Data Transmission Over Telephone Circuits

BY [redacted]

(b)(3)-P.L. 86-36

Unclassified

Many of the higher-speed cryptographic machines developed at NSA are required to operate over telephone channels. Because the telephone plant was not designed for digital devices, data transmission equipments are often used as intermediaries between these machines and the transmission medium. This article reviews the fundamentals of digital signals, describes the problems involved in transmitting digital data over telephone circuits, and presents some of the data-transmission techniques employed in COMSEC equipments.

INTRODUCTION

Digital data in binary form is perhaps the simplest and most rugged language for electronic machines from the standpoint both of processing and of relaying information. Samuel Morse first used it in 1832 with the hand telegraph, and it has since been extended to machine telegraphy, the dial pulses in telephony, and more recently to digital computers, telemetry, and secure communication systems. Since Morse slowly hand-keyed "What God hath wrought . . ." over a forty-mile wire line, high-speed data has spanned continents and oceans by radio and telephone. Yet basically little has changed in the interim except the "keying" rate and the communication facilities.

The problems involved in transmitting electrical signals over telephone circuits for any distance are essentially these: the channel favors certain of the various frequency components which in combination constitute a signal in such a way that some arrive at the destination sooner than others, and all are not equally attenuated; and the signal may be further corrupted by noise picked up in transit. These difficulties affect voice and digital signals alike, but the ear is oblivious to the delay variations and tolerant to some extent of the other two. With high-speed data, however, all these factors--but especially delay distortion—have to be carefully considered.

GENERALIZED DIGITAL COMMUNICATION SYSTEM

A generalized digital communication system is shown in Fig. 1. The "raw" information fed to the input processor could conceivably be temperature fluctuations or speech. These are converted by the input processor into a series of discrete pulses, arbitrarily called the "message." The transmitter portion of the transmission equipment suit-

(b)(3)-P.L. 86-36

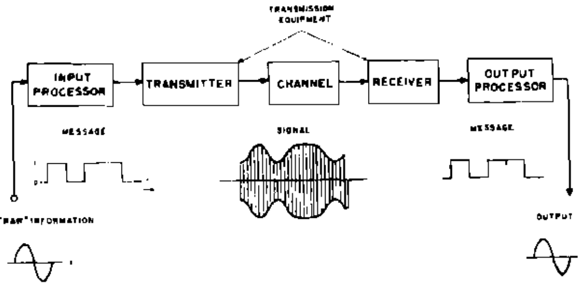


Fig. 1.

ably processes the message for transmission over the channel. Because of noise and the transmissive properties of the channel, the received message may not necessarily be that which was sent. The receiver portion of the equipment operates on the incoming signal in such a way as to minimize the disruptive effects of the channel and presents a replica of the message to the output processor. [1]

The two ends of the chain (the input and output processors) have been discussed in a previous article in the *Journal* [2]; we shall deal here with the properties of the parts between—the transmission equipment and the channel.

DIGITAL SIGNALS: FUNDAMENTAL CONSIDERATIONS

Digital signals represent information by a sequence in time of discrete symbols or signaling elements. Each signaling element is a choice from a finite set of alternatives. The simplest embodiment of a digital signal is one offering a choice between two possibilities for each signaling element: the binary signal. With a ternary signal there are three such possibilities; with a quaternary, four; and so on. If the transitions from one state to another only occur at prescribed time intervals the signal is *synchronous* or *time-quantized*. Signaling elements may be represented by the magnitude of a current pulse (telegraphy), the frequency of a sub-carrier frequency (Frequency-Shift-Key Telegraphy), the relative phase of a sub-carrier frequency (Collins Kineplex) [3], the position or duration of a pulse (Pulse Position or Pulse Duration Modulation) or a combination of the above.

One common message format, called *baseband*, is shown in Fig. 2a. It consists of a synchronous stream of binary elements of equal duration, with no intervals between them. In this case we say that the signaling element has a 100% duty factor; i. e., the product of the pulse

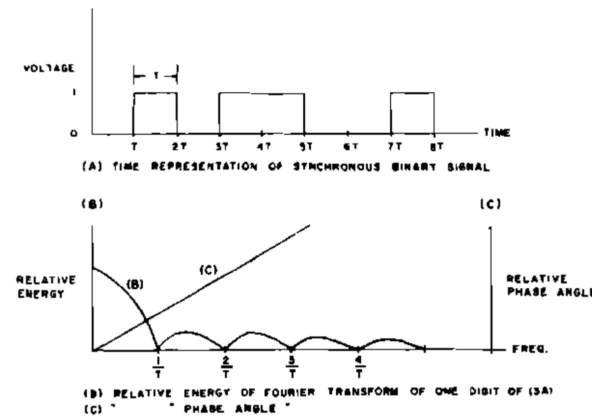


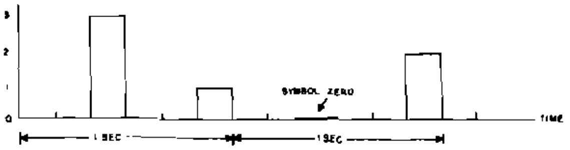
Fig. 2.

duration by the pulse repetition frequency (PRF) is unity. The stream shown below would be described as a synchronous quaternary signal of 33%-duty-factor pulses.

A standard measuring unit of signaling speed or modulation rate that will apply to any of the possible digital signals is sorely needed by the data community. For digital data the *baud*, borrowed from telegraphy, is probably the one most widely employed. It is defined as being equal to the reciprocal of the shortest signaling element's duration (in seconds). A speed of one baud is therefore one element per second, but one element per second is not necessarily one baud. The unit *bits/sec* is also often used, although, as defined by Shannon [1] it is a unit of information, measuring the *entropy* (uncertainty) of an information source or the capacity of a channel [4]. When denoting speed, bits/sec can be defined as the product of the number of binary choices or bits per symbol by the number of symbols per second. To illustrate an ambiguity that often arises when these units are employed indiscriminately, let us refer to the quaternary signal sketched over-leaf. Here the modulation rate is six baud, or four bits/sec —assuming all choices to be equally probable—, or two symbols/sec. Only in the baseband case are bauds, bits/sec, and symbols/sec synonymous.

From the examples, we note that speed expressed in baud may not always convey what a system is doing, but rather indicates its maximum capability and so can be directly related to the bandwidth re-

(b)(3)-P.L. 86-36



quirements of a channel—which is doubtless the reason why it is preferred by communicators. Bits/sec, on the other hand, if properly applied, is a truer measure of actual performance.

From now on, we shall consider only synchronous binary signals, and express speed in bauds. It is, of course, true that many data systems are not binary or even time-synchronous—for instance, the Start-Stop Teletype but the simplicity and the prevalence of the case chosen make it peculiarly suitable for elementary discussion.

In the foregoing, the message has been represented as a time function[1]; to evaluate the effects or requirements of the transmission medium, however, it becomes necessary to express the message characteristics in the parameters of the medium. These are generally given in terms of the circuit's steady-state loss and phase characteristics; i. e., how the telephone circuit attenuates sinusoidal signals of various frequencies and how it affects their transmission velocities. The former is called the amplitude or frequency-response of the circuit and the latter the delay or phase-response.

Through Fourier-transform calculus we can describe each signaling element of Fig. 2a. as the sum of an infinite number of sine waves of various periods ($2\pi/\omega_n$), amplitudes (A_n), and initial phases angles (ϕ_n). Or by the same token the entire message may be transformed into a series of sine waves.[2] With the signaling element so represented, we may readily compare its frequency and phase characteristic with that of the channel on a common basis of A_n and ϕ_n. A plot of the A_n and ϕ_n versus frequency obtained by transforming one digit of Fig. 2a—is shown in Fig. 2b. The figures indicate that the bandwidth or frequency space of this signaling element extends to infinity and its phase characteristic is linear. But what is the minimum bandwidth needed to accommodate baseband at a speed of $1/T$ bauds?

Nyquist has proved that the minimum bandwidth needed for distortionless transmission at $1/T$ bauds is $1/2T$ cycles per second. He further showed that the frequency spectrum of baseband, if partitioned into bands $1/2T$ wide centered at multiples of $1/T$, contained the same information [5]. From any one of these "fundamental bands" the original signal could be uniquely recovered; recovery from n fundamental bands ($n = 2, 3 \ldots$) only yields a more exact replica of the signaling element. This gave telegraphers of 1928 and the present-day data transmission engineers an important criterion as to the minimum bandwidth needed for distortionless transmission of digital signals.

[1] For Fig. 2a it is:

$$f(t) = H(t - T) - H(t - 2T) + H(t - 3T) - H(t - 5T) + H(t - 7t) - H(t - 8T)$$

Where $H(t - x) = \begin{cases} 0, & t < x \\ 1, & t \geq x \end{cases}$

[2] So that the message of Fig. 2a could also be written:

$$f(t) = A_0/2 + A_1 \sin(\omega_1 t + \phi_1) + A_2 \sin(2\omega_1 t + \phi_2) + \ldots A_n \sin(n\omega_1 t + \phi_n),$$

where: $\omega_1 = 2\pi/KT$, KT being the duration of the message,
A_n = amplitude of frequency $nf_1 = n\omega_1/2\pi$, $n = 1, 2, 3, \ldots$
ϕ_n = initial phase of frequency nf_1

TELEPHONE CIRCUITS AS DATA COMMUNICATION CHANNELS

Message-grade telephone circuits are currently being exploited almost exclusively for use as high-speed data channels: program circuits from 5 to 15 kc are available on a limited basis at premium tariffs, as are 48 kc, 240 kc, and video channels. We shall restrict ourselves, however, to the message-grade circuits, which consist of loaded and non-loaded 19, 22, 24, or 26 AWG cable, found in the local plant, a 4 kc channel in a cable carrier, and open wire lines—to name a few—combined in various and sundry proportions, plus the associated central-office equipment and line repeaters.

The majority of these facilities in the telephone plant have been in existence since long before the era of high-speed digital machines and, unfortunately, do not possess the characteristics needed for optimum data transmission. Modifications of the telephone plant that would improve its data performance would likewise help speech transmission, but would be economically unsound, considering the present preponderance of speech traffic and the very slight degree of improvement that would be achieved. For this reason data transmission equipment designers strive to pattern their systems to the existing plant. Considerable effort is now being expended in this area by the military establishment and by commercial firms. A much publicized data-transmission system is the Collins TE206 Data Transceiver, which is rated at 2400 bits per second over a normal 3 kc channel. This system is currently being field-tested on message-grade telephone circuits, but as of this writing no conclusive performance data are available.

The telephone-circuit channel characteristics that will be given primary attention in what follows are bandwidth, linear distortion, and noise; non-linear distortion and level control will already have been taken care of because of their importance in telephony.

(b)(3)-P.L. 86-36

Bandwidth Considerations

The bandwidth requirements of a digital signal were shown to be dependent on the modulation rate and on how faithful a replica of the signaling element was desired. To illustrate what happens when the bandwidth is reduced, let us consider the case of sending baseband at $1/T$ bauds over a channel $1/2T$ cycles in bandwidth. The channel,

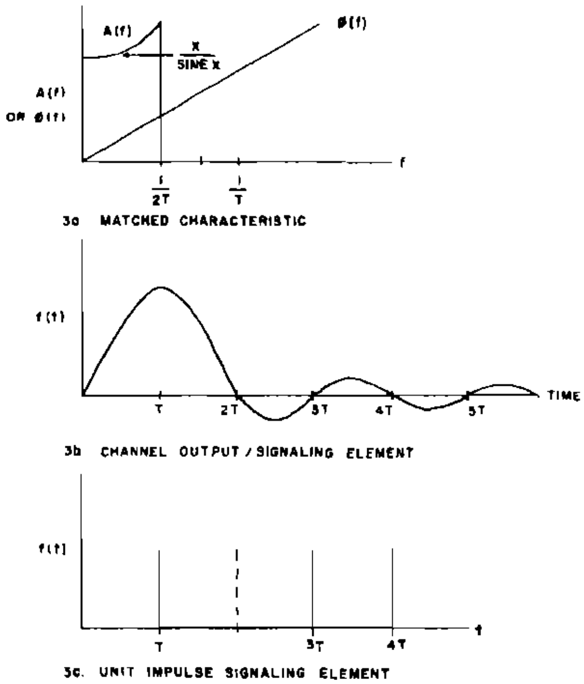


Fig. 3.

shown in Fig. 3a., is ideal, in that its frequency and phase response have been tailored to this signaling element. The channel output, per signaling element, has the form $(1/T)(\sin \pi t/T)(\pi t/T)$ and, as shown in Fig. 3b, is zero at integral multiples of T; successive inputs can occur only at integral multiples of T seconds— since the signal is synchronous— and will yield identical outputs. The received message will then consist of the sum of such waveforms.

If the channel output is examined or synchronously sampled at intervals of T seconds, the effect of all preceding inputs at the sampling instant for the pulse being considered is zero; furthermore, the signal is at its peak value, thereby maximizing the signal-to-noise ratio. Transmission at this rate over a channel of less bandwidth would result in intersymbol interference and thereby reduce the signal-to-noise ratio.

The choice of a different type of signaling element, e. g., a unit-impulse as shown in Fig. 3c, would not alter the output signal if the amplitude response of the channel were rectangular instead of skewed.

This constitutes an elementary proof that distortionless transmission in the Nyquist sense can be attained by signaling over a channel, matched to the signaling element, whose bandwidth is one-half the modulation rate. In practice, the characteristics of the ideal channel do not exist, and modulation rates less than one-half of the Nyquist maximum for a given channel bandwidth are used.

Linear Distortion Considerations

"Linear distortion" refers to both the delay and the amplitude distortion caused by a transmission facility whose amplitude and phase characteristics are not compatible with those of the transmitted signal, owing to imperfect filters and equalizers or to the fact that the effective bandwidth of the telephone circuit itself decreases with line length. Such departures cause the various frequency components of the signal to be unequally attenuated, and also to undergo unequal transmission delays. The net effect at a receiving terminal is "echoes"; i. e., the amplitude and zero crossing of the received signal are altered, thereby hindering the decision process, especially in the presence of noise. The correlation between echoes and the steady-state transmission characteristics was first published by H. A. Wheeler, and the reader is referred to his work [6] for a detailed presentation of this subject.

One way of undoing linear distortion is by performing the inverse process by means of a handicapping system of amplitude and delay equalizers that appropriately attenuate and delay each frequency component of the signal. This insures that the received component parts correctly combine in both amplitude and phase into the desired signal, by making the over-all delay and loss characteristics flat. Equalization in effect matches the channel to a particular signal. Very often, however, this may not offer the most practical solution: for example, baseband transmission over long distance telephone circuits is seldom feasible, because most of its energy is concentrated at the low frequencies, just where the channel characteristics, shown in Fig. 4, are (b)(3)-P.L. 86-36 [redacted]

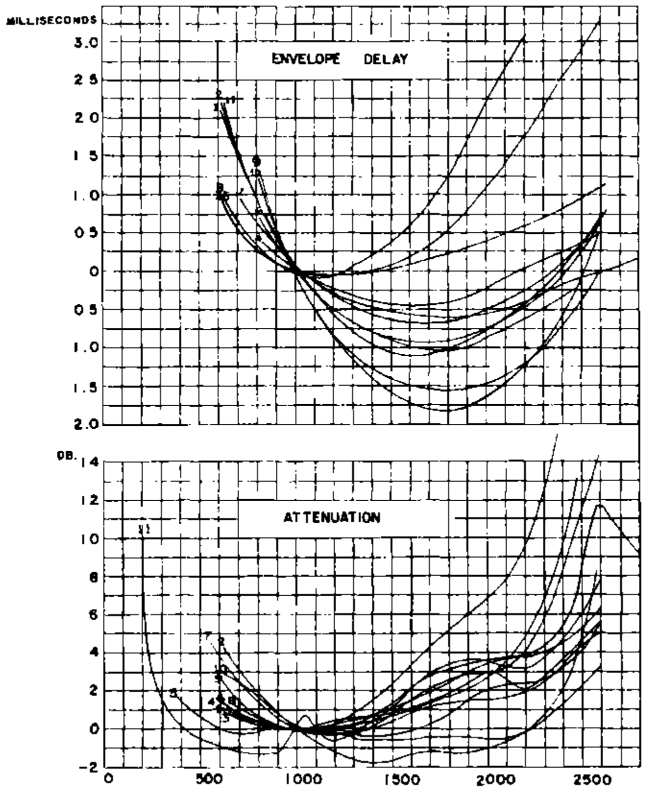


Fig. 4.

poorest. Equalizing below 700 cps, though possible, is impractical. But by frequency-translating a fundamental band of the signal to a more suitable position relative to the channel characteristic, using some modulation method,[3] we can readily overcome this difficulty. Even after modulation some equalization is generally required, but the amount is comparatively slight.

Noise Considerations

One definition of noise is that it is any extraneous fluctuation tending to interfere with the proper and easy perception of the signal. Noise in the over-all communications system comes from many different sources in varying degrees. The main types may be classed as *man-made*, *atmospheric*, and *thermal*.

Man-made noise covers a multitude of sins ranging from crosstalk to extraneous interference by auto ignitions. Atmospheric conditions, such as lightning and temperature-humidity variations, are nature's way of plaguing our efforts at communication by introducing noise impulses and diurnal changes into a circuit's frequency characteristics. Thermal noise is an inherent characteristic of all the components in the communications system from the wireline to the electronic hardware.

Man-made (dial pulses) and certain kinds of atmospheric noise share the qualities of abruptness and transience and are usually classed together as "impulse noise" by the communications engineer. Because of its nature, impulse noise defies mathematical analysis, so that most studies of it are empirical. If the signal spectrum it corrupts is narrow in comparison to the spectrum of the impulse, it resembles thermal or *white* noise in many respects. In any case, our attempt at providing an adequate defense against it leaves much to be desired. Thermal noise can be described, but in a probability sense only; for example, if all of the possible voltage levels of a thermal noise signal were plotted versus their relative frequency of occurrence, the resulting curve would be Gaussian with a mean value of zero, and the value of the second moment about the mean (variance) would be a measure of the noise power. In the case of synchronous binary signals, if the noise level exceeds half the peak-to-peak signal-level at the time of sampling, an error may result.

One obvious way to reduce the effects of noise and linear distortion introduced along the transmission path is to regenerate the signal before the damage becomes irreparable. In voice telephony, despite good repeater design, noise effects are cumulative, thereby limiting the number of tandem links, but synchronous binary signals can be

[3] A process which varies some parameter (amplitude, frequency, or phase) of a character frequency, in accordance with the information.

regenerated without error if the total perturbation never exceeds half the peak-to-peak signal amplitude; all that is required is a binary decision and a retransmission of a given waveform. Since the noise-level limits cannot be guaranteed, errors may occur despite regeneration, and will increase monotonically with the number of links in tandem. The error rate, however, will be considerably lower than in the nonregenerative case. But unless such regenerative repeaters could be provided by the telephone company they would be of little value.

Timing Considerations

As has been previously implied, timing is an important factor in the decision process at a regenerative repeater or receiver. This was illustrated previously in the example of binary signaling through a minimum bandwidth "matched channel;" in the case of linear distortion with the resulting echoes or "crossover jitter," accurate timing is even more essential because sampling margin is decreased. Timing plays an additional role in on-line additive-key cryptosystems because synchronous operation of remote key generators during the absence of incoming signals must be maintained. Both these considerations could be satisfied by extremely accurate clocks (timing sources) at the transmitter and the receiver, which after initial alignment would run synchronously for the cryptoperiod. Such devices are at present impractical because of their cost and size. Other schemes, which make use of the received message, have been successfully employed in COMSEC equipments. Three of these will be discussed.

One method used to maintain synchronism, called *time-base recovery*, derives the receiver timing by differentiating, rectifying, and filtering the received signaling elements. This is possible since synchronous elements inherently possess timing information. Such a means of frequency recovery is simply instrumented, but has the disadvantage that signal failure or long periods of steady mark or space will defeat it.[4] Although the latter condition is very improbable in cryptographic transmission, the former is common in HF radio transmission during severe fading. The effectiveness of time-base recovery is dependent on the "memory time" or "Q" of the timing-recovery circuitry and the stability of the transmitter clock. Time-base recovery is used in ciphertext autokey systems which are self-synchronous cryptographically, and in additive-key systems where "crypto-synchrony" is established on a push-to-talk basis, so that long term frequency stability is not required.

[4] An even more direct and flexible means of time-base recovery is possible with a synchronous AM signaling element called "Dipulse" (described in the next section). Here the receiver timing can be derived directly by limiting the signal.

A second means of synchronization employs a local timing source in each repeater and the receiver. This consists of an accurate clock which is automatically phased or synchronized by the incoming signal. Recovery systems of this type can maintain system synchronism for relatively long periods in cases of signal loss, but are more expensive to instrument.

A third method transmits the timing information on a separate channel simultaneously with the normal output signal. Although it simplifies the receiver, it is costly from a transmission bandwidth and power standpoint. The last two methods are common in cases where synchrony must be maintained for the cryptoperiod.

TRANSMISSION TECHNIQUES

The data transmission techniques used in COMSEC equipments may be divided into two broad categories: time-division modulation (TDM) or serial methods and frequency-division modulation (FDM) or parallel methods. For most of the applications within the telephone plant, TDM suffices and is the choice on the basis of terminal-equipment size and economy, whereas FDM equipments are used for high frequency radio applications and are larger and more expensive [7].

TDM

Various time-division techniques are used for the transmission of binary data over telephone circuits, the selection being dependent on signaling speed and the particular transmission facility. Some of these preponderantly used in COMSEC systems, with speeds ranging from 1500 to 50,000 bauds, will be described below.

Baseband Transmission

As the name implies, this refers to direct transmission of binary pulses over the telephone circuit. Baseband transmission in the 25–50 kilobaud range is possible on non-loaded cable pairs at distances up to 20 miles without regeneration. Here the channel characteristics can be made substantially compatible with those of the baseband signal by the use, above 3 kc, of simple amplitude equalization to offset the slowly falling channel response. Although the channel greatly attenuates frequencies below 300 cps, the resulting linear distortion is negligible at these speeds, because such a small percentage of the signal energy is affected.

Synchronous Amplitude Modulation

A common method of synchronous amplitude modulation, by which the signal is frequency-translated to match the channel, utilized in COMSEC systems, is "dipulse"; its signaling element and spectrum are illustrated in Fig. 5. In this technique, the baseband signal amplitude

(b)(3)-P.L. 86-36

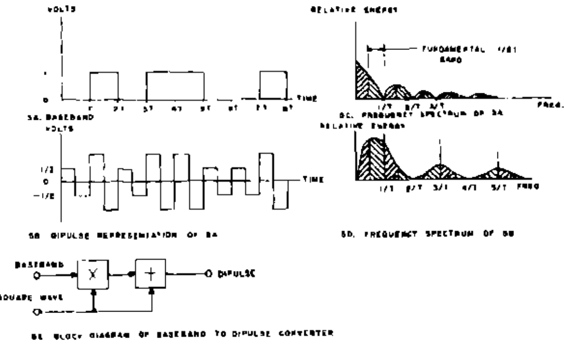


Fig. 5.

modulates a synchronous carrier whose frequency is equal to the modulation rate. At the receiver the baseband can be recovered by a conventional AM envelope detector, and since timing information is always present in the form of a synchronous carrier, it can easily be extracted, as noted earlier. Although this technique requires about twice as much channel bandwidth as the baseband case, its frequency translation properties make it desirable on circuits having adequate bandwidth but a poor low-frequency response.

Synchronous Phase Modulation

An embodiment of phase modulation termed "diphase" is also used when the low frequency response of the circuit is not suitable for baseband transmission, and offers somewhat the same advantage as dipulse in regard to frequency recovery. In diphase the basic signaling element is one cycle of a synchronous carrier frequency that is equal to the modulation rate, whose phase is varied 180° in accordance with mark-space information of the message. The salient features of diphase are illustrated in Fig. 6. Here we note that its spectrum resembles that of dipulse, except that the carrier is suppressed.

Baseband, dipulse, and diphase techniques have been successfully applied to COMSEC equipments operating up to 50 kilobauds.

Vestigial Sideband

Vestigial sideband transmission, proposed originally by Nyquist [5], offers the advantages of frequency translation in fitting the signal to the channel at a small (as compared to dipulse and diphase) bandwidth price, but at the expense of vulnerability to noise.[8] Although both dipulse and diphase signals are in a sense frequency-translated versions

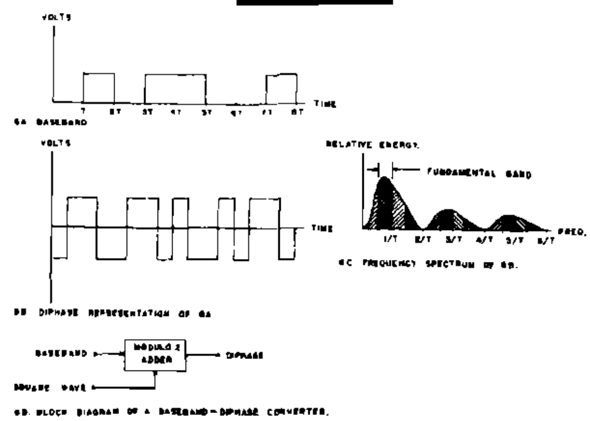


Fig. 6.

of the baseband signal, the amount of translation is governed by the signaling rate, and may not optimally match the signal to the channel. This is not the case with vestigial sideband.

A vestigial sideband signal consists of one sideband of a normal AM signal plus a modified version of the other, with the carrier frequency so selected that the signal is translated to the linear portion of the channel characteristic.[a] (To clarify the bandwidths implied: if a fundamental band requires one unit, AM will require two units but VSB only about 1.15.) The various stages of vestigial sideband generation are shown in Fig. 7. As indicated, the baseband signal is first pre-shaped to limit its frequency-spread before modulation. The baseband is recovered by conventional AM detection.

FDM

Frequency division modulation entails more than the frequency multiplexing used in carrier telephony and teletype. The latter is simply a device by which many messages are channelized or "stacked" in frequency for transmission purposes, whereas FDM is a method used to combat delay distortion encountered in HF radio and wire-line transmission of high-speed data. In FDM a serial binary stream is converted to many slower-speed parallel streams, which in turn are frequency-division multiplexed. If an input message at r bauds is

[a] The picture portion of a broadcast television signal is a common example of a vestigial-sideband transmission

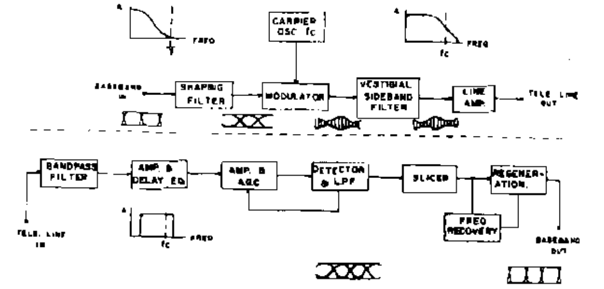


Fig. 7.

converted to n parallel streams, the modulation rate per transmission channel in FDM is then $R = r/n$ bauds. Since the modulation rate per channel is low if n is large, the effect of echoes due to linear distortion or multipath is less pronounced, but the narrow band channels require good frequency stability and level control in both the terminal equipment and the transmission facility in order to prevent crosstalk and intermodulation distortion.

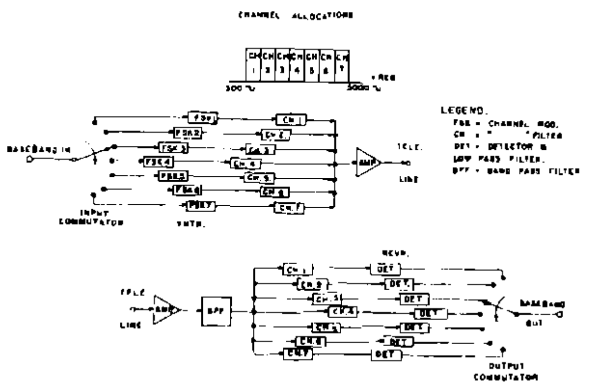


Fig. 8.

A typical FDM system is shown in Fig. 8. The message is time-demultiplexed by the input commutator, and each of the derived channels excites a frequency-shift oscillator in the voice frequency band such that a "mark" and a "space" correspond to oscillations of frequency f_{m_1} and f_{s_1} respectively (where the numerical subscripts denote the channel frequencies which are spaced throughout the voice band at intervals of 200 cps). The FSK outputs are linearly added for transmission, and the receiver performs the reverse process.

Frequency division modulation techniques have been successfully used in COMSEC systems at 1500 bauds over existing message-grade circuits in the telephone plant. For "attended" operation their salient advantages over TDM are decreased vulnerability to delay distortion and to noise, although this may be offset somewhat by crosstalk due to intermodulation distortion and system alignment.

CONCLUSIONS

The ideal data-transmission preparation equipment, neglecting considerations of cost, should be a flexible device that periodically scans the channel and selects the optimum transmission technique to match it. Until the advent of such a device, reliable data transmission over message-grade circuits on a random-call basis in the existing telephone plant appears to be limited to about 1000 bauds; higher-speed systems will require special circuits at higher tariffs. High capacity FDM transmission systems for use on voice circuits, though available, are not a panacea, because they are not readily compatible with serial-data processing systems, and are costly. Serial transmission techniques such as vestigial sideband, on the other hand, are economical and directly compatible with serial-data processing systems. The advent of low-cost high-speed data channels will certainly make them highly competitive.

In conclusion the author desires to acknowledge his debt to Mr. E. A. Enriquez, whose unfinished draft of an article on this same subject he has consulted extensively and repeatedly.

BIBLIOGRAPHY

[1] C. E. Shannon, "A Mathematical Theory of Communication," *Bell System Technical Journal*, July 1948.

[2] [redacted] "The ABC of Ciphony," *NSA Technical Journal*, July 1956.

[3] M. L. Doelz and E. T. Heald, "Binary Data Transmission Techniques for Linear Systems," *Proceedings of the IRE*, May 1957, p. 656.

[4] [redacted] "Some Notes on Information Theory," *NSA Technical Journal*, Vol. IV, No. 1 (this issue).

[5] H. Nyquist, "Certain Topics in Telegraph Transmission Theory," *American Institute of Electrical Engineers*, Vol. 47, p. 617, April 1928.

[6] H. A. Wheeler, "The Interpretation of Amplitude and Phase Distortion in Terms of Paired Echoes," *Proceedings of the IRE*, June 1939.

[7] [redacted] and [redacted] "Ionospheric Propagation," *NSA Technical Journal*, Vol. III, No. 4, October 1958.

[8] S. Goldman, *Frequency Analysis, Modulation, and Noise*, McGraw-Hill, 1948, p. 167.