NASr-21(02)

MEMORANDUM
RM-4997-NASA
SEPTEMBER 1966

MULTIPLE-ACCESS TECHNIQUES FOR COMMUNICATION SATELLITES: Digital Modulation, Time-Division Multiplexing, and Related Signal Processing

C. R. Lindholm

PREPARED FOR:
NATIONAL AERONAUTICS AND SPACE ADMINISTRATION

The RAND Corporation
SANTA MONICA • CALIFORNIA

NASr-21(02)

MEMORANDUM
RM-4997-NASA
SEPTEMBER 1966

MULTIPLE-ACCESS TECHNIQUES FOR COMMUNICATION SATELLITES: Digital Modulation, Time-Division Multiplexing, and Related Signal Processing

C. R. Lindholm

This research is sponsored by the National Aeronautics and Space Administration under Contract No. NASr-21. This report does not necessarily represent the views of the National Aeronautics and Space Administration.

The RAND Corporation
1700 MAIN ST. • SANTA MONICA • CALIFORNIA • 90406

September 1966 RB-4997

RM-4997-NASA, Multiple-Access Techniques for Communication Satellites: Digital Modulation, Time-Division Multiplexing, and Related Signal Processing, C. R. Lindholm, RAND Memorandum, September 1966, 101 pp.

PURPOSE: To provide background data for evaluation of digital techniques as a method of multiple access to a communication satellite.

FINDINGS: Basically, time-division multiple access is easy to achieve, but the present nondigital nature of most traffic makes analog methods less expensive. Time-division methods will be used when the balance changes to include more digital traffic. Amplitude and frequency modulation and phase-shift keying in digital transmission systems can be achieved using either coherent or non-coherent schemes. In the multiple-access application, however, it has not yet been determined that radio-frequency coherent systems are practical. Indications are that an all-coherent multiple-use system will be successful if the satellite furnishes a timing signal that can be used universally and if the feedback system is used to control timing. In time-division multiplexing, the only serious problems are deriving timing and controlling inter-symbol interference. Error correction codes are complex and expensive if used on a system basis; however, error correction on an individual channel basis is both practical and valuable. Error detection and retransmission are techniques that can be applied with excellent performance and at relatively low cost to non-real-time communication systems such as computer-to-computer links. Delta modulation as an analog-to-digital conversion technique is proving more efficient than pulse code modulation. It does not require the frame synchronization of pulse code modulation, it is less expensive, and it is well-suited to speech encoding. Companding techniques may be used externally to PCM or ΔM encoders and decoders, or the encoders can be designed to respond nonlinearly during the encoding and decoding processes.

BACKGROUND: This study is part of RAND's research for NASA on communication satellites. The digital techniques described in the study are presented in a context that will permit system analysis and comparison of digital techniques with other possible solutions to the multiple-access problem.

LV

PREFACE

In recent years RAND has been studying various aspects of the communication satellite multiple-access problem for the National Aeronautics and Space Administration. As part of that study, this Memorandum examines specific features of digital techniques: time-division multiplexing, error correction codes, analog-to-digital conversion methods, and synchronization problems. This information is presented in a context which will enable satellite communications systems engineers to design digital, or time-division, multiple-access systems. It will also permit system analysis and comparison of digital techniques with other possible solutions to the multiple-access problem. The Memorandum is intended as background material for further study of digital techniques.

SUMMARY

Methods of multiple access to a communication satellite have been broadly divided into three categories: frequency division, time division, and common spectrum. The second of these three, time division, employs techniques which are principally digital in nature. This Memorandum discusses those techniques which are digital or are related to time-division methods for performing multiple access.

Amplitude and frequency modulation and phase-shift keying are discussed with emphasis on performance parameters. Time-division multiplexing is reviewed, and some possible system difficulties are enumerated. Error correction codes are scrutinized from a systems viewpoint and found to be relatively poor when measured in decibel units. The reasons are examined, and applications for such codes are offered.

The two principal analog-to-digital and digital-to-analog conversion techniques--pulse code and delta modulation--are explained. The emphasis is again on the performance parameters characterizing system error rate or signal-to-noise ratio.

A short section discusses synchronization, listing the principal work done by this part of the system. Specific techniques for synchronization are not discussed since they relate so closely to the rest of the system design and vary greatly in complexity.

The main goal has been to offer elementary explanations of important digital techniques and to provide (graphically where possible) the relations needed to evaluate system performance. To this end, many tables and graphs are included.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

SYMBOLS

A	$\sqrt{E/n_o}$
a	amplitude of an RF sinusoid
B	bandwidth, Hz
c_i	i^{th} check bit (redundant bit)
E	signal energy for one bit = W·T
E[ ]	expected value of [ ]
e	number of errors correctable by a block code
f	frequency
i_i	i^{th} information bit
k	number of information bits
L	number of digits in PCM encoding of 2^L levels
M	number of channels being multiplexed
n	total number of bits in a binary block code
n_o	one-sided power spectral density, watts/Hz
P(error)	probability of a single bit error
P(word)	probability of a word error
P(x|y)	probability that x was received, given that y was transmitted
Prob( )	probability of event in parentheses
p	probability of error for a single bit of a block (short notation for P(error))
R	information rate, bits/sec
R_e	bit errors per unit time
r	rate of a code (a measure of redundancy) = k/n
S/N	signal-to-noise ratio
$(S/N)_{ch}$	channel signal-to-noise ratio
$(S/N)_e$	error induced signal-to-noise ratio

$(S/N)_Q$	quantization signal-to-noise ratio
$(S/N)_T$	total signal-to-noise ratio
T	time duration of one bit
t	time
U	threshold value for detection
V	output of an envelope detector
W	average power
Δ	quantization amplitude increment (also used with subscripts)
τ	a time duration
τ_{av}	mean time to undetected bit error
ω	$2\pi f$

ABBREVIATIONS

CC	control circuit
FDM	frequency-division multiplex
FM	frequency modulation
FSK	frequency-shift keying
NCFSK	noncoherent FSK
PCM	pulse code modulation
PM	phase modulation
PSK	phase-shift keying
TDM	time-division multiplex
ΔM	delta modulation

BLANK PAGE

I. INTRODUCTION

This Memorandum provides background theory and data for the design and evaluation of digital and time-division approaches to the communication satellite multiple-access problem. The study is not limited to discussion of any particular satellite design or configuration but deals with the broad field of digital communications. Interest in digital techniques is stimulated primarily by rapid improvement in the state-of-the-art of computer technology, which has reached a point where very complex processing may now be applied to communication systems at relatively low cost and with low space, weight, and power requirements. Thus, for many applications digital approaches may now offer advantages over analog schemes.

Error correction methods are discussed in the Memorandum. Although ten years ago they were a mathematical novelty and rather impractical to implement, today sophisticated codes are not only powerful but easily realizable. A quantitative measure of system performance for such codes is derived, and certain examples are evaluated.

A topic which is definitely related to this study is the class of random-access discrete address (RADA) communication systems.* Preliminary work has shown, however, that these systems are sufficiently complex to merit a study of their own, so they are not discussed here.

Digital techniques have been studied separately from analog techniques because of differences in basic language, in communications

---

*Sometimes referred to as spread spectrum and pulse address systems.

objectives, and in performance criteria. But it is recognized that in most applications it will be necessary to compare digital and analog systems for situations in which the traffic load is the same for both. Such a comparison is very useful in choosing the optimum system for a given application, and one goal of this Memorandum has therefore been the production of relationships which can yield the quantitative values needed in such a comparison.

No system comparison can take place without an analysis of the mixture of analog and digital traffic because the ultimate value of either system depends greatly on the nature of this mixture, i.e., on the ratio of analog traffic to digital traffic. It is appropriate, therefore, to stress the importance of digital data in the communications world. It is not possible to do more than speculate on the relative volumes of digital and analog traffic, but while both are increasing, the present rate of increase is much greater for digital. Certainly this is caused in part by the increasingly automated nature of today's world. In addition, security and privacy in transmission are more simply achieved through digital means. Though digital transmission is not currently used widely in commercial applications, there appears to be much demand for the protection it provides. This demand is due to the competitive nature of business and commerce and to an increasing desire for privacy on circuits in general.

II. DIGITAL TRANSMISSION SYSTEMS

INTRODUCTION

A digital transmission system is shown in Fig. 1, and its elements are identified and explained here. In most cases the source of a digital signal is binary. The data may be presented to the transmitter in either of two forms: serial (i.e., with one bit succeeding the previous one on a single input line) or parallel (with several bits presented on several input lines). The function of the encoder is to prepare these data in a manner appropriate to the subsequent demodulation process and to the channel characteristics.

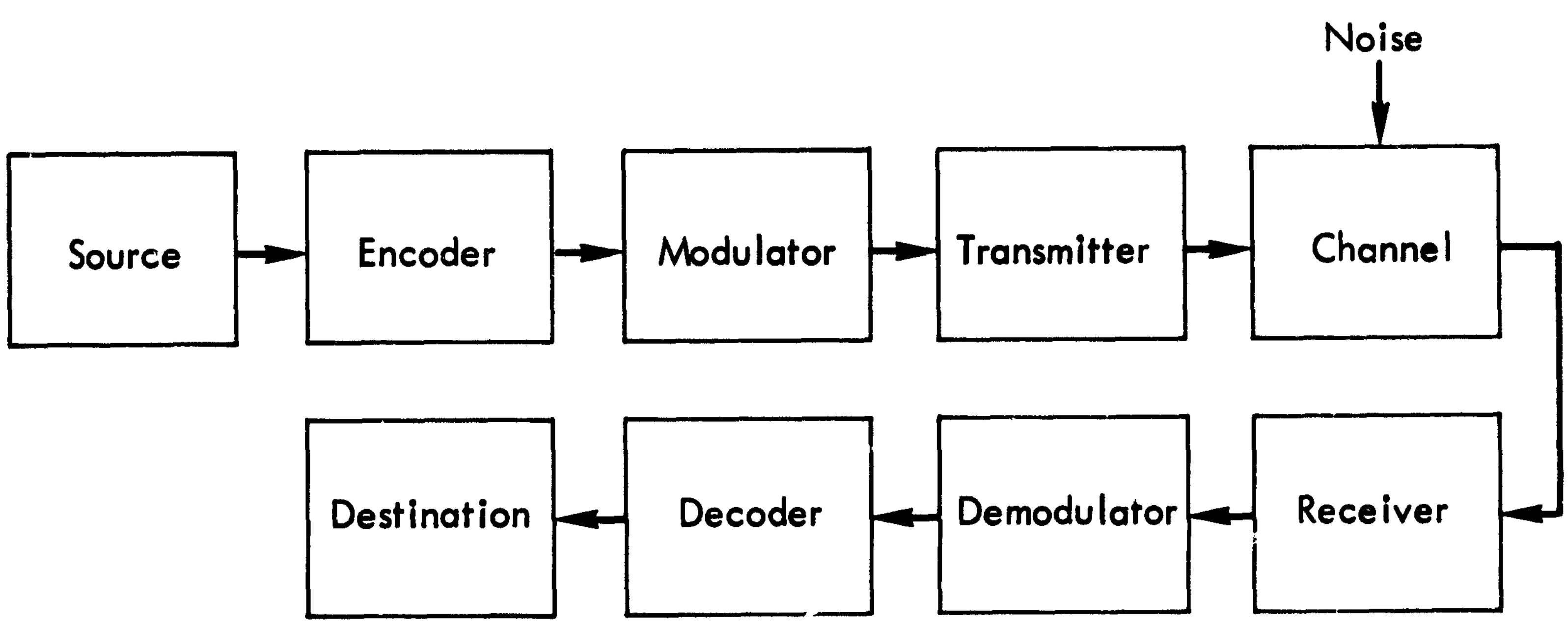


Fig.1—Digital transmission system

The modulator places the encoded digital information upon a carrier, and this information is amplified in the transmitter and sent over the channel to the receiver. Before reception, the transmitted signal is unavoidably corrupted by noise. Thus, the demodulator must attempt reconstruction of the signal and, because of the noise, will make errors which are acted upon by the decoder. The resulting signal is sent to the destination with errors as left by the decoder.

For example, suppose the source is a common business machine punched card. The data are carried by means of the areas in the card in which holes are or are not punched. There are many ways in which such a card can be transmitted. For example, the Hollerith card, which consists of 12 rows and 80 columns and is commonly known as an "IBM" card, can be scanned by rows, by columns, or by 960 sensors simultaneously. Scanning by rows or columns can be serial (sequential) or parallel (having 80 binary outputs at once). The card contents may then take any of the following forms:

- 960 serial bits on 1 line.
- 12 serial bits on 80 lines.
- 80 serial bits on 12 lines.
- 960 parallel bits on 960 lines.
- 27 serial bits on 36 lines (column binary format).

These can all be considered as various electrical encodings of one basic data source: the punched card.

The channel is characterized by its effect on transmitted digits. When the transmission is binary, noise may add to or subtract from the signal in such a way that transmitted "1's" may become "0's" or vice versa. When the probabilities of 1's becoming 0's and 0's becoming 1's are equal, the channel is called a "binary symmetric" channel. Several actual channels are so characterized in this report, and those modulations which have this property are specifically mentioned.

A theoretical advantage of the binary symmetric channel is that no knowledge of the a priori distribution of 1's and 0's is required to compute error probability. For nonsymmetric channels such knowledge is required.

FIDELITY CRITERIA FOR DIGITAL TRANSMISSION

The signal-to-noise ratio is a common measure of analog transmission system performance, but in digital transmission such a concept becomes less meaningful. More useful in practice is bit error probability. For independent errors, this one parameter is adequate to describe performance, while for nonindependent errors, such as occur on telephone lines and in certain VHF and UHF scatter transmission modes, where bursts of errors are prevalent, additional parameters are required to characterize the errors. For the communication satellite links of concern to this Memorandum, the statistical independence of contiguous errors is well verified, and one parameter suffices. Studies of digital reception techniques show that the important parameters characterizing digital reception in the presence of white gaussian noise,* and thus determining error probability, are:

- The detection technique used.
- The received signal energy per bit, E.
- The accompanying noise-power spectral density, n_o.

The ratio of the latter two, denoted by E/n_o, together with the modulation and demodulation technique being used, determines the error

*A stationary stochastic process having gaussian amplitude distribution and uniform spectral density.

probability. On fading links other parameters are also necessary, but as long as near-horizon paths are not used, such considerations will not affect satellite paths as they do overland microwave or tropo-scatter propagation. Thus, for each of the modulation and demodulation schemes considered, the relation between E/n_o (the ratio of input signal energy to noise spectral density) and p (bit error probability) will be described.

Other terms used to measure the fidelity of a digital circuit are:

- o Mean time between undetected bit errors, τ_{av}.
- o Bit errors per unit time, R_e.
- o Word error or character error probability, P(word).

The first of these is the relation between probability of undetected error and mean time to error, which is

$$\tau_{av} = \frac{1}{R \cdot p}$$

where R is the signalling rate in bits per second. The number of bit errors per unit time, denoted by R_e, is merely the reciprocal of the mean time to undetected error; hence,

$$R_e = R \cdot p$$

Word or character error probability refers to situations in which a group of bits occurs as a unit to denote a word or character. When all errors are statistically independent, the word or character error probability is found by summing the probabilities of those combinations of bit errors causing character or word errors. Let p denote the bit error probability, and let n bits compose a word or character; then the word or character error probability is

$$P(\text{word}) = \sum_{k=1}^{n} \binom{n}{k} p^k \left(1 - p\right)^{n-k}$$

where $\binom{n}{k}$ is the binomial coefficient. This is the probability of one or more errors in the n bits which make up the character and may be more simply calculated by subtracting from unity the probability of no bit errors:

$$P(\text{word}) = 1 - \left(1 - p\right)^n$$

Examples

o The common teletype code consists of 5 bits carrying the 32 teletype symbols and functions. Thus, 5 bits form a character, and a bit error probability of 10^{-4} results in a character error probability of $1 - \left(1 - 10^{-4}\right)^5 \approx 5 \cdot 10^{-4}$.

o Thirty-two bits form a "word" in many modern computers. For most applications, every bit must be correct. In a noisy communication system having a bit error probability of 10^{-3}, the probability of word error is $1 - \left(1 - 10^{-3}\right)^{32} \approx 0.0315$; that is, on the average 1 word in 30 is incorrect.

ENCODING

Any operation on digital data which alters its form for transmission over a practical channel is called encoding. The encoding of digital data may produce, for example, a signal which is analog in nature when the channel is well suited to analog signal transmission, or it may produce a high-rate digital stream including redundancy if the channel is suited to high-rate digital transmission. In the first example, the purpose is usually to reduce bandwidth when the channel is so restricted. In the second example, the addition of redundancy is for the purpose of enabling correction of errors in transmission. For practical purposes good encodings can eliminate errors in the transmission due to noise and even due to component failure if the system is appropriately designed. Thus, encoding is a signal processing technique of manifold application.

The simplest encoding is merely a transcription of the data into a binary electrical signal, forming a two-valued electrical equivalent of the data. In a unit time interval, one binary digit may be so represented, and successions of time intervals may represent a serial data stream. If the data are not binary, several bit intervals may be required to transmit the single nonbinary value. For example, decimal data require four binary digits. The next simplest encoding might be ternary or three-valued representations, the next quaternary, and so on. Such multivalued encodings are called "m-ary" encodings, and their transmission will be discussed later in the study. Most of the results and literature to date concern the binary (m = 2) case, and some of the m-ary systems have not been thoroughly explored. In the limit of large m, the electrical signal takes so many values that it may be considered analog, and in fact m-ary encodings are used because they are most appropriate to a channel of limited bandwidth.* An m-ary valued signal can represent $\log_2 m$ bits per unit time.

---

*An amplitude modulated signal can be constructed to occupy a bandwidth no greater than the information bandwidth of the modulation.

A discussion of encodings which introduce redundancy is presented in detail in Section IV.

DIGITAL MODULATION OF A CARRIER WAVE

The means available for placing digital signals upon a carrier are very similar to those available for analog signals. One distinction should be observed, however. With analog signals a sinusoidal carrier can be modulated in amplitude or angle, and in analog systems a strong similarity is noted between frequency and phase modulation. For example, the same RF waveform is obtained by frequency modulation (FM) with some signal or by phase modulation (PM) with the time integral of the same signal. For analog signals, the time integral of a signal is another similar signal, but with digital signals such a relation is not valid. So digital FM and PM are distinctly different in their formation, behavior, detection, error probabilities, etc. Thus, perhaps more so than with analog signals, digital signals can form three essentially different modulations of a carrier:

1. Amplitude or carrier-keyed modulation.
2. Phase modulation, also known as phase-shift keying (PSK).
3. Frequency modulation, also called frequency-shift keying (FSK).

The term "keying" is derived historically from the amplitude on-off carrier keying used with the first digital signals of telegraphy.

As with analog signals, detection of digital signals may take place either coherently (implying knowledge of carrier frequency and phase, and using cross-correlation) or noncoherently (using autocorrelation or envelope detection). Some of these systems and their characteristics are described here.

Amplitude Modulation

Amplitude modulation (AM) is frequently used to send m-ary digital information. During each time interval, one of m carrier amplitudes is transmitted. By the usual definition, an "amplitude" is the magnitude of a sinusoid; thus, the correspondence between data and amplitude should yield a positive value. However, this notation has come to be generalized in discussing AM so that now any time function which multiplies a sinusoid is called an "amplitude modulation" of the sinusoid. This more general view includes suppressed carrier schemes. Thus, AM with an amplitude of 0 or 1 is binary, and the average carrier power depends on the mean value of the amplitude. If instead a data 0 corresponds to an amplitude of +1 and a data 1 to an amplitude of -1, and symbols are equiprobable, the average carrier power is constant. The latter will be considered AM even though it is completely equivalent on transmission to π radian phase-shift modulation, i.e.

$$\pm \cos \omega t = \sin \left(\omega t \pm \frac{\pi}{2}\right)$$

The distinction between binary AM and binary PSK is thus not necessary until discussing the receiver, where the methods of detection become different. For m-ary ($m > 2$) AM and PSK no such direct correspondence exists.

AM is most commonly detected by an envelope detector. The envelope of $\pm \cos \omega t$ is constant, and hence such a signal is not recovered with an envelope detector. The AM detectors for $\pm \cos \omega t$ are identical with the detectors for PSK since the first function of the detector is to identify and generate $\cos \omega t$. Then a product detector followed by a low-pass filter performs the detection; thus,

$$2\cdot[\pm \cos \omega t]\cdot\cos \omega t \xrightarrow[\text{product detector}]{} \pm 1 \pm \cos 2\omega t \xrightarrow[\text{low-pass filter}]{} \pm 1$$

whereupon detection is accomplished. For conventional AM, however, the envelope detector may be and often is used. Being the most economical detector to build, it will be found in many systems in which its relatively poor performance under noisy conditions is not detrimental.

The principal value of AM lies in its minimization of channel bandwidth. No modulation has a potentially lower bandwidth since AM systems can be constructed with an RF bandwidth equal to the input information bandwidth.

Since telephone lines are of restricted bandwidth, AM (especially vestigial sideband and single sideband AM) has been widely used for data transmission. The noise on telephone lines tends to be impulsive and the corresponding errors bunched, and hence special coding to correct or detect errors is required. The channels of concern in this report are affected primarily by gaussian noise of uniform spectral density. Thus, the problems of interest center about the now classical case of signals submerged in gaussian noise.

<u>Probability of Error for AM Systems</u>

Denote by $P(1|0)$ the probability that a 1 was detected but a 0 transmitted. Then the total probability of error is given by

$$P(\text{error}) = P(1|0)\ P(0) + P(0|1)\ P(1) \qquad (1)$$

where P(0) and P(1) are the a priori probabilities of 0 and 1 respectively. For convenience, assume that the a priori probability of a 0 or 1 is .5.* To compute error probability for an envelope detector, choose a threshold U. If the value of detected voltage is greater than U, this will imply that a 1 was received. Values less than U signify that a 0 was received. The error probability, assuming gaussian noise, is then given by

$$P(\text{error}) = \underbrace{\tfrac{1}{2}\int_U^{\infty} x \exp\left[-\frac{x^2}{2}\right] dx}_{\tfrac{1}{2}\cdot \text{Prob}\,[\text{Noise} > U]} + \underbrace{\tfrac{1}{2}\int_0^{U} x \exp\left[-\frac{x^2+a^2}{2}\right] I_o(ax)\,dx}_{\tfrac{1}{2}\cdot \text{Prob}\,[(S+N) < U]} \qquad (2)$$

where $I_o(ax)$ is the Bessel function of the first kind of order 0 and imaginary argument. The probability densities shown are those of envelope of noise alone and of noise plus a sine wave, $a\cdot\cos\omega t$. In both cases, rms noise is taken to be unity. This expression is minimized for the value U satisfying

$$I_o(aU) = e^{-\frac{a^2}{2}} \qquad (3)$$

as can be verified by simple differentiation. Thus, when a mean-signal-power to mean-square-noise-power ratio ($a^2/4$) is given, values of the optimum threshold U and error probability P(error) are as given in Table 1.

---

*This channel is not a binary symmetric channel.

Table 1

AM ENVELOPE DETECTION PROPERTIES

S/N	U	P(error)
10	3.5	2×10^{-3}
15	5.5	2×10^{-6}
20	10.0	1×10^{-22}

Equation (3) is used to determine the threshold, and Eq. (2) is evaluated as follows:

$$P(\text{error}) = \tfrac{1}{2} \exp\left[-\frac{U^2}{2}\right] + \tfrac{1}{2}\left[1 - Q(a,U)\right] \qquad (4)$$

where Q is Marcum's Q function tabulated in Ref. 1. It is here assumed, of course, that 0 and 1 are equally likely.

Complete treatments of error probabilities for m-ary AM systems have been made for the important practical case of high-frequency radio. Here, in addition to noise, there is a fading of the signal due to multipath interference and/or ionospheric fluctuation. Again, the application of AM is forced by the narrow bandwidths available in this portion of the spectrum. Such limitations are less likely in communication satellites, especially in view of the linearity requirements imposed by AM. For this reason, the expressions for error probabilities for m-ary AM systems are not discussed here. (See Refs. 2, 3, and 4.)

Frequency Modulation

Frequency modulation is often used for digital traffic because it is easy to implement, effective in reducing the effects of noise, theoretically convenient to model, and easily used to form either video or radio frequency signals. It also forms a realization of the symmetric channel in the binary case.

FM is modelled in the binary case by a carrier at frequency 1 or 2, depending on whether the bit to be sent is 0 or 1. The corresponding receiver is a pair of detectors looking for signal energy near frequencies 1 and 2. See Fig. 2. To ensure correct detection, the two frequencies must be sufficiently separated so that their spectra do not overlap, or, equivalently, so that when one channel is occupied the other is empty. In general then, the two detectors operate at a spacing of at least B, where B is the full width of the signal on one frequency. This latter quantity is discussed in greater detail later. Of principal theoretical importance is the assumption that noises in the two channels are uncorrelated and that intersymbol interference is negligible. For separately filtered gaussian noise the assumption is valid, but it would not hold for impulse disturbances of broad bandwidth. Again, these are unlikely to be found on ground-to-satellite communication links. One possible source would be radar interference.

Error Probabilities for the Binary Case

Three detectors are available for digital FM signals:

1. The common FM discriminator or ratio-detector as used with analog modulated signals.
2. Two bandpass filters followed by coherent detectors.
3. Two bandpass filters followed by envelope detectors.

The first is used only incidentally for digital signals since it has inherently poorer characteristics than the other two. The principal

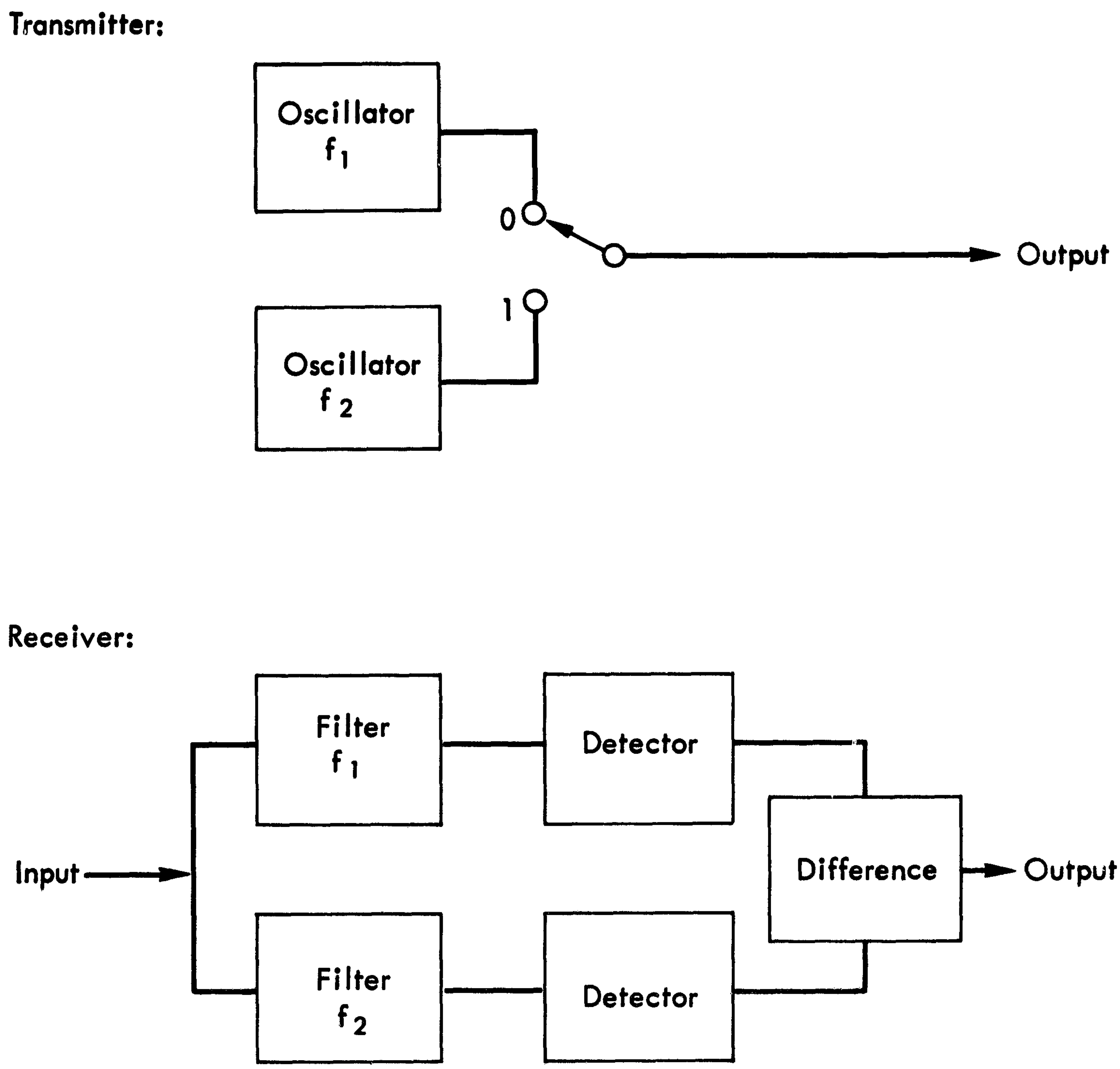


Fig. 2—Digital FM transmitter and receiver

difficulty is that the bandwidth for noise, and hence the output noise power, is appreciably greater than in the other two systems, where the noise is bandlimited before detection.* Also, since the first detector is highly nonlinear, it is nearly impossible to model it accurately enough to permit a good calculation of error probabilities.

The other two detectors are simpler to implement and have simpler theoretical properties. For coherent detection in gaussian noise (see Fig. 3), the expression for error is (see Appendix for derivation)

$$P(\text{error}) = \int_A^\infty \frac{1}{\sqrt{2\pi}} \exp\left[- x^2/2\right] dx$$

where $A^2 = E/n_o$, E is the mean signal energy, and n_o is the mean-square noise power per Hz. This equation is plotted in Fig. 4.

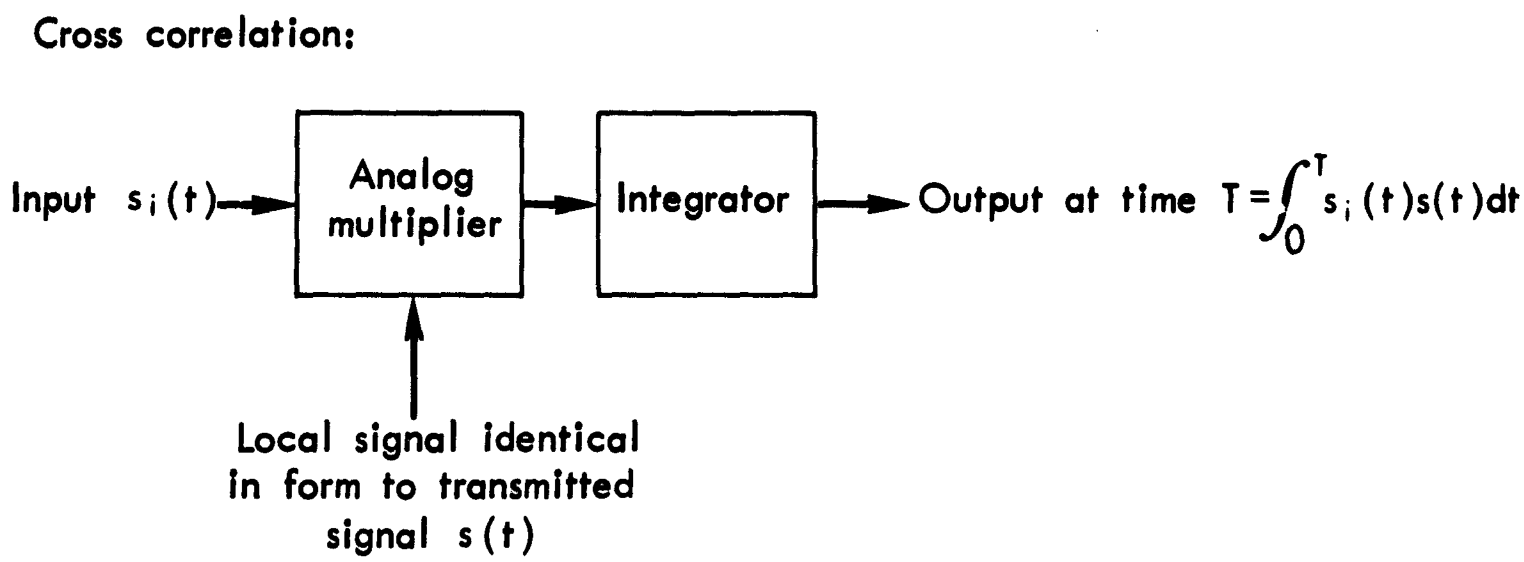
Noncoherent reception takes place with envelope detectors on the outputs of each of the filters in Fig. 2. In this case the error probability is given by (see Appendix)

$$P(\text{error}) = \tfrac{1}{2} \exp\left[- \frac{A^2}{2}\right]$$

This equation is also plotted in Fig. 4. A comparison of the two curves in Fig. 4 shows that for a large E/n_o ratio the error probability for coherent FSK is nearly an order of magnitude better than for noncoherent FSK (NCFSK). This difference decreases as E/n_o becomes small, that is, as P(error) becomes large. As the signal vanishes into the noise, the error probabilities for all systems approach 0.5.

---

*This is not to say that a special discriminator could not be built for digital signals, but only that the ordinary one is not optimum.



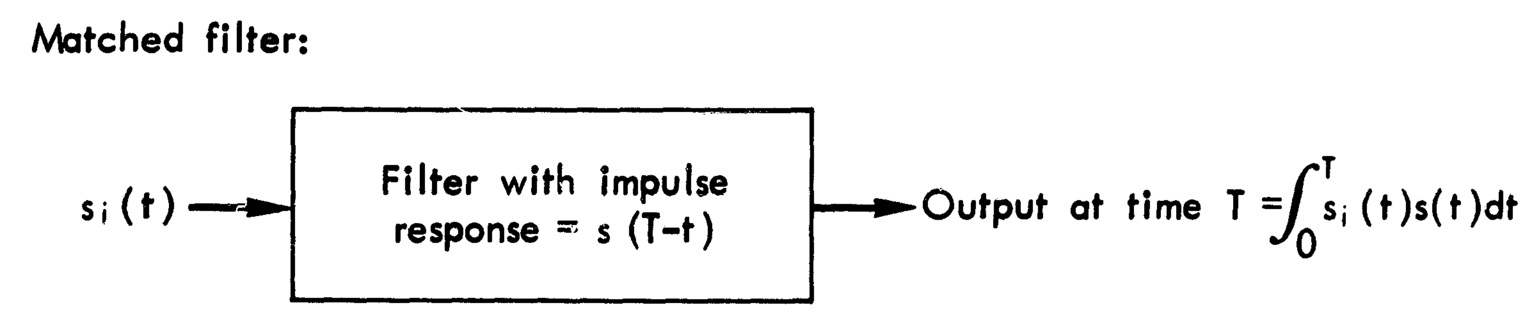


Fig.3—Two equivalent forms for a coherent detector

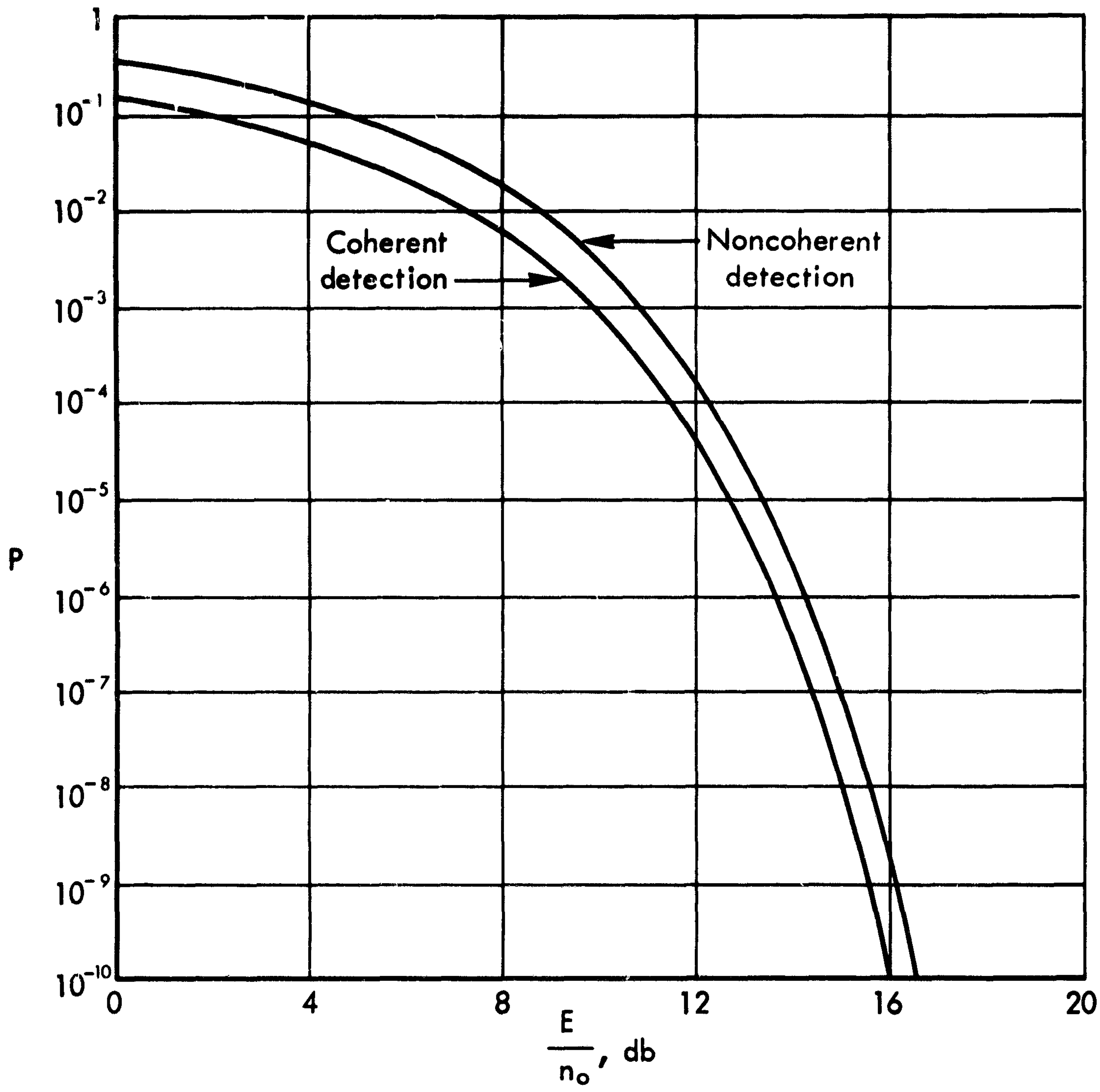


Fig. 4—Probability of error for FSK

At the receiver such a system becomes equivalent to a coin tossing game. These curves are useful later in assessing the value of error correction codes.

Error Probabilities for the m-ary Case*

The binary FM system may be generalized to the m-ary case with ease. With m frequencies available, choosing one corresponds to the transmission of $\log_2 m$ bits of information. The detector is an array of m filters (see Fig. 5), each followed by an envelope detector or a coherent product detector. The detector having greatest output is presumed to have the signal. Errors are made, therefore, when the output of one of the nonsignal filters exceeds the output of the signal filter. Results of these error probabilities are available in the literature.[2,5] Difficulties associated with the "greatest-of" detection make it difficult to implement the m-ary detector as described. In its place, a decision device may be placed on each detector output, and when a unique 1 is received, its channel is recorded, and multiple 1's and all 0's are rejected as errors (see Fig. 6). The performance of such a system is even worse than that of noncoherent systems since the decision device operates on the content of a single channel only. Such a system has the undesirable feature of a threshold setting as in AM.

The m-ary systems described above assume that a single frequency is transmitted during each time interval. It is, of course, possible to transmit k of the m frequencies simultaneously. Instead of m

*These are also called multifrequency systems or multitone systems.

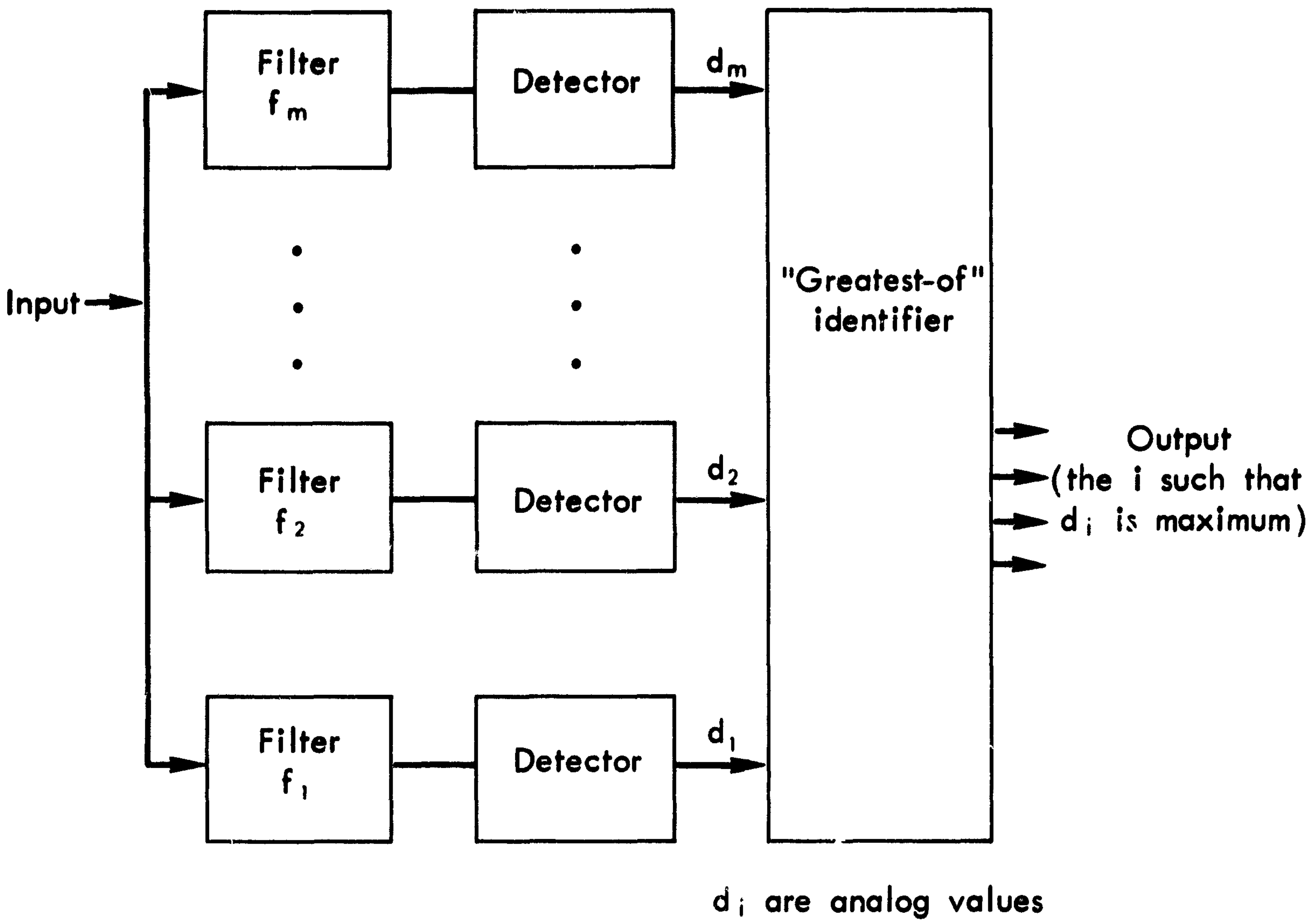


Fig.5—m-ary FSK receiver

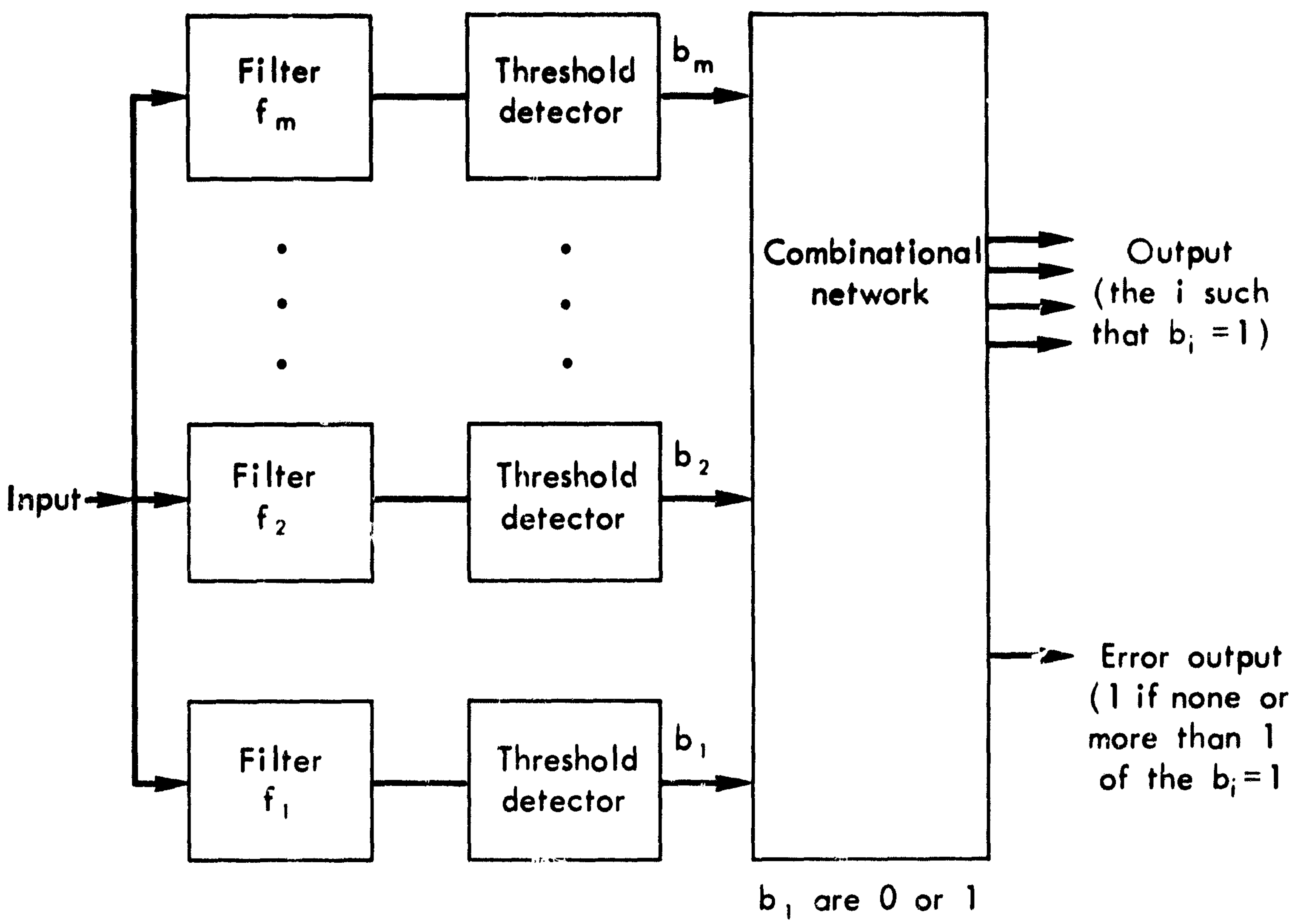


Fig. 6—Modified m-ary FSK receiver

different received signals, one has $\binom{m}{k}$ such signals, and hence $\log_2\binom{m}{k}$ bits of information. But now the energy at each frequency is less by a factor of k than before, so the error probability at the detector output is increased. The fact that exactly k signals are expected can aid the receiver in detecting the error condition.

Phase Modulation

Phase modulation has become increasingly important in communication systems. Though generally more difficult to implement than some other methods, its theoretical performance is enough better to justify the cost. It is being used in spacecraft communications to an increasing degree. Once coherent detection has been provided, phase detection offers lower error probabilities than any other binary detection technique. In this sense it is ideal.

Generation of binary phase modulation is simple. One of two sources which are 180° out of phase is chosen according to whether the bit to be transmitted is 0 or 1 (see Fig. 7). By making the carrier frequency and bit rate bear a rational (or integral) relationship, it is possible to generate signals with desirable spectral properties.

In addition to pure coherent detection, one may use an autocorrelation technique known as differentially coherent PSK, whereby the current phase is determined by comparison with the previous phase, which is used as a reference and saved in a storage element such as a delay line. However, since the reference is noisy, the error probability is greater than for a coherent system.

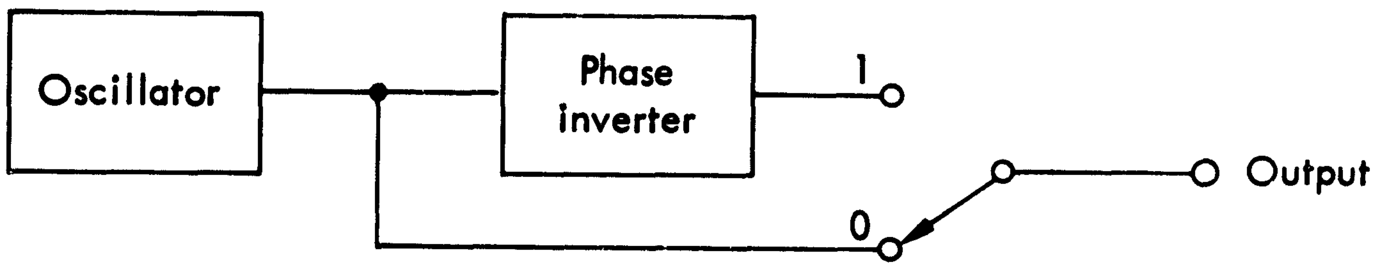


Fig. 7—Generation of phase modulation

The equivalence of $\pm 90^{\circ}$ phase modulation with suppressed carrier AM was mentioned earlier. When multiple phases are used, as in m-ary phase modulation, there is no longer any such relationship. Multiple PSK can be detected by coherent techniques, whereby the phase is actually measured and categorized, or by the differential technique mentioned.

Error Probabilities for Phase Modulation

In the case of coherent detectors the error probability is determined by (see Appendix)

$$P(\text{error}) = \int_{A\sqrt{2}}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{x^2}{2}\right] dx$$

where $A = \sqrt{E/n_o}$. This is shown in Fig. 8 as a function of E/n_o, the ratio of signal energy to mean noise-power density. On the same graph is shown the corresponding relation for differentially coherent PSK determined from the equation[8]

$$P(\text{error}) = \tfrac{1}{2} \exp\left[-A^2\right]$$

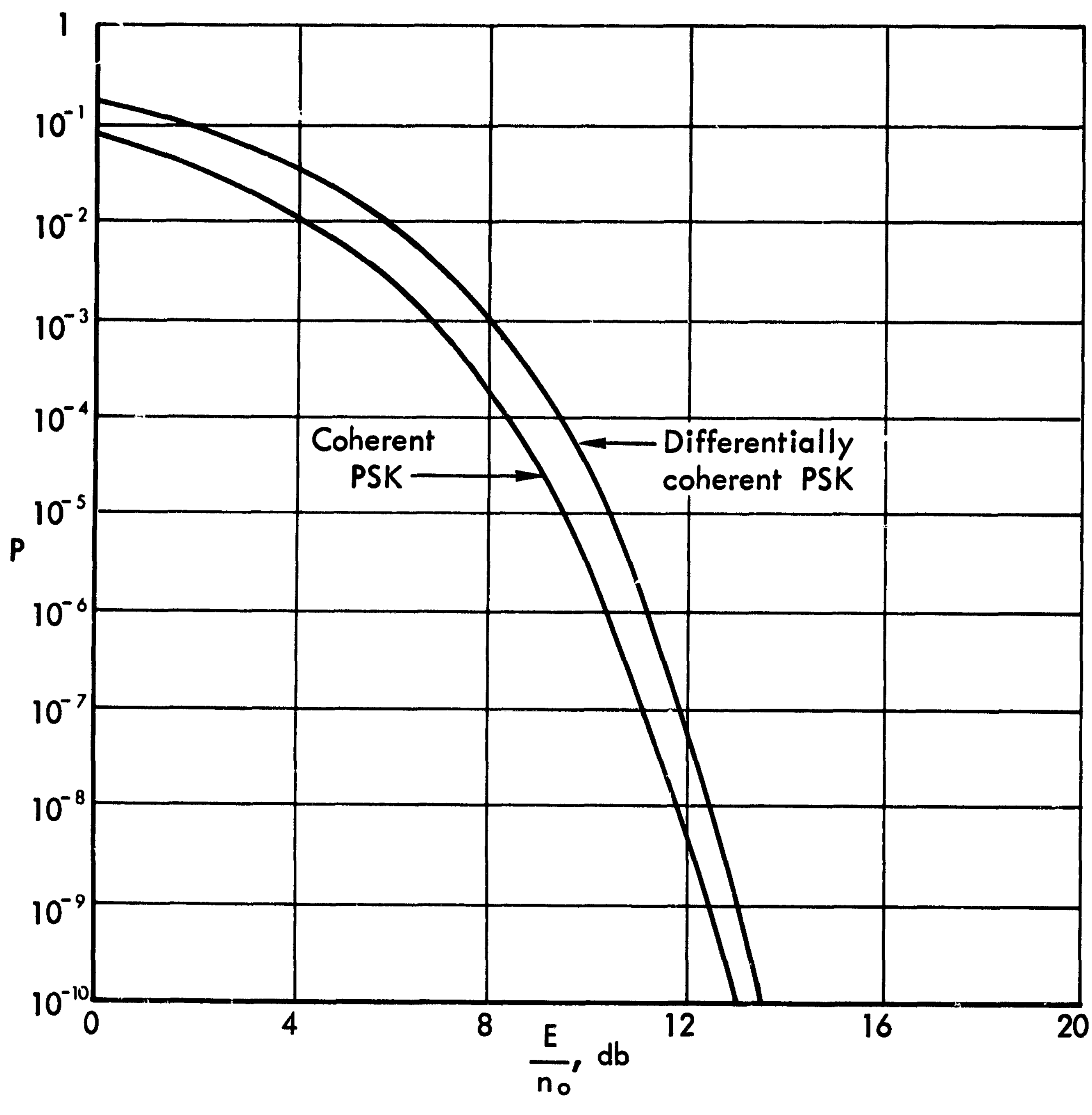


Fig. 8—Probability of error for PSK systems

where again $A = \sqrt{E/n_o}$. The two are seen to approach one another as the signal level becomes large. This corresponds to stabilizing the phase reference for the differentially coherent system.

Corresponding relations for m-ary PSK are derived in the literature.[2,6,7] These relations are based on theoretical models which are frequently difficult to achieve in hardware. The results are therefore approximate. An ordinary phase detector merely converts phase to amplitude, which must then be discriminated. Or, alternatively, the correct phase may be fed to each of m phase detectors. The detector having maximum output will usually correspond to the correct signal phase. Again, problems of choosing thresholds and of deciding maxima result in error situations. This emphasizes the fact that many of the problems associated with amplitude modulation also occur with FM or PM.

An early successful multiphase, differentially coherent system was Collins' "Kineplex."® It was designed to place data over conventional telephone circuits and boasted rates of 3300 bits per second over a long-distance private line circuit. In a military version using eight phases (AN/GSC-4), the bit rate was increased to 5400 bits per second.

PULSE CHARACTERISTICS, BANDWIDTH AND FILTERING

In the cases discussed in this Memorandum, one or more bits are being transmitted during a fixed time interval, T. Most characteristics of reception do not depend explicitly on the value of T. For example, talk of detecting pulse energy avoids mention of time T and average power W of the pulse. But when a system is being discussed, the information rate in bits per second does place a constraint on T, as does the allowed bandwidth. In addition, bandwidth becomes important in frequency-division multiplex circuits.

It is well known that finite duration and finite bandwidth are incompatible relations. From a practical viewpoint, finite duration pulses must be sent over finite bandwidth channels, and the theoretical relations must be carefully examined. This "finiteness," however, is a practical one; that is, if energy has vanished below residual noise levels in the system, it can be regarded as being zero though in fact it is not.

If minimum bandwidth is a constraint, which is not always the case, a pulse shape

$$f(t) = A\left(1 + \cos 2\pi\frac{t}{T}\right), \quad |t| < \frac{T}{2} \text{ where } A^2 = \frac{2}{3}\left(\frac{E}{T}\right)$$

$$= 0 \qquad\qquad\qquad , \quad |t| \geq \frac{T}{2}$$

has the finite duration T, energy E, and also has an acceptably low bandwidth. It is shown in Ref. 8 that 95 percent of the power is contained within a base bandwidth 1.11 (1/T). Similarly, 99 percent is contained within 1.40 (1/T). Of all easily generated pulse shapes suggested, the "raised cosine," as this one is called, has the most compact spectrum. The rectangular pulse, by comparison, has 95 percent of its power within 2.11 (1/T) and 99 percent within 10.03 (1/T) bandwidth. Thus, at the 99 percent level a rectangular pulse has over seven times the bandwidth of a raised cosine.

The penalty of excessively narrow filtering is intersymbol interference. That is, when the spectrum of a pulse is constrained, the pulse lengthens. Being longer, it affects neighboring pulses, making their probability of error greater. This intersymbol interference is a serious problem in constrained bandwidth systems and usually forces an increase in required signal-to-noise ratio at the detector.

One filter exists which theoretically has no intersymbol interference. This is the so-called "matched" filter, whose output peaks for the desired signal only. The matched filter is implemented by constructing a device whose transfer function is the complex conjugate of the signal spectrum.[9] The output waveform is then the auto- or cross-correlation of the desired or undesired signals respectively. Such a matched filter is frequently assumed to be used in deriving expressions for error probability in the output of a detector. An equivalent way to characterize a matched filter is to say that its impulse response is the same as the time-reversed pulse shape it is matching. Thus, for symmetrical pulses, such as the raised cosine or rectangular shape, the impulse response corresponds exactly to the desired pulse shape, and the same matched filter can be used for both generation and detection of pulses.

III. TIME-DIVISION MULTIPLEXING

INTRODUCTION

Multiplexing is the generic term for ways of combining two or more separate information sources (subchannels) onto a single channel. The classical way of doing this, exemplified by carrier telephony, is frequency-division multiplex (FDM). This technique is discussed in another Memorandum in this series.[10] Gaining wide acceptance where signals are inherently digital is the technique of time-division multiplex (TDM). In this technique, each input subchannel is assigned a time slot on a periodic basis and has exclusive use of the output channel during that time. The output channel is then connected to another input subchannel, and the cycle continues until all inputs have been serviced and the process starts over again. The commutation process need not be uniform. Any input subchannel can be assigned any duration in the output channel, depending on its information rate. This scheme is generally unsatisfactory since the provision of non-uniform time periods is not easily implemented. But time slots can be subdivided into some small useful quantum, and subchannels can be assigned various multiples of these quanta according to their needs. Such a scheme is used in telemetry systems where some slowly varying quantities need time slots only rarely, while wider bandwidth signals require more frequent time slots.

Historically, time-division multiplexing was suggested to overcome transmission problems caused by nonlinearities in amplifiers on multihop microwave systems. Unfortunately, the pulse performance of vacuum tubes was so poor that it proved easier to improve the linearity than to retain the pulse approach (TDM). A frequency-division multiplexed baseband is especially subject to crosstalk and distortion arising from nonlinear amplification characteristics. When the dynamic amplitude range of a transmission system is constrained for any reason, such crosstalk becomes likely. In such a channel, digital and TDM techniques become valuable. Adding to this value is the impressive state-of-the-art of digital systems as developed in the computer industry. By far the best combination, and extremely easy to implement, is time-division multiplexing of a number, M, of digital channels. This is the main concern of this section, though a few paragraphs are devoted to the time-division multiplexing of analog signals.

TIME-DIVISION MULTIPLEXING OF BINARY DIGITAL CHANNELS

Assume that M binary digital sources, each carrying information at rate R, are to be combined onto a single baseband channel. The combined channel contains information at rate M·R. A digital multiplexer consists of a counter and a set of gates arranged to introduce, one at a time, the signals from the M lines onto the single output line, as shown in Fig. 9. The result is a single digital signal which can be encoded, transmitted, and received in the same manner as any other digital source. The signal statistics in the baseband are changed insignificantly as compared to those in the individual channels. This is not true of analog basebands. Demultiplexing takes place in the inverse manner, as shown in Fig. 10. The only additional requirement is a synchronization signal enabling the counter to be set to the appropriate output at the correct time for that output. An actual

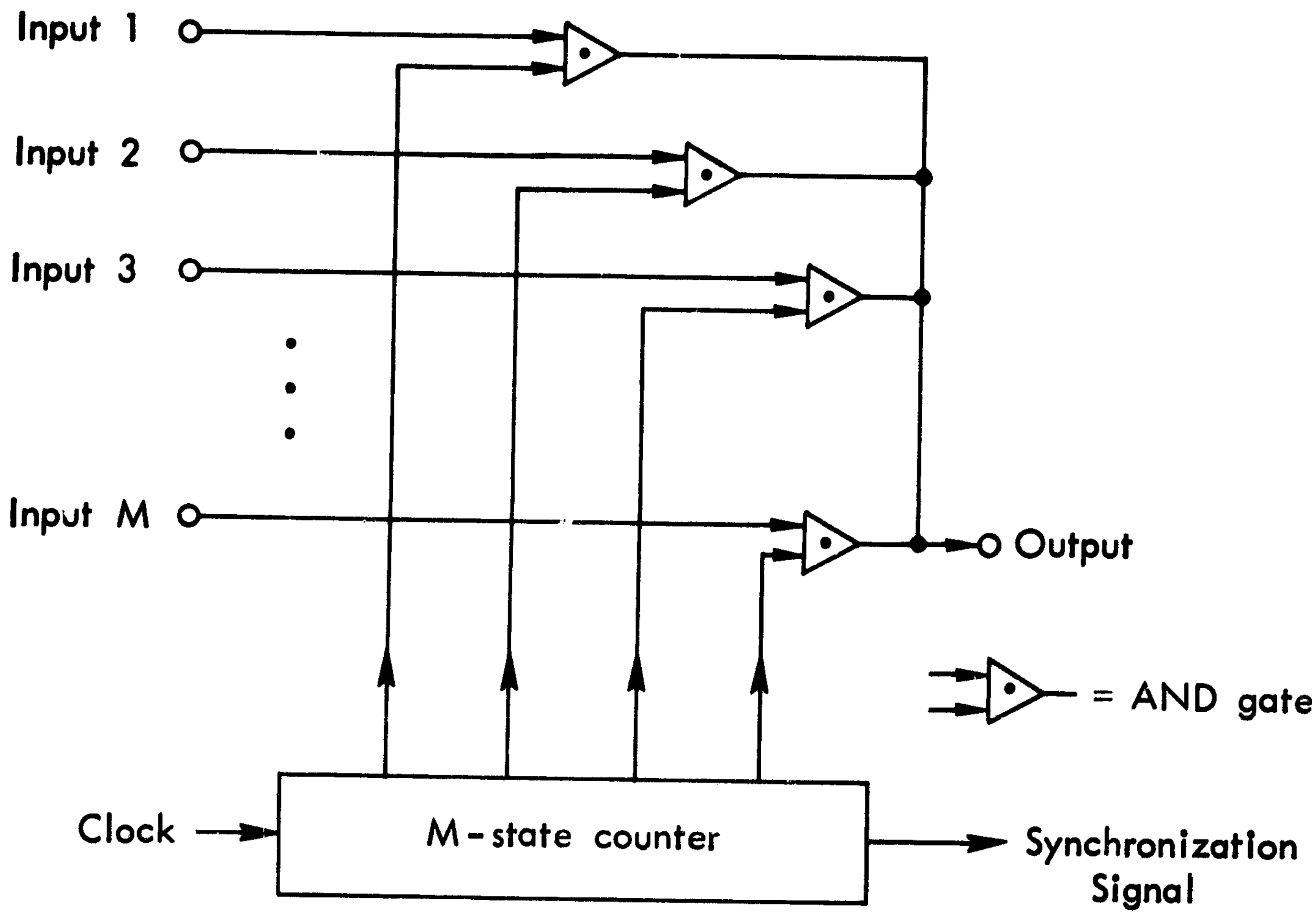


Fig. 9—Digital signal multiplexer

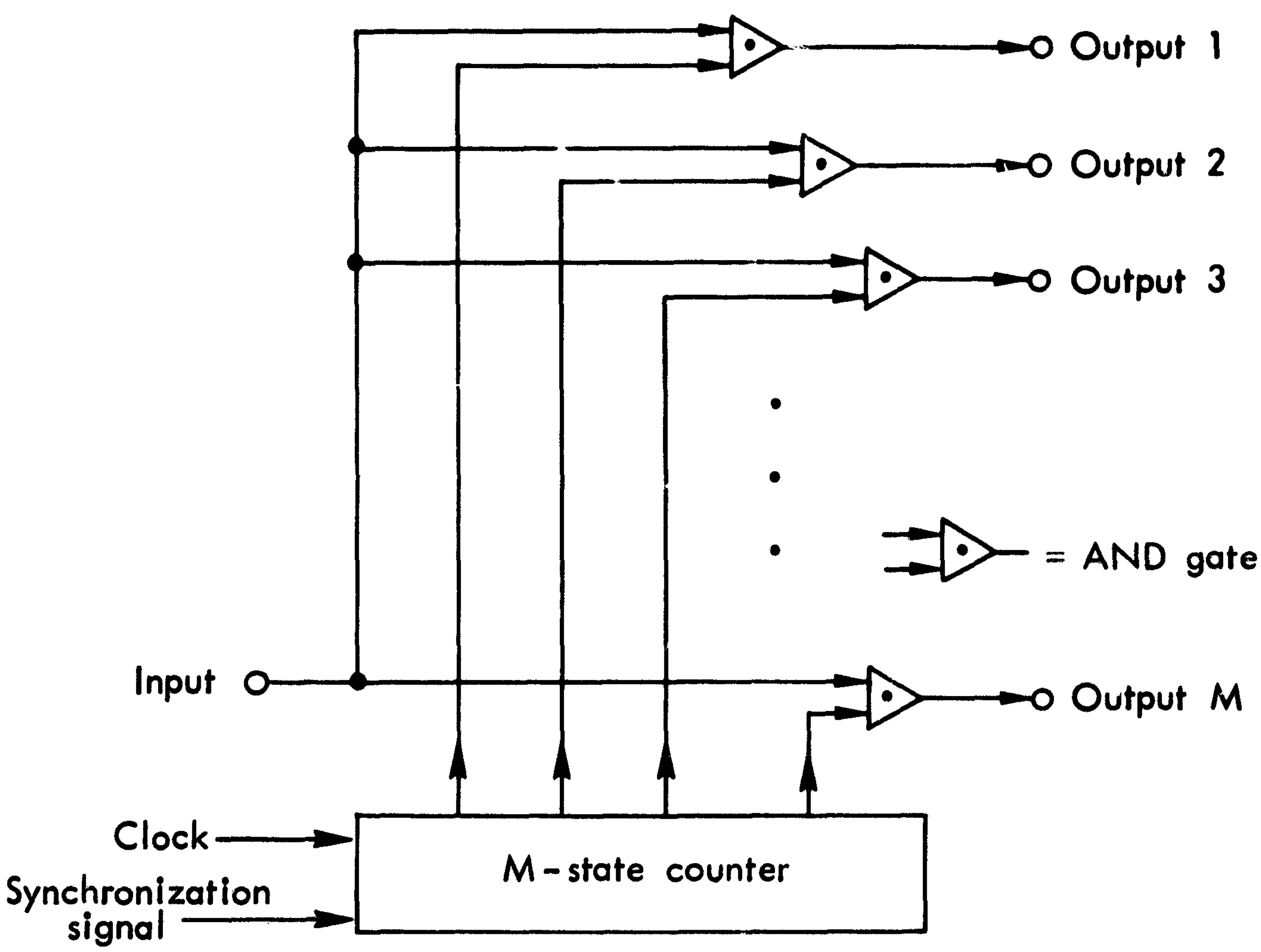


Fig. 10—Digital signal demultiplexer

demultiplexer would contain circuitry for lengthening the individual output waveforms by a factor of M, so the channel lines contain the rate R at minimum bandwidth.

An alternative multiplex-demultiplex combination is shown in Fig. 11. Here the channels are introduced into an M-bit shift register in parallel form at a single instant. The M bits are then shifted out in serial form. The inverse operation is performed by shifting the bits into the shift register in serial form and reading out in parallel form at the correct time. Again, the outputs are short pulses which must be lengthened.

Synchronization in these cases consists of identifying a channel at the receiver and providing a clock which remains essentially in phase with the incoming pulse train. Except in high-noise situations, these requirements are easily met. The general problem of synchronization is discussed in Section VI.

TIME-DIVISION MULTIPLEXING OF ANALOG CHANNELS

A basic property of analog channels is their finite bandwidth. Indeed, if channels are to be multiplexed, say by FDM, the bandwidth is deliberately restricted by filters. Such bandlimited signals are amenable to a discrete time representation by means of the sampling theorem:

> A bandlimited signal of bandwidth B Hz is equivalent to a sequence of samples (values) of the signal chosen at intervals 1/2B seconds apart.

These discrete time samples are equivalent in the sense that the original signal can be exactly constructed from them. Now consider M such bandlimited channels. If a sample of each is taken every 1/2B

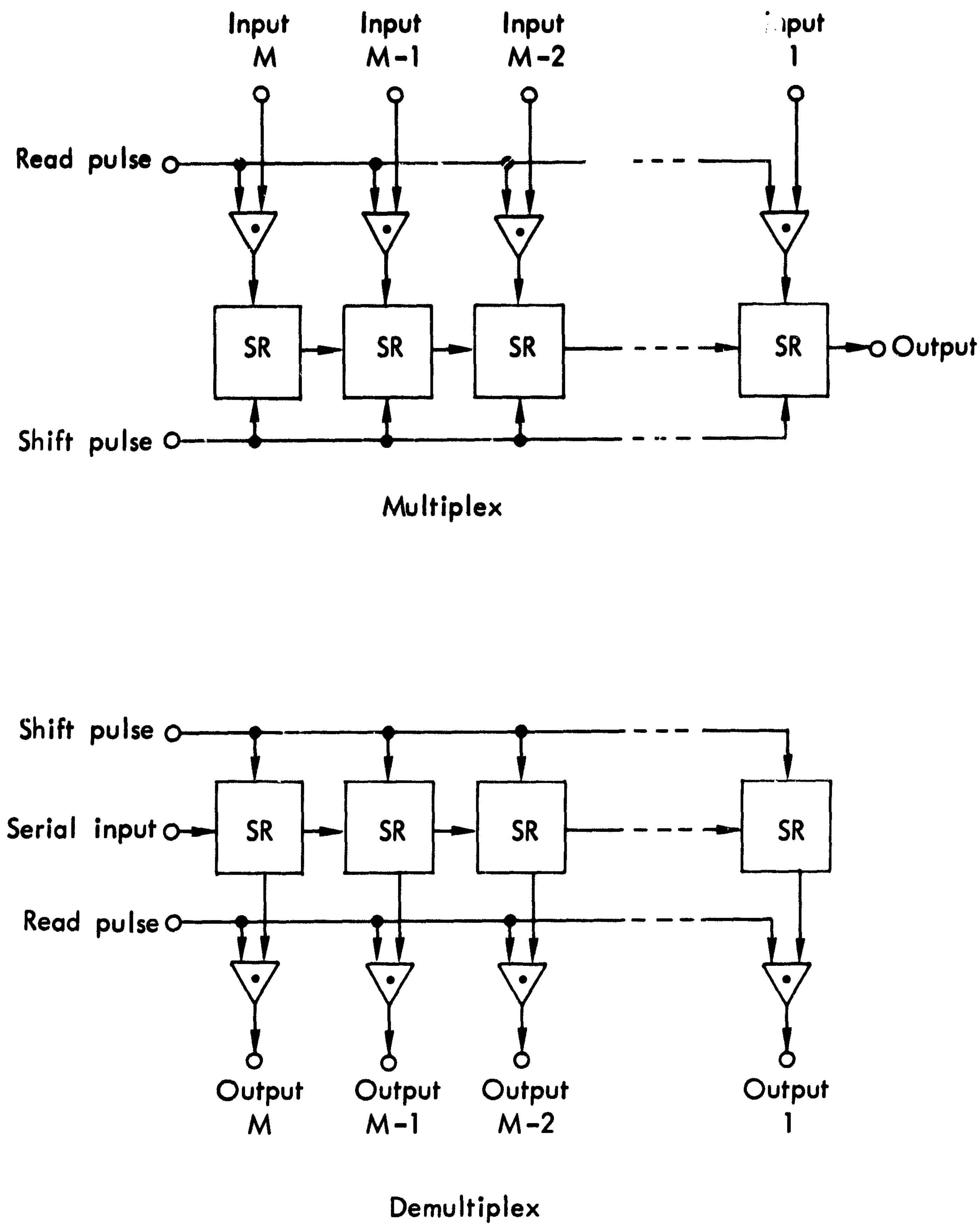


Fig. 11—Shift register techniques for digital multiplexing

seconds and if all M samples are merged into the time interval 1/2B, then time-division multiplexing has been performed. Samples are frequently transmitted as pulses having an amplitude proportional to the sample value. Then if such pulses are accorded a duration 1/2MW, all M will be transmitted in the time 1/2W. The process then continues cyclically in time. The bandwidth of the resulting transmission depends on the pulse shape used to transmit the samples, and it is not less than 2W per channel in any practical case.

These statements are illustrated in Fig. 12, where an analog signal, its sample equivalent, and its corresponding pulse form are shown. When the ratio t/τ is equal to M, there is adequate time available to place M channels in time-division multiplex.

Time-division multiplexing is often used as a prelude to analog-to-digital encoding since it makes it possible to use a single high-speed encoder instead of the M encoders which would be required if each channel were individually digitized and the resulting digital channels multiplexed. This fact is noted in Section V where the encoding of analog data into digital form is discussed.

PROBLEMS UNIQUE TO TDM

Time-division systems will suffer from any effect of memory in the components through which the signal passes. The most common source of such memory is filtering done on the signal waveform. The result is to cause one symbol to be perturbed by one or more previous symbols. This is called "intersymbol interference" and generally acts to introduce crosstalk between channels, much as poor filtering in FDM results in crosstalk. Since, however, the problem frequently originates in those

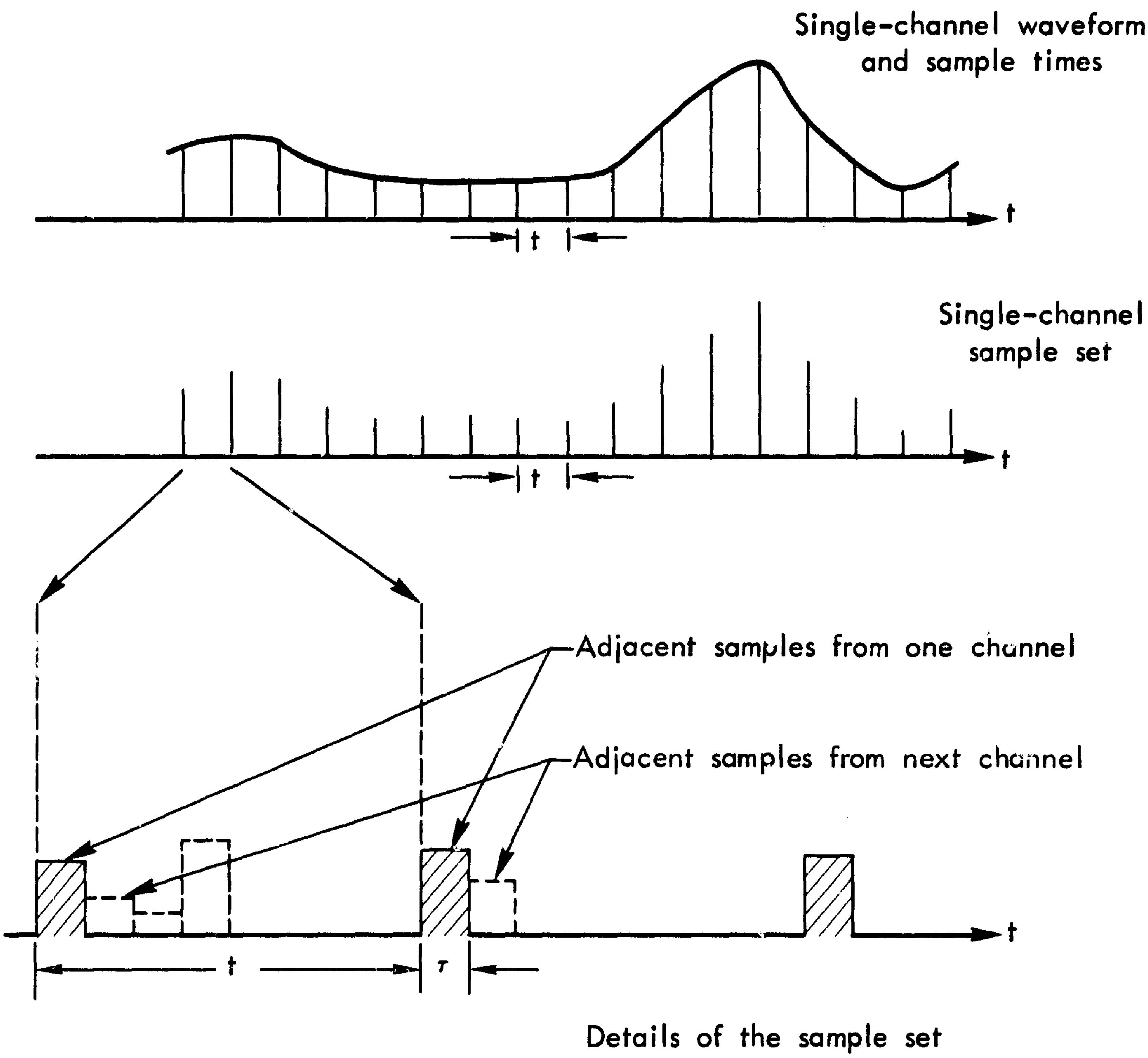


Fig. 12—Time-division multiplexing of analog channels

portions of the system under control of the designer, it can often be removed by equalization techniques. Theoretically, a matched filter has no intersymbol interference, so it can be used in the system with relative impunity. In early TDM systems, adjacent symbols were separated by "guard slots" of symbol-free time, during which the filter memory would decay. Such considerations play more of a role in analog sample systems, where a large amplitude pulse may be followed by a very low amplitude pulse whose amplitude must be accurately discriminated. The memory decay time must therefore be quite short compared to symbol length. In binary systems, this is less of a consideration, though at or near threshold the probability distributions of the signal plus noise are influenced by fragments of adjacent signals. In summary, this problem is of importance principally when the communication channel unavoidably filters the signal, as for instance over telephone lines. In such cases, careful equalization is often used to minimize the problem. An increased signal-to-noise ratio is another cure in binary systems. Added signal power enables setting detection thresholds higher, thus eliminating the chances that adjacent bits will cause incorrect detection of desired bits.

Another problem often mentioned concerning time-division systems is the synchronization requirement. Detailed inspection shows, however, that an identical problem exists in FDM or, in fact, in any situation where unique decipherability is not possible from transmitted redundancies. The pertinent needs are reviewed in Section VI, but they are really quite simple, and in stationary satellite systems they are nearly negligible.

CONCLUSIONS

Time-division multiplexing of digital as well as analog signals is easily done with modern digital circuitry. It is being proposed for many new telecommunications exchanges because of the increased importance of digital data traffic.[11,12] These new systems perform routing, switching, signalling, and accounting on the basis of message content in digital form. The absence of working systems is, in fact, attributable to the newness of solid-state equipment and the satisfactory performance of FDM for analog traffic. Since future traffic appears to be strongly digital, TDM assumes great importance.

IV. ERROR CORRECTION AND DETECTION

INTRODUCTION

All transmission channels contain inherent noise, or error causing phenomena. Decreasing this noise usually involves increased transmitted power, while in some communication situations (multipath, for example) short-term outages with no signal at all must be tolerated. Thus, it is rarely possible to send a completely error-free raw message without undue extra cost. One way to handle this situation is to send each message several times. This may be done serially or by parallel means. For example, diversity reception schemes use the redundancy offered by separate RF paths. Repeating the message at least three times, for instance, allows choice of the most commonly repeated characters. Natural languages such as English contain much redundancy in spelling and use excessive characters to contain ideas.

Coding for error correction has recently become practical with the development of codes possessing an implementable algebraic structure. Correcting very frequent errors is still difficult and expensive, but with careful system design few errors are expected, and these few can be corrected with relatively simple equipment. The use of coding techniques calls for the introduction of redundant digits travelling with the message. These extra digits require an additional share of transmitted power. Message digits are therefore robbed of more power, but there can be a net advantage when the normal error rate is low. For example, if the bit error rate is less than about 10^{-3}, the inclusion of single or double error correction can be as effective as increasing power by a few decibels. Some quantitative examples are included in this section, following a general discussion of the relevant system parameters.

A variety of encoding and decoding schemes exist, many of which are still in the experimental stage. Discussion here will be limited to block coding, convolutional encoding, and sequential decoding. Since it is impossible to explain these in detail in such limited space, it will be necessary merely to sketch briefly how the system works and to refer the reader to the literature for the many details. Of principal interest here are techniques which hold promise for low implementation cost per decibel of saved transmitted power. All techniques described in the literature can be implemented even though some may require equipment which is still being developed. These techniques are of less immediate interest. Of greatest value currently are binary cyclic block codes of medium length, i.e., several hundred bits per block.

BINARY BLOCK CODES

A binary block code uses incoming information bits in groups or blocks, of size k, and the code is described by the number of total bits in a block, n, the number of information bits, k, and sometimes by either the "distance" or error correction capability of the code.

The redundant digits (n-k in number) are chosen to make the complete n-bit word as far different from other n-bit words as possible. This difference is measured by the Hamming distance, which is defined as the total number of places (bits) in which two words disagree.

For example, the words expressed by 110010 and 101010 have a Hamming distance of 2 since they differ in only the second and third positions. A single error "moves" a code word to one of n adjacent places, each at distance 1 from the transmitted word. Such an error creates a fuzzy or uncertain area about the correct word. If a systematic way can be found to locate code word positions such that their fuzzy regions are disjoint, the original message can be reconstructed if only one error occurred. The same reasoning can be applied to larger numbers of errors. The result is to form larger fuzzy regions, and code words (information digits plus redundant digits) must be chosen which are further apart in Hamming distance. The greater the error correction capability required, the larger the number of redundant digits which must be added. A code is described by giving some parameters in a parenthetical grouping. An (n,k,e) code has a total of n bits; k of these are information bits, and the redundancy enables correction of e or fewer errors. In Table 2 a (7,4,1) code is shown. The reader can verify that any single error still gives a word which is closer, in the Hamming sense, to the original than to any other word. It will also be noted that every word is at least a distance of 3 from every other word. Simple reasoning shows that if every word of a code is at least 2e + 1 from every other word, then e or fewer errors can be corrected.

The rate r of a block code is the ratio of information bits to total bits, $r = k/n$. Generally speaking, a code becomes increasingly efficient (i.e., $r \rightarrow 1$) as n is made larger and larger. When n is several hundred, for instance, it becomes impossible to inspect bits

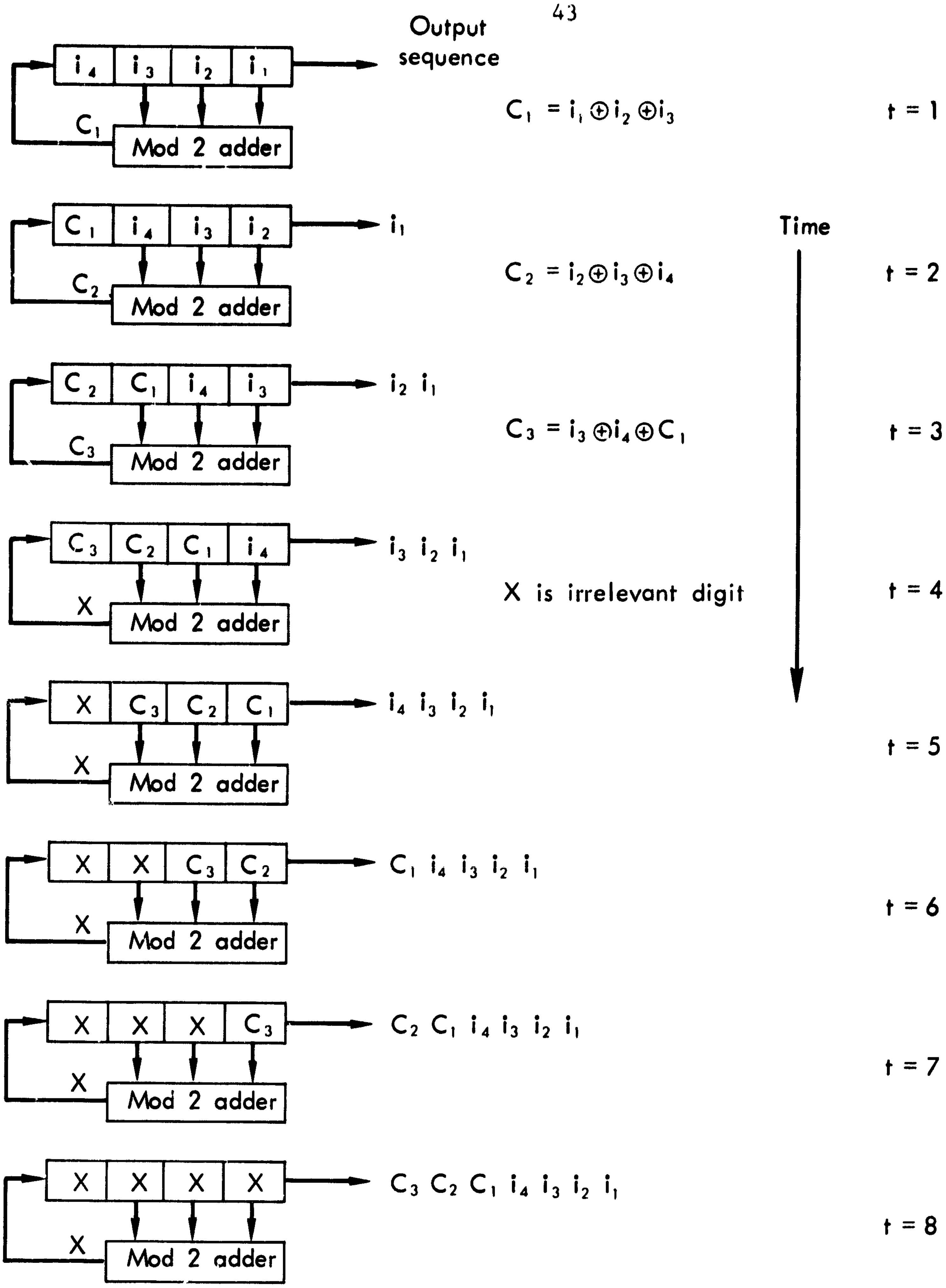
Table 2

A (7,4,1) CODE

Information Bits i_1 i_2 i_3 i_4	Redundant Bits c_1 c_2 c_3
0 0 0 0	0 0 0
0 0 0 1	0 1 1
0 0 1 0	1 1 0
0 0 1 1	1 0 1
0 1 0 0	1 1 1
0 1 0 1	1 0 0
0 1 1 0	0 0 1
0 1 1 1	0 1 0
1 0 0 0	1 0 1
1 0 0 1	1 1 0
1 0 1 0	0 1 1
1 0 1 1	0 0 0
1 1 0 0	0 1 0
1 1 0 1	0 0 1
1 1 1 0	1 0 0
1 1 1 1	1 1 1

individually, so unless an algebraic or arithmetic way of handling the messages is available, little utility can be expected. With the discovery of a class of codes by Bose and Ray-Chaudhuri(13) (and independently in France by Hocquenghem(14)), and their subsequent implementation by Peterson,(15) a truly powerful error correction and detection capability became available. Subsequent studies have improved and simplified these codes until now they are used in commercially available equipment. Since the required equipment is digital, they also gain from improvements due to integrated circuitry and its consequent reliability. Thus, it can be expected that compact system packages for encoding and decoding error correction codes will soon be found. As good as these codes are, however, they are still far from theoretically optimum, and work on codes therefore continues.

Any elementary discussion of the implementation of modern codes requires a basic understanding of modern algebra and digital logic design. This is unfortunate because it is a cause of the slow use and commercial availability of such equipment. The currently popular block codes are based on a cyclic structure in which the various cyclic permutations of the bits of a code word yield other code words. The cyclic structure is convenient since cyclic registers (i.e., shift registers with feedback) are easy to build and possess a well developed theory. Cyclic encoders operate by loading the register with k information bits, then cycling it until n bits have emerged. The first k bits which emerge are the information bits, while the next n-k are the redundant check bits whose presence guarantees the desired correction capability. Figure 13 shows the steps involved in generating the output word of 7 bits from 4 input bits. The corresponding decoder is more complicated, rising in complexity exponentially with the number of errors to be corrected. The increased complexity arises because the number of possible error patterns increases as the number of errors increases. Thus, there are exactly $\binom{n}{e}$ possible error patterns, given that exactly e errors occurred. Since the number of errors is not constant, the decoder must in fact be able to correct any of $\sum_{j=1}^{e} \binom{n}{j}$ error patterns. The redundant bits must be chosen to enable a unique determination of exactly which error pattern occurred.



Load 4 new digits at this point ($t = 8\frac{1}{2}$)

$\oplus$ denotes the sum modulo 2 of the indicated quantities. $A \oplus B = 1$ if A and B are different; $A \oplus B = 0$ if they are the same. This is the same as binary addition without carry.

Fig. 13—The seven stages (plus initial loading) in the formation of the redundant digits of the (7, 4, 1) cyclic code of Table 2

Thus far the amount of computation necessary to do multiple-error correction with block codes has restricted e to values of 1 or 2. More detailed correction requires much more complicated equipment and iterated techniques with many more computations per block. With e so restricted, block length is also restricted since the average number of errors per block $(n \cdot p)^*$ must be less than e. When the need becomes serious enough to justify the equipment or block length, then the complexity inherent in other schemes (such as sequential decoding, which is described later) deserves consideration. Block codes that correct few errors offer, of course, less startling improvements in system performance, as will be seen.

Block codes offer advantages in systems where bursts of errors are encountered, and their use in such environments is already well established. But in communication satellite use, the fading channel appears to be relatively unimportant, and most errors are due to receiver noise or gaussian-distributed external noise. For error burst correction, the Bose-Chaudhuri codes are quite popular, though a wide variety of codes apply to this form of noise distribution.

MEASURING THE PERFORMANCE OF ERROR CORRECTION CODES

The communication systems engineer normally expresses system performance in terms which can be closely related to transmitter power. To enable such a measure to be applied to error correction codes, several factors must be included. First, the information, or

---

*p is the bit error probability, which is assumed to be constant for all bits. The binary symmetric channel is assumed.

input, bit rate must be held constant independent of code parameters. Second, a measure of uncoded noise statistics must be included. Third, the modulation method must be specified. Finally, the output error rate must be specified. To see how these parameters enter, consider the comparison shown in Fig. 14. This shows a system with and without error correction. When all parameters are held constant except, say, the transmitter power, the change in the latter necessary to offer identical overall performance will measure the effectiveness of the coding and decoding applied. While easy to state, this is not easy to apply, and in general no formula exists giving this measure in terms of code parameters. The steps necessary to establish the numbers are explained below.

For block codes, the number of information bits forms a unit called a word. The probability of error P(word) for a word is the parameter of interest to the recipient. In the uncoded case, this is established in terms of the individual bit error probability p and the binomial distribution:

$$P(\text{word}) = \sum_{j=1}^{k} \binom{k}{j} p^j (1 - p)^{k-j} = 1 - (1 - p)^k \tag{5}$$

The bit error probability is in turn expressed through the p versus E/n_o curve for the modulation method used. Thus,

$$p = f(E/n_o) \tag{6}$$

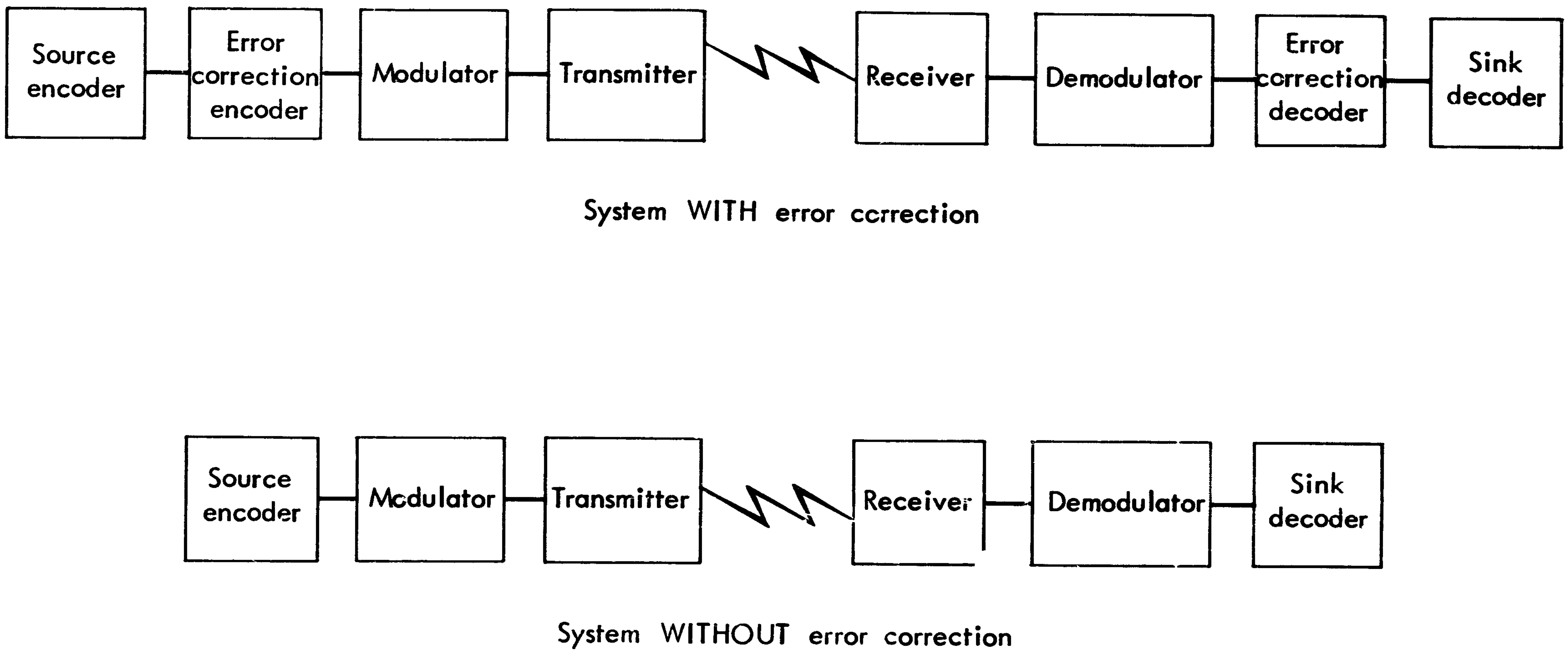


Fig. 14—System configurations for calculation of error correction performance

Coding is now introduced, changing the number of transmitted bits per unit time from k to n. Thus, the energy per bit is reduced to

$$E' = \frac{k}{n} E \tag{7}$$

Hence, the bit error rate at the output of the demodulator is now

$$p' = f(E'/n_o) \tag{8}$$

But now the probability of word error is zero unless the number of errors exceeds the correction capability e of the code. This is expressed by

$$p'(\text{word}) = \sum_{j=e+1}^{n} \binom{n}{j} p'^{j} (1 - p')^{n-j} \tag{9}$$

This expression is not exactly true for an actual code since some of the e + 1 errors may still be corrected by the decoder (it cannot guarantee to correct all e + 1 errors). Furthermore, a few of the error patterns do not affect information bits. The indirectness of the performance calculation should now be apparent. The following are the steps to be performed:

1. From the original system calculations, find p by using Eq. (6).

2. From p, find P(word) by Eq. (5).

3. Using the code parameters, find the p' in Eq. (9) which yields the same P(word) as in Step 2. In general $p' > p$.

4. For this p', find the new E'/n_o from Eq. (8), $E' < E$.

5. The improvement measured as a power ratio is given by E/E'.

This procedure is sufficiently involved that few such calculations have been made for actual codes. Instead, the performance is measured in terms of how much lower the error probability will be with coding for the same channel bit rate. This violates a fundamental rule in comparing systems which must hold information rate, not bit rate in the channel, constant. On inspection, the spectacular improvements often noted and attributed to coding are usually found to be available without coding by raising power only a few decibels. To show this, the above procedure was carried out for a few selected codes. The results are summarized in Table 3. The column entitled "Improvement" gives the additional relative power required of an uncoded system to match the coded system for the specified word error rate. These improvements assume noncoherent FSK as the modulation method. For other modulation techniques, the numbers are similar; that is, improvements are relatively insensitive to the actual modulation method.

For a fixed probability of error per received bit, increasing the block length n proportionately increases the expected number of errors. This fact discourages the development of really long codes since complexity increases approximately exponentially with the number of errors being corrected. The only argument in favor of long codes is that for a fixed error correction capability, the number of check bits decreases in proportion to the total number of bits, thus increasing the rate r of the code. The situation is different with burst noise situations, but as has been mentioned, these are not applicable to the ground-satellite-ground links being considered here.

Table 3

IMPROVEMENTS DUE TO SELECTED CODES

Code Parameters					Improvement (NCFSK) db	
Block Length n	Info Bits k	Errors Corrected e	Distance d	Description	Output Word Error Prob = 10^{-3}	Output Word Error Prob = 10^{-6}
31	26	1	3	BCH[a]	1.04	1.24
63	51	2	5	BCH	1.46	1.80
255	207	6	13	BCH	1.92	2.57
23	12	3	7	Golay	2.05	2.44
41	21	4	9	Prange-Mattson	2.11	2.62

[a]BCH = Bose-Chaudhuri-Hocquenghem

NON-BLOCK CODES: CONVOLUTIONAL ENCODING AND SEQUENTIAL DECODING

Redundancy must be inserted to correct errors. The method used ought to be simple to implement and to analyze, but the overall goal is the correction of errors. Thus, the decoder which operates on the redundancy to detect and correct errors is the focal point. The encoder is usually simpler since it operates with all inputs being certain. Sequential decoding is an alternative to block decoding, although for practical reasons there is an "active" block of incoming bits being saved. This method is also called "probabilistic decoding," indicating the importance of probability estimates to the decoding technique. Convolutional encoding is one means used to insert the redundancy, and as mentioned above, it is the simplest process.

Consider an incoming stream of message bits. The output of the encoder consists of the same bits, shortened in time to make room for the redundant bits. The redundant bits are inserted at a rate of 1 for every k information bits. Thus, a rate reduction of k/(k + 1) is established. If k = 4, for instance, the output is 5 bits for every 4 input bits. However, the output bit is not merely a function of the 4 current information bits, as in a block code. It may depend on them, but in general it depends on several of M bits in the past. The encoder is said to have a memory M bits long, and M may have a value of 100 in a practical system.

A complicated series of events takes place at the decoder. On the basis of the past M information bits decoded, together with the actual recomputed redundant bits, the incoming bit is examined. If it is agreed that there is high probability that the current bit has been correctly decoded (or that it has been verified if it is a redundant bit), the decoder rests, passing the bit on to the memory. However, any lack of agreement at this point results in a reexamination of the least likely bit accepted. This bit is changed, and the new situation is examined. If an improvement occurs, the change may be kept or other bits may be examined with an eye to changing them to improve the probability that the current bit is valid. As an older bit sinks further into the memory, its weight in current decision-making is reduced until finally it reaches the end of the line and is emitted as a supposedly correct bit. When M is large, the chance of a bit reaching the end of memory without having its correctness challenged diminishes rapidly if the redundant bits depend upon its value with very great frequency. This process of challenging the old bits in the decoder is the chief characteristic of probabilistic decoding.

Note that no information bit is made available until M bits have been stored and have helped play role in the decoding process. Then, due largely to economic reasons, the M + 1 bit can no longer be retained. The computations involved in this procedure call for a special-purpose computer capable of executing many thousands of instructions (the number is dependent on M) in the time period during which a single bit is received. Also, storage is required to keep track of the probabilities at each stage so that errors can be recovered quickly and updated. For a detailed discussion of the theory and one practical form, see Refs. 16 and 17.

To apply sequential decoding to a system, it is necessary to know relationships between bit error rate in the channel and the probability of error at the decoder output. These should, of course, be functions of the memory length M, the redundancy as measured by k, etc. Unfortunately, the complexity of existing systems makes useful calculations very difficult. In most applications, the goal has been to reduce errors to zero* rather than some low value, and the system was made sufficiently complex to do this within the measurement interval. The principal reason for this is that once an error has been accepted, it affects subsequent decoding adversely. Decoding becomes meaningless once an error has been accepted. Sequential decoding has therefore been applied to systems having two-way capability and using information feedback. Once decoding becomes unreliable it must be terminated and reestablished with correct information.

ERROR DETECTION

Thus far discussion has centered on error correction in communication channels. In many applications error detection, followed by message retransmission, is a desirable alternative. This is especially true where a feedback channel exists which can be used to signal retransmission and where non-real-time input permits holding of messages. In communicating contents of punched cards, teletype characters, or magnetic tape, buffers are commonly used at each end, and until the buffer is emptied, there is in principle no reason any number of

*P(error) = 10^{-20}, which is zero for all practical considerations.

attempts cannot be made to get a correct message. Error detection is generally much easier to implement than error correction and at the same time more powerful.

On analog channels operating in real time, error detection can be used to eliminate the reading or use of incorrect bits. In transmission of speech by pulse code modulation (PCM), for example, inhibiting reading and holding the previous sample value for an additional period is a better practice than decoding a sample containing random errors. Error detection can be done by feeding the information bits as received into an encoder at the receiver and comparing the parity bits resulting with those received. If a match occurs, either the message is correct or enough errors have occurred to convert one message into another completely and correctly. Proper choice of code can reduce the latter probability to an arbitrarily small value. Peterson and Brown[18] give an excellent discussion of the capabilities of modern codes for error detection.

V. ANALOG-TO-DIGITAL CONVERSION AND ITS INVERSE

INTRODUCTION

In this Memorandum, emphasis has been placed on signals whose origin is either fundamentally digital (e.g., business machine, teletypewriter, or computer signals) or has somehow been digitized. In this section, some of the processes for converting analog signals to digital ones are investigated, and related problems of fidelity, bandwidth, bit rate, etc., are discussed.

All analog signals have bandwidths which are limited even though very large. In practical situations, the bandwidth may be considered to to be that necessary to account for, say, 99 percent of the energy. Then, although the missing 1 percent will be lost or garbled, or often folded over onto the 99 percent, thus distorting it, the net effect on the received signal is usually not objectionable. It is by such a means, or by straightforward filtering, that voice bandwidth is taken to be 4 kHz.

As soon as the bandwidth is well defined, the digitizing process can begin by passing to the mathematically equivalent discrete time-sampled version of the analog signal. This can be done since, according to the sampling theorem, samples at a rate of 2B samples per second completely define an analog signal restricted to a band which is B Hz wide. This process results in discrete time samples, each of which has a continuum of amplitudes. A completely discrete signal results when these amplitudes are quantized.

Unlike the sampling process, quantization does introduce distortion. This distortion can be made arbitrarily small simply by reducing the amplitude samples to finer and finer intervals. This distortion, being highly correlated with the signal, has effects on the listener which are substantially different from those due to random noise. In particular, when there is no signal, there is no distortion or background noise. On a voice channel, this quantization has the effect of a granularity and interferes with the subtler aspects of speech associated with voice recognition. Common practice in commercial design has been to set quantization distortion power equal in magnitude to the noise power resulting from bit errors, so that each causes half the total system noise at the greatest maximum external noise threshold. As a design value this is probably not bad, but better measurements of the subjective effects of quantization are needed to enable this parameter to be more accurately specified.

Quantization distortion can be estimated quantitatively by computing the mean power in the error. If the levels are uniformly spaced in amplitude, the mean-square quantization error power is given for a wide variety of signal distributions by

$$W_Q = \frac{1}{12} \Delta^2$$

where Δ is the amplitude of a single level. If the levels are not uniformly spaced, it is necessary to evaluate the probability of error by a more elaborate procedure. Piecewise linear quantizers are convenient to implement. These are characterized by an input-output relation consisting of line segments with different slopes joined together. Such a relation is shown in Fig. 15.

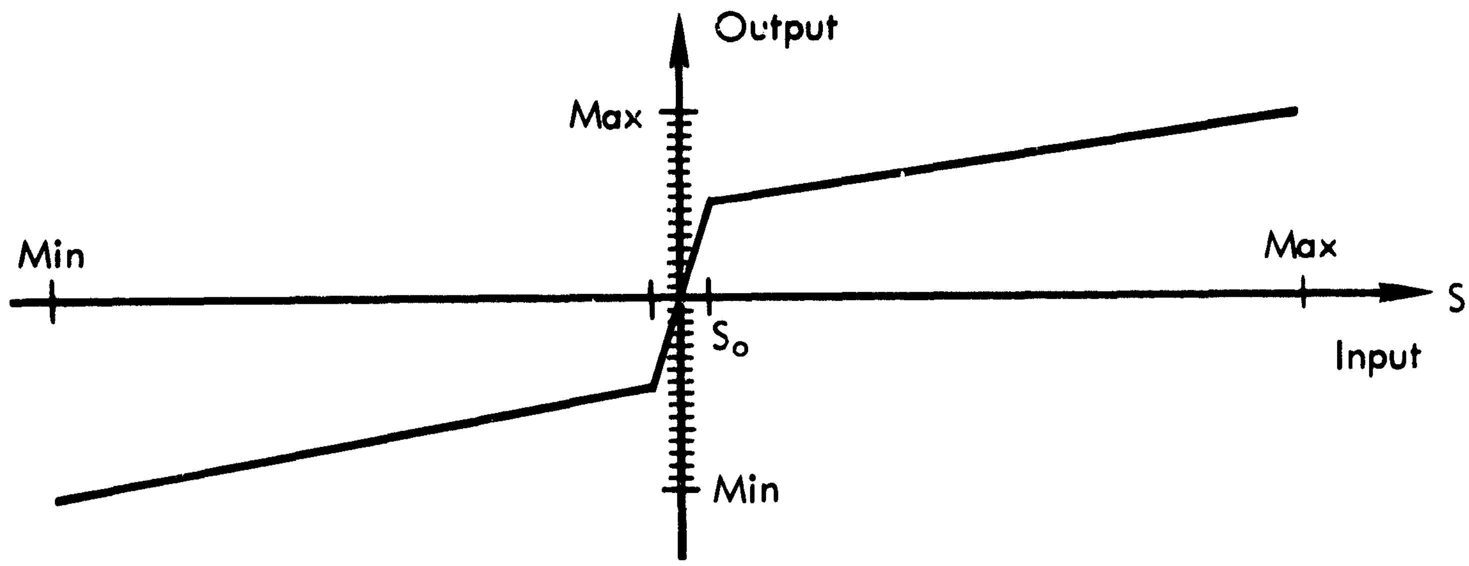


Fig. 15—Nonlinear encoder

The quantization noise power is still calculated by a simple relation. Thus,

$$W_Q = \frac{\Delta_1^2}{12} \cdot \text{Prob}\ (|S| \leq S_o) + \frac{\Delta_2^2}{12} \cdot \text{Prob}\ (|S| > S_o)$$

where Δ_1 and Δ_2 are the increments in signal corresponding to one level output difference in the regions $|S| \leq S_o$ and $|S| > S_o$ respectively. An encoder of this form is used with speech in cases in which peaks are often several times the rms amplitude, but Prob $(|S| > S_o)$ is near .5 even though Δ_2 may, for example, be equal to $20\Delta_1$.

DIGITIZING SPEECH OR SIMILAR ANALOG WAVEFORMS

There are two basic methods of forming a quantized digital signal from an analog signal for transmission over digital communication systems. The first is pulse code modulation, and the second is delta modulation (ΔM). Neither is a modulation in the conventional sense of the term, but merely a conversion process which makes an analog waveform into a digital form. The resulting digital forms are substantially different. PCM represents analog samples in the binary number system in the form of L-digit words, while ΔM provides a signal which can be processed by an integration-filtering operation. PCM requires synchronization so that the words can be decoded and the amplitude samples can be reconstructed, while ΔM is inherently asynchronous. This simplifies system design somewhat.

With speech waveforms, periods of silence are common when the speaker is listening or pausing, and periods of time with no input to either PCM or ΔM result in distinctive output waveforms. In some cases, it is preferable that no output occur when no input occurs. This can be implemented by special digital converters which reorganize the output into the desired form.

PULSE CODE MODULATION

Pulse code modulation was first used commercially around 1948, though it was invented nearly ten years earlier.* Early encoders used special electron beam tubes with an encoding matrix etched on a collector electrode. The signal was used to deflect the beam to an appropriate position. For ease in etching, a Gray code** was used, which was then converted to conventional binary form. From this early start,

---

*The history is described in Ref. 19.

**A Gray code has the property that small amplitude changes result in small Hamming distance changes. In straight binary notation, for example, an amplitude change of 1 from 31 to 32 causes 6 bits to change (i.e., 011111 becomes 100000).

simpler techniques resulted, and today many new techniques are being announced, solid-state technology having made many different approaches feasible. To cite one example, television signals have been encoded with tunnel diode comparators and high-speed logic.[20]

One common type of encoder is shown in Fig. 16. The signal sample is placed at one input to a comparator circuit. Then, by logical feedback, the digital-to-analog converter is made to produce an equivalent sample, quantized to the nearest level. The result of the comparator output is then used as the output signal. The receiver decoder is an identical digital-to-analog device, and in the absence of errors, will produce the sample value that best matches the true sample at the transmitter. This use of the receiver element in feedback circuitry will also be noted in delta modulation circuits.

The encoding process for a 64-level (6-bit) encoder can be described briefly. Suppose the input sample has an amplitude corresponding to the 37th level. The control circuit (CC) begins automatically by setting a reference of $1/2 \times 64 = 32$ levels. The comparator output is then 1. This enters the CC and creates the first bit of the output. The CC, having received 1, retains the 32 level and calls for $1/4 \times 64$ more. The new reference is then 48 levels, and the comparator output is 0. The CC discards the 16-level addition and supplies 8 levels, making a total of 40 at the reference point. The comparator output is again 0, the 8 disappears, and 4 is added. The output now becomes 36. The comparator yields a 1, the 4 is held, and 2 is added. The reference of 38 causes 0 in the output, the 2 is dropped and 1 added, whence the comparator output is 1, and the encoding operation is completed. The comparator output

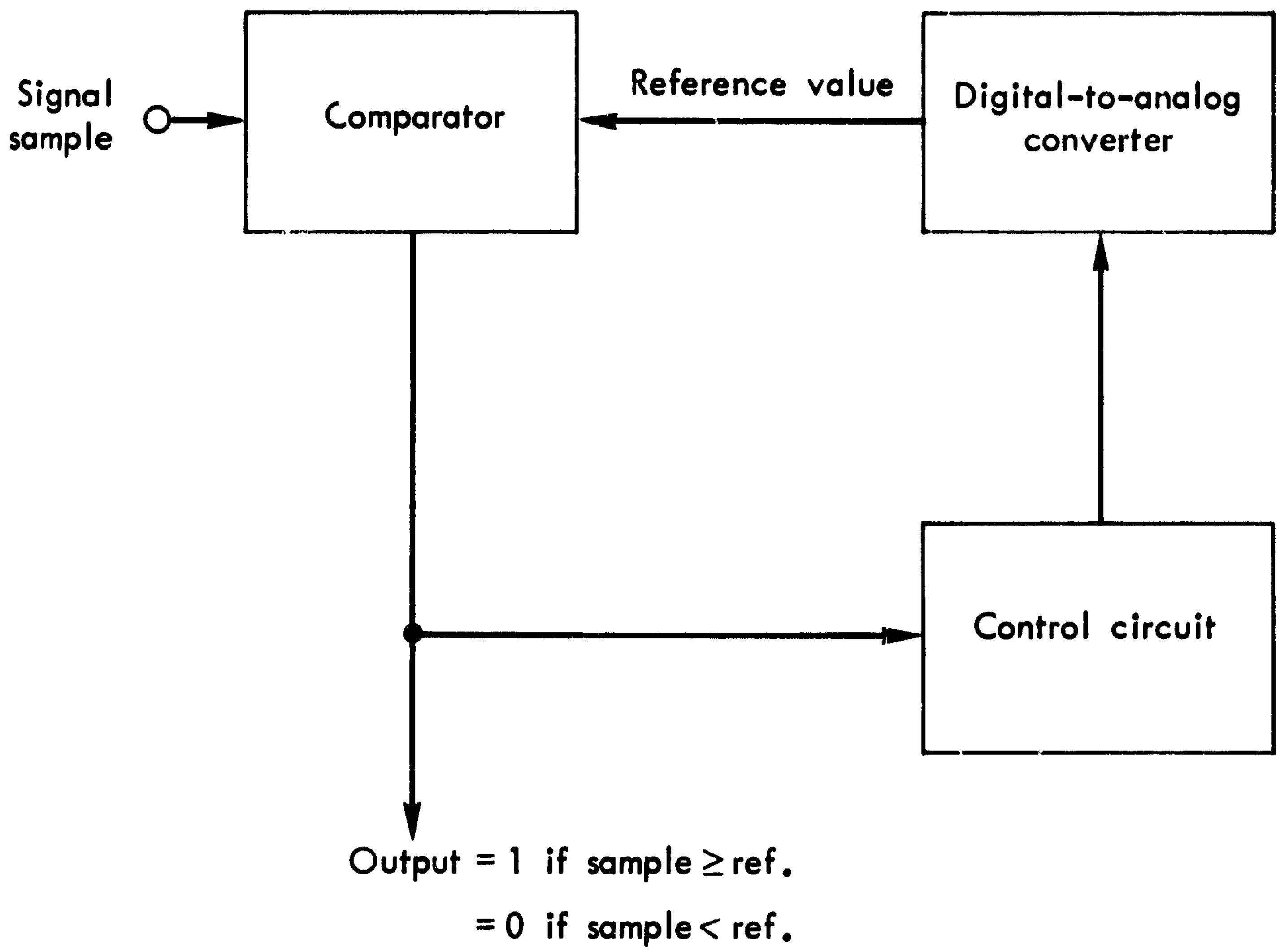


Fig. 16—PCM encoder—simplified form

sequence is 100101, and the binary number corresponds to 37. Note the simple algorithm: Try 1/2 the preceding level each time, drop the preceding level if comparator output is 0; do this 6 times in a 64-level encoder.

The digital-to-analog converter may take many forms. One which might have been used in the above example is shown in Fig. 17. The solid-state switches* are usually in the form of specially biased transistors which conduct either to the ground or to a reference voltage, according to the logic level presented at their input. The output is the binary weighted sum of those switches which are connected to the stable reference voltage supply.

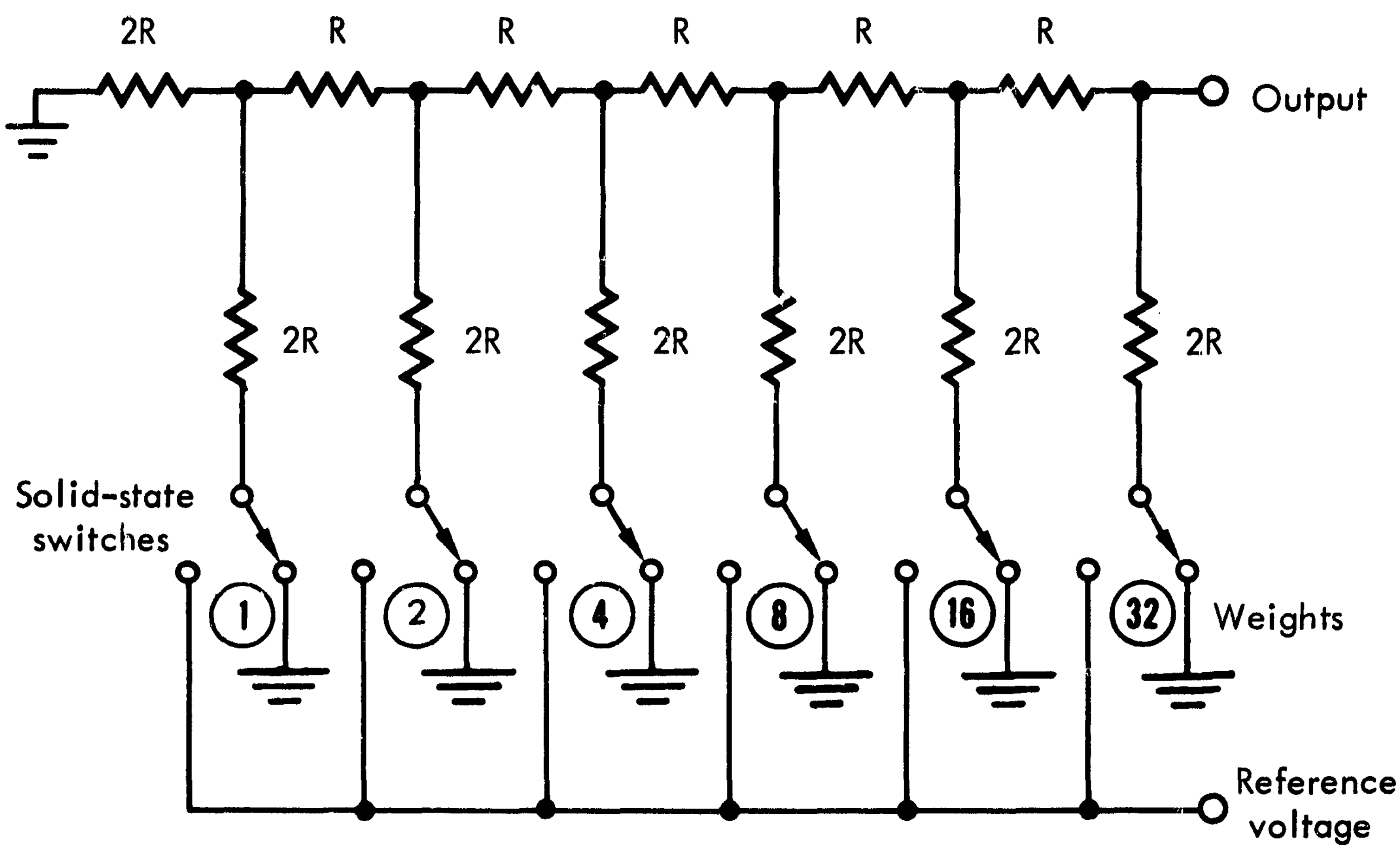


Fig. 17—Digital-to-analog converter

*In Fig. 17, switch arms are controlled by control circuit logic in the transmitter and by shift register contents in the receiver.

The simplest receiver consists of a shift register into which the binary digits are shifted. Then when all are present, the parallel outputs of the registers operate the switches of the digital-to-analog converter, forming the analog value at the output. This analog value is held on a capacitor until the next sample is available. Timing is supplied by synchronization circuitry. Figure 18 shows the receiver.

Performance of PCM Systems

As already mentioned, PCM systems suffer two sources of distortion. The first occurs in the quantization process and results in the granularity effect. Increasing the number of digits used to represent the sample value will reduce this effect, and commercial encoders capable of 15-bit or .003-percent accuracy are available. Such accuracies are not needed for speech, but they are often valued in critical telemetering applications. For speech, seven or eight binary digits

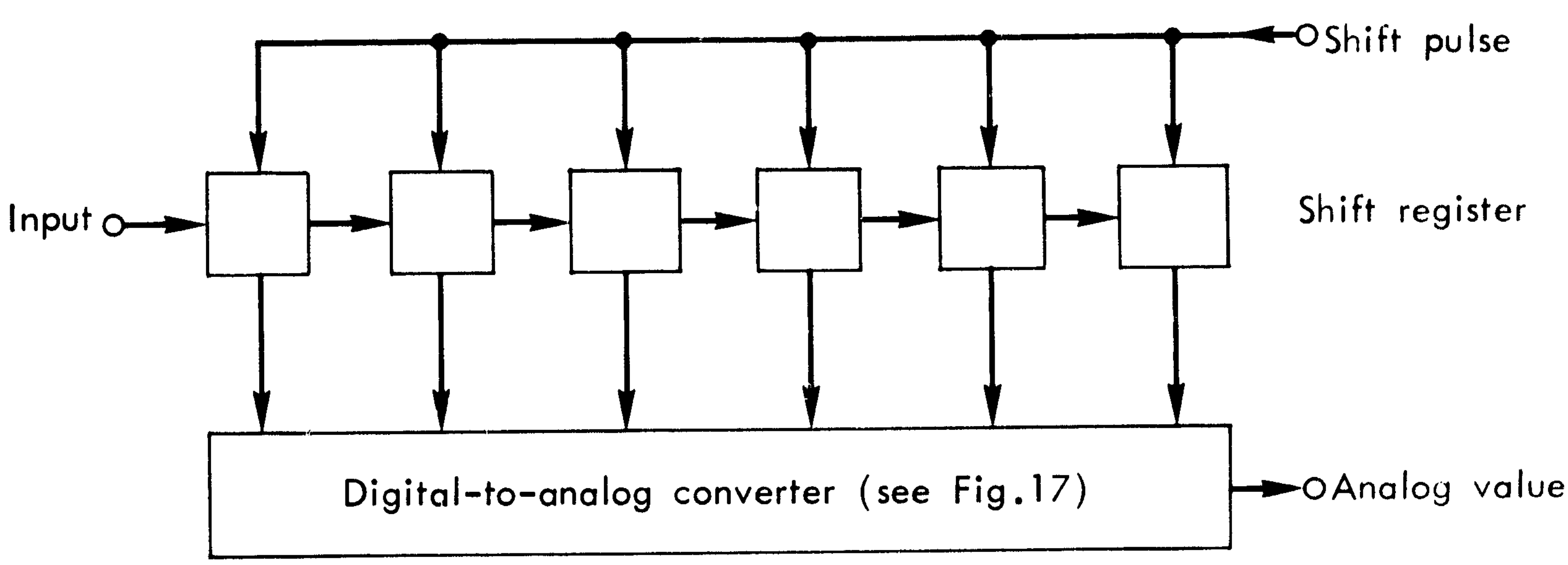


Fig. 18—PCM decoder

are adequate for good quality. An optimum, or matched, compressor-expander (often called a "compander") will make all levels equally likely, and the apparent mean signal to mean distortion power is given by[21]

$$\left(\frac{S}{N}\right)_Q \approx 4^L$$

where L is the number of digits used. In decibels, S/N is approximately six times the number of digits. The relation is approximate but holds very accurately for cases in which $L > 3$, which are those of most interest. The system designer chooses L so that an adequate quality signal is provided in the absence of error, that is, in the strong signal situation.

The second source of distortion in the received signal is errors in the received bits. This noise is present with or without a signal and reflects the effect of random errors in the transmission channel. When the probability of error per bit, p, is constant and independent from bit to bit, it can be shown[21] that the ratio of mean signal to mean error-noise powers is given by

$$\left(\frac{S}{N}\right)_e = \frac{1}{4p(1 - p)} - 1$$

For large signal-to-noise ratios, $p \ll 1$, and the expression simplifies to

$$\left(\frac{S}{N}\right)_e \approx \frac{1}{4p}$$

For $p = 10^{-5}$, for example, $(S/N)_e = 44$ db. Curves of the output signal-to-noise ratio versus the input signal-to-noise ratio measured in the information bandwidth can be prepared by using the appropriate detection characteristic (from Section II or IV) and eliminating the error probability. The result is shown in Fig. 19, in which the solid curve involves only $(S/N)_e$ and noncoherent FSK is used. When distortion is included, the sum $(S/N)_T$ will follow the dashed curve which depends on the number of levels. In Fig. 19 the abscissa is equal to $2L\ (E/n_o)$. Thus, it can be seen that for PCM, S/N levels out to a plateau fixed by the quantization distortion. The use and significance of companding will be discussed later in this section, where it will be noted that speech is particularly in need of such a processing prior to encoding.

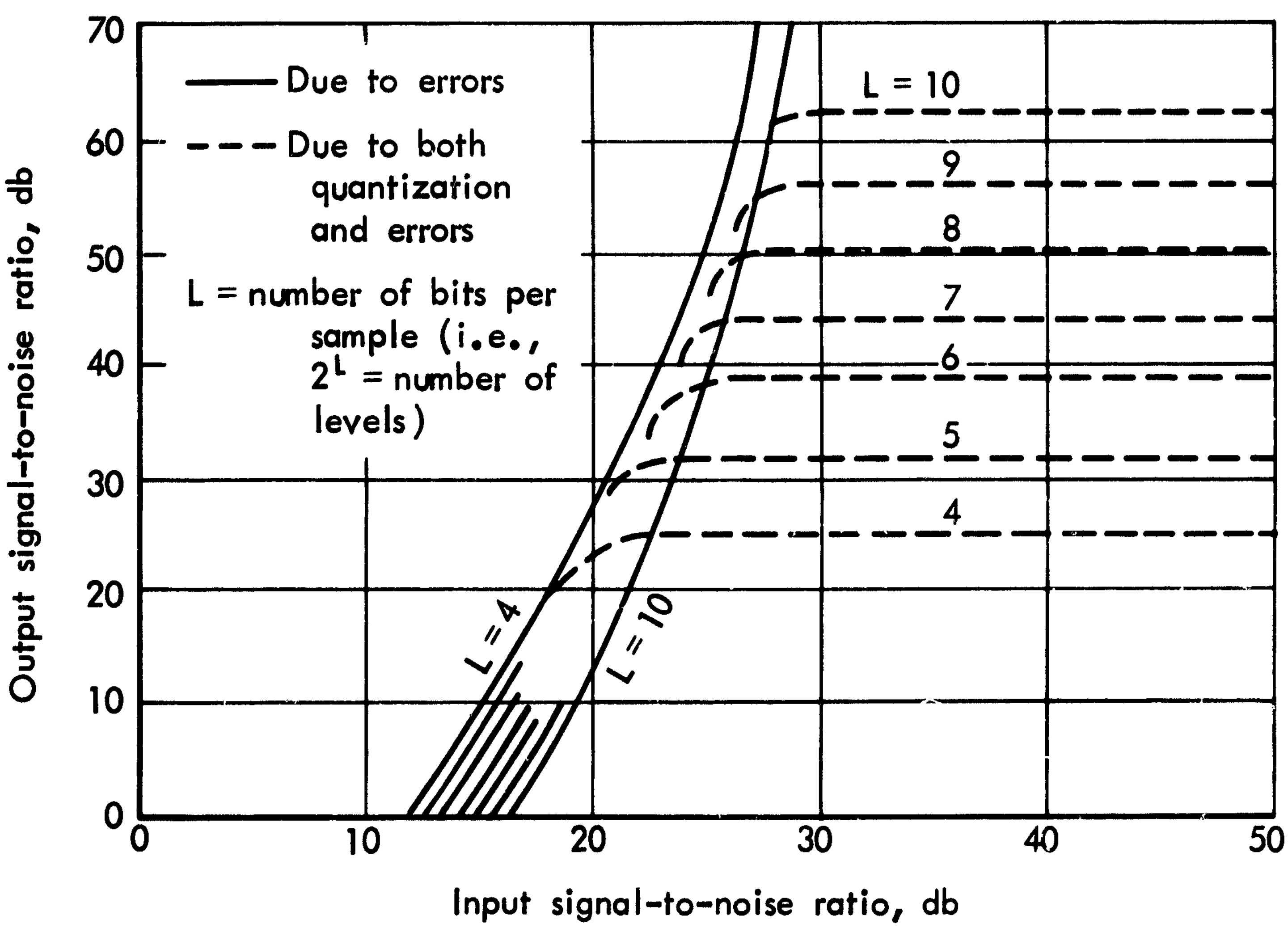


Fig. 19—PCM input-output characteristics

DELTA MODULATION

Sometimes the main goal of an analog-to-digital converter is merely a digital stream containing the information for later restoration into the original analog form, and the intermediate digital form need not play a role as such. In such a situation (encoding and decoding speech is one) only a digital stream of some sort is required. Delta modulation bypasses the sampling and conversion processes by constructing a digital pulse stream which can be operated upon directly, e.g., by means of integration and filtering, to produce an analog output which is equivalent to that at the transmitter. The integration and filtering network is sometimes moderately complicated, but the principles can be described by taking a simple special case.

Consider the output of the simple RC integrator of Fig. 20, when the input is a ±1 digital waveform. The duration of one pulse period T is short compared to $\frac{1}{RC}$, so pulse energies accumulate on the capacitor. A long series of +1 pulses will result in C charging to the peak value of 1. Changing to a sequence of minus pulses will cause the capacitor output to decrease, ultimately approaching -1, depending on the time

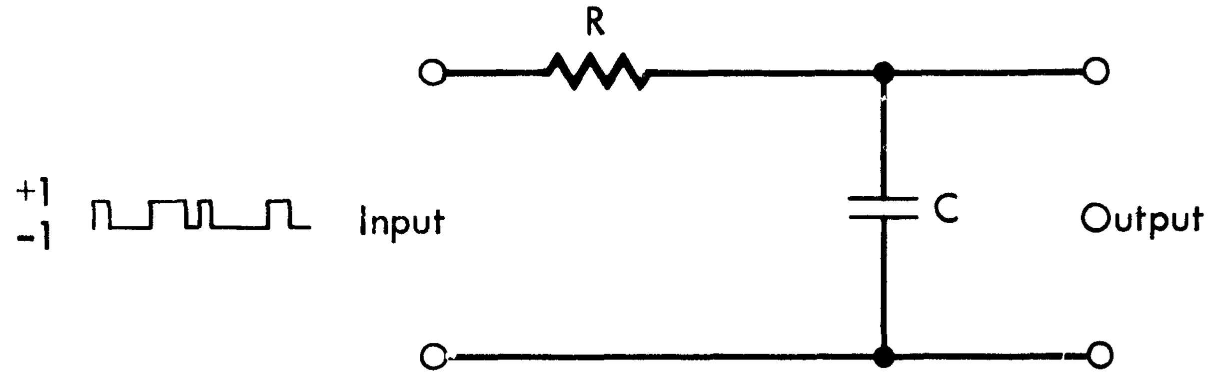


Fig.20—A simple RC integrator

constant. For combinations of ±1 in various sequences, different rising and falling values will be noted across C. Thus, any analog waveform restricted to ±1 range may be approximated with a suitable pulse train, assuming proper choice of RC. Now note how this circuit is used in the feedback system shown in Fig. 21.

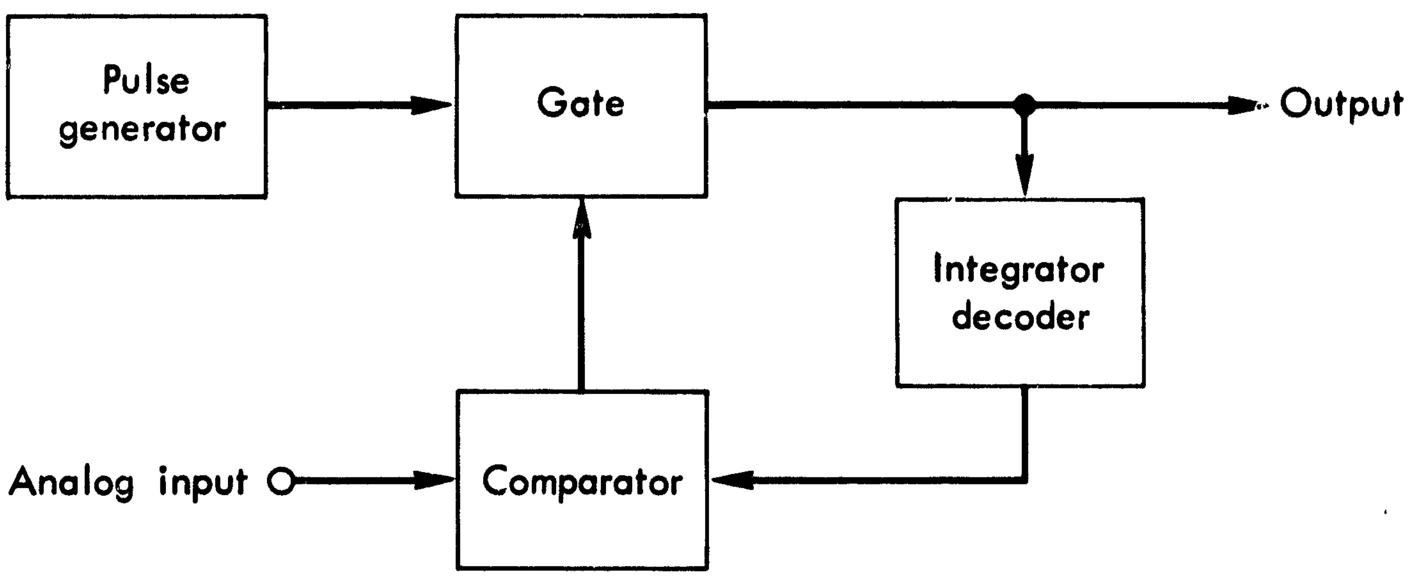


Fig.21—Delta modulation encoder

Here the purpose is to control the pulse output so that upon integration the pulse waveform produces a close approximation of the input analog waveform. The receiver consists of an identical integrator, which should then yield this best digital approximated waveform. There are several drawbacks in this system which can be corrected in many different ways. If the analog waveform is speech, frequent high-peak values are encountered which a simple system, as shown above, cannot accommodate without using an unreasonably high pulse rate. Different techniques have been suggested for combating this problem. One technique is to adjust the size of pulses entering the integrator to make it a function of the number which have occurred in succession.[22]

Another approach differentiates the speech before it enters the comparator and performs a second integration at the receiver to recover the speech. Other improvements relate to the 0 input waveform. In the system described, a 0 input yields a sequence of alternating +1's and -1's. With logic devices such a sequence can be mapped into all -1's. Thus, in the absence of speech, an all -1 sequence is transmitted rather than a +1, -1 sequence. This technique accomplishes nothing as far as most channels are concerned, since sending +1's or -1's results in identical signal content. But in AM transmission systems, all -1's may correspond to an idle condition at the transmitter, and power savings are thus realized.

Performance of Delta Modulation Schemes

There is no simple relationship between input and output in a ΔM system as there is in PCM. It is therefore more difficult to estimate the performance of ΔM systems in a meaningful quantitative manner. It is more common to cite qualitative experiments or to point to qualitative features of a system. The simple systems, whose performance can be most accurately described, are poor enough that more sophisticated systems, whose performance can only be measured, replace them. For the most part, the decoding integrator-filter is the key element. How a ΔM system may be evaluated in general is discussed here, though only the properties of the simplest such system can be numerically evaluated.

Low-frequency waves are sampled many more times per cycle than high-frequency waves. This leads to a quantization characteristic which favors low frequencies and becomes worse at the high-frequency end of the signal spectrum. This is sometimes countered by differentiation of the input waveform, which rearranges the relative content of high and low frequencies; then integration at the receiver corrects the condition. Integration is thus performed twice, with separate time constants, at the receiver. More sophisticated techniques, which operate on the digital content as well as the signal, are too complex to analyze in detail here.

For the simple system of single integration, the ratio of mean signal power to mean-square quantization noise power of a sine wave of frequency f at the output of the filter can be shown[23] to be

$$\frac{S}{N} = 0.04 \frac{f_i^3}{f_o f^2}$$

where f_i is the pulse frequency and f_o is the filter cutoff frequency, which is equal to $(2\pi RC)^{-1}$. The sine wave amplitude must obey a constraint imposed by slope limiting:

$$A \leq \frac{f_i}{2 f} \cdot \delta$$

where δ is the step amplitude. This latter constraint is fortunately appropriate to speech with its decreasing amplitudes at higher frequencies.

To assess noise due to transmission errors the following procedure may be applied. Since the demodulator is a linear circuit, the noise output can be evaluated by calculating the response to a waveform resulting from noise only. So assume that noise causes bits to be changed from 0 to 1 with probability p_o, and from 1 to 0 with probability p_1. On the binary symmetric channel $p_o = p_1$, of course.

A member of the noise ensemble might resemble Fig. 22 for a high error probability. Knowing the spectral density of this waveform, the output noise in the demodulator can be evaluated. But the spectral density is very easy to calculate. Assume that noise-caused errors are independent. Then, calling the waveform x(t), $E[x(t)] = p_o - p_1$, the process being discrete stationary.

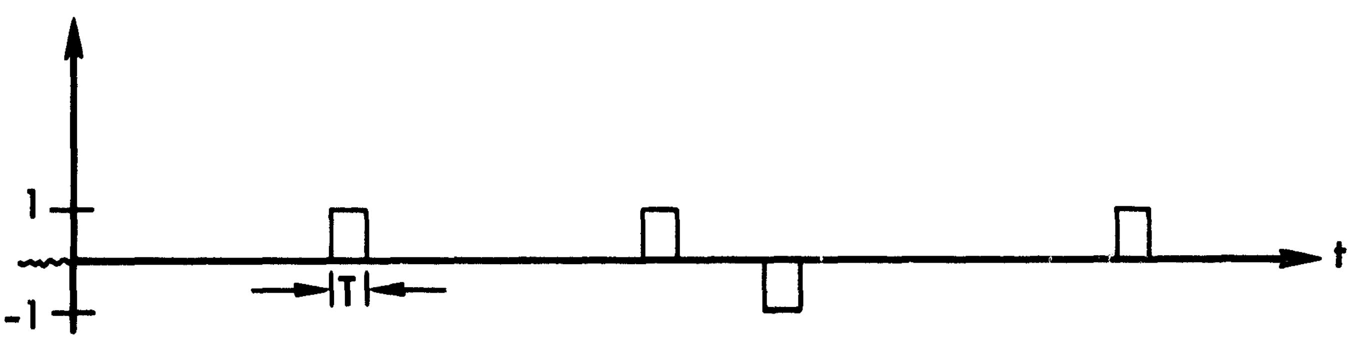


Fig. 22—Member of a noise ensemble (delta modulation)

The autocorrelation is

$$R(t,\tau) = E\left[x(t)\cdot x(t+\tau)\right] = (p_o + p_1) \quad \text{for } t \text{ and } t+\tau \text{ in same interval}$$

$$= (p_o - p_1)^2 \quad \text{for } t \text{ and } t+\tau \text{ in different interval}$$

The corresponding power spectral density is

$$S(\omega) = 2\pi(p_o - p_1)^2\, \delta(\omega) + \left[p_o + p_1 - (p_o - p_1)^2\right] \cdot 2T \frac{\sin \omega T}{T}$$

Ignoring the DC term and noting that $\omega T \ll 1$ for all frequencies of interest in the output, the simplified and constant power spectral density is

$$S(\omega) \approx 2T\left[(p_o + p_1) - (p_o - p_1)^2\right]$$

The noise bandwidth of the simple RC integrator of Fig. 20 is $B = \pi/(2RC)$. The total noise-power output (including frequencies too low to be audible) is then

$$N = S(\omega) \cdot B = \frac{\pi T}{RC}\left[(p_o + p_1) - (p_o - p_1)^2\right]$$

At a frequency f, the slope-limited demodulator permits a maximum amplitude of

$$A = (2\pi f T)^{-1}$$

Then a sine wave signal-power to mean-noise-power ratio becomes

$$\frac{S}{N} = \frac{RC}{(2\pi)^3\ f^2\ T^3\ K}$$

with $K = \left[(p_o + p_1) - (p_o - p_1)^2\right]$. On substituting typical numbers, S/N at frequencies 300, 800, and 3000 Hz is 72.2, 63.8, and 52.2 db respectively, for $p_o = p_1 = 1/2 \times 10^{-3}$, $T = 20\ \mu sec$ (corresponding to a 50-kHz pulse rate) and RC = 3 msec.

COMPANDING

Instantaneous companding is used in PCM voice encoding systems to help optimize the signal-to-noise ratio and channel information rate. It is well known that speech has a nonuniform amplitude distribution which is far from gaussian in character. A uniformly distributed input waveform optimizes channel rate and offers maximum S/N and minimum distortion. Companders attempt to change these undesirable characteristics of voice in a reversible way. This can be accomplished in several ways. One efficient method is to build the necessary characteristic directly into the encoder. The encoder then operates with nonlinearly spaced levels. Another method is to perform a nonlinear operation directly on the voice before it reaches the encoder. The desirable result of this operation, however it is performed, is to change the peak-to-average amplitude ratio of voice from about 10 to 1, to a factor closer to 2 to 1. Subjectively, the effect is to reduce the apparent noise due to quantization by making more levels participate in transmitting the voice signal.

Companders are not used in ΔM systems as separate elements, though their effect may be simulated by processing the speech and using nonlinear encoders. The output of the simple demodulator-integrator cannot supply the abrupt peaks occurring in speech. Thus, they are usually clipped, and some distortion results.

The performance of a compander is sometimes expressed by comparing the signal-to-noise ratios with and without companding for a fixed amount of peak clipping. Improvements varying from 8 to 15 db are sometimes quoted for such performance. A logarithmic compressor and exponential expander can be approximated by a diode-resistor attenuation network, while similar results can be obtained from the nonlinear digital encoder. A two-step approximation to the logarithm can be obtained in the latter by making the magnitude of subsequent reference levels depend on the decision made at the first digit.

Table 4 shows how the peak-to-average power ratio varies for different distributions. From this table it is clear that speech has an extraordinary number of peaks and that companding to a uniform distribution, for example, could yield an improvement of $(18 - 4.8) \approx 13$ db.

Table 4

WAVEFORM PEAK-TO-AVERAGE POWER CHARACTERISTICS

Distribution	Waveform	Peak-to-Average Power Ratio Value	db	Peak Definition
Delta	Square wave	1	0	Natural
Sinusoid	Sine wave test tone	2	3	Natural
Uniform	Carefully companded	3	4.8	Natural
Gaussian	Gaussian noise, Multichannel baseband	10.9	10.4	Exceeded 0.1 percent
Speech	Single talker	63	18	Exceeded 0.1 percent

ENCODING FDM BASEBANDS

It is technically feasible to encode a signal composed of frequency-division multiplexed channels. When these channels are voice channels, some desirable properties result. First, it is noted that the statistical properties of voice are such that an FDM signal composed of M active voice channels is nearly gaussian for large M and has a peak-to-rms ratio appreciably less than a single-voice channel.* With linear (i.e., uncompanded) encoding, therefore, fewer bits are needed per channel for the same channel output characteristics. A difference occurs when quantization noise is considered. With single-channel encoding, the quantization noise is strictly correlated with the channel signal. In multichannel encoding, the quantization noise power is, essentially, distributed uniformly over all channels including

*These properties are discussed in detail in Ref. 10.

those which are idle. Thus, the effect is more like gaussian noise or crosstalk, and the granularity mentioned for single channels is not present.

Currently, instantaneous companding of an entire baseband is possible only with nonlinear encoders. Separate analog companders introduce serious distortion which appears as crosstalk in all channels.

Simple relations can be derived for the channel signal-to-noise properties when frequency-division multiplex is encoded by PCM. These relations require knowledge of the statistical factors concerning the multiplexing of speech. The error and quantization noise are assumed to be uniformly distributed over all channels. Taking only quantization noise into account, the channel signal-to-noise power ratio is

$$\left(\frac{S}{N}\right)_{ch} = 3M \cdot 4^{L} \frac{\overline{P}_{TT}}{P_c}$$

where M is the number of channels, L the number of digits, $\overline{P}_{TT}$ the mean test-tone power, and P_c the peak single-channel power. Some of the numbers involved in this relation are summarized in Table 5. Of greatest interest are the decreasing number of digits required and the corresponding decrease in the individual channel rates. These figures are for channels in which no companding takes place. For large numbers of channels, companding offers no advantage since the baseband is already satisfactory statistically. For only a few, say 12, voice channels, companding might offer a modest improvement by reducing the number of bits required per sample by one or two.

Table 5

PCM ENCODING A BASEBAND OF VOICE CHANNELS

Channels	Bits/Sample	(PTT/P_c) db	Channel Output S/N db	Rate Per Voice Channel, Kilo-bits Per Second
1	8	-14.0	39	64
1	9	-14.0	45	72
12	7	-18.0	40	56
60	6	-19.2	39	48
60	7	-19.2	46	56
240	6	-21.3	43	48

VI. SYNCHRONIZATION

All of the schemes discussed in the preceding sections require some kind of time reference. Providing such a reference is the task of the synchronization circuitry. In analog systems, synchronization is needed if coherent detection is used. In frequency-division multiplex a pilot tone is usually transmitted, from which all demodulating frequencies are derived by synthesis techniques. In time-division and digital systems, synchronization circuitry performs the following jobs:

1. It provides a clock (bit rate) for timing all receiving operations.
2. It provides a time reference so specific bits can be associated with specific meaning.
3. It corrects for timing deviations in the propagating medium.
4. It maintains all the above when one or both terminals are in relative motion with respect to each other.

In communication satellite systems, 3 and 4 are especially important and in some ways unique. Successful performance of digital or time-division systems will depend on the successful solution of the problems presented by these requirements. Tasks 1 and 2 are required of any digital system and are normally performed by relatively simple circuitry. A clock can be established by using the received waveform and filtering it to derive an average frequency, which is used to lock a local clock oscillator. Time reference is sometimes established by looking for specific transmitted patterns sent as "synchronization signals."
Once such a time reference is established (e.g., at the beginning of a transmission), only a loss of signal would make regaining it necessary.

In other words, if the clocks at receiver and transmitter are independently keeping perfect time and ticking in synchronism, there need be only one synchronization signal sent. But such good clocks are not yet available, so reference information is transmitted periodically over the channel. The period depends mainly on the memory which the design engineer is willing to provide in the circuitry.

Communication satellite systems offer some unusual opportunities for synchronization. Because all stations using a given satellite see the same output, though propagated over different paths, the satellite can provide a common clock reference as well as a recurring timing pulse. This is probably better than making one station a "master" and all others "slaves" because only one propagation path is involved for each station, and it is essentially the same one both to and from the satellite. In the master-slave system, all slaves must continuously correct for both paths. Providing a stable clock in a satellite has already been proved practical with navigation satellites. Stabilities near 10^{-10} are achieved.

Synchronization is relatively simple to achieve in a communication satellite system when all stations have the capability of viewing their own transmissions as they appear to the intended recipients. Using this apparent feedback, timing can be refined appreciably over any nonfeedback system in which only estimates can be made of proper timing relationships.

A quantitative measure of synchronization performance is difficult to secure. Sometimes initial time to first synchronization is important. In time-division multiplex systems, the ability to measure time accurately becomes important since it affects the number and width of guard slots, which are the time intervals allotted to accommodate errors in timing. But precision time measurement takes a long time since measurement accuracy is generally proportional to the square root of the number of independent observations used. Synchronization signals which are subject to serious perturbations due to propagation effects or relative motion of stations must offer good signal-to-noise ratios if this accuracy is to be exploited.

Unfortunately, there are few generalities which can be written concerning synchronization systems. Sometimes they affect a major portion of the transmission, as in start-stop teletype systems, for example. Sometimes they occupy their own channel. Systems referred to as "asynchronous" are frequently found to be merely "self-synchronous." Different authors use the term "synchronism" independently to refer to different levels of synchronization. For example, a start-stop teletype system is called asynchronous because each character contains its own synchronization information independent of other characters. However, the system is still definitely synchronous, as any attempt to receive 100-word-per-minute transmissions on a 60-word-per-minute machine will show. Thus, a more accurate description of such a system would be that is it "character self-synchronous." Earlier, delta modulation was called asynchronous; a more accurate description would be to call it a "non-word-synchronous" system while PCM is "word-synchronous."

VII. CONCLUSIONS

The purpose of this Memorandum has been to review digital transmission techniques, especially those which may be applicable to the multiple-access problem. However, a study of time-division techniques for multiple access primarily involves applying those known digital techniques to a specific system. Basically, time-division multiple access is easy to achieve, but at the moment the nondigital nature of most traffic makes analog methods cheaper. This is due to the extra cost of analog-to-digital-to-analog conversion equipment. When the traffic balance changes to include more digital traffic, or when there is a distinct digital flavor to the multiple-access requirement, time-division methods will be used.

Reviewing modulation techniques, it was noted that coherent and noncoherent schemes can be used. In the multiple-access application, however, it is not determined that radio-frequency coherent systems are practical as yet, and this is an area which should be studied further. Early indications are that success can be achieved with an all-coherent multiple-use system if the satellite furnishes a timing signal which can be used universally and if the available feedback signal is sensibly used to control timing.

Time-division multiplex is especially easy to implement with currently available computer hardware. The only serious problems are deriving timing, which is a function of synchronization circuitry, and controlling intersymbol interference due to the filtering of waveforms. Both these problems can be solved but require special care.

Error correction codes are disappointing as a system improvement technique. To offer a useful performance, they must be complex and correspondingly expensive to implement. They are likely to be most useful on individual subchannels in which a lowering of error rate is required for some special application, in computer-to-computer links, for example. For analog transmission by digital means, error rates less than 10^{-6} are not required. A system designed to this specification may fail to meet the requirements of individual subchannels. Codes can then be applied to these subchannels to give whatever improvement is required. Here it is generally impractical, or even impossible, to improve an individual channel by raising power. Thus, error correction on an individual channel basis is practical and valuable. Error detection and retransmission are techniques which can be applied to non-real-time communication systems such as computer-to-computer links with excellent performance and at relatively low cost.

The area of analog-to-digital conversion has been dominated for many years by pulse code modulation, which is a direct digital representation of the analog data. Newer, and becoming more popular, is the family of techniques known as delta modulation. The latter lacks the frame synchronization requirements of PCM and is also simpler and cheaper to implement. It is well suited to speech encoding, having a distortion characteristic which is well matched to human speech. Companding techniques may be used externally to PCM or ΔM encoders and decoders, or the encoders can be designed to respond nonlinearly during the encoding and decoding processes and thus perform the companding operation.

Appendix

DERIVATION OF ERROR PROBABILITY EXPRESSIONS

CASE I: COHERENT RECEPTION (AM, FM, PM)

In Table 6, the characteristics of signal, noise, matched filter (i.e., coherent detector), and filter output components S and N are shown for each of the three modulation methods. The resulting error probability expressions are also given. Because all columns are so nearly identical, the coherent AM case is described here in detail; the others follow by identical calculations.

A coherent detector multiplies the incoming signal (and noise) by the known signal waveform and integrates from 0 to T. The output is scaled by 1/T for convenience. At time T an output S due to signal integration and N due to noise integration appears. N is a random variable, S a constant. An input gaussian noise spectral density of $n_o/2$ is assumed to represent n(t) through an autocorrelation function: $E[n(t)\cdot n(t+\tau)] = (n_o/2)\ \delta(t-t')$. Such noise through a linear filter will give rise to a gaussian random variable of mean zero and of variance as calculated for the AM coherent case.

Since

$$N = \frac{1}{T} \int_0^T n(t) \cos \omega_o t \, dt$$

it follows

$$\sigma_N^{\ 2} = E[N^2] = \frac{1}{T^2} \int_0^T \int_0^T E[n(t) \cdot n(t')] \cos \omega_o t \cos \omega_o t' \, dt \, dt'$$

Table 6

COHERENT DETECTION SYSTEMS

	AM*	FM	PM
Signal s(t) =	$(\delta_{i1} - \delta_{i0})\ a \cos \omega_o t$	$\delta_{i0}\ a \cos \omega_1 t + \delta_{i1}\ a \cos \omega_2 t$	$a \sin[\omega_o t + (\delta_{i1} - \delta_{i0})\frac{\pi}{2}]$
Filter	$\frac{1}{T}\int_0^T [\text{Input}] \cos \omega_o t\ dt$	$\frac{1}{T}\int_0^T [\text{Input}]\ (\cos \omega_1 t - \cos \omega_2 t)\ dt$	$\frac{1}{T}\int_0^T [\text{Input}] \cos \omega_o t\ dt$
Signal component at filter output S =	$\frac{1}{2}\ a\ (\delta_{i1} - \delta_{i0})$	$\frac{1}{2}\ a\ (\delta_{i1} - \delta_{i0})$	$\frac{1}{2}\ a\ (\delta_{i1} - \delta_{i0})$
Noise output, N, variance	$\frac{n_o}{4T}$	$\frac{n_o}{2T}$	$\frac{n_o}{4T}$
Prob(error)	$1 - F\left(\sqrt{\frac{2E}{n_o}}\right)$	$1 - F\left(\sqrt{\frac{E}{n_o}}\right)$	$1 - F\left(\sqrt{\frac{2E}{n_o}}\right)$

*Notation: $\delta_{ij} = \begin{matrix} 1 & i = j \\ 0 & i \neq j \end{matrix}$

Inserting the autocorrelation function, integration is immediate, yielding

$$\sigma_N^{\ 2} = E[N^2] = \frac{n_o}{4T}$$

In the FM case, statistical independence of the noise through the two filters is assumed so their cross-correlation vanishes. The output signal value, S, is in each case given by a/2.

The decision is made on the basis of the random variable S + N. If positive, a 1 was transmitted; if negative, a 0 was transmitted. This method is shown elsewhere to be optimum with respect to a variety of criteria.[(24)] An error is made if N changes the sign of S + N. This probability is given by

$$\text{Prob(error)} = \int_{\frac{a}{2\sigma_N}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$$

Denote by F(x) the normal distribution function

$$F(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$$

Then, for all coherent binary modulation schemes

$$\text{Prob(error)} = 1 - F\left(\frac{a}{2\sigma_N}\right)$$

Now the signal energy is given by $(1/2)\ a^2\ T$, so the error probability can be expressed in terms of the canonical variable E/n_o as shown in the table.

CASE II: NONCOHERENT RECEPTION (FM)

For noncoherent reception, the AM, FM, and PM detectors all take different forms. Thus, three different analyses must be performed. For AM this has already been sketched in the text. Here the error probability expression for noncoherent FM is derived. The corresponding relation for differentially coherent PSK is more complicated and will not be reproduced.

The detector for noncoherent FSK is described in Fig. 23.

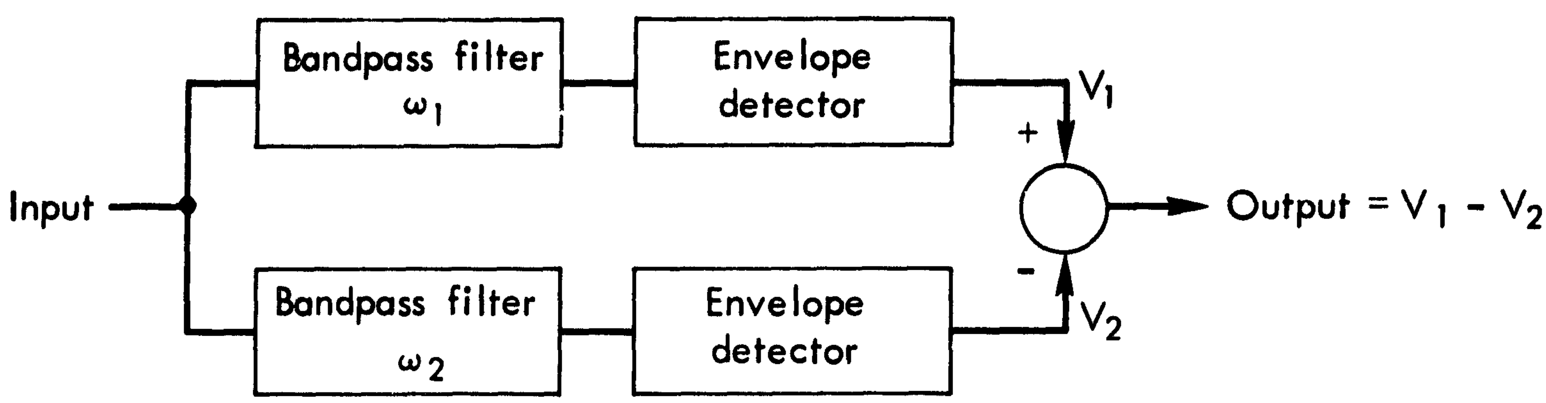


Fig.23—Noncoherent FSK detector

The detection criterion is again symmetric. If V_1-V_2 is positive, the signal is assumed to have been at frequency ω_1, while if a negative output is observed, the signal is assumed to have been at ω_2. Thus, an error occurs if either condition is untrue. Symmetry again makes it possible to treat either case alone. The error probability is given by

$$\text{Prob(error)} = \iint\limits_{V_1-V_2\geq 0} f(V_1,V_2)\, dV_1\, dV_2$$

where $f(V_1,V_2)$ is the joint probability density function for V_1 and V_2, given that the signal was actually at ω_2. The filter spacings are assumed to be sufficiently great that the noises through each are uncorrelated. Then the joint density function is merely the product of the individual density functions

$$f(V_1) = \begin{cases} \dfrac{V_1}{\sigma^2} e^{-\frac{V_1^2}{2\sigma^2}} & V_1 \geq 0 \\ 0 & V_1 < 0 \end{cases}$$

where σ^2 is the variance of the gaussian noise into the envelope detector, and

$$f(V_2) = \begin{cases} \dfrac{V_2}{\sigma^2} e^{-\frac{V_2^2 + a^2}{2\sigma^2}} I_0\left(\dfrac{a\,V_2}{\sigma^2}\right) & V_2 \geq 0 \\ 0 & V_2 < 0 \end{cases}$$

Here a is the amplitude of the sine wave carrier and σ^2 is as before. The calculation now proceeds smoothly.

$$\text{Prob(error)} = \int_{U_2=0}^{\infty} \int_{U_1=U_2}^{\infty} \frac{V_1 V_2}{\sigma^4} e^{-\frac{V_1^2+V_2^2+a^2}{2\sigma^2}} I_0\left(\frac{aV_2}{\sigma^2}\right) dV_1\, dV_2$$

$$= \int_0^{\infty} \frac{V_2}{\sigma^2} e^{-\frac{2V_2^2+a^2}{2\sigma^2}} I_0\left(\frac{aV_2}{\sigma^2}\right) dV_2 = \frac{1}{2} e^{-\frac{a^2}{4\sigma^2}}$$

The latter relation follows directly by substituting the integral expression for $I_0(x)$ and changing to polar coordinates. For a bit duration T, the bandwidth of the filters can be taken as 1/T. Thus $\sigma^2 = 1/T \cdot n_o$. Thus,

$$\frac{a^2}{4\sigma^2} = \frac{1}{2} \cdot \frac{a^2}{2} T \cdot \frac{1}{n_o}$$

and the exponent is simplified to $1/2 \cdot E/n_o$, where E is the signal energy and n_o the one-sided spectral density of the noise. Therefore

$$\text{Prob(error)} = \frac{1}{2} e^{-\frac{1}{2}\left(\frac{E}{n_o}\right)}$$

REFERENCES

1. Marcum, J. I., Table of Q Functions, The RAND Corporation, RM-339, January 1950.

2. Bennett, W. R., and J. R. Davey, Data Transmission, McGraw-Hill Book Company, Inc., New York, 1965.

3. Turin, G. L., "The Asymptotic Behavior of Ideal M-ary Systems," Proc. IRE, Vol. 47, No. 1, January 1959, pp. 93-94.

4. Sussman, S. M., "Simplified Relations for Bit and Character Error Probabilities for M-ary Transmission Over Rayleigh Fading Channels," Trans. IEEE, Vol. COM-12, No. 4, December 1964, pp. 207-209.

5. Akima, H., The Error Rates in Multiple FSK Systems and the Signal-to-Noise Characteristics of FM and PCM-FS Systems, National Bureau of Standards Technical Note 167, 1963.

6. Cahn, C. R., "Performance of Digital Phase-Modulation Communication Systems," Trans. IRE, Vol. CS-7, No. 1, May 1959, pp. 3-6.

7. Arthurs, E., and H. Dym, "On the Optimum Detection of Digital Signals in the Presence of White Gaussian Noise--A Geometric Interpretation and a Study of Three Basic Data Transmission Systems," Trans. IRE, Vol. CS-10, No. 4, December 1962, pp. 336-372.

8. McAuliffe, G. K., et al., Pulse Transmission Study, Final Report, Westinghouse Electric Corporation, Baltimore, 1959.

9. Turin, G. L., "An Introduction to Matched Filters," Trans. IRE, Vol. IT-6, No. 3, June 1960, pp. 311-329.

10. Reinhart, E. E., Multiple-Access Techniques for Communication Satellites: Analog Modulation, Frequency-Division Multiplexing and Related Signal Processing Methods, The RAND Corporation, RM-5117-NASA (to be published).

11. Inose, H., et al., "The Subscriber-Line Circuit and the Signaling and Tone System for an Experimental Time-Division Exchange Featuring Delta Modulation Techniques," Trans. IRE, Vol. CS-10, No. 4, December 1962, pp. 397-407.

12. Dumousseau, C., "An Integrated PCM Network," Trans. IEEE, Vol. COM-13, No. 1, March 1965, pp. 42-49.

13. Bose, R. C., and D. K. Ray-Chaudhuri, "On a Class of Error Correcting Binary Group Codes," Information and Control, Vol. 3, No. 4, March 1960, pp. 68-79.

14. Hocquenghem, A., "Codes correcteurs d'erreurs," Chiffres, Vol. 2, No. 3, September 1959, pp. 147-156.

15. Peterson, W. W., Error-Correcting Codes, John Wiley and Sons, Inc., New York, 1961.

16. Fano, R. M., "A Heuristic Discussion of Probabilistic Decoding," Trans. IEEE, Vol. IT-9, No. 2, April 1963, pp. 64-74.

17. Lebow, I. L., et al., "Application of Sequential Decoding to High-Rate Data Communication on a Telephone Line," Trans. IEEE, Vol. IT-9, No. 2, April 1963, pp. 124-126.

18. Peterson, W. W., and D. T. Brown, "Cyclic Codes for Error Detection," Proc. IRE, Vol. 49, No. 6, January 1961, pp. 228-235.

19. Deloraine, E. M., and A. H. Reeves, "The 25th Anniversary of Pulse Code Modulation," IEEE Spectrum, Vol. 2, No. 5, May 1965, pp. 56-63.

20. High Speed PCM Coder for Television Signals, Kokusai Denshin Denwa Company, Ltd., Tokyo, September 1963.

21. Mayer, H. F., "Principles of Pulse Code Modulation," in L. Marston (ed.), Advances in Electronics, Vol. III, Academic Press, Inc., New York, 1951.

22. Winkler, M. R., "High Information Delta Modulation," IEEE Conv. Rec., Part 8, Vol. 11, March 1963, p. 260.

23. deJager, F., "Delta Modulation--A Method of PCM Transmission Using the 1-unit Code," in W. Jackson (ed.), Communication Theory, Academic Press, Inc., New York, 1953.

24. Middleton, D., Introduction to Statistical Communication Theory, McGraw-Hill Book Company, Inc., New York, 1960.