

THE UNIVERSITY OF MICHIGAN
OFFICE OF RESEARCH ADMINISTRATION
ANN ARBOR

DIGITAL COMMUNICATION STUDIES

PART I: COMPARATIVE PROBABILITY OF ERROR
AND CHANNEL CAPACITY

PART II: ASYNCHRONOUS TIME-MULTIPLEXING OF
MULTICHANNEL SPEECH SOURCES

Technical Report No. 133

4251-2-T

Cooley Electronics Laboratory
Department of Electrical Engineering

By: M. P. Ristenbatt
T. G. Birdsall
T. Felisky
S. B. Weinstein

Approved by:


H. W. Farris

ORA Project 4251

Contract No. DA-36-039 sc-87172
U. S. Army Signal Supply Agency
U. S. Army Signal Research and Development Laboratory
Fort Monmouth, New Jersey

March 1962

ACKNOWLEDGMENTS

The authors wish to acknowledge the contributions of Mr. R. A. Carlsen, who wrote Appendix B, and was responsible for the queuing theory approach of Section 3.1. Thanks is due Mr. A. J. Viterbi of Jet Propulsion Laboratory for supplying a tabulated form of computed error probabilities for coherent block codes. Credit is due to the members of the Signal Corps Multichannel Security Division for calling out attention to pertinent reports and work.

TABLE OF CONTENTS

	<u>Page</u>
ACKNOWLEDGMENTS	ii
LIST OF ILLUSTRATIONS	v
LIST OF SYMBOLS (PART I)	vii
LIST OF SYMBOLS (PART II)	ix
ABSTRACT	xi
PART I: COMPARATIVE PROBABILITY OF ERROR AND CHANNEL CAPACITY OF DIGITAL COMMUNICATIONS SYSTEMS	1
1. GENERAL CONSIDERATIONS	2
1.1 General Block Diagram	2
1.2 Source Coding	3
1.3 Error Correcting and Detecting	4
1.4 General Detection Considerations	5
2. DETECTION CONSIDERATIONS	6
2.1 Optimum Receivers	6
2.1.1 Optimum Coherent Receivers	6
2.1.2 Optimum Incoherent Receivers	7
2.2 Signal Choice	7
2.2.1 Coherent Operation	9
2.2.2 Incoherent Operation	11
2.3 Example of Signal Selection in Binary Case	12
3. PROBABILITY OF ERROR CALCULATIONS AND RESULTS FOR DIGITAL SYSTEMS	14
3.1 Binary Systems	15
3.1.1 Binary Simplex	18
3.1.2 Coherent Orthogonal	21
3.1.3 Incoherent Orthogonal	23
3.2 Nonbinary Systems	25
3.2.1 Error-Coded Binary Signals	25
3.2.2 M-Symbol Alphabet	27
3.2.2.1 General P_{ch} Calculations	27
3.2.2.2 Basic M-ary Symbols	34
3.2.2.3 Block-Coded Binary Signals	34
3.2.2.4 M-Phases	35
3.3 Comparative Results and Conclusions	38
3.3.1 Bit-Error Comparison	40
3.3.2 Word-Error Comparisons	44
3.3.3 Information-Rate Comparison	47
3.3.4 Comparison Conclusions	50
4. CHANNEL CAPACITIES AND MAXIMUM RATES	55
4.1 Capacity Relations for Digital Systems	55
4.1.1 Continuous Channels	56
4.1.2 Discrete Channels	57
4.2 Maximum Information Rates if Data Rate is Variable	62
4.2.1 The Small S/N Region	64
4.2.2 Large Signal-to-Noise Ratio	64
4.3 Conclusions	66

TABLE OF CONTENTS (Cont.)

	<u>Page</u>
PART II: ASYNCHRONOUS TIME-MULTIPLEXING OF MULTICHANNEL SPEECH SOURCES	69
1. INTRODUCTION	69
1.1 Background and Previous Work	70
2. DESCRIPTION OF ASYNCHRONOUS TIME-MULTI- PLEXING SCHEME	73
3. PERFORMANCE OF THE BUFFER	76
3.1 Queueing-Theory Approach	77
3.2 Specialized Approach	79
3.3 Freeze Out Fraction (FOF)	81
3.4 Freeze Out Runs	81
3.5 Starting Transients	84
3.6 Channel Idle Time	84
3.7 Wait Time Distribution	85
4. RECONSTRUCTION OF WAVEFORMS	90
4.1 Theoretical Sampling Relations	90
4.2 Reconstruction Methods	91
5. PERFORMANCE OF SYSTEM	93
5.1 Choice of Parameters	93
5.2 Digit Rate	94
5.3 Comparisons	95
6. CONCLUSIONS	98
APPENDIX A: DETERMINATION OF THE MAXIMUM NUMBER OF BINARY DIGITS PER SECOND THAT CAN BE TRANSMITTED OVER A CHANNEL OF BANDWIDTH W	100
APPENDIX B: THE EQUIVALENT SOURCE	108
APPENDIX C: ANALYSIS OF FREEZE OUT RUNS	119
LIST OF REFERENCES	124
DISTRIBUTION LIST	127

LIST OF ILLUSTRATIONS

<u>Figure No.</u>	<u>Title</u>	<u>Page</u>
1	General block diagram of digital communication system	2
2	Optimum coherent receivers for signals known exactly in white Gaussian noise	6
3	Two types of optimum incoherent receivers	8
4	Probability of error vs. E_b/N_0 for binary signals	13
5	(This figure omitted)	-
6	General depiction of probability of error calculation binary, symmetric decisions in terms of probability densities	18
7	Two basic receivers for binary coherent operation	19
8	Distribution diagram for simplex binary symmetric situation	20
9	Distribution diagram for the coherent orthogonal, binary symmetric situation	22
10	Sketch of probability distributions for the incoherent, orthogonal AM case	24
11	P_w versus E_b/N_0 for coherent simplex and incoherent-orthogonal error-coded binary signals	28
12	P_{ch} versus E/N_0 for M-phase CPSK and DCPSK	39
13	$P_B(M)$ versus E_b/N_0 for coherent, bi-orthogonal signals, comparing binary with M-ary	42
14	$P_B(M)$ versus E_b/N_0 for incoherent, orthogonal signals comparing binary versus M-ary	43
15	P_B versus E_b/N_0 for M-phase (CPSK and DCPSK)	45
16	P_w versus E_b/N_0 of 12-bit word for three types of <u>binary</u> transmission	46
17	P_w versus E_b/N_0 for 12-bit word with bi-orthogonal signals	48
18	P_w versus E_b/N_0 for 12-bit word with incoherent orthogonal signals	49
19	P_w versus E_b/N_0 for 12-bit word with CPSK and DCPSK	51
20	Information rate/bandwidth for M-ary bi-orthogonal signals (coherent)	52
21	Information rate bandwidth for M-ary incoherent orthogonal signals	53
22	(This figure omitted)	-

LIST OF ILLUSTRATIONS (Cont.)

<u>Figure No.</u>	<u>Title</u>	<u>Page</u>
23	C/W versus E_b/N_0 for binary simplex, coherent orthogonal, and incoherent orthogonal signals	62
24	Maximum information rate per bandwidth for binary signals	65
25	Maximum information rate per bandwidth for M-ary coherent orthogonal signals	66
26	Maximum information rate per bandwidth for M-ary incoherent orthogonal signals	67
27	Depiction of sampling for extremal coding	73
28	Block diagram of transmitter for asynchronous multiplexing of multichannel speech using extremal coding	74
29	Depiction of the probabilities α_k and β_k along a time axis	79
30	Freeze out fraction as function of input rate u and storage capacity M	82
31	Channel idle time versus buffer length	85
32	Probability density curves of waiting time in buffer	88
33	Probability density curves of deviation of waiting time in buffer from its mean value	89
34	Portion of $(\sin x)/x$ curve suitable for maximum rate transmission	104
35	More practical curve ($2TW=4$) for maximum rate transmission	105
36	Buffering in an asynchronous multiplexed communication system	109
37	Notation for $f(x/\xi)$	110
38	Notation for $f(\xi)$	110
39	Notation for $P'(k, \tau/\xi)$	112
40	Notation for $P(k, \tau/\xi_1, \xi_2 \dots \xi_N)$	112

LIST OF SYMBOLS FOR PART I

In General:

x	=	input
y	=	output
(1)(0)	=	binary states
C	=	channel capacity in bits/sec---both continuous and discrete; maximum information rate at which one can send information over a channel and still reduce the P_e to 0 by external coding means
D	=	vector space dimensions per symbol time T , for a given signalling method; also called minimum "signal dimensions"
D'	=	vector space dimensions of a space corresponding to an en- semble of successive symbols which occur in time T'
E	=	energy per symbol
E_b	=	energy per bit
I	=	the information rate of digital systems in bits/sec
I'	=	the information rate of digital systems in bits/symbol
I_M	=	the maximum information rate for digital systems, when taking <u>both</u> P_e and bandwidth into account
M	=	number of alternative signals or symbols = alphabet size
N	=	number of successive symbols which can be placed in a time T' , and theoretically have no intersymbol influence
N_o	=	single-sided noise power per unit bandwidth = N/W
P_e	=	general term referring to probability of "making an error", or probability of confusing one symbol for another
P_B	=	probability of <u>bit</u> error
$P_{ch}(M)$	=	probability of error for an M -ary symbol or character
P_w	=	probability of <u>word</u> error, where a "word" refers to a group of bits; word length is determined by source data
P_{eT}	=	total probability of error---after corrective coding, occurs only when data is in bits.

LIST OF SYMBOLS FOR PART I (Cont.)

P_{ij}	=	transition probabilities of signals, between input and output of channel, due to noise
$p_i(y_j)$	=	<u>probability density</u> of y_j ; (output from j th filter) given that signal i is sent
$P(1 0)$	=	conditional probability of 1, conditional on 0
$s(t)$	=	a signal time waveform
S	=	a signal's average (or rms) power. Thus, variance of a signal
$S(f)$	=	Fourier transform of $s(t)$
T	=	length of baud in digital (alternative) transmission = length of any alphabet symbol
T_b	=	length of bit or equivalent bit length
y	=	y_M or $y_{M/2}$ = output of <u>desired</u> matched filter in P_e calculations
y^*	=	normalized y
ρ_{ij}	=	time correlation between $s_i(t)$ and $s_j(t)$
ψ	=	cumulative Gaussian function
Φ	=	particular error function = $\frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-t^2/2} dt$

LIST OF SYMBOLS FOR PART II

α_k	=	probability that exactly k buffer stages are occupied just after a buffer clock pulse
β_k	=	probability that exactly k buffer stages are occupied just before a buffer clock pulse
FOF	=	freeze out fraction = the percent of total samples from all of the sources which are rejected because the buffer is full when they arrive
M	=	capacity of the buffer in "number of samples"
N	=	the total number of sources for the multichannel situation
$p(k, u)$	=	the probability of k samples among the N sources in a time interval τ
$P(0)$	=	a column vector of initial state probabilities
$P_n(s\tau)$	=	the probability of n items in buffer (nth state) at time $s\tau$
P_n	=	state probabilities just before the occurrence of a buffer clock pulse
$P_r(c)$	=	the probability of a run of exactly c clock intervals, during which one or more samples from some source are frozen out during each interval
$P(d f \geq 1)$	=	probability that exactly d units are frozen out, in a given interval, provided that one or more was frozen out in a previous interval
$P_\theta(k)$	=	probability that k samples are in the buffer immediately before the arrival of the (k+1)th sample at time $\theta\tau$
$P_\theta(k \text{get in})$	=	the probability of a sample encountering k samples already in the buffer when it enters at time τ , if it is not rejected as freeze out
Q	=	the number of quantizing levels for the speech samples
R	=	the digit rate in bits/sec
R'	=	the average rate of occurrence of samples from an <u>active</u> speech source

LIST OF SYMBOLS FOR PART II (Cont.)

R''	=	the average rate of samples from a speech source when one takes into account the percent of "off" time of the source
Δt_M	=	the maximum amount of time displacement of samples which may occur before undesirable properties in the reconstructed speech occur
$[T]$	=	the matrix of transitional probabilities between "states," and can be found using a state diagram
τ	=	the time interval at which the samples are regularly removed from the buffer, if available
u	=	the average rate (taken across sources) of occurrence of samples in a time τ
w	=	the wait time for the $(k+1)$ th sample, which arrives at time $\theta \tau$

ABSTRACT

Part I is an organized presentation of the theoretical advantages and disadvantages of alternative modulation and coding methods for digital communications. White Gaussian noise with optimum receivers is assumed. The signals considered are: simplex, bi-orthogonal, coherent-orthogonal, incoherent-orthogonal, and M-phase. The probability of error and bandwidths for existing and contemplated signals of these types are compared. Both a constant and a variable data rate comparison are made.

The data for these comparisons were obtained from the literature and from calculations. Graphs and tables are used to depict the comparisons.

M-ary bi-orthogonal signals are deemed best for low signal-to-noise ratios if probability of error is stringent. Binary and M-phase signals appear best for medium and large signal-to-noise ratios.

This treatment should help structure the selection of modulation and coding methods for proposed communication systems. Also, since many signalling methods are considered, some fundamental principles, terminology, and methods of comparing are suggested.

Part II is an analytical evaluation of asynchronous time-multiplexing for multi-channel speech signals. This asynchronous system applies where a number of speech sources participate in a common channel or trunk, and it is proposed that the source-to-channel sample assignment be asynchronous. The analysis here includes the performance of the buffer and a predicted digit rate for the asynchronous operation.

The buffer analysis was accomplished by using a specialized probabilistic approach, which may be regarded as a special case of the more general queuing theory. The digit rate analysis assumes the use of extremal coding on the speech source. Based on the parameters assumed for the proposed asynchronous system, it is predicted that a multi-channel situation consisting of 100 channels will use only one third of the digit rate of synchronously multiplexed PCM, and about 5 db loss in quantization signal-to-noise ratio can be expected.

Successful operation of this system would mean a more efficient use of the channel and possibility of temporary overloads. Based on the analysis here it is felt that the theoretical advantages of asynchronous time-multiplexing are sufficiently profitable to warrant a serious feasibility demonstration of this technique.

PART I: COMPARATIVE PROBABILITY OF ERROR AND CHANNEL
CAPACITY OF DIGITAL COMMUNICATIONS SYSTEMS

In order to choose or design a communication system sensibly it is necessary to know the advantages and disadvantages of the alternatives which are available. It will be the objective in this first section to present both existing knowledge and proposed comparative methods in the selection of digital communication system modulation and coding methods. Attention has been paid to sensibly organizing the material, and it is presented as briefly as possible. It is the intention that with the knowledge to be summarized below, one might sensibly proceed toward selecting methods for fulfilling a given communication system's requirements.

It should be stressed that the emphasis in this report will be overwhelmingly on theory. The state of the art, and the practical considerations which may or may not make certain systems possible will receive cursory mention as these were not seriously studied under this work.

Because of the fact that new comparative calculations are published almost weekly in this communications area, it serves a very useful purpose to take cognizance of these results and to attempt to assay their significance. Hence some of the material in this section will naturally be reporting on published work, while other material will be original with the authors.

In many cases, statements will be made without proof since derivations may be followed in the pertinent reference. This section is undertaken only because we feel that it serves a very necessary function before getting into more detailed and original contributions.

After discussing the general considerations for digital communication systems, the detection picture will be presented. The two major topics then are: (1) probability of error calculations for various conditions and (2) channel capacity relations. A good deal will be said about channel capacity for digital systems.

1. GENERAL CONSIDERATIONS

Under "General Considerations" we will depict the general block diagram for digital communications systems and discuss generally the possibilities (and trades) for improving the performance in each of the "box" areas. In addition to quantitative analysis, there will be a collection of comparisons and qualitative conclusions.

1.1 General Block Diagram

The block diagram for a digital communication system can be depicted as shown in Fig. 1.

By "source coding" we refer to those operations which convert the actual message source into a sequence of symbols. A common example is the sampling, quantizing, and subsequent encoding (PCM) of any analog voltage. "Error coding" refers to putting some constraints between the symbols so that certain combinations of errors can be detected (and possibly corrected) in the receiver. This is a form of using intentional redundancy to protect against channel noise. "Modulation" is concerned with the actual choice of waveform for a particular symbol, and the RF transmission of the symbol. In the receiver, of course, the mirror image of each of the above functions occurs.

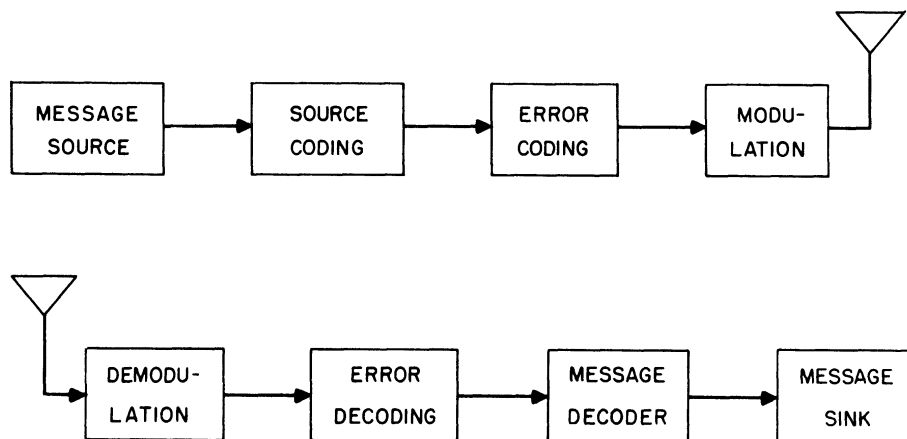


Fig. 1. General block diagram of digital communication system.

In this report we will regard the modulation-channel-demodulation complex as the detection situation. This is sensible since it is impossible to separate the effects of signal selection, power, etc., on the channel probability of error.

We wish, then, to consider the alternatives in digital communication system design as being separable into the three areas of: (1) message (or source) coding; (2) error-correcting (detecting) coding; and (3) detection improvement. In any system design concentration on one or on another of these areas will probably be most profitable to improve the system. The "trades" involved and the costs of each improvement are the subject of the ensuing material.

1.2 Source Coding

The standard method of source coding (for digital communications) is to sample an analog signal at the Nyquist rate ($2W$ samples per second), quantize the resultant samples, and digitally encode them in some fashion. With such a method it is clear that one is attempting to reproduce (in the receiver) nearly the exact original waveform.

For many analog signals (in particular, speech) one can reduce the apparent information content of the source by removing redundancies. For example, it is well known that good quality speech should require only about 50 bits/sec., but the sampling method above results in 30,000 bits/sec. The discrepancy may be interpreted as due to redundancy of the (source) speech signal, and bandspreading involved with digital encoding.

Clearly, if the redundancies can be reduced, a speech source will require less communication facility. Using the Vocoder principle, one can reduce the requirement to 2,000 bits/sec.

An approach different from the Vocoder method is to operate on Shannon samples in some fashion. "Delta modulation" schemes do this by sending only the "difference between present and integrated past samples." An extension of this is to use a number of past samples and predict the next sample value. Then the difference between the actual and the predicted sample value is sent. Experimental results using such predictions are reported in a companion report, "Some Computer Experiments Using Predictive Coding On Speech," TR-132.

Another source coding situation arises when the source information occurs in bursts. If delays are tolerable, one can try to smooth out the data so that a transmitter can operate closer to the average rate of the source. This aspect will appear in Part II with considerations of "asynchronous time-multiplexing" where a number of sources must be multi-

plexed.

1.3 Error Correcting and Detecting

In error coding the general objective is to introduce constraints between the symbols at the transmitter so that channel errors may be detected and possibly corrected in the receiver. It is useful to consider that error codes include both error-correcting and error-detecting (without correcting) situations.

Probably the earliest work on error codes was by Hamming (Ref. 1) who dealt with three codes: the parity check code, a single error correcting (which later became known as the Hamming code), and a single error correcting-double error detecting code. Such early codes were devised under the assumption that the occurrence of a single error was most likely and that the errors occurred at random and were independent of one another.

For the more realistic situation where errors occur in bursts (in adjacent digits) a new series of codes, called "cyclic codes," was devised. This includes the Fire codes and the Bose-Chaudhuri codes. The name "cyclic" comes from the fact that the code words can be generated in a cyclic fashion by feedback shift registers.

The present status of error codes is that a number of codes have been mathematically demonstrated, but that the application art lags somewhat behind this. Recently there have been profitable applications using shift-register encoding (Ref. 2) to implement error detecting .

By the use of such modern codes one can obtain efficient codes (high ratio of information-to-check digits for a given error rate) while at the same time providing for practical operating equipment (shift-register methods).

A summary of probability of error calculations with error codes will be given in Section 3.2.1. The most authoritative (recent) description of codes is Ref. 8.

One firm conclusion about error codes is that they are definitely profitable in situations where a feedback facility exists, that is, where one can ask for repeats. Then one can use a maximum of error detecting and a minimum of correcting. A large body of work has been published investigating this situation (Ref. 9).

Such feedback situations are clearly applicable to the transmission of non-real-time digital source data. However, no methods have been found to exploit the feedback-error-coding situation when the digital data stem from real-time sampled analog data (such as speech). The difficulty lies in dealing with the delays.

1.4 General Detection Considerations

Referring to Fig. 1, the "detection situation" is characterized by the modulation technique, the channel, and the demodulation. Each of these three items is concerned with identifying the right signal after reception. Since the effect of channel noise on probability of error is dependent on signal selection and receiver decision method, one should not separate these three items.

In general, the two things that can be done to improve error-rate performance in the detection area are:

- 1) Use proper detectors for the chosen signal whenever possible. (When the channel disturbance is additive Gaussian noise these will be optimum.)
- 2) Use as much "energy per decision" as necessary. Since the probability of error (with white Gaussian noise, using optimum detection) depends only on signal energy and channel noise (and not on signal time characteristics), one should integrate as much "energy per decision" as necessary. In other words, the greater the energy per decision, the lower the probability of error (P_e).

There are basically two ways to increase energy per decision in a digital system: (1) narrow-banding, and (2) increasing alphabet size.

In narrow-banding, one simply spends more time per symbol to obtain the increased energy. This of course means that the data rate is decreased accordingly.

One can increase energy per decision without decreasing the data rate by increasing alphabet size. Two ways of increasing the alphabet size are: (1) use M-ary basic symbols instead of binary, or (2) use "grouped" binary symbols. In either case the energy per decision is increased.

Although it will be found in Section 3.3.4 that such increasing of alphabet size may not be efficient, it may nevertheless be necessary if one has a given S/N, P_e and required data rate. When increasing alphabet this way, one is in effect trading S/N with increased complexity.

The detection situation will be discussed in more detail in Section 2.

2. DETECTION CONSIDERATIONS

Although the general conclusions about detection were summarized in Section 1, we now treat this in more detail. First, optimum receivers are discussed, then signal selection, and finally, the very important binary case is used as an example.

2.1 Optimum Receivers

In a digital communication system various alternative signals are sent at the transmitter in intervals of T seconds. Hence, for each interval, the receiver detector must decide which signal was sent. Receiver detectors will be either coherent or incoherent, depending upon whether propagation and equipment conditions are stable enough to use RF phase in the decision process.

2.1.1 Optimum Coherent Receivers. It has been shown (Ref. 38, etc.) that, if the channel noise is white Gaussian noise, the optimum method to perform the digital detection process is the crosscorrelator or matched filter detector. If one does not choose to incorporate a priori knowledge of the occurrence of the symbols, then it can be shown that this type of detector is the best that one can do, either on a signal-to-noise ratio criterion or an a posteriori probability criterion (Ref. 3). Thus in general the receiver will take the form shown in Fig. 2.

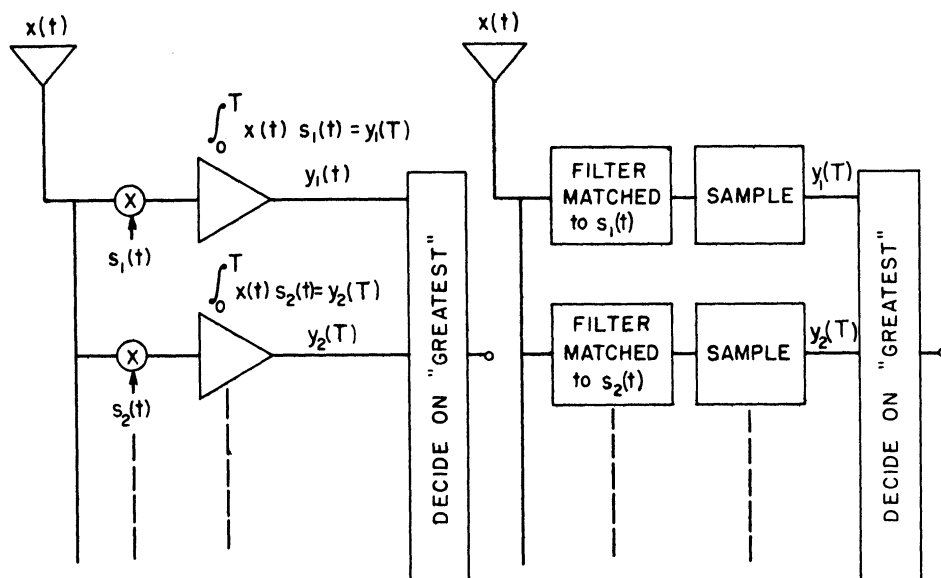


Fig. 2. Optimum coherent receivers for signals known exactly in white Gaussian noise.

Since these are coherent receivers, it is implied that one knows the phase and frequency of the RF carrier, and one is truly operating with the exact $s(t)$.

In Fig. 2(a) the incoming signal is first multiplied by the coherent local replicas. Then each product goes through an "integrate-and-dump" operation. In Fig. 2(b) the matched filter and sampler perform the identical function.

After the integration, the decision device (called the "comparer") specifies the output on the basis of which signal is greatest at $t = T$ (end of each symbol). Such receivers are called "maximum likelihood receivers."

It should be emphasized that the receivers of Fig. 2 are optimum, regardless of the choice of signals. However, one affects the detection situation by the choice of signals (considered below in Section 2.2).

2.1.2 Optimum Incoherent Receivers. The incoherent case refers to the situation where one does not know the RF phase sufficiently, and hence can work only with the video. Two types of optimum incoherent receivers are shown in Fig. 3---both involving matched filters. In (a) the carrier-band matched filter is followed by an envelope detector and sampler. The outputs of the sampler are then compared. This receiver is "optimum" for incoherent operation in white Gaussian noise (Ref. 18).

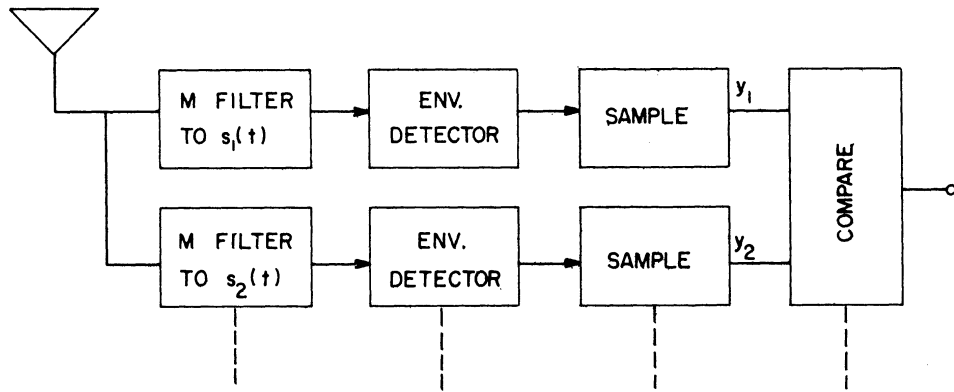
An alternative (but equivalent) incoherent method, shown in Fig. 3(b), is to multiply each alternative signal by both a sine and cosine function of the exact carrier frequency, and put both of these through a baseband matched filter (or correlator). Then one takes the square root of the sums of the squares, and compares these for the difference signals. This method requires knowledge of the exact carrier frequency, and is equivalent to the carrier-band matched filter receiver (Ref. 22).

Both of these receivers work on the envelope of the received signal, and hence all considerations depend on energy alone, since phase cannot be utilized. Also, the decision device (called the comparer) decides on the basis of which signal is greatest at $t = T$. Such a receiver is also a maximum likelihood receiver.

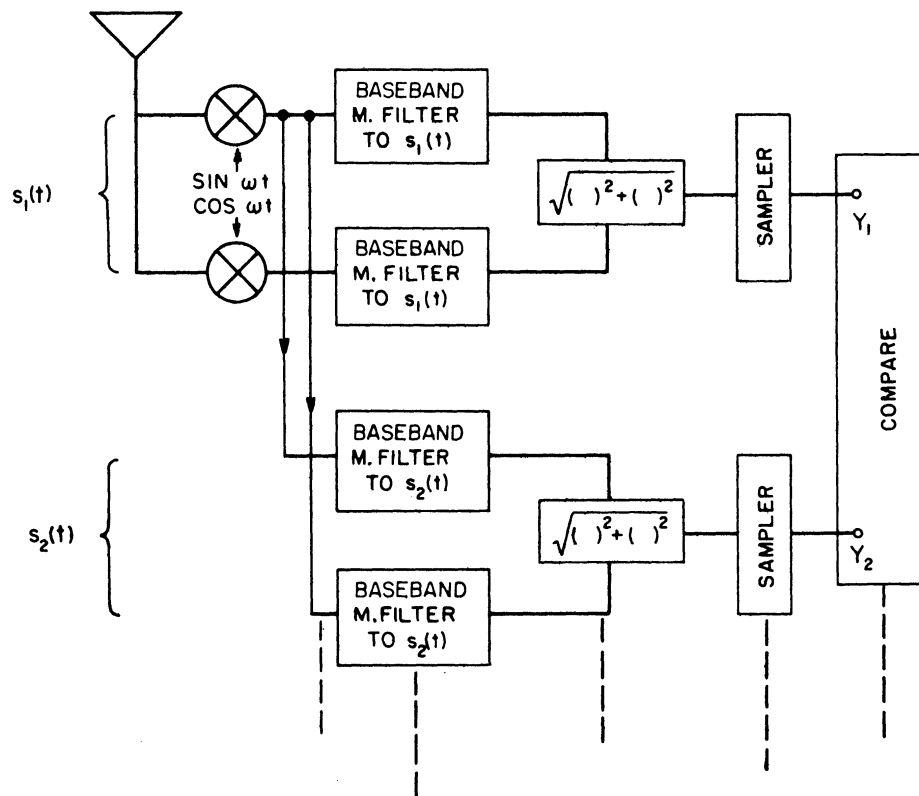
It should be emphasized that the receivers of Fig. 3, like those of Fig. 2, can be shown to be optimum regardless of the choice of signals. However, one can improve the detection by choosing the signals properly.

2.2 Signal Choice

We now deal with the question of what signals are best for use in conjunction with



(a) Carrier-band matched-filter receiver



(b) Base-band matched-filter receiver

Fig. 3. Two types of optimum incoherent receivers.

the receivers of Figs. 2 and 3. First the coherent case is treated.

2.2.1 Coherent Operation. "Coherent operation" implies that propagation and equipment conditions are stable enough to use RF phase in the receiver "decision" -- even at low signal-to-noise ratios. For such coherent operation, it is sensible to regard the possible signals as belonging to one of two classes: (1) "far-apart" or "distinguishable" signals, or (2) "close-packed" signals.

Such signals as "orthogonal" signals are considered "far-apart" because the major objective is to have the symbols as distinguishable as possible in the presence of noise. These signals are of interest when probability of error is a significant problem. In "close-packed" signals (such as M- phase with $M > 8$, or "phase-plus-amplitude" modulation) the major objective is to obtain efficient information transmission; here it is implied that S/N is high enough so that P_e is within requirements.

In this report we are concerned mainly with the far-apart signals. Hence we are concerned with those cases where probability of error is a significant problem. The following far-apart signals will be considered: (1) simplex, (2) bi-orthogonal, and (3) orthogonal.

That set of signals which are as "far-apart" as possible are usually termed "simplex" because of their analogy with vectors of a geometrical simplex.

Simplex signals (Ref. 3), viewed as points in signal space, correspond to a geometrical regular simplex where every vertex is equally distant from every other vertex. Of these, there are M signals, the geometrical simplex appears in an M-1 dimensional space. Two simple examples of such simplex signals are: (1) two binary signals, whose vectors are 180° out of phase --- $s_1(t) = -s_2(t)$, and (2) a ternary signal, where the vectors are at 120° phase with each other. Another simplex (baseband) signal set is a maximal (pseudo-random) sequence and its time translates (by one clock pulse).

Since simplex signals are the "farthest apart," such signals are optimum for coherent operation.

Although simplex signals appear optimum, orthogonal signals are near-optimum [and approach the optimum (simplex) for large alphabet M]. Signals are orthogonal, over an interval T, if:

$$\int_0^T s_i(t) s_j(t) dt = 0 \quad i \neq j$$

$$= E_i \quad i = j \tag{1}$$

where: E_i = total "symbol" energy of signal $s_i(t)$.

If the signals all have equal energy (a common assumption) then:

$$\int_0^T s_i^2(t) dt = E \quad \text{for any } i \quad (2)$$

Signals may also be orthogonal in the frequency sense, and this would be given by

$$\begin{aligned} \int_{-\infty}^{\infty} S_i(f) \overline{S_j(f)} df &= 0 & i \neq j \\ &= E & i = j \end{aligned} \quad (3)$$

where:
$$S_i(f) = \int_0^T s_i(t) e^{-j2\pi ft} dt$$

Thus it is seen that signals can be orthogonal in two senses: time or frequency.

A common example of "signals orthogonal in time" is simply time waveforms which do not overlap. Signals which do overlap in time can also be made orthogonal. The most common examples of "signals orthogonal in frequency" are FSK signals. Here the frequencies of the different signals are chosen to be effectively nonoverlapping.

Consider for a moment the role that orthogonality plays in affecting the probability of error for a digital system. Equation 1 says that, when signals are orthogonal, the time correlation or pairwise correlation coefficient between signals is zero. In general one writes the correlation coefficient as simply:

$$\frac{1}{E} \int_0^T s_i(t) s_j(t) dt = \rho_{ij} \quad (4)$$

One can see why the simplex signals mentioned above are superior to orthogonal signals for small M. Consider the binary case: if $s_1(t)$ is orthogonal to $s_2(t)$, the $\rho_{12} = 0$; but if $s_1(t) = -s_2(t)$, the $\rho_{12} = -1$. Clearly a "pairwise" correlation coefficient of -1 is better than 0.

Consider now an M-signal simplex alphabet. To calculate the total P_e is exceedingly complex, and has not been done directly. However, one can find the simplex P_e from the coherent orthogonal P_e by considering an energy increase (see Section 3.2.2.1). Of course, the more negative the pairwise correlation, the better.

It can be shown (Ref. 3) that the most negative value that can be obtained for the

pairwise correlation coefficient between members of an M-signal alphabet (where all ρ_{ij} 's are equal) is given by:

$$\rho_{ij} = -\frac{1}{M-1} \quad (5)$$

Here is the reason, then, that as M grows large, the simplex signals approach the orthogonal ones in detection effectiveness---since, as M increases, ρ_{ij} approaches 0.

Concerning the comparison between simplex and orthogonal signals, another point should be made. From an engineering viewpoint a very effective set of signals consists of using orthogonal signals and their negatives, called "bi-orthogonal" signals. For an M-signal alphabet one needs only M/2 correlators or matched filters. Also, the bandwidth is reduced, since now one has only M/2 "different" signal constructions.

The above signals have been based on the idea of being quite distinct or far apart, for purposes of distinguishing them in noise. Such signals are necessary when the S/N ratio is such that a significant probability of error (P_e) occurs.

If the S/N ratio is high enough so that P_e requirements are easily met, then signals other than far-apart signals may be desirable. One common signal which is partially correlated (for $M > 4$) is the coherent M-phase signal. Because of the increasing correlation between signals as M increases, $M = 8$ is probably the maximum which one would consider (Ref. 19) for intermediate S/N.

In addition to coherent M-phase signals, one can operate "differentially-coherent" where one requires coherence only over a symbol length. These signals will be discussed further in Section 3.2.2.

In conclusion, for coherent operation where P_e is significant, simplex signals are best, bi-orthogonal next, and finally orthogonal signals. When P_e is within requirements, signals other than the above far-apart signals become valuable; i. e., M-phase signals and others.

2.2.2 Incoherent Operation. For any incoherent operation there is no possibility of using phase information (by definition). Hence neither simplex nor bi-orthogonal (nor M-phase signals) are possible. This leaves orthogonal signals as the optimum signals for incoherent operation.

2.3 Example of Signal Selection in Binary Case

Since the binary case is the most frequently-used case, and its P_e curves are well known, we will note the results here to serve as an example of the above discussion.

For the coherent binary case, it is well known that the probability of error (in white Gaussian noise with optimum receivers) is given (Ref. 17) by:

$$P_e = P_B = \Phi\left(\left[(1 - \rho_{12}) E/N_0\right]^{\frac{1}{2}}\right) \quad (6)$$

where: $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_x^{\infty} e^{-\xi^2/2} d\xi$

E = signal energy

N_0 = single-sided white noise power per unit bandwidth

ρ_{12} = correlation coefficient between $s_1(t)$ and $s_2(t)$

P_B = bit error rate .

Note that for binary the simplex signal [$s_1(t) = -s_2(t)$] is identical to the bi-orthogonal signal. Curve (a) of Fig. 4 shows P_B versus E_b/N_0 for the simplex signal $s_1(t) = -s_2(t)$, where $\rho_{12} = -1$.

If one were to use orthogonal signals in this binary coherent case ($\rho_{12} = 0$), one would obtain curve (b) of Fig. 4.

Finally, in the incoherent case, where one should use orthogonal signals, the P_B curve is shown in curve (c). The basis for these curves will be exhibited in Section 3. 1.

In conclusion, we have noted that, for optimum detection situations, the P_e depends only on the energy (E) and the noise power density (N_0). It does not depend on other characteristics of the signal, such as BW, etc.

For coherent signals it was noted that (for a given M) simplex signals have the best P_e behavior, bi-orthogonal next, and then orthogonal. For incoherent operation orthogonal signals are best. Note that, for the coherent case, orthogonal signals become nearly as effective as simplex ones, as M increases.

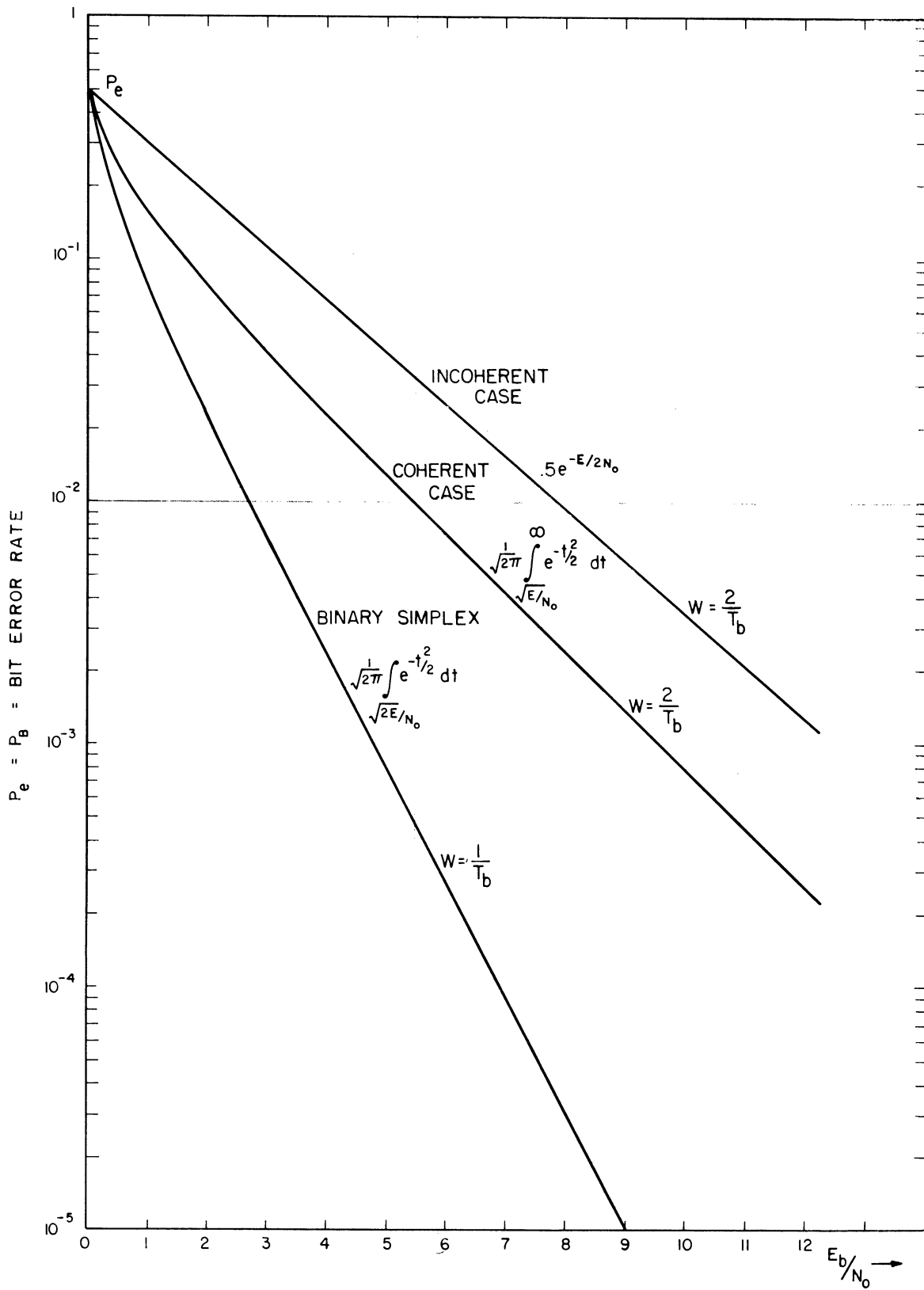


Fig. 4. Probability of error vs. E_b/N_0 for binary signals.

3. PROBABILITY OF ERROR CALCULATIONS AND RESULTS FOR DIGITAL SYSTEMS

For digital systems, the probability of error is the measure of effectiveness regarding channel noise. Of course there may be other sources of noise, such as quantization noise, timing errors, etc. However, to begin any realistic comparison, the channel noise probability-of-error is most basic.

In this section we will exhibit briefly the mathematical calculation for the probability of error and compare the results for various conditions of operation. The first section will be devoted to binary systems and the second section to nonbinary systems.

There are three types of error probabilities in digital systems, depending upon the original data and the size of alphabet: bit error probabilities, word error probabilities, character error probabilities. If the original data are bits (such as a series of binary numbers) then the bit error rate (P_B) is the pertinent criterion. If the information is sent by means of characters (M-ary basic symbol or block binary), the character error probability (P_{ch}) should be converted back to equivalent bit-error (P_B). On the other hand, if the original data are words (such as teletype or sampled analog signals) then the word error probability (P_w) is the pertinent criterion. If the information is sent by bits, the equivalent P_w should be calculated. If M-ary characters are used to transmit the words, the P_{ch} may or may not be equal to P_w , depending on the value of M.

In this report the symbol " P_e " will be a general term (which includes all the others) to denote "probability of error." Depending upon the case, it will be either a bit, a word, or a character error rate. It will always be marked accordingly.

A system will be called "binary" or "nonbinary" according to the "order of the decision process." If a decision is made at the end of every bit, and is not later revised (based on an error correcting code, for example), the system is considered "binary." With this definition, the nonbinary systems can be of two types: (1) error (correcting) coded binary signal, where intersymbol constraints are evaluated after the first binary decisions; and (2) M-ary symbol alphabet, where basic symbols (characters) other than binary (ternary, quarternary,

etc.) are used, and a decision is made on each of the M-ary characters. In turn there are two ways to realize an M-ary symbol alphabet: basic M-ary symbols and block coded binary. We feel this categorization is sensible, and has the most consensus in this field. (See Table I.)

The above items can be regarded as alternatives available to a system designer. Of course, within these alternatives there are still a number of further choices. For example, if binary, one can operate either coherently or noncoherently. If coherent, one can use PSK, or coherent FSK; if noncoherent, one can use noncoherent FSK or AM.¹ Because of all these permutations it is difficult to set up a consistent and sensible categorization. However, the chart in Table I attempts to do this, along with providing information to be discussed later.

In general the "practical" signals which are optimum in some sense are (see Section 2.2): (1) simplex, (2) bi-orthogonal, (3) orthogonal, and (4) M-phase. We will consider the P_e relations for digital systems in terms of these categories.

One can reach a general P_e equation for both the bi-orthogonal and the orthogonal set of signals. However, for the simplex set, no equation has been found so far, but the P_e can be related to the coherent orthogonal case. For the binary signal, however, the PSK signal is the simplex signal, and this P_e can of course be calculated. For this reason, we will first depict the P_e calculations for binary, treating it as a sole case. Remember that, for the bi-orthogonal and orthogonal cases, the binary P_e could be obtained as a special case of the M-symbol general relations of Section 3.2.2.

Another justification for proceeding in this way is that the binary case is the most common in practice, and therefore deserves special attention.

3.1 Binary Systems

As mentioned, binary systems are taken to include all those systems in which a binary decision is made at the end of each received bit, and is not changed thereafter (i. e., by a correcting code). In the following the symbols 1 and 0 will represent the two binary states (a +1 and -1 would be better for coherent situations, but this is clumsy to write).

For binary cases there are three different signals: (1) Simplex---PSK, where $s_1(t) = -s_2(t)$; (2) Coherent-orthogonal---coherent FSK, and (3) Incoherent-orthogonal---Incoherent FSK and AM. To depict the P_e calculation, we will consider each of these three cases.

¹Incoherent AM will mean carrier "off-on" operation in this report.

DECISION LEVEL	Signals in Use (or contemplated)		Signal Type	P _e Eq. No.	Theoretical Bandwidths		Constant Digit Rate Bandwidth (in terms of T _b)	Channel Capacity Comments		
					Controlled Environment	Uncontrolled Environment		Minimum Signal Dimensions D	Max C in terms of W	
BINARY	COH		CPSK*	Simplex	(6) & (13)	$\frac{1}{T_b}$	$2/T_b$	$1/T_b$	1	2W
			DCPSK**	Simplex	-	$\frac{1}{T_b}$	$2/T_b$	$1/T_b$	1	2W
			Coh. FSK	Coh. --Orthog	(16)	$\frac{2}{T_b}$	$3/T_b$	$2/T_b$	2	W
	INCOH		Incoh. FSK	Incoh. --Orthog	(25)	$\frac{2}{T_b}$	$3/T_b$	$2/T_b$	4	W/2
			Incoh. AM	Incoh. --Orthog	(25)	$\frac{2}{T_b}$	$3/T_b$	$2/T_b$	4	W/2
NONBINARY	Error-Coded		Incoh. AM Incoh. AM	Incoh. --Orthog Incoh. --Orthog	-----	-----	-----	-----	-----	-----
	M-symbol	M-ary Basic Symbol	CPSK	M-phase	(52)	1/T	2/T	$\frac{1}{T_b \log_2 M}$	2	$W \log_2 M$
			DCPSK	M-phase	-	1/T	2/T	$\frac{1}{T_b \log_2 M}$	2	$W \log_2 M$
			Kinoplex	Bi-orthogonal	(45)	M/2T	$\frac{M+2}{2T}$	$\frac{M}{2T_b \log_2 M}$	M/2	-----
			Coh. FSK	Coh--Orthog	(35)	M/T	$\frac{M+1}{T}$	$\frac{M}{T_b \log_2 M}$	M	$\frac{2W}{M} \log_2 M$
	Alphabet	Block-Coded Binary	Shift-reg. encoded	Simplex	-	$\frac{M-1}{T}$	$\frac{2M-2}{T}$	$\frac{M-1}{T_b \log_2 M}$	M-1	$\frac{2W}{M-1} \log_2 M$
			Reed-Muller encoded	Bi-orthogonal	(45)	M/2T	$\frac{M+2}{2T}$	$\frac{M}{2T_b \log_2 M}$	M/2	-----
				Coh-orthog	(35)	M/T	$\frac{M+1}{T}$	$\frac{M}{T_b \log_2 M}$	M	$\frac{2W}{M} \log_2 M$
				Incoh-orthog	(40)	M/T	$\frac{M+1}{T}$	$\frac{M}{T_b \log_2 M}$	2M	$\frac{W}{M} \log_2 M$

*CPSK-Coherent phase shift keying

**DCPSK-Differentially coherent phase shift keying

Table I. Depiction of various digital signals and their parameters.

Although the items appearing in this table will be dealt with in the following material, it is profitable to explain terms here.

Controlled Environment Bandwidth -- The bandwidths here refer to the bandwidth occupancy of a signal, if the adjacent communicators use the same signal and if their frequencies are controlled (and integrally related) with respect to the given signal. (Note that phases need not be integral). The bandwidths here state how close signals can be placed and not (theoretically) interfere with each other, under the stated conditions. The ideal signals take more bandwidth, but nevertheless they do not interfere.

Uncontrolled Environment Bandwidth -- The bandwidths

here assume that the spectrum of the ideal signal will be cut off at some point. With a "square-wave" signal pulse, the numbers here assume that only the main lobe of the (sin x)/x spectrum is passed. Such spectrum limiting will of course alter the square-wave signal pulse. This is called "uncontrolled environment" because now there need be no relation between communicator's signals.

Considering the spectrum of M-ary coherent cases, the bandwidth here includes the main lobe of the outermost (in frequency) signals. For coherent FSK there are M lobes; for block coded binary there is only one major lobe.

Constant Digit Rate Bandwidth -- This bandwidth is sim-

ply the controlled-environment W written in terms of T_b.

For constant rate, T_b is constant. Hence, these bandwidths are convenient for constant rate comparisons.

Signal Dimensions (D) -- D specifies the minimum dimensions (in a signal space representation) per symbol time for a given signalling method. Section 4.1.2 relates this minimum dimension to a minimum TW product (per symbol time) for a given signalling method. This in turn is used to find a capacity associated with each signalling method.

Maximum C -- Since the C for a given signalling method varies with E_b/N₀, the maximum C here is the highest (or asymptotic) value of C. For these values the P_e's of the symbols approach zero.

For binary signalling the following probability equation for probability of error can always be written:

$$P_e = P_B = P(1|0) P(0) + P(0|1) P(1) \quad (7)$$

where: P_B = probability of a bit error
 $P(1|0)$ = probability of specifying a 1 when a 0 is present.

If one assumes that the a priori probability of a 1 equals the probability of a 0, then $P(0)$ and $P(1) = \frac{1}{2}$. Furthermore it will be assumed that the penalty for either wrong decision is equal, so that $P(1|0) = P(0|1)$. Then the probability of error is given by

$$\begin{aligned} P_B &= P(1|0) \frac{1}{2} + P(0|1) \frac{1}{2} \\ &= P(1|0) = P(0|1) \text{ for symmetric,} \\ &\quad \text{binary situation} \end{aligned} \quad (8)$$

The major issue then, for all such symmetric binary systems, is to evaluate either $P(1|0)$ or $P(0|1)$. In general, this can be done by considering the a posteriori probability distributions. Consider that one is using two correlators or matched filters, as shown in Fig. 2. Then let:

$$\begin{aligned} p_0(y_1 - y_0) &= \text{prob. density of the output } (y_1 - y_0) \text{ if symbol 0 is sent} \\ p_1(y_1 - y_0) &= \text{prob. density of the output } (y_1 - y_0) \text{ if symbol 1 is sent} \end{aligned} \quad (9)$$

These a posteriori distributions, for a symmetric case, are shown sketched in Fig. 6. For all symmetric cases (and if the two distributions are of the same shape), the cut-point "x" of Fig. 6 will occur where the two curves are equal. The cut-point is the threshold, above which one always decides 1, and below which one decides 0.

Thus the "area of confusion," given by Eq. 8, is shown by the two marked areas of Fig. 6. One can evaluate either of these areas in the symmetric case. From these curves it can be seen that the P_B of Eq. 8 can be evaluated by:

$$\begin{aligned} &[\text{let } y = (y_1 - y_0)] \\ P_B &= P(1|0) = P(0|1) = \int_{-\infty}^x p_1(y) dy = \int_x^{\infty} p_0(y) dy \end{aligned} \quad (10)$$

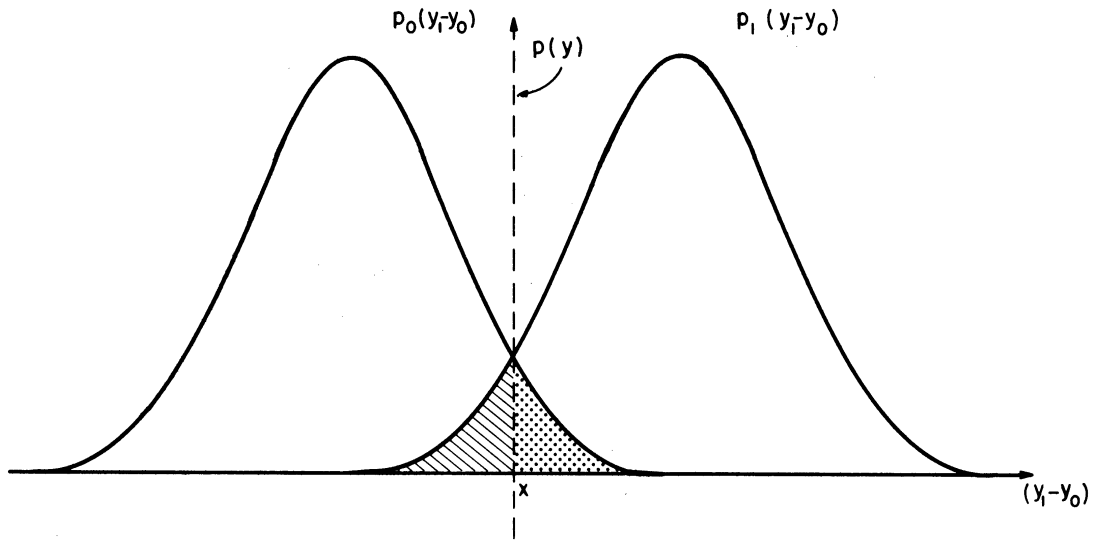


Fig. 6. General depiction of probability of error calculation for binary, symmetric decisions in terms of probability densities.

In general, for binary situations in white Gaussian noise, the P_e calculation involves the procedure just described. For each special situation the cut-point may vary, and the distributions may vary. With this general procedure we now show the relations for simplex, coherent orthogonal, and incoherent orthogonal.

3. 1. 1 Binary Simplex. In binary simplex, $s_1(t) = -s_2(t)$. Two basic receivers for such signals are shown in Fig. 2. In addition, for coherent binary cases, one can reduce the number of correlators (or matched filters) by a factor of two. One does this by correlating (or match-filtering) to the "difference of signals." Figure 7 shows the correlator version: (a) is repeat of Fig. 2(a), and (b) shows multiplication by $s_1(t) - s_2(t)$. Note that the decision device in Fig. 7(a) is "choose greatest," while in Fig. 7(b) it is "check polarity." Although correlators are shown here, matched filter versions may also be used. For purposes of calculating P_e we will use the one-correlator receiver of Fig. 7(b), although one could use either.

The objective is to obtain $p_1(y)$ and $p_2(y)$ (see Fig. 6). First note that if $s_2(t)$ is sent:

$$\begin{aligned}
 y(T) &= \int_0^T [s_2(t) + n(t)] [s_2(t) - s_1(t)] dt \\
 &= \int_0^T s_2^2(t) dt - \int_0^T s_1(t)s_2(t) dt + \int_0^T n(t) [s_2(t) - s_1(t)] dt
 \end{aligned}$$

Since, in binary simplex, $s_1(t) = -s_2(t)$:

$$y(T) = 2E + \int_0^T n(t) [s_2(t) - s_1(t)] dt \quad (11)$$

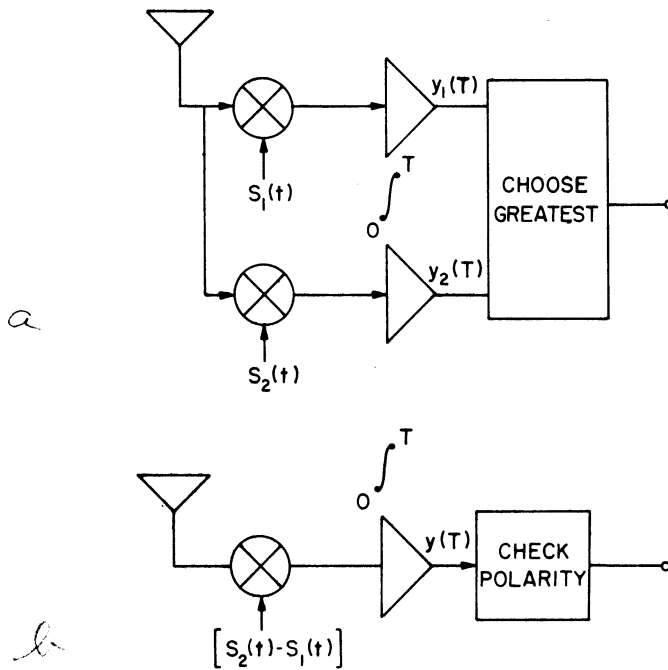


Fig. 7. Two basic receivers for binary coherent operation.

Looking at the "statistical" effect of $y(T)$, it is seen that the first term contributes a mean ($2E$), but no variance (since it is a constant). The second term will have a variance, and a zero-mean [assuming $n(t)$ has zero-mean]. If $n(t)$ is Gaussian, the second integral will be Gaussian since the integral of a Gaussian process is Gaussian (Ref. 10). By a variety of means, one can show that the σ^2 of this second integral is $2EN_0$ (Ref. 20).

If $s_1(t)$ were sent the mean would be $-2E$ with the same variance ($2EN_0$). Before depicting these curves, it is convenient to normalize for purposes of comparison and ease of writing equations. Consider that the mean is normalized to $\pm E$. Then the variance will be $EN_0/2$.

Since the cut-point of this binary symmetric situation will be at zero, the normalized distribution diagram will appear as in Fig. 8 (where $y^* = \frac{y_2 - y_1}{2}$).

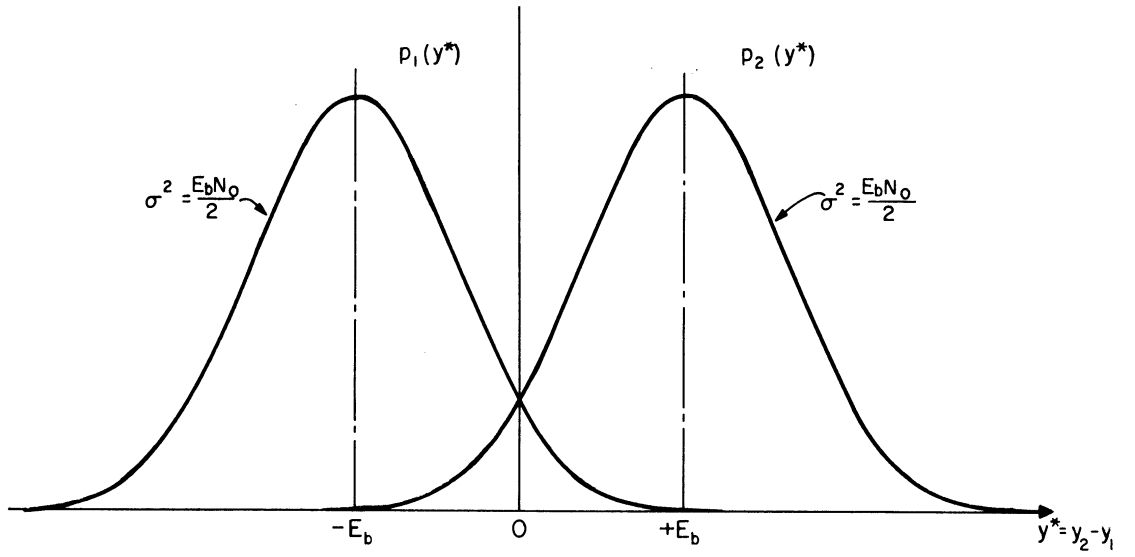


Fig. 8. Distribution diagram for simplex binary symmetric situation.

The equation for either of the curves will be:

$$p(y^*) = \frac{1}{\sqrt{2\pi} \sigma} \exp \left[-\frac{1}{2} \left(\frac{y^* \pm E_b}{\sigma} \right)^2 \right] = \frac{1}{\sqrt{\pi E_b N_o}} \exp \left[-\frac{(y^* \pm E_b)^2}{E_b N_o} \right] \quad (12)$$

$$\text{where: } \sigma^2 = \frac{E_b N_o}{2}$$

Then, using Eq. 10:

$$(\text{let } y = y^*)$$

$$\begin{aligned} P_e = P_B &= \int_0^{\infty} p_1(y) dy_1 = \int_0^{\infty} \frac{1}{\sqrt{\pi E_b N_o}} \exp \left[-\frac{(y + E_b)^2}{E_b N_o} \right] dy \\ &= \frac{1}{\sqrt{2\pi}} \int_0^{\infty} \frac{\exp \left(-\frac{y^2}{2} \right)}{\sqrt{\frac{2E_b}{N_o}}} dy \end{aligned} \quad (13)$$

This is the equation that was plotted in curve (a) of Fig. 4 (Section 2.3).

PSK. A direct way to obtain $s_1(t) = -s_2(t)$ is to use phase-shift keying (PSK).

For this case, the received signals can be written:

$$\begin{aligned} s_1(t) &= + A \sqrt{2} \sin \omega t \\ s_2(t) &= - A \sqrt{2} \sin \omega t \end{aligned} \quad (14)$$

Then, the following are true:

$$\begin{aligned} E_b &= A^2 T_b \\ \sigma^2 &= \frac{A^2 T_b N_o}{2} \end{aligned} \quad (15)$$

This elementary calculation for energy per symbol (and the ones to follow) must be regarded with caution. Such square wave calculations would be precise only if an infinite BW were available (although signals may be relatively closely spaced in frequency; see Section 3.3 for discussion of BW). The amplitude of the waveforms over a symbol will usually be an $A(t)$ rather than a constant A . The energy calculation would then be based on this $A(t)$.

We will plot most curves of this report in terms of E_b/N_o , where E_b/N_o = the signal energy per bit divided by the noise power per cycle.

3.1.2 Coherent Orthogonal. For the orthogonal case the relations of either Eq. 1 or 3 must be fulfilled; i. e. , the correlation between signals is zero.

For this case, one can again have the two basic receivers shown in Fig. 7 or their matched-filter versions. Because the signals are now orthogonal (instead of having a negative correlation as in the above simplex case) the $p(y)$'s have the same shape, but are now "closer together." Using the receiver of Fig. 7(b), the mean of the output distributions $p(y)$ will be $\pm E_b$, and the variance $E_b N_o$. The $p(y)$ diagrams for this case are shown in Fig.

9. Using Fig. 8, it is seen that here the variance is twice that of the simplex case.

Based on this $p(y)$, and using Eq. 10, the P_e is given by:

$$P_e = P_B = \frac{1}{\sqrt{2\pi}} \int_{\sqrt{\frac{E_b}{N_o}}}^{\infty} e^{-y^2/2} dy \quad (16)$$

This equation was plotted in (b) of Fig. 4 (Section 2.3).

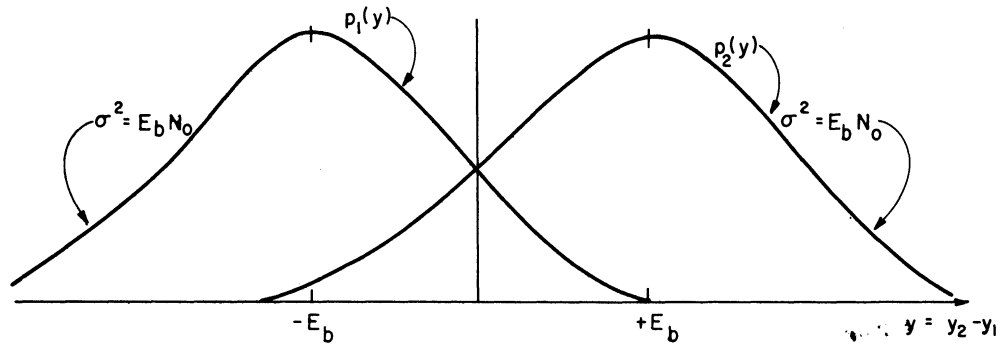


Fig. 9. Distribution diagram for the coherent-orthogonal, binary symmetric situation.

Coherent FSK. A way to obtain coherent-orthogonal signals would be to use coherent FSK. The signals for this case can be written:

$$\begin{aligned}
 s_1(t) &= \sqrt{2} A \sin \omega_1 t \\
 s_2(t) &= \sqrt{2} A \sin [\omega_2 t + \theta] = \sqrt{2} A \sin [(\omega_1 + \Delta\omega)t + \theta] \quad (17)
 \end{aligned}$$

where: θ is kept constant

In order for these waveforms to be orthogonal it is necessary that the following equation be fulfilled:

$$\int_0^T s_1(t) s_2(t) dt = 2A^2 \int_0^T \sin \omega_1 t \sin [(\omega_1 + \Delta\omega)t + \theta] dt = 0 \quad (18)$$

this equation is zero if:

$$\Delta\omega = \frac{M}{T} 2\pi$$

where M, N are integers

$$\text{and } (2\omega_1 + \Delta\omega) = \frac{N}{T} 2\pi \quad (19)$$

These conditions are independent of θ . If $2\omega_1 \gg \Delta\omega$ and $2\omega_1 T = K2\pi$, only the first condition of Eq. 18a need be considered. For such usual cases, then:

$$\omega_2 = \omega_1 + \frac{M}{T} 2\pi \quad M = 1, 2, 3, \dots \quad (20)$$

Thus the coherent frequencies can be spaced $1/T$ apart, under the usual conditions, for orthogonality.

The E for this case would be the same as that of Eq. 15.

3.1.3 Incoherent Orthogonal. For the incoherent-orthogonal case the signals are orthogonal, but incoherent operation of the receiver is used. The two possible optimum receivers are shown in Fig. 3.

Above, the P_e was investigated by finding the $p(y)$'s, and (using symmetry) one needed only to evaluate one "tail" of a Gaussian distribution, since there was symmetry about the cut-point. For the incoherent case the distributions are not symmetrical about the cut-point, as will be seen.

For incoherent orthogonal the two distributions of interest concern the envelopes. The distribution of the filter output which is not matched to the signal is given by the distribution of the envelope of narrowband, white Gaussian noise (i. e. , the Rayleigh distribution). Thus, (Ref. 5):

$$p_0(y_1) = y_1 e^{-y_1^2/2} \quad (21)$$

where: $p_0(y_1)$ = probability density of output y_1 if symbol 0 is sent.

The distribution of the output whose filter is matched to the signal is given by the distribution for the envelope of a sine wave plus noise, sometimes referred to as the Rice distribution.

Again, using Ref. 5, this is:

$$p_1(y_1) = y_1 e^{-\frac{y_1^2 + A^2}{2}} I_0(y_1 A) \quad (22)$$

where: $p_1(y_1)$ = probability density of output y_1 if symbol 1 is sent

A = peak-signal-to-rms-noise voltage ratio = $\sqrt{2E_b/N_o}$

$I_0(z)$ = Bessel function with imaginary argument

$$= \sum_{n=0}^{\infty} \frac{z^{2n}}{2^{2n} n! n!}$$

Equations 21 and 22 are the densities of interest for both incoherent AM and FSK.

However, the two cases are different because of cut-point considerations.

Incoherent AM. For incoherent AM (or "off-on") there is only one matched filter [a single one of Fig. 3(a), or a single pair of Fig. 3(b)], and the distributions appear as in Fig. 10. Note that the abscissa y_1 in this diagram is the output of the single matched filter, and not the difference $(y_2 - y_1)$ as in Figs. 8 and 9.

If symmetric errors are desired, one sets the point x so that the two areas shown are equal. If one chooses symmetric operation, the proper x can be found by the use of tables, and the following procedure:

Find x such that:

$$1 - Q(0, x) = 1 - e^{-x^2/2} = Q(A, x) \quad (23)$$

$$\text{where: } Q(A, x) = \int_x^\infty y e^{-y^2/2} e^{-A^2/2} I_0(Ay) dy$$

Then:

$$P_e = P_B = e^{-x^2/2} \quad (24)$$

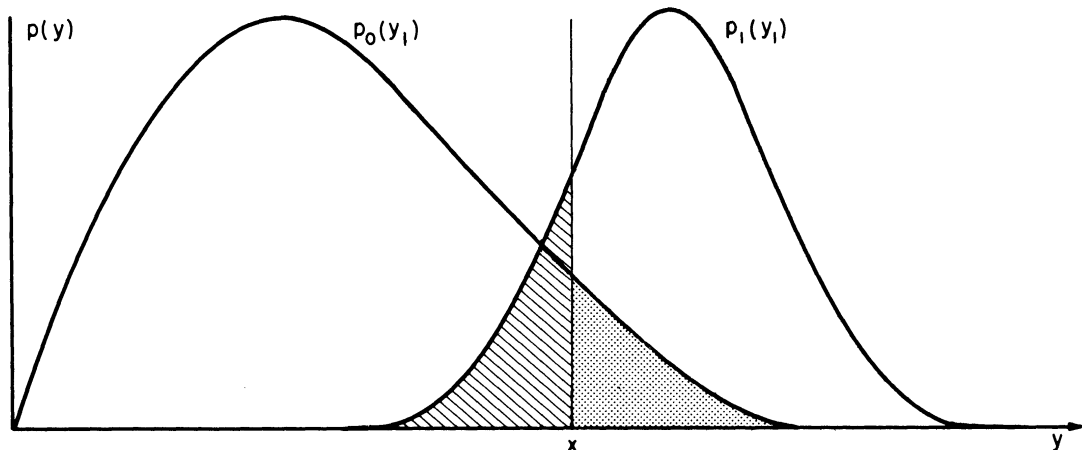


Fig. 10. Sketch of probability distributions for the incoherent, orthogonal AM case.

Incoherent FSK. For incoherent FSK the distributions of Eqs. 21 and 22 hold, but it is difficult to sketch the distributions in the fashion of Figs. 8 and 9. The P_e calculation will be shown in Section 3.2.2.1; we simply note the result here.

For symmetric errors, incoherent (orthogonal) FSK results in:

$$P_e = P_B = .5 e^{-\frac{E_b}{2N_0}} \quad (25)$$

This was the graph plotted in (c) of Fig. 4 (Section 2. 3). One will obtain this P_e if one uses incoherent FSK, which is an incoherent-orthogonal case.

3.2 Nonbinary Systems

By the earlier definition a system is nonbinary if the decision process involves more than one "bit." Thus, there are two nonbinary situations: (1) intersymbol constraints are imposed on the original binary decisions---called "error-coded binary signals," and (2) M-ary signals are used (see Table I).

Suppose that, with a given S/N ratio, the P_e is unacceptable. One can then consider remaining binary and increasing the symbol time (which decreases the data rate), or going to a nonbinary situation. Suppose it is necessary to keep the information rate constant; then it is necessary to go to the nonbinary systems. Of the two possibilities, error-coded binary and M-symbol alphabet, we consider first the error-coded binary.

3.2.1 Error-Coded Binary Signals. Since error-coded binary signals involve groups of binary digits, the first probability calculation that can be made will be a word probability (P_w)---that is, the probability that an error-coded word is incorrect. This result is immediately applicable if the original data are words of the same length as the "number of information digits" in the error-coded signal.

If the original data are independent bits, one would like to find the equivalent information-bit error-rate from the P_w . Unfortunately this is exceedingly difficult and we are not aware of any such calculations. Consequently we will be restricted here to a calculation for the P_w in an error-coded situation.

In error-coded binary, a message word of m information bits is encoded into an n-bit code word, containing $k = n-m$ redundant digits. Clearly, with this increase in numbers of transmitted bits, the net information rate must decrease, or conversely, the bit transmission rate must be increased, with consequent increase in bandwidth needed if information rate must be held constant.

After the coding, the binary code words are then transmitted by one of the schemes already discussed in Section 3. 1. Thus, the bit-error probabilities P_B are determined by the modulation-detection method, while the word-error probabilities P_w are determined by the

error-encoding and decoding scheme used. Therefore, the relations below are given in terms of bit-error probability, which have already been described.

As a comparison, the word-error probability for an n-bit uncoded word is simply:

$$P_w(n) = 1 - (1 - P_B)^n \quad (26)$$

where: P_B is the bit-error probability.

The word-error probability (P_w) for a general error-coded word (where the errors are assumed independent) has been worked out (Ref. 11). The assumption is made that the bit-error probability P_B is known and is independent for each digit. This is the case for a channel perturbed by white Gaussian noise, but does not necessarily hold for cases when the channel is disturbed by impulse noise.

The word-error probability is given for a code word of n bits, containing m information bits and k check digits---denoted as (n,m). Thus, there is a total of 2^m information words and 2^n message words, leaving r redundant words, where

$$2^n - 2^m = r \quad (27)$$

or
$$2^{n-m} - 1 = \frac{r}{2^m} \frac{\text{redundant words}}{\text{information word}} .$$

Therefore, the code should be able to correct at most 2^{n-m} errors, since each information word can have a total of 2^{n-m} message words associated with it.

The best code will correct all combinations of errors up to s-tuple, plus some j_{s+1} of (s+1)-tuple errors, where

$$\sum_{i=0}^s \binom{n}{i} + j_{s+1} = 2^{n-m} \quad (28)$$

The total probability of correctly receiving the word, or receiving a word with correctable errors, is:

$$P_c = \sum_{i=0}^s \binom{n}{i} P_B^i Q_e^{n-i} + j_{s+1} P_B^{s+1} Q_e^{n-s-1} \quad (29)$$

where: $Q_B = 1 - P_B$

So the probability of a word received in error is:

$$P_w(n, m) = 1 - \left[\sum_{i=0}^s \binom{n}{i} P_B^i Q_B^{n-i} + \sum_{s+1}^n P_B^{s+1} Q_B^{n-s-1} \right] \quad (30)$$

This expression can be evaluated for a particular code and modulation scheme.

Figure 11 shows the word probability of error (P_w) for a (16, 5) and a (32, 5) code.

Both incoherent (orthogonal) and coherent (simplex) binary detection are shown. These curves were calculated using Eq. 30 with the P_B 's given in Fig. 4. Also, the information data rate and the average power of the transmitter are kept the same as for the uncoded transmission. This is necessary to make a sensible constant data rate comparison. This means that, as E_b is increased, the E_p per bit decreases.

As seen in Fig. 11, there is relatively little advantage in this type of error code (for constant data rate and average power) for $m = 5$.

It must be emphasized that Eq. 30 is for codes where the errors are assumed independent. For more modern "burst correcting codes," we know of no equivalent relation, and no basic work was done on this in this contract. The reason for this, as mentioned in Section 1.3, is that the real power of error codes lies where feedback may be utilized. Since speech channels do not usually adapt themselves to such cases, error-coded signals are not of greatest importance to us.

3.2.2 M-Symbol Alphabet. The other general way to obtain better P_e 's, in contrast to the error-coding schemes described above, is to use communication signals which are M-ary rather than binary. Again, referring to Table I, there are two cases: (1) M-ary basic symbols and (2) block coded binary symbols. Although, from the theoretical error point of view these two cases are the same, there is good reason, from the equipment point of view, to separate the cases.

The essential profit of going to M-ary symbols lies in achieving the "integration of more energy per decision" (see Section 1.4).

In this section we will consider the probability-of-error calculations for M-ary signals in general; then the different ways to implement the M-ary basic symbol and block coded binary signal will be noted.

3.2.2.1 General P_{ch} Calculations. If one calculates the error-probability for an M-ary signal, the result is a character error rate rather than a bit error rate (P_B)---hence

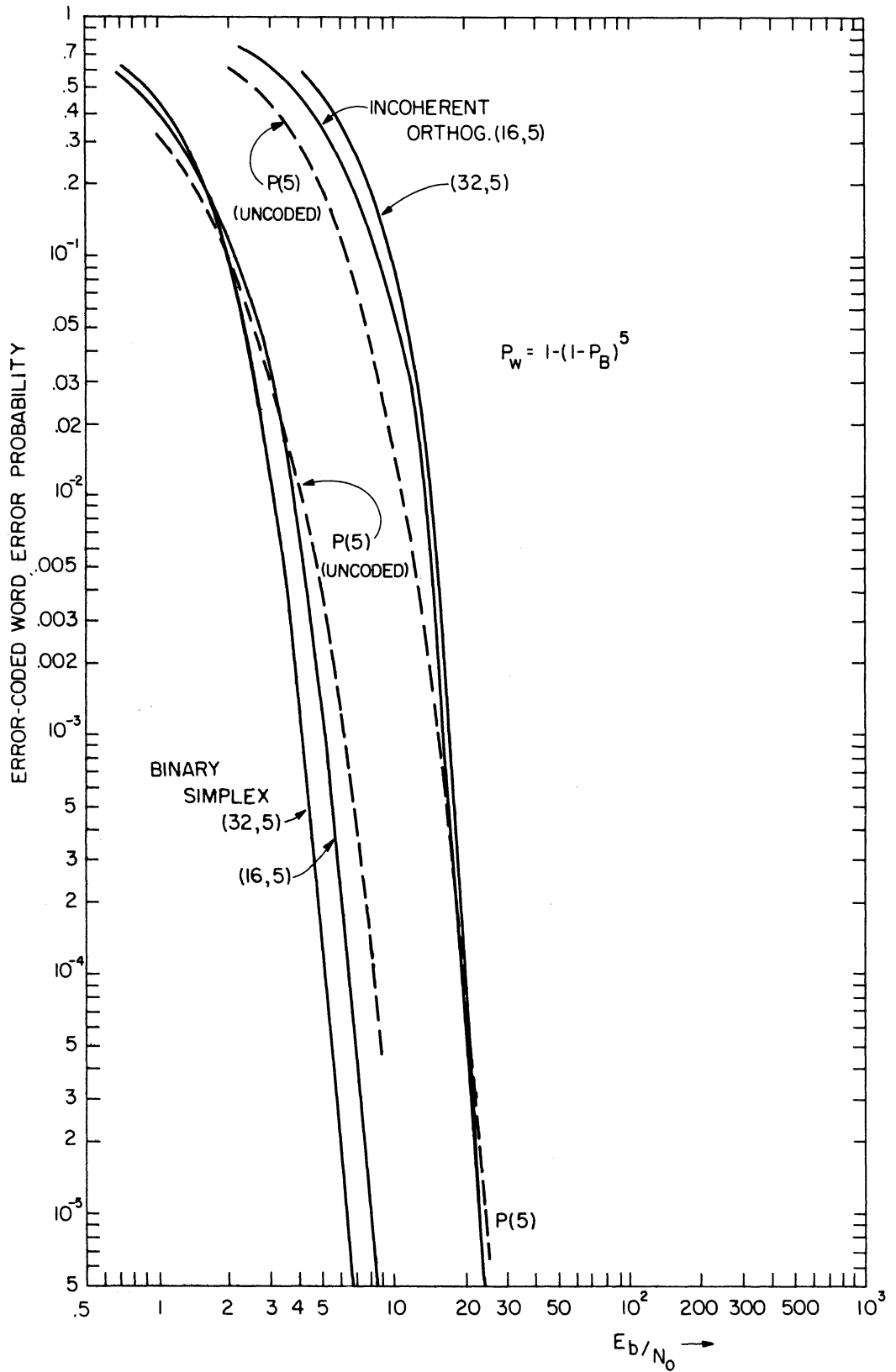


Fig. 11. P_w versus E_b/N_0 for coherent simplex and incoherent-orthogonal error-coded binary signals.

the use of the symbol P_{ch} . This is reasonable since an M-ary symbol can be viewed as equivalent to a "group of bits."

If the original data are words, one should convert the P_{ch} to a P_w , if they are not already equal. If the data are bits, one should convert the P_{ch} to equivalent bit-error-rate (P_B).

In this section, then, we will be dealing with the P_{ch} . For the comparisons of Section 3.3.1, the P_{ch} 's will be converted to equivalent P_B and P_w .

For M-ary symbols, the 4 types of signals to be studied are:

1. Orthogonal---coherent and incoherent.
2. Bi-orthogonal signals (require coherence).
3. M-phases---coherent or differentially coherent.
4. Simplex signals (see Section 2.2).

We now discuss P_{ch} calculations for these four cases.

Orthogonal Signals. The various types of orthogonal signals were described in Sections 3.1.2 and 3.1.3. We will first note the P_{ch} calculation for the coherent orthogonal case.

The receiver used for all of the signals below will be the "optimum" receiver, as discussed in Section 2.1. For M-ary signals, optimum receivers use the "maximum likelihood detector;" that is, one chooses the greatest correlator output out of the M possible outputs. The P_{ch} calculation then proceeds as follows.

Consider that signal M is sent. Then a certain voltage y_M (determined partly by noise) occurs at the output of the Mth filter. The probability of a correct decision $P(c)$ for that y_M is then determined by the probability "that all other y_i 's are less than y_M ." Then, if one integrates over all y_M , the result is $P(c)$.

Let:

$$p_M(y_M) = \text{the probability density of the output of the correct Mth filter when signal M is sent}$$

$$P(y_1, y_2, \dots, y_{M-1} < y_M) = \text{the probability that the outputs of all the other filters are less than } y_M. \quad (31)$$

Then $P(c)$ can be written:

$$P(c) = \int_{-\infty}^{\infty} p_M(y_M) P(y_1, y_2 \dots y_{M-1} < y_M) dy_M \quad (32)$$

where: $P(c)$ = probability of correct decision = $1 - P_{ch}$

Now, since signals s_j are orthogonal, the corresponding noise outputs y_j are independent, and hence:

$$\begin{aligned} P(y_1, y_2, \dots, y_{M-1} < y_M) &= \prod_{j=1}^{M-1} P(y_j < y_M) \\ &= [P(y_j < y_M)]^{M-1} \end{aligned} \quad (33)$$

If white Gaussian noise is assumed,

$$P(y_j < y_M) = \int_{-\infty}^{y_M} p_M(y_j) d(y_j) = \int_{-\infty}^{y_M} \frac{1}{\sqrt{\pi N_0}} e^{-\frac{y_j^2}{N_0}} d(y_j) \quad (34)$$

where: $p_M(y_j)$ is the probability distribution of the output of filter j , which is not matched to the incoming signal, when signal M is sent

Now substituting the Gaussian expression for $p_M(y_M)$ and $P(y_j < y_M)$, and writing $P_{ch} = 1 - P(c)$, the P_{ch} for orthogonal-coherent cases is given by the following (for simplification let $y = y_M$ = output of desired matched filter):

$$P_{ch} = 1 - \frac{1}{\sqrt{\pi N_0}} \int_{-\infty}^{\infty} e^{-\frac{(y-E)^2}{N_0}} \left[\int_{-\infty}^y \frac{1}{\sqrt{\pi N_0}} e^{-\frac{y_j^2}{N_0}} d(y_j) \right]^{M-1} dy \quad (35)$$

where: E = energy per character or symbol.

N_0 = single sided noise power density

In general this integral cannot be evaluated in closed form. However, computer calculations have been published (Refs. 36, 37 and 12). It should also be noted, for $M=2$, that this equation provides the same result as the binary coherent, orthogonal case treated in Sections 2.3 and 3.1.2 (for $M=2$, $P_{ch}=P_B$). In Section 3.1.2 a slightly different P_e derivation is referred to.

To find the appropriate E for the sine wave signal, Eq. 15 of Section 3.1.2, may be used. Note that this E is now energy per character (as opposed to energy per bit E_b).

Incoherent Orthogonal. For the M-signal incoherent system (for example, a multiple-FSK system using M incoherent frequencies), filters matched to the carrier pulse shape may be used (see Fig. 3), followed by envelope detectors and a decision device which chooses the filter having the greatest output at the end of the transmission interval. Because of the envelope detectors, the probability distributions at the outputs will be those of either a noise envelope, or a signal-and-noise envelope.

Assuming a maximum likelihood detection, the general equation (32) still holds. But now, using the distribution for the envelope of "noise alone" (Ref. 5) the probability of y_M exceeding any y_j is given by:

$$P(y_j < y_M) = \int_0^{y_M} y_j e^{-y_j^2/2} dy_j = 1 - e^{-\frac{y_M^2}{2}} \quad (36)$$

Also, the distribution $p_M(y_M)$ for "envelope of signal plus noise" is given by (Ref. 5):

$$p_M(y_M) = p_M(y) = y e^{-\frac{y^2 + 2E/N_0}{2}} I_0 \left(y \sqrt{\frac{2E}{N_0}} \right) \quad (37)$$

where: E = energy per character or symbol
 N_0 = single-side noise power density

Using these relations in Eq. 32, the P_{ch} is given by (Ref. 11):

$$P_{ch} = 1 - \int_0^\infty \frac{y}{N_0/2} e^{-\frac{y^2 + 2E}{N_0}} I_0 \left(\frac{y/\sqrt{2}}{N_0/2} \right) \left[1 - e^{-\frac{y^2}{2}} \right]^{M-1} dy \quad (38)$$

The term inside the bracket may be integrated and expanded into the binomial series:

$$\left[1 - e^{-\frac{y^2}{2}} \right]^{M-1} = \sum_{r=0}^{M-1} (-1)^r \binom{M-1}{r} e^{-\frac{ry^2}{2}} \quad (39)$$

Then one can complete the second integration, and the result is:

$$\text{Incoh-Orthog: } P_{ch} = \frac{1}{M} e^{-\frac{E}{N_0}} \sum_{r=0}^{M-1} (-1)^r \binom{M-1}{r} e^{\frac{E}{N_0} r} \quad (40)$$

where: E = energy per character or symbol

N_o = single-side noise power density

For $M=2$, ordinary FSK, this reduces to $P_B = 1/2 e^{-E_b/2N_o}$, the expression previously derived for uncoded binary signalling.

The above expression (40) holds for any orthogonal system employing envelope detectors, since only the assumptions needed are that the M signals are orthogonal, envelope-detected, and perturbed by white Gaussian noise.

Again, the proper expression for the E of the sine wave signal is as given in Eq. 15.

Bi-Orthogonal Signals. A signal which is better than coherent orthogonal signals (for P_e purposes) is a set of M -ary orthogonal signals and their negatives called, "bi-orthogonal signals." For example, Kineplex can be considered as using such a set, along with differentially-coherent detection. Such signals require coherence since one must be able to distinguish between a signal and its negative.

Two "natural" examples of bi-orthogonal signals are:

1. The Kineplex signal, where the cosine and sine of a single frequency are orthogonal, and two phases of each wave are available (a signal and its negative)---natural in frequency.
2. A block coded binary (see below, and Ref. 12), where coded blocks of binary digits are made orthogonal to each other (natural in time), and the two phases are available so that a negative for each signal is available.

We now show the calculation of P_{ch} for this case. The basic idea is similar to that of the orthogonal case (Eq. 32), except that now one must specify that $y_{M/2}$ have the proper sign (since there will be a signal which is the negative of the $M/2$ th signal that produced $y_{M/2}$).

Thus the basic equation for P_{ch} is the following (for simplification let $y = y_{M/2}$ = output of desired matched filter):

$$P_{ch} = 1 - \text{Prob} \left[\left\{ |y_j| < |y| \right\} \text{ and } y > 0 \right] \quad \text{for } j = 1, 2, \dots, (M/2 - 1) \quad (41)$$

$$= 1 - \int_0^{\infty} p_{M/2}(y) [P(|y_j| < |y|)]^{\frac{M}{2} - 1} dy \quad (42)$$

$$\text{Now: } P(|y_j| < |y|) = \int_{-y}^{+y} p_{M/2}(y_j) dy_j \quad (43)$$

Using this, the P_{ch} can be shown to be (Ref. 12):

$$P_{ch} = 1 - \frac{1}{\sqrt{\pi N_0}} \int_0^{\infty} e^{-\frac{(y-E)^2}{N_0}} \left[\frac{1}{\sqrt{\pi N_0}} \int_{-y}^{+y} e^{-\frac{y_j^2}{N_0}} dy_j \right]^{\frac{M}{2} - 1} dy \quad (44)$$

$$\text{Bi-Orthogonal: } P_{ch} = 1 - \frac{1}{\sqrt{\pi}} \int_0^{\infty} e^{-\frac{(y - \sqrt{E/N_0})^2}{N_0}} [\text{erf}(y)]^{\frac{M}{2} - 1} dy \quad (45)$$

where: E = energy per character or symbol

N_0 = single-side noise power density

This expression has also to be evaluated on a computer; curves have been published for block coded binary cases, and these are immediately convertible into the general cases.

It may be noted that the binary simplex signal [$s_1(t) = -s_2(t)$] is also a bi-orthogonal case. Hence, when $M=2$, the above expression is equivalent to:

$$P_{ch} = P_B = \frac{1}{\sqrt{2\pi}} \int_0^{\infty} e^{-\frac{y^2}{2}} \frac{1}{\sqrt{\frac{2E_b}{N_0}}} dy \quad (46)$$

which is the equation used in Section 2.3 and 3.1.1.

Bi-orthogonal signalling requires only half the bandwidth and half the correlators (or matched filters) required by orthogonal coherent signalling. For these reasons, such signals have great attractiveness from the engineering standpoint.

Comparative curves will be given later.

Simplex Signals. From the purely theoretical standpoint, the best set of signals when considering P_e are the simplex signals (Section 2.2).

It is remembered that (Ref. 3) simplex signals, viewed as points in signal space, correspond to a geometrical regular simplex where every vertex is equally distant from every other vertex. If there are M signals, the geometrical simplex appears in an $M-1$ -dimensional space. In one-dimensional space the signal is a binary PSK wave; in two-dimensional space, a 3-phase wave at 120° ; in 3-dimensional space ($M=4$) no signal is easily describable, but the simplex is easily constructed (Ref. 21).

However, the P_e calculation for M -ary simplex signals is difficult to achieve in closed form. This is because the noise outputs of the unmatched filters (see Fig. 2) are not independent. However, one can obtain the P_{ch} 's from the coherent orthogonal case since one can show (Ref. 16) that use of a simplex set instead of an orthogonal set is equivalent to an increase in energy of $M/M-1$. Hence, if one has the curves for coherent orthogonal, one can immediately plot them for simplex signals.

Remember that, for the binary case, $s_1(t) = -s_2(t)$ is a simplex signal.

3.2.2.2 Basic M-ary Symbols. The general P_{ch} relations for M -ary signals were given above. As mentioned before, two ways of obtaining the M -ary signals are: (1) basic M -ary symbols, and (2) block-coded binary. First we consider basic M -ary symbols.

For M -ary basic symbols one can have:

1. Simplex signals, constructed generally so that the M points correspond to a geometrical simplex.
2. Bi-orthogonal, sine and cosine of given frequency orthogonal, and negative of each frequency.
3. Orthogonal Coherent (M -ary, coherent FSK).
4. Orthogonal Incoherent (M -ary, incoherent FSK).
5. M -phases, coherent or differentially coherent (see below).

The M -phase signal is included here because we will later use this signal as an example of a close-packed signal, (see Section 2.2). For the other signals, the P_{ch} equations given above can be used directly, by determining the proper E for each situation.

3.2.2.3 Block-Coded Binary Signals. Block-coded binary signals represent a very practical way to implement "more energy per decision." The practicality comes from the fact that binary operations are involved, both in the transmitter and the receiver.

Also, it is possible to get the "best" signals rather easily this way, as we now note.

M-ary simplex signals can be obtained by using linear shift register sequences (Ref. 14) to form the binary block. For this, one uses (with an $L = 2^n - 1$ long sequence) the entire length of the sequence, and the L signals are obtained from using the sequences one clock step delayed from each other. It is well known that the correlation between such signals is $-1/L-1 = -\frac{1}{M-1}$ (Ref. 14).

Orthogonal block codes can be generated in a number of ways, and shift-register implementations are also possible (Ref. 12).

Bi-orthogonal signals (discovered by Muller and Reed, Ref. 15) also can be readily generated by shift-register means (Ref. 12). The Digilock communications system uses such bi-orthogonal codes. It is a telemetry data link where 5-bit words are shift-register encoded into (16, 5) bi-orthogonal codes (32 possible words).

To obtain the P_{ch} for any binary block-coded symbol one uses the appropriate equation, 35, 40, or 45. Again it is necessary to find the appropriate E . It must be stressed that the E in each of these equations refers to the energy per complete character. Hence the energy per bit E_b must be multiplied by the "number of bits per character."

3.2.2.4 M-Phases. A set of M-ary signals, somewhat different from those considered above, is the use of M-phases in coherent situations. For our purposes, these signals represent one example of a "close-packed signal" (if $M > 8$). These signals are unique in that they only require 2 dimensions, no matter what the level of M (see Table I). One significance of this lies in the ultimate capacity that one can achieve when S/N is very high.

A comment on M-phase signals, relative to the "simplex, bi-orthogonal, and orthogonal" category, is pertinent here. Each of these three signals (treated above) is essentially based on the idea of being relatively far apart, so that a certain degree of immunity to noise is achieved by sacrificing bandwidth. The basic worth of these signals is thus intuitively obvious, when conditions are such that one must resort to an M-ary signalling situation.

M-phase signals, on the other hand (for $M > 2$), are based primarily on the idea of conserving bandwidth at the cost of some S/N ratio (i. e., poorer P_{ch}); also, when E/N_0 is high (so that P_{ch} requirements can be met), the M-phase signal is a better signal (BW-

wise) than uncoded binary.

Two methods of operation are possible with M-phase signals: coherent (CPSK) and differentially coherent (DCPSK). In coherent operation the receiver must have a phase coherent signal (phase coherent with the transmitted signal). The receiver then appears exactly as in Fig. 2, where the $s_j(t)$ are gated sinusoids at the various phases.

For differentially coherent (DCPSK) operation, the information is extracted from "phase now relative to phase in prior symbol." Here it is only necessary that phase coherence be held over one symbol length. This operation (used by Kineplex) is more easily realized in practice than strictly coherent receivers. One result of this operation is that errors tend to occur in pairs.

For calculating P_{ch} of M-phase signals a different procedure from those above must be considered, since now the noise outputs of the "unmatched" filters are not independent (the "simplex" outputs are also not quite independent). For this reason it is of interest to note the P_{ch} calculation for M-phase. A brief sketch of a P_{ch} calculation for M-phases (taken from Ref. 13) will now be given.

Coherent M-Phases. Again we are interested in the P_{ch} for a "maximum likelihood receiver." Such a receiver would be as shown in Fig. 2, where the correlator or matched filter is matched to each phase.

Each signal of an M-phase system can be written as

$$s(t) = \sqrt{2S} \cos(\omega_0 t + \theta) \quad (47)$$

where: S is the received signal power,
 ω_0 is the angular center frequency, and
 θ is chosen from a discrete set $2\pi k/M$,
for $0 \leq k \leq M-1$.

The detection problem amounts to deciding which of the M possible transmitted phases is the received phase. Assuming the M phases to be equiprobable, the receiver chooses the θ which has the maximum a posteriori probability at the detector. This can be calculated (Ref. 13) by determining the probability that the received signal phase falls within the section

$$\frac{-\pi}{M} < \theta < + \frac{\pi}{M} .$$

Assuming that the received signal is narrowband, perturbed by additive Gaussian noise, the interfering noise can be written:

$$n(t) = I_c \cos \omega_0 t + I_s \sin \omega_0 t \quad (47a)$$

where: I_c and I_s are low-frequency random variables with zero mean.

The total received signal is then

$$\begin{aligned} x(t) &= s(t) + n(t) \\ &= (\sqrt{2S} + I_c) \cos \omega_0 t + I_s \sin \omega_0 t \end{aligned} \quad (48)$$

assuming that the desired phase θ is zero. The probability density of the phase of $x(t)$ can then be written, using the method of Rice (see also Ref. 10, p. 166):

$$p(\theta) = \frac{1}{2\pi} e^{-\frac{S}{N}} \left[1 + \sqrt{4\pi \frac{S}{N}} \cos \theta e^{\frac{S}{N} \cos^2 \theta} \cdot \psi(\sqrt{2 \frac{S}{N}} \cos \theta) \right] \quad (49)$$

where $\psi(x)$ is the cumulative Gaussian function:

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

Because the S/N ratio here is an input S/N, the conversion to E/N_0 is given by:

$$\frac{E}{N_0} = \frac{S}{N} \quad (50)$$

Then $p(\theta)$ appears as:

$$p(\theta) = \frac{1}{2\pi} e^{-\frac{E}{N_0}} \left[1 + \sqrt{4\pi \frac{E}{N_0}} \cos \theta e^{\frac{E}{N_0} \cos^2 \theta} \psi \left(\sqrt{2 \frac{E}{N_0}} \cos \theta \right) \right] \quad (51)$$

which is the probability density of the phase of the received signal. Integrating this density over $-\frac{\pi}{M} < \theta < \frac{\pi}{M}$ gives the probability that the received signal phase falls within the sector bounded by the decision levels $\pm \frac{\pi}{M}$; i. e., the probability of error for an M-phase system is

$$P_{ch} = 1 - \int_{-\pi/M}^{\pi/M} p(\theta) d\theta \quad (52)$$

This can be integrated by numerical or graphical means to give curves of error probability versus E/N_0 for given M (see Ref. 13).

For reasonably large x , the integral $\psi(x)$ can be approximated by the first terms of an asymptotic expansion; thus

$$\psi(x) \cong 1 - \frac{e^{-x^2/2}}{\sqrt{2\pi} x} \quad (53)$$

Using this in the expression for $p(\theta)$, the result is

$$p(\theta) \cong e^{-\frac{E}{N_0} \left[\sqrt{\frac{E}{N_0}} \cos \theta e^{\frac{E}{N_0} \cos^2 \theta} \right]} \cong \sqrt{\frac{E}{N_0}} \cos \theta e^{\frac{E}{N_0} (\cos^2 \theta - 1)} \quad (54)$$

Since $1 - \cos^2 \theta = \sin^2 \theta$, this can be written as

$$p(\theta) \cong \sqrt{\frac{E}{N_0}} \cos \theta e^{-\frac{E}{N_0} \sin^2 \theta} \quad (55)$$

Curves based on previous integrations of Eq. 52 (Ref. 24), which show P_{ch} vs. E/N_0 , are given by dashed lines in Fig. 12 (taken from Fig. 3, Ref. 13). Note that the ordinate is P_{ch} ; if bit error rate is desired, the P_{ch} must be converted to P_B . Also, the abscissa is in terms of E , where $E = \log_2 M E_b$.

Differentially Coherent M-Phase (DCPSK). For differentially coherent systems, the P_e calculation is quite involved. Here we will present only the results (Ref. 21 p. II-23) for comparison with the CPSK case above. The P_{ch} vs. E/N_0 curves are shown by the solid lines in Fig. 12. Comparing DCPSK with CPSK, it is seen that, for a P_{ch} of 10^{-4} , the db loss is 3.6 db for $M = 16$, and drops to 1 db for $M = 2$.

3.3 Comparative Results and Conclusions

Since the various signals and operation methods dealt with above represent alternatives available to a system designer, we wish here to draw sensible comparative graphs showing the relative advantage of the alternatives wherever possible. Note that in some cases

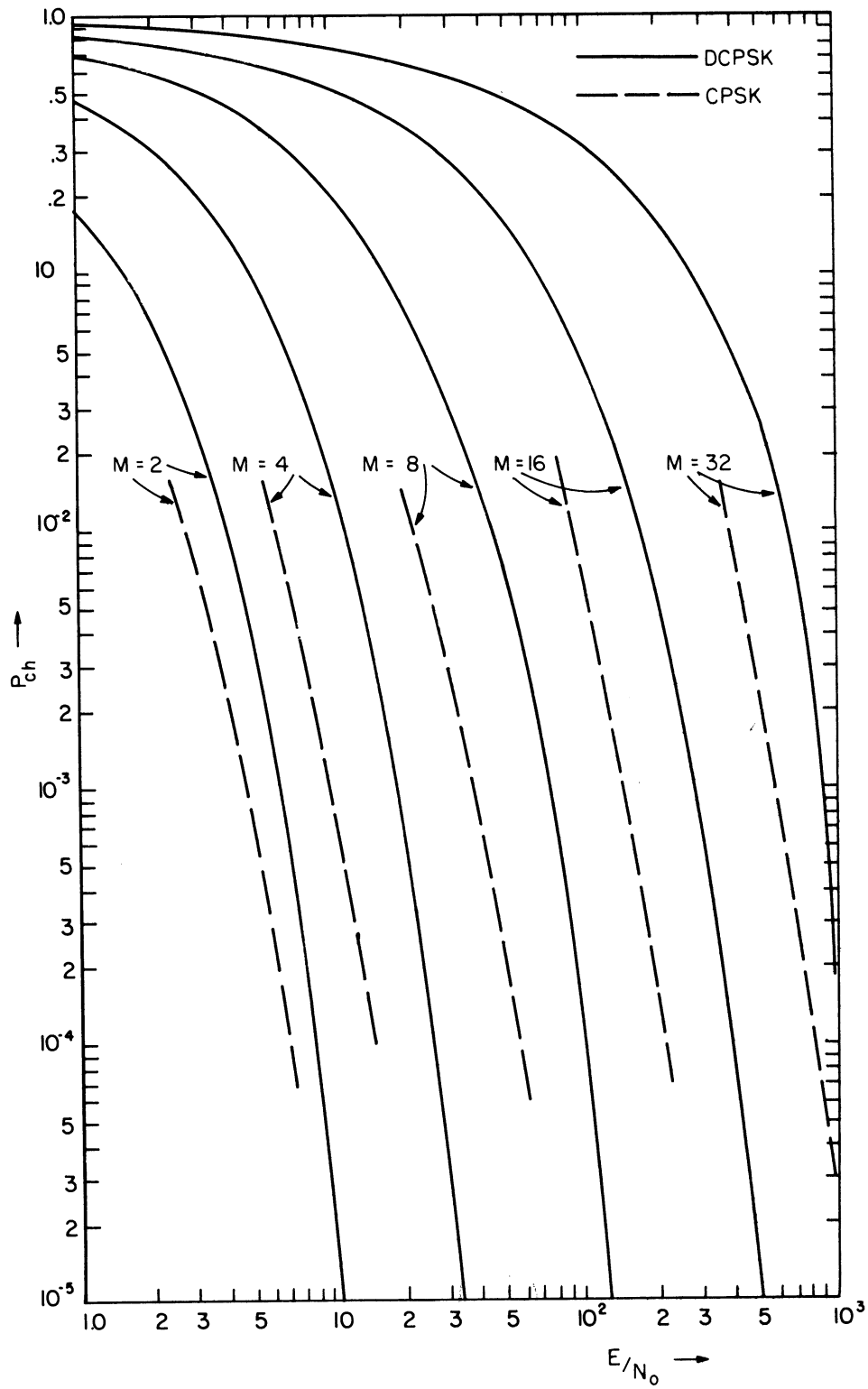


Fig. 12. P_{ch} versus E/N_0 for M-phase CPSK and DCPSK.

this is too laborious and so desired comparisons cannot be made without further research.

The objective in the following graphs is to show the P_e versus E_b/N_o for the various types of signalling. Also, the "theoretical" bandwidth for each case will be noted. The viewpoint will be that there is a certain power, S , available, and that this cannot be increased. The other three variables then are P_e , data rate, and bandwidth. In all cases we will make essentially "constant data rate" comparisons; this is occasioned by plotting all curves in terms of "energy per bit/noise power per unit bandwidth." If the energy per bit is constant (a vertical line on the graphs), and power is constant, then a constant data rate will result.

The comparisons that one makes depend somewhat on the priority of P_e versus bandwidth, assuming a constant data rate. In Sections 3.3.1 and 3.3.2 it is implied that P_e is more critical than BW, and the P_e 's of the signals are compared. Section 3.3.3 compares signals in terms of conserving BW at the expense of P_e 's.

When comparing the advantages of the previous alternative signals, where P_e is deemed essential, it is sensible to consider two separate cases: (1) bit-error comparisons, and (2) word-error comparisons. If the basic information is a sequence of independent binary bits, (or if one is comparing different types of systems) the proper criterion is the probability of error of the bits, and one is interested in how the various alternatives described above affect this bit probability of error. On the other hand, if the basic information consists of such things as teletype or any sampled analog data, where the basic information is already groups of binary bits, then the important criterion (and the criterion which the user will want to talk about) is the P_w of the "word."

Consequently, the following comparisons will be made on these two separate bases. We begin with the bit-error comparison.

3.3.1 Bit-Error Comparison. The first item of interest is to consider the alternative types of signals when one is transmitting straight binary. The possible signals are: simplex, coherent orthogonal, and incoherent orthogonal. A graph of P_B vs. E_b/N_o for each of these three types of signals was shown in Fig. 4 of Section 2.3. One can find, for a given N_o , what P_B can be obtained for a given energy if one uses either of the three types of signals.

Error-Coded vs. Binary Signals. We simply wish to note here that the probability of bit-error comparison between error-coded and binary signals is not yet possible. Quite laborious calculations would be involved in attaining this. It will be remembered that in Eq. 30 of Section 3.2.1 the item discussed was the probability of a word error--- that is,

the probability that the amount of digits in error is greater than the correcting capability of the error code. In order to convert this P_w to a bit P_e it is necessary to know the "average number of information digit errors per error-coded word error." This is quite a laborious calculation and requires effort devoted singly to it. This calculation was not deemed of sufficient worth to the goals of the contract to warrant the necessary effort.

M-Symbol Alphabet vs. Binary. The first graphical comparison will compare the bit P_e for M-symbol alphabet vs. binary.

As mentioned previously, the chief occasion for considering M-ary signalling as opposed to binary is when the P_e is not good enough (for binary) with the given E_b/N_o and data rate. If one uses M-ary symbols ($M > 2$) one is in effect increasing the energy per decision. If one keeps the data rate and the energy per bit constant, one can obtain improved P_e 's at the sacrifice of a slight delay (until decision on M-ary symbol is possible), some equipment complexity, and increased bandwidth.

Figure 13 shows the comparison of binary versus M-ary for coherent, bi-orthogonal signalling. The basic data were obtained from the computer calculations of Eq. 45 as reported in Ref. 12. Since the basic data are character probabilities, these data were converted to bit errors by multiplying by $\frac{2^n-1}{2^{n-1}}$ (where $n = \log_2 M$). This factor is approximately "the average number of bit errors per word error." The reasons for choosing bi-orthogonal cases as the coherent comparison (instead of orthogonal or simplex) are: (1) if one can use coherence, the bi-orthogonal signalling gives better P_e 's than orthogonal signals; and (2) the bi-orthogonal signalling is "more practical" than either orthogonal or simplex, since only $M/2$ correlators and half the bandwidth are required.

Figure 14 shows the binary-M-ary comparison for incoherent orthogonal signalling. These data were obtained from Eq. 40, again multiplying the character error-rate by $\frac{2^n-1}{2^{n-1}}$ to obtain P_B .

Figure 15 compares binary-M-ary for coherent M-phases. The data here were obtained by multiplying the P_{ch} of Fig. 12 by $\frac{2^n-1}{2^{n-1}}$ to obtain the approximate P_B , and converting the E to E_b by using $E = \log_2 M E_b$.

The "theoretical BW's" shown on each of the above curves are taken from the values in Table I, and deserve comment here. These BW's would be realistic under the following conditions: if a large number of communicators all used the same signalling meth-

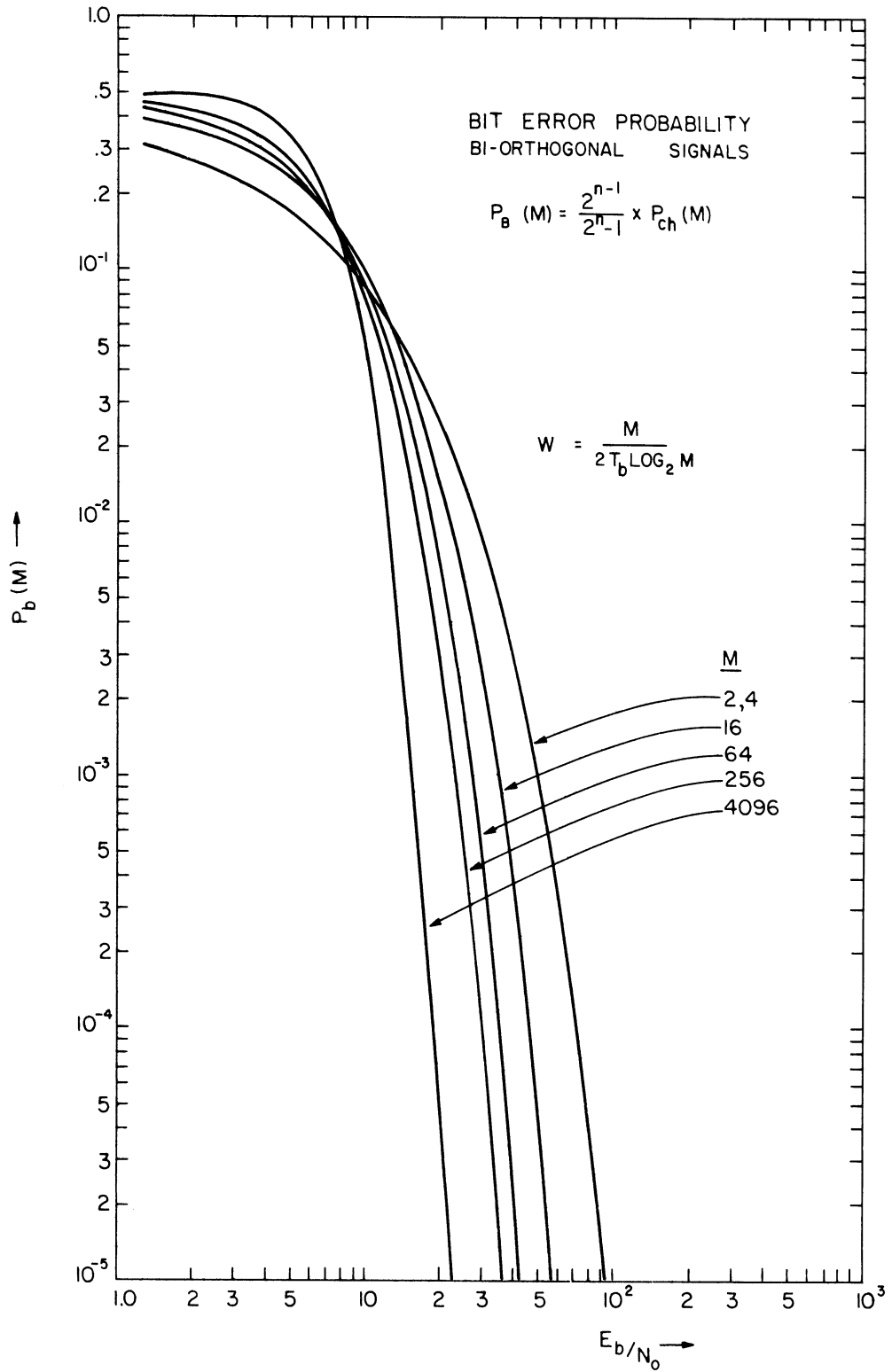


Fig. 13. $P_B(M)$ versus E_b/N_0 for coherent, bi-orthogonal signals, comparing binary with M-ary.

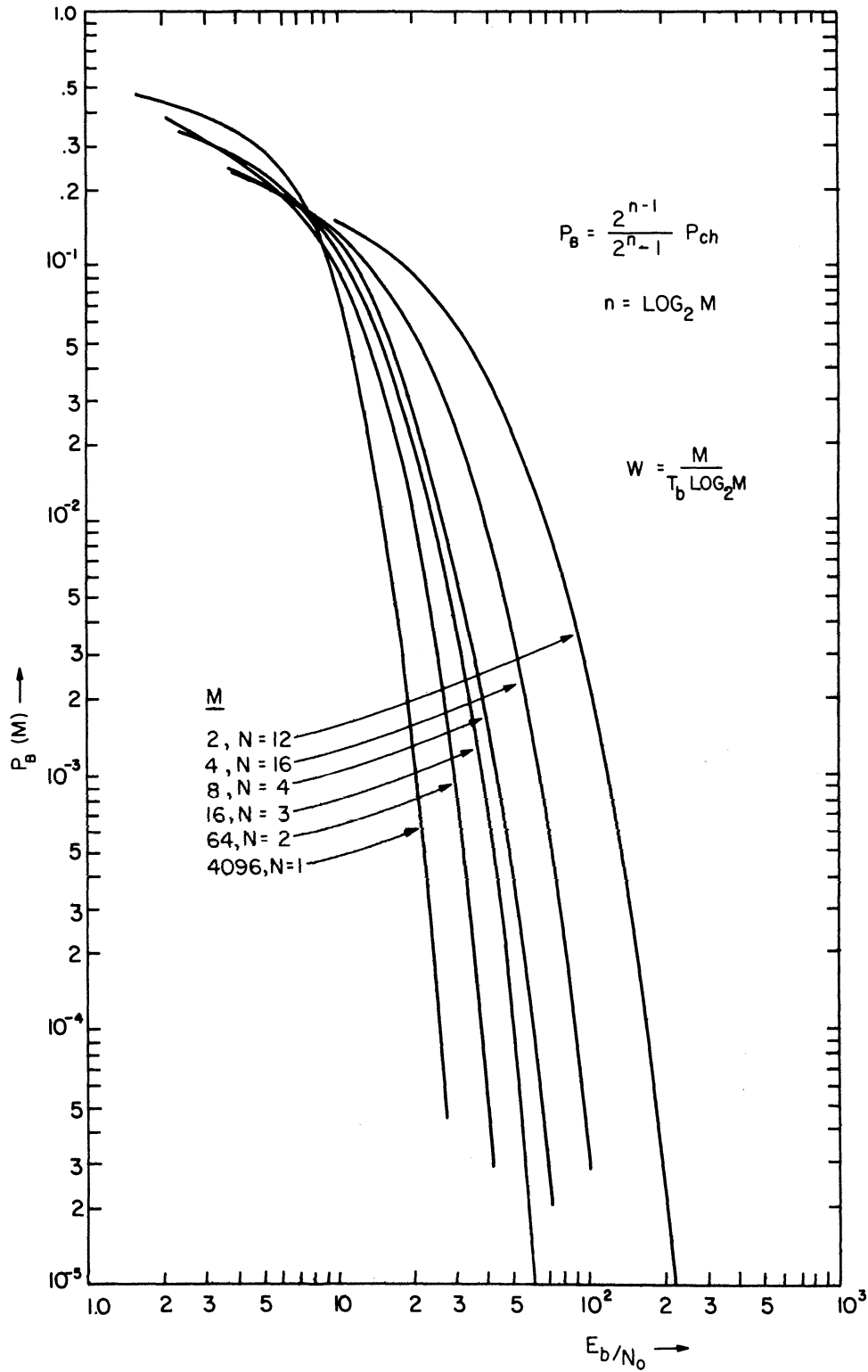


Fig. 14. $P_B(M)$ versus E_b/N_0 for incoherent, orthogonal signals comparing binary versus M-ary.

od, and controlled their frequencies with respect to each other, each channel would consume the BW shown. We might call this the "controlled-environment" BW. These BW's are established by writing an orthogonality integral, to make "crosstalk" between channels theoretically zero. Note that, while adjacent channels may be noninterfering with this BW, each channel actually uses a much larger spectrum (although the major energy is in the central lobe), so that, for example, square-wave video is attainable. Although the practical use of these BW's is limited, they do serve well for comparison purposes.

A more practical BW for each signal is also given in Table I. Here the criterion is that the main lobe of each signal is used, and everything outside filtered severely. Note that, if one uses these BW's in practice, previous theoretical calculations must be altered. For example, orthogonal signals will not be quite orthogonal; video pulses will be $\frac{\sin x}{x}$ type, rather than square-wave; and simple energy calculations for signals (Eq. 15) will have to be altered. The actual practical BW's of the various signals in an "uncontrolled environment" require substantial work, and must be tailored to the situation since results depend upon the assumptions.

3.3.2 Word-Error Comparisons. If the original data are words, it is the word-error probability which is of interest. We wish to compare the P_w 's of M-symbol alphabet versus binary.

In order to compare word errors sensibly, one needs "equivalent words." That is, it makes little sense to compare the error-rate of an n-bit word with that of an m-bit one. Therefore all the comparisons below are made in terms of a 12-bit word. As suggested in Ref. 22, the 12-bit word is good for M-ary comparisons because $2^{12} = 4^6 = 8^4 = 16^3 = 64^2 = 4096$.

In the following curves the P_w for a 12-bit word is plotted vs. E_b/N_o ---the "energy per bit" over N_o . If the data rate is assumed constant, and a maximum power S assumed, then the energy per bit remains constant.

Figure 16 shows the 12-bit P_w for the three types of binary operation. The P_w is related to the P_B by:

$$P_w = 1 - (1 - P_B)^n = 1 - (1 - P_B)^{12} \quad (55a)$$

The P_B 's for these curves were taken from Fig. 4.

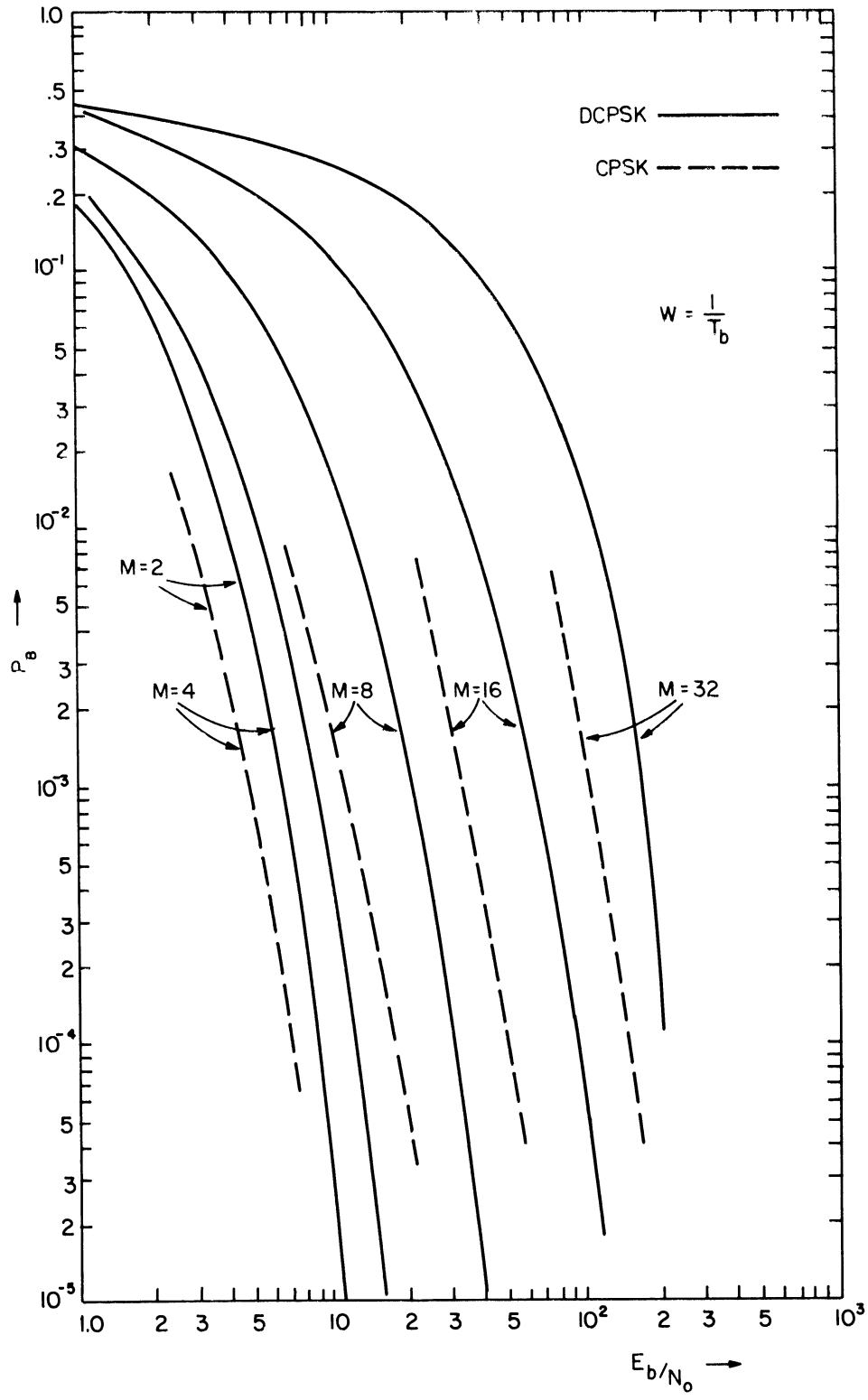


Fig. 15. P_B versus E_b/N_0 for M-phase (CPSK and DCPSK).

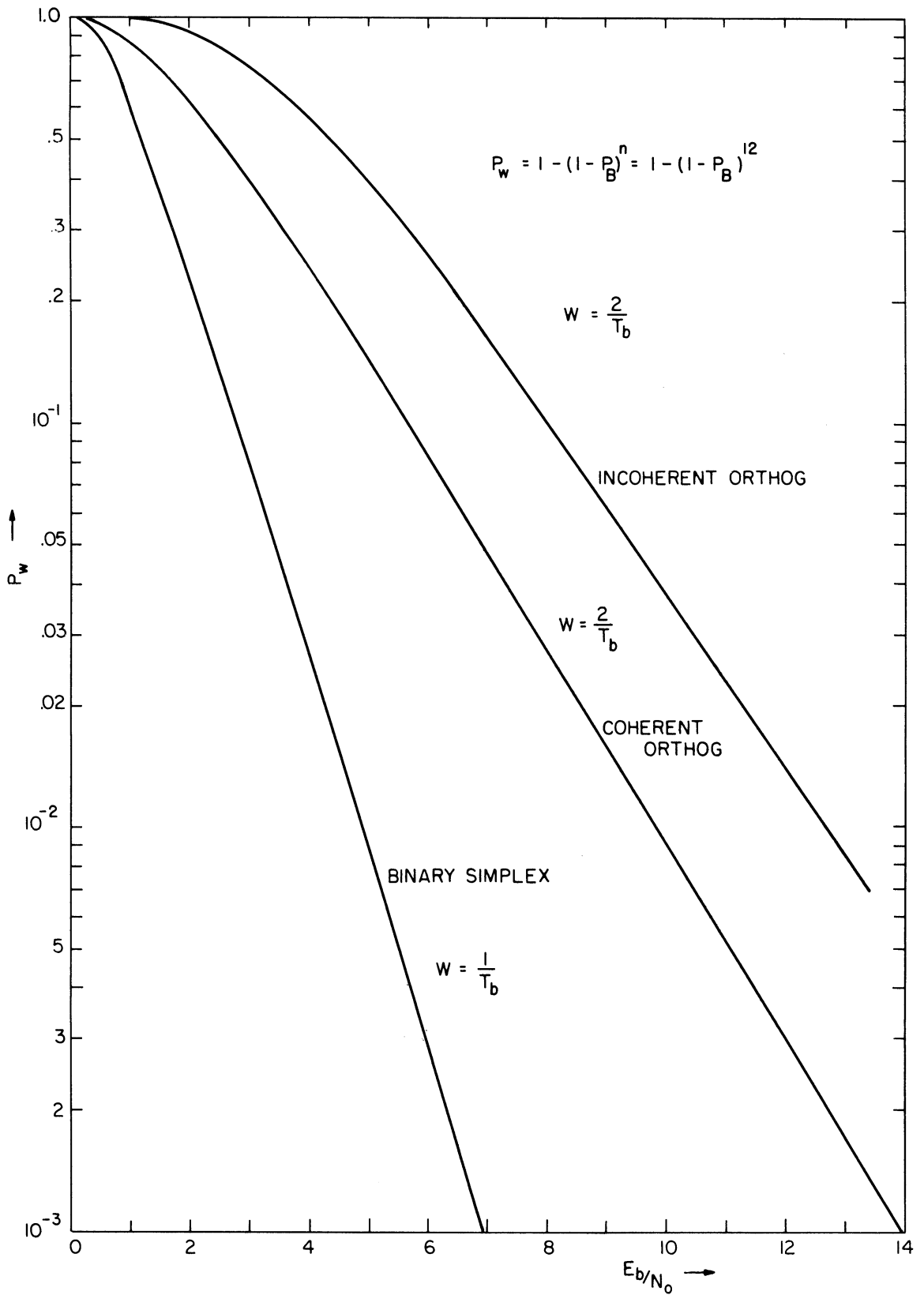


Fig. 16. P_w versus E_b/N_0 of 12-bit word for three types of binary transmission.

Figure 17 shows the binary-M-ary comparisons for a 12-bit word using bi-orthogonal signalling. This would be the coherent comparison. For these curves the P_{ch} is calculated from Eq. 45, and P_w obtained by using:

$$P_w = 1 - [1 - P_{ch}(M)]^N \quad (56)$$

where: $2^{12} = M^N$

The curves of Fig. 17 were taken from Ref. 22, Fig. 3.

Figure 18 is the incoherent (orthogonal) comparison. The P_{ch} are calculated using Eq. 40, and then Eq. 56 is used to obtain P_w . The data plotted in Fig. 18 were not recalculated, but were taken from Ref. 22, Fig. 4 after due checking.

Figure 19 shows the 12-bit P_w vs. E_b/N_o for M-phase signals. This curve was found by using Eq. 56 and obtaining the P_{ch} from Fig. 15. Since the M-phase signal requires coherence, it is sensible to compare Fig. 19 with Fig. 17.

A remaining comparison is error-coded versus binary. Since a calculation for a 12-bit word was too laborious, only a 5-bit word was used. Curves for a (16, 5) and a (32, 5) error code, both coherent and incoherent, were shown in Fig. 11. Note that, in this comparison, the bandwidths are not the same. The bandwidth expansion for error-coding for the same data rate is equal to the digit expansion (16/5 and 32/5) of the code.

3.3.3 Information-Rate Comparison. The curves of the above two sections are most directly useful where P_e is the essential consideration, and where BW is more or less flexible. Those curves suggest what one must do if a given E_b/N_o and given data rate do not provide a suitable P_e ; also, the curves depict the cost in bandwidth as M increases.

If the major consideration is conservation of bandwidth, and P_e is somewhat flexible, then the quantity of interest is the "information-rate per BW" vs. E_b/N_o . Such curves would depict how to get a given information-rate through a specified BW, but the high information rate may have a great many errors. Figures 20 and 21 show "information-rate/BE" vs. E_b/N_o for the two cases shown before (bi-orthogonal, incoherent orthogonal).

The information rate (I), as opposed to data rate (R), is a measure of the average number of correct symbols (or bits) per unit time in a noisy situation. If there is no noise, the information rate equals the data rate.

For the orthogonal cases, the information rate is given by (Ref. 6):

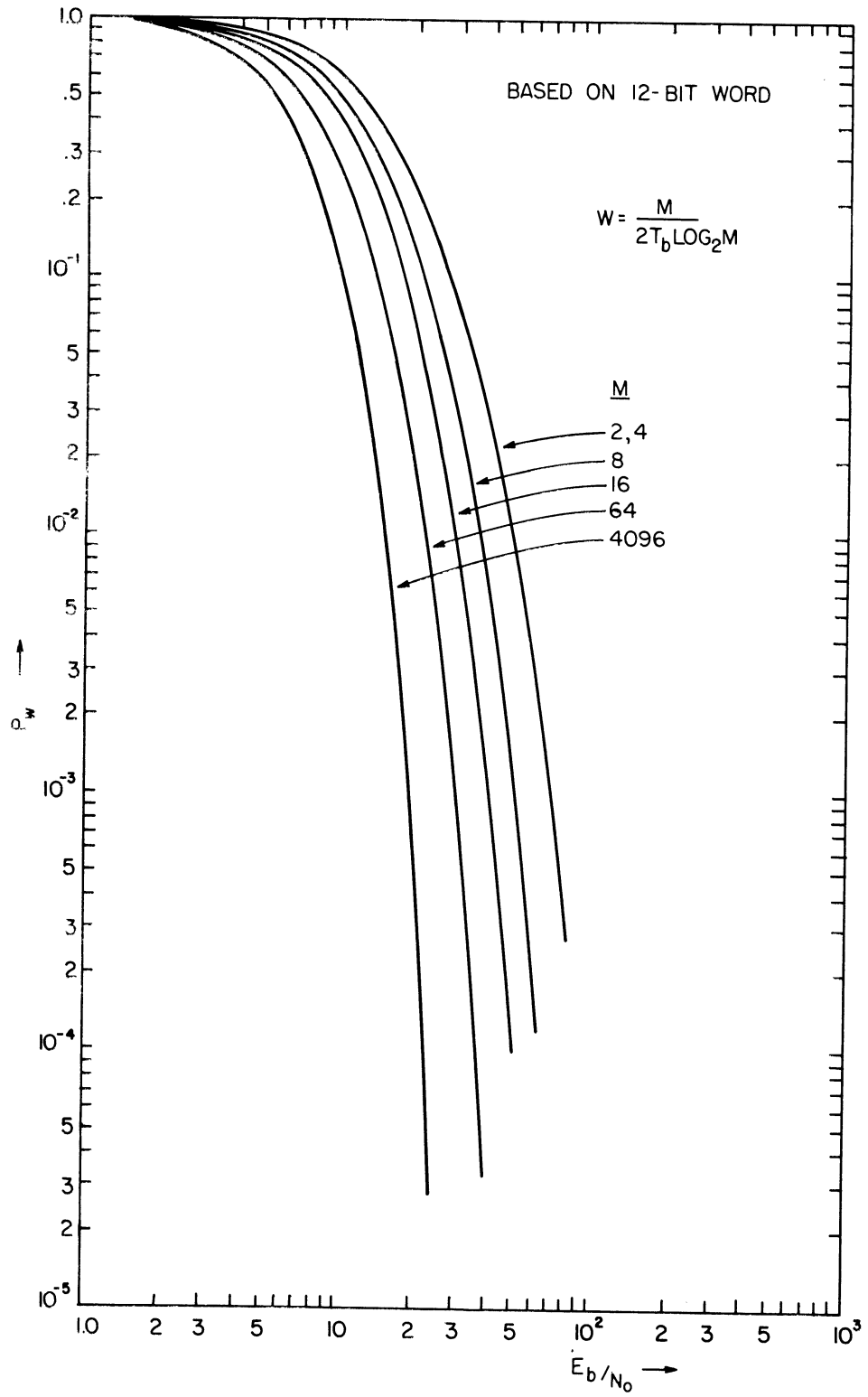


Fig. 17. P_w versus E_b/N_0 for 12-bit word with bi-orthogonal signals.

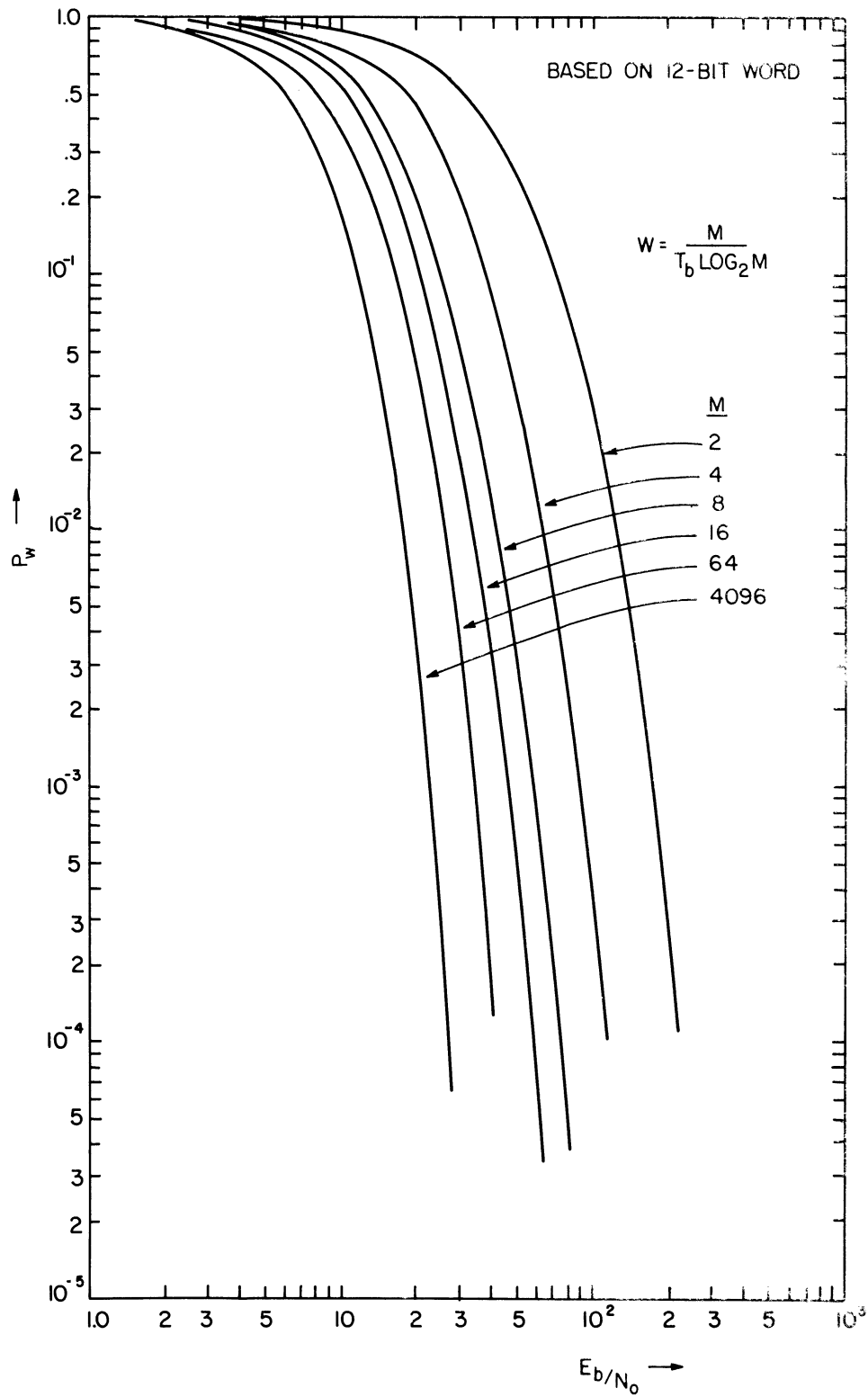


Fig. 18. P_w versus E_b/N_0 for 12-bit word with incoherent orthogonal signals.

$$I = \frac{1}{\log_2 M \times T_b} \left[\log_2 M + P_{ch} \log_2 \frac{P_{ch}}{M-1} + (1-P_{ch}) \log_2 (1-P_{ch}) \right] \text{ bits/sec} \quad (57)$$

where: I = information rate = average number of correct symbols

$\log_2 M \times T_b = T$ = length of general M-ary symbol

T_b = length of equivalent bit

P_{ch} = probability of an M-ary character error.

For the binary (orthogonal) case, this reduces to the well-known equation:

$$I = \frac{1}{T_b} \left[1 + P_B \log_2 P_B + (1 - P_B) \log_2 (1 - P_B) \right] \text{ bits/sec} \quad (58)$$

The value of I was divided by the "theoretical controlled-environment BW's" shown in Table I (and on the previous curves) to obtain the data for Fig. 21, the incoherent orthogonal case. The $P_{ch}(M)$ for each curve was obtained from Eq. 40 of Section 3.2.2.

For the bi-orthogonal case, available data in tabular form from Ref. 12 was used to obtain the I . Again the BW's shown in Table I were used.

3.3.4 Comparison Conclusions. To review, curves have been plotted of P_B , P_w , and I/BW vs. E_b/N_o for bi-orthogonal, incoherent orthogonal, and M-phase cases. Since the abscissa is always E_b/N_o , it is implied that one is making a constant data rate comparison. This follows if one assumes that the power S is constant; then, for a constant data rate (constant T_b), the E_b will be constant. We will be making variable rate comparisons in Section 4.3, when discussing "capacities."

For a constant rate comparison, it is sensible to think of two regions, low E_b/N_o and high E_b/N_o . If the E_b/N_o is high, the problems are fewer; that is, one can meet the P_e requirements with a bandwidth efficient signal (such as M-phase). If the E_b/N_o is low (but greater than 1), then one must emphasize either P_e or W . Hence we will consider two cases for low E_b/N_o : (1) P_e is most important, and W relatively expendable; and (2) W most important, and P_e relatively expendable.

If P_e is most important (for low E_b/N_o where $1 < E_b/N_o < 20$), and one is restricted to a given power S , then it is clear that one must use an M-ary orthogonal signal sufficient to give the desired P_e . From Figs. 15 and 19 it is seen that M-phase signals where $M > 8$ do not provide suitable P_e 's. From Figs. 13, 14, 17 and 18 it is seen that,

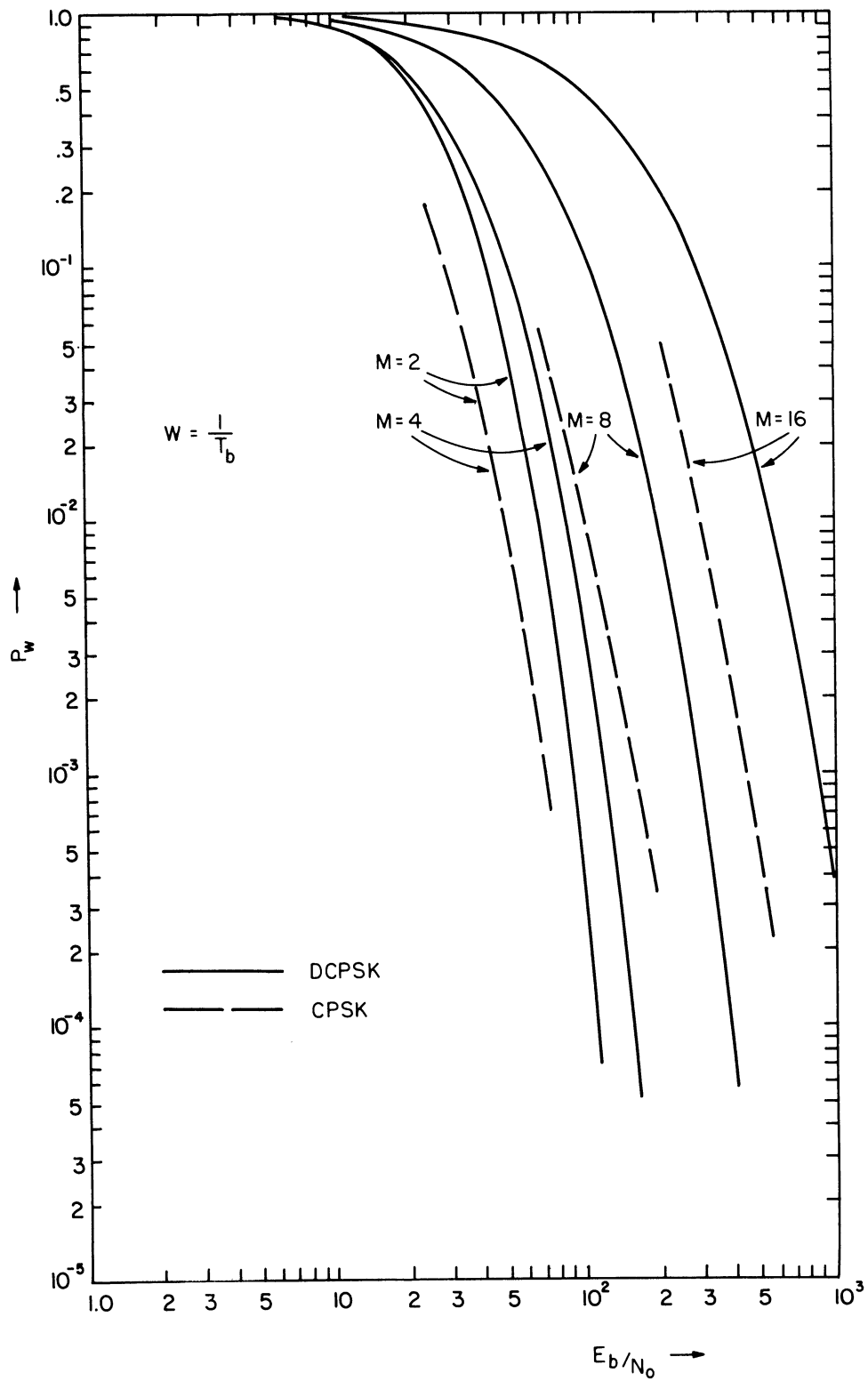


Fig. 19. P_w versus E_b/N_0 for 12-bit word with CPSK and DCPSK.

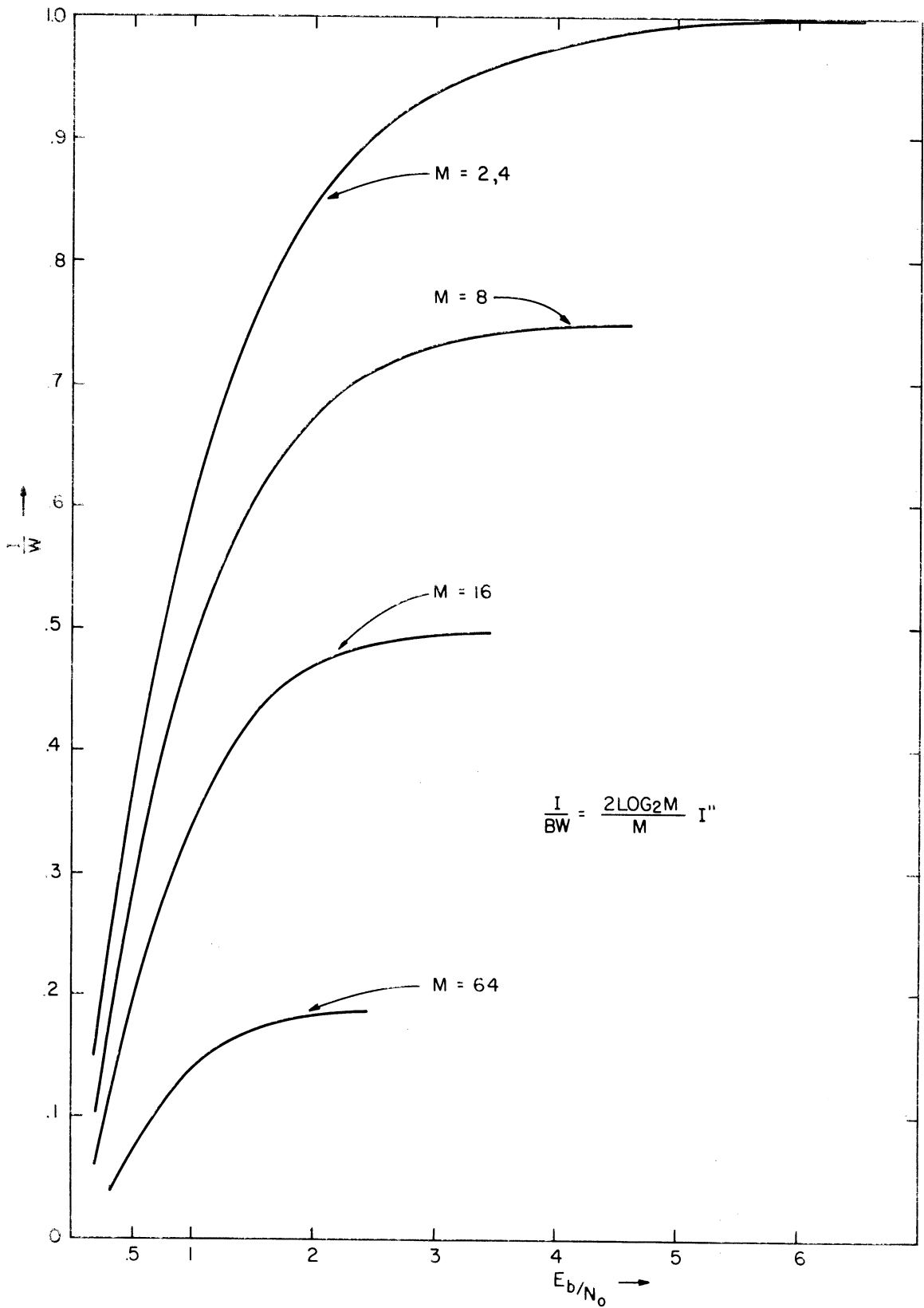


Fig. 20. Information rate/bandwidth for M-ary bi-orthogonal signals (coherent).

for a constant E_b/N_0 , increasing M with orthogonal signals improves P_e substantially. However, Figs. 20 and 21 give a measure of the cost in bandwidth of increasing M .

If one can consider increases in S (for a constant data rate), the steepness of the curves in the figures shows that this is more profitable than increasing M . For example, in the coherent case (Fig. 13), going from $M = 2, 4$ to $M = 4096$ is equivalent to a 3.5 db increase in power (S) for $P_e \approx 10^{-5}$. In the incoherent case (Fig. 14), going from $M=2$ to $M=64$ is about equivalent to increasing S by 6.8 db. Thus, to improve P_e for a constant data rate, increases in S (if possible) are more profitable than increasing M ; and this is more true for coherent than for incoherent operation.

If the bandwidth W is most important (for low E_b/N_0) then one must emphasize Figs. 20 and 21. Here the cost of using large M probably becomes prohibitive. From the figures, it appears that an M of 2 or 4 is suitable for coherent operation, and an M of 4 or 8 for incoherent operation. If W is crucial, one should probably avoid going beyond $M=8$, even for the incoherent case.

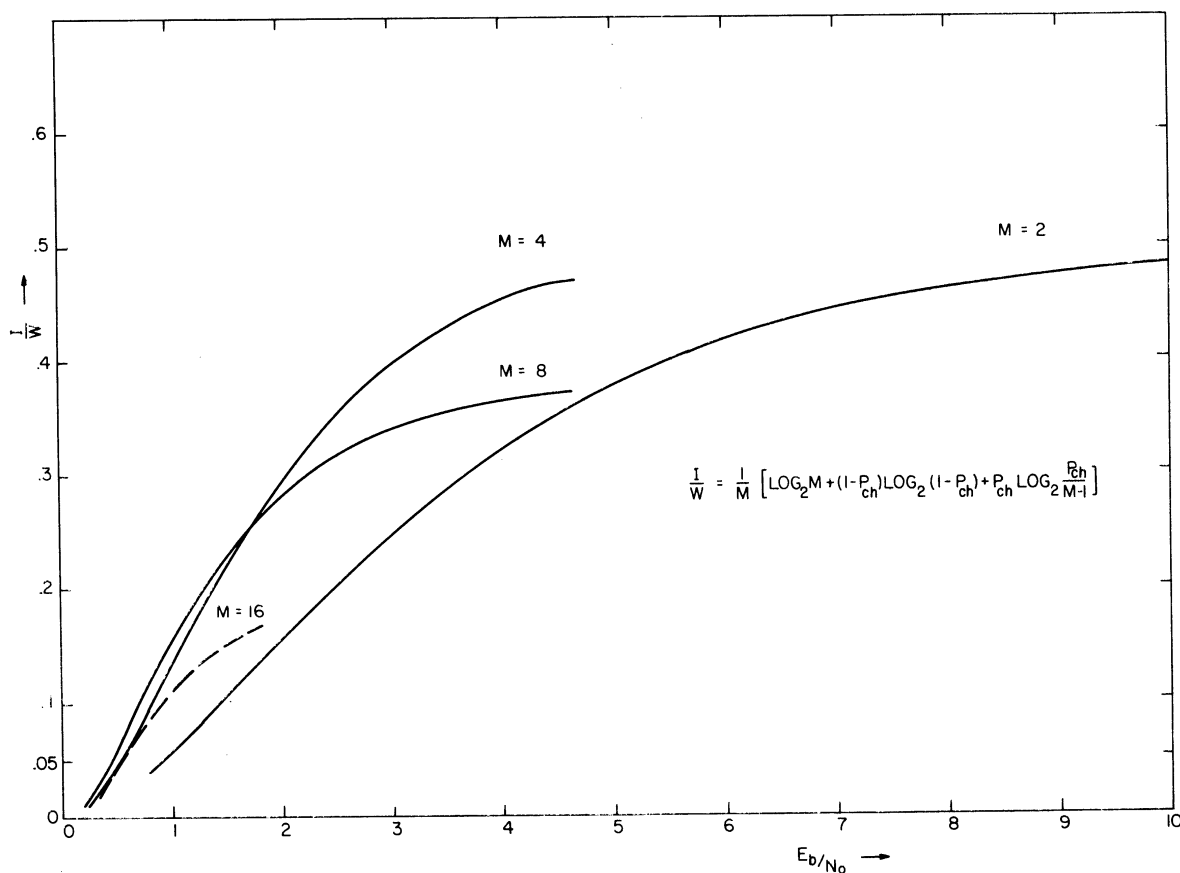


Fig. 21. Information rate/bandwidth for M -ary incoherent orthogonal signals.

Finally, if $E_b/N_o > 20$, so that P_e of the order of 10^{-5} can be obtained with "correlated signals," it appears clear that one should use signals that stress "high information/W" rather than "distinctness" (simplex, orthogonal, etc.). The signals of interest then are "close-packed," of which M-phase with $M > 8$ is one example. Figures 15 and 19 show the P_e 's for M-phase signals. Considering that M-phase signals become increasingly correlated with $M > 8$, the $M = 8$ signal appears profitable for "intermediate" E_b/N_o (20 to 50).

If $E_b/N_o > 50$, one should consider, in addition to increasing M of M-phase, the signals which encode both amplitude and phase. Also, other close-packed signals may be considered.

These are the conclusions, then, for a constant data rate comparison of some common (and feasible) signalling methods. Many design situations will not be clearly separable into the regions treated above; for all such cases one must balance the advantages and disadvantages.

4. CHANNEL CAPACITIES AND MAXIMUM RATES

Although the P_e calculations treated before are probably the most basic evaluator for a digital communications system, there may be cases where bandwidth conservation is more essential; for such cases the curves of Figs. 20 and 21, showing "information-rate per W" are of interest. Any consideration of information-rate per W is related to "channel capacity," which is now familiar to all communications engineers. This section might well then be considered an extension of the ideas in Figs. 20 and 21.

Channel capacity (C) is the maximum rate at which one can send information over a channel, and still (theoretically) reduce the P_e to zero (by external coding means). Although there is general agreement on this concept, there is much disagreement on the way in which C is calculated, as we shall note below.

The practical significance of C is that it should provide a sensible upper bound for comparing various practical signalling alternatives, especially with respect to utilization of W. In a sense, P_e considerations are outside the sphere of C, since the theory holds that "it is possible" to reduce the error towards zero at the rate C by external means.

In Section 4.1 a channel capacity for digital systems is portrayed. In Section 4.2 we suggest a "practical" method of viewing "maximum rates" (related to channel capacity) in digital systems.

4.1 Capacity Relations for Digital Systems

In general, when beginning consideration of channel capacity, one considers various types of possible channels (Ref. 6). Thus there may be discrete channels or continuous channels. In a discrete channel it is assumed that the input and the output can be represented by some discrete number of events. For a continuous channel it is assumed that the output and input can take on a continuum of values. There also may be mixed systems such as discrete-to-continuous or continuous-to-discrete. There is also the element of discrete time or continuous time. Although most physical channels may be time-continuous, they can often be satisfactorily represented by means of discrete time models (which is our case).

The basic calculation used to find the channel capacity for a given assumed model consists of evaluating the "average mutual information" between input and output sequences of events. This average mutual information depends on the probability distribution of the input sequence of events as well as the conditional probability distribution which characterizes the noise properties of the channel. For any given channel, then, the channel capacity (given the channel probability characteristics) is defined as the "maximum average mutual information which would result if one maximized over the possible probability distributions of the input" (Ref. 6).

4.1.1 Continuous Channels. One of the most familiar capacity formulas deals with a bandlimited continuous channel in which there is additive white Gaussian noise. The restrictions are that the input be a bandlimited time function and that its average power not exceed a given value S . For this situation, the channel capacity is given by the familiar relation:

$$C = W \log\left(1 + \frac{S}{N_0 W}\right) \text{ bits/sec} \quad (57)$$

where: C = channel capacity, when input is a band-limited time function with average power S
 W = signal (and hence channel) bandwidth.

This equation means that if one has a given average signal power S , if the channel noise is flat and white Gaussian of power N_0 per unit bandwidth, if a bandwidth W is available, and if one agrees to use as input a time function limited to bandwidth W , then the maximum rate at which one can possibly send information with practically zero error is given by the C of Eq. 57.

This equation can be applied to a given channel if the channel gain is reasonably controllable or constant (or slowly varying), and if one is going to use continuous input waveforms. One of the bounding conclusions concerned with the proof of this equation is that only with the use of signals having the statistics of white noise could one possibly hope to achieve this channel rate; this is how the probability distribution of the input enters the capacity relation.

Even if the actual channel being dealt with is discrete, Eq. 57 serves as the ultimate capacity, since the capacity for a given continuous channel will always be equal to or greater than the capacity for a corresponding discrete channel (where bandwidth, etc., are

considered equal). Consequently it is sensible to use Eq. 57 as the ultimate comparison even though discrete signals are used.

4. 1. 2 Discrete Channels. For the discrete case, the channel capacity is determined by the particular situation. A common case is "uniform" channels. A channel is uniform if:

$$\begin{aligned} P_{ij} \text{ all equal} & \quad (\text{for } i \neq j) \\ P_{ii} \text{ all equal} & \quad (58) \end{aligned}$$

where: P_{ij} = transition probabilities of signals between input and output of channel, due to noise.

It may be noted that orthogonal signals and simplex signals in a discrete channel result in a uniform channel, but bi-orthogonal signals do not strictly do so.

The channel capacity, in bits/symbol, for such uniform channels (Ref. 6) is given by:

$$I' = \log_2 M + P_{ch} \log_2 \frac{P_{ch}}{M-1} + (1-P_{ch}) \log_2 (1-P_{ch}) \frac{\text{bits}}{\text{symbol}} \quad (59)$$

where: P_{ch} = uniform probability of error for each of the M-ary decisions.

M = alphabet size.

It is very important to note that Eq. 59 lists the capacity in bits/symbol, whereas Eq. 57, for the continuous channel, lists the capacity in bits/second.

To convert the "bits/symbol" capacity of Eq. 59 to a "bits/second" capacity comparable to Eq. 57, (and this appears desirable) it is necessary to incorporate W and symbol time in a sensible fashion. Since we are interested in a capacity relation, we desire to use a minimum TW product in any calculation.

Immediately below, a minimum TW product for various signals is used which is related to the "dimensions" of the signals (see Table I). The capacity of discrete (digital) systems in bits/second is then found, using the minimum TW product (in terms of dimensions).

The final objective will be to compare various signals in terms of their capacity/bandwidth. Thus, we will plot C/W vs. E_b/N_o for the various signals. This is the practical interpretation of the theoretical statement that "an infinite number of dimensions is necessary

to specify a signal at a 'time' density of $2W$ per second." This gives a measure of the "best one can do" in terms of utilizing bandwidth, given a type of signalling.

Although we feel this is a sensible way to use a capacity relation, there are certainly other ways, and the general disagreement has led to some confusion about "capacity." One common capacity comparison treats the capacity as a single number consisting of the value of C when W approaches ∞ (C_∞). This usage has value for ultimate comparisons and concepts, but it is difficult to derive conclusions about any "real world" choices. Another capacity comparison (for discrete systems) uses theoretical W 's (as in Table I) of signals in the continuous capacity relation of Eq. 57. We feel somewhat dubious about this, since the theoretical BW 's specify "how close one can put controlled signals," and not "how much W is used by each signal." However, this usage does provide comparisons for controlled environments (each communicator using same type of signal, and coherent with respect to each other), even though this is somewhat unrealistic.

As mentioned, we propose here a digital capacity (C) in bits per second which is arrived at by using minimum TW products. Although time-bandwidth products of general signals are somewhat arbitrary (see Ref. 23), there is good reason for relating minimum TW products to the dimensions of any given signalling method. By "dimensions," of course, we refer to the signal space representation of the signals.

The signal space representation of signals is based on the Shannon Sampling Theorem, which says that, in an interval T , there are $2TW$ numbers required to specify the signal (if the signal is limited to W). These numbers can be considered as coordinates in a D -dimensional space (where $D = 2TW$).

Consider now the vector space for a large ensemble of successive symbols; this space will have dimensions D' , and D' is given by:

$$D' = 2WT' = 2WT \times N \quad (60)$$

where: T' = NT = time length of ensemble of successive (possibly overlapping) symbols
 T = time length of one symbol
 N = number of symbols in T'
 D' = dimensions of total space corresponding to T' .

If one uses the criterion that succeeding symbols have no intersymbol interference, the vec-

tor space representation for a given symbol time is an orthogonal subspace within the total space of D' dimensions. The number of signals which can possibly be sent in time T' is then given by:

$$N = \frac{D'}{D} = \frac{2WT'}{D} = \frac{2WNT}{D} \quad (61)$$

where: D = dimensions per symbol time T , called "signal dimensions."

The minimum TW for a given signalling method is then given by:

$$\min \{TW\} = \frac{D}{2} \quad (62)$$

It now remains to find the signal dimensions D for various signalling methods.

An M -ary coherent-orthogonal signal requires M dimensions per subspace, since this operation can be viewed as orthogonal subspaces within the symbol subspaces of dimension D .

Therefore, $D=M$ for coherent-orthogonal.

For incoherent-orthogonal operation, each symbol subspace requires $2M$ dimensions since the fact that phase information is not useful means that each "element" requires 2 orthogonal dimensions. Since there are 2 dimensions per symbol, there are $2M$ dimensions per subspace. Hence $D = 2M$.

Simplex operation has a signal dimension of $D = M-1$. For M -phase signalling ($M > 2$) the dimensions D remain 2, no matter what M is. For convenience, Table II shows the signal dimensions for the various methods (also listed in Table I).

M-ary Signal Type	Dimensions D
M-phase ($M > 2$)	2
Bi-orthogonal (Coh.)	$M/2$
Simplex (Coh.)	$M-1$
Coherent-Orthogonal	M
Incoherent-Orthogonal	$2M$

Table II. Minimum Signal Dimensions.

It should be noted that using minimum TW products based on signal dimensions as in Eq. 62 is a theoretical "capacity" type of calculation. If one is precise, one can never put absolute bounds on W and also on T. However, if one puts "practical" bounds on W (say -50 db on a filter response) one can in principle approach Eq. 62 if one allows overlapping signals (but nevertheless practically non-interfering). The above material is more fully explained in Appendix A, and practical realization problems are discussed.

There are the following three implications as one tries to achieve the TW products of Eq. 62 and Table II (see Appendix A):

- (1) The signals must overlap.
- (2) The signals have an envelope which approaches a $\frac{\sin x}{x}$ curve.
- (3) If $\min \{ TW \}$ were to be realized, a correlating receiver is no longer necessary; if $M = 2$, a precise sampler is required, and if $M > 2$ a combination (or block-coded) sampler is sufficient.

It is of interest to compare the minimum TW products here with the theoretical TW products of existing signals as given in Table I. For this purpose the uncontrolled environment bandwidths of Table I should be used, since these are the "limited" bandwidths. For coherent bi-orthogonal and orthogonal, the signal TW's are about twice the capacity TW's. For incoherent orthogonal the ratio of signal TW to capacity TW is $M+1/M$. Consequently, the "capacity" proposed here is a relatively realistic bound for a given signalling method.

We now use the minimum TW product of Eq. 62, along with the signal dimensions of Table I or II, to obtain the capacity of discrete systems in bits/second, given the signalling method. Using Eq. 59, and the minimum T for a given W, the result for coherent simplex is:

$$\text{Coherent Simplex: } C = \frac{I'}{\min T} = \frac{2W}{M-1} \left\{ \log_2 M + P_{ch} \log_2 \frac{P_{ch}}{M-1} + (1-P_{ch}) \log_2 (1-P_{ch}) \right\} \text{ bits/sec} \quad (63)$$

where: W = bandwidth as determined by response being about (-50 db)

M = alphabet level or number of symbols

C = capacity in bits/second.

For general M-ary coherent orthogonal:

$$C = \frac{2W}{M} \left\{ \log_2 M + P_{ch} \log_2 \frac{P_{ch}}{M-1} + (1-P_{ch}) \log_2 (1-P_{ch}) \right\} \text{ bits/sec} \quad (64)$$

For M-ary incoherent orthogonal:

$$C = \frac{I'}{T} = \frac{W}{M} \left\{ \log_2 M + P_{ch} \log_2 \frac{P_{ch}}{M-1} + (1-P_{ch}) \log_2 (1-P_{ch}) \right\} \quad (65)$$

It should be emphasized that, since we have used the limiting TW product, the equations above should truly represent capacities for the discrete cases.

Using Eqs. 63, 64, and 65, the capacities of the various alternative signals, starting with binary simplex and going up to the M-ary signals, have been plotted in Fig. 23. Since the ordinate is chosen to be C/W, the plot actually shows capacity per bandwidth for each type of signalling. An important difference between the C/W curves here and the I/W curves of Figs. 20 and 21 should be noted: in Figs. 20 and 21 the W was based on occupancy ---that is, on how closely similar signals can be placed. Here the W is strictly limited (-50 db). Hence the W here would be comparable to the uncontrolled environment W's of Table I. Also, Fig. 20 and Fig. 21 refer to actual signals, while here a bound is being used with no particular signal being specified.

To emphasize, the curves here deal with the capacity per limited bandwidth if one is operating a discrete system with the given signal type and E_b/N_o ratio. The cases of Figs. 20 and 21 essentially show "what the information rate is if one agrees to operate discretely and has a bandwidth occupancy and an E_b/N_o ratio."

Although there is about a factor of two between information rate/W and capacity/W for coherent signalling methods, we feel the term "capacity" is justified because of the minimum $\{TW\}$ concept. Thus this capacity may be regarded as a relatively practical bound. There are two important features to C as used here: (1) it is a bound based on a minimum TW product; (2) it is a single channel concept, so does not depend on the environment of other channels (as do the controlled W values in Table I).

From the figures, it is seen that binary simplex affords the largest C/W---the limiting value is 2. For coherent orthogonal the limiting value is 1.0, and the curve for

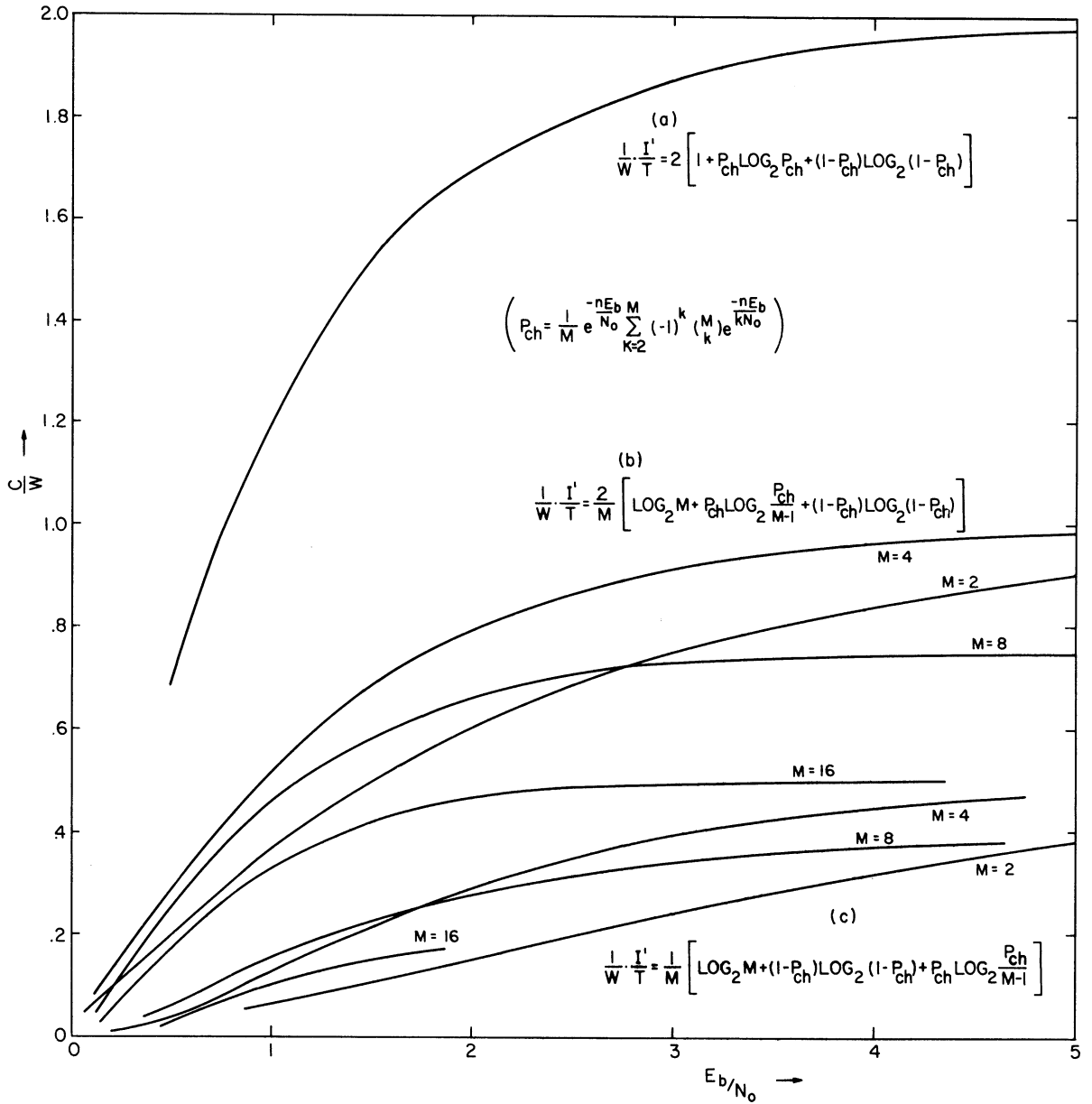


Fig. 23. C/W versus E_b/N_0 for binary simplex, coherent orthogonal, and incoherent orthogonal signals.

$M=4$ stays above all others for the entire range of E_b/N_0 . It is seen that, as M increases, C/W becomes very restricted.

For the incoherent case, $M=8$ is as good as $M=4$, in the low range. Again, as M continues increasing, the C/W becomes restricted.

4.2 Maximum Information Rates if Data Rate is Variable

All the preceding material plotted either P_e or I/W vs. E_b/N_0 . The various curves can be used to find either the P_e or the information rate per W for the various condi-

tions. Under each signal type, one specific objective was to compare binary versus M-ary operation.

Constant rate comparisons can be made by reading vertically at any E_b/N_o point. For constant E_b/N_o , the data rate R is constant if the power S is constant, since $R = 1/T_b$. Consequently, assuming that S is always fixed, all the preceding curves make "comparison sense" for constant data rates.

We now consider the case where S is still fixed, but R is variable. The objective is to sensibly compare the various signal types for R variable. The first significant change is that we now plot curves versus S/N , rather than E_b/N_o , since S is fixed, but E_b is not.

For the variable R case, the general aspects are as follows. Improved P_e 's can be obtained by slowing the data rate (longer T_b), and consequently reducing bandwidth. The cost, of course, is the slower data rate. The alternative is to use M-ary signals. The choice between these two alternatives depends on the "economics" involved.

Here we will adopt the criterion that "one should use as fast a data rate (low T_b) as possible, commensurate with P_e and W ." Thus we assume that one is given P_e and W , and adjusts the R to as high a value as possible. We would then plot maximum information rate per W versus S/N .

In order to do this, it is necessary to break the S/N axis, in any plot, into two regions: (1) the low signal-to-noise ratio case, and (2) the high signal-to-noise ratio case. In the low signal-to-noise ratio case the minimum T_b chosen, under a given S limitation, is determined by the probability of error allowable. For the high signal-to-noise ratio case, the T_b is limited and determined by the minimum TW product (Section 4.1.2). In other words the P_e is constricting for low S/N , and TW product is constricting for large S/N . For low signal-to-noise ratio cases, the probability of error as depicted by the capacities in Fig. 23 above become so high as to be of little practical worth; hence it is sensible to put a probability of error limitation on the low signal-to-noise ratio cases. For the high signal-to-noise ratio cases the probability of error becomes increasingly better and hence minimum T_b (for maximum rate) will be restricted by the TW product (discussed in Section 4.1.2).

The objective here will be first to present maximum rates for binary simplex and incoherent orthogonal under average power restriction. M-ary coherent and incoherent orthogonal will then be treated. In the curves of this section we plot "maximum rate" per bandwidth versus signal-to-noise ratio: the maximum rate is given in bits/second. The direct

import of this is that one can evaluate how efficiently one utilizes the bandwidth under the given P_e and W restrictions.

4.2.1 The Small S/N Region. As mentioned above, when the signal-to-noise ratio is small, the duration T_b must be made long enough to meet the acceptable error criterion or probability of error. Thus in this region the error rate will be constant with energy-to-noise ratio and hence the energy, the N_o and the I' of Eq. 59, is a constant. This can be written:

$$I_M = \text{Max Rate} = \frac{I'}{\text{Min } T} = \frac{I'N_o}{E} W(S/N) \quad (66)$$

where: I_M = information rate in bits/second when taking both P_e and W into account.

Curves of I_M/W can be found as follows: using previous curves, choose a P_{ch} ; find I' corresponding to that P_{ch} ; then evaluate $E_b/N_o = \frac{E}{\log_2 MN_o}$ from previous graphs. With this equation, and plotting I_M/W vs. S/N , one obtains the constant- P_e portions of the curves shown in Figs. 24, 25, and 26.

4.2.2 Large Signal-to-Noise Ratio. As the signal-to-noise ratio is increased, one arrives at the place where the acceptable probability of error is met, and now the maximum rate is determined by Eqs. 63, 64, and 65 given before. Hence the right sides of the curves in Figs. 24, 25, and 26 correspond to the same information plotted in Fig. 23, except that now we plot I_M/W vs. S/N rather than C/W vs. E_b/N_o . Note that C and I_M differ only in that I_M is constrained by P_e until $I_M = C$. From there on, $I_M = C$.

To plot I_M/W , one first assumes a given value of S/N . Convert this to E/N_o by multiplying by $\min \{TW\} = D/2$. With the resulting E/N_o , find the P_{ch} from previous graphs, using $E_b/N_o = \frac{E}{\log_2 MN_o}$. Then one can use Eqs. 63, 64, and 65 to find I'/T .

It should be noted that a low signal-to-noise region for one P_e criterion is different from that for another P_e criterion. Also, for M -ary, one sees that a low signal-to-noise ratio for binary may be a high signal-to-noise ratio for M -ary.

To emphasize, we have divided the maximum-rate situation into two areas: small signal-to-noise region and large signal-to-noise region. These regions are dependent upon the probability of error which must be met. The status or comparison between the small and large signal-to-noise regions is depicted in Table III. We contend that this depiction of maximum rates is sensible in terms of practical meaning for "maximum rate."

S/N increasing →

Small S/N	Large S/N
Error rate constant	Error rate begins to drop
T just long enough to give error rate	I' (no. of correct) increases as S/N increases
Little penalty for restricting to binary	Maximum rate limited by TW product of signal. Can do much better by going to M-ary (especially close-packed signals)

Table III. Conclusions about small and large S/N regions.

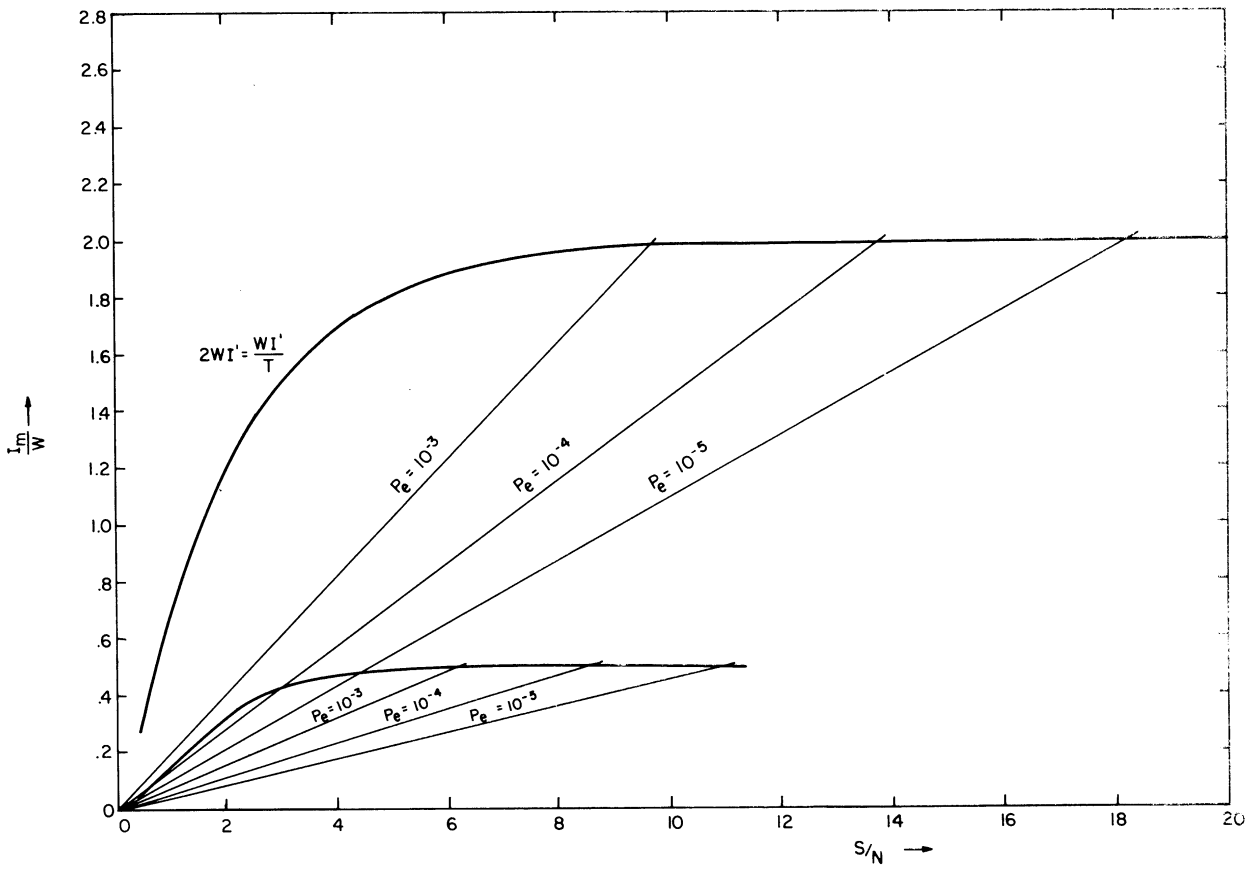


Fig. 24. Maximum information rate per bandwidth for binary signals.

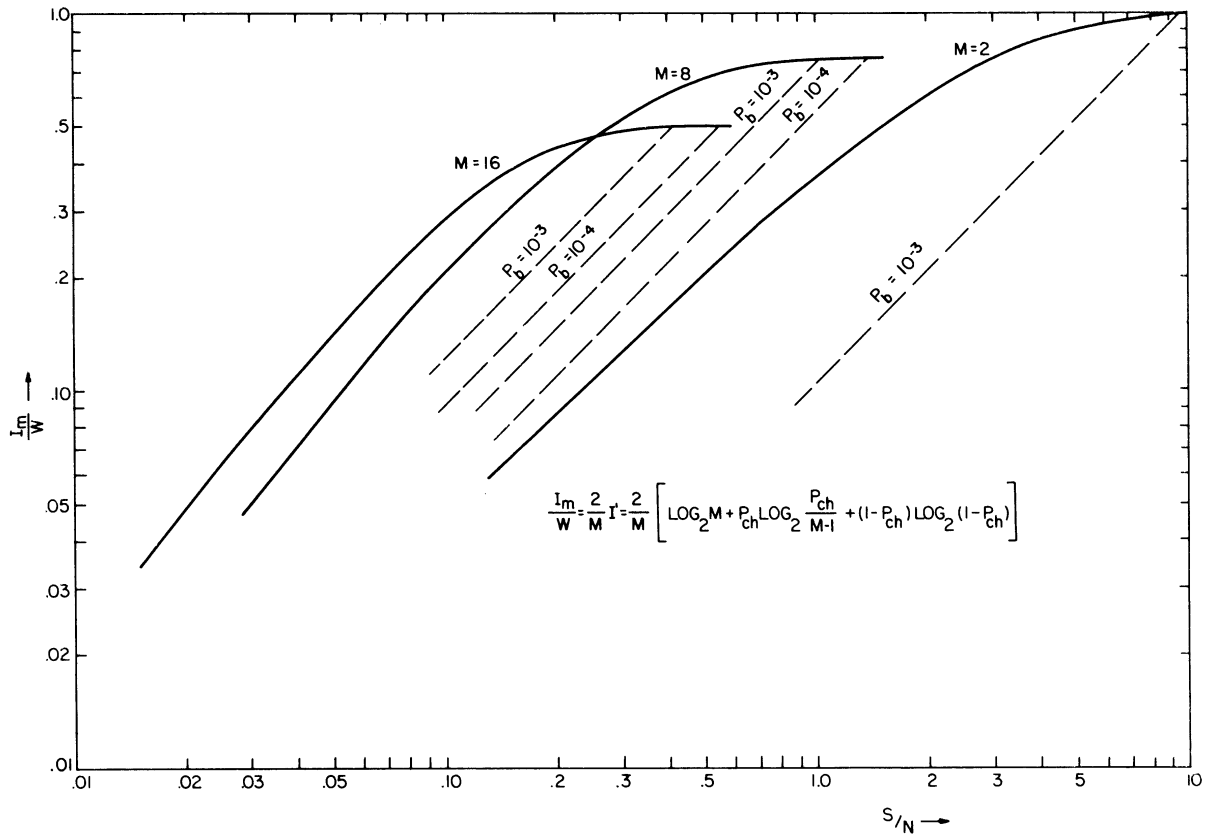


Fig. 25. Maximum information rate per bandwidth for M-ary coherent orthogonal signals.

In comparing these results to the P_e calculations in the previous section (Section 3) one should remember that there P_e was plotted vs. E_b/N_0 . Here we are interested in maximum rates, and we ask the question, "How well are we doing in terms of what one could do?" Also, in this comparison we are assuming a power limitation.

The variable rate conclusions from Figs. 25 and 26 are similar to the constant rate conclusions. If S/N for a given P_e is low ($S/N < 1$), M-ary operation becomes desirable (practically necessary). However, if $S/N > 10$, either $M = 2$ or a more closely-packed signal becomes desirable.

These figures should serve the purpose of depicting "how well one is doing" with respect to "how well one could do," taking both P_e and W restrictions into account.

4.3 Conclusions

In Section 4.1, the capacity relations for digital systems were considered, where a capacity for a given signalling method was studied. The viewpoint taken was that a certain

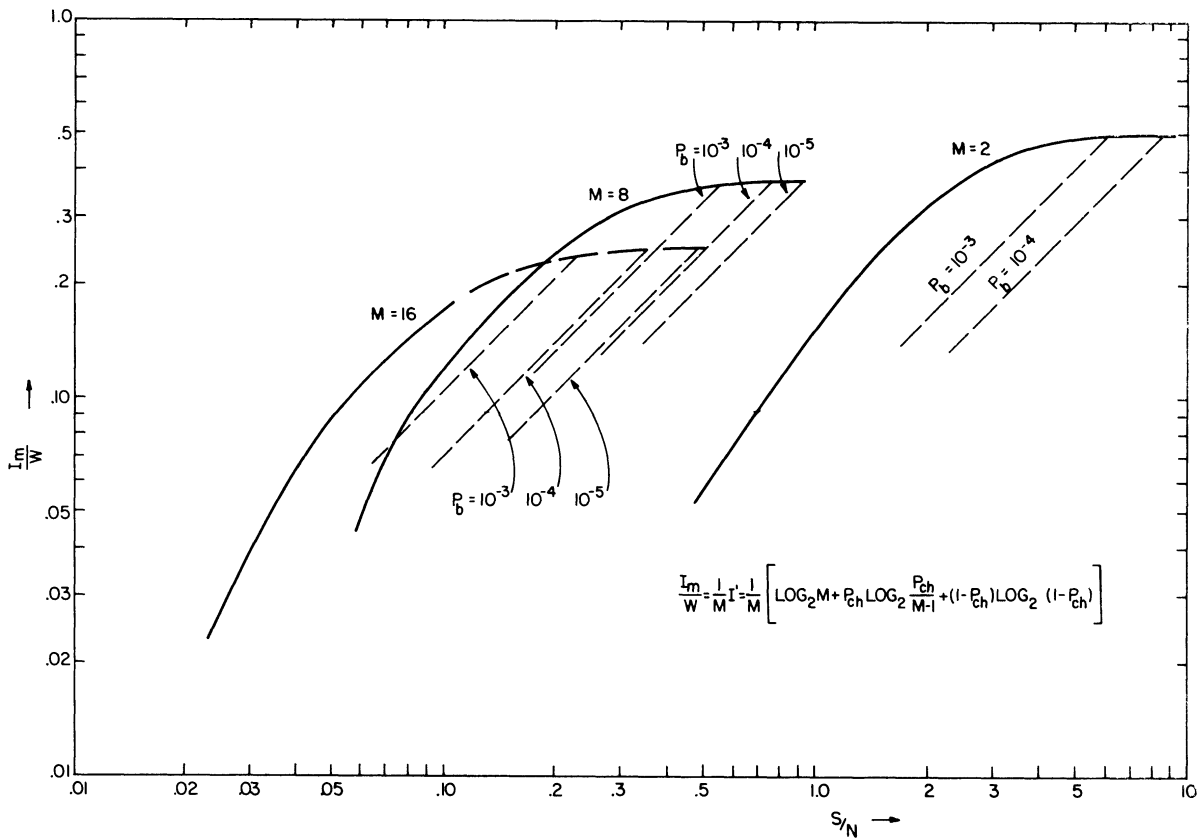


Fig. 26. Maximum information rate per bandwidth for M-ary incoherent orthogonal signals.

effective W was available, and the capacity was determined by using the minimum T for that W . The minimum $\{TW\}$ was related to the signal dimensions. The resulting capacity relations are a measure of the bandwidth utilization, irrespective of the channel P_e . However, the capacity concept implies that the P_e can be reduced towards zero (by external means), and the rate still maintained.

It was seen, in Fig. 23, that binary simplex offers the highest capacity (for far-apart signals). For coherent orthogonal signals, $M=4$ is slightly better than $M=2$ for the low E_b/N_0 ; but both approach the same value at high E_b/N_0 . For incoherent operation, $M=4, 8$, or 16 is preferable to $M=2$ for low E_b/N_0 , but $M=2$ or 4 is best for high E_b/N_0 .

The above ordering is under the criterion of bandwidth utilization, given a constant rate comparison.

For a "variable data rate comparison," the curves of Section 4.2 are applicable. Since a communicator is always faced with both a P_e and a W restriction, it is proposed that

the maximum rate plot be divided into two regions: (1) low S/N region, where P_e is constraining; and (2) high S/N where minimum $\{TW\}$ is constraining. Such curves appear in Fig. 24 for binary simplex, and in Figs. 25 and 26 for M-ary coherent and incoherent orthogonal. Given a W , a P_e , and S/N, and the signalling method, one can use these curves to establish the maximum rate (I_M) for that set of conditions. We feel this approach is a sensible way to view maximum rates; furthermore, the concept of maximum rate implies a variable data rate. In other words, if one chooses the highest I_M/W for a given P_e , W , and S/N constraint, one also chooses the "best" data rate.

To summarize, then, one can make two comparisons, depending on the design situation: (1) a constant rate comparison, and (2) a variable rate comparison.

A constant rate comparison is implied if both S and data rate (R) are specified. If so, the next step is to decide whether P_e or W is most important. The conclusions for these conditions were given in Section 3.3.4. In general, "far-apart" M-ary signals are required if E_b/N_0 is low for a given necessary P_e . Since increasing M increases the W , it appears desirable (considering both P_e and W) to use an M of 2 or 4 for coherent operation, and an M of 4 or 8 for incoherent operation. Also, the benefits of any increase in S should be emphasized, due to the steepness of all the P_e curves.

If E_b/N_0 is high for a given P_e requirement, one should use M-phase ($M > 8$) or some other close-packed signal.

A variable data rate comparison is implied if one is free to choose the data rate. In this case, one can stay within both P_e and W requirement. Here it is sensible to choose that data rate which gives the highest "information-rate per bandwidth" for the given P_e . The proper M is determined by the S/N, P_e and W conditions, as depicted in Figs. 25 and 26. P_e 's of about 10^{-5} and $S/N < 1$ specify an $M > 2$. Of $S/N > 10$, either $M = 2$ or a more closely-packed signal is desirable.

PART II: ASYNCHRONOUS TIME-MULTIPLEXING OF
MULTICHANNEL SPEECH SOURCES

1. INTRODUCTION

The material of Part I essentially concerned the evaluation and comparison of different signalling methods in digital channel transmissions. In Section 1.2 of Part I, it was noted that source coding is a potential area for improving the transmission of speech signals.

The problem of interest here is how can one improve the transmission of multi-channel speech signals, assuming that the information is to be transmitted digitally. Of course all the previous material still applies to the various possible methods of sending the information across the channel, so that this material is complementary to the previous material, and not an alternative to it.

If the information to be communicated is speech, then an area of immediate interest is the "source coding" of such speech to improve the information efficiency. For example, speech compression methods can be interpreted as a source coding method. (The relation of this to our topic will be discussed below.) Whenever the speech information comes from telephone communicators, source coding can effect improvements in two respects: (1) by reducing the redundancy in the speech signal itself, and (2) by altering the fact that, since each talker is part of a two-way conversation, each communicator is using this allotted part of the total facility only one half the time (for radio and cable links).

Since various speech codings for reducing redundancy have been proposed and investigated by others, our main interest here will be in the second aspect mentioned---improving the "channel utilization" when speech sources are part of a two-way conversation. This aspect can realistically be improved only when the facility is handling a number of speech sources so that one can take advantage of a "statistical effect" among them. In the material to follow we will assume use of an existing proposed speech encoding method, and then analytically evaluate the buffer situation and the use of this system in a multichannel

speech situation. The pertinent aspect is that each speech source will seek access to the channel only when it is "on" and when it has information (a sample) to transmit. Such operation will be referred to as asynchronous time-multiplexing.²

The objective of this material will be to provide analytic evidence of the fact that it is both profitable and (at least analytically) feasible to perform the coding and multiplexing function in an asynchronous manner---that is, by adopting the principle of allowing each source to seek access to the channel only when it has information to transmit. Although this idea is not new, it is only relatively recently that codings have been tested which make it possible to run a system asynchronously.

1.1 Background and Previous Work

For cases where multichannel speech is to be sent digitally, a familiar method is to use time-multiplexed PCM. Considering only the multiplexing function, we regard this as synchronous multiplexing since a given "slot" in the channel-time assignment is always given to a particular speech source (once it is connected by the "central"). In asynchronous time multiplexing the "source-channel time" relation is not fixed, but depends on the instantaneous conditions.

A good example of an operational, asynchronously multiplexed system is the "Time Assignment Speech Interpolation" (TASI) system, which is used by the Bell Telephone System on the Atlantic Oceanic cable (Refs. 26 and 27). In this system the asynchronous multiplexing is effected by a "fast-switching" of the source-to-line connection; during a given connection, conventional, regularly-spaced samples are sent. Consider that a talker's time axis is divided into "talk bursts." Whenever a talker is silent in excess of a stated time, an end of talk burst is assumed and the talker's line is switched to another (momentarily) waiting party. The original talker will get another line when he begins talking again. Thus the asynchronous multiplexing is effected by this fast-switching means, and the effective capacity of the oceanic cable has been about doubled. Although some "freezeout" inevitably occurs, this system is working successfully and the tolerable freezeout fraction designed

²The reader is cautioned that the term "asynchronous multiplexing" is used by some authors to denote "co-channel net operation." There the objective is to allow a "net" of communicators to contact one another at will. The "asynchronous" aspect, then, lies in enabling any communicator to contact any other over a common channel. In the problem here, a number of sources (after going through a "central") share a common (trunk) facility. The asynchronous aspect here lies in the assignment of channel time slots to any given source; hence it is called "asynchronous time-multiplexing."

into the system is about 0.5 percent. It is found that the speakers are not aware of the freeze out when it is kept to this fraction.

The speech coding method to be used in our analysis of the feasibility of asynchronous multiplexing is the "extremal coding" method (Ref. 28), also called "selective amplitude sampling" (SAS, Ref. 29). Both extremal coding and selective amplitude sampling refer to a sampling procedure wherein one samples irregularly, at the "extremum" of the speech signal. Thus one sends only the values of the relative maxima and minima of the waveform. It is necessary then both to send the value of this extreme and in some way note the time of occurrence of the extreme. In the receiver either an interpolation function, or a "boxcar" detector followed by a filter, may be used to reconstruct a good approximation of the actual speech signal. Before evaluating this coding procedure in the multichannel situation, it will be necessary to analyze the action of the buffer which is a required part of any practical asynchronous system.

It is also of interest to note that the ideas behind asynchronous multiplexing treated here are very much related to the ideas of "demand matching" which have been espoused in concept a number of times (Refs. 30, 31, and 32). The general idea in demand matching is that a source have access to the channel only when it has information to transmit. This is meant to incorporate removal of both redundancy and the nonuniform flow of information from the sources. In an ideal multichannel case, demand matching would result in all the information from all of the sources being handled by the channel; further, all the information from the multiple sources would be statistically stable enough that the channel could operate at a uniform rate (without a buffer) even though the individual sources were creating information at a sporadic random rate. In practice one cannot realize this ideal, and hence two difficulties arise: first, not all of the information can be transmitted (a situation usually called "freezeout" for each individual source), and second, the information rate of the total ensemble of sources will not be uniform, so that some buffering will be required. Nevertheless, the advantages to be gained with a tolerable freezeout fraction and a somewhat varying channel rate are sufficient to allow serious exploration of various implementations of demand matching.

Finally, it is of interest to note the comparison between speech compression methods and the extremal coding or selective amplitude sampling method to be used here. Speech compression techniques are a very profitable way to encode speech signals, but do

not derive any additional profit by being multichannel. In speech compression techniques one studies seriously the mechanism which creates speech and then attempts to encode efficiently. In our case, we consider a multichannel case of some time waveform with a given frequency spectrum. One resulting difference is that the more general coding techniques offer at least a possibility of being applicable to analog waveforms in general (which have the same frequency spectrum, etc.), whereas the speech compression method is rather strictly limited to speech itself. This is not to say that the coding methods studied here will necessarily work for signals other than speech, but they nevertheless have the possibility of doing so, and this should be kept in mind. Another aspect is that speech compression methods may be considered as an ideal realization which may still require a great deal of equipment before being used on a multichannel system. On the other hand, the techniques of interest here may be considered as intermediate coding where the encoding is not as efficient but has more hope of being realized practically.

In the material that follows we will first describe qualitatively the asynchronous multiplexing scheme with the use of a block diagram. Then the performance of the buffer will be mathematically analyzed, since this is the most crucial aspect for the multichannel case. It should be noted that the performance of the buffer is relatively independent of the actual coding scheme, except that the coding scheme must sensibly fulfill the "model" used in the analyzing of the buffer action. We conclude with comments on reconstruction of the waveforms and the performance of system comparisons.

2. DESCRIPTION OF ASYNCHRONOUS TIME-MULTIPLEXING SCHEME

Although one can describe asynchronous time-multiplexing schemes in general terms of demand matching, here we will (for concreteness) portray a system which involves extremal coding (also called SAS). It should be remembered, however, that the buffer analysis of Section 3 applies to any encoding which reasonably fits the model used in that analysis. Extremal coding is a suitable choice since it reasonably fits that model. Figure 27 depicts extremal coding, where a time waveform is sampled at the occurrence of relative maxima or minima. It may be noted that these maxima and minima occur quite irregularly. For

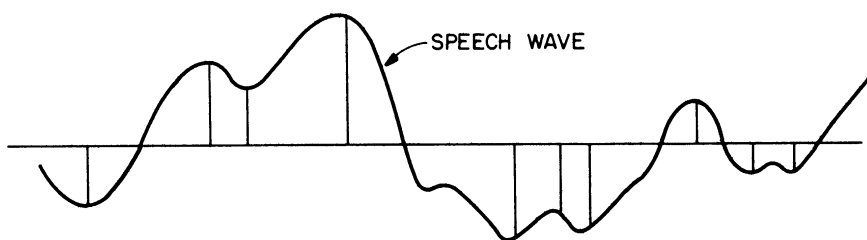


Fig. 27. Depiction of sampling for extremal coding.

a multichannel system, it will be seen that the following information would have to be transmitted: (1) amplitude of the extremum, (2) time of occurrence (in some fashion), and (3) source identification number.

If one wishes to encode a number of multichannel speech sources with extremal coding, and then send them asynchronously---that is, let the sources seek access to the channel only when they have an extremum to send---it is seen that a buffer will be required to smooth out the inevitable variations in arrival of information. If a buffer were not used, a great many extrema would be rejected or "frozen out." A block diagram of a system which can be allowed to run asynchronously is shown in Fig. 28.

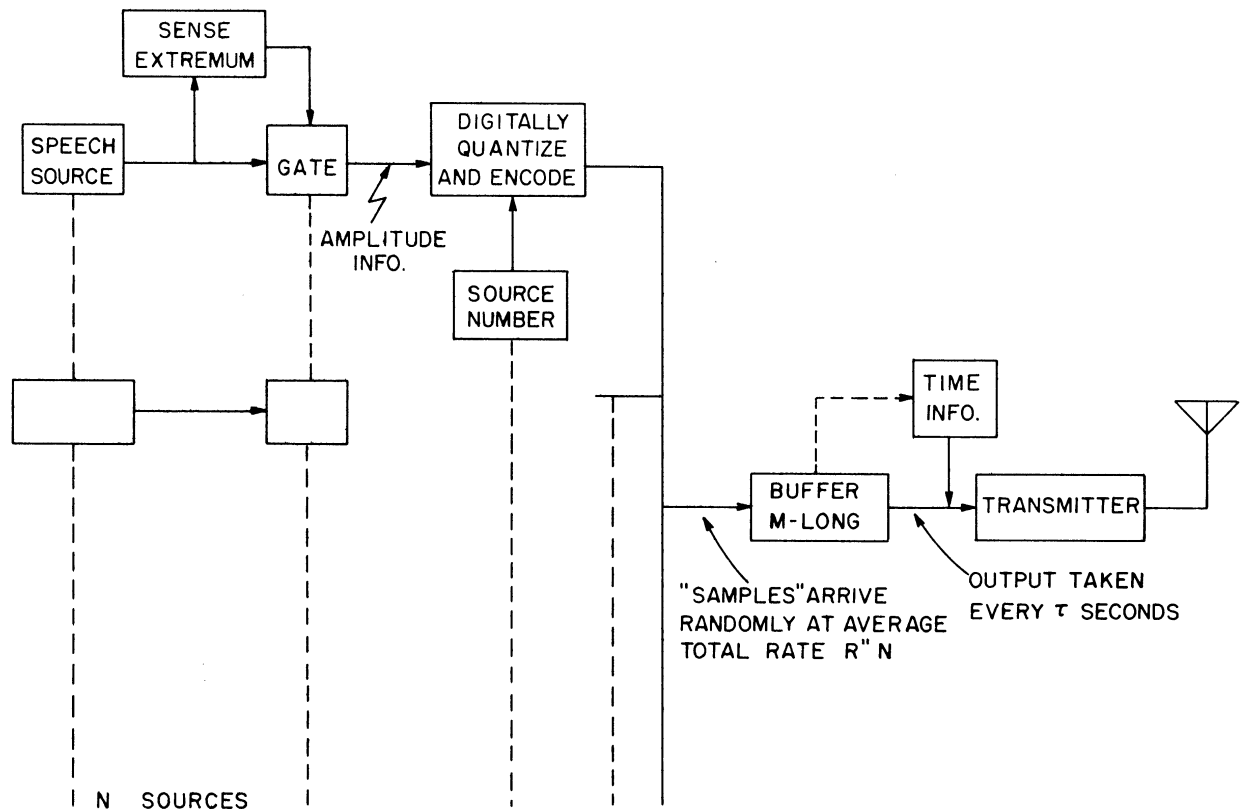


Fig. 28. Block diagram of transmitter for asynchronous multiplexing of multichannel speech using extremal coding.

In Fig. 28 it is seen that only the amplitude at an extremum is allowed to pass through the gate. This amplitude is then quantized and digitally encoded. Source number should be added at this point. Since the information (extremum) occurs at random, the essential idea is to allow the buffer to accept extrema from the individual sources whenever they occur with no prior ordering. Outputs will be taken from the buffer (if any are present) at a regular clocked interval---every τ seconds. (Hence, the multiplexing is asynchronous but the channel transmission is synchronous.) It is seen that, for some fraction of time, the channel will have nothing to transmit (buffer empty). The percent of time in which this happens is called "channel idle time."

On the other hand, the M-long buffer may be full when "a sample" arrives from a source. In this case the sample is rejected, and the total percent of such samples (from all sources) is called the "freeze out fraction" (FOF). Clearly the FOF will depend upon both the buffer length (M) and the operating parameter u (the ratio of average number of buffer input

samples to average number of output samples per unit time). The evaluation of FOF as a function of these two parameters will be the most important objective of Section 3. For acceptable operation, the FOF must be kept to a very low level.

Since each "sink" at the receiver end must know the time of occurrence of the sample, it is necessary to insert the "buffer delay time" onto the other information. This can be done at the buffer, as shown in Fig. 28. The accuracy of the "time" information must be commensurate with the encoding method used (see Section 5). For extremal coding it should be sufficient to send the buffer delay time in integers of τ (for a sufficiently large number of channels).

As mentioned, the essential problem in dealing with the feasibility of any such system lies in analyzing the action of the buffer. It is quite necessary to assure that only an acceptably small fraction of the information samples be frozen out. In addition, a number of other properties of such asynchronous operation can be analyzed by considering the performance of the buffer.

3. PERFORMANCE OF THE BUFFER

As noted above, the buffer will accept samples from any of the sources, so long as it is not "full." Samples are removed from the buffer at regular intervals---every τ seconds.

In order to do any analysis of the buffer operation it is necessary to have available the probability distribution for the number of samples going into the buffer. A sample is a quantized (and digitized) amplitude of the extremum with source number added (time information must be added later; see Fig. 28).

Since the output time axis of the buffer is divided into intervals of τ seconds, it is useful to consider the input time axis as also divided into τ intervals. Then, the question of interest is, "What is the probability of k units occurring (from all of N sources) within an interval τ ?"

The analysis of the buffering situation below will assume that the probability distribution for k events of the equivalent source (of all the N sources) occurring in a time interval τ is given by the Poisson relation:

$$p(k, u) = e^{-u} \frac{u^k}{k!} \quad (67)$$

where: $p(k, u)$ = prob. of k extrema among the N sources
in a time interval τ

u = average rate (taken across sources) of
occurrence of extrema in time τ .

It may be noted that u can be interpreted as the ratio of the average buffer input rate to the average output rate (since the output clock rate is one, in terms of τ).

This probability distribution results if one assumes a Bernoulli model for the group of N sources divided into τ intervals. That is, assume that (1) the occurrence of a sample in one source is independent of all other sources, and (2) the probability of occurrence of a sample in a time interval τ (for any given source) is constant; under these conditions a Poisson distribution, as in Eq. 67, will result for a number of sources taken to-

gether. Certainly the first assumption is justified, as is the second one if τ is sufficiently small. (τ is inversely proportional to N , the number of sources.) Nevertheless, it should be remembered that all results of the following analysis apply so long as the probability distribution of Eq. 67 holds.

A study of the range of applicability of Eq. 67 for cases relative to this problem was made, and reported in Appendix B. The study was made in terms of the "distribution of intervals between zero-crossings" of the individual sources. It was found that, if the interval distribution is exponential [$f(u) = e^{-u}$], the Poisson relation of Eq. 67 holds for N multiple is assumed to be Erlang type-1 [$f(u) = \lambda u e^{-\lambda u}$], it was necessary to allow $N \rightarrow \infty$ to show that N multiple sources have a "unit" distribution of Eq. 67. It will be the goal of further work to continue the study of the relation of N to the Poisson distribution assumed in Eq. 67. Although experimental evidence would be very helpful here, it is exceedingly difficult to obtain.

In any event the distribution of events for the equivalent source which feeds the buffer will be taken to be Eq. 67.

3.1 Queueing-Theory Approach

The analysis of any buffering situation can usually be interpreted as a basic queueing-theory problem. We will present briefly the basic ideas associated with this. However, it will be found that in this case results are found much more easily by considering a special case of a queueing problem, which neglects transients and is relatively simple. Therefore the bulk of the material will be devoted to this special analysis.

Consider for a moment the buffering problem as a queueing-theory problem; then one speaks in terms of "state probabilities." That is, the buffer is said to be in the n th state when n samples are in storage. Thus the n th state probability (P_n) is the probability of exactly n samples in the buffer. Thus if one evaluates P_M (where $M =$ stages of buffer), one can find the average freezeout fraction---that is, the average number of samples from the equivalent source which will be turned away. The probability P_0 , which is the probability that the buffer is empty, is related to the channel idle time, that is, the fraction of the total time that the transmitter is idle.

If one is interested only in steady-state conditions, then the following equations must be solved, to get any P_n (written in matrix form):

$$P_n(s\tau) = [T] P_n(s-1)\tau \quad (68)$$

where: $P_n(s\tau)$ = probability of n items in buffer at time $s\tau$ (column matrix)
 $P_n(s-1)\tau$ = probability of n items in buffer at time $(s-1)\tau$ (column matrix)
 $[T]$ = matrix of transitional probabilities, which can be found using a state diagram and Eq. 67.

For any steady-state buffer operation of the type here, the state probabilities (P_n) are cyclo-stationary in time; that is, the P_n 's are cyclic with the time intervals τ . Therefore:

$$P_n(s\tau) = P_n(s-1)\tau = P_n(s-m)\tau \quad \text{for } m=0,1,2, \dots \quad (68a)$$

Consequently Eq. 68 appears as:

$$P_n] = [T] P_n] \quad (68b)$$

In addition, for an M-long buffer, one has the equation:

$$\sum_{n=0}^M P_n = 1 \quad (69)$$

Equations 68a and 68b result in a set of M+2 equations in M+1 unknowns. Selecting any M equations from the matrix of (68b) and using Eq. 69, one can solve directly for the state probabilities as indicated in the appendix. This is the queueing-theory approach, then, to the steady-state buffer situation.

It should also be noted that use of the queueing-theory approach above allows one to solve for transients in the buffering situation, i. e. , the probabilities of given states after s intervals, assuming that one started with a buffer of given state probabilities. In this case the more general equation is:

$$P(s\tau) = [T]^s P(0) \quad (70)$$

where: $P(s\tau)$ ≡ a column vector of state probabilities after s intervals
 $P(0)$ = a column vector of initial state probabilities
 $[T]$ = the transition matrix.

3.2 Specialized Approach

Although the queueing theory just described is more general, it is simpler to obtain the steady-state buffer conditions (which are of interest here) by a more specialized approach. Consider that a clock pulse occurs at every τ interval, at which time an output is taken from the buffer. The technique used will be to write a recursion relation between state probabilities just after the clock pulse to determine the state probabilities prior to the clock pulse. The quantities of interest can then be related to these state probabilities.

We will now portray this more specialized analysis of the buffering situation. The quantities of interest will be the freeze out fraction, analysis of freeze out runs, channel idle time, and the wait time distribution of the input samples.

The state probabilities describing the occupancy of the storage stages cannot approach stationarity after long operation, but will approach "cyclo-stationarity" (mentioned earlier). That is, the probability of a given state will be a function of the time that has elapsed since the last clock pulse. A way to analyze this queueing problem is to single out two sets of state probabilities. Let:

- α_k = probability that exactly k buffer stages are occupied, just after a clock pulse; in other words, after one sample has been removed, if it was available;
- β_k = probability that exactly k buffer stages are occupied just before a clock pulse ($\beta_k = P_k$ of Eqs. 68 and 69).

The definitions of these two particular distributions are illustrated in Fig. 29

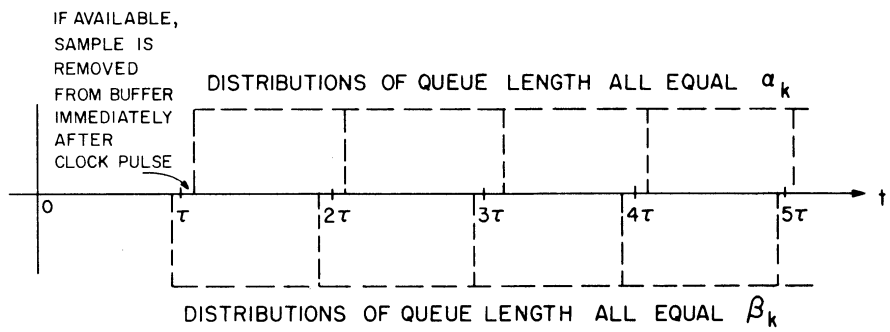


Fig. 29. Depiction of the probabilities α_k and β_k along a time axis.

Note that cyclo-stationarity means that the probability distribution of queue lengths for any instant of time within a clock interval will be the same from clock interval to clock interval (since there is no reason to favor a particular clock interval). Since one sample is removed from the buffer at each clock pulse, the following relations are valid:

$$\begin{aligned}\alpha_0 &= \beta_0 + \beta_1 \\ \alpha_k &= \beta_{k+1}, \quad k \geq 1\end{aligned}\tag{71}$$

With the help of the Poisson distribution of Eq. 67, it can be shown (see Appendix B) that:

$$\left. \begin{aligned}\beta_0 &= e^{-u} \alpha_0 \\ \beta_1 &= u e^{-u} \alpha_0 + e^{-u} \alpha_1 \\ \beta_2 &= \frac{u^2}{2} e^{-u} \alpha_0 + u e^{-u} \alpha_1 + e^{-u} \alpha_2 \\ &\vdots\end{aligned}\right\} \text{etc.}\tag{72}$$

Then

$$\begin{aligned}\beta_1 &= \alpha_0 - \beta_0 = (1 - e^{-u}) \alpha_0 \\ \alpha_1 &= e^u [\beta_1 - u e^{-u} \alpha_0] = e^{-u} [\alpha_0 - e^{-u} \alpha_0 - u e^{-u} \alpha_0] = \alpha_0 [e^u - u - 1] \\ \beta_2 &= \frac{u^2}{2} e^{-u} \alpha_0 + u e^{-u} \alpha_1 + e^{-u} \alpha_2 \\ \alpha_2 &= e^u [\beta_2 - \frac{u^2}{2} e^{-u} \alpha_0 - u e^{-u} \alpha_1] = [e^u - u] \alpha_1 - \frac{u^2}{2} \alpha_0\end{aligned}\tag{73}$$

In general, by continuing this iteration,

$$\alpha_k = [e^{u-u-\delta_{1,k}}] \alpha_{k-1} - \sum_{n=2}^k \frac{u^n}{n!} \alpha_{k-n} \quad \begin{matrix} 1 \leq k \leq M-1 \\ M \leq k \end{matrix}\tag{74}$$

$$= 0$$

where: $\delta_{1,k}$ = the Kronecker delta.

Thus, for a given input rate, u samples per clock interval, one can find the probability of any queue length (or state) at the beginning of a sampling interval in terms of the probability α_0 (of zero length).

If the queue length at the beginning of a clock interval is not permitted to exceed $M-1$ (implying that the buffer has a capacity of M samples), and since all α_k 's can be re-

lated to α_0 by Eq. 74, α_0 can be determined from the following equation:

$$\sum_{k=0}^{M-1} \alpha_k = 1, \quad (75)$$

since the sum of the α_k 's encompasses all possibilities for queue length at that time. Each α_k is a simple multiple of α_0 , and is found by multiplying α_0 by the appropriate factor. The distribution for queue lengths at the beginning of a clock interval is then determined.

3.3 Freeze Out Fraction (FOF)

The values of α_0 found through use of the above procedures are substituted into a simple equation for freeze out fraction. As previously explained, freeze out occurs when there are M pieces in storage and more pieces arrive before one is "dumped" at the next clock pulse. The freeze out fraction is written as:

$$\begin{aligned} \text{FOF} &= 1 - \frac{\text{av. no. of samples emerging from buffer each clock interval}}{\text{av. no. of samples entering buffer each clock interval}} \\ &= 1 - \left[\frac{1 - e^{-u} \alpha_0}{u} \right] = \frac{u - 1 + e^{-u} \alpha_0}{u} \end{aligned} \quad (76)$$

This expression leads to relations with FOF as a function of u and buffer capacity M, plotted as the family of curves of Fig. 30. In the design of a nonsynchronous multiplexed system the curves can be used to find the best compromise among the three parameters u (input rate), M (buffer capacity), and FOF (freeze out fraction).

It is seen that, as M increases, the FOF drops appreciably as u is decreased. A u of 0.8 is chosen as a suitable calculation point for the system evaluation (Section 5). Also, in that section a tolerable FOF of .05 percent (.0005) will be selected.

3.4 Freeze Out Runs

Figure 30 evaluates the expected fraction of total samples (from all the sources) which will be frozen out. In addition to this number it is of interest to have some information on the distribution of rejected samples from a given (single) source. This is a difficult problem and has not been solved completely. However, the following probabilities have been solved, and bear on this problem.

First, consider the distribution of freeze out "runs:"

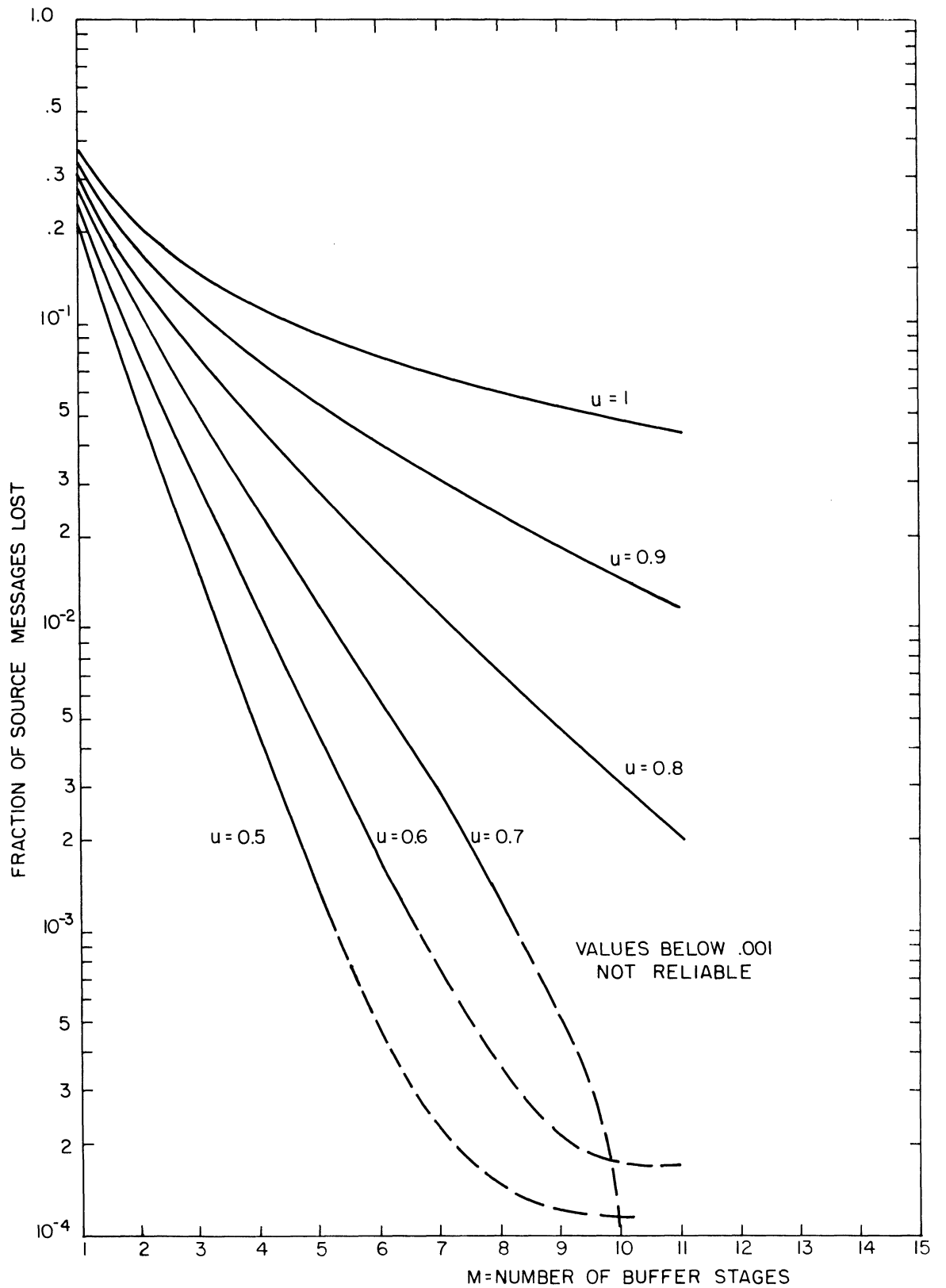


Fig. 30. Freeze out fraction as function of input rate u and storage capacity M .

let: $P_r(c)$ = probability of a run of exactly c clock intervals
 (during which one or more samples from some
 source are frozen out during each interval).

It can be shown (see Appendix C) that:

$$P_n(c) = (1 + u)e^{-u}[1 - e^{-u}(1 + u)]^{c-1}, \quad c \geq 1. \quad (77)$$

Next, consider the probability of a given number of units frozen out, provided that "one or more" was frozen out in the previous interval.

Let: $P(d | f \geq 1)$ = probability that exactly d units are
 frozen out, in a given interval,
 provided that one or more was frozen
 out in previous interval; i. e., the in-
 terval of interest is within a run.

Then it is shown in Appendix C that:

$$P(d | f \geq 1) = \frac{e^{-u} u^{d+1}}{(d+1)!} \quad (77a)$$

Table IV presents evaluations of $P_r(c)$ and $P(d | f \geq 1)$ for $c=1$ through 5 and $d=1$ through 5, for $u = 0.8$.

c or d	$P_r(c)$	$P(d f \geq 1)$
1	80.9 %	80.88 %
2	15.4 %	14.38 %
3	3.0 %	3.53 %
4	0.6 %	0.71 %
5	0.1 %	0.50 %

Table IV. Evaluations of $P_r(c)$ and $P(d | f \geq 1)$.

The $P_r(c)$ results indicate that samples arriving at least 5 intervals apart will be very unlikely to be members of the same freezeout run. With 10 or more channels partici-

pating, the probability of two consecutive freeze outs from the same channel becomes very small. There are other possibilities for consecutive freeze out besides a single freeze out run---for example, two freeze out runs separated by a moderate number of "no-freeze out" intervals. However, the rapid fall in $P_r(c)$ for increasing c is indicative of the trend for other possible realizations of two or more consecutive freeze outs from a given channel, so that for $u = 0.8$ and $M > 10$, little deviation from average FOF rate is expected.

The significance of the $P(d)$ results is that the number of lost samples during a freeze out run can be expected to be somewhat more than the number of clock intervals constituting the run, as is expected. One may expect that the ratio of frozen out pieces to clock intervals in the freeze out run will be less than 1.5.

3.5 Starting Transients

When the system begins operation, the buffer will be empty and thus a starting transient can be expected. The most significant aspect of this transient would be the lack of any freeze out for a period of time. No detailed analysis has been made of starting transients because such transients would enhance rather than degrade the freeze out situation. Transients can be calculated using the matrix equation (70).

3.6 Channel Idle Time

The objective of "channel idle time" is to estimate the fraction of clock intervals during which no samples are transmitted. Since the normalized average rate of entry of samples into the buffer is usually less than 1 ($u = 0.8$ is suggested), and since some samples are lost through freeze out, there will be a number of clock pulses occurring when the buffer is empty and nothing can be transmitted. The fractional quantity of such pulses is just β_0 , the probability of the buffer being empty at the end of a clock interval. Thus:

$$\text{Fraction of unused channel slots} = \beta_0 = e^{-u} \alpha_0 \quad (78)$$

Channel idle time is plotted in Fig. 31 as a fraction β_0 vs. M . The plot shows that minimization of channel idle time requires a high input rate, large buffer capacity, and consequently longer delays. These results are not unexpected; information theory has always recognized, particularly in the method of coding by large blocks, that the attainment of a smooth flow of information from a system whose input is irregular requires considerable delay and storage.

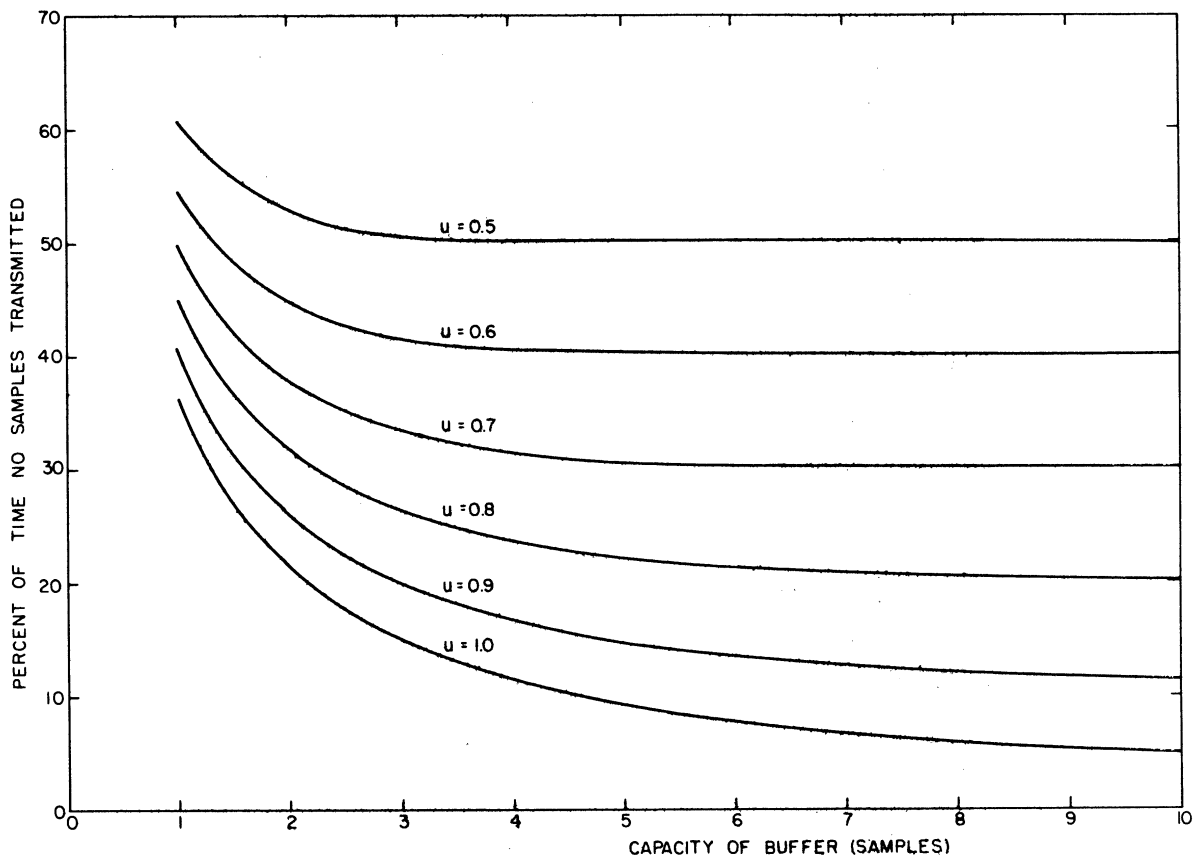


Fig. 31. Channel idle time versus buffer length.

In Fig. 31 it is seen that, for a u of 0.8, the idle time is about 20 percent for buffer lengths > 10 . Although this may seem inefficient, it must be remembered that in synchronous multiplexing the idle time is greater than 50 percent for telephone communicators (over radio and cable links). It will be seen, in Section 5, that asynchronous multiplexing combined with a speech coding (such as extremal coding) provides a more efficient multi-channel system.

Also, consideration should be given to the possibility of sending low-priority non-real time data during such idle times.

3.7 Wait Time Distribution

For the buffered system it is of interest to examine the distribution of waiting times for samples entering the buffer. Because of the cyclo-stationary pattern of buffer occupancy, this distribution will not be a smooth function. A figure is also provided in which the variation of waiting times from the expected values is plotted. The most important consideration, as far as time placement of samples is concerned, is not the average delay of a

message, but rather the unpredictable perturbation of individual samples from their correct relative positions in time. The noise thus generated must not be excessive.

Figure 32 clarifies the measurement of waiting time for a sample entering at a fraction θ of a time interval τ and encountering k pieces already in the buffer.

Derivation of Distribution

Consider the following notation:

$P(A, B)$ = joint probability of the events A and B.

$P(A|B)$ = probability of A conditional on B.

In the discussion of freeze out fraction, the following distribution for samples entering the buffer was used:

$$p(k, u) = e^{-u} \frac{u^k}{k!} \quad (79)$$

where: $p(k, u)$ = probability of k samples arriving during the interval τ , when the average rate of arrival is u .

When considering "waiting times," it is necessary to break the time axis into intervals smaller than τ . Assume θ is the fraction of τ at which a particular sample enters, so that θ varies from 0 to 1. Assume further that the distribution of samples entering during a period $\theta\tau$ remains Poisson. Then the following equation is applicable:

$$p(k, \theta u) = e^{-\theta u} \frac{(\theta u)^k}{k!} \quad (80)$$

where: $p(k, \theta u)$ = probability of k samples arriving during the interval $\theta\tau$.

It is desired to find the waiting time probability density vs. waiting time, with the rate u and the buffer capacity M as parameters. To do this it is first necessary to find the (discrete) probability of waiting time with u regarded as a constant. Then θ will be allowed to vary continuously (from 0 to 1) and a probability density will be defined.

The technique used is to evaluate the probability of waiting time for the $(k+1)$ th sample, assuming that it arrives at a time $\theta\tau$. First,

$$P_{\theta}(k) = \sum_{n=0}^k a_{k-n} \frac{(u\theta)^n e^{-u\theta}}{n!} \quad (81)$$

where: $P_\theta(k)$ = the probability that k samples are in the buffer immediately before the arrival of the $(k+1)$ th sample.

α_{k-n} = the probability that $(k-n)$ samples are in the buffer at the very beginning of the τ interval.

It may be noted that the wait time w for the $(k+1)$ th sample, arriving at time θ , will be:

$$w = k + (1-\theta) \text{ clock intervals } (\tau\text{'s}) \quad (82)$$

In order to find the probability of waiting time, it only remains to note that k may not exceed M , since additional arrivals ($k \geq M$) will be frozen out. Thus one needs to evaluate a conditional probability---the probability of waiting time conditional on "no freeze out," i. e., conditional on our test sample getting into the buffer. This is given by the usual equation:

$$P_\theta(k|\text{get in}) = \frac{P_\theta(k, \text{get in})}{P_\theta(\text{get in})} \quad (83)$$

where: $P_\theta(k|\text{get in})$ = probability of a sample encountering k samples already in the buffer when it enters at time θ , if it enters at time θ

$P_\theta(k, \text{get in})$ = the joint probability of getting into the buffer and encountering k samples already there, at time θ , and is equal to $P_\theta(k)$ since only those values of k permitting entry into the buffer by a new arrival will be considered ($k \leq M-1$).

$$P_\theta(\text{get in}) = \sum_{k=0}^{M-1} P_\theta(k) = \text{sum of probabilities of all possible values of } k \text{ (at time } \theta \text{) that do not prevent entry of a new sample.} \quad (84)$$

Since $P_\theta(k)$ has been defined in Eq. 81, a simple substitution of Eqs. 81, 82 and 84 into Eq. 83 yields the desired "wait time" distribution:

$$P_\theta(w-1+\theta|\text{get in}) = P_\theta(k|\text{get in}) = \frac{\sum_{n=0}^k \alpha_{k-n} \frac{(u\theta)^n e^{-u\theta}}{n!}}{\sum_{k=0}^{M-1} \sum_{n=0}^k \alpha_{k-n} \frac{(u\theta)^n e^{-u\theta}}{n!}} \quad (85)$$

where: $k = w-1+\theta = \text{an integer.}$

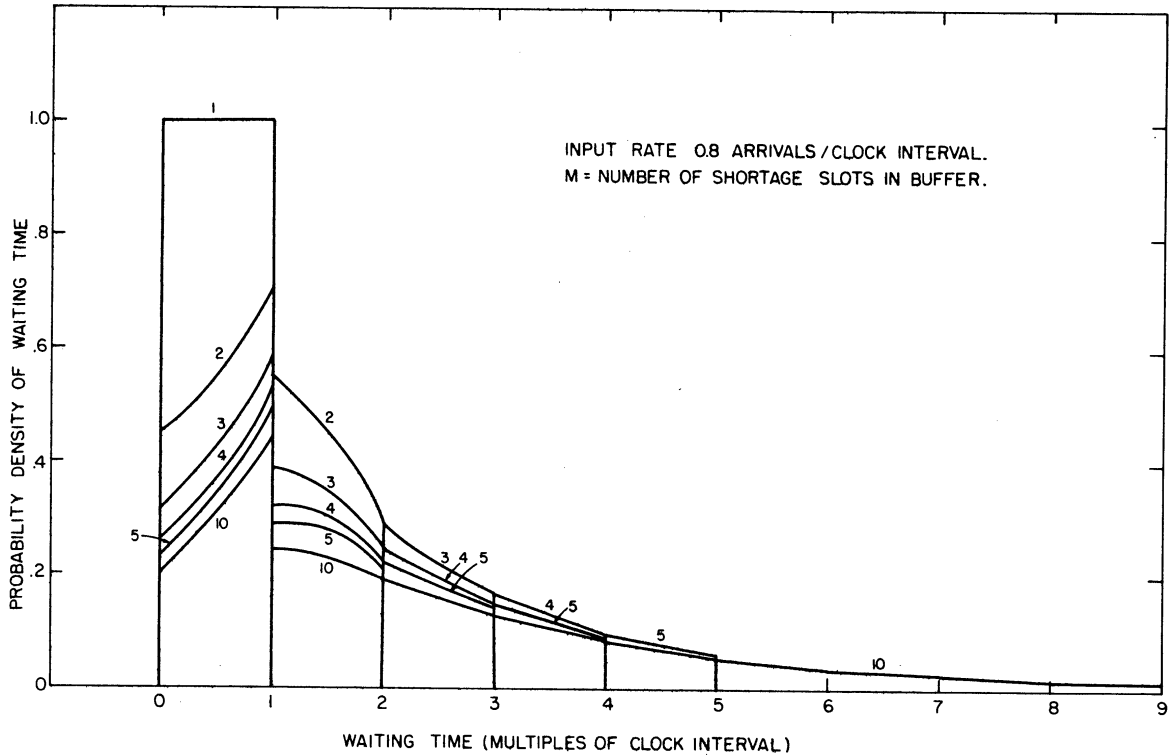


Fig. 32. Probability density curves of waiting time in buffer.

The distribution can be evaluated for one θ and all k 's $\leq M-1$, yielding the probability of M discrete waiting times. A probability density for continuous θ is, however, the physical problem. If the interval τ in which θ is defined is considered to be of unit length (i. e., we measure time in τ 's), the formula (85) will give the probability density for the value of θ specified. If several sets of calculations for different values of θ are made, smooth curves may be drawn through the resulting points to give the smooth probability density curves. Figure 32 is a plot of these curves for $M = 1, 2, 3, 5, \text{ and } 10$, and input rate 0.8 pieces/clock interval. Figure 33 presents the same curves, but with each shifted by an amount equal to its mean waiting time. Means were graphically computed from Fig. 32. Since perturbation of waiting time from the mean value is of principal importance, Fig. 33 will be the more valuable.

For buffers with capacity greater than 1, the most likely range of waiting times remains between 0 and 1, with emphasis toward 1. This situation is strongly implied by the

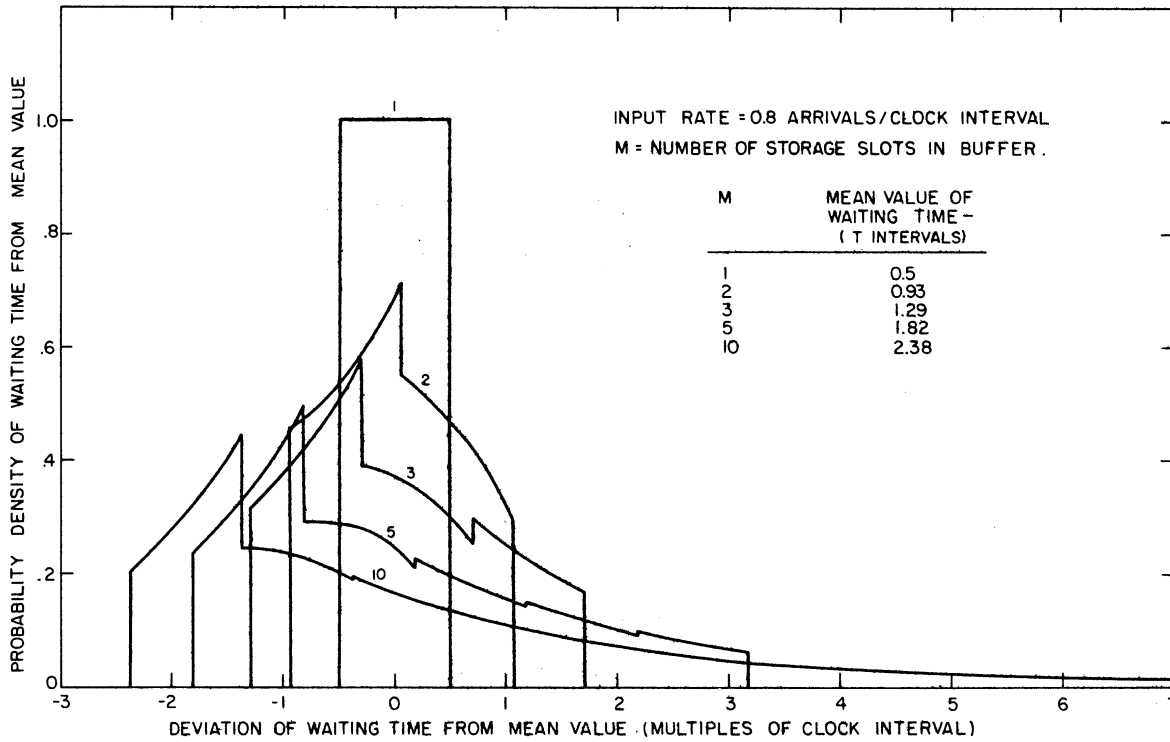


Fig. 33. Probability density curves of deviation of waiting time in buffer from its mean value.

physical picture, in which a sample arriving just after a clock pulse is likely to encounter no samples at all in the buffer and can thus expect to wait almost 1 clock interval.

The above graphs depicting the buffer performance will be referred to when evaluating system performance in Section 5.

4. RECONSTRUCTION OF WAVEFORMS

Before proceeding to evaluate the asynchronous system performance, using the above buffer analysis, we consider here the reconstruction of waveforms when the samples are irregular. It is noted that, with the extremal coding being considered here, one must reconstruct the waveform from extremely irregular samples for each source. First the theory for regular and irregular sampling is reviewed, and then comments about practical reconstructions will be given.

4.1 Theoretical Sampling Relations

The theory for uniform sampling is based on the familiar Shannon Sampling Theorem for bandlimited waveforms, where the waveform can be theoretically reconstructed using the relation:

$$f(t) = \sum_{n=-\infty}^{\infty} f\left(\frac{n}{2W}\right) \phi_n(t) \quad (86)$$

where: $\phi_n(t)$ = sin x/x type composing function

W = bandwidth of the bandlimited waveform.

Here the uniform samples are taken at times $t = n/2W$ ($n=0, \pm 1, \pm 2, \dots$). The composing function is given by:

$$\phi_n(t) = \frac{\sin 2\pi W(t - \frac{n}{2W})}{2\pi W(t - n/2W)} \quad (87)$$

For uniform sampling, the same composing function (except for a delay time) multiplies every sample. An ideal low-pass filter (with an ∞ delay) would reconstruct the $f(t)$ from the uniform samples.

There is similar theory available for nonuniform sampling (Ref. 33). One important difference for this nonuniform sampling theory is that one considers "energy-limited" signals, rather than "time-limited" signals (as is done in the Sampling Theorem). Noting this difference, the following theorem is stated in Ref. 33: "If the sample values at a finite

set of arbitrarily distributed sample points $t = \tau_p$ ($p = 1, 2, \dots, N$), are given, a signal $f(t)$ with no frequency component above W cps is defined uniquely under the condition that the 'energy' of the signal $\int_{-\infty}^{\infty} f^2(t)dt$ is a minimum." Note that the exact duration of the waveform is not required. The reconstruction of the signal is given by:

$$f(t) = \sum_{p=1}^N f(\tau_p) \psi_p(t) \quad (88)$$

$$\text{where: } \psi_p(t) = \sum_{q=1}^N a_{qp} \frac{\sin 2\pi W(t-\tau_q)}{2\pi W(t-\tau_q)}$$

The coefficients a_{qp} are obtained from the inverse of a matrix, where the matrix has elements given by:

$$\frac{\sin 2\pi W(\tau_p - \tau_q)}{2\pi W(\tau_p - \tau_q)}, \quad p, q = 1, 2, \dots, N. \quad (89)$$

Although such nonuniform sampling theory is of potential interest for codings like extremal coding, it is not clear how it may be applied to extremal coding. We have no conclusions as to the application of this nonuniform theory (for our purposes), but simply note its existence here for future study.

4.2 Reconstruction Methods

Here we wish to note two methods for reconstructing speech waveforms from the nonuniform samples of extremal coding. They are: (1) inset a standard interpolation function between samples, and (2) low-pass filter the "boxcarred" samples. Both of these have been experimentally tested, and shown to yield acceptable quality.

Interpolation Function Method

In general, the waveform between two arbitrary samples (a_i and a_{i+1}), using an interpolation function $F(X)$, can be written:

$$S(X) = a_i + (a_{i+1} - a_i) F(X) \quad X_{a_i} \leq X < X_{a_{i+1}} \quad (90)$$

where: $F(X)$ = interpolation function

$$F(X_{a_i}) = 0; F(X_{a_{i+1}}) = 1.$$

Writing in terms of time:

$$S(t) = a_i + (a_{i+1} - a_i) F\left(\frac{t-t_i}{t_{i+1}-t_i}\right) \quad t_i \leq t < t_{i+1} \quad (91)$$

where: $F(t_i) = 0$; $F(t_{i+1}) = 1$.

A particular interpolation function experimented with extensively (Ref. 28) for extremal coding is of the type:

$$F(X) = X^2(3-2X) \quad (92)$$

This function provides a continuous first derivative at the sample points (extremum), and is found to produce good speech. In the work of Ref. 28 the experiments were done with a computer; in practice, one may have to devise an analog "inserter" to use the interpolation function approach.

Filtered Boxcar Method

A second method of reconstruction shown to give acceptable speech (Ref. 29) is to low-pass filter the "boxcarred" samples. This is certainly a relatively simple method. Its significance for our purposes is that the reconstruction of the waveform should not be a serious obstacle to implementing an asynchronous system with extremal coding.

5. PERFORMANCE OF SYSTEM

It will be the objective here to choose sensible parameters and then evaluate the digit rate for asynchronous multiplexing when extremal coding is used on the individual speech sources. This rate will then be compared both to standard companded PCM, and to another efficient multichannel method.

It should be remembered that the qualitative system operation was described in Section 2, and that Fig. 28 shows a block diagram of the transmitter side.

5.1 Choice of Parameters

In the following calculations we will assume that the number of sources (N) to be handled by the multichannel system is 100. This number is chosen for the following reasons:

- (1) With N of the order of 100 we are certain that the Poisson distribution (Eq. 67) is a good approximation for the "equivalent source" provided by the 100 sources (even considering that sources are active only half the time).
- (2) With $N \geq 100$, the τ interval is sufficiently small that one need only send "time" information specifying the "number of buffer intervals (τ) delayed." Also, the fixed delay, M_τ , improves as τ is made smaller.
- (3) An N of 100 is physically reasonable, since present Signal Corps capabilities go up to 96.

From the buffer analysis of Section 3, a u of 0.8 is chosen. This was chosen so as to incur a reasonable idle time (see Fig. 30) while at the same time getting very small FOF for reasonable sized buffers. For buffer length, an M of 16 was chosen. This requires 4 binary digits to specify the buffer delay time, and the corresponding FOF is less than .05 percent. Considering that the freeze out runs are distributed among the 100 sources, we feel that this parameter choice ($M = 16$, $u = 0.8$) is conservative.

The remaining parameters are selected with reference to the experimental tests

of extremal coding in Refs. 28 and 29. The parameters and explanatory comments are listed as follows:

(1) $Q = 64 =$ quantizing levels for the extremum samples.

It was found that, with this number of levels, the decoded speech was definitely comparable to 40,000 bits/second companded PCM.

(2) $R' = 1500$ samples/second = average rate of occurrence of extremum from an "on" source. Although Ref. 28 finds an R' of 1400, Ref. 29 uses 1600. We have chosen the compromise.

(3) $R'' = 750$ samples/second = average rate of extremum, considering that each source is "off" 50 percent of the time. This is the rate we shall use in the following calculations. This factor should still allow us to envisage a Bernoulli model so that Eq. 67 is valid.

(4) $\Delta t_M = .125 \times 10^{-4}$ seconds = necessary accuracy of location (in time) of the extremum. It was found that, when this was exceeded, a quaver occurred in the decoded speech (Ref. 28).

These parameters, and the preceding ones, will be used to calculate the digit rate.

5.2 Digit Rate

Before calculating the digit rate it is necessary to determine the extent of time information required to stay within Δt_M (above). This is done by evaluating τ , the length of buffer output intervals (see Section 2). The τ can be found using:

$$\tau = \frac{u}{R''N} = \frac{0.8}{7.50 \times 10^2 \times 1 \times 10^2} = .107 \times 10^{-4} \text{ seconds} \quad (93)$$

where the parameter values described above have been used. Thus τ is less than Δt_M for the given parameters. Consequently, one need only identify the number of τ intervals by

which the sample was delayed (in the buffer). Thus, the time information requires only M levels ($\log_2 M$ bits).

Note also, with an M of 16, that the steady-state delay ($M\tau$) is only 171 micro-seconds. In a two-way speech conversation, this delay can be as great as half a second without discomfort.

We can now evaluate the digit rate for the proposed parameters. First, the following bits are required per sample:

Quantity	Bits Required	
Amplitude of Extremum	6	
Source Number (N=100)	7	
Time information	4	
	17 bits/sample	(94)

To convert this to digit rate in bits/second, one divides Eq. 91 by Eq. 93. The result is:

$$R = \text{digit rate} = \frac{17}{1.07 \times 10^{-5}} = 1.59 \times 10^6 \text{ bits/second} \quad (95)$$

It may appear that 17 bits/sample is large since a synchronous multiplex (PCM with uniform sampling) requires only 6. However, it must be remembered that the extremum samples have an average rate of only one fifth of uniform sampling, and that the extrema occur only when a source is active. It will be seen that the balance of these items lies in favor of the asynchronous method.

5.3 Comparisons

It is now of interest to estimate the quality associated with the asynchronous operation, and to compare the digit rate with existing and proposed alternative methods. Table V depicts the comparison of extremal coding for $N=100$ with time-multiplexed, companded PCM (and with log-differential PCM). We now discuss this comparison.

Based on the extremal coding listening tests of Ref. 28, it is concluded that the quality obtained with the digit rate of Eq. 95 is comparable to 40,000 bits/second, companded PCM. Using some curves in Ref. 34 (also reported in Ref. 35), 40,000 bits/second companded PCM results in about 21 db of quantization signal-to-noise ratio.³ (Note that this

³This quantization signal-to-noise ratio $[(S/N)_q]$ should not be confused with the channel S/N ratio of Part 1, Section 4.2. The channel S/N (related to E/N_0) is a measure of the channel disturbances, whereas $(S/N)_q$ is a measure of the disturbance due to digital modulation (quantization).

value is an estimated average over a 25-db signal level range.) The $(S/N)_q$ for a present military system (6 bits at 8 kc) is about 26 db (from the same curves). Consequently, it appears that the reduced digit rate of asynchronous multiplexing (Eq. 92) may incur about a 5 db reduction in $(S/N)_q$; we feel this is quite acceptable. The digit rate is reduced by a factor of 3.

In Table V, the time-multiplexed, companded PCM (at a 48,000 bit/second rate per source) is used as the reference. It may be suggested that one might use "time assignment" (such as in TASI---see Section 1.1) on the time-multiplexed PCM signal and decrease the multichannel digit rate by a factor of 2. On the basis of a brief study of this, we concluded that the equipment complexity of this is much too great for radio or cable links (i. e., economic factors for radio links are not comparable to those for oceanic cables).

Another possibility for handling multichannel speech digitally consists of using a speech coding of the "delta-modulation" type (with uniform sampling). It is of interest to compare the asynchronous case with this type. One of the most efficient methods tested thus far (Ref. 35) is "Log-Differential PCM." Although this log-differential scheme was not tried on speech itself, experimental results using sine waves are reported in Ref. 35, in which it is essentially concluded that four-digit log-differential PCM sampled at 9600 cycles per second is about equivalent to six digits companded PCM. Thus one can regard the four-digit PCM as effecting a saving of two digits per sample. Thus, for $N=100$, the digit rate would be 3.84×10^6 bits/second. This would provide an estimated $(S/N)_q$ of 26 db, as in the six-bit PCM case. Another case would be to use a three-bit log differential case. Here the estimated $(S/N)_q$ is about 17 db. Each of these cases is noted in Table V.

	Bit Rate for N=100	Bit Rate Ratio (to PCM)	Assumptions	Quality Estimates	Remarks
6-Digit Companded PCM (Time-Multiplexed)	4.8×10^6	1	Time-multiplexed PCM Each source sampled at 8 kc Companded	Good military telephone quality $(S/N)_q \approx 26$ db	This is our standard of comparison for the other two systems. Multiplex PCM is fairly complex in equipment.
Asynchronous Time-Multiplexing with Extremal Coding (Buffer Time Encoded)	1.59×10^6	3	$R' = 1500$ samples/sec. from each active source. Each source "on" about one half the time. $N = 100 =$ total number of channels. Four digits used to encode time information --- in the form of identifying the buffer delay for that sample. Time need only be specified within a time interval τ . ($\tau \approx .125 \times 10^{-4}$). Freeze out rate of .05 % is acceptable for extremal coding.	Quality is comparable to 4 digit, companded PCM sampled at 10 kc. (Based on listening tests in Ref. 28) $(S/N)_q \approx 21$ db	Automatic use is made of the fact that the channels are active about one half the time; thus, one obtains an automatic "time assignment." Such an asynchronous system permits an "elastic" overload, if required. Non-real time data may be put in the channel "idle time." Receiver extrapolation fitting can be expected to be complex in equipment, if a "boxcar" detection is not satisfactory. A device to sense extremes of a waveform would have to be developed. In total, this appears to be a good system to pursue.
4-bit Log-Differential PCM (Time-Multiplexed)	3.84×10^6	1.25	Sine wave experimental tests of Ref. 35 are good indicator for actual speech transmission. Some method would be provided to prevent channel noise from seriously affecting the performance of the delta-modulation.	$(S/N)_q \approx 26$ db Based on sine wave experiments of Ref. 35. $(S/N)_q \approx 17$ db Based on sine wave experiments of Ref. 35.	Each channel requires a "differential-encoder." System offers possibility of acceptable speech at "standard" transmission rates.
3-bit	2.88×10^6	1.67			

Table V. Comparison of three methods of digitally sending multichannel speech.

6. CONCLUSIONS

A comparison for the three systems---time-multiplex PCM, asynchronously time-multiplexed extremal coding, and time-multiplexed log-differential PCM---is shown in Table V. It is seen that the asynchronous system offers a substantial bit-rate improvement over that of time-multiplex PCM and log-differential PCM. The asynchronous system would consume only one third of the bandwidth (or corresponding improvement in channel S/N) of that of PCM. One cost of this improvement is about a 5-db loss in quantizing S/N ratio. It should be remembered that conservative figures were used throughout to obtain the estimated values. Further, extremal coding has been investigated by two groups and found to produce good quality speech.

For a number of reasons we conclude that asynchronous multiplexing has great potential. We note the following:

- (1) Automatic use is made of the fact that the speech sources are "off" half the time. In other words, since the channel is only being used when an extreme occurs, it naturally follows that one automatically obtains the advantages of time assignment.
- (2) The extremal coding should not be as seriously disrupted by channel noise as the delta-modulation methods.
- (3) Nonreal time data may be conveniently sent during the idle time of the asynchronous system. Thus, with the system parameters used for Table V, one obtains a potential data capability of 20 percent of the total channel capability. Thus one obtains not only a more efficient multichannel speech system, but also the capability of sending irregular data.
- (4) The asynchronous method allows one to temporarily add more sources than designed for. The degradation in performance would be gradual (or "elastic"). This is not possible for any synchronous multiplexing system.

Probably the main obstacles to actual implementation of this asynchronous method are: (1) guaranteeing that the designed freeze out fraction is not noticeable; (2) assuring that time jitter of the extremal samples is kept within tolerance; and (3) developing suitable coding detectors if boxcarring-plus-filtering does not give sufficient quality. We feel that each of these obstacles can be handled with reasonable effort.

In conclusion, it is believed that the advantage of (3) above over PCM, which appears from conservative calculations, and the other advantages listed above, are sufficient to justify serious feasibility studies of the asynchronous method. The next step could be either a simulation investigation, or an actual experimental set-up. Each of these appears to be quite extensive. The main objective of the above work has been to establish that a sufficient gain is available, so as to warrant extensive further efforts.

APPENDIX A: DETERMINATION OF THE MAXIMUM NUMBER OF
BINARY DIGITS PER SECOND THAT CAN BE TRANSMITTED
OVER A CHANNEL OF BANDWIDTH W

(T. G. Birdsall)

A.1 Introduction

There are several studies in the literature of minimum time bandwidth products which can be achieved for various waveforms. This discussion considers what appears to be the same problem of minimum time bandwidth product, but it differs from other studies sufficiently to justify performing the detailed work here explicitly, without reference to other material. Previous studies typically take two forms: (1) determining the minimum time bandwidth product for the duration of the signal when the bandwidth of the signal is measured on specific points of the waveform and spectrum (such as the 3-db points or the 6-db points of the spectrum); (2) placing very stringent conditions on one or the other. That is, the duration may be measured from the very start to the very end of the waveform but the bandwidth measured at some moderate point such as the 6-db point, or conversely, the spectrum may be limited stringently and the duration of the waveform measured at some moderate point. This is a reflection of the relationship between time and bandwidth for a single waveform. The spectrum of a definitely time-limited waveform is itself unlimited, and conversely a strictly bandlimited waveform cannot be strictly time-limited.

The purpose of this section is to determine the limitations on the time rate of signalling for block codes such that there is no interblock interference in the demodulation process. The logic of the analysis is based on the assumption that the bandwidth limitation placed upon the communication system is very stringent, like that laid down by the FCC in an assignment of the channel. It is assumed that the communicator does not have control of the adjacent channel transmission; therefore, it is assumed that the bandwidth limitation is of the type that assumes that the energy spilled out of the band, if any, is some low figure like 40 db or 60 db. For this reason the calculation will be based on the mathematical assumption that the band is absolutely limited, and in practice the only energy spilled out of the band will be the result of the practical generation of signals which approximate to a high degree

the mathematically bandlimited signals (but this approximation is only good to 1 percent or 1/10 percent in the rms spectrum). The second assumption is that the reception will be sufficiently well-controlled that, in determination of the rate of signalling, we may consider the effective duration of a signal to be the quotient of the number of signals that are sent over some long time, divided by that long time; that is, the duration of the signal is the average amount of time that is assigned to any single signal. Specifically, this is shown in Eq. 96.

$$T = \lim_{N \rightarrow \infty} \frac{\text{duration of } N \text{ symbols}}{N} \quad (96)$$

The following three sections will deal with three specific cases of signalling. All of these assume that the transmission and reception are time synchronous. The first two will deal with RF phase coherent reception for two-polarity transmission, and orthogonal transmissions, and a third section will deal with phase incoherent reception and orthogonal transmission. In each section a maximum signalling rate, the maximum number of digits per second, will be determined under the given transmission and reception limitations, using the criterion that there will be no interblock interference caused by spillover of energy of one block of symbols to the next. This is important because the basic assumption in computations of information rate and error probabilities made in the analysis of block codes is that each block can be considered as an entity, without regard to previous or subsequent transmissions.

A. 2 Time Synchronous Block Codes, Phase Coherent Reception, Two-Polarity Transmission

The analysis in this section is the direct application of Shannon sampling theorems, Ref. 10, for strictly bandlimited waveforms. If we consider any waveform, $x(t)$, which is strictly bandlimited, this theorem tells us that, if the values of the waveform are specified at instants of time separated by equal intervals of $1/2W$ seconds apart, then the entire waveform has been specified. This is shown in Eq. 97.

$$x(t) = \sum_{k=-\infty}^{\infty} x\left(\frac{k}{2W}\right) \frac{\sin \pi(2Wt - k)}{\pi(2Wt - k)} \quad (97)$$

$$I(t) = \sum_{k=1}^N v_k I_k \frac{\sin \pi(2Wt - k)}{\pi(2Wt - k)} \quad (98)$$

The sampling functions, or "reconstruction functions" as they are often called, are commonly

known familiarly as $\frac{\sin x}{x}$. Our interest here is in binary transmissions which can take on either polarity, plus or minus. Let us assume that the amplitude has been specified at the fixed level v . We could consider transmitting a waveform $I(t)$ such as shown in Eq. 98, composed of N information digits, with no transmission preceding or following it. Since each of the reconstruction functions is itself strictly bandlimited to the band W , and in fact, uniformly occupying the entire bandwidth, then the transmitted waveform $I(t)$ is also strictly bandlimited. In a strict sense the duration of this information waveform is infinite; however, if we specify some small amplitude and measure the duration out to the occurrence of this small amplitude, we will find that it occupies a duration somewhat longer than $\frac{N}{2W}$. As we increase the number of information digits that we consider to be transmitted, then the fraction of the duration that exceeds $\frac{N}{2W}$ becomes vanishingly small and the time duration that can be assigned to each single binary symbol is $\frac{1}{2W}$.

$$\lim_{N \rightarrow \infty} \frac{\text{effective duration of } I(t)}{N} = \frac{1}{2W} \quad (99)$$

We have thus demonstrated theoretically a form of signalling for which the $2WT$ product is unity. We must check to see that there is no interblock interference in the demodulation period. If one were to sample the received waveform composed of the attenuated transmitted waveform, plus added noise, at the synchronous times corresponding to the sampling points as they were impressed on the signal, one would recover the information digit. The preceding and subsequent information digits would have no effect on this sampled value, since their reconstruction functions are going through zeros at this instant of time. In addition, one would have added to the information digit the instantaneous value of noise occurring at that time. Since the $2WT$ product of the information waveform is unity, no advantage is available by the use of crosscorrelators or matched filters, since the signal-to-noise ratio improvement factor for such devices is simply the $2WT$ product, here unity.

It is obvious that such a signalling system could not, in practice, use the $\frac{\sin x}{x}$ functions, but as a limitation it is still valid, since one could use actual functions close to the $\frac{\sin x}{x}$ which would spill energy outside of the bandwidth to as small a degree as necessary. The price that would be paid is complexity of the modulation equipment and the delay incurred (which typically might be 10 or 20 signal durations). In most communication applications such delays are insignificant. This unity $2WT$ product might also be accomplished by other tech-

niques than using $\frac{\sin x}{x}$ transmissions and encodement in time. However, if one attempts to get a smaller $2WT$ product than unity, an analysis of the sampling theorem demonstrates that intersymbol interference must occur and no filtering technique will be available to remove such interference, since the waveform can contain at most $2W$ independent bits of information per second if this is going to be impressed on it by binary techniques.

A more general but more abstract consideration of this limitation can be obtained by viewing the vector space representation of a transmitted waveform. Let us consider a sufficiently long duration, $2WT$, which is much larger than one. This is the signal space which is available for the transmission of information. We would like to find subspaces in this space, each subspace corresponding to a specific bit of information, such that these subspaces are mutually orthogonal, i. e. , every vector in a subspace is orthogonal to any vector outside of the subspace. When we have the freedom to use two-polarity transmissions, since the information is binary, we can look for orthogonal co-ordinate systems in this large dimensional space and use each axis of the co-ordinate system for one bit of information. The value of the co-ordinate on each individual axis is chosen either plus or minus a specific quantity. There are an infinite variety of co-ordinate systems that one could lay down in the space, but by pure counting of the dimensionality of the space, any given orthogonal co-ordinate system contains exactly $2WT$ vectors. Thus the amount of information that can be transmitted is $2WT$ bits using a total duration of T seconds, and therefore one can transmit $2W$ binary symbols per second for an average duration of $1/2W$ for each symbol.

Figure 34 shows the central portion of a $\sin x/x$ reconstruction function which might be used for maximum rate transmission. Figure 35 shows a pulse function which is four times as long, that is, a $2WT$ product of 4, with the energy just over 50 db outside of the $2/W$ duration of the pulse. These are included to show that the prime engineering problems will be encountered in attempting to get short duration signals which are individually time-concentrated only in the last factor of about four, in attempting to get short, effective duration symbols.

A. 3 Time Synchronous Blocks, Phase Coherent Reception with Orthogonal Transmission

We can attack the study of orthogonal transmissions in two ways: first by demonstrating a possible transmission method very similar to that of Section A. 2, and second, by establishing a definite proof of maximum signalling rate by using the vector space representation. It is obvious that if one considers a form of binary pulse position modulation

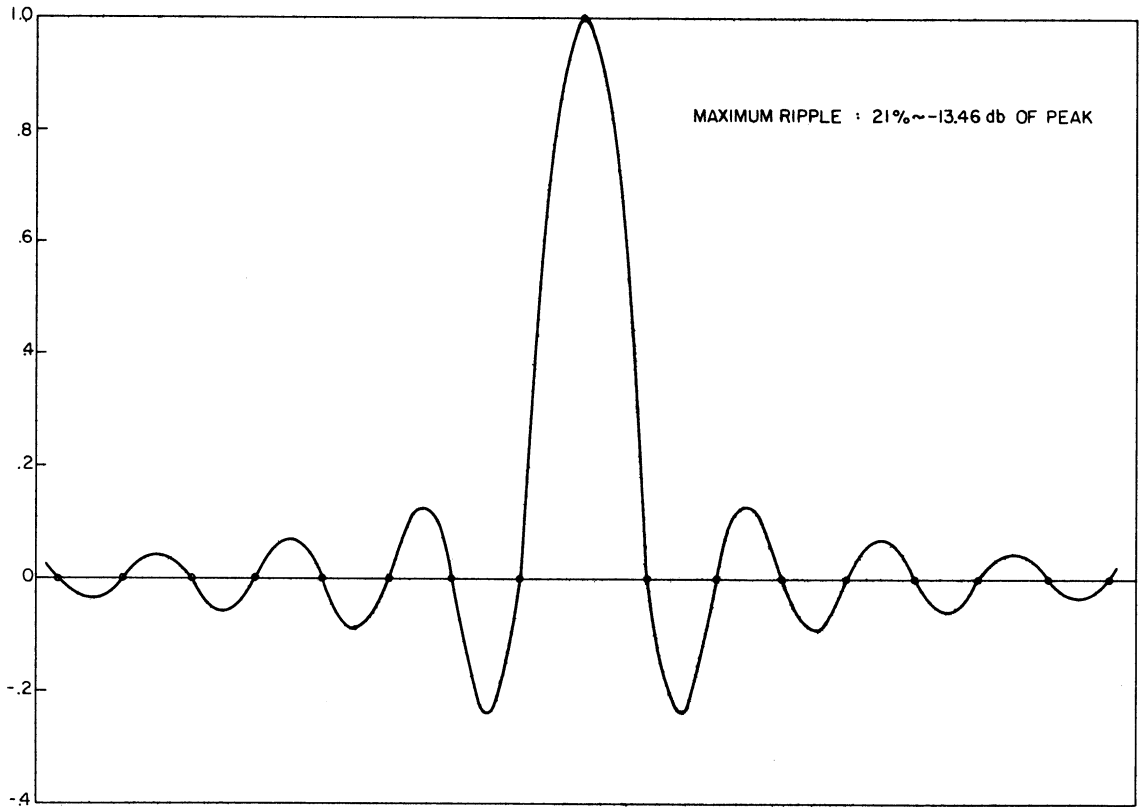


Fig. 34. Portion of $\frac{\sin x}{x}$ curve suitable for maximum rate transmission.

wherein the time axis is considered in pairs of points, all points separated $1/2W$ seconds apart and each pair therefore occupying $1/W$ seconds, the orthogonal binary encodement can be accomplished by sending a $(\sin x)/x$ pulse centered either at the first time of the pair or at the second time of the pair. These two possible waveforms are orthogonal to each other, as are all $(\sin x)/x$ waveforms separated by integral multiples of $1/\text{twice}$ their bandwidth. Thus it is seen that it is possible to transmit orthogonal signals at a rate of W digits per second.

In order to see that this is the maximum rate possible, we appeal to the vector space representation. Here a long duration, T , of transmission is considered, which can be effectively represented in a $2WT$ vector space. There are exactly $2WT$ vectors in every orthogonal complete set, and if we divide this space into orthogonal subspaces and require that each subspace contain two orthogonal vectors, then there will be at most $2WT$ divided by 2, or WT subspaces. Therefore the maximum rate of signalling corresponds to the number of subspaces possible over a long time, divided by that time: W subspaces per second. This

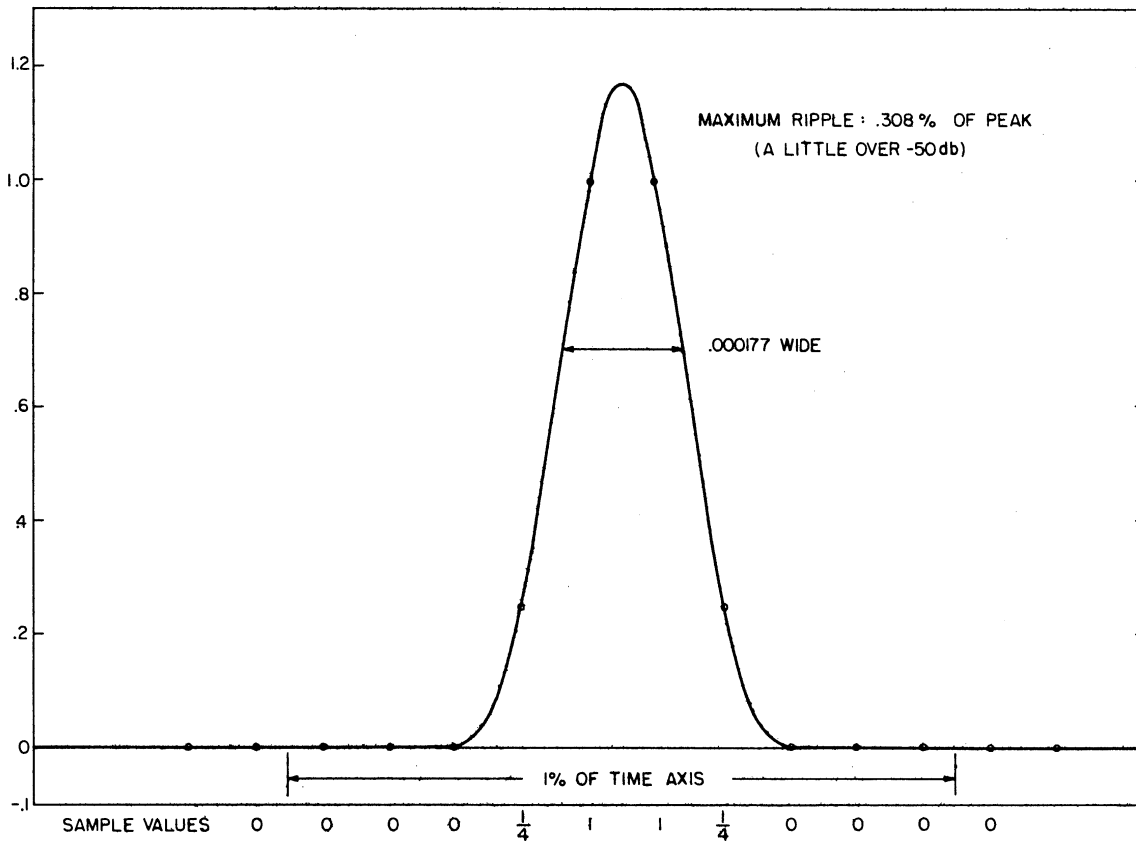


Fig. 35. More practical curve ($2TW=4$) for maximum rate transmission.

means a maximum of W binary orthogonal signals per second. Although the above reasoning may appear quite geometric or pictorial, it is based on the rigorous logic of the requirement that the total transmitted waveform be composed so that the demodulation of a given symbol is not unaffected by the information bit or modulation of previous and subsequent symbols. In the vector space representation this corresponds to a requirement that every possible component in one subspace be orthogonal to every other possible component in the other subspaces.

A. 4 Time Synchronous Blocks, RF Phase Incoherent Reception with Orthogonal Transmissions

In dealing with RF incoherent demodulation one does not have the freedom to draw a direct analogy between an RF waveform and a low-pass waveform, as was done in the previous two sections. Therefore this discussion will be completely limited to the vector space representation of the transmission and demodulation process. With an RF coherent demodu-

lation one can represent each single waveform, and in particular each pure sinusoidal waveform, as occupying a one-dimensional subspace, that is, a pure single vector in the space. If one considers incoherent demodulation, then the demodulator will be affected similarly by the reception of any phase of a pure sine wave component, and indeed can be shown to be a function of the "radius" in a two-dimensional subspace. This is customarily shown in studies of narrowband transmissions and noise (where "narrowband" means that the total bandwidth of the transmission is small compared to the center bandwidth, a condition always met in conventional communications). If the part of the waveform carrying one bit of information is decomposed into the two phase coherent components, the first coherent with a given phase of center frequency and the second at an orthogonal phase of center frequency as shown in Eq. 100,

$$x(t) = x_1(t) \cos \omega t + x_2(t) \sin \omega t \quad (100)$$

then one can transform to variables similar to polar co-ordinates by the use of the "cosine of two variables" formula, to obtain

$$x(t) = \sqrt{x_1^2(t) + x_2^2(t)} \cos[\omega t + \phi(t)] \quad (101)$$

This is often simply written as

$$x(t) = R(t) \cos[\omega t + \phi(t)] \quad (102)$$

where the radius R , or envelope R , is the square root of the sum of the squares of the two phase coherent components. A time synchronous, but RF phase incoherent, demodulator is one which is sensitive to the value of R and insensitive to the individual phase components x_1 and x_2 . That is, it is insensitive to the phase angle ϕ . If one wants to signal using this type of reception, one has to choose two waveforms which will be distinguishable at the receiver. The usual means of accomplishing this is to choose two different frequencies. In the vector space this is represented by assigning one two-dimensional hyperplane to one of the signals and a second and orthogonal hyperplane to the other signal. Thus each pair of binary signals consumes four dimensions, and the total number of such signalling subspaces that can be utilized in a long transmission of duration of T seconds is $2WT$ divided by four, yielding an average rate of $W/2$ signalling subspaces per second.

A. 5 Summary

This section has attempted to demonstrate limitations on the maximum signalling rate which can be used for binary transmissions where (1) the channel assignment is strict; (2) the communicator has no control over adjacent transmissions; and (3) the signalling waveforms need not be individually limited to short times, but merely concentrated in time, and cause no intersymbol interference because of the lack of limitation in time. The results are a direct consequence of the strict band limitation, as is pointed out in the Shannon Sampling Theorem. The method of deriving the results has been to consider the average rate of formation of allowable independent orthogonal subspaces over long durations of transmission. The ultimate limitations that are derived are in terms of the familiar $2WT$ product and, for the three cases, the $2WT$ values take on the minimum values of 1, 2, and 4. In actual practice one will find that the difficulty of obtaining improvement will increase steeply as one gets within a factor of about 4 of these limiting values. In these cases very stringent specifications must be applied to the precision of the time synchronization and the precision in formation of the actual signalling waveforms. Thus these limitations are seen to be valid: although they are physically approachable, they may not be viewed as practical goals by many users.

APPENDIX B: THE EQUIVALENT SOURCE

(R. A. Carlsen)

INTRODUCTION

In an asynchronous time-multiplexed communication system, N sources feed samples into a clocked buffer (Fig. 36). The purpose of this appendix is to show the relationship between the individual sources and the equivalent source indicated in the figure. It turns out that the equivalent source is Bernoulli (i. e., that the number of samples "generated" in an interval τ is Poisson-distributed) when a large number of independent sources is assumed.

In the first section, we define a model for the single source. Then, on the basis of this model, the probability of exactly k samples in an interval τ is determined for the single source. This result is then used to determine, in Section B. 2, the probability of k samples in the interval τ , given a large number of independent single sources which becomes, in the limit, the Bernoulli equivalent source. The discussion is summarized, in Section B. 3, by considering an alternative approach which leads to the same result. That is, the N sources are modeled as trials in a Bernoulli experiment.

The approach presented in the first two sections is instructive for at least two reasons. First, the end result is obtained by progressing from the single source whose characteristics are known to the equivalent source; second, some insight is gained into the problems encountered when N , the number of sources, is restricted to small values.

B. 1 The Single Source

The distribution of the intervals between extremes has been experimentally determined by Mathews.⁴ However, this distribution alone is not an adequate description of the statistics of the speech waveform. For example, we have no knowledge of the correlation between successive intervals. It is reasonable, therefore, to fill in the missing data by assumption (the procedure which is always followed in the construction of an analytical model).

⁴See Ref. 28 and Fig. 37.

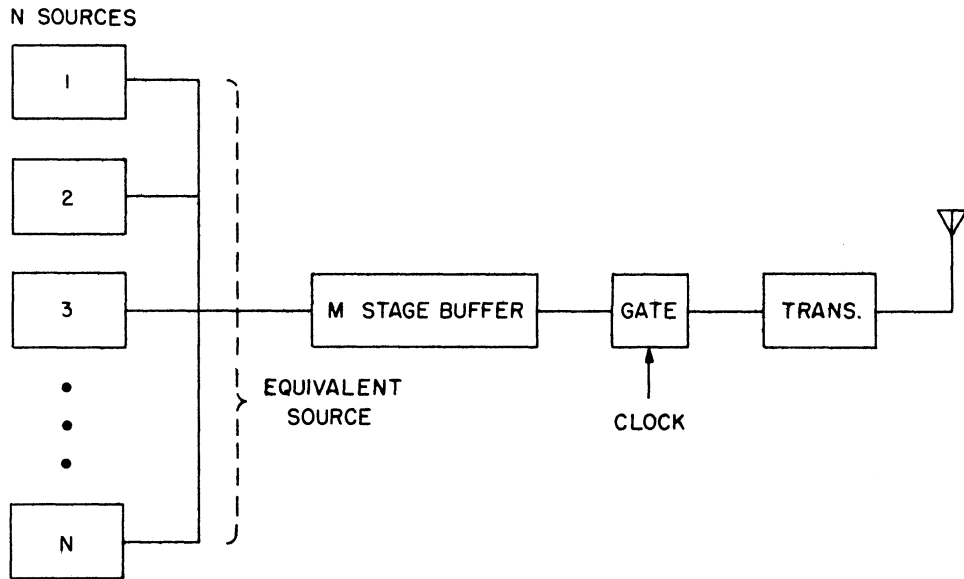


Fig. 36. Buffering in an asynchronous multiplexed communication system.

The implication is that the validity of the model will be verified by experiment.

Perhaps one might question the wisdom of defining a model for the speech process. At least a partial answer to this objection is the advantage which can be gained by knowing the precise characterization of the process which has been substituted for the original source. For example, the speech source is modeled as a stationary, simple Markoff process.

As a first step in defining a model for the speech process, stationarity is assumed. The individual speech source is obviously nonstationary, as the speaker is silent approximately half the time. Perhaps we can think of a single source as being equivalent to two system users, one occupying the silent periods of the other. Second, the distribution of the intervals between extremes will be assumed to be of the form

$$f(u) = \lambda u e^{-\lambda u} \quad (103)$$

where: λ = the average sample generation rate for a single source

u = the mean interval between extremes.

This function is a reasonable approximation of the experiment data and is, in addition, mathematically convenient. This function, Eq.103, is called the Erlang type-1 distribution and is obtained by the convolution of two exponential density functions. An exponential source means

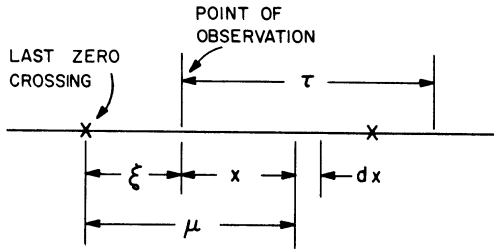


Fig. 37. Notation for $f(x/\xi)$.

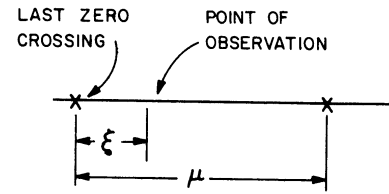


Fig. 38. Notation for $f(\xi)$.

simply that the intervals between samples are exponentially distributed. For example, an Erlang type-1 source can be obtained by removing every other sample from an exponential source.

It is clear that the extremes of the speech waveform are the zero-crossings of the derivative of the original waveform. Therefore the distribution for the intervals between the extremes will be referred to as the "zero-crossing distribution" with the understanding that the derivative of the speech signal is being discussed.

In addition to stationarity, it is assumed that the intervals between zero-crossing are independent. Assuming independence between intervals and the zero-crossing distribution (Eq. 103) suggests that one might model the speech source as a simple Markoff process.⁵ Although this additional assumption is not utilized in this appendix, it does establish a firm basis for additional analytical work. For example, Faureau⁶ has shown experimentally that the output of a three-pole low-pass filter with a white Gaussian input has a zero-crossing distribution of the form assumed here for the speech process.⁷ Thus, once the zero-crossing distribution for bandlimited Gaussian noise is solved, an analytical solution based upon a Markoff model would be an approximate solution for the current problem and an exact solution for a modified application of the system (e. g. , the transmission of pseudo-noise coded signals).

⁵Middleton, Ref. 39, pp. 45-48.

⁶Ref. 40, and Fig. 38.

⁷In fact, McFadden (Ref. 41) has shown that the intervals between zero-crossings are nearly independent for the three-pole low-pass filter.

When the simple Markoff model and an Erlang type-1 zero-crossing distribution have been specified, the next step is to determine the statistical properties of the source. If the interval between zero-crossings, u , is normalized with respect to the mean interval between zeros ($1/\lambda$), the assumed zero-crossing density function becomes

$$f(u) = u e^{-u} \quad (104)$$

where: u = the normalized interval between zeros.

Given a point of observation (ξ) relative to the last zero-crossing, the probability of the next zero-crossing in the range x , $x + dx$ (where x is the elapsed time since the point of observation) is given by

$$f_1(x/\xi) = \frac{f(u)}{\int_{\xi}^{\infty} f(u) du} \quad (105)$$

where: $f_1(x/\xi)$ = the probability density of a zero-crossing, given ξ since the last zero (see Fig. 37).

This is a conditional zero-crossing density which, for the zero-crossing density assumed in Eq. 104, is given by

$$f_1(x/\xi) = \frac{\xi + x}{\xi + 1} e^{-x} \quad (106)$$

One can easily show that for an exponential source, $f(u) = e^{-u}$, the above conditional density function, $f_1(x/\xi)$, is still exponential. That is to say, for an exponential source the distribution of the intervals between an arbitrary point of observation and the next zero-crossing is equal to distribution of the intervals between consecutive zero-crossings. That is equivalent to saying that probability of a zero-crossing from an exponential source in an incremental interval, du , is equally likely and independent from interval to interval. Since the zero-crossing density function is independent of the point of observation for the exponential density function, the probability of k zeros in an interval τ is Poisson-distributed (i. e., the single source is Bernoulli). Furthermore, N independent exponential sources in parallel are equivalent to a single exponential source with a generation rate N times that for a single source. As a result, an equivalent Bernoulli source is obtained when a finite number of independent exponential sources is assumed. However, the assumption that the zero-crossing density function

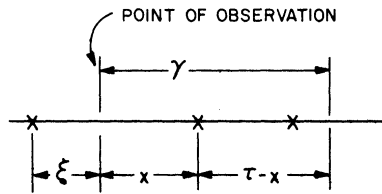


Fig. 39. Notation for $P'(k, \tau/\xi)$.

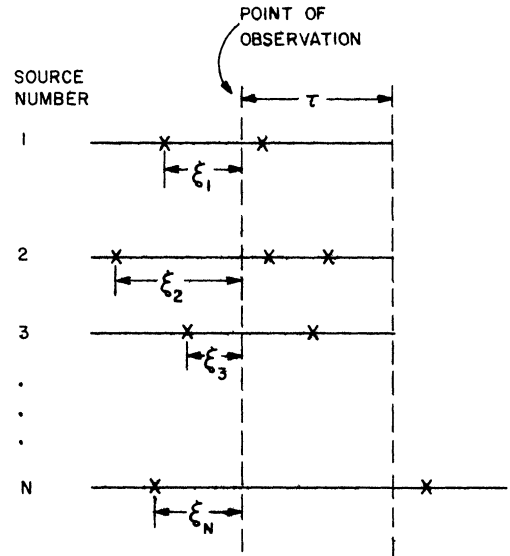


Fig. 40. Notation for $P(k, \tau/\xi_1, \xi_2 \dots \xi_N)$.

is exponential is a poor approximation of the experimental data.⁸

The objective is to obtain the probability, $P'(k, \tau/\xi)$, of k zero-crossings in an arbitrary interval τ , given ξ . It turns out, as indicated in Section B. 2, that the expected value of $P'(k, u/\xi)$ is of interest. However, it is first necessary to determine the probability density function, $f(\xi)$, for the intervals ξ . Suppose (as in Fig. 38) that the interval between two zeros is known to be u . The point of observation will always fall between two such zero-crossings. Now, the point of observation, ξ , is selected at random (i. e., each interval $d\xi$ between the indicated zeros is equally likely); the probability of starting the observation between ξ and $\xi + d\xi$, given u , is

$$f(\xi/u) d\xi = \frac{1}{u} d\xi \quad (107)$$

as the point of observation is uniformly distributed between 0 and u . However, the interval u is a random variable having the density function given by Eq. 104. Therefore,

$$f(\xi)d\xi = \int_{\xi}^{\infty} f(\xi/u) f(u) du d\xi \quad (108)$$

$$f(\xi) = e^{-\xi}$$

Now, one finds $P'(k, \tau/\xi)$ by finding the probability of k zeros in the interval τ .

⁸See Fig. 39 and Ref. 28.

This is determined by knowing the probability density for the intervals between every second, third, . . . zero-crossing. Since the intervals between zeros are independent, the desired density functions are obtained by convolution.

$$\begin{aligned}
 f_2(x/\xi) &= f_1(x/\xi) * f_1(x/0) \\
 f_3(x/\xi) &= f_1(x/\xi) * f_2(x/0) \\
 &\cdot \\
 &\cdot \\
 &\cdot \\
 f_n(x/\xi) &= f_1(x/\xi) * f_{n-1}(x/0)
 \end{aligned}
 \tag{108a}$$

where: $f_n(x/\xi)dx$ = the probability of the nth zero-crossing in the interval $x, x+dx$, given ξ since the last zero

$f_n(x/0)dx$ = the probability of nth zero-crossing in the interval $x, x+dx$, given an initial zero-crossing.

The probability of k zero-crossings or less is obtained by integrating these density functions for x greater than τ .

$$P'(k < n, \tau/\xi) = \int_{\tau}^{\infty} f_n(x/\xi) dx
 \tag{108b}$$

That is, we obtain the probability that the n th zero has not occurred. For $n=1$, in the above equation, we have the probability of no zero in the interval τ . The probability of k or more zeros is therefore

$$P'(k \geq n, \tau/\xi) = \int_0^{\tau} f_n(x/\xi) dx
 \tag{108c}$$

Thus, the probability of exactly $k=n$ zeros in the interval is given by

$$P'(k=n, \tau/\xi) = P'(k \geq n+1, \tau/\xi) - P'(k \geq n, \tau/\xi)
 \tag{108d}$$

These results agree with McFadden's equivalent expression for $P(k=n, \tau/0)$.⁹ These computations can be simplified by taking advantage of the following relationship:

⁹J. McFadden, Ref. 41, Appendix I.

$$P'(k, \tau/\xi) = \int_0^{\tau} f(x/\xi) dx P'(k-1, \tau-x/0)$$

$$P'(k, \tau/\xi) = \int_0^{\tau} f(\tau-x/\xi) dx P'(k-1, x/0) \quad (109)$$

The prime identifies the result with a single source.

The probability of no zeros in the interval τ , given ξ , is

$$P'(0, \tau/\xi) = \int_{\tau}^{\infty} \frac{\xi + x}{\xi + 1} e^{-x} dx \quad (\text{see Eq. 106})$$

$$P'(0, \tau/\xi) = \frac{\xi + \tau + 1}{\xi + 1} e^{-\tau} \quad (110)$$

$P'(k, \tau/0)$ is of particular interest as the point of observation is always at a zero-crossing.

One can show, by convolution, that

$$P'(k, \tau/0) = e^{-\tau} \left[\frac{\tau^{2k}}{2k!} + \frac{\tau^{2k+1}}{(2k+1)!} \right] \quad (111)$$

The expected value of $P'(k, u/\xi)$ is

$$P'(k, u) = \int_0^{\infty} P'(k, u/\xi) f(\xi) d\xi \quad (112)$$

$$\text{or } P'(k, u) = \int_0^{\tau} \int_0^{\infty} f(\tau-x/\xi) f(\xi) d\xi P'(k-1, x/0) dx \quad (113)$$

$P'(k, u)$ is the probability obtained by measuring across an ensemble of sources or from a long observation of a single source (an ergodic source is assumed). The computations are summarized by:

$$P'(0, \tau) = e^{-\tau} [1 + \gamma\tau]$$

$$P'(1, \tau) = e^{-\tau} \left[\frac{\tau^1}{1!} + \frac{\tau^2}{2!} \right] + \gamma e^{-\tau} \left[-\tau + \frac{\tau^3}{3!} \right]$$

$$P'(2, \tau) = e^{-\tau} \left[\frac{\tau^2}{3!} + \frac{\tau^4}{4!} \right] + \gamma e^{-\tau} \left[\frac{-\tau^3}{3!} + \frac{\tau^5}{5!} \right]$$

$$P'(k, \tau) = e^{-\tau} \left[\frac{\tau^{2k-1}}{(2k-1)!} + \frac{\tau^{2k}}{2k!} \right] + \gamma e^{-\tau} \left[\frac{-\tau^{2k-1}}{(2k-1)!} + \frac{\tau^{2k+1}}{(2k+1)!} \right] \quad (114)$$

$$\text{where: } \gamma = \int_0^{\infty} \frac{e^{-\xi}}{\xi+1} d\xi \cong .62$$

The above equations may be written as

$$\sum_{k=0}^{\infty} P'(k, \tau) = 1$$

$$P'(k, \tau) \geq 0 \quad \text{all } k \text{ and } \tau$$

$$\lim_{\tau \rightarrow \infty} P'(k, \tau) = 0 \quad \text{all } k$$

$$\lim_{\tau \rightarrow 0} P'(k, \tau) = 1 \quad k = 0$$

$$\lim_{\tau \rightarrow 0} P'(k, \tau) = 0 \quad k \neq 0$$

The expected number of zero-crossings, k , in an interval τ , is

$$u' = \sum_{k=0}^{\infty} k P'(k, \tau)$$

$$u' = e^{-\tau} \left[\tau + \frac{\tau^2}{2!} + 2 \left(\frac{\tau^3}{3!} + \frac{\tau^4}{4!} \right) + 3 \left(\frac{\tau^5}{5!} + \frac{\tau^6}{6!} \right) + \dots \right]$$

$$- \gamma e^{-\tau} \left[\tau + \frac{\tau^3}{3!} + \frac{\tau^5}{5!} + \dots \right]$$

which for small τ becomes

$$u' \cong (1 - \gamma)\tau \tag{115}$$

All the above results have been obtained using the normalized zero-crossing distribution.

Therefore, τ is replaced by $\lambda \tau$ in all the above equations for the actual source having a generation rate of λ samples per unit of time.

B.2 The Multiple Source

Since it is the multiple source which is of interest, consider the N identical (but independent) sources shown in Fig. 40. We want to determine the probability of k samples from the equivalent source in the interval τ , given the intervals $\xi_1, \xi_2, \dots, \xi_n$. The point of observation is just prior to a clock pulse, and the interval τ ends just prior to the next clock pulse. The probability of having no zero-crossings from the equivalent source during

the interval τ is

$$P(0, \tau/\xi_1, \xi_2 \dots \xi_n) = P'(0, \tau/\xi_1) P'(0, \tau/\xi_2) \dots P'(0, \tau/\xi_N) \quad (116)$$

In the same way, the probability of just one zero-crossing from the equivalent source is

$$\begin{aligned} P(1, \tau/\xi_1, \xi_2 \dots \xi_n) &= P'(1, \tau/\xi_1) P'(0, \tau/\xi_2) \dots P'(0, \tau/\xi_N) \\ &+ P'(0, \tau/\xi_1) P'(1, \tau/\xi_2) \dots P'(0, \tau/\xi_N) \\ &+ \dots \\ &+ P'(0, \tau/\xi_1) P'(0, \tau/\xi_2) \dots P'(1, \tau/\xi_N) \end{aligned} \quad (117)$$

For larger values of k similar expressions are obtained by considering all the possible combinations of N channel outputs which will produce k zero-crossings.

If the number of sources, N , is large enough to justify an ensemble average for $P(k, \tau/\xi_1, \xi_2, \dots, \xi_N)$, one can take advantage of the results of the previous section, Eq. 114. For example, the expected value of $P(0, \tau/\xi_1, \xi_2, \dots, \xi_N)$ is

$$\begin{aligned} P(0, \tau) &= \int_0^\infty \int_0^\infty \dots \int_0^\infty P'(0, \tau/\xi_1) P'(0, \tau/\xi_2) \dots P'(0, \tau/\xi_N) \\ &f(\xi_1, \xi_2, \dots, \xi_N) d\xi_1 d\xi_2 \dots d\xi_N \end{aligned} \quad (118)$$

Since the sources are assumed to be independent, the joint distribution for the intervals $\xi_1, \xi_2, \dots, \xi_N$ can be written as

$$f(\xi_1, \xi_2, \dots, \xi_N) = f(\xi_1) f(\xi_2) \dots f(\xi_N) \quad (119)$$

and, as a result

$$P(0, \tau) = \int_0^\infty P'(0, \tau/\xi_1) f(\xi_1) d\xi_1 \int_0^\infty \dots \int_0^\infty P'(0, \tau/\xi_n) f(\xi_n) d\xi_n$$

or

$$P'(0, \tau) = C_0^N [P'(0, \tau)]^N \quad (\text{see Eq. 114}) \quad (120)$$

In the same way, the probability of $k=1, 2, 3, 4, \dots$ can be shown to be given by

$$P(2, \tau) = C_1^N P'(2, \tau) P'(0, \tau)^{N-1} + C_2^N P'(1, \tau) P'(0, \tau)^{N-2}$$

$$P(3, \tau) = C_1^N P'(3, \tau) P'(0, \tau)^{N-1} + C_2^N P'(2, \tau) P'(1, \tau) P'(0, \tau)^{N-2} \\ + C_3^N P'(1, \tau)^3 P'(0, \tau)^{N-3}$$

$$P(4, \tau) = C_1^N P'(4, \tau) P'(0, \tau)^{N-1} + C_2^N [P'(3, \tau) P'(1, \tau) \\ + P'(2, \tau)^2] P'(0, \tau)^{N-2} + C_3^N P'(2, \tau) P'(1, \tau)^3 P'(0, \tau)^{N-3}$$

$$+ C_4^N P'(1, \tau)^4 P'(0, \tau)^{N-4}$$

That is to say, one can obtain $P'(k, \tau)$ for a single source and use this result and a knowledge of the combinations of sources which will produce a total of k zeros in a clock interval τ .

The taking of the ensemble average (i. e. , a large number of sources) is the assumption which must be avoided in order to study the advantages of asynchronous multiplexing for a small number of sources.

It is instructive to consider the probability of k samples in the interval τ as the number of sources, N , approaches infinity. Since the N sources are independent, the expected number of zero-crossings per clock interval from the equivalent source is

$$u = N u'$$

$$u = N(1 - \gamma)\lambda\tau \quad (\text{see Eq. 115})$$

Therefore,

$$\lambda\tau = \frac{u}{N(1 - \gamma)} \quad (121)$$

When this substitution is made in Eq. 120, we have

$$P(0, u) = C_0^N \left[\exp \left\{ -\frac{u}{N(1-\gamma)} \right\} \left(1 + \frac{\gamma u}{N(1-\gamma)} \right) \right]^N$$

It is easily shown that

$$\lim_{N \rightarrow \infty} P(0, u) = e^{-u} \quad (122)$$

Likewise,

$$\lim_{N \rightarrow \infty} P(1, u) = u e^{-u}$$

$$\lim_{N \rightarrow \infty} P(2, u) = \frac{u^2}{2!} e^{-u}$$

⋮

$$\lim_{N \rightarrow \infty} P(k, u) = \frac{u^k}{k!} e^{-u}$$

That is, the number of samples generated by the equivalent source is Poisson-distributed.

Thus, the equivalent source is Bernoulli when a large number of independent, simple Markoff, Erlang type-1 sources is assumed.

B. 3 The Bernoulli Experiment

The same result can be obtained by considering the transmitter as performing a Bernoulli experiment. In essence, the transmitter is looking at N channels in the interval τ which, for large N , is small enough that the probability of more than one sample from a single source is negligible. That is, a look at each source during an interval τ constitutes a Bernoulli trial (i. e., only a binary outcome is possible). If, in addition, it is assumed that the N sources are independent and that probability of a sample is constant from source to source in each interval τ (i. e., that the expected number of samples per clock period is constant), the assumptions underlying the Poisson approximation for a series of Bernoulli trials are satisfied. Thus, an equivalent Bernoulli source is obtained when (1) the assumptions governing a Bernoulli experiment are satisfied (from interval to interval), (2) a finite number of independent exponential sources is combined in parallel, or (3) the assumptions and model of Section B. 2 are applicable.

APPENDIX C: ANALYSIS OF FREEZE OUT RUNS

(T. G. Birdsall)

Summary of Freeze Out Runs

Equations and definitions:

$P(c)$ \equiv probability of a run of exactly c clock intervals containing
some (total) source inputs frozen out

$$P(c) = (1 + u) e^{-u} [1 - e^{-u}(1 + u)]^{c-1} \quad c \geq 1$$

$$E(c) = \frac{1}{P(c=1)}$$

$P(d)$ = probability of d source inputs being frozen in a clock inter-
val, during the second and third, etc. clock interval of a
run of freeze outs

$$P(d) = p(d+1, u) = e^{-u} \frac{u^{d+1}}{(d+1)!} \quad d \geq 1$$

Numbers for $u = .8$

<u>Runs of Length c</u>	<u>Blocks of Size d in a Given Interval</u>
$P(1) = 80.9 \%$	$P(1) = 80.88 \%$
$P(2) = 15.4 \%$	$P(2) = 14.38 \%$
$P(3) = 3.0 \%$	$P(3) = 3.53 \%$
$P(4) = 0.6 \%$	$P(4) = .71 \%$
$P(5) = 0.1 \%$	$P(5) = .50 \%$
$E(c) = 1.24$	

Note: These conditional equations are independent of M , the buffer size.

Detail

An attempt on the problem of freeze out follows (assume N large):

I) The first step is to consider a different problem--that of the distribution of consecutive freeze outs from all sources combined--and then to tackle the problem of a single source (voice channel) and its individual freeze out.

We know the distribution of "blocks" of freeze outs, that is, the number of inputs dropped between successive clock pulses. This is

$$\text{Prob}(0 \text{ digits dropped} \mid \text{some dropped}) = \frac{\beta_{M+d}}{\sum_{i=1}^{\infty} \beta_{M+i}}$$

Since there is only a finite number of nonzero α_k terms, we can write each β_{M+d} as:

$$\beta_{M+1} = e^{-u} \left[\alpha_{M-1} \frac{u^2}{2!} + \alpha_{M-2} \frac{u^3}{3!} + \dots + \alpha_1 \frac{u^M}{M!} + \alpha_0 \frac{u^{M+1}}{(M+1)!} \right]$$

$$\beta_{M+2} = e^{-u} \left[\alpha_{M-1} \frac{u^3}{3!} + \alpha_{M-2} \frac{u^4}{4!} + \dots + \alpha_0 \frac{u^{M+2}}{(M+2)!} \right]$$

$$\beta_{M+d} = e^{-u} \left[\alpha_{M-1} \frac{u^{d+1}}{(d+1)!} + \alpha_{M-2} \frac{u^{d+2}}{(d+2)!} + \dots + \alpha_0 \frac{u^{M+d}}{(M+d)!} \right]$$

$$= e^{-u} \sum_{i=1}^M \alpha_{M-i} \frac{u^{d+i}}{(d+i)!}$$

$$= e^{-u} \frac{u^d}{d!} \sum_{i=1}^M \alpha_{M-i} \frac{d!}{(d+i)!} u^i$$

$$= p(d, u) \sum_{i=1}^M \alpha_{M-i} \frac{u^i}{(d+1)(d+2)\dots(d+i)}$$

Note: β_{M+d} is strictly monotone decreasing in d since each coefficient of α_k is progressively smaller for $u < 3$ (so it certainly is for $u \leq 1$). This makes the last equation look like a poor choice; thus the following is preferable:

$$\beta_{M+d} = u^d e^{-u} \sum_{i=1}^M \alpha_{M-i} \frac{u^i}{(d+i)!} = \frac{u^d e^{-u}}{\Gamma(M)} \sum_{i=1}^M \gamma_{M-i} \frac{u^i}{(d+1)!}$$

These can be readily calculated for $u = .8$, for which we already know α_k and $\Gamma(M)$. Thus

$P(d \text{ digits dropped at once} \mid \text{there is freeze out})$

$$= \frac{u^d e^{-u} \sum_{i=1}^M \gamma_{M-i} \frac{u^i}{(d+i)!}}{\sum_{g=1}^{\infty} u^g e^{-u} \sum_{i=1}^M \gamma_{M-i} \frac{u^i}{(i+j)!}} = \frac{u^d \sum_{i=1}^M \gamma_{M-i} \frac{u^i}{(d+i)!}}{\sum_{i=1}^{\infty} \sum_{j=1}^{\infty} u^{i+j} \frac{1}{(i+j)!} \gamma_{M-i}}$$

A slightly different method is to calculate the probability of d_2 simultaneous freeze outs, assuming at least one freeze out in the previous τ -clock interval. In this case we know that there is one buffer cell empty at the beginning of the input interval (that is, $\theta = 0^+$)

$$\begin{aligned}
 p(d_2 \mid \text{previous freeze out}) &= p(d_2 + 1, u) \\
 &= e^{-u} \frac{u^{d_2+1}}{(d_2+1)!},
 \end{aligned}$$

while the probability of no subsequent freeze out is the probability of 0 or 1 arrival:

$$p(\text{no FO} \mid \text{previous interval FO}) = e^{-u}(1 + u)$$

These are independent of buffer size.

For example, for $u = .80$, we have the following:

$$\begin{aligned}
 p(\text{no further FO} \mid \text{FO in previous interval}) &= 80.88 \% \\
 p(1 \text{ FO} \mid \text{FO in previous interval}) &= 14.38 \% \\
 p(2 \text{ FO in block} \mid \text{FO in previous interval}) &= 3.53 \% \\
 p(3 \text{ FO in block} \mid \text{FO in previous interval}) &= .71 \% \\
 p(\geq 4 \text{ FO in block} \mid \text{FO in previous interval}) &= .50 \%
 \end{aligned}$$

This implies that there will be definite runs of freeze out once single freeze out occurs. If we further manipulate this, we get

$$\begin{aligned}
 &P(1 \text{ interval contains FO's} \mid \text{previous interval contained FO's}) \\
 &= 1 - e^{-u}(1 + u) \\
 &= e^{-u}(e^u - 1 - u) \\
 &= e^{-u} \gamma_1
 \end{aligned}$$

Iterating this, we get

$$\begin{aligned}
 &P(c \text{ subsequent intervals all contain FO's} \mid \text{initial interval} \\
 &\text{contains FO's}) = (e^{-u} \gamma_1)^c = e^{-cu} \gamma_1^c
 \end{aligned}$$

This is a geometric progression. Summarizing formally:

<u>Def.</u>	A freeze out run of length c occurs when there are source inputs frozen out in exactly c consecutive clocking intervals.
<u>Corr.</u>	During a freeze out of run length c , only the first source input arriving after a clock pulse will be accepted by the buffer.

Equations:

$$P(c = 1) = e^{-u}(1+u)$$

$$P(c \geq 2) = 1 - e^{-u}(1+u) = e^{-u}(e^u - u - 1)$$

$$P(c \geq 3) = e^{-2u}(e^u - u - 1)^2$$

$$P(c \geq c^*) = e^{-(c^*-1)u}(e^u - u - 1)^{(c^*-1)}$$

Therefore

$$\begin{aligned} P(\text{exactly } c) &= P(\geq c) - P(\geq c+1) \\ &= [e^{-u}(e^u - u - 1)]^{c-1} - [e^{-u}(e^u - u - 1)]^c \\ &= [e^{-u}(e^u - u - 1)]^{c-1} [1 - e^{-u}(e^u - u - 1)] \\ &= [e^{-u}(e^u - u - 1)]^{c-1} [(u+1)e^{-u}] \end{aligned}$$

$P(\text{freeze out run of length } c) = (u+1)e^{-u}[1 - e^{-u}(1+u)]^{c-1}$
--

Detailing:

$$P(c = 1) = e^{-u}(1+u)$$

$$P(c = 2) = e^{-u}(1+u) [e^{-u}(e^u - u - 1)]$$

$$P(c = 3) = e^{-u}(1+u) [e^{-u}(e^u - u - 1)]^2$$

To check: These should add to 1.0000.

$$\begin{aligned} \sum_{c=1}^{\infty} P(c) &= e^u(1+u) \sum_{i=0}^{\infty} [e^{-u}(e^u - u - 1)]^i = \frac{e^u(1+u)}{1 - [e^{-u}(e^u - u - 1)]} \\ &= \frac{e^u(1+u)}{1 - [1 - e^{-u}(u+1)]} = 1 \end{aligned}$$

The average value of a geometric probability density is $\frac{1}{1 - \text{ratio}}$, so

$$E(c) = \frac{1}{1 - [e^{-u}(e^u - u - 1)]} = \frac{1}{e^{-u}(u+1)} = \frac{1}{P(c=1)}$$

$E(c) = \frac{1}{P(c=1)}$

Example: $u = .8$

$$P(c=1) = .8088$$

$$E(c) = 1.24$$

$P(c=1) = .8088$	80.9 %
$P(c=2) = .1542$ (slide rule)	15.4 %
$P(c=3) = .0285$	3.0 %
$P(c=4) = .0056$.6 %
$P(c=5) = .0011$.1 %
$P(c=6) = .0002$	--

It is difficult to say directly how many inputs are frozen out during the initial interval of an FO run, but we have established above that the probability distribution in the remaining intervals is

$$P(d \text{ FO's}) = p(d+1, u).$$

LIST OF REFERENCES

1. R. W. Sanders, "Digilock Telemetry System," National Symposium on Space Electronics and Telemetry, Section 6.3, September 28-30, 1959, San Francisco, California, pp. 1-10.
2. Private communication and "Low-Density Parity Check Codes," Gallager, R. G., PGIT, Vol. I, T-8, No. 1, January 1962
3. G. L. Turin, "An Introduction to Matched Filters," PGIT, June 1960, pp. 311-329.
4. Davenport and Root, Random Signals and Noise, McGraw-Hill, 1958, p. 156.
5. S. O. Rice, "Mathematical Analysis of Random Noise," BSTJ, Vol. 23, July 1944, pp. 282-332.
6. R. M. Fano, Transmission of Information, MIT Press and John Wiley, 1961.
7. A. Viterbi, "Classification and Evaluation of Coherent Synchronous Sample Data Telemetry Systems," 1961 Western Electronic Show and Convention, Session 14-3, August 22-25, 1961.
8. W. W. Peterson, Error Correcting Codes, MIT Press and John Wiley, 1961.
9. J. J. Metzner and K. C. Morgan, "Coded Feedback Communication Systems," 1960 National Electronics Conference, pp. 250-256 and bibliography.
10. Davenport and Root, Random Signals and Noise, McGraw-Hill, 1958.
11. Siegfried Rieger, "Error Rates in Data Transmission," Letter to the IRE Proc., May 1958, pp. 919-920.
12. A. J. Viterbi, "On Coded Phase Coherent Communications," JPL Technical Report No. 32-25, JPL Laboratory, California Institute of Technology, August 15, 1960.
13. Charles R. Cahn, "Performance of Digital Phase Modulation Communications Systems," PGCS, May 1959, pp. 4-6.
14. T. G. Birdsall and M. P. Ristenbatt, "Introduction to Linear Shift-Register Generated Sequences," Cooley Electronics Laboratory Technical Report No. 90, The University of Michigan, Ann Arbor, Michigan, October 1958.
15. I. S. Reed, "A Class of Multiple Error Correcting Codes and the Decoding Scheme," PGIT-4, September 1954, p. 38.
16. V. A. Kotel'nikov, Translation by R. A. Silverman, The Theory of Optimum Noise Immunity, McGraw-Hill Book Co., 1959.
17. R. M. Lerner, "Modulation and Signal Selection for Digital Data Systems," Paper in National Electronics Conference, 1960, pp. 2-14.
18. G. L. Turin, "The Asymptotic Behavior of Ideal M-ary Systems," Letter to the Editor of IRE, January 1959, p. 93.

LIST OF REFERENCES (Cont.)

19. J. M. Brumbaugh, L. C. Morris, et al, "Future Astro-Communications Techniques," Technical Report RADC-TN-60-142, August 1960, Applied Research Defense Electronic Products, RCA, ASTIA No. 243698.
20. G. A. Roberts, D. R. Rothschild, R. A. Carlsen, J. L. Daws, Jr., T. G. Birdsall, and M. P. Ristenbatt, "An Introduction to Pseudo-Random Systems, Vol. I: Basic Concepts and Techniques," Cooley Electronics Laboratory Technical Report No. 104-I, The University of Michigan, Ann Arbor, Michigan, May 1960 (S).
21. John G. Lawton, "Investigation of Digital Data Communications Systems," Technical Report RADC-TR-161-58, Cornell Aeronautical Laboratory, January 3, 1961, ASTIA No. 256584.
22. K. H. Powers, Lectures in Communications Modulation and Detection Techniques, University of Pennsylvania Summer Course, 1961 (Notes to be published).
23. R. M. Lerner, Lectures on Communication System Theory, Chapter 11, "Design of Signals," McGraw-Hill Book Co., 1961.
24. A. Norton, E. L. Shultz, and H. Yarbrough, "The Probability Distribution of the Phase of the Resultant Vector Sum of a Constant Vector Plus a Rayleigh Distributed Vector," Journal of Applied Physics, January 1952, pp. 137-141.
25. R. W. Sanders, "Communication Efficiency Comparison of Several Communication Systems," Proc. IRE, April 1960, pp. 575-588.
26. Internal Memo to Cooley Electronics Laboratory Project 4251 File, E. P. Gould, Subject: TASI System.
27. BTL Record 1960 and AIEE, July 1959, Transactions on Communication and Electronics.
28. M. V. Mathews, "Extremal Coding for Speech Transmission," PGIT, Vol. IT-5, September 1959, No. 3, p. 129.
29. L. R. Spogen, et al, "Speech Processing by the Selective Amplitude Sampling System," Journal of the Acoustical Society of America, Vol. 32, No. 12, December 1960, p. 1621.
30. R. Filipowsky, "Ein Mehrkanal-Impulsübertragungsverfahren mit automatisch veränderlicher Impulszahl," Osterreichische Zeitschrift für Telegraphen-Telephon-, Funk- und Fernstechnik, No. 9-10, 11-12 (Sept. -Oct., Nov. -Dec.), 1955.
31. R. J. Filipowsky, "Digital Data Transmission Systems of the Future," PGCS, March 1961, p. 88.
32. Monthly Letter No. 10 on Project 04251, Signal Corps Contract No. DA 36-039 sc-87172.
33. J. L. Yen, "On Nonuniform Sampling of Bandwidth Limited Signals," IRE Trans. on Circuit Theory, December 1956.
34. Bernard Smith, "Instantaneous Companding of Quantized Signals," BSTJ, 36 (1957) 653.
35. R. L. Miller, "The Possible Use of Log Differential PCM for Speech Transmission in Unicom," Globecom-5, Chicago, Illinois, May 22-24, 1961.
36. A. H. Nuttall, "Error Probabilities for Non-Orthogonal M-ary Signals under Phase-Coherent and Phase-Incoherent Reception," Technical Report TR-61-1-BF, Litton Systems, Inc., June 15, 1961.

LIST OF REFERENCES (Cont.)

37. D. M. Green and T. G. Birdsall, "The Effect of Vocabulary Size on Articulation Score," Cooley Electronics Laboratory Technical Report No. 81, The University of Michigan, Ann Arbor, Michigan, January 1958.
38. "The Theory of Signal Detectability," by W. W. Peterson and T. G. Birdsall (U). Part I. "The General Theory," June 1953. Part II. "Applications with Gaussian Noise," July 1953.
39. D. Middleton, An Introduction to Statistical Communication Theory, McGraw-Hill, 1960.
40. R. P. Faureau, "Evaluation of Complex Statistical Functions by an Analog Computer," IRE Convention Record, Part IV, pp. 31-37, 1956.
41. J. A. McFadden, "The Axis Crossing Intervals of Random Functions, II," PGIT, pp. 14-24, March 1958.

DISTRIBUTION LIST

(1 copy unless noted otherwise)

OASD (R&E), Rm 3E1065
ATTN: Technical Library
The Pentagon
Washington 25, D. C.

Chief of Research and Development
OCS, Department of the Army
Washington 25, D. C.

Chief Signal Officer
ATTN: SIGRD-6
Department of the Army
Washington 25, D. C.

Chief Signal Officer
ATTN: SIGOP-5
Department of the Army
Washington 25, D. C.

Chief Signal Officer
ATTN: SIGAC
Department of the Army
Washington 25, D. C.

Chief Signal Officer
ATTN: SIGPL
Department of the Army
Washington 25, D. C.

Director, U. S. Naval Research Laboratory
ATTN: Code 2027
Washington 25, D. C.

Commanding Officer and Director
U. S. Navy Electronics Laboratory
San Diego 52, California

U. S. National Bureau of Standards
Boulder Laboratories
ATTN: Library
Boulder, Colorado

Commander
Aeronautical Systems Division
ATTN: ASAPRL
Wright-Patterson Air Force Base, Ohio

Commander, Air Force Cambridge Research
Laboratories
ATTN: CRO
L. G. Hanscom Field
Bedford, Massachusetts

Commander, Air Force Command and
Control Development Division
ATTN: CCRR & CCSD
L. G. Hanscom Field
Bedford, Mass.

Commander, Rome Air Development
Center
ATTN: RAALD
Griffiss Air Force Base, N. Y.

Commanding General
U. S. Army Electronic Proving Ground
ATTN: Technical Library
Fort Huachuca, Arizona

Commander, Armed Services Technical
Information Agency
ATTN: TIPCR
Arlington Hall Station
Arlington 12, Va. (10 copies)

Chief, U. S. Army Security Agency
Arlington Hall Station
Arlington 12, Va.

Deputy President
U. S. Army Security Agency Board
Arlington Hall Station
Arlington 12, Va.

Commanding Officer
U. S. Army Signal Equipment Support
Agency
ATTN: SIGMS-ADJ
Fort Monmouth, N. J.

U. S. Continental Army Command LnO
USASRD
Fort Monmouth, N. J. (3 copies)

Corps of Engineers Liaison Office
USASRD
Fort Monmouth, N. J.

AFSC Liaison Office
Naval Air R&D Activities Command
Johnsville, Pa.

Marine Corps Liaison Office
USASRD
Fort Monmouth, N. J.



DISTRIBUTION LIST (Cont.)

(1 copy unless noted otherwise)

Commander, Air Force Command and
Control Development Division
ATTN: CRZC
L. G. Hanscom Field
Bedford, Mass.

Commanding Officer
USASRDL
ATTN: Dir of Research/Engineering
Fort Monmouth, N. J.

Commanding Officer
USASRDL
ATTN: Tech Documents Center
Fort Monmouth, N. J.

Commanding Officer
U. S. Army Signal R&D Laboratory
ATTN: SIGRA/SL-NDM
Fort Monmouth, N. J.

Commanding Officer
USASRDL
ATTN: Logistics Division
For: B. J. Keigher
Fort Monmouth, N. J. (9 copies)

Commanding Officer
U. S. Army Signal R&D Laboratory
ATTN: Tech Info Div
Fort Monmouth, N. J. (3 copies)

Chief of Naval Operations
U. S. Navy Department
ATTN: E. H. Stuermer (OP-07T1C)
Washington 25, D. C.