AUTOMATIC DETECTION OF RADIO-SIGNAL INITIATION WITHIN THE H.F. BAND

A thesis submitted for the
degree of Doctor of Philosophy

by

Kannazuthu Puthanmadam Rajappan, M.Sc., D.I.C.
Electrical Engineering Department,
Imperial College of Science & Technology,
London, S.W.7.

April 1965.

ABSTRACT

This thesis describes an investigation into the problem of designing an electronic system capable of automatically searching selected portions of the High Frequency (H.F) spectrum for possible new transmissions, and into the limitations and practical feasibility.

When there is a priori knowledge of the various signal parameters such as type of modulation, frequency and time of transmission etc., various well known techniques can be used for the detection of the desired signal. But in the present problem under investigation, no such a priori knowledge may exist. Therefore we have to adopt special means of detection.

Preliminary experiments have shown that the technique of scanning the selected portions of the H.F. spectrum, measuring the energies contained therein, and comparing these with some preset thresholds is one of great potential accuracy; but that certain problems associated with noise and fading need to be resolved.

Experiments have been performed to study the fading effects and the effect of various types of noise in order to fix a threshold for detection. Experiments have also been performed to determine the probability of detection both as function of signal to noise ratio and signal duration.

Finally the basic design considerations of an electronic system for detecting new signals in the H.F. spectrum is discussed in detail. The proposed system is straight forward and needs very little refinement. The important and most striking feature of the system is the possibility of saving vast amounts of man-power which would otherwise be needed for the monotonous and repetitive job of tuning a receiver for long hours of day and night.

ACKNOWLEDGMENTS

The writer wishes to express his gratitude to his supervisor, Professor E. Colin Cherry, for his guidance and encouragement during the course of this work.

The pursuance of this research work was made possible by the Admiralty (A.S.W.E.) who provided a Research Assistantship as well as funds for the construction and purchase of electronic apparatus.

The assistance of Mr. A. White and others on the technical staff of the Imperial College is gratefully acknowledged. Thanks are due to Mrs. Janette Dean for her speed and accuracy in typing this thesis.

Finally the writer is heavily indebted to many friends and colleagues in the Communication Section for many valuable discussions and especially to Mr. A. H. Robinson and Mr. R. L. Wiley for their devoted participation in the work proof reading. Their assistance was invaluable.

CONTENTS

Section			Page
ABSTRACT			2
ACKNOWLEDGMENTS			4
CONTENTS			5
1.	INTRODUCTION AND BACKGROUND		9
	1.1	Introduction	9
	1.2	The Nature of H.F. Signals	9
	1.3	The Need for Monitoring	11
	1.4	Present Day Methods of Monitoring	12
	1.5	The Objects of Present Investigations	13
2.	RADIO SURVEILLANCE PROBLEM		15
	2.1	General Problem	15
	2.2	Two Possible Strategies	16
	2.3	Scanning the H.F. Spectrum	18
	2.4	Spectrum Analyser Sensitivity and Resolution	20
	2.5	Calculation of I.F. Bandwidth	25
	2.6	A Mathematical Model	27
	2.7	An Estimator	31
3.	SPECTRAL REPRESENTATIONS OF SIGNALS		35
	3.1	Introduction	35
	3.2	The Fourier Spectral Representation	36
	3.3	Time Bandwidth Problem	39
	3.4	The 'Uncertainty Principle'	43

4.	THE THEORY OF SIGNAL DETECTION	44
	4.1 General Discussion	44
	4.2 Bayes' Theorem	47
	4.3 Hypothesis Testing	50
	4.4 Neyman-Pearson Method	52
	4.5 Applications in our Problem	56
	4.6 Sequential Observation Process	60
	4.7 Summary and Conclusions	61
5.	PRELIMINARY EXPERIMENTS	62
	5.1 Introduction	62
	5.2 Circuit Arrangements	66
	5.3 Setting up of the Equipment	67
	5.4 Preliminary Tests	67
	5.5 Results and Discussions	68
	5.6 Threshold System	69
6.	EFFECTS OF FADING	74
	6.1 General Discussion	74
	6.2 Subjective Tests	81
	6.3 Preliminary Results	81
	6.4 Noise - Its Sources	83
	6.5 Effects of Noise	86
	6.6 Experimental Techniques for Evaluation of Fading	89
	6.7 Results and Conclusions	90
7.	MAIN EXPERIMENTS	93
	7.1 Introduction	93

	7.2	Determination of Threshold	94
	7.3	Experimental Technique	97
	7.4	Experimental Method for Fixing the Threshold Level	99
	7.5	Evaluation of the Present Strategy for Detection	99
	7.6	Experimental Set-up	100
	7.7	Experimental Technique	102
	7.8	Results and Conclusions	103
	7.9	Determination of Detection Probability	106
	7.10	Final Remarks	110
8.		SUMMARY AND CONCLUSIONS	112
	8.1	Summary	112
	8.2	Advantages of Threshold System	115
	8.3	A Note on Signal Extraction	116
	8.4	Suggestions for Future Research	117
	8.5	Conclusions	119
APPENDIX 1			120
APPENDIX 2			129
LIST OF REFERENCES			144

"If you are able to measure something, then you are in a position to talk about it"

Lord Kelvin.

INTRODUCTION AND BACKGROUND

1.1 Introduction.

The rapid and automatic search of wide frequency bands for the location and identification of H.F. signal sources is becoming increasingly important day by day. Equipment capable of conducting rapid surveillance is required at air traffic centres, monitoring stations, missile launch and tracking stations etc.

The investigations described in this thesis form a detailed study of the problem, its limitations and practical feasibility. In order to appreciate the complex nature of the problem and why such studies are warranted, it is useful to recount some of the present problems in the H.F. transmission.

1.2 The Nature of the H.F. Spectrum

The past few years have seen an enormous increase in the use of Radio Frequencies for broadcasting and other purposes. A mere glance at the Frequency List[3] published by the International Telecommunication Union, Geneva, will convince anyone of this staggering fact. During the early stages of the last war the medium wave broadcasting bands were already becoming overcrowded. Since then the use of these frequencies for broadcasting has still further increased both in number and power of transmitters used. With the great increase in the demand for frequencies, there has been considerable widening of knowledge of

the properties of some of the frequency bands and their suitability for various types of services.

Among these various bands a relatively narrow but important subdivision of the radio spectrum between 3 - 30 Mc/s, usually referred to as the High Frequency (H.F.) Spectrum, is used for various communication purposes. It was not until the potential advantages of the ionospheric propagation in the H.F. band were realised that the use of this particular part of the spectrum was seriously considered. One of the advantages was that long-distance communication was possible, with lower transmitter powers and much cheaper and simpler aerial systems than with the L.F. band. Added to this there was another important consideration, viz: the greater bandwidth available in the H.F. band made it highly desirable for those services wanting to communicate economically over long distances and stimulated its use for long distance fixed and mobile services.

However, there are also some disadvantages in the use of the H.F. Spectrum. Because of the variation in the characteristics of the ionosphere, with the time of the day, season of the year and the phase of the sun spot cycle, to maintain communication it is necessary for an H.F. circuit to be assigned more than one frequency. Again the changes in the ionosphere from sun spot reduce the useful bandwidth of the H.F. spectrum. Fading is yet another disadvantage. Radio circuits are prone to the selective type of fading which results from multipath transmission. Moreoever, the transmitted signals are also

liable to interference from atmospherics and other causes and to interruptions from ionospheric storms and 'Dellinger' fade-outs. Due to the fact that high frequencies can be used effectively for long distance communication with relatively low power emission, their potential interference range is also very large since the number of stations which can use the same frequency without mutual interference is limited.

Nevertheless, the economic advantages to be gained by using frequencies in the H.F. spectrum have led to the expansion of long-distance fixed, maritime mobile, aeronautical mobile and broadcasting services. Even though the initial development was mainly in connection with the long distance fixed service, which is still the major user of the H.F. Spectrum, the requirements for international point to point radio telephone and radio telegraph channels are still increasing. The use of the H.F. spectrum by broadcasting services, the land, maritime and aeronautical mobile services has greatly increased, especially during the last decade.

1.3 The Need for Monitoring

The ever growing number of radio signals appearing within the H.F. and other bands has created a pressing need for techniques whereby the presence or absence of one or more of these signals may be automatically located and detected. Systems for implementing these

techniques are bound to be complex and gigantic but have important civilian and military applications.

In such communication systems it is necessary to 'monitor' or 'track' the signals so that the message of interest can be located, detected and perhaps recorded for various purposes. Each monitoring mission has its own requirements depending upon the purpose for which the monitoring is used, therefore each system has its own peculiarities. It is not always new equipment; sometimes it is only a different arrangement.

A civilian organisation like the British Broadcasting Corporation uses a monitoring system the function of which is to intercept and summarise news items broadcast by foreign stations.

In military missions the monitoring is used for gathering and dissemination of intelligence including early warning systems.

1.4 Present Day Methods of Monitoring

At present the location and subsequent detection of such signals are largely carried out by well trained crews using an ever increasing amount of expensive receivers and other equipment. For example, the B.B.C. monitoring station at Caversham uses over 50 communication receivers and other subsidiary equipment. In addition man power is needed. The members of the crew locate the signals of interest by a combination of routine manual tuning of the receivers together with

intuition based on experience, trial and error methods and luck. Such methods were adequate until a few years ago.

1.5 The Objects of the Present Investigations

Today with the mounting costs of weapons and equipments there is an increasing necessity to save money and manpower. A system for automatic detection of signals of interest would not only have to be capable of detecting the signal, but also would have to be able to extract and record the message. Many years of research will be required before such a human replacement can be achieved. Nevertheless, we can make a humble start by building a limited system. Ours is such an attempt.

We begin with the statement of the problem in the next section. A possible strategy for the detection of unknown signals within the H.F. band and a mathematical formulation of the problem are presented in this section. Spectral representations of signals and the factors governing the relations between the frequency band, time of transmission and the ability to recognise variations of signal intensity are described in section three. Section four opens with a general discussion on the problem of signal detection. Various methods of hypothesis testing procedures with special reference to our problem are also discussed here. Sections 5 and 6 deal with the basic experiments conducted, while section 7 describes a method of calculating the probability of

detection of a signal as a function of signal to noise ratio and a limited experimental investigation for the evaluation of the proposed system for detection on new signals. We conclude, in Section 8, with a summary and suggestions for future work in this area.

2. THE RADIO SURVEILLANCE PROBLEM

2.1 The General Problem

The general problem is to set up a strategy of decision whereby we can ascertain if any one or more of a small number of new transmissions has started up, each possibly of short duration. The decision, in order to be useful, must be made in the shortest possible time. In the following argument we assume that if T is the time taken to search a bandwidth W_A and t is the duration of a transmission, then $t \gg T$.

Here are two problems which are closely related.

1. The strategy problem: the success of which depends upon the minimum time in which a new transmission is detected.
2. The detection problem involves hypothesis testing and extracting the signal from contaminating noise during the search.

There may be several stations, of course, starting up, but they are presumably independent. So we need consider finding only one of them. If $\overline{S}_o$ = average number of new stations starting up per unit time.

$\overline{S}$ = average number of stations correctly observed/unit time.

Then the probability $P = \overline{S}/\overline{S}_o$ = relative frequency of correct observation and it is this which determines the success of the strategy. The probability P is compounded of 2 probabilities.

1. If the search strategy does not permit the whole H.F. band W_A to be observed during the minimum detection time but only a fraction W, then,

$\frac{W}{W_A}$ = chance of any new station coming under the observation is P_o.

Again if P_d = chance of correct detection if any one of the new signals, (depending upon the efficiency of the detection techniques), then

$$P = \frac{\bar{S}}{\bar{S}_o} = \frac{W}{W_A} P_d = P_o P_d \qquad (2.1)$$

We now realise that this is a real time problem.

The only way we can be sure to observe any new station from the moment it starts up is to have concurrent observations of all possible stations over the whole band W_A. If we do not require to identify the exact station itself and record its message but only to observe whether any stations start up in some band, then the same argument is true i.e. all these bands must be observed concurrently. Each would require a receiver and an observer. The number required is clearly too large.

Let us now consider two strategies.

2.2 Two Possible Strategies

1. A large number of observers N, each with a narrow band

receiver. This has the following disadvantage:-

A very large N for reasonable p_d . However all stations would be under constant observation all the time and $P_0 = 1$ in equation (2.1).

2. A reduction in N or an increase in the bandwidth of the receivers, or both, by adopting some form of scanning.

The choice of any one of the above will depend upon the particular requirement. Thus with strategy 1 the aim would be to have a higher chance of being correct at the expense of missing some stations completely.

With strategy 2 the aim will be to look for all or as many as possible of the new stations, but the probability of correct detection be low.

Before we decide on the strategy we must establish the 'loss' to be associated with each type of error, since certain errors may be much more costly than others. For example, in a radar search problem, overlooking a target is usually much worse than the detection of a false echo; in such an unbalanced situation, we would decide that a target existed even though there was reasonable doubt. On this line of argument we will look for as many new transmissions as possible even though the probability of false alarms may be high.

Thus we will adopt strategy 2, viz. scanning the spectrum.

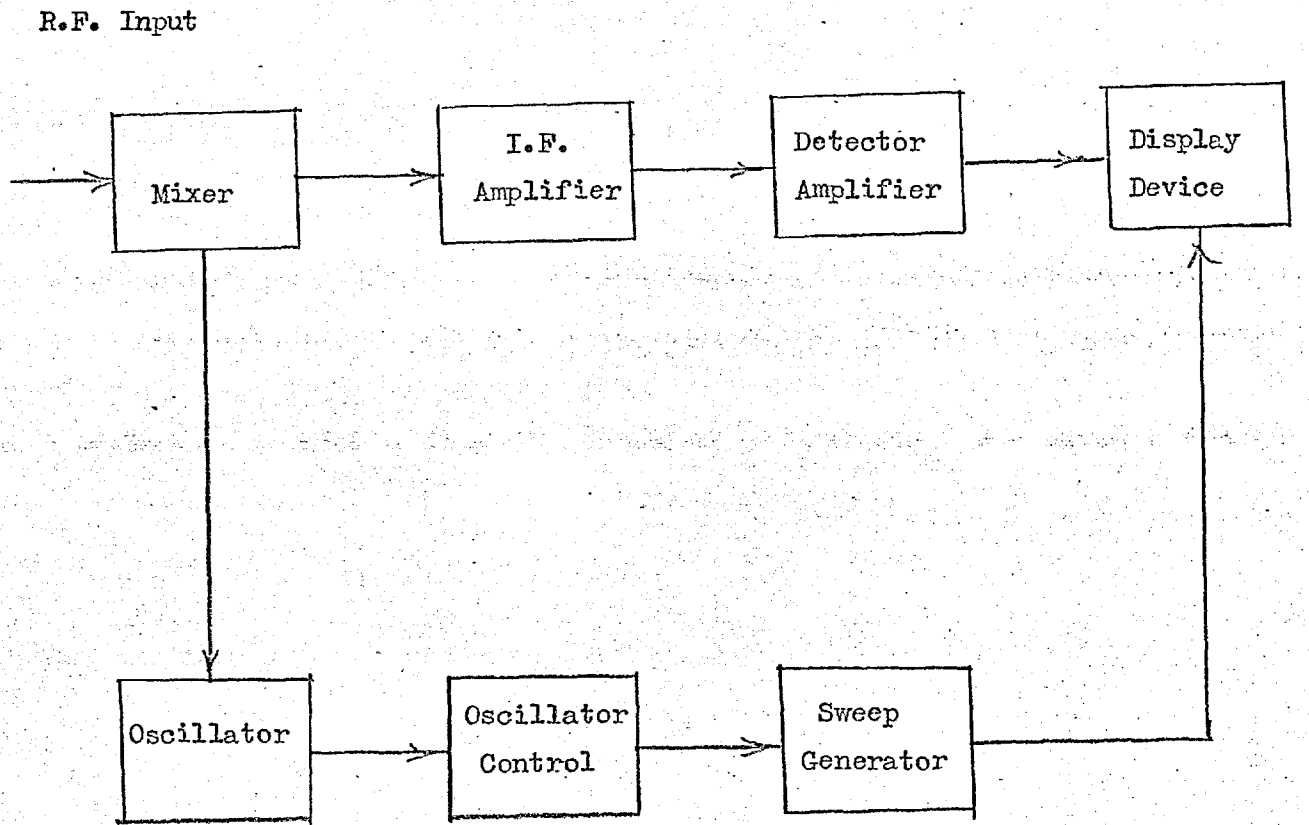
2,3 Scanning the H.F. Spectrum

The scanning can be done using spectrum analysers. A specially adapted spectrum analyser is commercially readily available. This is known as the Panoramic Adaptor and is used with a communication receiver. (See Figure 2.1).

The panoramic adaptor is a radio spectrum analyser with a device for providing a display of the signal distribution in a selected portion of the various frequency bands of the receiver. It is essentially a narrow band superhetrodyne receiver which is repeatedly swept in frequency over selected portions of the H.F. Spectrum. At the same time the horizontal deflection of the spot on the cathode ray tube moves in synchronism with the sweep. The vertical deflection of the spot is proportional to the output voltage of the receiver. The resultant display is a plot of amplitude versus frequency over the H.F. band of interest.

Much useful information can be obtained from such a display. In recent years the radio frequency spectrum analyser has evolved from a special purpose device into a highly refined and versatile measuring instrument. The presence or absence of signals of interest, their frequencies, relative amplitudes etc., may be determined from the display of the spectrum. The spectrum thus displayed can also be considered as the energy distribution of various signals.

At combinations of sweep speed, sweep time and bandwidth, the displayed trace can undergo the following changes:-



CONTINUOUS TUNING SPECTRUM ANALYSER.

FIG. 2.1.

1. Reduction in peak amplitude.
2. An increase in effective bandwidth.
3. A displacement of the frequency of peak amplitude in the direction of the frequency sweep.
4. An assymmetrical display with the trailing side of the response curve being stretched more to one side.
5. A damped oscillatory discharge at the trailing edge of the response.

Barber[1], Hok[18], Williams[34] and many others have studied this problem in some detail under some restrictive conditions and have arrived at some useful results. Nevertheless it is important to look at this problem in some detail since this will be directly connected with our immediate task of arriving at a suitable sweep rate.

2.4 Analyser Sensitivity and Resolution

It was stated before that the analyser is an instrument that displays on a rectangular co-ordinate field a plot of input strength versus frequency. Because all but the simplest unmodulated sinusoidal (CW) signals contain multiple frequency elements of differing amplitudes - in other words a complex spectrum rather than a single frequency - the spectrum analyser is most useful for our purpose.

A CW signal produces a trace on the spectrum analyser screen as shown in the Figure 2.2. If the sweep speed of the analyser is very slow, the trace is a plot of the band pass characteristics of the

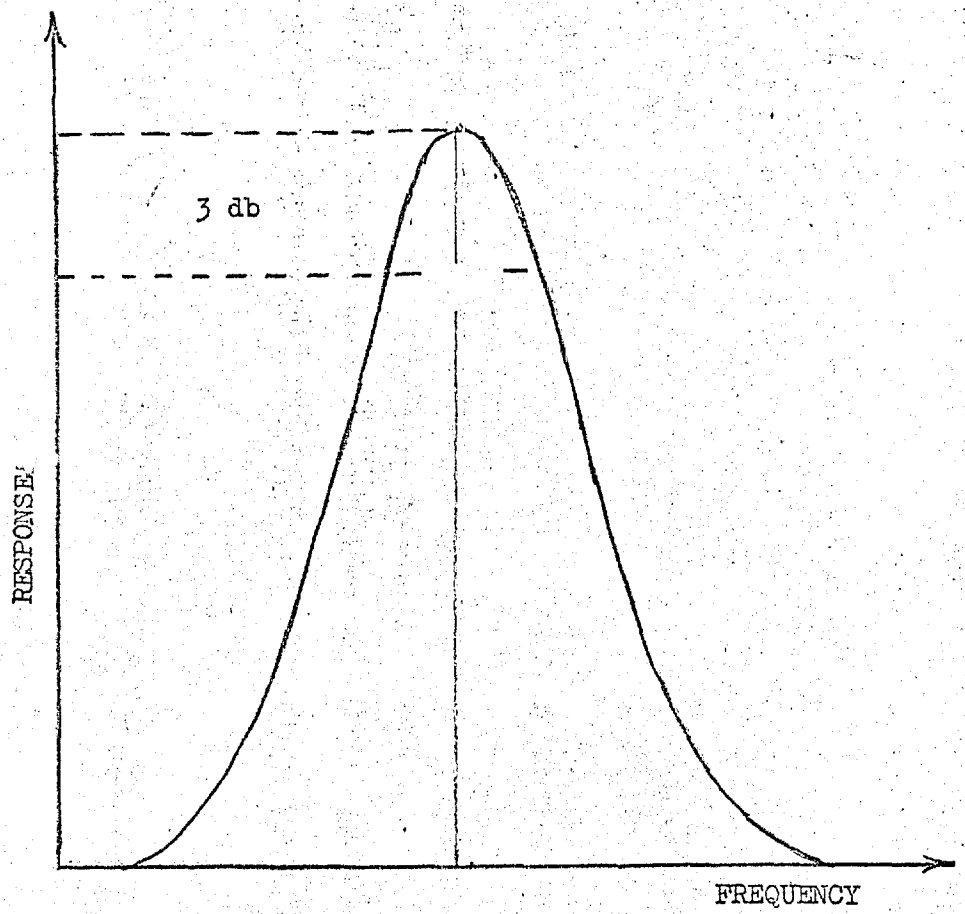


FIG. 2.2.

narrow band I.F. Amplifier. This may be explained as follows. Only one line exists in the spectrum of a CW signal. Heterodyning the single line with the swept wave from the swept local oscillator again results in a one line spectrum. Now, however, this line is being swept in frequency with time. Delivering this signal to the narrow band amplifier is equivalent to applying a signal of constant amplitude to the amplifier and varying its frequency. Plotting the output would result in the frequency response curve of the amplifier.

Since the plot on the screen of the CRT is nothing more than the amplitude versus frequency output of the narrow band amplifier, it is the frequency response curve or band pass characteristic curve of amplifier. If the sweep is rapid, the response of the tuned circuits may be too slow to allow the swept input to fully charge and discharge the amplifier circuit and the amplitude becomes a function of the sweep. As the sweep rate is increased, the pip becomes reduced in amplitude and is widened out.

We now introduce the term resolving power R to represent the ability of the narrow band I.F. amplifier to produce a distinct sharp response at a given sweep frequency. R is taken to be the displayed width of a CW pip at 3 db points. (See Figure 2.2). Thus, if the pip is 5 KC wide at 3 db points, the resolution R is said to be 5 KC. Two pips of equal amplitude and separated in frequency by R as so defined would barely be distinguishable from each other on the screen.

Both the resolving power and the sensitivity of analyser are

degraded when the sweep rate is increased due to the limited reponse time of the tuned circuits. This is shown in Figure 2.3. It is therefore necessary to find out these limitations.

First of all we have to make a valid assumption. This is about the response curve of an I.F. amplifier which is composed of several synchronously-tuned stages. If we look at the response curve it closely resembles that of a Gaussian error function.

It can be shown that (see Appendix) the response curve, though complicated, depends upon a single parameter K that involves the bandwidth of the filter and the rate of change of the instantaneous frequency of tone. K is defined as follows -

$$K = \frac{F}{TB^2} \qquad (2.2)$$

Where F = the sweep bandwidth in cycles per second

T = the time of the sweep

and B = the bandwidth of the I.F. amplifier of the receiver.

It will be observed that K is dimensionless.

Thus, given the band F to be scanned and any acceptable 'distortion factor' K , the I.F. slot bandwidth is

$$B = \sqrt{\frac{F}{kT}} \qquad (2.3)$$

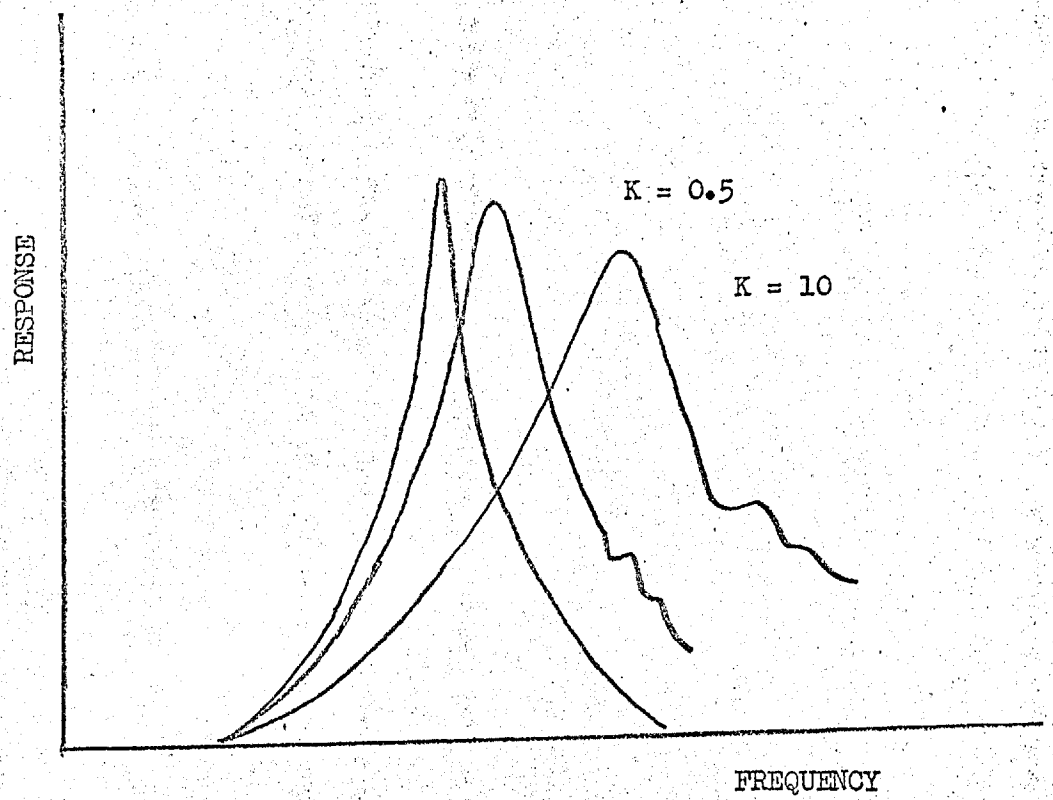


FIG. 2.3.

2.5 Calculation of I.F. Bandwidth

It is generally agreed by many authors that when K is less than 0.5 the distortion in the displayed spectrum is negligible. However, if either the sweep rate is increased or bandwidth decreased K rises and the distortion becomes evident. Initially, a loss of sensitivity occurs. Barber[1] has experimentally shown that when K is 10, the reduction in sensitivity is 4.5 db. According to Saul et al[30], the reduction is about 6 db. The latter also states that the amplitude of the response relative to that at 0 sweep rate is given by,

$$A = (1 + 0.195\ K^2)^{-1/4} \qquad (2.4)$$

On the basis of the above argument we arrive at the following conclusion.

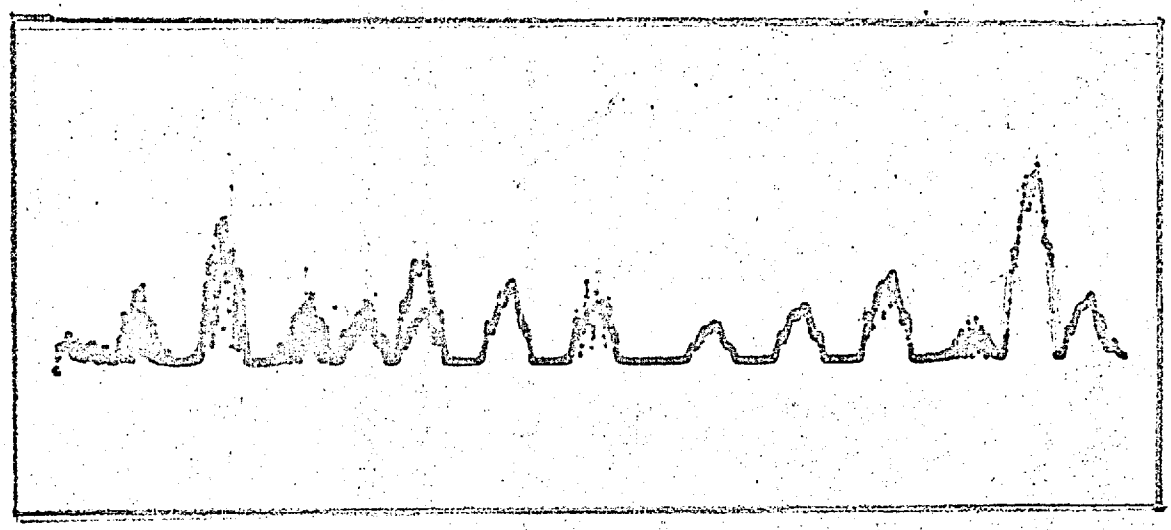
Suppose we sweep 100 Kc/c in 0.2 of a second.

Then $$\frac{F}{T} = \frac{10^5}{0.2} = 5 \times 10^5$$

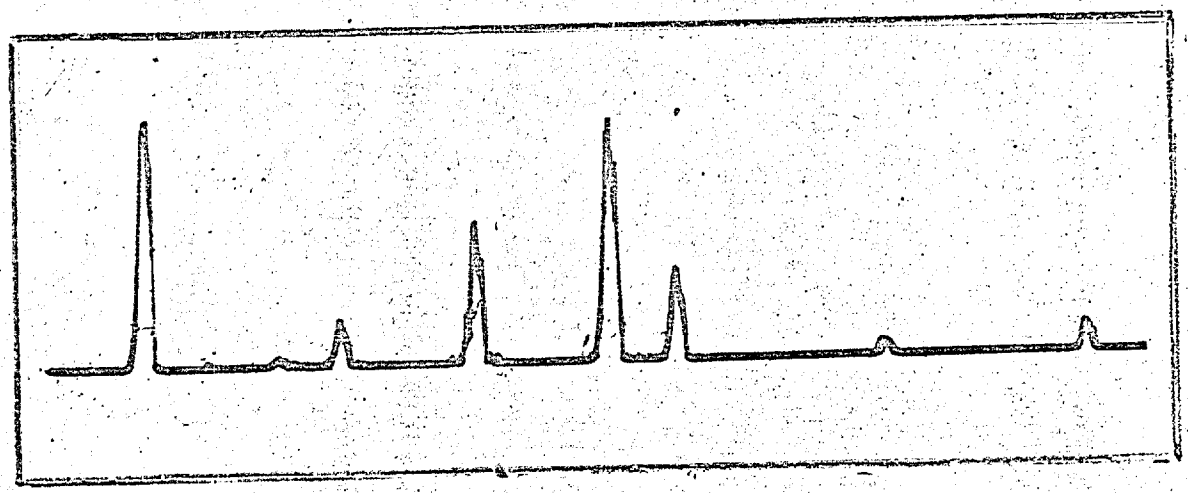
Now to restrict the value of K close to 0.5, we can safely choose an I.F. bandwidth of

$$\sqrt{\frac{5 \times 10^5}{0.5}} = 1 \text{ Kc/s.}$$

The displays of real and artificial spectra are shown in Figure 2.4.



AN EXAMPLE OF A SPECTRUM OF 100 Kc/s BANDWIDTH



THE SPECTRUM GENERATED BY AN ARTIFICIAL SPECTRUM GENERATOR

FIG. 2.4.

2.6 A Mathematical Model

In this discussion our attention is exclusively devoted to the simplest and most immediate question to ask of an H.F. spectrum with a given bandwidth. The question is 'How many signals are present at a particular time?'. With previous knowledge of the spectrum we can then say whether there is any increase or decrease in the number of signals present. The problem of estimating the number of signals that are present in a noisy H.F. spectrum when the signal amplitudes are statistically distributed is investigated. The problem can be stated in the following mathematical language.

We consider a random variable the distribution of which has a known mathematical form, but contains a certain number of unknown parameters. We are given a sample of observed values and it is required to use the sample values to form estimates of parameters.

Suppose we are given a sample of a population. It is assumed that the distribution of the sample has a given mathematical form, but involves a certain number of unknown parameters. Under these assumptions there will be an infinite number of functions of the sample values that might be proposed as estimates of parameters. Now how can we best use the data to form estimates? Another question which is more important is: What do we mean by 'best estimates' ?.

We may answer this by saying that the best estimate is the one which comes closest to the true value of the parameter to be estimated.

But we have to remember that every estimate is a function of a sample value and therefore is an observed value. Consequently we have no means of predicting the individual value assumed by the estimate in a given particular example, so that we cannot judge the goodness of the estimate from individual values, but only from the distribution of the values which it will assume in the long run. In other words, we can judge the goodness of estimate only from its sampling distribution. When greater part of this distribution is concentrated in some small neighbourhood of the true value, there is a great probability that the estimate will only differ from the true value by a small quantity.

The above concepts can be directly applied to our problem as follows.

A sample of the H.F. spectrum may contain thousands of signals. We can consider the sample to consist of C cells of which S contain signal plus noise and the remaining (C-S) contain noise alone. The signal amplitude distribution function is known to within the P parameter distribution vector $A = (a_1, a_2, \ldots\ldots\ldots\ldots a_p)$. The noise characteristics are assumed known. The observable quantities are R_i, $(i = 1,2 \ldots\ldots\ldots C)$. These observables are proportional to the energy content of the corresponding cells.

The general problem is characterised by $p + 1$ parameters, viz, p unknown signal parameters and the random variable S.

Our problem here is to estimate S.

In this particular solution of the problem we make the following assumptions.

1. Each cell contains one and only one signal.

2. No two signals overlap.

3. Many signals are assumed present.

For the estimation of $(p + 1)$ parameters we should like to form a set of observations, v_j, $(j = 1, 2, \ldots\ldots p = 1)$ where v_j are functions of R_i, $(i = 1, 2, \ldots\ldots C)$ and therefore of the parameter vector A and the number S. Since there are p unknown signal amplitude distribution parameters $A = (a_1, a_2 \ldots\ldots a_p)$ and the parameter S is unknown, we form the following function:

$$V = V(R_1, R_2, \ldots\ldots\ldots R_C) = V(A,S)$$

$$= v_1(A,S), \; v_2(A,S), \ldots\ldots v_{p+1}(A,S)$$

In this manner the random variables v_j form a model from which it is possible to estimate S.

Estimators of these $(p+1)$ parameters must, in general, be obtained from $(p+1)$ equations, each of which is a function of the observable quantities R_i. In this particular case only one parameter estimator is required, namely, S, and therefore we need only one equation.

The density function of R_i is complicated because it depends

upon whether a signal is present or not.

The basic problem with which we are confronted is to specify some estimation procedure and to analyse its performance. A satisfactory solution to this problem is considered to be the specification of an estimation procedure whose performance is good with some reasonable performance criterion that is compatible with the frame of reference of this problem.

In this particular problem of interest C and S may be very large numbers. If 100 signals are distinctly observed in the noisy H.F. spectrum, we are interested to know if there are a total of 110 or 120 actually present. In other words, we wish to know if there are 10 or 20 additional signals buried in noise. It appears that a suitable measure of estimator performance in this case is the mean and variance of the fractional error of $\hat{S}$.

Let $\hat{S}$ denote the estimate of S. Then ρ the fractional error can be written as,

$$\rho = 1 - \frac{\hat{S}}{S}$$

The mean and variance of then fractional error are,

$$E\rho = 1 - E\frac{\hat{S}}{S}$$

$$\text{and } \text{Var}\rho = \text{var}\frac{\hat{S}}{S}$$

2.7 Estimator of S

The minimum error estimator of S, $\hat{S}$, is obtained by setting threshold in each cell and comparing the energies therein with these present thresholds. This is described in detail in Sections 4 and 5. We can now write down the estimator in the following form

$$\hat{S} = U(X_i - T_i)$$

where X_i are the measured values of R_i the T_i are the thresholds in the corresponding cells, and $U(X)$ the unit step function.

The estimator can be obtained by the equipment shown in the block diagrams 2.5, 2.6 and 2.7.

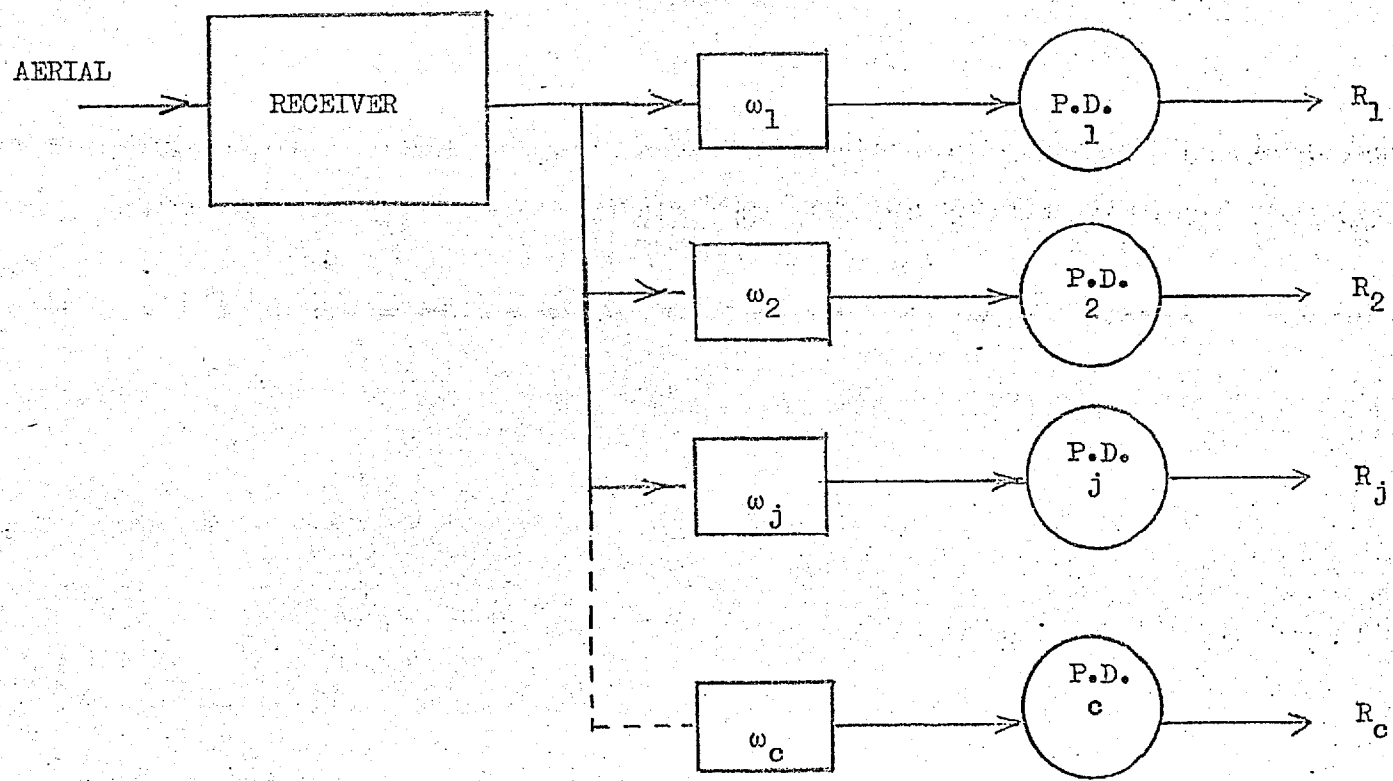
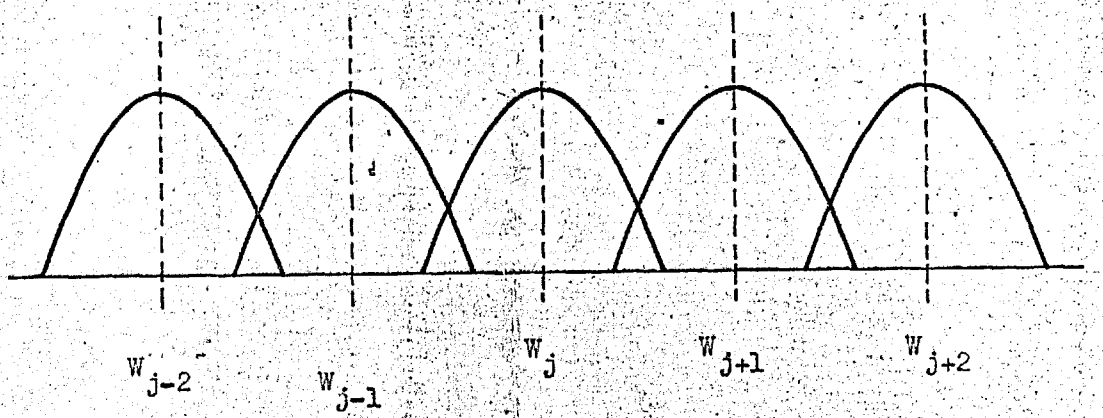

AERIAL
RECEIVER
ω_1
ω_2
ω_j
ω_c
P.D. 1
P.D. 2
P.D. j
P.D. c
R_1
R_2
R_j
R_c

FIG. 2.5.



REGION OF SPECTRAL COVERAGE.

FIG. 2.6.

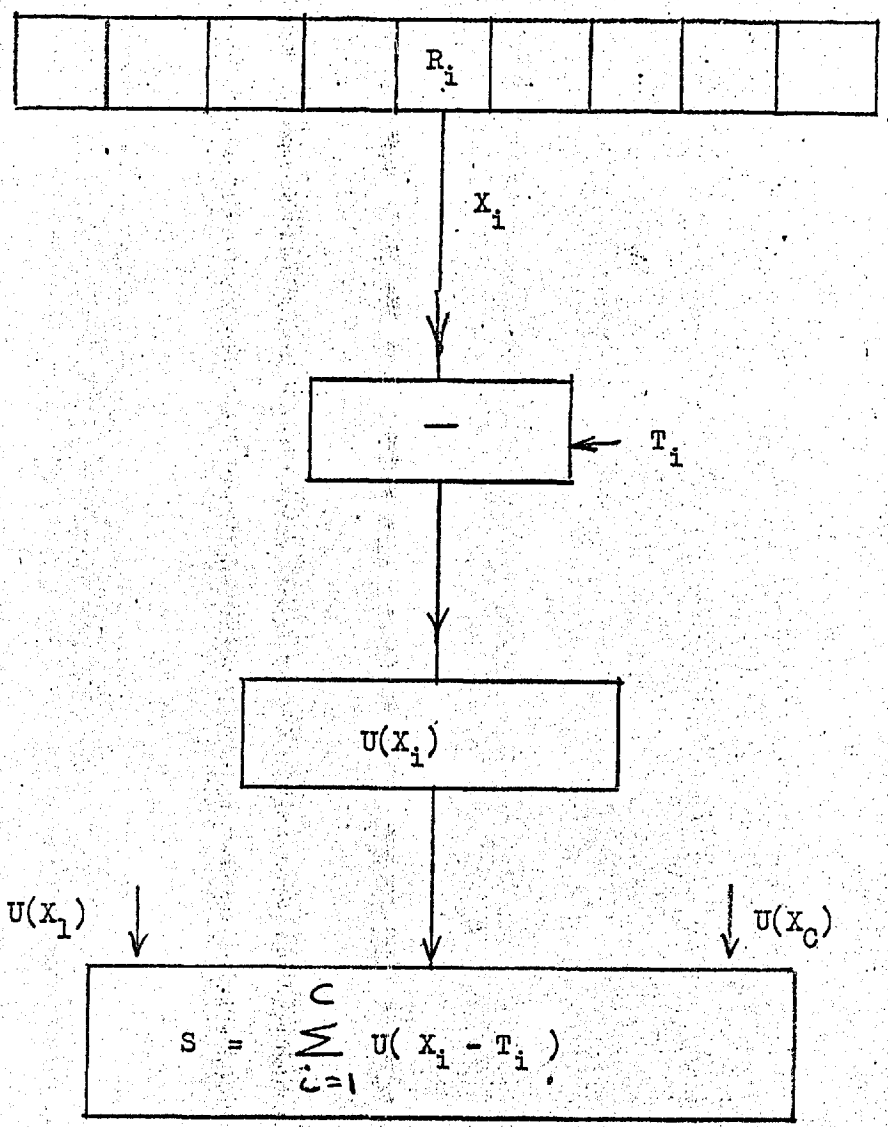


FIG. 2.7.

3. SPECTRAL REPRESENTATION OF SIGNALS

3.1 Introduction

In the last section we saw that under certain limiting conditions, scanning the High Frequency Spectrum of interest was the best strategy for our purposes. We also arrived at some useful results regarding the scanning rate, the bandwidth of the scanning filter which in our case is the bandwidth of the I.F. amplifier and various other factors. In this section we will look at the problem from a slightly different angle.

Suppose we are scanning a spectrum of bandwidth W_A. Within this band there may be many signals. Each one will have a certain bandwidth and certain duration. Let us represent these by w_1, w_2, ... , w_n and their corresponding durations by t_1, t_2, ..., t_n. Let us further assume that T represents the time taken for one complete scan of the spectrum and F the bandwidth of the scanning filter.

Now we ask the following question. How long during a particular scan do we have to look at each one of the signals before we can detect them? The answer is not simple. A short discussion on the basic properties of various signals, the relation between their bandwidths and durations etc. may not be out of place here before we return to answer the question.

3.2 The Fourier Spectral Representation

Hartley[17] stated that in order to transmit a certain amount of information a definite product, bandwidth X time is required. This statement is the logical outcome of Nyquist's discovery in 1924.[28] Looking back, it would appear that both Nyquist and Hartley's work was in part anticipated by the Fourier relations to which Hartley makes several references. By these, the spectrum $e(\omega)$ is related to its signal $e(t)$ as follows:

$$e(\omega) = \int_{-\infty}^{\infty} e(t)\, e^{-j\omega t}\, dt$$

and

$$e(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e(\omega)\, e^{j\omega t}\, dt$$

These formulae taken together comprise a Fourier transform pair and apply equally to finitely discontinuous waves and infinite waves. Very closely related to the Fourier Transform pair is the phenomenon of reciprocal spreading.[15] By this phenomenon, if one member of the Fourier pair is increased by a certain factor, then the other member of the pair is decreased by exactly the same factor. If we apply this principle to the spectrum and the time duration of a signal it is easily seen that if the spectrum width of a certain signal is deliberately increased by a certain factor, then the time duration of the signal is decreased in the same factor. The converse applies, if the width of the spectrum is deliberately decreased. In other

words, the product of the spectrum bandwidth and the time of duration of a signal is constant. This is Hartley's result. This constant is a function of signal only, and if an elementary signal is suitably defined, may be equal to unity.(15)

The analytical methods introduced by Fourier continually play an important role in the problems of physics and engineering. Even though Fourier originally derived his results for problems in heat conduction, the most extensive application of Fourier analysis has been in the fields of acoustics and electrical engineering. This property gives us a method for determining the frequency components present in a time-varying function. For example, tones from musical instruments produce a displacement of air molecules. This displacement is a time varying function subject to Fourier analysis.

Sinusoidal functions occupy a unique position in electrical engineering. They are easy to generate, and the steady-state response of a linear system to sinusoidal excitations can be found very easily. However, they are the favourite functions of engineers primarily because arbitrary periodic functions with a few restrictions can be expanded into Fourier series of harmonic sinusoidal components, and transient nonperiodic functions can be expressed as Fourier integrals. We can therefore discuss arbitrary waveforms in terms of their frequency spectra.

The Fourier Spectral representation owes much of its popularity

to its association with frequency since the frequency domain is well suited for specifying the transmission characteristics of signals. A spectral representation using sinusoidal components is appropriate whenever we are seeking an invariant under translations in time. Thus this type of representation is also useful in describing the time statistics, such as power spectrum, of a stationary stochastic signal. Also when we are interested in the time averaged properties of signals, the classical Fourier representation is useful. But such spectral representations are not well suited for discriminating between signals or for detecting the occurrence of a particular signal from corrupting noise.

The type of signals we use in communication are characterised by their epoch (that is, the time at which they occur) as well as their waveform or structure relative to the epoch. The ordinary signals like speech, music, etc., are best described as consisting of time varying frequency components and it is inconvenient to represent them as pure time functions or pure frequency functions.

Gabor[14] tried to overcome this difficulty by analysing the waveform not into timeless sine waves but rather into some elementary signals which are essentially concentrated about the time and frequency axes.

Inspired by the science of Quantum Mechanics, he drew a parallel between the uncertainty product of momentum and position and

the uncertainty product of time and frequency. Gabor's starting point was the 'Sampling Theorem', which states that a signal which in some sense lasts for a length of time T and in some sense occupies a bandwidth W can be characterised by a collection of 2 TW numbers. He then argued that a complicated signal can be thought of as a collection of elementary signals to each of which one and only one of the above numbers could be assigned. He then selected an elementary signal waveform that would have a minimum duration - bandwidth product, hopefully unity. This signal has an envelope of the form e^{-x^2} which determines the duration t of the signal and under the envelope is a 'carrier' which determines the frequency, f, around which the spectrum of the signal is distributed. Within any interval of time, t, the signal is pictured as being the sum of elementary signals occupying that interval, whose frequencies f are separated by 1/t . The next succeeding time interval of the signal is constructed of elementary signals 'occupying' that interval and so on. He then obtained a set of functions which are suitable for characterising signals such as speech which has a structure both in time and frequency.

3.3 Time Bandwidth Problem

The definition or duration of a signal used by Gabor is basically a radius of gyration. This is inconvenient from several points

of view. Amongst these is the need to use a single sided spectrum in order to achieve a reasonable definition of duration. However, the precise meaning of a bandwidth duration product remains unclear in the case of complicated signals for which TW is much greater than one. Thus the statement that a signal which lasts for time T and occupies a bandwidth W can be characterised by $2TW$ numbers can be only partially true. For example, a function cannot be limited to a duration T and bandwidth W in the sense that $s(t)$ the time function, and the corresponding spectrum both vanish outside of intervals of width W and time T respectively. In fact a signal and its spectrum cannot both vanish in any finite interval of time and frequency respectively.

By assuming that a signal is bandlimited, that is, its double spectrum $S(f)$ vanishes outside the interval of frequencies up to W, Lerner[20] gives exact meaning to a characterisation of bandwidth duration in terms of $2TW$ numbers.

It is well known that such a signal can be represented by a sum of Sin x/x functions.

$$S(t) = \sum a_n \frac{\text{Sin } \pi (2Wt - n)}{\pi (2Wt - n)}$$

To determine the a_n it is only necessary to examine the signal at times t_n separated by intervals of $1/2W$.

$$a_n = S(t_n) = S\left(\frac{n}{2W}\right)$$

By applying the above sampling procedure, 2 TW numbers of a_n can be determined from a segment of the signal of length T . But we have to remember that these 2 TW numbers do not characterise the band limited s(t) for all time nor do they completely characterise s(t) during the interval T unless all the a_n which are not included in the set of 2 TW numbers are zero. In fact the sin X/x function falls off so slowly that a_n would be determined at sample points well outside the interval T may profoundly affect the signal inside the interval T .

Thus even though it is correct to say that in a sense a band limited signal is characterised completely by 2 TW samples per second it is also true to say, in general, that the samples taken during any finite period of time T , may only poorly approximate the signal except at sampling points.

The above means of expressing the TW product has qualitative merit in estimating the number of coefficients that may be required to specify a signal. The exact value will depend upon

a) the arbitrary definition of duration employed;

b) the particular signal being analysed.

Gabor tried to build a complicated waveform out of elementary signals having unity duration bandwidth product and attached too much importance to a minimum TW product. But it appears that when

we consider about a collection of elementary signals, the arrangements of arrays of these elementary signals at intervals of time and frequency are more important than the precise values of their individual TW products.

In short, if a function representing a signal has a spectrum which is strictly limited to a finite bandwidth, the time representation of the function will trail on indefinitely. In theory, therefore, it is impossible to have a function of strictly finite bandwidth and finite time duration. Thus we conclude that we cannot define both frequency range and time interval of a signal exactly. An accurate definition of the time of occurrence requires a wide frequency range and an accurate definition of exact frequency requires a very long time of observation. The only way we could define a signal of a particular frequency occurring at a particular time is just the probability of occurrence in terms of bandwidth and the interval of observation. It is also clear that the greater the bandwidth and the time, the narrower will be the probability range.

Nevertheless, it is possible to have a function representing a signal which has almost all its energy content within a finite time duration and within a finite bandwidth. In fact, most of the communication signals can be represented by functions of this type. Our strategy for detection of various signals is based on this well known concept.

3.4 The "Uncertainty Principle" (8)(14)

Finally, when we try to apply the "uncertainty principle" from quantum mechanics to signal analysis, the signal takes the place of the elementary particle. Thus the question: "Where is the particle and what is its momentum?" becomes: "What is the frequency of the signal and when does it occur?" We can answer this by saying that the intervals of time and frequency in which signal energy is found cannot be reduced simultaneously without limit and in this manner the epoch and frequency are uncertain. However, there is no uncertainty about the spectral shape or waveform as these are completely known.

4. THE THEORY OF SIGNAL DETECTION

4.1 General Discussion

The theory of signal detectability is purely a mathematical model having its roots in probability theory and decision theory. The theory starts with the fundamental assumption that the environment in which the signals occur is a noisy environment. That is to say, at any time a signal is presented, by the very nature of the environment in which it occurs, there must be noise associated with this signal. Thus, the task of detecting a signal becomes one of discrimination between signal plus noise and noise alone. In any given observation interval, a particular observed waveform may have arisen with some probability from signal plus noise or it may have arisen with some other probability from noise alone. Following the observation of an interval a judgment must be made as to whether or not the signal arose from signal plus noise or from noise alone. Because of the probabilistic nature of this problem the judgment can never be made with absolute certainty. Consequently the task of detecting the signal becomes one of testing statistical hypotheses.

Statistics is concerned with the treatment of empirical data by means of the theory of probability. The need for probability theory arises because of the unpredictable nature of the phenomena under observation. Many engineering processes yield data which

appear as if they were generated by some chance mechanism. From experience with the game of chance, if the process were to be observed many times, we would expect to see the relative frequency of occurrence of a particular event of the process approximates a fixed quantity thus exhibiting the property of statistical regularity. Statistical analysis of experimental data is often followed by human decisions and actions - lots of manufactured goods are accepted or rejected, missiles are fired at suspected targets and so forth. Therefore much of the latest developments of statistical decision theory has been concerned with adapting statistical procedure to decisions and actions usually with the object of minimising the average total costs or risks involved.

In radar, for example, the data are a sequence of voltages measured at the terminals of the aerial. Errors in the data may be caused by random noise, the properties of which may be explained with the aid of probability. The observer must take certain actions on the basis of interpretation of the corrupted output voltages. The task of statistical decision theory is to find a way to process the randomly-varying data in such a way as to permit the adoption of maximally efficient course of action.

In experimental research, the object is sometimes to estimate parameters; but more often the ultimate purpose is to use the estimate in some way. One may wish to compare a new set of readings with that

of a standard one and perhaps come to some definite conclusion. We consider that two populations of readings are being compared: those which are standard and those newly obtained readings. The question is whether the mean of the second population is less than or greater than the first. To answer this we set up a null hypothesis, namely, the hypothesis that the two means are the same. On the basis of the sample from the second population, we shall accept or reject the null hypothesis. What exactly we do in a given situation will depend on the criterion for decision. As the transmitted signal or message passes through the communication channel and is amplified in the receiver it is corrupted by the channel noise as well as the receiver noise. The message thus arrives in a distorted form to the observer who assesses the message content on the basis of the voltage amplitudes. In many cases, as in the case of long distance communication, the intensity of the signal is so small that its very presence becomes uncertain. Therefore the first problem the observer has to solve consists in assessing whether a signal is contained in observed voltage or whether the latter is generated by noise alone. The random nature of the amplitudes of the voltages consisting of the input signals indicates that their complete description will be contained in the appropriate distribution laws. These laws will be different according to whether signal is present, and at the output of the receiver there is a mixture of signal and noise or whether there is

only noise. In the problem of detection, this difference shows up in the values of parameter distributions.

Therefore, the problem of detecting signals in a noisy environment reduces to testing on the basis of observed values of the signals received, a hypothesis on the values of the parameters of its distribution. Such problems are studied extensively in mathematical statistics.[10],[12]* We shall consider methods of solution of this problem.

4.2 Baye's Theorem

As is well known, the probability that the events A and B occur simultaneously, which we denote by $P(A,B)$, satisfies the following formula,

$$P(A,B) = P(A)\,P_A(B) = P(B)\,P_B(A) \qquad 4.1$$

Where $P(A)$ is the probability of event A,

$P_A(B)$ is the conditional probability of the event B, given that A has occurred and the terms

$P(B)$ and $P_B(A)$ are defined similarly.

The relation 4.1 can be written in the following form:

---

* Other relevant references are Bendat[4], Lee[19], Middleton[24], Mood[25], and Wilks[36].

$$P_B(A) = \frac{P(A,B)}{P(B)}$$

$$= \frac{P(A)\, P_A(B)}{P(B)} \qquad 4.2$$

Now suppose we have a complete set of mutually exclusive events.

$$A_1, A_2, \ldots\ldots\ldots A_K \qquad 4.3$$

The above events are such that one of them must occur, but two or more of them cannot occur simultaneously. If B is an event which can only occur if an event of 4.3 occurs, then the probability that B occurs is given by the expression,

$$P(B) = P(A_1)\, P_{A_1}(B) + P(A_2)\, P_{A_2}(B) + \ldots\ldots + P(A_K)\, P_{A_K}(B)$$

$$= \sum_{K=1}^{K} P(A_k)\, P_{A_k}(B) \qquad 4.4$$

If it is known that event B has occurred, then it is natural to ask which of the events A_k occurred jointly with B. In this case the events $A_1, A_2, \ldots\ldots\ldots A_K$ are called <u>hypotheses</u>, whose probabilities have to be determined. From the equations, 4.2 and 4.4 we get,

$$P_B(A_k) = \frac{P(A_k)\, P_{A_k}(B)}{P(B)}$$

$$= \frac{P(A_k)\, P_{A_k}(B)}{\sum_{k=1}^{k} P(A_k)\, P_{A_k}(B)} \qquad 4.5$$

The above expression was first propounded by Bayes[2] in the eighteenth century and is generally known as Bayes' theorem. In the above expression,

$P(A_k)$ is called the a priori probability of the hypothesis A_k. (the probability 'before the experiment')

$P_B(A_k)$ is called the a posteriori probability of the hypothesis A_k (the probability 'after the experiment', that is, after the event B has occurred).

and $P_{A_k}(B)$ is called the <u>likelihood function</u>.

The Bayes' theorem is powerful tool and is rigorously valid under certain conditions. Given only the probability distribution of the underlying parameter, namely, the a priori probability and the likelihood function, there is a way of finding that probability of a particular population gave rise to the sample. The difficulty is that we almost never know the necessary distribution. In many cases, the underlying parameter is not even subject to probability. Suppose, for example, that we had taken a sample from a population and found its mean to be 10 . If the sample were sufficiently large, we could be confident that the mean of the population was not far from 10 ; but suppose that the sample were small and we wished to make some inferences concerning the mean of the population - the

a posteriori probability, for example, the mean was as small as 7. This is the kind of question we shall attempt to answer in mathematical statistics, but Bayes theorem is not applicable because we cannot fix a number to the a priori probability that the population mean is 7 or less than 7. We do not generally know these a priori probabilities except in artificial cases such as classical problems of balls and urns.

4.3 Hypothesis Testing

To introduce the basic notions of hypothesis testing, it is helpful to consider the simplest case of a single alternative. Suppose a population is known to have either the probability density $P_o(x)$ or $p_1(x)$, and that it is desired to test on the basis of one observation whether the true density is $p_o(x)$ or $p_1(x)$. We designate by H_o the hypothesis that $p(x) = p_o(x)$ and by H_1 the alternate hypothesis that $p(x) = p_1(x)$. We shall call H_o the null hypothesis. The rejection of H_o will be equivalent to the acceptance of H_1. To test H_o we choose a number C and make an observation x_1. If x_1 is less than C, we accept H_o, and if x_1 is greater than C, we reject H_o (see Figure 4.1).

Two kinds of errors are possible in a single alternative case. We may reject H_o when it is in fact true; that is, the population

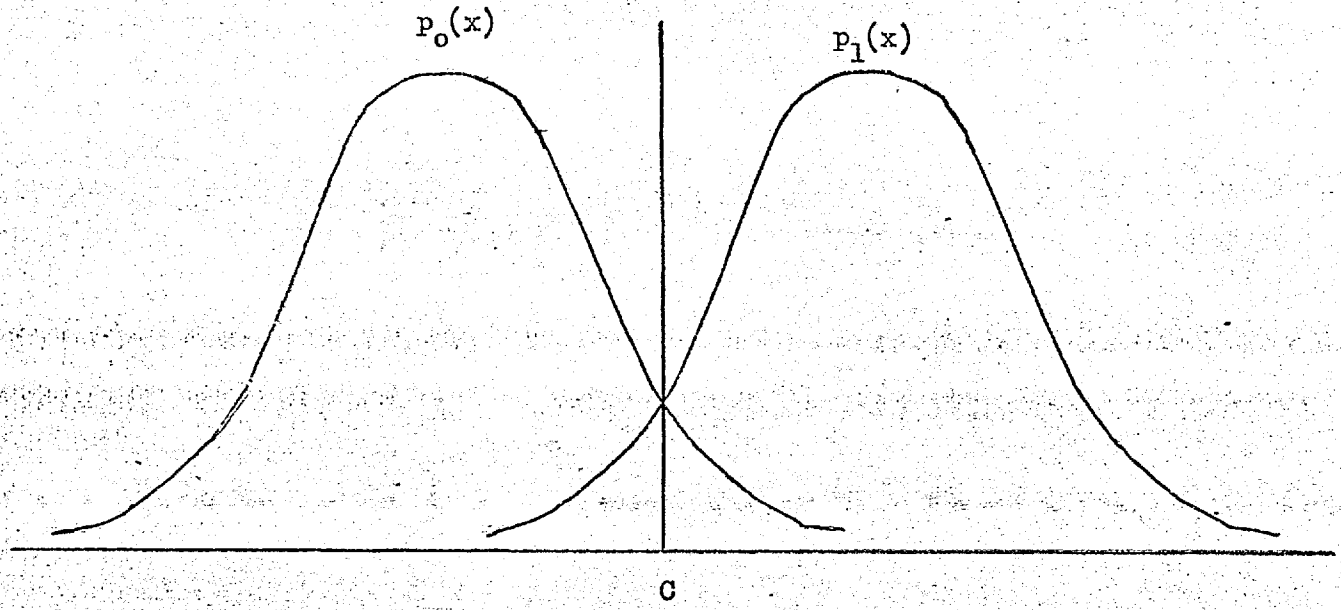


FIG. 4.1

may have $p_o(x)$ as its distribution even though the observed x exceeds C. This is called Type 1 error of the test and its probability of occurrence is

$$\int_C^\infty p_o(x)\,dx \qquad 4.6$$

This probability is often called the level of the test. A second possible type of error is the acceptance of H_o when it is false. That is, the observation may be less than C even though the true population distribution is $p_1(x)$. This is called the Type 11 error of the test and its probability is

$$\int_{-\infty}^C p_1(x)\,dx \qquad 4.7$$

The interval $x < C$ is called the acceptance region for the null hypothesis and the interval $x > C$ the critical region.

4.4 The Neyman-Pearson Method[27]

We shall now consider two other methods of solution to the problem. One of them has been developed by Neyman and Pearson, the other by Wald. The initial ideas of the two approaches are the same and they are present below.

Let z be a random quantity which can be multi-dimensional

with probability density $p(x; \alpha)$ depending upon the parameter α. The functional form of the probability density is determined but the values of the parameters are unknown. As to the possible value of the parameters, two hypotheses, H_o and H_1, are permissible. These are two kinds: Simple and Composite. They are termed simple when the hypothesis H_o consist of assuming $\alpha = \alpha_o$ while H_1 consists in assuming $\alpha = \alpha_1$. They are termed composite when at least one of them allows a parameter α to assume several values. An example of the later situation is the following situation. According to the hypothesis H_o we have $\alpha = 0$, and according to the hypothesis H_1 we have $\alpha > 0$. We shall be concerned, in what follows, with simple hypotheses.

Which of the two hypotheses is valid is assessed on the basis of results of drawing sample values of $\mathcal{X}$, which has a given probability density $p(X; \alpha)$. A sampling point will be denoted by $X = (x_1, x_2 \text{ ----- } x_m)$. The set of all such possible points belongs to an m-dimensional sampling space. The statistical methods considered for testing hypotheses consists of dividing the sampling space into subspaces. In Neyman's and Pearson's theory there are two subspaces and Wald's sequential analysis three. If sampling point is found in one of them, the hypothesis H_o is rejected; if it is found in another, the hypothesis is accepted. Finally in the case of sequential analysis, if the sampling point

is found in the third subspace, the experiment is continued owing to insufficient data.

The random nature of the sampling point may lead to two types of wrong decisions as to the validity of the hypotheses.

1. We may accept the hypothesis H_1 although it is H_o that is valid. This is known as the error of the first kind.
2. We may reject hypothesis H_1 in spite of the fact that it is valid; this is known as the error of second kind.

We will now illustrate what has been said above in terms of detection theory. The recipient of the message, having measured the amplitudes of the voltages at the output of the receiver must establish whether they are caused by the action of noise only, i.e. the hypothesis H_o or whether they contain useful signal in addition to noise, i.e. hypothesis H_1. Suppose the signal is actually present; as the recipient of the message has no previous knowledge of this, he may, not having detected the signal, mistakenly reject the the correct hypothesis. On the other hand suppose there is no signal; the recipient having "read" the non-existing message may mistakenly take noise for signal. It is impossible to have the two probabilities of both errors small at the same time. It is natural therefore to fix the probability of mistaking noise for signal, which is the probability of error of the first kind, and to attempt to choose the critical region in such a manner that the probability of

loss of signal which is the error of second kind to be minimum. Neyman and Pearson deals with the solution of this problem.

We have seen that there are two hypotheses. The first is H_o which consists in assuming that sampling has been carried out from a general set with probability density $p(X; \alpha_o)$, i.e. that the parameter α which uniquely defines the probability density has the fixed value α_o. And the second is H_1, which assumes that the parameter is equal to α_1.

Let p_1 be the probability of error of the first kind. We shall define the critical region as the set of points X that satisfy the inequality

$$p(X; \alpha_1) \geqslant C_p(X; \alpha_o) \qquad 4.8$$

where C is a constant and is found from the condition that the probability p_1 be equal to its assigned value. It may be shown that such a choice of critical region is optimum in the sense that the probability of error of the second kind will be a minimum. Let us introduce the term

$$L(X) = \frac{p(X; \alpha_1)}{p(X; \alpha_o)} \qquad 4.9$$

This is known as likelihood ratio in statistics. Then according to the basic theorem of Neyman and Pearson's theory the optimum criterion for deciding between the two hypotheses will be expressed as

follows.

Assume $L(X) \geqslant C$

Then hypothesis H_1 is accepted. Hypothesis H_o is accepted otherwise. The probabilities of errors arising as a result of adopting this criterion are expressed by integrals

$$p_1 = \int_{L(X) \geqslant C} p(X;\ \alpha_o)\, dx \qquad 4.10$$

and

$$p_1 = \int_{L(X) < C} p(X;\ \alpha_1)\, dx \qquad 4.11$$

The domain of integration in each case is the subset of the corresponding sample space defined by one of the inequalities

$L(X) \geqslant C$, or $L(X) < C$.

The order of multiplicity of the integrals is equal to the number of diamensions of the random quantity X.

4.5 Applications in Our Problem

We shall now consider the application of Neyman and Pearson's theory to the problem of detecting a signal from a noise background.

The recipient of the message has n voltage amplitudes available. On the basis of these voltages, knowing the statistical characteristics of signals and noise, he has to establish whether signals are present or absent. The best method of doing this, according to Neyman and Pearson, consists of substituting the observed values of the amplitudes in the likelihood ratio and comparing its value with the threshold C . This means that the optimum receiver must comprise a device to accomplish on the input voltage the operations indicated by the functional form of the likelihood ratio.

As a result of using this method, for a given probability of error of first kind, which we will call "False alarm" probability, we seek to reduce the probability of error of second kind. We usually reckon not in terms of the latter probability but in terms of probability of the opposite event: the detection of the signal.

We will denote the probability of false alarm and the probability of detection by F and D respectively. They will be expressed in terms of the probability densities of noise $P_N(X)$ and of signal and noise combined $P_{SN}(X)$, in the following manner:

$$F = \int_{L(X) \geqslant C} P_N(X)\, dx \qquad 4.12$$

$$D = \int_{L(X) \geq C} P_{SN}(X)\, dx \qquad 4.13$$

The above procedure gives the highest value of the probability D for a given probability F.

The evaluation of these probabilities is absolutely essential for an assessment of optimum methods for extracting signal from noise.

The evaluation of F and D are not very often easy and the difficulties arise mainly from the complex structure of the region of integration which is defined as the set of points for which the likelihood ratio $L(X)$ is larger than the threshold C, as well as from the high order multiplicity of the integrals.

We will now illustrate typical features of the integrals expressing the probabilities F and D.

First of all we will isolate the parameter of the probability density, the values of which distinguish the various hypotheses from each other. In the case of a random signal we shall take the quantity q namely the ratio of the standard deviations of signal and noise, σ_s and σ_n. This is nothing but signal to noise ratio; if the signal is not random, the parameter will be s which is the ratio of the signal amplitude to standard deviation of noise. Then the hypothesis H_o consists in assuming that there is no signal

($q = 0$ or $S = 0$) and that the voltage observed is caused by noise alone. According to the alternative hypothesis there is a signal, which is characterised by a fixed value of a parameter: q_1 for random signal and s_1 for non-random one. In many cases the value of the parameter is unknown to the message recipient and therefore strictly speaking we have to test the simple hypothesis $H_o(q = 0)$ against the composite one $H_1(q > 0)$. Due to practical limitations this is not usually done. The theory of simple hypothesis testing is adequate and the following argument provides justification.

In the above simple case of hypothesis testing, the critical region is chosen to maximise D for a given value of F. The method of choosing the critical region will remain the same if the parameter of the alternative hypothesis, H_1, is changed, but for the same region and the same value of F, the value of D will be different.

Thus as q (or s for non-random case) approaches zero, corresponding to H_o, the value of D will, in general, get correspondingly smaller.

Conversely, for larger values of q, D will increase.

Let the minimum acceptable detection probability D_1 be assigned on the basis of various relevant considerations and let the corresponding value of q be q_1. Then for values of q, q_1 and

the assumption implied above that D is a monotonic function of q, D will be greater than D_1, for the same predetermined value of F.

Thus if the simple hypothesis testing procedure is adequate for $q = q_1$, the detection performance of the system will be superior for all $q > q_1$.

The above is an intuitive approach and does not provide a rigorous proof of the justification of using the theory of simple hypothesis testing. Since these are sufficient for most engineering purposes we accept this approach.

4.6 Sequential Observation Process

The concept of sequential observations developed by Wald(35) permits a decision to be withheld in doubtful cases, and new observations to be taken until unambiguous readings are obtained. In this system, the transmission is not fixed in advance but is determined. Recently Middleton(23) and various others(16) have studied the properties of sequential observation in various fields of communication. It is shown that the 'Sequential Observer' is, on the average, more efficient in saving time for a given reliability or increasing the reliability for a given time than the 'Neyman-Pearson observer which is an optimised single-threshold system. The result is intuitively reasonable because the sequential observer

reduces guessing which, as is well known, destroys information.

4.7 Summary and Conclusions

Thus in Neyman-Pearson method an observer, on the basis of the data received during a given time interval t, decides whether the signal is present, by using a rule which maximises the probability of correct decision D for a fixed false alarm probability F. It is proved in mathematical statistics that Neyman-Pearson observer makes his decision by using just the 'Single threshold' decision rule, here the threshold T is determined from a fixed value of F. Any other decision rule will lead to a smaller value of D (for fixed F and t).

Wald's sequential observer analyses the data continuously and reserves his decision until new data arrives so that his decision rule is completely different. However, the mathematical theory of the sequential observer is characterised by great complexity and hence, we will be using only the decision rule with one threshold, interpreting it in the spirit of the Neyman-Pearson method.

Note

The terms "Neyman-Pearson" method and "Wald's" method are not used here in the strict statistical sense.

5. PRELIMINARY EXPERIMENTS

5.1 Introduction

It was mentioned in Section 2 that scanning the H.F. spectrum was most suited for our purposes. In this Section we describe some of the preliminary experiments which were performed for detecting new signals by making use of scanning techniques.

We are basically interested in the detection and location of a new signal. The appearance of a new signal will be indicated by a change in the spectral display. It is therefore reasonable to assume that the appearance of a new signal will be indicated by a change in the energy of the spectrum. Thus an increase in the energy of the spectrum will indicate the appearance of a new signal and a decrease in energy will indicate the disappearance of a signal. Of course, noise and fading will introduce errors. This is dealt with in Section 6.

The experimental set up is shown in Figure 5.1. We choose a spectrum of bandwidth W_A and scan it in T seconds. At the end of T seconds, the panoramic adaptor returns to the starting point and scans the spectrum again. This process continues.

Now if we start integrating the power, which is proportional to the voltage squared, the output of the integrator at the end

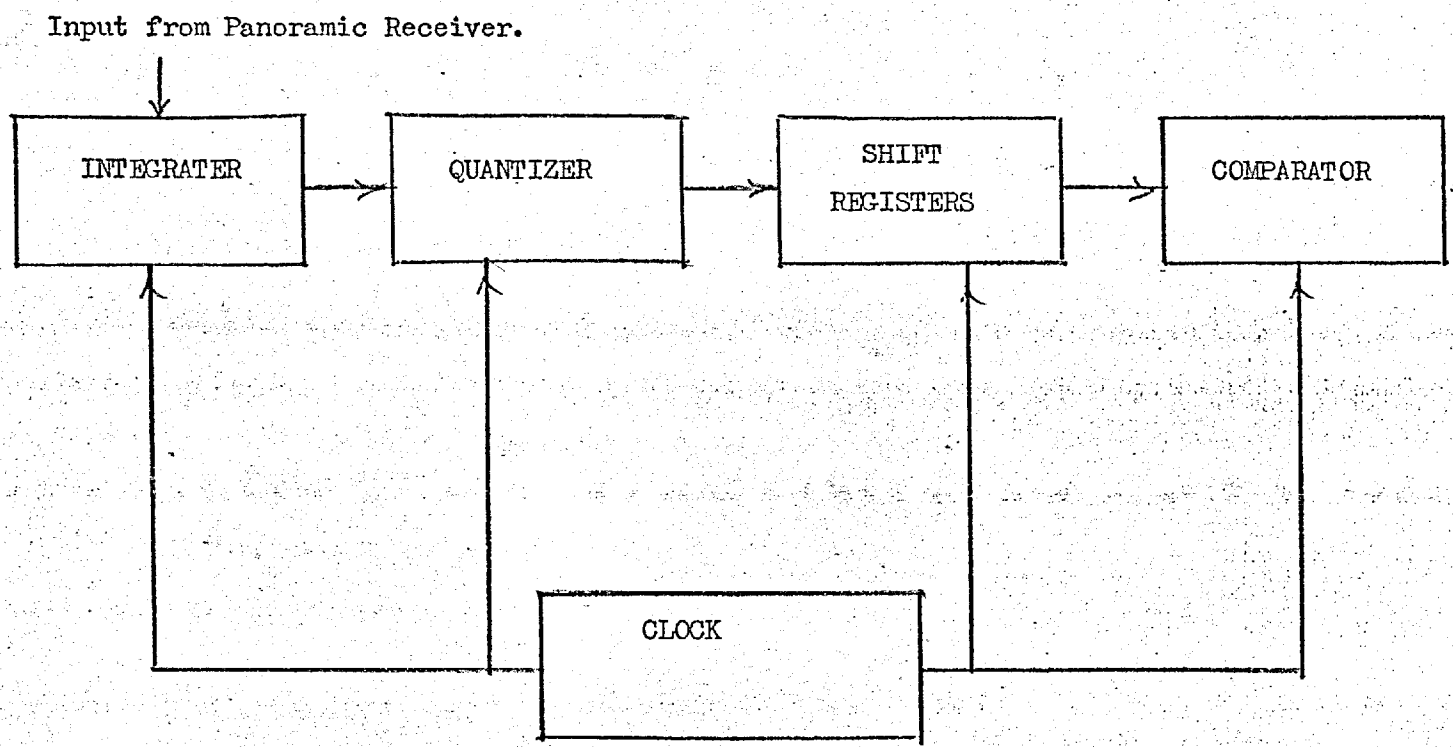

Input from Panoramic Receiver.
INTEGRATER
QUANTIZER
SHIFT
REGISTERS
COMPARATOR
CLOCK

FIG. 5.1

of T seconds represents a sample of the total energy in the spectrum W_A which was scanned during the time T.

If there are a number of signals in W_A, then in order to achieve greater accuracy in measuring the energy, we can subdivide the spectrum into equal portions, $w_1, w_2, \ldots\ldots\ldots\ldots, w_n$. We can also divide the total scanning time T equally into $t_1, t_2, \ldots\ldots\ldots, t_n$ such that t_1 is the time taken to scan w_1, t_2 to scan w_2 and so on. Let $t_1 = t_2 \ldots\ldots\ldots = t_n$.

From the diagram 5.2, it is easily seen that the energy contained in any one of the sub-spectrums, say w_i, at the end of scanning period t_i, can be written as,

$$E_i(t_i) = \int P\, dt$$

where P represents the power in w_i.

The change in the energy of the sub-spectrum w_i after two successive scans can be represented by

$$E_i(t_i + T) - E_i(t_i)$$

which is the quantity of interest in our experiments.

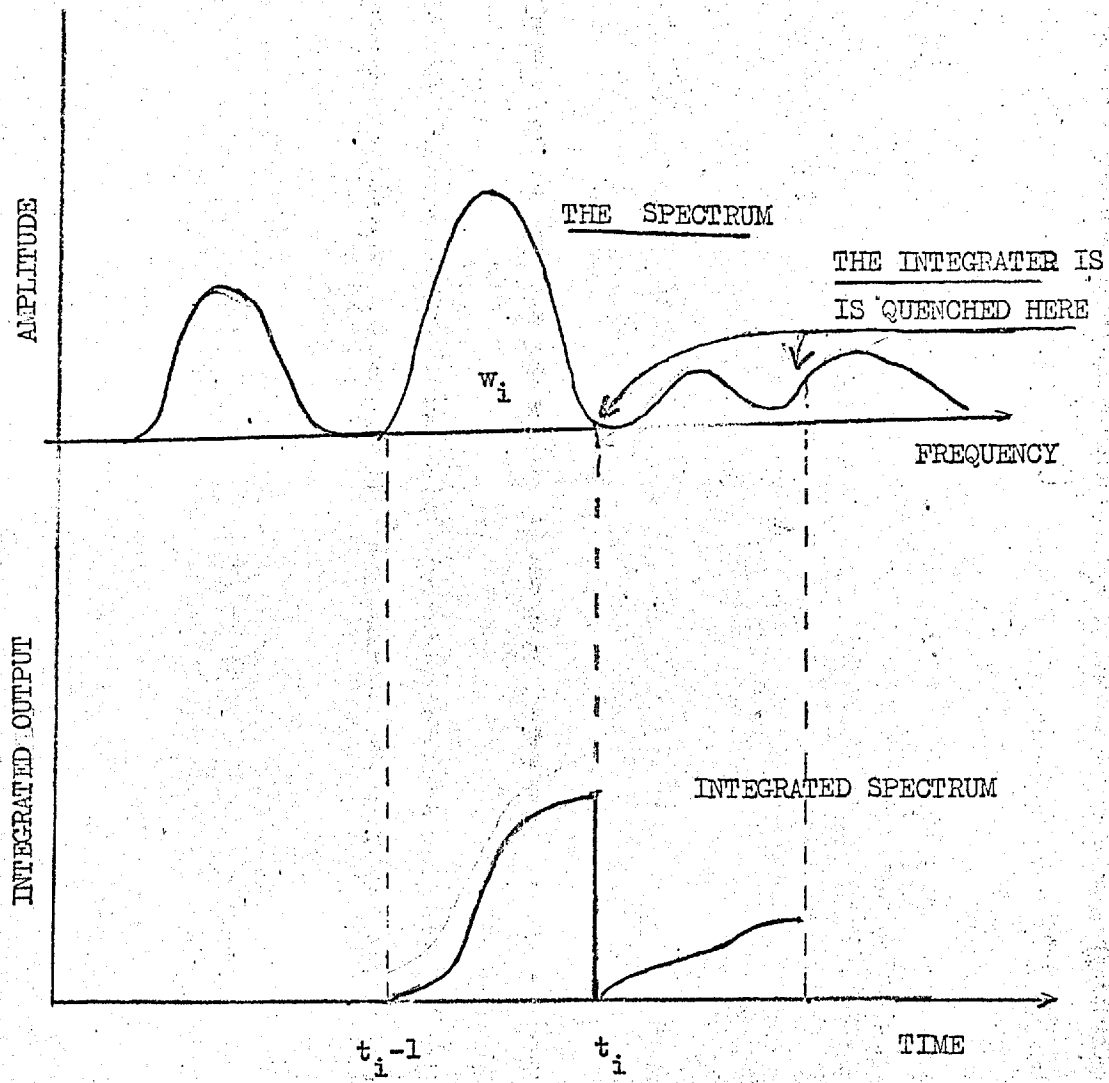

AMPLITUDE
THE SPECTRUM
THE INTEGRATER IS
IS QUENCHED HERE
w_i
FREQUENCY
INTEGRATED OUTPUT
INTEGRATED SPECTRUM
t_i-1
t_i
TIME

FIG. 5.2.

5.2 Circuit Arrangements

We can arrange the circuits so that after completing the scan of bandwidth w_i, the integrator quenches instantaneously. The next band w_2 is integrated for a period t_2 and the integrator is quenched again and so on until the whole band W_A is scanned after which the cycle repeats itself. After the energy has been integrated it is quantised, converted into 3 binary digits using a decimal to binary converter and, at the end of period t_1, the three digits, representing the energy in w_1, are stored.

The arrangement for storing was as follows. 33 shift register stages were arranged in three rows so that each column consisted of 3 stages. The 11 shift register stages in each row were inter-connected so that the digits in the store could be shifted from register no. 1 to register no. 11 successively under the influence of clock pulses.

The 3 digits from w_1 were stored in the three stages of column 1 until the digits representing the energy of bandwidth w_2 were ready to be stored. The digits already stored there were then shifted to column 2. When the digits from w_3 were ready to be stored, the digits in columns 1 and 2 were shifted to columns 2 and 3 respectively. This process continued until scanning of w_1 was completed for a second time. The energy in columns

1 and 11 then represented the samples of energy of the same spectrum, but T seconds apart. A comparator was used to observe and change in energy.

5.3 Setting up of the Equipment

We now describe the various steps involved in setting up the equipment. A variable D.C. voltage supply source was used for this purpose. The voltage source was variable from 0 - 5 volts and was fed into the input terminals of a potential divider. The potential divider had eight output terminals each of which was connected to a decision circuit. The decision circuits consisted of conventional Schmitt trigger circuits. Eight levels in steps of 0.5 volts needed for 3 binary digit outputs was thus set. Finally the equipment performance was tested as follows.

5.4 Preliminary Tests

For various test purposes, a specially designed spectrum generator was used. This was capable of supplying different frequencies within 1 - 2 Mc/s the amplitudes of which could be varied. The spectrum generator was adjusted to give an output of 3μ volts and this was fed to the input of the R.A. 17 receiver.

The panoramic adaptor was adjusted to scan 100 Kc/s in 0.2 seconds. The sweep bandwidth was adjusted for 1.2 Kc/s. Thus with the help of the circuit arrangement shown in Figure 5.1, it was possible to measure the energy of the spectrum in blocks of 10. The synchronising clock pulses were obtained from the panoramic adaptor.

The general performance of the equipment such as the shifting of digits under the control of clock pulses, decimal to binary conversion, the proper functioning of the comparator, etc. were tested with various input levels at the receiver. Some tests were also carried out with actual signals from an aerial.

5.5 Results and Discussions

The above tests, limited as they were, clearly showed that the equipment was functioning properly, as expected, and the technique for measuring the energy was quite adequate under ideal conditions.

However, the tests also showed some limitations of the method. It is well known that a two level signal is less subject to deterioration, or corruption by noise and fading, than a multi-level signal. This was evident during those tests with real signals from the aerial when the false alarm rate was fairly high. Since the main interest here is to know if there is a change in energy,

and not by how much the energy has changed; multi-level quantization of the signal is not of any special advantage. Instead, using a two level signal and a threshold system is more advantageous because it is economical, more efficient and above all much simpler.

We thus conclude that when we are looking for a signal of unknown frequency and waveform, and also when the time of arrival of the signal is not specified, the only way of tackling the problem appears to be one of making observations at the receiving end for changes in energy level against a predetermined threshold level. On the basis of these arguments, the following threshold system is proposed. It is simpler from an operational point of view and the basic units, which are transistorised, have been proved very reliable.

5.6 Threshold System

The simplified block diagram of the threshold system showing the inter-connections of various units is presented in Figure 5.3.

Basically the system consists mainly of decision circuits, monostable circuits, integrators and other auxiliary units. These are arranged as shown in the diagram. A bandwidth of W_A is swept in time T, in steps of equal bandwidths w_i, t_i being

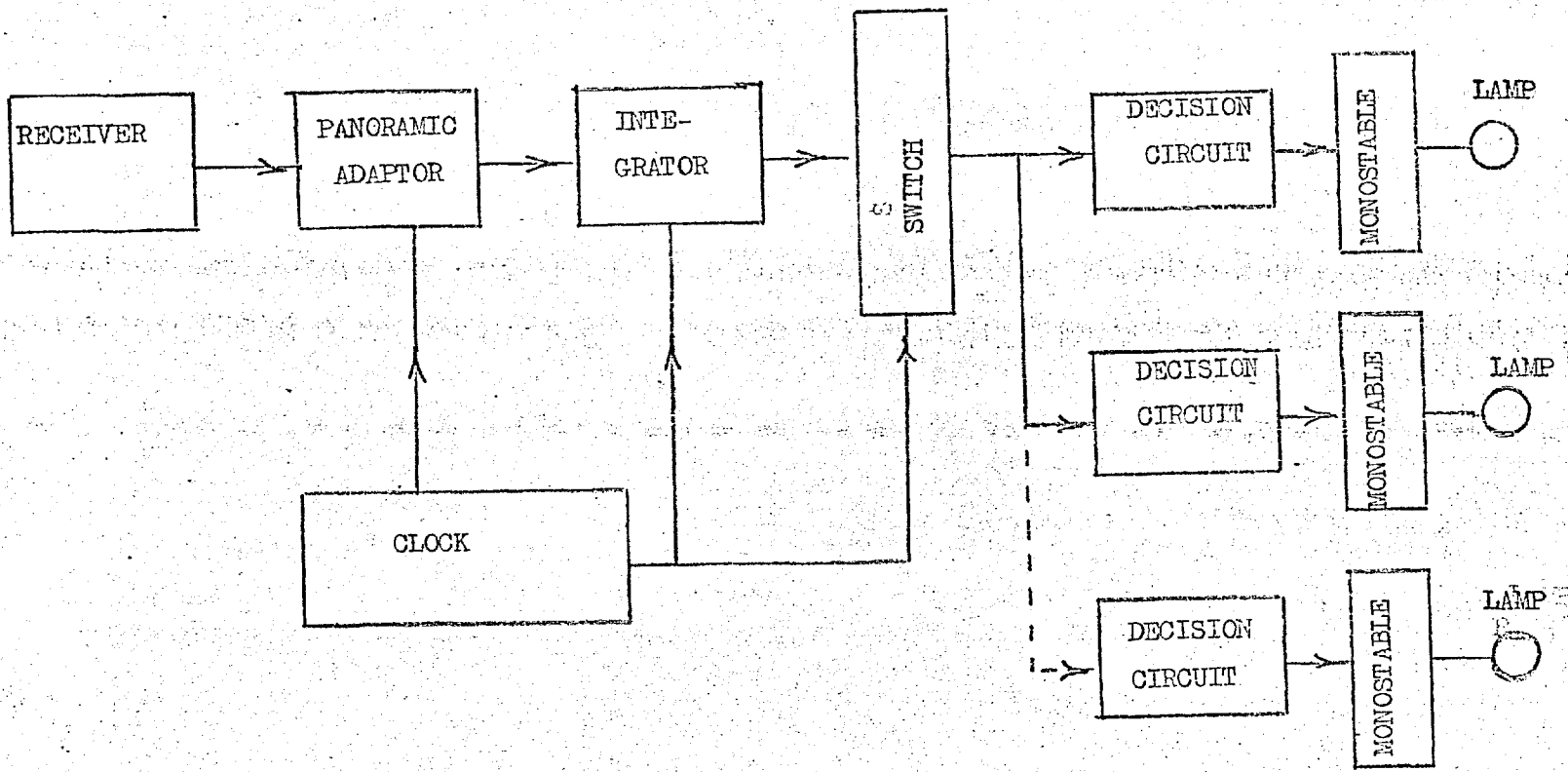

RECEIVER
PANORAMIC ADAPTOR
INTE-GRATOR
S SWITCH
DECISION CIRCUIT
MONOSTABLE
LAMP
DECISION CIRCUIT
MONOSTABLE
LAMP
DECISION CIRCUIT
MONOSTABLE
LAMP
CLOCK

FIG. 5.3.

the time taken to sweep the bandwidth w_i. This is achieved in the same manner as described in Section 5.2. The main change here is that after the completion of each sweep period t_i, the integral of the signal voltage which represents the energy in the spectrum w_i, is passed to a decision circuit. The thresholds for each of the decision circuits are preset to values slightly above the average values of the signal voltages in each w_i. (The method for setting these thresholds is described in Section 7). Thus the action of the threshold system is as follows : any voltage exceeding the threshold level T_i gives rise to a standard unit signal, while a voltage smaller than the level T_i is not recorded, that is, gives a 'zero' signal. In this manner, the signal as a result of passage through a threshold system, may be said to be 'quantized' in that it assumes only two values, which can be termed the 'one' and 'zero' values.

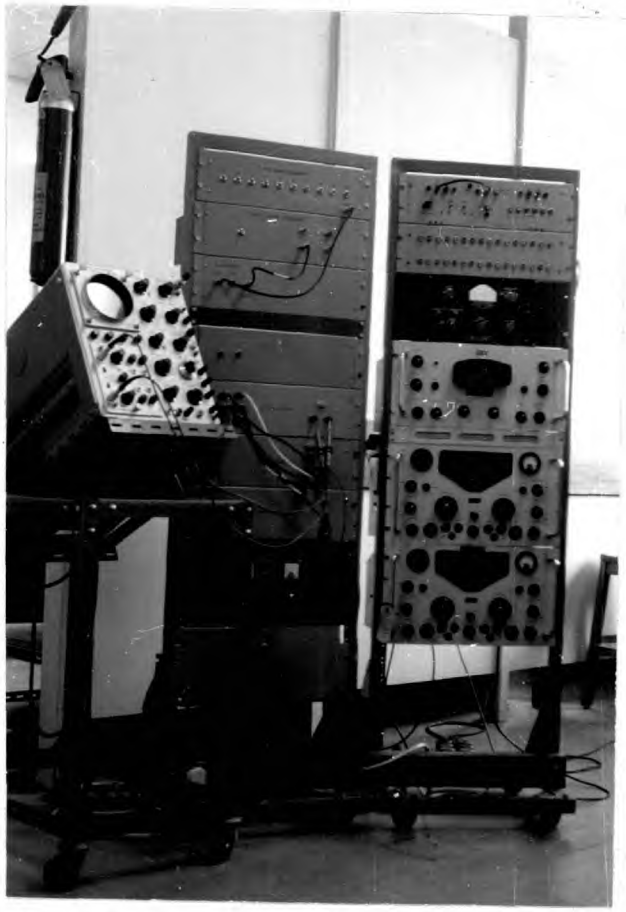
The outputs of the decision circuits are then fed into separate monostable circuits each of which include a circuit arrangement for lighting a lamp whenever a 'one' appears at the various outputs of the decision circuits. Thus the appearance of a new signal results in a train of pulses which last for the duration of the signal and cause the lamp to light for this period. The human observer can then tune to w_i to extract the signal.

In the actual experiments performed W_A was divided into

10 and the time of sweep selected was 0.2 seconds. There were provisions for changing either or both of these.

The procedure for setting the threshold levels is described in Section 7. This is one in which considerable latitude exists. Fading and noise are the most important factors which affect the threshold setting. These are considered in detail in the next Section.

The details of the above system and the design considerations are described in Appendix 2.



GNENERAL VIEW OF THE THRESHOLD SYSTEM

FIG. 5.4.

6. EFFECTS OF FADING AND NOISE

6.1 General Discussion

Communication systems of all types can be reduced to the single diagram shown in Figure 6.1. The transmission of useful information, by which we mean not only explicit messages but also any information of interest to the recipient, such as the presence or absence of a signal, is accomplished by means of electromagnetic or acoustical waves or by making use of other physical phenomena. Whatever the chosen physical carriers of information, at the transmitting end of the communication line, suitable signals are transmitted along the communication channel. The types of signals are varied, from a single frequency to complex ones, having arbitrarily wide and complicated spectra. As these signals pass through the communication channel the useful signals are open to corruption by noise, particularly interference from other radio signals and other forms of distortion. The signals therefore suffer various forms of degradations at the receiving end. The two main types of degradation are:-

1. Various types of fading.
2. Additive noise.

Some disturbance of the signal is unavoidable and economic design of the equipment aims only at reducing it to a tolerable proportion.

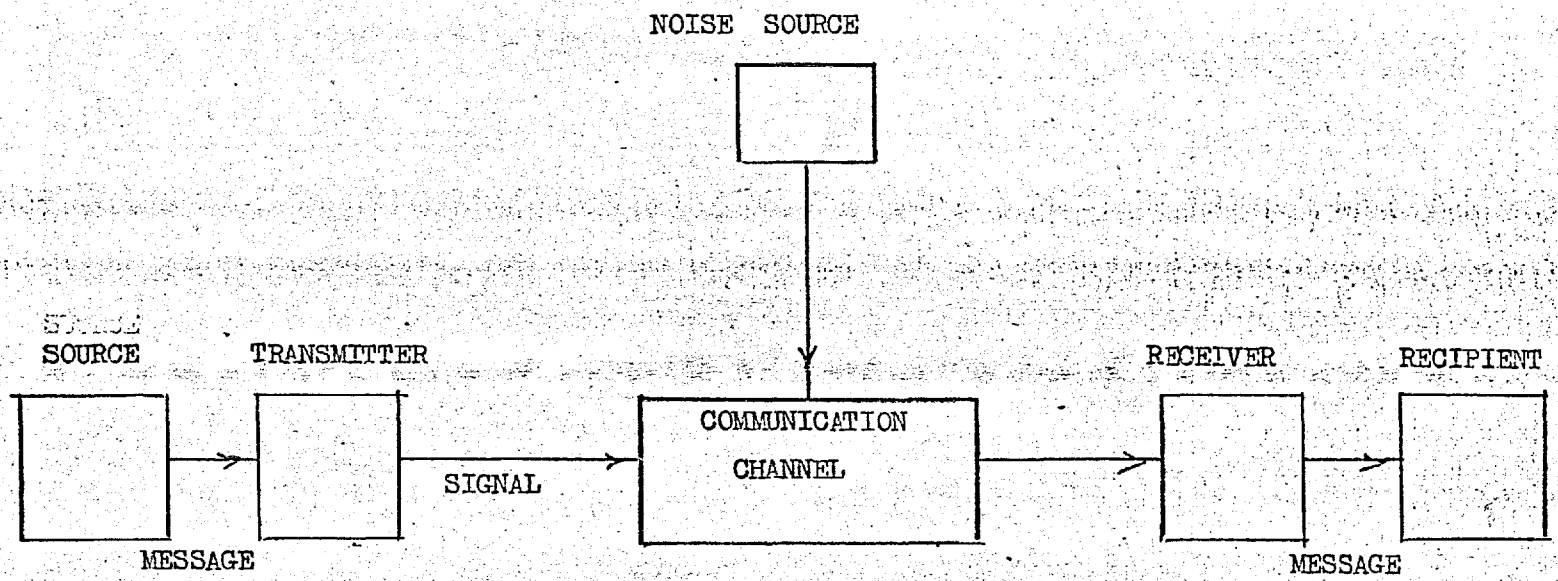

NOISE SOURCE
SOURCE
TRANSMITTER
SIGNAL
COMMUNICATION
CHANNEL
RECEIVER
RECIPIENT
MESSAGE
MESSAGE

FIG. 6.1.

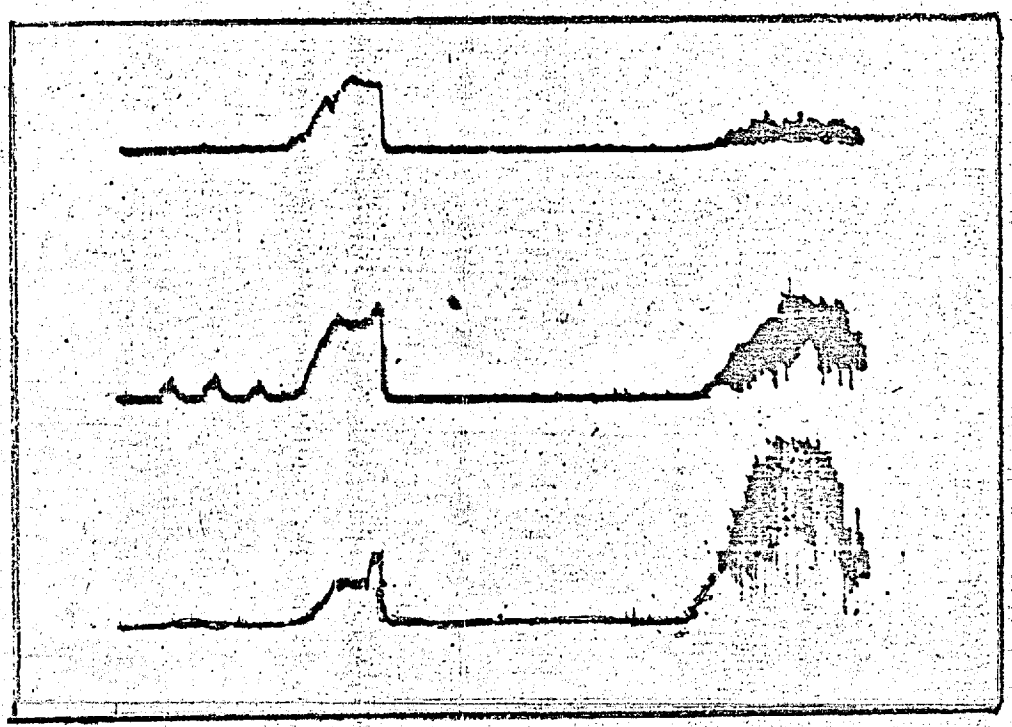
What is tolerable, of course, depends upon the use made of the received signal. Tolerances of noise, distortion and interference have been determined by operating experience or subjective tests for various signals and sources of distortion and the data on this is readily available.

It is well known that a fading signal is usually characterised by rapid amplitude variations. These variations complicate the measurement of energy in the H.F. spectrum.

Noise is another cause of error in our system. Even when the signal is powerful enough to make noise negligible, fading introduces errors. When both fading and noise are present, they are interrelated; for, combating of fading by AGC is less successful the greater the noise intensity.

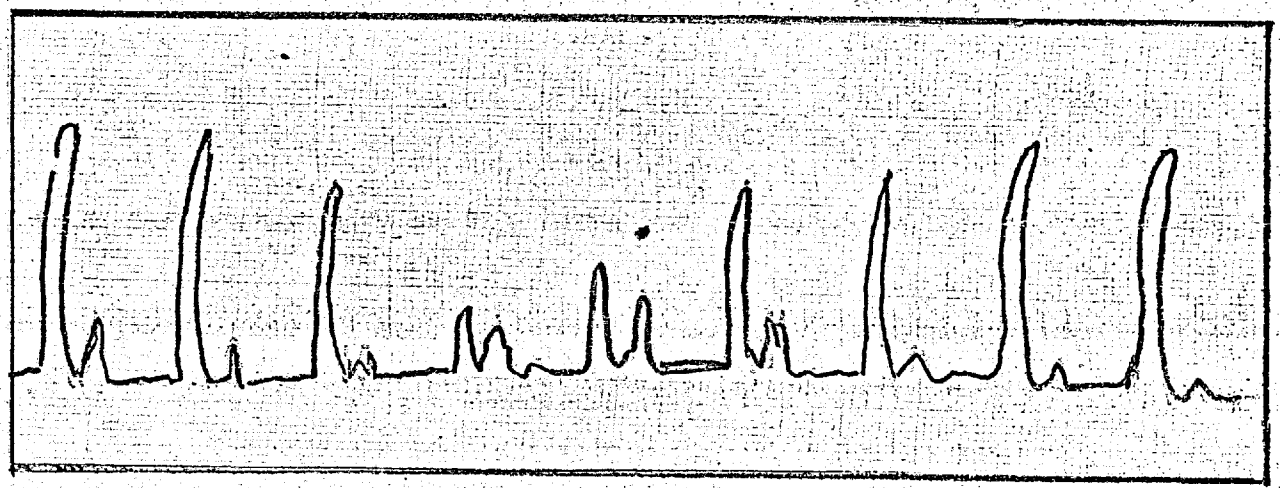
When continuous wave high frequency transmissions within the H.F. band are closely examined, it is found that signals fade differently throughout the H.F. band. This phenomenon known as selective fading is most troublesome in ordinary double side band transmissions. When selective fading occurs, it causes variations in the H.F. spectrum resulting sometimes in the complete loss of the signals. These are illustrated by Figure 6.2, 6.3 and 6.4. Figure 6.5 shows a steady spectrum.

The effect of fading has been observed for many years and it can be partially eliminated by adopting various well-known techniques.



AN EXAMPLE OF A FADING SPECTRUM

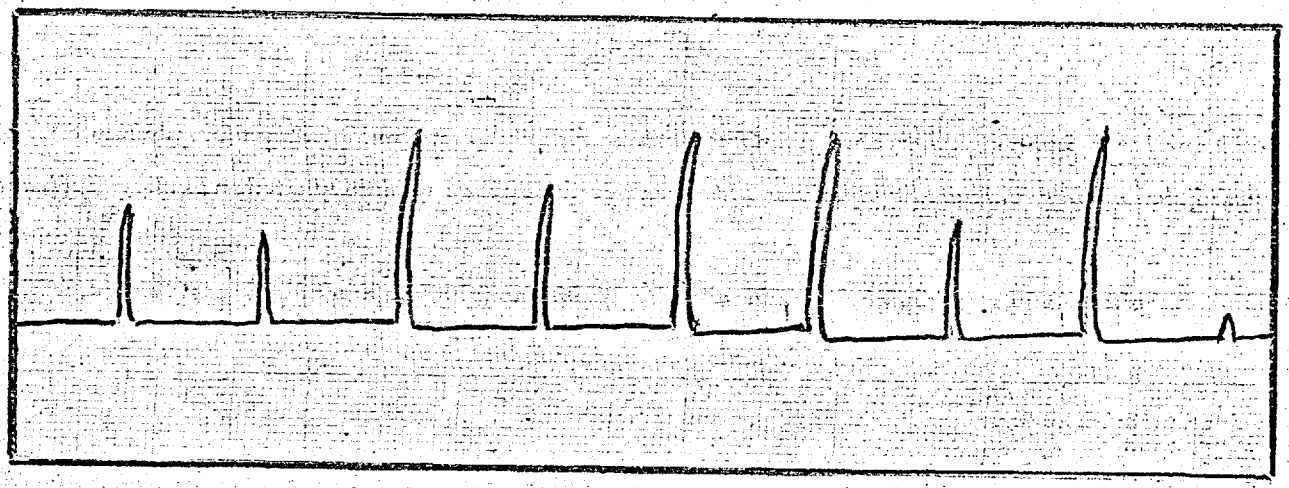
FIG. 6.2.



AN EXAMPLE OF A FADING SPECTRUM

(4.7 Mc/s)

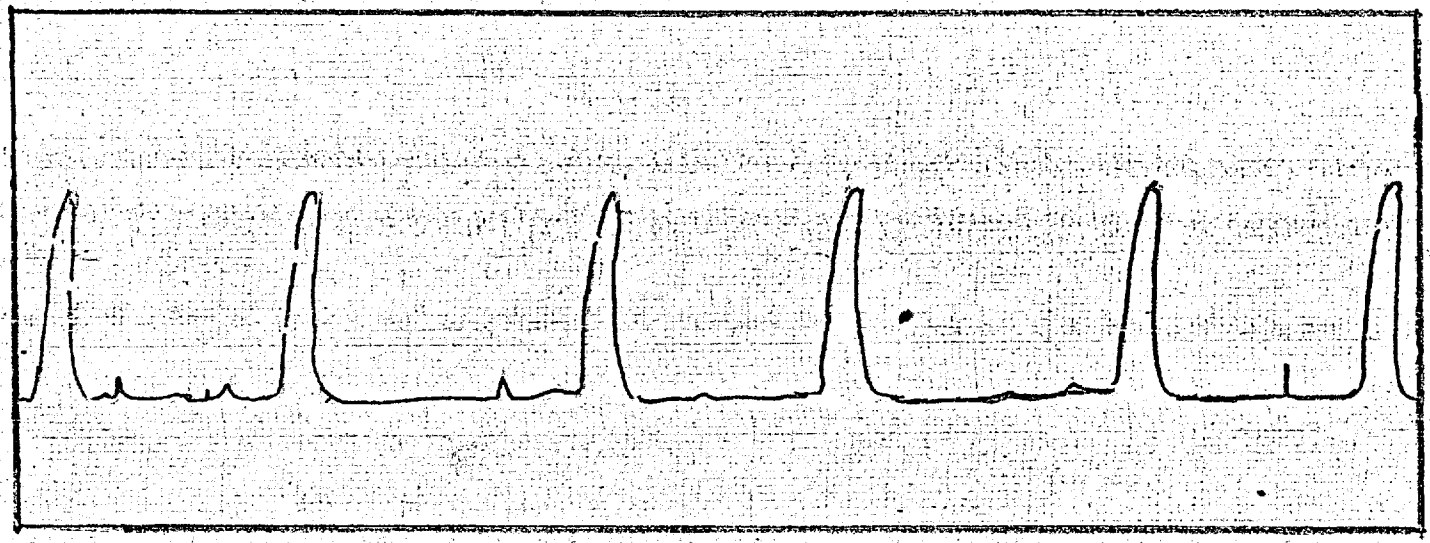
FIG. 6.3.



AN EXAMPLE OF A FADING SPECTRUM

19 Mc/s

FIG.6.4.



AN EXAMPLE OF A STEADY SPECTRUM.

FIG. 6.5.

Nevertheless, fading is still a basic problem in international broadcasting and other A.M. services and one that cannot be alleviated by simple expedients such as increasing transmitter power.

6.2 Subjective Experiments

A study of the main effects of fading was found to be very important in our study of H.F. spectrum and therefore an attempt was made to observe the effects systematically. The preliminary investigations were made by listening to the quality of various broadcasts, and observing the fading spectrum on the screen of the panoramic adaptor.

These tests were carried out during various hours of day and night and lasted over some two months. The investigations covered a wide frequency range. The method was subjective and did not give accurate quantitative results.

6.3 Preliminary Results

Among the various observations such as deterioration of the quality of both and music transmission, the following important and relevant points were also noted:

1. The fading though slow sometimes, was very deep and rapid

most of the time.

2. Frequently the fading was not particularly audible, but was readily seen on a CRO screen. These subjective tests also showed that certain signals suffer more from the above effect than equally strong signals from other directions.

The following distinct types of fading were observed:

a) Rapid fluctuations in the instantaneous signal strength. These fluctuations are caused by interference with the same signal arriving at the aerial by different paths. The effect is called multipath interference fading. This type of fading very often leads to complete loss of signal during time intervals that are long even when compared with the slowest components of the signal.

b) The instantaneous fluctuations in signal strength occur about a mean value of signal amplitude that changes relatively so slowly that its values must be compared at instants separated by minutes to hours before any significant difference can be perceived.

c) A third type of fading that is much slower than that just described was also noted. The 'hourly' mean signal levels are different from day to day just as they are from hour to hour in any one day. The mean

signal energy level changes from day to day and from month to month. Besides this, there are also seasonal variations in the mean signal level.

The last two types of fading do not affect our immediate results and therefore attention was devoted exclusively to study the first type of fading.

6.4 Noise - Its Sources

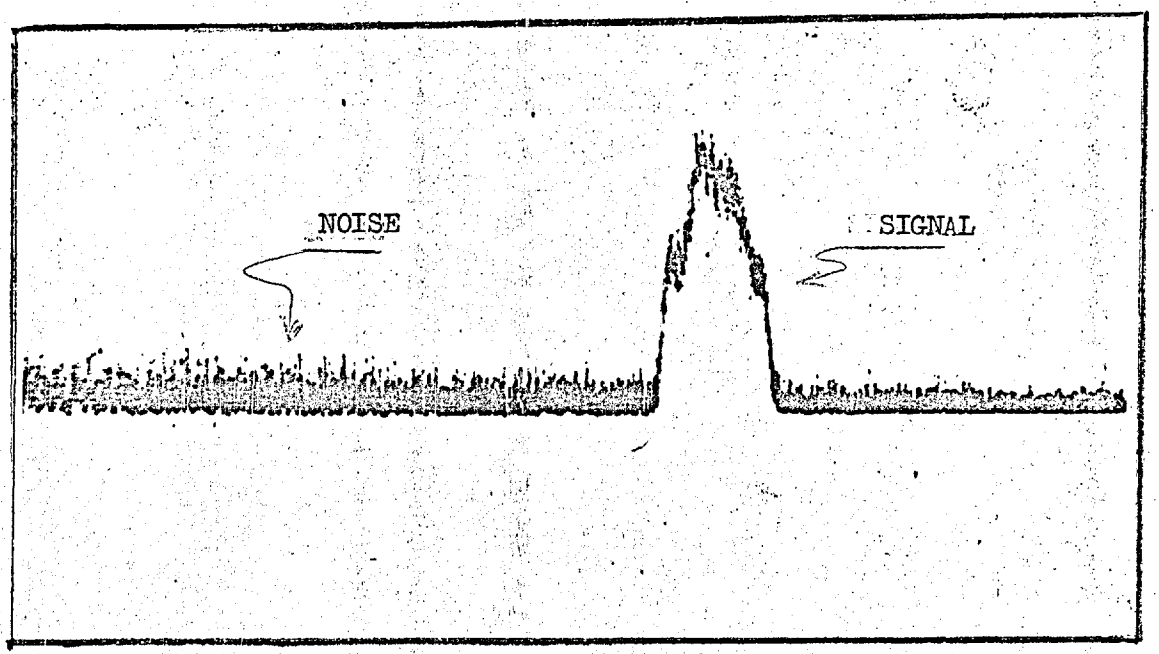
Certain sources of noise are peculiar to radio systems. Atmospherics and extraterrestrial noise[15] are examples of the above. Atmospherics are a major factor at low radio frequencies, but above 30 Mc/s their effect is generally lower than that of the noise generated in the radio receiver. This is partly accounted for by the fact that atmospheric pulses are relatively long with an energy distribution which decreases at higher frequencies, and partly because the characteristics of radio wave propagation limit the reception of the higher-frequency atmospherics from distant storms. Thunder-storms are very frequent in tropical regions and atmospheric noise generally decreases with increasing latitude; it also varies with time of the day and season. Some data are available showing the estimated levels of atmospheric noise[33], but the data on

which they are based are meagre[32]. Background atmospherics have the general character of random noise, with a peak amplitude proportional to the square root of the receiver bandwidth. Atmospherics, however, in general, closely resemble impulsive interference.

Extraterrestrial noise, which is mainly significant at frequencies between about 20 and 100 Mc/s, has its origin in astronomical sources which include the sun, stars, interstellar gas and some invisible concentrated sources called 'radio stars'[29]. The magnitude of such a noise source may be expressed in terms of an equivalent thermal noise voltage by allotting an effective 'noise temperature' to the aerial.

Interference from electrical machinery cannot in practice be completely suppressed[37]. Such interference, which is similar to impulse noise, differs from random noise in that the components of its spectrum are coherent; its peak amplitude is directly proportional to the receiver bandwidth, and not to its square root as in the case of background atmospherics. Usually impulse noise decreases with increasing frequency, although not always so rapidly as if its voltage were inversely proportional to frequency.

Figure 6.6 illustrates the noise in a spectrum of 10 Kc/s bandwidth.



AN EXAMPLE OF A NOISY 10 Kc/s SPECTRUM

FIG. 6.6.

6.5 Effects of Noise

The performance of a radio system depends upon the signal to noise ratio at the input of the receiver and the minimum acceptable ratio depends greatly upon the type of service, for example, telegraph, telephone, broadcast relay, and upon multiplexing and modulation techniques. Bandwidth noise exchange can be exploited by suitable methods of coding or modulation according to the well known formula[31] -

$$C = W \log_2 (1 + P/N) \text{ bits/sec}$$

where

C = Channel capacity

W = Bandwidth of the signal

and P and N are the average signal and noise powers respectively.

In almost all modulation systems there is a critical input signal to noise ratio (improvement threshold) below which the output signal to noise ratio falls rapidly. Thus frequency modulated systems can give a much reduced output noise level compared with amplitude modulated systems for a given input, but only for signals that exceed the peak noise level.

In the case of double-side band amplitude modulation, when

the noise level is high enough for the carrier to be fully modulated for most of the time by noise, the output signal level is reduced[13]. Similar considerations apply to systems using pulse modulation. On the other hand, single-side band amplitude modulation systems have no improvement thresholds, because the strong local carrier supplied to the detector provides a standard of coherence for demodulating the sidebands.

Of prime importance in any discussion about sources of error is the form of the interfering noise. Generally speaking, noise in the communication channel can be divided into two categories:

a) White Noise

b) Impulse Noise

Of course, both of these are band-limited in real channels. White noise is characterised by a flat frequency spectrum with equal contributions from all frequencies and with the various components having random phases. For calculation purposes it is generally sufficient to know only its average power and that the noise is 'white'. On the other hand, the term 'impulse noise' is not so clearly defined. Conceptually, impulse noise is made up of a collection of randomly sized impulses. However, by convention, impulse noise has come to mean the noise which reaches the receiver as a result of the presence of various short disturbances

in the communication path.

These two types of noise vary in their importance from one communication medium to another. White noise is negligible on wire circuits, while on radio circuits, particularly during a fade, it may be a most serious source of difficulty. Impulse noise is the greater problem on wire circuits.

Since the exact statistical form of white noise is known its ideal effect on a communication signal is fairly simple to compute. On the other hand, a similarly accurate description of impulse noise is not yet available. This is partially because not as much data has been recorded for impulse noise but it is probably more attributable to the way the noise is categorised. Although it is assumed that all such noise originate as sharp spikes, other short-time phenomena also cause the prime disturbances which appear in the circuit as 'impulse noise'. The result of this is that accurate statistical descriptions of impulse noise are not available and, even when they become so, they will probably be so complex that they will be difficult to use.

Even though accurate descriptions are not available, a number of interesting and useful facts are known about impulse noise in voice frequency telephone networks. In the first place, they have very low average power on the whole. Secondly, the noise

which appears will be correlated with itself over a time at least of the order of the reciprocal of the noise bandwidth and often for considerably longer than this. Finally, the noise often reaches very large values, when it occurs, so that errors are very likely to appear at that time, no matter how good the filter and detection process.

6.6 Experimental Technique for Evaluation of Fading

The R.A. 17 receiver along with the panoramic adaptor was used for this purpose. A pen recorder was connected to the specially adapted panoramic adaptor. A selected spectrum containing a broadcasting station was then recorded continually. Each scan lasted 0.2 seconds and continuous recording of the spectrum was made for 2 - 3 minutes. This was repeated every 15 minutes. These tests were carried out for 8 - 10 hours both during day and night.

The variations in the amplitude of a particular broadcasting station were then studied; usually the variation in the amplitude was measured every 0.2 sec.

6.7 Results and Conclusions

The measurements of the amplitudes of the fading signal for a period of 15 minutes were taken as a block. The mean value and the standard deviation was evaluated for each of the blocks. Each of the values of standard deviation was expressed as a percentage of the mean value about which it was calculated.

The values so obtained are shown in Figures 6.7 and 6.8 for one typical day and no substantial variation of the mean level is apparent throughout the day.

For the experiments carried out, the period of rapid fading was found to be 0.02 - 10 seconds, whereas that of slow fading was seen to be between 10 minutes to 1 hour. Both were found to be random in character.

McNicol[26] who has done some early work in this field, has derived an expression for the amplitude for the rapidly fading signal. However, we are interested only in determining a safe threshold for our equipment and this was achieved.

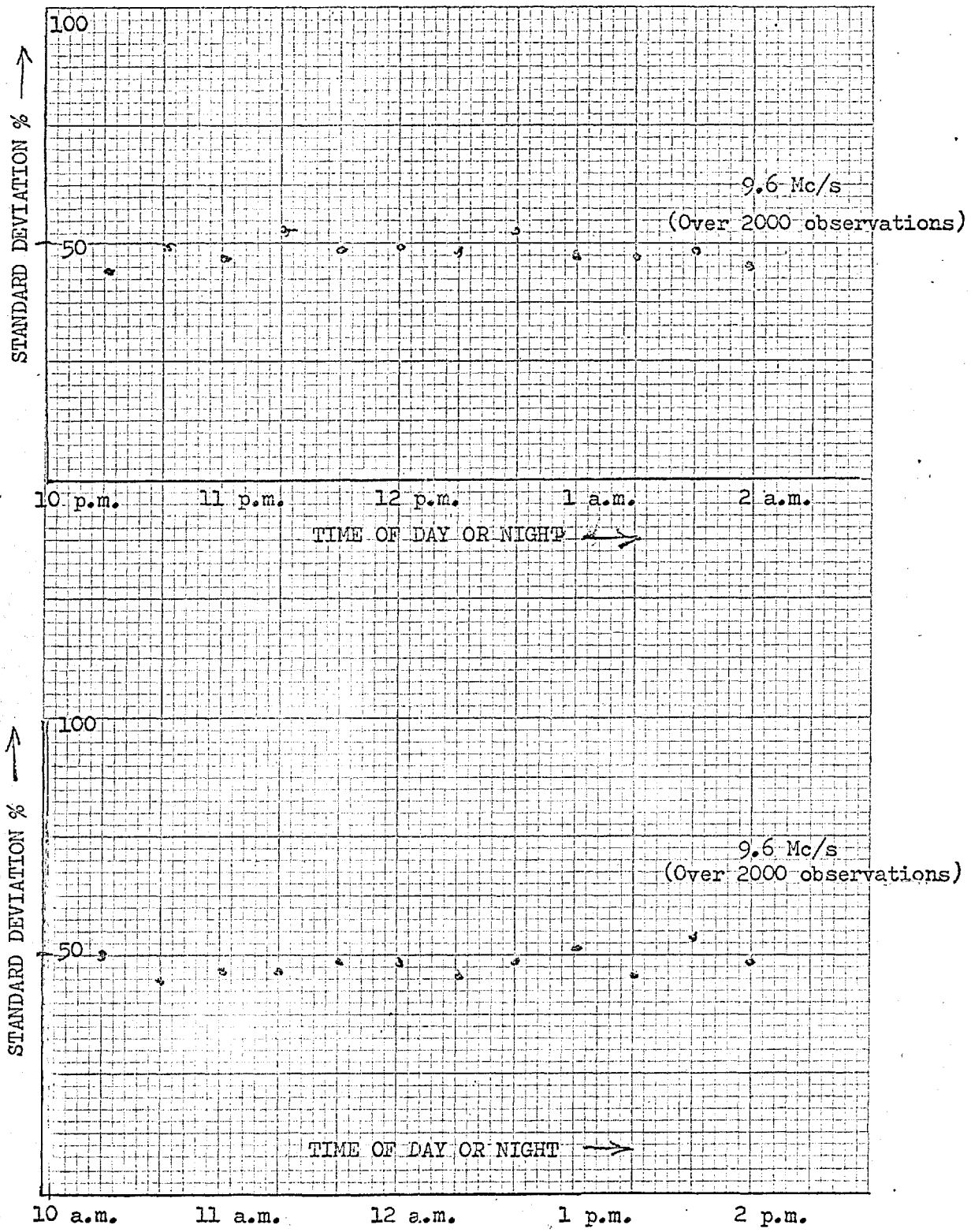

100
STANDARD DEVIATION %
9.6 Mc/s
(Over 2000 observations)
50
10 p.m.
11 p.m.
12 p.m.
1 a.m.
2 a.m.
TIME OF DAY OR NIGHT
100
STANDARD DEVIATION %
9.6 Mc/s
(Over 2000 observations)
50
TIME OF DAY OR NIGHT
10 a.m.
11 a.m.
12 a.m.
1 p.m.
2 p.m.

FIG. 6.7

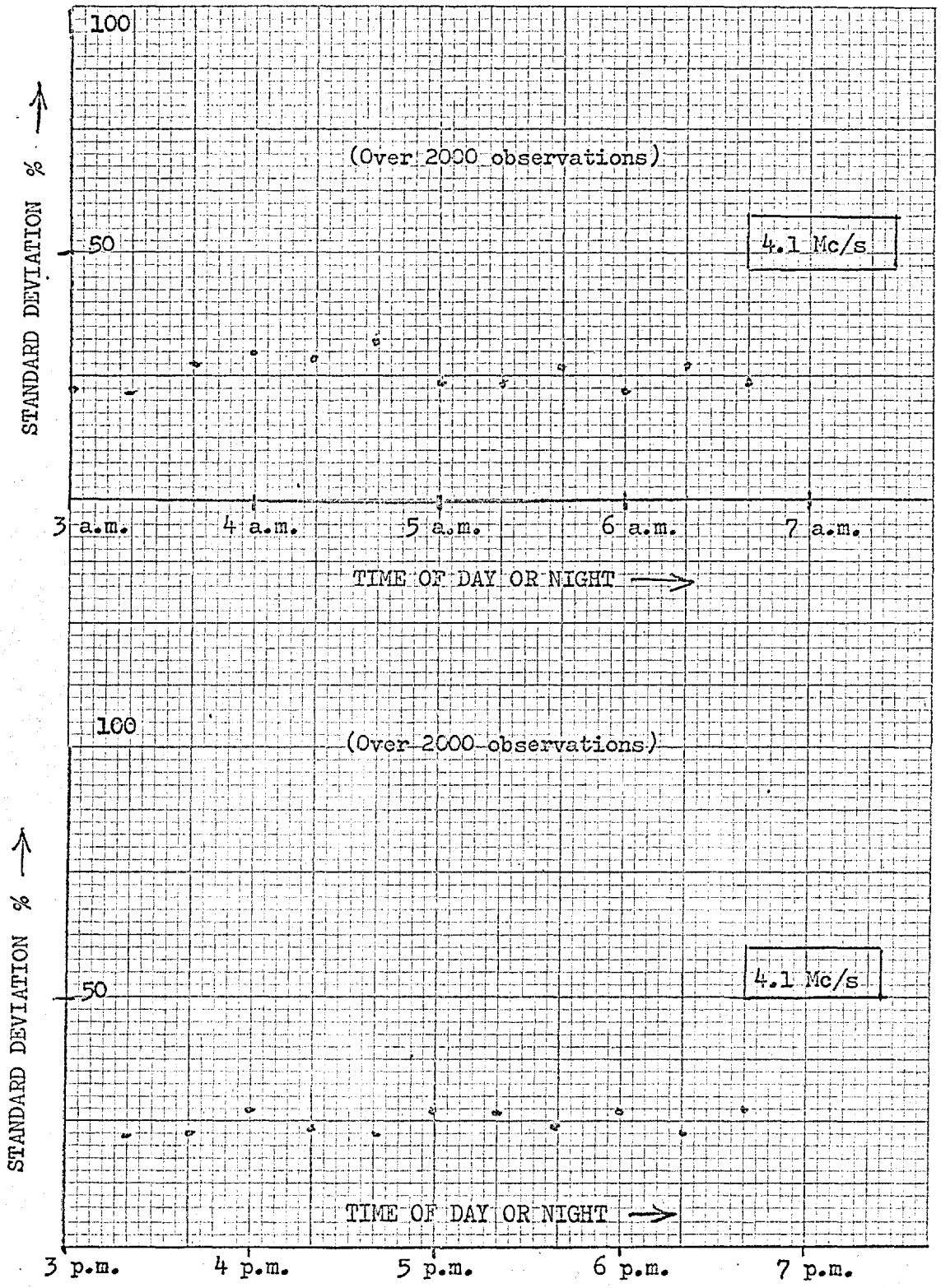

100
STANDARD DEVIATION %
(Over 2000 observations)
50
4.1 Mc/s
3 a.m.
4 a.m.
5 a.m.
6 a.m.
7 a.m.
TIME OF DAY OR NIGHT
100
(Over 2000 observations)
STANDARD DEVIATION %
50
4.1 Mc/s
TIME OF DAY OR NIGHT
3 p.m.
4 p.m.
5 p.m.
6 p.m.
7 p.m.

FIG. 6.8

7. MAIN EXPERIMENTS

7.1 Introduction

We begin this Section with a detailed description of the techniques adopted for fixing the threshold level. The latter part of this Section presents an account of a limited experimental investigation for the evaluation of the present system.

It was mentioned in Section 5 that the procedure by which the threshold level is set is one in which considerable range exists. The main factors which greatly influence the threshold setting are noise, fading, and, above all, the degree of accuracy that is required.

It may be worth mentioning at the outset that the experiments carried out in Sections 5 and 6 and the observations made were only meant for exploring a reliable method of setting a threshold level under extreme conditions of fading and noise. To minimise the effects of some of the above disturbances, some important channels are provided with highly directional diversity aerials and elaborate receivers; even so, these are at times very badly affected. The experiments reported here were carried out using only a vertical aerial.

7.2 Determination of Threshold

As mentioned above, the experiments carried out in Sections 5 and 6 were mainly for gaining a better understanding of the minimum signal energy that can be detected and to find a means of fixing a threshold for detection.

The basic problem of fixing a minimum threshold becomes a complex one when noise and fading occur and an elegant theoretical analysis is seldom possible. When a human observer has to judge visually or aurally whether a signal is present or not, it is clear that some of the psycho-physical properties of the eye or ear are bound to influence the signal threshold. For example in the visual observation of a signal on a radar screen enough light must be produced on the screen to make the display visible. In other words there is a brightness limit for the visual detection of the signal. Moreover, there is a contrast limitation. Also there is a minimum signal level, which varies from person to person, below which one cannot hear or see. Such limitations are very often very important although little can be said theoretically about these limitations.

However, when the function of the observer is reduced to counting or measuring, then the question of a minimum detectable signal becomes a definite statistical problem for which a mathematical or theoretical analysis can be attempted. The success of

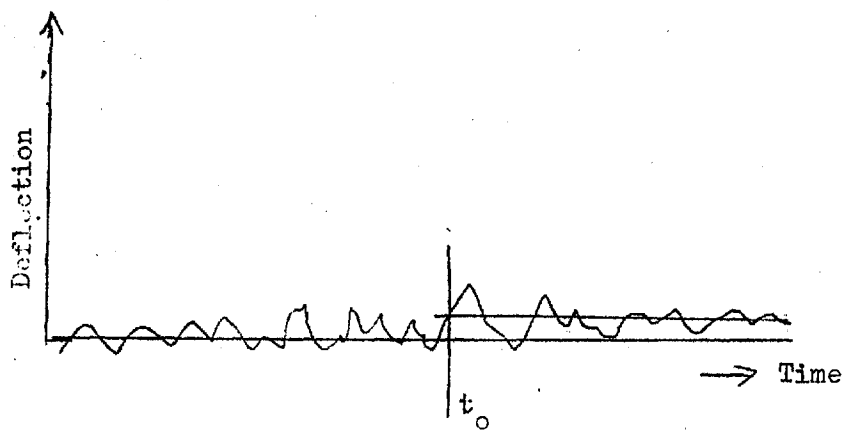
such an analysis has shown that for a wide range of experimental conditions, the essential limitation for detectability of a signal is due to the statistical nature of the problem.

The problem now arises as to how to define a minimum detectable signal or signal threshold. Here we have to remember that we can speak of a minimum detectable signal only when,

a) The number of observations are limited

b) The time of observation is limited.

Let us consider for example the simple case of detecting a small current by means of a sensitive galvanometer. Due to the thermal agitation of the mirror of the galvanometer will have an unavoidable Brownian motion around its equilibrium position.



Therefore it might seem that when a current at $t = t_0$ produces a deflection small compared with the mean 'jitter' of the mirror, such a current could not be detected. This is true only when the number of observations are limited. When the time is not limited, the average position and therefore also a (slow) change of average position of the mirror can be determined as accurately as desired.

Strictly speaking, therefore, the noise or thermal motion does not limit the detectability of a signal. It is only when there is a limited observation time for small signals that we have to make a guess as to whether the signal is present or not and there is a definite probability that the guess is right. Of course, for increasing signal strength this probability will rapidly increase and approach unity. We can therefore safely make the following definition.

The minimum detectable signal is that signal for which the probability of correctly guessing its presence or absence is, say, 90%; this means that on an average the guess will have proved right in nine out of ten cases. The minimum detectable signal defined in this manner will decrease as the observation time increases and will, in principle, approach zero as this time goes to infinity. It is not necessary that the time of observation be

actually the time during which the human observer looks or listens to the signal and remembers the results of observation. It is often possible to let the detecting system perform observations automatically and present a suitable average to the human observer.

Thus we have the idea that the determination of a minimum detectable signal is essentially a game of chance when only finite time is available. This is clearly recognised in the experiments that follow.

7.3 Experimental Technique

The experimental technique used was essentially the same as described in Section 6. The general layout and a block diagram of the inter-connections of various units are shown in diagram 7.1.

On the basis of the results of the experiments reported in Section 6, it was easy to see that the relative frequency of changes in the signal energy level approximated a fixed quantity thus showing a statistical regularity. Our aim in the following experiment was to find a reliable threshold level based on the mean energy level.

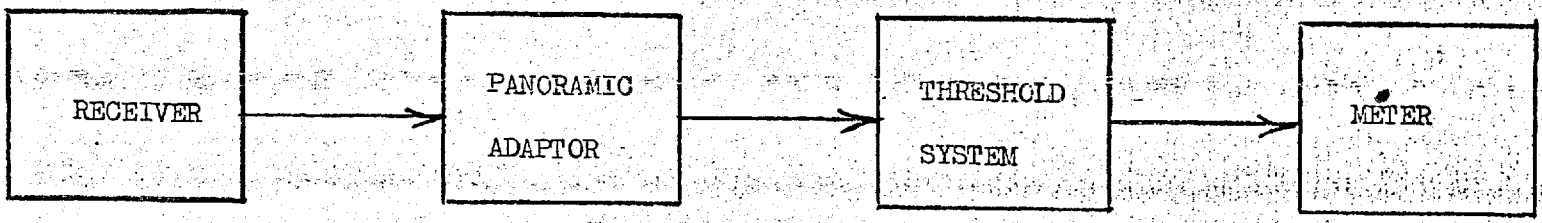


FIG. 7.1.

7.4 Experimental Method for Fixing the Threshold Level

The experiments were very simple in principle, but long and time consuming.

A particular portion of the spectrum, w_i, was chosen. The energy in this spectrum was measured after every integration period. Thus the recordings of the energy level within w_i were made continuously for short periods of 5 - 10 minutes at a time during various hours of day and night. From these readings the mean value of the energy in the spectrum was calculated.

In practice, it was found that the threshold level has to be raised by 3 db above this mean value. This was in order to keep the false alarm rate to a minimum. It was also noticed during these experiments that any threshold level above the value quoted caused some signals to be missed altogether.

It was also noticed that the variations of this threshold level were within 10% of the mean.

7.5 Evaluation of the Present Strategy for Detection

A limited experimental investigation of a technique for evaluating our strategy for the detection of a new signal will now be described. The technique was simple and straightforward. Nevertheless, the work was found to be very interesting because it demonstrated

that the present technique of signal detection could be exploited in actual practice with great accuracy.

Although it was not difficult to establish in the laboratory the actual conditions of reception, the experiments described below were conducted under limited conditions.

7.6 Experimental Set-up

The actual arrangement of the experimental apparatus is given in the block diagram 7.2.

An amplitude modulated signal was fed into an adding circuit for the purpose of superimposing noise. The output of the adding circuit was then fed into two identical communication receivers connected in parallel. One of the receivers was operated by a human observer for the detection of signals by manual turning. The other receiver was connected to our threshold system.

The experiment was performed with the help of 3 persons. All of them were qualified electrical engineers working in the field of communications. One of them, C, was in addition, a Radio Amateur. All these persons were familiar with present day monitoring techniques. A, B and C were situated in separate rooms.

Before the experiment a suitable threshold level was chosen

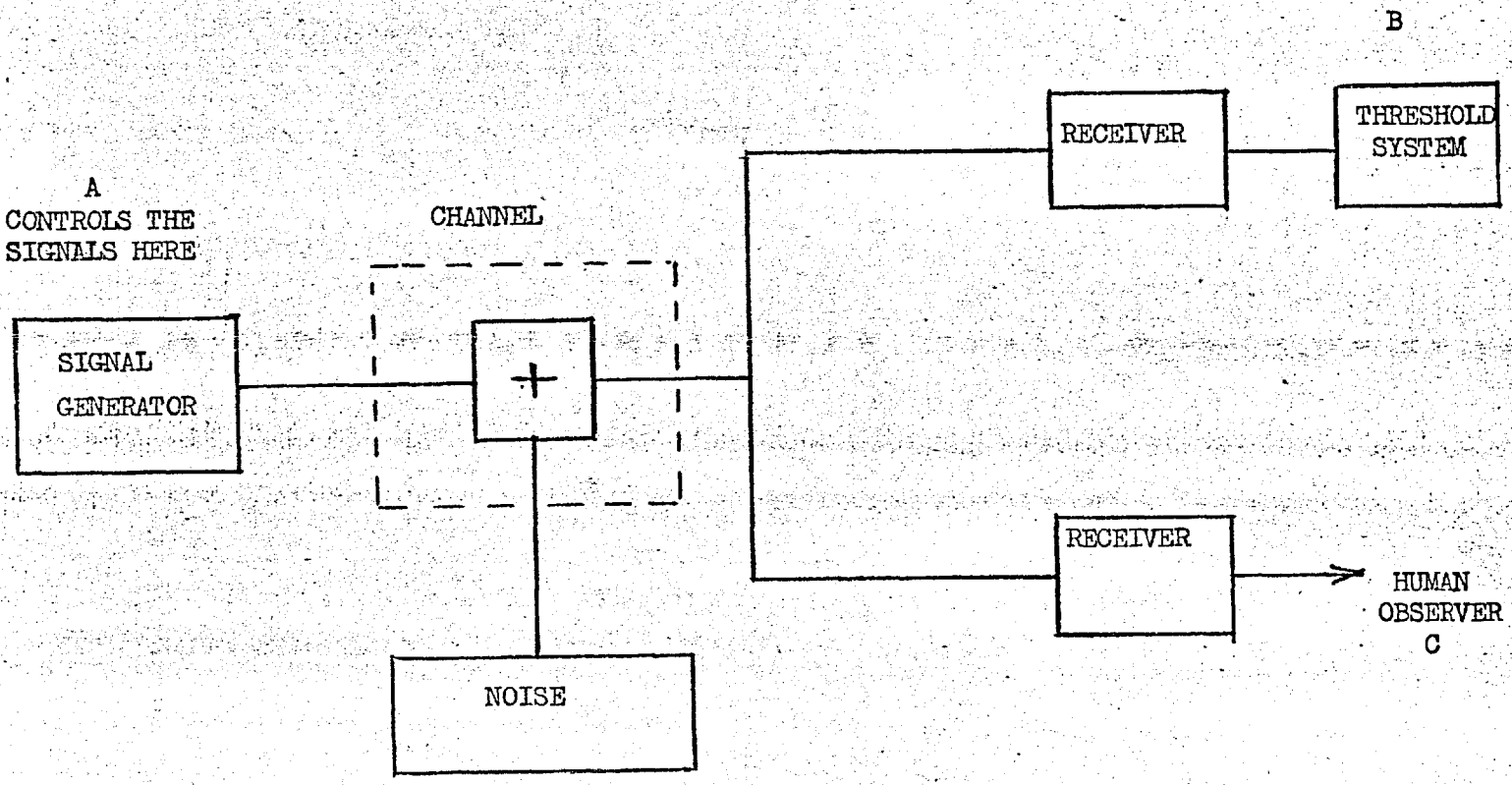

B
RECEIVER
THRESHOLD SYSTEM
A
CONTROLS THE SIGNALS HERE
CHANNEL
SIGNAL GENERATOR
+
RECEIVER
HUMAN OBSERVER
C
NOISE

FIG. 7.2.

in the case of B and it was decided to try and detect any signal above this threshold. The input to the receivers was maintained at an arbitrary level of 1 microvolt.

7.7 Experimental Technique

The actual experiment was done as follows. A, who controlled the signal, would send a message to both B and C: 'During the next two minutes a signal will appear in such and such a band'. The task of both B and C was then to detect the signal. B observed whether the threshold system indicated the presence of a new signal and C tuned the receiver manually to detect the new signal. After the end of two minutes both B and C recorded whether or not they had detected a signal.

The experiment was then repeated with signals of various frequencies and durations.

In the next experiment B and C were told to search given bands during short test intervals to find any new signals existing during that time. Later they were told to detect any new signals besides the ones present at the beginning of the experiment.

Finally an experiment was carried out in which attention was exclusively devoted to the human observer.

In this experiment a 100 kc/s bandwidth consisting of known signals was selected. At infrequent intervals which varied from 2 to 3 minutes, a new signal would appear somewhere in this bandwidth. The signals were of 10 seconds duration. The task of the observer was then to detect the signal. The total number of signals thus detected during a test session of half an hour was then recorded. Similar tests were carried out for four test sessions.

The results are shown in Figure 7.3.

7.8 Results and Conclusions

Strictly speaking, the results of the above experiments cannot be analysed in terms of efficiency relative to an 'ideal observer'. However, the results provided sufficient useful information about the relative merits of the two systems, namely, detecting the signal by a human operator adopting the manual tuning method and detection by threshold system.

It was evident from the experiments that in a given bandwidth the human observer failed completely to detect signals of less than 3 seconds duration where as the threshold system was capable of detecting signals even when their duration was less

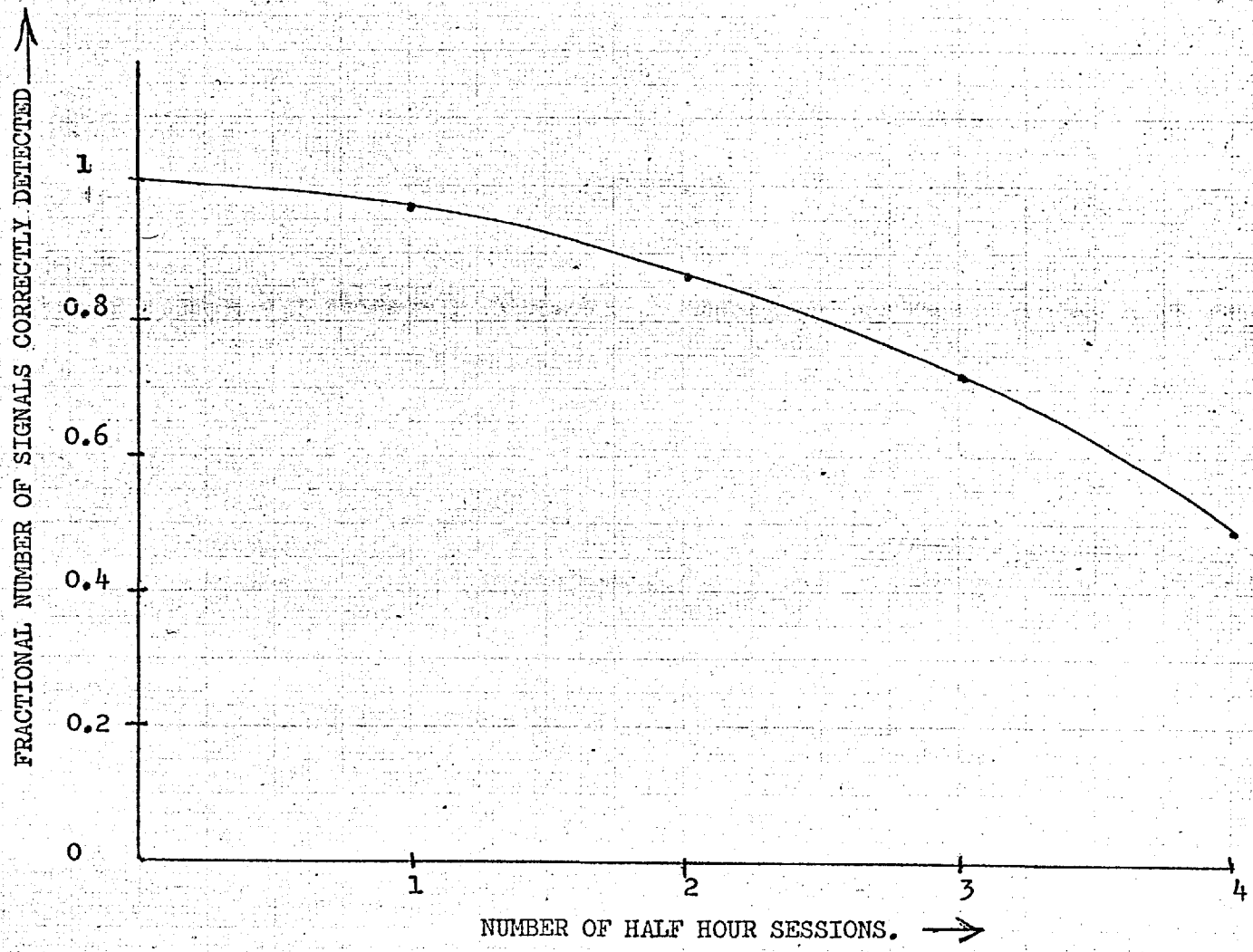

FRACTIONAL NUMBER OF SIGNALS CORRECTLY DETECTED
1
0.8
0.6
0.4
0.2
0
1
2
3
4
NUMBER OF HALF HOUR SESSIONS.

FIG. 7.3

than 1 second. Whether a signal of duration of 1 second or less can convey any useful information is a different question and we will not discuss the matter here.

Both the human observer and the threshold system were capable of correctly detecting the signal during short test sessions, provided the signal duration was more than 3 seconds. However, in longer test sessions, the human observer missed some signals completely.

The above tests were carried with noise added to the signals and the signal to noise ratio was maintained at about 10 db. At lower signal to noise ratios, the threshold system did not appear to be as efficient as the human operator. This problem is investigated in the next Section.

The performance of a human observer in a situation where a weak and infrequent signal has to be detected after variable time intervals has been a matter of great interest to experimental psychologists for some time. Broadbent(7), Mackworth and Taylor(21), Blair and Kaufman(5) and various other have studied this problem from a behavioural point of view. Our experiments also indicated results almost on the same lines. The studies have clearly indicated that the capacity of the human observer to detect a signal rapidly declined with time. Even in a two hour period of watch, the first half hour gave substantially better

performance than any of the succeeding half hours (See Figure 7.3). This drop in efficiency of the human observer can be very serious and costly in situations such as radar operators keeping watch for aircraft and ships appearing on their screens and the human operator monitoring the H.F. spectrum for possible new signals.

The threshold system on the other hand gives a steady performance and is not subject to fatigue or any other deteriorations which affect the human observer.

7.9 Determination of Detection Probability

We now describe the results of experimental investigations of a technique for evaluating the probability of signal detection as a function of signal to noise ratio.

The experimental set up is shown in Figure 7.4. The communication channel was fed via an adding circuit for superimposing noise on the input signal. Thus the receiver input signal was a known mixture of signal and noise. A minimum threshold level was found using the method described in the earlier Section. Even though the threshold level was well above the mean value of the signal, due to the random nature of the noise, the noise voltage sometimes crossed the threshold level and thus gave rise to

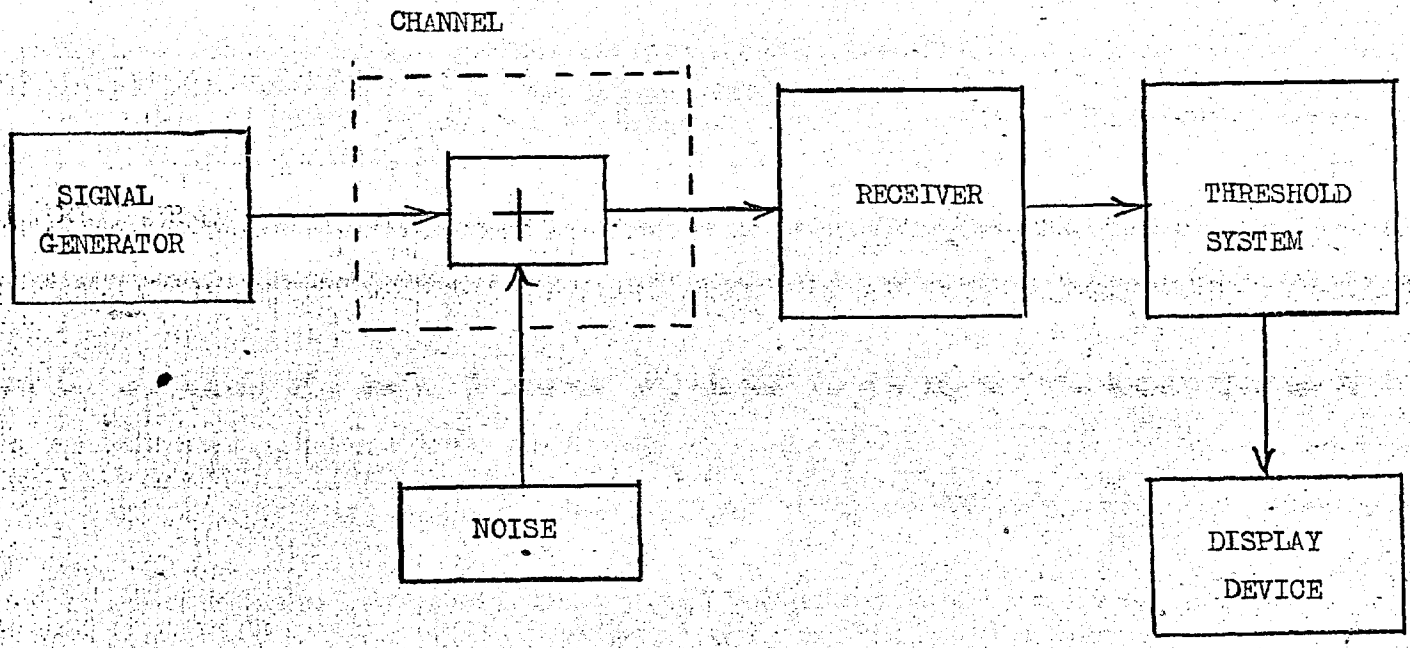

CHANNEL
SIGNAL GENERATOR
RECEIVER
THRESHOLD SYSTEM
NOISE
DISPLAY DEVICE

FIG. 7.4

false alarms.

The experiment consisted of conducting a survey of the number of times the measured voltage crossed the threshold level due to the characteristics of the noise without the presence of any additional signal. Each survey lasted for 10 minutes and surveys were made on ten different occasions.

The signal to noise ratio was taken as the peak signal power to noise power, and this was varied by keeping the gain of the input signal circuit constant and changing the power of the added noise.

A number of tests for various values of signal to noise ratio were carried out.

The results are shown in Figure 7.5.

The above experiments were repeated with real signals from the aerial as follows.

A spectrum containing a fading signal was selected. A threshold level was found for this spectrum. The signal power was measured and the noise power was also measured when the signal disappeared from the spectrum. Thus a rough estimate of the signal power relative to the noise was obtained.

The number of times the noise level caused false alarms was then recorded and the experiment described in the previous section was repeated for various values of signal to noise ratio choosing

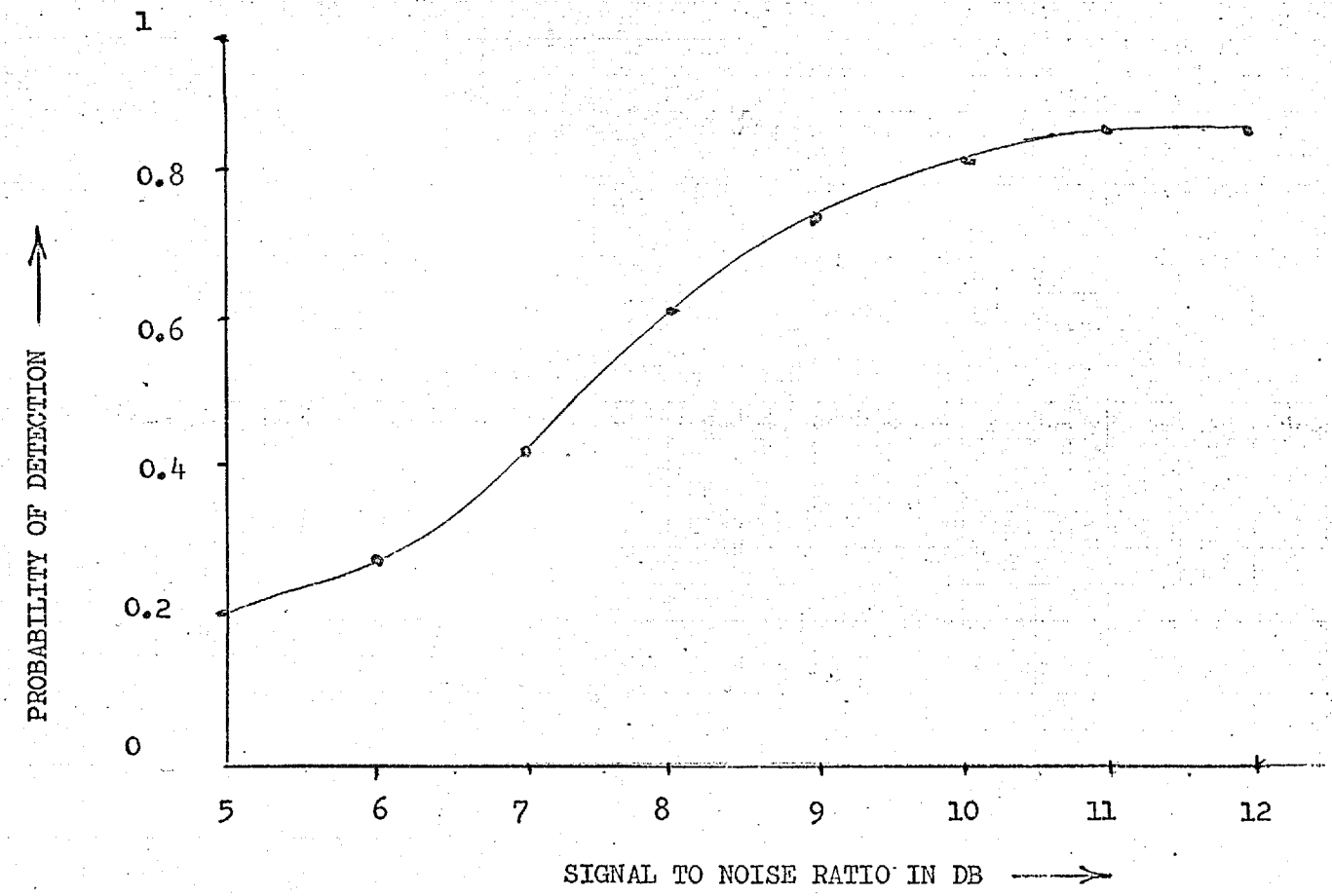

1
0.8
0.6
0.4
0.2
0
PROBABILITY OF DETECTION
5
6
7
8
9
10
11
12
SIGNAL TO NOISE RATIO IN DB

FIG.7.5

signals having varying signal strength from different parts of the spectrum. The results are shown in Figure 7.6.

7.10 Final Remarks

A method has been described for determining the probability of detection of a new signal. The results clearly show that at higher values of signal to noise ratio, 10 db in the case of artificial signals and 12 db in the case of real signals from the aerial, the probability of correctly detecting the signal is about .85 . However, it is not very often that we have such large signals.

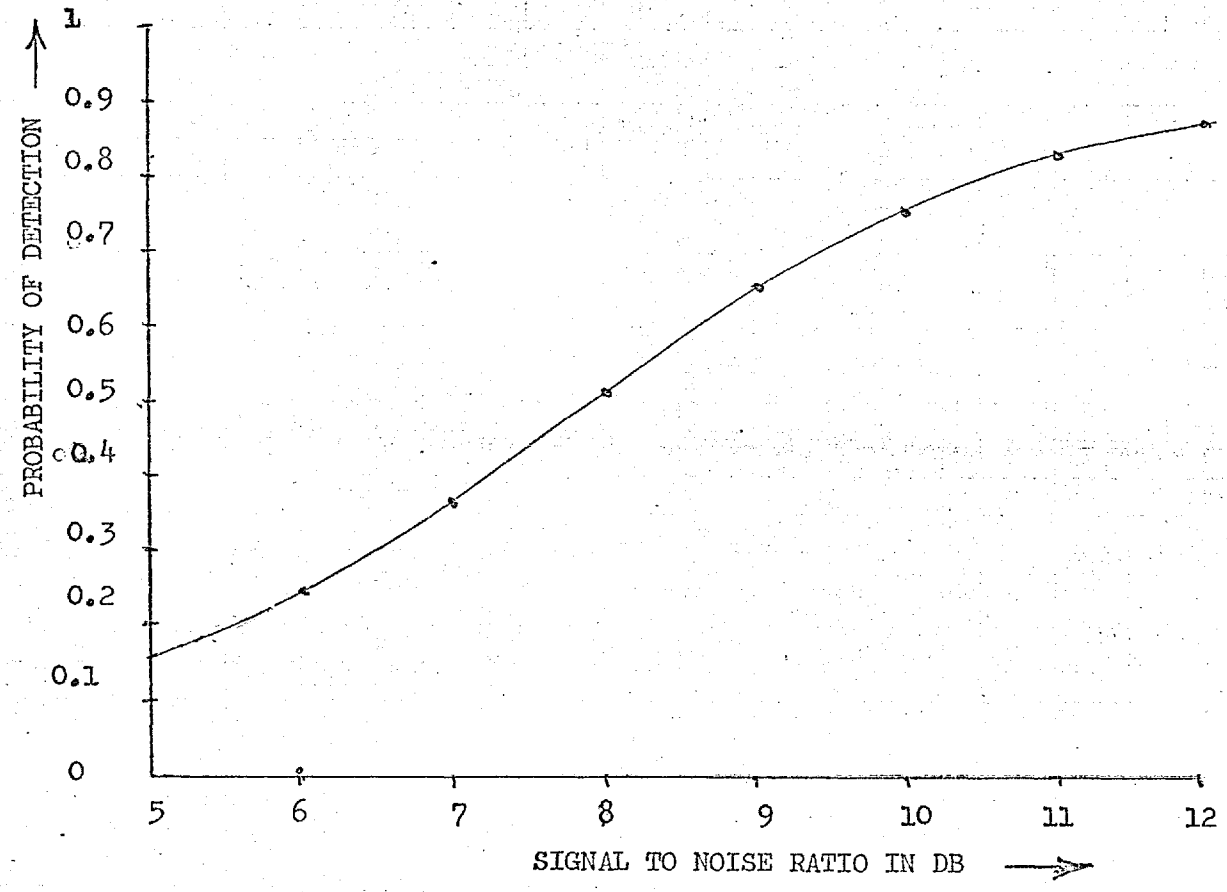

PROBABILITY OF DETECTION
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
5
6
7
8
9
10
11
12
SIGNAL TO NOISE RATIO IN DB

FIG. 7.6

8. SUMMARY AND CONCLUSIONS

8.1 Summary

The problem of finding the best strategy for the detection of a signal of unknown parameters within the H.F. band, and the details of a proposed system capable of detecting such a signal have been described.

The technique is based on measuring the energy of the spectrum of interest and comparing it with some previously set threshold. If the energy level is above the threshold then we conclude that a new signal has been detected, and vice versa.

The preliminary set of experiments described in Section 5 clearly show that the energy in the spectrum could be measured with considerable accuracy.

The two basic forms of data representation are analogue and digital. The common analogue form makes use of physical or electrical variations which resemble the observed data. Thus an analogue system is defined as one in which some signal attribute such as amplitude or frequency is proportional to, or a function of, some measurable parameter. The digital form uses abstract codes having some arbitrary but logical relationship to the observed data. Such a digital system based on the binary coding technique

also offers great accuracy and versatility.

In our experiments, since we were not interested in the absolute measurement of energy but only in changes in energy, it was decided that digital techniques were best suited for our purpose.

The behaviour of signals and noise was studied in specific situations and a considerable amount of research was done on the analysis of such signal amplitudes. In particular, detailed investigations were made of the various problems in setting a threshold.

In this connection experiments were performed to study the effects of fading. This involved the measurement of the amplitude distribution of a fading signal at various hours of day and night. A pen recorder was used for this purpose. The mean and variance of the signal amplitudes were used as guide in fixing the threshold.

The final set of experiments were devoted to the determination of the probability of detection. This was found to be a function of both signal to noise ratio and signal duration.

The success of the strategy depends on a high probability of detection of a signal of short duration. To obtain some measure of this success, the above experiments were found quite adequate.

A mathematical model of the system is also presented in Section 2. Here the problem of estimating the number of signals that are present in a noisy H.F. spectrum, when the signal amplitude is

distributed and the distribution is known within certain parameters, is analysed.

In most cases it is possible to formulate the model in more than one form and the one chosen will depend on the ease of solution. In our case, the ease of solution has been the main consideration in the selection of the model.

In the experimental system so far described, the bandwidth selected for scanning was 100 Kc/s and the sweep time was chosen to be 0.2 seconds. In a real situation, several receivers will be used concurrently, each sweeping a different band of the spectrum. It was noticed during these experiments, that on certain occasions, when the spectrum was not crowded, we could safely increase the bandwidth to 1 Mc/s with an associated sweep time of 2 seconds. Despite this, the selection of 100 Kc/s bandwidth was influenced by the following main factors.

a) The regulations published by the International Telecommunication Union, Geneva, demand that two broadcasting stations should be separated by at least 4 Kc/s. Because of this ruling, a bandwidth of 10 Kc/s cannot contain more than two or three broadcasting stations. Under these conditions the spectrum is not very crowded, and the appearance of a new signal can be easily detected.

b) It would be easier, from the human observer's point of

view, to locate the change indicated by an alarm in a block of 10 rather than 100 which would have been the case had a bandwidth of 1 Mc/s been chosen.

8.2 Advantages of Threshold System

The Threshold System appears to be very versatile. As the individual member of the crew would attempt to tune a fixed narrow band receiver in order to detect the presence of a new signal, so the above system compares the energy in the spectrum with a pre-set threshold level and, if the energy exceeds the threshold level, it sounds an alarm. This calls the attention of the human observer who can then tune that specific narrow band in order to extract and possibly record the message.

The description of the system given in Section 5 may give the impression that the apparatus is complicated and difficult to handle. However, the apparent complications arise from the difficulty in explaining the inter-relation between the units and the reason for the choice of the arrangement. The individual units are simple from an operational point of view and have proved to be extremely reliable.

The main advantage of the Threshold System is that readily

available commercial equipment can be used, thereby reducing the cost.

Above all the greatest advantage is the saving of manpower. Men will be released from a monotonous and repetitive type of duty involving the task of tuning a receiver for long hours of day and night. The staff so released could be used in other important and interesting technical work.

8.3 A Note on Signal Extraction

In the fundamental problem of signal detection, an observation is made during a fixed interval of time, and a decision is then made as to whether the interval contains only background noise or a signal and noise. In this basic problem only two alternatives exist; noise is always present whereas the signal may or may not be present during the specified observation interval.

In actual practice, when we want to extract a useful message from corrupting noise, the problem becomes somewhat modified. In the practical reception of signals, whether audio, television, or radar, the final link in the channel - the human observer - makes a large contribution to the signal extraction. A human being can

give more weight to an intelligent message than to corrupting noise and he can even interpolate in the gaps where the noise makes the signal undetectable. As soon as a signal can be recognised as belonging to a customary type of communication signal, such as speech, the human perception mechanism gives greatly preferred weighting, on the average, to known signals as compared with noise and decreases, to a large extent, the probability of noise being detected as a signal. Cherry[9] who has done some pioneering work on the subject of human perception, has considered the problem of detecting signals in special types of noise, generally other speech signals. Broadbent[6] also deals with this problem in detail.

Thus the 'perception selectivity' which is a remarkable faculty of the human observer plays a very important role in signal extraction. Until now no reasonable machine exists which can replace such a human observer except in some very simple cases. Under these circumstances, when we try to replace the human observer in our system the provision of such ingenuity in the form of some machine becomes quite complex and consequently very costly.

8.4 Suggestions for Future Research

Several interesting points were raised during the course of

the investigation. These and other ramifications are described below.

In our experiments we had selected a bandwidth of 10 Kc/s and the energy contained therein was taken as a basis for fixing the threshold. It appears that if the selected bandwidth is reduced to a value approximating the bandwidth of the signals considered, there is a possibility of detecting weaker signals. This problem is worth considering with economic factors in mind.

Another avenue of research in this area lies in the potentialities of adopting correlation techniques for detecting the presence of new signals. This immediately raises the problem of storage and application of digital computer techniques.

In the past most communication systems have been designed and developed piecemeal as aggregates of special purpose links; but in future we can expect greater emphasis on integrated, general-purpose communication networks, with automatic traffic routing and control. It appears that binary digits are best suited for an appropriate representation for input and output messages in such general-purpose networks. Since digital computers are bound to play an important role in these tasks, it may be rewarding to divert research in these channels.

8.5 Conclusions

The primary purpose of this study has been to investigate the practical aspects of a proposed strategy for detecting signals of unknown parameters and of short duration within the H.F. band, and the possibility of building an electronic system which can detect the signals automatically. The conclusion reached was that the technique is practicable, but there are some reservations. A highly efficient threshold estimation technique must be employed to overcome the effects of fading, noise, etc.

In the course of the present study an interesting digression was made. This is the possibility of applying statistical techniques for estimating the number of signals that are buried in noise, given a mixture of signal and noise. This has wide applications in radar. It is hoped that some of the points raised may be of value to others working in this field. In any event the method of attack should be generally applicable, and at least from this point of view, it is hoped that the study will be useful.

APPENDIX 1

Determination of the degradation of analyser sensitivity and resolution as a function of sweep rate.

We assume that analyser sensitivity curve is given by Gaussian error function. This is true to a good approximation for an intermediate frequency amplifier composed of several synchronously-tuned stages. Then the normalized steady state voltage amplitude response can be written as

$$H(\omega) = \exp\left(\frac{-\ln 2}{2\pi^2}\right)\left(\frac{\omega - \omega_o}{B}\right)^2 \qquad (1)$$

where B is the 3 db bandwidth in cycles per second and ω_o is the angular frequency.

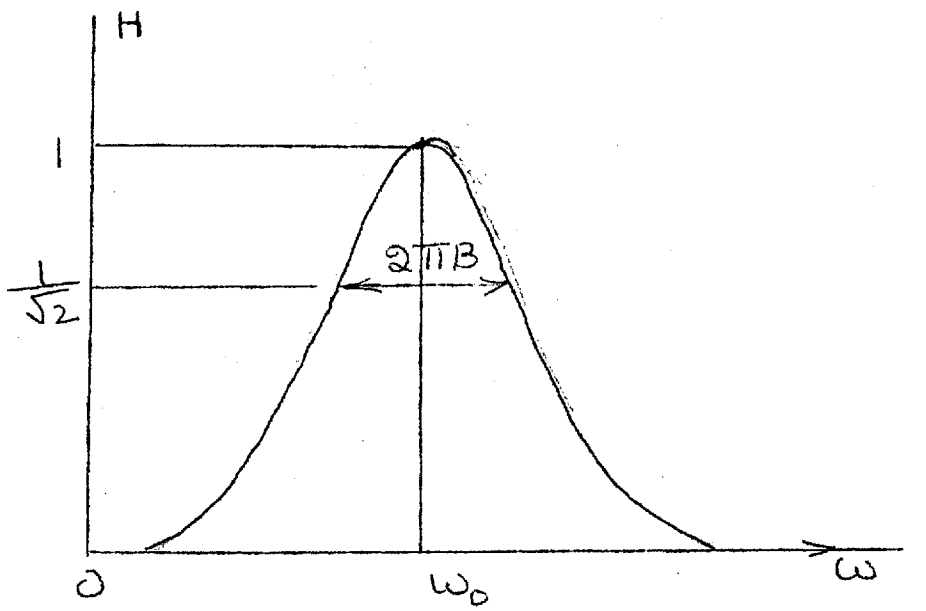


Figure 1

The swept signal voltage may be represented by

$$e(t) = \epsilon^{j\theta} \tag{2}$$

Where θ is the instantaneous angular frequency and is given by

$$\omega_i = \frac{d\theta}{dt} \tag{3}$$

We now assume a linearly swept signal as shown in Fig. 2.

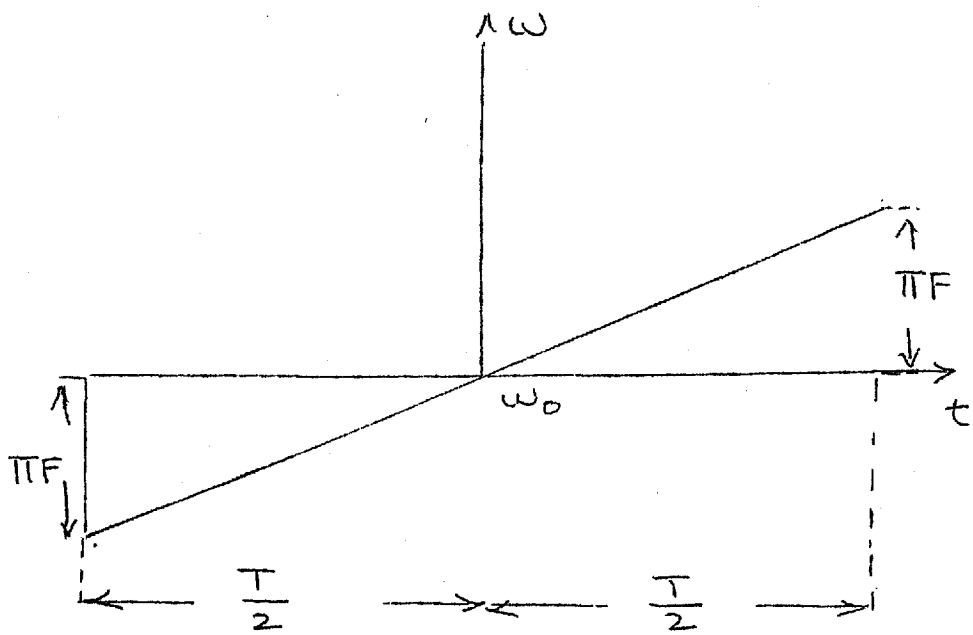


FIG 2

Where $\frac{F}{T}$ is the sweep rate in cycles per second^2 and $\omega = \omega_o$ at $t = 0$.

We can now write the following

$$\omega_i = \omega_o + \frac{2\pi F}{T} t \tag{4}$$

Hence from equation (3)

$$\theta = \int \omega_i \, dt$$

$$= \omega_o t + \frac{2\pi F}{T} t^2 \qquad (5)$$

Substituting this value for θ in equation (2) we have,

$$e(t) = \exp j \left(\omega_o t + \frac{\pi F}{T} t^2\right) \qquad (6)$$

The frequency spectrum of e(t) is given by the Fourier transform

$$E(\omega) = \int_{-\infty}^{\infty} e(t) \; e^{-j\omega t} \, dt \qquad (7)$$

$$= \int_{-\infty}^{\infty} e^{j \left(\omega_o - \omega\right)t + \frac{\pi F}{T} t^2} \, dt \qquad (8)$$

Now equation (8) can be evaluated in the following manner:

Let us consider the integral

$$\int_{-\infty}^{\infty} e^{j(at^2 + bt)} \, dt \qquad (9)$$

We can complete the square in the exponential as follows.

$$at^2 + bt = a\left(t + \frac{b}{2a}\right)^2 - \frac{b^2}{4a} \qquad (10)$$

$$\text{Let} \quad x = t + \frac{b}{2a} \qquad (11)$$

$$\text{Then} \quad dx = dt \qquad (12)$$

Substituting in equation (9) we obtain

$$\int_{-\infty}^{\infty} \in^{j(at^2 + bt)} \, dt$$

$$= \in^{\frac{-jb^2}{4a}} \int_{-\infty}^{\infty} \in^{jax^2} \, dx$$

$$= 2 \in^{\frac{-jb^2}{4a}} \int_{0}^{\infty} \in^{jax^2} \, dx \qquad (13)$$

The above expression is a definite integral and can be evaluated using Pierce's Short Table of Integrals, viz,

$$\int_{0}^{\infty} \in^{-a^2x^2} \, dx = \frac{1}{2a} \sqrt{\pi} \qquad (14)$$

It is now easily seen that equation (13) becomes

$$\int_{-\infty}^{\infty} \in^{j(at^2 + bt)} \, dt = \sqrt{\frac{j\pi}{a}} \in^{\frac{-jb^2}{4a}} \qquad (15)$$

Now making appropriate substitution in equation (8) we get

$$E(\omega) = \sqrt{\frac{jT}{F}} \exp - j \frac{T}{4\pi F} (\omega_o - \omega)^2 \qquad (16)$$

The output voltage transient of the Gaussian amplifier is given by inverse Fourier transform

$$V(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} E(\omega) H(\omega) \in^{j\omega t} \, d\omega \qquad (17)$$

$$= \frac{1}{2\pi} \sqrt{\frac{jT}{F}} \int_{-\infty}^{\infty} \in^{-j\frac{T}{4\pi F} (\omega_o - \omega)^2}$$

$$\in^{-\left(\frac{\log 2}{2\pi^2 B^2}\right)(\omega-\omega_o)^2} \in^{j\omega t}\, d\omega \qquad (18)$$

We will now try to solve the above equation.

First of all we make the following substitutions:

$$c = \frac{T}{4\pi F} \qquad (19)$$

$$d = \frac{\ln 2}{2\pi^2 B^2} \qquad (20)$$

$$\text{and} \quad k = \frac{1}{2\pi}\sqrt{\frac{jT}{F}} \qquad (21)$$

Equation (18) can now be written in the following simplified form.

$$V(t) = k \int_{-\infty}^{\infty} \in^{-d(\omega-\omega_o)^2} \in^{jc(\omega_o-\omega)^2} \in^{j\omega t} d\omega \qquad (22)$$

$$= k \in^{-\omega_o^2(d+jc)} \int_{-\infty}^{\infty} \in^{\left\{-(d+jc)\omega^2 + 2d\omega_o + j(2c\omega_o+t)\right\}\omega} d\omega \qquad (23)$$

Let $Z_1 = d + jc \qquad (24)$

$$Z_2 = 2\,d\,\omega_o + j(2\,c\,\omega_o + t) \qquad (25)$$

Substituting these in (23) we have

$$V(t) = k \in^{-\omega_o^2 Z_1} \int_{-\infty}^{\infty} \in^{-Z_1\omega^2 + Z_2\omega} d\omega \qquad (26)$$

Now

$$-Z_1\omega^2 + Z_2\omega = -Z_1\left(\omega - \frac{Z_2}{2Z_1}\right)^2 + \frac{Z_2^2}{4Z_1} \tag{27}$$

Therefore

$$V(t) = k\,\epsilon^{-\omega_o^2 Z_1}\,\epsilon^{\frac{Z_2^2}{4Z_1}} \int_{-\infty}^{\infty} \epsilon^{-Z_1\left(\omega - \frac{Z_2}{2Z_1}\right)^2} d\omega \tag{28}$$

Now let

$$W = \omega - \frac{Z_2}{2Z_1} \tag{29}$$

Whence

$$dW = d\omega \tag{30}$$

$$\therefore V(t) = k\,\epsilon^{-\omega_o^2 Z_1}\,\epsilon^{\frac{Z_2^2}{4Z_1}} \int_{-\infty}^{\infty} \epsilon^{-Z_1 W^2} dW \tag{31}$$

A gain applying equation (14) the integral factor of equation (31) can be written as

$$\int_{-\infty}^{\infty} \epsilon^{-Z_1 W^2} dW = 2\int_{0}^{\infty} \epsilon^{-Z_1 W^2} dW = \sqrt{\frac{\pi}{Z_1}} \tag{32}$$

We have thus for equation (31)

$$V(t) = k\sqrt{\frac{\pi}{Z_1}}\,\epsilon^{-\omega_o^2 Z_1}\,\epsilon^{\frac{Z_2^2}{4Z_1}} \tag{33}$$

$$= \frac{1}{2\pi}\sqrt{\frac{j\pi T}{F(d+jc)}} \quad X$$

$$\exp\left\{\frac{2d\omega_o + j(2c\omega_o + t)^2}{4(d + jc)} - \omega_o^2(d - jc)\right\} \tag{34}$$

Now we have to remember that we are actually interested in determining the envelope A(t) of the real part of V(t). Since V(t) is complex in our example it may be represented by

$$V(t) = |V(t)| \in^{j\phi(t)} \tag{35}$$

$$\therefore \text{Re}\{V(t)\} = |V(t)| \cos\phi(t) \tag{36}$$

which has the envelope

$$A(t) = |V(t)| \tag{37}$$

If we now look at equation (34), it can be seen that

$$V(t) = Z \exp\left[\phi_1(t) + j\,\phi_2(t)\right] \tag{38}$$

where Z is complex and $\phi_1(t)$ and $\phi_2(t)$ are real. Hence from equation (37)

$$A(t) = |Z| \exp\phi_1(t) \tag{39}$$

$$\text{Now } Z = \frac{1}{2\pi}\sqrt{\frac{\pi T}{F}} \times \sqrt{\frac{j}{d+jc}} \tag{40}$$

$$\therefore |Z| = \frac{1}{2\pi}\sqrt{\frac{\pi T}{F}}\,(c^2 + d^2)^{-1/4} \tag{41}$$

also,

$$\phi_1(t) + j\,\phi_2(t) = \frac{\left[2d\omega_o + j(2c\omega_o + t)\right]^2}{4(d + jc)} - \omega_o^2(d + jc) \tag{42}$$

Whence

$$\phi_1(t) = \frac{-dt^2}{4(c^2+ d^2)} \tag{43}$$

Now substituting equations (41) and (43) into equation (39), we obtain

$$A(t) = \frac{1}{2\pi}\sqrt{\frac{\pi T}{F}}\ (c^2 + d^2)^{-1/4} \exp \frac{d^2}{4(c^2+d^2)}\ t^2 \tag{44}$$

We now observe that A(t) has the form of Gaussian error function. Let a be the maximum amplitude of A(t). Thus

$$a = \frac{1}{2\pi}\sqrt{\frac{\pi T}{F}}\ (c^2 + d^2)^{-1/4} \tag{45}$$

Substituting equation (19) and (20) into equation (45) we have

$$a = \left[1 + \left(\frac{2 \ln 2}{\pi}\right)^2 \left(\frac{F}{TB^2}\right)^2\right]^{-1/4} \tag{46}$$

$$= \left[1 + 0.195 \left(\frac{F}{TB^2}\right)^2\right]^{-1/4} \tag{47}$$

The resolution of the spectrum analyser is the apparent bandwidth of the output transient as viewed on the screen and is as follows:

$$R = \frac{F}{T} \times \Delta t \tag{48}$$

Where Δt is the time required to sweep between the 3 db points of the transient envelope. Now looking at equation (44) we can write,

$$\frac{1}{\sqrt{2}} = \exp \frac{-d}{4(o^2+ d^2)} \left(\frac{\Delta t}{2}\right)^2 \tag{49}$$

whence,

$$\Delta t = \left[(8 \ln 2) \left(\frac{c^2 + d^2}{d} \right) \right]^{\frac{1}{2}} \tag{50}$$

Substituting equations (19) and (20) into equation (50) we have

$$\Delta t = 4\pi B \left[\left(\frac{T}{4\pi F} \right)^2 + \left(\frac{\ln 2}{2\pi^2 B^2} \right)^2 \right]^{\frac{1}{2}} \tag{51}$$

Therefore from equation (48)

$$\frac{R}{B} = \left[1 + \left(\frac{2 \ln 2}{\pi} \right)^2 \left(\frac{F}{TB^2} \right) \right]^{\frac{1}{2}} \tag{52}$$

$$= \left[1 + 0.195 \left(\frac{F}{TB^2} \right)^2 \right]^{\frac{1}{2}} \tag{53}$$

—o—

APPENDIX 2

EXPERIMENTAL APPARATUS

The apparatus required for the experiments described earlier was quite simple. Apart from standard laboratory equipment like Noise Generator, Vacuum Tube Voltmeter etc., the following are worth mentioning.

The receiver used for the experiments was type RA 17 communication receiver supplied by Racal Electronics. This is designed for use as general purpose receiver capable of providing a high order of sensitivity, selectivity and stability. The receiver covers a frequency range from 1 - 30 Mc/s.

The panoramic adaptor type R.A. 66 (manufactured by Racal) was used along with R.A. 17 receiver, for providing a visual display of the signal in any selected frequency band lying between 1 and 30 Mc/s. The equipment was capable of displaying a maximum spectrum width of 1 Mc/s and a minimum of 100 Kc/s. The swept spectrum can be seen on a cathode ray tube. There was provision for varying the sweep speed.

Integrator

The integrator and quenching circuits used were straight forward and the circuit diagram is shown in Figure A.2.1.

The circuit diagram of a Schmitt Trigger decision circuit is shown in Figure A.2.3. It is used occasionally in switching circuits as a zero crossing detector and in our equipment is is used to give a fixed output level for a predetermined input level. The predetermined input level is set by special circuit shown in Figure A.2.2. The sensitivity and stability of this device is excellent.

The circuit shown provides an output voltage of 10 volts with a rise time of 0.4 μ seconds. The maximum operating rate is about 300 Kc/s.

Artificial Spectrum Generator

Figure A.2.11 shows the block diagram of the Artificial Spectrum Generator used for test purposes. A photograph of the spectrum generated by this device is shown in Figure 2.4 on page 26 under the heading "Spectrum Generated by Artificial Spectrum Generator". The common circuits were used in designing the oscillators. There was provision for generating one or more signals of various frequencies at the same time.

Storage Facilities

Storage facilities form an essential form of our equipment described earlier in Section 5. There are two basic requirements for the storage system.

a) It must have sufficient capacity.

b) It must be capable of retrieving the stored information in a very short time.

With these in view the following types of storages were considered.

1. Magnetic Tape

If we use magnetic tape a very precise equipment is needed. Also tape noise was found to be a serious problem.

2. Magnetic Drum

It is possible to get delays of very long duration with magnetic drum. But the auxiliary equipment needed for storage and retrieval of the signal is quite complex.

3. Punched Tape

This is a good way of storing information and has the advantage that storage is permanent because of its non-erasable property; but it needs a tape punch and tape reader and was found too inconvenient for our purposes.

4. Bi-stable Devices

It was decided to use bistable devices such as shift registers.

This type of storage is volatile and erasable. The capacity is one binary bit for a stage and the stage is kept conducting or cut off to represent 1 or 0 respectively. The principle of moving digits successively through a series of bistable devices under control of a train of clock pulses is a well known computer technique and in the past few years shift registers using transistors have become very common. High switching speeds, small size, cheapness, low power consumption, long operating life and very high reliability are the main characteristics of such devices.

The circuit diagram is shown in Figure A.2.4. Figure A.2.5. shows the circuit used in the decimal to binary converter and comparator.

Decimal to Binary Converter

In order to convert the 8 levels of voltages into binary digits a decimal to binary converter was used. The block diagram is shown in Figure A.2.6.

Later on it was found that a simple diode network was sufficient for the purpose.

Comparator of Binary Digits

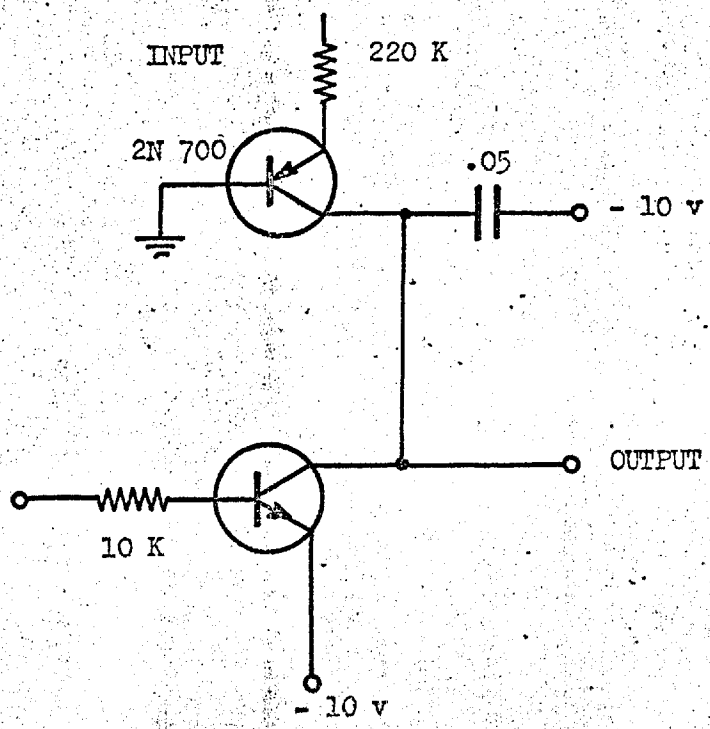
This is an arrangement for comparing two sets of binary

numbers. If the digits are identical there is no output. The circuit arrangements are shown in Figure A.2.7.

Figure A.2.8. shows the circuit diagram of inverter amplifier and Figure A.2.9. shows the diagram of Gate circuit.

Monostable Circuit

The monostable used in the system gives one output pulse for every input pulse. This was used to light a lamp in our system. The output may be taken from either collector. The circuit used is given in Figure A.2.10.



INTEGRATER AND QUENCHING CIRCUIT

FIG. A.2.1

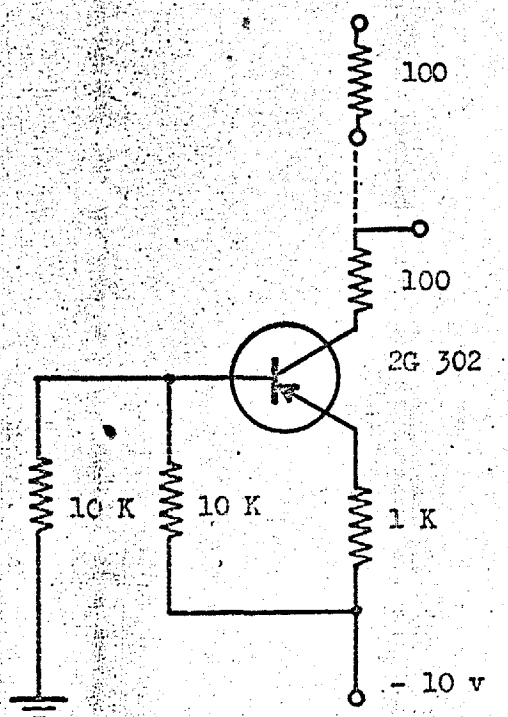
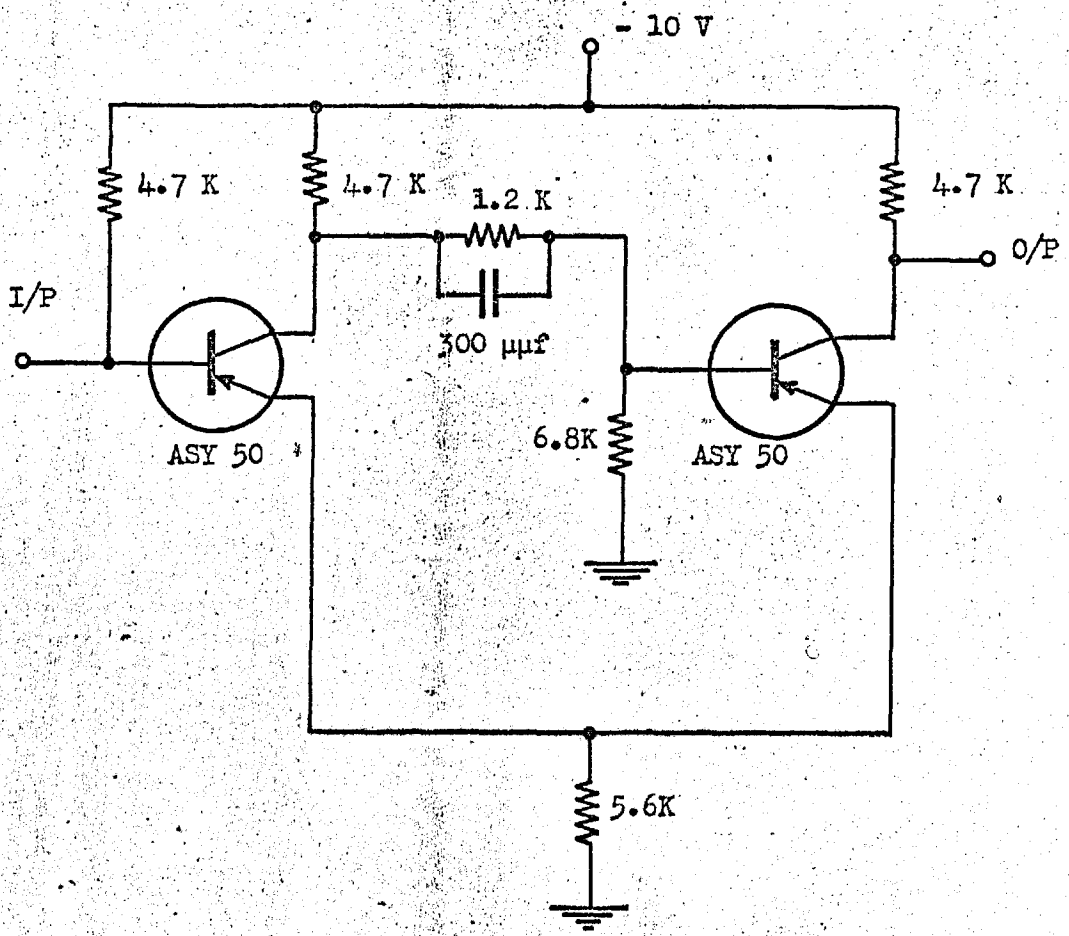
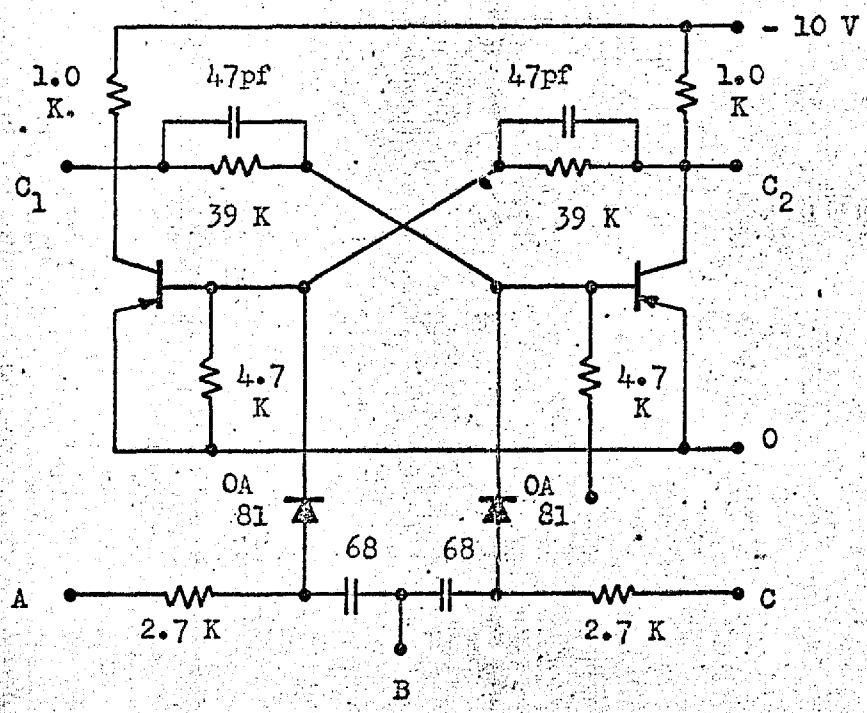


FIG.A.2.2.



SCHMITT TRIGGER CIRCUIT

FIG. A.2.3.



1) With medium-speed transistors, e.g. 2G302, the circuit is capable of operating at 2 Mc/s.

2) For operation as a binary counter, connect A to C_1 and and C to C_2, and use B as input.

3) For operation as a shift register stage, connect A to C_2 of preceding stage and C to C_1 of the preceding stage, and apply shift pulses to B.

4) The circuit triggers or shifts on positive-going pulses.

COUNTER, SHIFT REGISTER STAGE.

FIG. A.2.4.

BISTABLE CIRCUIT

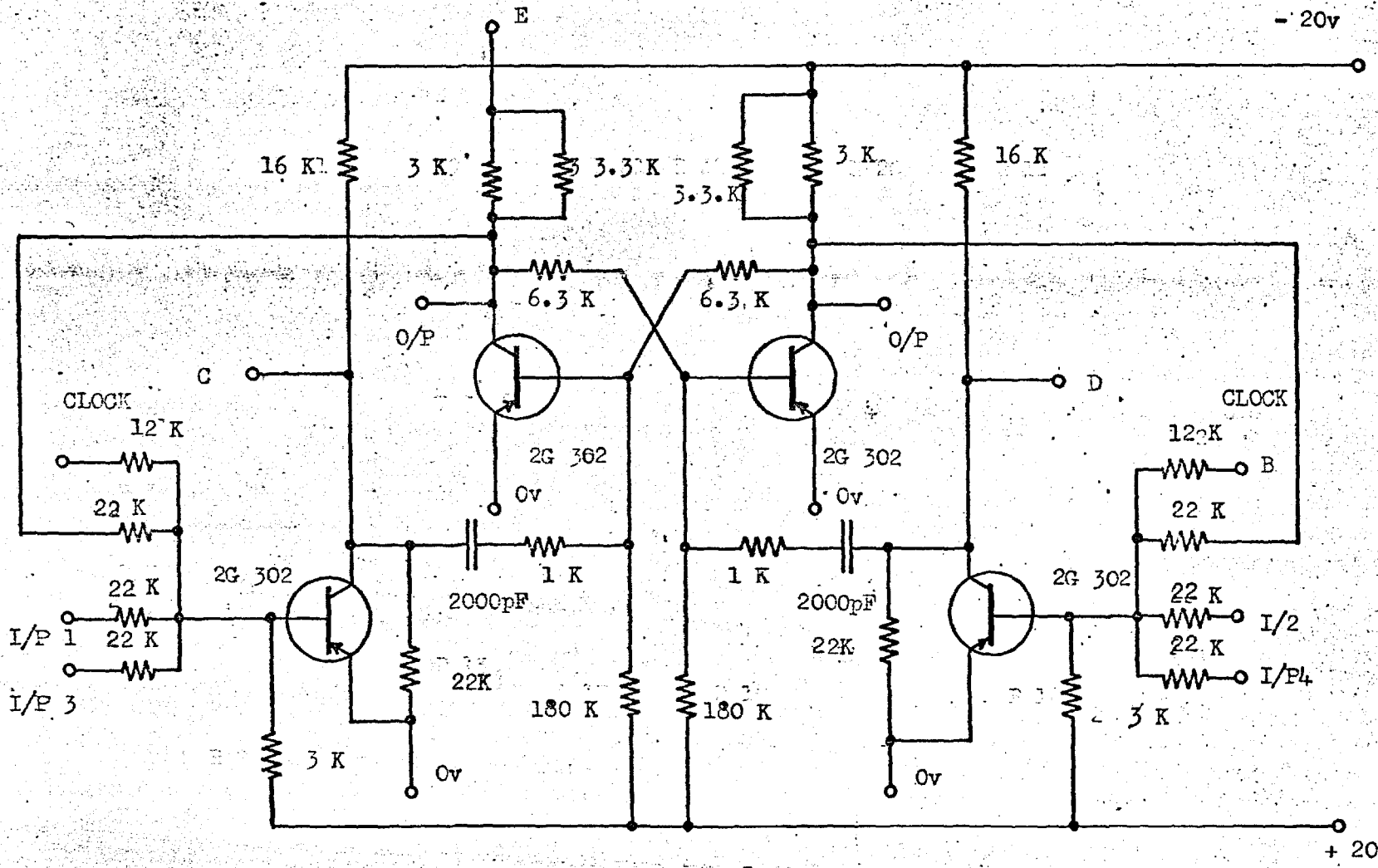


FIG. A.2.5.

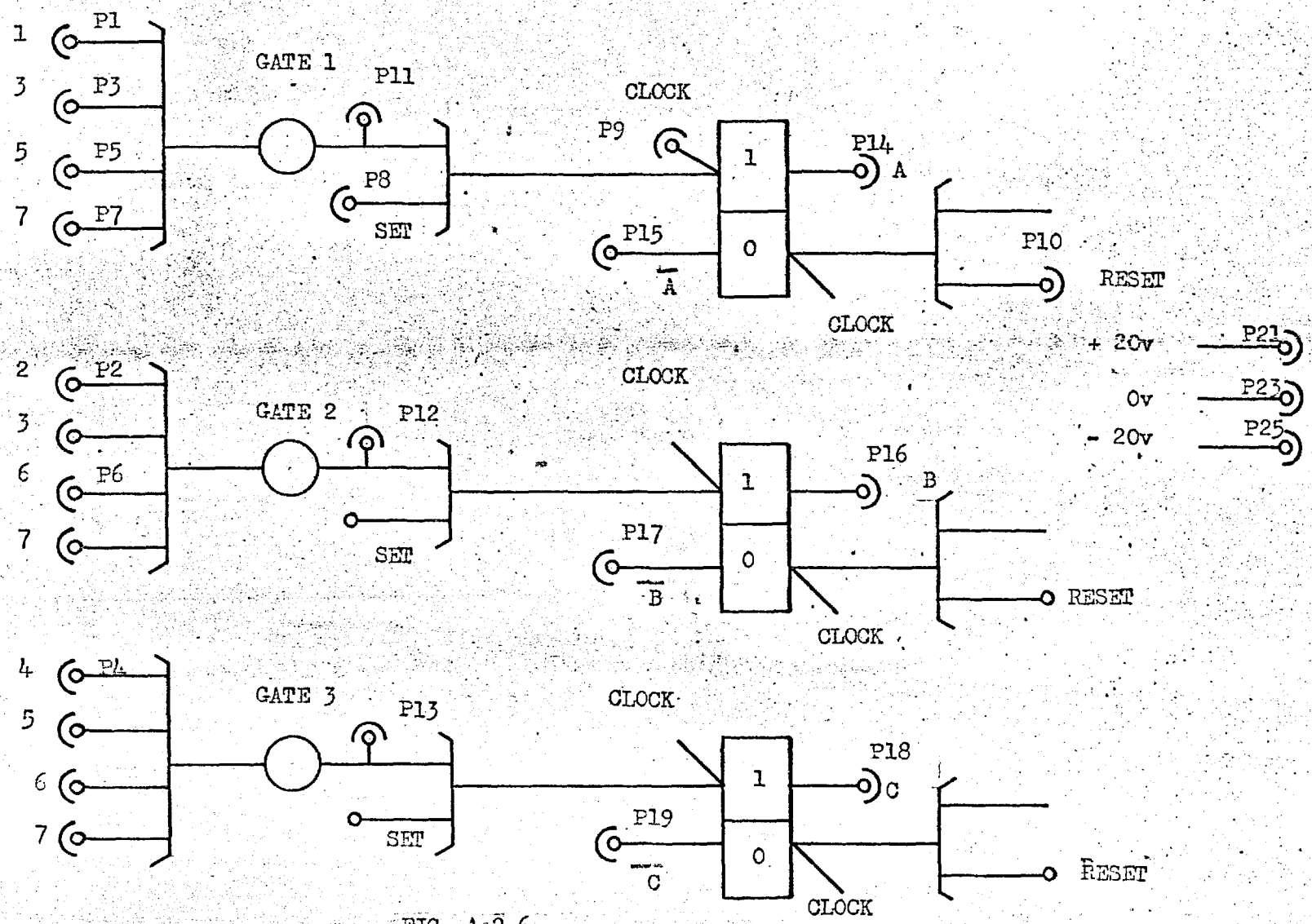


FIG. A.2.6

DECIMAL TO BINARY CONVERTER

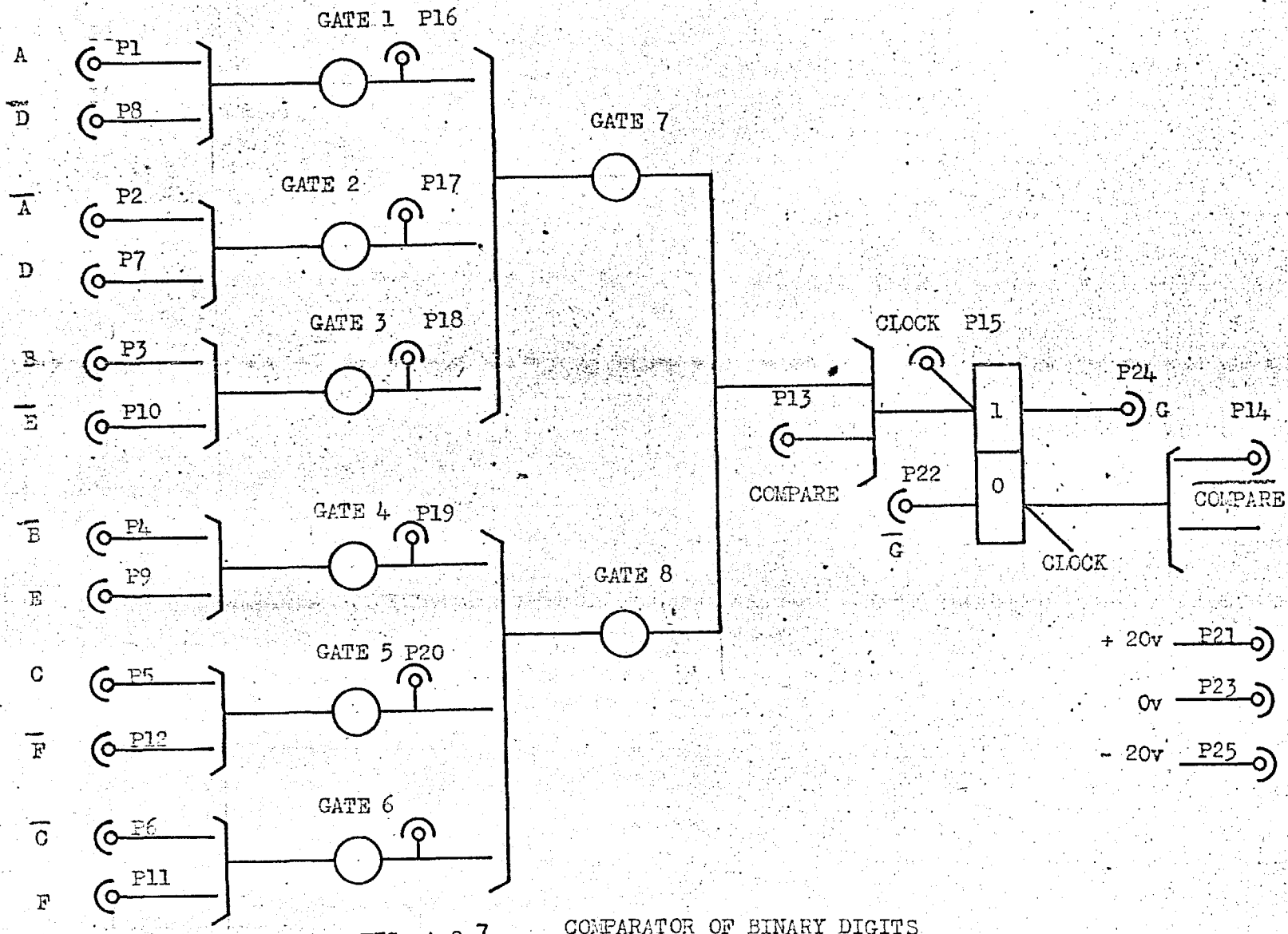
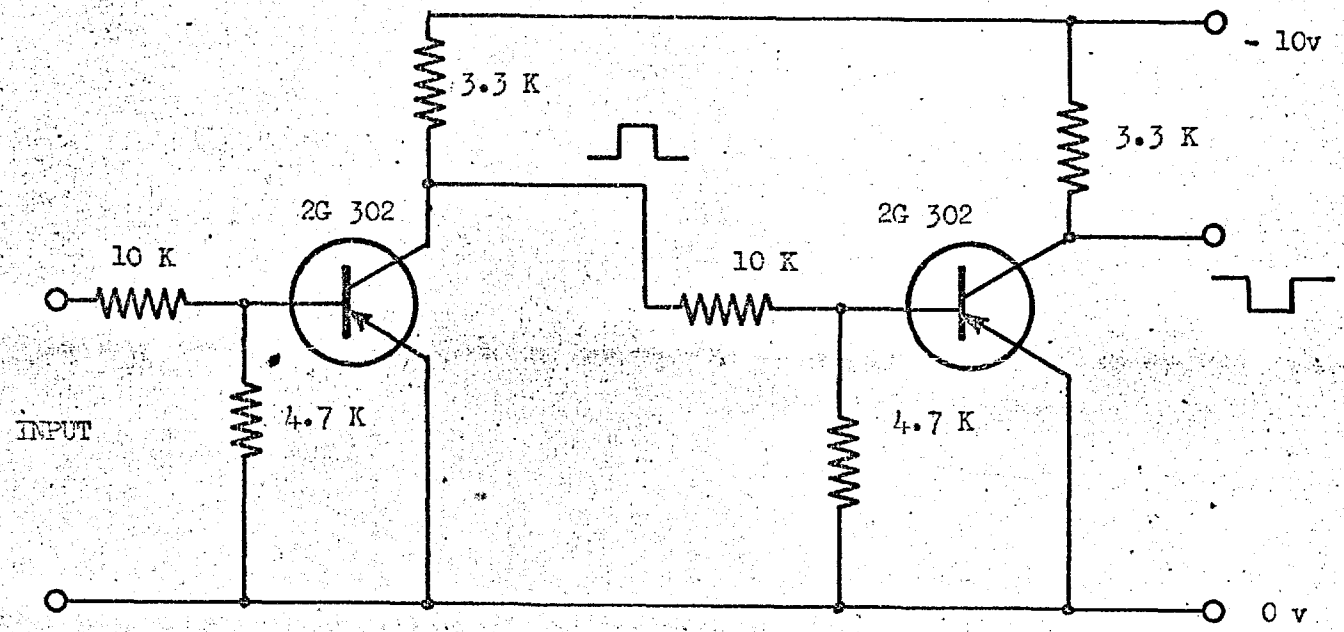
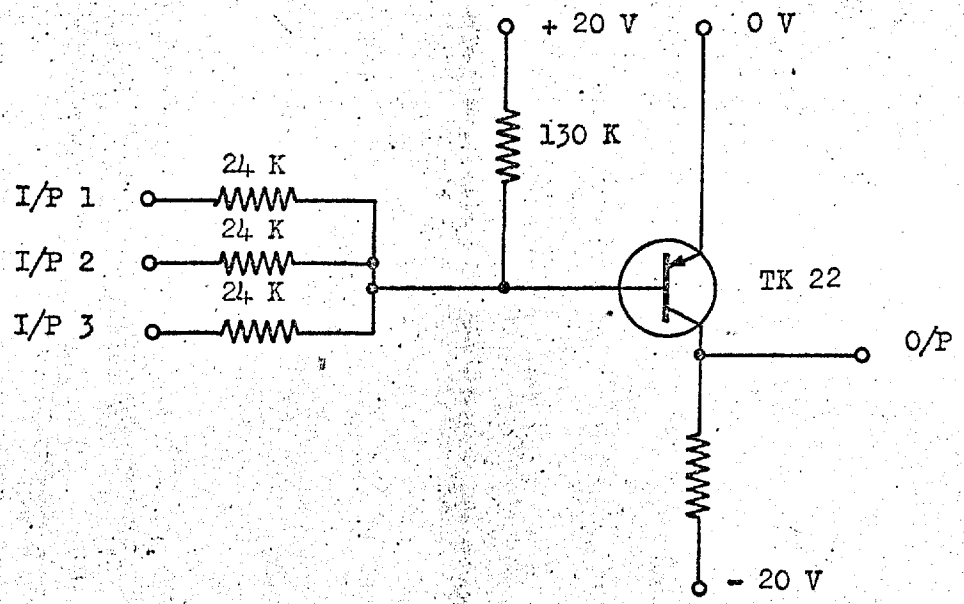


FIG. A.2.7. COMPARATOR OF BINARY DIGITS



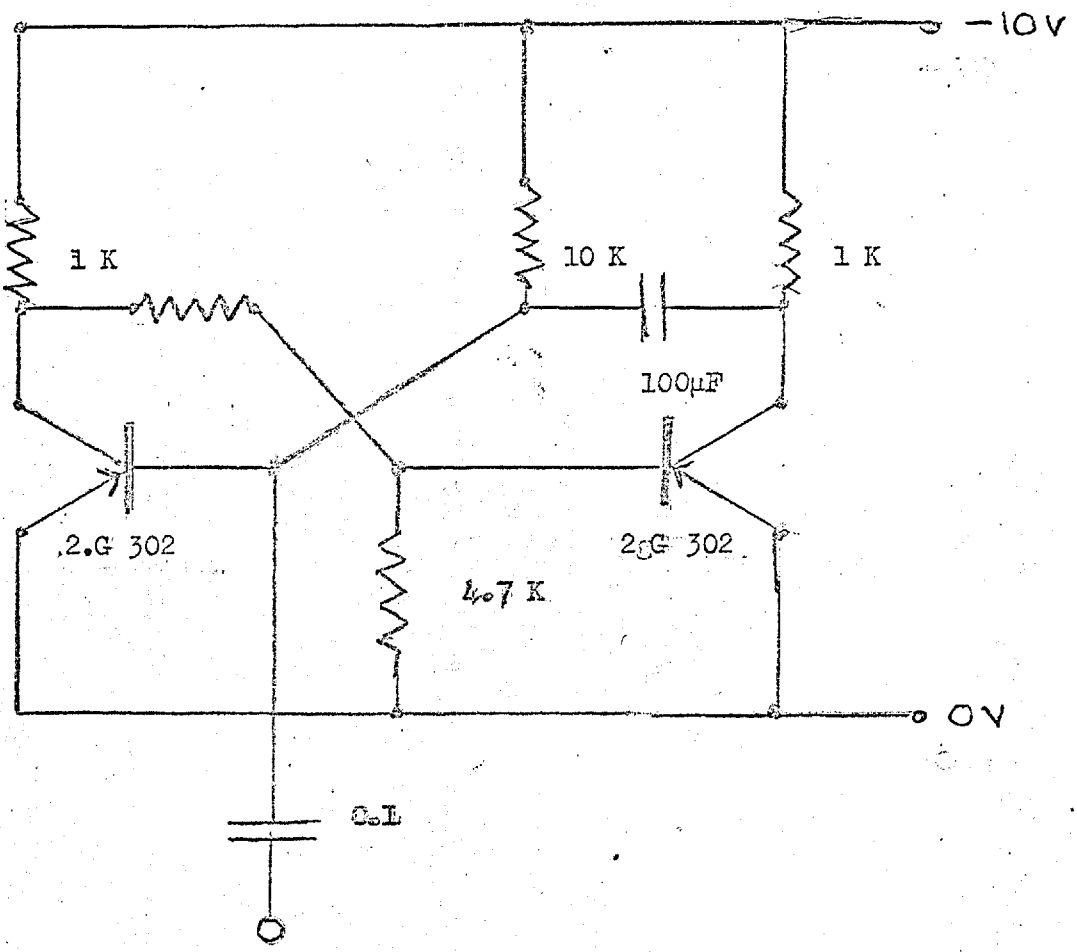
INVERTER AMPLIFIER.

FIG. A.2.8.



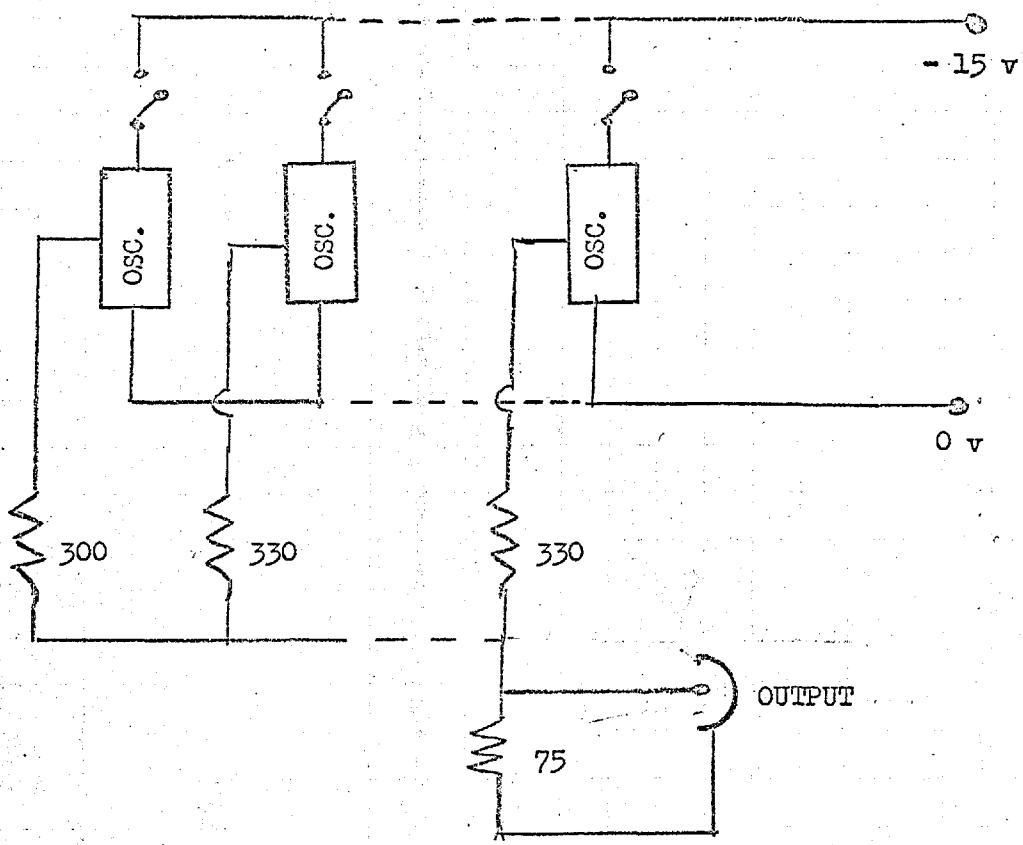
TRL 'NOR' CIRCUIT.

FIG. A.2.9



MONOSTABLE.

FIG. A.2.10



ARTIFICIAL SPECTRUM GENERATOR

FIG.A.2. 11

REFERENCES

1. Barber, N.F., "The Optimum Performance of a Wave Analyser", Electronic Engineering, 21, 175, (1949).
2. Bayes, T., "An Essay Towards Solving a Problem in the Doctrine of Chances", Phil. Trans. Royal Society, 1, (iii), p. 370, (1763).
3. Berne List. (International Frequency List), International Telecommunication Union, Geneva, (1961).
4. Bendat, J.S., "Principles and Applications of Random Noise Theory", John Wiley & Sons, New York, (1958).
5. Blair, W.C. and Kaufman, H.M., "Command Control", General Dynamics Corporation, Electric Boat Division, Groton, Conn., Report No. SPD 59-082.
6. Broadbent, D.E., "Perception and Communication", Pergamon Press, London, (1958).
7. Broadbent, D.E., "Vigilance", British Medical Bulletin, (Experimental Psychology), 20, p. 17, (1964).
8. Cherry, E.C., "On Human Communication", John Wiley & Sons, New York, (1957).
9. Cherry, E.C., "The Cocktail Party Problem", Discovery, pp. 32-35. (March 1963).

10. Cramer, H., "Mathematical Methods of Statistics", Princeton University Press, Princeton, U.S.A. (1963).

11. Davies, I. L., "On Determining the Presence of Signals in Noise", J.I.E.E., 99, Part III, p. 45, (1952).

12. Davenport, W.B. and Root, W.L., "The Theory of Random Signals and Noise", McGraw-Hill Book Co., New York, (1958).

13. Fubini, E.G. and Johnson, D.C., "Signal to Noise Ratio in A.M. Receivers", Proc. Inst. of Radio Engineers, 36, p. 1461, (1948).

14. Gabor, D., "Theory of Communication", Proc.I.E.E., Vol. 93, Part III, p. 429-437, (1946).

15. Goldman, S., "Frequency Analysis, Modulation and Noise", McGraw-Hill Book Co., New York, (1948).

16. Harris, B., et al. "Optimum Decision Feedback Systems", Record of the I.R.E. National Convention Record, Part II, p. 3, (1957).

17. Hartley, R.V.L., "Transmission of Information", B.S.T.J., 7, pp. 535-563, (1928).

18. Hok, G., "Response of Linear Resonant Systems to Excitation of a Frequency Varying Linearly with Time", J.App. Phys., 19, 232, (1948).

19. Lee, Y.W., "Statistical Theory of Communication, John Wiley & Sons Inc., New York, (1960).

20. Lerner, R.M.,"Means for Counting 'Effective' Numbers of Objects or Durations of Signals", Proc. I.R.E., Vol. 47, p. 1653, (1959).

21. Mackworth, J.F., and Taylor, M.M., "A Survey and Analysis of Vigilance", Canadian Joul.Psychol., 17, p. 302, (1963).

22. Middleton, D., "An Introduction to Statistical Communication Theory", McGraw-Hill Book Co. Inc., New York, (1960).

23. Middleton, D., "Statistical Criteria for the Detection of Pulsed Carriers in Noise", 1, 11, J.Appl.Phys., 24, p. 371, 379, (1953).

24. Middleton, D., "Statistical Theory of Signal Detection", I.R.E. Trans., on Inform. Theory, IT-3, 26, (1954).

25. Mood, A.M., "Introduction to the Theory of Statistics", McGraw-Hill Book Co. Inc., New York, (1950).

26. McNicol, R.W.E., "The Fading of Radio Waves of Medium and High Frequencies", Proc. I.E.E., 96, art III, p. 517, (1949).

27. Neyman, J., and Pearson, E.S., "On the Use and Interpretation of Certain Test Criteria for Purposes of Statistical Inference", Biometrika, 20A, 175, p.263, (1928).

28. Nyquist, H., "Certain Factors Affecting Telegraph Speed," B.S.T.J., 3, p.324, (1924).

29. Ryle, M., "Radio Astronomy", The Physical Society Reports on Progress in Physics, 13, p.184, (1950).

30. Saul, Robert, and Luloff, Elaine, "Designing a Spectrum Analyser", Electronic Industries, p. 66, (April 1959).

31. Shannon, C.E., "A Mathematical Theory of Communication," B.S.T.J., 27, p.379 and 623, (1948). Also in book form with Weaver, W., University of Illinois Press (1949).

32. Thomas, H.A., "Some Measurements of Atmospheric Noise at High Frequencies", Proc. I.E.E., 97, Part III, p.335 (1950).

33. Tremellen, K.W., and Cox, J.W., "The Influence of Wave-Propagation on the Planning of Short Wave Communication", Journal I.E.E., 94, Part IIIA, p. 200, (1947).

34. Williams, M., Edward., "Radio Frequency Spectrum Analysers", Proc. I.R.E., 8, (1948).

35. Wald, A., "Sequential Analysis", John Wiley & Sons, New York, (1948).

36. Wilks, Samuel S., "Mathematical Statistics", John Wiley & Sons, New York, (1962).

37. Whitehead, S., "A Tentative Statistical Study of Domestic Radio Interference", Joul. I.E.E., 90, Part III, p.181, (1943).

BOOLEAN TWO-TERMINAL ANALYSIS AND SYNTHESIS

S. Okada

General Dynamics/Electronics

Research Division

Rochester, New York

K. P. Rajappan [1]

Mathematical Department, University of Rochester

K. P. Young[2]

I. INTRODUCTION

In 1939 the applicability of the incidence matrices to switching circuit design was exemplified in O-Trc (see References). In 1954 the cut-set method, multi-contact case, and the existence of the pseudo-ties with a factor such as xx' and pseudo-cuts with a pair of terms such as $x \smile x'$ were pointed out in O-tasc 281 and 289. After R. Gould achieved excellent results by the conventional use of the loop matric method in G-gsc and G-ncn (see also HiG-LODEC 196). This matric method was further developed as a rigorous method in OMY-bim. In this method the consideration of the non-Abelian group of the invariant transformations of variables of the given Boolean function reduced the amount of calculation and provided many useful graphs. In 1960 the solution was programmed for an IBM650 computer by Y. Moriwaki.

C. Saltzer showed the relationship between the topological coboundary operation (cut-set method) and the product form of the Boolean functions in S-atcas, and W. Mayeda's paper M-ssg described a loop basis method. The present paper reports the further results on this topic after OMY-bim.

II. A NEW BOOLEAN LATTICE FOR BOOLEAN FUNCTIONS

In the minimization of Boolean functions, the N-cube of 2^N points was used for N variables. However, if the K-cube ($K = 2^N$) of 2^K points and its subgraph are used, the standard sum forms a subset of its atoms, and the standard product forms a subset of the dual atoms (B-LT 163), and the elements of height 2, 3, . . . are $\binom{K}{2}$, $\binom{K}{3}$, . . . The Karnaugh map method can be expressed on this lattice.

III. ENUMERATION OF LOOPS IN GRAPHS

L. Lunelli and E. Biondi enumerated the number of loops passing a branch of a complete graph of n nodes in LB-el:

$$M(n) = \sum_{q=3}^{n} \binom{n-2}{q-2} (q-2)!$$

Based on this combinatorial reasoning, the Boolean functions of complete graphs of n nodes are easily given. If the branches of nodes 12, 13, . . . are called a, b, . . ., then the functions for the branch a are given:

$n = 3$, $M(3) = 1$, $f = bc$,
$n = 4$, $M(4) = 4$, $f = bd \smile ce \smile bef \smile cdf$,
$n = 5$, $M(5) = 15$,

$f = be \smile cf \smile dg \smile bfh \smile bgi \smile ceh \smile cgj \smile dei \smile$ [illegible] $\smile$

$bfij \smile bghj \smile ceij \smile cghi \smile dehj \smile$ [illegible]fhi.

The branch a itself can also be included. If any parallel branches are substituted by their equivalent single branch, any graph is a subgraph of the complete graph of the same number of nodes. Therefore, for example the branch a, b, . . . consists of parallel branches A, B and C; D and E, . . . We obtain the Boolean function by substituting $a = A \smile B \smile C$, $b = D \smile E$, . . . into $F = a \smile f$. The result of this substitution is $F = (A \smile B \smile C) \smile (D \smile E)$ (. . .). . . (. . .). The nonexistence of a branch, for example, g is expressed by $g = 0$.

IV. REALIZATION OF GRAPHS FROM THEIR CUT-SET MATRICES (OY-Arc)

For any oriented graph, there always exists a nonoriented graph obtained by the elimination of orientation. Hence, a routine procedure of realizing graphs from any cut-set matrix consisting of only 0 and 1 (mod. 2) is given here. It can be easily generalized to a partly or fully oriented set of branches.

The cut-set matrix $B^a_{\varkappa}$(a, b, . . . = 1, 2, . . . B; $\varkappa$ = 1, 2, . . . , A) of which the rows are linear combination of node coboundaries is generally

[1] Formerly General Dynamics/Electronics, Research Division, Rochester, New York.

[2] Formerly Microwave Research Institute of Brooklyn Polytechnic Institute, New York.

Appeared in Proc. of 2nd Symposium on Switching Circuit Theory and Logical Design American Institute of Electrical Engineers, September 1961.

flat (B < A). Realization of the branch loop incidence matrix $C_p^{\varkappa}$ (p, q = 1, 2, . . . , C; B+C=A) is done exactly in the same manner starting from the step 1''.

1) Left-diagonalization. By a linear combination (mod. 2) of rows and only an interchange of columns, the left square submatrix B^a_b can be brought into the diagonal form $B^a_b = 1^a_b$ (1^a_b: unit matrix), and the remaining right submatrix B^a_p will be called the right core.

1') Shunt-elimination. If a flat cut-set matrix $B^a_{\varkappa}$ has a set of two or more equal columns or a collection of such sets, then each set expresses a set of branches in parallel. Therefore, all columns of parallel branches are eliminated by leaving the first column of each set and deleting all other equal columns. Thus $B^a_{\varkappa}$ is reduced to $B^a_{\varkappa'}$. A record of the deleted branches with their corresponding undeleted parallel branches should be kept.

2') Transition to its null-factor $C_{q'}{}^{\varkappa'}$. If the right core $B^a_{p'}$ is adopted as an upper core $C^a_{p'} = B^a_{p'}$ and the unit matrix C^q_p is attached under it, the resulting lower-diagonalized

$$C^{\varkappa'}_{p'} = \begin{bmatrix} C^a_{p'} \\ C^{q'}_{p'} \end{bmatrix}$$

forms the null-factor (mod. 2) of $B^a_{\varkappa'}$, hence the loop matrix. As is well-known: $B^a_{\varkappa'} C^{\varkappa'}_{p'} = 0$. Then this tall $C^{\varkappa'}_{p'}$ is transposed to a flat matrix $C_{p'}{}^{\varkappa'}$.

1'') Series-elimination. If a flat loop matrix $C_{q'}{}^{\varkappa'}$ possesses sets of equal columns (branches in series similar to 1')), all columns of the series branches should be deleted, and a record should be kept of the deleted branches. Thus it is reduced to $C_{q'}{}^{\varkappa'}$.

2'') Transition to its null-factor $B^{a''}{}_{\varkappa''}$. This is done by transposing the flat $C_{q'}{}^{\varkappa''}$ to a tall $C^{\varkappa''}_{q'}$ and adopting its upper core $C^{a''}{}_{q'}$ as the right core $B^{a''}{}_{q'}$ and attaching a unit matrix $B^{a''}{}_{b''}$; thus $B^{a''}{}_{\varkappa''}$ is obtained.

The steps 1' - 2'' should be repeated until all equal columns of $B^{a(i)}{}_{\varkappa(i)}$ are eliminated. If a realizing graph is a series-parallel type, the matrix $B^a_{\varkappa}$ is reduced to a single unity by a finite number k of the above four operations. This graph of series-parallel order k is realized by the records of steps 1' and 1''. Otherwise, all series-parallel parts are eliminated. However, because the following process realizes the graphs independent of whether the graph is series-parallel or not, the above four steps can be skipped.

Characteristics of cut-set matrices. For any cut-set, a Jordan curve can be drawn which "cuts" all branches of the cut-set at any points of the curve. Each branch can be located freely as far as one terminal is inside and the other is outside. This Jordan curve is called the ambit of this cut-set. Each branch of the tree is called a trunk. Then the necessary and the sufficient condition that any flat matrix $B^a_{\varkappa}$ of only 0 and 1 elements can form a cut-set matrix of a possible graph, is that all ambits of B semidiagonalized rows, namely uni-trunk cut-sets, are realizable without mutual crossing of these ambits on a Euclidean plane (ambit theorem). If all ambits with cutting branches are realized without mutual crossing, the graph is immediately realized by the following simple process. At first, Realization of ambits will be described. On a finite Euclidean plane, ambits of cut-sets of the right core B^a_p are realized one by one. The first two ambits of $B^1{}_p$ and $B^2{}_p$ can be drawn so as not to include each other without loss of generality, because the case where one is contained in the other can be transformed to the above case without change of cutting branches. This is easily seen by the aid of stereoprojective mapping of the whole plane. Such an inversion of outside and inside of an ambit is called the Jordan inversion. From the third ambit, all possible realizable cases should be considered, this can be determined by elementary combinatorial properties of cut-set branches relative to the ambits, since each branch can have exactly two ends, and a new ambit should never cross the already drawn ambits. When the realization at any step is not unique, the figure of ambits that simultaneously expresses all possible cases prevents exponential growth of figures; especially in the frequent case of two possibilities, a revolving door can be built in a part of wall of an ambit, and the door is separated into two semicircles by a partition vertical to the ambit wall where one ambit is drawn in one of the semicircles. By revolving this ninety degrees, the inner ambit can be either inside or outside of an ambit. Such doors can be combined in any complicated manner. The order of realization of such ambits does not change the final realizability. Hence, if at some intermediate step any ambit becomes unrealizable, this forms a rigorous proof of the nonexistence of any graph which has the given $B^a_{\varkappa}$ as cut-sets. In the a. c. case, the unrealizable rows of cut-sets can be realized by ideal transformers as exemplified in O-ats. Hence, this yields a general method of determining the minimum necessary number of ideal transformers for a given matrix $B^a_{\varkappa}$. If all ambits are realized, one trunk for each ambit should be crossed with the ambit, and thus the tree part is supplemented and then the following process realizes the graphs.

Termination to graph. The outer area of all ambits corresponds to a datum node (ground of height zero) or the root as a tree, which should be drawn in any position within this area, and all ends of branches in this area should be extended and connected to this datum node. An area of an ambit, excluding areas of all other inner ambits, is called the terrace. Then each terrace corresponds to a single node which can be drawn as any point within this terrace and all branches terminating in this terrace should be connected to this node. An ambit that has no inner ambits forms a terrace by itself.

V. TOPOLOGY OF BOOLEAN TWO-TERMINALS

The switching network theory developed here is independent of Ohm's law or any other branch causality characteristic expressed as the metric of the branch number-dimensional space. Therefore, it can be called the affine network theory (O-tasc 267, 275, 279; O-ats). From this point of view, the node-branch incidence matrix is expressed as

$B^{\alpha}{}_{\varkappa}$ where $\alpha, \beta, \gamma = 1, 2, \ldots, \alpha^0$;

α^0 = no. of nodes; $\varkappa, \lambda, \mu, \nu = 1, 2, \ldots, A$;

$A = \alpha^1$, α^1 = no. of branches. For example,

$$B^{\alpha}{}_{\varkappa} = \begin{bmatrix} 1 & . & . & 1 & 1 & . & . & . \\ . & 1 & . & . & . & 1 & 1 & . \\ . & . & 1 & . & . & . & 1 & 1 \\ . & . & . & 1 & . & 1 & . & 1 \\ . & . & . & . & 1 & . & 1 & . \\ 1 & 1 & 1 & . & . & . & . & . \end{bmatrix} \quad \text{(mod. 2)} \qquad (1)$$

Each row can be regarded as a base covector expressed by a pair of the initial and terminal hyperplanes O and B^{α}. Geometrically it will be called the layer and will be denoted with a lower arrow:

$$\underrightarrow{b}^{\alpha} = \underrightarrow{OB}^{\alpha} = B^{\alpha}{}_{\varkappa}\, \underrightarrow{a}^{\varkappa} \; , \qquad (2)$$

where $a^{\varkappa}$ is the base layer of the A-dimensional affine space. Then the coboundary operator δ of a node α corresponds to the determination of the A-component of the layer $\underrightarrow{b}^{\alpha}$ by a multiplication of the base vector $\vec{a}_{\lambda}$:

$$\delta\alpha = \delta \underrightarrow{b}^{\alpha} = (\underrightarrow{b}^{\alpha}\, \vec{a}_{\lambda})\, \underrightarrow{a}^{\lambda} = B^{\alpha}_{\lambda}\, \underrightarrow{a}^{\lambda} \; . \qquad (3)$$

If $\underrightarrow{b}^{\alpha}$, $\vec{a}_{\lambda}$, and $\underrightarrow{a}^{\lambda}$ are expressed by $\underrightarrow{\alpha}$, $\vec{\lambda}$, and $\underrightarrow{\lambda}$, then, for example,

$$\delta 1 = (\underrightarrow{\alpha}\, \vec{\lambda})\, \underrightarrow{\lambda} = B^{\alpha}{}_{\lambda}\, \underrightarrow{a}^{\lambda} = \underrightarrow{a}^{1} + \underrightarrow{a}^{4} + \underrightarrow{a}^{5} \qquad (4)$$

$$= \underrightarrow{1} + \underrightarrow{4} + \underrightarrow{5}. \qquad (5)$$

Each column of the matrix $B^{\alpha}{}_{\varkappa}$ is regarded as the affine (parallel) projection of the base vector $\vec{a}_{\varkappa}$ on B-space:

$$(\vec{a}_{\varkappa}\, \underrightarrow{b}^{\alpha})\, \vec{b}_{\alpha} = \vec{b}_{\varkappa} \; , \qquad (6)$$

where $\underrightarrow{b}^{\alpha}$ is the base layer of B-space. Then the boundary operator ∂ of a branch $\varkappa$ corresponds to this affine projection:

$$\partial \vec{\varkappa} = \partial \vec{a}_{\varkappa} = (\vec{a}_{\varkappa}\, \underrightarrow{b}^{\alpha})\, \vec{b}_{\alpha} = B_{\varkappa}{}^{\alpha}\, \vec{b}_{\alpha} \; , \qquad (7)$$

for example,

$$\partial \vec{1} = (\vec{1}\, \underrightarrow{\alpha})\, \vec{\alpha} = \vec{1} + \vec{6} \; , \quad \underrightarrow{\alpha} = \underrightarrow{b}^{\alpha} \; . \qquad (8)$$

The rank B of the matrix $B^{\alpha}{}_{\varkappa}$ is $\alpha^0 - 1$; and

$$B = \alpha^0 - 1 \; . \qquad (9)$$

The number of all sets of nodes including the node $\dot{1}$ and excluding node $\dot{6}$ is

$$H = 2^{\alpha^0 - 2} = 2^{6-4} = 2^4 = 16 \quad \text{for (1).} \qquad (10)$$

If the coboundaries of these H sets of nodes are calculated, it follows that

$$B^{h}{}_{\lambda} = B^{h}{}_{a}\, B^{a}{}_{\lambda} \; , \qquad \begin{matrix} h = 1, 2, \ldots, H, \\ a = 1, 2, \ldots, B. \end{matrix} \qquad (11)$$

For example,

$$B^{\dot{1}'}{}_{\lambda} = B^{\dot{1}'}_{\lambda} = B^{\dot{1}'}{}_{a} B^{a}{}_{\lambda} = [\,1 \ldots .]\, B^{a}{}_{\lambda} = [1..11....] \; ,$$

$$B^{\dot{2}'}{}_{\lambda} = B^{\dot{2}'}{}_{a} B^{a}{}_{\lambda} = [\,11 \ldots .]\, B^{a}{}_{\lambda} \qquad = [11.1111..] \; .$$

In this case,

$$B^{\dot{1}'}{}_{\lambda} \leq B^{\dot{2}'}{}_{\lambda} \quad \text{for all } \lambda \; ,$$

and the inequality holds for at least one λ. Then the covector $B^{\dot{1}'}{}_{\lambda}$ is said to be smaller as a covector:

$$\underrightarrow{b}^{\dot{1}'} < \underrightarrow{b}^{\dot{2}'} \quad \text{or} \quad \dot{1}' < \dot{2}' \; .$$

When there is such an order relation, the larger covector separates the graph into more than two parts, and hence larger covectors should all be deleted. In the Boolean product expression, the elimination of the larger Boolean factor is automatically included in the general Boolean algebraic theory. Furthermore, a covector $B^{h}{}_{\lambda}$ that includes a pair or pairs of "make" and "break" contacts of the same "relay" topologically forms a cut-set but not a Boolean cut-set, hence it will be called a pseudo-cut. All pseudo-cuts are deleted. The remaining covectors $\underrightarrow{b}^{h'} = B^{h'}{}_{\lambda}\, \underrightarrow{a}^{\lambda}$ form all of the cut-sets that separate node 1 from node 6. Then the result can be expressed as the product sum form of the Boolean function of the given switching network. Furthermore, if two branches λ and $\lambda + 1$ are governed by the same relay R, there holds

$$B^{h}{}_{\lambda} \smile B^{h}{}_{\lambda+1} = B^{h}{}_{R}. \quad \text{(O-tasc 281 (5, 14))}$$

Thus, each factor of product form of the Boolean function forms only a subset of the coboundary of all possible combinations of $\alpha^0 - 2$ nodes. This forms the foundation of the Boolean two-terminal synthesis based on the node-coboundaries.

VI. SYNTHESIS

Suppose that the Boolean function is given either in the standard product or minimum product form. Then, all other possible product expressions will be obtained. However, because the number of contacts of the standard relization of the minimum product is known, only those expressions that have at most this number are sufficient to consider.

Under this upper bound, the realizability of the circuits beginning from the single contact case can be investigated one by one by a gradual increase of contacts under consideration of the pseudo-cuts.

REFERENCES

B-LT - G. Birkhoff: Lattice Theory, Amer. Math. Soc. Col. Pub. 25 (1940, rev. 1948) A. M. S. Providence, R. I.

G-gsc - R. Gould: The Application of Graph Theory to the Synthesis of Contact Networks, Proc. Internat. Symp. on Theory of Switching, Part I, Ann. of Comp. Lab. Harvard Univ. vol. 29 (1959) 244-92.

G-ncn - R. Gould: A Note on Contact Networks for Switching Functions of Four Variables, IRE Trans. PGEC-7. 2(Sept. 1958) 196-9.

HiG-LODEC - R. A. Higonnet and R. A. Grea: Logical Design of Electrical Circuits (1958) McGraw-Hill, N. Y.

LB-el - L. Lunelli and E. Biondi: Enumeration of Loops in Graphs, Rend. Institut Lombardo di Scienze e Lettere 91 (June 27, 1957) 903-11, 912-26. In Italian.

M-ssg - W. Mayeda: Synthesis of Switching Functions by Linear Graph Theory, IBM J. 4. 3 (July 1960) 321-8.

O-ats - S. Okada: Algebraic and Topological Foundations of Network Synthesis, Proc. Symp. on Modern Network Synthesis of Polytechn. Inst. of Brooklyn, N. Y. (April 13-5, 1955) 283-322.

OMY-bim - S. Okada, Y. Moriwaki and K. P. Young: Realization of Boolean Polynomials Based on Incidence Matrices, Proc. of 9. Eastern Joint Comp. Conf. Boston (Dec. 1-3, 1959) 120-32.

OO-NF - S. Okada and R. Onodera: Theory on Interlinked Electromagnetic Networks and Fields in the Tensor Geometry of Linear Space, Memoirs of the Unifying Study of the Basic Problems in Eng. Sc. by Means of Geometry, ed. by K. Kondo, vol. I (1955) Tokyo. p. 1-112.

O-tasc - S. Okada: Topology Applied to Switching Circuit Theory, Proc. Symp. on Information Networks of Polytedm. Inst. of Brooklyn, N. Y. (April 12-4, 1954) 267-90.

O-Trc - S. Okada: On a Theory of Relay Circuits, Lecture note at Nippon Elec. Co. Tokyo (May 22, 1939) p. 24 in Japanese.

OY-Arc - S. Okada and K. P. Young: Ambit Realization of Cut-Set Matrices into Graphs, Res. Rep. PIBMRI-801-60 (Nov. 21, 1960) Polytechn. Inst. Brooklyn, N. Y.

S-atcas - C. Saltzer: Algebraic Topological Methods for Contact Network Analysis and Synthesis, Quart. App. Math. 17. 1 (April 1959) 173-83.