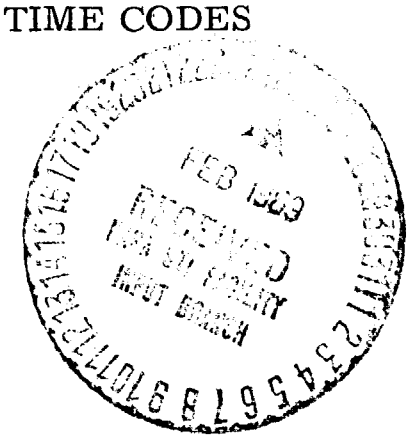
AD/COM

First Quarterly Report

for

INVESTIGATION AND ANALYSIS OF TIME CODES



October 6, 1967

Contract NAS 5-10540

Prepared for

National Aeronautics and Space Administration
Goddard Space Flight Center
Greenbelt, Maryland

ADVANCED COMMUNICATIONS

INFORMATION MANAGEMENT

RESEARCH

DEVELOPMENT

ENGINEERING

ADCOM, INC.

WESTERN DIVISION
PALO ALTO, CALIFORNIA
(415) 328-0200

808 MEMORIAL DRIVE
CAMBRIDGE, MASSACHUSETTS 02139
(617) 868-1000

WASHINGTON BRANCH
COLLEGE PARK, MARYLAND
(301) 779-4666

First Quarterly Report

Contract NAS5-10540

ERRATA and Elucidations

Page	
49	Figure 4.5, remove waveforms (E) and (F).
62	Paragraph 4.1.4.3, replace "Figs. 4.14 and 4.15" with "Figs. 4.12 and 4.13."
75	7th line from top, replace "shimmer" by "slimmer."
83	Note that the significance of advantage (d) cited for pseudorandom sequences lies simply in the availability of the choice of any one of the three multiplexing methods. In other words, the use of a pseudorandom sequence does not deny the designer the choice of whichever of the three multiplexing methods he may prefer for a particular application.
93	2nd line from bottom, replace "Ref. 1*" by "Ref. 5." Footnote becomes Ref. 5 on page 137.
97	11th line of text from top should read "is the geometric representation of the signals of Fig. 4.27b, is meant to apply for all $\tau \leq T/2$. For $\tau > T/2$, the performance is equivalent to that resulting when $\tau = T/2$."
98	Figure 4.29(b), the dimensioned distance along the horizontal axis between the Noisless '0' and the Noiseless '1' is changed to read $\sqrt{(1-2\alpha) \cdot \frac{2ST}{N_o}}$ The decision zone boundary is then located a distance $\sqrt{\frac{1-2\alpha}{2}} \cdot \sqrt{\frac{2ST}{N_o}}$ from the vertical axis.

Errata and Elucidations, continued .....

Page

99 Eq.(4.19) should read

$$P_E = \Phi \sqrt{\frac{1-2\alpha}{2}} \cdot \sqrt{\frac{2ST}{N_o}}$$

Line 2 of text should read "signalling is $-10 \log_{10}[(1-2\alpha)/4]$. The relative...."

100 Table 4.3, under "Relative Detectability, dB" for PDM should read:

$$10 \log_{10}[(1-2\alpha)/4]$$

101 Figure 4.30 first curve title should be "Optimum, Antipodal Signalling," and PDM, $\alpha = 0.182$, not 0.365.

106 Last paragraph, first line change "Ref. 1" to "Ref. 5."

109 Line 10 from top, change "while" to "whole."
Last line on page, change "Ref. 1" to "Ref. 5."

114 Table 4.6, substitute corrected Table attached hereto for one in book.

125 2nd paragraph, first line, change "Ref. 1" to "Ref. 5."

131 End of 4th paragraph "error-control capability." Add:

"One should note, however, that in the case of data which are not processed in "Real" time (i.e., if data are to be recorded for processing later), coarse time data are not automatically available. Some special provision, such as the recording of local coarse time on a spare track of the data tape, must be made if it is to be preserved for later use in error control."

4th paragraph, line 8 should read: "often readily available to the user....."

A-1 APPENDIX A, Figures A.1 - A.10, Source:

First Semiannual Report for Time Code Study
(15 September 1964 - 28 May 1965) Contract #NAS5-9739 by Electronic Engineering Co. of California, Santa Ana, California for the Goddard Space Flight Center, Greenbelt, Maryland 20771.

AD/COM

First Quarterly Report

for

INVESTIGATION AND ANALYSIS OF TIME CODES

October 6, 1967

Contract NAS 5-10540

Prepared for

National Aeronautics and Space Administration
Goddard Space Flight Center
Greenbelt, Maryland

Approved by Elie J. Baghdady

Élie J. Baghdady
Technical Director

ADCOM, Inc.
808 Memorial Drive
Cambridge, Massachusetts

G-110

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

TABLE OF CONTENTS (Continued)

TABLE OF CONTENTS (Continued)

LIST OF ILLUSTRATIONS

LIST OF ILLUSTRATIONS (Continued)

LIST OF ILLUSTRATIONS (Continued)

1. INTRODUCTION

This document constitutes the First Quarterly Report by ADCOM, Inc. to the Goddard Space Flight Center on Contract NAS5-10540.

The ultimate objective to be served by this contract is the evolution of concrete criteria for the definition of an optimum system concept for time code transmission and distribution, taking into account fundamental limitations on performance imposed by the characteristics of transmission media and of the signal design and the limitations on acceptable cost and complexity of implementation techniques. The implementation considerations of the present study are to include the characteristics of wireline, tropospheric line-of-sight and tropospheric beyond-the-horizon media, the relative merits of various modulation and demodulation techniques and the properties of various baseband waveform designs for representation of the time code information. The signal and channel considerations determine the optimum transmission and reception technique; the characteristics of the timing information source-user pair determine the final encoding, decoding and processing requirements of the timing information for efficient utilization of the available transmission link and for error control.

The Work Statement consists of two major parts. Part (a) covers a number of evaluation criteria that were not considered in previous studies of time code systems, including the basic characteristics of transmission media, signal reception and detection, time code error detection and correction, and recording and processing of timing information for different users. Part (b) calls for an integrated use of the results of all analyses performed to date on this and previous programs to determine the optimum system of time code generation, transmission, distribution, recording and processing as a function

of time code format, transmission distance and recording medium. Comparison of the existing systems with the optimum should then reveal the total performance degradation in these systems and the degree of achievable improvement over the performance of current systems.

1.1 General Outline of the Report

The organization of this report is as follows.

The basic considerations for establishing an optimum time code system concept are taken up in Section 2. In this discussion the overall requirements for time code system definition are divided into:

(a) Transmission link design (baseband waveform, modulation/demodulation technique, baseband waveform regeneration at receiver).

(b) Coder-decoder design (for efficient time data encoding and transmission error control).

(c) Recording and processing to meet individual user requirements.

For convenience in the identification of the signal design and transmission problems, the time-code transmission and distribution channels are classified on the basis of the range or separation between source (or transmitter) and user (or receiver). The channel characteristics that affect timing signal transmission accuracy are then identified as: fading statistics, frequency response characteristics, coherence bandwidth and fading rate. These channel characteristics are then shown to affect the selection of time code waveform, signal modulation and receiver processing of time-code signals.

In Section 3, a summary of essential transmission characteristics is presented for the wireline, LOS and BH tropo channels for the various link lengths of interest in time code transmission.

Section 4 is devoted to signal design for reliable and efficient binary time code representation. Signal design is subdivided into baseband waveform design and the selection of carrier modulation technique. Properties of applicable baseband waveforms are considered from the viewpoint of time code design, including

> Time characteristics: Waveforms, synchronization requirement (acquisition, tracking), effect of drop-outs of specific lengths, regenerability, decoding requirements,
>
> Spectral characteristics: Spectral density, bandwidth enclosing a desired percentage of bit energy, dc content, and
>
> Recording requirements: Frequency characteristics and sensitivity to recording noise and drop-out effects.

Consideration of modulation and demodulation requirements includes: suitability of specific modulation techniques, sensitivity to specific propagation effects, filter distortion effects, intersymbol interference, and receiver noise.

Finally, the existing NASA and IRIG time codes are considered to establish their adaptability to the different types of channels.

Conclusions of the present report, and the relation of the report contents to the Work Statement are presented in Section 5.

A number of appendices are also included to supplement or provide background information for the discussions in the main text.

2. BASIC CONSIDERATIONS FOR OPTIMUM TIME CODE SYSTEM CONCEPT

For the purpose of the present study, the basic functional configuration of a time code system is illustrated in Fig. 2.1. Accordingly, the problem of system design can be divided into

(a) Coder-decoder design for ensuring efficient time data encoding and providing adequate measures for error control and time information regenerability at the receiver in the face of possible occasional link failures causing time data dropout;

and

(b) Transmission link design, including the choice of baseband-waveform representation of time code, and the modulation and receiver signal processing and demodulation techniques for the most efficient utilization of available power and bandwidth within the limitations imposed by the channel convolutional and additive disturbances, and by baseband and predetection recording characteristics.

The time information coder-decoder design is determined by the available timing information at the source, by the user requirements and by the statistics of errors introduced in the transmission link and the corresponding error control necessary for satisfying the user.

In the present study the basic characteristics of time codes as determined by timing information source-user pairs will be stated in a general form suitable for the very specific requirements of evolving a time code system concept. Beyond that, the emphasis will be on the transmission link aspects of the system as pointed out in Fig. 2.1.

2.1 Considerations for Transmission Link Design

Time code communication is a form of information transmission. As with all other types of information transmission systems, the essential technical

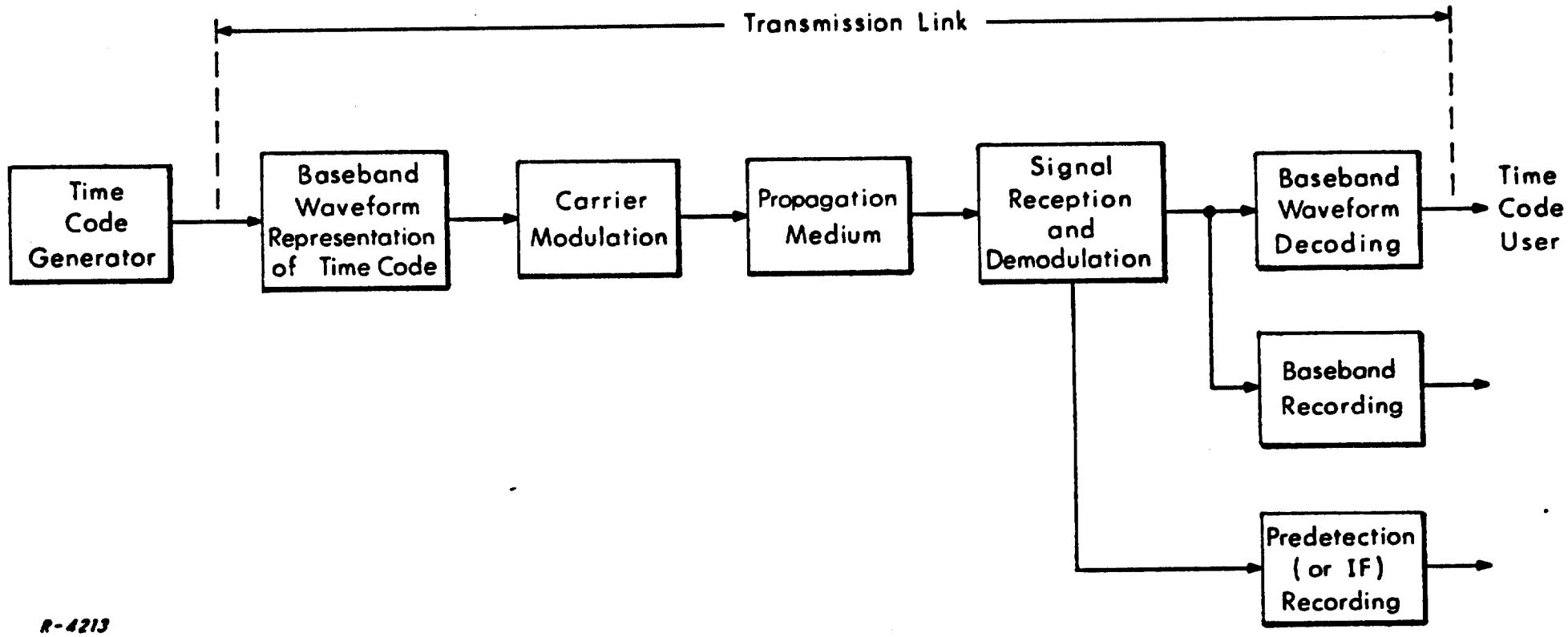


Fig. 2.1 Basic Functional Configuration of a Time Code System

considerations for the evolution of an optimum time code system are:

(a) characteristics of the information source-user pair;

(b) characteristics of the contemplated or available transmission channel; and

(c) selection of baseband waveforms and of carrier signal modulation for efficient and reliable transmission of the desired information within the limitations imposed by the channel characteristics, by regulations on the use of the channel and by limitations on available signal power.

The characteristics of the information source-user pair basically determine what it is that must be transmitted and the ultimate criteria that must be applied for judging the transmission performance. In another respect, the information rate determined by the source-user pair defines the lower limit for the transmission link capacity that must be provided to ensure reliable transmission of the specified information. The cost of the actual link employed or contemplated for the purpose at hand, when stated per unit of the source-user information rate, provides a measure of the "economic efficiency" or "cost effectiveness" of the link.

The characteristics of the available or contemplated transmission medium determine such elementary requirements as the operating frequency, the means for efficient coupling of signal source and signal receiver to the medium, and the transmitter power and receiver sensitivity requirements. In the present study, these elementary considerations will be assumed to be adequately resolvable in a practical situation, and attention will be focused on the convolutional and additive disturbances introduced by the transmission medium and how these disturbances influence the design of the signal waveform and modulation, and the choice of receiver signal processing techniques.

Finally, the nature of the time code information limits the desirable type of baseband waveforms and modulation techniques to a few digital varieties, regardless of the details of the sequences selected for the various parts of the time code.

2.2 Fundamental Information Characteristics of Time Codes

A time code is in essence a waveform that yields, upon user processing, a sequence of labeled time markers. The marker locations specify time, while a digital label relates them to a time standard. This information is updated at some frame rate of issuing labels, such as once per second.

Thus, the absolute minimum characteristics of a long waveform that can be classed as a time code are that

(a) It contain sufficient information for user to generate regularly spaced makers from it.

(b) It provide sufficient information in the waveform for user to be capable of labeling the markers with respect to reference time.

For illustration, consider the NASA and IRIG codes described in Figs. A-1 through A-10 in Apprendix A. Examination of each of the code waveforms shown there reveals that each frame contains

(1) A frame marker, conveyed by a sequence of alternating 1 (wide pulse) and 0 (narrow no-pulse), (such as 1010101010 in Fig. A-1) at the beginning of each frame.

(2) Subsequences of narrow pulses (for binary 0) and wide pulses (for binary 1), each subsequence providing a coded label for the main marker (such as the number of seconds, minutes or days identified by it) and the frame subdivisions.

It is important to emphasize that the unambiguous and accurate identification of the time markers is the most critical aspect of time code reception, the frame marker and the "clock" for subdividing each frame being perhaps the most important. From prior knowledge of code structure and proper reception of significant parts of the code (including both markers and labels) missing labeling information (as a result of transmission error or of deliberate occasional omission to enhance the marker energy) could be readily guessed.

2.3 Fundamental Characteristics of Transmission Channels

A classification of time code channels on the basis of link length is presented in Table 2.1. The three transmission media specified by the Work Statement for inclusion in the present study are each listed where it might provide a natural choice.

The channel disturbances that will afflict the signal can be subdivided into additive and propagation disturbances. Additive disturbances interfere with the signal, but propagation effects cause the signal to acquire random fluctuations that impose limitations on the performance of a communication link in three different ways:

(a) Signal outages or drop-outs, each resulting from a "fade" or drop of the received signal strength below the threshold of acceptable performance in the presence of independent additive disturbances.

(b) Limited bandwidth over which the fluctuations experienced by signal components at different frequencies (or by phase or frequency-reference components and other components of the transmitted signal at different frequencies) will exhibit sufficiently high cross-correlation to keep the resultant signal distortion below tolerable bounds.

(c) A channel fluctuation rate that sets a nonzero lower limit on the frequency content in the signal baseband spectrum if interference between the channel fluctuations and the desired signal baseband waveform is to remain negligible.

The signal fadeout probability is a measure of propagation reliability that is determined by the statistics of the signal fluctuations caused by variable multipath transmission. Other gross transmission parameters of principal importance in characterizing the effects of a randomly time-variant medium upon broad classes of signals are:

Table 2.1

IDENTIFICATION OF TIME CODE SYSTEMS WITH ELIGIBLE TRANSMISSION MEDIA

Service Area of Time Code System	Nominal Link Length	Eligible Transmission Media
Intra-mural time distribution	1 mile or less	Wireline Line-of-sight for the longer hops
Intra-center time distribution	1 to 10 miles	Wireline may be available Line-of-sight (preferable)
Inter-station time distribution	10 to 30 miles	Wireline may be available Line-of-sight for 10 to 30 miles
Intra-Range and ground-to-low-orbital time distribution	30 to 150 miles	Line-of-sight to 30 or 40 miles Beyond-horizon tropo for 30 to 150 miles Line-of-sight for ground-to-low-orbital vehicle
Earth-to-orbital or space vehicle	Exceeding 150 miles	Line-of-sight

(a) The multipath-spread (differential-delay, time-smear or delay spread) parameter is a measure of the duration of the channel impulse response, or of the maximum delay difference between the first and last significant paths. Thus, the multipath spread is a measure of the smearing in time experienced by a time-localized signal (pulse stretching for pulsed signals). Multipath spread is completely determined by the relative geometries of the possible paths – which determine the possible "spread" in path lengths – and by the relative velocities of propagation over the various paths.

(b) The coherence bandwidth is the maximum frequency range over which a desired degree of correlation (or phase coherence for some applications) is maintained among the transmission fluctuations of component sine waves. Thus all frequency components of a signal would fluctuate practically in step if they were contained within the coherence bandwidth, and the channel differential propagation effects upon link performance would then be considered negligible, or within tolerable bounds.

(c) The fading rate (fading bandwidth, frequency smear or doppler spread) parameter is a measure of the bandwidth of the received waveform (the channel response) when the input to the channel is a pure sine wave. Thus, this parameter measures the nominal width of the dispersion in frequency experienced by each individual frequency component in the transmitted signal; it is a measure of the nominal rate at which the fluctuations in the channel, and hence the received signal perturbations, occur. When the test signal at the channel input is a pure sine wave, the fading rate of the channel response to this test signal is completely determined by the relative motions of the various multipath-generating physical entities (reflectors, scatterers, boundary surfaces, etc.) and by the changes (if any) in the relative velocities of propagation over the various paths.

(d) The Decorrelation time is a measure of the time separation that must exist between two impulse excitations to yield a correlation coefficient of 1/e (or less) between the envelopes of the responses to the impulses. Pulses or parts of the signal waveform spaced in time by a correlation time constant may be expected to exhibit practically independent fluctuation patterns.

The above parameters are not all independent because the coherence bandwidth and the delay spread are inversely proportional to each other, as are the fading bandwidth and the decorrelation time. These interrelationships are helpful because some of the parameters are easier to determine for a particular situation than the others. For example, delay spread can be determined from the geometry of the paths, and this is often a simple calculation. The coherence bandwidth is most frequently computed in terms of the delay spread. We first recall that the coherence bandwidth is an upper bound on the permissible bandwidth for the usual types of signal and demodulation methods in order to keep performance degradation caused by frequency-selective fading in transmission below tolerable limits. Therefore, evaluation of the coherence bandwidth must be based on a criterion of performance degradation that is suitable for the intended application. Computations based on reasonable engineering criteria of noticeable degradation in probability of error binary signal reception show that the coherence bandwidth should be taken as roughly one-tenth or less of the reciprocal of the differential delay (or difference between the times of arrival of the earliest and the latest significant paths). The decorrelation time is determined by the dynamics of the intervening medium, which determine the speed with which the multipath geometry and individual path characteristics will change.

3. CHARACTERIZATION OF TRANSMISSION CHANNELS

The purpose of this section is to provide a summary of the characteristics of the channels mentioned in the preceding section for transmitting time-code signals. The characteristics of interest here are those that must be considered in the design of signal (selection of baseband, or time-code waveform, and of type of modulation) and the selection of receiver techniques for reliable transmission of time-code information.

3.1 Characterization of Wireline Channels

Wireline channels are used extensively for local point-to-point communication links serving individuals and organizations. These links may be applied to voice, radio, television and data communication. Data link channels are used for remote inputs and outputs to computers, high-speed facsimile, and other data.

Three categories of wireline channels may be identified; namely, exchange (local) channels, short haul (up to 400 miles) and long haul (400 to 3000 miles). All three categories will be considered here, although ground-to-ground distribution of time codes may be of interest only for distances below 400 miles at present.

For purposes of code selection, only a characterization of digital transmission errors on wireline channels is meaningful because access to the transmission facility is presumed possible only through terminal equipment provided by the telephone company. This is done in Appendix B. However, for purposes of selecting the baseband waveform, a system-function characterization is essential.

3.1.1 System-Function Characterization of Wireline Channels

Although local exchange channels may consist only of wireline equipment, short- and long-haul channels usually consist of concatenations of wireline and radio channels. However, the most frequently employed radio channels in the VHF and higher ranges normally possess coherence bandwidths well in excess of many times the nominal 4-kHz bandwidth of a basic local exchange channel. Moreover, the fading rates of the commonly used radio channels (excluding the ground-to-space situations) are usually well below the lower nominal cutoff frequency of the 4-kHz basic channel. Consequently, only the drop-out or momentary outage effects of radio channels will be normally encountered in 4-kHz wireline channels. Otherwise, the system-function characteristics of a 4-kHz wireline channel are essentially determined by the system functions of the wireline parts of the overall link.

In long-haul and intercontinental links employing HF channels, the system function characteristics of HF channels may well dominate over the wireline channel characteristics.

Wireline channels not including HF links in the overall system are characterized by essentially deterministic, time-invariant frequency response and impulse response functions. The sufficiency of each of these system functions for characterizing the channels for most signal transmission purposes is ensured by the fact that the overall link including modulators, demodulators and other signal processing and amplifying operations normally behaves in a predominantly linear manner.

The measured[1,2] attenuation and delay characteristics of a nominal 4-kHz wireline channel are illustrated in Figs. 3.1 and 3.2 and in Table 3.1.

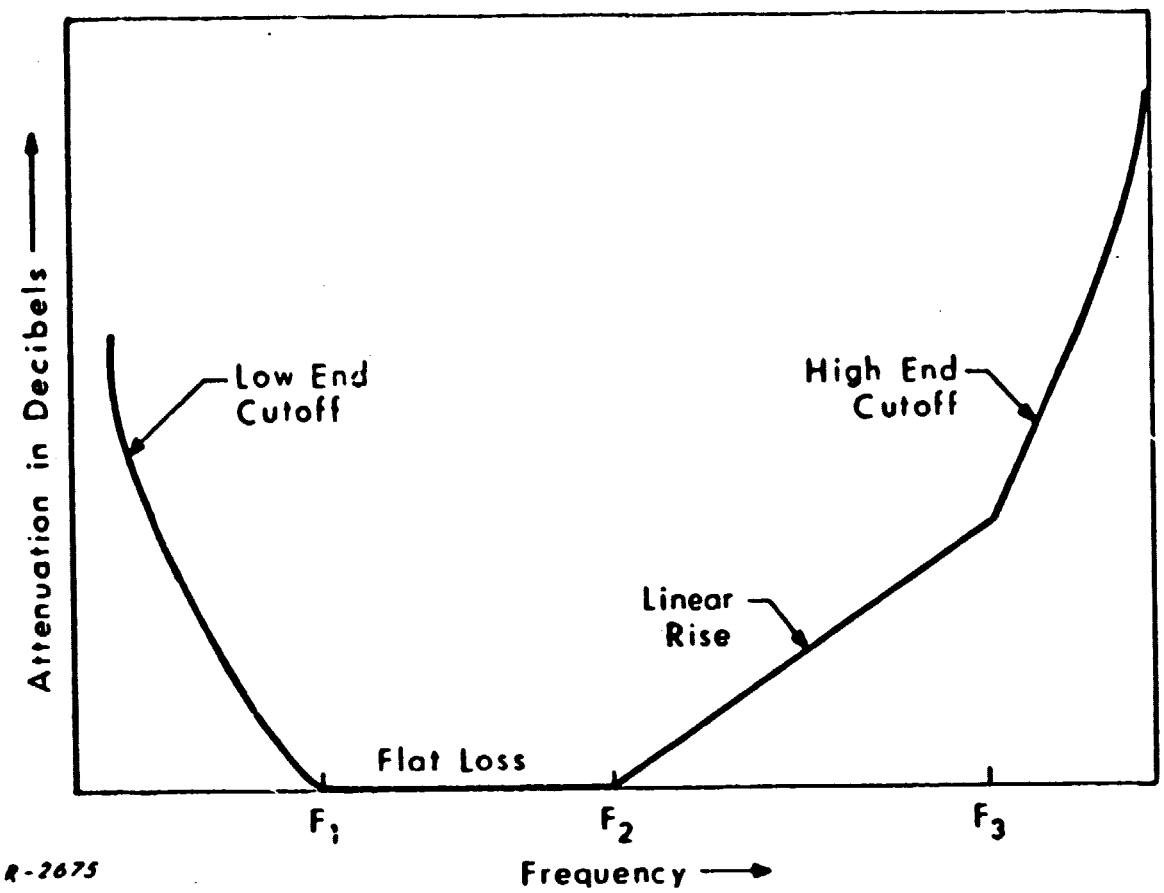


Fig. 3.1 Relative attenuation characteristic of telephone circuits.

Table 3.1

PARAMETER VALUES FOR FIG. 3.1

	Roll-off below F_1 (dB/octave)	F_1(Hz)	F_2(Hz)	Linear rise constant (dB/100 Hz)	F_3(Hz)	Roll-off above F_3 (dB/octave)
Exchange	15	240	1100	0.50	3000	80
Short Haul	24	300	1075	0.60	2950	90
Long Haul	27	280	1150	0.61	2850	80

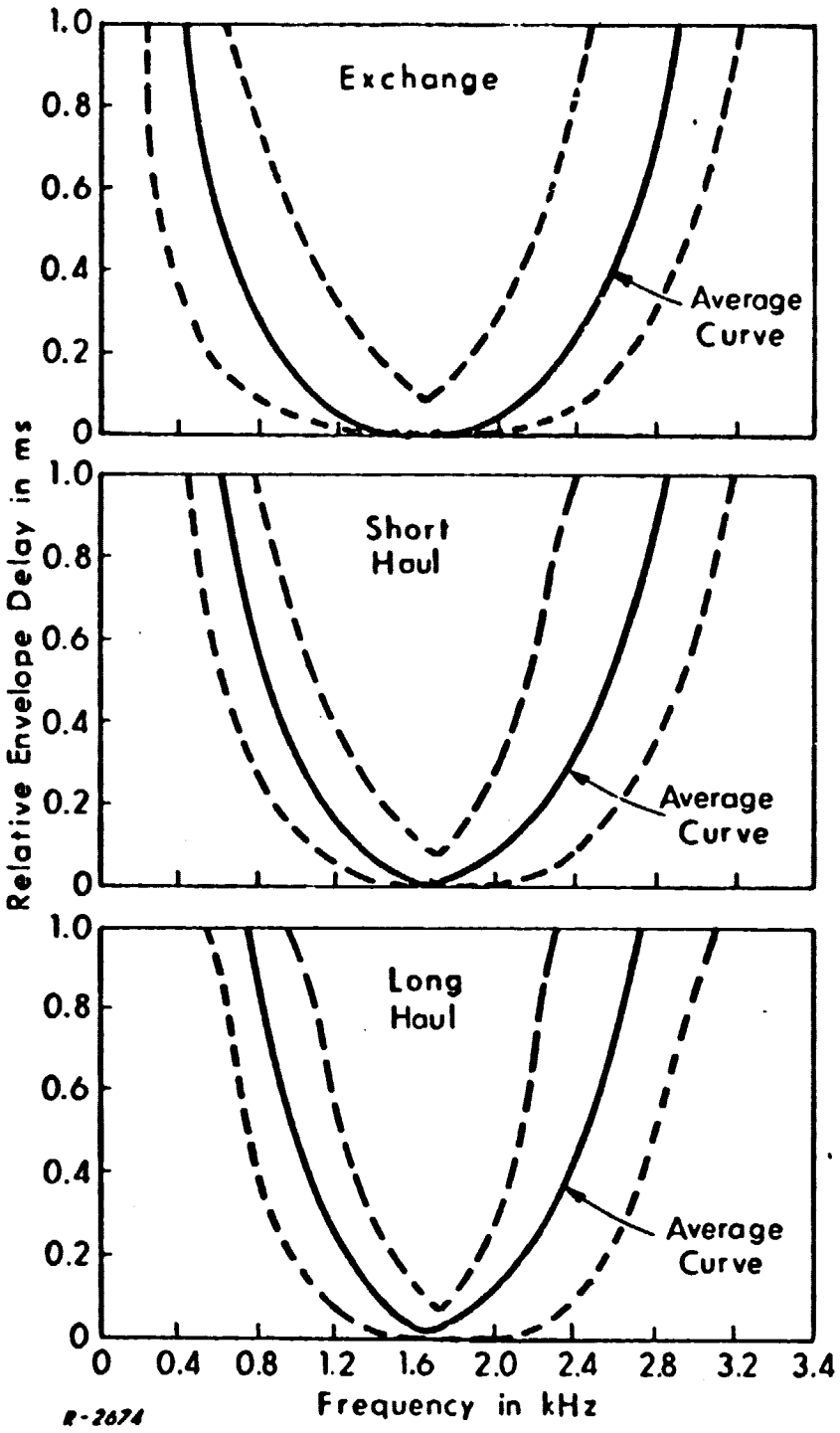


Fig. 3.2 Envelope delay distortion -- locus of 90 per cent of circuit characteristics.

The attenuation characteristic is divided into four regions by the frequencies F_1, F_2 and F_3. Outside the band (below F_1 and above F_3) attenuation increases rapidly. Between F_1 and F_2 the frequency response is essentially flat, while attenuation increases linearly from F_2 to F_3. Roll-off below F_1 is less in exchange systems than in short- or long-haul systems due to use of single-sideband carrier systems in the long-distance channels. Roll-off above F_3 is high in all systems due to wire capacitance, inductive loading, and use of carrier systems on long-distance connections.

The envelope delay curve of a wireline channel is parabolic in shape with a minimum somewhere in the middle of the band. Delay spread over the band is usually emphasized by setting the minimum delay equal to zero in the plots. Average relative envelope delay curves are given[1] by the solid lines of Fig. 3.2, with the 90% locus given by the dotted lines. The frequency at which the curve is tangent to the abscissa is the frequency of minimum delay (FMD), and it lies in the range from 1200 to 2000 Hz. Note that there is wide variation in the three types of channels in the shape of the delay spread as a function of frequency.

Measurements[3] of the impulse response of wireline channels yield nominal durations (and hence "multipath spreads") of less than 15 msec. As stated earlier, the duration of the impulse response bears directly on the possibility of intersymbol interference.

Wireline channels are also plagued by occasional sudden drops in signal level, called line drop-outs or amplitude hits. The drop in signal level may amount to more than 10 dB, may persist for several hundred milliseconds, and may repeat up to 20 times in a one-second interval. Frequency shifts in the tens of Hz, as well as incidental slow FM with deviation up to a few Hz, are also encountered.

For efficient utilization of available wire equipment, a number of voice channels are frequency multiplexed over a single wire pair. The multiplexing scheme is as follows. Twelve voice channels are SSB modulated onto a carrier to make up one group. Five groups SSB modulated onto a higher carrier make up a super group. Similarly, 10 super groups make a master group. Three master groups are multiplexed for FM radio (microwave) links. Each voice channel occupies 4 kHz of bandwidth, including guard bands. Frequently, the usable bandwidth (about 300 Hz to 3 kHz) is not adequate for a particular customer's needs; for example, a high speed computer data link. In such a case, the wireline company leases to the customer a wideband channel consisting of a number of voice channels. These wideband channels are called Tel-Pac lines, and are available in various bandwidths. Tel-Pac A is 48 kHz wide, covering 12 voice grade channels, and is equivalent to one group. Tel-Pac B is 96 kHz wide, covers 24 voice channels, and equals two groups. Other multiplexing and Tel-Pac data are given in Table 3.2.

3.1.2 Wireline Noise Characterization

The types of noise of major concern in wireline channels are:

(a) Impulse noise, induced by lightning discharges and by other impulsive radiations in urban areas

(b) Switching noise

(c) Intermodulation and cross-talk from other channels

(d) Thermal-type noise that may cumulate nonlinearly as a result of demodulation and remodulation in reconstructive repeaters.

A major source of interference in baseband wirelines is atmospheric noise or "sferics". These are produced by lightning discharges whose instantaneous power is well above the highest-powered signal source. The frequency

Table 3.2

WIRELINE VOICE CHANNEL MULTIPLEXING

SSB Modulation		No. Voice Channels	Bandwidth
1 group	= 12 voice channels	12	48 kHz
1 super group	= 5 groups	60	240 kHz
1 master group	= 10 super groups	600	2.4 MHz

3 master groups are multiplexed for FM radio (microwave) links

TEL-PAC - WIDEBAND DATA LINK CHANNELS

Type	Bandwidth	No. of Voice Grade Channels	Equivalent
A	48 kHz	12	1 group
B	96 kHz	24	2 groups
C	240 kHz	60	1 super group
D	960 kHz	240	4 super groups

spectrum of a lightning discharge as recorded by a receiver will depend upon the distance to the source of the sferic, the characteristics of the path between the source and receiver and the response of the receiver. Figure 3.3 shows the frequency spectrum of the radiation component at a distance of one mile. Here the signal peaks in the vicinity of 5 kHz. As the signal is propagated, the lower frequencies suffer greater attenuation so that the spectrum is shifted to 8-10 kHz at large distances.

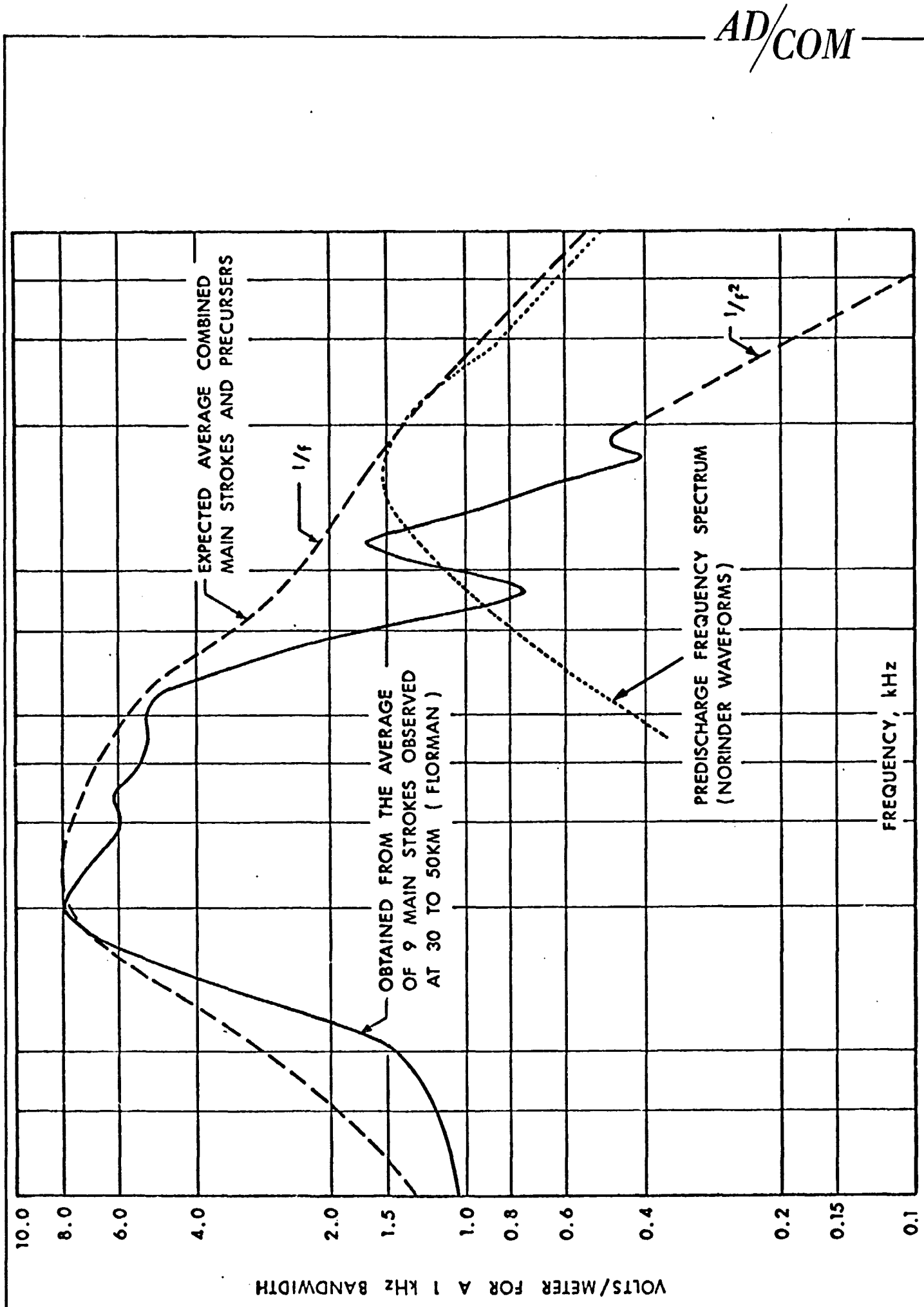


Fig. 3.3 Frequency spectrum of the radiation component equivalent field intensity at a distance of one mile.

3.2 Characterization of Line-of-Sight Channels in the 100 MHz - 30 GHz Range

Line-of-sight transmission to or from the earth's surface differs from free-space transmission because of the presence of

(1) earth's surface with varying degrees of terrain roughness,

(2) earth's atmosphere, and

(3) obstacles, surface and airborne.

It is convenient to distinguish at the outset between surface-to-surface and surface-to-air or surface-to-space (and the reverse) transmission.

Additive disturbances in line-of-sight channels consist mainly of receiver noise, galactic noise, noise from the sun, and noise due to atmospheric absorption, all of which may be approximately characterized as white Gaussian noise inside any given RF bandwidth. These types of noise have been discussed estensively in Chapter 15 of E.J. Baghdady (Ed.), Lectures on Communication System Theory, McGraw-Hill, New York, 1961.

3.2.1 Propagation Characteristics of Surface-to-Surface Links

The intervening terrain and atmosphere bring about reflection, diffraction, refraction, scattering and absorption, all of which have been studied extensively in terms of simplified models. Detailed methods for predicting signal levels in the presence of the above effects are available but will not be discussed here. Of main concern here are the shorter term effects associated with variable multipath transmission which influence the selection of baseband waveform and RF modulation and reception for time codes.

Two general types of fading are encountered in so-called microwave line-of-sight links over distances ranging from a few miles up to about 50 miles.

The first is a relatively rapid type of fluctuation in the received signal strength whose cause is usually traced to interference among two or more slowly varying replicas of the desired signal arriving via different paths. This is called multipath-interference fading. The second is a relatively slow fluctuation in the mean value of the received signal whose cause is usually traced to changes in the constitution of the medium traversed by the signal which give rise to varying degrees of attenuation of the signal in transit between transmitter and receiver. This is called attenuation fading.

The mechanisms that enable the transmitted signal to arrive at the receiving site via more than one path are of two types:

(1) Reflection at intervening surfaces, and

(2) Atmospheric stratifications near the earth's surface.

The various microwave line-of-sight multipath mechanisms are illustrated in Fig.3.4.

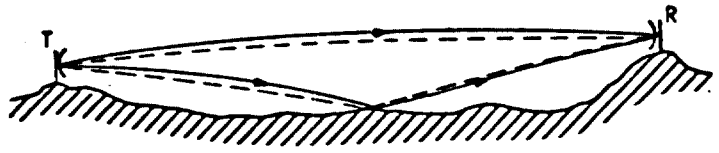
3.2.2 System Function Characterization of Surface-to-Surface Line-of-Sight Transmission

The various multipath propagation and other fading mechanisms described in Sec.3.2.1 are strongly influenced by the siting of the antennas and the conditions of the intervening atmosphere.

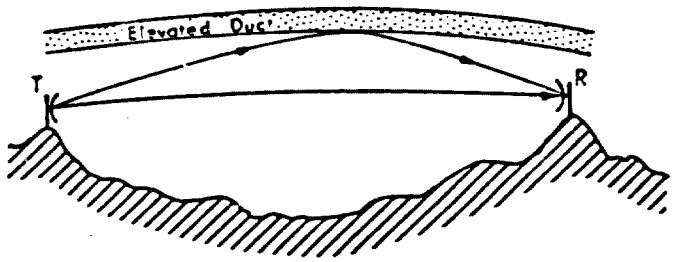
Where possible, the terminals of a line-of-sight link must always be sited to provide

(a) sufficient path clearance; and

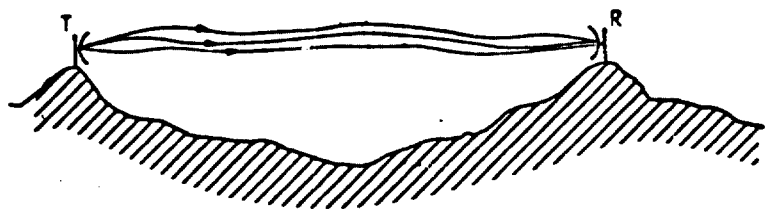
(b) transmission over terrain with very low reflection coefficient; in particular avoiding intermediate bodies of water.



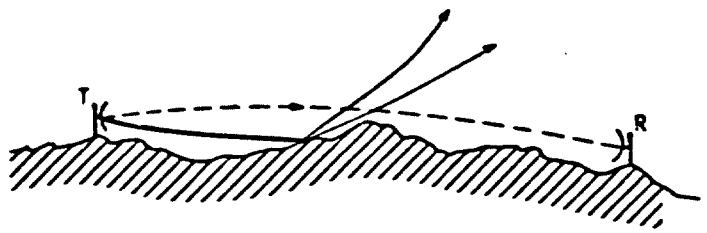
(a) Reflection at surface of intervening terrain



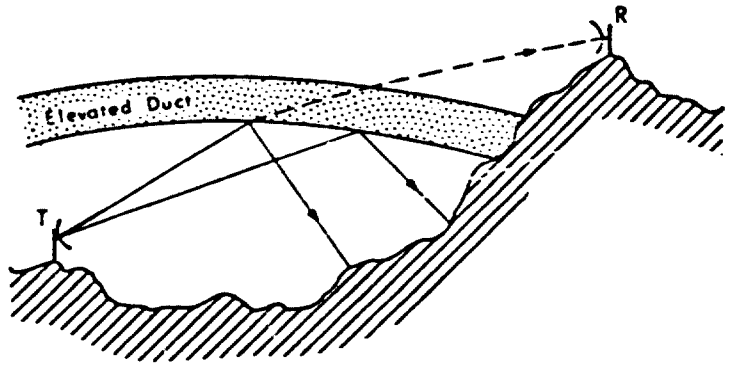
(b) Reflection by elevated duct



(c) Atmosphere multipath



(d) Effect of insufficient ground clearance and upward bending of the rays



(e) Deflection by elevated duct

Fig. 3.4 Multipath and signal deflection mechanisms in tropospheric line-of-sight surface-to-surface transmission.

But even if the necessary siting precautions are taken (and this is sometimes not possible for (b)), atmospheric multipath and duct-attentuation fading remain a serious threat, particularly on clear summer nights with insufficient wind and convection currents to mix the air, and when nonuniform distributions of temperature and humidity create steep gradients of variation in the dielectric constant of the intervening atmosphere.

For purposes of system-function characterization of the channel, we therefore recognize the following facts:

(i) The most common case of multipath is one in which two or three replicas of the desired signal are received. The amplitude ratios of the received replicas may be close to unity.

(ii) More than three replicas of the desired signal are sometimes encountered, and not too infrequently.

(iii) On a typical well-planned 30-mile link, most somewhat localized, frequency-selective deep fades are brought about by the interaction of quite a few components, perhaps four to six. Among these components, those with relatively long delays usually are appreciably smaller in magnitude than those with shorter delays. The relatively weak and longer delayed components are important in determining the shape of the path-loss versus frequency curve near the maxima of loss. Deep fades resulting from multipath on links with adequate path clearance are always frequency selective with near-Rayleigh fading statistics. A steady relatively nonselective component of loss of about 10 dB usually accompanies frequency-selective deep fades.

(iv) The various received replicas have different angles of arrival in the vertical plane, usually fractions of a degree above the line-of-sight for atmospheric multipath, fractions of a degree below the line-of-sight for ground and water surface reflections and sometimes below it for atmospheric multipath as, for example, under substandard atmospheric conditions causing upward bending of the rays.

(v) Attenuation fading does not offer sufficient statistical diversity with frequency, antenna spacing, polarization, etc., to be effectively combatable by diversity techniques.

(vi) The multipath propagation characteristics appear to be independent of polarization.

If we assume truly specular paths, the resultant signal y(t) at the receiving end, in response to x(t) at the transmitting end may in general be expressed as

$$y(t) = \sum_{n=1}^{N} a_n x(t-\tau_{d,n}) \tag{3.1}$$

where $\tau_{d,n}$ is the transmission delay of the n^{th} path, and a_n is the associated attenuation factor. If the a_n's are properly defined, this expression holds independently of whether or not each path results in distortionless transmission of x(t) (i.e., whether or not the individual paths provide uniform attenuation and uniform phase shift across the frequency band occupied by x(t)). If x(t) is of sufficiently short duration, then the a_n's and $\tau_{d,n}$'s will be relatively time-invariant over the duration of x(t), and we can write

$$Y(j\omega) = X(j\omega) \sum_{n=1}^{N} a_n e^{-j\omega\tau_{d,n}}. \tag{3.2}$$

The system function of the line-of-sight transmission medium may therefore be expressed in the form

$$H(j\omega, t) = \sum_{n=1}^{N} a_n(t)\, e^{-jn\omega\tau_{d,n}(t)} \tag{3.3}$$

in which the dependence of a_n and $\tau_{d,n}$ upon time is explicitly brought out.

For purposes of illustration, consider the likely case in which only two paths are important. Here

$$H(j\omega, t) = a_1(t)\, e^{-j\omega\tau_{d,1}(t)} + a_2 e^{-j\omega\tau_{d,2}(t)} \tag{3.4}$$

A phasor construction of $H(j\omega)$ is shown in Fig.3.5. This resultant phasor operates on the transmitted signal. Zeros of transmission will occur if

$$a_1 = a_2$$

and

$$\omega\tau_1 = \omega\tau_2 + (2n + 1)\pi, \quad n = 0, 1, 2, \ldots$$

or

$$\omega = \frac{(2n + 1)\pi}{(\tau_1 - \tau_2)}, \quad n = 0, 1, 2, \ldots$$

Figure 3.6 gives a plot of $|H(j\omega, t)|$ of Eq.(3.4) as a function of ω for fixed t. With changes in delay times and additional paths, one obtains the classical selective fading effects.

Estimates of coherence bandwidth can be made on the basis of Fig.3.6 to be about $10^{-2}/(\tau_2 - \tau_1)$ Hz or less. Numerically, a delay spread of about 12×10^{-9} sec has been encountered under severe fading conditions on a typical, well-planned 30 mile link operating at 4 GHz. This indicates a coherence bandwidth of about 1 MHz which is totally adequate for time code transmission.

Fading rates of fractions of a Hz are typical of surface-to-surface line-of-sight links, which need not present a threat to the transmission of time codes.

Thus, the most serious propagation disturbance that can be expected in the transmission of time codes over line-of-sight links is drop-outs caused by deep signal nulls.

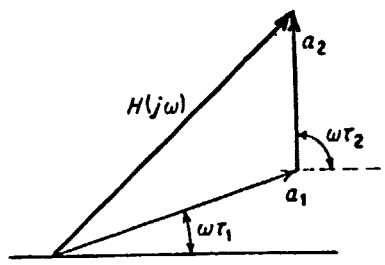


Fig.3.5 Phasor diagram for two-path line-of-sight propagation.

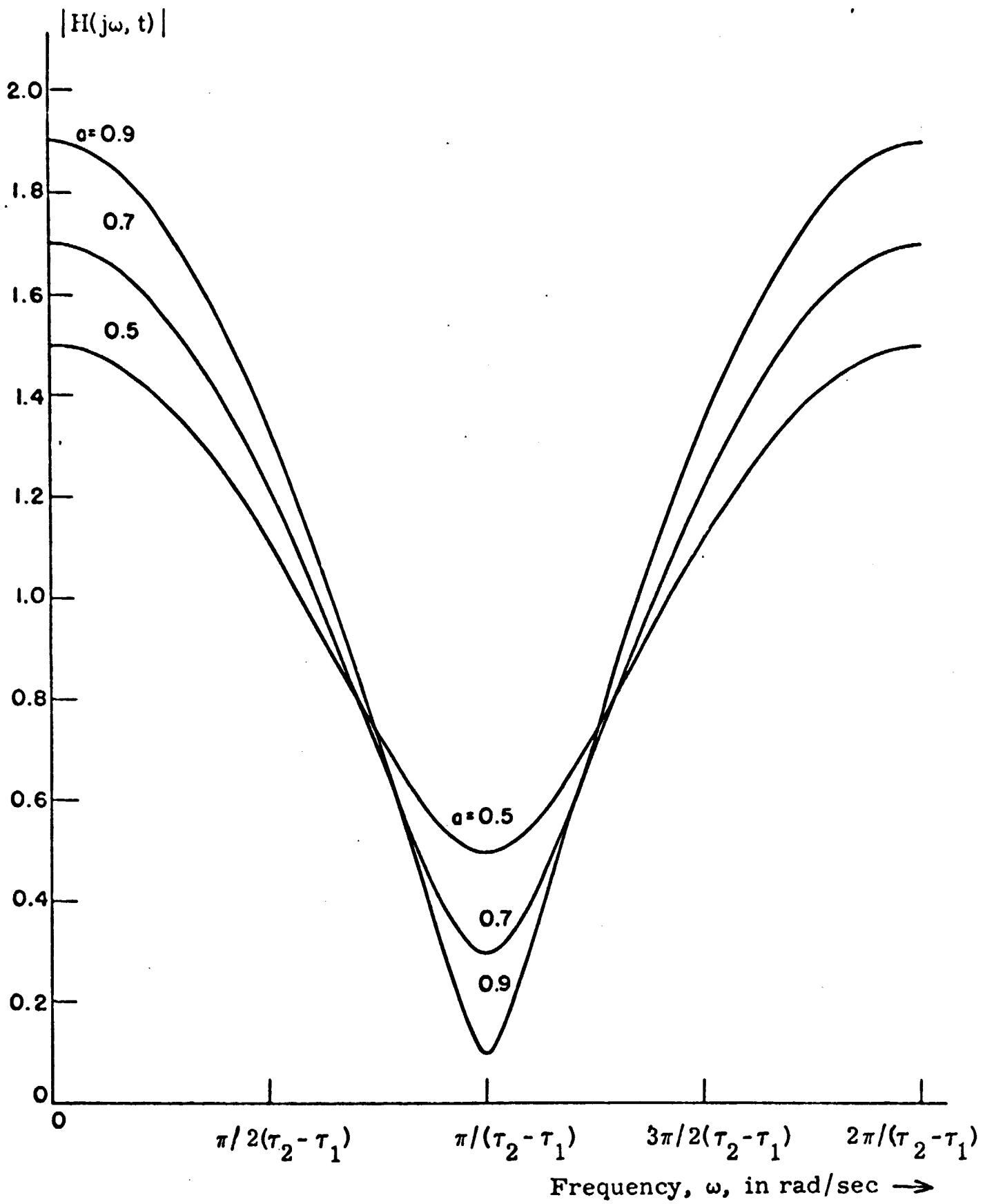


Fig.3.6 Instantaneous transfer function for resultant propagation over two simultaneous specular paths.

3.2.3 Characterization of the Surface-to-Space (or Air) Line-of-Sight Medium

A number of special problems may result from the motion of an aircraft or spacecraft terminal in a line-of-sight link. In fixed terminal surface-to-surface links the attainment of synchronous digital communication is not difficult both because of the stability of the path and the availability of highly stable oscillators at both terminals. However, when one terminal is in an aircraft or a spacecraft the attainment of synchronous digital communications is considerably more difficult because of the changing path length, time-variable doppler shifts and the probable greater instability of oscillators in the moving craft. Changing path length affects synchronization in time, and time-variable doppler shifts affect synchronization in frequency.

Another problem that results from the motion of a terminal is increased doppler spread or fading rate. This increased doppler spread comes about from the fact that communication between air or space and ground is likely to involve contributions from paths that start in widely different directions from the craft. The differences in original ray directions result in different doppler shifts for the different paths since doppler shift varies as the cosine of the take-off angle relative to the direction of motion of the craft. To take an extreme case, an airplane traveling at Mach 3 and transmitting at 135 MHz will cause a doppler shift in the direction of airplane motion of around 405 Hz. If propagation is achieved with multipath components having take-off angles whose cosines differ by 1/3, then a 135 Hz beat will occur in the fading; i.e., the fading rate will be as high as 135 Hz.

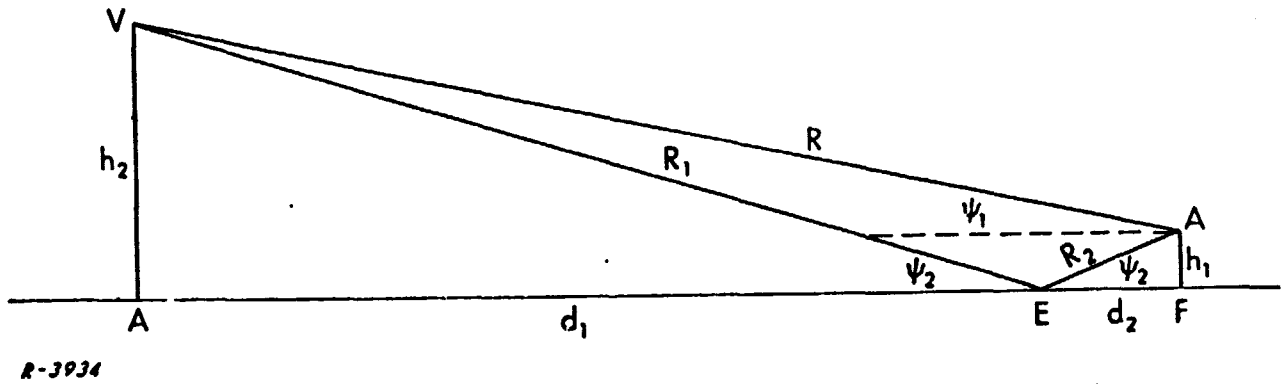
Finally, bombardment of a moving aircraft by air particles results in an additive component of noise called precipitation static.

Representative geometrical configurations for surface-to-space or to air links are illustrated in Figs. 3.7 and 3.8.

At elevation angles above the horizon in excess of about 5 degrees the principal cause of multipath is terrain features in the environs of the ground station, such as buildings, hills or mountains, power lines, depressions, tree groves, etc. It can be shown, however, that the differences in delays and doppler shifts among the various paths are generally too small to cause any noticeable effects on the communication link performance.

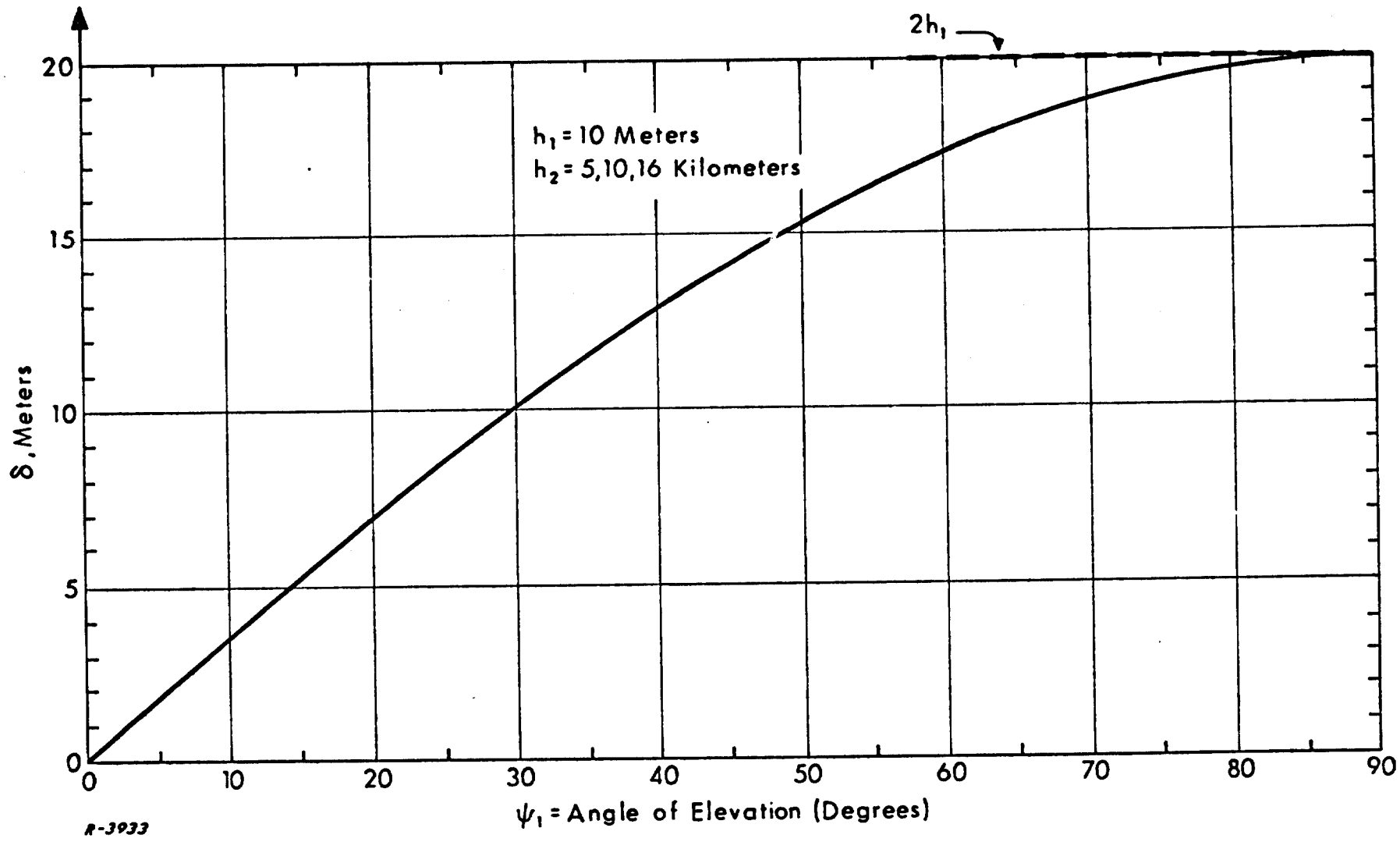
At elevation angles below 5 degrees above the horizon, communication with a distant craft will be subject to the effects of at least one secondary path whose characteristics depend upon the electrical properties of the ground reflecting surface (worst for a body of sea water) or of a signal deflecting elevated duct or layer. The importance of such multipath depends on the significant doppler spread, delay spread and high values of multipath amplitude ratio that may result, especially from elevated layers.

Limitations on the size of the craft antenna may result in a very broad antenna beamwidth. This makes it difficult to suppress secondary-path signals by antenna directivity. Even antennas of moderate beamwidth will pick up reflections off the earth's surface when pointing at a low elevation angle.



(a) Flat-Earth Geometry for Two-Path Propagation

Fig. 3.7 Multipath in Surface-to-Air Propagation



(b) Path Difference, δ, as a Function of Elevation Angle ψ_1, for a Flat-Earth Surface

Fig. 3.7 Multipath in Surface-to-Air Propagation

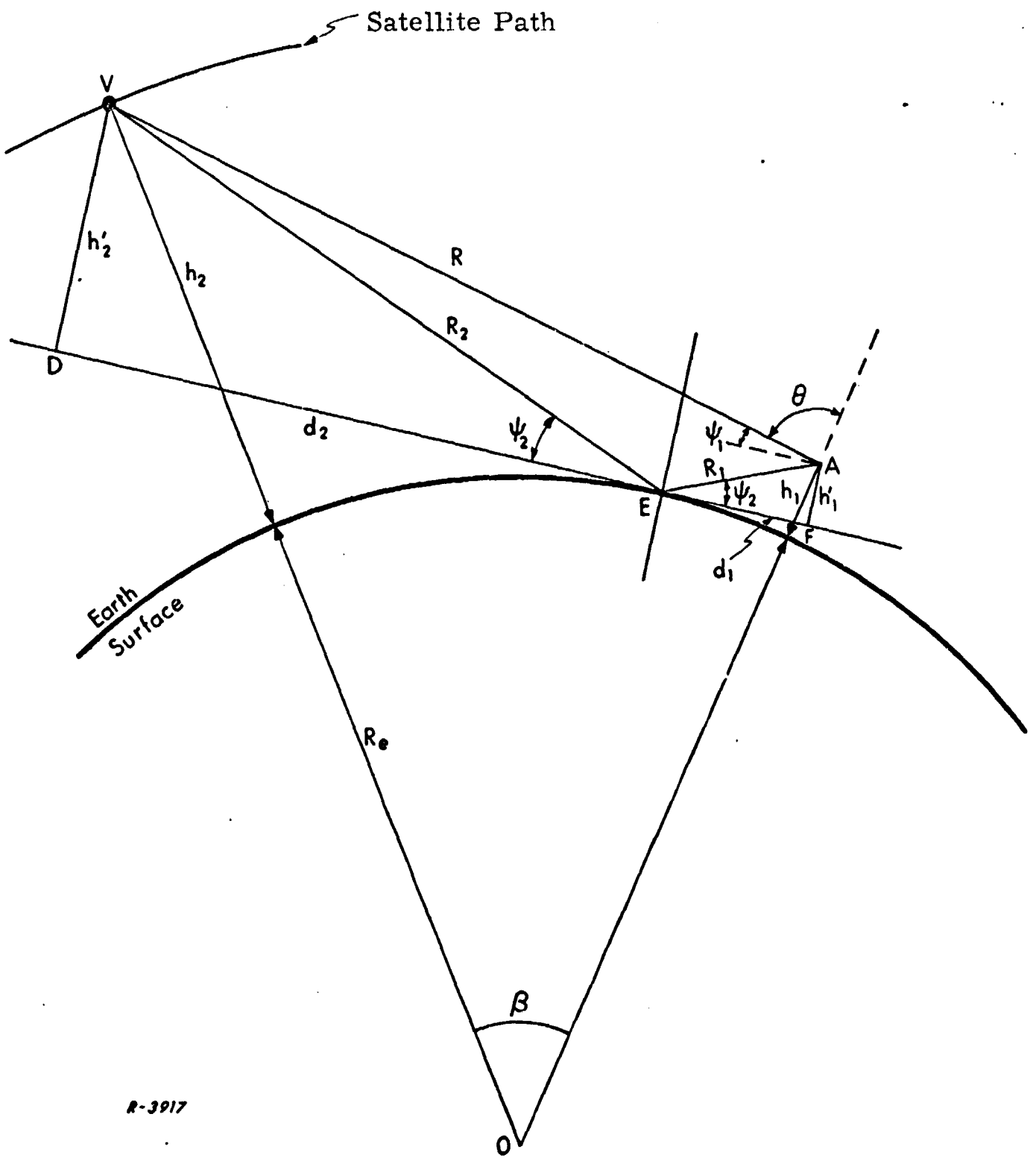


Fig. 3.8 Geometry of Reflection from a Smooth Spherical Earth

Direct-path length (or slant range) = $R_1 = VA$

Indirect-path length = $VE + EA$

Path length difference $\delta = (VE + EA) - VA$

Ground range $d = d_1 + d_2$

In general, the secondary-path signals may be caused by four mechanisms:

a) Atmospheric disturbances such as scatter effects or inhomogeneous refraction effects;

b) Atmospheric ducts and inversion layers;

c) Reflections from the earth's surface (sea or land) in the neighborhood of the ground antenna; and/or

d) Reflections from prominent objects (large structures, aircraft, high mountains).

Multipath due to atmospheric disturbances (item a) usually results in very short delay spreads. The importance of atmospheric multipath rests upon the fact that the paths are nearly equal in strength, and hence will result in nulls in several tens of dB's. Atmospheric multipath may be encountered even with highly directive antennas, although undesired reflections of types (b), (c) and (d) are rejected by such antenna patterns.

The secondary signals caused by multipath add vectorially to the direct signal at the receiving antenna, producing a resultant received signal whose phase and amplitude differ from those of the direct signal. Because the geometry of the propagation paths may be changing with time, the different path lengths and attenuations will change correspondingly, resulting in:

a) Variable delay spread of the received signal replicas; and

b) Variable doppler spread or fading rate.

The attendant fluctuations of the signal strength (fading) and phase due to destructive and constructive interference of the various signal replicas also result in SNR fluctuations, and link outages due to signal drop-outs below detector thresholds.

For situations in which the aircraft and ground site are within distances sufficiently small so that a flat earth can be assumed, the geometry in Fig. 3.7 applies. A plot of path-length difference, δ, as a function of elevation angle ψ_1 for h_1 = 10 meters and h_2 = 5 km is shown in Fig. 3.7(b). The same plot applies also for h_2 = 10 and 16 km. Note that for $\psi_1 = 0$, $\delta = 0$, and that as $\psi_1 \rightarrow \pi/2$, $\delta \rightarrow 2h_1$ monotonically. The delay difference between the paths is given by δ/c, where c is the velocity of propagation.

In the case of fixed transmitter and receiver antennas the path difference is fixed, and no alternation between constructive and destructive interference is observed. But as the aircraft moves, the elevation angle changes and the relative phases and amplitudes of the two waves will change causing alternation between maxima and minima of the resultant signal strength.

The effect of irregularities of the surface of the earth on the reflection of the ray impinging upon it will depend on the wavelength of the radiation and the grazing angle ψ_2. The criterion used to determine if the earth could be considered smooth or rough is the Rayleigh criterion (Ref. 1, p. 411). The Rayleigh criterion states that a rough surface can be considered smooth if

$$h_o \sin \psi_2 < \lambda/8$$

where

h_o = height of obstacle

ψ_2 = grazing angle of ray

λ = wavelength considered

For example, at $\psi_2 = 5^o$, terrain whose obstacles do not exceed 3 meters may be considered smooth for waves at 136 MHz, while the figure for waves at 2 GHz is 0.2 meter.

The over-all effect of roughness is to reduce the effective reflection coefficient ρ to the range $0.1 < \rho < 0.3$, resulting in a reduction of fade depths, as well as to introduce some further fluctuation if the roughness is time-variant (e.g., due to sea-wave motion).

Surface-to-synchronous satellite transmission can be expected to portray three basic effects in addition to the regular effects of the atmosphere, namely

(a) Multipath transmission;

(b) Doppler frequency shift; and

(c) Doppler spread.

The most probable multipath condition is again the two-path case. Fig. 3.8 shows the reflection geometry for a spherical Eartn. It is assumed that h_1 is small compared to h_2. Under this assumption, d_1 will be small compared to d_2 because the reflection point E will be relatively close to F. The second assumption we make is that within relatively small areas, the Earth is sufficiently flat so that the tangent AF can be considered very nearly at right angles to both OE and OF. Plots of the resulting delay difference, τ_d, as a function of the elevation angle ψ_1 are given in Fig. 3.9 for

h_2 = 35,890 km (synchronous altitude)

h_1 = 10 m, 10 km and 16 km

3.3 Characterization of Beyond-the-Horizon (BH) Tropospheric Channels in the 100-10,000 MHz Range

3.3.1 Propagation Mechanisms of BH Channels

Tropospheric beyond-the-horizon propagation is strongly affected by the antenna characteristics of the terminals. The propagation mechanisms encountered are as follows:

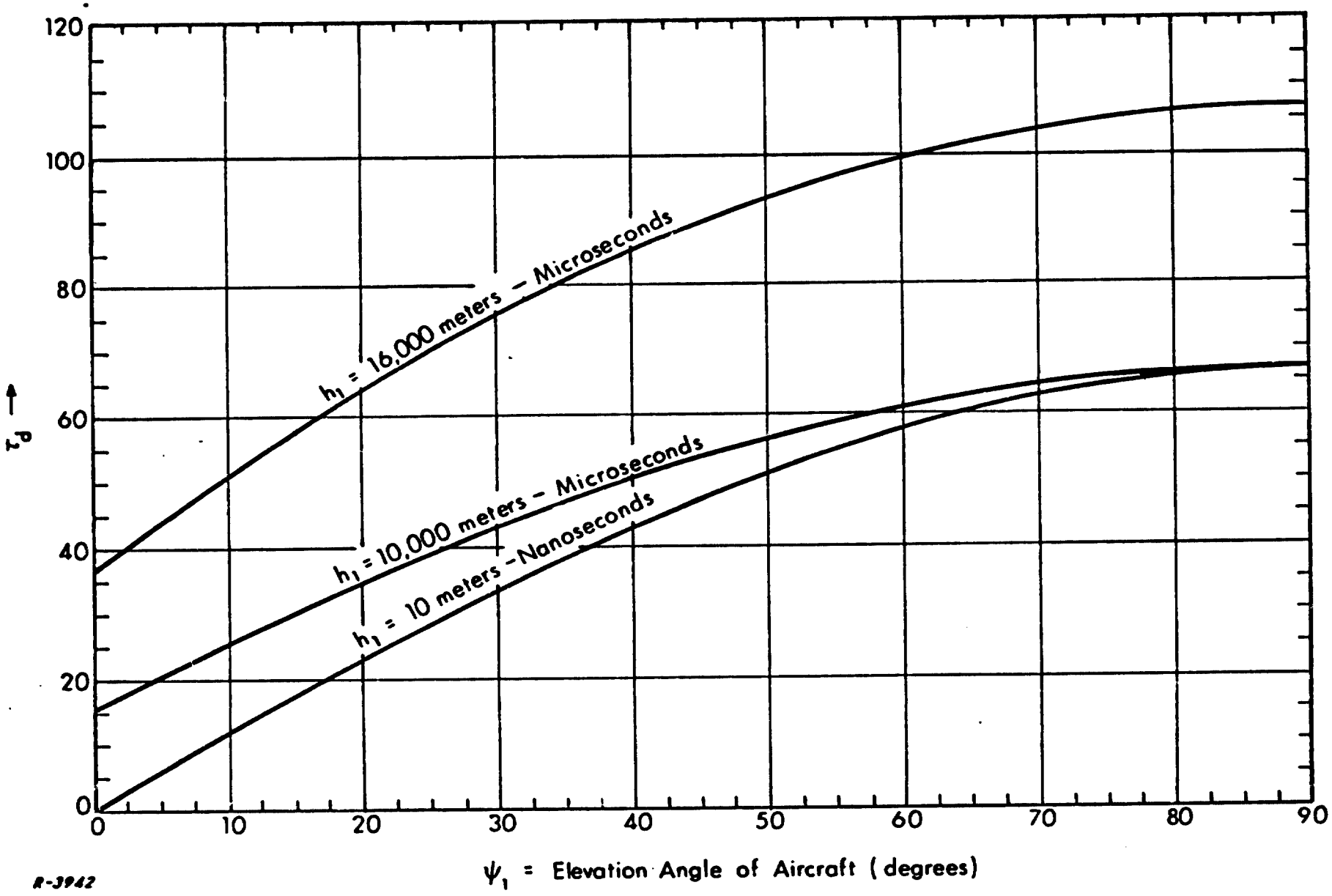


Fig. 3.9 Multipath Delay Difference, τ_d, vs. Elevation Angle, ψ_1, of Aircraft for Synchronous Satellite Medium (h_2 = 35,890 km)

Diffraction, usually intermediate between smooth spherical-surface and knife-edge diffraction. The diffracted field decays exponentially with distance beyond horizon, and is the dominant component of signal in the region from the horizon to up to 75 miles beyond.

Fields resulting from diffraction are generally stable, except when there are several "hot" spots on a knife-edge, in which case a variable multi-path condition arises because of variable atmospheric conditions that vary the illumination of the hot spots.

Troposcatter, caused by time and spatial fluctuations (blobs, turbulent convection transfer of water vapor, etc.) of refractive index in the volume defined by the intersection of the transmitting and receiving antenna patterns. Scatter accounts for relatively high signal levels from about 75 to 500 miles beyond the radio horizon, with considerable fading having a Rayleigh envelope distribution. It is dominated by the diffraction component from radio horizon out to about 75 miles beyond, and becomes dominant beyond that point. Signal received in response to a sinusoidal-carrier test signal is well-modeled by a gaussian random process. Field strength is about 100 dB below free-space value for a 300-mile range. Range of fluctuations is about ±16 dB relative to rms value, which is consistent with the nominal range for a gaussian model.

The fading rate (= width of fluctuation spectrum) depends somewhat on the component of mean wind transverse to the path, and varies linearly with frequency of emitted signal (suggesting a doppler mechanism for fading fluctuations).

Ducting and Reflection from Elevated Layers - Ducts and reflecting layers are caused by temperature inversions in the atmosphere. A duct acts as a wave guide, sustaining wavelengths with nominal cutoff of

$$\lambda \leq 0.014\ d^{3/2},\ \text{cm}$$

where d = duct height in ft. Depths of atmospheric ducts are generally of the order of 100 ft. Radiations of greater wavelength than $0.014\ d^{3/2}$ are also trappable but they suffer greater attenuation. Attenuation within a duct is small (about 5 dB per 100 miles for ground ducts) but may be high at interface of walls of duct. Components arriving via ducts are practically specular in character.

The frequency selective properties of the (specular) components contributed by the nonscatter mechanisms can generally be neglected because

(a) RF bandwidths of practical interest (<20 percent of RF carrier) are narrower than the frequency selectivity of ducting, layer reflection and diffraction; and

(b) Frequency-selectivity effects of scatter component and of the resultant of all components present dominate those of the individual nonscatter components.

3.3.2 Statistical Properties of Tropospheric BH Signal Models

From the mechanisms of multipath propagation beyond the horizon, briefly described above, and from empirical data, it is possible to sum up the major characteristics of the tropospheric beyond-the-horizon channel as follows:

(a) Signals received beyond the radio horizon appear to be characterized approximately by a pure scatter component having the properties of a narrowband gaussian process, plus (frequently for path lengths below 150 miles, and occasionally for path lengths above 150 miles) a "non-fading" specular component caused by diffraction, and/or atmospheric ducting. The word "non-fading" is in quotes because the specular component will actually be found to fluctuate rather slowly when viewed over a sufficiently long interval of time. More than one such component are sometimes encountered.

(b) The nature of the multipath structure, and the dependence of the received signal characteristics upon antenna beam geometries cause:

(i) the depths of fade to be usually less than about 30 dB;

(ii) the coherence bandwidth to vary inversely with the cube of the hop length, with values ranging from as low as a few tens of kHz to as high as several MHz; and

(iii) several types of diversity to be effective, including space, frequency, angular and time, but not polarization diversity.

(c) The fading rate varies from fractions of a Hertz to a few Hertz between stationary terminals. Since the major contributions to the received signal come from "rays" within a relatively small spherical angle about the axis of an airborne antenna beam, the increase in doppler spread due to aircraft motion may not raise the effective doppler spread beyond 10 Hz.

3.3.3 Gross Transmission Parameters of Tropospheric BH Channels

For tropospheric propagation, the parameters defined in Sec. 2.3 depend on the path length, the conditions of the intervening troposphere and the mechanism of multipath transmission (i.e., the presence of one or more specular paths, and/or of a second scatter volume). The time variability of the geometry of the propagation paths in the link and the turbulence of the atmosphere account for the fluctuations in the multipath conditions that degrade system performance. Thus, the gross transmission parameters are themselves random variable that can only be specified in statistical terms.

The coherence bandwidth of tropospheric transmission is determined by the spread of multipath delays of the significant contributions from various parts of the scattering volume. This spread of delays depends upon the effective scatter

volume, which in turn is determined by the antenna beamwidths relative to the average scattering angle (defined as $\bar{\theta}$, the ratio of the distance d to the 4/3 earth's radius) and/or the scattering pattern of the blobs. With broad antenna beams, the multipath is controlled by the scattering blobs, and the bandwidth over which the cross-correlation coefficient between frequency components of the signal will exceed 1/2 is given by

$$B_{corr} \approx 25/d^3 \text{ MHz} \qquad \alpha >> \bar{\theta}_s \equiv 3d/160 \qquad (3.5)$$

in which d is the distance in hundreds of miles over the earth's surface between the transmitting and receiving antennas, and α is the vertical beamwidth for the (assumed identical) transmitting and receiving antenn :. The bandwidth over which the various frequency components of a signal may be expected to fade essentially in step, thus maintaining a high degree of coherence, is one tenth, or less, of B_{corr}. Thus,

$$B_{coh} \approx 2.5/d^3 \text{ MHz} \qquad \alpha >> \bar{\theta}_s \qquad (3.6)$$

The diversity bandwidth may be taken as

$$B_{div} \approx 30/d^3 \text{ MHz} \qquad \alpha >> \bar{\theta}_s \qquad (3.7)$$

It is sometimes found convenient also to express the above bandwidths in terms of the reciprocal of the delay difference of the upper and lower extreme paths of the common volume geometry determined by the antenna beamwidths. This estimate yields bandwidths of the correct order when the antenna beamwidths are narrow. As noted above, with broadbeam antennas the scattering blobs or atmospheric irregularities rather than the beamwidths influence the delay spread, or maximum nominal multipath delay difference. This is

because with wide beamwidths, a wide range of scattering angles is encountered within the common volume, and the scattering efficiency decreases rapidly with increasing scattering angle.

Thus, with antenna beamwidths that are small compared with the average scattering angle, the bandwidth formulas may be expressed either in terms of the delay difference $\Delta\tau$ between the extremal paths defined in the vertical plane by the antenna beamwidths, or in terms of distance separating transmitting and receiving antennas. The bandwidth B_{corr} for a correlation coefficient of 1/2 is

$$B_{corr} \approx \frac{4}{\alpha d^2} \text{ kHz} \qquad \alpha << \bar{\theta}_s \equiv 3d/160 \tag{3.8}$$

where d is again in hundreds of miles, and α is the antenna beamwidth in radians. The diversity bandwidth may be taken as slightly larger than B_{corr}, say $5/\alpha d^2$ kHz, and the coherence bandwidth is approximately 1/10 of B_{corr}.

As stated earlier, the fading bandwidth, or fading rate of a tropomedium is on the order of a few Hz.

3.3.4 Additive Disturbances in Tropospheric BH Channels

The dominant additive disturbance in the 100-10,000 MHz range is internal receiver noise with some contribution from galactic noise at the 150-300 MHz end.

4. SIGNAL DESIGN FOR BINARY TIME CODE REPRESENTATION

The design of signal for waveform representation and transmission of a time code consists of

(a) design of baseband waveform for representing a binary sequence expression of the time code, and

(b) selection of modulation method for carrier or sub-carrier transmission of the time code baseband waveform.

In this section, various ways in which a binary code sequence can be represented in terms of a rectangular pulse waveform will be described and their time and spectral characteristics and methods of generation and decoding will be discussed. The properties of binary maximal-length linear sequences that make them particularly suitable for representing time markers will also be presented. The sensitivity of various promising waveform representations of time codes to channel disturbances, filter distortion, recording noise and distortion and the suitability of the current NASA and IRIG code waveforms to transmission over the types of channels described in Sec. 3 will also be discussed.

4.1 Rectangular Waveform Representations of Binary Sequences

A number of rectangular waveform representations of binary sequences will now be discussed. A description will be given of each waveform including special time and spectral properties and methods of generation and decoding. To avoid confusion in terminology, the terms 'mark' and 'space' will be used for the binary sequence while 1's, 0's, and possibly -1's, will be reserved for the waveform representation.

Two types of rectangular-pulse representations are first distinguished: unmodulated and modulated. We consider first waveforms of the unmodulated type. Waveforms of the modulated type are discussed in Sec. 4.1.5.

4.1.1 Time Characteristics of Unmodulated Rectangular Waveforms

A number of possible unmodulated-pulse representations are shown in Fig. 4.1. The encoding procedure is given in Table 4.1. Additional waveforms of the unmodulated type are presented in Fig. 4.2. The most conventional of the waveforms is the NRZ in which a mark is represented by a 1 and a space by a 0. In bipolar, a space is represented by a 0, the first mark is represented by a +1, the second by a -1, the third by a +1, etc., alternating between ±1 between each mark. In transition encoding, a transition represents a mark and no transition represents a space. A transition is a shift from a "0" to a "1" or vice versa. In dicode, the first mark in a run is represented by a +1 and the first space in a run by a -1. Both marks and spaces after the first bit in a run are represented by 0. A run is an uninterrupted sequence of marks or of spaces.

In duo-binary, a space is represented by a 0, and a mark by either ±1. To determine which, observe the polarity of the last mark and count the number of spaces since. If the number of spaces is even, maintain the same polarity; if the number of spaces is odd, invert the polarity. In the example given in Fig. 4.1, the first mark is arbitrarily chosen to be +1. The first and second marks are separated by one (odd) space, so the polarity of the second mark is inverted to -1. Second and third marks are separated by zero (even) spaces, so polarity is maintained at -1. Spaces are always represented by zeros.

The three remaining codes utilize split-bit waveforms where two different levels are used to represent a single bit. The relative time within the bit allotted to the two levels is usually half and half. In RZ (return to zero)

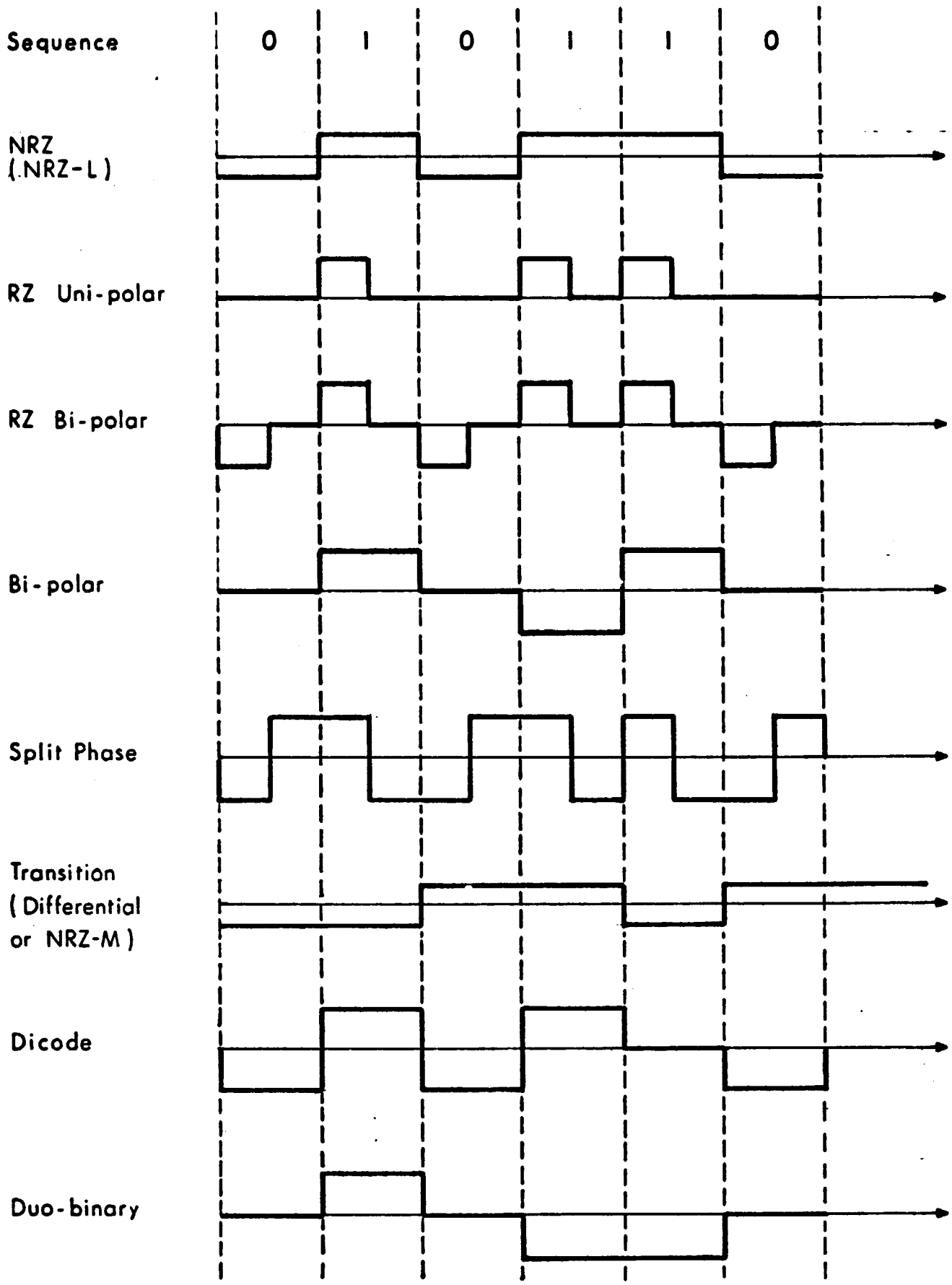


Fig. 4.1 Various Unmodulated Rectangular Waveform Representations of a Binary Sequence

Table 4.1

CODING CHARACTERISTICS

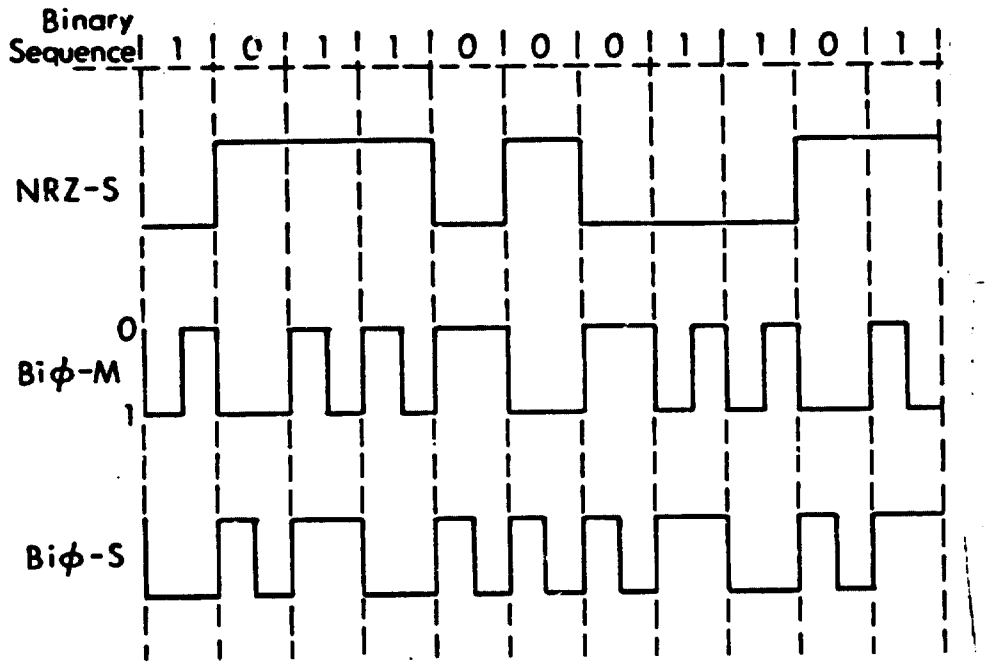
Waveform	Mark	Space
NRZ (NRZ-Level)	1	0
Bi-polar	±1 alternating between each mark	0
Transition (NRZ-Mark)	Transition	No Transition
Dicode	First in run: +1 All subsequent in same run: 0	First in run: -1 All subsequent in same run: 0
Duo-binary	±1 inverting if number of spaces since last mark is odd.	0

Split-bit Waveforms	First Half of Bit	Second Half of Bit	First Half	Second Half
RZ - Unipolar	1	0	0	0
RZ - Bipolar	+1	0	-1	0
Split Phase(Biϕ-L)	+1	-1	-1	+1

Definitions:

Run: A consecutive sequence of marks or of spaces.

Transition: A change from a 0 to a 1 or vice versa.



Transition Encoding Space
"One" is represented by no change in level
"Zero" is represented by a change in level

Bi-Phase-Mark (or Manchester 1)
A transition occurs at the beginning of every bit period
"One" is represented by a second transition 1/2 bit period later
"Zero" is represented by no second transition

Bi-Phase-Space
A transition occurs at the beginning of every bit period
"One" is represented by no second transition
"Zero" is represented by a second transition 1/2 bit period later

Fig. 4.2 Additional Rectangular Waveform Representations of a Binary Sequence

unipolar, for a mark the first half-bit is 1, and the second half-bit a zero. A space is represented by a zero. In RZ bipolar a mark is the same as for RZ unipolar, but for a space the first half-bit is -1, and the second half-bit a zero. In split-phase the first half-bit for a mark is usually +1, and the second half-bit -1. For a space the first half-bit is -1, the second half-bit a +1. Occasionally in split-phase, the above conventions for mark and space are reversed.

4.1.2 Synchronization Characteristics of Unmodulated Rectangular Waveforms

One of the important properties in the consideration of any waveform representation of a binary sequence is the ease of bit synch extraction. Bit synchronization is required to ensure proper detection of any binary signal. The bit synch extraction properties will be discussed for the waveforms.

4.1.2.1 Bit Synchronization for NRZ Waveforms

In the NRZ (nonreturn to zero) waveform, no discrete frequency components appear; that is, there is no spectral line at the bit rate. Therefore, there is no signal energy whatsoever at the bit repetition frequency. With no energy at the pertinent frequency, the bit rate cannot be recovered by simple filtering or phase-locked loop techniques. There are two possible solutions.

The first is to generate energy artificially at the bit rate by periodically inserting a repetitive bit (say a zero) into the sequence. This method has the added advantage that if the additional bit is inserted between complete words, bit and word synchronization may be accomplished simultaneously. However, the additional bits reduce the code efficiency. In addition, the method does not perform well in the presence of noise. For these reasons, periodic insertion of synch bits is not recommended.

The second possibility takes advantage of the information contained in the transitions or changes that occur between a mark and a space. To make use

of this information, a positive pulse is generated at each transition. This in turn is used either to synchronize a free-running multivibrator, or to trigger a one-shot. If marks and spaces are equally probable, there will be a synchronization or trigger pulse on the average for 50 percent of the bits. The pulse at each transition may be generated as in Fig. 4.3. A train of positive and negative going pulses is generated by differentiating the binary waveform. These may be converted to a single polarity by separating the positive from the negative, inverting one and recombining. Alternatively, the bipolar pulses may be passed through a full wave rectifier.

4.1.2.2 Bit Synchronization for RZ Waveforms

The RZ (return to zero) signals contain information inherent to the waveform that is sufficient to obtain bit synchronization. The RZ unipolar waveform is partially self-synchronizing. A sine wave at the bit rate may be obtained by filtering. This may then be tracked in a phase-locked loop that drives a pulse generator (see Fig. 4.4). The generator will produce correctly spaced pulses of the desired width that will either synchronize a free-running multivibrator or trigger a one-shot. RZ bipolar is completely self-synchronizing. By full wave rectification, a square wave at the bit rate is obtained.

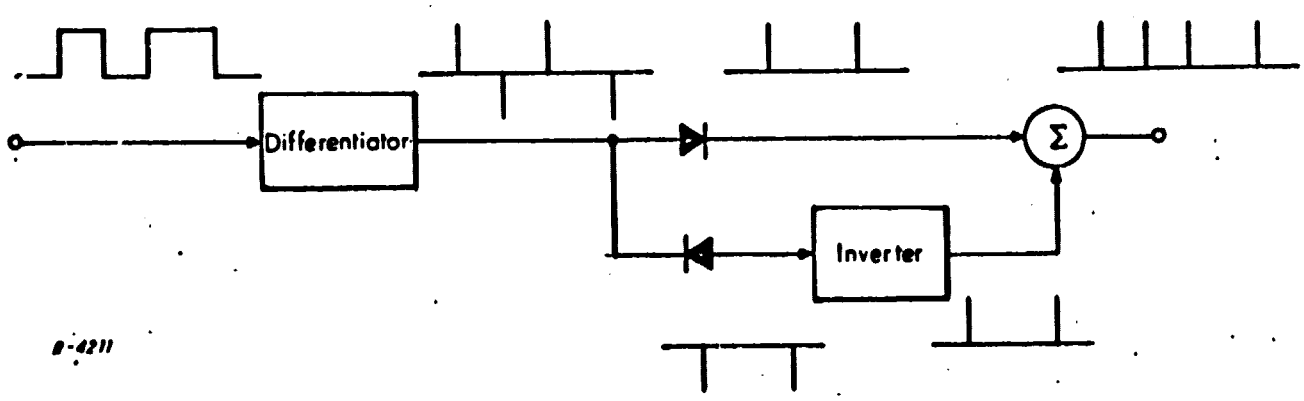


Fig. 4.3 NRZ Synch Pulse Generator

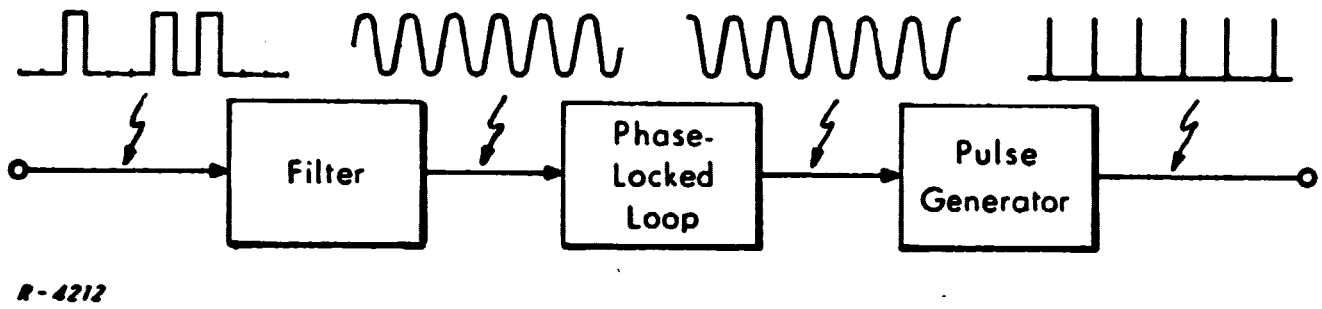


Fig. 4.4 RZ Unipolar Synch Pulse Generator

The cost of the self-synchronizing properties in both unipolar and bipolar RZ is an approximate doubling of the bandwidth for a given bit rate. Also, the separation in signal space of adjacent values of any three level signal is greater than that of a two level signal. Therefore the bipolar RZ signal will suffer more in the presence of noise.

4.1.2.3 Bit Synchronization for Split-Phase Binary Signals

Figure 4.5 is the block diagram of a split phase waveform synchronizer. The waveform at point (A) is indicated for the time period over which the six bit sequence 010001 has been transmitted.

The waveform produced at point (B) by the differentiation of the input data signal consists of a sequence of positive-going and negative-going spikes. The nature of the split-phase signal is such that there is a spike produced in the middle of every bit period. In addition, a spike is generated as the beginning of the period of any bit having the same binary value as the bit preceding it. Full-wave rectification and pulse shaping (hard limiting and 'stretching') result in the waveform at (C) .

The 'phase detection' operation is done by cross-multiplying waveform (C) with the hard-limited output of the VCO (waveform (D)). This provides an error signal that may effectively be averaged over several bit periods by the lowpass filter and applied to the VCO to lock the cycle of (D) in step with that

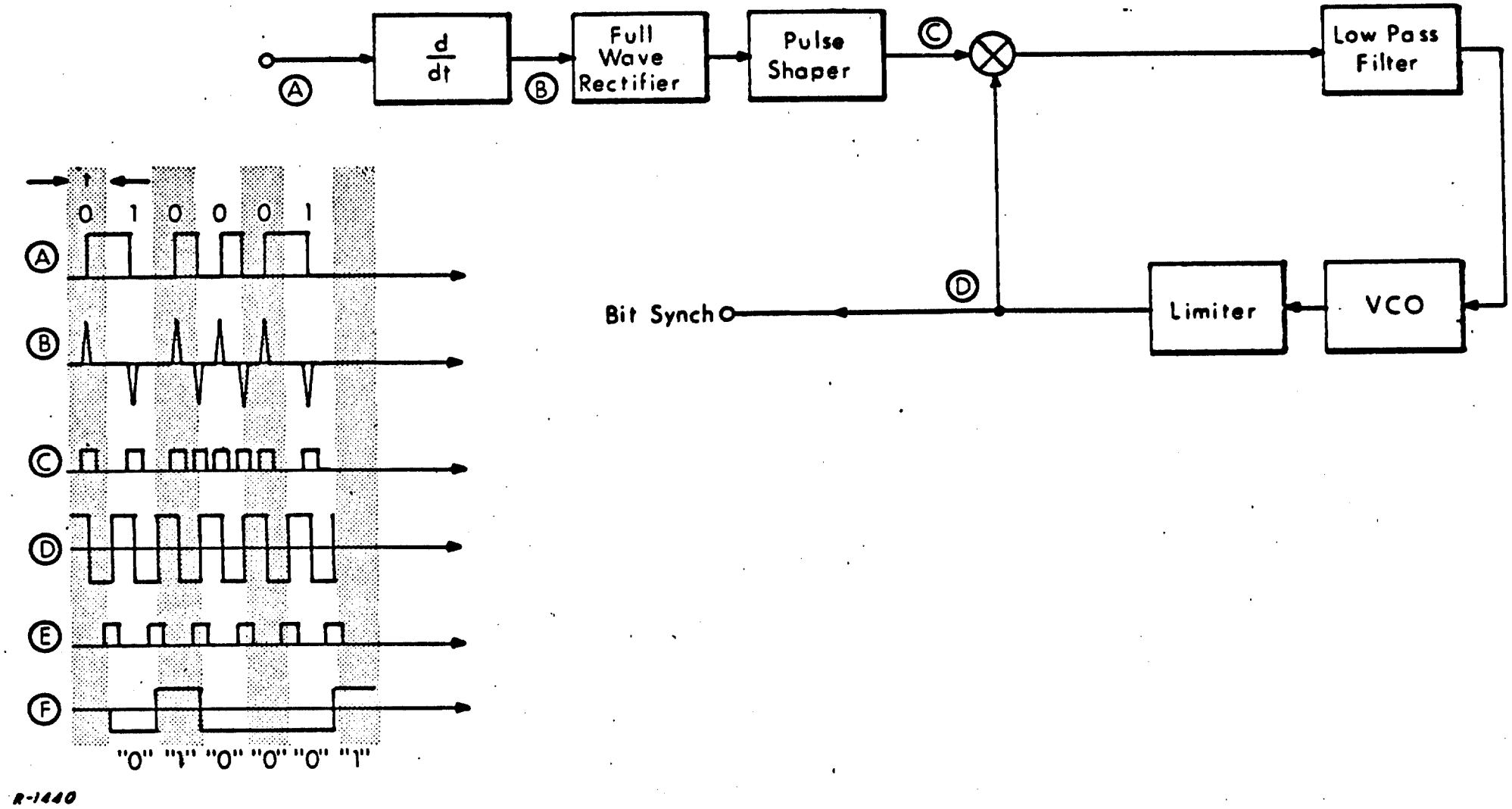


Fig. 4.5 Phase-Locked Loop Application for Split Phase Bit Synchronization

of (C) as shown. When the waveforms are so synchronized, the error signal produced is zero. But note that if waveform (D) is shifted slightly to the right, the effect of the six midpulse markers in (C) is to add together to produce a positive error signal: the two additional markers will produce a negative indication, which subtractively reduces this positive error signal. Since the number of such 'wrong' markers may never exceed the number of midpulse markers, the magnitude and polarity of the error signal may properly be used to lock the VCO output to (C) as indicated.

4.1.2.4 Bit Synchronization for Other Waveforms

The transition-encoded waveform has identical characteristics with the NRZ, and the possible synchronizing procedures are the same. The remaining waveforms – bi-polar, dicode, and duo-binary – while they are three-level rather than binary, have similar characteristics to NRZ, and very nearly identical spectra. As such, the synch procedures for these waveforms are also the same as for NRZ.

4.1.3 Generation and Decoding of Unmodulated Rectangular Waveforms

In the previous sections, the rules for generating the waveforms and methods of synchronization have been discussed. In this section, methods of generation and decoding, including functional block diagrams, will be discussed. The binary sequence input, abbreviated "Bi. Seq.", is a signal of +1 volt or 0 volts. The clock signal is a narrow pulse, occurring at the beginning of each bit. "FF" is a flip-flop with inputs S (set), C (compliment), and R (reset), and outputs "1" and "0". The "1" displays a +1 during the SET condition and 0 during RESET. The "0" output displays simultaneously the compliment of the "1". A dc inverter converts a (0, +1) sequence to a (0, -1). An ac inverter converts (0, +1) to a (+1, 0) sequence.

4.1.3.1 Generation and Decoding of RZ Waveforms

The two RZ waveforms may be generated and decoded by the methods of Figs. 4.6(a) and (b). In the unipolar generator, the clock triggers a one-shot multivibrator generating a train of half-bit width pulses. These are gated on by the mark; the signal remains zero for spaces. In the dicoder, the half-bit width marks set a flip-flop which is reset after each full bit by the clock. This results in a full-bit width output pulse from the FF.

In the bipolar generator, two half-bit wide pulse trains are generated. A positive going pulse train is triggered by the clock, and a negative going train is generated by dc inverting the positive pulses. The marks from the binary sequence gate on the positive pulses and the ac inverted spaces gate the negative train. When summed, they form the bipolar waveform for the decoder, the marks set a flip-flop while the ac inverted spaces reset the FF. The results are zero and one level marks and spaces of full-bit duration.

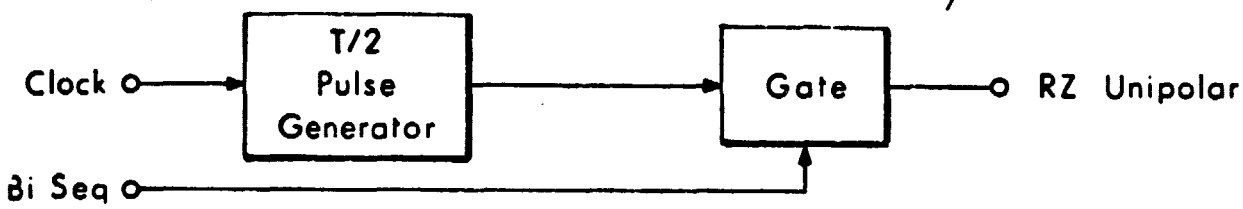
4.1.3.2 Generation and Decoding of Binary Waveforms

The bipolar waveform exhibits a 0 for a space and a ±1 for a mark, alternating between + and - between each mark. In the bipolar generator (see Fig. 4.7), a trigger pulse is generated at the beginning of each mark by comparing the binary sequence with the clock pulse in an "AND" gate. This pulse triggers a +1 volt pulse of width T, a similar -1 volt pulse generator, and compliments a flip-flop. The "1" and "0" outputs of the FF alternately gate on the +1 and -1 pulses, which appear only during marks. During a space the entire system is dormant and no output (zero) appears.

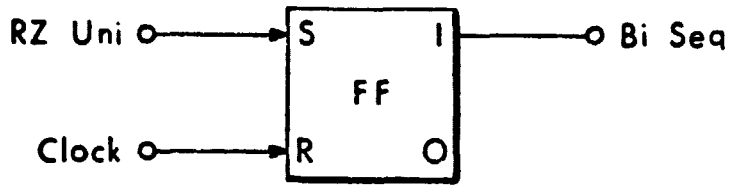
In the decoder, a full wave rectifier returns the bipolar waveform to the standard binary form.

4.1.3.3 Generation and Decoding of Split Phase Waveforms

The split phase waveform may be generated and decoded as in Fig. 4.8. The clock triggers a generator with push-pull output (±1) whose pulse width is

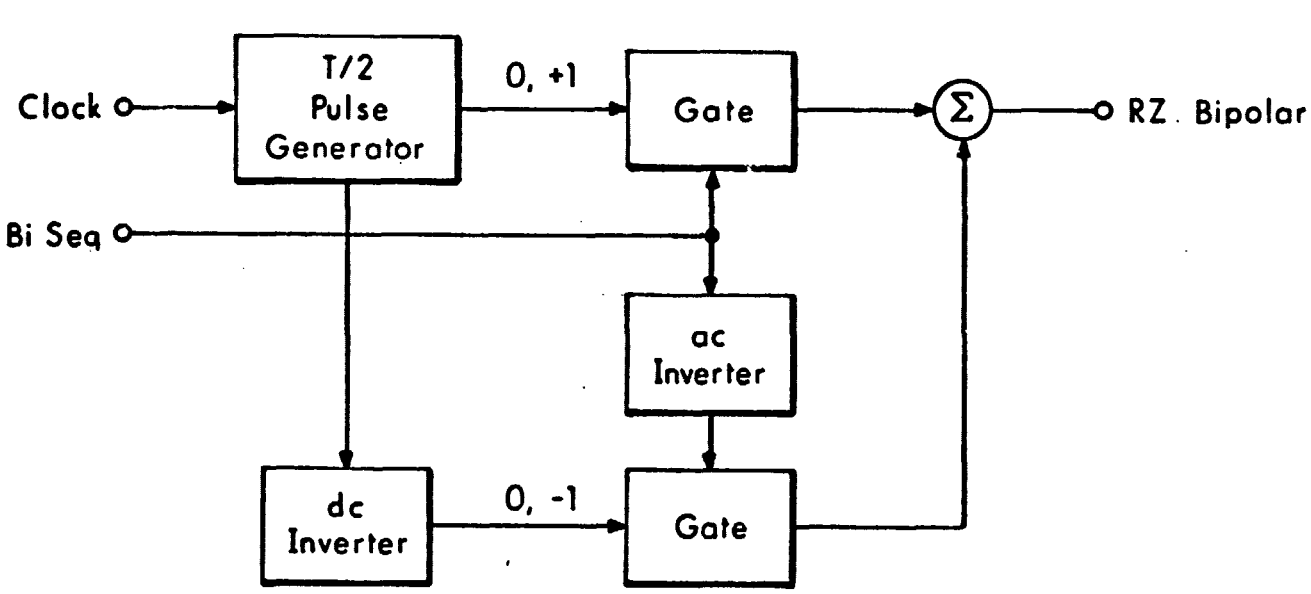


Generator

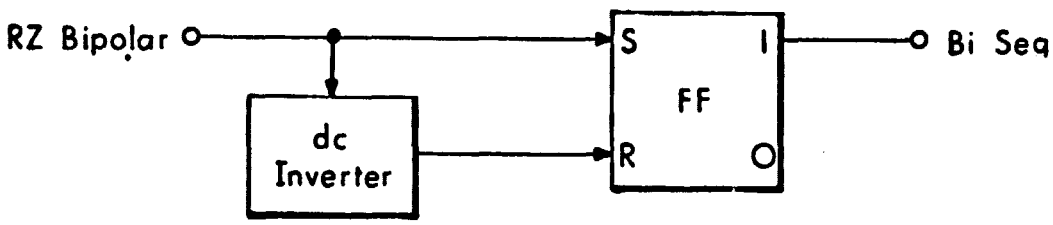


Decoder

(a) Unipolar



Generator



Decoder

R-4203

(b) Bipolar

Fig. 4.6 RZ Generation and Decoding

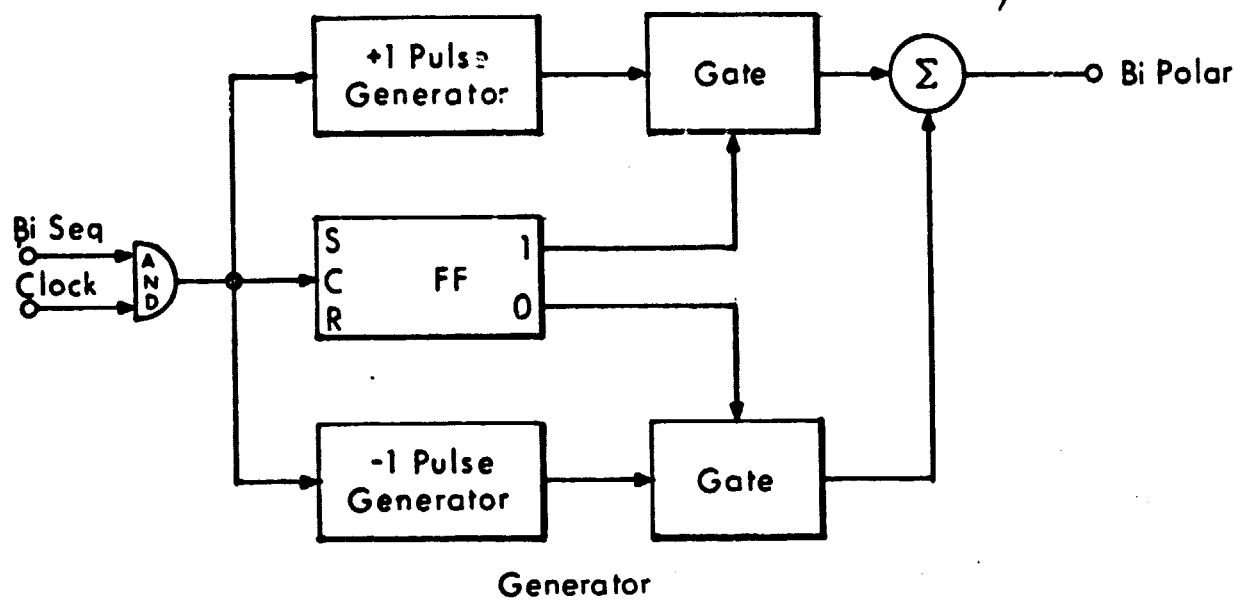
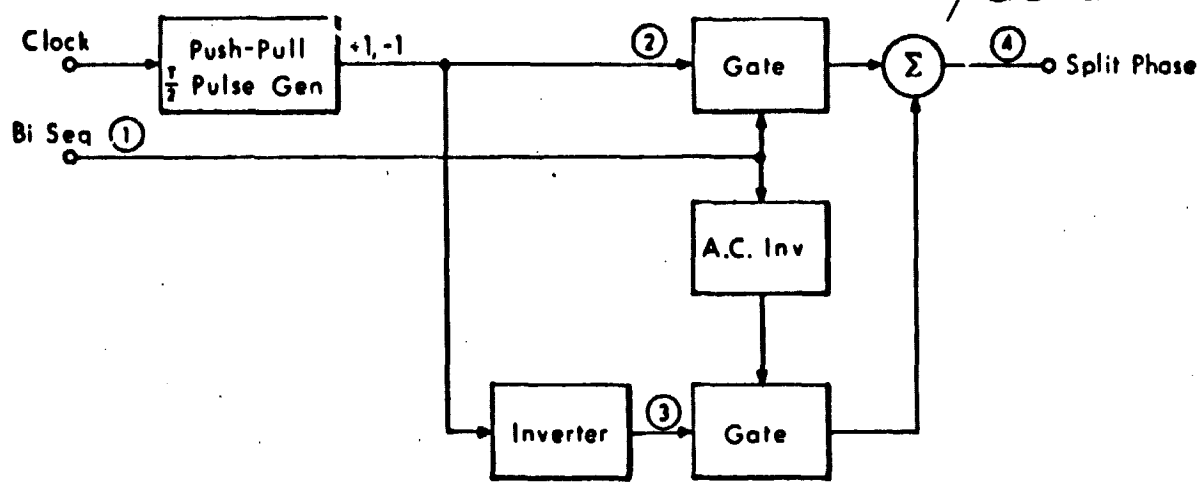


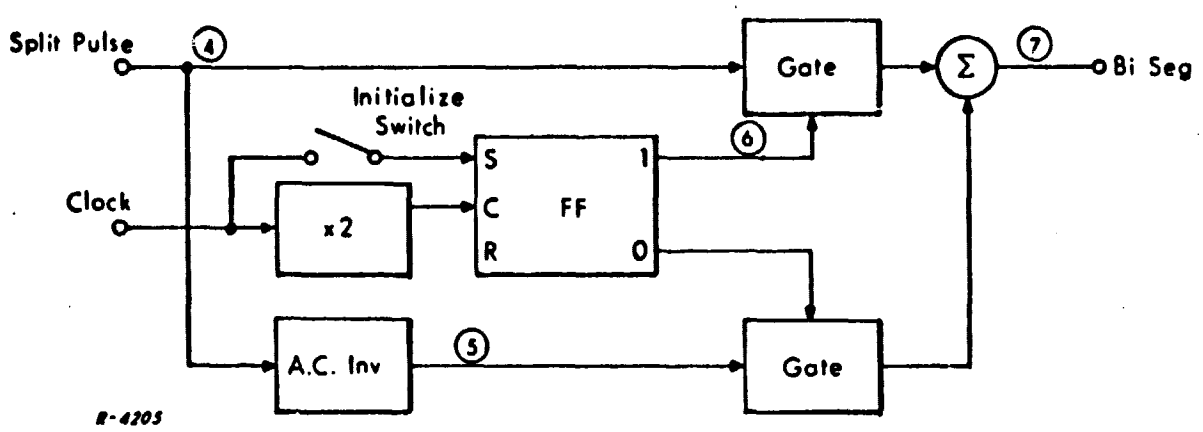
Fig. 4.7 Bipolar Generation and Decoding

half the bit period. This is a square wave, (2), at the bit rate with amplitude limits of +1 and -1. The square wave is gated by the mark, (1), while the inverted square wave, (3), is gated by the ac inverted space. Note that since the square wave has no dc component, it is immaterial whether we speak of an ac or dc inverter. The summed gate outputs make up the split phase waveform (4).

In the decoder, the clock pulse train is doubled in frequency and used to compliment a flip-flop. The FF output, (6), is used to gate on the normal split-phase waveform (4) during the first half of each bit, and the inverted split-phase waveform (5) during the second half of each bit. The summed gate outputs, (7), give us the original binary sequence of (1) except for level. If for some reason the 0, 1 level is required, we need only attenuate by 1/2 and dc re-reference the waveform of (7). In order to ensure that the binary sequence is not inverted during decoding, it is necessary that the flip-flop start in the correct state. This may be accomplished by feeding the unmultiplied clock pulse to the SET input of the FF by momentarily depressing the initialize switch at the beginning of the message.



Generator



Decoder

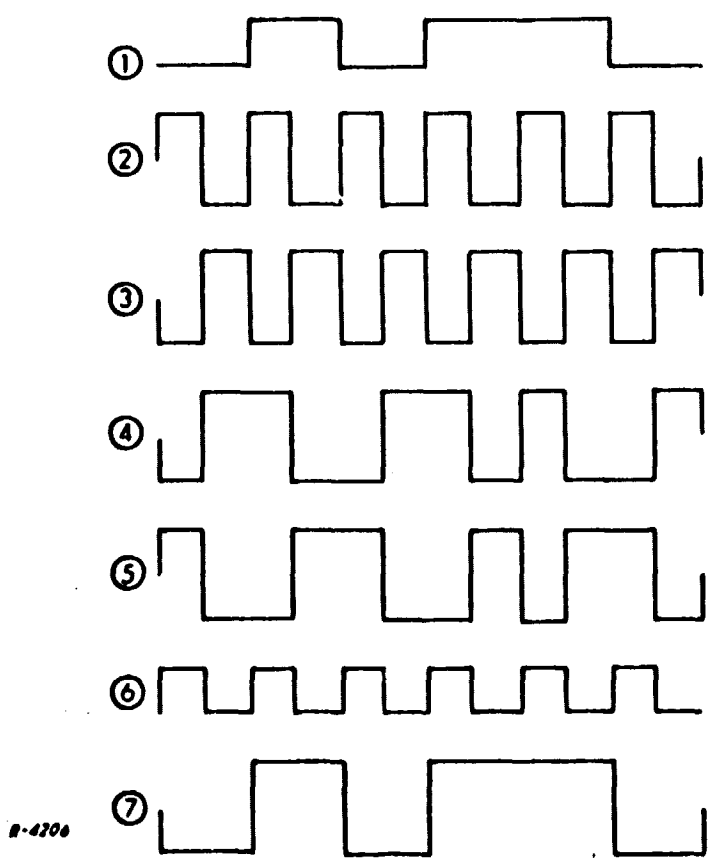


Fig. 4.8 Split Phase Generation and Decoding

4.1.3.4 Generation and Decoding of Transition Encoded Waveforms

In transition encoding for each "mark" a transition is sent; for each "space", no transition is transmitted. A transition is a shift from a mark to a space or vice versa. Transition encoding may be implemented by the method depicted in the block diagram of Fig.4.9.

Each coded bit is compared with a clock pulse in an "AND" gate, and for each "1", a complement pulse is sent to a flip-flop. A delay line delays the clock pulse by half a bit width so that the comparison may be made at the center of a mark or space, rather than near either end. The flip-flop then exhibits a transition for each "mark" and remains unchanged for a "space." Its output is the transition encoded signal. The decoder compares each bit with the previous in an "Exclusive OR" gate. An "Exclusive OR" gate has an output if either input is present but not if both are. Its output is a "1" if the two consecutive bits are different (transition occurred) and "0" if they are the same (no transition). The original binary sequence is thus recovered. Note that the one-bit delay, used in comparing the preceding bit with the current bit, is just a one-stage shift register consisting of two flip-flops.

Since the decoder looks only for the presence or absence of a transition, either the encoded signal or its complement will be decoded properly, as they both have transitions in the same phases. Therefore, inversion in the receiver (as for example in demodulating a PSK signal), has no effect on the decoded message. However, with transition encoding there is a double error effect. If one bit is erroneously decoded, it affects the presence or absence or a transition both at its beginning and end. There will, in general, then be an error in two bits of the binary sequence. In addition, the complete message requires one additional bit, as it is impossible to tell if a transition occurred until the beginning of the second bit. The addition of this bit, however, causes a negligible increase in the length of any practical message.

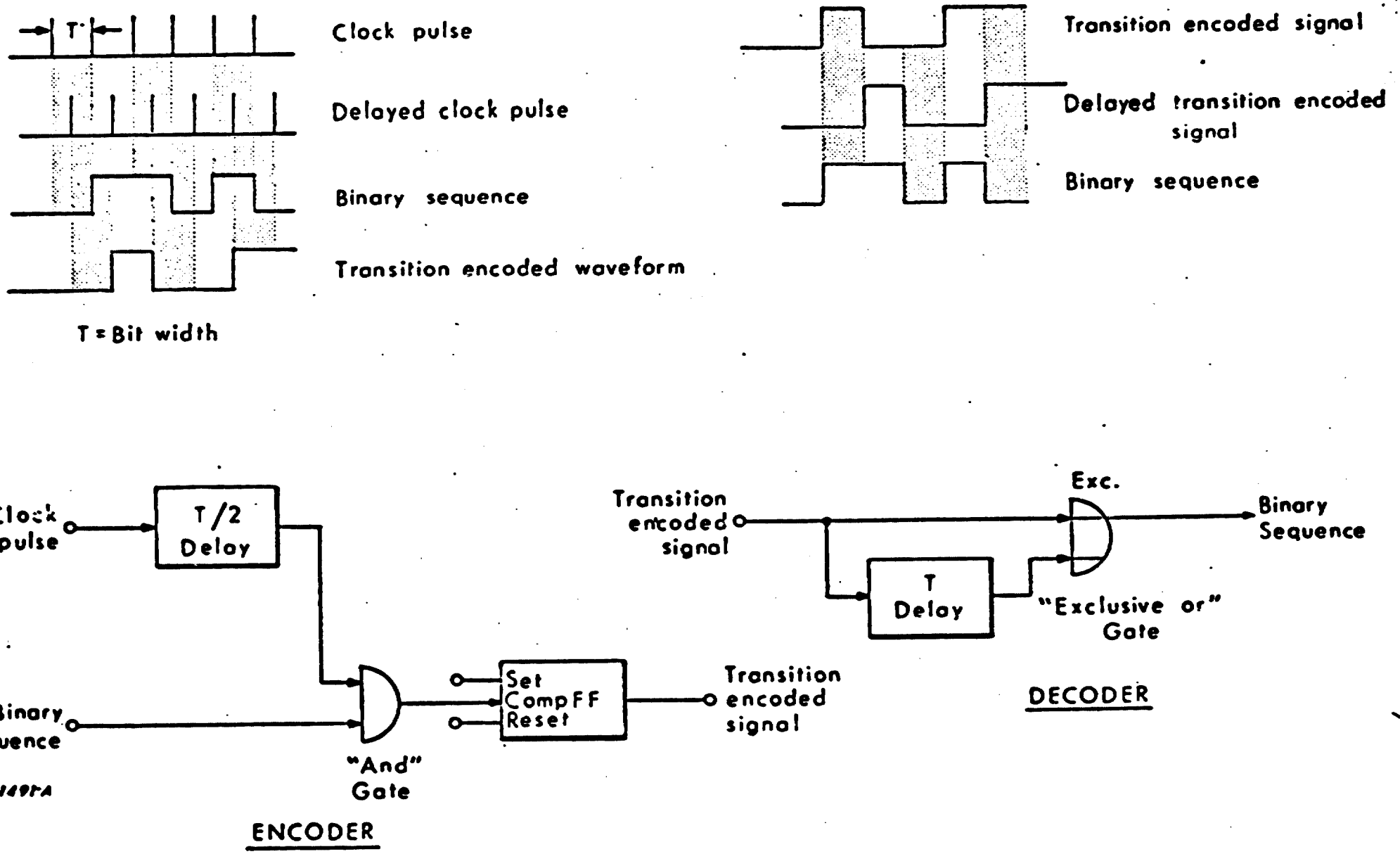


Fig. 4.9 Generation and Decoding of Transition Encoded Waveforms

4.1.3.5 Generation and Decoding of Dicode Waveforms

The dicode waveform exhibits a +1 for the first mark in a run and a 0 for all subsequent marks in the same run. A run is an uninterrupted sequence of marks or of spaces. A -1 is exhibited for the first space in a run, and a 0 for all subsequent spaces in the same run. Pulses are conserved by this method, and some bandwidth reduction is obtained. Dicode may be generated as shown in Fig. 4.10. Two pulse generators are used which are triggered only by the leading edge of a positive-going pulse. The first mark in a run triggers the +1 pulse generator that puts out a single pulse of width T (one bit width) and value +1. All subsequent marks in the run are ignored, as the generator only looks for a positive-going leading edge. Similarly, the inverted first space in a run triggers the -1 pulse generator that puts out a single pulse of value -1.

In the decoder, a +1 sets a flip-flop that displays a 1 output until an inverted -1 arrives to reset it to 0.

4.1.3.6 Generation and Decoding of Duo-binary Waveforms

In duo-binary space is represented by a 0, and a mark by a ±1 depending upon the number of spaces occurring between the current mark and the last mark. If the number is even, polarity is the same; if odd, the polarity is inverted. A certain amount of bandwidth reduction is obtained in this manner. In the duo-binary generator, Fig. 4.11, flip-flop No. 1 keeps track of the number of spaces since the last mark. It is reset by each mark and complimented by each space. These inputs are "AND"-ed with the clock pulse since the flip-flop responds to pulses, not levels. If its output is 1, the number is odd; and if its output is 0, the number is even. FF No. 2 determines the polarity of the current mark output. At each mark, the output of FF No. 1 is sampled and if it is a 1 (odd number of spaces), FF No. 2 is complimented. The mark pulse resetting FF No. 1 is

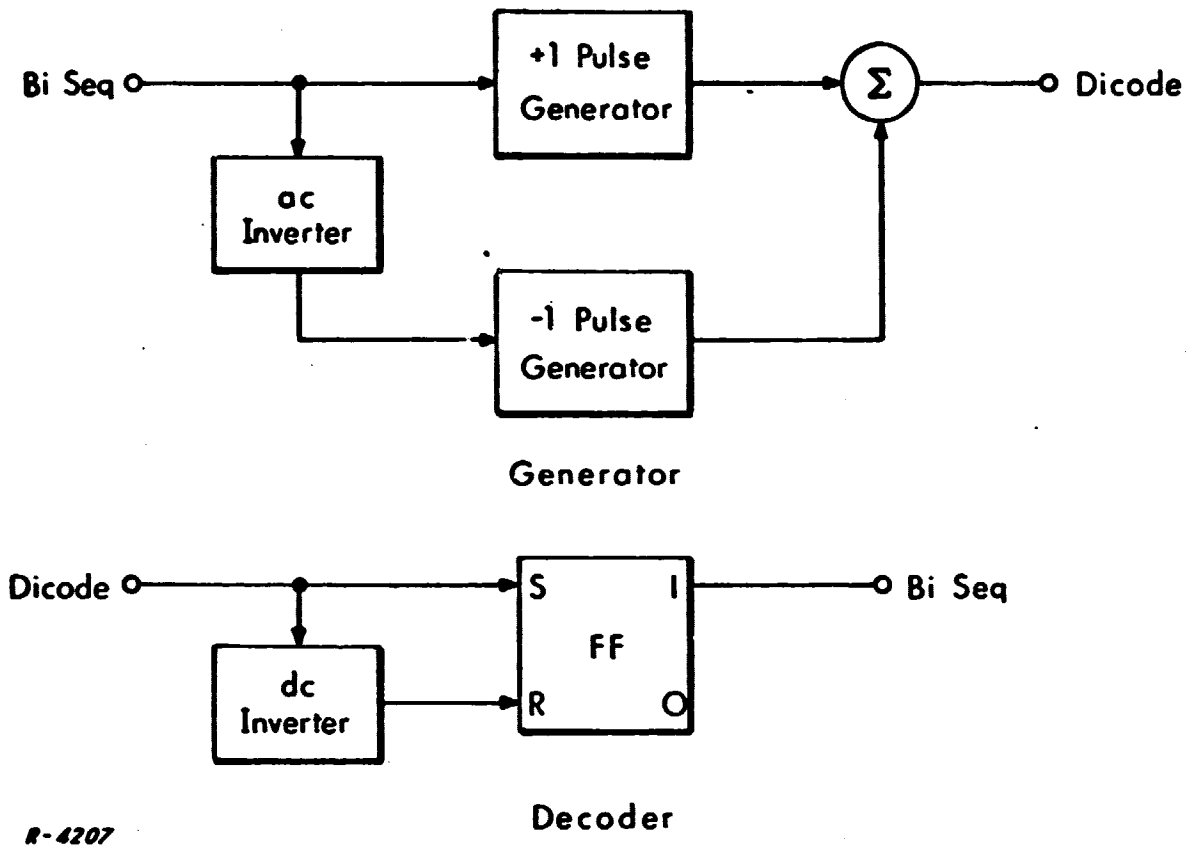
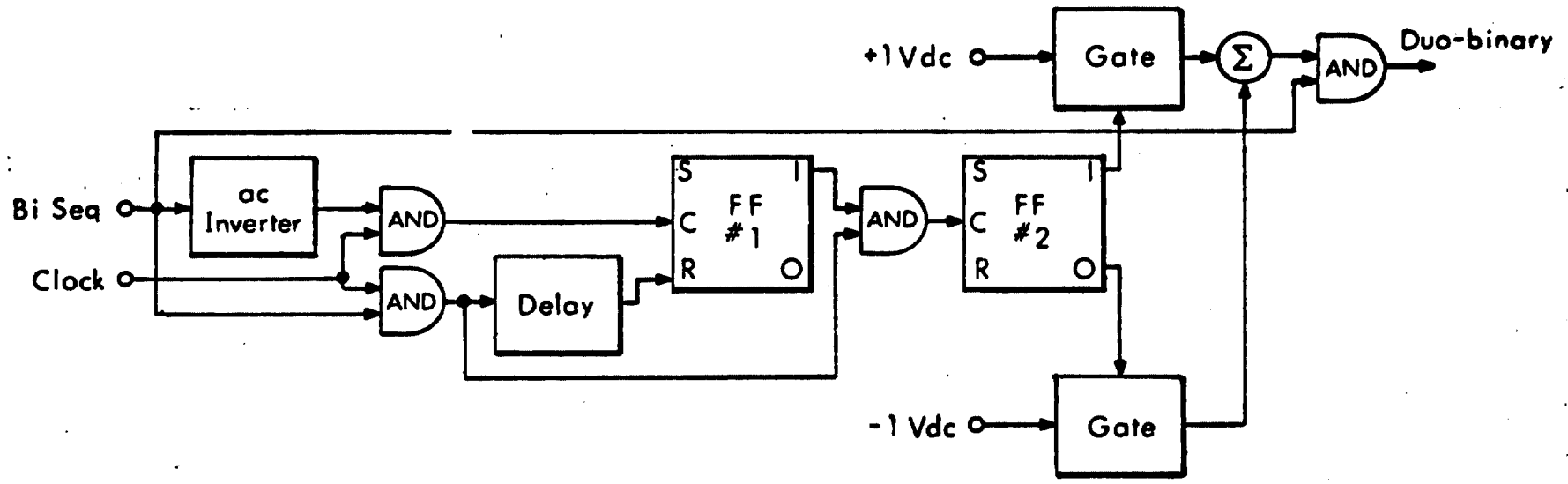
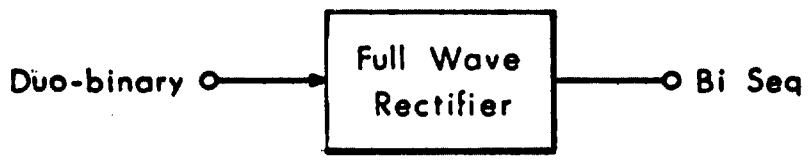


Fig.4.10 Generation and decoding of dicode waveforms



Generator



Decoder

R-4208

Fig. 4.11 Generation and decoding of duo-binary waveforms

delayed by a small fraction of a bit to prevent the FF from being reset before the mark pulse has had a chance to sample its output. FF No.2 then applies either a +1 volt or a -1 volt dc level to the output, depending upon its state--+1 for SET, -1 for RESET. This is gated on by the mark from the binary sequence. The final AND gate is gated off by a space and a zero output is displayed.

4.1.4 Spectral Characteristics of Unmodulated Rectangular Representations

4.1.4.1 Spectral Density Functions

The spectral occupancy of the various waveforms is of interest when considering transmission through any band-limited channel and recording the waveform on magnetic tape. The low-frequency response of a tape recorder does not go to zero frequency, so the spectral density of the waveform near 0 Hz is of special interest. The high-frequency cutoff of the recorder will limit the highest bit rate that can be accommodated.

The spectral densities of the NRZ, bi-polar, transition-encoded, dicode, and duo-binary waveforms are given by the $[\frac{\sin x}{x}]^2$ curve of Fig.4.12 where T is the bit width, and f is the frequency in Hz. Unity transmitted power is assumed. Figure 4.12 also gives the spectrum of RZ-unipolar, provided we observe that T is now the pulse width, or half the bit width. Note that this has the effect of doubling the scale of the abscissa and halving the ordinate. The bandwidth of the RZ unipolar waveform, then, is double that of NRZ for a given bit rate.

Split-phase and RZ-bipolar spectra are given by the $[\frac{\sin^2 x}{x}]^2$ curve of Fig.4.13, where again T equals the bit width. Note that the bandwidths of split phase and RZ bipolar are also double that of NRZ for a given bit rate.

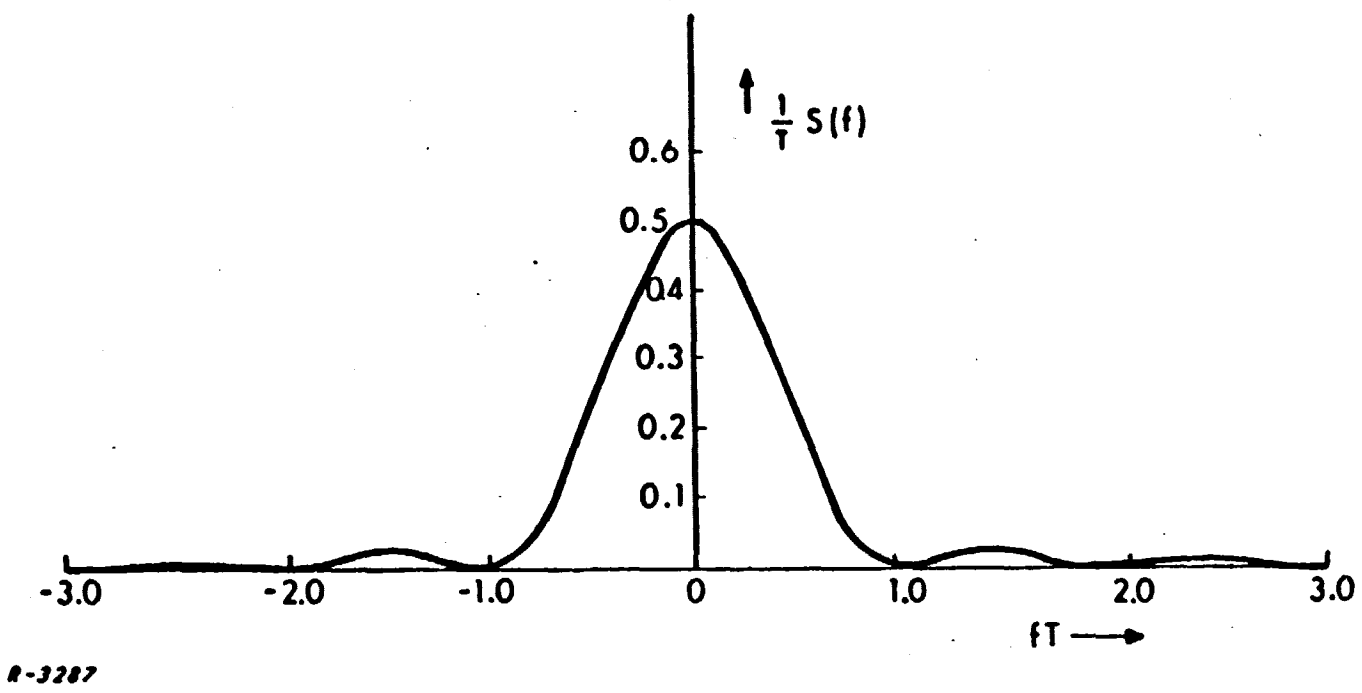


Fig. 4.12 Spectral Density Function of Rectangular Pulse of Duration T.

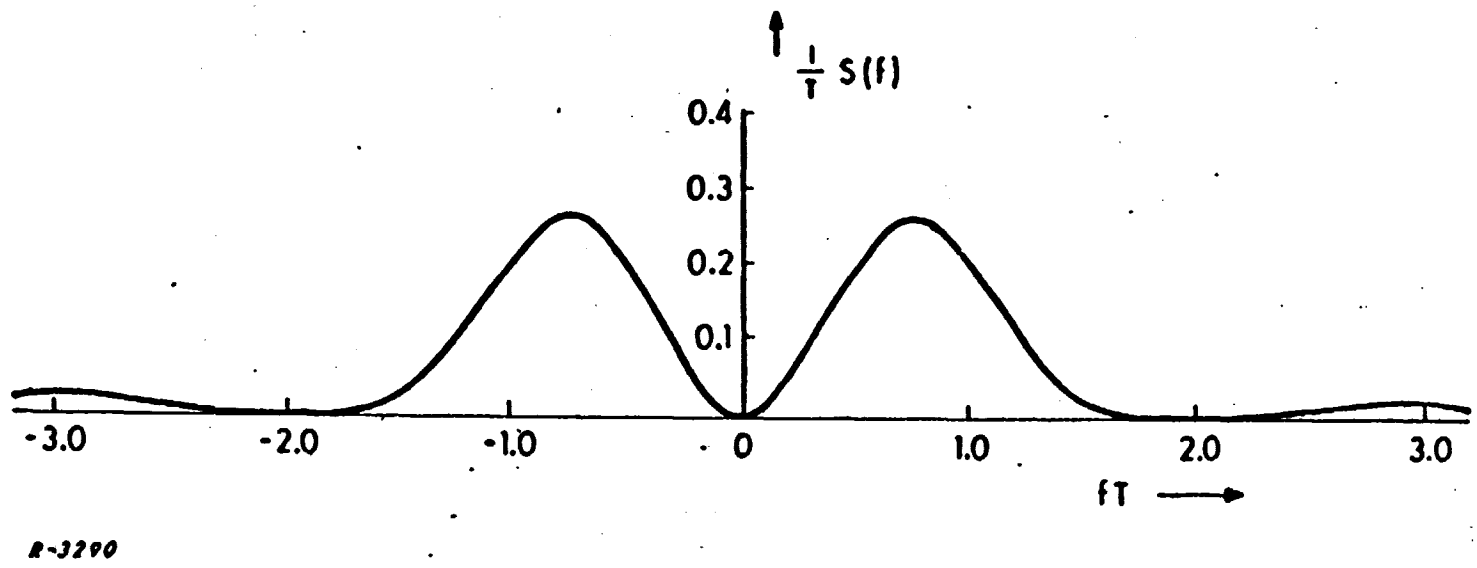


Fig. 4.13 Spectral Density Function of Split Phase and RZ Bipolar

4.1.4.2 Bandwidth Enclosing a Specified Percentage of Bit Energy

Since the spectra of all the waveforms extend to infinity, 100 percent of the energy cannot be transmitted through any finite channel. It is of more practical interest, then, to inquire what the bandwidth of a channel must be to contain a specified percentage of the signal energy, say 90 or 95 percent. This is given by the integral of the curves of Figs. 4.12 and 4.13, and is plotted in Figs. 4.14 and 4.15 respectively. Figure 4.14 gives the data for NRZ, bipolar, transition, dicode, and duo-binary, where T is the bit width and f is the frequency in Hz, and for RZ unipolar where T is half the bit width. Figure 4.15 gives the data for RZ bipolar and split phase where, again, T is the bit width. From Fig.4.14 we see that the bandwidth required to contain 90 percent of the bit energy of an NRZ waveform signal is 0.78 times the bit repetition rate. To contain 95 percent we require 1.62 times the bit rate. For RZ unipolar 90 and 95 percent require 1.56 and 3.24 times the bit rate respectively. From Fig. 4.15, a split phase signal requires 1.5 times the bit rate for 90 percent of the energy and 2.36 times for 95 percent.

4.1.4.3 Spectral Density Near 0 Hz

The behavior of the spectral density function of a waveform near 0 Hz is of interest in generating RF signals with separable carrier components and in tape recording. From Figs. 4.14 and 4.15 we observe that the spectral densities of split phase and RZ bi-polar go to zero parabolically at 0 Hz, but attain their peak values at 0 Hz for the other waveforms.

4.1.5 Pulse Modulation Representations of Binary Sequences

Continuing the study of waveform representation of binary sequences, we shall next discuss pulse modulation--the representation of the binary data by varying some parameter of a pulse. Pulse modulation may be divided into three categories: 1) pulse amplitude modulation (PAM), 2) pulse duration modulation ()DM), and 3) pulse position modulation (PPM).

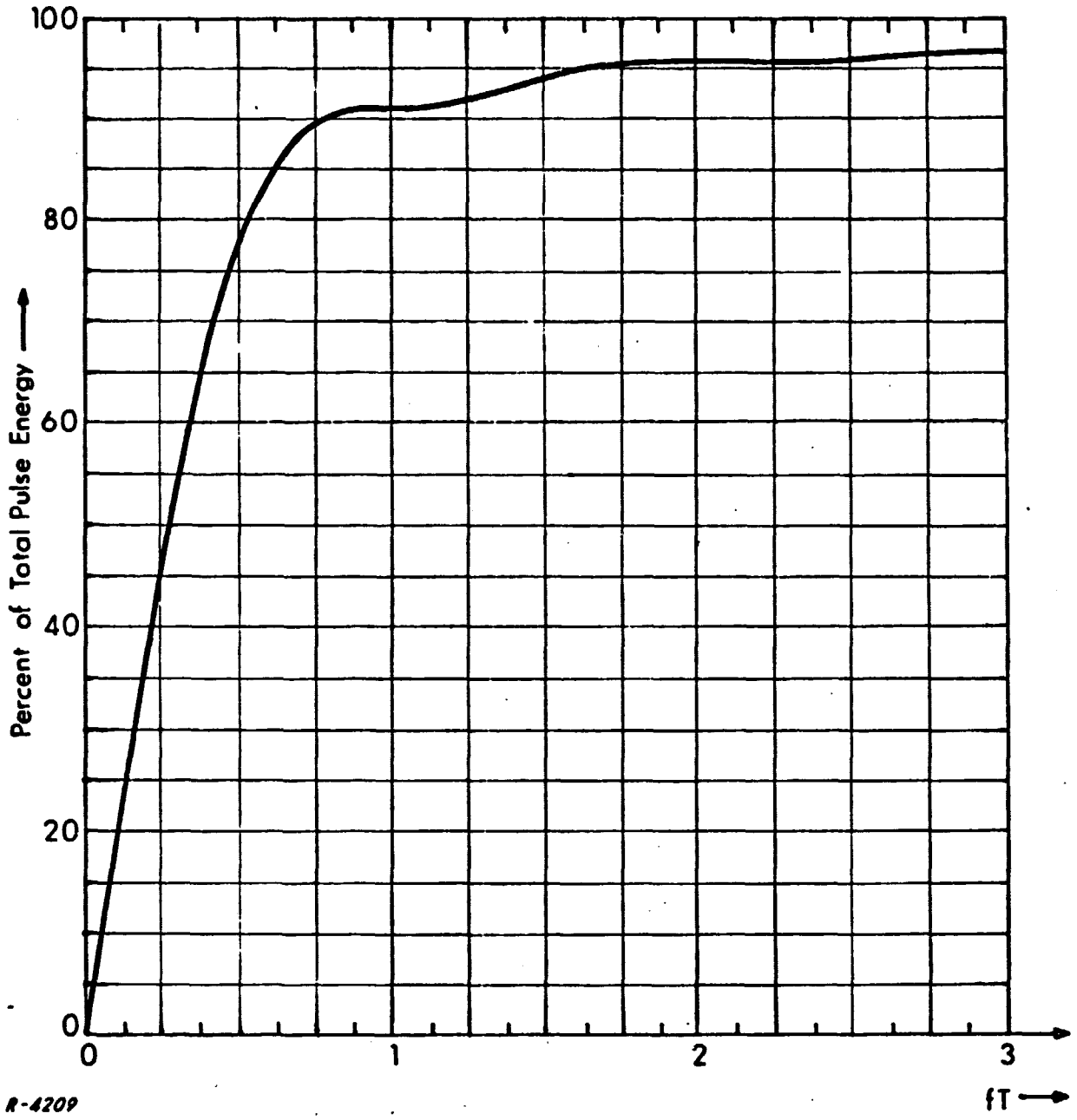


Fig. 4.14 Plot of Percent of Total Pulse Energy Enclosed Within a Frequency Range of ±f Hz About the Center of the Spectral Density Function, for a Rectangular Pulse of Duration T.

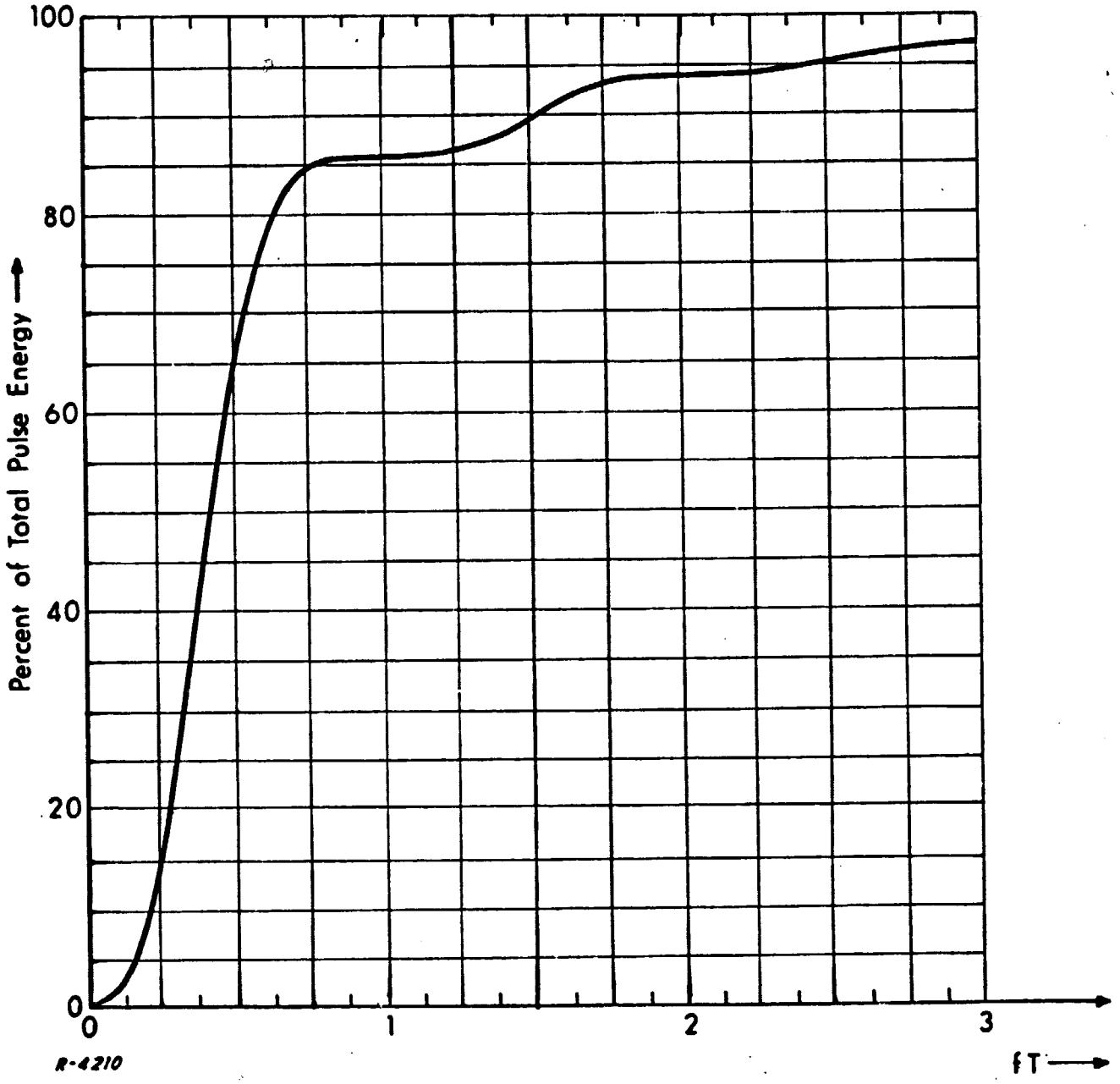


Fig. 4.15 Plot of Percent of Total Bit Energy Enclosed Within a Frequency Range of ±f Hz About the Center of the Spectral Density Function, for Rectangular Split-Phase and RZ Bipolar Pulses with Bit Duration T.

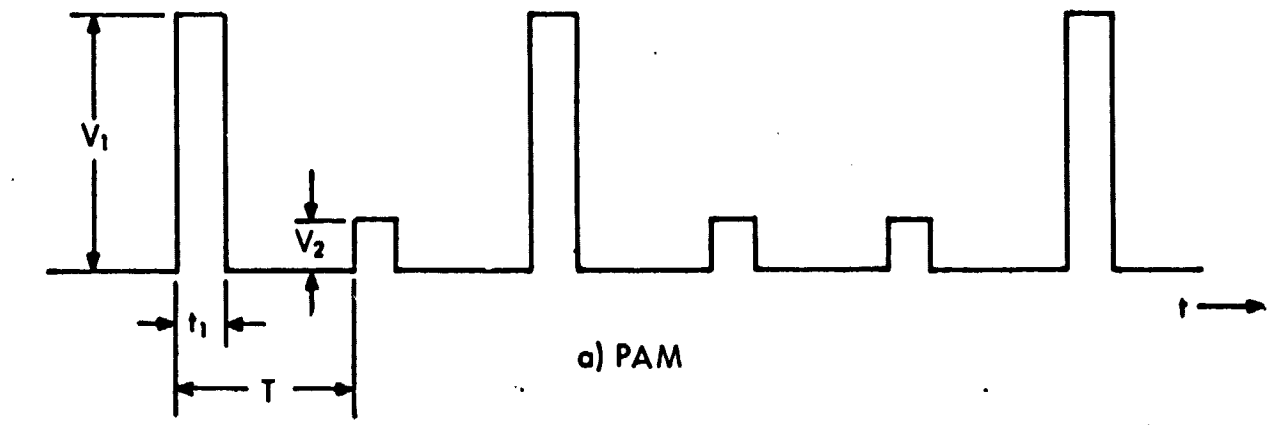
4.1.5.1 Pulse Amplitude Modulation

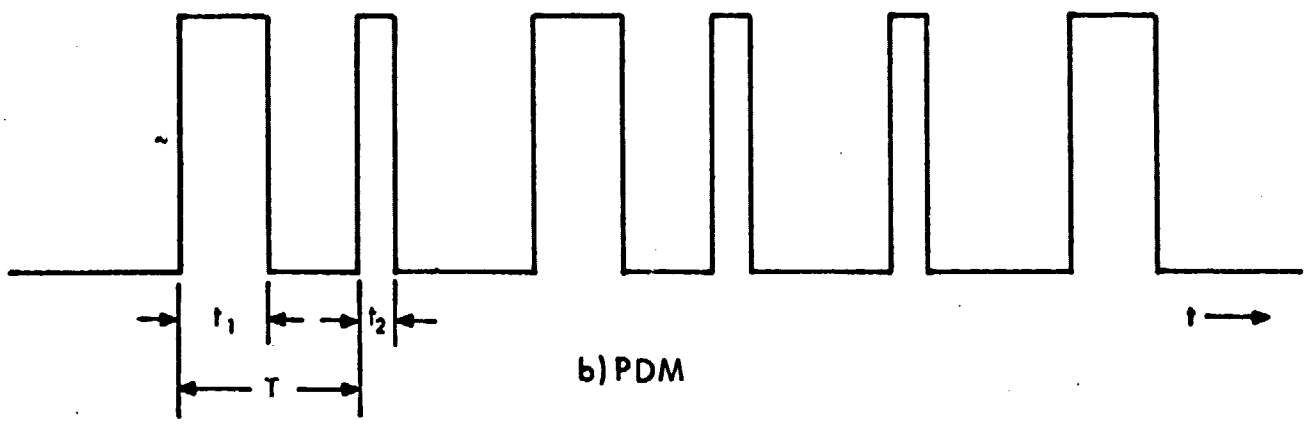
In pulse amplitude modulation (PAM) a train of pulses of fixed width and constant separation is modulated in amplitude. In binary PAM, the magnitude of each pulse is constrained to be one of two levels. Usually the larger level represents a mark and the smaller, a space. A typical PAM waveform is given in Fig.4.16(a). The spectrum of the PAM signal is given in Fig.4.17.

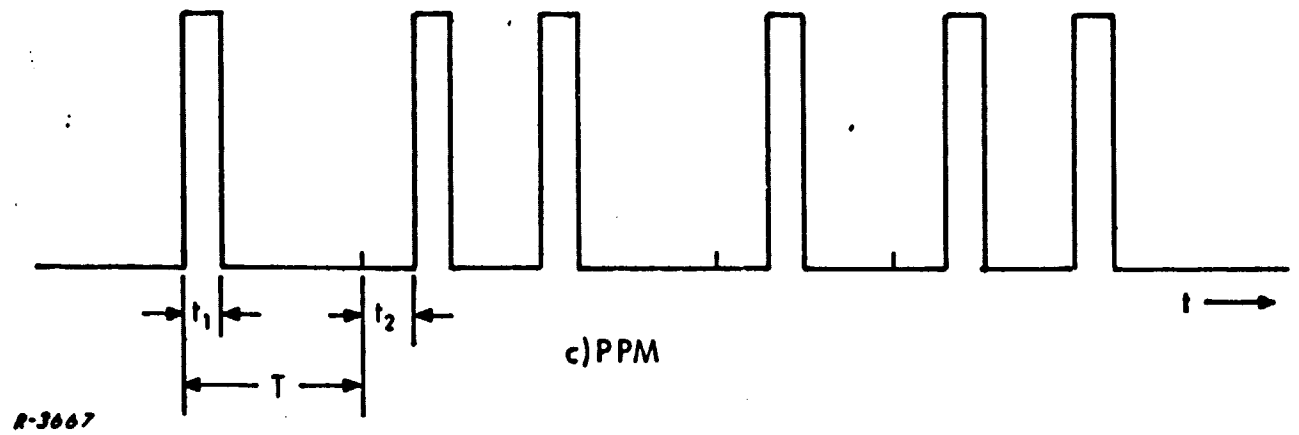
The PAM signal may be generated by the method of Fig. 4.18(a). A synchronous pulse generator generates a pulse train at the same frequency as and coincident with the binary input data. That is, a pulse is generated for each binary "bit", whether it is a mark or a space. These pulses are limited by two limiters in parallel to the two desired output voltage levels V_1 and V_2. Note that either V_1 and/or V_2 can be negative. The V_1 level is gated on by the binary "one", while the V_2 level is gated on by the inverted binary "zero". The two gate outputs are then summed to form the PAM signal.

A similar signal may be obtained somewhat more simply by analog means. However, the analog method is considered inferior as it does not maintain a close tolerance on the output levels. A block diagram of an analog PAM generator is given in Fig.4.19(a). A synchronous pulse generator generates the pulses as before. In this method the pulses are fed to a variable-gain amplifier whose gain is controlled by the binary modulating signal. This amplifier might be a transistor (or vacuum tube) with the pulse train applied to the base (grid), the binary modulation applied to the emitter (cathode) and the output taken from the collector (plate). The amplifier output is a PAM signal.

In detection (see Fig.4.18(b)), the amplitude-modulated pulses are tested for amplitude level in a comparator. As in all digital systems, the bit rate is recovered and used to synchronize the comparator output. The output is not allowed to change except when a new bit is expected.







The sequence 101001 is represented in all cases

Fig. 4.16 Rectangular Pulse Modulation Waveforms

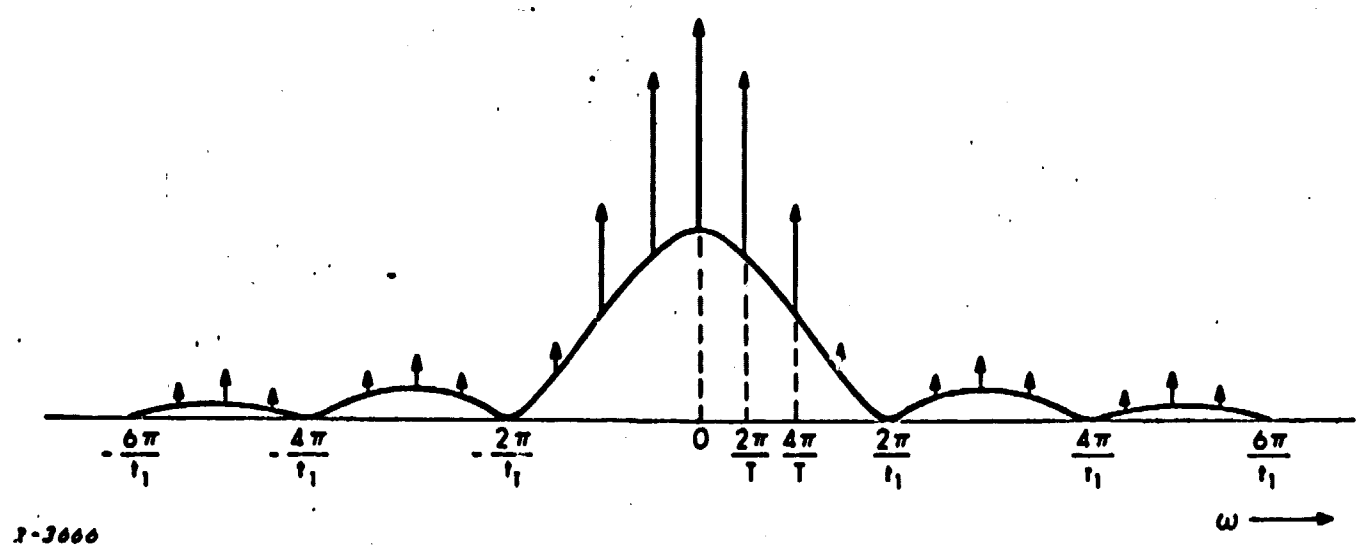
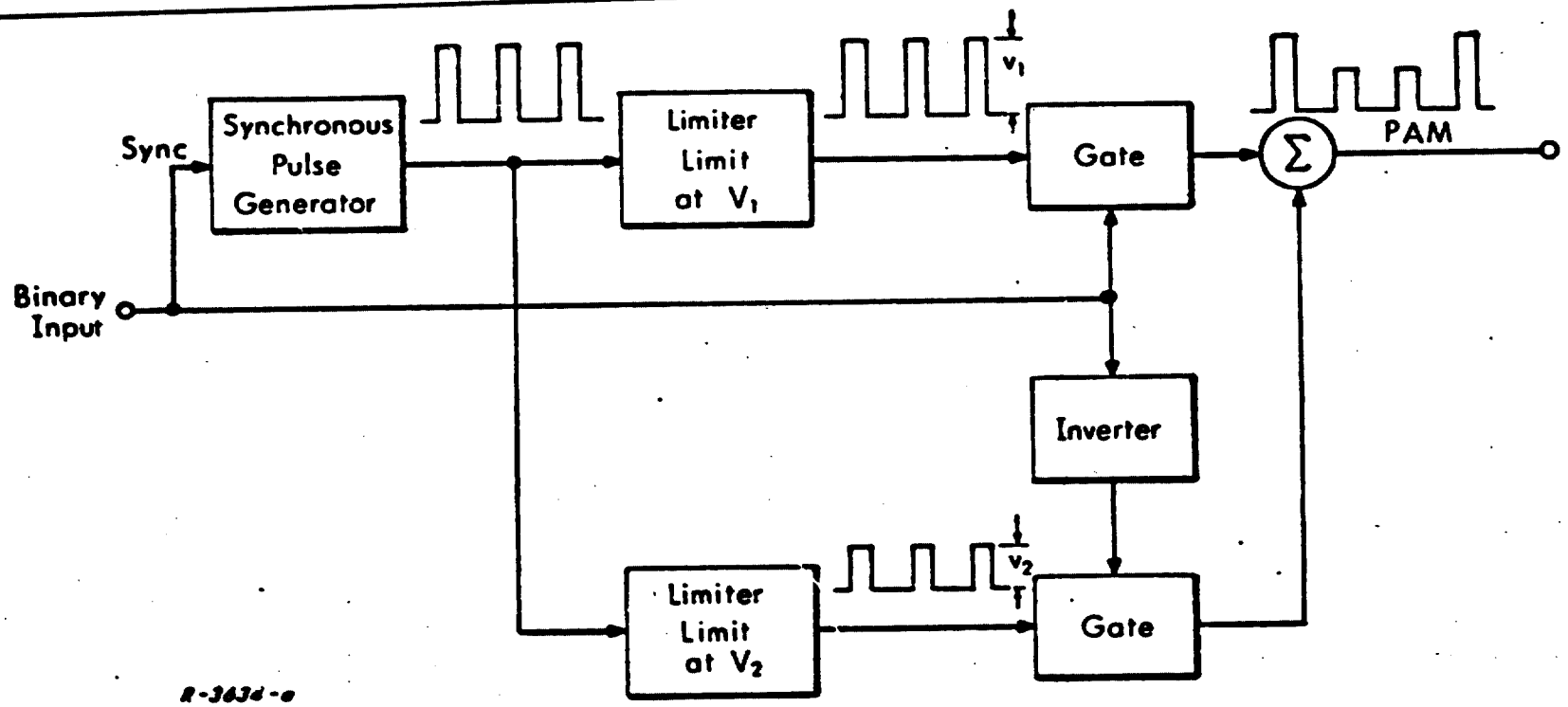


Fig. 4.17 PAM Spectrum

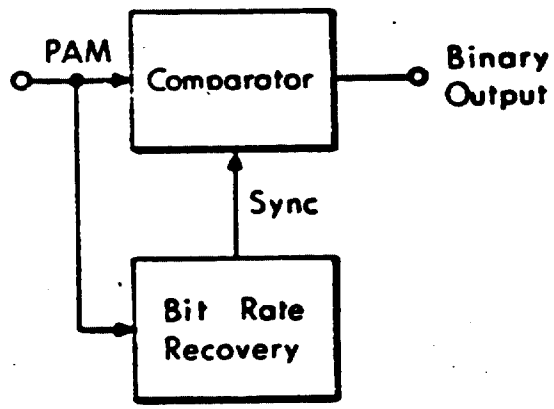
4.1.5.2 Pulse Duration Modulation

In pulse-duration modulation (PDM) a train of pulses of constant amplitude is modulated in width. Time between leading (or trailing) edges is fixed. In binary PDM, the duration of the pulse is constrained to be one of two widths. Binary PDM is used extensively in the current time code systems. A typical PDM waveform is given in Fig. 4.16(b).

The PDM signal may be generated by the methods of Fig. 4.20(a). As in PAM, a synchronous pulse generator is used to generate a pulse train, one pulse for each mark or space. However, in PDM, rather than being used directly, these pulses are used to trigger two blocking oscillators (B.O.'s) in parallel. Each B.O. puts out a train of pulses of fixed width and constant amplitude. The pulse widths from the two B.O.'s are t_1 and t_2 respectively, $t_1 \neq t_2$. One B.O. output is gated on by the binary "mark", while the other is gated on by the inverted space. Normally the mark will be used to control the longer of the two pulse widths. The two gate outputs are then summed to form the PDM signal.

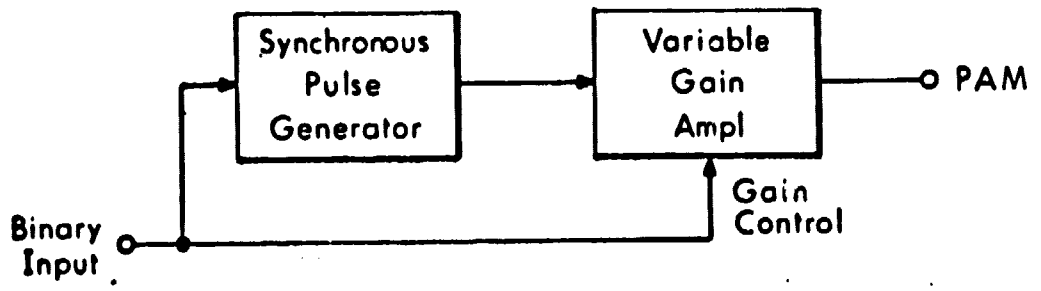


a) Modulator

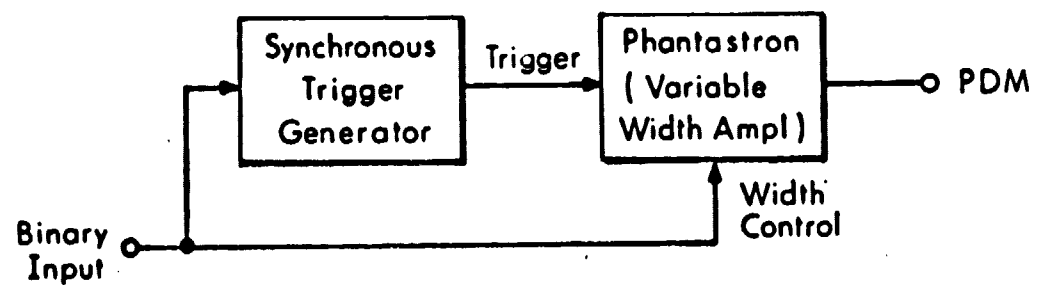


b) Detector

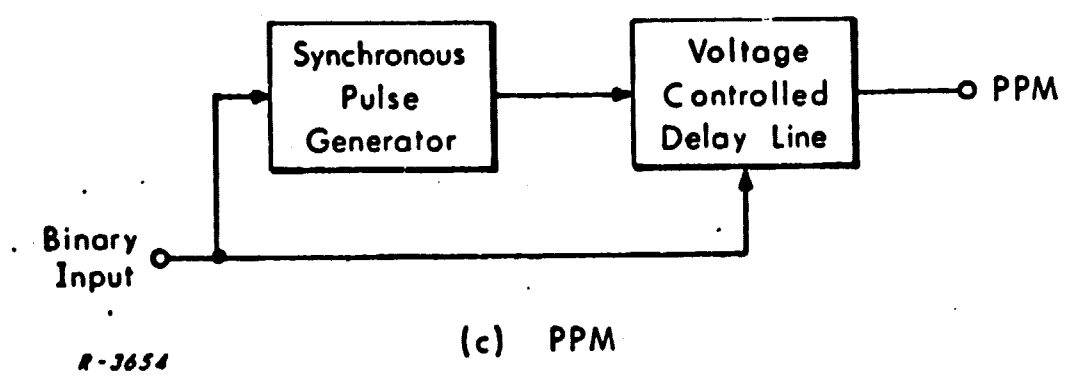
Fig. 4.18 Generation and Detection of Pulse Amplitude Modulation



(a) PAM



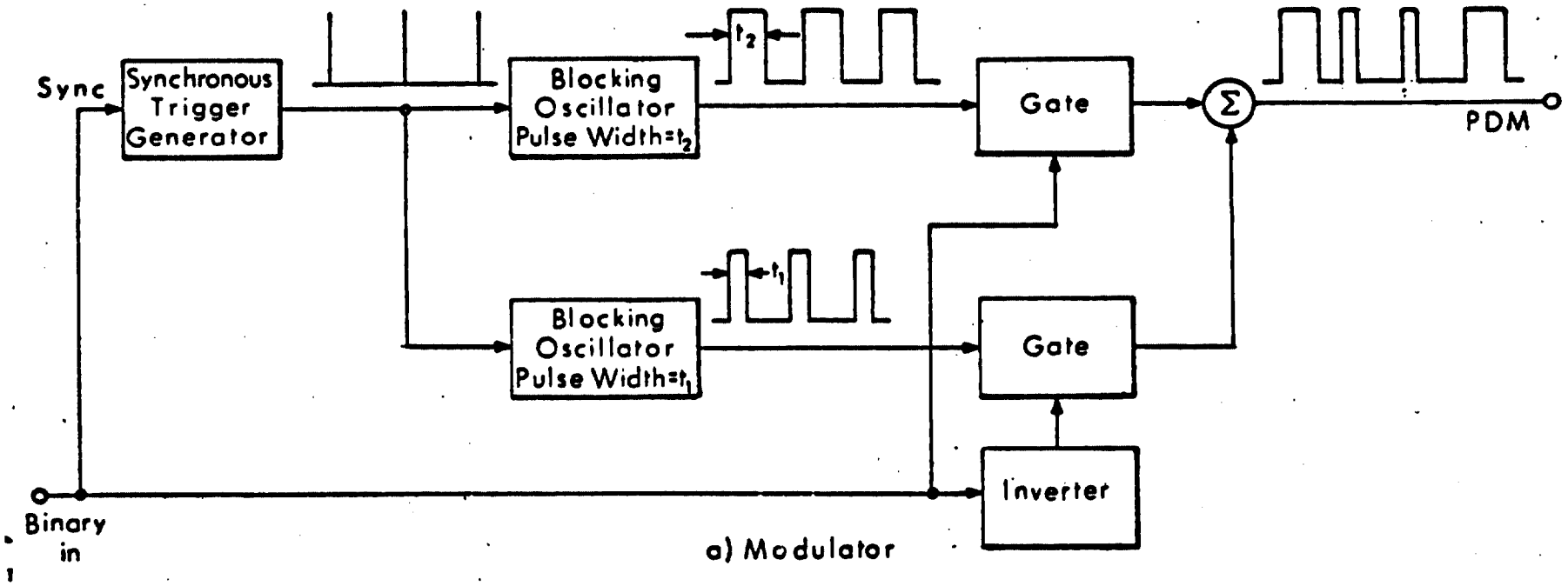
(b) PDM



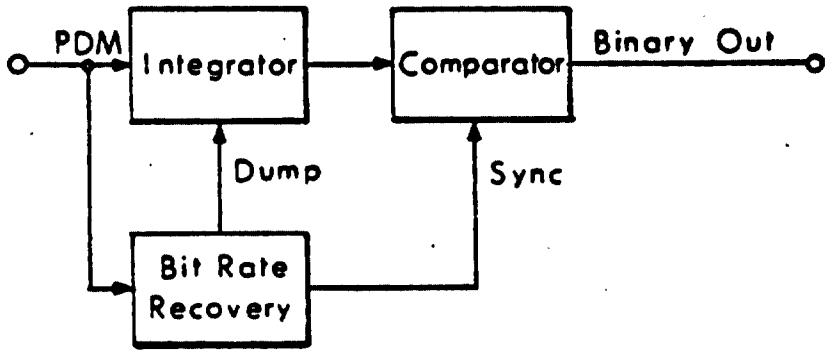
(c) PPM

R-3654
S-737

Fig. 4.19 Alternate Pulse Modulation Schemes--Analog Methods



a) Modulator



A-39K

b) Detector

Fig. 4.20 Generation and Detection of Pulse Duration Modulation

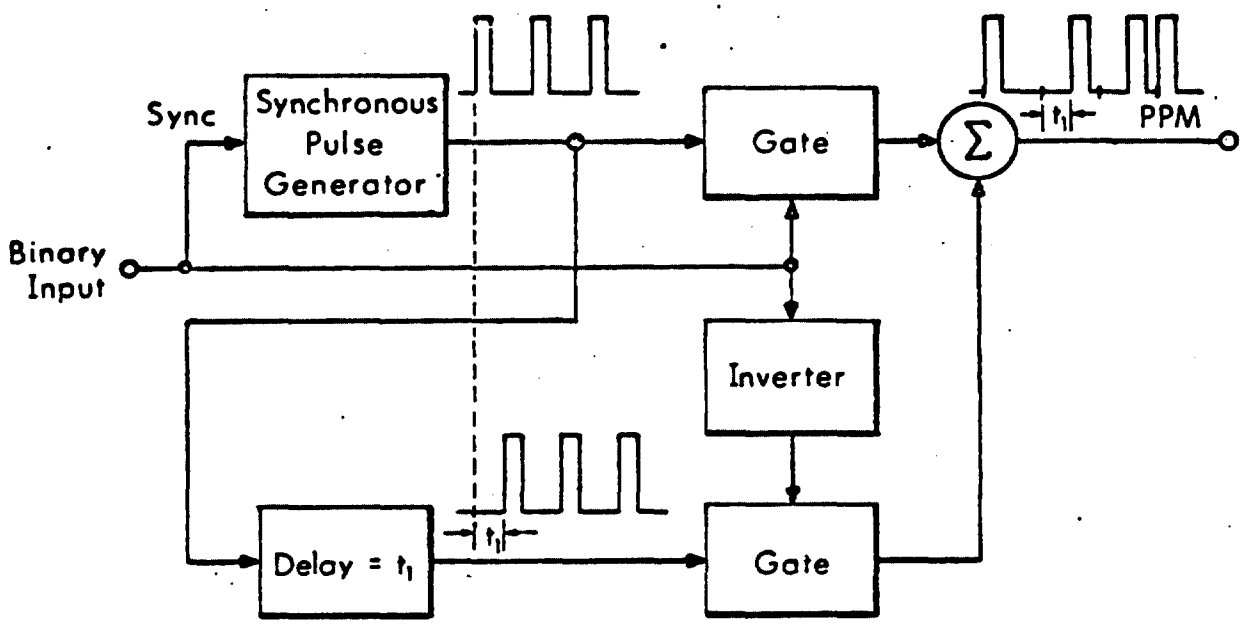
As in PAM, the PDM signal may be generated by analog means, with similar cost. A close tolerance is not maintained on the pulse widths. A block diagram of this method of PDM generation is given in Fig.4.19(b). The synchronous trigger pulses are generated as in the method of Fig.4.20(a), but are used to trigger a variable width generator such as a phantastron circuit. The binary modulation data is used to control the width. The phantastron or other variable width generator output is then a pulse-duration modulated signal.

A block diagram of the PDM detector is given in Fig.4.20(b). The pulse-duration modulated signal is then fed to an integrator. The bit rate is recovered and used to start and stop the integration at the correct time. The wider pulse will result in a larger charge on the integrator than will the narrower pulse. The integrator output is tested by a comparator that displays a "one" if the integrated output exceeds a certain threshold, and a "zero" otherwise. The comparator is also synchronized by the bit rate recovery unit to test the integrated output at the proper time.

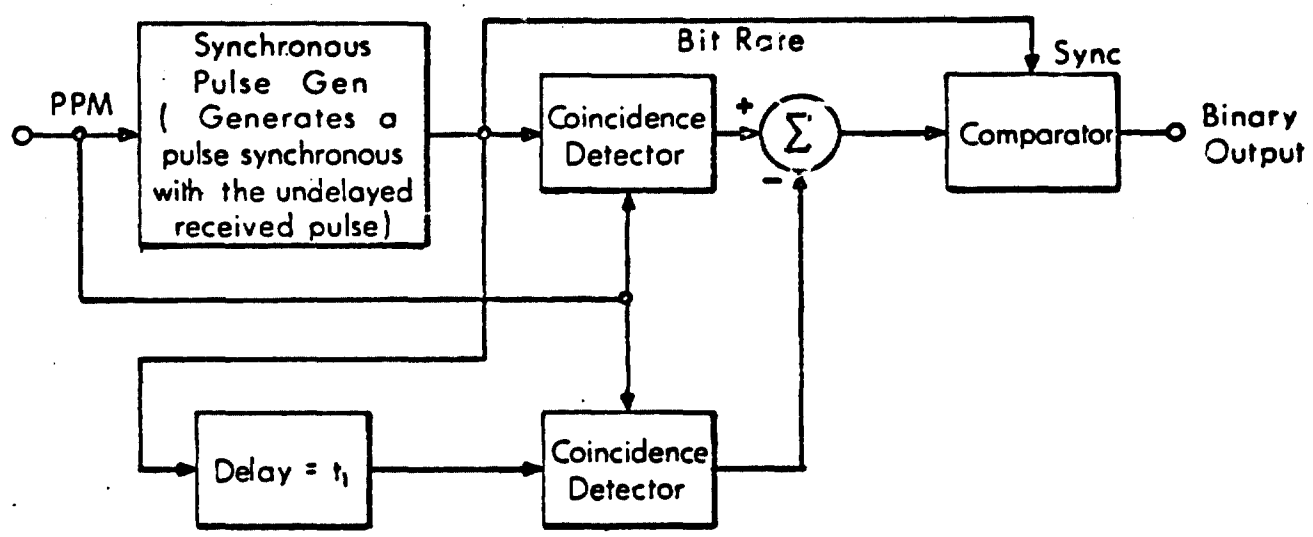
4.1.5.3 Pulse Position Modulation

Pulse Position Modulation (PPM), the third of the series of pulse modulation schemes, utilizes the relative position of each pulse in the train to convey the information. The pulses are of constant amplitude and fixed width, and contain variations only in separation, or position of the pulses. A typical PPM waveform is given in Fig. 4.16c.

Figure 4.21a gives one method of generating the PPM signal. The synchronous pulse generator generates a pulse train as in the other two pulse modulation techniques. A delay line gives a fixed delay of t_1 to the pulse. A sample of the undelayed pulse train as well as the delayed train is retained. The undelayed train is gated by one of the binary units (usually the mark) while the delayed pulse is gated by the other. The sum of the two gated outputs forms the PPM signal.



a) Modulator



b) Detector

R-3636
5-735

Fig. 4.21 Generation and Detection of Pulse Position Modulation

There is a method for PPM that is analogous to the "analog" method of generation of PAM and PDM. It makes use of a voltage controlled delay line, which, while not beyond the state of the art, does present some problems in physical realization. As it is neither better nor simpler than the method of Fig. 4.21a, it is presented here for completeness only. A block diagram is given in Fig. 4.19c. The synchronous pulse generator output is fed to the voltage controlled delay line whose delay is controlled by the modulating signal. The output is the PPM signal.

Detection is accomplished by the method of Fig. 4.21b. The PPM signal is then used to synchronize a pulse generator that generates a pulse train coincident with the position of the undelayed received pulse. Its output is then delayed by t_1 in a delay line. The delay line output is then a pulse train coincident with the position of the delayed received pulse. These two locally generated replicas of the two possibilities of the anticipated received signal are compared with the actual received signal in two coincidence detectors. The coincidence detector outputs are zero unless both inputs are present simultaneously. These two outputs are subtracted and the difference is compared with a threshold (say zero volts) in the comparator. The synchronous pulse generator output contains the necessary bit rate information so it is not necessary to recover the bit rate separately. This is, as before, used to synchronize the comparator output.

4.1.5.4 Spectra of PDM and PPM

The spectrum of a PDM signal is determined by the width of the shortest pulse in the sequence. The spectrum of a PPM signal is determined by the shortest spacing between pulses or the pulse width, whichever is shorter. Usually the pulse width will be shorter than any of the spacings so this will normally be the determining factor. The spectral density plot of both PDM and PPM given in Fig. 4.12 can therefore be applied to estimate the bandwidth occupancy with appropriate interpretation of the parameter T in each case.

4.2 Time Marker Representation by a Maximal-Length Linear Sequence

It was pointed out in Sec. 2.2 that the time markers – particularly the main frame marker – are the most essential parts of a time code. In essence, a time marker is a localized occurrence in time that is intended to define a time instant sharply. In existing time code systems, the "occurrence" is an edge of a rectangular pulse. The more vertical such an edge is the less sensitive will its position be to instantaneous values of additive disturbances. But a perfectly vertical edge would strictly require an infinite transmission and receiver bandwidth. Therefore, actual pulse edges rise gradually and their positions are sensitive to the instantaneous fluctuations of additive noises. Specifically, the magnitude of the slope of a pulse edge is directly proportional to the bandwidth occupied by the pulse. For relatively weak additive gaussian noise, the rms jitter caused by the noise in the position of an edge with a finite slope is inversely proportional to the magnitude of this slope.

An alternate approach to the realization of a precise marker is to employ a waveform that occupies the entire time interval allotted for the extraction of the time marker and possesses a high, sharply peaked autocorrelation function; i.e., an autocorrelation function that drops rapidly in value as a function of deviation from zero shift. In this way, the process of detecting the marker at the receiver will be a cross-correlation operation that involves a "smoothing" filtering operation on the noise over the full time interval, thus ensuring maximum smoothing out of the effects of the attendant noise. In order to establish what waveforms would provide the desired autocorrelation properties we shall consider a few suggestive examples.

Consider first a single pulse, repeated at intervals equal to the frame length. The autocorrelation function of a rectangular pulse of width Δt is a triangular pulse of width $2\Delta t$ at the base. Evidently, for good resolution, the "slimmer and taller" the pulse, the better. But peak power limitations on equipment place a bound on how high the pulse can be. Moreover, a limitation

on how slim the pulse can be is placed by the requirement of reasonably long integration times to smooth out the noise (or reduce its contribution to the output of the cross-correlator) and enhance the collected signal energy. The combined effect is a limitation on the amount of energy that can be packed into the pulse, or on how high the peak of the autocorrelation function can be relative to the background noise.

Increased peak of correlation (or packed energy) with a shimmer correlation pulse can be secured by using a group of separate short pulses and integrating the contributions of the successive pulses in the group. However, if the spacing between successive pulses is sufficient to prevent false lock on a secondary peak, the resultant peak factor will in most practical cases be so high that the available energy will still be severely limited.

The combination of the property of high signal energy that is characteristic of a continuous waveform with an autocorrelation function that is similar to that of a pulse train can be achieved by using a pseudorandom waveform to provide the desired marker waveform. For this reason, pseudorandom waveforms will be singled out for special consideration in the present section. A particularly suitable type of signal with the desired properties can be generated from maximal-length linear sequences.

4.2.1 Correlation Properties of Linear, Recurring, Maximal-Length Sequences

The autocorrelation function R(k) of a linear, recurring, maximal-length sequence is characterized by the following two general properties:

(a) R(k) is a discrete (or point) function of the discrete variable k that represents the number of digits through which the shift is made.

(b) R(k) is a two-level function, possessing one value for $k = 0$ and another for all $k \neq 0$.

Quantitatively, a binary sequence can be expressed as a succession of selections from two possible levels, as in Fig. 4.22. The exact values of

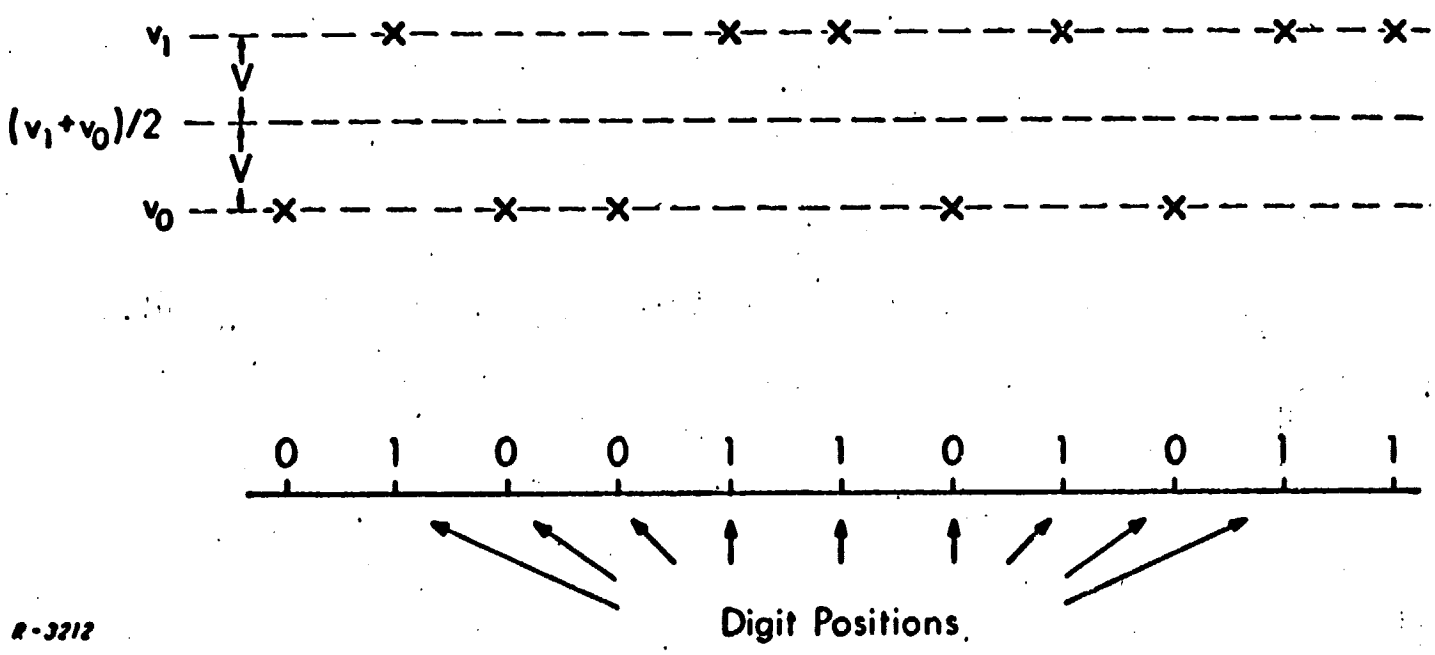


Fig. 4.22 Quantitative Representation of a Binary Sequence as a Succession of Selections From Two Possible Levels, v_o for 0 and v_1 for 1.

R(k) for $k = 0$ and $k \neq 0$ then depend upon the relative quantitative significance of the two possible states in the sequence. In general, the binary symbols 0 and 1 may be conveyed by the constant physical quantities v_o and v_1 respectively. The most important special cases are the ON-OFF situation in which v_o is the zero quantity and v_1 is some nonzero quantity; and the Phase-Reversal situation in which $v_o = -V$ and $v_1 = +V$, $V \neq 0$. We shall derive the expression for the autocorrelation function of any general recurring binary sequence of v_o's and v_1's, and apply the result to the case of a maximal-length, linear, recurring sequence of v_o's and v_1's and its ON-OFF and Phase-Reversal special cases.

With reference to Fig. 4.22, we first note that the levels v_o and v_1 of the sequence can be expressed as

$$v_1 = \frac{v_1 + v_o}{2} + V$$

and

$$v_o = \frac{v_1 + v_o}{2} - V$$

where

$$V = \frac{v_1 - v_o}{2} \qquad (4.1)$$

In this way, the sequence of v_o and v_1 levels can be expressed as a sequence of subtractions and additions of the quantity, V, from or to the mean level $(v_1 + v_o)/2$. Therefore, the sequence of v_o and v_1 levels is equivalent quantitatively to the sum of a sequence in which the same level $(v_1 + v_o)/2$ appears at every digit position plus a sequence of -V's and +V's corresponding to the v_o's and v_1's, respectively.

Thus, we have for the autocorrelation function, R(k), of a recurring sequence of v_o's and v_1's, of fundamental length p,

$$R(k) = \sum_{n=1}^{p} \left[\left(\frac{v_1 + v_o}{2} \right) + \alpha_n \right] \left[\left(\frac{v_1 + v_o}{2} \right) + \alpha_{n+k} \right] \qquad (4.2)$$

where the α_n's represent the "alternating" +V and -V component of the sequence. Expansion of the product under the summation sign yields

$$R(k) = \sum_{n=1}^{p} \left(\frac{v_1 + v_o}{2} \right)^2 + \sum_{n=1}^{p} \alpha_n \alpha_{n+k}$$

$$+ \left(\frac{v_1 + v_o}{2} \right) \left[\sum_{n=1}^{p} \alpha_n + \sum_{n=1}^{p} \alpha_{n+k} \right] \qquad (4.3)$$

The first sum on the right is

$$\sum_{n=1}^{p} \left(\frac{v_1 + v_o}{2} \right)^2 = p \left(\frac{v_1 + v_o}{2} \right)^2$$

The second sum on the right, in Eq. (4.3), is the autocorrelation function $R_{\alpha\alpha}(k)$ of the "alternating" component sequence. For any recurring sequence of fundamental length, p,

$$R_{\alpha\alpha}(k) = \sum_{n=1}^{p} \alpha_n^2 \quad , \quad k = 0, p, 2p, \ldots$$

$$= pV^2 \quad , \quad k = 0, p, 2p, \ldots \qquad (4.4)$$

and

$$R_{\alpha\alpha}(k) = \sum_{n=1}^{p} \alpha_n \alpha_{n+k}$$

$$= [A(k) - D(k)] V^2 \qquad (4.5)$$

where we have used the notation A(k) and D(k) for the numbers of agreements and disagreements among the α_n and α_{n+k} sequences of +V and -V. In general, $R_{\alpha\alpha}(k)$ may assume different values for different k. But from the Correlation Property of maximal-length, linear recurring sequences, we have

$$A(k) - D(k) = -1 \quad \text{for all integer } k \neq 0, p, 2p, \ldots \tag{4.6}$$

which signifies that $R_{\alpha\alpha}(k)$ assumes only one value for all integer $k \neq 0$ or any integer multiple of p.

Finally the third group of terms on the right in Eq. (4.3) is

$$\left(\frac{v_1 + v_2}{2}\right) \left[\sum_{n=1}^{p} \alpha_n + \sum_{n=1}^{p} \alpha_{n+k}\right] = (v_1 + v_2) \sum_{n=1}^{p} \alpha_n$$

since each sum ranges over one complete period of the same recurring sequence. Moreover, for a sequence of +V's and -V's,

$$\sum_{n=1}^{p} \alpha_n = V[N_p(+V) - N_p(-V)] \tag{4.7}$$

where

$N_p(+V)$ = Number of +V's in a period of p positions

and

$N_p(-V)$ = Number of -V's in a period of p positions

But maximal-length, linear, recurring sequence possesses a Subpattern-Occurrences Property that states that there are $(p + 1)/2$ 1's (here +V's) and $(p - 1)/2$ 0's (here -V's), in a fundamental period of p positions. Consequently, for such a sequence,

$$N_p(+V) - N_p(-V) = 1 \tag{4.8}$$

and

$$\sum_{n=1}^{p} \alpha_n = V \tag{4.9}$$

Substitution of the above results into Eq. (4.3) yields for any general recurring sequence of v_o's and v_1's having a fundamental period length of p digits,

$$R(k) = p\left(\frac{v_1 + v_o}{2}\right)^2 + [A(k) - D(k)]\left(\frac{v_1 - v_o}{2}\right)^2 + \frac{v_1^2 - v_o^2}{2}[N_p(v_1) - N_p(v_o)] \quad (4.10)$$

For a maximal-length, linear recurring sequence, we make use of Eqs. (4.4), (4.6) and (4.7) in conjunction with Eq. (4.10) to obtain for $k = 0, p, 2p, \ldots$

$$R(k) = p\left(\frac{v_1 + v_o}{2}\right)^2 + p\left(\frac{v_1 - v_o}{2}\right)^2 + \frac{v_1^2 - v_o^2}{2}$$

$$= \left(\frac{p+1}{2}\right)v_1^2 + \left(\frac{p-1}{2}\right)v_o^2 \quad (4.11)$$

and, for $k \neq 0$ or any integer multiple of p,

$$R(k) = p\left(\frac{v_1 + v_o}{2}\right)^2 - \left(\frac{v_1 - v_o}{2}\right)^2 + \frac{v_1^2 - v_o^2}{2}$$

$$= \left(\frac{p+1}{4}\right)v_1^2 + \left(\frac{p-3}{4}\right)v_o^2 + \left(\frac{p+1}{2}\right)v_1 v_o \quad (4.12)$$

The simplified result in Eq. (4.11) follows also from the fact that there are $(p + 1)/2$ v_1's and $(p - 1)/2$ v_o's in one fundamental period of a maximal-length sequence. The R(k) expressions for ON-OFF ($v_o = 0$, $v_1 = V$) and Phase-Reversal ($v_o = -V$, $v_1 = V$) maximal-length, linear sequences are given in Table 4.2.

Table 4.2

Correlation Functions of Maximal-Length, Linear Recurring ON-OFF and Phase-Reversal Sequences

	R(k) for 0 or V Sequence	R(k) for -V or +V Sequence
$k = 0, p, 2p, \ldots$	$(p+1)V^2/2$	pV^2
All integer $k \neq 0, p, 2p, \ldots$	$(p+1)V^2/4$	$-V^2$

The autocorrelation coefficient of a recurring binary sequence of fundamental period p digits is

$$\rho(k) = R(k)/R(0) \tag{4.13}$$

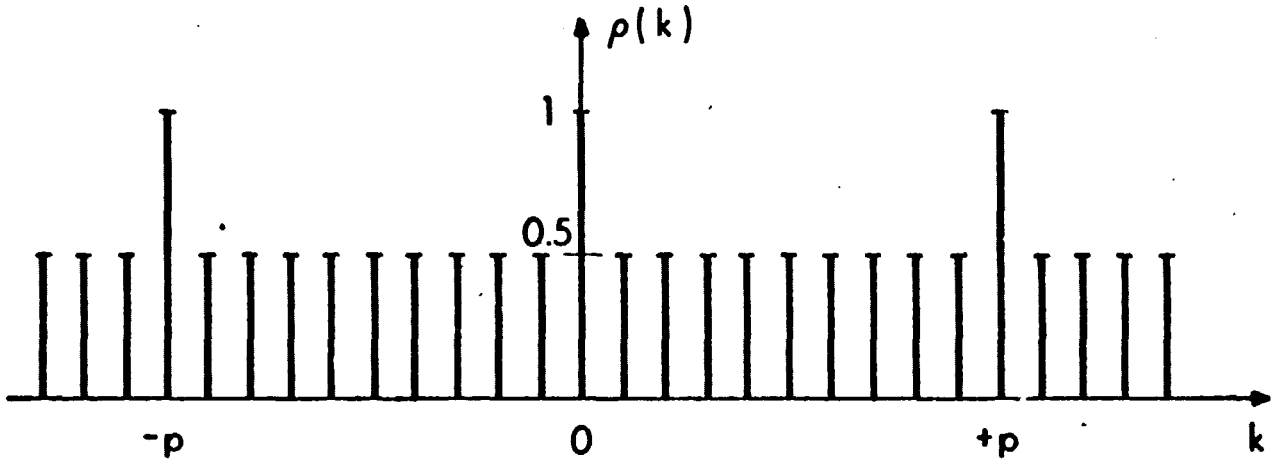
If we apply this to an ON-OFF maximal-length, linear sequence, we obtain, with the aid of Table 4.2,

$$\begin{aligned} \rho(k) &= 1 \quad \text{for } k = 0, p, 2p, \ldots \\ &= 1/2 \quad \text{for all integer } k \neq 0, p, 2p, \ldots \end{aligned} \tag{4.14}$$

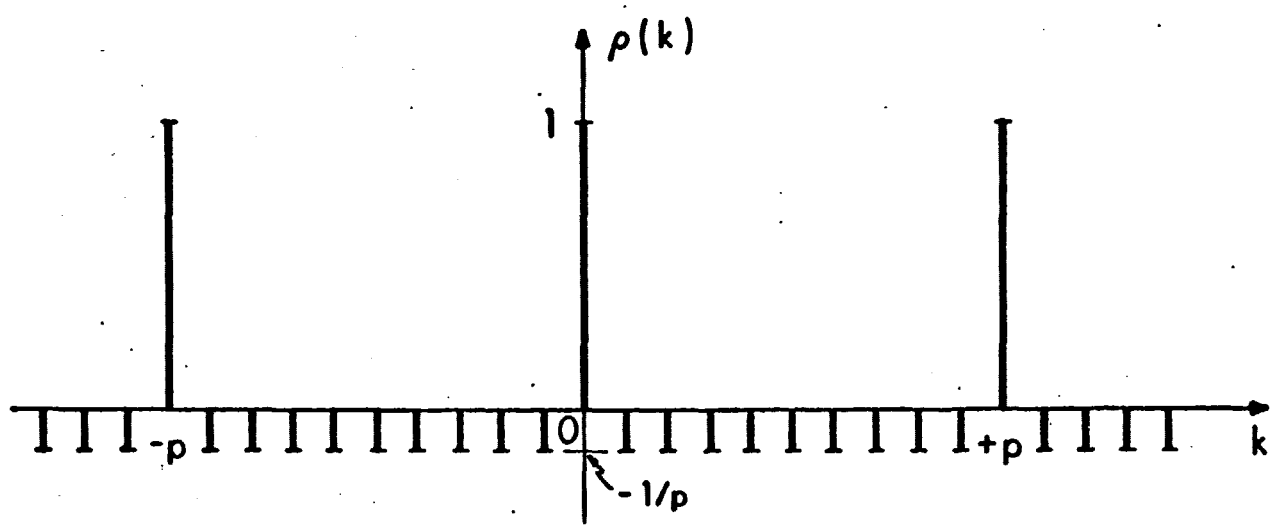
Similarly, for a Phase-Reversal maximal-length, linear sequence, we have

$$\begin{aligned} \rho(k) &= 1 \quad \text{for } k = 0, p, 2p, \ldots \\ &= -1/p \quad \text{for all integer } k \neq 0, p, 2p, \ldots \end{aligned} \tag{4.15}$$

The correlation coefficients given by Eqs. (4.14) and (4.15) are illustrated in Fig. 4.23. A sequence having the correlation coefficient defined by Eq. (4.15) and illustrated in Fig. 4.23(b) is often called a pseudo-noise sequence.



(a) $\rho(k)$ for ON-OFF Sequence



(b) $\rho(k)$ for Phase-Reversal Sequence

R-3213

Fig. 4.23 Correlation Coefficients for ON-OFF and Phase-Reversal Maximal-Length, Linear Recurring Sequences with p-digit Periods

4.2.2 Application of PR Sequences to Time Marker Representation

The advantage of using a phase-reversal representation of a binary, maximal-length, linear sequence is evident from Fig. 4.23. In application to time code marker representation, a time interval allotted to the establishment of a time marker is subdivided into subintervals of minimum width equal to the minimum pulse width (or no-pulse width) acceptable (e.g. the minimum in current use in NASA time code). Each subinterval is then filled with a rectangular pulse of positive or negative polarity according to the requirements of the desired pseudorandom sequence.

If the desired pseudorandom sequence is made to phase-reversal modulate a coherently related clock frequency, then part or all of the frame interval can be employed for the sequence that determines the main frame marker, and the coherent clock frequency can be recovered by means of a clock PLL aided by the local sequence generator. The zero crossings of the recovered clock can be used to subdivide the time frame in the desired manner.

The advantages of using pseudorandom sequences for time markers can be attributed to:

(a) Use of correlation detection with significant integration time for smoothing out the noise effects;

(b) An extremely advantageous autocorrelation shape for the marker carrying waveform;

(c) Full occupancy of the available time interval by the signal waveform, which maximizes the energy packed into the marker;

(d) The ability to multiplex the desired time label information by time division, frequency division or code multiplex;

and

(e) Simplicity of pseudorandom sequence generators.

An important practical consideration affecting the use of pseudorandom sequences is the time required to effect acquisition. The type of autocorrelation function illustrated in Fig. 4.23(b) has the disadvantage of not providing any indication during search process of which way or how far to shift the local replica of the delayed event in order to bring about the desired time coincidence. It is impossible to learn anything from each trial except that exact correlation has or has not been achieved. Consequently, in search of a systematic acquisition procedure two straightforward approaches may be considered. In the first, a number of trials equal to the total number of digits, p, in the code are performed, each involving integration over a fraction of the total length of the code. The code shifts that yield the highest indications of correlation are then selected for further trials (possibly over longer fractions of the code). The acquisition time will then exceed or approximately equal p times the integration time per trial. In the second approach, trials are performed, each involving integration over a fraction of the code length sufficient to provide decisive indication of correlation. In this the average acquisition time is given by p/2 times the integration time per trial. If the correlation properties of the code are to be fully realized, averaging in each trial must occur over the entire length of the code. But if the available signal energy relative to the background noise density is sufficiently high, reductions in the integration time may be allowed, resulting in a proportionate reduction of the acquisition time.

The aforementioned methods may require p trial correlations, while in principle, the fundamentals of maximal length sequences show that the sequence carries only $\log_2 p \approx N$ bits of information and hence should, in the absence of noise, be completely determinable from one trial inspection of N digits. This, however, represents an extreme lower bound and is unrealistic because it assumes noise-free conditions.

When the signal energy relative to the background noise density is sufficiently high, the number of trials can be significantly reduced if the code is

made up of several short subcodes that can be acquired individually. Several methods exist by which short sequences can be combined into a long sequence whose period is the product of the periods of the subsequences. If the lengths of periods of the subsequences are relatively prime (i.e., have no common divisors except ±1), they can be combined linearly or nonlinearly (modulo 2 combinations of sequences of zeros and ones) to produce the required long sequence. For example, three sequences (subcodes) with relatively prime periods p_a, p_b, p_c can be combined to form a new sequence with period $p_w = p_a p_b p_c$, which can be acquired in at most $\sum_{i=a}^{c} p_i$ trials by direct correlation with each of the subsequences separately. This technique, however, has the drawback that while lining up each component subsequence, the full energy of the incoming signal is not employed, since the maximum height of the cross-correlation function between the combined sequence and one of its components is less than the height p of the autocorrelation function of the combined sequence. Thus, detectability is traded for speed of acquisition when such combined-sequence techniques are used.

The acquisition time for each of the components depends on the number of elements over which the average is taken during each trial correlation. If this number is small, there may be considerable self-noise from the code appearing along with the random receiver noise. Correlation would have to be performed over the entire code period p_w to eliminate self-noise and to achieve the idealized correlation function of the code. As the length of the averaging interval is decreased, the acquisition time will decrease, but a point will be reached when the increased noise bandwidth and increased code self-noise begin to cause acquisition indications.

4.3 Recording Requirements

It is frequently desired to record coded waveforms on magnetic tape. It is therefore of interest to examine the effects of recorder characteristics on the various waveforms discussed in the preceding sections; in particular, sensitivity to recording noise, recording dropout effects, and spectral effects.

4.3.1 Sensitivity to Recording Noise

Empirical data has shown that most of the noise in a magnetic tape recorder originates in the recorder electronics. For example, in a particular high quality commercial recorder, ADCOM determined that the noise from the electronics alone, and the total system noise including noise from the recording head, differed by less than one db. The electronics is designed to compensate for the non-flatness of the amplitude vs. frequency response of the recording head. The electronic compensator response curve is therefore the inverse of the recording head response. A typical recording head response is given in Fig. 4.24 Figure 4.25 gives electronic compensator response provided by manufacturer, while Fig. 4.26 gives noise response measured by ADCOM of the above-mentioned recorder. The frequency response of the head, of course, changes with the tape speed. The response of the electronic compensator (and hence the noise response) is therefore switched any time the tape speed is changed. The curves of Figs. 4.24 through 4.26 are for a tape speed of 120 inches/sec. The shape of the curves remains the same for any other tape speed, changing only the abscissa by a scale factor.

It may be noted from Figs. 4.25 and 4.26 that the electronic compensator response and the noise response are high at both ends of the passband, and relatively low in the middle. Very significant improvement can therefore be

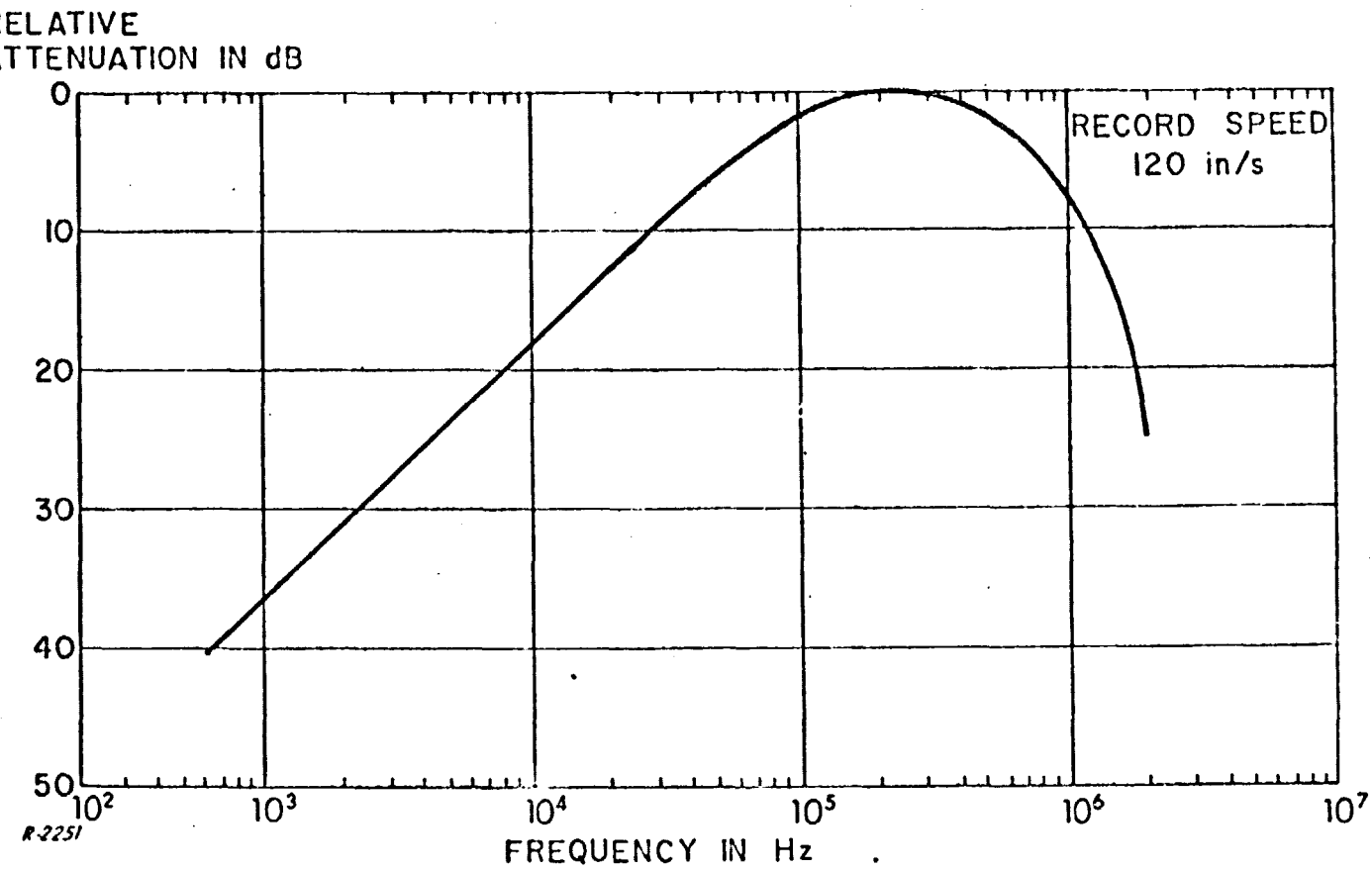


Fig. 4.24 Tape Recorder Head Characteristic

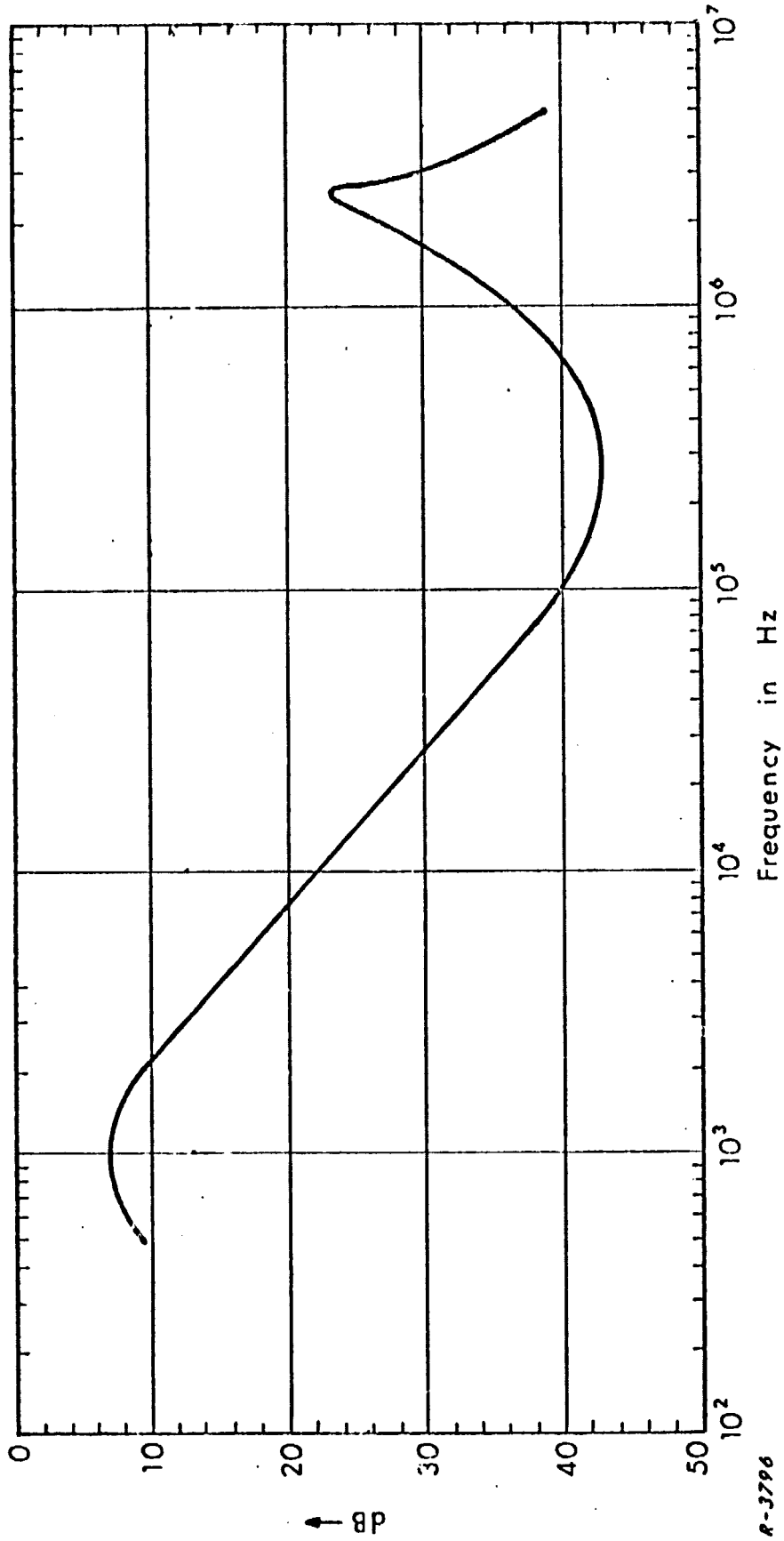


Fig 4.25 Tape Recorder Electronic Compensator Characteristic

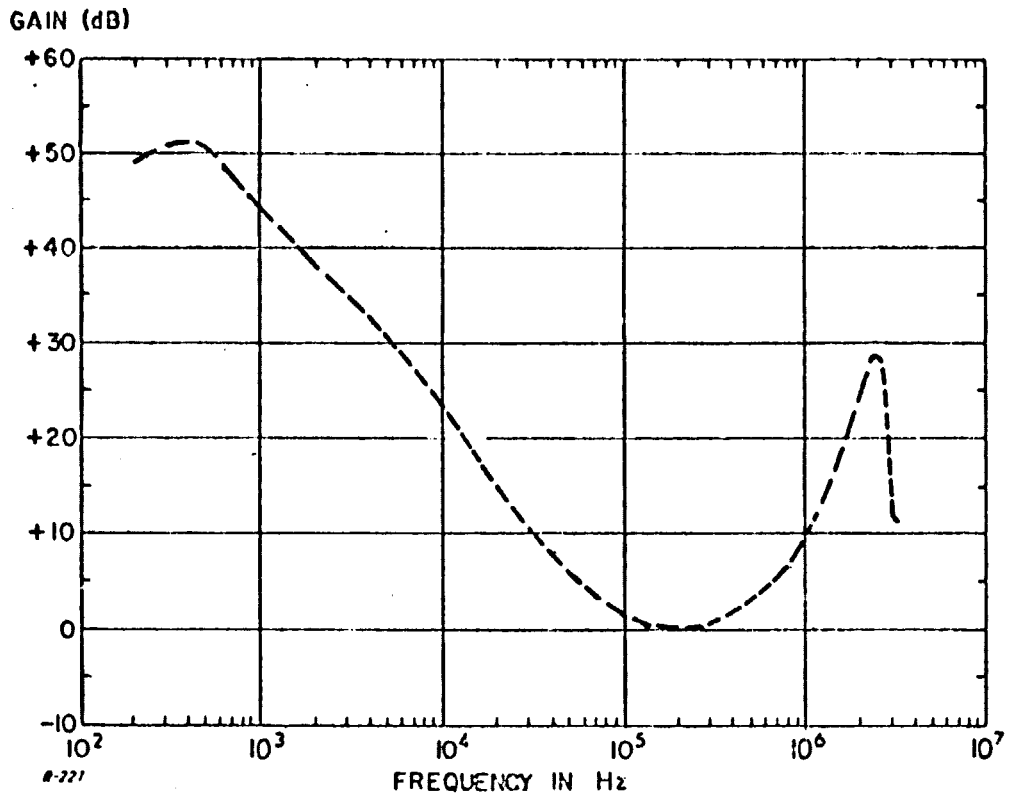


Fig. 4.26 Noise Spectrum of Recorder

obtained by narrowing the bandwidth from both ends. From Figs. 4.12 and 4.13 it is clear that the distortion caused by low-frequency cutoff will be less severe in the case of the split-phase and RZ-bipolar waveforms than with the other types. However, narrowing the bandwidth from the high frequency side will create less distortion in NRZ and the other full bit width waveforms than in the split-bit waveforms.

4.3.2 Recording Dropout Effects

Recording dropout is defined as the momentary loss of signal due to mechanical reasons or imperfections. Dropouts come from two sources - an imperfection in the tape (a spot with no magnetic coating), and from the tape moving slightly away from the recording head due to air currents, vibration, etc., as the tape moves over the head. With the state of the art as it is today in the manufacture of tapes, it may be said with certainty that a dropout due to tape imperfection will not be long enough to lose a single bit. However, dropouts due to separation of tape and head sufficiently long to cause loss of data have been experienced. The signal attenuation due to separation of the tape from the head is approximately 55 dB recorded wavelength. The recorded wavelength is the length of tape required to record one cycle of signal, and is given by

$$d = \lambda \frac{V}{c}$$

where

d = recorded wavelength

λ = wavelength in free space

V = tape speed

c = speed of light

For wideband recorders, recorded wavelenghts of as short as 0.1 mil may be encountered. (A mil is 10^{-3} inches.) Although today's recorders are designed to keep the tape continuously in touch with the head, clearly when the above dimensions are present, fluctuations will occur. In fact, a considerable

amplitude fluctuation will occur even under ideal conditions. The fact that the higher the bandwidth, the more severe the dropout favors the NRZ and other full bit width waveforms from the viewpoint ot tape dropout effects.

4.3.3 Spectral Effects

4.3.3.1 Low-Frequency Effects

The most important properties of the spectrum of a waveform when considering magnetic tape recording are its spectral density around 0 Hz and at high frequencies. The magnetic tape recorder does not have a frequency response extending to 0 Hz. As a result, frequency components in the waveform at or near dc will be lost in the recording process, with a resulting distortion in the recorded waveform. It is therefore desirable if possible to find a waveform with no dc content. Waveforms with no dc content are the split phase and the RZ-bipolar. Unfortunately, both of these have other undesirable properties.

The response of a tape recorder also is limited in its high frequency response or bandwidth by the speed of the tape across the head. Neglecting the recorder electronics, doubling the tape speed will in general double the frequency response of the recorder. For a given tape speed, then, the lowest bandwidth waveform may be recorded at the fastest bit rate for a fixed probability of error, or conversely, will have the lowest probability of error for a fixed bit rate. Unfortunately the only two waveforms with no dc content occupy approximately double the bandwidth of the others. On strictly a bandwidth basis, and neglecting dc content, one might suppose that NRZ could be recorded at about twice the bit rate of split phase for a fixed tape speed and error rate. Empirical data has shown this factor to be between 1.3 and 1.4 rather than 2, due primarily to the dc content of the NRZ or other full bit width waveform. This implies that the bandwidth factor still predominates, but not by as great a factor as might be supposed.

4.3.3.2 Phase Linearity

The requirements for good reproduction of a waveform are that not only the amplitude response but also the group delay be flat with frequency. If one frequency component of the waveform is delayed with respect to another in recording and playback, the waveform will be distorted. Since the recording head is not phase linear, any good recorder will have delay equalization circuits to compensate for this nonlinearity. These circuits have the inverse or complementary characteristics of the recording head, and are generally composed of all-pass lattice networks.

4.4 Immunity of Different Waveform Modulations to Various Degrading Factors

4.4.1 General

In this section we will compare three different types of time code waveform modulation in order to determine their relative immunity to three specific and serious sources of signal degradation. The three waveform types are shown in Fig. 4.27; observe that the waveforms in Fig. 4.27 are envelopes. These envelopes may contain a few cycles of some accurate reference carrier, which the user may extract and use to generate a continually verifiable fine time source for subdividing basic units of time to the necessary precision (as in the NASA and IRIG carrier level shift time codes). On the other hand these baseband signals may contain no carrier (as in the dc level shift versions of these NASA and IRIG codes). The three basic bit-encoding modulations PDM, PPM, and PAM, will be compared for their relative resistance to the following:

i) Jitter due to low-frequency cutoff. The primary effect to be studied is the disturbance in pulse edge location or in carrier zero-crossing location brought about by transients smearing over from the preceding pulse.

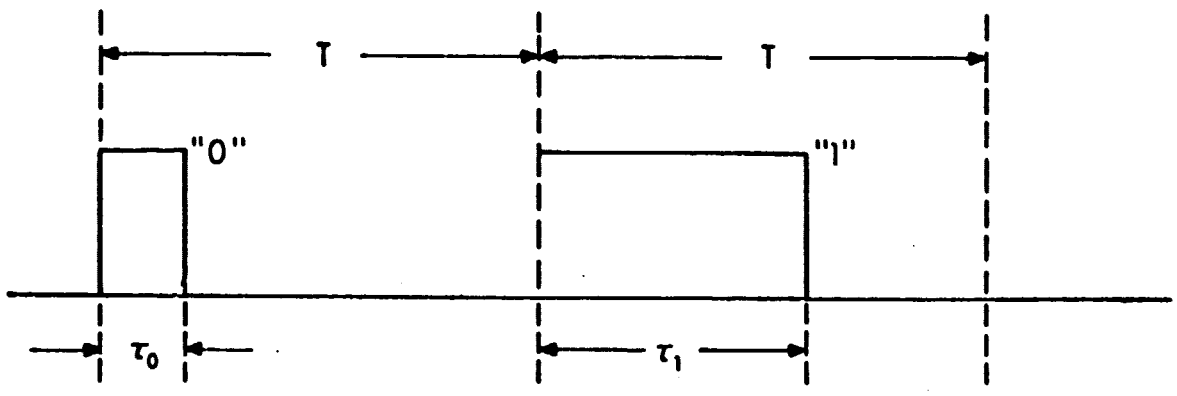
ii) Jitter due to Additive Random Noise. A random fluctuation in carrier zero-crossing location or pulse edge location without serious distortion of the basic waveform will be caused by additive random noise which is small compared to the signal waveform amplitude.

iii) Bit Errors Due to Additive Random Noise. During periods of transmission link outage, the signal-to-noise ratio may not be high enough to assure virtually error-free bit decoding. The three-bit modulation approaches may be compared to determine relative immunities to errors of this sort.

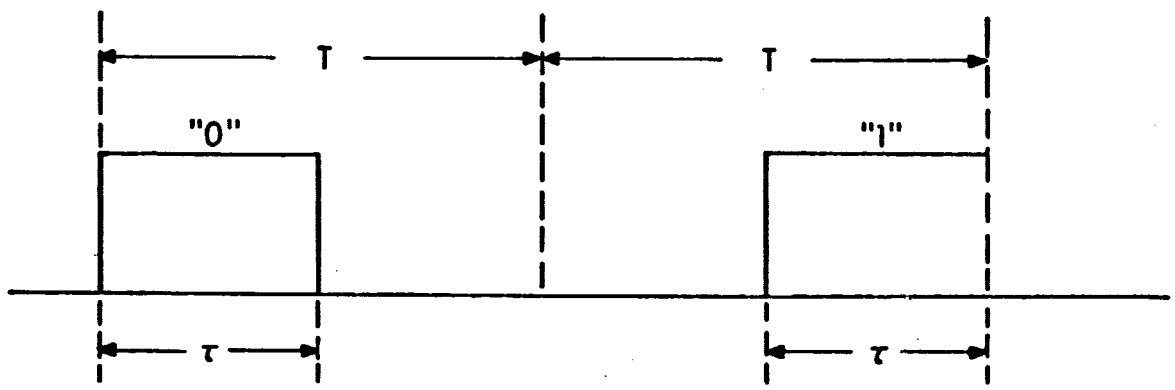
The basic analysis has already been performed for binary PDM in Ref. 1*. The analysis has here been extended to encompass the other two bit encoding modulations.

---

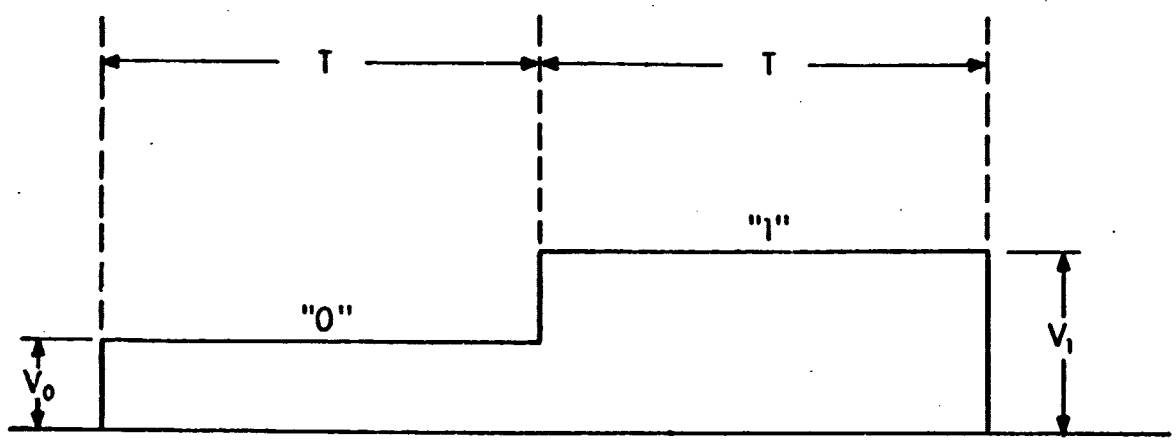
* Revised Draft, First Quarterly Progress Report, Time Code Study, Task VII of Command System Study for the Operation and Control of Unmanned Scientific Satellites.



(a) PDM (Pulse Duration Modulation)



(b) PPM (Pulse Position Modulation)



(c) PAM (Pulse Amplitude Modulation)

R-4178

Fig. 4.27 Various Forms of Baseband Waveform Modulation

4.4.2 Analysis

4.4.2.1 Relative Immunity to Bit Errors

In order to properly compare different waveform encodings for relative performance, it is necessary to establish common constraints among all the signals. Our comparison will be made with all signals constrained to have a maximum pulse signal power S. In Appendix B it is shown how the optimum detection of binary signals in the presence of additive 'white' gaussian noise of single-sided spectral density N_o watts/cycle of bandwidth is describable by a simple geometric representation. In this representation the two signals, 'mark' and 'space', are represented by points in two-dimensional space. The decoded 'point' is displaced from its true value in the figure by a random shift along the line between the two points. If this random shift, which is gaussian with zero mean and unit variance, should cross the decision boundary, an error will be made.

Figure 4.28 shows the geometric representations of binary PAM and binary antipodal signalling. For a given value of ST, the maximum allowable signal energy, antipodal signalling yields the lowest possible decision error probability (see Fig. 4.28a)

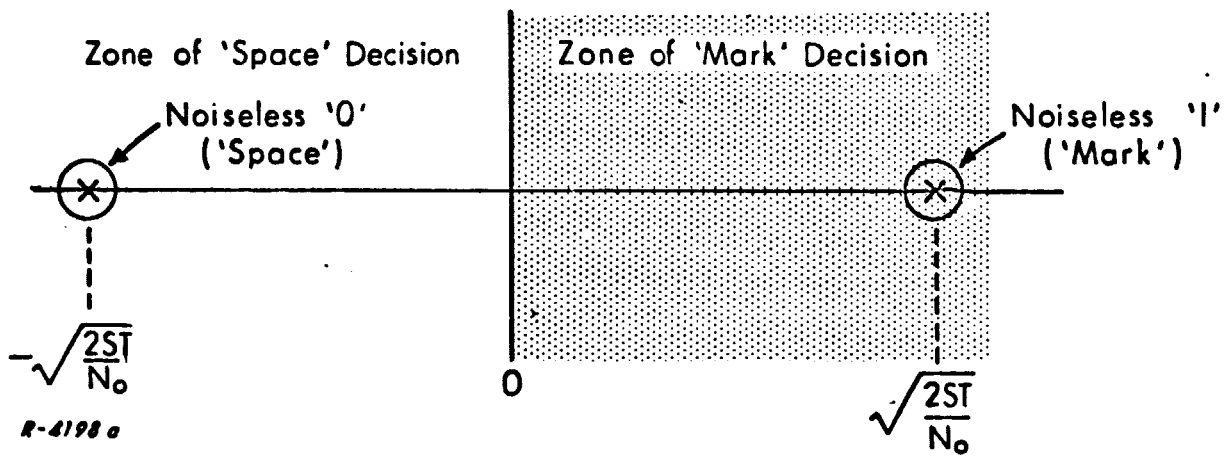
$$P_E = \Phi\left(\sqrt{\frac{2ST}{N_o}}\right) = \Phi(\sqrt{R_{max}}) \qquad (4.16)$$

where

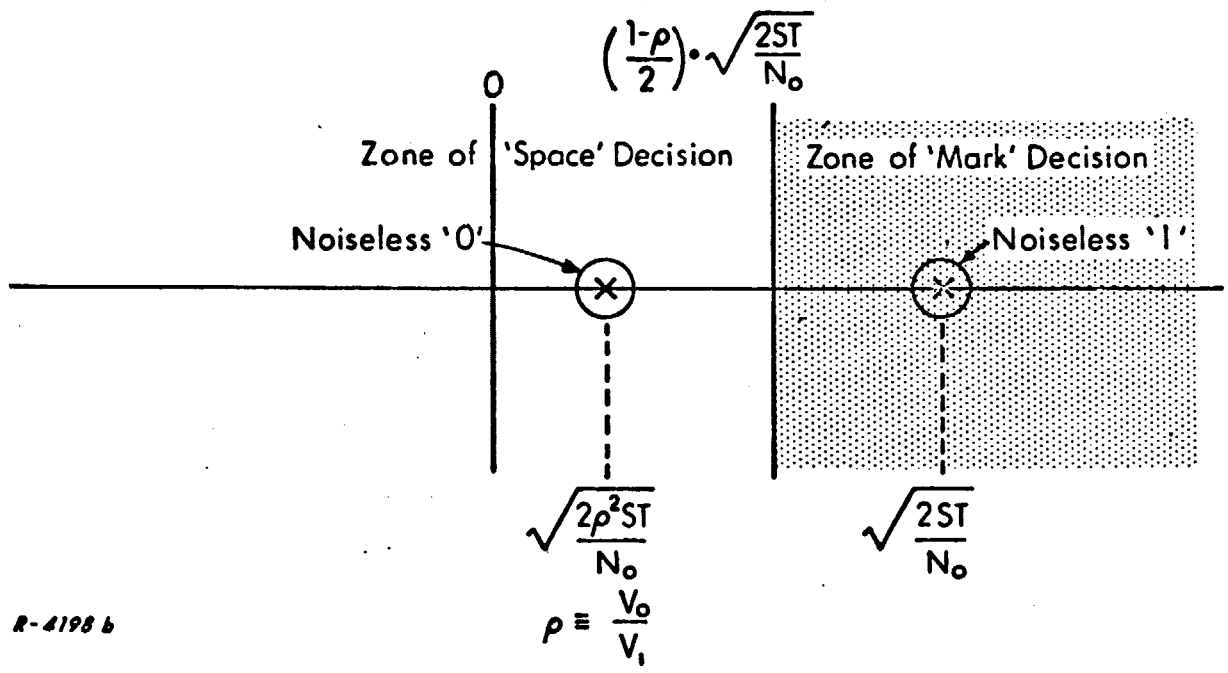
$$\Phi(x) \equiv \int_x^\infty \frac{1}{\sqrt{2\pi}} e^{-y^2/2} \, dy$$

The quantity $\Phi(x)$ is readily found in tables of the normal (gaussian) probability distribution.

For binary PAM (see Fig. 4.28b) the error probability must be expressed as a function of ρ, the voltage ratio of 'space' signal level to 'mark' signal level.



a) Geometric Representation of Decision Zones - Binary Antipodal Signals



b) Geometric Representation of Decision Zones - Binary PAM

Fig. 4.28 Geometric Representations of Binary PAM and Binary Antipodal Signaling

When $\rho = 0$, we have on-off keying as a special case of PAM. The error probability is:

$$P_E = \Phi\left[\left\{\frac{1-\rho}{2}\right\} \cdot \sqrt{\frac{2ST}{N_o}}\right] \qquad (4.17)$$

Comparison of P_E as given by Eq. (4.17) with P_E as given by Eq. (4.16) indicates that the decision error performance of binary PAM is exactly the same as that of binary antipodal signalling when the signal power of the latter is reduced by a factor of $(1-\rho)/2$. Thus, PAM is seen to be at a disadvantage of $-20 \cdot \log_{10}[(1-\rho)/2]$ dB with respect to optimum binary signalling.

In the appropriate picture for PPM, the points are placed upon orthogonal axes. Since any portion of the total bit interval T which the two signals share in common does not contribute to the making of a decision, Fig. 4.29a, which is the geometric representation of (b) is meant to apply for all $\tau > T/2$ will give performance equivalent to that obtainable when $\tau = T/2$.

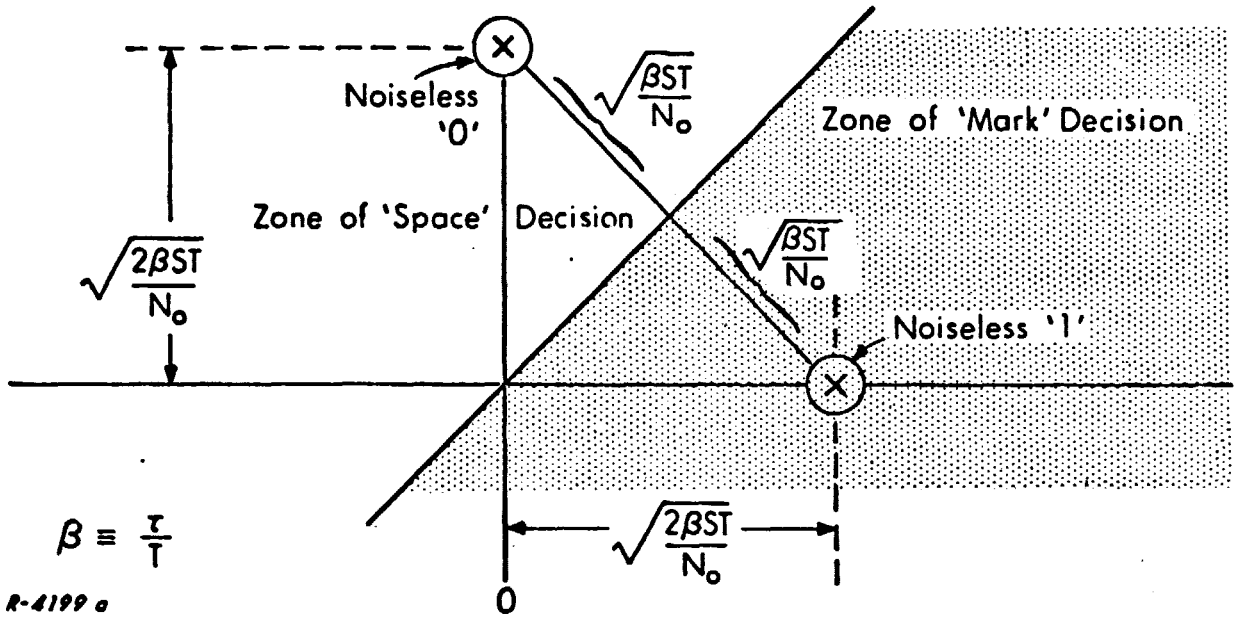
In terms of the parameter $\beta \equiv \tau/T$:

$$P_E = \Phi\left(\sqrt{\frac{\beta ST}{N_o}}\right) = \Phi\left[\sqrt{\frac{\beta}{2}} \cdot \sqrt{\frac{2ST}{N_o}}\right], \quad \beta \leq 1/2 \qquad (4.18)$$

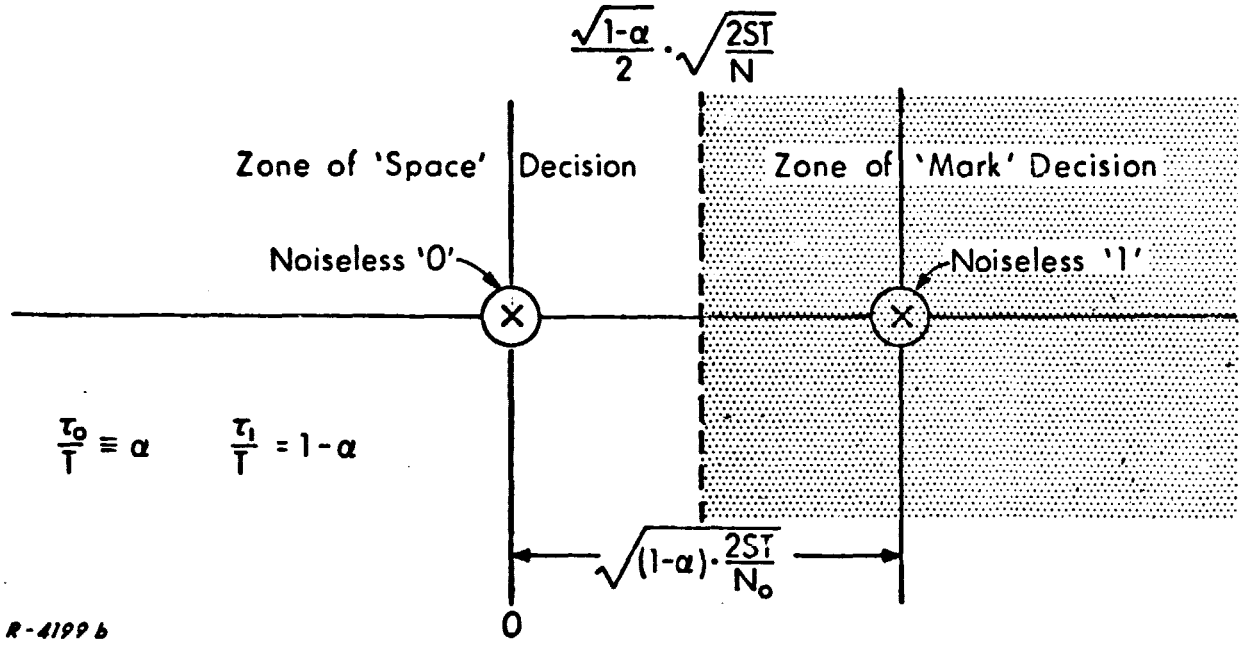
The degradation with respect to the optimum is seen to be $-10\log_{10}(\beta/2)$. In order to complete our comparison we must also introduce a bandwidth factor. Note that with PAM, pulses are of full duration T. On the other hand, the PPM pulses have relative duration β, and therefore the PPM waveform has relative bandwidth $1/\beta$.

In the analysis of PDM encoding, one must realize that the maximum detectability factor (simply the duration difference) for a given value of relative bandwidth is obtained when $T - \tau_1 = \tau_o$. Thus we have assumed this condition to hold in the PDM representation of Fig. 4.29b*.

* It is interesting to observe that this condition does not hold true in the NASA BCD time codes, where relative durations of 0.2T and 0.6T are used. Using durations of 0.2T and 0.8T would not affect bandwidth, but it would increase the detectability by a factor of (0.8 - 0.2)/(0.6 - 0.2) = 1.8 dB.



a) Geometric Representation of Decision Zones - Binary PPM ($\beta \leq 1/2$)



b) Geometric Representation of Decision Zones - Binary PDM

Fig. 4.29 Geometric Representations of Binary PPM and Binary PDM

It is seen that:

$$P_E = \Phi\left(\sqrt{\frac{1-\alpha}{2}} \cdot \sqrt{\frac{2ST}{N_o}}\right) \tag{4.19}$$

Thus the relative detectability of PDM compared to binary antipodal signalling is $-10\log_{10}[(1-\alpha)/4]$. The relative bandwidth factor is $\frac{1}{\alpha}$, $\alpha \leq \frac{1}{2}$.

Table 4.3 summarizes the relative signal detectability and bandwidth characteristics of these three waveform modulations. The standard for detectability is antipodal signalling, while the standard for bandwidth is a pulse of duration T (PAM).

It may be seen that none of the other bit modulation schemes can do any better than approach to within 6 dB of antipodal signalling performance. This can be accomplished with PAM only by going to the on-off keying extreme ($\rho = 0$). It can be achieved with PPM by letting β go to 1/2; however, twice the bandwidth of PAM is required. Using PDM, one can only get close to this -6 dB by making α very small, i.e., by making the 'space' pulse quite short. In the limit, as $\alpha \rightarrow 0$, PDM merges into on-off keying. For reasonable bandwidth occupancy factors, PDM must fall somewhat short of optimum. For example, if the relative bandwidth is 3, the relative detectability of PDM is -7.8 dB.

The curves of P_E vs. $2ST/N_o$ displayed in Fig. 4.30 graphically compare all the modulation schemes. The detectability, $2ST/N_o$, has the dimensions of a signal-to-noise ratio. It is in fact the value of peak available signal to rms noise level achievable by optimum 'matched' filtering in the antipodal signalling case. It may also be thought of as twice the ratio of the signal power S divided by the noise power contained in the band defined by the reciprocal of T.

It is to be observed in conclusion that the conclusions of this section are equally applicable to all these bit encoding techniques both in their dc forms and in their forms as pulses of reference carrier.

Table 4.3

COMPARISON OF RELATIVE DETECTABILITY AND BANDWIDTH CHARACTERISTICS OF PDM, PPM, AND ANTIPODAL SIGNALLING

Waveform Modulation	Relative Detectability, dB	Relative Bandwidth Occupancy
Antipodal Signalling	0 dB	Depends upon signal structure, e.g., = 1 for PSK = 2 for split phase
PAM $\rho = \frac{V_o}{V_1}$	$20\log_{10}[(1-\rho)/2]$ dB	1
PPM $\beta = \tau/T$	$10\log_{10}\beta/2, \quad \beta = 1/2$	$1/\beta$
PDM $\alpha = \tau_o/T$; $1-\alpha = \tau_1/T$	$10\log_{10}[(1-\alpha)/4]$	$1/\alpha$

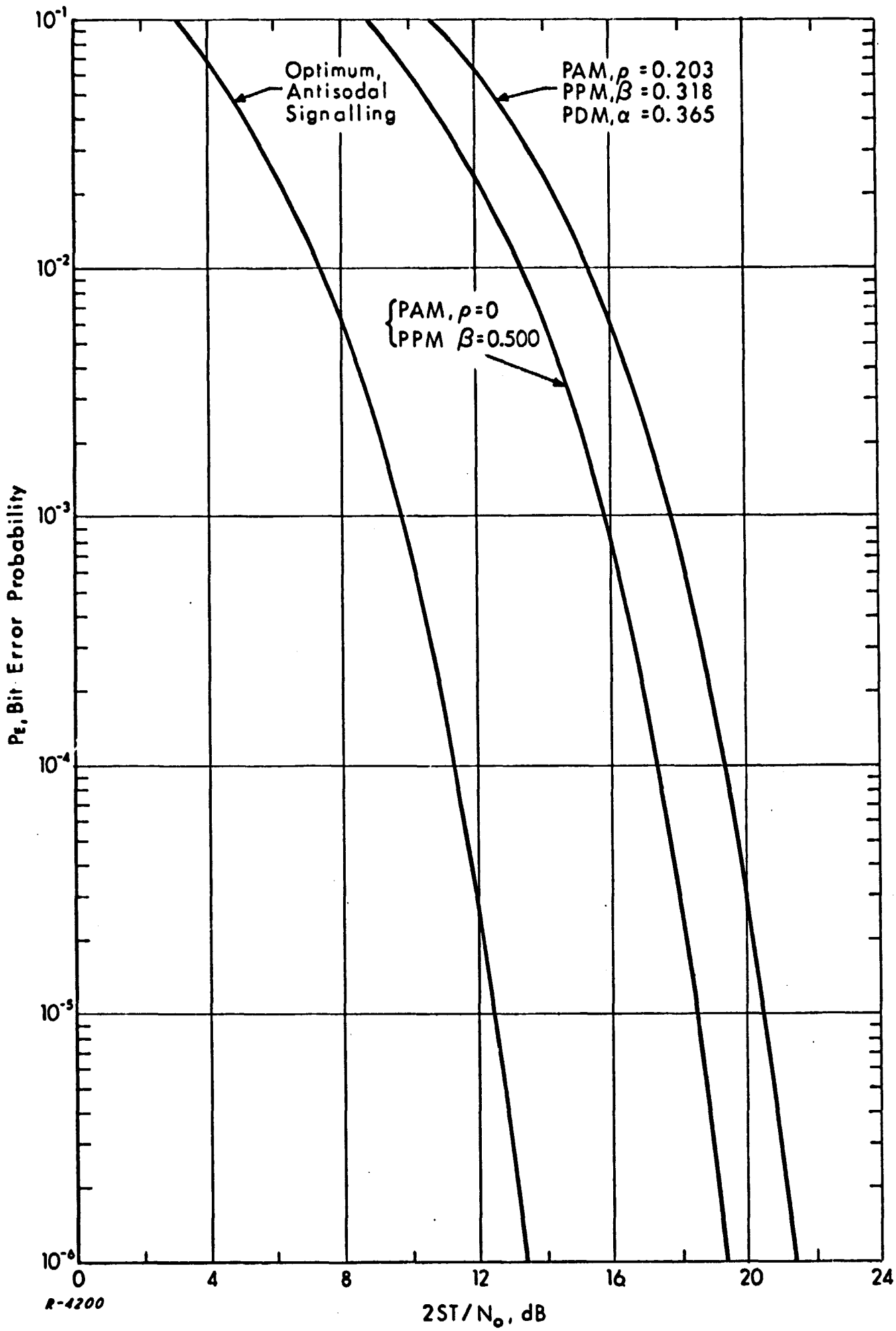


Fig. 4.30 Comparison of Decision Bit Error Probabilities for Different Bit Modulation Techniques

4.4.2.2 Relative Immunity to Jitter Caused by Additive Noise

The noise is assumed to be of small amplitude relative to the bit signal. Its effect will be to shift the locations of zero crossings of pulses of carrier, and to shift the locations of leading edges of dc pulses, thereby introducing a small 'jitter' into the waveform. The basic shape of the signal waveform will not be significantly altered.

In order to evaluate the jitter-producing effect of additive noise upon a zero crossing of a cycle of carrier within a time code waveform, consider that the carrier $\sqrt{2S}\ \sin \omega_c t$ (signal power = S) is approximated by:

$$\sqrt{2S}\ \sin \omega_c t \approx \sqrt{2S}\ \omega_c (t - t_n) \tag{4.20}$$

where t_n is the time of occurrence of the n^{th} zero crossing.

If at that moment in time the additive noise process has assumed the voltage value ν, the zero crossing will be slightly shifted in time to a location t' determined by:

$$\sqrt{2S} \cdot \omega_c (t' - t_n) + \nu = 0 \tag{4.21}$$

The time error $\Delta t \equiv (t' - t_n)$ introduced by the noise is:

$$\Delta t = \frac{-\nu}{\sqrt{2S} \cdot \omega_c} \tag{4.22}$$

Equation (4.22) establishes the important fact that the random variable of time shift is linearly related to the random noise, and therefore possesses a scaled version of the noise statistical distribution, i.e., if the noise is normally distributed the time shift is normally distributed. We may express the rms time jitter in terms of the ratio of carrier signal power to mean-squared noise power:

$$R \equiv \frac{S}{\overline{\nu^2}} \tag{4.23}$$

Thus,

$$\Delta t_{rms} = \frac{1}{\sqrt{2R} \cdot \omega_c} \tag{4.24}$$

We may similarly deal with dc pulses to obtain an expression for jitter. The pulse is assumed to be of amplitude $\sqrt{S}$. Observe that a pulse having absolutely vertical rising and falling edges cannot have these edge locations jittered by the presence of additive noise. We therefore model the pulse as having a linear leading edge with a 'risetime' of τ_r; i.e., $1/\tau_r$ is indicative of the highest frequency passed by the system.

The leading edge of the pulse has a slope of $\sqrt{S}/\tau_r$. Thus for a noise voltage of ν the time shift is:

$$\Delta t = -\frac{\nu \tau_r}{\sqrt{S}} \tag{4.25}$$

In terms of

$$R' \equiv \frac{S}{\nu^2} \tag{4.26}$$

we have:

$$\Delta t_{rms} = \tau_r / R' \tag{4.27}$$

In the preceding analysis no assumptions about the statistical properties of the noise have been made. If the noise is considered to be gaussian with zero mean (no dc component) then interesting probabilistic descriptions of the noise-induced jitter may be formulated. A gaussian random variable with zero mean is described by its rms value, the standard deviation σ. The probability that a random sample of such a variable will be greater than $k\sigma$ or less than $-k\sigma$ (i.e., fall outside of the $k\sigma$ points) is given by

$$\theta(k) = 2\Phi(k) = 2 \cdot \int_k^\infty \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \, dx \tag{4.28}$$

An excellent approximation to $\theta(k)$ is:

$$\theta(k) \approx \frac{\sqrt{2} \cdot e^{-k^2/2}}{\sqrt{\pi} \cdot k} \qquad k >> 1 \qquad (4.29)$$

Equation (4.29) always provides a conservative estimate of $\theta(k)$; the approximation closely approaches equality for large values of k. By applying the results of Eq. (4.29) and Eq. (4.24) one can describe the probability of noise-induced jitter exceeding some specified amount in a pulse of sinusoidal carrier:

$$\Pr\{|\Delta t| > \alpha\} \approx \frac{1}{\sqrt{R}\,\alpha\,\omega_c\sqrt{\pi}}\, e^{-R\alpha^2\omega_c^2} \qquad (4.30)$$

An analogous result obtains for dc pulses:

$$\Pr\left\{\frac{|\Delta t'|}{\tau_r} > \alpha\right\} \approx \frac{\sqrt{2}}{\alpha R' \cdot \sqrt{\pi}}\, e^{-\alpha^2 R'^2/2} \qquad (4.31)$$

In applying Eqs. (4.30) and (4.31) to the different waveforms to estimate jitter values, we must bear in mind that the carrier-containing PDM and PPM signals will certainly have carrier signal running during 'off' periods of the pulse waveform. The NASA carrier level shift codes, for example, run a carrier 10 dB down from the pulse power level. The carrier is continuously present, and therefore may be coherently tracked and extracted.

Knowing the signal-to-noise ratio and the quiescent carrier level, the systems analyst may utilize Eq. (4.30) or Eq. (4.31) to directly compute the probability that the jitter will exceed specified limits either during the 'on' periods of pulse application or during the 'off' periods.

Figure 4.31 plots Eq. (4.30) for a 1 kHz carrier frequency.

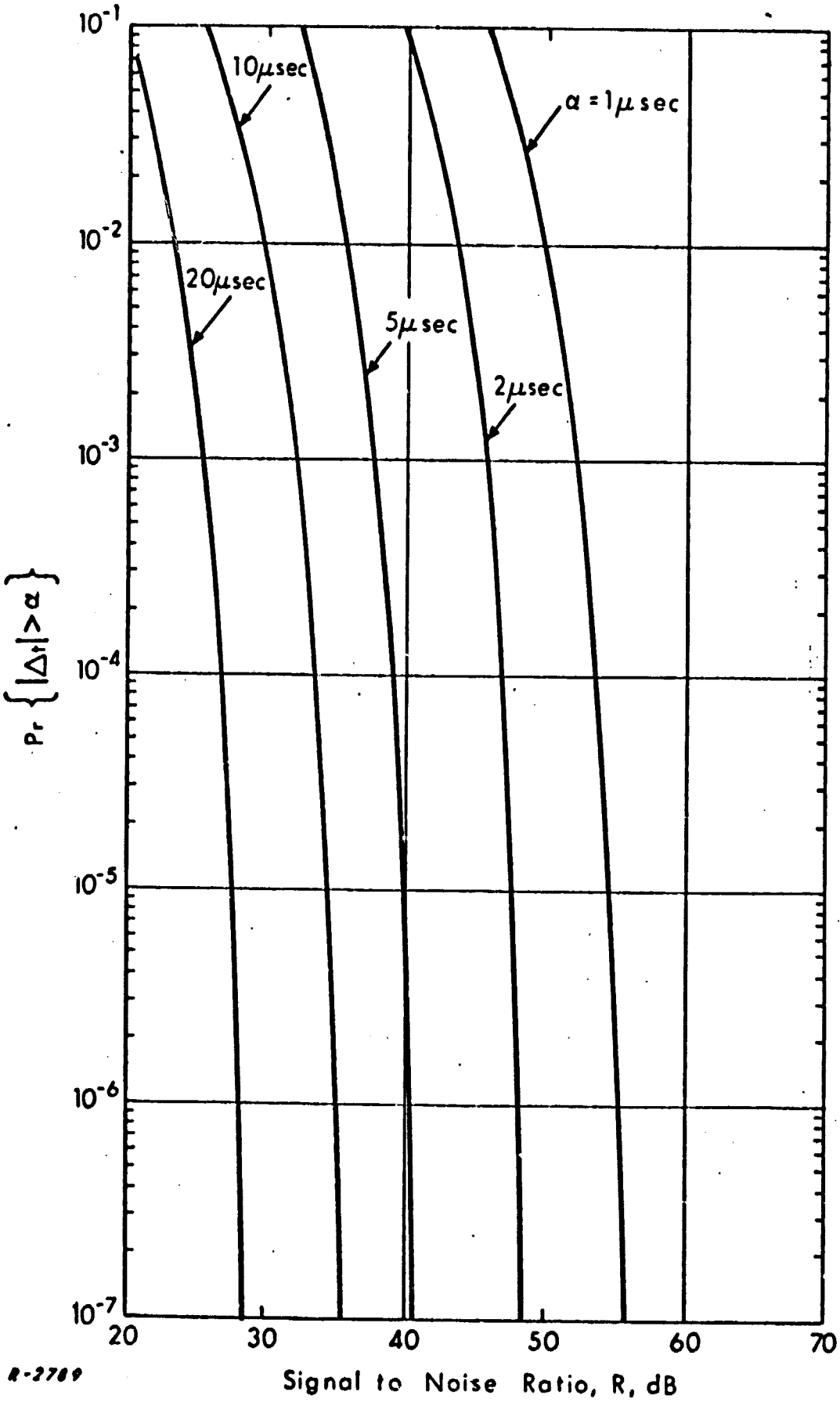


Fig. 4.31 Probability of Jitter Caused by Additive Gaussian Noise Exceeding Certain Limits; Carrier Level Shift Code, f_c = 1 kHz

4.4.2.3 Jitter Due to Low-Frequency Cutoff for Different Waveform Modulations

4.4.2.3.1 Introduction

The word "jitter" is herein used to denote the fine timing errors affecting time code waveforms. In those codes containing a reference carrier within their waveform structures, jitter is manifested as a random fluctuation in the locations of the zero crossings of the carrier with respect to their correct locations in time. If the code is composed of dc pulses jitter will show up as perturbations in the rising and falling edges of these pulses. In this section we shall consider, for PDM, PAM and PPM the magnitude of jitter created by low-frequency cutoff in the receiving-recording system.

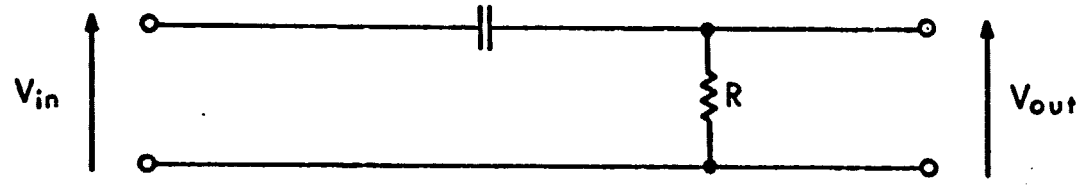
4.4.2.3.2 Jitter Due to Low-Frequency Cutoff: Carrier Level Shift Codes

In time codes of the carrier level shift variety, the pulses denoting coded time information are created by the stepping up of the carrier above its quiescent (continually running) level. The repeated stepping up and stepping down of the carrier generates transients in the system; lingering transient response corrupts the carrier and introduces jitter.

A simple model of low-frequency cutoff is shown in Fig. 4.32. Tape recording systems, which lack dc response, may introduce this sort of cutoff into a system.

In Ref. 1 the complete analysis of jitter was carried out for PDM carrier level shift codes. Results of this analysis are summarized in Table 4.4, which was worked out for a ratio of "on" carrier to "off" carrier of 10 dB. The quantity t_x is the jitter (maximum fluctuation) in the location of the first zero crossing of a pulse. (Strictly speaking, this quantity depends upon the entire past history of pulses, but in fact it is significantly affected only by the pulse immediately preceding.) The curve of t_x vs f_1 (the "corner" frequency

a) Equivalent Circuit



b) Laplace Transform:

$$H(s) = \frac{s}{s+\omega_L} \qquad \omega_L = \frac{1}{RC}$$

c) Asymptotic Plot of Magnitude of Filter Response

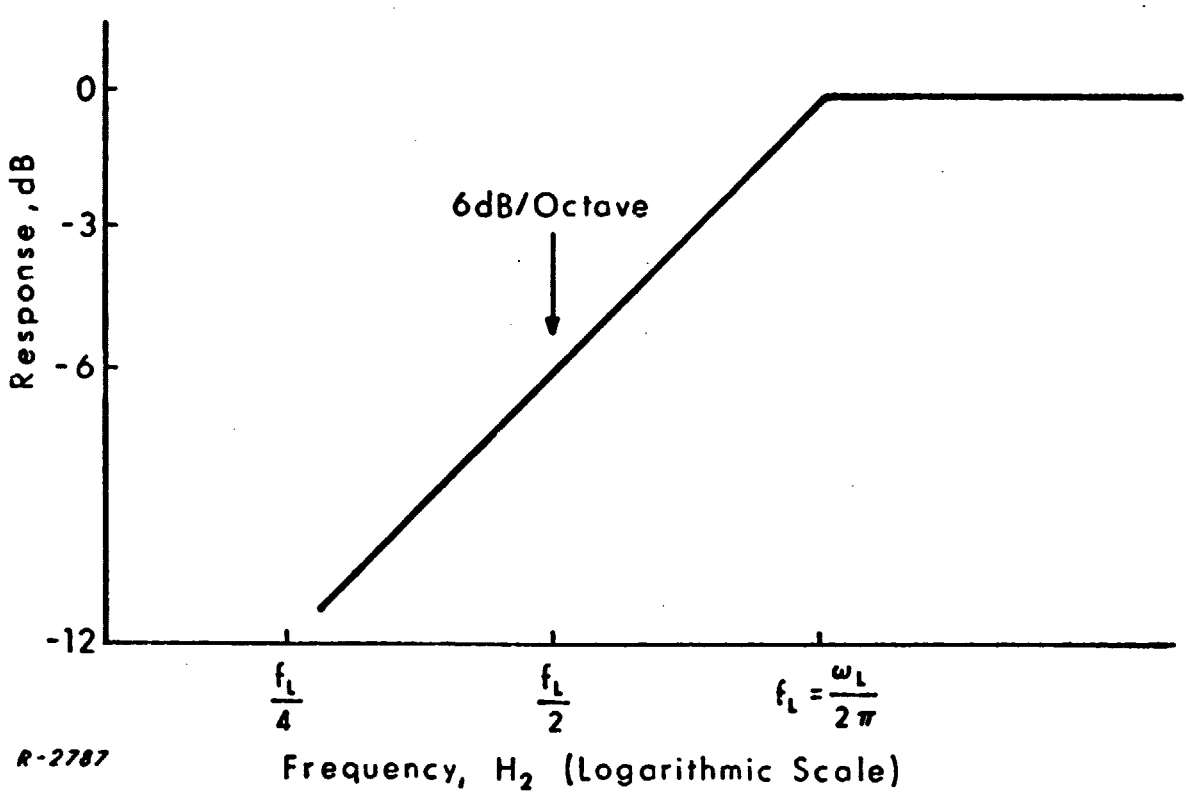


Fig. 4.32 Simple 6 dB/Octave Cutoff Function.

Table 4.4

Maximum Value of t_x as a Function of Critical Value of f_1 Producing It for Several PDM Carrier Level Shift Codes

Code	Max t_x	Critical f_1
NASA 36-Bit BCD (1 kHz carrier)	4.68 μs	39.8 Hz
NASA 28-Bit BCD (100 Hz carrier)	9.33 μs	0.80 Hz
NASA 20-Bit BCD (100 Hz carrier)	0.03 μs	0.007 Hz
IRIG Standard Format A (10 kHz carrier)	0.93 μs	796.0 Hz
IRIG Standard Format B (1 kHz carrier)	9.33 μs	79.6 Hz
IRIG Standard Format C (100 Hz carrier)	18.7 μs	1.59 Hz
IRIG Standard Format C (1 kHz carrier)	0.19 μs	1.59 Hz
IRIG Standard Format D (100 Hz carrier)	0.15 μs	0.013 Hz
IRIG Standard Format D (1 kHz carrier)	0.00 μs	---
IRIG Standard Format E (100 Hz carrier)	93.3 μs	7.96 Hz
IRIG Standard Format E (1 kHz carrier)	0.93 μs	7.96 Hz
IRIG Standard Format F (100 Hz carrier)	1.86 μs	0.16 Hz
AMR A2 (1 kHz carrier)	0.00 μs	---
AMR B2 (1 kHz carrier)	0.09 μs	0.80 Hz
AMR C2 (1 kHz carrier)	1.86 μs	15.9 Hz
APGC Eglin Signal # 4 (1 kHz carrier)	4.68 μs	39.8 Hz

of the filter) has a single hump; the coordinates of this maximum are the values appearing in Table 4.4.

The magnitude of jitter is determined by the shortest "off" period between "on" bursts of carrier. Clearly then, the PPM versions of any of the PDM codes listed in Table 4.4 would have jitter values orders of magnitude smaller. The reason for this is that with PPM ($\beta < 1/2$) the time between "on" bursts will be at least a bit interval T; with the NASA 36 bit BCD code, this interval may be as small as 0.4 T. For the same reasons PAM versions of these codes will have negligible jitter; PAM is simply step-up keying for while bit periods with respect to the quiescent carrier.

4.4.2.3.3 Jitter Due to Low-Frequency Cutoff: dc Level Shift Codes

Low-frequency cutoff will produce jitter in dc level shift codes just as it does in carrier level shift codes. A pulse having absolutely vertical rising and falling edges cannot possibly have any jitter introduced into the locations of these edges by the presence of transient terms. Such pulses can only occur, however, in systems having infinite bandwidth. We have therefore chosen to model the level shift pulse as having a linear leading edge with a "risetime" of τ_r, i.e., $1/\tau_r$ is indicative of the highest frequency passed by the system.

In carrier level shift codes jitter is a function solely of low-frequency cutoff (the results of Sec. 4.4.2.3.2 would not be significantly changed by the introduction of any reasonable high frequency cutoff above the carrier frequency into the filter function). For the dc level shift codes, the interplay of both high and low frequency response is of concern; the low-frequency cutoff determines the residual transient and the high frequency response determines the sensitivity (slope) of the leading edge.

The jitter for these codes is best expressed in terms of the dimensionless ratio t_x/τ_r. Table 4.5 summarizes the results of analysis in Ref. 1 in the

Table 4.5

Maximum Value of (t_x/τ_r) as a Function of Critical Value of f_1 Producing It for Several PDM dc Level Shift Codes

Code	$(t_x/\tau_r)_{max}$	Critical f_1
APGC Eglin Signal # 7	0.167	0.22 Hz
AFFTC Edwards Center Format A	0.316	367.5 Hz
AFFTC Edwards Center Format B	0.167	27.4 Hz
AFFTC Edwards Center Format C	0.167	0.27 Hz
IRIG B: dc Level Shift Version	0.474	47.4 Hz

evaluation of jitter in PDM dc level shift codes. For the same reasons previously enumerated, the PPM and PAM versions of such codes will display jitter which is quite small as a fraction of risetime.

4.4.3 Conclusions

We may observe in conclusion that jitter due to low-frequency cutoff is not a very substantial problem in PDM time codes; the values in Tables 4.4 and 4.5 are "worst case" values, and those which would arise in actual system usage would tend to be much lower. Jitter will be even more unimportant when PAM and PPM bit encodings are employed.

4.5 Analysis of the NASA and IRIG Time Codes

4.5.1 Basic Characteristics of Codes

The NASA and IRIG time code group is composed of ten codes -- three NASA pulse-duration modulated (PDM) binary-coded decimal (BCD) codes, the NASA serial decimal code, and six IRIG PDM BCD codes. The three NASA PDM codes are the NASA 36 bit, 28 bit, and 20 bit BCD codes. Sample waveforms are given in Figs. A.1 through A.10.

The 36 bit code is shown in Fig.A.1, Appendix A, and has a frame time of 1 sec and 100 bits/frame. We shall define a marker as a characteristic of the waveform that marks a particular instant in time, such as the beginning of a frame or the leading edge of a bit. A label will be defined as a piece of encoded data uniquely identifying a particular mark in time. Accordingly, the NASA 36 bit code contains a labeled marker once every second at the beginning of each frame, and an unlabeled marker once every 10 msec at the leading edge of each bit. The time at the beginning of the frame in seconds, minutes, hours, and days is labeled during the frame in PDM BCD. The frame is divided into 10 subframes of 10 bits each. The first four bits of the first subframe give units of seconds in BCD form (1, 2, 4, and 8). The remaining six bits of the first subframe are marker bits used for clocking only. Similarly the second subframe shows tens of seconds; the 3rd, units of minutes; the 4th, tens of minutes; the 5th and 6th, units and tens of hours; the 7th, 8th, and 9th, units, tens, and hundreds of days. The 10th subframe is divided half and half between bits that identify the sending station, and reference or frame synchronization bits. The narrowest pulse width is the 2 msec binary "0". The widest pulse width is the 6 msec binary "1". The narrowest off-time is the 4 msec between the end of a binary "1" and the beginning of the next bit. The narrowest pulse width or the narrowest off-time in the waveform, whichever is smaller, determines the spectrum. Since the spectrum extends to infinity in frequency, no

finite bandwidth channel can include 100% of the signal energy. It is therefore more pertinent to speak of the bandwidth required to include a given percentage of the bit energy. This is obtained from the integral under the spectral density curve. The frame width, bit width, width of the narrowest and widest pulses and the narrowest off time, the bandwidth required to include 90%, 95%, and 98% of the bit energy, as well as the length of the frame sync pulse train in time and percentage of frame width are given for each of the ten NASA and IRIG codes in Table 4.6.

The NASA 28 bit code has a frame time of one minute and 120 bits/frame. Twelve subframes of 10 bits each identify in PDM BCD form the time at the beginning of the frame to a precision of one minute. The "precision" is the smallest division in time that a code will display. Precision should not be confused with accuracy which, of course, is much higher. Zeros are implied for all units of time smaller than the degree of precision. The NASA 20 bit code has a frame time of one hour and 60 bits/frame. Six subframes of 10 bits each identify in PDM BCD the frame time to a precision of one hour. All three NASA BCD codes use 5 bits for frame sync.

The NASA serial decimal code has a frame time of ten seconds and 100 bits/frame. There are 10 subframes of 10 bits each which identify in serial decimal form the time to a precision of 10 seconds. In serial decimal encoding, the number of pulses in a subframe determines the value of the data, in straight decimal form. All pulses are the same width. One sync pulse per subframe is always transmitted; two sync pulses mark the beginning of a frame. For example, the number 6 would be transmitted as 7 pulses (6 data + 1 sync) in all subframes except the first subframe in the frame. There it would be 8 (6 data + 2 sync). The remaining pulses in the subframe are omitted altogether.

In the IRIG codes, subframes are used but for synchronization purposes only, All the data is transmitted contiguously. All IRIG codes use two bits for frame sync and one bit for subframe. The space remaining after the data has been transmitted is used for control functions and optionally to transmit time

Type of Code	Frame Width	Bit Width	Narrowest Pulse	Widest Pulse	Narrowest Off Time	Bandwidth for Given % of Energy			Reference Width (Frame Synchronization)	
						90%	95%	98%	Time	% of Frame
NASA Codes										
36 bit	1 sec	10 msec	2 msec	6 msec	4 msec	391.5 Hz	811.5 Hz	2.5 KHz	50 msec	5
28 bit	1 min	0.5 sec	0.1 sec	0.3 sec	0.2 sec	7.83 Hz	16.23 Hz	50.0 Hz	2.5 sec	4.16
20 bit	1 hr	1 min	12 sec	36 sec	24 sec	0.06525 Hz	0.13525 Hz	0.4167 Hz	5 min	8.33
Serial Decimal	10 sec	0.1 sec	50 msec	50 msec	50 msec	15.66 Hz	32.46 Hz	100.0 Hz	0.4 sec	4
IRIG Format										
A	0.1 sec	10 msec	2 msec	8 msec	2 msec	391.5 Hz	811.5 Hz	2.5 KHz	20 msec	2
B	1 sec	10 msec	2 msec	8 msec	2 msec	391.5 Hz	811.5 Hz	2.5 KHz	20 msec	2
C	1 min	0.5 sec	0.1 sec	0.4 sec	0.1 sec	7.83 Hz	16.23 Hz	50.0 Hz	1 sec	1.67
D	1 hr	1 min	12 sec	48 sec	12 sec	0.06525 Hz	0.13525 Hz	0.4167 Hz	2 min	3.33
E	10 sec	0.1 sec	20 msec	80 msec	20 msec	39.15 Hz	81.15 Hz	250.0 Hz	0.2 sec	2
F	5 min	5 sec	1 sec	4 sec	1 sec	0.783 Hz	1.623 Hz	5.0 Hz	10 sec	3.33

Table 4.6 NASA and IRIG Code Characteristics

of day in straight binary seconds. IRIG A has a 0.1 second frame time, 100 bits/frame, and a precision of a tenth of a second. IRIG B has a one second frame, 100 bits/frame, and one second precision. IRIG C has a one minute frame 120 bits/frame, and one minute precision. IRIG D has a one hour frame time, 60 bits/frame and one hour precision. IRIG E has a 10 second frame time, 100 bits/frame, and a 10 second precision. IRIG F has a 5 minute frame time, 60 bits/frame, and a one minute precision.

4.5.2 Suitability for Transmission over Wireline, LOS and BH Tropo Media

It was pointed out in Section 3 that the coherence bandwidths of the LOS and BH tropospheric media were generally in the order of a megaHertz or more, and that the fading rates of these media were in fractions of a Hz for surface-to-surface links and in tens and perhaps hundreds of Hz for surface to air or to low-orbital-craft links. For practical purposes, the higher fading rates can be assumed to be rarely in excess of 10 Hz. These observations suggest that no significant limitations on the performance of time code systems of the NASA and IRIG varieties need be encountered over such links because of coherence bandwidth. The fading rate on the other hand may be expected to degrade performance on the codes with longer pulse widths. Performance limitation due to fading is dependent upon whether the width of the longest pulse in the code exceeds the fading time. The fading time is equal to the reciprocal of the fading rate, and in this case is a few seconds for surface-to-surface links and is measured in tens of milliseconds for surface-to-air links. From Table 4.6 we note that the NASA 20 bit code (36 sec), the IRIG D (48 sec), and the IRIG F (4 sec), and possibly the NASA 28 bit (0.3 sec) and the IRIG C (0.4 sec) will be subject to performance limitations due to the fading rate of the channel. For the remaining codes, only the probable drop-outs or outages pose a threat to the transmission reliability.

The situation is however different with wireline media because of the attenuation and delay characteristics of these media. To start with, there is

a gap in the transmission from zero Hz to about 300 Hz. Single nominal 4 kHz channels also cut off effectively above about 3300 Hz. In order to assess the general effects of these characteristics some further analysis is required, as follows.

4.5.2.1 Baseband Transmission

The IRIG and NASA baseband code spectra are given by the rectangular waveform spectrum of Fig. 4.12. Passing these through a wireline channel will cause a loss of the dc content below about 300 Hz. Figure 4.14 gives the percentage energy contained in a specific bandwidth. This may be interpreted as the percentage energy lost if the frequency response of the channel is cut off below a specific bandwidth. Using 300 Hz and the data from Table 4.6, we find that the NASA 36 bit, and the IRIG A and B lose about 84% of their energy, and that all the others lose better than 99%. It may be safely said that the IRIG and NASA codes are not suitable for transmission over a wireline channel at baseband.

4.5.2.2 Subcarrier Transmission

While none of the codes are suitable for wireline transmission at baseband, the spectrum of any of the codes may be shifted in frequency by use of a suitable subcarrier. The bandwidth of a wireline channel is from 300 Hz to 3300 Hz, or about 3 KHz. From Table 4.6 we observe that all NASA and IRIG codes have a 98% bandwidth spectrum of 2.5 KHz or less. Note that this is the double sided spectrum which is applicable if conventional DSB AM is used to modulate the subcarrier. Therefore any of the codes may be transmitted over a wireline channel by use of suitable subcarrier. For a 300 Hz to 3300 Hz wireline, a 1.8 KHz subcarrier is optimum for the three high bandwidth codes. For the others, any subcarrier between about 400 Hz and 3 KHz is acceptable.

4.5.3 Acquisition Properties of Time Codes

4.5.3.1 General

The process of acquiring frame sync on a time code must be gone through when timing data is initially received, and subsequently whenever transmission link outages have caused loss of frame sync. In order to evaluate properly the timing system acquisition characteristics, it is necessary to clearly define the following properties:

1. Probability of Correct Acquisition, P_C

   This is the probability that, starting from an out-of-sync condition, the system will acquire sync correctly and announce this fact.

2. Probability of False Acquisition, P_{FA}

   This is the probability that the system will produce the announcement that correct sync has been acquired when in fact an out-of-sync condition exists. P_{FA} is merely the complement of P_c : $P_{FA} + P_C = 1$.

3. Mean Time to Acquisition, T_A

   This is the average time it takes from the beginning of the acquisition process to its successful completion. Other quantities describing the statistics of acquisition time may also prove to be valuable, such as the probability of it exceeding certain values somewhat larger than the mean.

4. Mean Time to Recognize False (or Lost) Acquisition, T_D

   This is the average time elapsed from the initiation of a false announcement of acquisition (or from loss of frame sync) until the error is recognized and acquisition procedures are reinitiated. This characteristic will be of importance in command systems where incorrectly labelled data may be processed in real time to produce erroneous command information.

5. Rate of Occurrence of Wrongly Initiated Reacquisition Procedures, N_F

   This is a kind of "false alarm" rate: the occurrence of a few errors raises the fear that sync has been lost (though in fact it has not) with the result that timing information is considered invalid until reacquisition has been successfully achieved.

There are interesting tradeoff relations involving these quantities and the signal-to-noise ratio (or bit error rate). The exact quantitative forms that such tradeoffs would assume depend upon the time code, the transmission link characteristics, and the exact nature of the decision strategies adopted by the user. Exhaustive descriptive analyses of time code acquisition properties are therefore ruled out by virtue of the large magnitude of such a task.

Before proceeding with a more limited analytic study of time code acquisition for a few carefully selected cases, it is most useful to list quantitatively the general tradeoff trends that should exist among the abovementioned properties, regardless of actual user implementation as follows.

(i) As the user decision strategy is held fixed, and the signal-to-noise ratio increases, P_C will increase and T_A will decrease, N_F will decrease, while T_D will not tend to change appreciably.

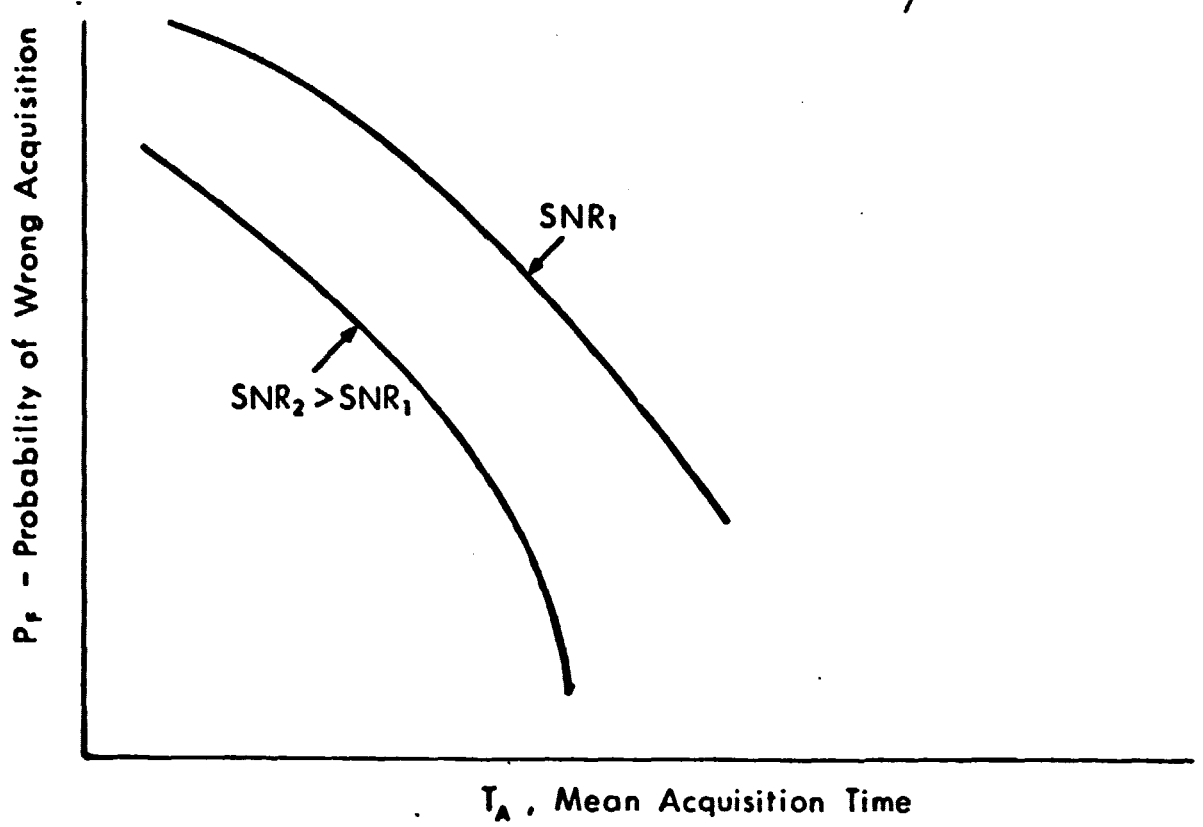
The reason for this is that as the link becomes "cleaner", the occurrence of erroneously decoded elements (or those for which decoder confidence in correctness is low) will decrease. With this more perfect data, P_C must increase, and T_A will decrease since higher confidence in frame location will often eliminate the necessity of waiting for further data to bring confirmation. The lower frequency of bit errors must reduce N_F, as these unnecessary reacquisition attempts are triggered by confidence-destroying error patterns. The recognition of false acquisition depends upon the fact that the resultant pseudo-frames will differ so strongly from proper frames as to promptly generate an alarm; this mechanism should not be much affected by the addition of a few bit errors.

(ii) For a fixed signal-to-noise ratio, one can through variations in decision strategy increase P_C at the cost of increasing T_A. Likewise, one can decrease N_F if willing to tolerate a greater T_D.

An increase in P_C without any increase in signal-to-noise ratio may only be achieved by raising the threshold level of confidence at which a decision will be made; this requires an increased T_A to provide the necessary integration to increase confidence. Similarly, the use of a more strongly confirmed out-of-sync indicator as a trigger to initiate reacquisition will reduce N_F but will increase T_D.

(iii) The curve of P_F vs T_A for a given signal-to-noise ratio will lie totally above the corresponding curve for a higher signal-to-noise ratio (see Fig. 4.33a). The curve of N_F vs T_D for a given signal-to-noise ratio will lie totally above the corresponding curve for a higher signal-to-noise ratio (see Fig. 4.33b).

The user may conceivably possess some sort of a threshold raising/lowering dial on his equipment, which enables him to control the threshold for acquisition announcement and generate the P_F vs T_A curve. If the signal-to-noise ratio increases, then for a given T_A a lower P_F may be obtained than before, or for the same P_F a lower T_A is required. Thus, the second curve lies below the first. Similar considerations apply to the N_F vs T_D curve.



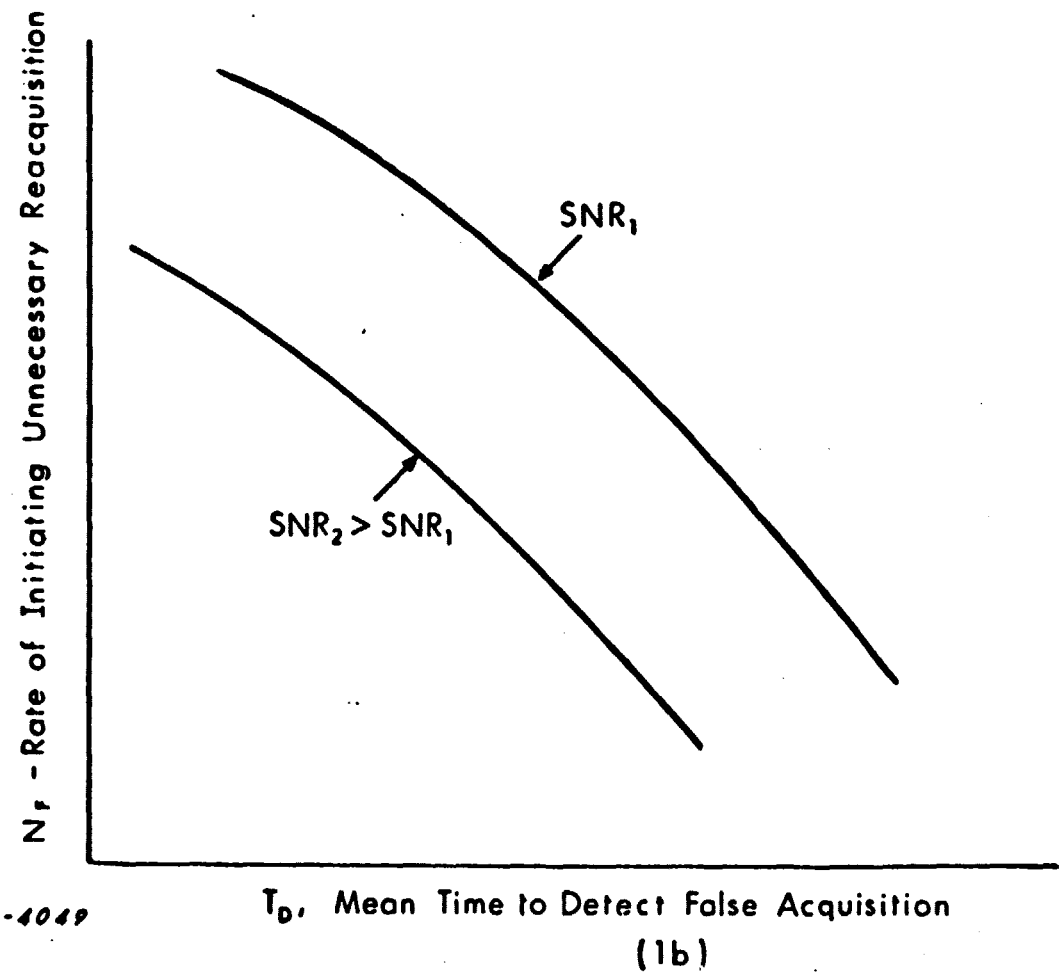


Fig. 4.33 Variation of Timing System Acquisition Characteristics with Signal-to-Noise Ratio

4.5.3.2 Analysis

To see how these qualitative results might actually work out in a real situation, we have chosen first to look at the NASA BCD codes. The 36 bit and 28 bit BCD carrier level shift code frames are divided into 10 parts by position identifier markers (1's) occurring every ten index periods. In any such section bit position Number 1 is the position identifier bit. In the section containing the reference marker, the next four bits are reserved for identification of the sender; five consecutive 1's following this constitute the actual reference marker (see Figs. 4.34 and 4.35). The other sections have five 0's in these positions.

Let us first consider that frame acquisition is announced upon having detected five consecutive 1's; the marker immediately following is considered to be the on-time marker of the next frame. During the acquisition process a sliding bit-by-bit search of each five bits in a row must be made. The dominant cause of false acquisition is the correct decoding of a position identifier followed by the BCD number 7 (three 1's) followed by an index marker "0" incorrectly decoded as a "1", thus spuriously generating the reference marker sequence.

If the bit error probability is represented by the symbol p, and we consider (as a conservative approximation) that during the scan for acquisition the encountering of one of the above-described "dangerous" configurations will arise and will constitute the total hazard to false acquisition* then

$$P_{FA} \approx p(1 - p)^4 \sim p \qquad \text{for } p \ll 1 \tag{4.32}$$

If it is considered that false acquisition is noted by failure of the five 1's to reappear exactly one frame interval later, then assuming that the BCD digit is still a "7", i.e., it was not the seconds digit if we were dealing with the 36 bit code, etc., the false acquisition will be detected one frame later

* Other false acquisition would depend upon multiple error patterns and would have terms involving p^2 and higher powers of p; they can generally be ignored relative to (1).

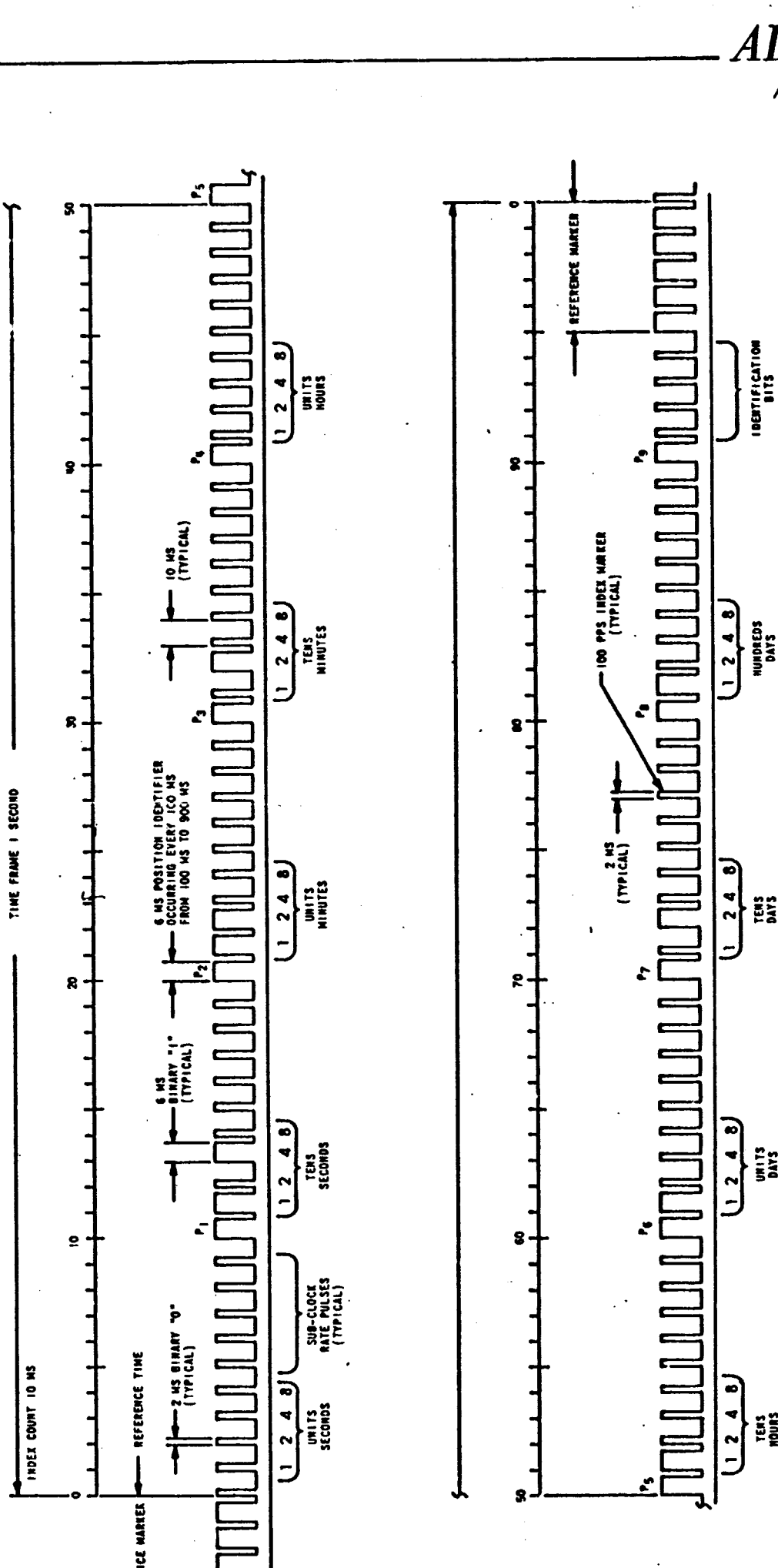


Fig. 4.34 NASA 36-Bit BCD Time Code Carrier Level Shift, F_C = 1 kHz

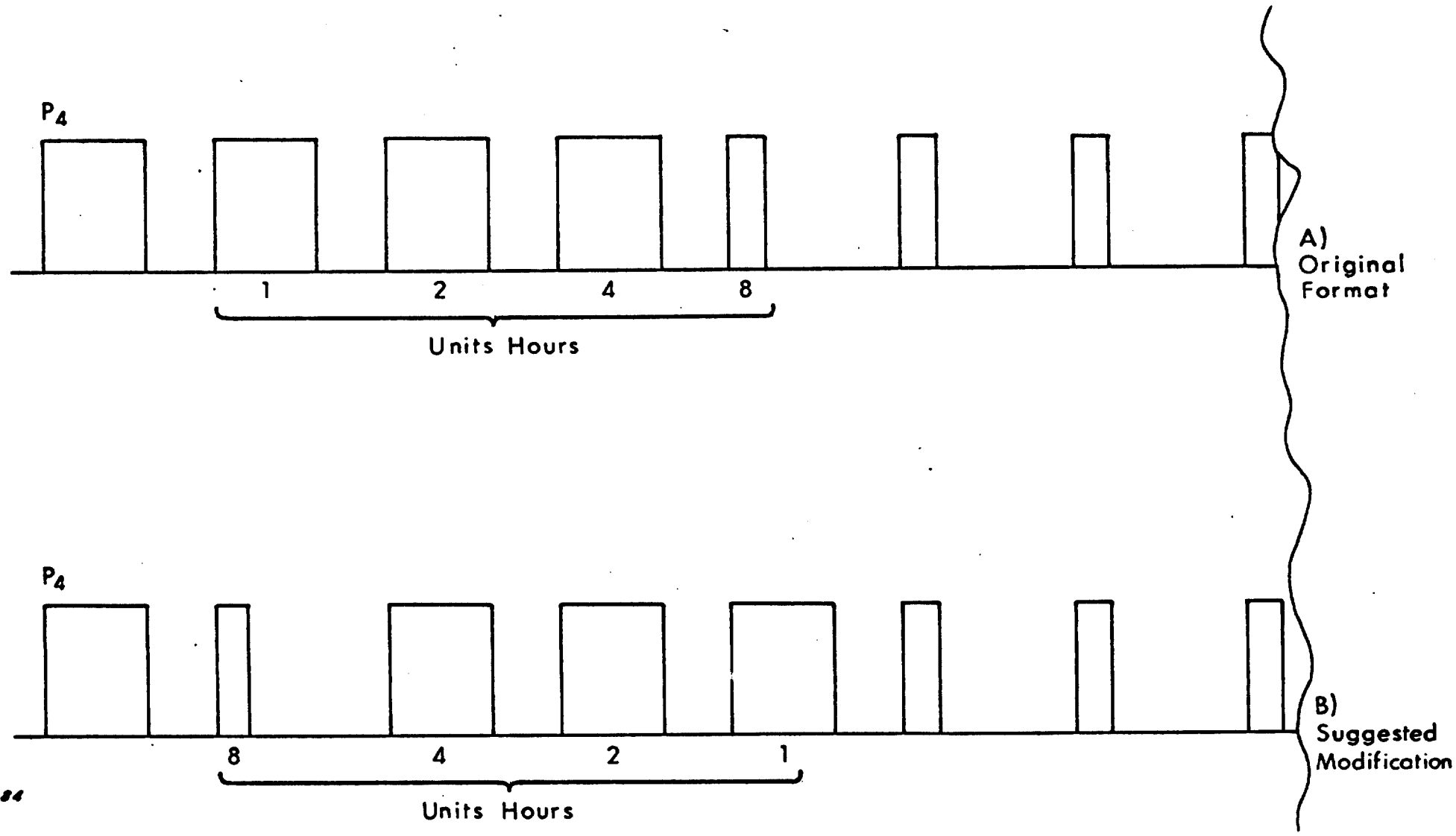


Fig. 4.35 Section of NASA 36-Bit BCD Time Code Format, Illustrating Modification to Reduce False Sync Occurrence

with probability 1 - p, two frames later with probability p(1 - p), ... m frames later with probability $p^{m-1} \cdot (1 - p)$. Thus

$$T_D = \sum_{m=1}^{\infty} p^{m-1}(1 - p) \cdot m = \frac{1}{1-p} \text{ frame intervals} \quad (4.33)$$

$$\sim 1 + p$$

T_D is seen to be very nearly one frame time for all $p \ll 1$.

The mean time to acquisition T_A consists of P_C times the "average" time of half a frame interval (on the average this process will be begun in mid-frame) plus the term $P_{FA}(T_D + 0.5)$. This latter term approximates the average time added by false acquisitions which must be subsequently connected by reacquisition. Thus

$$T_A \equiv \frac{1}{2} + p(1 - p)^3 \sim \frac{1}{2} + p \quad \text{frame intervals} \quad (4.34)$$

If it is assumed that the user considers the frame sync to be lost whenever the expected sequence of five 1's fails to appear, then a wrongly initiated reacquisition procedure will come about when any of the bits in the reference marker are decoded erroneously subsequent to valid acquisition of frame sync. Thus,

$$N_F \approx 1 - (1 - p)^5 \sim 5p \quad \text{false reacquisitions/frame interval} \quad (4.35)$$

A simple modification of user strategy would be to require that two successive identifications of the reference marker separated by a frame interval are required to confirm acquisition and that the reacquisition procedure will only be initiated upon two successive failures of the reference marker to appear as anticipated. We may then derive the following asymptotic expressions for small p.

$$P_{FA} \sim p^2 \quad (4.36)$$

$$T_D \sim 2 + 3p \quad \text{frame intervals} \quad (4.37)$$

$$T_A \sim 3/2 + 2p^2 \quad \text{frame intervals} \quad (4.38)$$

$$N_F \sim 25\,p^2 \quad (4.39)$$

By comparison of Eqs.(4.36) - (4.39)with Eqs.(4.32)-(4.35)we see that P_{FA} and N_F have been squared, i.e., decreased by several orders of magnitude the cost of attaining this improvement is an approximate doubling of T_D and an approximate tripling of T_A. More subtle user strategies, which are not profitable to analyze at this point, might involve the employment of more refined bit decision methods in which confidence "scores" are expressed as well as the actual binary decision. Acquisition decisions and initiations of reacquisition would be keyed to associated confidence scores crossing critical threshold levels.

It is instructive to consider the effect of the format modification of the NASA BCD codes suggested in Ref. 1 (see Fig. 4.35) upon its acquisition characteristics. Reevaluation of these characteristics under the same user strategy leading to Eqs. (4.32) - (4.35) leads to:

$$P_{FA} \sim p^2$$

$$T_D \sim 1 + p^2$$

$$T_A \sim 1/2 + p^2$$

$$N_F \sim 5p$$

Note that the format change affects the squaring of P_{FA} without significantly changing the other quantities.

4.5.3.3 Comments

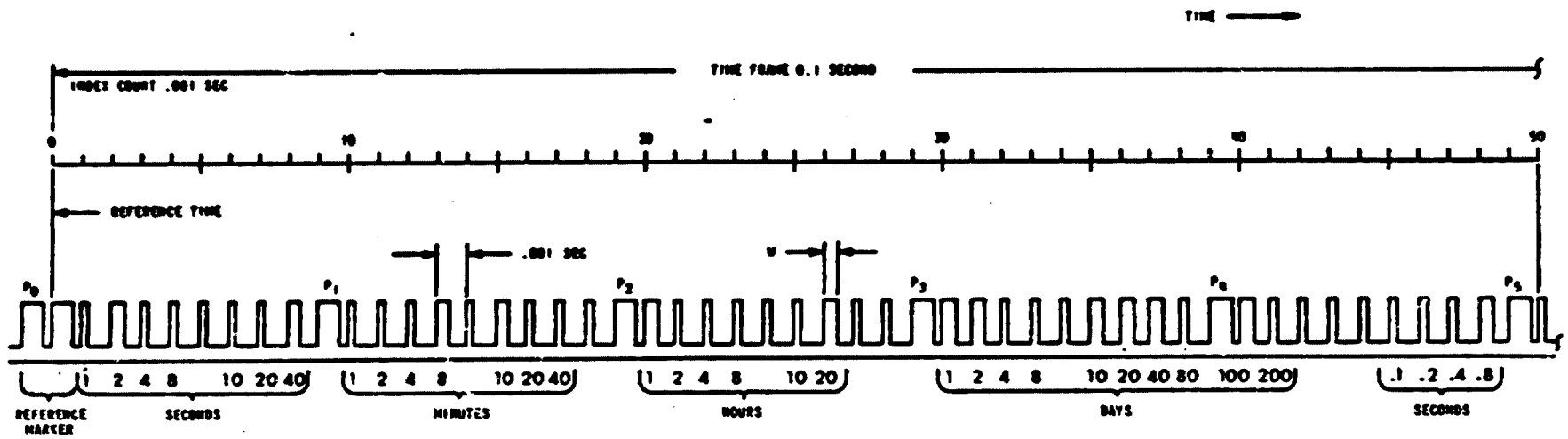
The one property inherent in a time code itself which will contribute favorably to timing system acquisition performance is the degree to which the reference marker differs from the string of markers within the frame with which it is most likely to be confused. For the NASA BCD codes this consists of a difference of duration of 0.4 index counts in a single bit; the modified format guarantees a duration difference of 0.4 index counts in two bits. The situation is somewhat worse for the IRIG family of time codes, of which format

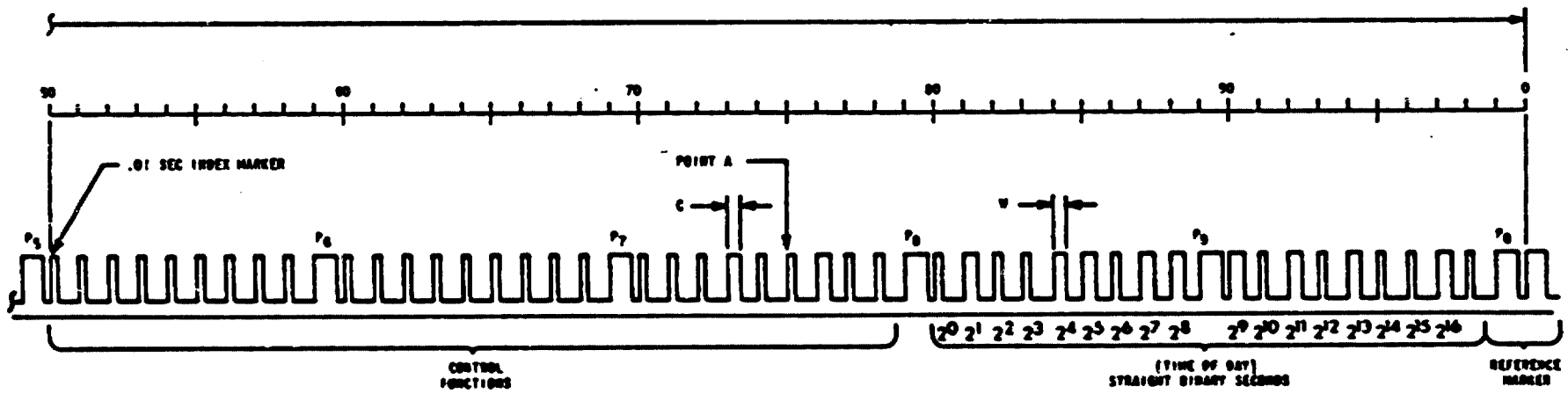
A (Fig. 4.36) is typical, where the reference marker consists of a 0.8 index count long pulse following a 0.8 count long position identifier. Any odd-valued grouped-binary digit will place a 0.5 count long binary '1' following a position identifier, representing a duration difference of 0.3 index counts in a single bit.

In terms of this energy difference factor, if we consider the standard NASA BCD format as a reference, the modified NASA format is 3.0 dB better, and the IRIG format is 1.25 dB worse. It is to be expected* that a set of acquisition characteristics and tradeoffs similar to those obtained with a timing system using the standard NASA BCD format may be obtained with one using the modified NASA format but with 3 dB less signal-to-noise ratio available; to obtain similar acquisition performance using the IRIG format would require an excess signal-to-noise ratio of 1.25 dB with respect to the standard NASA BCD format.

We now comment in more detail upon the possible fruitful approaches which a user might employ for the detection of a false acquisition. Consider first the false acquisition of the IRIG Format A carrier level shift code (Fig. 4.37). Position #30 is mistakenly thought to be position #0. As a result the grouped binary numbers read out will be unrelated to the correct values. Positions 30 - 41 should provide the coded day of the year; pseudo-positions 30 - 41 will contain bits designated for control functions and will doubtless give a wrong indication. If it is assumed that the user has locally available a sufficiently good clock to provide him with day of the year, hours and minutes, then within less than half a frame time after wrong acquisition three warnings of acquisition failure have been provided through wrong readings in these grouped binary values. It will similarly be true with the NASA BCD codes that the incorrect

* This intuitive statement may be justified somewhat more rigorously through arguments based upon a geometric interpretation of decision theory, for the case of gaussian noise.





P=POSITION IDENTIFIER, 0.8MS DURATION
W=WEIGHTED CODE DIGIT, 0.5MS DURATION
C=CONTROL ELEMENT (EXAMPLE) 0 5MS DURATION
DURATION OF INDEX MARKERS = 0.2MS

R-2805

TIME AT POINT A: 21:10:42+.0+.07+.005
= 21 HR., 10 MIN., 42.075 SEC ON DAY 173

Fig. 4.36 IRIG Standard Format A Carrier Level Shift, F_C = 10 kHz

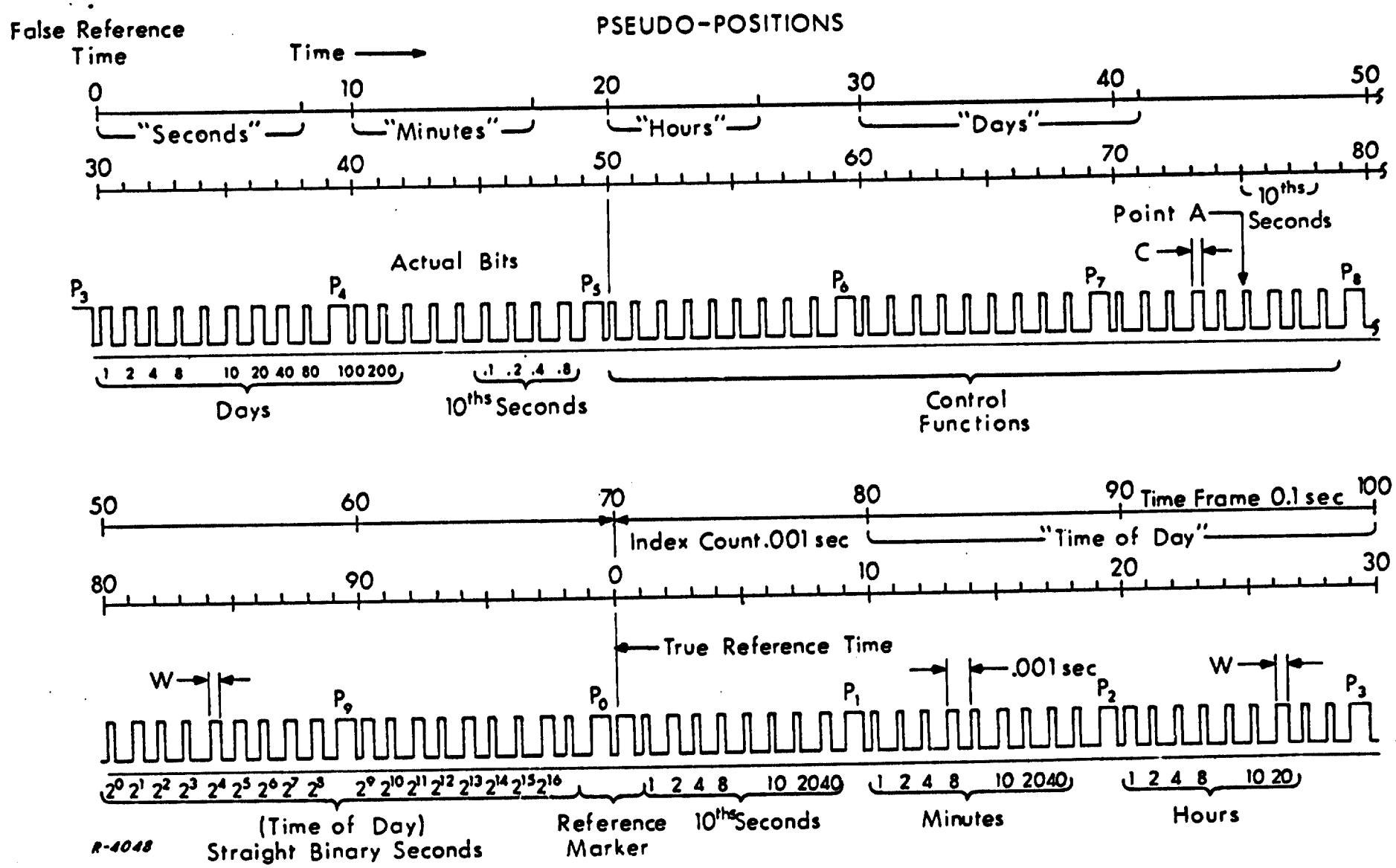


Fig. 4.37 Incorrect Acquisition of IRIG Carrier Level Shift Time Code, Format A

nature of the coarser digital time parameters provides sufficient information to indicate the need for initiating reacqusition procedures well before a frame time has expired.

Once the code has been acquired, the digital time must change predictably, stepping along one frame time for each successive frame. It requires little user complexity to generate and store the pattern of bits conveying time anticipated in the next received frame. When a bit-by-bit comparison indicates that a given error threshold has been exceeded, the signal is given that frame sync has somehow been lost and must be reacquired. In general, the user's basic approach should be to compare what he gets as a time code frame with what he expects to get, and look upon the presence of a strong discrepancy as evidence that frame sync is not present.

4.6 Possible Uses for Idle Bit Positions

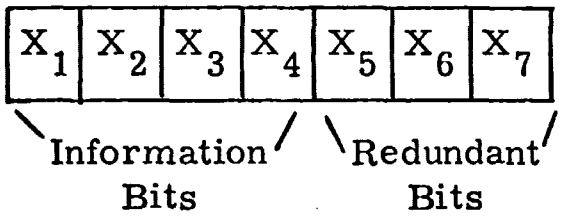
4.6.1 Error-Correction Coding

The presence of index markers in the NASA and IRIG codes having no specified function permits of the use of these bit positions for the purpose of supplying redundant parity checks upon information-bearing bits in the time code frame.

The most appropriate error-control codes to consider for use in conjunction with time codes are block codes. The essentials of block error-control codes are best explained by a simple example (see Fig. 4.38). The code displayed is one form of the (7, 4) single-error-correcting code.

The redundant bits X_5 - X_7 are dependent upon the information bits X_1 - X_4 through the parity check relationships in Fig. 4.38b. These are checks for even parity; (i),for example, states that the first, third, fifth and seventh bits contain an even number of 1's (zero is considered to be an even number). As noted, the validity or failure of the three parity relations upon reception immediately pinpoints the location of a single error or indicates an error free block if that is the case.

(a) BLOCK OF 7 BITS



(b) PARITY CHECKS (ESTABLISHED ON TRANSMISSION)

i) $X_7 = X_1 \oplus X_3 \oplus X_5$

ii) $X_6 = X_2 \oplus X_3 \oplus X_7$

iii) $X_5 = X_4 \oplus X_6 \oplus X_7$

(c) PARITY CHECKS (PERFORMED UPON RECEPTION)

i) $Q_1 = X_1 \oplus X_3 \oplus X_5 \oplus X_7$

ii) $Q_2 = X_2 \oplus X_3 \oplus X_6 \oplus X_7$

iii) $Q_3 = X_4 \oplus X_5 \oplus X_6 \oplus X_7$

(d) SYNDROME = NUMBER TELLING ERROR LOCATION

$$S = 2^2 \cdot Q_3 + 2^1 \cdot Q_2 + 2^0 \cdot Q_1$$

For error in X_j, $S = j$.

For no errors, $S = 0$.

Fig. 4.38 (7, 4) Single-Error-Correcting Code

Decoding will fail if two errors affect a block. Imagine that X_1 and X_5 are both in error. Performance of parity checks upon reception will yield $Q_1 = 0$, $Q_2 = 0$, $Q_3 = 1$; thus, $S = 4$, and the fourth bit will be changed. In fact, it may be shown for this code that no matter how badly perturbed by errors a block may be it may always be changed into a sequence satisfying all parity checks by changing one (and only one) bit.

The code cited has the capability to render the information-bearing output error free as long as no more than one bit error per block is presented to the user. The occasional presence of an 'uncorrectable' error pattern will lead to an output of a bad information block.

Taken down to its most essential elements, error-control coding is a means of increasing the reliability, i.e., decreasing the probability of error, of an arbitrary stream of bits.

It is most important to reaffirm a fact previously noted: the stream of bits within successive time code frames is far from arbitrary. It is in fact highly predictable from frame to frame. As a result, once a frame has been properly acquired, correction of occasional errors is best made by comparison with the bits in the preceding frame. Just about any likely pattern of errors can be detected and corrected in this way. Moreover, as noted in the preceding section, the more coarse elements of digital time, e.g., month, day, hour, minute, are readily available to the user from his own sources before the time code reception process has even been begun, and constitute an important source of error-control capability.

It is therefore concluded that the use of error-control coding is unnecessary. Error-correction power is limited only by the need to avoid 'correcting' the substantial error patterns which arise as a result of lost frame sync, and which must not be confused with the low-density error patterns due to random noise degrading a frame. Our recommendation regarding the most sensible usage for idle bit positions is therefore to employ them as a means of strengthening reliability and speed of frame acquisition, i.e., use them to contribute to

the strength and reliability of the reference marker. This amounts to saying that it is desirable to build a highly reliable frame sync sequence into the code. Important research has been documented on the topic, and significant results have been obtained. We will cover this topic in a later section.

4.6.2 A Note on Frame Synchronization Sequences for Time Codes

As stated above, it is our conclusion that unused bit positions in the NASA and IRIG code frames might well be used to build a more reliable reference marker, i.e., a frame sync sequence. Frame sync sequences are frequently employed in communication systems where the data must be properly broken into frames to be usable.

The general manner in which digital frame sync sequences may be employed in a data transmission system is briefly described below. It is described in terms of the simpler user implementation in which binary decisions are made upon each received bit; the extension to strategies in which confidence expressing scores are assigned to bits may easily be visualized.

Consider that the receiver continually searches a specified n bit positions among the last m received bits. In the absence of any bit errors these n positions will contain the bit pattern of the frame sync sequence at an instant in time which may be used to indicate the time of frame sync.

The multiplying-summing scheme indicated in Fig. 4.39 multiplies each of the proper n bits (which are considered to have a value of either +1 or -1) with the stored value that position would have right when the sync sequence is in place. The result is that for every bit period an integer correlation score between -n and +n appears as output. Frame sync is announced when the correlation score exceeds some preset threshold value.

In the design of frame sync sequences and frame acquisition procedures, one must note that there are three distinct phases of operation which

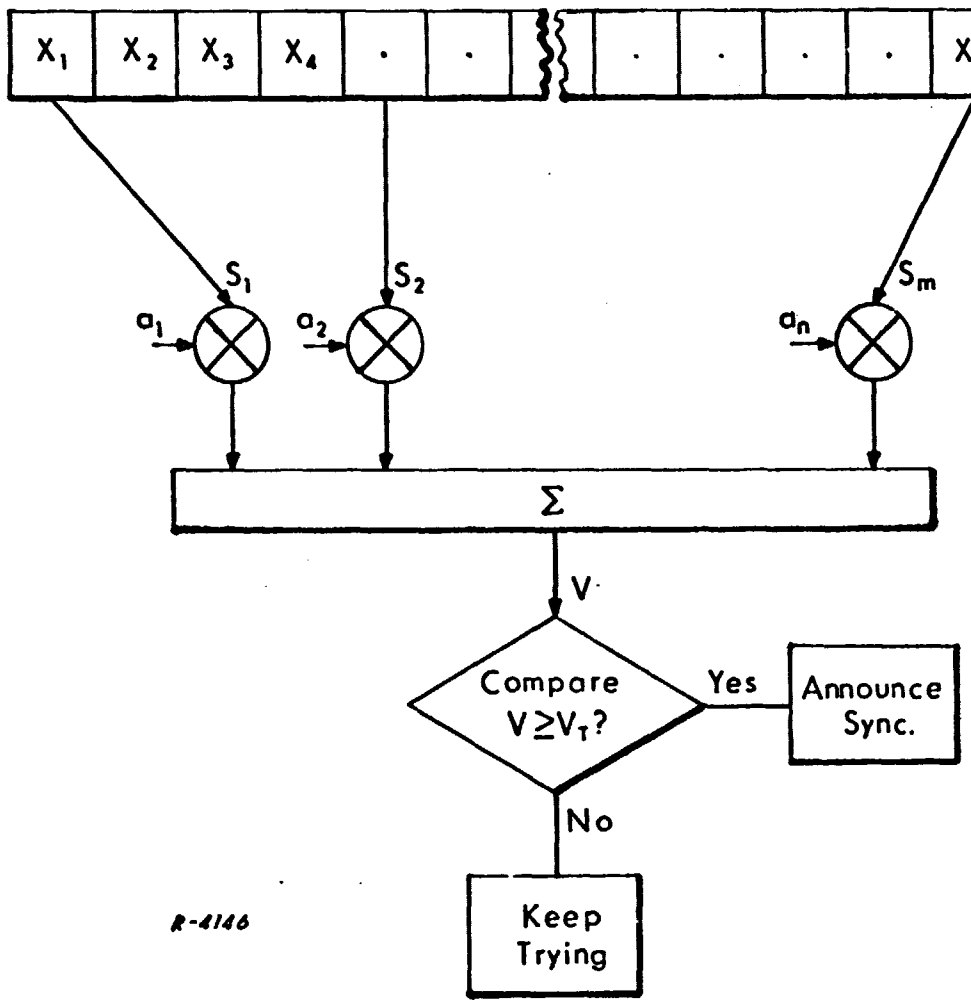


Fig. 4.39 Basic Frame Acquisition Procedure

must be considered. First, there is the random phase where the bits in the test positions are merely data bits. If $\frac{n+T}{2}$ or more of the random bits match up with members of the sequence, false frame sync will result. This sort of false sync occurrence depends upon the threshold chosen and upon the sequence length but is invariant to the details of design of the sequence. Next there is the overlap region which occurs as the time for correct sync draws nearer and the test positions are filled partially with random data and partially with bits of the sync sequence. Sequence design plays a role in reducing accidental high correlations in this region. Finally there is the in-sync position at which the sync sequence, possibly perturbed by errors, fills the test positions. This time sync will fail to be identified only if $\frac{n-T}{2}$ or more errors have disrupted the sequence, independent of the actual design details of the sequence.

The procedure in designing frame sync sequences is thus to first select a sequence length consistent both with the number of bit positions available and the tolerable amount of complexity, and then to choose among the 2^n possible sequences the one which minimizes the occurrence of a false high correlation in the overlap region. Once the sequence is selected the threshold value sets the tradeoff relationship between the occurrences of false frame sync and missed sync.

In order to clarify the above comments by way of example, observe that the frame sync sequence for the NASA BCD time codes consists merely of five consecutive 1's. There is no choice of thresholds, since one must insist upon five-out-of-five agreements to avoid false sync on a BCD "7". There is also a serious overlap region problem. Note that in the 36 bit BCD code (Fig. 4.39) that if reception is first picked up on the second bit of the reference marker, or if reception is picked up earlier but the first '1' gets decoded as a '0', then if the units seconds digit is an odd number, a false acquisition one bit position away from the correct one will be formed.

5. SUMMARY OF PRINCIPAL CONCLUSIONS

The question of optimum time code system has been divided in this report into

(a) basic characterization and encoding of the timing information,

(b) basic characterization of the available transmission media and the assessment of their limitations, and

(c) signal design for representing the desired timing information in the manner most immune to transmission distortion and disturbances; specifically,

(i) the selection of the waveform representation of the time code sequence, and

(ii) the selection of the modulation method for taking full advantage of available signal power, bandwidth and transmission channel characteristics.

In addition to the essential material presented in this report on the above aspects of the complete time code system problem, a number of significant conclusions have been reached. These are:

— The major transmission hazard with all channels considered is signal drop-out, or "outage", due to fading. Otherwise, the coherence bandwidths of tropospheric line-of-sight and beyond-the-horizon channels are adequate for all currently conceivable time code transmission requirements, but the frequency characteristics of wireline channels make baseband transmission of time codes by any of the many possible rectangular waveform representations simply not feasible. Transmission over wireline channels requires the use of special subcarrier modulation in order to situate the time code spectrum optimally within the transmission band of the wireline channel. Channel fading rates are generally low, but may well cause transmission degradation if the pulse widths are too long.

— Many interesting rectangular waveform representations exist that are suitable for binary time codes. Among the most promising from the viewpoint of simplicity and performance under conditions of limited available bandwidth, limited peak power and limited time marker transmission interval is the class of pseudo-random sequences.

— The current NASA and IRIG time codes do not represent optimum code and signal design within the existing limitations on signal power, bandwidth and marker transmission interval.

— Improvements are possible through appropriate code and waveform design.

— Error correction coding is not necessary.

— With PDM, PPM and PAM, jitter caused by additive noise and by low-frequency cutoff effects is of the order of microseconds for (S/N) ratios in excess of 20 dB. The low-frequency cutoff effects are somewhat larger for PDM than for PAM and PPM.

— For PAM, PDM and PPM, with a peak-power constraint, the average probability of bit error due to additive white gaussian noise is inferior to the performance of an optimum coherent antipodal system by the equivalent of more than 6 dB in S/N ratio. For a given bandwidth, PDM is inferior to PAM and PPM.

The principal questions for the remainder of the present program are:

(a) the integration of the total background material presented in this report and other results established in previous studies to evolve a practicable optimum time code system concept; and

(b) the comparison of the performance of the optimum time code system with the performance of the existing NASA and IRIG systems in order to determine the margin of improvement possible over the existing systems.

REFERENCES

1. A.A. Alexander, R.M. Gryb, and D.W. Nast, "Capabilities of the Telephone Network for Data Transmission, "Bell Sys. Tech. Jour., Vol. 39, pps. 431-476, May 1960.

2. R.L. Townsend, and R.N. Watts, "Effectiveness of Error Control in Data Communication Over the Switched Telephone Network," B.S.T.J., Vol. 43, pps. 2611-2638; November, 1964.

3. H.L. Yudkin, "Some Results in the Measurement of Impulse Noise on Several Telephone Circuits," Proceedings of the National Electronics Conference, Vol. 16, pps. 222-231, 1960.

4. D.E. Kerr, Propagation of Short Radio Waves, Rad. Lab. Series, Vol. 13, McGraw Hill Book Co., New York, 1951.

APPENDIX A

Figures A.1 – A.10

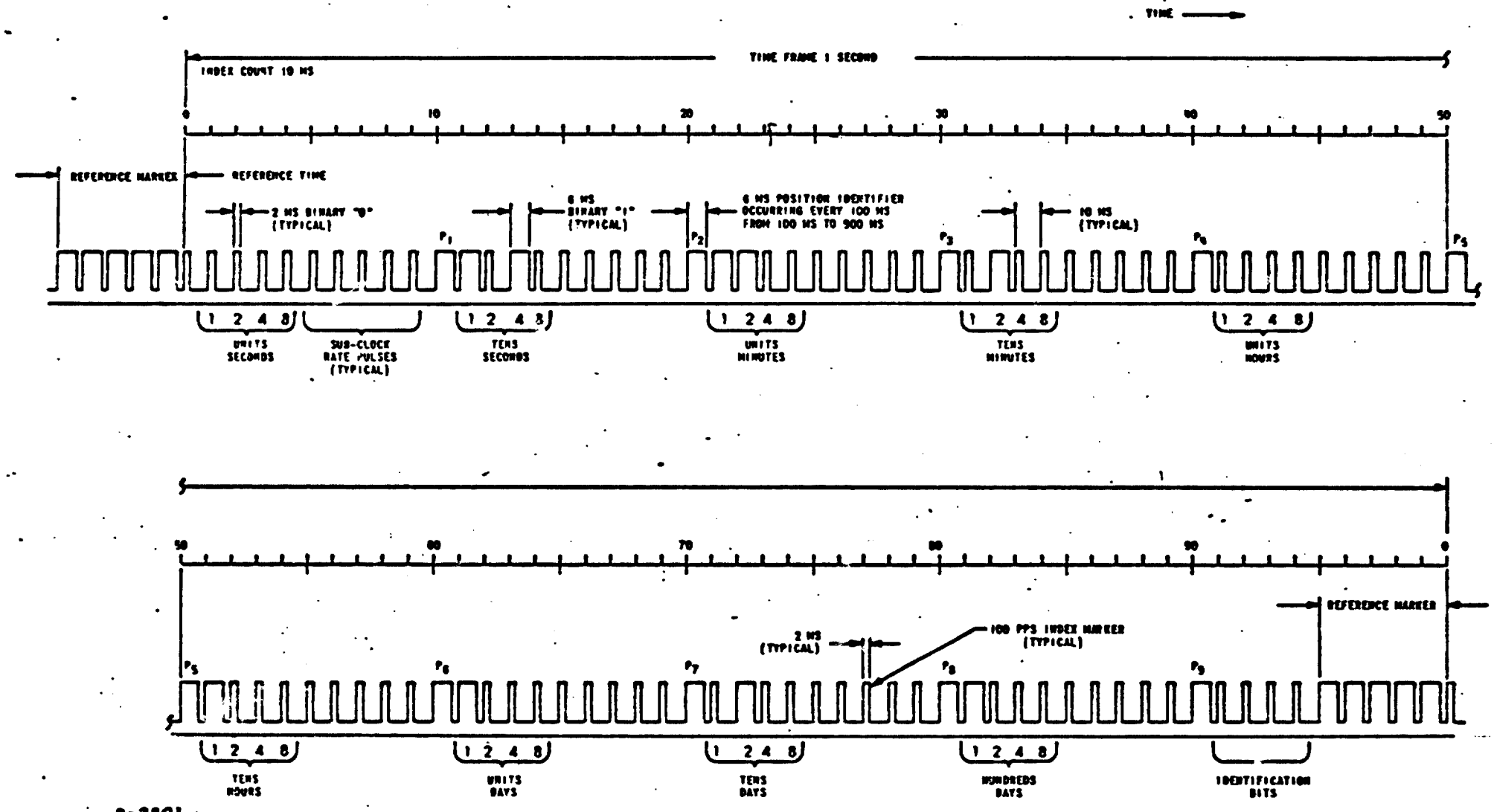


Fig. A.1 NASA 36-Bit BCD Time Code Carrier Level Shift, F_c = 1 kHz.

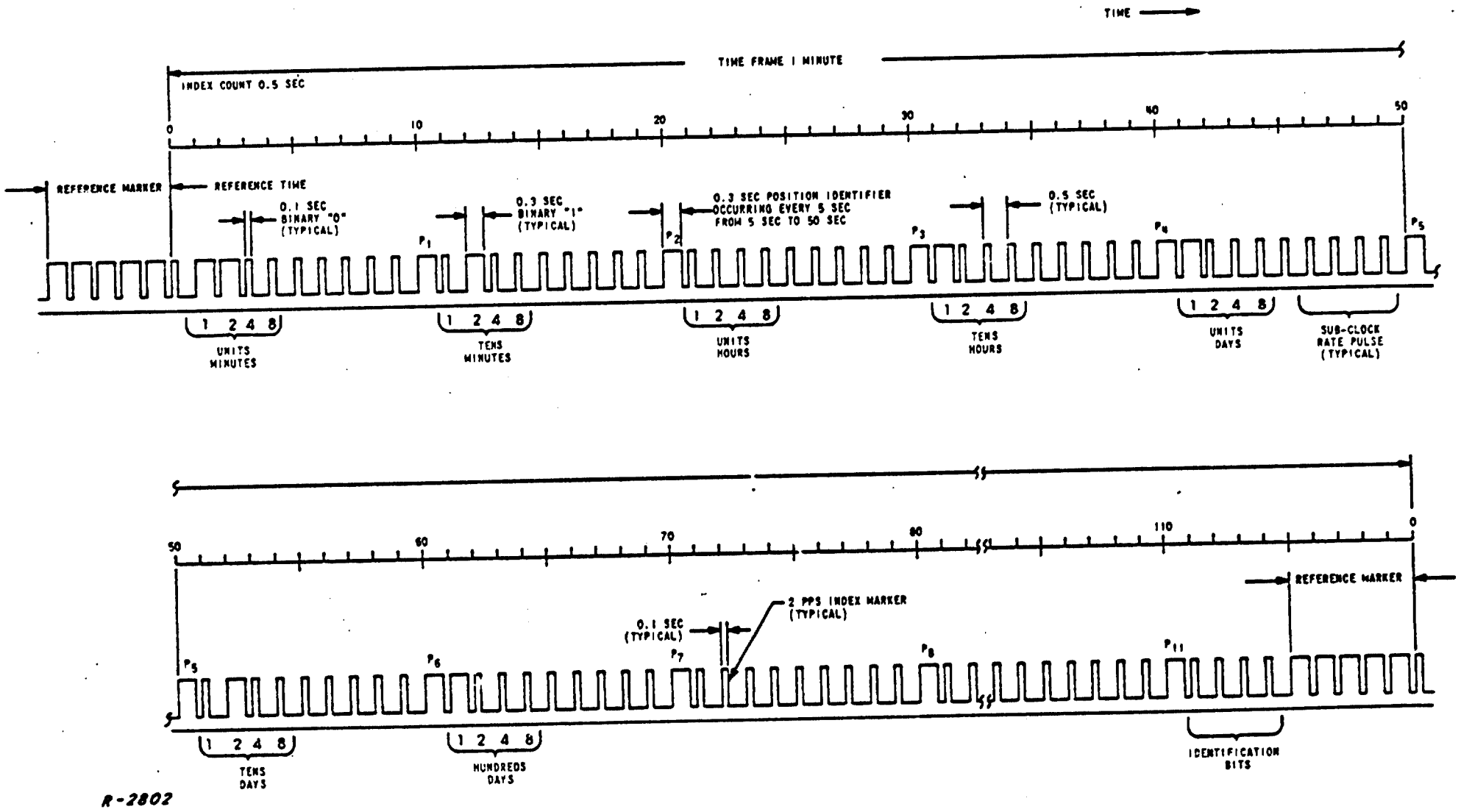
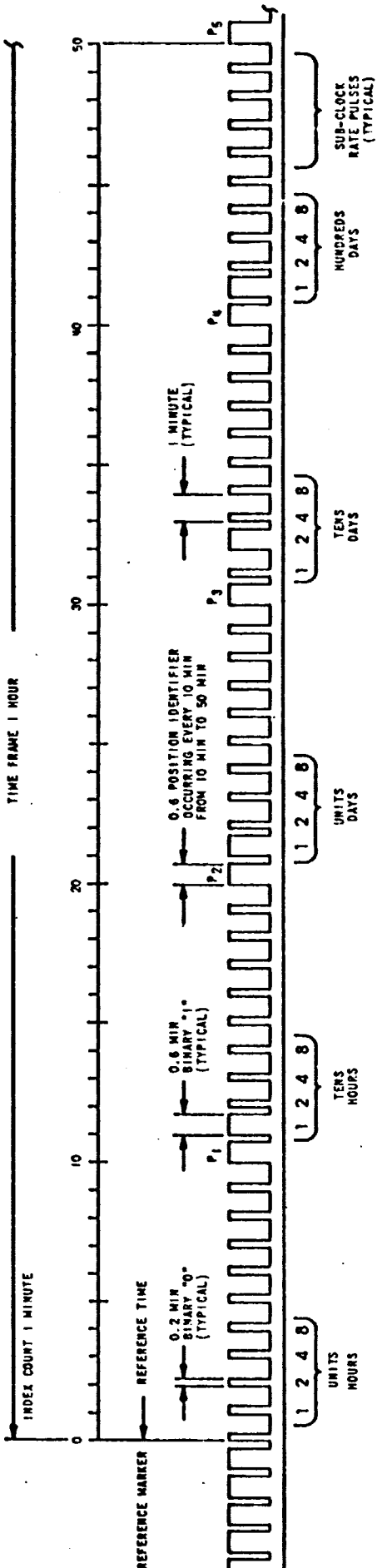


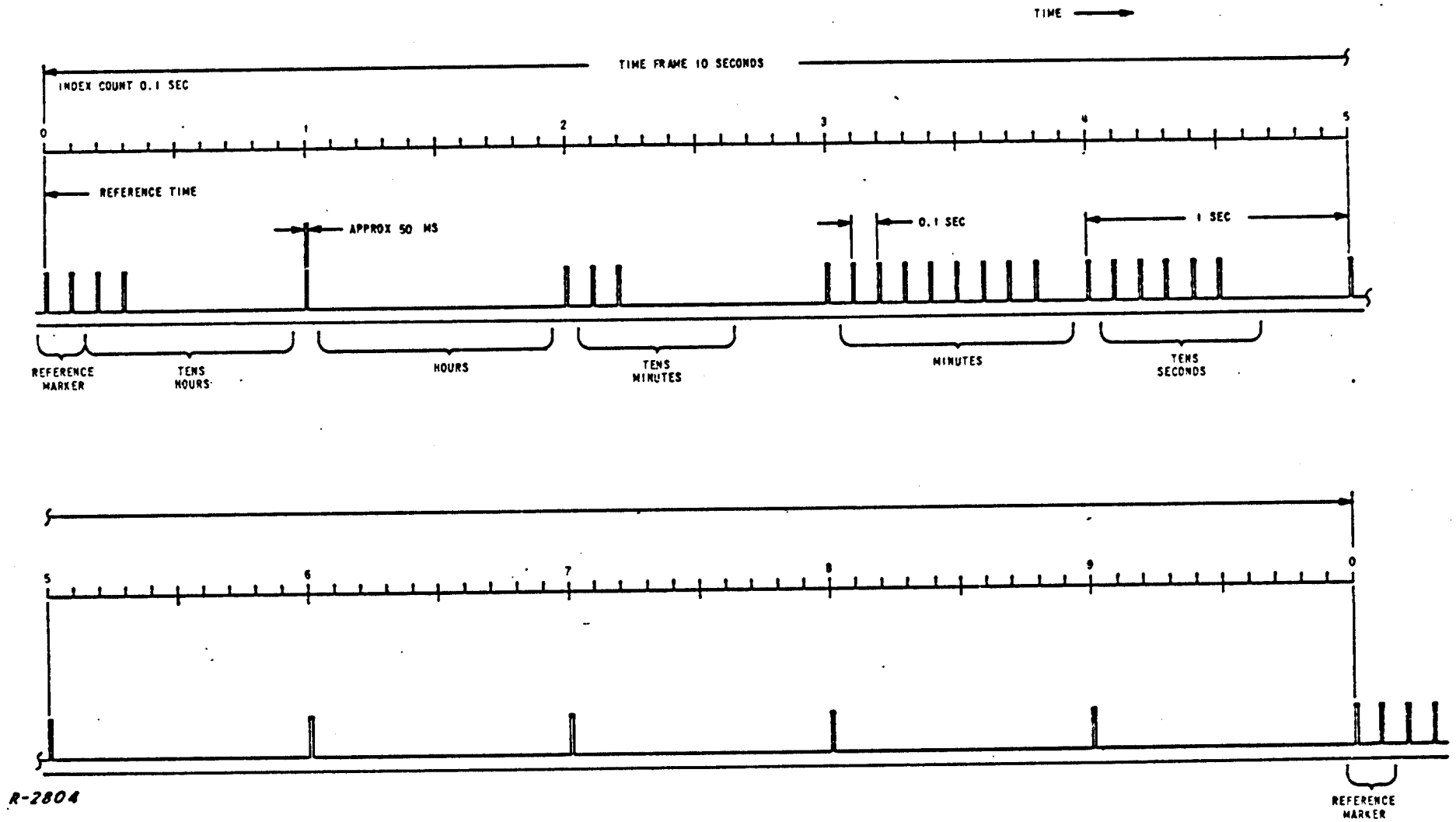
Fig. A.2 NASA 28-Bit BCD Time Code Carrier Level Shift, F_c = 100 Hz.



TIME AT REFERENCE: 121 DAYS, 10 HOURS

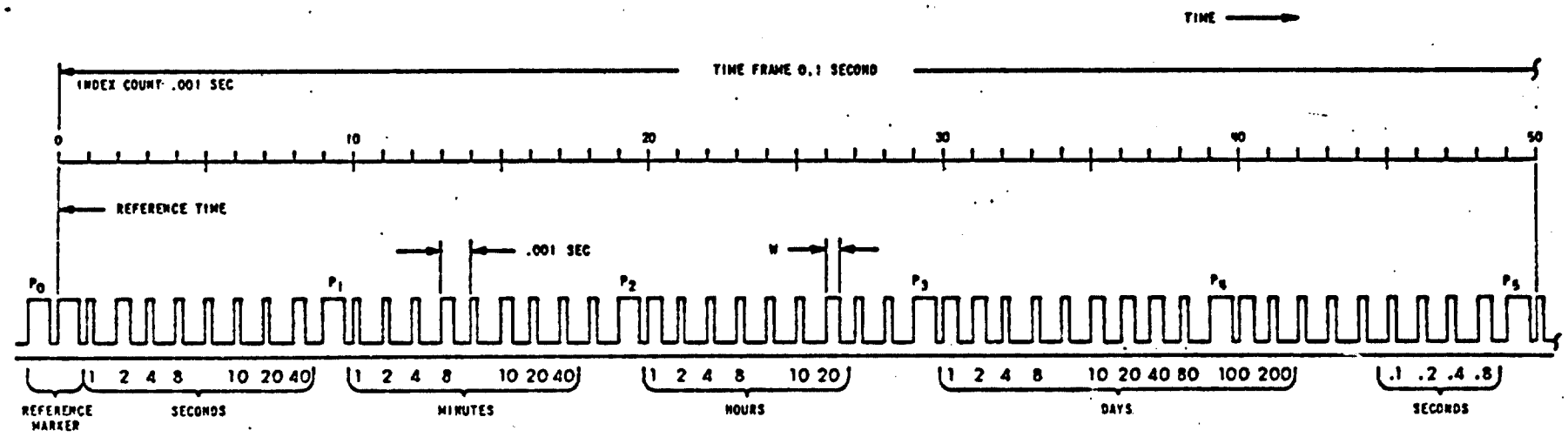
r Level Shift, F_c = 100 Hz.

Fig. A.3 NASA 20-Bit BCD Time Code Carrier L



TIME AT REFERENCE: 20HR .., 28MIN .., 50SEC

Fig. A.4 NASA Serial Decimal Time Code dc Level Shift.

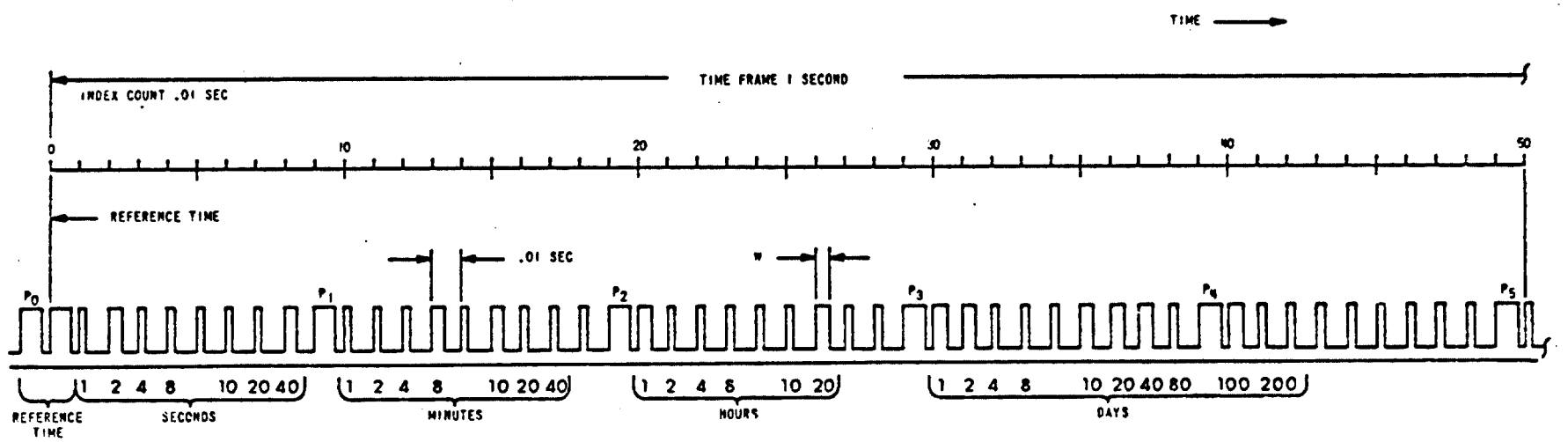


P=POSITION IDENTIFIER, 0.8MS DURATION
W=WEIGHTED CODE DIGIT, 0.5MS DURATION
C=CONTROL ELEMENT (EXAMPLE) 0.5MS DURATION
DURATION OF INDEX MARKERS = 0.2MS

R-2805

TIME AT POINT A: 21:18:42+.8+.07+.005
= 21 HR., 18 MIN., 42.875 SEC ON DAY 173

Fig. A.5 IRIG Standard Format A Carrier Level Shift, F_c = 10 kHz.

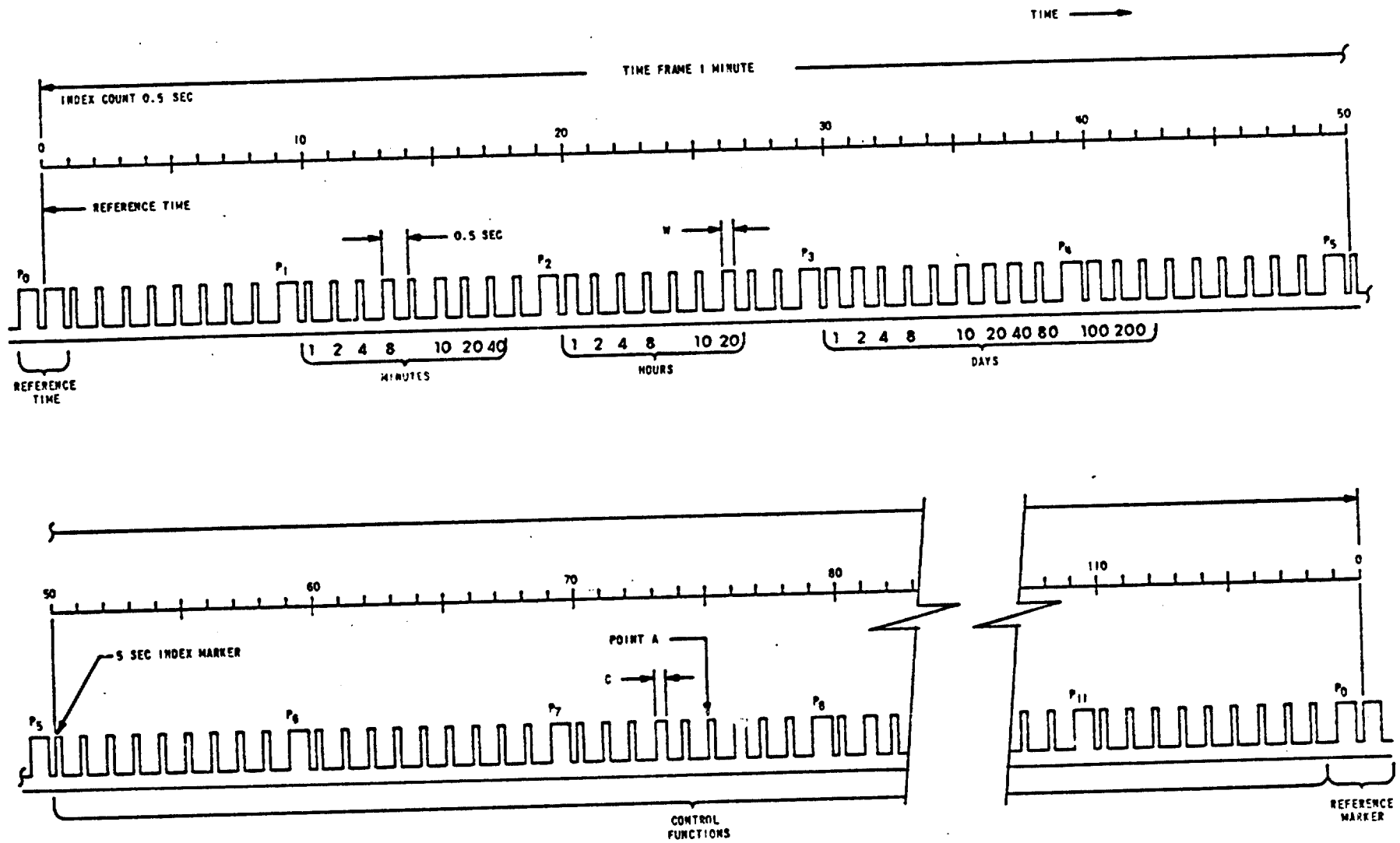


P = POSITION IDENTIFIER, 8MS DURATION
W = WEIGHTED CODE DIGIT, 5MS DURATION
C = CONTROL ELEMENT (EXAMPLE), 5MS DURATION.
DURATION OF INDEX MARKERS, 2MS

R-2806

TIME AT POINT A = 21:18:42 +.7 +.05
= 21 HR., 18 MIN., 42.75 SEC. ON DAY 173

Fig. A.6 IRIG Standard Format B Carrier Level Shift, F_c = 1 kHz.

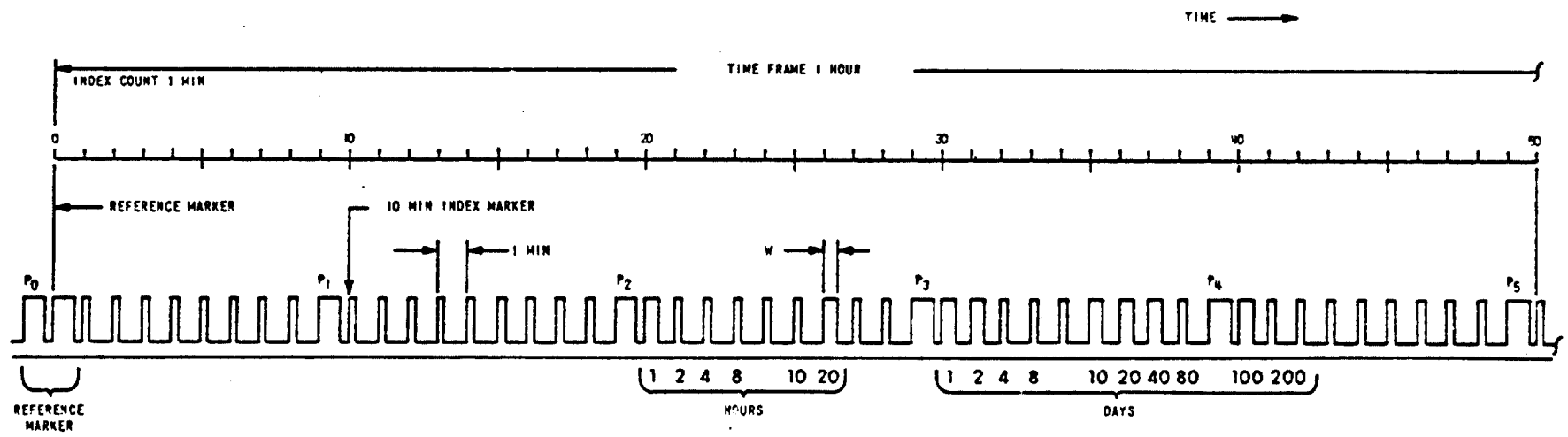


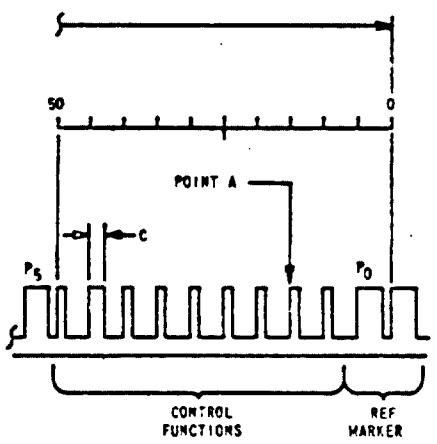
TIME AT POINT A = 21:18 + 37.5 SEC
= 21 HR., 18 MIN., 37.5 SEC ON DAY 173

P=POSITION IDENTIFIER, 0.8SEC DURATION
W=WEIGHTED CODE DIGIT, 0.25SEC DURATION
C=CONTROL ELEMENT (EXAMPLE), 0.25SEC DURATION
DURATION OF INDEX MARKERS = 0.1SEC

R-2807

Fig. A.7 IRIG Standard Format C Carrier Level Shift, F_c = 100 Hz or 1 kHz.



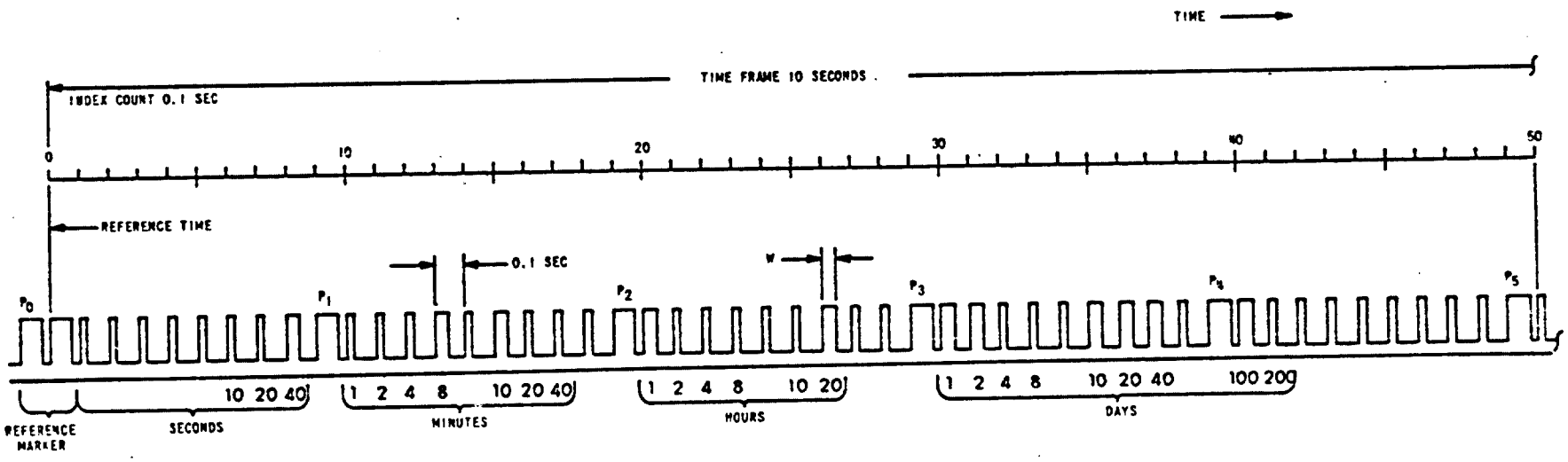


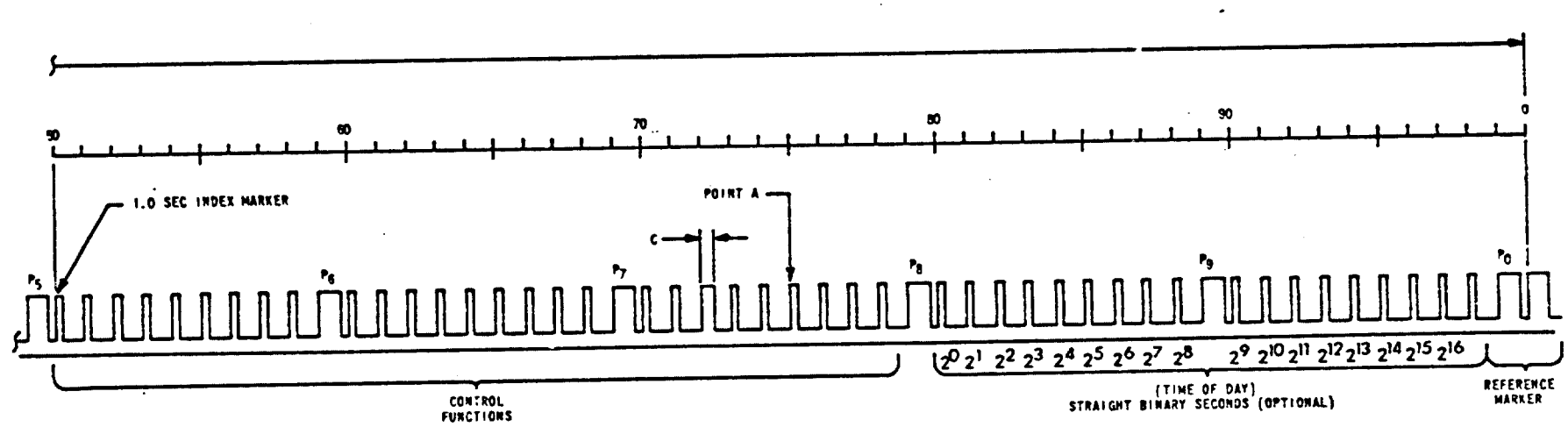
P=POSITION IDENTIFIER. 4BSEC DURATION
W=WEIGHTED CODE DIGIT. 30SEC DURATION.
C=CONTROL ELEMENT (EXAMPLE). 30SEC DURATION
DURATION OF INDEX MARKERS = 12SEC

R-2808

TIME AT POINT A = 21HR. 57MIN.
ON DAY 173

Fig. A.8 IRIG Standard Format D Carrier Level Shift, F_c = 100 Hz or 1 kHz.



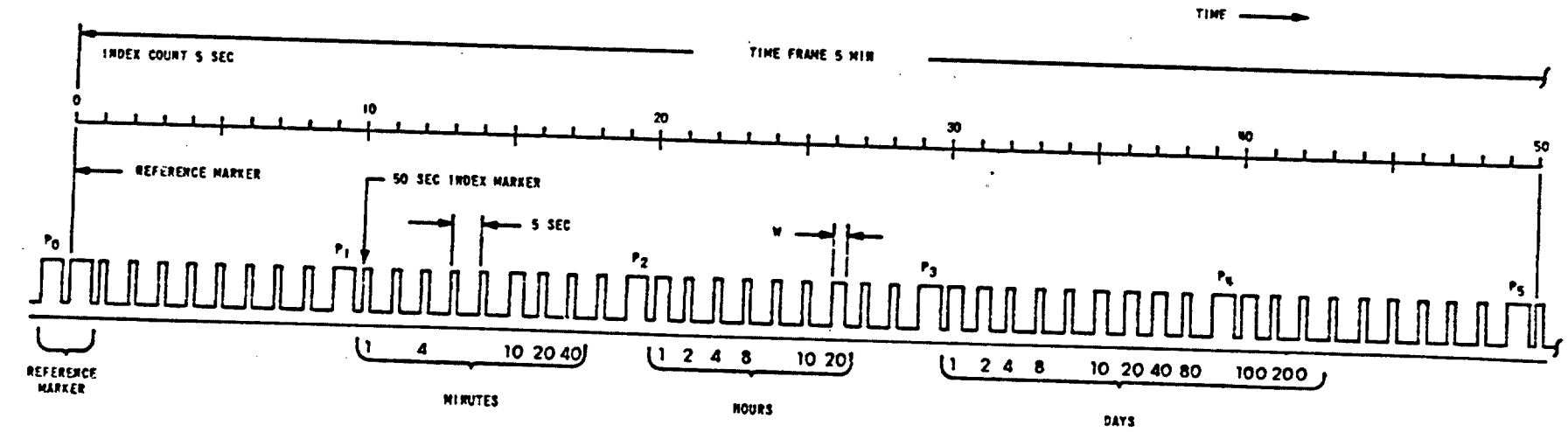


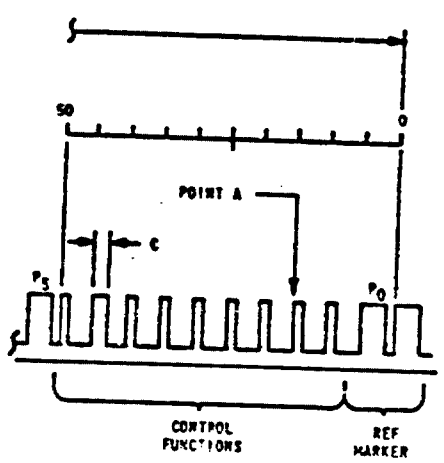
P=POSITION IDENTIFIERS. 8OMS DURATION
w=WEIGHTED CODE DIGIT. 5OMS DURATION
C=CONTROL ELEMENT (EXAMPLE). 5OMS DURATION
DURATION OF INDEX MARKERS=2OMS

R-2809

TIME AT POINT A = 21:18:40+7.5
= 21 HR., 18 MIN., 47.5 SEC ON DAY 173

Fig. A.9 IRIG Standard Format E Carrier Level Shift, F_c = 100 Hz or 1 kHz.





P=POSITION IDENTIFIER, .8 SEC DURATION
W=WEIGHTED CODE DIGIT, 2.5 SEC DURATION
C=CONTROL ELEMENT (EXAMPLE), 30 SEC DURATION
DURATION OF INDEX MARKERS = 1 SEC

TIME AT POINT A = 21 HR. 16 MIN. 45 SEC
ON DAY 173

R-2810

Fig. A.10 IRIG Standard Format F Carrier Level Shift, F_c = 100 Hz.

Appendix B

A so-called maximum likelihood decision may be performed upon the data sequence $\vec{r}$ which is known to be a noisy replica of either of the two messages '0' or '1' by forming the ratio of conditional probability densities and comparing it to a threshold separating the decision zone for a '1' from that for a '0':

$$L(\vec{r}) \equiv \frac{p(\vec{r} \text{ received}/\text{ '1' sent})}{p(\vec{r} \text{ received}/\text{ '0' sent})} \underset{\text{'0'}}{\overset{\text{'1'}}{\gtrless}} T \qquad \text{(B-1)}$$

Thus, if $L(\vec{r}) \geq T$ the decision is '1'; if it is $< T$, the decision is '0'. The maximum likelihood decision criterion is superior to all other criteria in the following sense: for some specified probability of erroneously deciding '0' when '1' was really sent the maximum likelihood procedure minimizes the probability of erroneously deciding '1' when '0' was in reality sent. The optimality of the maximum likelihood decision criterion may be stated in an alternate manner. If separate costs, or penalties, are assigned to the making of a '0' - '1' or a '1' - '0' error, then the operation upon the data stream which minimizes the cost of decision-making must be the maximum likelihood decision process. In particular, when '0' and '1' are equiprobable, the maximum likelihood decider will cause decision errors to be symmetrical, and will provide the minimum possible error probability at its output.

If one performs a maximum likelihood decision upon pulse-duration-modulated bits such as are found in NASA and IRIG codes, the time that the signals have in common contributes no useful information to the decision, and should be ignored. This is illustrated graphically in Fig. B-1 below. The proof follows.

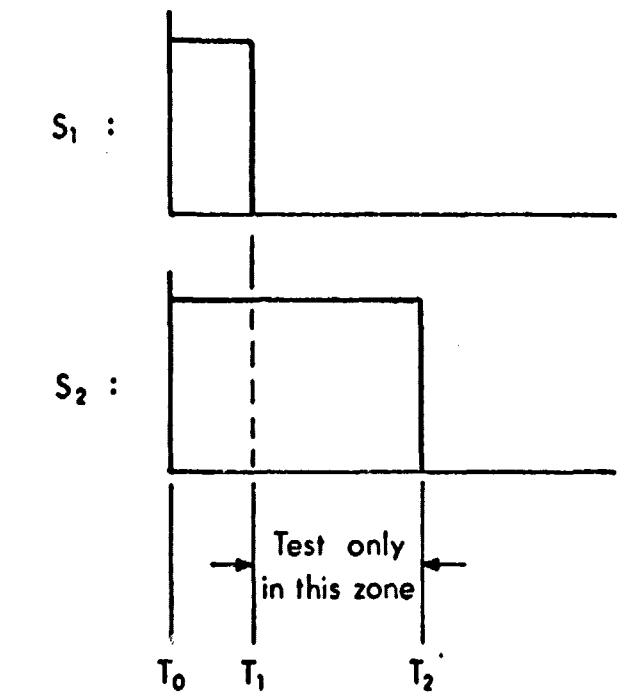


Fig. B-1 Optimum Hypothesis Testing for PDM Binary Signals.

The received data $\vec{r}$ may be broken into two segments: that part $\vec{r}$, occurring in the interval $T_0 \leq t \leq T_1$ which is common to both symbols and the segment $\vec{r}_2$ occurring during $T_1 < t \leq T_2$. Since the underlying signal in $\vec{r}_1$ is the same no matter which symbol was actually sent:

$$p(\vec{r}_1 \text{ received}/\text{'1' sent}) = p(\vec{r}_1 \text{ received}/\text{'0' sent}) = p(\vec{r}_1) \qquad \text{(B-2)}$$

The likelihood ratio may be expressed as:

$$L(\vec{r}) = \frac{p(\vec{r} \text{ received}/'1' \text{ sent})}{p(\vec{r} \text{ received}/'0' \text{ sent})} = \frac{p(\vec{r}_1, \vec{r}_2/'1')}{p(\vec{r}_1, \vec{r}_2/'0')}$$

$$= \frac{p(\vec{r}_2/'1') \cdot p(\vec{r}_1/\vec{r}_2, '1')}{p(\vec{r}_2/'0') \cdot p(\vec{r}_1/\vec{r}_2, '0')} \qquad \text{(B-3)}$$

But we have already seen in Eq. (B-2) that $\vec{r}_1$ is independent of which signal was actually sent. Thus:

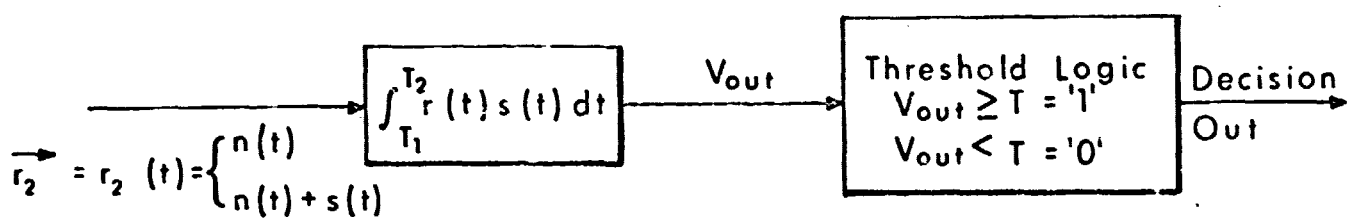
$$L(\vec{r}) = \frac{p(\vec{r}_2/'1')}{p(\vec{r}_2/'0')} \qquad \text{(B-4)}$$

The optimum decision criterion requires using only the data segment $\vec{r}_2$.

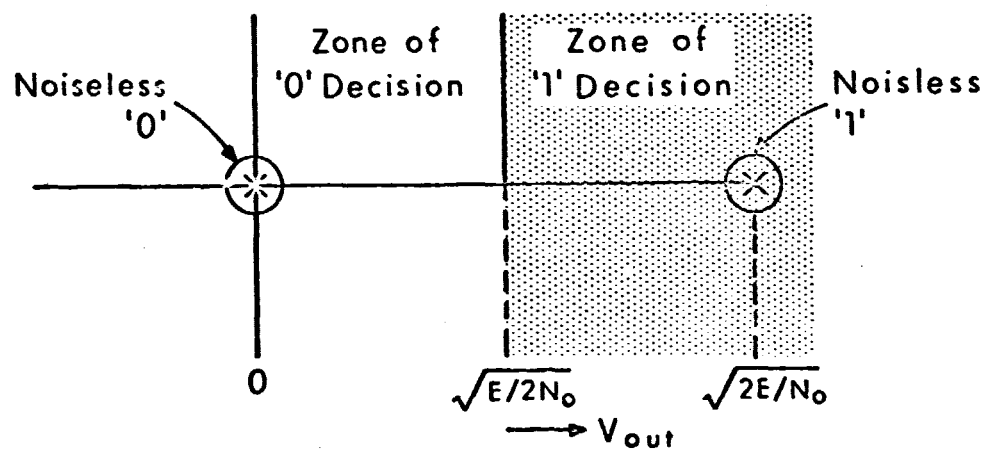
The application of Eq. (B-4) to the case of pulse-duration-modulated equiprobable signals perturbed by additive white gaussian noise with spectral density N_o watts/cycle (white noise has this spectral density over all frequencies; noise which is wideband relative to the signal is usually well approximated by white noise) leads to the correlation detection scheme of Fig. B-2a. Figure B-2b shows a graphical interpretation of the decision rule. In the absence of noise, the output submitted to threshold logic is either zero or $\sqrt{2E/N_o}$ volts, where

$$E = S(T_2 - T_1) \qquad \text{(B-5)}$$

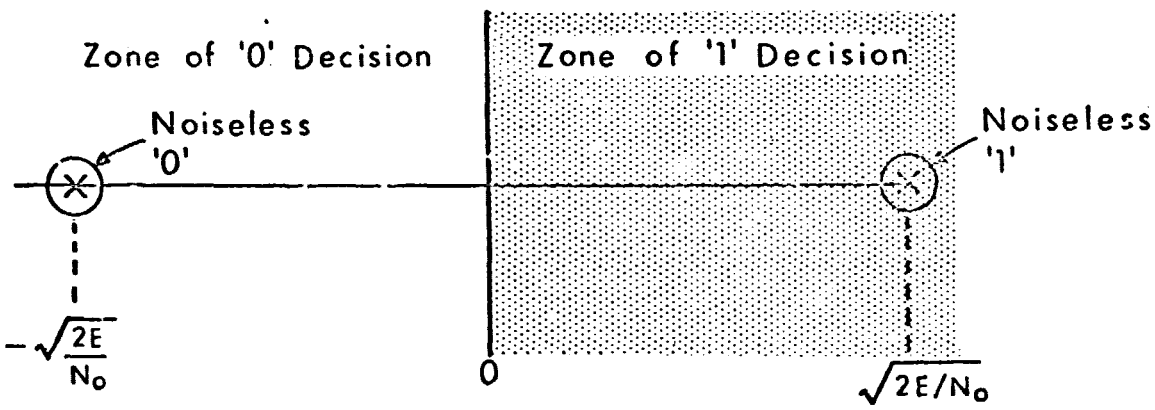
S being the average power in the signal



a) Maximum Likehood Procedure For PDM Signals.



b) Geometric Representation of Decision Zones — PDM Signals



c) Geometric Representation of Decision Zones - Binary Antipodal Signals

R-2791

Fig. B-2 Maximum Likelihood Decision Procedures.

The output is displaced from its true value by a gaussian voltage with unit standard deviation. The decision boundary is symmetrically placed midway between these points, and the probability of error, which is the same no matter which symbol was actually transmitted, is simply the probability that a unit gaussian variable exceeds $(2\sqrt{2E/N_o})^{-1}$ volts:

$$P_E = \Phi\left(\sqrt{E/2N_o}\right) \tag{B-6}$$

where: $\Phi(x) \equiv \int_x^\infty \frac{1}{\sqrt{2\pi}} e^{-y^2/2}\, dy$

The optimum binary signal design consists of antipodal signaling i.e., '1' is denoted by f(t), and '0' by -f(t). Figure B-2c shows the greater separation in the decision space which obtains from making one signal the negative of the other. The error probability is:

$$P_E = \Phi\left(\sqrt{\frac{2E}{N_o}}\right) \tag{B-7}$$

Appendix C

A NOTE ON THE SPECTRAL CHARACTERISTICS OF CARRIER LEVEL SHIFT TIME CODES

The NASA and IRIG carrier level shift time codes are composed of pulse trains built up of periodic bursts of carrier. In the evaluation of the spectral characteristics of these time codes, one should realize that the duration-modulated character of the pulses causes the regions of the frequency spectrum containing pulse energy to in general be subject to change with the occurrence of every index marker. It is therefore more meaningful to investigate the spectral content of the component pulses rather than attempt the formidable task of analyzing the spectral composition of the code averaged over many frames. In addition, one may always extend results describing system response to single pulse to describe system response to multiple pulses by superposition.

The spectrum of the shortest pulse in the code which we are considering to be indicative of the spectral requirements of the code, has already been derived.*

For a pulse consisting of n cycles of sinusoid of frequency f_c beginning and ending with 0^o phase, the energy density spectrum (scaled for unity peak amplitude) is

$$S_1(f) = \left[\text{sinc}\left[n\pi(1 - f/f_c)\right] - \text{sinc}\left[n\pi(1 + f/f_c)\right]\right]^2 \qquad \text{(C-1)}$$

where

$$\text{sinc } x \equiv \sin x/x$$

The two terms in Eq. (C-1) may be combined to form a single term:

$$S_1(f) = \frac{4}{(1+x)^2}\,\text{sinc}^2\left[n\pi(1 - x)\right] \qquad x \equiv f/f_c \qquad \text{(C-2)}$$

---

* G-57, Task VII, Final Report, Appendix A

Or in terms of the pulse duration τ:

$$S_1(f) = \frac{4}{[(1 + f/f_c)]^2} \operatorname{sinc}^2[(f_c - f)\tau/2] \qquad (C-3)$$

The usual 'narrowband' approximation to a pulse of sinusoid is:

$$S_1(f) \approx \operatorname{sinc}^2[n\pi(1 - x)] \qquad (C-4)$$

It may be noted that Eq. (C-3) differs from the usual approximation for the spectrum of a 'narrowband' ($n \to \infty$) pulse of sinusoid by virtue of the multiplicative term $4/[1 + (f/f_c)]^2$ out in front. For a narrowband pulse the lobes of the sinc^2 function virtually fade away to nothing by the time f/f_c is significantly different from 1.

Qualitatively speaking, the exact expression differs from the sinc^2 approximation in that the sidelobes below the main lobe have higher energy content and those above have higher content. The main lobe will assume a skewed energy distribution with more energy below f_c than above.

Figure C-1 compares exact and approximate spectra for the short pulse of the NASA BCD and IRIG carrier level shift time codes (n = 2). The main lobe (500 Hz to 1500 Hz for the NASA 36-bit BCD code) carries slightly more than the 90 percent of total energy carried in the main lobe of the sinc^2 approximation for a narrowband pulse. This may be seen by observing that for $f < f_c$, we have $x < 1$, and $4/(1+x)^2 < 1$. Consequently, for all $f < f_c$, $S_1(f)$ will be greater than the approximation, and for $f > f_c$, $S_1(f)$ will be less than the approximation. As a result, those lobes lying below the main lobe in frequency will have more energy than the corresponding lobes generated by Eq. (C-4), and those above will have less energy.

The main lobe assumes a skewed energy distribution with more energy in its lower half than in its upper half. The total energy in the main lobe is slightly greater than in the approximating sinc^2 function. This may be proven by noting that the heights of the approximate $S_1(f)$ curve given by Eq. (C-4) at the

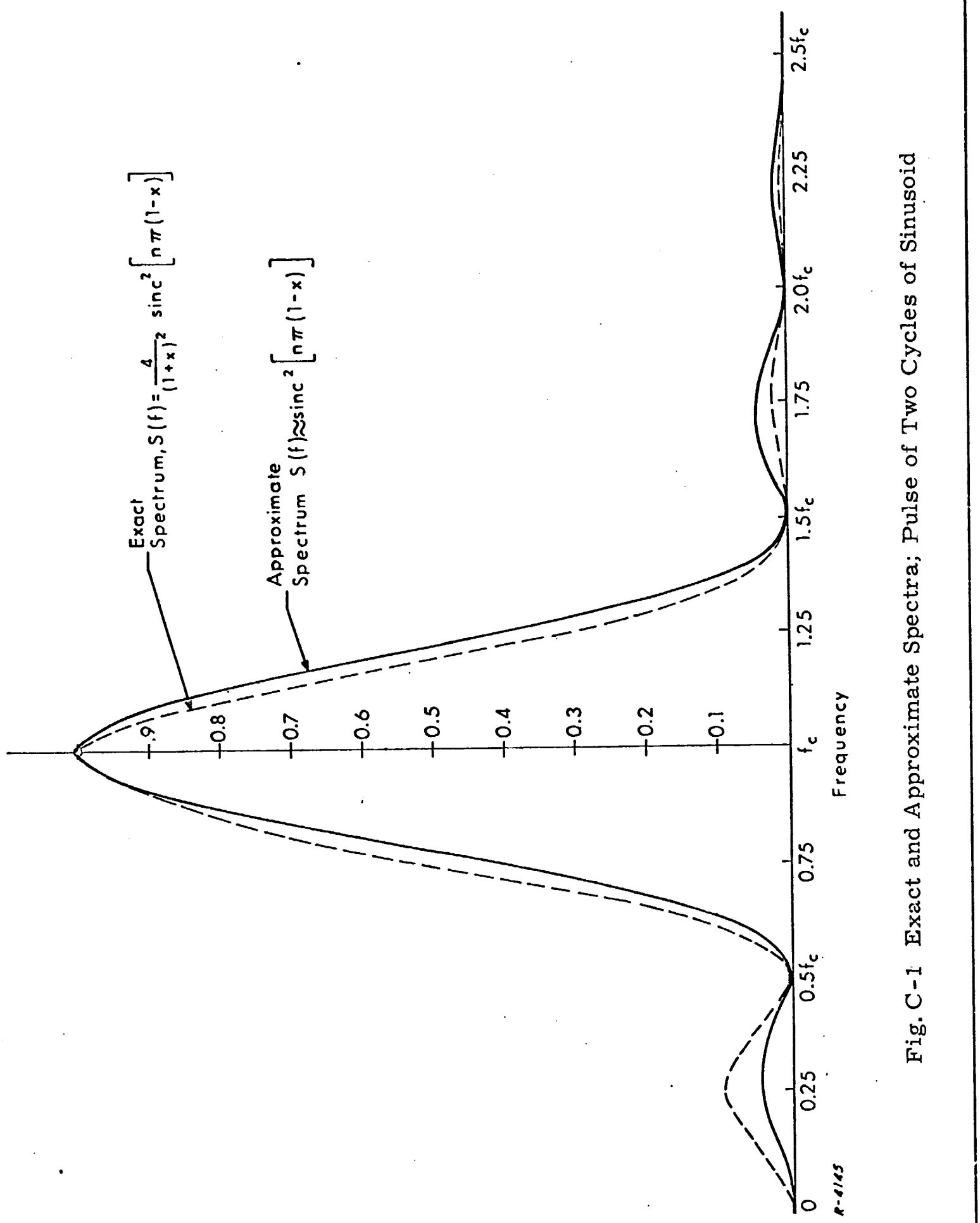


Fig. C-1 Exact and Approximate Spectra; Pulse of Two Cycles of Sinusoid

points $x = 1+\alpha$ and $x = 1-\alpha$ are equal and equal to $\text{sinc}^2 n\pi\alpha$. Thus, infinitesimal bands of frequency at these points contribute energy proportional to $2\,\text{sinc}^2 n\pi\alpha$. The exact $S_1(f)$ function yields a multiplier for $\text{sinc}^2 n\pi\alpha$ of:

$$\frac{4}{(2+\alpha)^2} + \frac{4}{(2-\alpha)^2} \quad .$$

It is easily seen that:

$$\frac{4}{(2+\alpha)^2} + \frac{4}{(2-\alpha)^2} \geq 2 \qquad \text{for} \quad 0 \leq \alpha < 1.$$