxygen can be continually combined in such a way that water and electrical energy result.

The output of a fuel cell is low voltage, high current dc. Individual cells can be stacked in both series and parallel arrangements the same as conventional batteries to increase voltage or current.

Figure 3. Small natural gas fuel cell produces 500 watts of electricity. Smaller units can use gasoline.

Batteries vs. Fuel Cells

Conventional batteries, it should be noted, are closely related to the fuel cell. However, the battery is self contained and must either be discarded or recharged when its stored energy has been used up. The fuel cell will continue to produce electrical energy as long as fuel is supplied.

The storage battery has advantages where high power is needed over a short period of time. The fuel cell is more applicable for needs of moderate power over longer times. A combination of the two, with the fuel cell charging the storage battery between uses, could capitalize on the strong points of both. It is also characteristic of the fuel cell that when no current is being drawn, it does not expend fuel.

The hydrogen-oxygen fuel cell was the first to be produced and received most of the early research effort. In its basic form as described, it is not too difficult to follow the progress of chemical events. However, even the first successful cells were more sophisticated than our example.

For instance, the gases hydrogen and oxygen do not readily interact at room temperatures. Some cells operate at much higher temperatures (250°F to 500°F) and use chemical catalysts in the electrodes to help the reactions along.

Hydrocarbon Fuels

There are many practical reasons for development to shift now to systems that operate on fuels other than hydrogen. The hydrocarbon fuels seem ideal because of their availability. A cell using fuel oil or natural gas, for example, could be operated almost anyplace in the world without serious problems of transporting of fuel. Methane, gasoline, kerosene and alcohol are among the hydrocarbons being investigated, along with the noncarbons ammonia and hydrazine.

The engineering problems associated with fuels other than hydrogen have placed many barriers in the way of a practical system. The direct conversion of hydrocarbons is particularly difficult. But a compromise is being used — a system now under field test with the Army breaks down hydrocarbon molecules and extracts hydrogen, which can be accepted by the fuel cell. However, the indirect process is not as efficient as direct conversions.

While the complication of manufacture and transportation of pure hydrogen makes other fuels desirable, the availability of pure oxygen also presents problems. Although a ready source of oxygen, air contains some undesirable elements. Current models of hydrocarbon-air fuel cells must process the air before it is used, removing carbon dioxide and stabilizing temperature and humidity.

Hydrocarbon-air systems now being tested include a package of two 35-pound man-carried units producing 500 watts at 32 volts. Another unit with an output of 3 kW can be carried by jeep or small truck.

Applications

The possible applications of the fuel cell are about as varied as the uses of electricity — but can be extended beyond that. The use of a practical and efficient direct energy conversion device as a substitute for heat engines in the generation of electrical power immediately leads to the application of power for electronic equipment. Large scale power generators are also considered. But it is not out of the question to consider this new technique as a substitute for other motive forces. Allis-Chalmers has put fuel cells to work powering a farm tractor, and the Army is now testing a ¾-ton truck powered by a fuel cell package developed by Monsanto Research Corporation. Boats, submarines and many other vehicles will also be the subject of increased research.

Generation of electrical power at home is also being looked at. The development of a power plant operating on natural gas has been launched, with Pratt & Whitney Aircraft carrying out the program. The company also developed the fuel cells to be used on the Apollo spacecraft.

The only moving parts in the fuel cell are found in the gas flow mechanism. As a result, the fuel cell is a very quiet machine—both mechanically and electrically — an ideal advantage to the military. They are capable of resisting a good deal of shock and acceleration, and are affected very little by radiation, all points in favor of space applications. And the water byproduct is actually a plus in manned space travel where astronauts must take with them all food and supplies.

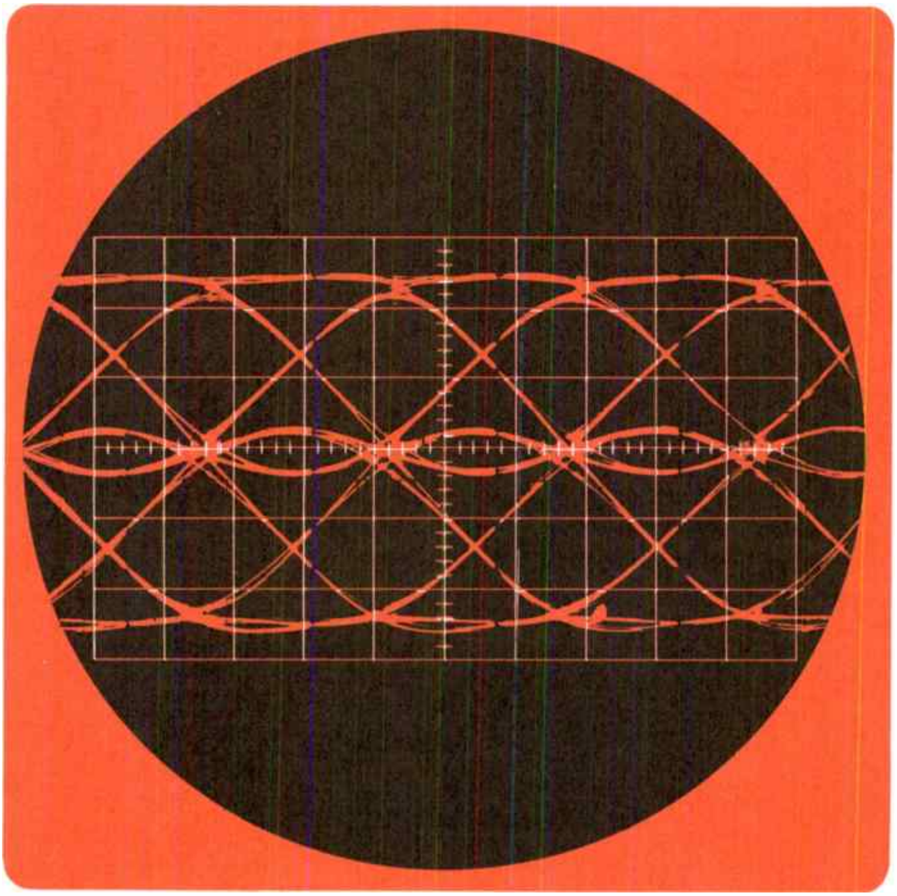
The telecommunications industry is primarily interested in fuel cells to provide power to remote locations. Brown, Boveri and Company, of Switzerland, has been operating a television relay station on fuel cell power for nearly two years.

Operators in this country have yet to put a commercial device to work in such application, but interest in the fuel cell and other direct conversion devices is high. As repeater equipment uses more solid state circuitry, thereby reducing power requirements, this type of power will become more realistic. In the meantime, engineering progress will account for increases in efficiency, reduction in size and weight, and the practicality of using common fuels.

The Lenkurt®
JANUARY 1969

DEMODULATOR

Circuit Conditioning Part 1



Data transmission speeds go higher and higher and line quality requirements increase correspondingly.

Quality, in a communications circuit, is relative. The same circuit may perform flawlessly for voice communications and not handle data traffic at all satisfactorily. Before the recent upsurge in the amount of digital traffic being handled by common carriers, this was not an especially serious problem. But nowadays, data is demanding an increasing amount of the telephone industry's attention. They are faced with the pressing requirement to improve their networks to more efficiently handle digital type traffic.

Since it is not feasible to replace all existing voice-grade circuits or to install whole new networks dedicated to data transmission, the obvious solution is to upgrade existing circuitry.

Two phenomena cause most of the difficulties encountered in attempting to send digital data over voice circuits. They are delay distortion and amplitude distortion. But both can be overcome by proper conditioning of the circuit. The Lenkurt 971B Adjustable Equalizer is designed specifically to cope with problems of distortion in voice circuits intended for data traffic. It corrects for both amplitude and envelope delay distortion, and conditions circuits to meet C–1, C–2, C–3 and C–4 standards.

Delay Distortion

Delay distortion of electrical signals is a type of distortion caused by the non-linear phase delay-versus-frequency characteristics of a communications circuit.

In voice-frequency transmission facilities, such distortion is caused mainly by the capacitive and inductive effects of transformers and amplifiers at the low frequencies. At the higher frequencies it is caused by loading coils and line capacitance. In carrier transmission facilities, channel bank, group, supergroup, and mastergroup transformers and filters are the main causes of delay distortion.

Delay distortion is not a problem until it begins to interfere with the ability of the communications receiver to understand the information contained in the signal. In the case of speech transmission, delay distortion has not been a problem, since the human ear is relatively insensitive to variations in phase-versus-frequency relationships.

However, digital signals are quite vulnerable to the effects of delay distortion. Data bits usually originate as rectangular-shaped pulses which are used to modulate a carrier at a particular keying rate for transmission over a communications circuit. The analagous AM or FM signals resulting from this modulation process are composed of many frequencies.

The envelope of these signals results from energy at the fundamental and harmonic frequencies adding together vectorially. If such a signal passes through a circuit with non-linear phase-versus-frequency characteristics, it becomes severely distorted. In fact, the signal energy may "spread out" to the point where adjacent pulses begin to interfere with one another. Under such conditions, the data receiver may not be able to properly detect the information content of each bit, (for example, a binary 1 may be detected as a binary 0, or vice versa).

When considering the cause of delay distortion, it should be noted that an appreciable impedance mismatch between line sections or between the line and office apparatus will influence the distortion characteristics of the facility. The presence of reflected currents sometimes causes an overall distortion much different from that which is indicated by analyzing the individual facility components. This is particularly noticeable at low frequencies or with short lengths of line, where the loss is relatively low, and reflection and interaction effects are consequently greater. In the case of long loaded cable circuits, manufacturing tolerances and subsequent treatment of cables can cause appreciable variation from the delay determined by formula or from typical curves. Another problem is the spacing of load coils which varies from the ideal because of the necessity for spacing coils according to the location of manholes or telephone poles.

Measurement Techniques

There are two common techniques used to measure delay distortion in communications circuits. One consists of measuring phase shift, either with a phase meter or by means of an oscilloscope Lissajous pattern – a method mainly used in the laboratory. It is a tedious procedure because phase shift-versus-frequency must first be computed and then point-by-point slope measurements made to derive the delay curve.

This method requires accurate frequency measurement and is suitable only when both terminals of the circuit are available at the same location. The method is also occasionally used for transmission circuits where two similar circuits exist so they can be "looped back" to the measuring point. Figure 1 is a curve of frequency-

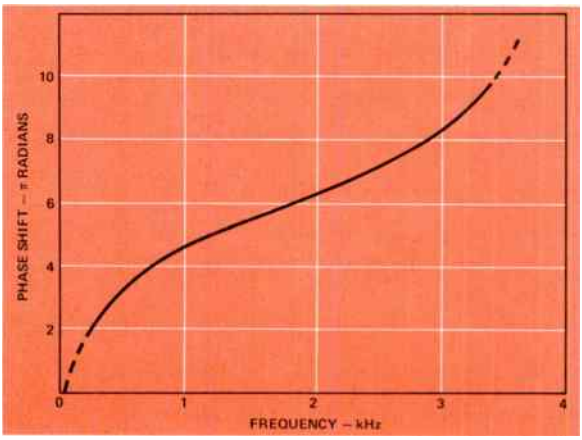


Figure 1. Frequency/Phase Shift Characteristic's of Typical Communications Channel.

versus-phase for a typical communications channel.

The second technique consists of measuring the envelope delay characteristics of a circuit rather than the frequency/phase-shift characteristics and is relatively easy to accomplish.

Envelope delay of a particular frequency is equal to the slope of the phase-frequency curve at the specific frequency. Therefore, envelope delay can be determined by measuring the phase shift of two incremental frequencies and then computing the slope. The degree of resolution of the measurement depends upon the incremental spacing of the two frequencies used to determine the slope characteristics. The closer the spacing the greater the resolution of the measurement.

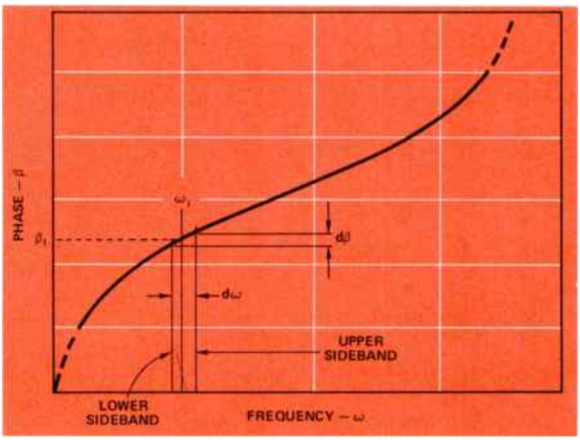
Measuring the slope at various points throughout the passband of a circuit provides an indirect indication of the phase-versus-frequency characteristics of the circuit.

However, it is not necessary to use two incremental frequencies to make such a measurement. The usual practice is to transmit an amplitude modulated carrier through the circuit under test, and measure the resulting phase shift of the modulation envelope. The result is the same as measuring the slope of two incremental frequencies, and the measurement concept employs much simpler computational techniques.

When transmitting an amplitude modulated carrier through a circuit, energy at the carrier and upper and lower sideband frequencies will be displaced in time according to the frequency/phase characteristics of the circuit. This is illustrated in Figure 2. It means that the envelope of the signal will shift in phase from the carrier by an amount equal to the mean value of the slopes of the two sidebands (assuming both sidebands are equal in amplitude). The amount of this phase shift is called *envelop delay*.

As a rule, the lower the modulating frequency, the more the envelope phase shift approaches the actual value of the slope of the phase shift at the carrier frequency. It can be assumed in practice that the envelope phase shift of the amplitude modulated carrier is equal to the slope of the frequency/

Figure 2. Envelope Delay at Amplitude Modulated Carrier Frequency ω_1.



phase curve describing the circuit at the carrier frequency. (See Figure 2). This means that if the amplitude modulated test signal is tuned or swept across the band of interest, the envelope phase shift detected at the receiving end of the circuit provides an indirect measurement of the frequency/phase shift characteristics of the circuit. A curve of envelope delay-versus-frequency for a typical communications circuit is shown in Figure 3.

Delay measuring test sets have been developed to measure envelope delay in communications circuits. An example of such a set is the Sierra Model 340B, shown in Figure 4. The set generates an amplitude modulated carrier that can be manually adjusted or electronically swept over a frequency range corresponding to voice band, standard group and supergroup circuits. The carrier is usually 50 percent amplitude modulated with one of three frequencies.

As mentioned previously, the lower the modulating frequency the more meaningful the measurement. However, variations in envelope delay are difficult to measure accurately if the frequency is too low and a compromise is usually reached. Three modulating frequencies, 25 Hz, 83-1/3 Hz, and 250 Hz, have become somewhat standard throughout the communications industry and are employed in most delay measuring test sets.

The test set also processes the modulated signal after it has traveled through the communications circuit. The phase shift encountered by the envelope with respect to the carrier is measured by a zero-crossing detector. The difference in phase is read directly on a digital display or meter as delay in microseconds. Typically, these test sets are capable of measuring delays as great as 20 milliseconds.

It is important to realize that the absolute envelope delay at a particular frequency is of no immediate concern in equalizing a circuit. It is only the *relative* delay that is important. This is the term used to express delay distortion performance requirements in communications standards. Relative delay is the difference between the envelope delay (in microseconds) measured at some frequency within

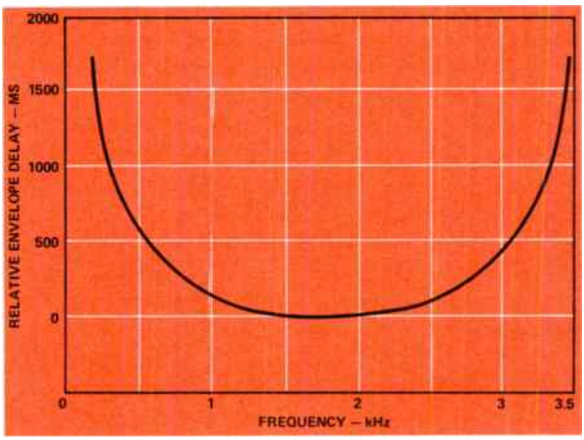


Figure 3. Typical Relative Envelope Delay – 46A Channel Equipment (Measured With Channel Units Back-To-Back).

Figure 4. Envelope Delay Test Set

(Courtesy Sierra Electronic Operation, Philco-Ford Corp.)

the band of interest and a reference delay and frequency established within the same band.

In performance standards, envelope delay distortion is usually expressed as a maximum difference in envelope delay (in microseconds) between two frequencies within the passband of the circuit. For example, the envelope delay standards for a telephone trunk circuit might read:

> "Envelope delay distortion shall not exceed:
> 80 microseconds between 1000 and 2600 Hz.
> 250 microseconds between 600 and 2600 Hz.
> 500 microseconds between 500 and 2800 Hz."

This means that the difference in envelope delay (in microseconds) between any two frequencies between 1000 and 2600 Hz cannot exceed 80 microseconds, and so on. Using a delay measuring test set, the envelope delay characteristics of a circuit can be determined very quickly to see if they meet these requirements.

Such a test can be accomplished by tuning the amplitude modulated carrier to a series of discrete frequencies across the band of interest (i.e. 1000 to 2600 Hertz), and manually recording the relative envelope delay reading at each frequency on the receiver of the test set.

Alternatively, the carrier can be electronically or manually swept across the band of interest and the envelope delay information recorded on an X–Y recorder or viewed on an oscilloscope.

Subtracting the minimum delay reading from the maximum delay reading provides a measure of the envelope delay distortion for the particular passband. If the difference between 1000 and 2600 Hz is greater than 80 microseconds, then the delay performance of the circuit will not meet the standards described in the example and the circuit must be equalized.

Delay Equalization

It is not always necessary to measure the delay characteristics of a circuit to equalize it for data. There is a technique for equalizing a circuit known as the "eye pattern" method which is useful when conditioning a circuit for a particular data modem.

With this method, a data modem such as the Lenkurt 26C which has its own test pattern generator is placed in the circuit with the modulator generating a random data pattern. An oscilloscope is connected to the data receiver (or demodulator) and synchronized with the receiver clock. The overlapping traces viewed on the oscilloscope provide an indication of the amount of phase distortion present in the signal.

Equalizers can be added to the circuit and adjusted until the distorted eye pattern improves. This technique, although it does not assure that the circuit is equalized to certain specified limits, does provide the fastest adjustment to optimum setting for the particular data modem involved.

The job of equalizing a telephone circuit for data transmission has been greatly improved through the use of delay measuring test sets and variable equalizers. Most of the delay measuring test sets provide analog output voltages that are proportional to the carrier frequency and envelope delay detected in the receiver. These voltages are used to drive an oscilloscope or an X–Y recorder to provide a visual display of the relative envelope delay-versus-frequency characterisitics of a circuit. Such visual information is extremely helpful when attempting to equalize a circuit.

There are several methods of setting up delay test sets to measure the envelope delay characteristics of a circuit such as *loop back*, *end-to-end with return reference*, and *end-to-end*, as shown in Figure 5.

The loop-back method is used primarily in the laboratory or in factory tests to measure the envelope delay characteristics of a circuit when both ends are available at the same location. The end-to-end with return reference method generally provides the most accurate measurement of a circuit because of better synchronization between the transmitter and receiver of the test set.

The first step in equalizing a circuit with a delay measuring test set is to establish the test arrangement. The end-to-end method is generally the most convenient.

In such an arrangement, an operator is required at each end of the circuit with an order wire facility so that the operators can talk to each other. At the transmit end of the circuit the terminal equipment is removed and the delay test set transmitter connected in its place. The test set receiver is then connected to the opposite end of the circuit. An X–Y recorder can be connected to the receiver to facilitate equalizing the circuit.

It is important to understand that the absolute delay within a circuit cannot be reduced. Equalizers correct

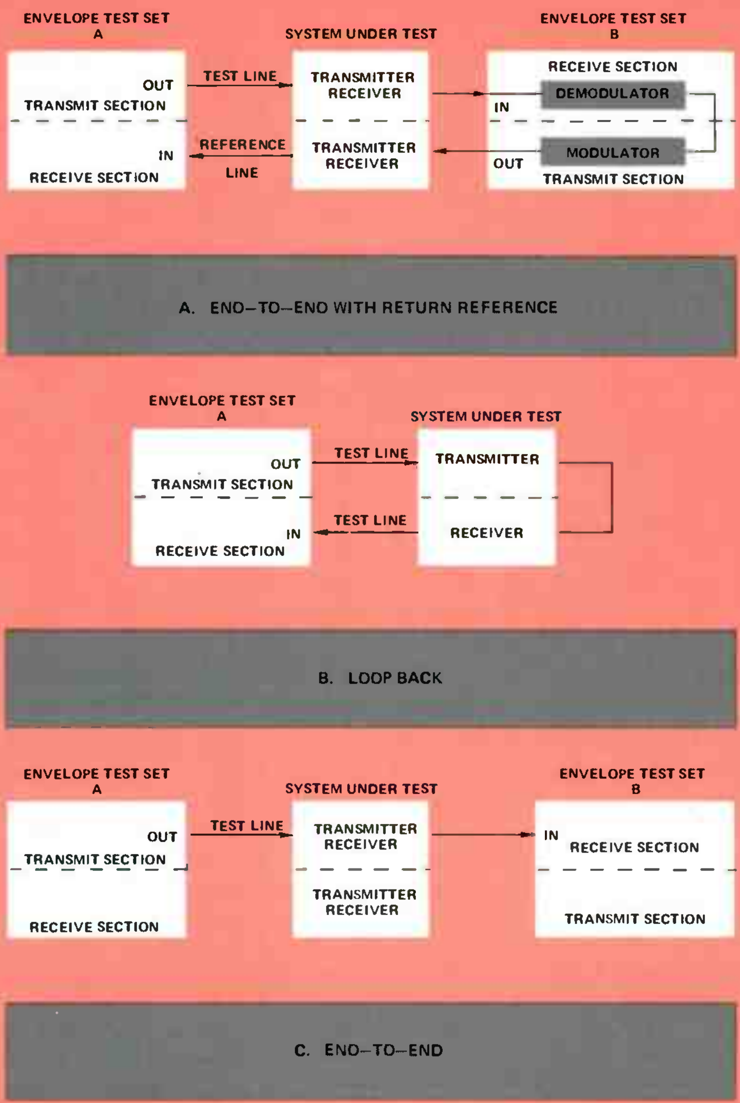


Figure 5. Envelope Delay Test Sets, Test Arrangements

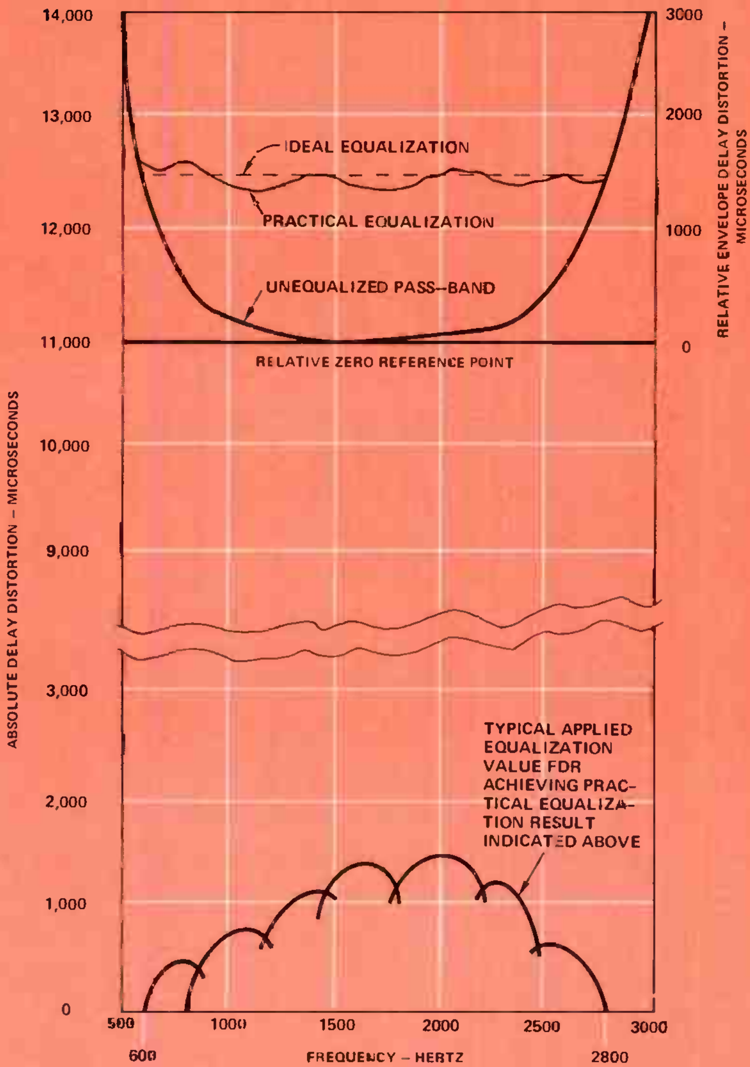


Figure 6. Relationship of Relative Envelope Delay Distortion To Absolute Delay Distortion.

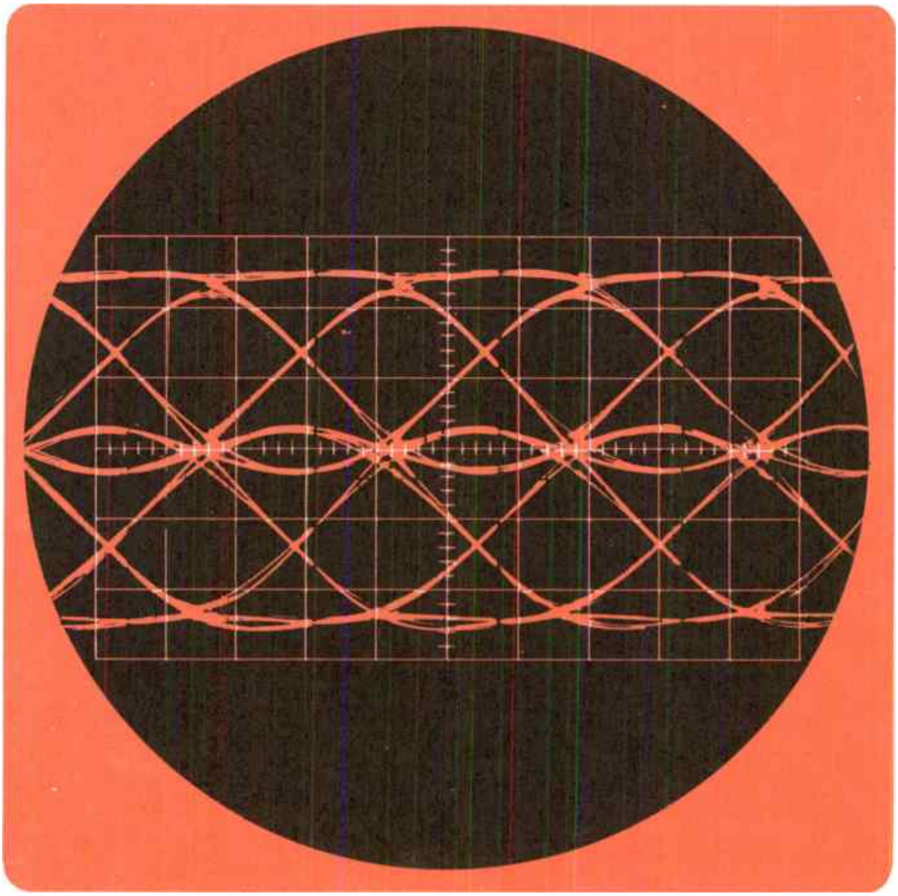
at a specific point in the circuit, for relative values of delay (and associated amplitude) distortion across the passband. This point is usually at the receiving end of the circuit. Fundamentally, relative delay distortion is corrected by adding extra delay to those areas in frequency where the signal distortion value is low. This effectively flattens the passband by setting the overall distortion levels, across the band, at the value of the worst (highest) condition. Even though system end-to-end absolute (total) delay has been increased slightly for certain frequencies, the conditioned signal will display very little effect from these forms of distortion.

Relative delay distortion is usually determined by establishing, individually, the lowest respective delay value and corresponding frequency within the passband as the reference point and comparing the values at other frequencies with this point of reference. For most applications, the reference point is arbitrarily set at zero microseconds of delay. In this way, measurements made at discrete frequencies within the passband of the circuit will all read as positive numbers. Occasionally, the point of zero reference is set at a specific frequency, such as 1500 Hz. In this event, values of distortion throughout the passband may appear both as positive and negative numbers with respect to the reference point. Figure 6 shows the relationship between relative and absolute delay distortion.

The February, 1969 DEMODULATOR, Circuit Conditioning, Part 2, will discuss amplitude distortion and the means for coping with it.

The Lenkurt®
FEBRUARY 1969
DEMODULATOR

Circuit Conditioning Part 2



Amplitude distortion creates unique problems for data transmission over VF facilities.

Even though amplitude and delay distortion are not *necessarily* related and do not *always* appear in the. same circuit, their causes are much the same. However, they are still two separate and distinct phenomena – consequently the means for correcting them will also differ.

By definition, amplitude distortion is a variation of loss or gain with frequency. It has two distinguishing characteristics – band edge roll-off and in-band ripple.

Band edge roll-off is usually caused by filters in a voice multiplexing system, or by loaded cable, or by high pass characteristics of transformers and series capacitors. In-band ripple is caused principally by impedance mismatches and their attendant reflection.

Data Distortion

For frequency division multiplexing (FDM) equipment such as the Lenkurt 25B or other low speed data transmission systems, band edge distortion in the voice channel may be so severe that it is impracticable to obtain more data channels by equalization. In-band amplitude ripple or delay distortion, however, is seldom so severe as to affect these low-speed channels. Therefore, equalizers for frequency division multiplexing equipment need only correct the amplitude response at the corners of the channels.

For higher speed data transmission – 1200 bits per second or more – delay distortion usually is controlling, and only moderate amounts of amplitude correction are required. Usually, just enough to correct a general amplitude slope through the frequency band is sufficient.

However, in a switching network such as Autovon or Autodin, where high speed data is to be transmitted over several tandem switched sections, it is necessary that both amplitude and delay distortion be tightly controlled in each link of the switching system. This is so that the overall characteristics of several sections in tandem (which are the sum of the characteristics of the individual sections) will meet the requirements for data transmission.

In cases like this, the in-band ripple of the amplitude response becomes important and must be equalized. As previously mentioned, this ripple is caused by impedance mismatch; conceivably it could even be caused by imperfect impedance matching of delay distortion correction devices.

Amplitude Equalization

Treatment for amplitude distortion usually accompanies delay equalization. As stated in Part 1 of this issue, the absolute delay within a system cannot be reduced from its total value at the point of origin. But, amplitude correction can be accomplished with gain devices.

The Lenkurt 971B Adjustable Equalizer is designed specifically for the correction of amplitude distortion as a function of frequency as well as delay distortion to enable VF circuits to meet stringent C3 and military DCA requirements for data transmission (as outlined in DCA Circular 330-185-1). See Table 1.

The 971B is usually operated from the receive end of the circuit. An exception would be in cases where intermediate switch locations occur in long-haul circuits that can be separated into shorter segments for switching. In such cases, for purposes of equalization, each segment is treated as a separate circuit.

Correction is accomplished by adding gain or loss in the areas of frequency where distortion exists to flatten the passband. Amplitude-versus-frequency distortion is reduced over a particular frequency range. This, in turn, decreases the distortion of the signals being transmitted in that range or, conversely, increases the data transmission rates.

A good adjustable equalizer system, such as the 971B, generally consists of independent delay and amplitude equalizers. Since the delay systems were described in Part 1 of this article, only the amplitude equalizer will be discussed here.

Basically, there are three methods of amplitude equalization. They involve the use of either "bump" sections, band edge booster sections or cosine sections.

The first method introduces variable compensation shapes (bumps) uniformly across the frequency range

Table 1

TYPE C3 CONDITIONING

Type C3 – For access lines and trunks associated with a Switched Circuit Automatic Network or common control switching arrangement

Access Lines
- The envelope delay distortion shall not exceed:
  - Between 1000 and 2600 Hz, a maximum difference of 110 micro-seconds
  - Between 600 and 2600 Hz, a maximum difference of 300 micro-seconds
  - Between 500 and 2800 Hz, a maximum difference of 650 micro-seconds
- The loss deviation with frequency (from 1000 cps reference) shall not exceed
  - Between 500 and 2800, –0.5 dB to +1.5 dB
  - Between 300 and 3000, –0.8 dB to +3.0 dB

    (+ means more loss)

Trunks
- The envelope delay distortion shall not exceed
  - Between 1000 and 2600 Hz, a maximum difference of 80 micro-seconds
  - Between 600 and 2600 Hz, a maximum difference of 260 micro-seconds
  - Between 500 and 2800 Hz, a maximum difference of 500 micro-seconds
- The loss deviation with frequency shall not exceed
  - Between 500 and 2800, –0.5 to +1.0 dB
  - Between 300 and 3000, –0.8 to +2.0 dB

NOTE: Conditioning in accordance with the above specifications is limited to:

Each Interexchange or local access line – between the customer's station and switching center

Each trunk – between switching centers

Extracted from Tariff F.C.C. No. 260 Page 143.2
– Effective August 7, 1967

to be equalized. It is effective for correcting moderate in-band distortion but has the problem of allowing too much interaction among the separate bumps in the line. However the bump method can deal with chosen subdivisions within the band on a one-by-one basis.

In facilities with steeply dropping band edges – such as loaded or submarine cable – equalization is aided by the use of a tuneable high and low frequency booster section with shape controls. Such devices (the Lenkurt 971B is one) are equipped for dealing with slope or band edge drop-off problems. But, unless another equalization technique is used, they cannot cope with mid-band ripples.

Cosine Equalization

Bump section and band edge booster sections are fairly conventional techniques for equalizing data circuits. There is, however, a method used by the Lenkurt Adjustable Equalizer called cosine equalization which is not so common.

Cosine sections introduce a set of harmonically related cosine wave shapes of gain versus the log of the frequency. In general they are considerably more effective than bump shapes for equalizing severe in-band amplitude distortion, especially since they require fewer sections.

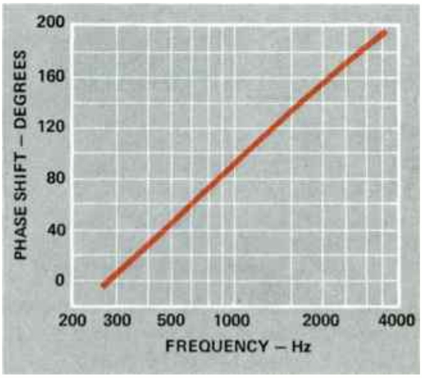


Figure 1. Cosine section phase shift-versus-frequency characteristic – Amplitude Equalizer.

In a conditioning system which employs cosine equalization techniques, the cosine equalizer consists of a number of identical phase-shifting networks connected in tandem. The networks are driven by compound transistors and unity-gain phase splitters the outputs of which are combined in a summing amplifier.

The 971B cosine equalizer provides five cosine shapes of amplitude versus logarithm of frequency. The shapes are continuously variable in amplitude and sign. Combinations of these shapes are used to cancel amplitude variations in the voice band of the circuit.

Figure 1 shows the phase shift/frequency relationship characteristic in a typical cosine stage. This relationship is almost linear. If the output voltage of the first phase shift stage is added linearly to the input voltage, the shift at 300 Hz will be zero and the summed voltage will be double that of the input voltage.

At 3000 Hz though, the phase shift is 180 degrees, and if the two voltages are added their sum will be zero. So, as the frequency is varied between 300 and 3000 Hz a half-cosine curve of the voltage-versus-log/frequency can be obtained. The curve will be on a logarithmic scale because of the logarithmic relationship between phase and frequency.

And, since the phase shift networks are in tandem, the output of the second stage (or section) will go from 0 to 360 degrees as frequency is varied between 300 and 3000 Hz. Output shifts of the next three stages will be 540, 720 and 900 degrees respectively. Consequently the five cosine stages will produce 1/2, 1, 1-1/2, 2 and 2-1/2 cosine curves. All five situations are represented in Figure 2.

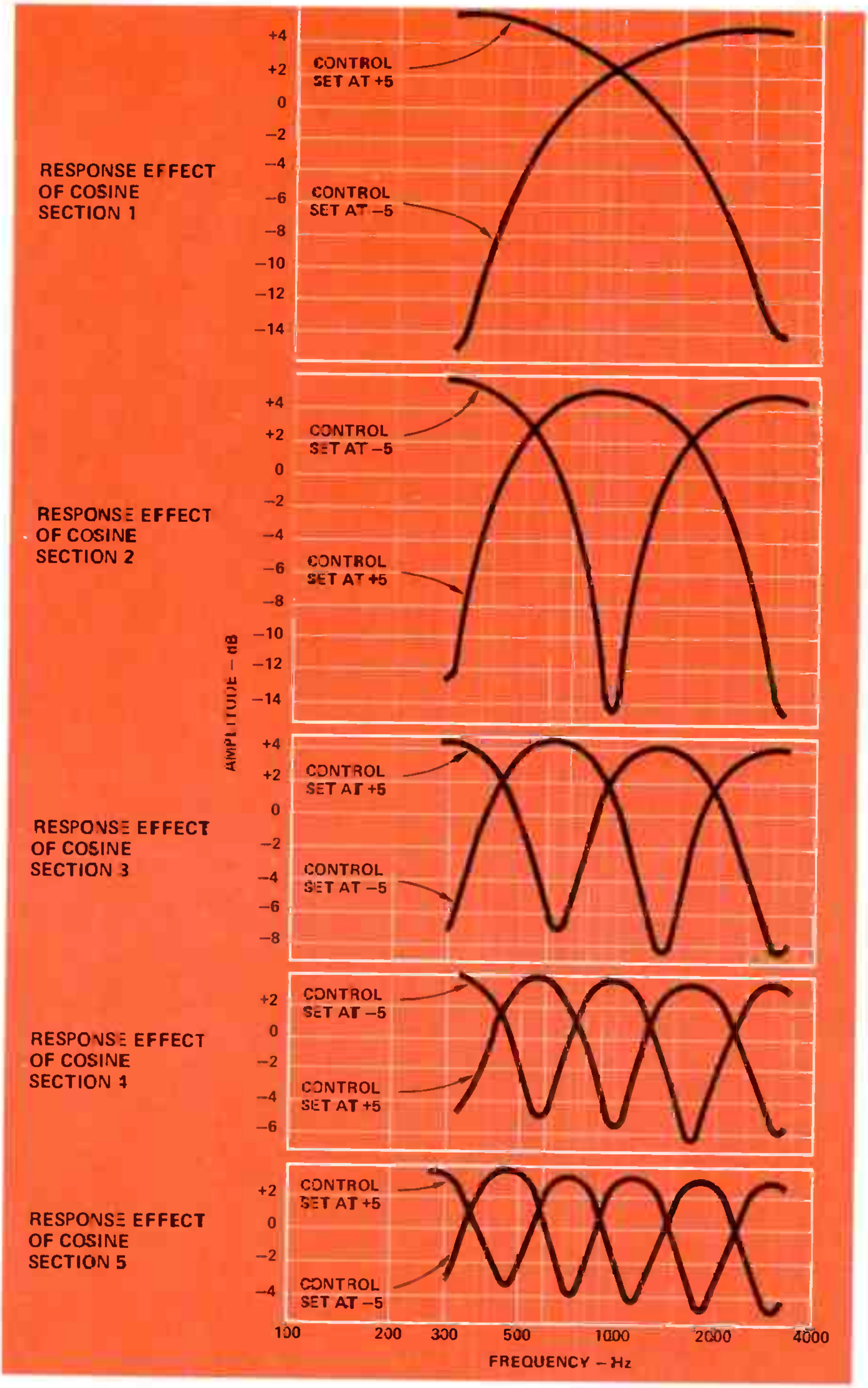


Figure 2. Response effects of Lenkurt 971B's five cosine sections

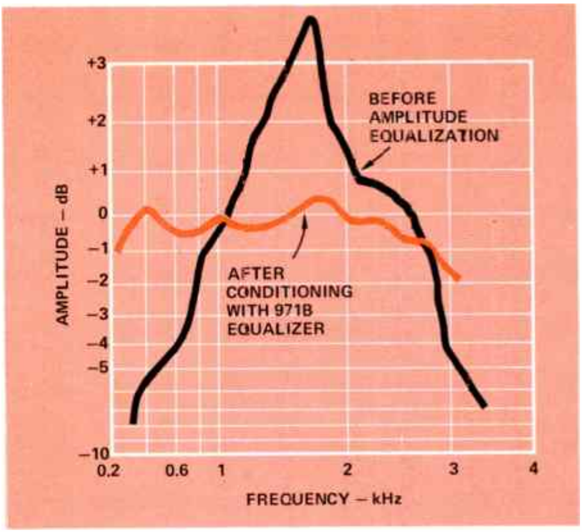


Figure 3. Actual amplitude distortion condition – before and after equalization. End-to-end measurement of VF circuit consisting of four links of 46A Carrier and one link of 3-kHz submarine cable.

Measurement Techniques

Equipment for measuring amplitude distortion is basically an oscillator and a VTVM; whereas an envelope delay test set is used for determining delay distortion. There are also some differences in technique.

Whereas the reference point for delay measurements was arbitrarily set at zero microseconds on the test set, a specific frequency – usually 1 kHz with a 0 dB reference point – is chosen for amplitude measurements. Consequently, values of distortion throughout the passband may appear as both plus and minus figures with respect to the reference level. But just as with delay measurements, a reference point is provided to which distortion in other portions of the band may be compared. From this comparison, it can be determined whether or not equalization is feasible or even required.

Absolute values of minimum distortion and loss within a system are not directly related to the problem of correcting for relative distortion and its effect on a signal. Even though they may not be of immediate concern to the equalizer system, absolute values are important for purposes of comparison, particularly in parallel circuits. In such cases, the circuit with the lowest distortion and loss figures is "built out" of the network by the equalizer to the value of the circuit having the highest figures.

One of the most expedient methods of measuring amplitude distortion involves the use of level tracers such as the Siemen's REL 3 K 211G. A level tracer provides an oscilloscope trace pattern of the pass-band amplitude condition. Hence, the effects of adjustments made on the equalizer can be readily seen and adjustment results can be recorded for reference. In this method, one level tracer's generator is used at the transmitting end of the line as a sweep signal source. The receiving portion of a second level tracer is connected to the output of the equalizer at the line's receiving terminal. The tracer is then adjusted to track the sweep of the transmitting unit. The result is a calibrated amplitude-versus-frequency curve of the pass band.

By using a level tracer in conjunction with an adjustable equalizer such as the Lenkurt 971B, off-line amplitude equalization is possible. This is done by placing a mirror image of the trace (inverse tracing) before the cathode ray tube of the level tracer and matching to it the shape of the equalizer circuit. This technique is used when, for some reason or another, end-to-end alignment is not possible.

Alignment for Amplitude Equalization

Just as with delay equalization, three basic pieces of equipment are required — an adjustable equalizer and two sweep level tracers.

The level tracers are connected to each end of the circuit and their controls are set for the desired transmit/receive levels as determined by the particular system requirements. Requirements differ widely depending upon the particular transmission facility being equalized.

To determine the system's general amplitude response characteristics, the line response should be first observed and recorded off the level tracer without the equalizer in the circuit.

The five cosine equalizer section controls are then adjusted to get the best possible (flat) response across the entire pass band. These controls are especially effective for minimizing in-band amplitude ripples. Each cosine section introduces a distinct positive or negative shape whose amplitude is directly proportional to the control knob setting.

At the start, each cosine control should be varied in both directions from its zero setting to determine the relative effect on the response for that particular section. Experimentation is usually necessary to determine the various combinations of cosine control settings which will best equalize the circuit.

From a simple operational point of view, it is not really necessary to know what each term is accomplishing except for a general understanding. This is because cosine shapes, up to the third or fourth terms, can be reckoned reasonably accurately from the gain/frequency response on the level tracer.

For higher terms, reduction of the total gain spread over the whole band usually provides sufficient adjustment standards. But when adjusting the cosine sections sequentially, if a point is reached where no appreciable reduction in gain spread is noted, this and all higher order sections should be left at zero. A new combination is then started by setting the first section at a new level.

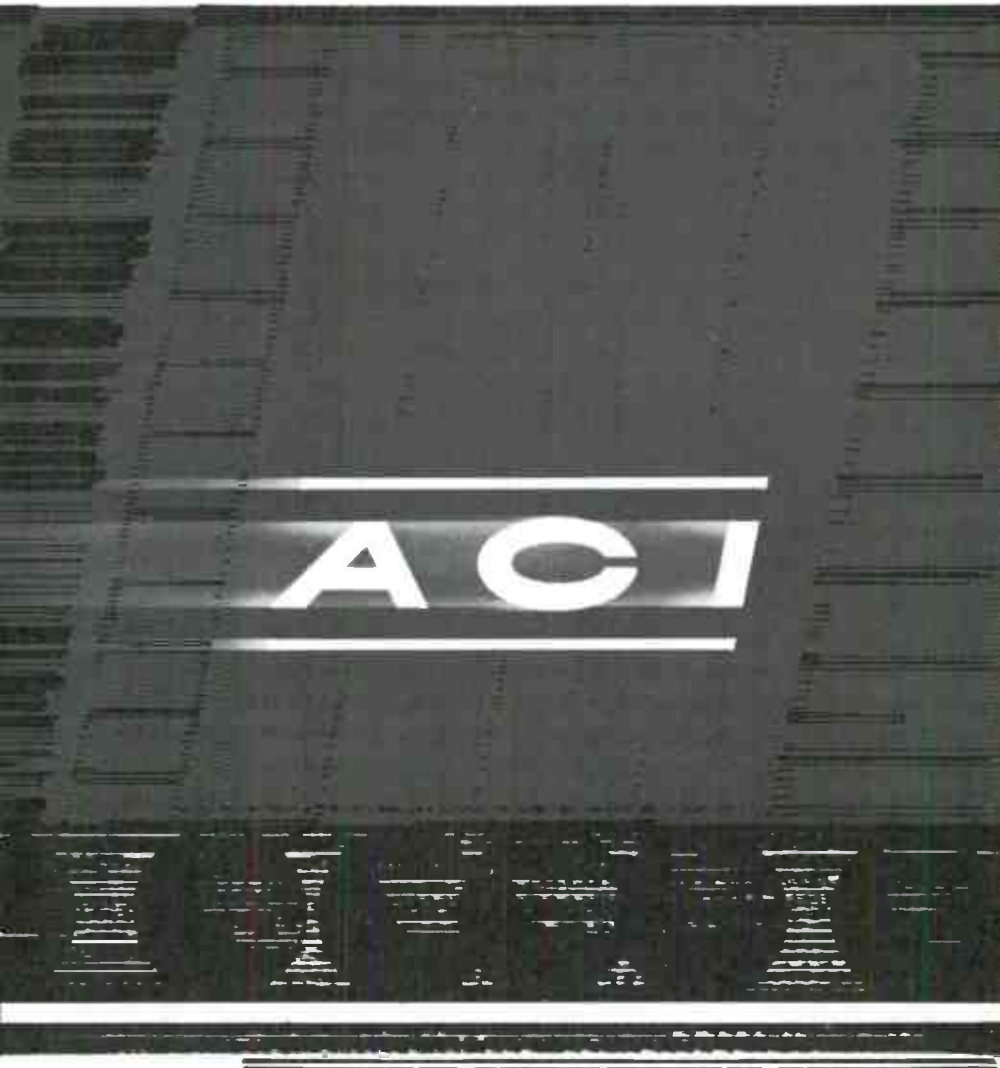
In obtaining the desired amplitude response, it may be necessary to juggle the high and low booster sections and the 0 dB, 1000 Hz reference values during alignment since cosine adjustments can change the high and low end response.

But if all goes well, and the desired amplitude response is obtained, the line should then be ready and capable of carrying digital information as well as analog signals.

The *Lenkurt*®

AUGUST 1968

DEMODULATOR



. . . . Automatic Car Identification aids the railroads in computerized traffic control

There are more than two million railroad cars in North America – boxcars, tankers, reefers, flatcars, cattle cars, gondolas and cars carrying trucks "piggyback". These millions of cars are owned by 130 different railroads and are scattered over 255,000 miles of track.

Simply accounting for all this rolling stock is a monumental task for a single line. The problem is compounded by the number of companies in any one area and further by long-standing agreements among the various carriers. Traditionally, railroads have comparatively free access to each others' cars for hauling freight. When making up a train in the marshalling yards, one railroad is likely to use cars belonging to any number of other lines. For some idea of the complexity and magnitude of the accounting problems involved it is only necessary to watch a passing freight train and count the number of different labels on the cars. For accounting and billing purposes, all these cars must be counted, identified, and have their destinations and cargoes noted.

All of this information is usually marked on the sides of the cars in some form of numeric code. Heretofore, the information was read and processed in the yards by men using pencils and tally sheets. The data was then carried to an office for computation and forwarding – by mail or telegraph – to the various companies

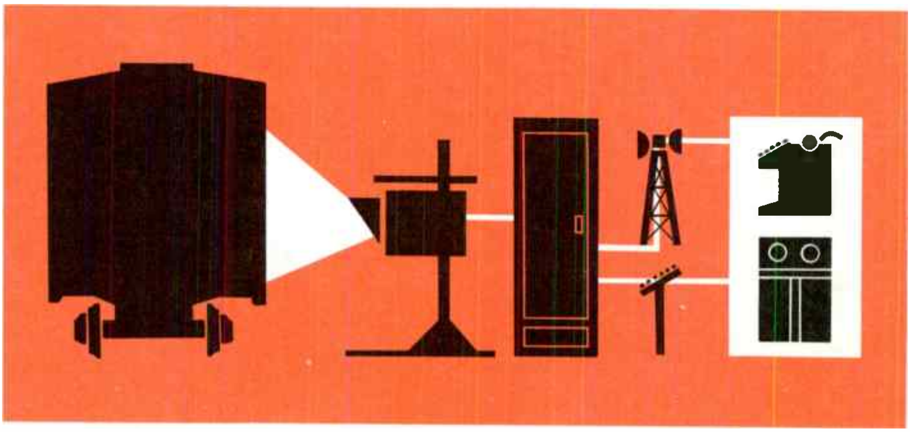


Figure 1. The basic ACI system consists of three components – the label (on the car), the scanner and the decoder. Optional equipment can be added to meet requirements for higher levels of system sophistication.

involved. This all added up to an extremely slow, inaccurate and expensive process.

Obviously, a fast, accurate and universal method of identification – probably electronic – was needed. To this end, the Association of American Railroads (AAR) turned to the electronics industry for some means of automatic car identification. Sylvania Electronics developed and manufactured an ACI system and gave it the trade name Kartrak (see Figure 1). The system has recently been adopted by the AAR as the industry standard.

Working in conjunction with computers and data communications modems, the Kartrak system is able to identify a passing railroad car and transmit the information to a computer center while the train is moving – at speeds up to 80 mph. Cars are counted, identified as to owner, type and serial number and the information forwarded – all automatically. Optional equipment for recording cargo, destination and direction of travel is also available.

The basic system consists of three components: label, scanner and decoder. The label is on the car itself and contains the pertinent data; the scanner reads the passing labels, and the decoder processes the information and feeds it to a communications system.

Retroreflective Labels

All of a car's identification data is encoded and condensed into a single multi-colored label. The car labels,

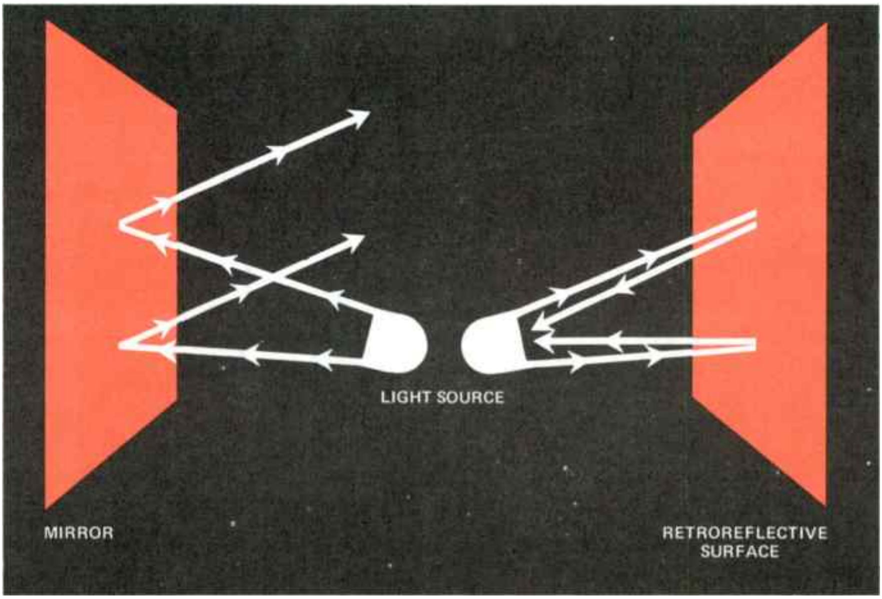


Figure 2. Retroreflection differs from ordinary mirror reflection in that light is reflected directly back to the source regardless of its incoming angle, or of any slanting of the surface.

LABEL MODULE CODE STRUCTURE

NUMBER	FIRST PART RED	FIRST PART BLUE	SECOND PART RED	SECOND PART BLUE
0	0	X	X	X
1	X	X	X	X
2	X	X	X	0
3	X	0	0	0
4	X	0	X	0
5	0	X	0	0
6	X	X	0	X
7	X	0	X	X
8	X	X	0	0
9	0	X	0	X
STOP 10	0	X	X	0
START	X	0	0	X

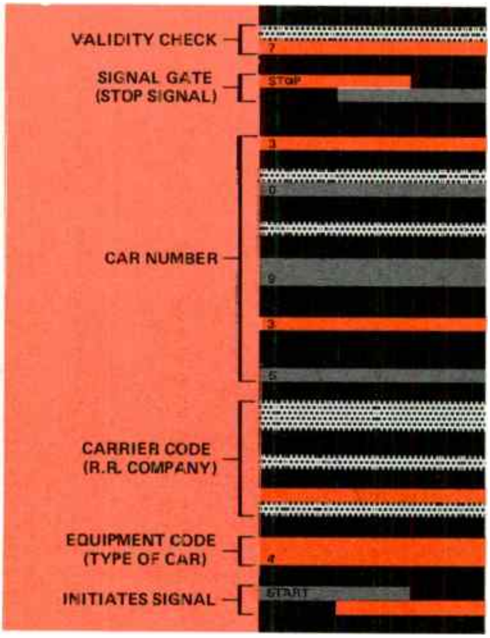


Figure 3. An ACI label has 13 parts, each divided into two strips. As the scanner reads the label from the bottom to the top, combinations of the colors red, blue and white are set in coded pairs to represent the numerical values in the accompanying chart. For example, the START symbol is represented by a red strip in the first part, and by a blue strip in the second. These are indicated by the x's in the Code Structure Chart. Reflected light from a white (checkered) strip triggers both the red and the blue sensors in the scanner and is represented in the chart by x's in both columns. (Note: The label strips appearing as grey in the illustration are bright blue in an actual ACI label.)

affixed to both sides of the car, are made from adhesive-backed, retroreflective sheeting very similar to that seen on automobile bumper strips and highway warning signs. Light striking such material is reflected back to its source along the same path it followed when transmitted. As illustrated in Figure 2, retroreflection is different from mirror reflection where light is reflected at an angle opposite to that from which it came, or away from its source; unless of course the mirror and the light source are perpendicularly opposed. A retroreflective surface, on the other hand, reflects light back to its source regardless of the light's incoming angle — in much the same manner as a corner reflector antenna works.

The label surface, developed by 3M Company, is coated with tiny glass beads — about 90,000 of them per square inch — each of which is its own

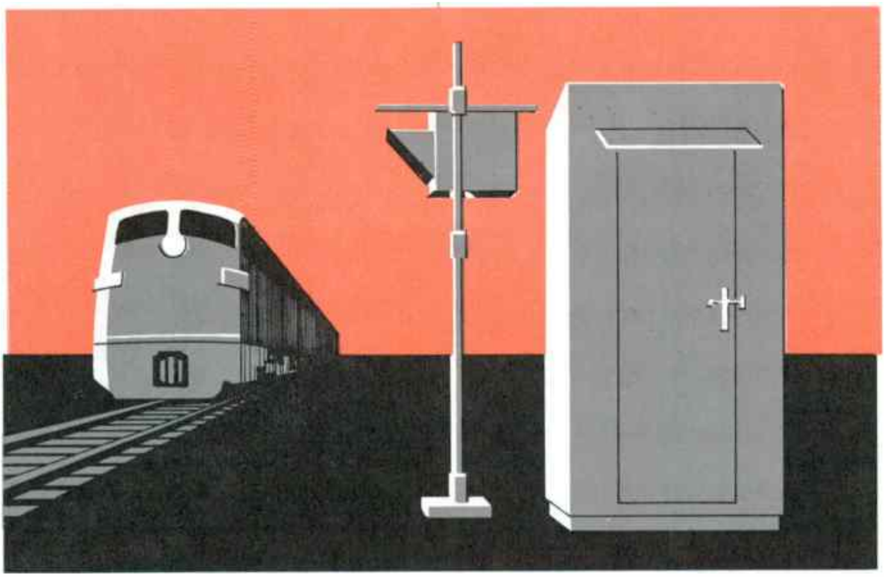


Figure 4. Mounted at trackside so as to provide maximum view of the passing trains, the scanner covers a vertical area of just under nine feet. Its protective housing allows the scanner to operate at temperatures ranging from -50^{o} to 150^{o} Fahrenheit.

optical system. En masse, the beads reflect light back to the wayside scanner — light that is 200 times more intense than normal reflected light from a colored object.

Each car label is made up of thirteen 1 x 6 inch colored strips of sheeting. The strips are arranged one atop the other in a color coded sequence and stuck to the side of the car. The color code consists of some combination of red, blue, black or white (Figure 3). Each strip is divided into two parts whose colors represent discrete numbers to the trackside scanner.

It's All Done With Mirrors

The heart of the system is the scanner. Its primary function is to read the labels from the passing cars — in fact, it reads each label up to four times as the train passes.

As illustrated in Figure 4, the scanning equipment is mounted at trackside in a weatherproof steel housing. Its mounting position permits the scanner to cover an area reaching from 16 inches to 9.5 feet above track level. This allows the scanner to pick up every passing label regardless of the type of car it is on or its location on the car.

The schematic diagram in Figure 5, shows the various operating components of the scanner. A 9,000 watt xenon lamp is the system's light source. The light is first routed through the system by a series of mirrors to the multifaceted scanning wheel, each facet of which is also a mirror. As the wheel spins at a high speed, it causes the light beam to move from the bottom of the scan to the top at the same high rate.

The light is projected at the label and then follows the same path back to the partially silvered mirror. From here, the light is focused through the lens to create an image of the label. This image is then transmitted through the slit plate. Due to the rapid rotation of the scanning wheel and the narrow aperture in the plate, only a small portion of the label's image passes through the plate at any one instant. Hence, a time sequential pulse train of light is created. This light train is analogous to the car label – that is, composed of bands of red, blue or white light. At this point, the light is optically filtered into two broad spectra defined as red and blue. Here, photomultipliers change the optical signals into electrical pulses for input to the decoder.

Decoder

Interpretation of the data from the scanner is the primary function of the decoder. It is basically a digital device composed of analog to digital (A/D) conversion circuits, digital logic and output circuitry. The A/D conversion circuitry changes the electrical input from the scanner into meaningful digital values for transmission as data. Logic circuits analyze incoming signals

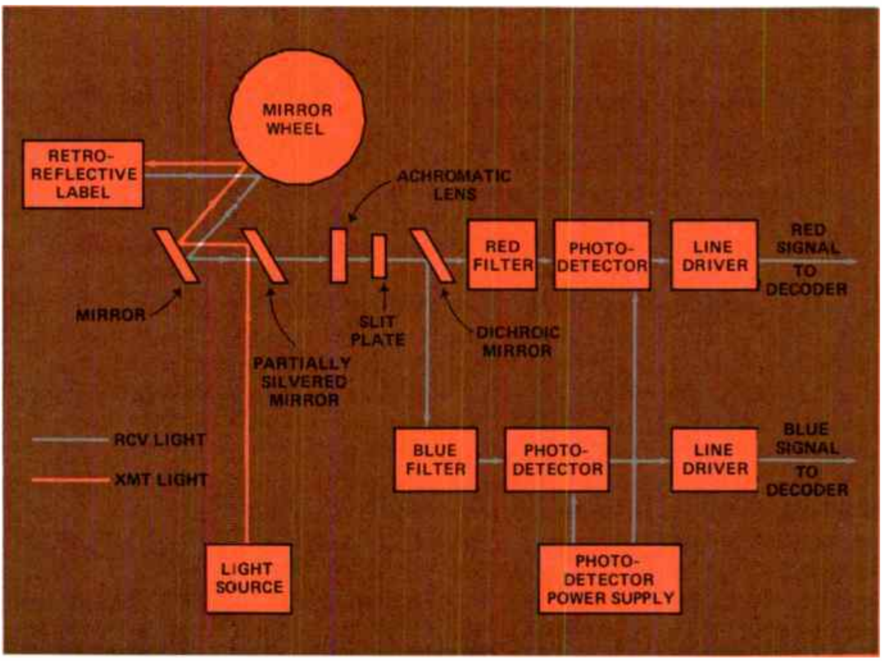


Figure 5. The scanner, represented schematically here, reads the passing labels, separates the color signals and relays them to the decoder.

and "decide" whether a proper label has been scanned and accuracy requirements have been met.

As outlined in Figure 6, the basic elements of the decoder's logic are the label data register, validity check circuitry, a bad label designator, and an unlabeled car detector. The output circuitry is separate.

The decoder also has a storage function sufficient to store the data from one label. As the incoming data is stored in the label data register, the validity check circuitry makes the necessary arithmetic calculations and stores the result. While this is going on, another part of the validity check circuitry monitors the input for the presence of Start and Stop digits (see Figure 3). Thus, the circuitry is able to indicate the presence of label data which meets the predetermined criteria for validity. Normally, the pattern is a Start indication followed by ten digits of information which is in turn followed by the Stop indicator. The thirteenth digit is derived from a mathematical calculation for determining the validity of the color code.

Positive and negative checks indicate that the information received in the label data register is correct or incorrect. When a positive comparison is indicated, the information is transmitted directly to the output circuitry. But, if a negative comparison occurs, the information in the label data register is transferred to the bad label

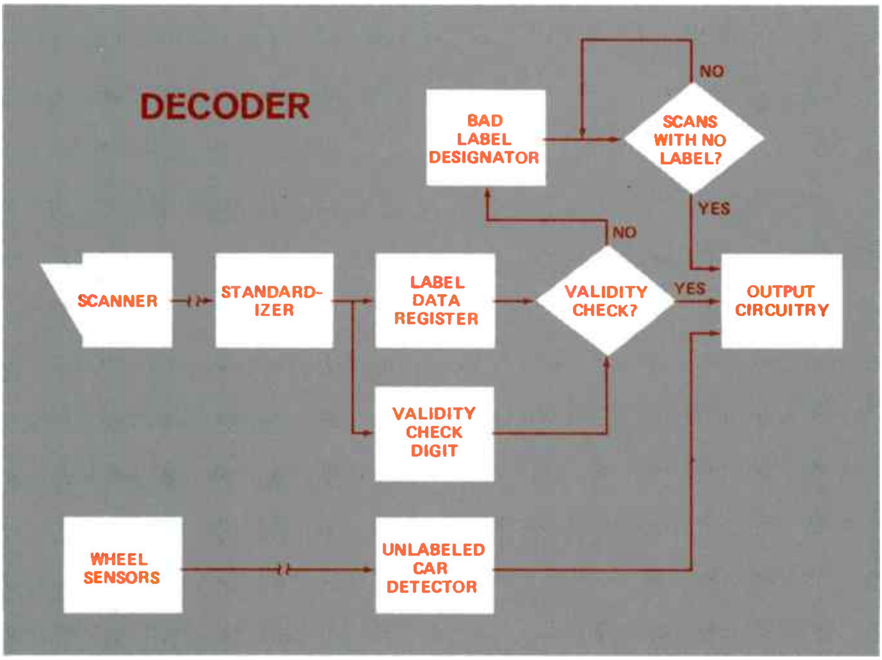


Figure 6. Primarily, the decoder's job is to receive the color signals and convert them into meaningful electrical impulses for transmission as data.

designator for storage. The data remains here during the several subsequent scans until either a positive check is obtained and the data can be forwarded or no positive check is received and special handling instructions are sent with the bad label designation. Such a designation is received at the computer terminal as a question mark.

Another device, the unlabeled car detector, works in conjunction with electronic wheel sensors on the track. The wheel sensor counts the passing cars, and relays a signal for each car to the unlabeled car detector. If this signal is not cancelled by the indication from the label data register that a label has been read, the unlabeled car detector generates a series of zeros to the output circuitry indicating that an unlabeled car has passed.

Output from the basic system is in either 5-level Baudot or ASCII code. The output circuitry interfaces with teletype equipment locally or with data modems for transmission of the data to some distant point. Lenkurt's new data set, the 25B, will handle ACI data as well as telegraph and telemetry traffic.

The output circuitry has the additional function of transforming data into a selected output code and data rate for either direct printout or further processing.

Optional Equipment

When railroad traffic density or train speeds dictate system sophistication beyond that of the basic ACI system, certain optional equipment may be implemented to meet these requirements.

Since time is an all-important factor in railroad traffic control, it is normally desirable to have times and dates recorded along with the other ACI information. The optional device for this function is a calendar clock which is incorporated into the decoder assembly. The calendar clock generates a total of seven characters of information. The first three indicate the date while the four following digits indicate time on a 24 hour basis. Hence, 927 1440 would be translated as Sept. 27 at 2:40 p.m. However, the calendar clock cannot be used alone. Another unit, called a message generator, must work with it.

Most data messages are in three parts – prefix, body and suffix. In this case the ACI data, generated by the scanner, is the body of the message. The prefix and suffix are added by the message generator. These may include such information as address group, time and date, control characters, direction of train approach or any other configuration of the 64 available alpha-numeric characters that a user might dictate.

Buffers

Sixteen characters of information are generated by each labeled car that passes a scanner. Normally, depending on the code structure, this is represented by about 150 binary digits. Given an average car length of 44 feet and a train speed of 80 mph, this comes to about 300 b/s as a required signaling speed. Calculation is further complicated by carriers bearing more than one label – such as piggyback cars. By AAR specification, labels on multiple carriers must be at least six feet apart. In this case, the peak signaling speed can reach 2400 b/s.

Obviously, if a 2400 b/s data system such as the Lenkurt 26C is in service, there is no transmission problem. But, if the existing railroad signaling equipment is not capable of such speed, some sort of buffering or

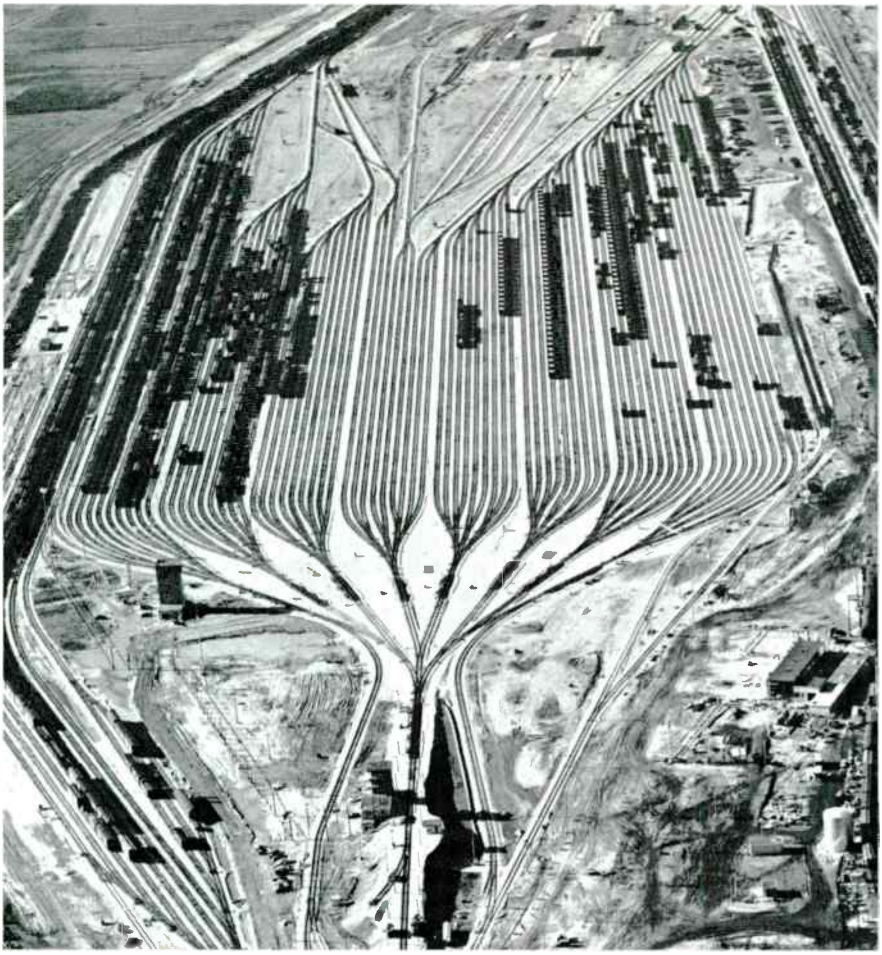


Figure 7. The complexity of marshalling yards such as this one illustrates the need for automatic car identification.

store-and-forward system becomes necessary.

The Kartrak ACI system has two kinds of buffer storage – magnetic tape or core storage. The magnetic tape buffer is an endless loop of recording tape which can store label data from as many as ten 200-car trains. Playback is on command, and if another train approaches while the buffer is operating in playback mode, label data from the new train will still be stored simultaneously.

Functionally, the core buffer is identical to the magnetic tape device. Their capacities, however, are different. Nominally, the core buffer can store data from only 500 labels; with optional attachments capacity can be increased to 1,000.

Other additional options increase the system's capabilities so that piggy-back, stacked-container and other special cars can be read and recorded.

Communications Interface

For data communications, most railroads use one of three transmission systems – point-to-point teletype, multipoint teletype or the IBM 1050 multipoint data communications system. In the teletype arrangements, telegraph wires are normally the transmission media, but dedicated, voice grade circuits can also be used. In the case of high speed data though, voice circuits are the rule.

In order that ACI can be used in conjunction with existing data communications systems – particularly the teletype arrangements – it is necessary to employ some method of control compatible with these common systems. Using a teleprinter control unit, a buffered ACI system is able to operate point-to-point over dedicated circuits. Stored data is transmitted from the buffers to centrally located teleprinter units. Codes for these units are either 5-level Baudot or ASCII. Output hardware can be either off/on keying units for driving a standard telegraph loop, or a regular EIA interface for use with data modems. The Lenkurt 25B data set is designed for use with either telegraph or data, or both.

Again, where the IBM 1050 or other computerized systems are used, access to the telephone network is via medium or high-speed data sets. Four Kartrak systems are scheduled for installation on a freight line in the Midwest. They will use Lenkurt 26C data transmission sets for communication with a Sylvania TCS-50 computer over dedicated lines. The 26C's bit rate of 2400 b/s enables the system to operate without buffer storage. Communication between ACI system and computer is direct and on a real-time basis.

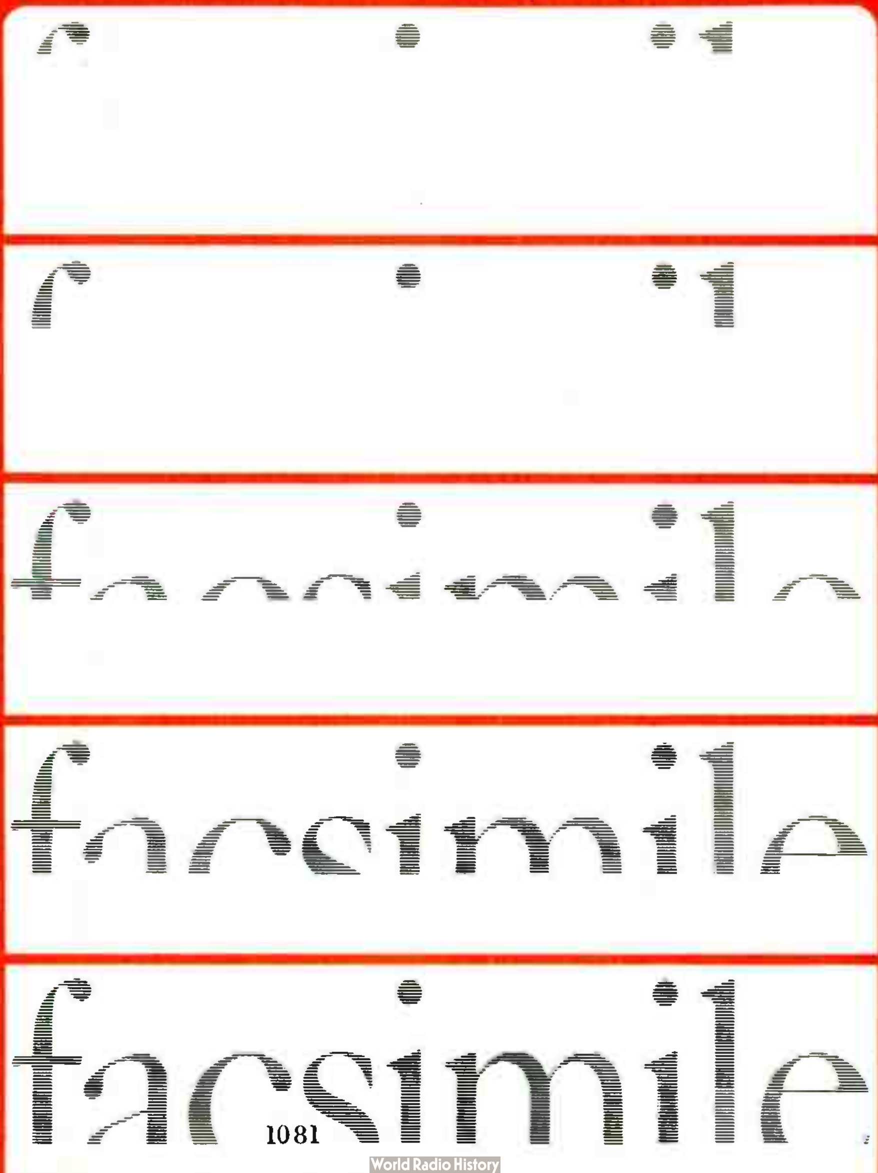
Instant Location

By 1970, all the nation's railroads should have a functioning ACI system. A traffic controller will be able to find any car on his line anywhere in the country, instantaneously. Computers will sort out all the complex billing and usage information and accounts will be credited and debited automatically.

Interestingly enough though, the really crucial and exacting job must be done manually. Somebody – some man – has to put all those labels on all those cars.

The Lenkurt
DEMODULATOR

SEPTEMBER 1969



facsimile

That old Chinese proverb about a picture being worth a thousand words is no less true in this day of highspeed electronic communications than it was in the early days of paper. As technology improves, the relative value of picture transmission may well exceed that of a thousand words.

The art of sending pictures, maps or whole messages over radio, telephone, telegraph and cable facilities is known as facsimile transmission, or "fax". The basic principles of fax have been understood for almost 100 years. Only in recent years, however, have transmission techniques improved to the point that this unique form of telecommunications can begin to be commercially exploited.

Technological advances in the field of data transmission have, incidentally, made it possible to develop fax systems which significantly reduce transmission time and thereby lower line cost. The expense associated with older transmission techniques has been the most limiting factor in the history of fax. A reduction of common carrier rates, in 1962, considerably improved the commercial outlook of fax, but future advances in data transmission technology will probably have a greater effect.

Background

The early applications of facsimile were essentially limited to the transmission of photographs, telegrams and weather maps. One of the oldest commercial users of fax is Western Union Telegraph Co., presently operating more than 30,000 units for the delivery of telegrams. Thousands of fax equipments are used by the military, weather bureaus and commercial airlines for the recording of weather maps. The advent of weather satellites has made this function more reliable and timely.

The transmission of photographs and weather maps requires high resolution and good quality reproduction. Even early fax systems were capable of this kind of performance – quality was not the problem. The problem, until recently, has been providing economical methods for the transmission of large volumes of lesser quality material, such as letters and line drawings. Many similar applications do not require high resolution.

The early uses of fax persisted because of the need to transmit certain kinds of high resolution graphic material from one place to another regardless of cost. Photographs for newspapers, emergency telegrams for people unreachable by phone, and weather maps to chart airline flights are examples of such indispensable services. The use of facsimile as a necessary tool in keeping the public informed was assured by the Federal Communications Commission when they reserved a portion of the radio spectrum specifically for the transmission of fax by newspapers.

In the light of its present development, it is appropriate to reassess facsimile's role in the field of electronic communications. In a fundamental way, fax provides some unique advantages over other methods of communication. Visual material is often less ambiguous than audio material; most people are better visually-oriented than audio-oriented. Because of this, visual material is not as often misperceived as material which is only "heard". There is also the added benefit of having a readily available permanent record for verification when accuracy is in doubt. Fax material can be more quickly copied for distribution or filing. It can also handle a broad scope of information without the problem of language barriers. Noisy rooms do not impair reception. Fax is as easy to use as a telephone. Another interesting facet is its capacity to transmit legally acceptable documents and signatures.

Low resolution, black and white machines are now being widely manufactured. The availability of these new devices and prognostications of a substantial market, has ushered in a new nationwide business — the facsimile franchising service field which promises to become a booming enterprise. Offices already exist in more than 200 cities across the U.S., making this new service available to any commercial activity wishing to participate. New offices are being opened with each passing week.

There are two types of franchising currently in use. One method has a central office to maintain a pick-up and delivery service to subscribing customers for a monthly rate. The second method invites the franchise holder to establish a secretarial service-type office where anyone desiring to transmit written information presents the information, by person or phone, and is charged on a cost-per-copy basis.

Basic Fax

The principle types of fax are gray-scale and black-and-white. Recently announced color fax is essentially an extension of gray-scale fax technology. All types require a revolving drum, upon which can be attached the copy to be transmitted, and a scanning device which "sees" the copy as it revolves. Line by line the scanner picks up the image, in the form of light impulses, as it sweeps across the page. The received light pattern passes through a precision optical system to a photocell which, in turn, converts the light impulses into an analog signal pattern. Both gray-scale and black-and-white fax can be transmitted over common communications channels by analog means.

Shades of Gray

The signal pattern, or waveform, from a wide-range gray-scale fax system differs from that of a black and white system. A wide-range gray-scale waveform may have an infinite number of tone values between black and white. The waveform produced is a precise electronic representation of the original image on the drum. This kind of waveform is both asynchronous and analog in nature. It is asynchronous because the waveform has not been "sampled" or "clocked" at measured time intervals (see figure 1) and is considered analog because its amplitude varies in direct proportion to the illumination spectrum of the original material. Asynchronous, analog fax systems are well suited for the transmission of pictures and other graphic material where good definition of a wide range of gray tones is necessary.

Good quality gray-scale transmis-

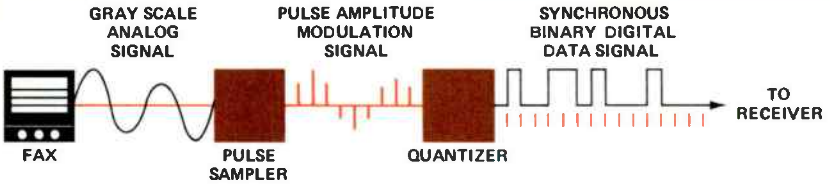


Figure 1. By using a "sampling" device, a continuous, analog signal can be converted into a synchronized, pulse amplitude signal. This signal can further be quantized and encoded into binary digital form.

sion is costly. Since it requires broadband facilities capable of high signal-to-noise ratios.

Fax in Black and White

A simpler system operates only in black and white — no shades of gray. This system employs a scanning device which only registers a black or white impulse and therefore generates a waveform with only two values corresponding to black and white. The digital binary waveform produced by this system is asynchronous and analog in nature because the time interval between two successive transitions corresponds directly to the original image. The application of the word analog is, thus, somewhat less meaningful when it concerns black-and-white systems; its common use is generally reserved for reference to gray-scale systems.

The binary nature of a black and white waveform makes it convenient for this form of fax to take advantage of present and future advances in data transmission techniques.

Fax Transmission

In telecommunications transmission there is an inverse relationship between transmission bandwidth and transmission time. This is particularly significant with fax. As bandwidth is increased, the time is decreased, and vice-versa. In general, the bandwidth necessary for fax transmission depends on the required resolution and the speed of transmission.

The relationship between resolution and bandwidth is such that doubling resolution requires four times the bandwidth — if transmission time remains constant. In the future, any significant innovations in effectively reducing transmission costs must reduce both transmission bandwidth and transmission time.

PCM and FAX

A significant new development in the telecommunications industry took place in 1937 with the invention of pulse code modulation (PCM). However, not until the advent of transistors and integrated circuits was this new idea given serious attention. In the 1950's, Bell Telephone Laboratories began to explore the PCM system.

The basic concept of PCM was entirely different from frequency division multiplex (FDM) which has served the industry for so many years. Unlike FDM, which provides frequency separation for all channels, PCM involves switching rapidly from one channel to another — producing a time division multiplex (TDM) system where each channel is allotted a separate time interval.

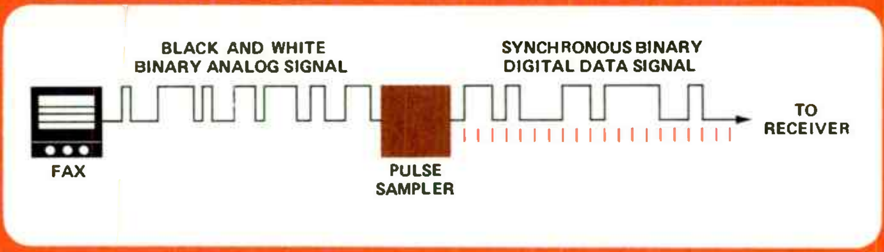


Figure 2. A "sampling" device applied to a black-and-white, binary signal, produces a synchronous, digital binary signal in one step. The binary nature of this mode effectively lowers transmission noise but introduces "jitter".

The Bell System designated their 24-channel time division multiplexing PCM system as T1 carrier. Although primarily designed for voice transmission, the T1 has proven to be ideally suited to the transmission of data or fax.

The use of PCM systems in the telephone network is increasing very rapidly. An example of this growth is the increasing number of Lenkurt's 91A (comparable to Bell's T1) time division multiplex PCM systems now in operation.

Lenkurt's 9003 and 9005 high-speed data terminals are designed for use over 91A (or T1) repeatered lines. Providing data rates of 50 Kb/s (and higher), these data terminals are well suited to the transmission of fax.

The PCM mode of transmission is not greatly troubled by signal distortion caused from noise. It has, however, another kind of distortion problem for black and white fax. When a black and white fax binary wave is synchronized by a sampling pulse it produces an error in the signal pattern which is called "time jitter" (see figure 3). This "jitter" occurs because the sampling pulse rarely takes place at the actual point of transition between black and white. As a result, the synchronized waveform rounds-off the original pattern, causing the intervals to be stretched or compressed, accordingly. The received image may therefore present a slightly fuzzy appearance due to this "time jitter", although legibility is not greatly impaired.

One advantage of binary data transmission is the fact that since there are only two possible waveform pulse values — one or zero, transmission distortion is relatively low and does not increase with distance. The efficiency of this system results from the fact that each positive pulse is regenerated at full amplitude, regardless of signal degradation up to that point. Likewise, every zero pulse is regenerated at zero amplitude.

Data Compression

Another promising field of investigation for bandwidth and time reduction, is the elimination of image redundancy. Nearly all graphic material has a dominant tone-value. For example, a page of typewritten material is essentially white — even when filled with words. This repetition of the same tone, over and over, is redundancy. The conclusion has been logically drawn that if only the non-redundant data could be transmitted (the black letters on a typed page, for example) a reduction in required bandwidth and time might be accomplished. Several

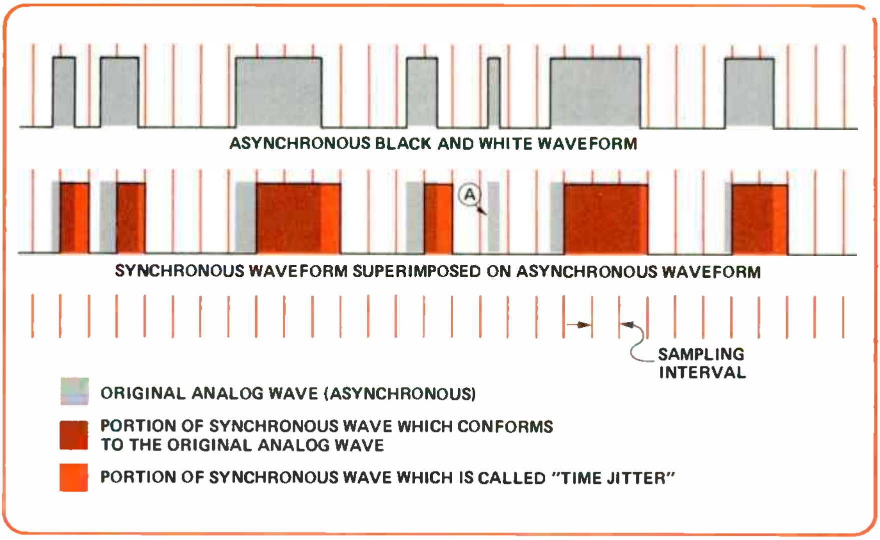


Figure 3. When a black and white analog waveform is synchronized by sampling pulses (vertical red lines), a certain amount of distortion occurs which is called "time jitter". This jitter results from the fact that the sampling eliminates changes in the original waveform which take place between pulses. The entire interval between pulses will read either "up" or "down" depending on the location of the original waveform at the time of the last pulse. Obviously, some complete white-to-black-to-white (and vice versa) changes can take place between pulses and therefore be completely eliminated from the synchronized wave (A).

techniques are being explored to eliminate redundancy and some methods are presently in use. One system of this type encodes the location of black-white and white-black transitions and transmits this information through digital transmission systems such as Lenkurt's type 26C or 26D data modems. Using Lenkurt's 26D, speed improvements as high as four to one have been achieved.

Present Applications and Future Possibilities

Early uses of facsimile have been enhanced and new uses are constantly being explored. Effective techniques for removing redundancy promise to place facsimile within the reach of many new applications.

Significant improvements in facsimile technology have reawakened the interest of police authorities around the globe.

Police departments which have installed facsimile systems are finding their use indispensable in communication of fingerprints and "mug" shots. There has long been a need for rapid identification of suspected criminals. Court imposed restrictions on police detention practices served to highlight this need. Rapid and positive identification not only makes it easier to identify wanted criminals, it has the added benefit of facilitating the release of innocent persons.

The Chicago and Los Angeles police departments have had, for some time, operational facsimile systems using high resolution data transmission equipment. This application requires good quality gray-scale reproductions and 200 line per inch resolution – twice the resolution of a typical copying machine.

The newspaper industry represents one of the largest single markets for facsimile transmission systems. Many newspapers are now making use of various available devices for this purpose. It has been found desirable to provide high resolution of this large format material with wide-range grayscale. It is necessary that newspaper pages be transmitted at very high resolution rates, on the order of 1000 lines per inch, to maintain sufficient standards of quality of half-tone picture material. Perhaps a system could be devised to transmit newspaper printed material at about 200 lines per inch, then slow down for half-tone picture material at 1000 lines per inch.

Facsimile systems have been used to transmit full size newspaper proof sheets from the main plant to remote satellite printing plants where the transmitted page is converted into a photographic negative which is then used for direct reproduction. At the present state of technological development, this application is still relatively expensive.

The future of facsimile transmission in the newspaper industry is very promising. Elaborate systems have been proposed which would require complete microwave networks capable of transmitting a standard newspaper page at 1000 lines per inch resolution within time limits of about four minutes per page between plants.

The Wall Street Journal already achieves fast service to parts of California and other Western states by using facsimile to transmit page proofs. Many banks are now using facsimile to speed stop-payment notices on checks by transmitting the notice from the bookkeeping office to the main banking office. Manufacturing companies are using facsimile equipment as an adjunct to their teleconferences in order to provide charts and diagrams so that each participant has precise references for the topics of discussion.

Courtesy of Dacom, Inc.

Figure 4. This black and white fax system transmitter produced by Dacom, Sunnyvale, California, makes use of Lenkurt's 26D data modem.

It has been reported that there are over 18,000 current users of facsimile systems. This list includes railroads, hospitals, schools, airlines and general business. It can be safely predicted that as new advances are made in compression techniques, bandwidth requirements will be reduced, costs will drop and new applications will become virtually endless.

The Lenkurt

FEBRUARY 1970

DEMODULATOR

HOLOGRAPHY

Holography, from the Greek root *holos* meaning whole, is a picture making process that captures the three-dimensional aspects of an object, rather than the flat, fixed-viewpoint of conventional photography.

In holography, three-dimensional images are formed from two-dimensional photographic negatives. In the recording process, a coherent wave source is split into two parts (Figure 1). Half the source (reference wave) strikes the holographic plate directly. The other half (object wave) illuminates the object to be recorded. Each point on the object reflects light onto the holographic plate. Having traveled different paths, the two waves are no longer in phase, and therefore reinforce or cancel each other as they converge on the holographic plate – producing an interference pattern (Figure 2). There is a unique interference pattern recorded over the entire holographic plate, for each point on the object.

Laser light is the most commonly used coherent source, but illumination may also be accomplished with electron waves, X-rays, microwaves, and acoustic waves.

Once the holographic plate has been exposed and processed, it is capable of reconstructing the original three-dimensional object. Reconstruction is accomplished by illuminating the hologram with the same frequency reference wave used for recording. Since the hologram records all the information that the object wave contains, the reconstructed image will display this information – the size and shape of the object; the brightness of every point on the object; and the position of the object in space, from all angles that are intercepted by the holographic plate during the recording process.

Basically, a hologram is a recording of two coherent waves. When the hologram is illuminated with one wave, the other wave is simultaneously reconstructed.

There are two fundamental types of hologram – transmission and reflection. In a darkened room, transmission holograms are illuminated from behind with monochromatic light (one color). Reflection holograms, however, with their built-in filters are illuminated with white light (all colors) from the side where the viewer is standing, so

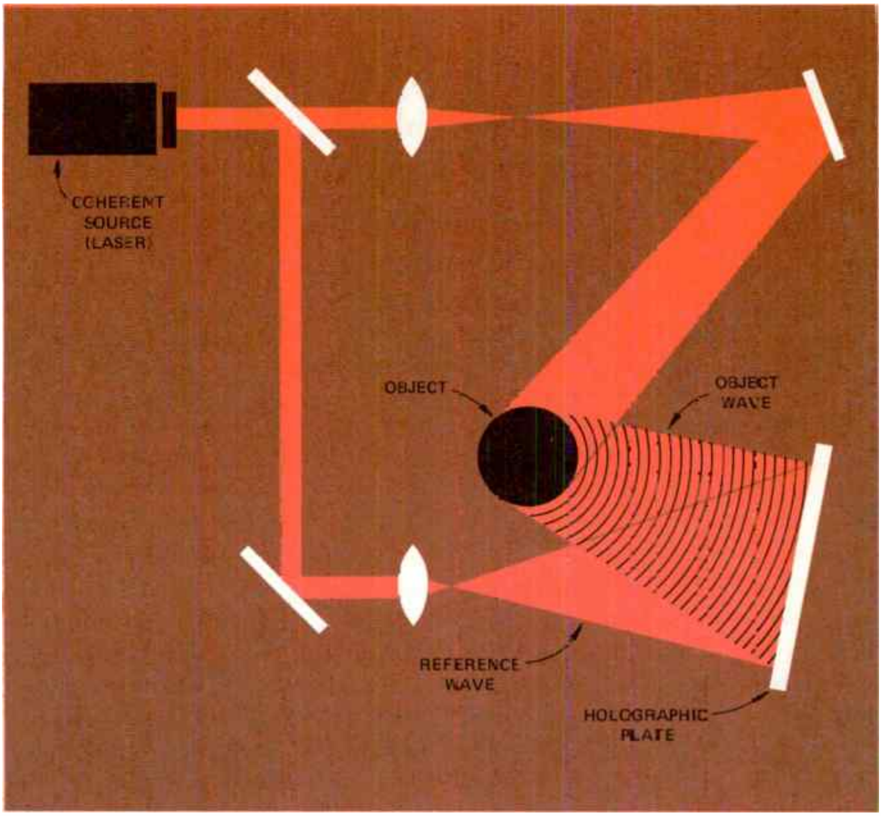


Figure 1. Holographic recording process using laser illumination. The laser beam is split, with the object waves illuminating the object and the reference wave providing a coherent background. The two waves interfere at the holographic plate forming the hologram.

the light is reflected from the hologram to the viewer. A reflection hologram is easier to handle because it may be illuminated in subdued lighting; although, it is not as dramatic as the transmission hologram.

Background

Holography is not a new concept. Dennis Gabor, in 1948, introduced the theory of holography. It was his hope that holography could be used to improve the resolution of electron microscopes. Unfortunately, limited by the intensity of the illumination, the photographic processing, and a disturbing background image, Gabor and the others experimenting with holography did not get the results they desired. The most important of these

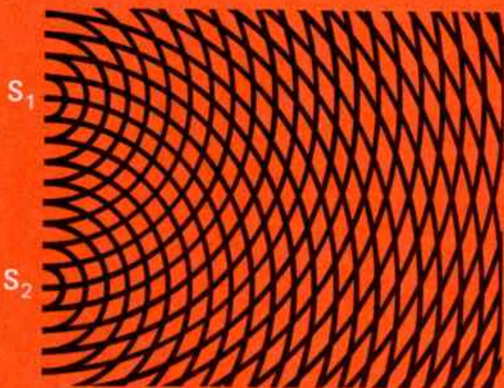


Figure 2. Interference pattern formed from two arbitrary point sources (S_1 and S_2). Place your eye near the left edge and look along the figure. The areas that look white show the cancellation of the waves; between these areas there is reinforcement.

limitations was the illumination intensity.

By 1962, with the development of the laser, intense coherent light became available – over-coming the illumination limitations and eliminating the background image. Laser light also made it possible to record objects which were not easily recorded in Gabor's original system. Objects with dark backgrounds and continuous tones could now be holographically recorded.

Volume Holograms

Gabor's original hologram theory did not use coherent light; therefore, the thickness of the recording emulsion was considered inconsequential and the hologram was viewed as two-dimensional. With the use of coherent laser light, however, the emulsion thickness became an important factor. Specifically, if the emulsion is thicker than the width of the interference fringes (see Figure 3), the object wave and the reference wave will interfere throughout the depth of the emulsion. This produces a volume hologram which is a stack of surface holograms, one atop the other.

If a reflection, volume hologram is made by using three colors of light (blue, green, and red), three holograms will be recorded within the same emulsion. When this hologram is illuminated by white light, each hologram will select the color from the white light to which it responds, and the result will be a three-dimensional color image.

Whole from Part

The entire image from transmission holograms can be reconstructed from a fragment of the original hologram. This is not surprising, since each point on the object is recorded as an interference pattern, or diffraction grating, over the entire hologram surface. This grating can be thought of as a Fresnel lens (see Figure 4). Such a lens focuses all the light falling on it at a particular point. A Fresnel lens also has the property that regardless of how many "rings" there are in the lens segment, or whether there is a complete ring, the light will still be focused at the same point. The resolution (clarity) of the point is determined by the number of rings used. Therefore, as a smaller and smaller section of each Fresnel lens or diffraction grating is used, the clarity is diminished, but the point will still be imaged.

For the same reason that a part of the hologram will reconstruct the whole image, the hologram is relatively insensitive to blemishes and dust parti-

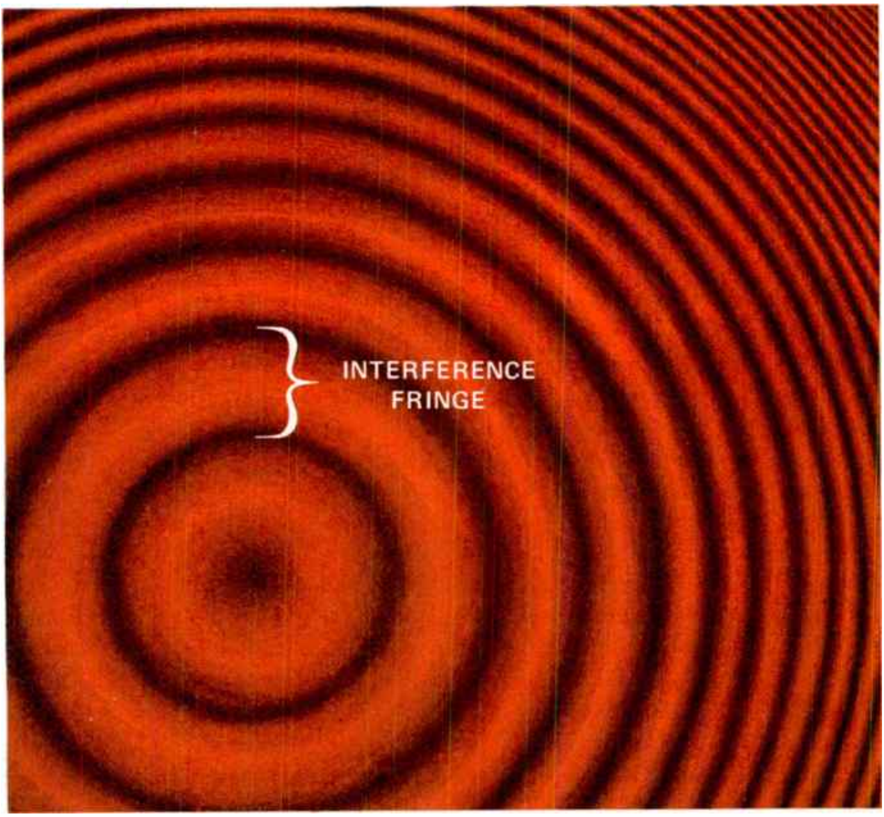


Figure 3. A concentric ring, diffraction grating, for one point on the object, covers the entire holographic plate. The center of the grating need not be on the holographic plate. Such a grating, with variable spacing, focuses the light falling on it at a particular point.

cles. If a blemish destroys part of a diffraction grating, but not all of it, the light will still be properly imaged. The image will not be destroyed unless the total surface of the hologram has been obliterated.

One Hologram Worth Many Words

The large depth of field of holograms is a great advantage in microscopic investigations. It is particularly valuable for examining moving, microscopic objects in a thick sample. A pulsed laser is used to "freeze" the movement in a sample so it can be recorded on the hologram. When the object is reconstructed and viewed through a microscope, different layers of the sample can be brought into focus – something that cannot be done with a photograph where the movement has been stopped with a strobe light. This principle has been used to analyze the size and distribution of particles in aerosols, liquids, and smog. The old adage, "a picture is worth a thousand words," can be extended – "one hologram is worth a great many pictures."

It was Gabor's hope that holography, being lensless photography, would improve the resolution of electron microscopes. Unfortunately, many of the problems that faced Gabor are still present today; since a magnification hologram has all the distortions observed with conventional lenses.

For holography at non-visible wavelengths, the situation is vastly different. For example, magnification is possible using X-rays for recording and visible light for reconstruction – producing a sharply focused, magnified image. However, the technical problems of X-ray holography are severe, and have prevented its practical realization. Obtaining an X-ray source with sufficient intensity and coherence is a major obstacle.

3-D Imaging

The three-dimensional aspect of holography has received widespread attention. Holography has several advantages related to 3-D imaging. Holographic reconstruction gives an image with high resolution and great depth of field. Parallax, as observed in reconstructed holographic images, allows the viewer to see around objects. Holograms also have some economic advantages such as full color images formed from less expensive black and white emulsions and no need for additional imaging optics.

Holographic three-dimensional imaging also has some disadvantages. The object must remain perfectly still for recording. The object motion can, however, be stopped by using a pulsed-laser for illumination. At the present, it is virtually impossible to take holographic pictures in daylight or under normal illumination because the object to be recorded must be illuminated with only one wavelength of light. Multi-color images are not promising, since these images require long exposures. The size of the object to be recorded is limited by the laser power. The largest hologram that has been made is 18 x 24 inches.

A 3-D holographic television system could be designed today, if available

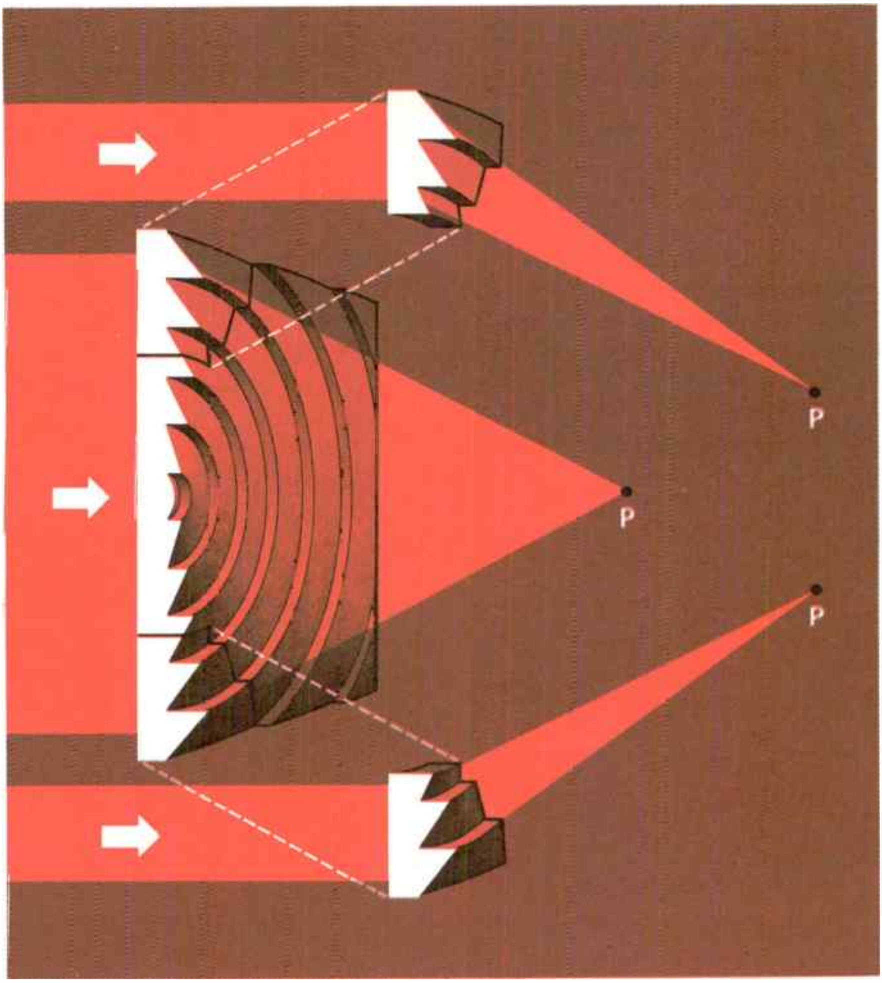


Figure 4. The light falling on a Fresnel lens will focus at a particular point (P) regardless of how large a segment of the lens is used.

components held tighter specifications. For example, a hologram 10 inches square and having 1,000 lines per millimeter has 6×10^{10} picture elements, compared with 2.5×10^5 for a conventional television picture. If the scan rates of present TV systems were maintained, a 10^5 increase in bandwidth would be necessary – a jump from 6 to 600,000 MHz. The entire radio-frequency spectrum, including the microwave region, would be inadequate to meet this requirement. If a suitable transmission method can be found, it may one day be possible to watch holographic 3-D television in the middle of your own living room.

The first attempt at holographic movies involved taking a series of pictures of still objects and then viewing them in rapid succession producing a sensation of motion – the animated cartoon concept. The term "true" is used to describe the newest 3-D movie because it is a motion picture of a moving object. (Figure 5). This reconstructed image is truly three-dimensional and may be viewed without additional lenses or filters.

Change Detection

Holography has many properties that make it natural for application to interferometry, a technique used to detect structural changes. For example, a hologram placed in its recording position will display the reconstructed image on top of the actual object. A subsequent slight movement of the object produces interference fringes between the object and image waves.

A hologram can be exposed more than once. The image waves, recorded at different times, can be simultaneously reconstructed and their interference pattern observed (Figure 2). Shock waves, for example, produced by projectiles passing through air or gas density changes can be easily recorded with pulsed-laser, double-exposure holograms.

A new method of holographic interferometry (HI) has been introduced which uses a single-exposure, two-wavelength laser pulse in the recording process; rather than the double-exposure, single-wavelength formerly used. This new technique can double sensitivity by proper choice of wavelengths and physical arrangement of components.

The field of nondestructive testing makes extensive use of HI. Holographic nondestructive testing (HNDT) is a method of detecting or measuring the significant properties or performance capabilities of materials, parts, assemblies, equipment, or structures without impairing their serviceability. HNDT is simply HI combined with suitable test-object stressing. Common types of stressing include temperature, pressure, sound, and vibration. HNDT is now a practical design and quality control tool for analyzing sandwich structures, tires, rubber-to-metal bonds, and many other objects. Large areas may be inspected quickly, a variety of flaws can be detected simultaneously, and a choice of several low level stressing methods is available.

Cryptography

A hologram is a coded message that can be decoded by using a coherent illumination source. If, however, the hologram were made with a diffuser such as ground glass placed between the object and the holographic plate, it could only be decoded using the same diffuser. If the same diffuser is placed so that it coincides with its own image, a sharp, clear picture of the original object is reconstructed – as if the diffuser were a clear glass.

The same diffuser must be used in both recording and reconstruction. A section of ground glass, 1 centimeter

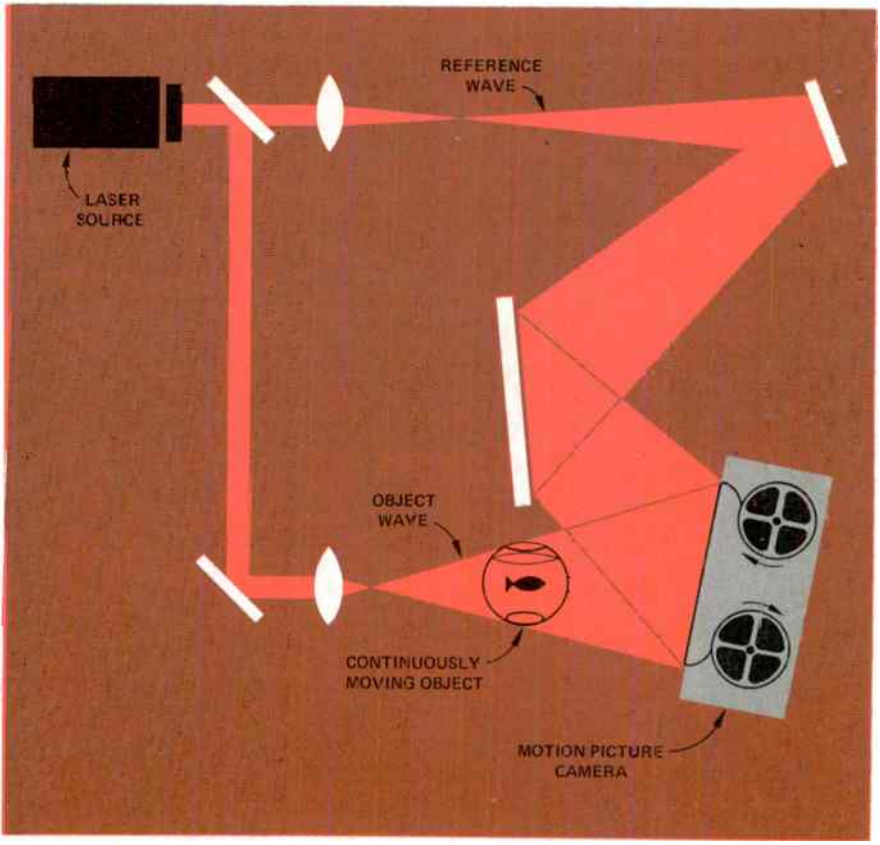


Figure 5. The basic system used to make the first "true" holographic movie.

square, can contain a billion distinct resolution elements, each of which retards the transmitted light. It is not probable that two ground glass samples would be similar enough to allow an image to form. With only one "key" to a cryptographic hologram, the message is considered secure.

Memories Can Forget

Holograms are useful for data storage because many holograms can be superimposed on the same photographic plate. Built-in redundancy is a must in today's computer systems – transmission holograms, with their insensitivity to blemishes and ability to reproduce an entire image from a small fragment, provide the necessary redundancy. Optical memories capable of storing 10^4 bits of data on each page are stored in 1 mm^2 on the hologram. A 10 cm^2 hologram stores 10^4 pages – totaling 10^8 bits per hologram.

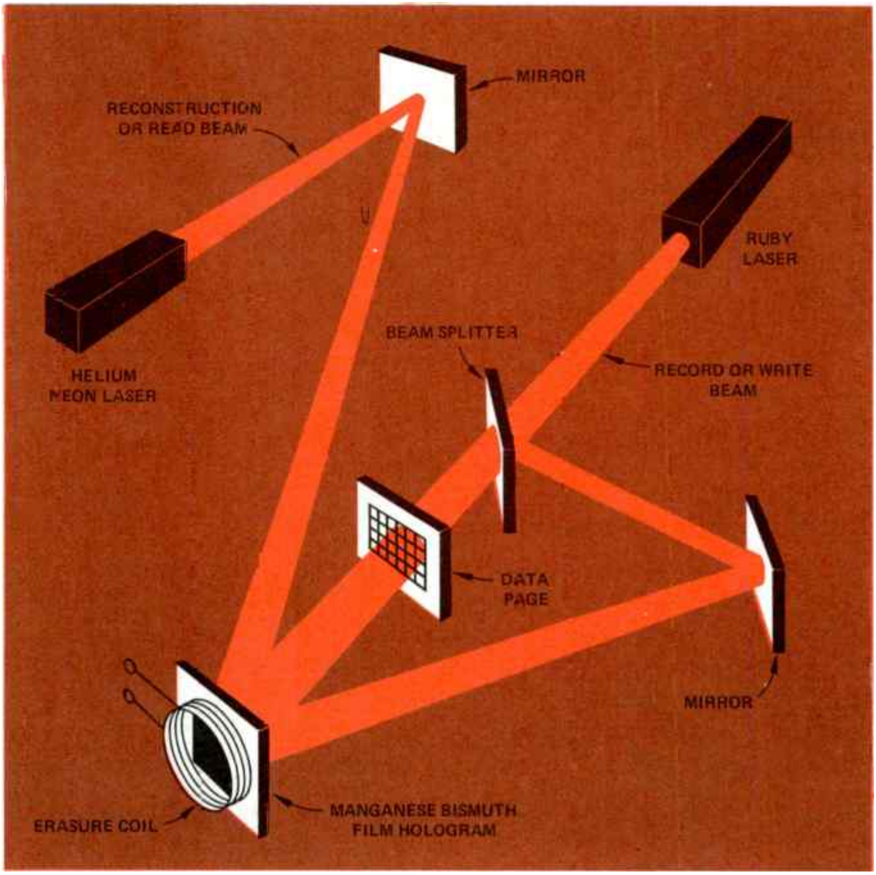


Figure 6. The simplified system for writing, reading, and erasing the magnetic hologram memory.

Until now, optical memories have been confined to the static, read-only type. This means, once made and inserted into the memory, the hologram can be read whenever necessary. This information cannot, however, be changed. Whenever a change is required, the old hologram has to be removed and a new hologram made and inserted into the memory.

A dynamic, read-write memory is now in the laboratory stage. These optical memories are capable of being written in the memory, stored, read at will, and then erased when new information must be entered. Ideally, these can be reused indefinitely.

These new holographic memories are made by depositing a single crystal layer of manganese bismuth (MnBi) on a base of mica. A strong magnetic field, applied to this film, causes the magnetic moments of the atoms to line up making it ready for use as a hologram. During the hologram recording process, the magnetic moments of the atoms move out of alignment in accordance with the laser light interference pattern. The recording time for such a hologram is about 10 nanoseconds. The hologram is erased by simply applying a voltage to a nearby coil, creating a strong magnetic field perpendicular to the hologram surface. This magnetic field realigns the magnetic moments of the MnBi crystal (Figure 6). The hologram memory has now forgotten the old information and is ready to record the new data.

Advertising

Reflection holograms have opened the door to hologram use in advertising. Not many people have lasers handy for reconstructing a transmission hologram image, but most people have a white light source, such as a flashlight or high intensity lamp, to illuminate a reflection hologram. Presently, the cost of producing hologram copies is rather high. Holograms, nevertheless, are beginning to appear in the print media.

Advertising holograms seem best suited for inclusion in magazines, although books and newspapers can also carry them. As for direct mail – mailing a hologram is no more difficult than mailing a photograph.

Although many uses have already been made of holograms, there are probably others that have not been explored. Advertising with holograms will provide public exposure to holograms – the more exposure, the more possible applications.

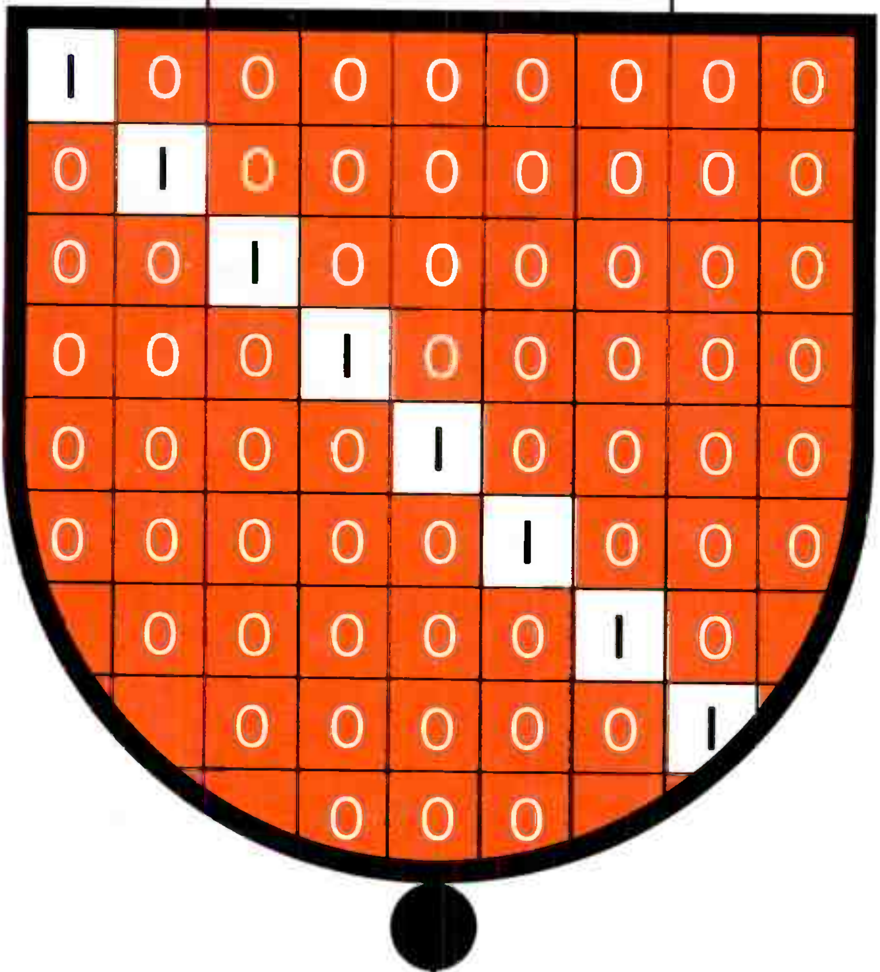
Commercial Applications

Holography is no longer just a laboratory curiosity, but now has applications in the fields of nondestructive testing, microscopic investigation, entertainment, information storage, and advertising. In fact, in the last two years, more than 500 companies in the United States alone, have made economic gains by taking advantage of this extraordinary form of photograhy, which permits the viewer to see around obstacles. Dr. Dennis Gabor, the inventor of holography, has forecast that by 1976 his brainchild will become a billion-dollar industry.

The Lenkurt®

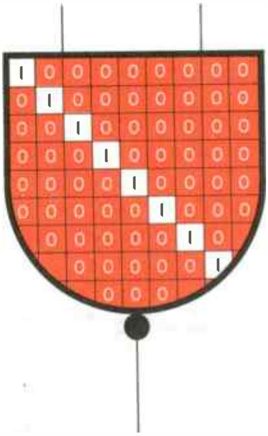
DEMODULATOR

DECEMBER 1969



BINARY LOGIC AND PCM

1101



Aristotle, in 330 B.C., to explain his philosophies, developed a logic system dealing with statements that were either true or false. In 1847, George Boole reduced Aristotle's logic to a mathematical shorthand that has become a universal logic language.

Binary logic is a way of thinking that can be applied to the design of *any* system where the "inputs" and "outputs" are just on-off actions. The invention of transistors led to the development of a series of logic modules capable of performing basic binary logic functions in electronic systems.

The complex PCM (pulse-code modulation) system can be broken down into subsystems whose inputs and outputs are simply on-off actions. This subsystem equipment is then designed using the principles of binary logic and implemented with corresponding logic modules.

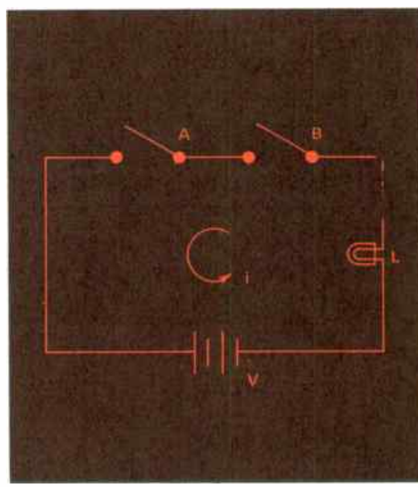


Figure 1.

Logic Modules

Basic logic modules are called AND gates, OR gates, and INVERTERS. These modules can be combined to obtain NAND and NOR gates. Logic building blocks called flip-flops can be made from these gates.

AND, OR, and INVERTER

Consider the circuit with two switches (A and B) connected in series, a voltage supply (V), and a light bulb (L) shown in Figure 1.

The light will be on, if, and only if, switch A *and* switch B are closed. The logic AND gate gets its name from this simple circuit analogy.

The "switching" circuit described above can be implemented with relays or diodes as well as with switches. All these circuits are cumbersome for the logic designer, so shorthand logic symbols have been developed. The logic symbol for the AND gate is shown in Figure 2.

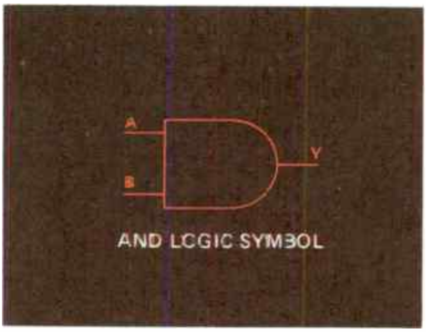
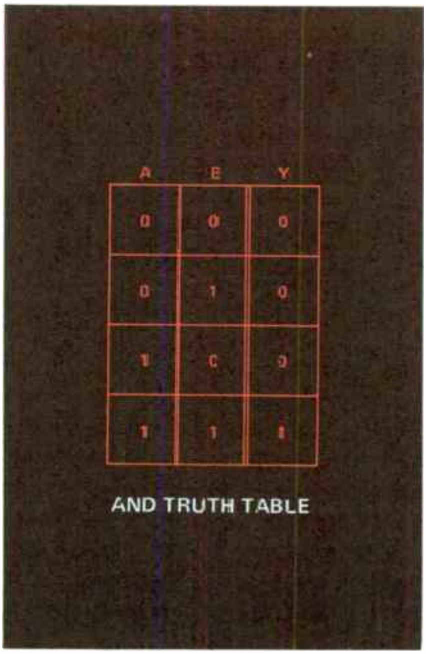


Figure 2.

Using a truth table and representing an "on" condition as a "1", an "off" condition as a "0", and the AND gate output as "Y", the combination of states for an AND gate is graphically displayed (Figure 3).



A	B	Y
0	0	0
0	1	0
1	0	0
1	1	1

AND TRUTH TABLE

Figure 3.

Using the "switching" circuit analogy again, put the two switches in parallel rather than in series (Figure 4).

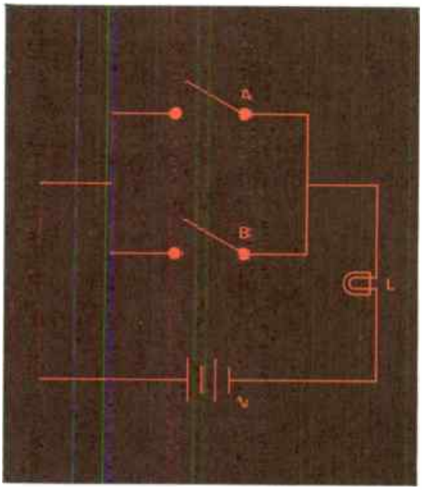
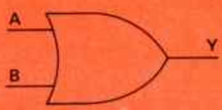


Figure 4.

With this arrangement, the current flows in the circuit and the bulb is on, if switch A *or* switch B *or* both are closed. The logic OR gate gets its name from this type of an arrangement. Figure 5 shows the logic symbol and truth table for an OR gate.

Logic functions can be implemented with diodes for electronic applications. But, diodes have two weaknesses. First, the output of diode AND and OR gates is attenuated. Second, diode gates are passive elements and unable to drive a network of gates.

Common emitter transistors have the ability to amplify a signal. By putting a transistor at the output of the diode AND and OR gate circuitry, the attenuated signal is restored to its original level. Because transistors are

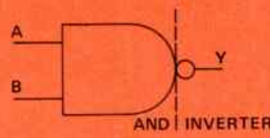


OR LOGIC SYMBOL

A	B	Y
0	0	0
0	1	1
1	0	1
1	1	1

OR TRUTH TABLE

Figure 5.



NAND LOGIC SYMBOL

A	B	Y
0	0	1
0	1	1
1	0	1
1	1	0

NAND TRUTH TABLE

Figure 6.

active devices, they are capable of driving a network of logic functions.

As well as solving the inherent problems of diode gates, transistors perform the logic function of inversion. Regardless of the input signal state, the transistor output will be inverted (a "1" becomes a "0" and vice versa). Transistors are known therefore, as INVERTERS.

NAND and NOR

The combination of a logic AND gate and a transistor INVERTER is called a logic NAND gate (for NOT–AND) (Figure 6).

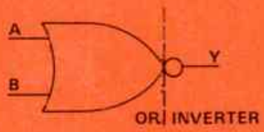
The output from a NAND gate will be negative, if, and only if, A and B are both positive.

A logic NOR gate is the combination of a logic Or and an INVERTER (for NOT–OR) (Figure 7).

If, and only if, both the NOR inputs are negative, the NOR output will be positive.

Integrated circuit technology has made NAND and NOR gates less expensive than the use of discrete components to construct NOT–AND and NOT–OR circuits.

For simplicity, the inputs to the logic gates have been limited to two,



NOR LOGIC SYMBOL

A	B	Y
0	0	1
0	1	0
1	0	0
1	1	0

NOR TRUTH TABLE

Figure 7.

A	B	C	Y
0	0	0	1
0	0	1	1
0	1	0	1
0	1	1	1
1	0	0	1
1	0	1	1
1	1	0	1
1	1	1	0

NAND TRUTH TABLE
FOR THREE INPUTS

Figure 8.

but in practice, the gates can have more than two. The same logic rules prevail. For a NAND gate, the output will be negative, if, and only if, all the inputs are positive. Similarly, for a NOR gate, the output will be positive, if, and only if, all the inputs are negative. The truth table for a NAND gate with three inputs is shown in Figure 8.

Flip-Flops

One of the most common circuit building blocks formed from groups of logic gates is a flip-flop – widely used for storing a single bit of information. The popularity of flip-flops is due to the following factors:

1. They are available in integrated circuits or can be built from readily available discrete components.
2. They are fast acting – can be made to change states in as little as a few nanoseconds (depending upon the propagation delay of the logic family).
3. They are active devices.

Truth tables rather than circuitry will be used to explain flip-flops. The designer is interested more in what happens to his signal, than how it happens. Having selected his logic modules from the same family (compatible power requirements, etc.), the designer works with a "black box" and its corresponding truth table. The flip-flops discussed are from the 930 DTL (Diode Transistor Logic) family used in the Lenkurt 91A PCM system.

The basic flip-flop is made of two NAND gates. It has two inputs (R and S) which determine what state ("0" or "1") the flip-flop will assume next, and two outputs (Q and $\bar{Q}$) which determine the flip-flop's present state. The Q and $\bar{Q}$ outputs from any flip-flop are always opposite states; if Q is "1", $\bar{Q}$ is "0" and vice versa.

The inputs and resulting outputs for an R–S flip-flop are shown in the truth table for a particular bit time (Figure 9). The output is a function of the flip-flop's outputs and its inputs at the previous bit time.

INPUTS AT BIT TIME t_n		OUTPUTS AT BIT TIME $t_{(n+1)}$	
R	S	Q	$\bar{Q}$
1	0	1	0
0	1	0	1
0	0	?	?
1	1	Q_n	$\bar{Q}_n$

R-S TRUTH TABLE

Figure 9.

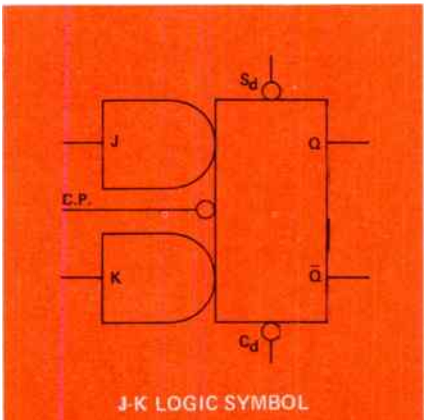
If the R input is "1" and the S input is "0", the Q output will be "1". If the input states are reversed, the output states will also be reversed. If the input states are both "1", the output will be unchanged from what it was at the previous bit time. The "?" in the truth table indicates the output is undesirable, and therefore to be avoided, when the input states are simultaneously "0".

Although it is possible to design a circuit such that the input states are never simultaneously "0", it is also possible to use a J–K flip-flop which tolerates all possible input combinations (Figure 10).

Regardless of what the output was, it will change to the opposite state, when both inputs are "1". The output will be unchanged, if both J and K are "0".

The J–K flip-flop is essentially two R–S flip-flops in series. The J–K inputs affect the flip-flop only when synchronized with a clock pulse – a steady stream of signals used to allow the input voltages to reach their final value. The direct set and clear inputs (S_d and C_d), on the other hand, operate directly on the output without being synchronized with the clock pulse. The first R–S flip-flop reacts at time "1" as shown in Figure 11; the second R–S flip-flop at time "2"; while the direct set or clear can react at anytime.

If a "0" is applied at C_d, the J–K flip-flop is placed in the clear state (Q=0). If a "0" is applied at S_d, the J–K flip-flop is placed in the set state (Q=1). The S_d and C_d inputs dominate the output even if synchronized with the J–K inputs.



INPUTS AT BIT TIME t_n J	K	OUTPUTS AT BIT TIME $t_{(n+1)}$ Q	$\bar{Q}$
1	0	1	0
0	1	0	1
0	0	Q_n	$\bar{Q}_n$
1	1	$\bar{Q}_n$	Q_n

J-K TRUTH TABLE

Figure 10.

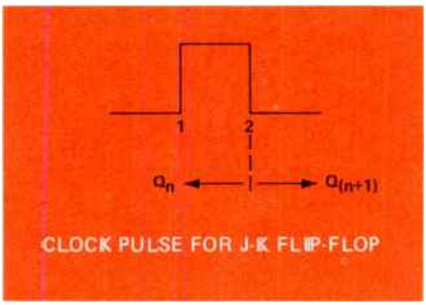


Figure 11.

Logic Modules for PCM Sampling

The sampler at both transmitting and receiving terminals in Lenkurt's 91A PCM system is basically a shift counter. This counter is the heart of the electronic mechanism that sequentially opens and closes the sampling gates for each channel – thereby multiplexing or demultiplexing the signals.

The number of stages in the shift counter is half the number of channels to be sampled; therefore, a 12 stage shift counter is needed to sample 24 channels.

Such a counter can assume 2^{12} = 4096 binary states. Only 24 states are required for sampling – the other 4072 states are undesirable and must be suppressed. The counter is operating properly when these undesirable states have been eliminated and the desired mode 1 operation (Figure 13) is sequentially sending out 24 separate pulses to the gates of 24 separate channels.

Mode 1 operation is accomplished by connecting 12 J–K flip-flops in series – one for each stage (Figure 14).

These flip-flops are driven by a clock pulse. With each clock pulse, the flip-flop state is shifted one stage to the right – the state of stage I at time t_1 will be the state of stage II at time t_2; etc. At stage XII, the Q output is fed back to the K input of stage I and $\bar{Q}$ is fed back to J of stage I. This "crossover" of output to input causes the state to reverse.

In a shift counter, the same state is shifted from one stage to the next with each clock pulse, reversing state when shifting from stage XII to stage I. Mode 1 fits this definition; there-

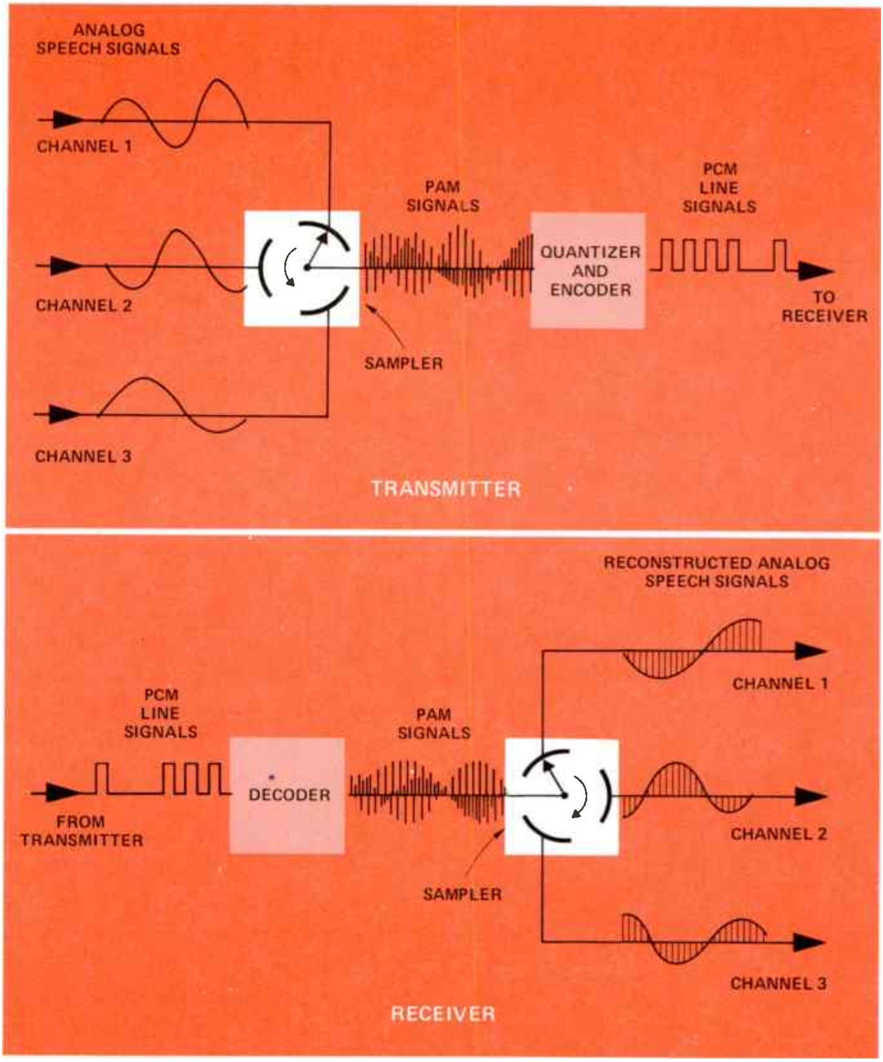


Figure 12. Simplified PCM system (only 3 of the 91A's 24 channels are shown).

fore, once the register assumes a mode 1 pattern, the counter will begin to cycle within this mode.

Applying power to the counter, the register may contain any one of the 4096 possible binary states (110001110000, for example). It is necessary to add either gate "X" *or* gate "Y", to suppress the undesirable states in the shift counter (Figure 14).

Gate "X" is actuated when both stage I and stage XII are in state "1", setting all the internal stages (II – XI) to "1". On the following clock pulse,

BIT TIMES	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII
1	0	0	0	0	0	0	0	0	0	0	0	0
2	1	0	0	0	0	0	0	0	0	0	0	0
3	1	1	0	0	0	0	0	0	0	0	0	0
4	1	1	1	0	0	0	0	0	0	0	0	0
5	1	1	1	1	0	0	0	0	0	0	0	0
6	1	1	1	1	1	0	0	0	0	0	0	0
7	1	1	1	1	1	1	0	0	0	0	0	0
8	1	1	1	1	1	1	1	0	0	0	0	0
9	1	1	1	1	1	1	1	1	0	0	0	0
10	1	1	1	1	1	1	1	1	1	0	0	0
11	1	1	1	1	1	1	1	1	1	1	0	0
12	1	1	1	1	1	1	1	1	1	1	1	0
13	1	1	1	1	1	1	1	1	1	1	1	1
14	0	1	1	1	1	1	1	1	1	1	1	1
15	0	0	1	1	1	1	1	1	1	1	1	1
16	0	0	0	1	1	1	1	1	1	1	1	1
17	0	0	0	0	1	1	1	1	1	1	1	1
18	0	0	0	0	0	1	1	1	1	1	1	1
19	0	0	0	0	0	0	1	1	1	1	1	1
20	0	0	0	0	0	0	0	1	1	1	1	1
21	0	0	0	0	0	0	0	0	1	1	1	1
22	0	0	0	0	0	0	0	0	0	1	1	1
23	0	0	0	0	0	0	0	0	0	0	1	1
24	0	0	0	0	0	0	0	0	0	0	0	1

Q OUTPUTS FOR STAGES

Figure 13. The 24 possible register patterns for mode 1.

stage I goes to "0" and all other stages are "1" as required for mode 1 operation. Figure 15 shows a sequence of patterns which starts with an arbitrary display when the power is applied and continues until the display matches mode 1.

If gate "Y" is used instead of gate "X", all the internal stages (II – XI) are set to "0" when stages I and XII are both "0".

For each of the 24 desirable states of the shift counter there is a readout gate made up of a two-input NAND

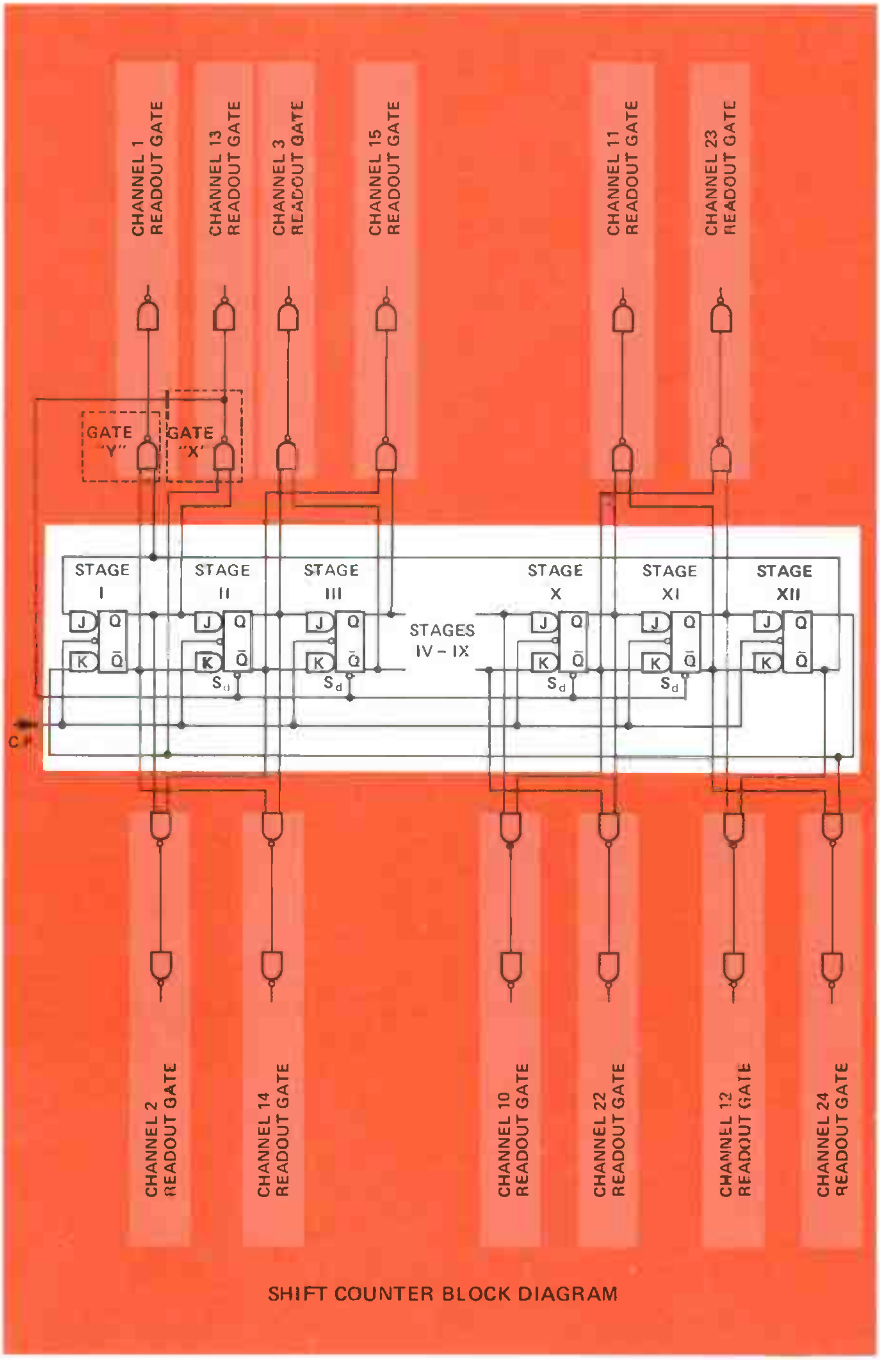
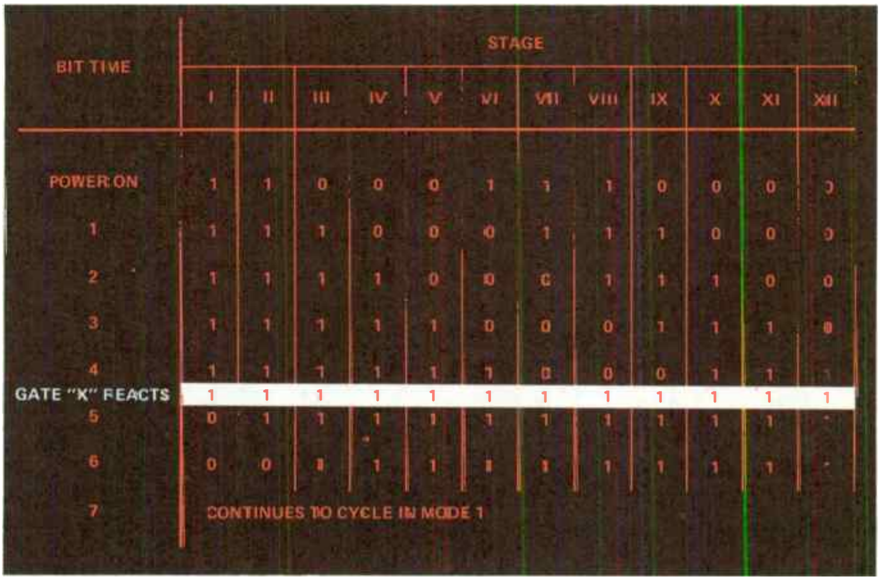

CHANNEL 1 READOUT GATE
CHANNEL 13 READOUT GATE
CHANNEL 3 READOUT GATE
CHANNEL 15 READOUT GATE
CHANNEL 11 READOUT GATE
CHANNEL 23 READOUT GATE
GATE "Y"
GATE "X"
STAGE I
STAGE II
STAGE III
STAGES IV – IX
STAGE X
STAGE XI
STAGE XII
J
K
Q
$\bar{Q}$
S_d
C+
CHANNEL 2 READOUT GATE
CHANNEL 14 READOUT GATE
CHANNEL 10 READOUT GATE
CHANNEL 22 READOUT GATE
CHANNEL 12 READOUT GATE
CHANNEL 24 READOUT GATE
SHIFT COUNTER BLOCK DIAGRAM

Figure 14.



BIT TIME	STAGE											
	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII
POWER ON	1	1	0	0	0	1	1	1	0	0	0	0
1	1	1	1	0	0	0	1	1	1	0	0	0
2	1	1	1	1	0	0	0	1	1	1	0	0
3	1	1	1	1	1	0	0	0	1	1	1	0
4	1	1	1	1	1	1	0	0	0	1	1	1
GATE "X" REACTS	1	1	1	1	1	1	1	1	1	1	1	1
5	0	1	1	1	1	1	1	1	1	1	1	[illegible]
6	0	0	1	1	1	1	1	1	1	1	1	[illegible]
7	CONTINUES TO CYCLE IN MODE 1											

Figure 15.

gate followed by a single input NAND gate. (Figure 14)

The register state can be determined by knowing where there is a transition from "0" to "1" or vice versa, or by knowing that there is no transition (all "0's" or all "1's"). By comparing the outputs of adjacent stages (white areas in Figure 13), the register transition points can be determined. This comparison is done with the readout gate. For each bit time only one readout gate will be in state "1" – indicating the position of the transition point or the lack of any transition.

When a readout gate is in state "1", it opens the corresponding channel sampling gate. The cycling of the 24 readout gates for the shift counter successively opens and closes the sampling gates of each of the 24 multiplexed channels in the 91A PCM system.

Figure 14 shows that the "X" and "Y" gates used for mode suppression of the counter are required for readout – allowing the mode suppression without additional logic modules.

Binary logic and the 930 DTL modules are also utilized in the quantizing and coding equipment for Lenkurt's 91A PCM system.

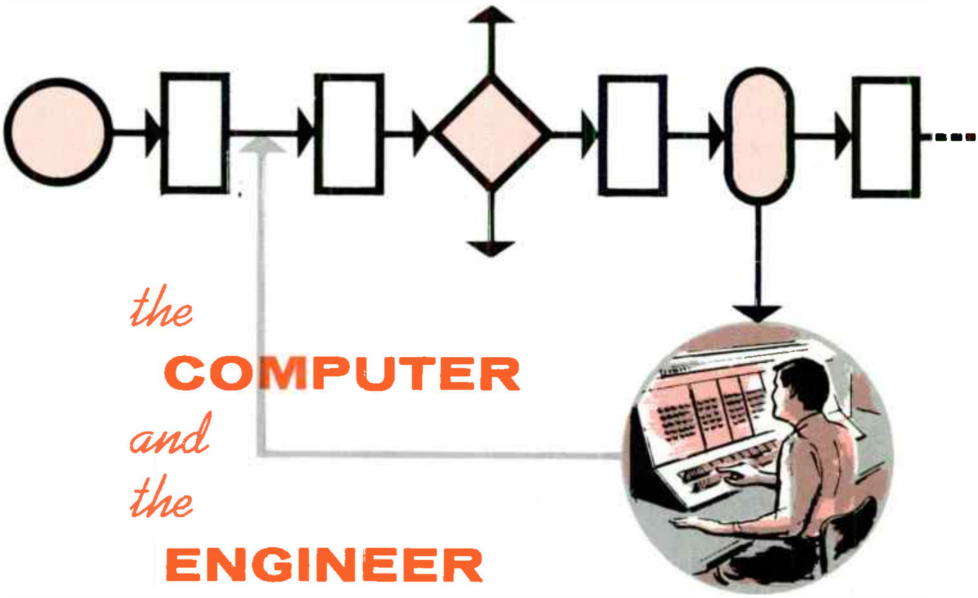
Framework for Expansion

PCM has achieved its present state in the communications industry because of efficient application of binary logic and the timely development of reliable, low cost logic modules.

Binary logic provides the framework for PCM's future expansion.

JUNE, 1966
VOL. 15, NO. 6

The Lenkurt DEMODULATOR



the COMPUTER and the ENGINEER

The computer has become a most valuable and necessary piece of equipment in the design laboratory of today's electronics firm. Through accuracy and speed in complicated mathematical calculations, the computer has made possible design techniques never before feasible.

This article concerns the scientific digital computer and its uses by the electronics engineer.

In a relatively short life, computers have grown out of desk-top calculating machines, relying on a complicated assortment of gears and mechanical linkage to perform their task, to electronic giants capable of thousands of mathematical calculations a second. Computers are called on for mammoth data processing duties, for controlling automated machines and, of prime interest to the electronics design engineer, to carry out complicated mathematical computations with a speed and accuracy never before possible.

Computer Types

Computers are of two basic types, analog and digital. There is very little similarity between the two, and they serve markedly different functions. The analog computer is used primarily as an analysis device, where the physical properties of a system—mechanical, electrical, or whatever—can be represented by analogous circuits in the computer. With amplifiers, potentiometers, and circuits performing such mathematical functions as differentiation and integration, the computer imitates the real item. Once an electrical model is devised, analyzing the system is particularly suited to the analog computer.

In operation, the system is tested under varying conditions by changing voltages in the computer, with all calculations carried out simultaneously. Results are displayed immediately on one of a number of readout devices: strip-chart recorders, X-Y plotters, or even oscilloscopes. The complexity of the model has little to do with the computer's problem-solving speed.

Digital computation, a more time consuming step-by-step process, is much more accurate, and is limited basically only by the preciseness of input data and the ability of the machine to store and use significant figures. Digital computers, with which this article is primarily concerned, are basically manipulators of stored information.

The computer (Figure 1) consists of a central processing unit, where the actual computation is done, a storage unit providing the machine with an electronic memory, and input and output devices with which the user can communicate with the machine.

The facts of a problem are held by the machine's memory — commonly a large grouping of magnetic cores which can retain thousands of single electrical pulses for later use. Instructions to the computer, called a *program,* are written by the operator to guide the machine in processing and changing the stored pieces of information in a prescribed way. In data processing operations, huge quantities of information are handled, facts sorted, relatively easy computa-

Figure 1. Typical scientific computer is the IBM 1620 II, used at Lenkurt. Behind the central processing unit is magnetic disc storage drive.

tions done, and the information reassembled in meaningful form — a printed page, a stack of punched cards, payroll checks, or perhaps utility statements.

Scientific Use

Scientific computers serve an entirely different function. The computer first sees the scientific or engineering problem as a small set of numerical facts; the answer will probably be no more than a short series of digits. But inside the computer an enormous amount of work has gone on. Mathematical computations which might have taken weeks — if possible at all — had they been done manually by the engineer, are now obtained from the computer in minutes.

The digital computer does all its work by breaking down complex mathematical functions into many simplified arithmetic operations, much as the mechanical desk calculator does multiplication by a series of additions. With the computer, difficult mathematical problems can be reduced to processes of addition, subtraction, and logical decisions, performed thousands of times per second.

The theoretical knowledge necessary to construct a computer had been available for many years, but it was not until reliable electronic components freed the computer from the limitations of the mechanical calculating machine that the "brain" began to grow.

The first computers were merely arithmetical manipulators able to do a sequence of calculations without further human control. They were complex, could handle huge quantities of data, and did it with speed never before imagined possible. But they were tied to limited specific processes, much like a platoon of girls operating desk calculators with instructions to carry out a single set of arithmetic operations and list them on a form.

Logic Added

An historic hurdle was overcome when logic functions were added — the computer now had the ability to make decisions. This is synonymous to having the desk calculator operators stop at some point in their work, analyze the figures, and proceed on one of a number of alternate paths. In its simplest form, logical response is to answer "yes" or "no" to a certain stimulus; that is, to take one of two possible paths. Judging the presence or absence of some variable condition, and making a decision based on this, is the function of the elementary logic circuit of the computer. In this way the modern computer makes decisions which will determine some later course of action.

Beyond the ability to make decisions comes the faculty of *learning* from experience. Some computers can analyze their own method of approaching a problem and improve on it the next time the same problem is presented.

The availability of electronic computers in the early 1950's for use by design engineers greatly relieved the tedious routines followed until that time. One of the common — and laborious — problems faced by telecommunications equipment designers is the creation of new filter networks. (About 90 percent of *all* networks are filters, defined as circuits to separate wanted signals from groups of signals.) Well-developed methods are available for making accurate mathematical models of network circuits, involving extensive use of complicated calculations with

32	16	8	4	2	1	NUMBER
0	0	0	0	0	1	1
0	0	0	0	1	0	2
0	0	0	0	1	1	3
0	0	0	1	0	1	5
0	0	1	0	1	0	10
1	0	1	0	1	1	43

Figure 2. Sample of binary number equivalents used in computers. Values increase in multiples of two, right to left.

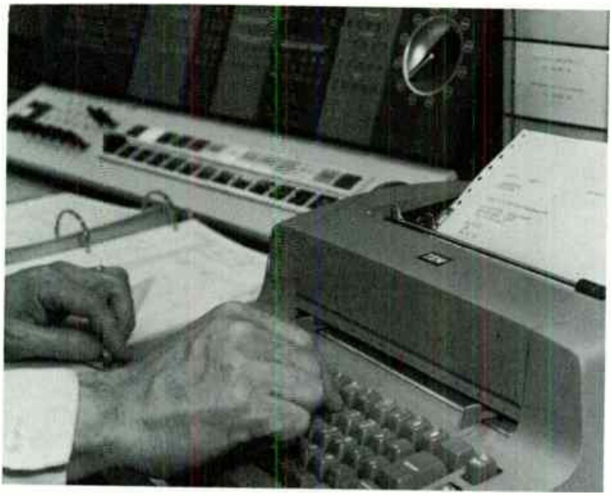


Figure 3. Electric typewriter allows engineer to supply data to computer as design problem progresses. Results of computation also appear on the typewriter.

many repetitive functions. This suits the digital computer perfectly.

Binary System

The digital computer, as opposed to the analog machine, deals with discrete numbers. Just as the mechanical calculator relies on the exact positioning of gears for computation, the computer positions binary pulses in a precise method, storing them in its electronic memory until they are to be used.

The binary number system in computer machine language differs from conventional decimal numbers only in concept. In customary numerical expression using the decimal (or more correctly the "denary") system, a number such as 837 is merely a convenient method of listing the number of units, tens, hundreds, etc. In each position there is a possibility of 10 different integers, 0 through 9. The number 43 really means $(4 \times 10^1) + (3 \times 10^0)$.

The binary system is based on the possibility of only two choices in each position, written as either 1 or 0. Examples are given in Figure 2. The number 43 is written 101011, or $(1 \times 2^5) + (0 \times 2^4) + (1 \times 2^3) + (0 \times 2^2) + (1 \times 2^1) + (1 \times 2^0)$. While this may seem a cumbersome notation, it is very appropriate for digital computers because it corresponds with the natural two-state ability of many electronic components—relays, magnetic core memory units, and flip-flop circuits.

Computer Languages

To communicate with the computer, the user must either *talk* in the basic machine language or, if he prefers, provide a means of translation in some other language. In modern computers, several programming languages are available to the user, depending on the specific purpose at hand. Special "compiler" programs already in the computer perform the translation. In this way the programming languages are *user ori-*

ented, rather than machine oriented, and allow the programmer freedom of expression and thought in terms more familiar to him.

A language developed primarily for scientific work is FORTRAN (FORmula TRANslation) which bears a close resemblance to the native tongue of the engineer and scientist — mathematics. A similar language more popular in Europe is ALGOL (ALGebraically Oriented Language). Add to these COBOL (COmmon Business Oriented Language), PL/1 (Programming Language/1), and a growing number of languages for specific disciplines.

The facts of the scientific problem, along with the program to be used to reach a solution, are fed to the computer from one of a number of input devices — magnetic tape, magnetic discs, punched cards, perforated tape, and the typewriter. The computer stores this information in its electronic memory for future use. Certain facts may be supplied directly by the engineer, operating an electric typewriter at the computer console (Figure 3). Answers also will appear on the typewriter when the problem is completed.

When a program is first written for a problem, a flow diagram is constructed expressing the logic of the approach. A non-mathematical example in Figure 4 illustrates decision making, computer style.

Iteration

The example also points to the repetitive ability of the digital computer, one of the best used qualities of the computer in scientific calculation. By the mathematical process of *iteration* (or

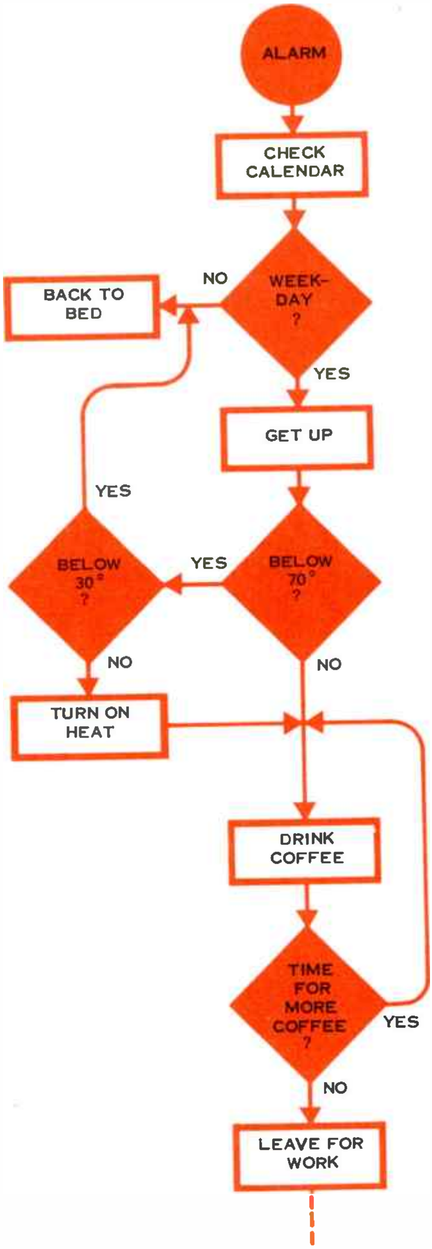


Figure 4. Simulated flow diagram illustrates logic used in planning computer program.

repetition) the computer can begin with a guess at the answer and then check it. If necessary, the computer will modify and repeat the original calculation. With each iteration, the machine answer approaches the correct value sought.

One very common example of an iteration, known as the Newton method, is illustrated graphically in Figure 5. The problem is to find the special value of a, the point where the curve crosses the x axis. The point also corresponds to $y = 0$. The computer makes a first guess x_1 for the value of a and then computes y_1. Since y_1 is not zero, the computer seeks a better approximation of a. It computes the slope (or tangent) of the curve at (x_1, y_1), then calculates where this straight line intersects the x axis. This value, x_2, is taken as a better approximation of a. The sequence is repeated until the value of a is known as accurately as the capabilities of the machine will allow. The iteration is then stopped.

Another example of iteration is the solution of the quadratic equation

$$x^2 + ax - b = 0$$

This can be rearranged to give

$$x(x + a) - b = 0$$

$$x = \frac{b}{x + a}, \text{ and finally}$$

$$x_{n+1} = \frac{b}{x_n + a}.$$

With $n = 0$, a guess for the value of x is entered and the calculation is carried out for the right side of the equation. If the answer is equal to x, our guess has been accurate. If not, the first answer is substituted for x_1, and the calculation

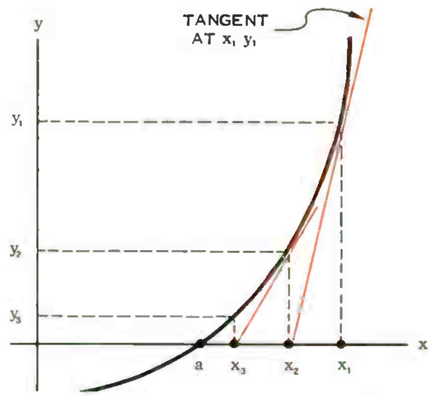


Figure 5. Computer uses iteration to find the point where curve crosses x axis.

is carried out again — approaching the true value of x at each iteration. Given the equation

$$x^2 + 6.3x - 14.8 = 0,$$

the iteration approach would proceed in this way:

$$\text{Let } x_0 = 0$$

$$x_1 = \frac{14.8}{0 + 6.3} = 2.349$$

$$x_2 = \frac{14.8}{2.349 + 6.3} = 1.711$$

$$x_3 = \frac{14.8}{1.711 + 6.3} = 1.847$$

$$x_4 = \frac{14.8}{1.847 + 6.3} = 1.817$$

$$x_5 = \frac{14.8}{1.817 + 6.3} = 1.823$$

$$x_6 = \frac{14.8}{1.823 + 6.3} = 1.822$$

$$x_7 = \frac{14.8}{1.822 + 6.3} = 1.822$$

The iteration, in this case, was stopped when four significant figures in the answer stabilized. This process, by its nature, will converge on the smaller of the two roots of the equation. The second root is found immediately by dividing the final term in the equation by the first root:

$$\frac{-14.8}{1.822} = -8.122$$

Monte Carlo Method

Another important mathematical practice appropriate to the iterative ability of the computer deals with problems of probability, and is called the Monte Carlo method. In the manufacture of electronics equipment, component values vary randomly within the specified tolerance range assigned. With the help of the computer it is possible to calculate what percentage of finished products will be within a nominal performance range. Using the Monte Carlo method, the computer generates values for the circuit components in a random manner, much the way a factory worker would select actual parts during assembly. Each time different component values are selected, the circuit performance is checked by the computer. From this, a distribution curve can be plotted, indicating the probable number of acceptable units. If the percentage is too low, tolerances must be tightened for components most likely to affect the circuit, and the process is repeated. In this way the design engineer can use his mathematical model of the circuit to advantage, predicting with great accuracy the performance of the finished units leaving the assembly line.

Filter Design

A typical filter network design problem will illustrate the procedure used by the engineer. A standard network and its loss-frequency curve are shown in Figure 6. The designer will work with a number of computer programs, each intended to help him solve particular problems and approach a practical end product. The first program used will aid him in selecting the optimum parameters defining the performance of the filter. After feeding the program into the computer's memory, the engineer enters tentative values for certain design characteristics, such as passband edge frequencies (f_1, f_2), the frequencies at the infinite loss points ($f_{1\infty}$, $f_{2\infty}$, $f_{3\infty}$, $f_{4\infty}$) and the passband ripple. Then, one by one, the designer supplies the computer with sample frequencies in the stopbands and passbands. The computer calculates the loss at each of these points from which an accurate loss curve can be plotted.

When satisfied with his basic design, the engineer must determine the components to be used. A second program is placed in the computer's memory. Using what is called *insertion-loss theory,* the engineer and the computer can now generate the component values needed to satisfy the original design characteristics.

The results of this second operation are used in a general analysis program to prove the design under practical conditions (till now the components were considered to have pure capacitance and inductance). The computer then produces values for total loss, phase shift, envelope delay, reflection coefficient and input impedance.

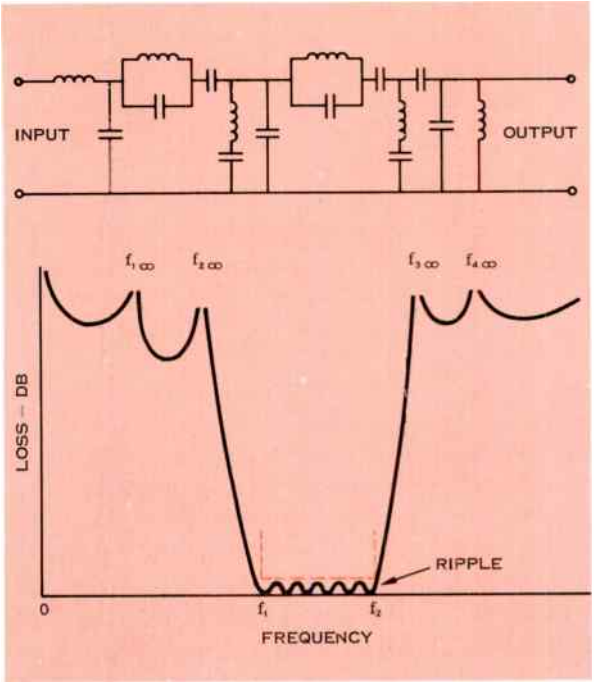


Figure 6. Filter circuit (top) will produce this type of loss-frequency curve. Engineer and computer determine optimum component values to provide proper passband frequencies, acceptable ripple, and infinite loss points.

A fourth computer step is used to examine tolerances, using the Monte Carlo method previously described. When all the design programs have been completed, a prototype model can be built, and the filter circuit is well on its way to becoming part of a new product.

Prior to the time when engineers had the work-horse capabilities of the computer at their disposal, a method known as *image-parameter* theory was used in filter design. Insertion-loss theory was just not practicable due to the vast amount of calculation necessary. But with the computer a slightly superior product can be produced using the insertion-loss technique, with the added advantage of the method's simplicity and straightforward approach.

Time Saving

Approximate times for performing the various steps may be listed, considering computers now found in most engineering facilities, like the IBM 1620II used at Lenkurt, but understanding that more modern and likewise faster machines are being produced constantly. The first program, used to select the basic performance parameters for a fairly complicated filter, takes the engineer and the computer together about 30 minutes. Component values are produced from the second program in about 10 minutes. The general analysis program takes an additional 10 to 15 minutes. Longest because of its many repetitive steps is the Monte Carlo tolerance analysis, running about 60 minutes. Total time: less than two hours.

The image-parameter method is not only slower by two to ten times depending on the complexity of the filter, but lacks the complete information obtained by the insertion-loss technique. In image-parameter design it is still necessary to construct a "breadboard" model to check dissipation characteristics and tolerances — all done on the computer with insertion-loss methods.

Other Uses

In addition to the design of networks, a great deal of time on the computer is spent analyzing one-time mathematical problems that do not lend themselves to generalized programming. Again, time-saving benefits of the computer are invaluable to the design engineer. As long as mathematics is a part of electronics design, computers will be a tool for the design engineer. And even today's methods, superior by many orders of magnitude to the labors of only a few years ago, will be supplanted as more sophisticated techniques are created, and man and machine learn better how to work together.

Probably one of the most promising developments in the last year is graphical on-line simulation of electronic circuits. Using a "light pen" the designer can draw his circuit on the face of a computer-coupled cathode ray tube. With proper commands to the machine's memory, he can analyze circuit functions under specific parameters. Using the light pen, almost instantaneous modifications are possible, allowing the engineer to redesign the circuit and see the results while still at the computer.

The same device is used in other disciplines to analyze the structure of bridges, buildings, and even three-dimensional mathematical representations — viewed from side, top, or bottom.

Time Sharing

There is a trend to centralize computer facilities for use by a number of subscribers on a time-sharing basis. Several large computer centers have already been established around the country, with users connected remotely through telephone communications facilities. Time-sharing computers can allow two-dozen or more users on the line at one time, efficiently sandwiching their programs together with such speed that the user seldom, if ever, notices a delay. The user gains the advantages of large computer capability, while only paying for the computing time he needs.

The scientific computer, modern-day design tool of the electronics engineer, will improve in its usefulness as quickly as man's inventiveness will allow. Along the way, the computer is not only absorbing a tremendous work burden, but is making possible increased quality and better performance in electronics equipment.

HERTZ?

In an effort to reach worldwide understanding in published scientific and technical work, standardization of terms is essential. The Institute of Electrical and Electronics Engineers (IEEE) standards committee recently accepted a number of new standards of electrical units and symbols, established in close cooperation with many international organizations. At the beginning of the year the *Demodulator* adopted the IEEE recommendations.

One of the most noticeable changes is the adoption of the name *hertz* as a unit of frequency. While *cycles per second* is still widely used, and technically correct, *hertz* is now preferred in that it is more understandable in all languages.

In general, symbols for units are written in lowercase letters, except when the name of the unit is derived from a proper name. Thus, it is *hertz,* but *Hz* when abbreviated; *decibel,* but *dB.*

Compound prefixes are to be avoided: $\mu\mu$F (micromicrofarad) becomes pF (picofarad); mμs is now ns (nanosecond); and the familiar kmc (kilomegacycle) should be GHz (gigahertz).

Prefixes for Metric Units

10^{12}	tera	T
10^{9}	giga	G
10^{6}	mega	M
10^{3}	kilo	k
10^{2}	hecto	h
10	deka	da
10^{-1}	deci	d
10^{-2}	centi	c
10^{-3}	milli	m
10^{-6}	micro	μ
10^{-9}	nano	n
10^{-12}	pico	p
10^{-15}	femto	f
10^{-18}	atto	a

Examples of Usage

decibel	dB
gigahertz	GHz
hertz (cycles/second)	Hz
kilohertz	kHz
megahertz	MHz
picofarad	pF
siemens (mho)	S ($1\ S = 1/\Omega$)

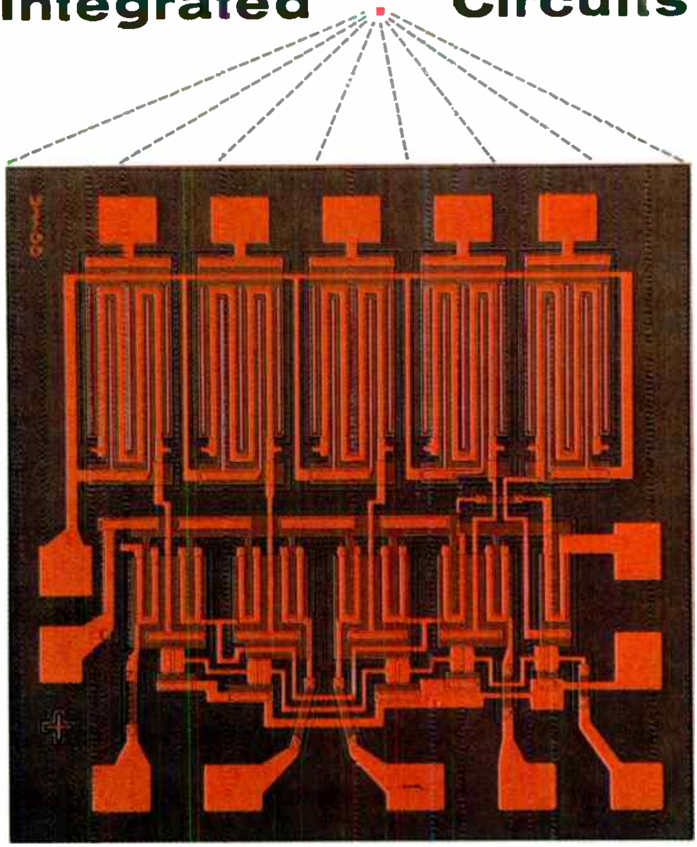
Readers may consult the IEEE Standard Symbols for Units (No. 260, January 1965) and other IEEE documents for more complete details.

SECTION V
SOLID-STATE TECHNOLOGY

DECEMEER, 1967
VOL. 16, No. 12

The Lenkurt DEMODULATOR

Integrated Circuits



Integrated circuits promise increased reliability and lower cost in addition to minute size.

Microelectronics, until very recently, was mainly concerned with packaging discrete components in as small a space as possible. First electron tubes and then transistors were packed tightly with diodes, resistors, coils, and capacitors to make miniaturized circuits. With vacuum tube technology, engineers were proud to get something like 6000 components into a cubic foot of space. Transistors replaced tubes, and upped the packing density to about 100,000 parts per cubic foot. And now, while it may seem like a lot of room, 10-million components might fit into that same cubic foot through the use of integrated circuits.

While all of this space saving is impressive, it is not size alone that makes integrated circuits attractive—in fact imperative—to the future of electronics design. The real savings come in manufacturing cost and reliability.

Reliability vs. Interconnection

In all microelectronics the basic motivation is to control complexity. More sophisticated needs bring more complex devices and gigantic growth in the number of parts. So great is this increase in individual items in the machines of todays electronics world that design engineers are faced with what they term the "tyranny of numbers".

The problem is one of reducing the number of components that must be individually manufactured, tested, interconnected, packaged and finally retested. Circuit reliability is inversely related to the number of individual devices and the necessary interconnection.

Early electron tube and transistor attempts at microelectronics solved many of the space problems, but did not substantially improve reliability. The number of interconnections remained the same. Integrated circuits, by their very nature, go to the root of the interconnection/reliability relationship.

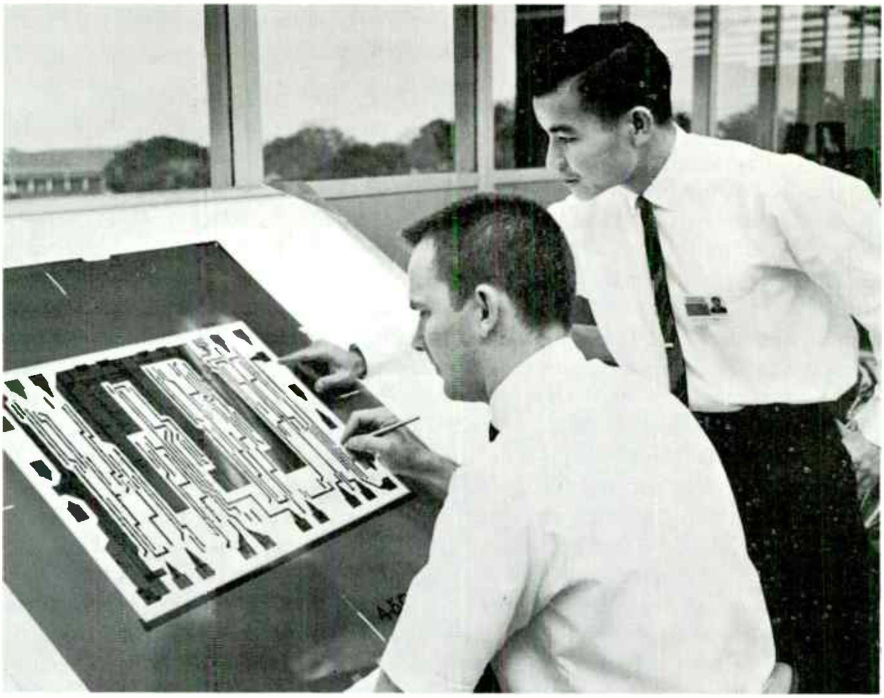
An integrated circuit contains a number of inseparably associated active and passive devices. In a monolithic integrated circuit, all of the circuit parts are fabricated within a single block of material; hybrid circuits include some discrete active devices attached to the integrated circuit components such as resistors and capacitors.

A Circuit on a Chip

Whatever the type of IC, connections, as such, are not used but are "built in" to the device itself. Instead of making a number of separate components and then joining them together to form functioning circuits, the monolithic IC approach is to construct all components at the same time on one "chip".

No new physical principles are involved in integrated circuits, but rather the innovation of mass fabrication techniques. Silicon integrated circuits are really just an extension of transistor technology. But with IC's, hundreds or thousands of circuits, each involving many transistors and other components, are produced at one time.

The creation of a monolithic integrated circuit begins with the basic circuit design. A breadboard model may even be constructed with discrete components to test circuit operation. Then



Courtesy Fairchild Semiconductor

Figure 1. Original art work for an integrated circuit must be prepared in exacting detail. Reduced, it will control fabricating process.

the IC equivalent is drawn representing all of the components and necessary interconnections in the circuit (Figure 1). A separate piece of art work, usually about 30 by 30 inches, is made to represent each of the many steps in the manufacturing process.

The circuit is reduced photographically about 500 times, and then is stepped and exposed repeatedly across a small glass plate. The image on the plate or mask can later be transferred to the surface of a silicon wafer for the fabrication of integrated circuits. In this way as many as 1500 identical circuits may be placed on one wafer—all produced at the same time, all with the same characteristics (Figure 2).

From the finished wafer each circuit or chip is separated, connecting wires are attached, and the IC is packaged in a convenient form.

Manufacturing Process

The raw material from which IC's are fabricated begins as a large ingot of carefully grown silicon. It is between 1 and 2 inches in diameter and about a foot long. The silicon is "doped" with very small but accurately controlled quantities of impurities which change the electrical properties of the material (the basis of all semiconductors).

Glass-like and very brittle, the ingot is cut with a diamond carborundum saw into wafers 12 mils (0.012 in.) thick. After additional processing and polishing, the finished wafer is no thicker than 6 mils.

This wafer will serve as a mechanical

base or substrate for future operations. Onto this substrate another layer of silicon is added by a process called epitaxial growth. The epitaxial layer, about 2 mils thick, has the same crystal structure as the substrate, but is different electrically because of different doping.

Now a series of steps is taken to carefully change the properties of the epitaxial layer in selectively masked areas (Figure 3). In this way the circuit, first drawn in an area almost a yard square in order to obtain good dimensional accuracy and now only slightly larger than a period, can be transferred to the wafer.

The wafer is first exposed to oxygen at high temperature, resulting in a layer of silicon dioxide, called the passivating layer. This layer is coated with a photosensitive resin and with the first mask in place is exposed to light. The exposed areas become hardened, but those under the mask can be rinsed clean of resin. Next the wafer is subjected to an acid which etches away the silicon dioxide layer not protected by the hardened resin.

Courtesy Sylvania Electric Products

Figure 2. Integrated circuit wafer is compared to half dollar. Each square is an entire circuit.

Diffusion of Impurities

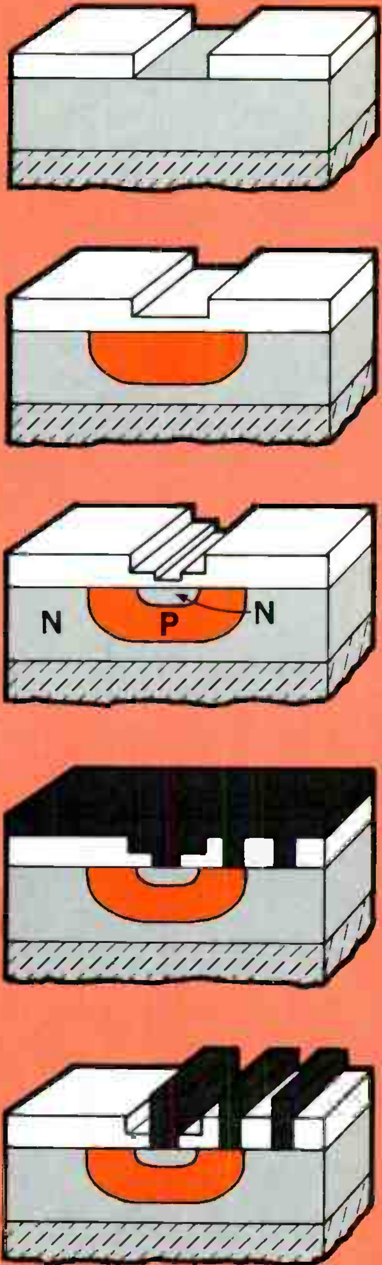
The result of the photo etching is a window through which selective parts of the chip may be exposed to a diffusion process. Diffusion is accomplished by "soaking" the wafer in a furnace with an impurity-rich atmosphere. These impurities will diffuse into the silicon only in the areas not protected by the silicon dioxide layer.

The routine is repeated over and over—photo resistive coating, masking, exposure, rinsing, etching and diffusion. With each step another function is added—a transistor emitter or collector, a resistor or capacitor.

The type of semiconductor produced is determined by controlling the types and amounts of impurities introduced to the silicon. One form of impurity will result in an excess of conduction electrons. This is called N-type silicon. Another impurity will leave a deficiency of conduction electrons, resulting in P-type silicon.

To create a diode, a junction is formed between P-type and N-type material—a PN junction. A transistor requires two junctions, and may be PNP or NPN.

While some values of resistance and capacitance can be diffused into the epitaxial layer, in many cases it is more practical and exacting to add these functions *on top* of the finished active devices. This may be done by a thin or thick film process. Minute quantities of metal are laid down through sputtering or evaporation for thin films or by a silk screen process for thick films. A very thin and narrow piece of metal makes an accurate resistor and can even be trimmed later to touch up the value for more critical tolerances. Capacitors likewise can be made with thin film. It is possible to use the silicon substrate



Window etched through silicon dioxide to expose epitaxial layer to diffusion.

Impurity is diffused into silicon. In the process, new oxide layer forms over entire surface.

Repeated etching and diffusion create other parts of device. NPN areas now form a transistor.

New windows are cut for placement of leads. Surface is covered with metal, typically aluminum.

Metal is etched away leaving leads to each part of device.

Figure 3. Simplified illustration of integrated circuit manufacture shows steps of etching, diffusion and metalization.

as one plate, and the thin-film layer as the other.

Economy of Duplication

All of these processes take place at the same time on *all* circuits on the wafer. It is here that the reduced cost of manufacturing is realized. What can be done to one circuit can be done to hundreds at the same time.

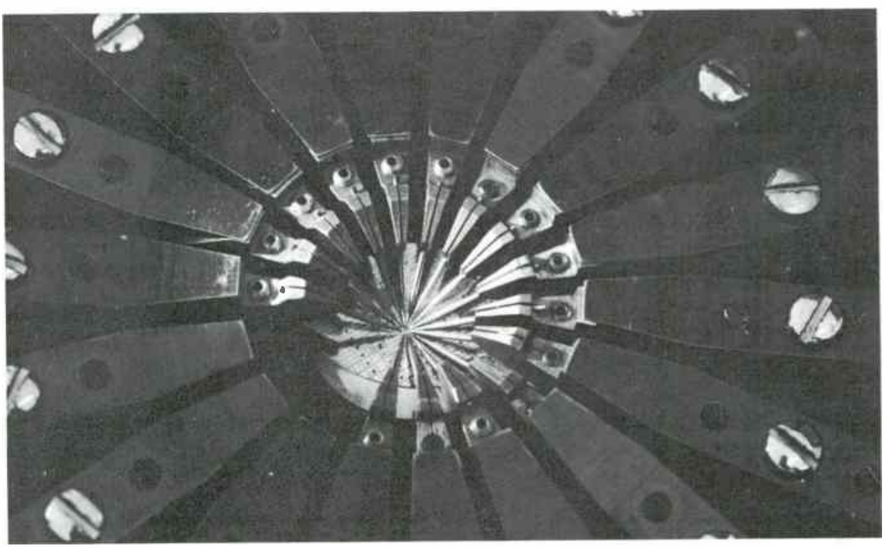
But not every circuit or chip on a wafer is going to be perfect. Seemingly minor defects caused by dust particles and other imperfections can be fatal to such small units. Manufacturing techniques are constantly being improved, but still the yield from a wafer is not 100 percent. And the percentage goes down with more complex circuits.

After all diffusion processes have been completed, but while the wafer is still intact, each circuit is automatically tested. The wafer is stepped through a computer-controlled device where probes determine whether each circuit is acceptable or not (Figure 4). If the circuit does not come up to specification, it is marked, fished out later and destroyed.

The chips, measuring about 0.04 inches on a side, are then separated using a diamond scribing point. They are now ready for packaging—one of the most expensive steps on the assembly line. Until now all manufacturing processes were carried out at the same time for all circuits on a wafer. In fact, many wafers are treated together as a group. At the packaging stage circuits must receive individual manual attention for the first time.

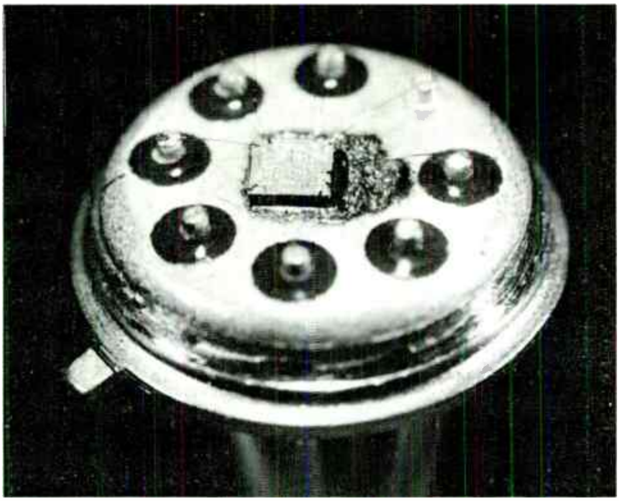
Packaging Methods

Before the chip can be sealed in one of the more than 250 types of packages, a means of connecting it to other component parts must be accomplished. One technique uses the bonding of tiny gold wires to the connection "pads" on the chip (Figure 5). These wires are



Courtesy Fairchild Semiconductor

Figure 4. Computer-controlled probes reach down to test individual circuits while still on the wafer. Imperfections in crystal structure, dust, and failure to meet mechanical and electrical tests all affect the yield from a wafer.



Courtesy Fairchild Semiconductor

Figure 5. Small gold wires connect IC to terminals of standard TO-5 package.

then attached to larger terminals for easy access.

Another scheme — generally called the "flip-chip" method — involves the building of small humps of metal on the pads. The chip can then be turned face down on a larger substrate or frame and bonded by soldering, thermocompression or ultrasonics (Figure 6). This process lends itself to automation and physical durability.

A method of building up heavier metal leads coming directly from the circuit is also being developed. These "beam leads" make external contact easier and also add structural strength to the entire device.

Once the IC has been packaged it must again be tested. Special tests are made for frequency response in linear devices and switching speeds for digital circuits. They are also subjected to a variety of mechanical tests, shock, vibration, acceleration and temperature changes to ensure dependability.

Operating Speed

When fast switching times are required—always a consideration in digital equipment—an interesting dependency on size makes integrated circuits valuable.

The ultimate factor limiting the operating speed of any electronic device is the velocity of electromagnetic propagation (the speed of light). In space this rate is approximately 186,000 miles per second, or about one foot in 10^{-9} seconds (one nanosecond). Electric current flowing through a conductor will have a speed somewhat slower than this, depending on the characteristics of the conductor.

If a switching circuit is to operate in the nanosecond range, the distance between circuit components must be measured in fractions of an inch. Integrated circuits make this possible.

Typical applications of the digital integrated circuit include many associated with computers: flip-flops, adders, gates, buffers, and memory cells.

The linear integrated circuit, as the name implies, is applicable to most amplifying chores and may be the building block for anything from a simple audio amplifier to a complex communications network. Linear IC's are found in analog computers, communications equipment and even hi-fi systems.

Again, it is not so much the major breakthrough in technology that is seen in the final equipment, as it is the ingenuity of the designer effectively using available tools. For example, Lenkurt engineers are currently investigating ways of using digital integrated circuits to perform analog functions—especially desirable for economic reasons as well as reliability.

A typical monolithic integrated circuit contains a number of transistors and associated passive components interconnected to provide a functional circuit. After packaging, the integrated circuit becomes a unit of some larger design—interconnected to perform the functions of an entire system or piece of equipment. And so numbers of integrated circuits are stacked in smaller and smaller spaces just as tube and transistor circuits were before them.

(IC's)^2

The next logical jump in technology is to combine a number of circuits onto a single chip, with the same advantage of reduced interconnections and manufacturing ease. Several such devices, using a technique known as medium or large scale integration (LSI), are now on the market (Figure 7).

Large scale integration is particularly applicable to computer and other digital technology, where the same type logic circuit may be used hundreds or thousands of times. LSI is capable of offering 100 logic gates on a single chip. This tends to reduce manufacturing and packaging costs. However, increasing the number of components on a chip results in decreasing cost only to a point. Beyond a certain level, circuit complexity tends to reduce the yield of usable chips on a wafer—a chip is only as good as any of its parts—and cost per component begins to rise.

LSI is having a definite effect on the circuit designer who has always had

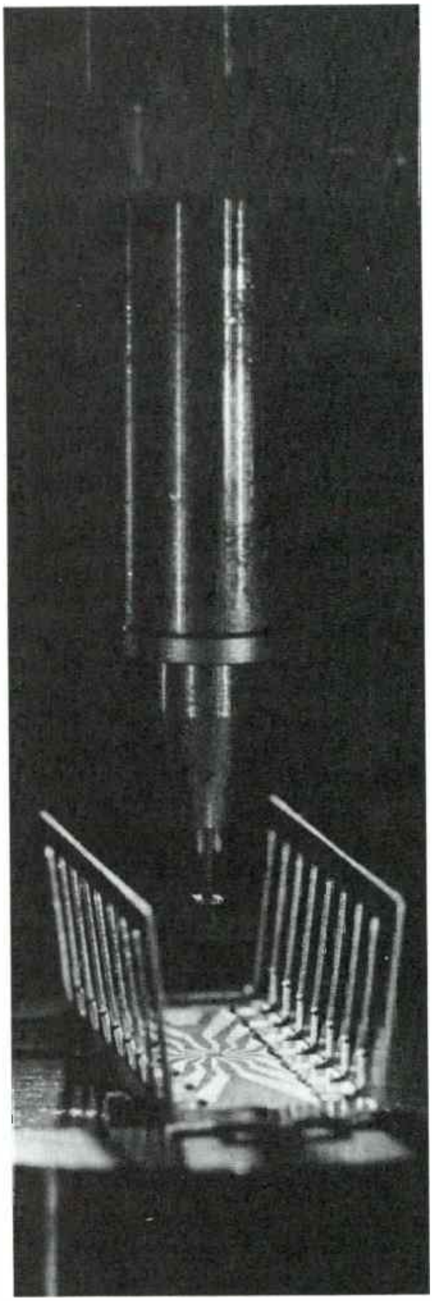


Figure 6. Recently developed technique shows Fairchild integrated circuit about to be mounted face down on frame. All contacts are made at one time.

freedom to choose components according to his own specifications. As entire circuits, and beyond that entire functions, become standardized in manufacture, the design engineer's ingenuity may have to be applied to the use of standard circuits instead of to the design of original units.

At least three types of LSI packages will probably be offered:

1. Stock items ready for use (mainly logic and memory units).
2. Basic units, such as gate circuits, which require final metalization. Interconnection will be provided according to the buyers specifications.
3. Custom units, where specifications demand original design.

LSI circuits at present are more appropriate for digital than linear systems because of the repetitive nature of the elements used. And because components on an LSI chip are closer together than discrete IC's, operating speeds of digital systems are further increased.

IC's and Communications

Miniaturization of electronic systems has been an obvious product of integrated circuits, even though it must be remembered that the more basic value of IC's is in cost and reliability. Com-

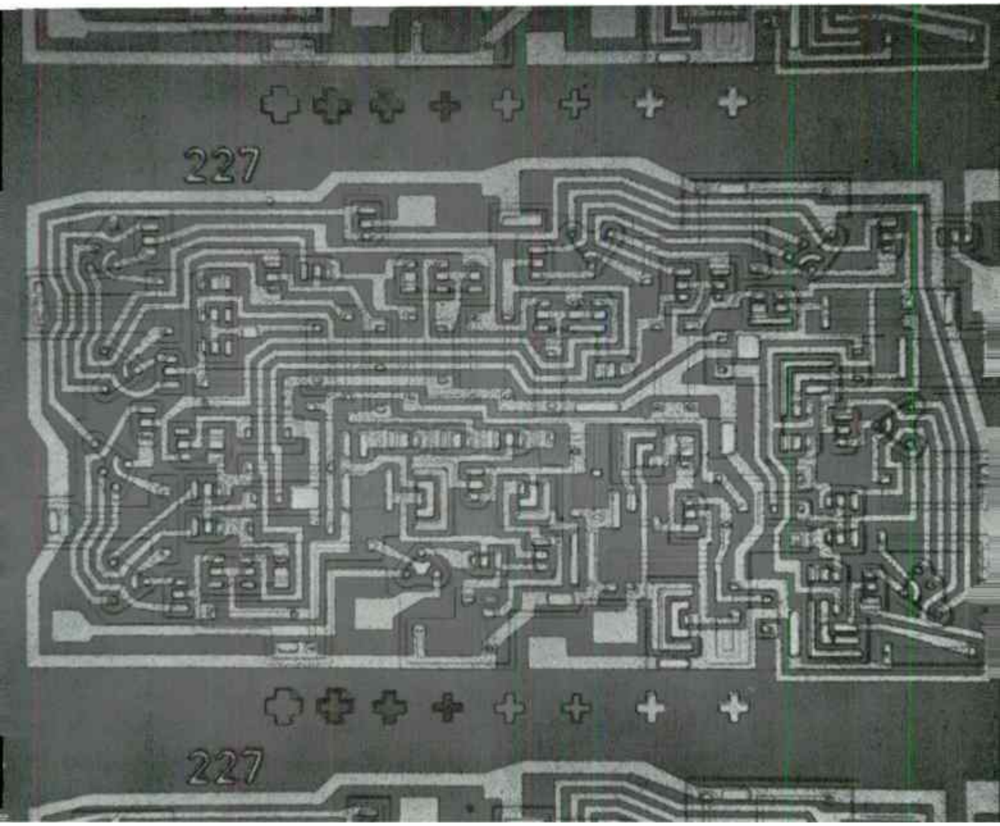
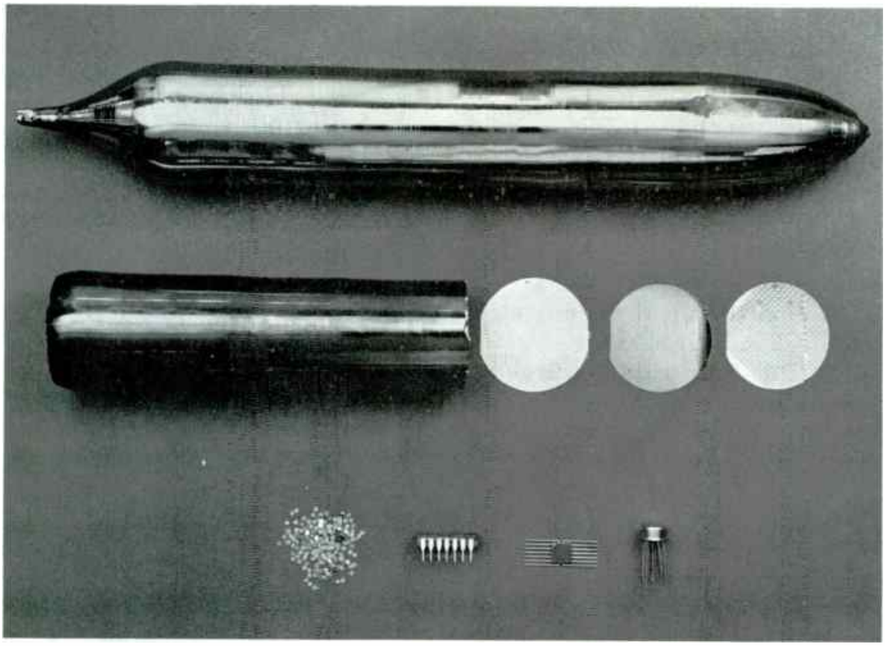


Figure 7. Sylvania IC contains 116 transistors, resistors and diodes to produce 40 gate circuits on each chip. Used in computers or digital communications equipment.



Courtesy Fairchild Semiconductor

Figure 8. Silicon ingot (top) is sliced, and then polished before integrated circuits are fabricated. From the finished wafer (middle, right) come hundreds of chips (bottom, left) ready for packaging. Shown are Fairchild's Dual In-Line, a flatpack, and a TO-5 can.

puters that would have filled rooms by previous technology now stand in the corner or even on a desk top. And the size advantage is now being applied effectively to the designs of new communications systems, especially as linear devices become more practical.

IC's, including transistors, diodes, resistors and capacitors, are so small that the manufacturer's assembly line is now characterized by rows of microscopes. But filter circuits—so important to the telecommunications industry—have remained a problem. These filters require very high quality inductors, large coils of wire that are heavy, bulky and relatively expensive.

A solution in constructing inductorless filters using integrated circuit technology has been proposed by Lenkurt engineers. While still in the experimental stage, the technique exhibits a great deal of promise for the future in keeping communications systems economical, reliable and small.

Simply stated, the method replaces each inductor in a complex filter network with a circuit that behaves electrically like an inductor. This circuit consists of one capacitor and a circuit called a gyrator. Low-sensitivity, high-Q circuits have been produced in a package roughly a quarter of an inch square. This compares in size to a coil of about 1.5 inches in diameter and 1 inch thick.

Data and PCM

The first area of telecommunications to be significantly affected by integrated circuits will be where digital techniques are used. Here, the extensive developmental effort placed in digital IC's for computers can be easily transferred to data transmission and its cousin pulse code modulation (PCM).

For data and for PCM, circuits are needed to deal strictly with digital pulses—signals that are either *on* or *off*. In PCM transmission, for example, binary digits are used to represent discrete values of a voice signal. In the first generation PCM system, bit rates of 1.5 Mb/s are used, requiring very fast sampling and switching. Digital integrated circuits are almost a must. And in future systems, with rates approaching 300 Mb/s, the need is even more pronounced.

Integrated circuits also provide a means of manufacturing extremely stable resistors as part of the miniature circuit. Because all resistors are made at the same time, and with their close proximity will be subjected to identical changes in environment, they are excellent components in critical communications applications.

Hybrids Sometimes Appropriate

Much of this discussion has centered around monolithic integrated circuits. Hundreds of circuits, including all active and passive components, are processed as a unit on one wafer. And the wafer is repeated hundreds and thousands of times. Obviously this is advantageous whenever great quantities of the same circuit are needed.

Many telecommunications equipment designs will use "off the shelf" IC's both of digital and analog variety. But some specific applications will not require the large quantities needed to justify monolithic integration.

In these cases hybrid IC's may be more appropriate. The equipment manufacturer could produce much of the circuit using in-plant thin or thick film capability. Other components—including active devices—would be added as separate units. Single transistors, or standard integrated circuits could be bonded to a unique circuit of the manufacturer's design. Bell Laboratories, for example, has used this approach for a tone-generating circuit in some Touch-Tone telephones.

Integrated circuits may also affect the design of future microwave radio equipment. With their fixed component relationships and close spacing, IC's may solve the problem of manufacturing circuits with uniform performance to operate at microwave frequencies.

The IC Promise

The entire electronics industry is facing a change of great magnitude—a physicist might call it a technological quantum jump. Integrated circuits as they stand promise price savings and reliability. Large scale integration will open roads to miniaturization still hard to visualize accurately. A hundred transistors on a chip, or a hundred-thousand on a wafer are foreseeable in the next five years.

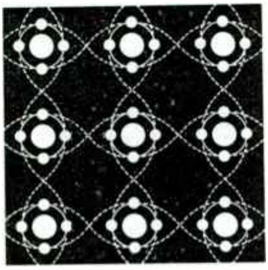
Whether it's a shoe-box size computer, a carrier system inside a telephone handset, or even wrist watch television—the integrated circuit is the gadget dreams are made of.

The Lenkurt

DECEMBER 1968

DEMODULATOR

solid state review



Today's profusion of novel electronic devices is but one result of the phenomenal growth in the solid-state art.

Every area of activity has its jargon and its catch-phrases. This is as true of industry as it is of politics or the arts. And, undoubtedly, the most widely heard phrase in electronics marketing today is "solid-state". The term spans the entire industry from consumer products to super-sophisticated communications systems – anything from a teen-ager's vestpocket radio to a phased-array radar acquisition system. The term is rapidly becoming a catch-all and losing its original meaning.

Originally, the term solid-state applied only to devices and circuits which were fabricated using principles of applied solid-state physics. New devices such as transistors and diodes were called solid-state to distinguish them from their gas-state or other non-solid-state predecessors.

Later, numbers of solid-state devices were incorporated into modules or subassemblies. These in turn were assembled into operating systems. The upshot of all this is the current practice of calling any apparatus that contains so much as a single transistor solid-state. We now have "solid-state" radios, TV receivers, electric shavers and even kitchen mixers. In fact, most people call a transistor radio simply a "transistor". Plainly then, with the help of advertising and marketing people, the term solid-state has taken on new applications and meanings.

Solid-State Physics

The term, solid-state, refers to the internal devices and circuits that make an apparatus function.

In general these devices belong to that unique electronic species known as semiconductors. As a rule, the materials found in electronics are either conductors or nonconductors. The latter are sometimes called insulators or dielectrics. All these terms are used to indicate a given material's ability to pass an electric current. In the case of conductors and insulators, the names are self-explanatory. But what about semiconductors?

Obviously, a material either passes electric current or it does not. It is difficult to imagine a material that only "half-passes" or *almost* conducts electricity. Nonetheless, this is the term we use to describe materials that, in their native state, will not allow current to pass. But when properly treated or "doped" with a foreign element they will carry current. Most often these elements are germanium or silicon doped with a small amount of arsenic, indium, or some other impurity.

"Solid-state" is not the only term in electronics which is widely misunderstood or loosely used. There is also considerable confusion surrounding the following:

Microcomponents refers to an assembly of very small, interconnected discrete components – active or passive – which forms an electronic circuit. Interconnection of the various leads is by soldering or welding. Microcomponents use no substrates.

Microcircuits, on the other hand, satisfy the same criteria as do microcomponents but they do employ substrates.

Active substrates are usually a working part of the electronic circuit which they support physically. They possess ohmic and electrical values.

Passive substrates perform no electronic function but provide physical support and a thermal sink for their circuits.

An **integrated circuit** is an electrical network – active or passive – composed of two or more circuit elements inextricably bound on a single semiconductor substrate.

Transistance, a function of transistors, is an electrical property that causes applied voltages to create amplification or accomplish switching.

Any element's capacity to allow the free flow of electrons depends directly on the atomic structure of the element. Elements such as copper or gold have rather loosely bound atoms in which some electrons are free to move. Consequently they make good conductors. Other materials like lead and air possess very tight atomic structures with no free electrons.

An atom consists of a nucleus with a number of surrounding electrons in concentric rings. In this respect, atoms can be thought of as planetary systems with the sun as nucleus and the planets as the circling electrons. It is the number of these electrons – specifically in the outer ring – which determines a material's willingness to allow the flow of electricity (Fig. 1).

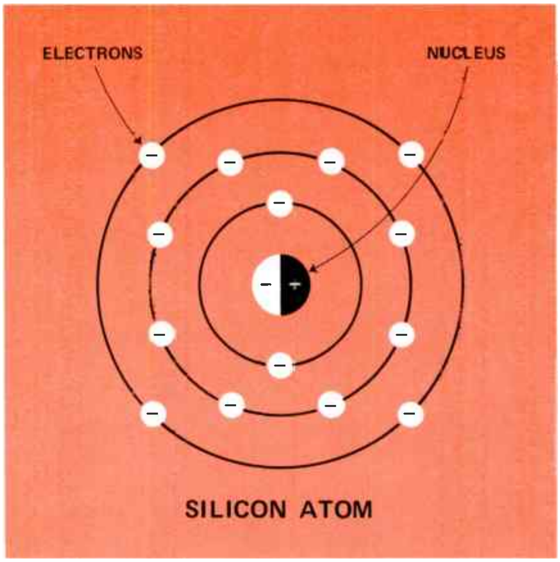
One of nature's phenomena is that eight electrons in the outer ring compose the firmest shell. These are the valence or free electrons. Elements or compounds possessing such strong atomic shells are not usually good conductors because they have no free electrons.

Semiconductor elements such as silicon and germanium have four valence electrons in their atoms. When two of these atoms combine, a shell of eight electrons forms a perfect covalent bond. In such materials, resistance is very high – almost to the point of insulation.

There are ways to reduce this resistance and make the materials act as conductors. One way is using heat to cause excitation among the electrons sufficient to break the covalent bonds and allow electron movement. The other is to dope the material with an impurity and cause the atoms to rearrange themselves.

It works like this. Silicon and germanium are crystals. Their crystalline structure results from the formation of perfect shells composed of eight electrons. This happens when any two silicon atoms – each containing four electrons – join together forming covalent bonds. The result is a very stable atomic structure, whose resistance is high.

Figure 1. Atoms have a maximum of two electrons in the first ring, or energy level, eight in the second, eighteen in the third and a definite number in each succeeding ring. Usually, the rings nearest the nucleus will possess the maximum number of electrons before more appear in succeeding rings. When the number of electrons in the outer ring is less than the maximum, they are called "free" or valence electrons.



Now, if a foreign substance whose atoms have an odd number of valence electrons – say 5 – is added to the silicon, four of them will pair with the four host electrons. The fifth, or odd electron will remain free to move about in the compound and carry electric current. See Figure 2.

In cases like this where the added impurity or donor provides the free electron, an N type, negatively charged, semiconductor is created. If, on the other hand, the crystal is doped with an impurity whose atoms contain only three electrons, a "hole" is created by the absence of a fourth impure electron to pair with the fourth silicon electron. This hole is contributed by the host or acceptor atom and may therefore be remembered as creating a P-type semiconductor whose charge is positive. In both cases, electron/hole activity provides the means for conduction. A working circuit can be made by joining an N type and P type together. This is called a junction, hence the terms junction diode and junction transistor.

Junctions

A diode is the result of joining two pieces of semiconductor material of opposite polarities. They consist of either PN or NP junctions. Add another element to the diode and it is a transistor – PNP or NPN. In either case, the PN or NP junctions function as emitter, collector or base in the transistor. Methods of construction and intended uses vary widely, but in any case, transistors consist of those three elements. In an NPN transistor, the N-type carriers are called the majority carriers, and the P-type, minority. Just the opposite is true in PNP devices.

In most cases, the names of various transistors are a result of the construction method used in their fabrication and the resulting shape. Some of the types are: grown-junction, alloyed junction, drift-type transistors, micro-alloy diffused-base transistors (MADT), mesa and planar transistors. Each of these performs differently from the others and consequently, application differs with each.

Mesa, and more recently, planar transistors are representative of the current level of sophistication in the solid-state art. They are characterized by higher voltage and current capacities and, in the case of planar types, higher operating frequencies. However, the current trend is away from mesa types.

New solid-state devices seem to be appearing almost daily. Some of the more noteworthy and better-known are grouped below in two broad categories – transistors and diodes.

Transistors

There are many different Field Effect Transistors (FET's) currently available – among them, JFET's (J for Junction), IGFET's (IG for Insulated Gate), and MOSFET's (MOS for Metal Oxide Semiconductor). FET's differ from regular junction transistors in their polarity and construction. Junction transistors are bipolar whereas FET's are unipolar. This means that they use only the majority carriers – holes and electrons – as a means of conduction. Junction transistors use both majority and minority carriers simultaneously.

In a JFET the interface between input and output (channel and gate) is a junction. The interface in an IGFET is insulated by a dielectric, and in a MOSFET the insulating material is an oxide.

They are all extremely fast digital-type transistors. Applications are anywhere that switching or any on-off function is required.

Diodes

Limited Space-charge Accumulation (LSA), Impact Avalanche Transist

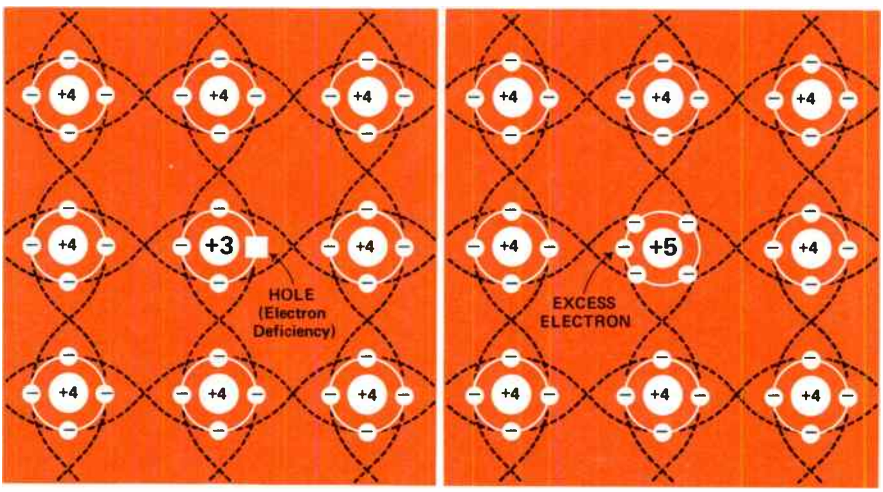


Figure 2. Depending upon the atomic structure of the impurity used, proper doping of semiconductor materials will create either "holes" or free electrons to act as conductors.

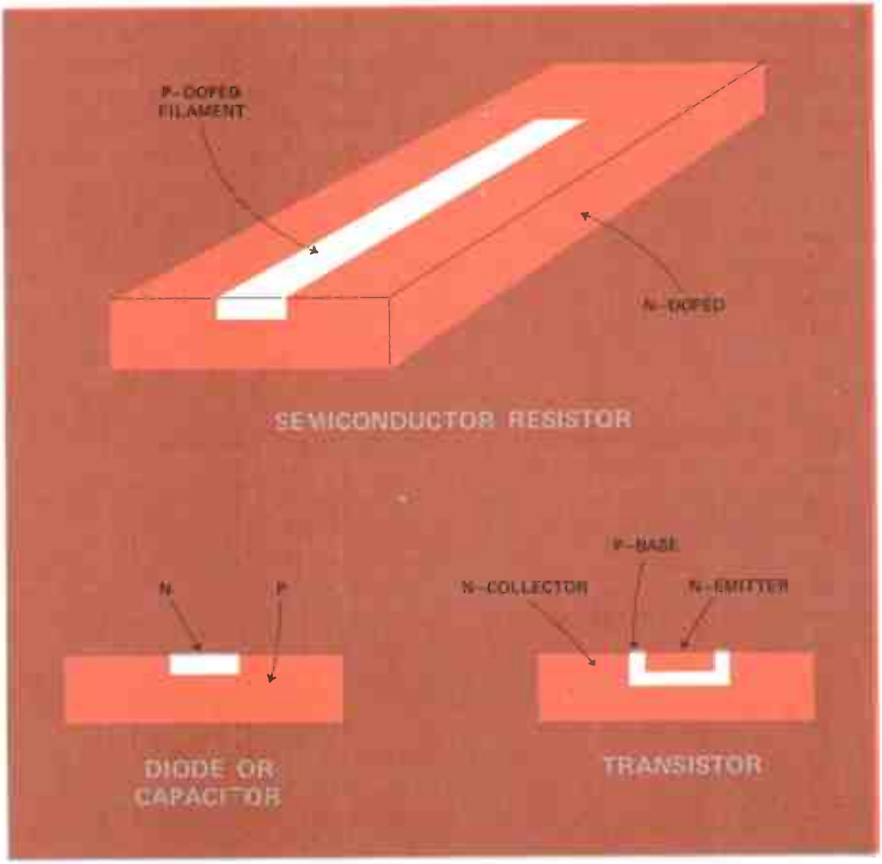


Figure 3. Semiconductor microelectronic devices

Time (IMPATT) and other avalanche-type diodes are semiconductor devices intended primarily for use as amplifiers and oscillators for microwave radio. This is also true of Gunn and other gallium-arsenide devices.

Schottky barrier diodes are metal semiconductor junctions having the property to allow current of only one polarity to pass. They are used in microwave mixers and frequency multipliers. Recently, Schottky devices have been improved by the addition of a silicon "collar" around the junction which enables the device to operate very near its theoretical efficiency limits. The diode itself is characterized by a molybdenum-gold Schottky contact.

IC's, A Logical Extension

Transistors, commonplace today, provided the foundation for all subsequent developments in component technology. Viewed from the vantage point of today's accumulated knowledge, they are basically simple devices. But without them, none of the devices discussed here would have been possible, nor, for that matter, would there be integrated circuits.

Now it seems a logical step to go from discrete semiconductor components to silicon-based integrated circuits containing these same components on a single chip.

Once it was established that an operational device could be fabricated from semiconductor material, it was only natural that some interest should be directed toward the idea that an entire circuit might be made the same way using the same materials.

Integrated circuit fabrication employs principles and techniques very similar to those used in the manufacture of discrete semiconductor devices. IC's are made from carefully grown and doped silicon. Within a single chip many individual components – transistors, diodes, and even resistors and diode-type capacitors – can be etched, diffused or otherwise fabricated. (For a detailed account of IC fabrication, see the December, 1967 *Lenkurt Demodulator*).

Recent developments in IC technology center around smaller, stabler and faster digital circuits and expanded applications for linear (analog) IC's. Linear circuits are now finding use in virtually every area of communications as well as in most solid-state consumer products.

Digital circuits find most of their application in data processing hardware where they are used as logic gates, memory cores and other functions peculiar to computer operations. They are also widely used in digital communications systems.

Linear circuits, on the other hand, have a much wider scope. They are used in every area of communications, telemetering, control and home entertainment systems. Linear IC's perform as amplifiers, oscillators and in scores of other more complex functions.

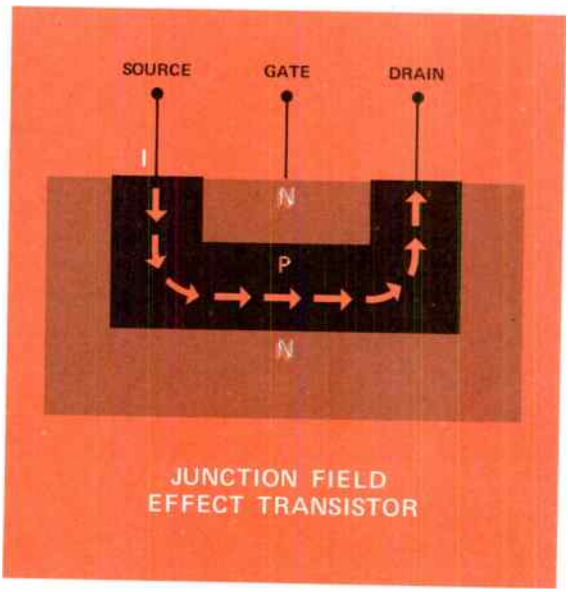


Figure 4. Current (I) flowing in the channel is controlled by varying the voltage across the gate. The applied voltage creates a field in the channel which has a "pinching" effect on the current. Thus, channel current can range from almost zero to full conduction. Since they are voltage controlled, unipolar FET's more closely resemble vacuum tubes than do their more common cousins which are bipolar and current controlled.

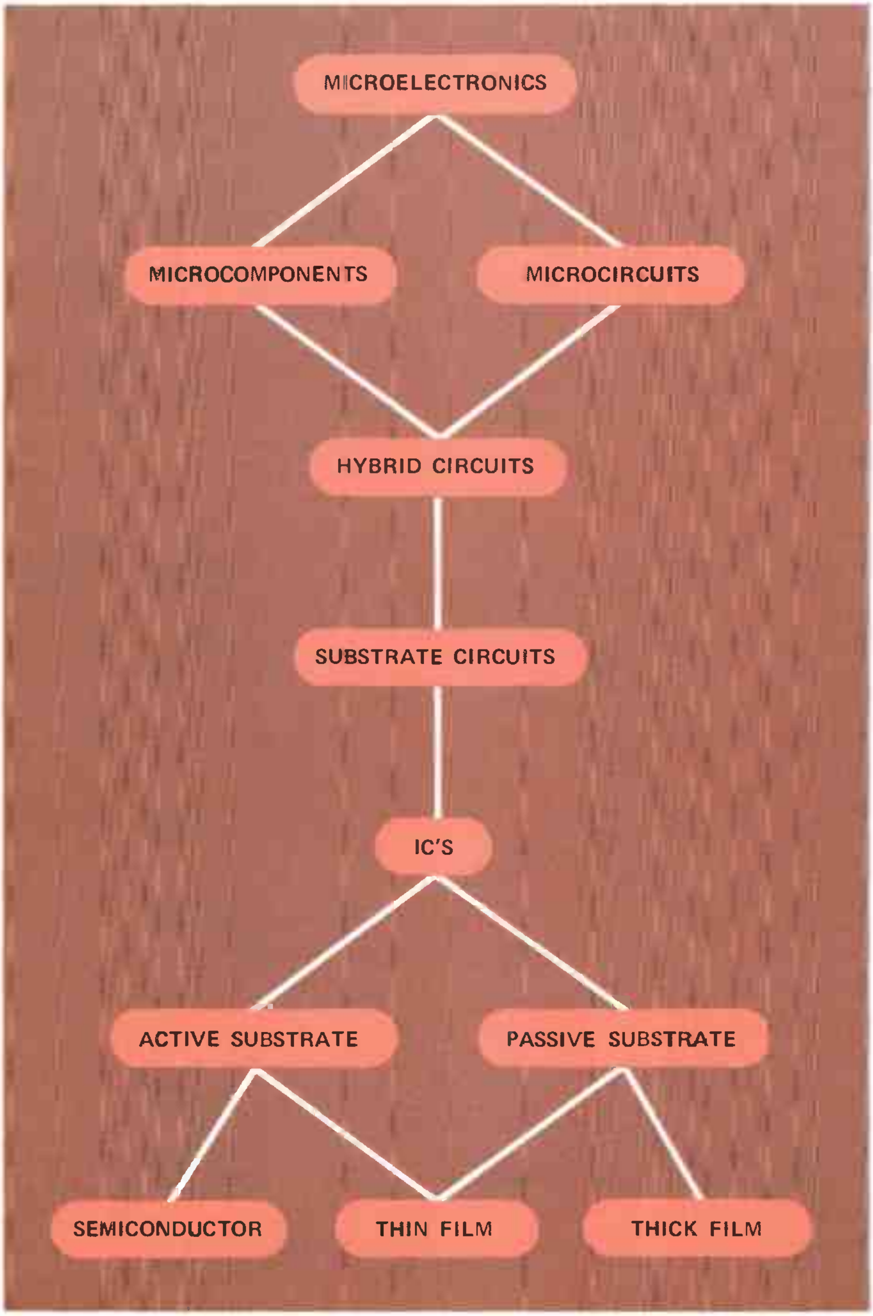


Figure 5. All of the various technologies under the broad umbrella of microelectronics depend on solid-state physics.

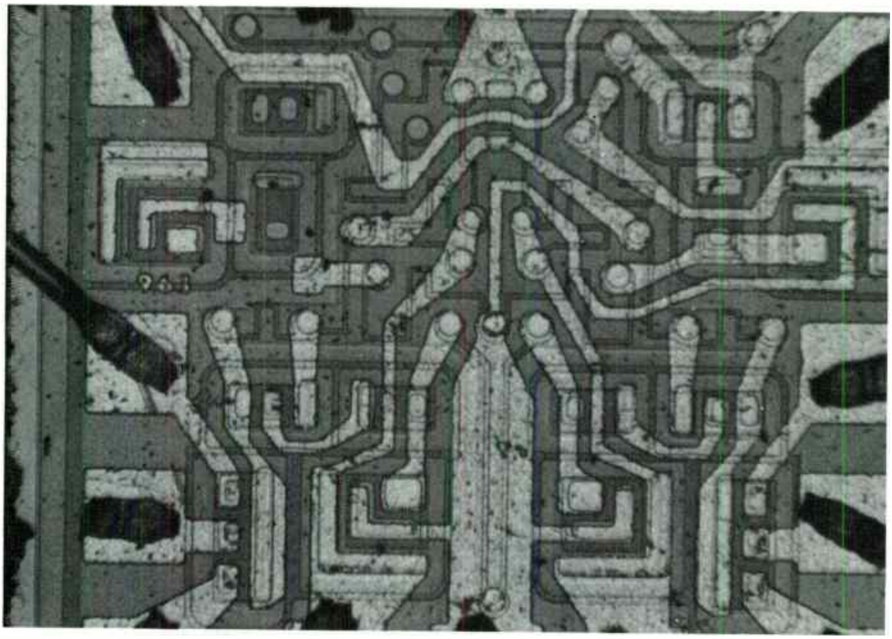


Figure 6. Called a Triple–3 Input Gate, this integrated logic circuit contains three triple input NAND gates. It is used in the Lenkurt 91A PCM Carrier System.

Since they manufacture both digital and conventional telecommunications systems, Lenkurt is of course deeply involved with solid-state development. Both digital and linear IC's are more and more widely used in Lenkurt systems. These systems are characterized by higher reliability, lower operating costs and longer life.

Current Trends

One recent development in the fabrication of IC's is growing square silicon wafers rather than round ones. The result is more usable chips per wafer since the loss incurred through wasted area has been eliminated.

Currently, the growing debate in semiconductor technology involves packaging and interconnection techniques. Proponents of flip-chips and beam lead devices are vying for acceptance by the industry at large.

Primarily, beam leads are favored for their superior uniformity, flexibility of application and overall reliability. Considerations such as thermal dissipation and power handling capabilities seem about equal. Flip-chips, on the other hand, allow for greater packaging density.

There also is an intriguing economic aspect of the boom in semi-conductor technology which is just emerging. As applications for IC's become standardized, more and more semiconductor manufacturers are providing more and more input to the finished product. They are developing their own packaging techniques and subassemblies, and in some cases they even supply whole integrated systems.

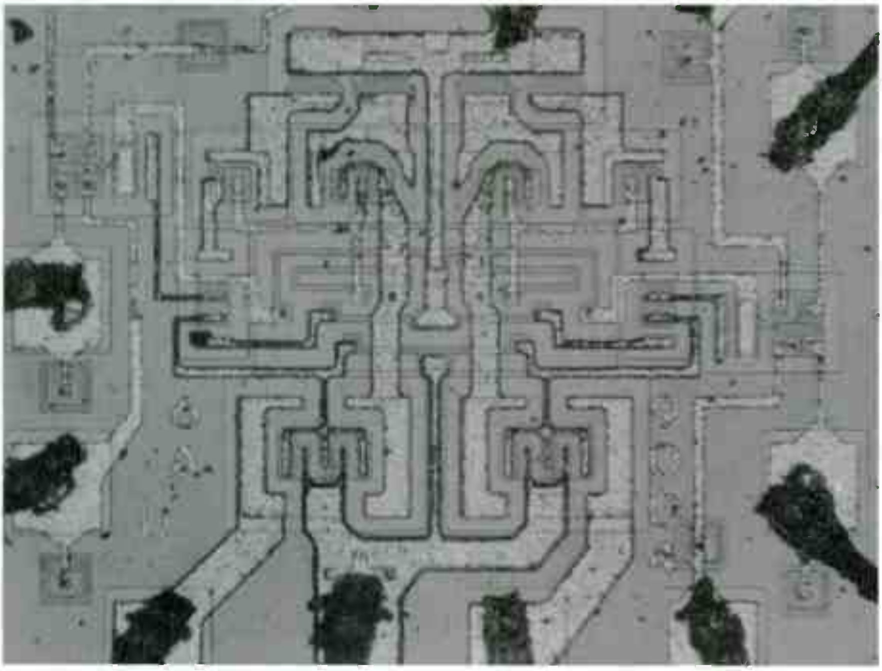


Figure 7. Also used in the 91A, this digital IC has two 4-input NAND gates.

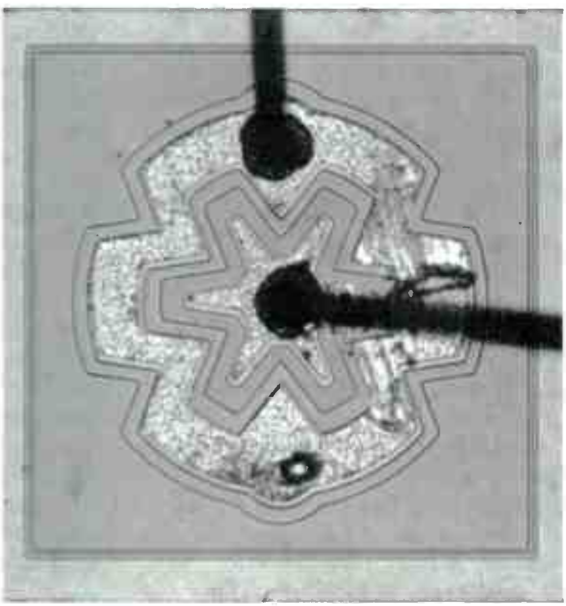


Figure 8. Using star geometry, semiconductor manufacturers are able to increase the peripheral length of the transistor. This enables it to handle greater power and current loads and still maintain low input and output capacitance.

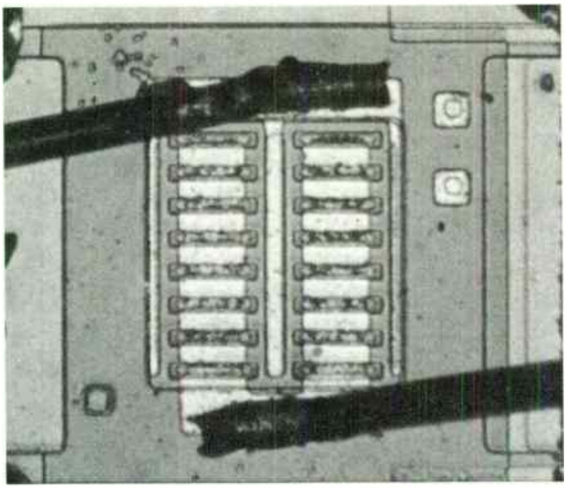


Figure 9. Characteristic of the present state of the art, multiple emitter type transistors are used at Lenkurt in wideband amplifier circuits.

Generally, electronics may be divided into three broad areas – design, component fabrication, and system assembly. Heretofore, design and system assembly were the function of the equipment manufacturer. The semiconductor vendor fabricated only the components to the specifications set by the manufacturer. This is changing. Makers of semiconductor components and circuits are penetrating deeper and deeper into the areas traditionally allotted to equipment manufacturers.

Is it possible that present manufacturers eventually will become only clearing houses or wholesalers for entire electronic systems manufactured totally by the erstwhile makers of semiconductors?

This prospect is not so farfetched as it seems at first blush. It has already begun to happen in the computer hardware manufacturing industry. Stripped of their packages and peripheral devices, there can be little operational difference between two computers whose internal circuitry is manufactured by the same company.

Linear IC's are finding ever wider application in the communications industry. If designers and manufacturers of communications systems become as dependent upon vendors of linear IC's as computer manufacturers are upon digital IC suppliers, they will find themselves in the same position.

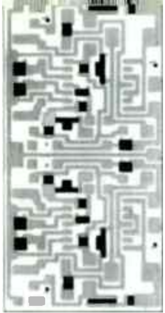
The fast-closing developmental gap between digital and linear IC's makes this possibility somewhat less clear in the communications sector of the electronics industry. But communications manufacturers are going to have to choose between buying more "ready-made" components, or, as some have in the past, making their own.

Beyond such business considerations, it is unwise – if not impossible – to attempt any meaningful predictions as to the future of solid-state technology. Conceivably, any of the circuits or components mentioned in these pages may be obsolete before the ink is dry.

The Lenkurt DEMODULATOR
OCTOBER 1968

Hybrid Microcircuitry

Part 1



Thick film, a modern application of an ancient art form

Since the advent of the transistor, the dominant trend in electronics has been toward more and more miniaturization. The advantages of microelectronics go beyond simple space saving, although size itself is certainly a considerable factor. Experience has shown that solid-state microcircuitry provides higher reliability, reduces problems of heat dissipation and readily lends itself to mass production. Thin and thick films, multi-layer printed circuits, integrated circuits – monolithic, medium, and large scale – are some examples of the relatively new technology of microelectronics.

In keeping with this trend, Lenkurt has been involved for some time with thick film process development as part of a general microelectronics program. Thick films were chosen for their wide range adaptability and relatively low initial expense. Also, thick film technology provides a springboard for ventures into other areas of microelectronics.

Thick -vs- Thin

Thick film is a microelectronic technique where the passive components of an electrical circuit such as conductors, resistors and capacitors are miniaturized by deposition onto a small piece of ceramic material. For general purposes, these circuits may be defined as passive thick film integrated circuits. They differ from thin films both in the thickness of the films themselves and in the method of deposition. Where deposition of thin films is done by sputtering or chemical evaporation in a vacuum chamber, thick films are deposited on ceramic substrates by a screening process using special resistive and conductive inks.

The actual process is simply a modification of the ancient methods used by potters and ceramists to decorate porcelain and clay. Using these methods, all of a circuit's passive components can be miniaturized to fit a small substrate. A substrate, in this case, is the small piece of white "china" on which a thick film circuit is deposited.

Ceramic substrates for the thick film process are typically 96% alumina and 4% glass; their dimensions range from less than one to several inches square and average about 0.025 inches in thickness, depending upon the circuit design. Recently, a beryllia substrate has also been used with considerable success.

Process flow in thick film fabrication is actually a series of interdependent – yet different – processes. First, the design of the circuit is laid out as original artwork on a precision drafting machine. The artwork is then reduced photographically and a silk screen replica of the circuit is made. Passive components such as conductors, resistors and capacitors are applied to a substrate which is then fired in a furnace. After the firing cycle, the resistive and conductive values are tested and adjusted. Finally, compatible discrete active components – transistors, diodes, IC's, – can be

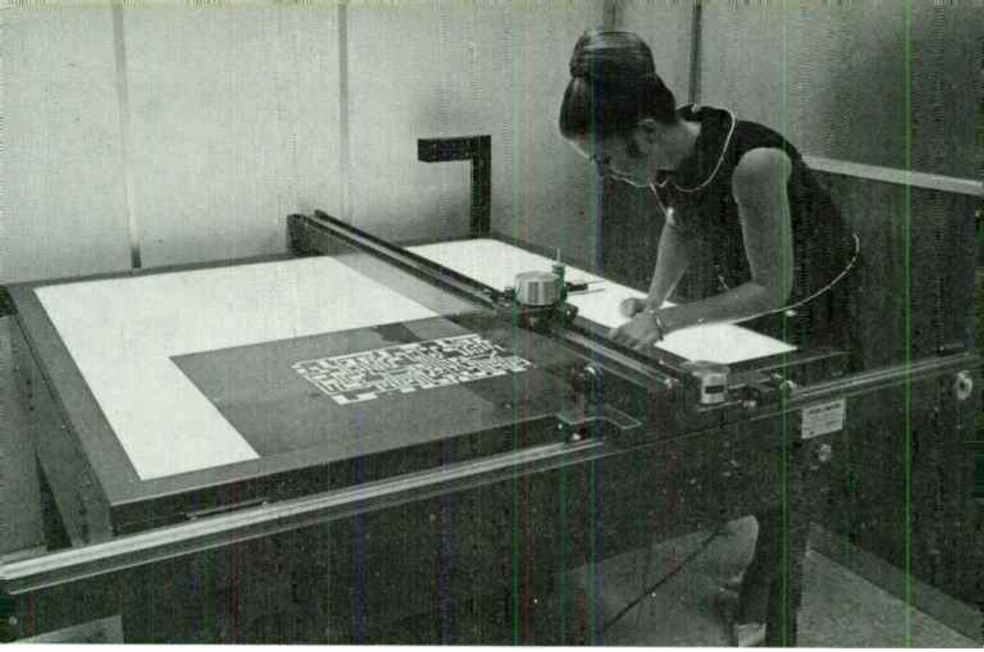


Figure 1. A precision drafting machine such as a coordinatograph is a "must" for development of microcircuitry. Note the digital light display in the background.

attached, transforming the passive circuit into a hybrid thick film circuit. The entire circuit is then packaged for use in a system.

Original Artwork

When a design request is received, the first requirement is to lay out a large working model of the circuit. Design requests normally contain all the necessary engineering data. The artwork is done on acetate or other translucent, stable drafting film using an electronic precision drafting machine called a coordinatograph (Fig. 1). The coordinatograph is accurate to 0.001 inch; movement in the x and y axes is registered electronically with coordinates appearing on a digital light display.

After the original artwork is completed, the acetate masters go to the photo lab for photographic reduction. Using a specially built camera with height adjustable over six feet, the circuit layout is reduced in size by a factor of up to ten. One effect of reduction is to increase accuracy of the design by the same factor – hence, tolerances of 0.0001 inch.

Once the film is developed, the resultant negative (or a positive) is used to make a silk screen stencil of the circuit. The stencil film itself is a photosensitive gelatin composition backed with a very thin sheet of polyester.

After applying the negative to the gelatin film, they are put in the plate making machine and exposed over a mercury vapor lamp. Due to the comparatively slow warm-up time of the lamp, exposure is measured in light units (lumens) rather than in time.

Exposure hardens that portion of the gelatin not protected by the circuit pattern on the negative. The negative and stencil film are separated and the unexposed gelatin is rinsed away. Having been hardened by exposure, the circuit pattern resists removal by 110° water used for rinsing (Fig. 2).

After rinsing, the remaining pattern is further hardened under cold water, and the stencil is ready for mounting on a screen. The screens used in the thick film process are not actually silk screens but are made from fine stainless steel wires. The mesh, or wires per inch of screen area, range from 105 to 325 depending on film thickness and pattern accuracy requirements.

Using a transparent alignment mask to aid in centering, the stencil film is mounted wet on the screen. Extreme care must be taken at this stage to avoid accidents which later might ruin the registration of the pattern on the substrate. Since the pattern is so small, it is conceivable that some of its elements might not adhere properly and turn or even fall partly through the mesh of the screen.

When the stencil film has been fixed to the screen, excess water is blotted off – but some is left so as to allow capillary action to draw the film even more firmly onto the screen.

Final drying of the screen is best done naturally at room temperature. Since this normally takes several hours, attempts have been made to speed up the process with forced warm air. But this often results in a too brittle film and increases the danger of losing some of the finer parts of the circuit pattern.

A different screen must be made for each different pattern – conductive, resistor or final glaze – which is to be applied to a substrate.

How Clean is Clean?

From drafting room to final packaging, all of the work involved in the thick film process must be done in a strictly controlled environment. Since any of the processes involved can be adversely affected by dust, pollen, fibers or even fingerprints, extensive efforts are made to maintain a clean atmosphere. All rooms are air-conditioned. Dust and other foreign matter are filtered from the air, smoking is not allowed and technicians wear lab smocks and silk gloves while working. In this respect a thick film lab resembles a hospital operating room.

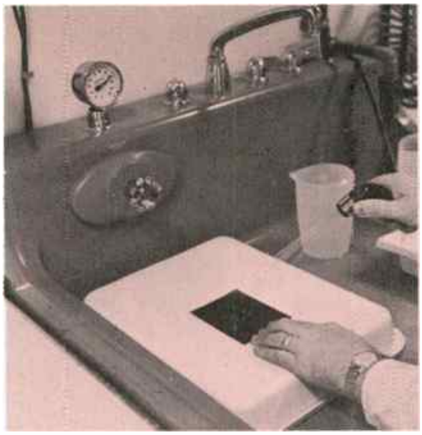


Figure 2. The resistor pattern can be seen emerging from the stencil film as the unexposed area is rinsed away. Using the temperature blender in the background, a steady water temperature of 110° is maintained.

The finished product begins to take shape when the conductor, resistor and glaze patterns are screened onto the substrates and the elements are fired in a conveyor furnace. Circuit patterns are applied to the substrate by a silk screen process using precious inks. The inks are actually slurries of noble metals suspended in a highly viscuous paste. The metals may be gold, platinum, palladium, silver, or an alloy of any of these. The paste is mostly an organic compound with an evaporative binder or catalyst added.

Because of their nature and expense ($1.29 to $6.40 per *gram*), the slurries – or inks – require special handling. Their containers are stored on a rack of slowly rotating rollers. The continual rolling action keeps the inks ready for use. Constant agitation keeps the metal particles in suspension rather than collecting in the bottom of the jar.

Deposition

Called simply a screener, the apparatus for applying the circuit patterns to substrates consists of a vacuum chuck for holding the substrate securely, a micrometer-adjustable mount for the screen itself, and an adjustable squeegee which pushes ink through the pattern of the screen. The squeegee blade is usually polyurethane of a specific hardness. Adjustments are in all three directions to ensure perfect interaction between squeegee, screen, and substrate. The distance between screen and substrate is called the "snap-off".

Substrate, screen, and squeegee must be perfectly parallel and the squeegee's motion steady and horizontal or the thickness of the deposited ink will vary. Since the electrical values are a function of the ink's volume, varying thickness will cause inconsistencies in resistance and conductance.

Ink is spread onto the screen around the circuit pattern with a small spatula, and the squeegee is moved across the screen. This action forces the ink through the screen onto the substrate and prints the desired pattern.

Normally, several test runs are required before alignment of all the various parts is perfect. All of the ink used for testing and alignment is kept and returned to the supplier for reclamation. Used screens and substrates used in test screenings are washed in acetone, or thinner. This rinse along with other used ink, is returned to the ink manufacturer for credit.

After testing and final adjustments, any number of substrates may be screened as long as the screen itself holds up and is not "coined" or otherwise damaged. "Coining" results when screen and substrate are too close and the screen's wires are bent leaving an impression of the substrate.

Alignment of screen and substrates for subsequent screenings is done by using small registration marks on the screen. They line up precisely with other similar points left on the substrate by previous screens.

Drying and Firing

Firing is done in a cycle. The sequence is tied directly to the temper-

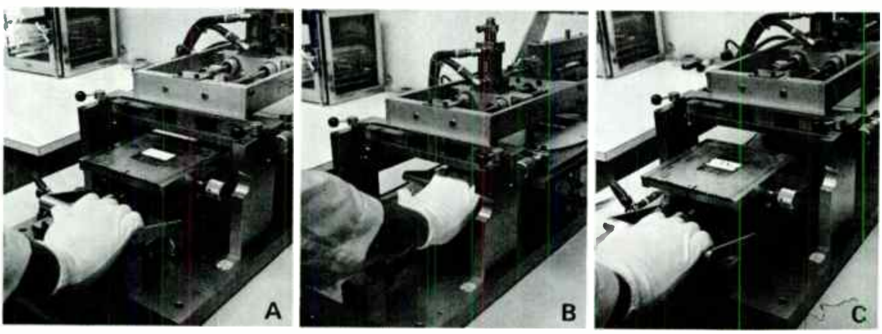


Figure 3. There are three steps in depositing the resistor pattern: (A) substrate is fitted to the vacuum check, (B) the substrate passes under the screen and the squeegee forces the ink through the screen and onto the substrate, and (C) the chuck is withdrawn with the pattern deposited on the substrate.

Figure 4. Using airbrasive powder for trimming resistors, the trimmer works in conjunction with an electrical bridge. The bridge measures resistance and enables the technician to trim resistors to within 0.1% of desired value.

ature profiles of the various component values of the circuit. Since resistance is inversely proportional to firing temperatures, components requiring higher firing temperatures are deposited and fired first. These will often be the conductor patterns. However, one or two resistor screenings are sometimes made before the conductor screening. This is done in special cases where resistor inks of very precious metals are used and higher firing temperatures are required.

The normal deposition sequence is; screening, settling, drying and firing. Immediately after screening, the substrates are set aside so that the ink can settle. Settling fills the voids and impressions left in the ink by the screen.

Drying is accomplished by leaving the freshly screened substrates in a temperature-controlled oven at 125° for about 15 minutes. The primary purpose of drying is to drive out the binder from the ink.

Firing temperature profiles vary (from 500° to 1050°C) according to desired electrical values. These values in thick films are a function of two factors, the volume of ink and the temperature at which they are fired. Since the resistance is inversely proportional to the ink's volume it can be increased by airbrasive trimming but never decreased. Consequently, resistor ink is likely to be applied some-

Figure 5. The conveyor type furnace can produce firing temperatures up to 1500°C.

what thickly. The term, airbrasive, derives from air-driven abrasive powder.

In some cases where more than one of a circuit's elements require the same firing temperature, resistor and conductor patterns will be screened and dried without individual firing. Then, after all the elements are deposited, the entire circuit can be fired.

Resistors and Capacitors

Typical sheet resistivities in thick film resistors range from 1 ohm/sq. to 1 megohm/sq. with thickness normally set a 1 mil. The formula for sheet resistivity is:

$$R = kt\frac{l}{w}$$

Where:

R *equals sheet resistivity.*
k *equals the resistivity constant of the ink.*
l *equals length of the resistor pattern.*
w *equals width of the resistor pattern.*
t *equals thickness (assumed to be constant).*

From this formula, it can be seen that resistance in thick films is a function of the shape of a resistor. When length is increased in relation to width, resistance value is proportionately increased.

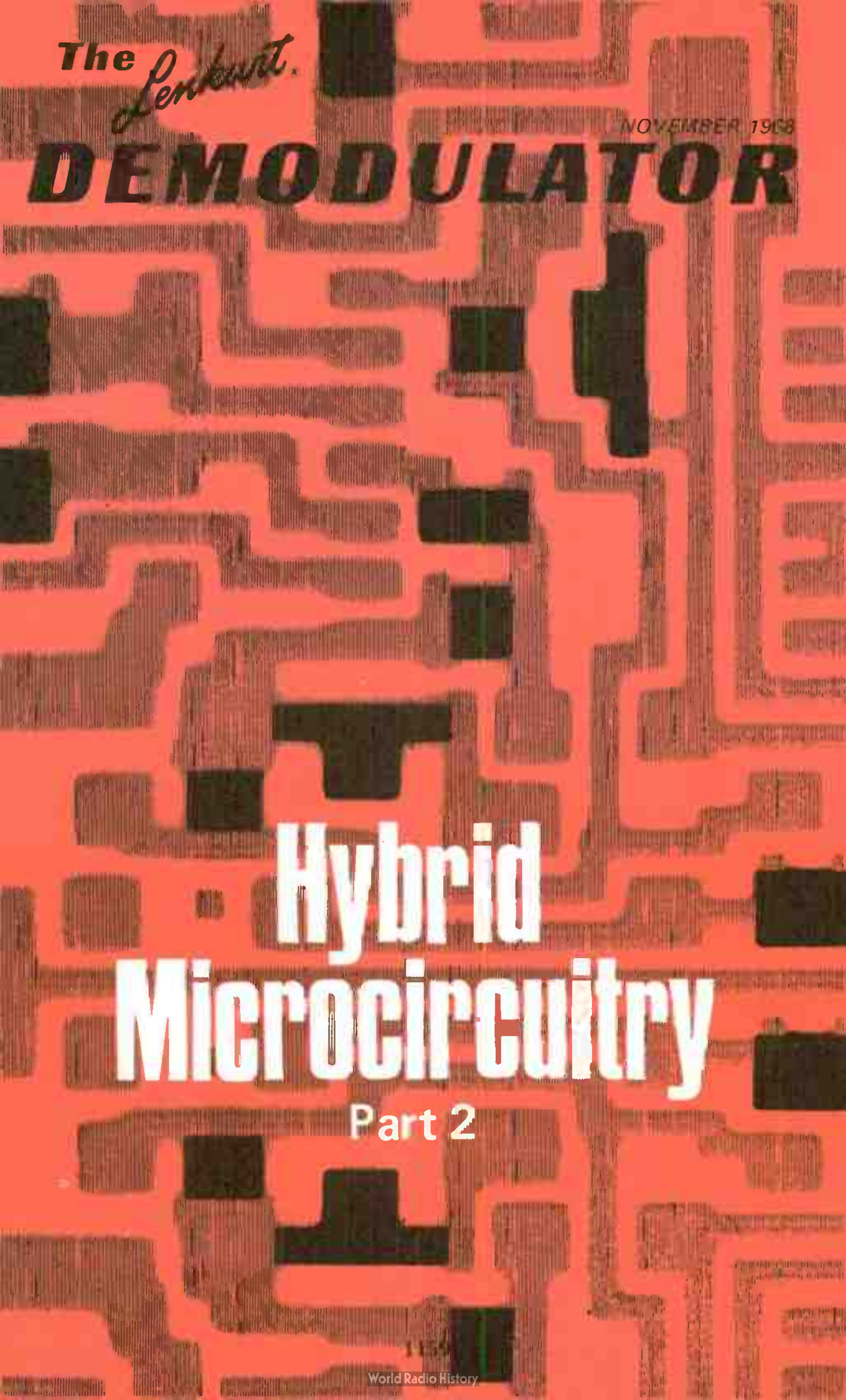
After firing, the distribution of values for any particular resistor pattern is normally ±10%. By firing resistors to only about 90% of the desired value, careful trimming can then increase accuracy to ±1%. Lenkurt engineers consistently produce thick film resistors with such value tolerances.

Though done somewhat less often, capacitors may also be fired into a thick film circuit. Capacitor inks are usually composed of oxides such as titania or barium titanate and multicomponent glasses. Inherent limitations of the inks used for capacitors make high capacitance values extremely difficult to achieve. Capacitances up to 100,000 picofarads per square inch can be achieved, but in the lower ranges – 10,000 pF/in.^{2} – capacitors are more stable and can successfully pass breakdown tests up to 500 volts. For values beyond these limits, or for special tolerance and stability requirements, it is necessary to attach discrete components or chips after the thick film process is completed.

Once all the passive components – conductors, resistors, and capacitors – are fired, the substrate is coated with a glaze and refired. If the device has been designed merely as a passive circuit, it is now ready for packaging. Final packaging is a simple process of dipping the substrate into some kind of epoxy-based plastic after the leads have been attached.

If it is an active circuit, the glazed substrate must be presoldered before active components can be attached. Presoldering leaves bumps of solder at the points where active component leads will be attached. Typical active components are transistors, diodes and IC's. After the active components have been attached and final tests made, the entire circuit is encapsulated and ready for use.

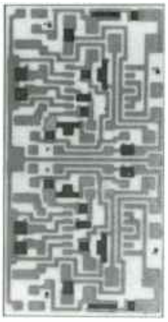
Editor's Note – The November Demodulator will discuss some of the applications and extensions of thick film technology, as well as some limitations.



The Lenkurt.
DEMODULATOR
NOVEMBER 1968

Hybrid Microcircuitry

Part 2



New techniques in microelectronics may lead to breakthroughs in microwave design and packaging.

Impressive as many of the current applications of thick film technology are, its potential uses and extensions are even more so.

Significant space savings and increased reliability have been achieved in several Lenkurt systems employing thick film circuits. Among them are the 82A and 83A Station Carrier Systems and Lenkurt's 91A PCM Cable Carrier.

At this point most are filter circuits such as the one in Figure 2. Beyond simple filter circuits, the heightening interest is in developing hybrid microcircuits for use at microwave frequencies. To do this, other, more complex areas of microelectronics must be explored.

One of the outstanding aspects of thick film technology is its commonality with other microelectronic techniques. Whether the medium is printed circuitry, thick film, thin film or integrated circuitry, the goal is essentially the same – increased reliability through miniaturization. And regardless of the means employed – from PC's to IC's – the initial steps are the same. The circuit design is first laid out as original artwork, then reduced photographically (Fig. 3).

Through this similarity, other microelectronic techniques such as thin film may be viewed as extensions of thick film technology. The thick film process is most effectively employed in the miniaturization of general purpose electronic circuits which may allow relatively loose tolerances. But thin films have accuracy requirements roughly ten times as stringent as those for thick films.

Stripline and Microstrip

Tight tolerances on this order are required of components intended for use in systems designed to operate at microwave frequencies. Typical of such components are stripline, microstrip and other microwave integrated circuits (MIC's).

Strictly speaking, integrated circuits have a different function at microwave frequencies than at the lower frequencies. At low frequencies, hybrid IC's simply provide the interconnections between the various active components in a circuit. But in the microwave range, they essentially become transmission lines and provide all the same functions as sections of transmission line – such as waveguide – can be made to provide.

Stripline and microstrip are techniques for miniaturizing circuits. As a replacement for conventional circuits, stripline and microstrip are especially effective at the higher frequencies where tight control of a line's characteristic impedance is essential.

Since a line's ability to carry a signal is a function of its electrical properties as well as its size and shape, it is a rather straightforward process to go from cable to stripline.

Basically, a coaxial transmission line consists of a round center conductor located concentrically inside a cylindrical outer conductor. The stripline

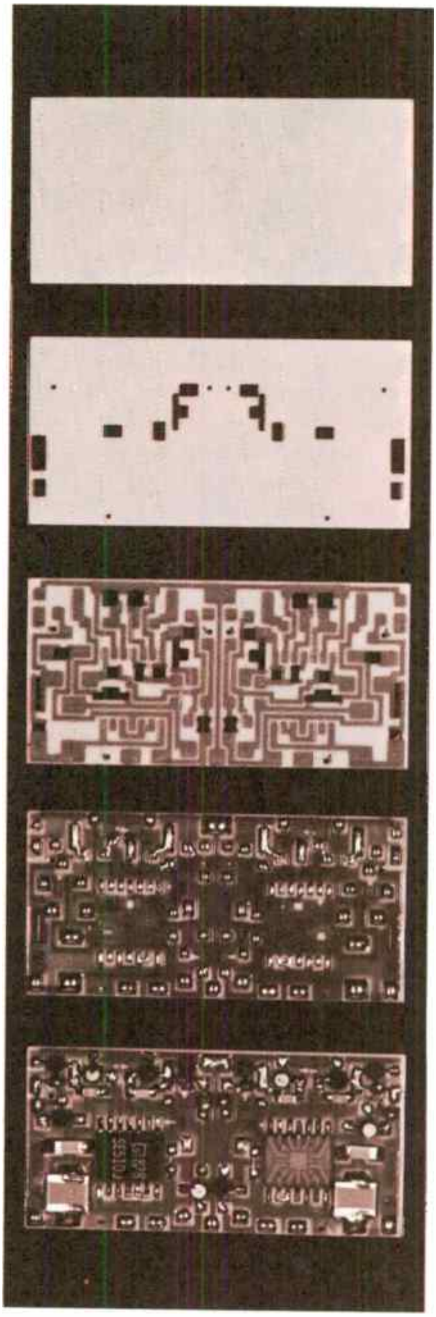


Figure 1. From the top down, these substrates show the various steps in fabricating a circuit using the thick film technique.

is much the same thing, only the shape and size are different. It consists of a conductor in a dielectric material (may be air) supported by two parallel ground planes in a kind of sandwich configuration.

The next step in this evolution is to simply lift the top off the sandwich and the stripline becomes a microstrip (Fig. 4).

Photoetching

Circuit patterns required for stripline and microstrip configurations generally require dimensional accuracies that cannot be achieved using direct screening techniques. Consequently, they are often produced by etching a metallic conductor layer on the surface of a suitable dielectric substrate. The substrate may be teflon-glass laminate, ceramic or glass. And the metallic surface layer is usually a copper sheet, silver applied by thick film process or evaporated thin films such as gold on chrome.

Photoetching begins with a large scale drawing of the required pattern which is later reduced photographically to a one-to-one facsimile. Large-scale original artwork permits greater accuracy and provides a more convenient working dimension.

After thorough cleaning and drying, the metallic surface to be etched is uniformly coated with a thin film of photoresist. The film is applied by dipping, spraying, roller coating or spinning the substrate. It is then fired to remove residual solvents before exposure.

A photographic image of the circuit pattern is produced on the surface by placing the 1:1 scale transparency in contact with the coated substrate and exposing it to ultraviolet radiation for a pre-determined time. For a negative acting resist, the action of the ultraviolet radiation is to change the molecular structure (polymerize) in the areas

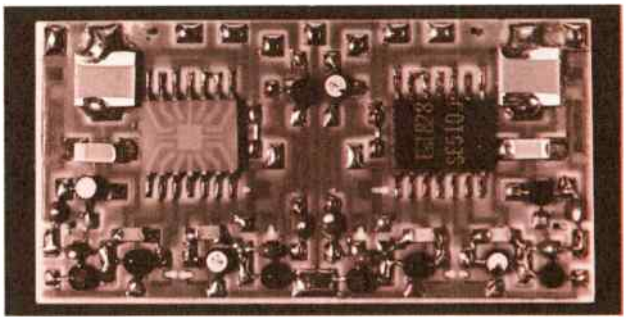


Figure 2. This filter circuit is an example of Lenkurt thick film technology. (Shown here actual size.)

of the photoresist corresponding to the transparent regions of the photographic image.

After processing in a developing solution, the polymerized areas remain as a tough chemical-resistant surface. The areas corresponding to the opaque regions of the image dissolve in the developer and expose the metallic surface underneath.

There is also a method using positive acting resist in which this action is essentially reversed. In this method, the circuit pattern on the photographic image is transferred to the metal surface of the substrate. The polymerized resist acts as protective covering and prevents etching of the covered areas. The unprotected regions are etched away, leaving the desired circuit pattern.

In cases where one metal is deposited atop another, it is possible to etch each one separately. This technique is called selective etching and is possible due to different etching properties of different solutions in relation to certain metals. Other methods are the photomask technique and selective electro-chemical plating.

Stripline and microstrip are two of several circuit configurations which can be produced by these methods. Furthermore, they can have active components such as semiconductors attached and become working hybrid microcircuits.

When perfected, stripline and microstrip may be used in microwave radio systems to help replace bulky waveguide plumbing.

Some Extensions

Another related area of microelectronics is concerned with developing solid-state devices for microwave output power. The primary emphasis is on components such as diodes and semiconductors rather than circuitry, although, the components obviously must work in or with microcircuits.

Already, some lower frequency systems such as Lenkurt's 71F2 Radio (2 GHz) are solid-state. But ultimately, engineers hope to replace the klystron with a small solid-state device in the higher microwave bands as well.

High frequency application is the major obstacle to the development of operational, all solid-state microwave devices. This is not universally true – devices such as switches, couplers, multipliers, filters and some amplifiers and oscillators have been successfully designed for use throughout the microwave spectrum. But operational, high frequency replacements for transmitter tubes are still in the future, albeit the near future.

Then why all the emphasis on microcircuitry and solid-state? Although the advantages of microcircuitry do not necessarily *outnumber* the disadvantages they do *outweigh* them. Improved overall performance is of course the greatest advantage offered by solid-state devices and microcircuitry. This is primarily a result of the near-perfect stability that can be achieved by using thick film and other hybrid integrated circuits in communications systems. Also, increased ruggedness of the components themselves adds to the overall reliability of the system.

A considerable cost advantage is realized since microcircuitry fabrication techniques are readily adaptable to mass production methods.

At the present state of the microelectronic art, there are still many problems. Higher line losses, parasitic capacitances, and other spurious modes resembling cross-talk in a telephone circuit are among the knottier ones currently facing the design engineer. Also, limited handling characteristics in terms of voltage and power present unique problems. Moreover, the high potential packaging density of miniature circuits on substrates can create problems of heat conductivity.

The problems, however, are not considered insurmountable. The art is still quite young – and even some of the problems are just now being recognized.

The problems are extremely complex, and vary with each new device. In all of them, however, the primary problem is the same and very elementary. It is simply to develop a small, solid-state apparatus capable of generating and manipulating sufficient signal power for continuous wave (cw) propagation at microwave frequencies.

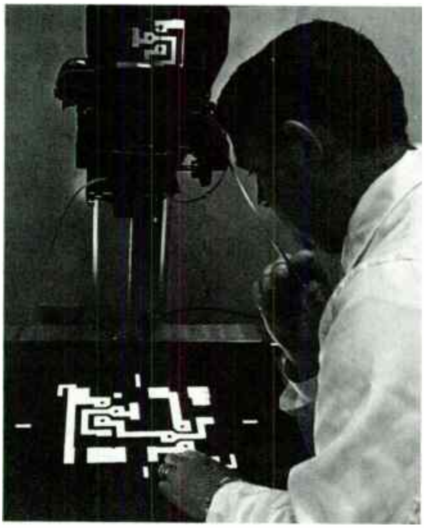


Figure 3. Photographic reduction of the circuit design is an indispensable part of any microelectronic technique.

Three Candidates

Progress is being made. Currently, some of the more likely devices are gallium-arsenide (GaAs) crystals such as Gunn diodes, limited space charge (LSA) devices, impact avalanche transit time (IMPATT) and other avalanche diodes.

A GaAs crystal can be made to oscillate at microwave frequencies simply by applying a high voltage (dc bias) across it. First though, to achieve the desired electron activity, the crystal must be "doped" with a foreign element such as sulphur or selenium.

What goes on inside the crystal is not so simple. But it is basically a series of electron movements which, taken cumulatively, cause an oscillation. This, in turn, can be used to create output power. The oscillation, called the Gunn effect, is named for its discoverer, J.B. Gunn. The device itself consists of a doped GaAs crystal and wafer mounted in a heat sink cavity and surrounded by dielectric material.

Gunn oscillators provide cw power with peak outputs ranging from a few milliwatts to about one kilowatt, depending on frequency. As the frequency goes higher, power output falls off. Low average power is one of the limitations of Gunn devices. Up to now, it hasn't been possible to generate sufficient average power output at high enough frequencies for microwave applications.

Gunn devices are further limited in their present usefulness by inherent problems of heat dissipation. They are reliable to only about 100°C. Beyond that, the entire device usually breaks down.

LSA devices, on the other hand, are not hampered so. Related to the Gunn effect, limited space-charge accumulation is another related phenomenon which may make possible power outputs of a watt or more in the microwave spectrum.

In the case of LSA devices, oscillation is the result of a negative resistance characteristic (see the *Demodulator*, Sept. 1968) in a high dc bias field.

Unlike the Gunn devices, the power output of an LSA oscillator is not dependent on frequency. This is because the accumulated space charge is not a function of the electron's transit time — an important fact in Gunn diodes. As the thickness of a Gunn diode's active region increases, the device's efficiency decreases correspondingly. This is not true of LSA devices.

An LSA oscillator can be many times as thick as a Gunn diode of the same frequency. Consequently, it can withstand much higher voltages and temperatures. It is therefore able to produce consistently higher average power outputs.

An IMPATT diode is a simple silicon PN junction capable of generating more than one watt of cw power at about 12 GHz. Here too, oscillation is a result of negative resistance caused by a combination of the crystal's internal emission (the avalanche effect) and electrons moving at saturation velocities.

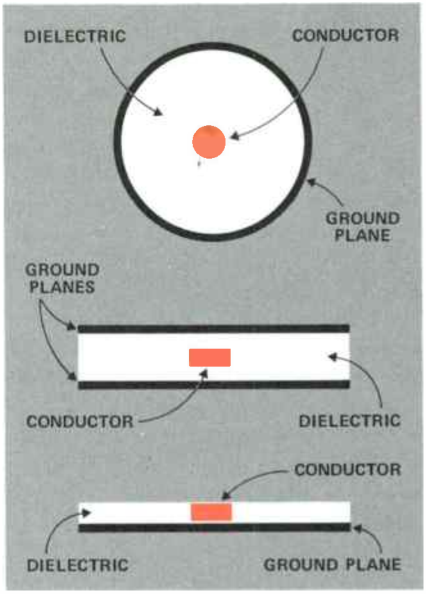


Figure 4. Transmission lines have undergone changes from coaxial cable (top) to stripline (center) and microstrip (bottom).

Right now, the biggest problem with IMPATT devices is the inherent noise which accompanies any avalanche device. Stabilizing cavities have been able to reduce avalanche noise to a useable level for some devices with limited applications at lower frequencies.

But, problems such as noise, heat and low voltage capacities are not insurmountable. One or another of these devices is destined to replace tube-type transmitters in the foreseeable future.

At the present state of the art, no one of the devices discussed here is capable of doing the job done by conventional tube-type power sources. However, some kind of hybrid circuit incorporating the features of many of these components is another area of interesting possibilities.

MIC's

Within certain frequency and power limitations, the most effective devices in use today are hybrid MIC's employing thin and thick film technology.

Microwave integrated circuits employing microstrip transmission lines have met with considerable success at the lower frequencies. But because of its small circuit size, transmission losses are high compared to waveguides or coax. This becomes a limiting factor in high power applications. One of the ironies of microelectronics is the fact that as engineers come closer to their performance goals through miniaturization of devices, the devices are more vulnerable to power and heat. Consequently, the ability to produce useable power output is limited by the technique itself. But again, these problems are considered as only temporary.

Smaller, stabler and more reliable microwave radio systems are but one promise offered by the art of microelectronics. Compact, portable radar systems is another. The idea of small, highly portable radar systems is, of course, very appealing to the world's defense agencies. This is especially true of airborne radar and other systems intended for use in battle zones where there is a high probability of damage to components. Miniaturization of components will allow duplication and redundancy sufficient to provide back-up operation in almost any situation.

Whether any one or a combination of several of the devices discussed here is finally adopted as the industry standard, the effects will be dramatic and far-reaching.

Each of the individual devices has much to offer, yet each is beset with problems. Therefore, the best bet for the future will probably be a device resulting from joint efforts involving several technologies.

Millimeter Communications?

What then? A simpler, less expensive microwave radio comes to mind first. Another – if somewhat more remote – possibility might be the capability to really open the millimeter portion of the spectrum for communications. A small amount of successful exploration has been done. Experimental, prototype, millimeter systems do exist.

If sufficient channels to meet the demands of the future are to be found, other areas of the spectrum must be opened. The increased bandwidth offered by the millimeter range can do much to further the work in these areas. Experimental efforts at using the millimeter band for communications have been generally successful. An LSA device is currently being used in guided wave PCM system operating above 50 GHz.

But none of the possibilities mentioned here will be realized until the problems are worked out. And in microelectronics, just as in other technologies, problems exist to be solved.

Success in one area often hinges on previous successes in related areas. So, what has been learned from thick film and IC fabrication will provide the foundation for success in the efforts toward microwave miniaturization.

The *Lenkurt.*

SEPTEMBER 1968

DEMODULATOR

negative resistance devices

Negative resistance and impedance — effective elements in modern circuit design.

The fascinating implications of negative resistance have intrigued scientists and engineers for decades. Near the turn of the century certain "freak" devices – among them the carbon arc – exhibited negative resistance properties and were actually put to practical use.

The carbon arc, the dynatron tube, and more recently the tunnel diode all fit the class of physical devices having negative resistance. Circuits using standard components can also be used to produce negative resistance and now have their own place in many new telecommunications applications.

In the dynatron tube, developed about 1918, the plate current increases when the plate voltage is reduced. A very similar negative resistance is found in the modern tunnel diode. The tunnel diode is a two-terminal device that has the ability to amplify because of a unique relationship of voltage to current over a portion of its operating range. The voltage-current curves in Figure 1 compare the ordinary resistor, a conventional junction diode, and the tunnel diode.

At a certain point on the curve of the tunnel diode, an increase in the applied voltage causes a decrease in the current. As long as the tunnel diode is operated within the limited voltage range indicated, and in a suitable circuit, the negative resistance effect may result in amplification over a range extending up to microwave frequencies.

In most ways, negative resistance can be thought of as being the reverse of positive resistance. Over the voltage or frequency range within which it is designed to operate, the negative resistance device will deliver energy to the circuit to which it is connected, in contrast to a positive resistance which absorbs and dissipates energy. In the tunnel diode, it is strictly an internal phenomenon which allows this. In the circuit devised to work as a negative resistance, it is essentially a positive feedback technique that achieves the result.

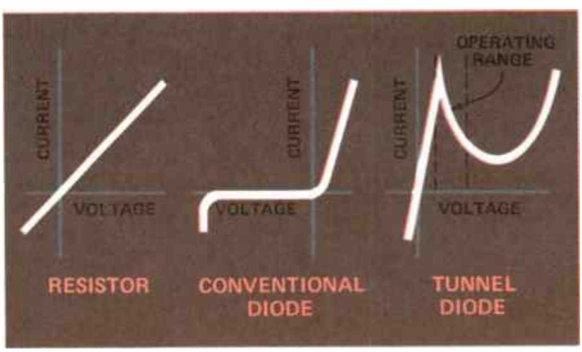


Figure 1. Comparison of the voltage-current relationships of resistor, conventional diode, and tunnel diode. The tunnel diode shows unique negative resistance characteristics over a narrow range of voltages.

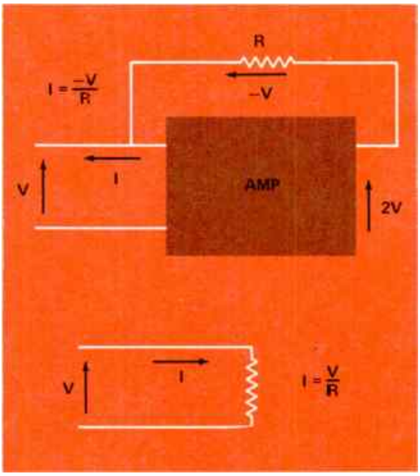


Figure 2. Feedback circuit produces negative resistance. If voltage gain is two times, current of -V/R will flow opposite to that in a conventional resistor, shown below.

Feedback Circuit

The awareness that positive feedback may be used to obtain negative resistance probably came with some of the first radio receivers. There, positive feedback circuits were used to form regenerative amplifiers. Basically, a regenerative amplifier is simply an arrangement where a controlled amount of signal is returned from the plate of a vacuum tube to the grid – or from output back to input. This signal is amplified over and over again, adding gain each time around.

A modern version of this feedback circuit – and an example of a simple negative resistance – is shown in Figure 2.

Impedance Converters

An extension of the first feedback circuits led to circuits more accurately described as negative *impedance* converters (NIC). In some respects the ideal NIC resembles a transformer with an unusual twist. If a positive impedance is applied to one end of the circuit, the negative of that impedance will be seen at the other (Fig. 3).

The reactance of a capacitor is negative and at any one frequency its value can be chosen so that this negative reactance exactly cancels the positive reactance of an inductor. This happens in every tuned circuit. But the capacitor is not behaving exactly like a negative inductor. On the other hand, if an inductor terminates one end of a NIC, the impedance seen at the other end will truly behave like the negative of an inductor at all frequencies, and is a proper negative impedance.

Working from this concept, engineers at the Bell Telephone Laboratories in the early 1940's saw the possibilities for a negative impedance telephone transmission repeater. Inserted in series with the line, the NIC would decrease the impedance, and therefore increase the line current and reduce transmission loss. The first successful version was built in 1948 and became known as the E1 repeater. It first appeared with vacuum tubes, and now transistor models are used extensively in the exchange plant of the Bell system. The E1 repeater is transformer coupled, locally powered, and designed to match the characteristic impedance of the cable.

A transistor version of the E1 repeater is shown in Figure 4. Positive

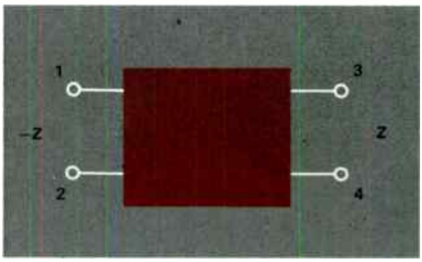
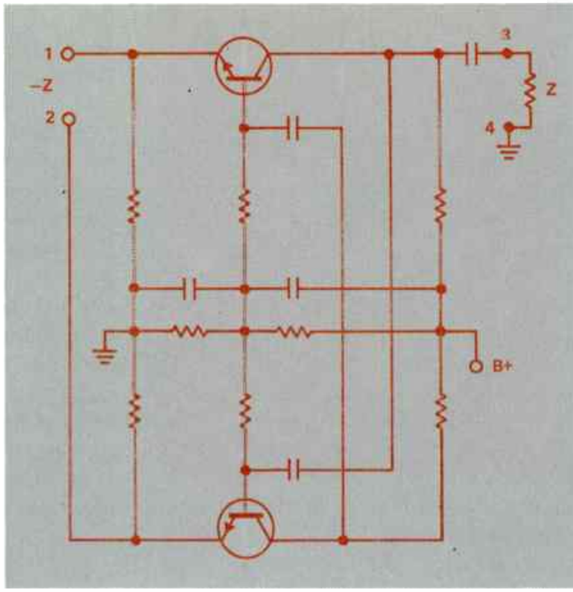


Figure 3. Positive impedance at terminals 3 and 4 will appear as negative impedance at 1 and 2.

Figure 4. The Bell E1 repeater uses positive feedback circuits to produce negative impedance characteristics. Terminals 1 and 2 are coupled to transmission line.



feedback in the circuit produces the negative impedance characteristics, while the negative feedback provides stability. The value of the impedance applied to terminals 3 and 4 appears at terminals 1 and 2 as the negative of that value. In practice, terminals 1 and 2 are effectively connected in series with the two sides of a telephone line and present to the line a series, or voltage-type negative impedance (Fig. 5).

This type converter – or booster as the application better suggests – is a very apt repeater for two-wire lines operating at voice frequency. It is relatively easy to operate at these frequencies, and because it is strictly an impedance device, the repeater provides gain to the transmission line in both directions.

Because of its nature, the E1 repeater is more likely to be found at exchange offices rather than along the line. In fact, for best impedance matching, it should appear at the midpoint of the transmission line.

NIB

More recent experimental efforts at Bell Labs are in the direction of a small two-way repeater powered by line current and spaced along a voice line much like inductive loading coils. It is, in fact, envisioned that the new

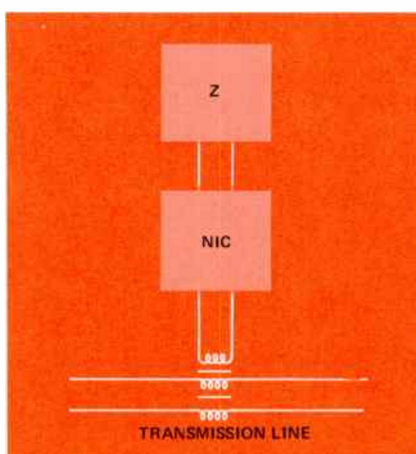


Figure 5. Connection of impedance Z and NIC to transmission line.

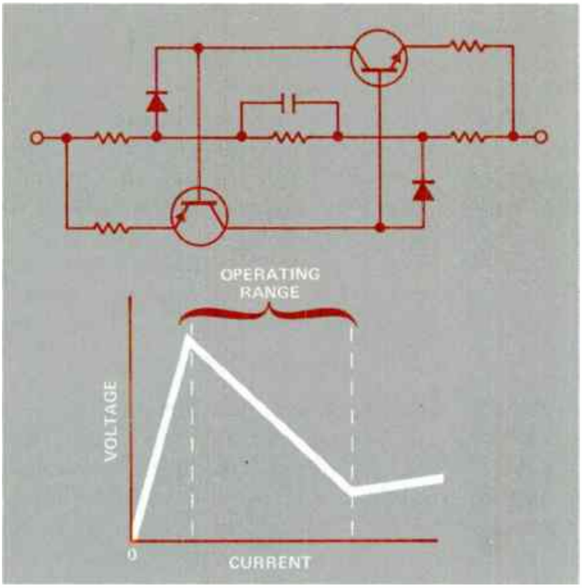


Figure 6. Bell negative impedance booster (NIB) operates in series with the line and is powered over the cable. Slope of the operating curve is stabilized by transistor emitter feedback. Diodes add to the operating linearity.

repeater could replace existing loading coils at about the same spacing.

The negative impedance booster (NIB), as Bell prefers to call it, is more a resistance than an impedance device (Fig. 6). It operates in series with the line and Bell says the NIB might someday be built into the cable during manufacture. The NIB, when perfected, would result in practically lossless and distortionless signal transmission for long rurul telephone lines, local inter-office trunks, and other voice frequency uses.

The NIB is markedly different from the E1 repeater. Spacing of the NIB would be at regular intervals down the cable, preferably not more than a quarter wavelength apart at the top frequency of the transmission band. For telephone speech a suitable spacing would be 12,000 feet. At that spacing, essentially flat, lossless transmission could be achieved up to 7 kHz – up to 18 kHz at 6,000-foot spacing.

Because no coils are used (either as inductors or as matching transformers) the circuits lend themselves well to integrated circuit techniques utilizing only resistors, capacitors and solid state devices. On long subscriber loops, cable with repeaters may be less expensive than the larger sized cable that would normally be necessary.

The main advantage of the NIB over conventional repeater amplifiers in telephone transmission is its simplicity in both construction and operation. Because it has no directional properties and boosts the signal in both directions it can be positioned almost anywhere on the line without concern for expensive hybrid transformers or level control. Additionally, the repeater provides almost automatic equalization by compensating for the increase in cable attenuation with frequency. Delay distortion in the experimental model was almost nonexistent.

Active Filters

When the transistorized NIC became available, there was interest in developing the device for service other than just cancelling losses in transmission lines. The NIC soon became a component in general circuit synthesis and was used in most early RC active filters. The NIC was the "active" element.

The NIC is also a most convenient way of making a good negative resistance device. This technique is especially useful in cancelling inductor losses in filters. Considerable work is being done at Lenkurt in this area.

The negative resistance device can be used to decrease the effect of resistance in a filter inductor, thereby increasing the Q of the circuit. Q is a convenient way to express the merit of an inductor and is derived by dividing the inductor reactance by the equivalent series resistance.

In the design of conventional LC filters, the quality of performance achievable over the pass band is limited primarily by the dissipation in the inductors. There are three ways commonly used to reduce this dissipation. The classical method is to place an equalizer network in tandem with the filter to correct the unequal losses across the frequency band concerned. As a result, the overall pass-band loss of the filter is at least 0.5 dB greater than the maximum loss in the filter without the equalizer.

Another method is to predistort the original filter design in such a manner that the pass-band loss characteristic with dissipation is close to the desired flat response. But from a technical point of view a third solution is almost ideal: the cancellation of dissipation using negative resistance.

Lenkurt Applications

A channel filter being investigated at Lenkurt involves this technique and could result in a dramatic sharpening of the edge of the pass band. Using seven ferrite inductors and five negative resistance devices, a very high Q is expected – perhaps 8,000 – with the band flat to within 0.1 dB.

The negative resistance circuit suggested for this use consists of two transistors and five resistors. The circuit is shown in Figure 8, along with the application of the circuit in a conventional filter.

Figure 7. Lenkurt-built negative resistance (left) for channel filter studies compared to Centralab package and Sylvania IC version. About actual size.

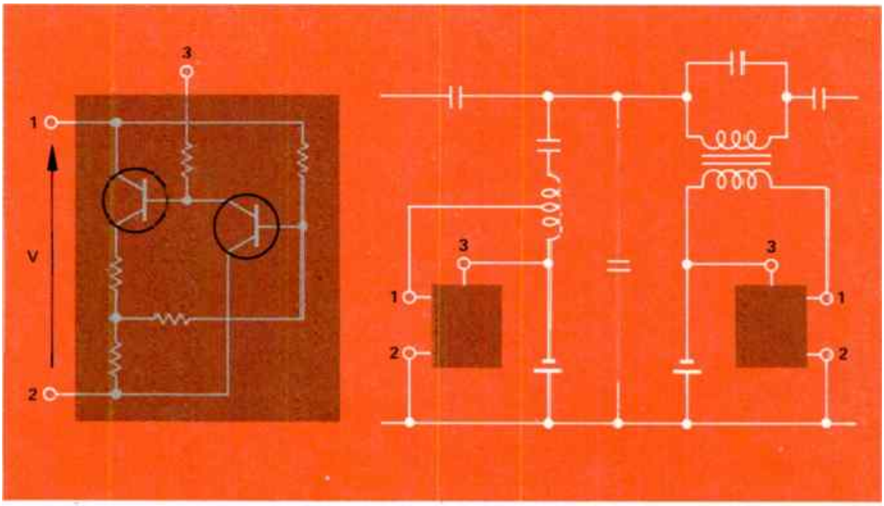


Figure 8. Negative resistance circuit (left) and how it is coupled to the filter network. Very high Q is expected from reduced dissipation in the inductors.

Negative resistance is being used to sharpen band elimination characteristics for critical data circuits. In the Lenkurt 25A data transmission equipment used by the news wire services, a special splitting filter is necessary for switching out portions of the band at various locations across the country. Engineers have said that it would have been impossible to achieve the high degree of sharpness necessary without using negative resistance techniques.

From reversing the sign of resistance (R to −R) and impedance (Z to −Z), it was historically a short step to *inverting* the impedance (Z to 1/Z – or more correctly R^2/Z). Here, another important contribution to filter design will be provided through the introduction of a device known as the gyrator which can perform this impedance inversion. The gyrator can transform the impedance of a capacitor into that of an inductor and so literally make a capacitor look like an inductor to the rest of the circuit. The capacitor, of course, is much smaller than the inductor.

Again, while the theory has existed for many years, and some circuits have been built, more recent developments have moved the gyrator closer to practicality. If perfected, the gyrator and an associated capacitor could be substituted for each bulky inductor in conventional filter networks, eliminating a great deal of size and weight. But for the moment, this relative of the NIC may be too costly for a favorable economic tradeoff.

Negative resistance and negative impedance circuits have moved from curiosities to practical devices – especially augmented by the development of transistors. Many of these applications are significant and will continue to be of interest in the expansion and improvement of telecommunications.

The Lenkurt

NOVEMBER 1970

DEMODULATOR

crystal filters ○ part one

Improved carrier design has followed closely on the heels of filter development. Crystal filter design has recently improved quality and reduced size of carrier and multiplex systems.

Filters have played such an important role in carrier communications that it is worth discussing their history and development, with an emphasis on crystal filters. The technique of carrier and multiplex system design points out the filter requirements and the need for certain types of filters.

Analogies are often used to describe filters, but comparing electronic filters to such things as a screen, a strainer, or a piece of paper used to separate solids and liquids may be misleading. A filter placed in an alternating current transmission path permits the free passage of certain frequencies and blocks others. The degree to which the filter blocks (attenuates) the signal is a function of the filter design and is measured in decibels.

Resonance

The principle that allows a filter to discriminate between frequencies, passing some while rejecting others, is known as resonance. The reeds and strings of musical instruments exhibit this natural phenomenon of resonance — the desire to vibrate at a single frequency. Filter operation can be explained using the principle of resonance and the analogy of a child jumping rope. If the child is jumping at the same frequency that the rope is turning, the child can jump into, with, and out of the turning rope with little difficulty. If, on the other hand, the child is not jumping at the rope's frequency, there will be interference and the child will not be able to pass through the turning rope.

Perhaps this analogy belabors the point, but it is important to understand the relationship between resonance and filter operation. An ac signal at the resonant frequency of the filter will be passed through the filter and other frequencies will be attenuated. If this analogy were to hold exactly, the characteristic curve of such a filter would look like Figure 1.

Electronic filters, however, do not fit this single-frequency characteristic curve, but rather have varying degrees

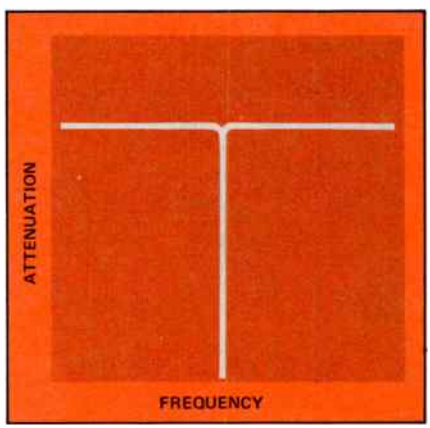


Figure 1. A filter that blocks all except its resonant frequency would have a characteristic curve as shown.

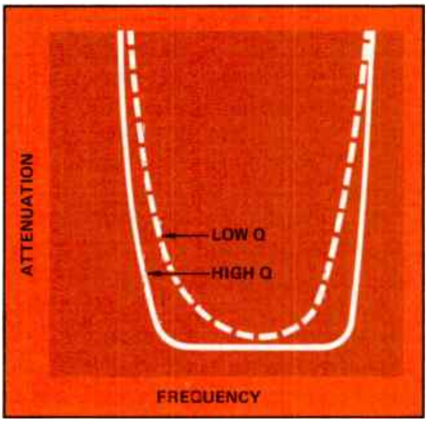


Figure 2. These curves show the passbands of the same filter made with resonators having different Q values. Note that high Q values must be used if the filter is to have a flat passband with sharp corners.

of sharpness, and actually pass a range of frequencies. This spreading of the pass region is a function of the filter resistance.

The term used in filter language to describe resonator sharpness is Q, which is inversely proportional to the resistance. High Q's indicate low resistance, high efficiency, and the ability of resonant circuits to obtain sharp, steep discrimination. Conversely, low Q's indicate a lack of sharpness. Figure 2 shows what happens to the passband response of a filter built with either low Q or high Q resonators. The actual Q value required will be determined by the bandwidth of the filter and also by its approximate center frequency. For example a bandpass filter having a 4-kHz passband will require lower resonator Q's to give the same flat passband if it is built at a center frequency of 10 kHz than if its center frequency is 10 MHz.

Bandpass Filters

An interesting point worth remembering is that not all resonant circuits are considered to be filters; however, all filters are combinations of resonant circuits. The art of designing filters involves using a number of resonant elements properly coupled together so that zero attenuation is approximated at all frequencies in the desired passband and maximum attenuation is achieved at all other frequencies. Figure 3 shows a typical LC filter as a series of LC resonators.

The low Q of early LC filters motivated Lenkurt, and others, to devel-

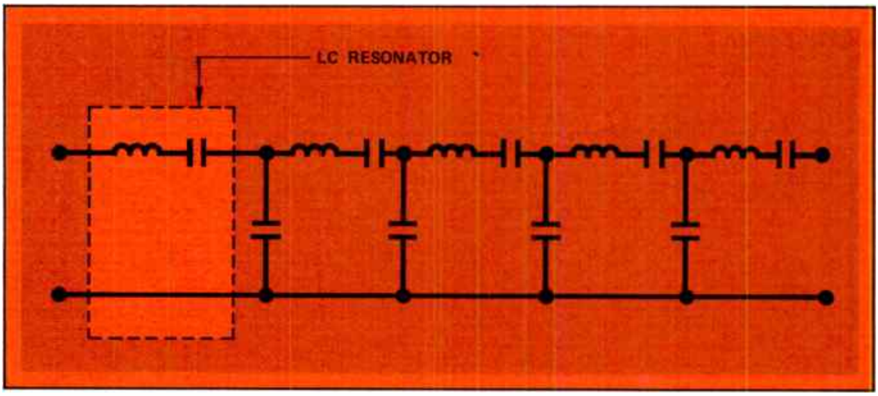


Figure 3. A typical LC filter is made up of a series of LC resonators.

op magnetic cores from powdered metals for low resistance inductors to improve LC filter sharpness. The rounded corners at the ends of the passband were eventually squared off by developing inductors with increasingly lower resistance ratings. It is still not possible to obtain stable inductors with Q's much greater than 500. Therefore, the center frequency of an LC filter for a 4-kHz wide voice channel is confined to values less than 100 kHz.

Early Crystal Filters

LC networks are not the only type filters capable of separating channels in carrier systems. Crystal filters are just as suitable and perform even better. Rather than using electrical components to form a resonant circuit, certain crystalline materials can be set in mechanical resonance by an electrical field. Such crystalline material is called piezoelectric and is such that mechanical vibrations can be excited in the crystal by ac signals.

A crystal resonator is simply a plate of piezoelectric material, usually quartz, with a metal electrode on each side of the plate. If these two electrodes are connected to an alternating current, the crystal will try to vibrate at the signal's frequency. And, if the crystal has been cut to the proper dimensions, it will be set in resonance by the signal.

The passband for crystal filters has steep sides and square corners. This sharpness of a crystal filter is the result of the higher Q's of the crystal resonators that comprise the filter. Q, the ratio of reactance to resistance, can be increased by lowering the resistance, as was done with early LC resonators, or by raising the effective reactance. A crystal resonator with the same resonant frequency as an LC resonator has essentially the same inherent resistance. But the effective inductance and capacitance of the crystal resonator differ in such a way that the reactance of the crystal resonator is much larger – resulting in a larger Q. Figure 4 shows typical values for an LC resonator and a crystal resonator designed for the same resonant frequency.

Crystal filters appear to be far superior to LC filters, yet crystal

	LC resonator	Crystal resonator
Frequency	10 MHz	10 MHz
Inductance	.01 mH	9.2 mH
Capacitance	25 pF	.028 pF
Resistance	4 ohms	3.8 ohms
Q	157	152,000

Figure 4. The table shows the equivalent values for an LC resonator and a crystal resonator operating in the same frequency range.

filters were not used in early carrier systems since performance is not the only criterion in filter selection. Since LC filters were improved to give satisfactory performance, the use of costly, bulky, high Q crystal filters could not be justified.

Changing Requirements

Cost and size reduction are being stressed in this age of miniaturization. Therefore, the high cost and large size of low-frequency crystal filters make them even more unsuitable for miniaturized carrier equipment. But, for reasonable performance, at least at frequencies below several megahertz, inductors for LC filters also require a volume incompatible with present-day miniaturization.

The size of the quartz plates used in crystal resonators, and consequently, the filters themselves can be reduced by increasing the frequency at which the signal is filtered. The resonant frequency of a quartz plate is inversely proportional to the plate size; therefore, the smaller the plate, the higher the frequency. The upper limit on the resonant frequency of a crystal is determined by the smallest plate that can be easily handled and will not shatter at the resonant frequency.

From this it seemed logical that one way to reduce the size of crystal filters was to design a carrier system that filters higher frequencies – designing the carrier system around the filter. A small, inexpensive, high-frequency crystal filter had to be developed before the new carrier system could be designed.

Crystal size is the limiting factor in resonant frequencies but the mode of vibration is also a determining factor. The vibration mode – whether flexure, extensional, face shear, or thickness shear – is determined by the crystal cut.

High-Frequency Filters

The cut considered for high-frequency crystal filters is the AT cut because it exhibits good frequency stability, as shown in Figure 5. The cut of the crystal determines the angle at which the plate (or wafer) is cut from the block of raw crystalline material and has nothing to do with the size of the crystal. Therefore, an AT cut crystal can come in a variety of sizes and consequently, a variety of resonant frequencies.

High-frequency AT cut crystals vibrate in a thickness shear mode. In this mode the surfaces under the electrodes move in opposite directions with both surfaces remaining parallel. Figure 6 illustrates the thickness shear mode of vibration.

In the thickness shear mode of vibration other modes and overtones of the fundamental resonant frequency can also be excited. In this case, experience has shown that rather than eliminating these unwanted modes and resonances, reduction of their intensity will accomplish the same results.

Energy Trapping

As early as 1946, methods of eliminating unwanted modes of vibration were being studied. W. S. Mortley published an article in 1946 on his experimental observations of what has since become known as energy trapping. Mortley's theory was given little attention when it was first published. It was not until the early 1960's that his original work began to have relevance and was given the proper attention in crystal filter design.

Prior to Mortley's experimentation, unwanted modes of vibration were eliminated by thinning the crystal plate toward the edges. This had the effect of reducing the intensity of all modes except the desired thickness

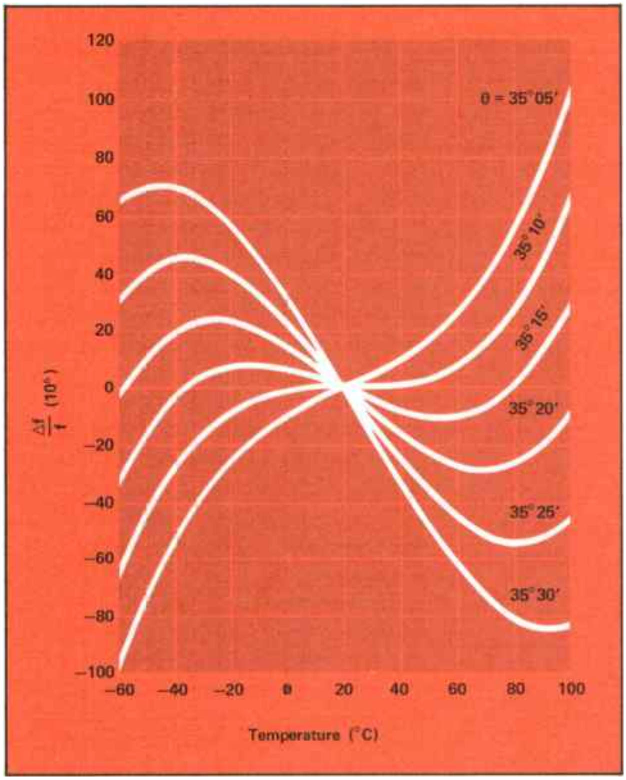


Figure 5. An AT cut crystal with an angle of 35° 10' has the best frequency stability.

shear mode. By reducing these unwanted modes, interference between them was avoided. This technique of thinning the crystal is known as shaping and works well in crystal resonators where there is an air-gap between the crystal plate and the electrodes. Mortley noticed, however, that shaping was not a successful technique for eliminating unwanted modes of vibration in crystals with deposited metal-film electrodes.

The reappearance of unwanted modes with deposited metal-film electrodes led Mortley to conduct some further experiments. From these he discovered that the intensity of some unwanted modes and also their frequency spacing with respect to the desired mode could be controlled by the mass of metal in the film and the shape of the electrodes.

From these observations it was suggested that the same laws that apply to the propagation of electromagnetic waves apply equally well to the transmission of mechanical shear waves in quartz.

The comparison of mechanical shear waves in quartz to electromagnetic wave propagation was carried even further. The filter can be compared to a waveguide. Just as it is possible to have a high coefficient of reflection from the junction of waveguides differing slightly in dimensions, quartz crystal also reflects waves where the crystal changes shape. Therefore, by putting the proper step in the crystal in the direction of wave

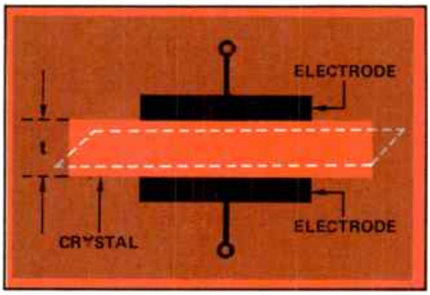


Figure 6. The dotted line shows the thickness shear mode of vibration in an AT cut crystal.

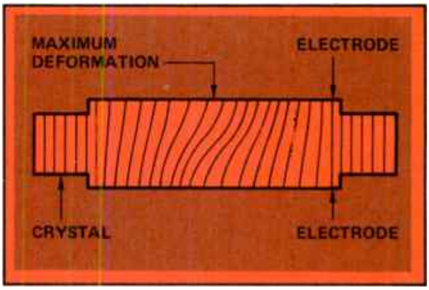


Figure 7. The shear waves are trapped in the thicker area under the electrodes.

propagation it is possible to trap the desired modes of vibration. Figure 7 shows the effect of dimensional changes on wave propagation.

For high-frequency filters, where the crystal is small and fragile, it is not practical to machine steps in the quartz. Crystal filters with metal-film electrodes deposited on a flat crystal plate have an inherent step. The metal-film on the outer surface of the crystal adds to the vibrating mass of the quartz plate, trapping the energy under the electrodes, because the boundary of the electrode acts like a dimensional step in the quartz plate. Those unwanted modes and resonances travel into the thinner unplated areas where they decay exponentially with distance.

Although Mortley had proposed this theory of energy trapping, the actual control of unwanted modes was more of an art than a science for many years after Mortley's paper. Bechmann in 1961 published a set of experimentally derived rules which give optimum dimensional ratios for the crystal plate and the electrodes to suppress the unwanted modes. In 1963 Shockley, Curran, and Koneval formulated their theory of energy trapping which was essentially the same as Mortley's and which confirmed Bechmann's experimentally derived numbers.

New Horizons

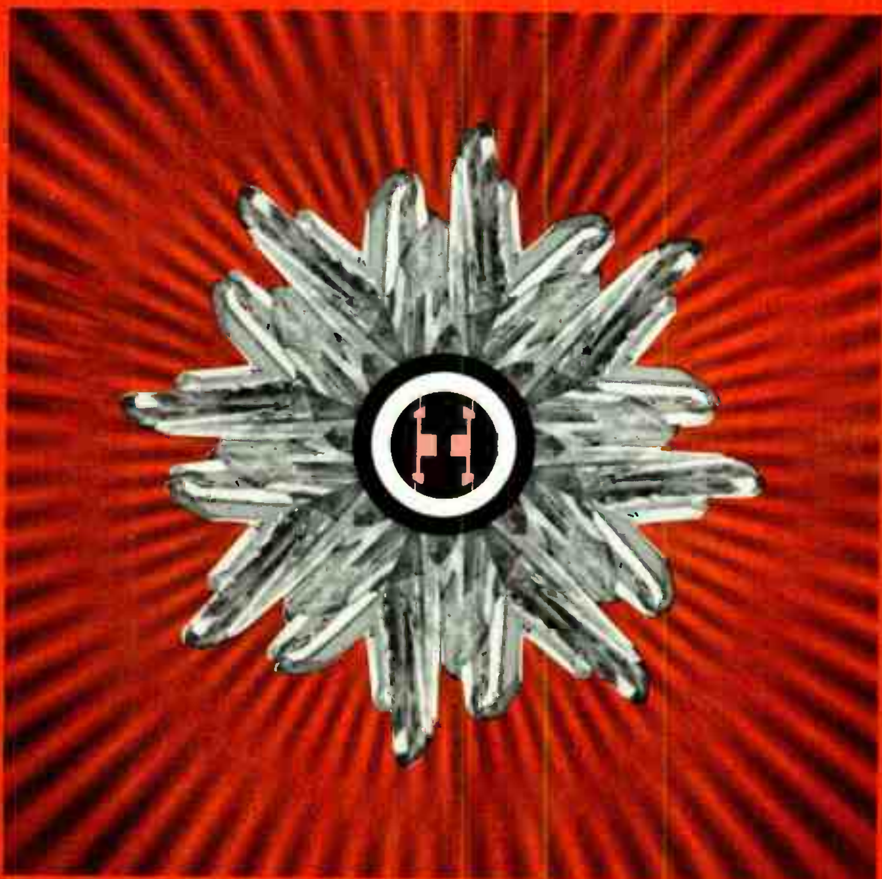
The theory of energy trapping has opened up a whole new area for high-frequency crystal filter design. This theory has led to the design of filters where multiple resonators are placed on a single quartz plate.

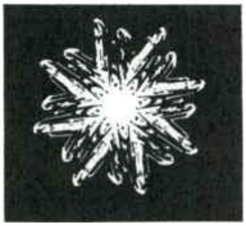
Crystal filter techniques have come a long way from the first high-cost, low-frequency, bulky designs rejected for carrier channel usage. Today multi-resonator crystal filters are responsible for reduced size and cost of improved carrier systems. These new multi-resonator crystal filters and their effect on carrier and multiplex system design will be the subject of the next DEMODULATOR.

The Lenkurt DEMODULATOR

DECEMBER 1970

crystal filters ○ part two





Multi-electrode crystal resonators have substantially reduced the size and cost of crystal filters and offer improved design of multiplex systems.

In multiplex systems using LC channel filters, as much as 75 percent of the volume is taken up by the filters. Size reduction and improved performance are possible with the new technology of high-frequency crystal filters. The advent of filter size-reduction served as the impetus to further miniaturization such as the introduction of integrated circuits to produce an even smaller, more economical multiplex package.

Typical high-frequency crystal filters have as much as a 20:1 size reduction compared with LC filters for similar operations. In addition to the size reduction, crystal filters operate with greater stability over a wider range of temperatures than LC filters. Whereas LC channel filters were designed with a center frequency usually below 100 kHz, the optimum center frequency for low cost crystal filters is about 8 MHz. Multiplex systems can be designed to take advantage of the small size and high stability of these high-frequency crystal filters.

Coupled Crystal Resonators

High-frequency crystal resonators use AT-cut quartz wafers that have metal-film electrodes plated on the surface of the wafer (see Figure 1). By applying the correct frequency signal to such a crystal resonator, the crystal will mechanically vibrate and electrically behave like a resonant circuit. By coupling several such resonators together, it is possible to build a filter that has a suitable frequency response for operation as a bandpass filter. When each wafer contains a single electrode pair, the resonators are coupled electrically with shunt capacitors to form crystal filters.

Energy trapping is used to minimize the unwanted modes of vibration in high-frequency crystal resonators with plated metal-film electrodes. When the mass of the electrodes is large enough, the desired mechanical vibrations are

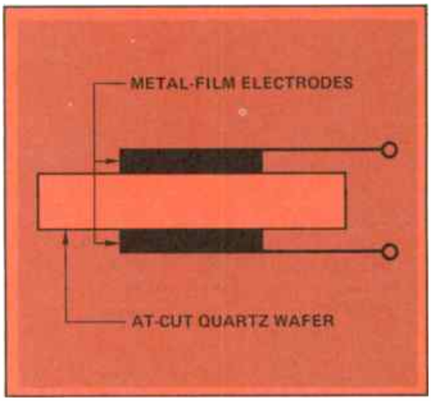


Figure 1. High-frequency crystal resonators have metal-film electrodes plated on the surface of AT-cut quartz crystal wafers.

confined primarily to the area under the resonator electrodes.

It is possible, however, to "let" the energy out into the unplated areas of the crystal wafer in a controlled manner by adjusting the amount of metal in the plated electrodes. The vibrations that do get into the unplated areas are vibrating at the resonant frequency of the crystal but the amplitude of the vibration decays exponentially with distance. It was discovered that these untrapped vibrations could be put to work rather than letting them decay. By placing another identical electrode pair close to the first one it is possible to set the second pair in resonance by these untrapped vibrations, since the resonant frequency of the second pair is the same as the first.

Electrode pairs are acoustically coupled when the resonant vibrations of one pair set another pair in resonance using the crystal as the only connection between them. Consequently, the crystal areas plated with metal-film electrodes form mechanically resonant systems, coupled by the transmission of energy through the unplated quartz wafer.

The principle of acoustical or mechanical coupling can be illustrated by common materials such as a thick sheet of foam rubber and two identical metal blocks resting on top of the foam. Application of an alternating force to the top of one block will set it in vibration and if the other block is close enough, it too will vibrate because the foam rubber is capable of transmitting the vibrations. The blocks must be close together to be coupled because most of the energy is trapped under the vibrating block and the little that does escape is quickly attenuated by the foam rubber. Elaborating on the mechanical coupling, a mechanical model can be developed for acoustically coupled crystal resonators. Figure 2 shows such a mechanical model.

The amount of coupling between adjacent resonators depends upon the dimensions of the resonators, the thickness of the metal electrodes, and the spacing between resonators. Consequently, by changing these three vari-

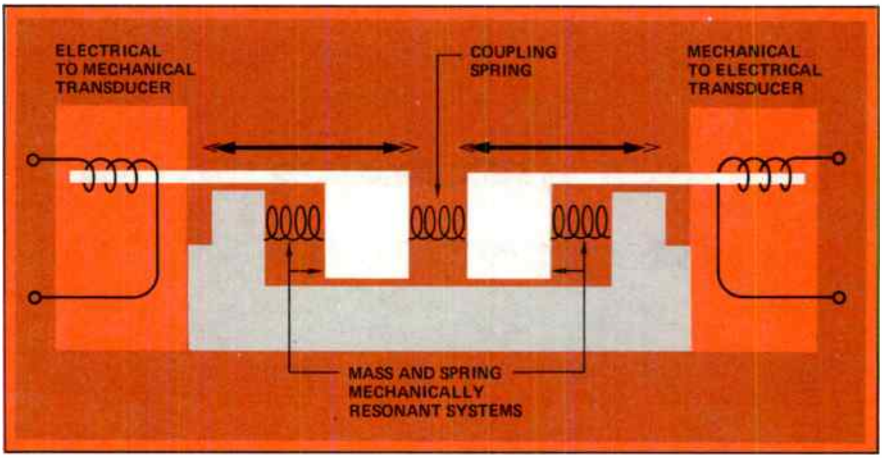


Figure 2. A mechanical model can be used to explain acoustically coupled crystal resonators.

ables it is possible to control the degree of acoustical coupling.

Multi-Electrode Resonators

When all the resonators of a bandpass crystal filter are put on a single quartz wafer, the filter is referred to as monolithic. Figure 3 shows a monolithic crystal filter. It was the discovery that individual crystal resonators could be acoustically coupled rather than just electrically coupled that has led to the substantial reduction in both size and cost of crystal filters. As many as ten resonators have been placed on a single crystal wafer. With the proper arrangement of the electrode areas, it is possible to make relatively complex filters in a monolithic form. The center frequency of such monolithic crystal filters is in the range from 5 - 150 MHz with passbands ranging from 0.001 - 0.1 percent of the center frequency. Figure 4 shows the improved performance of each additional stage of a monolithic crystal filter.

Although it is possible to place as many as ten resonators on a single crystal wafer, it is sometimes more reasonable to group several simpler multi-electrode resonators together. In this way the filter uses a combination of acoustical and electrical coupling between the individual resonators. Each of the simple monolithic structures are electrically coupled with a shunt capacitor. Such a filter, combining the advantages of both structures, is referred to as a polylithic crystal filter.

Lenkurt uses polylithic filters for its new 36A multiplex system. These filters have a center frequency above 8 MHz. Figure 5 compares the size of one of Lenkurt's polylithic filters, which occupies only one cubic inch, with a convential LC filter.

Computer-Aided Design

Computers are often used in the design of LC bandpass filters. The same technique is being used in the design of crystal filters. With the proper computer program and the desired center frequency and bandwidth the necessary component values for an LC resonator can be calculated.

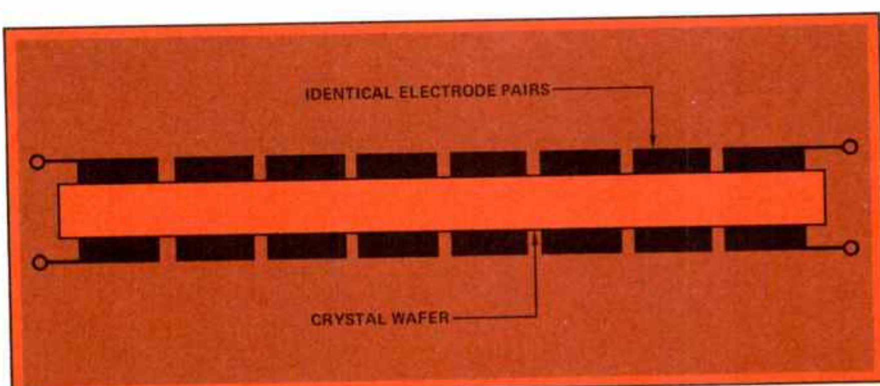


Figure 3. A monolithic crystal filter is a single crystal wafer with acoustically coupled resonators formed by plating identical electrode pairs on the crystal surface. Crystal filters are bi-directional; therefore, either end may be used as input or output.

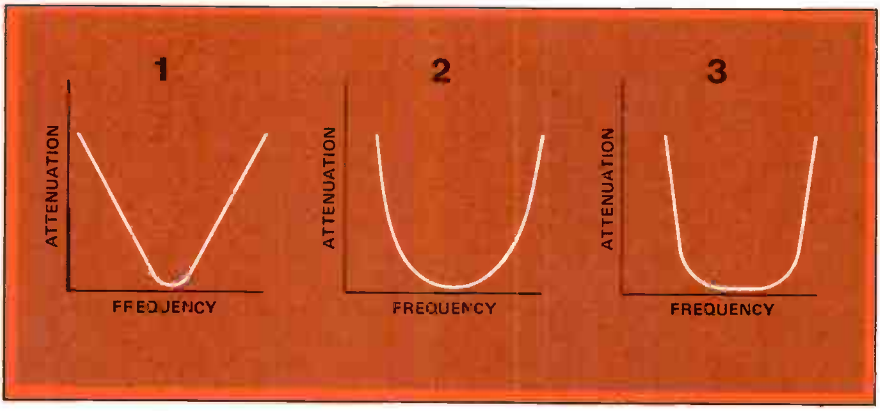


Figure 4. Each successive stage of a monolithic crystal filter improves the performance. The curves shown illustrate the successive improvement for three stages of a monolithic filter.

Figure 5. High frequency crystal filters, such as Lenkurt's polylithic filters, are reduced in size compared with LC filters.

In designing a crystal filter for a particular resonant frequency, the desired information from the computer is the crystal size and shape and the dimensions of the electrodes. For multi-electrode crystal resonators, the spacing of the electrodes is also necessary. Using established computer programs, it is possible to calculate the component values for an equivalent LC resonator and from these values, the crystal filter requirements.

The circuit shown in Figure 6 illustrates the four electrical components needed to produce an electrical resonator equivalent to a two resonator crystal. Knowing these four component values, the dimensions of the electrodes and their spacing can be calculated. The optimum dimensions of an AT-cut crystal wafer are a function of the coefficients of elasticity of the crystal material used. The thickness of the crystal wafer is a function of the desired center frequency, since the resonant frequency depends solely on the wafer's thickness. The vibrations of the resonator never reach the edges of the plate because they are trapped under the electrodes; therefore, the lateral dimensions cannot affect the resonant frequency.

Applications

The use of a two-step modulation scheme is one change in multiplex systems using high-frequency crystal filters. This new modulation technique translates the voice channel to an intermediate frequency (approximately 8 MHz) before filtering. In the second modulation step the filtered channel is translated to its appropriate frequency allocation for transmission.

Sophisticated mechanical design is also necessary when operating with radio frequencies. In the 8 MHz range, electromagnetic energy is being radiated by the wires and components used in the system. To minimize the possibility of electrical coupling between components due to radiated energy, sections of a high-frequency system are shielded from each other using metal plates and enclosures. Likewise the circuit path lengths used at this frequency are kept short to minimize the radiating and absorbing (transmitting and receiving) surfaces. Coaxial cable with its built-in shield, can be used where it is necessary to have long connecting wires and where external shielding is impractical. Figure 7 shows a multiplex channel unit from Lenkurt's 36A system which uses

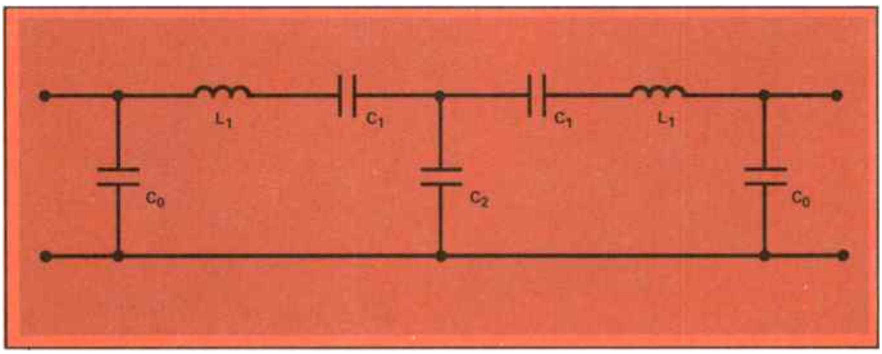


Figure 6. By knowing the values of the four electrical components illustrated, the equivalent two resonator crystal can be designed.

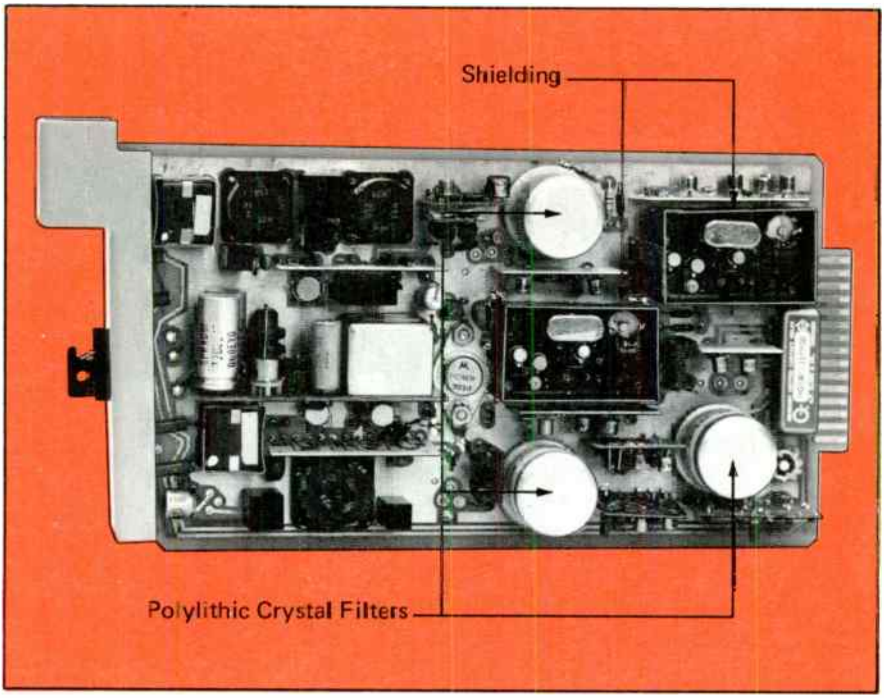


Figure 7. The channel unit for Lenkurt's 36A multiplex system illustrates the shielded, compact design used in high-frequency systems.

three polylithic crystal filters operating in the 8 MHz range. Such a unit is compact and shields the high-frequency sections.

Two polylithic crystal filters in Lenkurt's 36A system are used as channel bandpass filters – one for transmit and one for receive. The third is a narrowband carrier selection filter used to select the channel carrier from the multi-channel received signal. The recovered carrier is used at the receiver to demodulate the voice signal and to operate the signaling relay. The small size of these filters has permitted the placing of a complete 36A channel unit on a single printed circuit card.

What Next?

The development of monolithic and polylithic crystal filters with their stable, narrowband, high-frequency operation and the added advantages of simplicity, small size, and economy, has done a lot to push crystal filters closer toward their introduction into broader fields than the field of highly complex technical instrumentation.

INDEX TO VOLUME 1 AND VOLUME 2