DIGITAL TRANSMISSION SYSTEMS OPERATING OVER

HIGH FREQUENCY RADIO CHANNELS

by

WILLIAM HODGKISS

A thesis submitted for the degree of

Doctor of Philosophy

in the Faculty of Engineering

University of London

July 1980

TO MY FATHER

ABSTRACT

This thesis is concerned with methods of transmitting digital information over high frequency (HF) radio channels. The flexibility of radio as a means for communication arises from the fact that no physical interconnection is required between transmitter and receiver, whilst at the HF band the low level of transmission loss makes possible long ranges of communication at low power levels. The expanding need for digital communication facilities has resulted in the increased importance that is being attached to HF radio channels.

Techniques for the serial transmission of digital information are investigated. The basic binary modulation and demodulation methods, and the relationship of the bit error rate to the rate of transmission, are evaluated. The usefulness of adaptive equalisation in countering intersymbol interference caused by the multipath characteristic of the HF channel is also determined. Trials over a "real" HF link were carried out and reported on in the thesis.

The results of the trials show that the phase-shift-keying method of modulation gives superior performance to amplitude shift keying. Coherent detection in the form of a phase-locked loop is shown by the results to be more suitable to the HF environment than incoherent detection since it permits channel equalisation. Equalisation was found to be desirable on account of the presence of multipath which is frequently encountered in HF channels. The transmission bit rates tested ranged from 0.4 kilobits per second to 3.6 kilobits per second, and the overall average bit error rates, after equalisation, were measured as varying from 5.7×10^{-2} at the lowest bit rate to 1.4×10^{-1} at the highest rate. Equalisation was seen to improve the average error rate by a factor of 1.7 at a bit rate of 2.1 kilobits per second.

CONTENTS

ACKNOWLEDGEMENTS

I would first and foremost like to thank my supervisor, Dr. L.F. Turner, for his help and guidance throughout the course of the research and the preparation of this thesis. I would then like to thank Mr. John Pennington of the Admiralty Surface Weapons Establishment for his assistance both in the respect of advice given and for the provision of resources for the research.

Thanks are also due to many colleagues at Imperial College, especially Bob Andrews, William Edmondson, Alex Lax, Mike Perdios and Chen Hui Min, and I would like to pay particular tribute to Ian and Vivien Colyer for their support and friendship during this work.

I am also grateful to the Science Research Council and to my father for financial assistance, and finally I would like to express my appreciation to Miss Cynthia Collins for her swift and diligent typing of the manuscript.

List of Illustrations

-14-

LIST OF COMMON SYMBOLS AND ABBREVIATIONS

%	percentage
∞	infinite quantity
Σ	summation
$\int$	integration
$\delta(t)$	dirac-delta function
"0" and "1"	the binary digital symbols
f	frequency variable
f_p	plasma frequency
$f(x)$	function of variable x
$F_1(x,y)$	
$F_2(x,y)$	functions of variables x and y
h_v	virtual height
j	square root of minus one
n	integer variable
P_e	probability of error
t	time variable
w	radian frequency variable
w_c	radian carrier frequency
z	unit advance $= \exp(jwT)$
z^{-1}	unit delay $= \exp(-jwT)$

ASK	amplitude shift keying
ASK+C	ASK with additional carrier
ASK-SC	ASK with suppressed carrier
ASK-rec.det.	ASK detected by rectifier
b/s	bits per second
cc	cubic centimetre
cos	cosine
dB	decibel
d.c.	direct current
DPSK	differential phase shift keying
DSB	double sideband
erfc x	complimentary error function of x
FSK	frequency shift keying
GHz	gigacycles per second
HF	high frequency (radio)
Hz	cycles per second
IF	intermediate frequency
Kb/s	kilobits per second
KHz	kilocycles per second
Km	kilometres
KW	kilowatts
MHz	megacycles per second
mS	milliseconds
MSE	mean squared error
MUF	maximum usable frequency
pdf	probability density function
PLL	phase-locked loop
PSK	phase shift keying
rms	root mean square

sin	sine
S/N	signal-to-noise
SSB	single sideband
tan	tangent
UHF	ultra high frequency
VCO	voltage-controlled oscillator
VHF	very high frequency
W	watts

CHAPTER ONE

INTRODUCTION - A HISTORICAL PERSPECTIVE

At 12.30 p.m. on the twelfth of December 1901, at an old
military barracks in St. John's, Newfoundland, three faint clicks
were heard from a receiver, and the letter S appeared on a telegraph
tape machine. Thus was heralded a revolution in long-distance
communications because this brief message had been transmitted from
Poldhu, in Cornwall, by radio. A twenty-eight year-old Italian,
Guglielmo Marconi, had confounded the sceptics who said that radio
communication over the horizon was not possible.

In 1865 Clerk Maxwell had proved that electro-magnetic waves
could be generated by electrical oscillations and that they would
propagate in space with the speed of light. It took until 1888
though for Heinrich Hertz to perform the pioneering experimental work.
He showed that the waves did indeed have the same velocity as light
and that, like light, they underwent reflection, refraction, interference,
and polarisation. Further experiments by Sir Oliver Lodge in 1894 and
by Marconi in 1896 increased the range of the transmissions. However,
because of their similarity to light in propagating in straight lines,
it was thought to be impossible to transmit and receive electro-magnetic
waves over the horizon. It took Marconi's famous trans-Atlantic
experiment to disprove this.

The phenomenon was independently expained in 1902 by Heaviside
and Kennelly. They suggested that there was an electrical reflector in
the earth's atmosphere, which subsequently became known as the
Kennelly-Heaviside layer. It is now more commonly termed the E layer.

.In 1924, Sir Edward Appleton confirmed its existence by experiment, and found its height to be 60 to 70 miles above the earth. Later Appleton fired short radar-type pulses at the ionosphere and found that some reflection was taking place from a higher level of about 150 to 180 miles above the earth. This was christened the Appleton layer but is now called the F layer.

At the time of Marconi's experiment transmissions were radiated by spark discharge across the plates of a capacitor. The use of aerials had yet to be appreciated,and little progress was made over the next twenty years.

The valve transmitter amplifier was introduced in 1919, and during the early nineteen-twenties Marconi established the technique of using beamed-array aerial systems over long transmission distances. Good reception was shown to be possible between England and Australia. By 1927, with the success of the Empire Radio Link, it could be said that long-distance radio communications was well proven.

In the thirties progress was made with the establishment of radio-telephone links and with the development of techniques such as single-sideband transmission (SSB). The advantages claimed then by Reeves[1] for SSB over double sideband (DSB) still hold good today. In particular the advantages are:

(1) The halving of the bandwidth of the modulated signal, and

(2) the improvement in the received signal-to-noise ratio (this is because the transmitter power goes to one sideband).

The development of digital communications, in the form of telegraph and teleprinter services, proceeded at a much slower pace.

The limitations imposed by the HF channel were appreciated even though its behaviour was not widely understood at this stage.

Appleton continued to be the major researcher into the ionosphere at this time. Daily measurements of the diurnal changes in the E layer were taken[2], and the first method of critical-frequency prediction was to be formulated on these results[3]. Appleton also postulated the dependence of E-layer ionisation on the sunspot cycle[4]. In addition he tried to predict the maximum usable frequency (MUF) of the F layer (since for optimum reception it is necessary to transmit at just below this frequency). However, anomalies began to appear in the effect of seasonal changes and geographical locations on the F-layer critical frequency[5]. It was not until during the Second World War, when more stations investigating the ionosphere had been set up in other parts of the world, that sufficient data was available for world-wide MUF prediction of the F layer.

In the early thirties the aerial systems in use were mainly of the reasonant type, e.g. half-wave dipoles. Directivity was achieved by combining these sorts of aerials to give an interference pattern which led to the desired beam pattern. The effects of fading at the receiver were reduced by choosing the best signal from two or more aerials (or by a combination process). This method is still employed today and is known as Space Diversity reception. In the late thirties the non-reasonant array, particularly the Rhombic aerial, was introduced. Reasonant aerials are at a disadvantage when the operating frequency has to be varied during the day (or night) in order to obtain the best propagating conditions.

By the Second World War the state of HF communications was fairly advanced, except for the digital aspect. The deleterious efforts of the ionospheric medium on digital transmission had not been overcome

and, because of the availability and efficiency of cable links, there was little incentive to do so. However the physical vulnerability of these cable links in wartime conditions gave a boost to digital radio research. It was found that a two-tone system for the two digital bits gave a better performance than on-off keying but even so, apart from this,no real advance was made.

By the end of the War the HF prediction service provided world-wide coverage. Important work by Mitra[3] had improved forecasting.

As a result of congestion in the HF band the Administrative Radio Conference was convened in 1947. The whole of the frequency spectrum from 10 KHz to 10.5 GHz was allocated. A body called the International Frequency Registration Board was set up to oversee the frequency spectrum with particular regard to the HF band (which they defined as being from 4 to 27.5 MHz).

Developments in the digital field in the 1950's were boosted by Shannon's work at the end of the previous decade[6,7]. Van Duuren[8] produced the first error-correcting system which used an automatic request (from receiver to transmitter) for re-transmission. Also frequency diversity techniques were analysed[9] enabling the correct tone spacing to be employed in FSK systems.

As far as ionospheric forecasting was concerned the state of the art had improved considerably by the sixties. The previous method had involved the prediction of the sunspot number and, for a six month forecast of the MUF (this length of time being necessary in order to distribute the forecast data world-wide), the error could be as high as 30% by the sixth month. A new system was introduced by Naismith et al[10] which involved comparing the trend of the critical-frequency curves to that of a similar trend from an earlier period.

Advances in the hardware field in the period immediately preceding the present day revolved around the use of synthesised frequency generators. A synthesised frequency is generated by a series of phase-locked loops locked to a master crystal oscillator. This can be realised in all-digital form, and frequencies accurate to six-or seven-significant figures can be "dialled up" either manually or by remote control. The technique can be used both for the carrier-frequency generator and for the receiver. This together with the use of non-reasonant aerials and wideband amplifiers (which became available about the same time) enables the system operating frequency to be changed in a matter of seconds. Hence the diurnal changes in the optimum propagation frequency can be followed without long breaks while the transmitter and receiver are retuned. Having summarised the development of HF communications (much of the material for this summary has been taken from Betts[11]), the emphasis is now shifted towards the subject of the digital communication over HF, which forms the theme of this work.

Outlets for digital communication have expanded considerably with the introduction of the digital computer. This growth has been accelerated by the rapid advance in recent times in integrated-circuit techniques, which have facilitated the construction of increasingly powerful computers. Both industrial and military organisations have taken advantage of the enormous power of the computer for collating, processing and storing data, and for other purposes such as automation.

Computers are often required to have access to each other for the transfer and processing of information. Since the physical interconnection of one computer to another is usually impractical some means of communication must be provided. Communication systems are are required to enable data analysis and processing to be performed

where a main computer is not available. Further outlets for digital communication have been created by the digitisation of such processes as facsimile and speech transmission. (The digitisation of such processes has obvious applications for military users because of the security advantages.)

The advance in computer and integrated-circuit technology has been applied to the communication process itself, considerably simplifying it and allowing the implementation and sophisticated realisation of such techniques as channel equalisation, matched filtering and coding. This has also given a boost to digital forms of communication.

The result of the developments outlined above is a growing need for digital communication channels. There are three major means of communicating digital data. The first is via land line. Land lines are usually a combination of telephone lines and microwave relays. They tend to be of short ranges both because of the difficulty in connecting telephone lines between countries and because of the construction necessary for microwave waveguides. Furthermore land lines tend to be restricted to the more developed countries. A second commonly-used medium is that of satellites. This, however, is limited primarily to the interconnection of population centres, to military purposes, and to certain special-purpose users such as for American space research. By far the most flexible and accessible means of digital data transmission is radio. Its advantages stem from the fact that its "land lines" are always available, i.e. from the fact that electro-magnetic waves propagate in space and can be refracted by the earth's atmosphere to allow over-the-horizon communication. Hence it can connect mobile transmitters and receivers, and also remote geographical locations to each other and to central bases,

and it has a range of communication varying from line-of-sight to several thousand miles.

VHF radio or land line channels are usually satisfactory for line-of-sight transmission. The difficulty is in communicating over-the-horizon. When radio is⟍the medium transmission occurs
used and the ionosphere is
mainly as a result of two phenomena (according to operating frequency), these being by troposcatter and skywave or HF propagation.

Troposcatter takes place at VHF and UHF frequencies, and so has the advantage of⟍wide bandwidths and hence a large capacity for
potentially
carrying information. However severe transmission loss is present in troposcatter propagation (Bullington[34] gives 60 dB at 100 MHz over 100 miles). The usual explanation[34] for the troposcatter process is that energy is reflected or scattered by turbulent air masses in the troposphere which are positioned at the intersection of the beam widths of the transmitter and receiver aerials. The path loss is because only a fraction of the transmitted energy is "scattered" back to the receiver. Hence troposcatter is not very feasible at distances over 100 miles.

Skywave propagation takes place by what is equivalent to a reflection of the radio wave off the earth's ionosphere. Such propagation occurs at HF frequencies, and useful signals can be received at distances of up to several thousand miles. At night the loss in transmission is only some 3 to 6 dB below free space for each reflection[34]. During the day the loss is somewhat greater because of the presence of an absorbing layer in the ionosphere (see Chapter Two) but it is still small compared to that experienced in troposcatter.

We can therefore conclude that skywave propagation

constitutes the most useful method of digital data communication at all except line-of-sight distances. It has the advantages that transmitter and receiver can be set up anywhere at any distance of up to several thousand miles apart and that they do not require either any physical interconnection or access to a restricted message relayer such as a satellite. However because of its popularity as a transmission medium there is considerable channel crowding, and so it is essential to make the most efficient use of the available bandwidth. In the digital case, then, as much data as possible must be transmitted per channel.

We have thus defined the need for digital communication over HF, and have observed that it must be efficient. The objective of this work is to investigate the techniques required for such efficient digital communication over HF, and to establish the performance of such a system.

The first step in this investigation was as indicated in Chapter Two. The structure and properties of the ionosphere were examined so that the mechanism of HF propagation could be understood. The particular difficulties that were presented to the user of digital communications over HF could then be highlighted. In particular it was found that multiple-mode propagation (commonly termed multipath) which would cause intersymbol interference was likely to exist.

Following this, in Chapter Two, a review of the previous work in digital communications over the HF channel is presented. It is shown that most researchers have concentrated on countering the multipath problem by developing parallel sub-channel systems. It is shown that, while considerable research has been applied to these systems such

.that they are by now well developed, certain intrinsic disadvantages are present in them. As is indicated in the Chapter, comparatively little attention has been paid to the less complex serial-transmission process. Although it is apparent that this process is more directly vulnerable to multipath, it was found in the review that little attempt had been made to counter it by using such techniques as adaptive equivalisation.

As a result of the work carried out in reviewing the material dealt with in Chapter Two the programme of research was decided upon,and is detailed in the last part of Chapter Two. It was decided to evaluate the performance of the fundamental methods of serial-data transmission. It was felt among a number of important questions that required answering were:

(1) What are the best forms of modulation and demodulation?

(2) How does the error probability vary as the bit rate changes from, say, 0.5 Kb/s to 3.6 Kb/s.

It was thought necessary that such results should be established by real trials over a typical HF link, rather by the channel simulation technique that was common among previous researchers. The prospective user of the HF channel could then examine these results in the light of his needs.

Having decided on the programme of research, the test system required to perform such research is then described in Chapter Three. In particular we see from considering the ionospheric reflection process that it is reasonable to assume that the signal from the shortest multipath would always be of significant strength. In other words, for each

pulse transmitted, the first pulse to arrive at the receiver could
be taken as the signal with the remaining multipaths being treated
as echoes. (Such an assumption was justified by the results in Chapter
Four which showed that the first multipath to arrive was the strongest
35% of the time, and on average had a strength equal to 70% of the
strongest path.)

This assumption considerably simplified the design and operation
of the test system. It led to the adoption of a frame method of data
transmission. After each block of data had been transmitted a gap
was left equal to the expected maximum channel dispersion time. Then
a single pulse, or "set-up" pulse, was sent followed by another similar
gap, and then the next block of data. As a result of the two gaps the
set-up pulse would be received free from intersymbol interference.
Hence it could be used to establish bit synchronisation and decision
threshold levels and could be used as a polarity reference for PSK.
Further,the set-up pulse format allowed the simple implementation
of a decision-feedback equaliser. Because the first version of the
set-up pulse to be received was treated as the signal the channel
impulse response would be that present in the gap between this version
and the start of the next block of data. Thus the "setting-up" of
the equaliser was straightforward.

The experimental procedure is given in Chapter Four. It is
seen that the permutations of the various modulation modes and bit
rates were transmitted in a cyclic fashion by a series of three-minute
integration runs. The runs for each permutation were then aggregated
to give its overall error performance. The cyclic method helped
with averaging out the variations in the channel condition, and also
permitted histogram results to be presented. In all, approximately
100 hours of tests were performed over a period of one month from

from March to April 1977.

The results are then presented in Chapter Four. The dependence of the error probability on the bit rate, the comparison of the modulation modes, and the evaluation of the equaliser performance are shown. Also shown are estimations made of the channel multipath structure.

Finally, in Chapter Five, the conclusions are given. The channel conditions are summarised as are the performance of the various test parameters, viz modulation modes, bit rates, and component parts of the system. Lastly suggestions for further work are given. An improved form of equaliser is proposed, as is the use of coding to improve the error performance at low bit rates.

CHAPTER TWO

DEVELOPMENT OF THE RESEARCH PROGRAMME

SUMMARY

This chapter begins in Section 2.1 with a description of the properties of the HF channel. The structure of the ionosphere is given, the reflection process is explained, and it is shown how the maximum frequency that can support propagation depends on the electron density of the particular reflection layer. The phenomena of multipath and fading are also examined and the composition of the noise added by the channel is given. The section is concluded with a summary of the factors that affect digital transmission.

In Section 2.2 previous work in the field of digital HF is reviewed. As is indicated, most development has been concentrated on parallel sub-channel type of systems. Little research has been performed on the less complex serial transmission technique. In particular little attempt has been made to use adaptive equalisation over the HF channel.

As a result of this review the plan of the research was decided upon and this is presented in Section 2.3. The section shows that the work was to be concentrated on the serial data transmission process. The error performance of such fundamental parameters as the type of modulation mode, and the dependence of the error rate on the bit rate was to be obtained by tests performed over a real HF radio link.

2.1   THE PROPERTIES OF THE HF CHANNEL

In order to understand the problems of transmitting digital

data over the HF channel we must discuss its properties. We start

by describing the structure of the propagating medium; that is the

ionosphere itself.

2.1.i   The Ionosphere

The upper atmosphere becomes less dense the greater the distance

from the earth's surface. The ionosphere, in the region from 50 Km to

350 Km, is composed mainly of molecules and atoms of oxygen and

nitrogen and, to a lesser extent, of nitric oxide.   The molecules

tend to atomise as the height increases. The ionosphere is formed by

the ionisation of these molecules and atoms under the action of the sun's

radiation. To some extent, the ionisation of meteors makes a

contribution[12].

The composition of the ionosphere has been determined by vertical

and oblique radio soundings, and by the use of rockets and satellites[12].

It is generally divided into three regions labelled the D, E and F

regions. The reflecting media within these regions are not layers

as such but peaks of ionisation intensity within the relationship

of electron density to height above the earth. Typical distributions

at noon and midnight are taken from Davies[13] and are as shown in

Fig. 2.1.1.

The D region is that part of the ionosphere which is between

60 and 90 Km above the earth's surface. Its critical frequency (the

maximum carrier frequency at which a vertically-incident ray can be

reflected) is between 100 KHz and 700 KHz[12] because of the low

level of ionisation (see Section 2.1.2 for the dependence of the

critical frequency on the ionisation level). There are high concentrations of neutral particles and heavy ions present, and this causes absorption of wave energy as a result of particle collisions excited by the wave. The level of absorption can be shown to be inversely proportional to the square of the frequency at HF[12,13]. Hence, because of the amount of absorption that would take place, the D layer is not utilised for reflection. The signal carrier frequency is made sufficiently high for the ray to pass through the D layer and be reflected from the high layers.

The D region is a day-time phenomenon. At night, molecular recombination is no longer counteracted by the presence of the sun's radiation, and so the D layer effectively disappears. Now, waves propagating by multiple "hops" between the upper reflecting layers and the earth's surface have to traverse the D layer twice for each hop. Hence at night a larger number of hops are possible before the signal is attenuated below the level of the noise, and so transmission distances are greatly increased. For the same reason background noise and interference from other users increase at night.

The band from 90 Km to about 170 Km is known as the E region. Maximum ionisation is at around 110 Km[14], the electron density here being of the order of 10^5 electrons per cc at maximum. This layer does support propagation, the maximum distance for a single reflection being 2000 Km. There is still a significant concentration of heavy particles but, with the day-time critical frequency being 3 to 4 MHz, absorption is not high compared to the D region. At night the critical frequency drops by an order of magnitude owing to the fall in electron concentration (see Fig. 2.1.1). An approximate equation relating the

critical frequency F_E to the time of day by means of the solar zenith angle θ is given, in Davies[13], by

$$F_E = 0.9 \left[(180 + 1.44R) \cos θ\right]^{0.25}$$

where R is the sunspot number. This equation is independent of changes in season.

That portion of the ionosphere above 170 Km is termed the F region. In daytime, it splits into two parts termed the F_1 and F_2 layers. The F_1 layer is centred at around 200 Km. This layer is merged into the E layer as far as simple models of the ionosphere are concerned[12], and it is not generally considered on its own as a vehicle for transmission.

The F_2 layer is concentrated at about 300 to 320 Km. Here electron densities are around 10^6 electrons per cc at maximum, the resultant critical frequency being 5 to 10 MHz in our latitudes. The dependence of its critical frequency on external factors such as the solar zenith angle, sunspot number and the time of season has been of constant interest to ionospheric physicists. The critical-frequency variations of the other layers can be predicted by theoretical and experimentally-found laws, as shown in[12,13,14]. The F_2 layer however exhibits anomalies in the behaviour of its critical frequency as compared to the other layers. For instance, its critical frequency is higher in winter than summer, and is higher as a whole for the northern hemisphere than for the southern hemisphere[12]. Prediction of the F_2 layer is of importance because it plays the dominant role in long-distance HF communications. Its extra height and low level of absorption enable it to support single-reflection propagation for

distances of up to 4000 Km. Carrier frequencies of up to 50 MHz can be used[14]. These figures are for daytime; at night, the F_1 and F_2 layers merge at around 300 Km, the critical frequency dropping to 2 to 4 MHz at minimum. This single F layer is the chief propagator at night.

We now examine the process of ionospheric propagation.

2.1.2 The Method of Wave Propagation

The discontinuity in the ray path is popularly-termed reflection but it is, in fact, a process of refraction. The refraction of the ray arises from the continuous change in the refractive index with altitude of the particular layer, this being due to the changing electron density. The refractive index is a function of the carrier frequency f as follows:

$$u = \surd(1 - f_p^2/f^2) \tag{2.1.1}$$

where the plasma frequency f_p is related to the electron density N by:

$$f_p = \surd|Ne^2/(mE_0)|/2\pi \tag{2.1.2}$$

where e and m are the charge and mass of the electron and E_0 is the permittivity of free space.

(Note: the effect of the earth's magnetic field and curvature are ignored in this initial analysis.)

In order for the wave to be refracted by the earth its trajectory must become horizontal at some point. By applying Snell's Law to a medium exhibiting continuous change in its refractive index a

relationship between the index U_h (at the height of "reflection") and the initial angle of incidence θ of the ray upon the layer can be established (see[12,13]). It is

$$U_h = \sin \theta$$

Hence, at the point of "reflection", we have (from equations (2.1.1) and (2.1.2)):

$$\sin \theta = (1 - N_h e^2/4f^2\pi^2 E_0 m)^{\frac{1}{2}} \tag{2.1.3}$$

where N_h is the electron density at the level of reflection. On rearranging equation (2.1.3) the carrier frequency necessary for these conditions to be met is seen to be

$$f = \sqrt{(N_h e^2/4\pi^2 E_0 m)} \sec \theta \tag{2.1.4}$$

The critical frequency f_c is obtained from this equation by setting the value of θ equal to zero (i.e. vertical incidence) and using the value of the maximum electron density, N_{max}, say. Hence, from equation (2.1.4), the critical frequency is seen to be

$$f_c = (N_{max} e^2/4\pi^2 E_0 m)^{\frac{1}{2}} \tag{2.1.5}$$

This is the maximum frequency at which a vertically-incident ray is refracted back to earth. For a lower carrier frequency a lower value of the electron density satisfies equation (2.1.3) (for θ zero). If the frequency is higher than the critical frequency no value of N satisfies equation (2.1.3), and so the ray must pass through the layer, albeit with its path somewhat "bent" by refraction.

An important feature of the refraction process is that it can
be shown to be equivalent to mirror-type reflection from a precise
altitude. This height is termed the virtual height. By applying
the theorem of Breit and Tuve (as in$^{(12,13)}$) it can be shown that
the virtual height h_v is a function solely of the plasma frequency f_p
at the summit of the path, i.e.:

$$h_v = F(f_p) \qquad\qquad (2.1.6)$$

This equation implies that rays of any angle of incidence having
the same refraction height have the same virtual height (because f_p
is a function solely of the electron density). In effect they travel
a path through a medium with a refractive index of unity, and are
perfectly reflected from the same height.

Note: the relationship in equation (2.1.6) is determined
experimentally by the vertical-sounding techniques described by
Picquenard$^{(12)}$ and Davies$^{(13)}$. A typical example is reprinted from
Picquenard and shown in Fig. 2.1.2.

The results that have been established will now be used to
study the case of propagation between two points separated by a
fixed distance. It will be shown that, when the carrier frequency is
less than the MUF, two ray paths are possible for a single reflecting
layer (again the effect of the earth's magnetic field is ignored).

Now, by considerations of simple geometry, the distance D
between the transmitter and receiver can be related to the virtual
height h_v and to the angle of incidences of a general ray path by
(see Fig. 2.1.3):

$$\tan i = \frac{D}{2h_v} \qquad (2.1.7)$$

The carrier frequency f necessary for a ray, incident to the layer at an angle i, to have a plasma frequency at the path summit of f_p is, from equation (2.1.4),

$$f = f_p \sec i \qquad (2.1.8)$$

The elimination of i from equations (2.1.7) and (2.1.8) leads to

$$f^2 = f_p^2 \left(1 + \frac{D^2}{4h_v^2}\right) \qquad (2.1.9)$$

In order to obtain the MUF we choose a value for f and then see if equations (2.1.6) and (2.1.9) can be solved simultaneously. If so, a higher value of f is then tried (and the equations solved again) until there is no solution. This process is performed graphically as in Fig.2.1.4, (from Davies) since equation (2.1.6) is always experimentally established in the form of a curve. In Fig. 2.1.4 the MUF is the value of f for the line (i.e. equation (2.1.9)) tangential to the virtual-height plasma-frequency curve, at the point c, i.e. the MUF is 20 MHz. When the carrier frequency is less than the MUF, for example when f is 18 MHz, the h_v-f_p curve is intersected at two points. This means that two ray paths are present. The one corresponding to the higher virtual-height (i.e. the intersection at point b')is termed the high-angle ray, and the one having the lower virtual-height (point b) is called the low-angle ray.

This phenomenon is illustrated in Fig. 2.1.5. What is happening is that the summit of the ray path with the MUF does not traverse the

portion of the layer with the maximum electron density, but a portion lower down. Hence if the transmission frequency is lower than the MUF two ray paths exist where (i) a ray at a higher angle of incidence than the MUF ray spends a shorter time in the layer and is refracted through a small enough horizontal distance to arrive at the receiver (the low-angle ray) and (ii) a ray incident more vertically than the MUF ray is shifted through a large enough horizontal distance to arrive at the receiver after re-emerging from the layer (the high-angle ray). As the carrier frequency is increased these two paths get closer until complete convergence takes place when the MUF is reached.

We have seen that, for transmission between two points, two ray paths exist in a single layer whenever the carrier frequency is below the MUF. By taking into account the earth's magnetic field we will now show that each path has a pair of rays associated with it.

Any electro-magnetic wave, on entry to a magneto-ionic medium, splits into two waves termed the ordinary and extraordinary waves[12,13]. They traverse different paths in the medium because the expression for each of their refractive indices is different (the difference though being slight at HF). It is such that the ordinary wave is reflected at a greater altitude than the extraordinary wave. This means that there are two MUF's for each transmission mode, and this phenomenon can be observed on ionospheric soundings (see Fig. 2.1.2).

It is helpful at this point to summarise the features of ionospheric propagation propounded in sections 2.1.1 and 2.1.2.

We have seen that the ionosphere is composed of three regions which are, in order of increasing altitude;

(i)   the D region, which acts as an attenuator, and
        disappears at night,

(ii)   the E layer, which is a reflector but becomes
        ineffective at night, and

(iii)   The F region, or which the F_2 layer is the chief
        reflector for long-distance propagation.

The process of bouncing the wave off a particular layer has been

shown to be a process of refraction, but also to be equivalent to

mirror-type reflection from a virtual height. The critical frequency

of a layer was shown to be proportional to the square root of the

maximum electron density of the layer (equation (2.1.5)). For oblique

incidence, the MUF was shown to increase with increasing angle of

incidence (from equation (2.1.4)). It has been discovered that, by

using a carrier frequency below the MUF in transmitting between two

fixed points (this being the usual case since the MUF is a limiting

condition), two distinct ray paths, the high-angle and low-angle rays

exist in a single layer. Furthermore, each ray splits into two

upon entering the ionised layer under the influence of the earth's

magnetic field, and so, from a single layer four paths can be received.

Finally, it should be noted that, because the absorption (mostly

from the D layer) of the wave energy from excited particle collisions

decreases with increasing frequency, optimum propagation is obtained

for a carrier frequency value close to the MUF.

2.1.3 Multiple-mode Propagation (Multipath)

In addition to the four paths that can exist in each layer, further propagation paths arise because, as we have seen, there are two reflecting layers, the E layer and the F_2 layer. The shortest paths are by single reflection from the E and F_2 layers. Other paths arise from the ducting of the signal both between the layers and between one or both layers and the earth's surface. The time delays of the signals received can be calculated by simple geometry, and from a knowledge of the speed of light. The dispersion times are of the order of a few milli-seconds[12]. The single hop paths will tend to be the stronger because they:

(i)   have larger angles of incidence and so penetrate less deeply into the layers; thereby less absorption results,

(ii)   are only single reflections; and

(iii)   they only traverse the absorbing D region twice.

2.1.4 Fading

The random variation with time of the received signal envelope is termed fading. It is caused as follows.

When the transmitted signal has a constant carrier (or, in the digital case, when the pulse lengths are long compared to the channel dispersion time), the multipath components add up in a phasor fashion at the receiver. The envelope of the resultant has a Rayleigh distribution and the distribution of the carrier's phase is uniform[15]. When one path is particularly strong the distribution becomes Rician[15,16]. The time variations are caused by:

(i)   variations in the summation (by what is effectively
        interference) of the modes due to (a) the atmospheric
        layers shifting and (b) local changes in electron
        densities and absorption. Fading periods are of
        the order of seconds;

(ii)   flat fading (i.e. all propagation modes altering in
        strength in unison) caused by changes in D region
        absorption. Here fading periods are of the order
        of an hour or longer, and

(iii)   focussing both because of the propagation modes that
        arise in a single layer and because of local
        deformation of the particular ionospheric layer.
        Fading periods here are of the order of 15 to 30
        minutes[13].

The interference-type fading is frequency selective because it depends on the wavelength of the interfering signals[33]. It is sufficient to be marked over the bandwidth of a modulating signal.

When serial high-speed digital data transmission is used the individual multipaths will in general be resolved. Here we are considering the multiple-mode types discussed in section 2.1.3. The ray paths present in a single layer, from high-and low-angle rays and ordinary and extraordinary waves, have too small time differences to be resolved, and will result in the individual multipaths undergoing independent fading. Usually the low-angle ordinary ray is less attenuated than the others[12] and so the resultant fading characteristic is Rician[13]. Davies also indicates that the fading is slower and shallower than the overall envelope.

2.1.5 Background Noise

This is composed of different sources, as shown in Fig. 2.1.6 (which is taken from Billington[34]). The frequency band of interest is that around 10 MHz. Here atmospheric static and man-made noise are the prominent disturbances. Man-made noise is highly correlated with population density[14], and so this contribution depends on the geographical location of the receiving station. Atmospheric noise rises considerably at night; this can be attributed to the disappearance of the absorbing D region enabling long-distance propagation of static to take place. We note that, at 10 MHz, the receiver noise contribution is much less than the channel noise component. The result of the summation of the various sources is that the noise appears as short high amplitude pulses of random occurrence superimposed on a lower level of random noise[14].

2.1.6 Summary of the Factors that Affect Digital Transmission

Having described the properties of the HF channel we will now summarise them as they affect digital transmission.

The most important factor discriminating against digital transmission is that the received signal is likely to contain several components because of the multipath phenomenon. Clearly each path has a different length, and so each component has a different delay between transmitter and receiver. Hence if a short burst of carrier is transmitted the received signal is a series of echoes as the difference in the path delays manifest themselves. Now, with transmission rates ranging, say, for 1 to 4 Kb/s, bit rates periods are correspondingly 0.25 ms to 1 ms (using serial transmission). This is smaller than the usual figure discussed in section 2.1.3 for the channel dispersion time of a few milli-seconds. Hence inter-

symbol interference occurs. This is the major problem encountered
in transmitting digital information over the HF channel.

There are other problems that affect digital transmission.
For example multipath causes difficulties with carrier extraction
because the echoes all have different carrier phases. Also inter-
symbol interference makes bit synchronisation difficult. Fading has
a serious effect on amplitude-sensitive operations in the receiver such
as level extraction for decision-making in the case of ASK.

Having observed the type of conditions likely to prevail over
the HF channel we now see how previous researchers have attempted to
cope with them.

2.2   REVIEW OF THE PREVIOUS WORK

In this section we examine previous work that has been carried out on efficient data transmission over HF in order to formulate a plan of research.

The previous work is considered in the light of the constraints that ionospheric conditions and bandwidth limitations impose on the user of the HF channel. The type of channel conditions that are prevalent were referred to in the previous section. The bandwidth limitations (which affect the maximum bit-rate value) are because channels are internationally allocated in 3 KHz channels in order to prevent user congestion. The maximum binary bit rate than can be transmitted down a channel of width f Hz (using SSB) is 2 f b/s. (This result is a consequence of Nyquist's work[17]). Thus 6Kb/s is the maximum bit rate that can theoretically be used over an HF channel that employs binary signalling. This figure is impossible to attain in practice by binary means because ideal pulse shaping, bit synchronisation etc would be required[18]. A figure of 1 bit per Hz, i.e. 3 Kb/s for the HF channel, is acceptable as making "efficient" use of bandwidth in the binary case.

In the early telegraph-transmission circuits, the bit rate was of the order of 50 to 100 bits/secs (b/s). Betts[11] considered that 100 b/s was the maximum rate for reliable transmission in the presence of typical HF-channel multipath conditions. At bit rates of 50 to 100 b/s, the bit periods (10 to 20 ms) are considerably longer than the channel dispersion time (typically 3 ms) and so intersymbol interference is not a problem.

The beneficial effects of the use of long pulses to avoid inter-symbol interference were put to advantage in the first attempts

to transmit at high speed (i.e. at rates of the order of 3 Kb/s).
These attempts employed techniques that involved splitting the 3 KHz
channel into a number of sub-channels with the data being multiplexed
amongst them. The sub-channels produced their own data stream at a low
bit rate (50 to 100 b/s), and they were synchronised so that the
pulses from each sub-channel were generated exactly in parallel with
the other sub-channels. Hence the appearance of the composite data
stream after the sub-channel outputs had been summed was a series of
"pulses" at the sub-channel bit rate, each "pulse" being the n bits
from the n sub-channels in parallel.

A more detailed explanation of the operation of these parallel
sub-channel systems is given in Section 2.2.1. In this section
we go on to show that they have the basic disadvantages of system
complexity, vulnerability to frequency selective fading, and are
inefficient in their use of transmitter power. We then show how,
by describing examples of these systems, they have been extensively
developed within these limitations.

In Section 2.2.2 we examine the various methods of serial
transmission that have been propounded. Let us reiterate the problem
that is likely to occur from multipath When the bit rate is 3 Kb/s
the bit period is 0.3 ms. A typical HF channel may have several
multipath echoes spanning a time of 2 or 3 milliseconds; thus
intersymbol interference may extend over several bit periods when the
bit rate is 3 Kb/s. Therefore poor performance is likely without
special techniques to combat such interference. In Section 2.2.2
we see that one technique that has been developed is that of adaptive
equalisation. The channel is modelled at the decision instances as
a tapped delay line; in other words, effectively as a non-recursive
filter. The equaliser is a form of digital filter whose structure

is such that it is the inverse of the channel model. The equaliser
is adaptive in that it continually follows the changes that occur
in the channel model as the strengths of the various multipaths
fluctuate. Hence, if such a device is cascaded with the data stream
at the receiver, intersymbol interference is removed. Other types
of serial-transmission systems are also described in Section 2.2.2.

In Section 2.2.3 we review the results of error-rate measurements
that have been made over the HF channel. We see that few such
measurements have been taken and that they have almost all been obtained
using parallel sub-channel systems. No concerted evaluation of serial
transmission systems has been made.

We summarise the results of the review of the previous work in
Section 2.2.4, and we conclude that there is a need to examine the
basic methods of data transmission by serial means and to evaluate
them by experiment over a real HF link.

Finally, in Section 2.2.5, the detailed plan of research required
to perform such an investigation is given.

We start by examining parallel sub-channel systems.

2.2.1 Parallel Sub-channel Systems

As has been explained these systems operate by frequency-
division multiplexing the data between a number of sub-channels whose
individual bit rates (which are identical and synchronised to one
another) are of the order of 50 to 100 b/s. The data is then
transmitted in "frame" pulses which are the summation of the sub-
channel outputs. The direct effect of intersymbol interference is
avoided because these frame pulses are long compared to the channel
dispersion time. We will now examine how the data is modulated and
demodulated on the individual sub-channels.

The waveform on each sub-channel consists, in the simplest case, of the sub-carrier being amplitude modulated by the square pulses of the digital bit stream. Hence the spectra of the sub-channels are the familiar sinx/x functions centred on the sub-carriers. The sub-channels are packed together such that significant parts of each sinx/x function overlap. We will now show that, provided the sub-carriers are spaced correctly, a detection process can be used that avoids any co-channel interference.

Let w_T be the tone spacing and let $u_i(t)$ be the data-modulating waveform of the ith channel ($u_i(t)$ will be either "1" or "0"). Then the ensemble $S(t)$ of the sub-channels is*:

$$S(t) = \sum_{i=1}^{N} u_i(t) \cos i\, w_T t \qquad (2.2.1)$$

where N is the number of sub-channels.

The necessary detection process is the matched filter[**] realised in its correlation form. The correlator waveform for the nth sub-channel, $C_n(t)$, is given by:

$$C_n(t) = \cos n\, w_T t$$

The correlator output D for this channel is thus

$$\int_0^\tau C_n(t)\, S(t)\, dt$$

$$= \int_0^\tau \cos n\, w_T t \sum_{i=1}^{N} u_i(t) \cos i\, w_T t\, dt \qquad (2.2.2)$$

where τ is the integration time (which would usually be the bit period).

Now if the sub-carrier (or tone) spacing w_T is such that

---

*this analysis is for the signal at baseband

[**] unless otherwise stated this term means the optimum in white noise

$$\dot{w}_T = \frac{\pi}{\tau} \qquad\qquad (2.2.3)$$

we have that:

$$D = \frac{1}{2}\tau$$

for i = n, i.e. the sub-channel we are trying to detect, and

$$D = 0$$

for i ≠ n.

Hence no inter-channel interference occurs if equation (2.2.3) is satisfied.

The above condition is that of orthogonality and the sub-channel tones are said to be orthogonally placed when:

$$w_T = \pi/\tau$$

We have seen that the main advantage of the parallel sub-channel system is that the direct effect of intersymbol interference that would be present in serial transmission under multipath conditions is avoided. We can also see another advantage in that a matched filter detector is necessary to prevent any co-channel interference. The matched filter effectively improves the signal-to-noise ratio at the receiver as compared to the simple decision-maker[38]. We will now examine the disadvantages of the parallel sub-channel type of system.

1.   Inefficient use of transmitter power

The aggregation of the sub-channels is expressed in equation (2.2.1). The amplitude of S(t) is going to vary randomly from frame interval to frame interval according to the density of "1"'s

and "0"s in each frame (expressed in $u_i(t)$). Hence $S(t)$ is going to contain large peaks whenever there is a high density of "1"s in a particular frame. Now, for the reason given in Chapter 3, transmitter amplifiers have to be peak-power limited and so should contain a variable-gain circuit that fixes the peak of the modulated signal. Thus, for a signal such as the output of a parallel sub-channel system, the reduction in the average transmitted power as compared to the available power is likely to be significant. Darnell[32] found that the 16-channel system used in his experiments could only be transmitted at a power of 100 W using a 1 KW power amplifier. Such "de-rating" of the available transmitter power constitutes the most serious defect of parallel sub-channel modems, because of the reduction that results in the signal-to-noise ratio at the receiver.

2.   Vulnerability to frequency-selective fading

Frequency-selective fading occurs when multipath is present because the "summation" of the paths at the receiver is an interference-type effect. In order to explain the process let us consider that a single tone $\cos w_i t$ is transmitted. Let the carrier angular frequency w_i be varied over a range w_B, equivalent to the bandwidth of the channel. Consider if the time τ that the multipaths are dispersed over is such that the product $w_B \tau$ is of the order of a wavelength of the carrier frequency. Then, as the frequency varies in the range w_B, the amplitude of the summation of the multipaths is going to vary also. In practice

$$w_B = 2\pi 3 \cdot 10^3$$

and typically,

$$\tau = 2 \text{ ms}$$

$$\therefore \quad w_B \tau = 12\pi$$

$$= 6 \text{ wavelengths}$$

Hence there will be peaks and troughs in the effective channel frequency response, and so some sub-channels will experience low signal-to-noise ratios.

3.   Effects of multipath other than frequency-selective fading

It has been stated that the parallel sub-channel method avoids the effect of intersymbol interference experienced in serial systems. However the smearing effect of multipath does cause some intersymbol interference because of the integration process in the bit detection method. Also it has been shown by Bello[22] that such smearing of the frame pulses violates the sampling orthogonality principle of bit detection (i.e. there is a non-zero result in equation (2.2.2) for channels other than nth channel). Hence some co-channel interference occurs resulting, in Bello's example, in an irreducible error rate of 10^{-2}, for a 1 ms multipath spread (this error rate includes the effects both of intersymbol and interchannel interference).

These deleterious results of multipath can be alleviated by
if the orthogonality condition is to be maintained
detecting a central gated portion of each bit. such a solution, however, leads to a reduction of the overall bit rate, because the reduction in the detector integrator time means that less sub-carriers are available in the channel (from equation 2.2.3).

4.   Bit synchronisation requirement

Bello has computed the irreducible error rate likely to arise in the Kathryn modem[20] due to errors in bit synchronisation. Accurate bit synchronisation is important in order to prevent interchannel interference by preserving sampling orthogonality. Bello shows that, for the channel conditions used in his model, a synchronisation error of 0.1 ms means that the rate is still as high as 10^{-3}. The bit

length used in the Kathryn system was 13 1/3 mS.

5. System complexity

The complex nature of the parallel sub-channel system is apparent in the discussion of it so far.   The fact that a multiplicity of sub-channels each requiring a coherent sub-carrier at the receiver are used means that pilot tones are usually employed. Hence, a narrow-band filter is needed for each sub-channel at the receiver. Incidentally, Bello shows that Doppler effects arising from the ionospheric reflection process can lead to a significant irreducible error rate via the pilot tone.

Further complexities arise because each sub-carrier must be generated with the correct tone spacing, and each correlator detector must have the correct integration time (in order to prevent inter-channel interference). We have also seen how precise bit synchronisation must be.

We will now give details of examples of parallel sub-channel systems that have been developed. The first example was the Kineplex modem, details of which were published in 1958[19]. Twenty sub-channels were used, each frame being transmitted 75 times a second. Each tone was phase-shift keyed* to one of four phases, and so two bits per sub-channel were received from each frame. This led to an overall data rate of 3 kb/s.   Demodulation was accomplished by storing the phase of the previous bit of each tone in a high-Q resonator.   The reasonator output provided the correlating waveform for detection.

---

*It can be shown that the principle of sampling orthogonality derived in equations (2.2.1)to (2.2.2)applies to modulation by phase-shift keying as well as by amplitude-shift keying.

This is a form of incoherent demodulation, and so differential encoding and decoding were necessary. The total bandwidth of the original system was 3.4 KHz.

The disadvantages of the Kineplex method are two-fold. Firstly the detection process is complex. The high-Q reasonators were originally realised by using Ni-Span C bars, with the energy being coupled by transducers. For each sub-channel two resonators were needed (one for storing the phase of the previous bit, and one for storing the phase of the present bit) making 40 in all. Secondly the differential coding requirement meant that the error rate could have been double that obtained by using coherent demodulation.

The Kathryn modem[20,21], used 34 sub-carriers orthogonally spaced by 81.4 Hz with a frame rate of 75 Hz. The resultant bit rate was 2550 b/s. A one milli-second guard band was employed to mitigate the effects of multipath. The modulation and demodulation process was more sophisticated than in Kineplex, considerable attention having been paid to reduce the effects of frequency-selective fading. Each sub-carrier was PSK SSB modulated with:

(a) the main information bit sequence,

(b) the information from the sub-channel 17 tones away, in phase quadrature to that in (a),

(c) a pilot tone.

At the receiver this tone was isolated and then put to two uses:

(i) to phase correct its particular sub-channel (since, as it had the same frequency as the modulated sub-carrier, channel perturbations affected it in the same way), and

(ii) its amplitude, which would have varied from sub-
channel to sub-channel because of frequency-selective
fading, provided a weighting function so that maximal-
ratio combining could be used. Since dual space
diversity was employed there were four independently-
detected components for each bit, i.e. for each
receiver there was a component from (a) plus its
duplicate from (b). Thus there were four
orders of diversity. The final bit decision was
taken from the output of the combiner using
conventional phase detection (i.e. by correlation).

The Andeft design[23] operated at 4.8 kb/s by transmitting
66 sub-channels (separated orthogonally by 40 Hz) at a frame rate
of 75 b/s, in the frequency band from 400 Hz to 3 KHz. Frequency-
differential PSK modulation was employed. This technique involves
adding a pilot tone to each sub-channel of sufficient distance from
the sub-carrier such that the tone can be band-pass filtered at the
receiver without containing components of the modulated signal. This
frequency difference is then re-inserted so that the modulated signal
can be coherently detected. Andeft could also be operated at
2.4 kb/s, in which case the 4.8 kb/s-carrying capacity of the modem
was used to implement dual frequency diversity. This helped to
reduce the effects of frequency-selective fading.

For all these systems the general disadvantages expounded
earlier of parallel sub-channel systems apply, viz. inefficient use of
transmitter power; vulnerability to frequency-selective fading; and
the extent of system complexity necessary to realise each system

and to avoid degradation of performance by such problems as inaccurate bit synchronisation.

Having discussed the advantages and disadvantages of parallel sub-channel systems we will now examine the types of serial transmission systems that have been proposed for use over the HF channel.

2.2.2 Serial transmission systems

As was described in the introduction to this section one technique used to counter the effects of intersymbol interference in serial transmission is that of adaptive equalisation (see Chapter Three for a full explanation of this method).

Probably the earliest application of this idea was by DiToro et al in the Adapticom system[24]. This used a transversal (i.e. non-recursive in digital-filter terminology) filter as the equaliser. Its tap gains were set up from its measurement of the channel model. For this to be achieved the data stream was interrupted periodically (every 100 ms or 160 ms) to transmit a probe pulse. The system was tested over a simulated multipath channel.

Although Adapticom was developed for use over HF channels, neither it nor the results of the other research and development carried out in the sixties on equaliser schemes by Lucky and others (summarised by Lucky in [42]) has in general been applied to "real" tests over HF.

A novel method of dealing with multipath-caused intersymbol interference has been proposed and tested by Goutmann and Gutman. Goutmann[25] has argued that intersymbol interference can be regarded as being convolutional encoding. Thus their system[26] for combating HF-channel distortion incorporated a sequential decoder in the receiver.

The channel was tracked   to provide the decoder with knowledge of time variations in the "encoding" structures.   Hence the modem was adaptive to changes in channel conditions.

The disadvantages of such a method is that it operates after the hard decision has been made. Intersymbol interference is bound to force the level of some pulses at the decision instance to be close to the decision threshold. Hence, for these pulses, high error rates will occur when noise is present. In other words, the coding method may operate successfully in the absence of noise because it removes the results of intersymbol interference. In the presence of noise the error rate is likely to be higher than by using equalisation because, unlike equalisation, the decoder has not removed the effects of intersymbol interference. Although some form of equalisation was used in the coding scheme as outlined in[26], the same criticism applies if there was still some intersymbol interference present; and if there was not then the decoder would have been unnecessary.

Apart from this, problems arose with the decoder in operation because of the computational burden placed on it. At times when there were large-scale changes in the channel mode structure the decoder was unable to compute the encoding "tree" and so the system broke down. Then a probe sequence had to be sent to retrain the decoder, requiring some form of communication from receiver to transmitter.

, The technique of spread spectrum signalling has been applied to HF[27]. This technique involves transmitting data as bursts of wideband (a few hundred kilohertz at least) noise. The noise burst is generated using a filtered pseudo-random sequence and is detected at the receiver by correlation with a regenerated identical noise

waveform. The impulse-like auto-correlation function allows low
noise power to be used hence permitting the spectral density of
the wideband signal to be low enough not to add significantly to the
HF channel background noise (and hence to "interfere" with other users).
Effective gains in the signal-to-noise ratio of 11 to 19 dB over
conventional systems were recorded by Farrell and Andjargholi.

However spread spectrum appears doomed to fail over typical HF
channels in its serial form because of multipath. The effects of
frequency-selective-fading is liable to be acute because of the wide-
band nature of the noise signal and then there is the problem of
multipath symbol smearing. It is not easy to see how the intersymbol
interference of one "blob" of noise on another can be overcome.

A more promising scheme is that of frequency skipping
as proposed by Schmidt[28]. This obviated the effects of multipath
by dividing the allocated channel into sub-channels. Bits are transmitted
serially down each sub-channel in rotation, with each sub-channel
being vacated for an interval that just exceeds the channel dispersion
time. Hence the number of sub-channels necessary to implement this
technique is equal to one plus the nearest integer greater than the
result of dividing the channel dispersion time by the pulse width.

Although such a system prevents any intersymbol interference
existing it is very wasteful of bandwidth. For example, if the
channel dispersion time is 2 ms let the 3 KHz HF channel be divided
up into three sub-channels using 1 ms wide pulses. Then no inter-
symbol interference would exist, but each sub-channel would only be
in use for one-third of the time.

This concludes the review of the methods of data transmission over HF that have been proposed. We will now examine the results that have been obtained of actual error-rate measurements of HF links.

2.2.3 Error-rate measurements over HF links

The numbers of results that have been obtained over actual HF links are very limited, most designers preferring to use simulated channels. This, though much more convenient for the development of a particular system, makes comparisons of different designs and techniques difficult since, in effect, both system and simulator are being compared. The difficulty in forming accurate HF channel models is apparent from the discussion on ionospheric propagation in Section 2.1. The received signal is likely to be comprised of a multitude of independently. time-varying random components, with the addition of background noise having a non-gaussian nature. All of these random components are also non-stationary. Because of the problems in modelling the HF channel then, only examples of actual measurements of the HF channel will be discussed.

Kirsch, Gray and Hanna performed tests using the Kathryn modem over a 3000 Km link[29]. They obtained error-rate results varying from 10^{-2} to 10^{-6}. The S/N ratio was noted as being in the range from 20 dB to 50 dB. It was observed that for figures above 20 dB there was no clear dependence of the error probability on the signal-to-noise ratio. This illustrates the irreducible error-rate phenomenon found in the discussion on parallel sub-channel systems in Section 2.2.1.

McManamon et al. have tested a six sub-channel FSK system having an effective rate of 450 b/s over a 2800 Km link[30]. The

transmitter power was 1 KW.   Error rates of the order of 10^{-2}

were obtained. McManamon highlights the difficulty of justifying

the usefulness of the average error-rate measurement for a non-

stationary process such as the HF channel. Such a criticism also

applies to signal-to-noise measurements, and this topic is discussed

further in Chapter Five.

Chase[31] performed measurements with a 16 sub-channel DPSK

modem having a bit rate of 2.4 kb/s; the transmitter power was 4 KW.

Error rates of approximately 10^{-1} were obtained. Chase also describes

"Codem", a combined parallel sub-channel and coding system. This was

found in tests to give an improvement better than an order of magnitude.

The results of Goutmann and Gutman[26] were obtained when operating

over a 2400 Km link, with a transmitter power of 1 KW. Errors occurred

at a probability of 10^{-3} and 10^{-4} when testing their convolutional-coding

approach.

The spread spectrum system of Farrell and Andjargholi[27] was

found to give error probability results of approximately 10^{-5}.   To

achieve this only 100 W of transmitter power was required, and the

S/N ratio of the received signal was -4 dB. However the path of

transmission was only 15 miles long and so the propagation mode would

have been ground wave only, with skywave reflections being of small

amplitude in comparison. a HOku link using FSK at 1kb/s (SJ).
Goodman and Farrell have made error measurements over

Finally Darnell[32] has performed measurements over a 1000 Km

path using a capability of 1 KW of transmitter power. He compared the performance

of a parallel sub-channel DPSK modem having 16 sub-carriers and dual

in-band frequency diversity, with a serial four-phase DPSK modem,

both having bit rates of 1.2 kb/s. A Real-Time Channel Estimator

was used by Darnell to select the best-operating channel. Error
probabilities ranging from 10^{-2} to 10^{-3} were obtained for the parallel
sub-channel system, and a probability of 10^{-2} was found for the
serial system.

We now summarise the development that has been made on data
transmission over the HF channel.

2.2.4 Summary and conclusions

We have seen that, despite their intrinsic disadvantages described
in section 2.2.1, the parallel sub-channel type of system has been
well established by the development that went on in the 1960s. Error
rates of 10^{-3} have been obtained in tests over typical HF links.

Not nearly as much attention has been paid to serial-transmission
techniques. Spread spectrum seems unusable over a multipath channel
for the reasons given, whilst Goutmann and Gutman's approach of treating
intersymbol interference as a natural code gave problems when the
channel conditions changed. Indeed their results should be treated with
some scepticism in the light of the points made in Section 2.2.2.
Schmidt's method of frequency stepping, whilst avoiding intersymbol
interference, is very wasteful of bandwidth.

Serial transmission using conventional modulation techniques
and utilising adaptive equalisation to combat intersymbol interference
is the most fruitful area for further research. Such a method has
the following advantages:

(1)   The use of narrow pulses would tend to resolve any
        multipath into its individual components Hence
        such a pulse stream would not suffer from frequency-
        selective fading and also from the deep fades that
        result, where long pulses are used, when the

multipaths completely interfere with one another.

Provided that an adaptive equalisation technique

could be found to remove the effects of inter-symbol

interference then, under multipath propagation, serial

transmission should not experience the irreducible

error-rate condition that occurs when using the

parallel sub-channel approach.

(2)   The pulse stream would have an approximately constant

amplitude envelope, and so the available transmitter

power could be fully used.

(3)   The realisation of the basic system would be uncomplicated

as compared to the parallel sub-channel type.

Hence we can now state the objective and draw up of the plan of

the research.

2.3   OBJECTIVE AND PLAN OF THE RESEARCH

In view of the foregoing a decision was made (i) to concentrate the research onto serial methods of data transmission, (ii) to evaluate the performance with respect to such fundamental parameters as the type of modulation and demodulation and the dependence of the error rate on the bit rate and (iii) to obtain the results by tests carried out over a "real" HF link.

Then with these points in mind it was decided to construct and test a serial system so that the following aspects of serial data transmission could be examined:-

(1)   To compare the conventional methods of modulation, viz. amplitude, phase and frequency modulation.

(2)   To investigate the relative performance of coherent and incoherent demodulation with respect to the various methods of modulation.

(3)   To determine how the error performance varies with bit rate.

(4)   To study and evaluate the use of equalisation as a means of obviating multipath.

Fig 2-I.I   Ionospheric electron density profiles
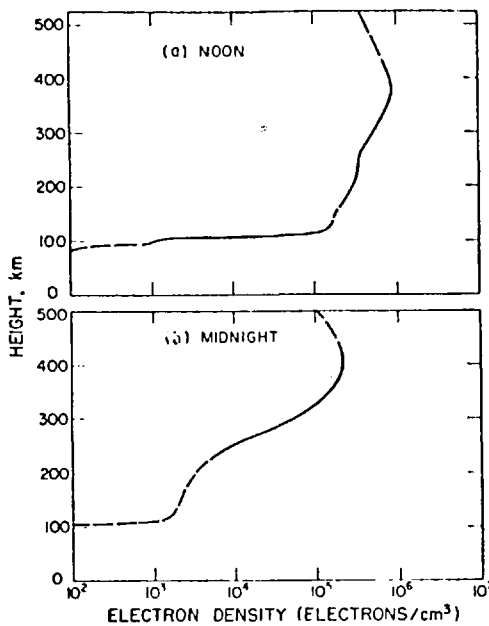
Fig 2-I.2 Typical relationship of plasma frequency to virtual height
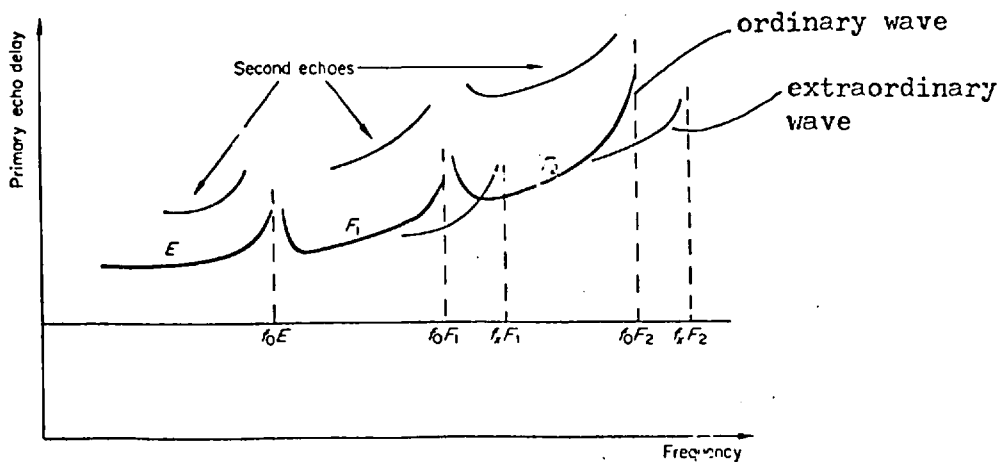
Fig 2-1.3   Geometry of the ray path
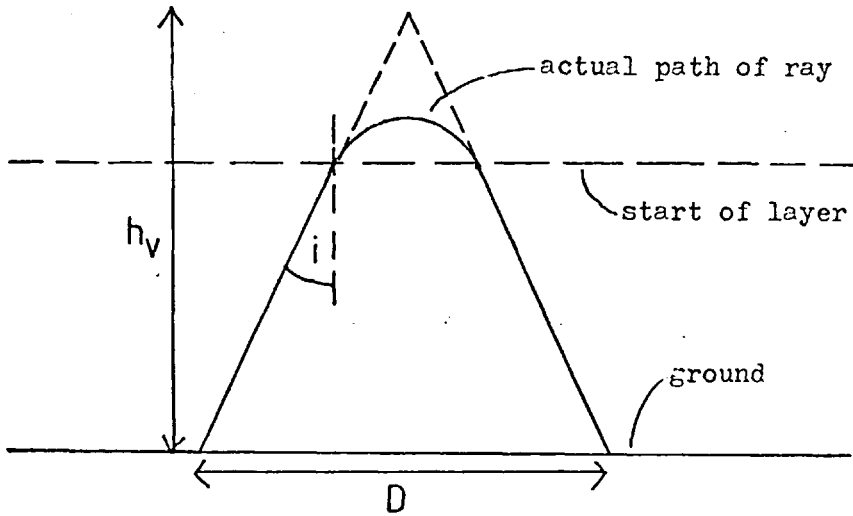
Fig 2-I.4   Variation of plasma frequency with virtual height with
super-imposed transmission line curves
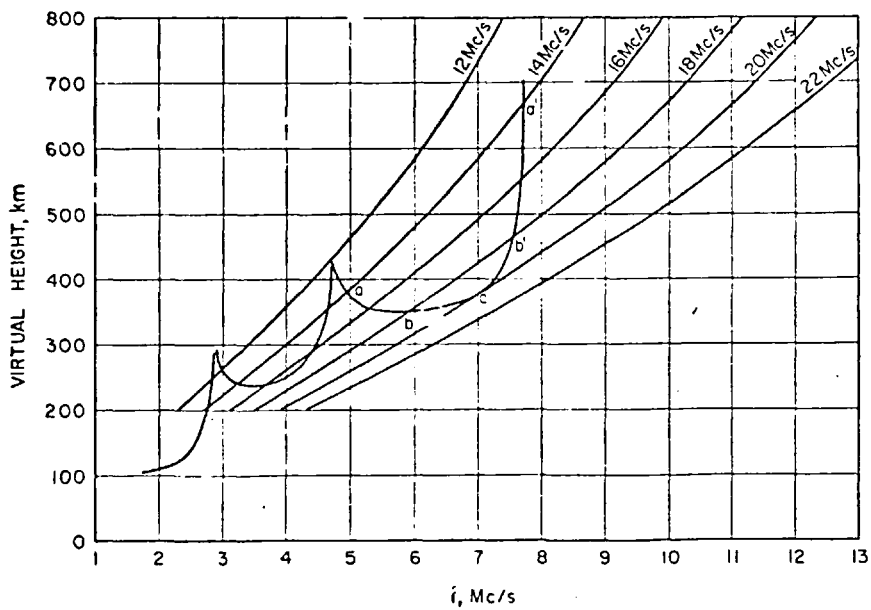
Fig 2-1.5   <u>High- and low-angle ray paths</u>
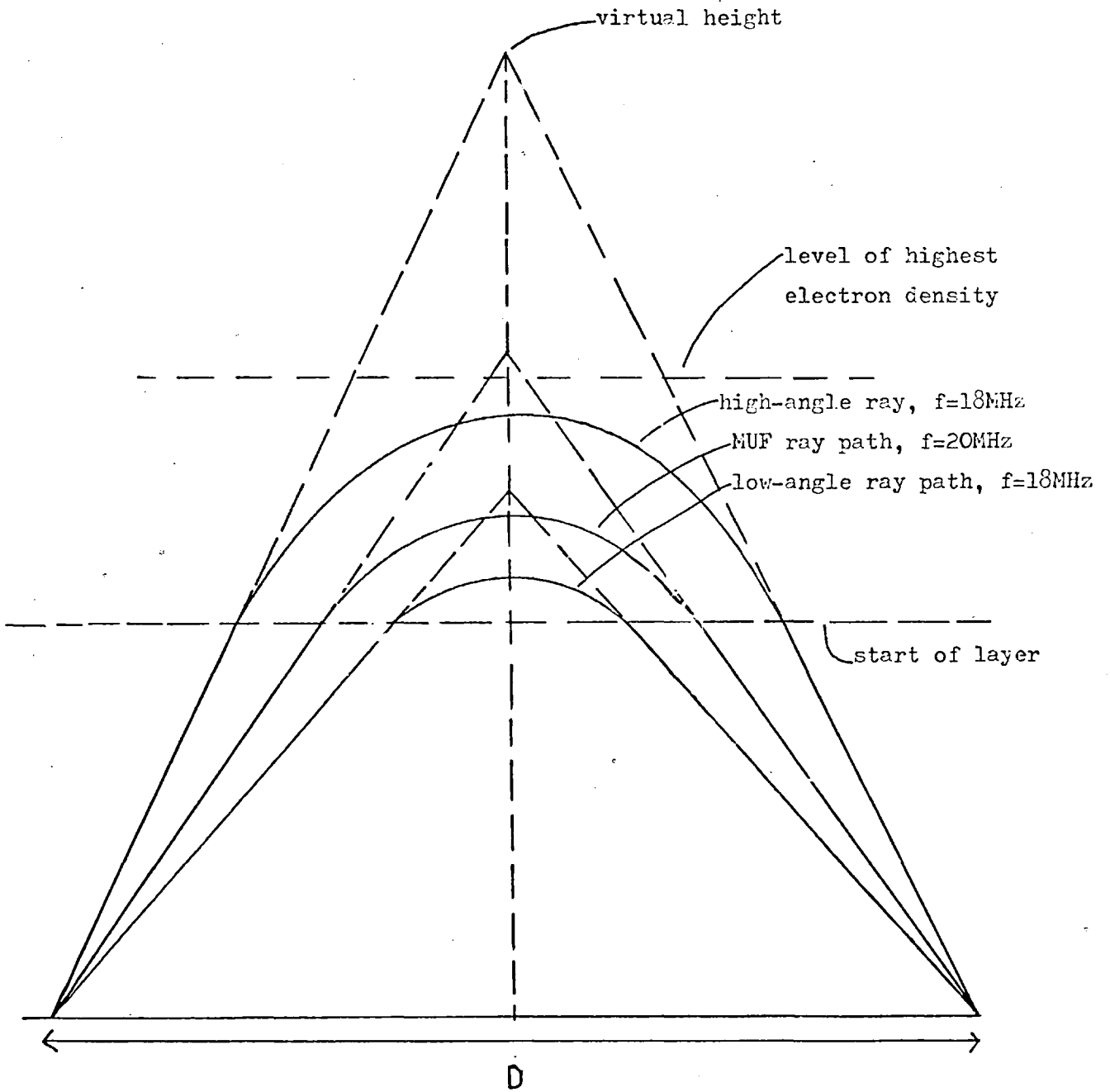
Fig 2-1.6   Spectral distribution of noise sources in the HF channel
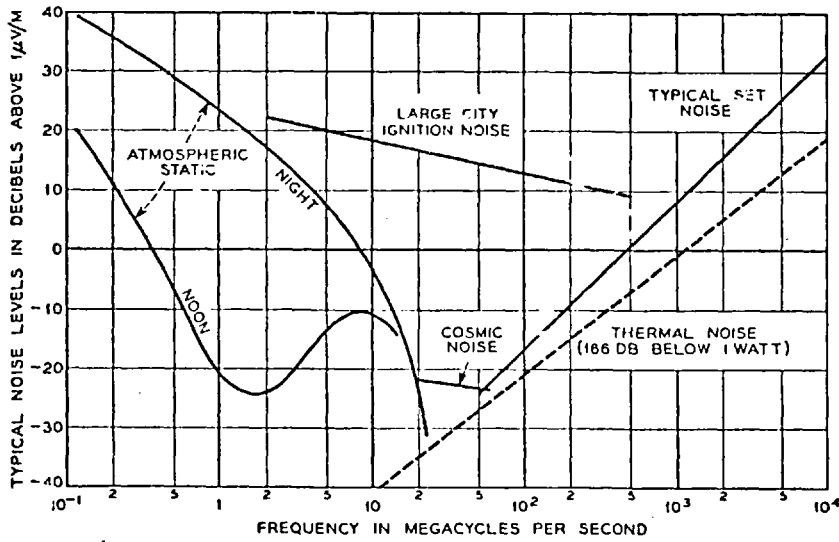
CHAPTER THREE

THE TEST SYSTEM

INTRODUCTION AND SUMMARY

The test system used for the purpose of investigating the questions raised in the previous chapter about serial transmission is described in this chapter. Block diagrams of the transmitter and receiver sections are shown in Figs. 3-0.1 and 3-0.2 respectively. The radio link is shown in Fig. 3-0.1 also.

The system was developed in two stages. Initially it was a serial-transmission system having a variable bit rate from 150 b/s to 3.6 kb/s. Two types of modulation were employed, viz. ASK with rectifier detection and DPSK with PLL detection. (Differential encoding and decoding had to be employed with the PSK mode because of the lack of indication of bit polarity at the receiver.) Initial tests were performed over a 100 Km link between Portsmouth and Imperial College. The purpose of these tests was to determine how essential an equaliser was, and generally to get the "feel" of HF channel conditions. The tests showed that multipath (three or four paths) was consistently present and that the dispersion time was frequently as high as four milli-seconds. This prevented transmission at rates above 250 b/s except for short periods when multipath was reduced or absent. At this bit rate the error rate was found to vary between 10^{-2} and 2×10^{-3}. If the rate was higher than 250 b/s the resultant intersymbol interference led to error rates of the order of 1 in 5 or less. More important was the effect on bit synchronisation. The method employed synchronised the bit timer on to the instances of the

threshold crossings made by each pulse. Hence it is apparent why high levels of intersymbol interference would affect such a system.

As a result of these tests, it was obvious that the intersymbol-interference problem would have to be overcome before reliable transmission at higher data rates could be achieved. In other words it was necessary to incorporate an adaptive equaliser into the system, and to find a better method of bit synchronisation.

The solution of the intersymbol-interference problem was greatly assisted by the investigations performed below of the channel multipath structure likely to arise over the main test link. This test link was of the distance typically needed for communication purposes of 1000 Km. It was found that, for such a link, the signal received from the shortest multipath was likely to be of consistently strong amplitude in comparison with the other multipaths. This enabled the received pulse stream to be regarded as a main signal followed by a series of echoes. It will subsequently be demonstrated how the problems of obtaining bit synchronisation at high data rates and of realising the equaliser were simplified as a result.

We now deduce the likely multipath structure of a 1000 Km HF link. The shortest-distance modes in ionospheric propagation over 1000 Km are by single-hop reflection from the E-layer and the F_2-layer. Now the transmission frequency should be close to the MUF of the link for lowest attenuation from absorption. Let us then evaluate the MUF for each of the two paths in order to see whether one of these modes does not arise because the transmission frequency would be between the two MUF's.

The MUF is calculated from eqn. (2.1.4). Since we are only interested in the comparative figures for the E and F_2-layers it can be written as:

$$MUF = K N^{\frac{1}{2}} \sec \theta \qquad (3.0.1)$$

where K is a constant, N is the maximum electron density in the layer, and θ is the angle of incidence of the path.

In order to calculate the angle of incidence from the path geometry the following values are assumed for the heights of the layers:

Virtual height of the E-layer is 110 Km

Virtual height of the F_2-layer is 320 Km

Since, in the HF link in question, the transmitter-receiver separation is 1000 Km the geometry of the two paths is as shown in Fig. 3-0.3. (The effect of the earth's curvature is ignored since we are making an approximate analysis.) θ_E and θ_F are respectively the angles of incidence of the E-layer ray and the F_2-layer ray. We have then:

$$\sec \theta_E = 4.6$$

$$\sec \theta_F = 1.8$$

Taking values for the electron densities as 10^5 electrons per cc for the E-layer and 10^6 electrons per cc for the F_2-layer (from section 2.1.1) the MUF's are, from eqn. (3.0.1):

$$MUF_E = 4.6A$$

$$MUF_F = 5.7A$$

A is a constant. Now the transmission frequency is usually made 0.85 of the overall MUF as a "rule of thumb"[13]. Hence, since the MUF's for the two layers are close in value, both propagation modes are liable to exist in this case where a 1000 Km link is used.

Having ascertained that single-hop reflections from both layers are likely we now estimate the relative powers at the receiving aerial of these two modes.

Loss of power from the signal in propagating by these two modes is due to the following factors:

in E-layer reflection

(a) radiated power loss (proportional to $1/d^2$, d being the path length)

(b) absorption in traversing the D region

(c) absorption in the E-layer reflection

in F_2-layer reflection

(a) radiated power loss

(b) absorption in the D region

(c) absorption in the traversion of the E-layer

(d) absorption in the reflection off the F_2-layer

Let us compare the degree of attenuation for the two layers. The level of absorption in the reflection process for the "E" ray is more significant both because of the more oblique incidence (resulting in the ray spending a longer time in the layer) and because of the higher number of molecular collisions excited in the E-layer. The amount of absorption in the D region is also greater

for the "E" ray because this level is proportional to the secant of the angle of incidence[12]. These factors are counter-balanced by the 1.3 dB relative loss of the "F_2" ray due to its greater path length and by the fact that it must traverse the E-layer twice, which causes some absorption. Overall, then, the "E" and "F_2" rays are likely to be of similar strength.

We must now see what the relative strengths of the multiple-hop modes are in comparison to the single-hop reflections. In fact the multiple-hop modes tend to be more attenuated because:

(a) there is some loss in the reflection from the earth's surface (about 6 to 10 dB for a single-"bounce" F_2-layer reflection assuming "good" ground)

(b) the path length is greater (hence more radiated power loss)

(c) the absorbing regions are traversed for a greater total distance (the lesser angle of incidence means that each passage through the absorbing region is shorter; however, this is more than counteracted because there have to be more passages than in the single-hop-mode case, e.g. four for a single "bounce" off the earth's surface)

(d) the necessary smaller angle of incidence for the multiple-mode case causes a greater depth of penetration into the reflecting layer. In fact particular modes may not arise because their angles of incidence are such that the rays from them are not reflected at all.

We have seen that single-hop reflections from the E- and F$_2$-layers are likely to be roughly equally attenuated, and to be attenuated less than any multiple-hop reflections. Hence the first signal to be received, from the E-layer single-hop reflection, is likely to be the equal strongest mode.

We will now investigate whether the signal received from this multipath mode is always going to be of significant amplitude. As was found in Chapter Two the signal received from a single reflection theoretically consists of the interference sum of four rays, viz. a pair of ordinary and extraordinary rays from each of the low- and high-angle rays. For the case of oblique incidence, the high-angle rays are significantly attenuated because of their longer passage through the ionised layer[12]; this effect will be more pronounced for the E-layer as compared to the F$_2$-layer both because of the more oblique angle of incidence and because of its higher absorption characteristic. For a similar reason, the extraordinary ray which is reflected from a smaller angle of incidence will be more attenuated than the ordinary ray. Thus the received signal should contain a specular component (from the low-angle ordinary ray) with random fluctuations. In practice, due to deformities in the layer, the signal can consist of more than one such reflection, but, in any case, the amplitude distribution is found to be Rician[13]. Davies states that the amplitude of the composite ray is given, in the expression for the Rician pdf, by

$$P(A) = (2A/A_R^2) \exp(-(A^2 + A_S^2)/A_R^2) \, I_0(2AA_S/A_R^2) \, ,$$

with $\qquad A_S/A_R = 2$ or more.

A_S is the specular component, A_R is the rms value of the amplitude variations, A is the amplitude, and $I_0(x)$ is the Bessel function of zero order and imaginary arguments. The extent of the variations are easier to appreciate if we equivalence them to the amplitude distribution of the case when a carrier is rectifier-detected in the presence of bandpass noise. The distribution at the detector in this case is also Rician (see section 3.7.2 and eqn. (3.7.7)). Equivalencing the carrier amplitude to the specular component A_S and the noise amplitude to the random variations expressed by A_R leads to a "signal-to-noise" figure of 9 dB for A_S/A_R = 2.

We have seen then that the signal arising from the shortest multipath, the first to arrive, is likely to be the equal strongest mode and should always be of significant amplitude. (Such an assumption was shown to be correct by the results of the trials in Chapter Four.) Hence as far as the test system was concerned, it was assumed that the signal arriving at the receiver would consist of a main pulse followed by a series of echoes. We now continue with a description of how this assumption enabled the problems caused by intersymbol interference to be overcome.

In the second stage of the development of the test system bit synchronisation was achieved by transmitting the data in frames, as follows. For each block of data transmitted a gap was left equal to the expected maximum channel dispersion time. Then a single pulse, termed the "set-up" pulse, was sent followed by another gap of the same length as before. Then the next block of data would be transmitted. This pattern was repeated every 40 mS. The extent of the interruption of the data stream meant that the effective bit rate was reduced by 25%. However the adoption of such a scheme in conjunction with the assumption that the signal arriving from the

shortest multipath could be taken as the main pulse meant that a distortion-free pulse was received every 40 mS. That the set-up pulse would be free from multipath distortion was because

(a) no interference could occur from echoes of the previous block of data due to the length of the gap before the set-up pulse

(b) no interference could occur from other multipath echoes associated with the set-up pulse because the pulse from the first multipath to arrive was to be taken as the signal

Bit timing was achieved, therefore, by locking the synchroniser onto the set-up pulse, or more particularly the transition of the decision threshold made by the leading* edge of the set-up pulse. All required timing information for the decision-maker, equaliser, and threshold extractor was derived with respect to the set-up pulse. A fuller explanation of the operation of the bit synchroniser is given in section 3.10.

The effect of intersymbol interference on the bit error rate could only be solved by some form of equaliser, and the adoption of the frame method of data transmission enabled an equaliser to be simply realised. The topic of equalisation is considered in section 3.11. Here we see that the operation of conventional equalisers requires the use of what are effectively micro-computers. This is necessary in order for the equaliser to perform the processing involved as it minimises intersymbol interference and follows the changes

_____

*Hence, in practice, only the first half of the set-up pulse is required to be free from multipath distortion.

in the channel conditions.

The conventional types of equaliser have to determine the channel impulse response in a continuous data stream where considerable intersymbol interference may well be present (hence the requirement of processing). In the test system, however, because it consisted entirely of echoes the channel impulse response was displayed in the gap left between the set-up pulse and the start of the next block of data instead of being submerged in a continuous data stream. Hence the impulse response could be easily measured and stored by the equaliser using a series of sample-and-hold and low-pass filters. It was assumed that the changes in the channel impulse response were of limited bandwidth. Then it was possible, given the correct choice of sampling rate (i.e. repetition rate of the set-up pulse) and filter cut-off frequency, for the sample-and-hold and low pass filter circuits to follow these changes continuously. This equaliser could be termed automatic in that it required a periodic training pulse, but adaptive in that it could continually follow the changes in the channel conditions. Having established the channel impulse response the equaliser operated in decision-feedback mode to remove the intersymbol interference. This involved deciding when a pulse had been received and then subtracting its echoes from the bit stream.

We have seen how a decision-feedback equaliser that did not require processing hardware was implemented in the test system. The factor that enabled this to be done, the use of a set-up pulse, was put to other advantages. These were that firstly the envelope of the height of the set-up pulse could be extracted in order to set decision thresholds. Secondly the set-up pulse could be used as a polarity reference for PSK. Hence PSK did not have to be differentially encoded and decoded in the tests, a process that doubles the error

rate.

We will now briefly describe the remaining component parts of the test system in its final stage of development. We will do this by summarising the contents of this chapter, section by section. Note that each section number refers to the same-numbered component "block" in Figs. 3-0.1 and 3-0.2.

In section 3.1 the format of the bit stream is given. We have seen already how the data was transmitted in the form of frames. The data itself consisted of a 255-bit pseudo-random bit sequence. At the receiver the identical sequence was generated and synchronised to the transmitted one for the purpose of making on-line error measurements (see section 3.1.3).

The chosen method of varying the bit rate was to use the narrowest pulse possible and then to reduce the rate by leaving blank spaces between bits. It was felt that the alternative method of widening the pulse would have left the received pulse stream vulnerable to multipath distortion at low bit rates, and that it was better to "resolve" the multipath.

In section 3.2 we show how the pulse-stream spectrum was shaped by synthesising the correct pulse shape in the time domain. The pulse spectrum had to satisfy two conditions for data transmission over HF:

(a) the spectrum bandwidth must be restricted to 3 KHz (the HF channel bandwidth) in order not to cause interference with other users;

(b) the spectrum shape must satisfy the Nyquist criteria in order that there be no intersymbol interference present at the decision instants [18].

We find that an appropriate spectrum shape is that of the raised cosine. Two spectra of this type were used and this, together with the bit-rate variations available, meant that eight settings of the bit rate in the range from 0.4 Kb/s to 3.6 Kb/s were tested.

We discuss the methods of modulation in section 3.3. They are of three fundamental types; Amplitude Shift Keying (ASK), Phase Shift Keying (PSK), and Frequency Shift Keying (FSK). We see that three types of ASK, each having different modulation levels, were tested. We also see that binary PSK is a form of ASK, and that it was effected by transmitting a positive pulse for the binary "1", and a negative pulse for the binary "0". Finally in this section we consider FSK. We find that it requires excessive bandwidth to give comparable noise performance with ASK and PSK at high bit rates. We also see that it is a non-linear form of modulation and hence that the use of a time-domain equaliser would be precluded. For these reasons FSK was not employed in the tests.

The pre-emphasis filter and the modulator are described in sections 3.4 and 3.5 respectively. The filter was necessary in its low pass function to remove the high-frequency components generated in the pulse shaper. Its part in the pre-emphasis, de-emphasis technique is explained in section 3.8. The form of modulation used was Double Sideband (DSB). This meant that the transmitted signal had 6 KHz bandwidth instead of the 3 KHz value prescribed for the HF channel that would have been present with Single Sideband (SSB) modulation. It was decided that the extra complexity of SSB together with the problem of reinverting the carrier at the receiver put it outside the scope of implementation of these experiments. The modulator also incorporated an automatic gain control. This was to

control the modulated signal such that the peak transmitted power was just below the maximum available from the transmitter power amplifier.

In section 3.6 the radio link over which the experiments were made is described. We see that it contained conventional components, and that a 1 KW transmitter power amplifier was used. The path length was 1000 Km, the transmitter being based in Portsmouth and the receiver in Wick.

We discuss the demodulation methods in section 3.7. All of the modulation modes were coherently detected, with a Phase-locked Loop (PLL) being used to extract the carrier. It was felt that reinserting the carrier by filtering a pilot tone would have been disadvantageous for two reasons:

(a) The difficulty in realising a bandpass filter of bandwidth small enough such that the output phase is not affected by low-frequency components of the modulating waveform[18].

(b) In a multipath environment the pilot tone may be "nulled" to zero amplitude or exhibit fast phase changes on occasions because of the interaction of the multipaths.

Provision was also made for the rectifier detection of the ASK mode of modulation. This was so that comparison could be made with coherent detection. (It was realised though that baseband equalisation* could not be used here because rectifier detection is non-linear.)

In section 3.8 we consider the filtering techniques that can be employed to improve the signal-to-noise ratio at the decision maker. First of all the matched-filter method was examined. Under single-mode conditions its performance in the presence of noise was found to be only a slight improvement over simple decision making without the filter.

*Hereafter the term "equalisation" refers to that at baseband unless stated otherwise

Under two-multipath conditions its performance was worse. In any case a pulse stream where matched-filter detection was used would have required equalisation across the whole of the bit interval. Instead a spectrum pre-emphasis, de-emphasis technique was incorporated into the test system. It was found by experiment that the signal-to-noise ratio was improved in gaussian noise by 1 dB on average.

Section 3.9 deals with the method used to detect the level of the envelope of the set-up pulse in the presence of noise. This was required in order to set the decision level. It was also necessary to extract the envelope of any residual d.c. present in the demodulated signal in order to subtract it from the signal. This residual d.c. arose in the modulation process of two of the ASK modes.

In section 3.10 we find the bit synchroniser (that has already been referred to) described in detail.

As has been similarly referred to, section 3.11 considers the types of equalisation available, and gives the structure of the one that was implemented in the test system.

In section 3.12 we find the decision-maker described. The correct levels of the decision threshold for when both coherent detection and when rectifier detection are employed are discussed.

Finally, in section 3.13, the methods of error detection and counting that were used are given. We have seen that the identical data sequence was generated at the receiver for error measurements to be made. Three simultaneously-performed error counts were made, viz. errors before equalisation,errors after equalisation and a "datum" error measurement. This latter measurement was made on the

set-up pulse and the first bit in each data block. These two

bits were free of intersymbol interference, and so give a measure

of the residual error rate.

We start this chapter by observing the format of the bit

stream.

3.1   THE BIT-STREAM FORMAT

In this section we examine the methods of varying the bit rate and see the form of the test data that should be transmitted. Finally we see described the frame system of data transmission.

3.1.1   Bit-rate Variation

The maximum bit rate that can be transmitted is determined by the channel bandwidth and the shape of the pulse spectrum chosen (see section 3.2). There are two basic methods of varying the bit rate down from this figure. These are:-

(a)   to alter the pulse width as the reciprocal of the date rate - pulse widening;

(b)   to leave blank gaps after each bit (keeping the pulse width constant) - pulse blanking.

In the presence of noise method (a) is superior. If the pulse is widened its bandwidth is reduced. Hence the receiver filters can be correspondingly narrowed to lessen the noise present in the pulse stream. In other words reducing the bit rate by pulse widening improves the signal-to-noise ratio and thus the error performance. Method (b) uses the same pulse width throughout, and so the noise level remains unchanged as the bit rate is lowered.

Method (b), that of pulse blanking, is less affected by multipath. The multipath situation likely to arise over the type of radio link involved is given by Darnell[32]. He snows that the individual modes would have delays separated by at least 0.5 mS. For a maximum bit rate figure of 4 Kb/s (say) the pulse width will be 0.5 mS and so the multipaths will be resolved. If the bit rate is reduced by widening the pulse,the pulse will become progressively

more distorted by the multipath so reducing the signal level. In
fact there will be occasions when the signals from each path add up
vectorially to zero at the decision instan^{t} , so causing large error
bursts. This situation cannot be rectified by a conventional equaliser
because it is not intersymbol interference. If pulse blanking is used
multipath will simply cause intersymbol interference. This can be
removed by an equaliser because the individual paths are resolved.

When an equaliser is not used pulse blanking is still likely
to perform better under multipath conditions. Higher bit rates will
be possible because its narrower pulses means that the dispersion will
be less than by using pulse widening. Also there is a possibility
that at higher bit rates still the decision instants will lie between
echoes of the previous pulses, and so no intersymbol interference
occurs.

Obviously it is preferable to compare both methods in a
series of field tests. However, as there would otherwise have been
too many variables, it was decided to use only one of these two methods
We have noted that when the bit rate is lowered by widening the pulses
the noise level is reduced but the individual pulses become increasingly
vulnerable to multipath-caused distortion. This effectively lowers
the signal level and can cause large error bursts. In other words
the error performance may well get worse at lower bit rates. Hence,
because of the expected predomination of multipath conditions, method (b),
that of pulse blanking, was employed in the test system.

A summary of the bit rate values used is given in
section 3.2.3.

We now examine the type of test data that should be
transmitted.

3.1.2   Form of the Test Data

The data transmitted for the error-rate measurements should

be similar to that sent for actual communications. Now the outputs

of digital data sources can contain long strings of "1"s or "0"s,

and can show a large long-term bias in favour of one particular bit

polarity. For example, in facsimile, a weather map has a much greater

expanse of white (equivalent to one of the bit polarities) than black

(the other polarity). Since both of the data characteristics mentioned

above can lead to problems in system design and performance it is usual

to "scramble" the data so that the transmitted bits have a random

distribution. Such a scrambler can be realised with a pseudo-random

bit sequence generator. The implementation with logic circuitry of a

scrambler and de-scrambler having a 7-bit sequence is shown in Fig. 3-1.1*.

If the message to be sent contains long strings of ones or

zeroes the scrambled data will be the pseudo-random bit sequence or its

complement. It is sensible, then, to use a pseudo-random generator

to provide the bit pattern for error-rate measurements.

Hence, in the test system, an 8-bit generator having a cycle

length of 255 bits provided the test data. During the error tests

the cycle of an identical generator at the receiver was synchronised to

---

*The asynchronous scheme illustrated in Fig. 3-1.1 will increase the
error rate of the message by a factor of three because the received
scrambled data is fed into the delay line of the de-scrambler. A
superior system would have the pseudo-random sequence generators at
receiver and transmitter operating in closed loop, both being
synchronised with respect to their cycles. The actual data is then
exclusive-ORed to the transmitter pseudo-random generator to scramble,
and the received data is exclusive-ORed to the receiver generator to
de-scramble .

.the transmitter generator. The errors could then be counted using a bit-to-bit comparison.

We now describe in more detail the frame format of data transmission that has been referred to in the introduction.

3.1.3   Frame Pattern of Data Transmission

In order to facilitate certain receiver operations the bit stream was periodically interrupted to transmit an isolated pulse termed the set-up pulse.

It was necessary that, at the receiver, there should be no mutual interference between the set-up pulse and the data when multipath was present. Referring to Fig. 3-1.2* we see illustrated the situation around the set-up pulse corresponding to the maximum expected channel dispersion time τ. The worst case echo (drawn in dashed lines) is a time τ after the main pulse. It is apparent from Fig. 3-1.2 then that the interval between either side of the set-up pulse and the start and finish of the data blacks should be just greater than the sum of the maximum expected channel dispersion time τ and one pulse width T, i.e.

$$T' \geqslant \Gamma + \tau$$

The gap T' was determined by the initial experiments between Portsmouth and London as 5 mS.

Because of the resultant interval of 10 mS when the data stream is interrupted there will be a redundancy factor in the information rate dependent on the transmission rate f_0 of the set-up pulse. This rate was again determined from experiment (see section 3.9) and was

---

*In this and subsequent pulse-stream diagrams the pulses are shown by triangles for ease of drawing.

set at 25 Hz, i.e. a period T_0 of 40 mS. Hence in such a system if the original bit rate is R the overall bit rates B is

$$B = R (T_0 - 2T')/T_0$$

Using the values decided for T_0 and T' (of 40 mS and 5 mS respectively) the overall bit rate is given by

$$B = 0.75R \qquad\qquad (3.1.1)$$

3.2   RESTRICTION OF BANDWIDTH BY PULSE SHAPING

Over HF radio users are allocated channels of fixed band-
widths (3 KHz) at specified carrier frequencies.   Hence the modulated
pulse stream must have its bandwidth restricted in order not to
interfere with other users. In this section we consider the pulse
shaping of ASK. Amplitude modulation is a linear process (as is
shown in section 3.3) and so the spectrum shaping can be performed at
baseband. When modulated the baseband spectrum is linearly transposed
to be centred on the carrier frequency. In the next section we will
see that binary PSK can be realised as a form of ASK. Hence the process
explained here for spectrum-shaping ASK can be applied to binary PSK.
As far as FSK is concerned we find in section 3.3 that it is a non-
linear process and has, in its serial form, an excessive bandwidth
requirement. We find that therefore this form of modulation was not
implemented in the test system. To summarise, in section 3.2.1
we consider the theory of the adopted scheme of spectrum shaping,
and in section 3.2.2 we see how the scheme was realised. Finally in
section 3.2.3, we see summarised the bit-rate values used in the
test system.

3.2.1   Theory of the Spectrum-shaping Method

As well as possessing limited bandwidth a further requirement
of the spectrum-shaping process is that the pulses must not exhibit
any mutual intersymbol interference at the instants of decision-
making. Three criteria that the pulse spectrum shape must satisfy
can be formulated from Nyquist's work[17] (e.g. by Bennett and Davey[18]).
One such shape that satisfies these criteria is that of the raised
cosine. Its one-sided frequency response H(f) is shown in Fig. 3-2.1

and is given by:

$$H(f) = 1 \qquad\qquad f < f_c - f_x$$

$$H(f) = (1 + \cos[\pi(f + f_x - f_c)/2f_x])/2 \quad f_c - f_x < f < f_c + f_x \qquad (3.2.1)$$

$$H(f) = 0 \qquad\qquad f > f_c + f_x$$

The ratio f_x/f_c is termed the roll-off factor R. Performing the Fourier transform on eqn. (3.2.1) (as done in Appendix 3.1) gives the eqn. h(t) of the pulse shape as:

$$h(t) = \frac{\sin 2\pi f_c t}{2\pi f_c t} \cdot \frac{\cos 2\pi f_x t}{1 - (4f_x t)^2} \qquad\qquad (3.2.2)$$

This function is drawn in Fig. 3-2.2. We see that the $\sin 2\pi f_c t$ term in eqn. (3.2.2) causes h(t) to have zero crossings when:

$$2\pi f_c t = n\pi$$

i.e. when
$$t = \frac{n}{2f_c}$$

n being an integer. Hence pulses may be added at intervals of $n/2f_c$ and sampled at their peaks without any interference being present from the sidelobes of previous or subsequent pulses.

The fastest rate that bits can be transmitted without any interference being caused is when they are added at intervals of $1/2f_c$. The maximum bit rate then is $\underline{2f_c}$.

A useful relationship to derive is that connecting the maximum bit-rate frequency, $2f_c$, and the frequency at which the spectrum

amplitude is just zero, F, to the roll-off factor $R(=f_x/f_c)$. Let us write:

$$2f_c = 2f_c \frac{(f_c+f_x)}{f_c+f_x}$$

$$= 2 \frac{(f_c+f_x)}{1+f_x/f_c}$$

Now $F = f_c + f_x$ and the roll-off factor $R = f_x/f_c$. Hence the bit rate B is given by:

$$B = \frac{2F}{1+R} \qquad\qquad (3.2.3)$$

We are interested in making efficient use of the available bandwidth. From eqn. (3.2.3) we see that the maximum bit rate that can be transmitted over a channel of basebond limit FHz is 2 Fb/s (it is termed the Nyquist rate). This gives R = 0. Then the pulse spectrum is square (because $f_x = 0$) and the pulse shape has the familiar sin x/x characteristic:

$$h(t) = \frac{\sin 2\pi f_c t}{2\pi f_c t} \qquad \text{(from eqn. (3.2.2))}$$

However, in practice, the use of such a pulse spectrum characteristic is not feasible. In the first case the square spectrum shape is impossible to realise because the appropriate filter would have an infinite delay. Secondly, the overlapping pulse tails from a composite pulse stream represent a divergent series which can add up to large values[18]. Then, if the decision-maker is not precisely timed, the "eye opening" will be much impaired and incorrect interpretations

of pulses may be made. Furthermore, since the pulse tails will add up to large values outside the sampling instances, the effective transmitted power will be much reduced.

In contrast overlapping tails of pulses formed from the raised cosine spectrum (having obviously a non-zero value of f_x) produce a convergent series in the sum of their amplitudes. This is because the denominator in eqn. (3.2.2) increases faster than linearly with t. Thus slight errors in timing do not produce significant amounts of intersymbol interference.

We have seen that a pulse stream having a raised-cosine spectrum can form a reliable means of transmitting digital data. The method of synthesising the pulses will now be considered.

The conventional means of forming a signal with a particular spectrum is to construct a filter with the correct frequency response and excite it with impulses. There are, however, considerable theoretical and practical difficulties in filter construction where the response has a function such as eqn. (3.2.1). A simpler and more flexible approach is to synthesise sampled versions of the pulse shapes in the time domain. These can then be fed through a low-pass filter to remove the sampling-frequency harmonics. Such an application of the Sampling Theorem as follows commends itself because the pulse spectrum is band-limited.

Assume the analogue waveform being sampled is g(t) and that its Fourier transform is G(f), i.e.

$$g(t) \leftrightarrow G(f)$$

The sampling waveform s(t) is square-topped of width τ. Its Fourier transform S(f) is given by:

$$S(f) = \tau \frac{\sin w\tau/2}{w\tau/2}$$

The sampled waveform is denoted by:

$$g_s(t) \leftrightarrow G_s(f)$$

It is a series of waveforms of $s(t)$ weighted by $g(nT_0)$ (T_0 being the sampling interval), i.e.

$$g_s(t) = \sum_{n=-\infty}^{+\infty} s(t-nT_0) \, g(nT_0)$$

as shown in Fig. 3-2.3.

From the square-topped sampling theorem[35], the spectrum of $g_s(t)$ is given by:

$$G_s(f) = \frac{\tau}{T_0} \frac{\sin w\tau/2}{w\tau/2} \sum_{n=-\infty}^{\infty} G(f-nf_0) \qquad (3.2.4)$$

where $f_0 = 1/T_0$ (i.e. it is the sampling frequency).

Let us assume that the waveform $g(t)$ being sampled is band-limited to a frequency f_c, and that the sampling frequency f_0 is such that:

$$f_0 > 2f_c$$

The spectrum $G_s(t)$ of the sampled waveform is then the product of the two functions drawn in Figs. 3-2.4(a) and 3-2.4(b).

Hence, if the sampled waveform is generated and fed through an ideal low-pass filter having a cut-off frequency between f_c and f_c-f_0, the resultant waveform at the filter output will have a spectrum $G(f)$ distorted by the multiplicative term in Fig. 3-2.4(b):

$$S(f) = \frac{\tau}{T_0} \cdot \frac{\sin w\tau/2}{w\tau/2} \quad \text{(including the } T_0 \text{ term into } S(f))$$

This distortion term can be removed by creating the sampled waveform $g_s(t)$ from the Fourier transform of $G(f)/S(f)$. Then:

$$G_s(f) = S(f) \sum_{n=-\infty}^{\infty} \frac{G(f-nf_0)}{S(f-nf_0)}$$

After low-pass filtering as above, we are left with the "n=1" term:

$$S(f) \cdot \frac{G(f)}{S(f)}$$

$$= G(f)$$

However, in a real transmission system, there are further factors that distort the pulse spectrum viz:

(1) the low-pass filter is not ideal in its amplitude and phase response and even less so is the IF filter of the communications receiver used in the transmission system (though in theory the same process as performed to remove the effects of $S(f)$ could be employed for these filters if it were possible to determine their exact responses)

(2) the HF channel itself is liable to cause some frequency distortion;

(3) the analogue waveform $g(t)$ (i.e. the pulse) has an infinite duration if the spectrum is band-limited, and the sampling theorem assumes an infinite sample train in order to form $\Sigma G(f-nf_0)$

exactly. In practice the generator used to

produce $g_S(t)$ must have a finite length and so

only a "windowed" version of g(t) will be formed*.

The resultant spectrum will be slightly misshapen

and spread. This spreading will be removed by

the transmitter and receive filters leading to

further slight pulse distortion.

In practice, then. it is not really necessary to derive

$g_s(t)$ from the Fourier transform of G(f)/S(f) instead of from Ĝ(f).

The distortion introduced by the multiplicative term S(f) can be

reduced to insignificant levels by making τ small so that

$\frac{\sin w\tau/2}{w\tau/2}$ is substantially unity over the frequency band: $|f| < f_c$

(i.e. the bandwidth of G(f)).

We have seen then how an analogue waveform having a band-

limited spectrum of specified shape can be generated by applying

the sampling theorem. We will now see how this method was used in

the test system to produce the pulse shapes. These were of two

varieties:

(a) raised-cosine spectrum, 100% roll-off

(b) raised-cosine spectrum, 33% roll-off

(a) 100% roll-off

The roll-off factor is thus such that in eqn. (3.2.1)

$$\frac{f_x}{f_c} = 100\%$$

*This is similar to the case of realising the spectrum shaping with

a filter having the correct shape but not having an infinite delay.

i.e.

$$f_x = f_c$$

Substituting for f_x in eqn. (3.2.1) gives the spectrum equation as:

$$H(f) = (1 + \cos \pi f/2f_c)/2 \quad , \quad f < 2f_c$$

$$= 0 \quad\quad\quad\quad , \quad f > 2f_c$$

This function is shown in Fig. 3-2.5.

The pulse shape equation is expressed by (from eqn. (3.2.2)):

$$h(t) = \frac{\sin 2\pi f_c t}{2\pi f_c t} \quad \frac{\cos 2\pi f_c t}{1-(4f_c t)}$$

i.e.

$$h(t) = \frac{\sin 4\pi f_c t}{4\pi f_c t(1-(4f_c t)^2)}$$

This is drawn in Fig. 3-2.6. Here we see that successive pulses can be transmitted at intervals separated by $1/2f_c$.

From eqn. (3.2.3) we obtain the bit rate B as a function of the band-limit frequency F as:

$$B = \frac{2F}{1+R}$$

Now the roll-off factor R is 100%. Therefore:

$$B = F \quad\quad\quad\quad\quad\quad (3.2.5)$$

In other words the bit rate using a 100% roll-off raised-cosine spectrum is 50% of the Nyquist rate.

The generated sampled version of the pulse $h_s(t)$, is shown in Fig. 3-2.7. We see that $h(t)$ is reproduced only in the limits:

$$1/2f_c < t < 1/2f_c$$

i.e. the main lobe. The level of energy in the sidelobes is small so that the amounts of intersymbol interference and spectrum spreading introduced by this approximation are minimal.

As can be seen in Fig. 3-2.7 the sampling interval was made $1/8f_c$. Ignoring the distorting effect of the windowing of pulse sidelobes the spectrum $H_s(f)$ of the sampled pulse is, from eqn. (3.2.4), given by:

$$H_s(f) = \frac{\tau}{T_0} \frac{\sin w\tau/2}{w\tau/2} \sum_{n=-\infty}^{\omega} H(f-nf_0)$$

Now τ, the sampling interval, is $1/8f_c$ as is T_0, the sampling rate. f_0 is the reciprocal of the sampling rate.

Hence:

$$H_s(f) = \frac{\sin 2\pi f/16f_c}{2\pi f/16f_c} \sum_{n=-\infty}^{\infty} H(f-n8f_c)$$

The spectrum of the sampled pulse is shown in Figs. 3-2.8(a) and (b). Fig. 3-2.8(a) illustrates

$$\Sigma H(f - n8f_c)$$

and Fig. 3-2.8(b) shows

$$\frac{\sin 2\pi f/16f_c}{2\pi f/16f_c}$$

(b) 33% roll-off

Here

$$R = \frac{f_x}{f_c} = \frac{1}{3}$$

hence

$$f_x = f_c/3$$

Substituting for f_x in eqn. (3.2.1) gives the spectrum function as:

$$H(f) = 1 \qquad\qquad , \quad f < 2 f_c/3$$

$$H(f) = (1+\cos 3\pi(f-2f_c/3)/2f_c)/2 \quad , \quad 2f_c/3 < f < 4f_c/3$$

$$H(f) = 0 \qquad\qquad , \quad 4f_c/3 < f$$

It is illustrated in Fig. 3-2.9.

As before the pulse shape is found from eqn. (3.2.2) by substituting for f_x (as $f_c/3$). Hence:

$$h(t) = \frac{\sin 2\pi f_c t}{2\pi f_c t} \frac{\cos 2\pi f_c t/3}{1-(4f_c t/3)^2}$$

and this is shown in Fig. 3-2.10.

From eqn. (3.2.3), the bit rate B as a function of the band-limit frequency F is given by

$$B = \frac{2F}{1+1/3}$$

$$\therefore \quad B = 3F/2 \qquad\qquad (3.2.6)$$

Hence the bit rate here is 75% of the Nyquist rate.

The shape $h_s(t)$ of the sampled pulse generated is shown in Fig. 3-2.11. Because of their significant size the first pair of sidelobes are generated as well as the main lobe.

The spectrum $H_s(f)$ of the sampled pulse is, since the sampling interval τ is $1/8 \; f_c$, again:

$$H_s(f) = \frac{\sin 2\pi f/16f_c}{2\pi f/16f_c} \; \Sigma \; H(f-n8f_c)$$

The spectrum of $h_s(t)$ is shown in Figs. 3-2.12(a) and (b).

We now consider the type of low-pass filter which for the two pulse spectrum shapes is capable of removing the sampling-frequency harmonic components and yet does not distort their spectra. In practice the choice made for the sampling pulse width τ ensured that the type of low-pass filter was not very critical. This was for two reasons:

(a) the sampling-pulse width was such that it equalled the sampling period, hence causing the zero crossing of the $\frac{\sin w\tau/2}{w\tau/2}$ function to be at the centre of the first harmonic of the sampling frequency (see Figs. 3-2.8 and 3-2.12);

(b) the sampling frequency was at least four times the pulse band-limit frequency. Hence there was a considerable "dead" band between $H(f)$ and $H(f-f_0)$.

These two factors ensured that the low-pass filter did not need to have a steep cut-off slope. A fourth-order Butterworth filter whose cut-off frequency was $4f_c$ was found to lead to minimal distortion of the pulse as well as providing sufficient attenuation. The method

adopted in the test system, however, was to employ a second-order filter whose cut-off frequency was F (the band-limit frequency of the two spectra). This caused some slight pulse distortion but was used because it was part of the technique described in section 3.8 to improve the signal-to-noise ratio.

The reason that two types of pulse shape were used are as follows. The 100% roll-off pulse has sidelobes of small amplitude as can be seen in Fig. 3-2.6. Hence the peak amplitude of a random stream of pulses of this shape will not be much greater than the height of an individual pulse. For the 33% roll-off factor the sidelobes are significant and hence in this case the peak amplitude of a random pulse stream is somewhat greater than the individual pulse heights. In a system where the transmitter is peak-power limited the eye opening at the receiver is higher with the 100% roll-off pulse than with the 33% roll-off pulse. The situation was exacerbated in practice because of the filter system used. Measurements showed that the 100% roll-off pulse gave an improvement over the 33% roll-off pulse of 2 dB for ASK and 4.5 dB for PSK. (As will be seen in section 3.3.1, for PSK negative pulses are transmitted for the binary "0". Hence the effect is greater for this mode.) To summarise then, the 33% roll-off pulse gives a higher bit rate (see section 3.2.3) but a 2 to 4 dB worse signal-to-noise ratio than the 100% roll-off pulse.

3.2.2   Synthesis of the Sampled Pulse Shapes

The stepped waveforms of the two types of sampled pulses can be considered as being the impulse responses of transversal filters whose tap coefficients correspond to the amplitudes of the steps. Hence such a filter can be used to synthesise the pulse shapes.

The realisation of the filter used to synthesise the 33% roll-off pulse is shown in Fig. 3-2.13. We see that at the input the digital bit stream is gated by a pulse whose width is that of one "step". This pulse is periodic at the bit rate (i.e. its period is $1/2f_c$ - see Fig. 3-2.11). The output of the gate is fed into a series of TTL shift registers which respresents the delay line of the filter. They are clocked by a frequency having the same period as the gating pulse. Hence the gating pulse ripples down the shift register chain whenever a "1" is present in the digital bit stream.

The outputs of the shift registers are connected to NOR and OR gates. These gates have open-collector outputs. They, in effect, present a short circuit to earth for a "0" and an open circuit for a "1". The filter taps are realised by connecting resistors corresponding to the step heights from these gate outputs to the virtual earth of a shunt-feedback amplifier. The virtual "earth" is at a fixed positive voltage. This amplifier acts to sum the taps as follows. Let us assume that all the shift register outputs are at "0". Now let a "1" appear in the digital bit stream so that the gating pulse causes a "1" to be rippled down the shift register chain. Then the pulse shape depicted will appear at the amplifier output.

A design feature is that the virtual earth of the amplifier sat at 1/2 of the supply voltage. Also a further resistor was connected from the supply voltage to the virtual earth to offset the amplifier output to zero. These two facts ensured that changes in the supply voltage did not produce any offset voltage at the amplifier output. This is an important factor when setting the level of modulation (see section 3.3).

The generation of the sampled pulse for the 100% roll-off case was similar but here only the main lobe was generated. Hence half the number of gates were needed.

We now summarise the bit-rate values that were used in the test system.

3.2.3   Bit Rate Values

The band-limit frequency F of the pulse spectra was made 3.2 KHz in order to use the highest bit rates in accordance with the HF channel bandwidth requirements. Hence the maximum raw bit rates used were:

(a) for the 100% roll-off pulse   B = 3.2 Kb/s (from eqn. 3.2.5)

(b) for the 33% roll-off pulse   B = 4.8 Kb/s (from eqn. 3.2.6)

In section 3.1.3 we saw though that the overall bit rate was 75% of the raw bit rate because of the frame format of data transmission. Hence the maximum bit rate figures were:

(a)   100% roll-off   B = 2.4 Kb/s

(b)   33% roll-off   B = 3.6 Kb/s

For the reasons given in section 3.1.1 the bit rate was reduced by leaving gaps, or blanks, after each pulse. Hence, if 1/N is the bit-rate reduction factor pulses are transmitted at intervals of $1/2f_cN$ (see Figs. 3-2.6 and 3-2.10). The bit rate figures used then are summarised in Fig. 3-2.14(a) and (b). As can be seen the bit rate was varied from 3.6 Kb/s to 0.4 Kb/s overall.

3.3   DISCUSSION OF THE METHODS OF MODULATION

The three basic methods of carrier modulation, viz.

amplitude, phase, and frequency, were to be compared in the experiments.

(As is usual in digital terminology they will be referred to as ASK,

PSK, and FSK respectively.)

In section 3.3.1 we consider ASK and PSK. We will see that

a binary PSK system can be implemented by treating it as bi-polar

ASK. We will see how the pulses generated by the pulse shaper are

combined to produce the type of modulation required. We will also

find the aspects of these different modes compared.

In section 3.3.2 FSK is examined as a method of serial binary

communication. We will see that it presents the significant

disadvantages (compared to ASK and PSK) of requiring greater bandwidth

and of being non-linear. It was concluded therefore that its testing

in the experiments was not justified.

3.3.1   ASK and PSK

Three types of ASK(each having different modulation levels),

and one type of PSK, were used in the experiments. First of all we

must consider what is meant by the term "modulation level" in order

to define the modes.

Let f(t) be a general modulating signal having f_{max}, f_{min}

and f_{pp} as, respectively, its maximum, minimum and peak-to-peak values.

Let it be biased about zero such that

$$f_{max} = - f_{min}$$

as shown in Fig. 3-3.1.

f(t) is modulated by multiplying it with the carrier cos wt, and the modulation level is adjusted by the addition of an amount of carrier V cos wt.   The modulated signed S(t) is then:

$$S(t) = [f(t) + V] \cos w_c t \qquad\qquad (3.3.1)$$

Two amplitude modulation (AM) modes can now be defined:-

(a) AM plus carrier

Here:

$$V \geqslant f_{pp}/2 \quad (i.e. -f_{min})$$

Then the signal is said to have a modulation level M given by:

$$M = \frac{f_{pp}}{V + f_{pp}/2} \times 100\%$$

For example, when

$$V = f_{pp}/2$$

the modulated signal has a modulation level of 100%.

The general situation of S(t) is shown in Fig. 3-3.2.

A feature of the "AM plus carrier" mode is that

$$V + f(t) \geqslant 0 \qquad \text{for all t}$$

because

$$V \geqslant - f_{min}$$

Therefore the modulated carrier experiences no π phase shifts which the zero crossings of the modulating signal would otherwise cause. Hence rectifier or envelope detection can be employed.

(b) AM with suppressed carrier

This mode arises when:

$$V = 0$$

The resultant mode is termed "suppressed carrier" because modulating signals such as a sine wave have a mean value of zero when biased as in Fig. 3-3.1. Then the spectrum of S(t) has no component centred on the carrier frequency because there is no dc content in the signal at baseband.

The use of the "AM with suppressed carrier" mode means that, for a given peak value of the modulated carrier, the maximum value of the modulating signal can be obtained at the receiver. However the modulated carrier contains 0 and π phase modes and so some form of coherent detection is necessary.

We can now define the different types of ASK and PSK used in the test system. Before doing so though it is helpful to consider how aspects of each type are likely to affect the operation of the receiver system in order to see why each was chosen.

(1) Ratio of eye opening to peak modulated signal level (the eye opening being the difference between the "1" and "0" levels at the sampling instants.)

One feature of each modulation mode to be compared is what signal-to-noise ratio does it produce at the receiver? For there to be a worthwhile evaluation of this some parameter of the modulated signal level must be fixed. It is obviously desirable that the

signal-to-noise ratio of the received signal is as high as possible.
Hence the transmitter power amplifier should be functioning close
to overloading. Now the maximum-power limitation in an amplifier
designed for a known load (the aerial) is that of peak power as
opposed to mean power. If the peak power rating is exceeded then
at the point where the amplifier output-stage voltage swing "touches"
the power supply rails the carrier is clipped. Carrier harmonics
would then be transmitted causing interference with other users.
Hence the level of the maximum part of the carrier envelope should
be fixed such that at this instance the transmitter power amplifier
is just below saturation. (In the test system this operation was
performed by Block 3-5 in Fig. 3-0.1.)

Hence in order to have a comparative measure for each mode
of the received signal-to-noise ratio we must consider the ratio
of the eye opening to the peak modulated signal (the eye opening
being that of the received signal).

(2) Average level of the modulated signal

A phase-locked loop was used in the receiver for carrier
extraction (see section 3.7). A PLL performs better the higher the
average signal-to-noise ratio. Since the peak power has to be fixed
the average power varies according to the type of modulation and
the value of the bit rate.

(3) Whether the type of mode has a residual dc level

As will be seen in the later sections the presence after
demodulation of any residual dc in the pulse stream affected certain
receiver operations. These are: firstly that the decision-making
threshold would be shifted; secondly the type of equaliser that was
realised treated any residual dc as intersymbol interference and

.therefore introduced level shifts of its own. Hence it was necessary to remove the d.c. and, as we will see in section 3.9 , this process as implemented amplified the noise present in the pulse stream.

(4) Whether the mdoulated carrier has uni-phase or bi-phase components

If the modulating signal has no zero crossings the carrier when modulated is of constant phase. Thus a carrier extractor such as a phase-locked loop need be stable only in the one phase. However if the modulating signal has zero crossings then since

$$- A \cos w_c t = A \cos (w_c t + \pi)$$

the modulated carrier has 0 and π phase components. Then the PLL will need to be stable in both phase modes.   Such a system is likely to "flip" between the two stable modes which would give rise to error bursts at the receiver.

Each of the modes that were tested, viz. the three types of ASK and the one type of PSK, are now described. The pulse stream diagrams for each mode are shown in Figs. 3-3.3   and 3-3.4. Fig. 3-3.3 illustrates the pulse-blanking method of reducing the bit rate (see Fig. 3-2.14 for the correspondence between the bit-rate reduction factor and the bit rate).   Fig. 3-3.4 illustrates the data-interrupt   period where the set-up pulse is transmitted (see section 3.1.3).   In these two diagrams the pulses are represented by triangles for ease of drawing. (The triangles correspond to the pulse main lobes.)   The diagrams are drawn to scale to produce a peak modulated carrier level of V. Hence the eye-opening levels can be appreciated.   (Modulation is assumed to be performed by straight multiplication by $\cos w_c t$ of the waveforms in the diagrams.)   The

receiver-operation considerations (1) to (4) referred to earlier are summarised for each mode in Fig. 3-3.5.

We now describe each mode:

(a) <u>ASK</u>

Conventionally, if T is the bit-rate period, the binary "1" is transmitted as:

$$A \sin w_c t \qquad 0 < t < T$$

$$0 \qquad \text{otherwise}$$

The binary "0" is transmitted by a space (i.e. no signal).

This ideal ASK system is termed On-Off Keying, and can be described as being formed by 100% carrier modulation with square pulses. However, as has been discussed, bandlimited pulses must be used and so pulses having raised-cosine spectrums were used in place of square pulses in the test system (section 3.2). Because these pulses have sidelobes there are zero crossings present in the unmodulated signal. Hence the modulated carrier contains 0- and π- phase components.

(b) <u>ASK plus carrier (ASK + C)</u>

This mode is identical with (a) except that some residual d.c. was added to prevent zero crossings caused by the pulse side-lobes. The resultant modulation level was 80%.

(c) <u>ASK-suppressed carrier (ASK-SC)</u>

Again the pulse stream is the same as used for the ASK mode. However it has now been biased so that the maximum and minimum

parts of the waveform equally span the zero level. Hence this mode qualifies for the "AM with suppressed carrier" term defined earlier.

(d) PSK

In a binary PSK system the carrier is shifted by π for the opposite polarity bit. Since

$$\cos(w_c t + \pi) = - \cos w_c t$$

binary PSK can be realised with a bipolar ASK format as illustrated in the pulse-stream diagrams. A "1" is transmitted with a positive pulse and a "0" is transmitted with a negative pulse.

The three ASK modes were realised in the test system by using one of the pulse-shape synthesisers described in section 3-2.2. The PSK mode used two, one for positive pulses, and one for negative pulses. The digital bit stream was multiplexed between the two according to the polarity of each individual bit.

We now discuss the FSK mode of modulation to see whether it presents a feasible means for high-speed serial data transmission over the HF channel.

3.3.2   FSK

The first step is to investigate whether frequency modulation is linear and so whether, in order to counteract the likely presence of multipath, FSK can be equalised.

Let $M[f(t), t]$ represent a general modulation process on a signal $f(t)$, and let $D[g(t), t]$ be a general demodulation process on a received signal $g(t)$.

The effect of multipath is to sum up delayed versions of the modulated signal. Let us assume that, for example, a two-path

situation exists. Then the received signal is:

$$A_1 M[f(t+t_1), t+t_1] + A_2 M[f(t+t_2), t+t_2]$$

t_1 and t_2 are the propagation delays of the two paths and A_1 and A_2 are the path strengths.

The linearity requirement of the modulation process is that, after demodulation, we have

$$D[A_1 M[f(t+t_1), t+t_1] + A_2 M[f(t+t_2), t+t_2], t]$$

$$= B_1 f(t+t_1) + B_2 f(t+t_2) + K \tag{3.3.2}$$

where B_1, B_2, and K are <u>constants</u>. Then there is no cross-coupling between $f(t+t_1)$ and $f(t+t_2)$ and so samples of $f(t+t_2)$ can be added or subtracted from $f(t+t_1)$, or vice versa. This is the action of orthodox equalisers.

For example, let us test coherently-detected AM for linearity using the above definition. We have the modulation process as:

$$M[f(t),t] = f(t) \cos w_c t$$

The received signal g(t) is

$$g(t) = A_1 f(t+t_1) \cos(w_c t + \theta_1) + A_2 f(t+t_1) \cos(w_c t + \theta_2)$$

where

$$\theta_1 = w_c t_1$$

and

$$\theta_2 = w_c t_2$$

and A_1 and A_2 are propagation constants as above.

The demodulation process is the operation

$$D[g(t), t] = g(t) \cos(w_c t + \theta_3)$$

followed by a low-pass filter to remove frequency forms of value $2w_c t$.

Hence, for the above example, this coherent demodulation process leads to

$$D[g(t),t] = A_1 \cos(\theta_1 - \theta_3) f(t+t_1)/2 + A_2 \cos(\theta_2 - \theta_3) f(t+t_2)/2$$

$[A_1/2]\cos(\theta_1 - \theta_3)$ and $[A_2/2]\cos(\theta_2 - \theta_3)$ are obviously constants, and so AM, coherently detected, is a linear process.

The same procedure will now be applied to FM. The modulation process here is given by

$$M[f(t),t] = \cos[w_c t + k\!\int f(t) dt]$$

where k is a constant.

The demodulation process determines the instantaneous frequency of a signal[36]. Hence:

$$D[\cos\theta(t),t] = A\frac{d\theta(t)}{dt}$$

Thus,

$$D[\cos(w_c t + k\int f(t)dt),t] = A\frac{d}{dt}(w_c t + k\int f(t)dt)$$

$$= Aw_c + Akf(t)$$

i.e. the original modulating signal is recovered, together with the constants A, w_c, and k.

Now, as in the example for AM, we wish to know

$$D\{A_1 M[f(t+t_1), t+t_1] + A_2 M[f(t+t_2), t+t_2],t\}$$

Let

$$R(t) = A_1 M[f(t+t_1), t+t_1] + A_2 M[f(t+t_2), t+t_2]$$

$$= A_1\cos(w_c t + k\int f(t+t_1)dt + \phi_1) + A_2\cos(w_c t + k\int f(t+t_2)dt + \phi_2)$$

where A_1 and A_2 are the multipath strengths and,

$$\phi_1 = w_c t_1 \quad, \quad \phi_2 = w_c t_2$$

Let us write

$$\theta_1(t) = w_c t + k\int f(t+t_1)dt + \phi_1$$

and

$$\theta_2(t) = w_c t + k\int f(t+t_2)dt + \phi_2$$

Then

$$R(t) = A_1 \cos\theta_1(t) + A_2 \cos\theta_2(t)$$

This function is drawn as a phasor diagram in Fig. 3-3.6. R(t) can be written in the form

$$R(t) = S(t) \cos\theta(t)$$

where

$$S(t) = f[A_1, A_2, \theta_1(t), \theta_2(t)]$$

and, by inspection of Fig. 3-3.6, $\theta(t)$ is given by:

$$\tan\theta = \frac{A_1 \sin\theta_1 + A_2 \sin\theta_2}{A_1 \cos\theta_1 + A_2 \cos\theta_2} \quad \text{(dropping the bracketed t)}$$

Now the demodulator output is given by:

$$D[\cos\theta, t] = \frac{A d\theta}{dt}$$

Let $\tan\theta = x$

where

$$x = \frac{A_1 \sin\theta_1 + A_2 \sin\theta_2}{A_1 \cos\theta_1 + A_2 \cos\theta_2}$$

Then differentiating $\tan\theta = x$ leads to:

$$\frac{d\theta}{dt} \sec^2\theta = \dot{x}$$

$$\therefore \frac{d\theta}{dt}(1 + \tan^2\theta) = \dot{x}$$

$$\therefore \frac{d\theta}{dt} = \frac{\dot{x}}{1+x^2}$$

Let us evaluate the expression for $d\theta/dt$ in sections.
Firstly:

$$1 + x^2 = 1 + \left[(A_1 \sin\theta_1 + A \sin\theta_2)/(A_1 \cos\theta_1 + A_2 \cos\theta_2)\right]^2$$

$$= \frac{[A_1\cos\theta_1 + A_2 \cos\theta_2]^2 + [A_1 \sin\theta_1 + A_2 \sin\theta_2]^2}{[A_1 \cos\theta_1 + A_2 \cos\theta_2]^2}$$

$$= \frac{A_1^2 + A_2^2 + 2A_1A_2 \cos(\theta_1 - \theta_2)}{[A_1 \cos\theta_1 + A_2 \cos\theta_2]^2}$$

Secondly:

$$x = \frac{A_1 \sin\theta_1 + A_2 \sin\theta_2}{A_1 \cos\theta_1 + A_2 \cos\theta_2}$$

$\therefore$ Differentiating this expression to find $\dot{x}$ leads to:

$$\dot{x} = \frac{\dot\theta_1[A_1^2 + A_1A_2 \cos(\theta_1-\theta_2)] + \dot\theta_2[A_2^2 + A_1A_2 \cos(\theta_1-\theta_2)]}{[A_1 \cos\theta_1 + A_2 \cos\theta_2]^2}$$

Now the required function $Ad\theta/dt$ (the demodulator output)
is given by

$$\frac{Ad\theta}{dt} = \frac{A\dot{x}}{1+x^2}$$

By substituting for $\dot{x}$ and $(1 + x^2)$ and simplifying we have:

$$\frac{Ad\theta}{dt} = \frac{\dot\theta_1 A[A_1^2 + A_1A_2 \cos(\theta_1-\theta_2)]}{A_1^2 + A_2^2 + 2A_1A_2 \cos(\theta_1-\theta_2)} + \frac{\dot\theta_2 A[A_2^2 + A_1A_2 \cos(\theta_1-\theta_2)]}{A_1^2 + A_2^2 + 2A_1A_2 \cos(\theta_1-\theta_2)}$$

Now

$$\theta_1 = w_c t + k\int f(t+t_1)dt + \phi_1$$

and

$$\theta_2 = w_c t + k\int f(t+t_2)dt + \phi_2$$

Hence

$$\dot{\theta}_1 = w_c + k\, f(t+t_1)$$

and

$$\dot{\theta}_2 = w_c + kf(t+t_2)$$

Substituting for $\dot{\theta}_1, \dot{\theta}_2, \theta_1,$ and θ_2 leads clearly to

$$\frac{Ad\theta}{dt} = f(t+\tau_1)\, F_1\left[f(t+\tau_1),\ f(t+\tau_2)\right] \, f(t+\tau_2)F_2\left[f(t+\tau_1),\ f(t+\tau_1)\right]$$

Comparing this to equation (3.3.2) shows that FM is a non-linear process and so not amenable to equalisation. The removal of intersymbol interference from $f(t + t_2)$ (say) would be precluded because it is modulated after demodulation by a complex function of $f(t + t_1)$ and $f(t + t_2)$. This cross-modulation would be even more complicated if three or more multipaths were present.

Since FSK cannot be equalised it will be at a severe disadvantage compared to the ASK and PSK modes in the presence of multipath. A theoretical investigation is now made to see how the performance in noise of FSK compares to ASK and PSK bearing in mind the bandwidth restrictions imposed on the HF channel.

In FSK, the two binary symbols are represented by two waveforms:

$$S_1(t) = A \cos w_1 t \qquad 0 < t < T$$
$$= 0 \qquad \text{otherwise}$$

$$S_2(t) = A \cos w_2 t \qquad 0 < t < T$$
$$= 0 \qquad \text{otherwise}$$

T is the bit period.

An FSK bit stream then is equivalent to an FM carrier where the modulating signal, r(t) say, is a square wave. The FSK modulated carrier S(t) can thus be represented by

$$S(t) = A \cos \left([w_c + w_1 r(t)] t\right)$$

where $r(t) = +1$ for one symbol and -1 for the other symbol. w_1 is the frequency deviation and is given by $w_1 - w_c (= w_c - w_2)$. w_c is the carrier frequency

In order for us to determine the noise performance of FSK, we first of all have to deal with a general FM signal, R(t), say. Let R(t) be written as:

$$R(t) = A \cos \left[w_c t + \theta(t)\right]$$

The term "Frequency Modulation" implies that the instantaneous frequency of R(t), w_i say, is given by

$$w_i = w_c + kf(t)$$

where f(t) is the modulating signal and k is a constant.

Now the instantaneous frequency of R(t) is the differential of the phase, i.e.:

$$w_i = \frac{d}{dt}\left[w_c t + \theta(t)\right]$$

$$= w_c + \frac{d\theta(t)}{dt}$$

Hence by comparing the equations for w_i we have:

$$kf(t) \equiv \frac{d\theta(t)}{dt}$$

$$\therefore \quad \theta(t) = k\!\int\! f(t)dt$$

Hence the expression for a basic FM signal is:

$$R(t) = \cos\left[w_c t + k\!\int\! f(t) \quad dt\right] \tag{3.3.3}$$

The expression for the signal-to-noise ratio of the demodulated signal, bandpass noise having been added to the modulated signal, will now be derived.

Fig. 3-3.7 is a block diagram of the FM demodulator. The bandpass filter has a bandwidth $2w_B$ equal to the bandwidth of the FM signal. The input to this filter is

$$A \cos\left[w_c t + k\!\int\! f(t)dt\right] + n(t)$$

the noise n(t) being "white" and "Gaussian".

The output B(t) of the bandpass filter is:

$$B(t) = A \cos[w_c t + k\int f(t)dt] + n_c(t) \cos w_c t + n_s(t) \sin w_c t$$

$$(3.3.4)$$

where n_c and n_s are independent low-pass Gaussian waveforms.

The action of the discriminator is to produce an output proportional to the differential of the phase of the input signal, i.e.

if the input is $\cos\theta(t)$

the output is $C \dfrac{d\theta(t)}{dt}$

where C is a constant.

From eqn. (3.3.3) the discriminator output for R(t) would be:

$$C \frac{d}{dt} [w_c t + k\int f(t)dt]$$

$$= Cw_c + Ckf(t) \qquad\qquad (3.3.5)$$

This expression contains a constant term Cw_c plus the baseband signal f(t) multiplied by a constant. Thus the discriminator has recovered the modulating waveform.

We have established the signal component in the discriminator output in eqn. (3.3.5). We now find the component containing noise when the input is a signal such as that expressed in eqn. (3.3.4). Let the modulating signal f(t) be zero. Then eqn. (3.3.4) becomes:

$$B(t) = A \cos w_c t + n_c(t) \cos w_c t + n_s(t) \sin w_c t$$

$$\equiv X(t) \cos[w_c t + Z(t)]$$

where

$$X(t) = [(A+n_c)^2 + n_s^2]^{\frac{1}{2}}$$

and

$$\tan Z(t) = \frac{n_s(t)}{n_c(t) + A}$$

The output of the discriminator is the instantaneous frequency:

$$w_1 = C \frac{d}{dt}[w_c t + Z(t)]$$

$$= Cw_c + C\dot{Z}(t) \qquad\qquad (3.3.6)$$

Clearly the noise component is $C\dot{Z}(t)$. This is solved (as before) by differentiating $\tan Z(t)$:

$$\frac{d}{dt}[\tan Z(t)] = \frac{d}{dt}[n_s(t)/(n_c(t) + A)]$$

$$\therefore \quad \dot{Z}(t)[1 + \tan^2 Z(t)] = \frac{\dot{n}_s(t)}{n_c(t) + A} - \frac{n_s(t)\dot{n}_c(t)}{[n_c(t) + A]^2}$$

Substituting for $\tan Z(t)$ and simplifying leads to:

$$\dot{Z}(t) = \frac{\dot{n}_s(t)[n_c(t) + A] - n_s(t)\,\dot{n}_c(t)}{[A + n_c(t)]^2 + n_s^2(t)}$$

Because of the unwieldy nature of this expression the usual approximation[36] will have to be made that A is much greater than $n_c(t)$, $n_s(t)$ and their derivatives. The expression for Z(t) now reduces to:

$$\dot{Z}(t) = \dot{n}_s(t)/A$$

In order to obtain an expression for the signal-to-noise ratio at the output of the FM demodulator we have to know the noise power of $\dot{Z}(t)$.

Differentiation corresponds to passing the signal through a filter having a transfer function jw. The spectral power density response of such a filter is:

$$G_0(w) = |jw|^2 \, G_i(w)$$

where $G_0(w)$ and $G_i(w)$ are the output and input spectral densities.

Hence the spectral power density of $\dot{Z}(t)$,$G_z(w)$, is given by:

$$G_z(w) = \frac{w^2}{A} \, G_s(w)$$

where $G_s(w)$ is the spectral density of $n_s(t)$. This is flat and low-pass. Let its density be N. Hence:

$$G_z(w) = \frac{w^2}{A^2} \, N$$

Referring to Fig. 3-3.7, the discriminator is followed by a low-pass filter. Its cut-off frequency w_m is equal to the maximum frequency component in the modulating signal f(t). w_m is less than w_B (the equivalent cut-off frequency of the bandpass filter before the discriminator) since, as will be shown, frequency modulation

increases bandwidth.

$G_z(w)$ is drawn in Fig. 3-3.8. The noise power N_o at the low-pass filter output is the shaded area. Its value is:

$$N_o = C \int_{-w_m}^{w_m} G_z(w) \; dw$$

$$= C \int_{-w_m}^{w_m} \frac{w^2}{A^2} N \; dw$$

$$\therefore N_o = \frac{2}{3} \frac{CN}{A^2} w_m^3 \tag{3.3.7}$$

Having found the noise level we wish to define a signal level for the digital case in order to derive a signal-to-noise ratio expression. In the digital case the "signal" is the eye-opening value E.

The modulating signal (f) in FSK is a pulse stream having a value of, say

$$w_1/k \text{ for the binary "1"}$$

$$-w_1/k \text{ for the binary "0"}$$

at the decision making instant . From eqn. (3.3.5) the eye-opening E at the discriminator output will be

$$E = Ck\left[\frac{w_1}{k} - \frac{-w_1}{k}\right]$$

i.e.

$$E = 2Cw_1$$

Let us define the signal-to-noise power in the digital case as

$$E^2/N_0$$

Hence the signal-to-noise ratio S_{FSK} for FSK is given by

$$S_{FSK} = E^2/N_0$$

$$= \frac{(2Cw_1)^2}{2C^2 N w_m^3/3A^2}$$

$$\therefore \quad S_{FSK} = \frac{6A^2 w_1^2}{N w_m^3} \tag{3.3.8}$$

An equivalent expression will now be derived for ASK in order to compare their noise performances.

As has been argued the limitation of the maximum possible signal in a normal transmitter power amplifier is that of peak power. Thus, in order to have the equivalent signal power of the described FSK system, the ASK modulated signal $f(t)$ must be, at the decision making instant,

$$f(t) = A \cos w_c t \qquad \text{for a "1"}$$

and

$$= 0 \qquad \text{for a "0"}$$

An ASK demodulator system is shown in Fig. 3-3.9. The bandpass filter now has a bandwidth equal to that of the modulating signal. The output of it is:

$$f(t) \cos w_c t + n_c(t) \cos w_c t + n_s(t) \sin w_c t$$

The coherent detector is equivalent in operation to multiplying the above expression by $\cos w_c t$ and then low-pass filtering to remove the $\cos 2w_c t$ and $\sin 2w_c t$ components. Hence the demodulated signal is:

$$f(t) + n_c(t)$$

The signal power is again taken as the eye-opening, E, squared, i.e. A^2.

The spectral density of the noise must be the same as for the FSK case, i.e. N. The noise power in $n_c(t)$ is therefore its spectral density, which is N times the total bandwidth $2w_m$.

Thus the ASK signal-to-noise ratio S_{ASK} is:

$$S_{ASK} = A^2/2w_m N$$

The comparative signal-to-noise ratios of the FSK and ASK cases are found by combining this equation with eqn. (3.3.8), thus:

$$\frac{S_{FSK}}{S_{ASK}} = \frac{6A^2 w_i^2 / N w_m^3}{A^2/2w_m N}$$

$$\therefore \quad \frac{S_{FSK}}{S_{ASK}} = \frac{12 w_1^2}{w_m^2} \tag{3.3.9}$$

Eqn. (3.3.9) shows that the higher the frequency deviation w_1 the greater the improvement in the signal-to-noise ratio. However, as will be shown shortly, increasing the frequency deviation increases the bandwidth of the FM signal. A channel such as HF radio is bandlimited and so the value of the modulating signal bandwidth w_m would

have to be reduced, leading to a lower bit rate being possible than for ASK. The next step in the analysis therefore is to establish an expression for the bandwidth of an FM signal as a function of the frequency deviation (w_1) and the modulating-signal bandwidth (w_m).

It is most easily obtained when the modulating signal is a sine wave. Taub and Schilling[37] take the modulating signal $v(t)$ as:

$$v(t) = A \cos\left(w_c t + k\int B \cos w_m t \; dt\right)$$

where $B \cos w_m t$ is the modulating signal, w_c is the carrier frequency and A and k are constants.

They write it as:

$$v(t) = A \cos(w_c t + \frac{w_1}{w_m} \sin w_m t)$$

w_1 being the frequency deviation. They now express it as an infinite series of Bessel functions $J_n(C)$, where:

$$C = \frac{w_1}{w_m}$$

C is term the modulation index. The series for $v(t)$ is then:

$$v(t) = J_0(C) \cos w_c t - J_1(C)\left[\cos(w_i - w_m) t - \cos(w_i + w_m)t\right]$$

$$+ J_2(C)\left[\cos(w_i - 2w_m)t + \cos(w_i + 2w_m)t\right]$$

$$+ J_3(C)\left[\cos(w_i - 3w_m)t - \cos(w_i + 3w_m)t\right]$$

$$+ \text{etc.}$$

This expression indicates that there are an infinite number of sidebands spaced on either side of the carrier at separations of w_m, $2w_m$, $3w_m$, etc. This implies that even if the modulating signal is bandlimited the FM signal has infinite bandwidth. Taub and Schilling have made a definition of bandwidth for FM as being that which contains 98% of the signal power (this is for a sine-wave modulating signal). By examining the table of Bessel functions they have shown that the equation for the FM bandwidth w_B is:

$$w_B = 2(w_1 + w_m) \qquad (3.3.10)$$

(This is known as Carson's Rule.)

Lathi[38] uses a heuristic argument to estimate the FM bandwidth for a general modulating signal having a bandlimited nature. By the sampling theorem a bandlimited signal can be represented by a square-topped sampling function. If $f(t)$ is bandlimited to f_m the Nyquist sampling period is $1/2\,f_m$. The spectrum of the FM carrier for this "staircase" sampled waveform consists of pulses having (sin x)/x shapes. These are centred on instantaneous frequencies corresponding to the pulse heights. The maximum instantaneous frequency is

$$w_c + k|f(t)|_{max}$$

$$= w_c + w_1$$

This spectrum is shown in Fig. 3-3.10.

By making the approximation that the energy of the (sin x)/x function is concentrated in the main lobe the entire FM signal spectrum

can be said to lie in the range

$$w_c - (w_1 + 2w_m) \text{ to } w_c + (w_1 + 2w_m)$$

where

$$w_m = 2\pi f_m.$$

The angular-frequency bandwidth w_B is thus

$$w_B = 2(w_1 + 2w_m) \tag{3.3.11}$$

or

$$w_B = 2(C + 2) w_m$$

However this equation does not hold for the case known as Narrowband FM. Here

$$w_1 << w_m$$

and so in the general expression for an FM signal of

$$\cos(w_c t + k\int f(t)dt)$$

we have

$$k\int f(t)dt << 1$$

for all t. Hence the Narrowband FM signal can be written as

$$\cos w_c t + k\int f(t)dt \sin w_c t$$

This is equvalent to a carrier and sidebands; in other words, an AM signal. If the spectrum of $f(t)$ is $F(w)$, the spectrum of

$$\int f(t)dt \quad \text{is} \quad F(w)/jw$$

Thus, if $f(t)$ is bandlimited to w_m, the FM signal bandwidth in the Narrowband case is $2w_m$. However eqn. (3.3.11) gives the result for the Narrowband case ($w_1 << w_m$) as $4w_m$. Hence this equation gives an exaggerated estimate of FM signal bandwidth at low modulation indices ($w_1 << w_m$ is equivalent to there being a low modulation index C because $C = w_1/w_m$).

In our particular case of comparing FSK with ASK we require from eqn. (3.3.9) that

$$\frac{12\ w_1^2}{w_m^2} > 1$$

i.e.

$$w_1 > 0.3_{approx} w_m$$

in order for FSK to be an improvement over ASK in the presence of noise. Let us make

$$w_1 = 0.5\ w_m$$

in order for FSK to be a significant improvement (5 dB). Then let us see what the FM signal bandwidth is likely to be.

Since w_1 is approximately equal to w_m the Narrowband FM approximation cannot be made. Let us therefore combined eqns. (3.3.10) and (3.3.11) to say that the FM signal bandwidth is

$$w_B = 2(w_1 + 1.5\ w_m)$$

Hence if $\quad W_1 = 0.5\ W_m$

$$W_B = 2(0.5\ W_m + 1.5\ W_m)$$

$$\therefore\ W_B = 4\ W_m$$

Now in the ASK case the bandwidth of the modulated signal is simply twice the bandlimit frequency of the modulating signal, i.e. $2\ W_m$.

Hence the FSK modulating signal has to have one-half of the bandwidth of the equivalent ASK modulating signal in order to give an improvement in signal-to-noise ratio of 5dB for the same modulated signal bandwidth*. In other words the channel capacity is reduced by 50% which is an unacceptable reduction. Furthermore the FSK data pulse would have to be twice as wide as that for ASK resulting in increased susceptability to intersymbol interference in multipath conditions.

The conclusion is that, for this investigation of the error performance at high data rates over HF radio using serial transmission, FSK is of no use. It cannot be equalised because of its non-linear nature, and its bandwidth requirement for significant noise-performance improvement over ASK (and even more so for PSK) is too great.

Having discussed the methods of modulation used in the test system we now examine the remaining parts of the transmitter hardware.

---

*In practice, because of the "threshold" effect in FM, at low signal-to-noise ratios the frequency deviation would have to be made somewhat greater. This would necessitate an even bigger decrease in the modulating-signal bandwidth.

3.4   THE PRE-EMPHASIS LOW-PASS FILTER

The primary use of this filter in the test system was to remove the pulse spectra centred on harmonics of the sampling frequency. As was stated in section 3.2.1 on the pulse shaper, a second-order filter ~~centred~~ with cut-off on the bandlimit frequency was found to be sufficient.

A secondary function of this filter was to improve the signal-to-noise ratio at the receiver by using its pass-band section to pre-emphasise the signal spectrum. This method is described in section 3.8.

3.5   THF MODULATOR AND AUTOMATIC GAIN CONTRO! (AGC)

The modulator was a straightforward double-sideband system (DSB). Hence the baseband signal (the waveforms in Figs. 3-3.3 and 3-3.4) was simply multiplied by the carrier cos $w_c t$. Single-sideband modulation was not considered to be feasible to implement in these experiments because of its complexity. Also there would have been the problem of carrier extraction at the receiver for demodulation purposes. DSB permits the use of a phase-locked loop.

The purpose of the automatic gain control was outlined in section 3.3. It was to compensate for the different modulation modes and bit rates*. Its output was such that the peak-modulated signal was always fixed to allow the transmitter power amplifier to be operating at just below saturation.

This concludes the discussion of the transmitter hardware. The radio link is now described.

---

*The AGC was connected after the modulator, and so also compensated for variations in the modulator output for different carrier frequencies.

3.6 THE RADIO LINK

The transmitter was situated at Havant, in Hampshire, and the receiver was situated at Wick, in the north of Scotland. The path length was approximately 1000 Km. The transmitter used a power amplifier whose peak-power capability was 2 KW (i.e. 1 KW average power for an unmodulated carrier). The transmitter aerial was of the horizontal log-periodic type, details of which are given in Appendix 3.2.

The receiving aerial was a vertically-mounted half-wave dipole tuned to 5 MHz. (The carrier frequencies used in the tests varied from 4 to 7 MHz.) The communications receiver was of standard type (being a Racal RA17). The receiving station was situated in an unpopulated area, so the level of man-made noise picked up from nearby surroundings was expected to be relatively small.

The receiver hardware is now discussed, starting with the demodulator system.

3.7   DEMODULATION SYSTEM

Here we see the methods of demodulation analysed. In
section 3.7.1 we deal with coherent demodulation of the three ASK
modes and the one PSK mode. Part (a) discusses the respective
performances of ASK and PSK assuming perfect carrier extraction. In
part (b) the structure of the carrier extractor used in the test
system, a phase-locked loop, is described. Its performance in noise
is given in order to see how closely it approaches the perfect carrier
extractor.

In section 3.7.2 we see that the ASK mode was also detected
incoherently, by full-wave rectification. We see that theoretically
rectifier detection leads to inferior performance in the presence of
noise as compared to coherent detection. We find that furthermore
it is a non-linear process and so see that a pulse stream demodulated by
this method cannot have any intersymbol interference present removed
by an equaliser. However, because it is an asynchronous system its
use was incorporated into the test system. It was felt that it would
provide a more reliable means of detection than a PLL and so be a means
of evaluating the PLL performance in the field trials.

The theoretical error performances of ASK and PSK coherently-
demodulated, and ASK rectifier-demodulated, are shown in Fig. 3-7.1.

3.7.1   Synchronous Detection

This section is split into two parts:

(a) The theoretical error performance for ASK and
PSK in the presence of Gaussian noise is
evaluated. Perfect carrier extraction is
assumed.

(b) The phase-locked loop used for carrier extraction
is analysed.

(a) Error performance of PSK and ASK assuming perfect carrier extraction

The input to the demodulator is modelled as in Fig. 3-3.9. The channel is assumed to add Gaussian noise $n(t)$ to the transmitted signal $f(t) \cos w_c t$, $f(t)$ being the pulse stream. (In practice the channel noise is not Gaussian (see chapter two), being composed mainly of interference. However, our assumption of the noise as being Gaussian simplifies this a priori evaluation of the demodulation techniques.) The bandpass-filter bandwidth is small compared to its centre frequency. Thus the noise at the filter output can be represented in the narrowband form, as shown in Schwartz[36], by

$$ n_c(t) \cos w_c t + n_s(t) \sin w_c t $$

where $n_c(t)$ and $n_s(t)$ are independent low-pass Gaussian processes each having the same bandwidth as the bandpass filter.

The total waveform at the filter output is:

$$ f(t) \cos w_c t + n_c(t) \cos w_c t + n_s(t) \sin w_c t $$

Perfect synchronous detection involves multiplying the above signal by $\cos w_c t$, and then low-pass filtering to remove components at the $2w_c$ frequency. The resultant signal is:

$$ f(t) + n_c(t) $$

We now give the error probability for the ASK mode (mode (a) in section 3.3.1) and the PSK mode.

ASK

Here f(t) has an amplitude at the sampling instant of A
(say) for the binary "1" and zero for the binary "0".

The error probability is determined as follows. The
probability distribution at the decision-making instant is that
of the level (at this moment) of f(t) added to the Gaussian
probability distribution of $n_c(t)$[37]. The optimum decision level
must be determined. Since the distribution shape is independent of
f(t) this level is a fixed fraction of the pulse amplitude which, by
considerations of symmetry, is one-half (see section 3.12). The
result for the error probability, P_e, is given in Taub and Schilling[37]
as:

$$P_e = \frac{1}{2} \text{ erfc } \frac{A}{2\sqrt{(2N)}}$$

where A is the pulse height, or eye opening in this case, N is the
noise power of $n_c(t)$, and erfc x is the complimentary error function.
Defining a signal-to-noise ratio S as being:

$$S = A^2/N$$

we obtain:

$$P_e = \frac{1}{2} \text{ erfc } \frac{\sqrt{S}}{2\sqrt{2}} \qquad\qquad (3.7.1)$$

This function is plotted in Fig. 3-7.1.

PSK

      As argued in section 3.3 all modulation modes must have the same peak modulated signal. With the bipolar pulse scheme used to implement PSK a "1" is therefore a positive pulse of height A and a "0" is a negative pulse having the same value (see Figs. 3-3.3 and 3-3.4). Thus the eye opening is 2A, and so the error probability is:

$$P_e = \frac{1}{2} \text{ erfc } \frac{2A}{2\sqrt{(2N)}}$$

Hence:

$$P_e = \frac{1}{2} \text{ erfc } \frac{\sqrt{(4S)}}{2\sqrt{2}} \qquad (3.7.2)$$

      Equations (3.7.1) and (3.7.2) show that PSK gives a four-fold (i.e. 6 dB) improvement in signal-to-noise ratio over ASK. Thus, in a plot of error probability against signal-to-noise ratio, the PSK curve is identical in shape to that of ASK but shifted to the origin by 6dB. Both error probability curves are plotted in Fig. 3-7.1.

      We now see the phase-locked loop carrier extractor described, and find how closely it resembles synchronous detection.

(b) The phase-locked loop carrier extractor

      The error probability curves evaluated for PSK and ASK assumed perfect carrier extraction. In practice the reference-carrier output from the PLL is of the form:

$$\cos[w_c t + \theta(t)]$$

where $\theta(t)$ is the phase error. It will be a noisy waveform if the input signal has additive noise. In this section we determine the

effect of $\theta(t)$ in worsening the error probability in additive noise.

First of all we attempt to derive theoretically the dependence of $\theta(t)$ on the value of the additive noise. We assume this noise is narrowband.

The realisation of the phase-locked loop is shown in block-diagram form in Fig. 3-7.2. The positive zero crossing of the carrier triggers a monostable via a hard limiter. Thus a constant-width pulse is generated in each cycle. The output of the monostable is fed into the phase comparator. Its equivalent operation is shown in Fig. 3-7.3. (a) and (b) show the out-of-lock positions. In (a) the mean value of I_{pc}, the current output of the phase comparator, is positive. This is fed into the loop filter shown in Fig. 3-7.4. The resultant increase in the voltage-controlled oscillator (VCO) input voltage makes its output frequency higher. Thus the VCO output waveform X_{VCO} is shifted to the left. In the opposite out-of-lock position (b) the converse occurs.

The effect of the loop filter capacitor C is to slowly integrate the effects of the input current I_{pc}. Its value can be assumed to be large so that the I_{pc} variations are conveyed via the loop filter resistor, R. Thus in the in-lock position shown in Fig. 3-7.3(c) the capacitor can be assumed to have a constant voltage throughout each cycle, of such a value that the VCO frequency is the same as the carrier frequency. The capacitor has charged or discharged so as to make the average value of I_{pc} zero. Thus the pulse X_m exactly straddles the negative-going edge of X_{VCO}. Clearly there is a constant phase difference between the VCO and carrier dependent on the monostable pulse width. It is offset in practice by pre-delaying the VCO waveform.

A significant advantage of this type of phase comparator occurs when the pulse width τ is made much less than the carrier period, T. Then I_{pc} is zero for most of the time, and so the loop filter is in effect normally disconnected from the phase comparator. Thus noise feedthrough is much reduced and so the internal noise in the PLL is almost solely produced by the VCO, unlike PLL's of the Costas type[40]. Also there is no path for the loop-filter capacitor to discharge through unless I_{pc} is non-zero. This is because its output is fed to the VCO by a high input impedance buffer. Hence the loop filter becomes an almost perfect integrator if the pulse width is made small (i.e. a perfect second-order system in PLL terminology).

The effect on the VCO waveform phase of carrier phase variations caused by additive noise is now analysed. It is assumed that the frequency content of these variations is small compared to the carrier frequency. Further it is assumed that the loop capacitor is infinitely large and is charged such that the VCO frequency in the absence of any output from the phase comparator is exactly that of the carrier frequency.

Let the VCO have the linear frequency-to-input voltage relationship illustrated in Fig. 3-7.5 of

$$f = k(V + V_0) \qquad\qquad (3.7.3)$$

where f is the VCO output frequency, V is the input frequency and k and V_0 are constants ($V_0 = f_0/k$ in Fig. 3-7.5).

If V varies during a cycle the period T of this altered cycle may be found by writing

$$\frac{1}{T} = k(V + V_0)$$

i.e. in general

$$\int_{t_1}^{t_2} k(V + V_0)dt = 1 \qquad (3.7.4)$$

where $T = t_2 - t_1$.

We will now determine the period of the waveform X_{VCO} represented in Fig. 3-7.6. The cycle period T may he defined as the interval between two repeatable parts of the oscillating waveform. For the waveform in Fig. 3-7.6 wc will use successive negative-going edges.

We now apply eqn. (3.7.4) to the VCO input waveform shown in Fig. 3-7.6 consisting of a step of height v on a steady level of amplitude V_c. We have then

$$k(V_c + v + V_0)\, \tau + k(V_c + V_0)(T - \tau) = 1$$

during pulse   not during pulse

with $T = t_2 - t_1$

i.e. $k(V_c + V_0)T = 1 - kv\tau$

Hence:

$$T = \frac{1}{k(V_c+V_0)} - \frac{v\tau}{V_c+V_0}$$

For $t > t_2$, the VCO input voltage remains constant and so $V = V_c$. The effect of the pulse is therefore to phase shift the VCO waveform because the period has been shortened only between $t_2 < t < t_1$. The value of this phase shift is found thus:

The period T_0 for the constant VCO input voltage V_c is, from eqn. (3.7.3), given by:

$$T_0 = \frac{1}{k(V_c+V_0)}$$

The phase shift, $\Delta\theta$, is given by

$$\frac{T_0-T}{T_0}$$

(n.b. the "units" of the expression are such that $\Delta\theta = 1$ is one full cycle).

Substituting for T_0 and T gives:

$$\Delta\theta = \frac{1/k(V_c+V_0) - [1/k(V_c+V_0) - v\tau/(V_c+V_0)]}{1/k(V_c + V_0)}$$

Hence:

$$\Delta\theta = kv\tau$$

The value computed above of the phase shift $\Delta\theta$ is that caused when the monostable pulse does not traverse either the positive- or the negative-going edges of the VCO waveform. Fig. 3-7.7 shows the phase shift for the complete range of positions of the monostable pulse. (The phase θ_i of the centre of the pulse is with respect to

the negative-going edge of the VCO waveform). As can be seen
the phase shift response is in two sections:

(1)   If $\theta_i(t) - \theta_o(t) > \tau/2T$.

then $\Delta\theta = kv\tau$

(2)   If $[\theta_i(t) - \theta_o(t)] < \tau/2T$

then $\Delta\theta = kv\tau \dfrac{[\theta_i(t) - \theta_o(t)]}{\tau/2T}$ (3.7.5)

where $\theta_i(t)$ is the carrier phase (the input) and $\theta_o(t)$ is the VCO phase
(the output).

In order to attempt to derive an equivalent circuit for the
PLL we now evaluate its step response. By step response we mean
that the PLL is in lock in the position shown in Fig. 3-7.3(c), i.e.

$$\theta_c(t) = \theta_i(t) = 0$$

and then $\theta_i(t)$ is shifted by an amount θ_s at $t = 0$, say. The step
response is the resultant function of $\theta_i(t)$ for $t > 0$.

The output phase response is a series of "steps" of height
$\Delta\theta$, the value of $\Delta\theta$ being according to Fig. 3-7.7. One step occurs
per cycle of the carrier. The response is most easily expressed in
a smooth curve so let us assume that the carrier cycle period T and
the height of the maximum phase step ($kv\tau$) are both small. (This
will be true in practice.) The step response is then as shown in
Fig. 3-7.8. Again it is in two regions:

(a) $\theta_0(t) < (\theta_s - \tau/2T)$

Then the phase shift response is in section (1) of Fig. 3-7.7 (see above). In time Δt there are $\Delta t/T$ cycles of the carrier. The phase is shifted by $kv\tau$ increments per cycle in this region. Hence the phase shift in time Δt is:

$$kv\tau \cdot \frac{\Delta t}{T}$$

Thus, because T is assumed small:

$$\theta_0(t) = \int_0^t \frac{kv\tau}{T} dt$$

$$\therefore \quad \theta_0(t) = \frac{kv\tau t}{T} \tag{3.7.6}$$

In this region then the phase response can be said to exhibit a fixed slew rate, of value $kv\tau/T$.

(b) $\theta_0(t) > (\theta_s - \tau/2T)$

The time t_0 when $\theta(t)$ equals $(\theta_s - \tau/2T)$ is found from eqn. (3.7.6):

$$\theta_s - \frac{\tau}{2T} = \frac{kv\tau t_0}{T}$$

$$\therefore \quad t_0 = \frac{(\theta_s - \tau/2T)}{kv\tau}$$

In this region ($t > t_0$) the phase step $\Delta\theta$ per cycle becomes a function of the carrier $[\theta_i(t)]$ to VCO $[\theta_0(t)]$ phase difference, the function having been derived in eqn. (3-7.5). From

this we have:

$$\Delta\theta = \frac{kv\tau[\theta_s - \theta_o(t)]}{\tau/2T}$$

Hence in the interval $t \rightarrow t + \Delta t$ the phase shift is:

$$\frac{kv\tau[\theta_s - \theta_o(t)]}{\tau/2T} \frac{\Delta t}{T}$$

Hence:

$$\theta_o(t) = \int_{t_o}^{t} 2kv[\theta_s - \theta_o(t)]dt + (\theta_s - \tau/2T)$$

This is solved by letting $\theta_o(t) = \theta_s - Ae^{Bt}$

Therefore

$$\theta_s - Ae^{Bt} = 2kvA \int_{t_o}^{t} e^{Bt} dt + \theta_s - \tau/2T$$

i.e.

$$\theta_s - Ae^{Bt} = \frac{2kvA}{B} (e^{Bt} - e^{Bt_o}) + \theta_s - \tau/2T$$

Equating the coefficients of e^{Bt} gives:

$$B = -2kv$$

Equating the constants leads to

$$\theta_s = Ae^{Bt_o} + \theta_s - \tau/2T$$

i.e.

$$A = \frac{\tau/2T}{e^{Bt_o}}$$

Hence:

$$\theta_o(t) = \theta_s - \frac{\tau/2T}{e^{-2kvt_o}} e^{-2kvt}$$

i.e.

$$\theta_0(t) = \theta_s - \frac{\tau}{2T} \; e^{-2kv(t-t_0)}$$

Hence overall

$$\theta_0(t) = \frac{2kv\tau}{T}t \quad , \quad \theta_0(t) < \theta_s - \tau/2T$$

$$= \theta_s - \frac{\tau}{2T} \; e^{-2kv(t-t_0)}$$

The equivalent circuit of the PLL is that of the delta-modulation system shown in Fig. 3-7.9. The delta modulator has an imperfect quantiser in order to be equivalent to region (2) of the PLL phase shift response (Fig. 3-7.7). If we compare the step responses of the PLL and this delta modulator we find they are identical.

In order to obtain the noise performance of the PLL theoretically then we have to derive the statistics of $\theta_i(t)$. Then we should be able to use existing theory concerning the delta modulator to establish the relationship between $\theta_0(t)$ and $\theta_i(t)$.

The function $\theta_i(t)$ arises because the input signal to the PLL is the sum of a carrier plus narrowband noise, i.e.:

$$A \cos w_c t + n_c(t) \cos w_c t + n_s(t) \sin w_c t$$

This is equivalent to

$$R(t) \cos[w_c t + \theta_i(t)]$$

where the PLL input signal $\theta_i(t) = \tan^{-1} \dfrac{n_s(t)}{n_c(t) + A}$.

However the statistics of $\theta_i(t)$ resulting from this function of the Gaussian noise waveform $n_c(t)$ and $n_s(t)$ are difficult to calculate because $\theta_i(t)$ is the argument in a trigonometrical expression. It is not possible for it to be outside the range 0 to 2π.

Because of this difficulty in calculating the noise performance an experimental approach was adopted. The basis of this approach is now outlined. As we have seen the situation in which we wish to find the PLL output is that in which its input consists of a carrier plus narrowband noise, i.e.

$$A \cos w_c t + n_c(t) \cos w_c t + n_s(t) \sin w_c t$$

The PLL output is $\cos^*[w_c t + \theta_0(t)]$, $\theta_0(t)$ being the phase noise caused by the narrowband-noise component present in the PLL input. Demodulation is performed by multiplying these two expressions and then low-pass filtering to remove the $\sin 2w_c t$ and $\cos 2w_c t$ components. The result is (ignoring the x 1/2 overall multiplying factor)

$$A \cos \theta_0(t) + n_c(t) \cos \theta_0(t) + n_s(t) \sin \theta_0(t)$$

Let us consider the $A \cos \theta_0(t)$ term. Since $\theta_0(t)$ is a random variable let us take the rms value $A\sqrt{\overline{\cos^2\theta_0(t)}}$.

With no additive noise

$$\theta_0(t) = 0$$

---

*It is in fact the fundamental of the harmonic series that gives rise to the square-wave output of the PLL.

and hence

$$A\sqrt{\overline{\cos^2\theta_0(t)}} \to A$$

With additive noise present

$$\theta_0(t) \neq 0$$

Therefore:

$$A\sqrt{\overline{\cos^2\theta_0(t)}} < A$$

Hence the factor $\sqrt{\overline{\cos^2\theta_0(t)}}$ is a measure of the "goodness" of the way the PLL extracts the carrier.

The value

$$A - A\cos\theta_0(t)$$

is the reduction in the demodulated-signal amplitude caused by PLL phase fluctuating. In order to determine the "perfectness" of the PLL we wish, therefore, to compare the value

$$A - A\sqrt{\overline{\cos^2\theta_0(t)}}$$

to the noise level present in the signal after demodulation from the $n_c(t)$ and $n_s(t)$ components.

This was determined using the experimental system shown in Fig. 3-7.10. The rms value of $\sin\theta_0(t)$ was measured because this has a mean value of zero (since $\theta_0(t)$ has zero mean). The value $A\sqrt{\overline{\cos^2\theta_0(t)}}$ was found as follows. The meter reading V_0 is given by:

$$V_o = \sqrt{[kA \sin \theta_o(t)]^2}$$

(where k is a constant of the multiplier and low-pass filter).

$$= kA\sqrt{1 - \overline{\cos^2 \theta_o(t)}}$$

$$= kA\sqrt{[1 - \overline{\cos^2 \theta_o(t)}]}$$

$$\therefore A\sqrt{\overline{\cos^2 \theta_o(t)}} = \sqrt{(A^2 - V_o^2/k^2)}$$

For the purposes of expressing the noise performance of the PLL, its noise term

$$A - A\sqrt{\overline{\cos^2 \theta_o(t)}}$$

is given as a ratio to the value of the demodulated signal component A, i.e. in the ratio R_{PLL} where

$$R_{PLL} = \frac{A - A\sqrt{\overline{\cos^2 \theta_o(t)}}}{A}$$

$$= \frac{A - \sqrt{(A^2 - V_o^2/k^2)}}{A}$$

$$= 1 - \sqrt{[1 - (V_o/kA)^2]}$$

Likewise the noise term from the additive noise present in the demodulated waveform,

$$n_c(t) \cos \theta_o(t) + n_s(t) \sin \theta_o(t) ,$$

is expressed as such as ratio. The noise power of the additive noise component is

$$[n_c(t) \cos \theta_0(t) + n_s(t) \sin \theta_0(t)]^2$$

Since $\theta_0(t)$ is slowly varying and uncorrelated to $n_c(t)$ and $n_s(t)$ this expression is equal to

$$[n_c(t) \cos w_c t + n_s(t) \sin w_c t]^2$$

Hence the rms value of the additive noise component is the voltage V_i in Fig. 3-7.10. The desired ratio R_N is:

$$R_N = \frac{V_i}{A}$$

In Fig. 3-7.11 then the PLL-noise term to signal ratio R_{PLL} is plotted against the additive noise to signal ratio R_N. The results show that the noise introduced by the PLL is insignificant and so the PLL used in the test system can be regarded as a satisfactory carrier extractor.

This type of PLL was developed because the proprietary types available as integrated circuits were not found to give adequate performance. Also they were usually stable in only the one phase mode. From Fig. 3-3.5 we see that three of the four ASK and PSK modes produced modulated carriers that have 0 and π phase components; the fourth mode, ASK + C, produced a modulated carrier of single phase.

The configuration of the test-system PLL described so far is stable in the one-phase mode (see Fig. 3-7.3). It was made stable in the C and π phases by doubling the VCO output frequency for the appropriate modulation modes, as in Fig. 3-7.12. If the

phase changes by π the monostable pulses still lock to alternate negative-going edges of the VCO waveform but at a position one cycle along (or back). Since here the VCO output frequency is twice the carrier frequency the X_{VCO} waveform is divided by two before being fed to the demodulator*.

We now discuss rectifier detection, because the ASK mode was demodulated by this process as well as, separately, by the PLL demodulator.

3.7.2   Rectifier Detection of ASK

First of all we see a general analysis mode of the case where an AM signal is detected by rectifier in the presence of narrowband noise.

Let the AM signal be $f(t) \cos w_c t$. The composite waveform at the detector input is

$$f(t) \cos w_c t + x(t) \cos w_c t + y(t) \cos w_c t$$

with $[x(t) \cos w_c t + y(t) \cos w_c t]$ being the usual representation of narrowband noise. The composite waveform can be express as:

$$R(t) \cos [w_c t + \theta(t)]$$

where

$$R(t) = \sqrt{\left\{ ([f(t) + x(t)]^2) + y^2(t) \right\}}$$

and

$$\tan \theta(t) = \frac{y(t)}{f(t) + x(t)}$$

---

*Such a PLL could be used in, say, a four-phase system by using a VCO running at four times the carrier frequency.

This waveform is then passed through a full-wave rectifier. Now, from the theory of narrowband noise (as in [36]), $x(t)$ and $y(t)$ have spectra that are low-pass, their cut-off frequency being usually equal to the bandlimit frequency of $f(t)$ (because of the bandpass filters that are present in a receiver system). This frequency will be much less than the carrier frequency. Hence changes in $R(t)$ and $\theta(t)$ will be small in comparison to the carrier frequency. This allows the rectified waveform of $R(t) \cos [w_c t + \theta(t)]$ to be expressed in the standard Fourier series

$$\frac{2R(t)}{\pi} - \frac{4}{\pi} \sum_{m=1}^{\infty} \frac{(-1)^m}{4m^2-1} \cos 2m \, w_c t$$

If the rectifier output is fed through a low-pass filter to remove the carrier-frequency harmonics we have the result as

$$2 \frac{R(t)}{\pi}$$

Now

$$R(t) = \sqrt{[f(t) + x(t)]^2 + y^2(t)}$$

Hence if $f(t) > 0$, for all t, this demodulation process will recover the original signal undistorted but corrupted by the noise terms $x(t)$ and $y(t)$.

The next stage is to analyse the effect of these noise terms in the expression for $R(t)$ when $f(t)$ is a digital signal. The probability of error will depend on the statistics of $R(t)$. It is a well-known result (e.g. see Schwartz[36]) that the probability density function of $R(t)$ has the Rician distribution:

$$P_f[R] = \frac{R}{N} \exp(-R^2/2N) \exp(-f^2/2N) I_0(Rf/N) \qquad (3.7.7)$$

where $R(t) \rightarrow R$ and $f(t) \rightarrow f$, N is the noise power $\overline{x^2(t)}$ $(= \overline{y^2(t)})$, and $I_0(x)$ is the modified Bessel function.

We are considering ASK. Let $f(t)$ be A at the decision instance for a "1". $f(t)$ is zero for a "0".

Hence the probability distribution of $R(t)$ when a "1" has been transmitted is:

$$P_A[R] = \frac{R}{N} \exp(-R^2/2N) \exp(-A^2/2N) I_0(RA/N) \qquad (3.7.8)$$

When a "0" has been transmitted $f(t) = 0$ and so eqn. (3.7.7) reduces to make the pdf the Rayleigh distribution:

$$P_0[R] = \frac{R}{N} \exp(-R^2/2N) \qquad (3.7.9)$$

These two functions are drawn in Fig. 3-7.13.

In order to calculate the probability of error for ASK it is necessary for us to know the value of the decision level that gives the best error performance. Let the level be x times the pulse height A, i.e. xA.

Then the probability P_0 of an incorrect detection of a "0" is:

$$P_0 = \int_{xA}^{\infty} P_0[R] dR \qquad (3.7.10)$$

The probability of an incorrect detection of a "1" is:

$$P_1 = \int_{0}^{xA} P_A[R] dR \qquad (3.7.11)$$

Assuming that an equal number of "1"s and "0"s are transmitted the overall error probability P_e is given by:

$$P_e = P_0/2 + P_1/2 \qquad (3.7.12)$$

The optimum decision level is that which minimises P_e so it is found by solving:

$$\partial P_e / \partial(xA) = 0$$

This is done by letting

$$P_0 = \int_{xA}^{y} F_0(r)dR \quad , \quad P_1 = \int_{z}^{xA} F_1(R)dR$$

and

$$G_0(R) = \int F_0(R)dR \quad , \quad G_1(R) = \int F_1(R)dR$$

Hence from eqn. (3.7.12)

$$P_e = \frac{1}{2} \left[G_0(y) - G_0(xA) \right] + \frac{1}{2} \left[G_1(xA) - G_1(z) \right]$$

$$\therefore \quad \frac{\partial P_e}{\partial(xA)} = -\frac{1}{2} \frac{\partial \left[G_0(xA) \right]}{\partial(xA)} + \frac{1}{2} \frac{\partial \left[G_1 xA \right]}{\partial(xA)}$$

$$= -\frac{1}{2} F_0(xA) + \frac{1}{2} F_1(xA)$$

Hence the optimum decision level is the solution of

$$F_0(xA) = F_1(xA)$$

i.e. from eqns. (3.7.10) and (3.7.11)

$$P_0 [xA] = P_A [xA]$$

From eqns. (3.7.8) and (3.7.9) we have

$$\frac{xA}{N} \exp(-[xA]^2/2N) = \frac{xA}{N} \exp(-[xA]^2/2N) \exp(-A^2/2N) I_0(xA^2/N)$$

which gives x as the solution of:

$$I_0(xA^2/N) = \exp(A^2/2N) \qquad (3.7.13)$$

This equation has to be solved numerically. (A way of approximating to the optimum threshold for hardware implementation purposes is shown in section 3.12.2.) As can be seen x is dependent on the signal-to-noise ratio A^2/N.

The probability of error can now be found as a function of x and of A^2/N.

From eqns. (3.7.9) and (3.7.10) P_0, the probability of an incorrect detection of a "0", is given by:

$$P_0 = \int_{xA}^{\infty} \frac{R}{N} \exp(-R^2/2N)dR$$

$$= [- \exp(-R^2/2N]_{xA}^{\infty}$$

Hence:

$$P_0 = \exp(-[x/2] [A^2/N])$$

From eqns. (3.7.8) and (3.7.11) P_1 is given by:

$$P_1 = \int_0^{xA} \frac{R}{N} \exp(-R^2/2N) \exp(-A^2/N) I_0(AR/N)dR$$

This can only be solved by means of the Q-function (which is tabulated by Marcum[39]):

$$Q(a,b) = \int_b^\infty x \exp[(-a^2-x^2)/2] I_0(ax)dx$$

Since $Q(a,0) = 1$,

$$\int_0^b x \exp[(-a^2-x^2)/2] I_0(ax)dx = 1 - Q(a,b)$$

Hence by writing P_1 in the form

$$P_1 = \int_0^{x\sqrt{S}} (R/\sqrt{N}) \exp[-S-(R/\sqrt{N})^2/2] I_0[(\sqrt{S})(R/\sqrt{N})] d(R/\sqrt{N})$$

where $S = A^2/N_1$ the signal-to-noise ratio, we have:

$$P_1 = 1 - Q(\sqrt{S}, x\sqrt{S})$$

Hence the overall error probability is, from eqn. (3.7.12), given by:

$$P_e = \frac{1}{2} \exp(-xS/2) + \frac{1}{2}[1 - Q(\sqrt{S}, x S)] \qquad (3.7.14)$$

This is plotted in Fig. 3-7.1, alongside the error-performance graphs for coherently-detected ASK and PSK. From the convergence of the two ASK curves we see that rectifier detection tends to become equivalent to coherent detection at high signal-to-noise ratios.

(This fact is proved by Schwartz, Bennett and Stein[33].) At low signal-to-noise ratios, though, coherent detection is shown in Fig. 3-7.1 to be clearly superior.

Having evaluated the noise performance of rectifier detection we will now see that it is a non-linear process. Let us use the same definition of linearity and notation as in section 3.3.2, and let us use the same two-multipath example. Hence the received signal is:

$$g(t) = A_1 M[f(t+\tau_1), t+\tau_1] + A_2 M[f(t+\tau_1), t+\tau_2]$$

$$= A_1 f(t+\tau_1) \cos(w_c t+\theta_1) + A_2(t+\tau_2) \cos(w_c t+\theta_2)$$

(as in the example for coherently-detected AM given in section 3.3.2).

This can be written in the form

$$g(t) = R(t) \cos(w_c t+\theta_3)$$

The rectifier detector gives an output proportional to R(t). By using the phasor diagram in Fig. 3-3.6 we have

$$R(t) = \sqrt{[f(t+\tau_1)\cos(w_c t+\theta_1)+f(t+\tau_2)\cos(w_c t+\theta_2)]^2 + [f(t+\tau_1)\sin(w_c t+\theta_1)+f(t+\tau_2)\sin(w_c t+\theta_2)]^2}$$

$$= \sqrt{\{f^2(t+\tau_1)+f^2(t+\tau_2)+2f(t+\tau_1)f(t+\tau_2)[\cos(w_c t+\theta_1)\cos(w_c t+\theta_2)+\sin(w_c t+\theta_1)\sin(w_c t+\theta_2)]\}}$$

Even after low-pass filtering the rectifier detector is obviously non-linear because the expression for R(t) is not of the form as in eqn. (3.3.2). Hence a rectifier-detected bit stream cannot be equalised by a conventional form of equaliser.

We have found that rectifier-detection has inferior noise performance as compared with coherent detection, and that it is non-linear and so precludes the use of a conventional equaliser. However, as mentioned in the introduction to this section, rectifier-detection is an asynchronous process and so provides a means of checking the reliability of a PLL detector. Hence in the test system it was used to demodulate the ASK mode (the ASK mode is illustrated in Figs. 3-3.3 and 3-3.4). The ASK mode was also detected separately by the PLL.

We now examine the methods of filtering that can be used to improve the signal-to-noise ratio at the receiver.

3.8   FILTERING TECHNIQUES FOR IMPROVING THE SIGNAL-TO-NOISE RATIO

The standard technique of improving the signal-to-noise ratio at the receiver is the matched filter. This is considered in section 3.8.1. The analysis shows that the improvement for the raised-cosine data pulse is small, and further that the performance of the filter deteriorates sharply in the presence of multipath. Also because of the requirement for it that an equaliser would have to operate across the entire pulse width the matched filter was not used.

The method implemented was that of signal spectrum pre-emphasis, de-emphasis which is described in section 3.8.2. The effectiveness of the technique was established experimentally, and the results showed that it led to some improvement for all the modulation modes.

First of all we discuss the matched filter.

3.8.1   The Matched Filter

The matched filter is evaluated for error performance in the presence of additive Gaussian noise both with and without multipath being present. The performance of the straightforward method of sampling the data pulse at its peak is given for comparison purposes. The pulse shape is assumed to be triangular, a shape which approximates fairly well to the raised-cosine derivatives used in the test system (see section 3.2). Such an approximation eases the algebra. We also assume the transmission method is ASK.

The matched filter is described by Lathi[38]. It is conveniently realised by correlation. If the noise is white the correlating waveform should be the complex conjugate to that of the

transmitted pulse. Thus the correlating waveform is the time inverse of the pulse and is therefore identical if the data pulse is symmetrical about its centre.

Fig. 3-8.1 shows the matched-filter correlator as realised for the triangular-shaped pulse. The integrator output is sampled at $t = T/2$ for decision making (T being the pulse width). The integrator is then reset for the next bit.

In order to compare the error performance of the matched-filter with the pulse-sampling method it is necessary to define a signal-to-noise ratio that relates to the error performance. The probability of error, P_e, in the presence of Gaussian noise is given by Taub and Schilling[37] as:

$$P_e = \frac{1}{2} \text{ erfc } \frac{V_p}{V_N 2\sqrt{2}}$$

where V_p is the ASK pulse height, V_N is the rms noise voltage and erfc x is the complimentary error function. Now the decision level is $V_p/2$. Thus the equation may be written as·

$$P_e = \frac{1}{2} \text{ erfc } \frac{V_p - V_p/2}{V_N \sqrt{2}}$$

$$= \frac{1}{2} \text{ erfc } \frac{V_p - V_t}{V_N \sqrt{2}}$$

where V_t is a general threshold level.

Thus an appropriate signal-to-noise ratio S can be defined as being:

$$S = [(V_p - V_t)/V_N]^2 \qquad (3.8.1)$$

The significance of including the threshold level in this expression for the signal-to-noise will become apparent during the analysis for the case when multipath is present.

The signal-to-noise ratios are now computed for the case when no multipath is present.

Matched filter - no multipath

The output of the matched-filter integrator (Fig. 3-8.1) at $t = T/2$ is the pulse energy E (Lathi[38]). The decision level is $E/2$. We must now find a value for the noise voltage for eqn.(3.8.1).

Lathi shows that the matched filter maximises the square of the signal-to-noise ratio at its output. This is given by

$$\frac{s_0^2(T/2)}{n_0^2(T/2)} = \frac{E}{N}$$

where $S_0(T/2)$ is the matched-filter output at the decision instance $T/2$. $\sqrt{n_0^2(T/2)}$ is the rms noise voltage value V_N that we want, and N is spectral density of the noise. $S_0(T/2)$ is the pulse energy E however. Hence

$$E^2/V_N^2 = E/N$$

$$\therefore \quad V_N = \sqrt{(EN)}$$

The value of E for the triangular pulse is given by

$$E = \int_{-\infty}^{\infty} s^2(t)dt$$

and by inspection of Fig. 3-8.1 this is:

$$E = 2V^2 \int_0^{T/2} (1-2t/T)^2 \, dt$$

$$= 2V^2 [t-2t^2/T + 4t^3/3T^2]_0^{T/2}$$

$$= V^2 T/3$$

For eqn. (3.8.1) we have then:

$$V_p = E, V_t = E/2,$$

and

$$V_N = \sqrt{(EN)}$$

$$\therefore \quad S = |(E-E/2)/\sqrt{(EN)}|^2$$

$$\therefore \quad S = \frac{E}{4N}$$

and

$$E = V^2 T/3$$

$$\therefore \quad S = V^2 T/12 \tag{3.8.2}$$

Pulse sampler - no multipath

The variables needed for eqn. (3.8.1) are $V_p = V$ (the pulse height), $V_t = V/2$, and V_N given by:

$$V_N^2 = 2f_B N$$

where f_B is the bandlimit frequency of the data pulse and N, as

before, is the noise spectral density. The triangular pulse is

used as an approximation in shape for a pulse having a raised-cosine

spectrum. Taking 100% as being the roll-off factor for the spectrum

the bandlimit frequency of a pulse of width T is 2/T (see Figs. 3-2.5

and 3-2.6).

Hence:

$$V_N^2 = 2. \frac{2}{T} . N$$

$$= 4N/T$$

Substituting for V_p, V_t and V_N in eqn. (3.8.1) leads to:

$$S = \frac{(V-V/2)^2}{4N/T}$$

therefore

$$S = V^2T/16N \tag{3.8.3}$$

A comparison of eqns. (3.8.2) and (3.8.3) shows that the matched

filter gives only 1.2dB improvement in the signal-to-noise systems.

The effect of multipath on the two detector systems is

now examined. We will assume the channel has a main path and a single

echo path having a relative delay of $\tau(<T/2)$, the echo path having

a transmission coefficient of -a with respect to the main path.

The situation is shown in Fig. 3-8.2. The results of the effect on

the signal-to-noise ratios are shown in Fig. 3-8.3 for different values

of τ and a. Interference from adjacent bits is ignored. Their effects

are assumed to be contained in the single echo pulse in this

appro ximate analysis. The graphs in Fig. 3-8.3 are plots of

eqn. (3.8.1). The threshold levels V_t and noise voltages V_N are the same as before for the two detector systems, but V_p is now a function of τ and a because of the interference from the echo path. The results display the generally inferior performance of the matched filter in the presence of multipath.

A further drawback of using the matched filter would be the difficulty of incorporating an equaliser. Intersymbol interference would have to be removed from the whole of the width of the pulse in order not to affect the correlator. The equaliser would therefore be larger and more complicated.

For these reasons therefore it was decided that the matched filter was of no use in a medium such as the HF channel where multipath is to be expected. Another approach was adopted and is now described.

3.8.2   Signal Spectrum Pre-emphasis, De-emphasis

A general treatment of this method is made by Schwartz[36]. The block diagram of such a system is shown in Fig. 3-8.4. The pre-emphasis filter, $H_1(f)$, operates on the signal before the noise is added. The de-emphasis filter operates on the signal plus noise.

Since the signal spectrum should be unaltered at the de-emphasis filter output,

$$H_2(f) = 1/H_1(f)$$

The average power at the pre-emphasis filter output is assumed fixed because it is proportional to the transmitted power. Let $G_n(f)$ be the noise spectral density and let $G_s(f)$ be the signal spectral density. Then the optimum network $H_1(f)$ that maximises the

signal-to-noise ratio at the de-emphasis filter output is given by:

$$[H_1(f)]^2 = \sqrt{[G_n(f)/G_s(f)]}$$
(3.8.4)

However, as was argued in section 3.3, a more realistic criteria of the transmitted power to fix is that of the peak power. Since there is no easy relationship between the peak power of a signal and its spectrum (or spectral density) the process of establishing the response of the optimum filter is not straightforward. In practice an experimental approach was undertaken, using eqn. (3.8.4) as a guide. This is now described.

First of all the positioning of the filters in the test system is explained. The filter in block 3-4, Fig. 3-0.1, had as its primary aim to remove spectra associated with harmonics of the sampling frequency of the system used to generate the pulses. Its cut-off frequency could have been positioned anywhere in the blank region between the baseband spectrum and the first harmonic (see Figs. 3-2.8 and 3-2.12). Its response in the pass band was utilised for the pre-emphasis of the pulse spectrum.

In the receiver the de-emphasis filter was positioned on block 3-8, Fig. 3-0.2. Again this filter had a primary purpose as a low-pass filter, to remove the double-carrier-frequency components associated with the demodulator. Its pass band was used for de-emphasis.

As mentioned above, eqn. (3.8.4) was used as a guide for the shape of $H_1(f)$. If we assume that the noise is Gaussian then $G_n(f)$ is constant. Hence from eqn. (3.8.4) we have:

$$[H_1(f)]^2 \propto \frac{1}{S(f)}$$

$S(f)$ is the pulse specturm. $(G_s(f) = |S(f)|^2)$. $H_1(f)$ is drawn in Fig. 3-8.5, together with the raised-cosine shape $S(f)$. We see that $H_1(f)$ should have a response up to the bandlimit frequency f_B as shown. Above f_B $H_1(f)$ should have a low-pass characteristic to remove the extraneous frequencies generated by the pulse shaper.

The de-emphasis filter $H(f)$ should have the inverse shape to $H_1(f)$ and also be low-pass above f_B.

Since $H_1(f)$ and $H_2(f)$ are effectively in cascade it is essential that their composite phase and amplitude responses are reasonably linear in order not to distort the pulse stream. It was decided therefore to make the composite response of the two filters exhibit the Butterworth maximally-flat response. Fig. 3-8.6 shows the characteristics of the individual sections. The sections with "peaks" are $H_1(f)$ since they approximate the shape shown in Fig. 3-8.5. They are second-order filters. The sections without peaks are $H_2(f)$. The cascaded filters have third, fourth, and fifth order Butterworth responses. The cut-off frequency of all the filters is equal to the bandlimit frequency of the signal spectrum. In Fig. 3-8.6 the two types of raised-cosine pulse spectra are also shown.

Experiments were performed using each of the three responses in turn. A fourth filtering system was employed whereby $H_1(f)$ and $H_2(f)$ were flat in the passband. The system used for these experiments is shown in Fig. 3-8.7. The peak signal at the modulator output was held constant for each different filter composition and modulation mode. Three different noise levels, each differing by 4dB, were used in each of the tests. In the results the levels are termed high, medium and low.

The simple way of determining the signal-to-noise ratio
level in digital transmission is by measuring the noise and the
eye-opening at the decision maker. However, because the higher-order filters
introduced some intersymbol interference, the eye-opening was indistinct.
The method employed to obtain the signal-to-noise ratio was to
measure the error rate for each setting. This was then converted to
a signal-to-noise figure using the theoretical error-probability
function.

Figs.3-8.8 and 3-8.9 show the results. The two modes used
were ASK and PSK, at the two highest bit rates for each of the pulse
spectra, viz. the x1 and x1/2 bit-rate reduction factors. For each
setting and for each noise level the signal-to-noise figure given is
relative to the reading from the filter system with no signal spectrum
pre-emphasis and de-emphasis. The vertical axis gives the effect of
each filter in dBs. Positive numbers indicate an improvement.

The critical factor affecting the results when using pre-
emphasis is whether the filter $H_2(f)$ reduces the noise more than the
filter $H_1(f)$ increases the peak signal at $H_1(f)$ output. Here the
peak signal is held constant. Therefore $H_1(f)$ is effectively reducing
the eye opening at $H_2(f)$ output. The peak signal is increased at
$H_1(f)$ output both by its effect on the height of the pulses and by
its effect of increasing the magnitudes of the pulse sidelobes because
of its non-linear amplitude and phase response. Both of these effects
get more pronounced the higher the order of the composite filter.

The graph shows that the 5th order filter combination performs
better than the 3rd or 4th orders. Furthermore, for all bar the 33%
roll-off x1 rate PSK mode, the 5th order filter combination is superior

to the filter combination with the flat passband. As can be seen

the improvement is most marked with the lower bit rate. Here there

is greater separation between bits and hence the peak signal at $H_1(f)$

output is increased less. ASK is better than PSK for the same reason.

For the 5th order filter combination the results for the

100% spectrum roll-off show approximately a 1dB advantage over those

for the 33% roll-off spectrum. Fig. 3-8.6 shows that the pre-emphasis,

de-emphasis filter stages correspond more closely to the 100% roll-off

spectrum. In other words the effect of $H_1(f)$ on the peak signal is

worse forthe 33% roll-off spectrum because of its greater high-frequency

content.

To sum up, then, signal pre-emphasis de-emphasis, when the

composite filter has a 5th order Butterworth characteristic, at worst

degrades the effective signal-to-noise ratio at the decision maker by

0.25dB (33% roll-off factor x1 rate, PSK mode) and at best improves

it by 1.75dB (100% roll-off x1/2 rate,ASK mode) . The system is

therefore beneficial, and this 5th-order filter combination was

implemented into the test system.

3.9   LEVEL DETECTION OF A TIME-VARYING SIGNAL IN THE PRESENCE OF
       ADDITIVE NOISE

First of all the applications in the test system for such a method are described. We then see the method itself discussed.

3.9.1   Applications of the Level Detector

In the test system level detectors were required for two purposes:

(a) For the purpose of setting the decision level for the ASK modes. As will be shown in section 3.12 this level is a function of the pulse height. In a radio channel the pulse height at the receiver is likely to have a substantial dynamic range because of fading. Hence an automatic system was necessary.

It had to be decided where in the pulse stream the pulse amplitude was to be measured. When multipath is present the data pulses are distorted by intersymbol interference. An equaliser can correct this distortion at the pulse centres. However the equaliser implemented in the test system was of the decision-feedback type. If the decision level is in a feedback loop with the equaliser the whole system is likely to latch into a "nonsense" state whenever error bursts occur. For this reason the "set-up" pulse height (see section 3.1.3) was measured because it was, in theory, not distorted by intersymbol interference. Its height was measured at every frame by the level detector so that the decision level could be established.

(b) For the purpose of removing any residual d.c. present in the demodulated bit stream in order to obtain the correct decision level.

Simply measuring the pulse height is adequate for the ASK mode. A "0" has zero amplitude and a "1" has the pulse amplitude. In the ASK + C and ASK-SC modes however, the "0" level is not zero. Let us consider the case of ASK + C (see Figs. 3-3.3 and 3-3.4). The "0" level, which is residual d.c., is 0.2V. The "1" level is V. Now, as will be shown in section 3.12.1, the optimum decision level is exactly half-way between the "0" and "1" amplitudes. Thus the decision level is

$$\frac{V-0.2V}{2} + 0.2V$$

$$= 0.6V$$

Since V is the "1" level, the pulse height, the decision level is proportional to the pulse amplitude.

We have assumed that no multipath is present. Let us now consider what happens to the decision level if we have a main pulse, $A_1 f(t)$, and an echo path, $A_2 f(t + \tau)$. $f(t)$ is the bit stream, τ is the relative delay between the two paths, and A_1 and A_2 are constants resulting from the multipath strengths and the phase difference caused by the time delay, τ. We assume that the time delay τ is such that it does not distort the main pulse at the decision level. The "1" level of the first pulse is thus

$$A_1 V + A_2\, 0.2V$$

The "0" level, the residual d.c., is:

$$A_1 0.2V + A_2 0.2V$$

The decision level is then (midway between these "0" and "1" levels)

$$\frac{(A_1 V + A_2 0.2V) - (A_1 0.2V + A_2 0.2V)}{2} + (A_1 0.2V + A_2 0.2V)$$

$$= A_1 0.6V + A_2 0.2V$$

The presence of the A_2 term shows that the decision level is not proportional to the main pulse in the presence of multipath. A similar result occurs for the ASK-SC case.

The difficulty of establishing the decision level for these two modes can be overcome by measuring the residual d.c. level, the "0" level. The decision level is then the simple function of being half-way between this level and the "1" level measured in (a). In practice the measured "0" level is subtracted from the bit stream which has the same effect. This is because the equaliser used in the test system would have treated any residual d.c. as multipath echoes.

As before the residual d.c. level had to be measured in the test system as a point where no multipath echoes could distort this measurement. Such a point was present just before the set-up pulse.

3.9.2   The Method of Level Detection

The method used to measure the pulse-height and the residual-d.c. levels is now described. The level detector is shown in Fig. 3-9.1. For the particular level to be determined the sample-and-hold is triggered at the instance of that level; in the case of the pulse amplitude this instance is the centre of the set-up pulse, and for the residual d.c. it is just before the set-up

pulse. (Each amplitude has its own level detector.) The output

of the sample-and-hold is fed into a low-pass filter with cut-off

frequency f_c. Now the levels to be measured vary with time due

to fading. The assumption is made that the spectrum of these

variations is bandlimited. Then the sampling theorem can be invoked

to choose the sampling rate, i.e. the repetition rate of transmission

of the set-up pulse, so that the envelope of these levels can be

extracted.

The two variables in this system are:

(1) the sampling rate, f_s

(2) the cut-off frequency, f_c, of the low-pass filter.

It is necessary to optimise these two variables in practice because

the output of the low-pass filter will contain noise components as

additive noise is present in the signal. This disturbs the sample-

and-hold circuit whose perturbations are in effect averaged out by

the low-pass filter.

The weighting of spectral density at the output of the

sample-and-hold circuit for a noisy input signal can be shown to be

inversely proportional to the sampling frequency[43]. Thus it is

advantageous to use a high sampling frequency to reduce the noise at

the sample-and-hold (and hence the filter) output. However this

reduces the bit rate because the necessary increase in the frequency

of transmission of the set-up pulse increases the proportion of

data-off to data-on intervals.

A similar sort of "trade-off" is present in setting the

filter cut-off frequency. The lower the value the less noise is present

at the filter output because the output noise is roughly proportional

to the filter bandwidth (the sample-and-hold output noise spectrum being fairly flat). However in practice the envelope variations of the levels are not bandlimited but are concentrated within a few Hz of zero frequency. Thus lowering the filter bandwidth means that the lower-amplitude higher-frequency variations are not extracted by the level detector.

In practice the values for f_s and f_c were established in preliminary trials. They were set at:

Sampling frequency $\qquad f_s = 25$ Hz

Filter cut-off frequency $\quad f_c = 5$ Hz

The repetition rate of 25 Hz for the set-up pulse transmission frequency meant the bit rate was reduced by 25% (see section 3.1.3).

3.10 BIT SYNCHRONISATION

A significant problem in a data transmission system is to ensure that the incoming pulse stream is sampled at the correct instances for decision making. Sophisticated systems, for example[41], have been developed which operate by iteration, varying the sampling position to maximise the eye opening.

A simpler method is to use a form of phase-locked loop which synchronises to instances of the pulse stream crossing the decision level. Such a scheme is outlined by Bennett and Davey[18]. Problems occur though when intersymbol interference is present. Equalisation only removes distortion at the decision instants , the pulse centres. The pulse edges remain distorted and so the threshold crossings are a function of the intersymbol interference. They are no longer a fixed interval from the pulse centres. Under such circumstances some means of obtaining an interference-free pulse must be found.

As was outlined in the introduction to this chapter the adoption of the frame format of data transmission led to an interference-free pulse being available in the form of the set-up pulse. Since the first multipath to arrive was taken as the signal the bit synchroniser could utilise the positive zero crossing of this first echo of the set-up pulse to lock onto.

The operation of the synchroniser is as follows. It generates a "window" having half the width of a pulse. It is positioned so that it can lock onto the leading edge of the set-up pulse. Fig. 3-10.1 shows the locked state. If the leading-edge transition of the decision level occurs in the first half of this window the bit synchroniser frequency is increased slightly; if the transition occurs in the second half of the window the freqency is decreased

slightly. Thus the window centre is locked to the average

position of the threshold transition. Outside the duration of

the window the feedback mechanism of the bit synchroniser is

disabled so that decision-level transitions by data pulses have

no effect. Hence bit "phase" information is received only at the

reception of each set-up pulse. We see from Fig. 3-10.1 that this

arrangement ensures that multipath smear does not affect synchronisation

provided that the second multipath to arrive is delayed at least one

half of one pulse width after the main pulse.

The PLL action of the synchroniser is identical in operation

to that outlined by Bennett and Davey. Crystal oscillators stable

and accurate with respect to another to better than one part in 10^5

are used at the transmitter to generate the bit-stream clock, and

at the receiver for the synchroniser. No problem with drift from

mismatching can then arise. The phase error mechanism is similar to

the carrier-extractor PLL described in section 3-8; each phase error

shifts the window by a fixed phase amount, in this case by a fixed

fraction of the pulse width. In the simplified diagram of the

synchroniser shown in Fig. 3-10.2, this phase amount is equal to

$1/N_1$. Divider N_2 is set according to the relation

between the pulse width and the frame period.

These were both constant in the test system. Divider N_1 can be varied

though, with the crystal-oscillator frequency being altered in inverse

proportion to maintain constant the frequency at the divider output.

Divider N_1 controls the inertia of the phase-error mechanism.

Perturbations of the threshold transition instance by the set-up

pulse were caused by additive noise in the pulse stream. The higher

the divider value, the higher is the inertia of the error

correction system, and hence the better is the synchroniser's ability

to average out these perturbations. However the inertia cannot be

too high because the synchroniser has to track time shifts caused by

the particular ionsopheric reflecting layer altering in altitude.

A compromise was established by experiment.

All timing in the digital part of the receiver was with

respect to the set-up pulse. For instance the waveform whose

frequency is f_B in Fig. 3-10.2 coincided in its phase with the decision

instances in the data block. It was therefore fed to the decision maker.

Other waveforms were derived in the divider chain N_1 and N_2 to control

the equaliser.

3.11   THE EQUALISER

As was remarked upon in the introduction to this chapter a series of tests was performed without an equaliser. These showed that multipath was consistently present which caused high error rates. It forced the bit rate to be dropped to around 250 b/s to prevent intersymbol interference. It was apparent that an equaliser had to be developed in order for data transmission to be reliable at bit rates of the order of a few kilo-bits per second.

In section 3.11 2 we see the standard types of equaliser reviewed. We see they require complex circuitry and/or a micro-computer to realise them. Such a realisation was not felt to be within the scope of this work and so an alternative system was evolved as described in section 3.11.3. Here we see how (1) the assumption of taking the first multipath to arrive as the signal path, and (2) the adoption of the frame format of data transmission were used to allow the implementation of an equaliser of the decision-feedback type.

First of all we derive the channel model. Throughout we assume that the modulation mode is ASK.

3.11.1   The Channel Model

We start by briefly reviewing the phenomenon of ionospheric propagation. HF radio is transmitted over the horizon by bouncing the signal off the ionosphere. Since the ionosphere consists of a number of layers of differing heights above the earth more than one signal arrives at the receiver. Each signal will have traversed different path lengths and so will have experienced different propagation delays. Hence the received signal for the transmission of

a single pulse is as shown in Fig. 3-11.1. With the centre of
the first pulse to arrive at the receiver taken as the origin the
values t_1, t_2, and t_3 represent the relative delays of the
subsequent pulses. (We have assumed here that the received signal
has been coherently demodulated). Let us see what factors the
magnitudes of the (coherently) demodulated multipath echoes depend on.

If A_0, A_1, A_2, and A_3 are the reflection coefficients of the
four multipaths shown in Fig. 3-11.1 then the received modulated
signal is

$$A_0 f(t)\cos wt + A_1 f(t-t_1)\cos w(t-t_1) + A_2 f(t-t_2)\cos w(t-t_2) + A_3 f(t-t_3)\cos w(t-t_3)$$

where $f(t)$ is the pulse and w is the carrier frequency.

The demodulation process involves multiplying the above
expression by $\cos(wt + \theta_0)$ and low-pass filtering to remove the
$\cos 2wt$ terms. θ_0 is a constant of the carrier extractor which will
tend to be an "average" of the phases of all the paths.

The result of demodulation is therefore proportional to

$$A_0 f(t)\cos\theta_0 + A_1 f(t-t_1)\cos(\theta_0 - wt_1) + A_2 f(t-t_2)\cos(\theta_0 - wt_2) + A_3 f(t-t_3)\cos(\theta_0 - wt_3)$$

Let us write $\theta_1 = \theta_0 - wt_1 - 2\pi n_1$, $\theta_2 = \theta_0 - wt_2 - 2\pi n_2$ and
$\theta_3 = \theta_0 - wt_3 - 2\pi n_3$ where n_1, n_2 and n_3 are integers such that

$$0 < \theta_1 < 2\pi \ , \ 0 < \theta_2 < 2\pi \quad \text{and} \ 0 < \theta_3 < 2\pi$$

Then the demodulated signal $d(t)$ can be written as:

$$d(t) = f(t)A_0 \cos\theta_0 + f(t-t_1)A_1\cos\theta_1 + f(t-t_2) A_2\cos\theta_2 + f(t-t_3) A_3\cos\theta_3$$

$$(3.11.1)$$

This equation shows that magnitudes of each of the pulse echoes after demodulation are dependent on the relative time delays and hence on the phases of each pulse carrier, as well as on each reflection coefficient.

Fig. 3-11.1 can be termed the analogue impulse response of the channel. In a digital system the demodulated waveform is sampled for decision-making purposes at instances separated by the period of the bit rate. Here,therefore, we are interested in the values of the analogue impulse response only at these instances around the signal pulse. Assume in Fig. 3-11.1 that the first path is used as the signal pulse. This pulse will be sampled for decision making at its centre, i.e. when:

$$t = 0$$

The digital impulse response is therefore that shown in Fig. 3-11.2, $s(nT)$ being the impulse response. This is the value of the demodulated waveform at the instance nT where T is the bit-rate period. It is important to note that the ditigal impulse response would be different if T was altered.

Let us now see how intersymbol interference arises. Let another bit be transmitted at time:

$$t = T$$

(relative to the receiver).

The decision maker will sample for this bit at this time; however the value $s(T)$ will also be present from the bit transmitted at time $t = 0$. The total value present at the decision maker is therefore, from eqn. (3.11.1),

the signal pulse   interference from previous pulse

$$f(0) \; A_0 \cos \theta_0 \quad + \quad f(T - \tau_1) \; A_1 \cos \theta_1$$

(The interference adds because coherent demodulation of amplitude modulation is a linear process as has been shown.)

This leads to the simpler method of treating intersymbol interference in a digital system; by using the theory and terminology of digital filtering.

The channel is then represented as the non-recursive digital filter shown in Fig. 3-11.3. The impulse response of this filter is identical to that of the channel response shown in Fig. 3-11.2 if the multiplying constants are (from eqn. (3.11.1))

$$i_0 = f(0) \; A_0 \cos \theta_0 \quad , \quad i_1 = f(T-t_1) \; A_1 \cos \theta_1$$

$$i_2 = f(2T-t_2) \; A_2 \cos \theta_2 \; , \quad i_3 = f(3T-t_3) \; A_2 \cos \theta_3$$

The impulse response of the filter in Fig. 3-11.2 is:

$$I(z) = i_0 + i_1 z^{-1} + i_2 z^{-2} + i_3 z^{-3}$$

using the z-transform approach. z^{-n} implies a delay of nT

The message transmitted, the digital bit stream, can also be represented by a z-transform M(z) as

$$M(z) = m_0 + m_1 \; z^{-1} + m_2 \; z^{-2} + \text{etc.}$$

since each bit is an interval T (i.e. z^{-1}) after the previous one. Each m_n is obviously "1" or "0".

The received signal can be given a z-transform also, of:

$$R(z) = r_0 + r_1 z^{-1} + r_2 z^{-2} + \text{etc}$$

Let us derive the relationship between the received signal R(z) and the transmitted signal M(z). This is obviously going to be a function of I(z).

The transmitted signal can be written as

$$M(t) = \sum_j m_j \, \delta(t-jT)$$

using the delta function for the pulse shape since we are only interested in the value of M(t) at the pulse centre (the decision-making instance).

The channel impulse response and the received signal can be written as, respectively,

$$I(t) = \sum_{k=0}^{P} i_k \, \delta(t-kT) \qquad \text{(n.b. the channel dispersion time}$$
$$\text{extends from } t = 0 \text{ to } t = pT)$$

$$R(t) = \sum_j r_j \, \delta(t-jT)$$

The value of R(t) at the time instance, jT is given by:

$$R(t) = r_j \delta(t - jT) = (m_j \, i_0 + m_{j-1} \, i_1 + m_{j-2} \, i_2 + \text{etc.})$$

i.e.

$$R(t) = \left[\sum_{k=0}^{p} m_{j-k} \, i_k \right] \delta(t-jT) \qquad\qquad (3.11.2)$$

This is the discrete time convolution of the arrays {M}

and [I}. If M and I are expressed by the z-transform the equivalent

operation is:

$$R(z) = M(z) \, I(z)$$

i.e. simple polynomial multiplication. (The action of polynomial

multiplication followed by the grouping the constants under each

power of z^{-1} performs the same operation as eqn.(3.11.2).)

Ideally R(z) should be M(z), the transmitted signal. but

distortion has taken place because of the channel response I(z).

We now analyse the standard equaliser structures that attempt to

remove this type of distortion.

3.11.2   Equaliser Structures

An equaliser is placed after the channel as shown in

Fig. 3-11.4. The equaliser has a z-transform H(z). The received

message is now:

$$R(z) = M(z) \, I(z) \, H(z)$$

If R(z) is to be M(z) then H(z) must be tne reciprocal

of I(z). Such an observation leads directly to one form of equaliser,

viz:

The recursive equaliser

Since I(z) is of the form

$$I(z) = i_0 + i_1 z^{-1} + i_2 z^{-2} \text{ etc}$$

then R(z) is M(z) if H(z) is of the form

$$H(z) = \frac{1}{1 - h_1 z^{-1} - h_2 z^{-2} - \text{etc}}$$

where

$$h_1 = -i_1/i_0 \ , \ h_2 = -i_2/i_0, \ h_3 = -i_3/i_0, \ \text{etc.}$$

(The channel response is assumed to be known. Methods of determining it will be dealt with later.) This structure of $H(z)$ (as shown in Fig. 3-11.5) is a recursive digital filter. The problem with this form of equaliser is that it is unstable if any of the h_ns become greater than unity. Hence a linear recursive filter cannot in genral be used in a time-varying medium such as the HF channel because any of the i_ns can become larger than unity. A method of alleviating this unstable property is to introduce a non-linearity into the system. One such method is the decision-feedback structure. This structure is examined later.

Because the recursive filter can be unstable the next step is to see if a non-recursive structure can form the equaliser.

The non-recursive or transversal equaliser

In this case $H(z)$ has the same form as the channel model, viz:

$$H(z) = h_0 + h_1 z^{-1} + h_2 z^{-2} + \text{etc}$$

in which case there are no instability problems. However we require that for

$$R(z) = M(z)$$

that

$$I(z) \ H(z) = 1$$

However, for the non-recursive form of $H(z)$,

$$I(z) \, H(z) = (i_0 + i_1 z^{-1} + i_2 z^{-2} + \text{etc}) \, (h_0 + h_1 z^{-1} + h_2 z^{-2} + \text{etc})$$

which clearly cannot be unity.

We have therefore to choose the h_ns to minimise some performance criteria such that:

$$I(z) \, H(z) = 1 + \sum_j (0) \, z^{-j}$$

In practice, to allow for a processing delay of z^{-d} say, we wish to make

$$I(z) \, H(z) = z^{-d} + \sum_j (0) \, z^{-1} \quad , \quad j \neq d$$

In order to establish some performance criteria it is necessary to define an error z-transform $E(z)$ that is the difference between the transmitted and received signals. It is given by:

$$E(z) = M(z) \, z^{-d} - R(z)$$

$$= M(z) \, z^{-d} - M(z) \, I(z) \, H(z)$$

$$= M(z) \, [z^{-d} - H(z) \, I(z)] \qquad (3.11.3)$$

$E(z)$ will be of the form:

$$E(z) = e_0 + e_1 z^{-1} + e_2 z^{-2} + \text{etc}$$

One performance criteria that can be used is that of the mean squared error given by:

$$MSE = e_0^2 + e_1^2 + e_2^2 + e_3^2 + \text{etc}$$

Hence in this case the h_ns in H(z) must be chosen so that MSE is minimised. It is done by making M(z) unity. The impulse response of the whole system is then formed. Eqn. (3.11.3) is evaluated and thus the MSE can be computed. The resultant expression for the MSE will be a function of the h_ns and the i_ns in H(z) and I(z). It is partially differentiated with respect to each h_n and set to 0 to minimise the MSE. A set of simultaneous equations will be formed and this enables each h_n to be found.

A second performance criterion that can be used is that of distortion D. This is the sum of the magnitudes of the error sequence E(z), i.e.

$$D = |e_0| + |e_1| + |e_2| + \text{etc}$$

(D is related to the binary eye opening).

Both the two criteria described tend to produce an overall impulse response of the form shown in Fig. 3-11.6(a). The channel for this example has two multipaths. Its impulse response is {I}. A five-tap equaliser which has a response array {H} is used. In this example the h_ns are found by minimising the MSE. The propagation delay through the equaliser is z^{-3}. The resultant response for {R} for the message {M} equal to a single impulse is the convolution of {I} with {H}. The error array {E} shown in the example is found by applying eqn. (3.11.3).

A further performance criteria that can be used is termed "zero forcing". Here the object is to force as many as possible of the e_ns in E to become zero. In the example shown in Fig. 3-11.6(b) a set of solvable simultaneous equations can be found for {H} in which all but the last e_n are forced to zero. (The same type of channel

and equaliser are used as for the MSE example except that z^{-d} is now unity.)

The transversal filter is one of the two most usual types of equaliser. The other is the form of recursive filter referred to earlier, the decision-feedback equaliser.

The decision-feedback equaliser

This is a particular case of introducing a non-linearity into the recursive-filter structure to curb its instability properties. Its form* is shown in Fig. 3-11.7. The non-linear element arises from the inclusion in the feedback parth of the decision maker. Hence the description "decision-feedback".

Its operation is as follows. Let the channel be of the form:

$$I(z) = i_0 + i_1 z^{-1} + i_2 z^{-2} + \text{etc.}$$

The equaliser taps are then such that $h_1 = i_1$, $h_2 = i_2$. $h_3 = i_3$ etc. Then if a decision is correctly made that a pulse has been transmitted (i.e. if at $t = 0$ the decision maker decides that R(z) is i_0) the feedback exactly cancels the effects of the channel impulse response echoes i_1, i_2, i_3 etc.

The problem with the decision-feedback equaliser is when incorrect decisions are made. If,say, a false "1" decision is made the attempt by the equaliser to cancel the effects of i_1, i_2 etc.(by subtracting them) would tend to cause additional errors.

---

*The equaliser shown is for when the channel has only post-echoes. When pre-echoes are present a linear transversal equaliser precedes the decision-feedback equaliser to remove their effects.

Channel identification

The channel impulse response can be measured directly by making the message M(z) a "1" followed by a string of "0"s. This method is not satisfactory though if noise is present. It is better to use a pseudo-random sequence as a test bit stream and use its auto-correlation properties to effectively average out the noise.

This method is illustrated in Fig. 3-11.8. Here m(t) is the transmitted pseudo-random sequence. It is assumed that this sequence is known at the receiver whence it is fed into the correlator via the delay nT. The integrator output is the cross-correlation function

$$R(nT) \;=\; \int r(t) m(t-nT) \; dt$$

where T is the bit-rate period and r(t) is the received signal. But r(t) is the convolution of the channel impulse response i(t) and the test sequence m(t) i.e.

$$r(t) \;=\; \int i(u) m(t-u) du$$

Hence R(nT) is given by:

$$R(nT) \;=\; \int \{\int i(u) m(t-u) du\} m(t-nT) dt$$

$$=\; \int i(u) \; \{\int m(t-u) m(t-nT) dt\} du$$

by changing the order of integration.

Let $v = t - nT$

$$\therefore \quad R(nT) \;=\; \int i(u) \{\int m(v+nT-u) m(v) dv\} \; du$$

The expression inside the curly brackets is the auto-correlation function of the pseudo-random sequence. If random this will be the delta-function $\delta(nT-u)$. Thus:

$$R(nT) = \int i(u)\delta(nT-u)du$$

$$\therefore \quad R(nT) = i(nT)$$

Thus the channel impulse response can be found, a series of integrators and delays being needed to evaluate $i(nT)$ for all n. If the channel is varying in time the data will have to be interrupted periodically to repeat the test sequence. An equaliser system whereby the channel is evaluated by a test sequence, with the equaliser tap settings being fixed during data transmission, is known as an <u>automatic</u> equaliser.

The method of interrupting the data to transmit a probe sequence periodically introduces redundancy. A more efficient method of determining the channel response is to use the data itself, as in Fig. 3-11.9. The real data is randomised with a scrambler so that the transmitted signal s(t) has the same characteristics as the m(t) sequence used above. The received signal r(t) is fed through the equaliser to the decision maker. The equaliser is assumed to have been set up as above by using a test signal to determine the initial channel characteristics. Then the decision-maker output s'(t) will be the same as s(t) except for decision errors. Hence by using the same correlation technique as in Fig. 3-11.8, with one correlator for each i_n in the channel response {I}, the channel can be found. The averaging process of correlation will usually prevent decision errors from affecting the correlator outputs too seriously.

The outputs of each correlator are used to adjust the equaliser tap settings. The changes in the channel structure will be followed because the correlators are operating continuously. Such an equaliser system whereby the data itself is used to evaluate the channel respone, with the equaliser taps being continually modified, is termed adaptive.

Conclusions

We have seen described the broad outlines of digital equalisers. Some examples of them are in papers by Hirsch and Wolf[44], Niessen and Willim[45], Lender[46], Monsen[47], and Bowen, George and Storey[48]. The two structures commoriy realised are the transversal system and the decision-feedback system.

The transversal equaliser has the disadvantage that it can never be the precise inverse of the channel. There is always some intersymboi interference present in the pulse stream after equalisation.

The decision-feedback equaliser requires less taps than the transversal filter because it has the inverse structure of the channel. Also, for ASK, it requires only half its taps to be operating because,in effect, just "1"s are equalised. Hence for both these reasons the decision feedback equaliser should suffer less from inaccurate tap settings. Against this is the inherent property of the decision-feedback structure to propagate errors from wrong decisions.

Comparisons of the two structures under simulated channel conditions, e.g. by George, Coll, Kaye, and Bowen[49] and Monsen[50] (further references are given by Lucky[42]), have in general shown the superiority in performance of the decision-feedback form. The decision-feedback equaliser has the further advantage of being easier

to implement. This is because the tap settings bear a direct relationship to the channel impulse response measurements, unlike the transversal filter whose tap settings require an arithmetical algorithm to determine them.

The decision-feedback structure was therefore implemented in the test system. As well as for the above reasons it seemed an appropriate choice because in the trials the first multipath to arrive was to be taken as the signal path. Hence the channel impulse response was a main pulse followed by echoes; no pre-echoes would be present. The test-system equaliser is now described.

3.11.3   The Equaliser Implemented in the Test System

The conventional type of decision-feedback equaliser requires extensive circuitry to realise it. A correlator is required for each channel response "tap", and analogue shift registers and multipliers are needed to realise the digital filter. It was not feasible to construct the equaliser in this form and so an alternative equaliser was evolved, as follows. First of all its way of evaluating the channel response is described, and then its method of removing intersymbol interference is given.

Channel identification

The scheme was adopted in the test system to periodically interrupt the data to transmit a single pulse (see section 3.1.3). This was done to facilitate the operation of various parts of the receiver including, as we shall now see, the equaliser. Since multipath from the channel was to manifest itself as post-echoes only the channel impulse response under such conditions is displayed in the gap between the set-up pulse and the first pulse in the new block of data. The situation is illustrated in Figs. 3-11.10. For a

digital equaliser the required type of impulse response is the values of the echo tail at bit intervals (T) after the main pulse, as shown in Fig. 3-11.10. The digital impulse response is the voltages V_1(=0 in the illustration), V_2, V_3 etc. at the intervals T, 2T, 3T, etc. after the main pulse.

In the test-system equaliser these voltages were measured by a series of level detectors of the type described in section 3.9, the sample-and-hold of each particular level detector being triggered at the appropriate number of bit periods after the set-up pulse. The sampling rate (because of the fixed transmission rate of the set-up pulse), and the cut-off frequency of the low-pass filters, were made the same as for the pulse-height and residual-dc level detectors. This was because the variations of the V_1, V_2, V_3 etc. envelopes would be the same.

Note: we can see why it was important that there was no residual dc because, otherwise, the channel level detectors would have treated this as "echoes" and so have made false measurements of the impulse response.

The method of operation of the equaliser in removing inter-symbol interference is now explained.

Equaliser operation

The block diagram of the equaliser for ASK is shown in Fig. 3-11.11.

The output of the filter of the nth level detector is V_n, which is the envelope of the voltage of the impulse response at the instance nT after the decision instance (the centre of the set-up pulse). Now let us assume that the set-up pulse in Fig. 3-11.10

represents a "1". Then if, for all n, the outputs V_n of each of the level detectors in Fig.3.11.11 are successively subtracted from the pulse stream the intersymbol interference from the set-up pulse is removed as far as decision instances of subsequent pulses are concerned. When the decision-maker detects each "1" this operation occurs. When a "0" is detected no pulse has been transmitted so the operation is not triggered because no intersymbol interference will have been generated. This type of equaliser is therefore equivalent to the decision-feedback form.

The equaliser's operation has been explained for ASK. It can also be employed for PSK because of the bipolar-ASK method used of generating PSK. For a "1" the equaliser's operation in removing intersymbol interference is identical to that for ASK. For PSK a "0" is transmitted by a negative pulse; hence its intersymbol interference products are removed by sequentially <u>adding</u> the level detector outputs to the pulse stream.

The equaliser was realised as shown in Fig. 3-11.11 using standard integrated circuitry. (The sample-and-hold trigger instances and the level detector outputs refer to the values in Fig. 3-11.10.) We see in Fig. 3-11.11 that the digital output of the decision-maker is fed to a digital shift-register chain. When a "1" is detected it "ripples" down the chain at each pulse of the clock, which has the period T of the bit rate. When a "1" is present at the output of the nth stage V_n is subtracted from the pulse stream, the appropriate analogue switch having been activated on. In this way the sequential subtraction of the pulse tails was affected. For PSK a separate chain of shift registers and analogue switches were incorporated.

In summary then, such an equaliser is of the decision-feedback type with tail cancellation only. It may be termed automatic in that the data has to be periodically interrupted to transmit a probe pulse, and be termed adaptive in that it continually monitors changes in the channel digital impulse response. This is because if the envelopes of the impulse response "taps" are in effect bandlimited in their amplitude variations, and if the sampling theorem is correctly applied to the level detector components, then these envelopes will be continuously followed, the filter outputs changing during the transmission of the data as they follow the sample-and-hold outputs.

3.12   THE DECISION-MAKER

The circuit of the decision-maker used in the test system
is shown in Fig. 3-12.1. The demodulated pulse stream and the
decision level are fed into a hard-limiter voltage comparator.
The output of the comparator is thus the pulse stream sliced into
bits.

However the polarity of the output bit stream is ambiguous
when the PLL is operating in its two-phase stable mode. This
occurred when the three modulation modes that produce modulated
carriers that had 0 and π phase components were being transmitted.
The polarity ambiguity was resolved in the test system by the remainder
of the circuit shown in Fig. 3-12.1. Here we see the sliced bit
stream is fed into an EXCLUSIVE-OR logic gate together with the
quantised output from the pulse-height level detector. If this level
is positive the slicer output is passed directly to the error detector;
if the level is negative the bit stream is first inverted. The
level is negative when the PLL is locked in its π mode; hence the
demodulated pulse stream appears as "upside down".

This bit polarity correction is elementary for the ASK and
ASK-SC modes since it is equivalent to taking the analogue modulus
of the demodulated pulse stream. For PSK this method of polarity
correction is most important. The pulse stream is effectively coded
in transmission such that, if a pulse is detected at the receiver as
being of the same polarity as the set-up pulse, it is a "1". If
the opposite then it is a "0". Since this technique was employed
in the test system it meant that differential encoding and decoding
was not necessary for PSK. Hence the error rate was not doubled as
a result.

The only point not touched upon in the description of the decision-maker is the setting for the decision level. This should be optimised for each of the modes to give the lowest error rate in the presence of Gaussian noise. The rest of section 3.12 is concerned with this. In section 3.12.1 we deal with the correct setting for the coherently-detected modes, PSK, ASK, ASK-SC and ASK+C.

The correct level for the rectifier detection of ASK is given in section 3.12.2. We find that it is a complex function of the signal-to-noise ratio, unlike the case for coherent detection where the optimum decision level is a constant fraction of the pulse height. We see that it is possible to derive a good approximation to the optimum level for rectifier detection. It results from a linear combination of the expected values of the "1" level and the "0" level in the presence of rectified noise. These two values were the outputs of the "0" and "1" level detectors (Block 3-9, Fig. 3-0.2) in the test system. Firstly coherent detection is analysed.

3.12.1   Decision Level for the Coherently-Detected Modes

The setting for the optimum level is well known in the case of coherent detection. Assume that V_1 is the level at the sampling instance for a "1" and V_0 is the level for a "0" (assuming the absence of noise). The distribution of the additive Gaussian noise around these levels is shown in Fig. 3-12.2, V_D being the decision level. The value of V_D that gives the least error probability is found as follows.

The probability of a false "1" is given by the area under $p(V-\ddot{v}_1)$ for

$$V_D < V < \infty$$

(p(V) is the distribution of the noise).

This area is

$$\int_{V_D}^{\infty} p(V-V_1)dV$$

Similarly the probability of a false zero is

$$\int_{-\infty}^{V_D} p(V-V_0)dV$$

Hence the total error probability is given by:

$$P_e = \frac{1}{2} \int_{V_D}^{\infty} p(V-V_1) + \frac{1}{2} \int_{-\infty}^{V_D} p(V-V_0)dV$$

(assuming an equal density of "1"s and "0"s).

Using an identical argument to that in section 3.7.2 the optimum value of V_D is given by:

$$p(V_1-V_D) = p(V_D-V_0)$$

i.e.

$$V_D = \frac{V_1+V_0}{2}$$

Hence with coherent detection the optimum decision level is midway between the "1" and "0" amplitudes.

The threshold levels for each of the coherently-detected modes is now given.

PSK

Here $V_1 = V_p$ and $V_0 = -V_p$, V_p being the pulse amplitude.

Hence $V_D = 0$.

ASK

This is on-off keying and so $V_1 = V_p$ and $V_0 = 0$.

Hence $V_D = V_p/2$.

ASK-SC, ASK+C

The residual level will have been removed by the residual dc extractor   Hence

$$V_D = V_p'/2$$

V_p' being the pulse height.

In all the ASK modes the pulse height amplitude V_p was measured by the level detector described in section 3.9.

The analysis so far has assumed perfect carrier extraction. In fact there should be a small correction factor for the ASK and ASK+C modes. The true "1" level is less than V_1 because the PLL detector was not perfect, the phase angle of the PLL output being perturbed by the presence of the additive noise. These perturbations may be construed as being an additional noise source having a wholly negative probability distribution. This distribution could be superimposed on the Gaussian one to derive a modified optimum for the decision level V_D. However the phase angle distribution is not known and in any case its effect has been shown to be small compared

to the additive noise in the signal (in section 3.7.1). The
decision threshold was therefore taken as being one-half of the
pulse-height level detector output.

We now turn to the derivation of the optimum decision level
for ASK, rectifier detected.

3.12.2 Decision Level for Rectifier Detection

Rectifier detection was used in the test system on the ASK
mode (where on-off carrier keying was used).

Rectifier detection distorts the Gaussian probability
distribution since the output of such a detector must always be
positive. The situation has been analysed in section 3.7.2. Here
we found that the optimum decision level was the solution for x
of:

$$I_0(xA^2/N) = \exp(A^2/2N)$$

where xA is the decision level, A is the pulse amplitude and N is the
noise power. This equation can be written as

$$I_0(xS) = \exp(S/2) \qquad\qquad (3.12.1)$$

S being the peak-signal-to-noise ratio.

The solution of this equation is drawn in Fig. 3-12.3.
As can be seen x is a function of the signal-to-noise ratio. There
is obviously a difficulty in deriving the solution to eqn. (3.12.1)
in a practical hardware system. A means of approximating to the
solution for the purposes of practical implementation is now given.

In section 3.9 a level detector which operates to extract the envelope of an amplitude was described. When rectifier detection is employed the output of this device will be the expected value, the first moment of the probability density function, of the sum of the amplitude in question and the rectified noise. Since the rectification of the sum of the amplitude plus noise produces a pdf which cannot go less than zero the expected value should increase as the signal-to-noise decreases like the threshold level does in Fig. 3-12.3. In fact it is possible to linearly combine the outputs of two level detectors, one operating on the pulse height plus rectified noise, the other on the rectified noise only, to give the optimum decision level to a very close approximation.

Let E_1 be the expected value of the pulse height A plus rectified noise. The probability density function for signal R consisting of a level A plus rectified noise of variance N is given by eqn. (3.7.8) as:

$$P_A[R] = \frac{R}{N} \exp(-R^2/2N) \exp(-A^2/2N) I_0(RA/N) \qquad (3.12.2)$$

The expected value E_1 of R is given by:

$$E_1 = \int_0^\infty R \, P_A(R) dR$$

$$= \int_0^\infty \frac{R^2}{N} \exp[(-R^2-A^2)/2N] \, I_0(RA/N) dR$$

Rice has solved this integral[16]. Substituting n = 1 into his eqn. (3.10.12) gives:

$$E_1 = \sqrt{(2N)} G(3/2) \, _1F_1(-1/2;1; -A^2/2N) \qquad (3.12.3)$$

$G(x)$ is the Gamma function and $_1F_1(x;y;z)$ is the hypergeometric function.

From standard tables $G(3/2) = \sqrt{\pi}/2$.

In Appendix 4B of the above paper Rice shows that

$$_1F_1(-1/2;1; - z) = \exp(-z/2) \left[(1+z) I_0(z/2) + zI_1(z/2)\right]$$

where $I_0(x)$ and $I_1(x)$ are Bessel functions.

Eqn. (3.12.3) then becomes

$$E_1 = \sqrt{(2N)}(\sqrt{\pi}/2)\exp(-A^2/4N) \left[(1+A^2/2)I_0(A^2/4N)+(A^2/2N)I_0(A^2/4N)\right]$$

Substituting $S = A^2/N$ (the peak-signal-to-noise ratio) and writing $\sqrt{N}$ in terms of S and A leads to

$$E_1 = A\sqrt{(\pi/2)}\left[(1+S/2)I_0(S/4) + (S/2)I_1(S/4)\right] \left[\exp(-S/4)\right]/\sqrt{S} \qquad (3.12.4)$$

This function is plotted in Fig. 3-12.3. We see that as S becomes large E_1 tends to the value A. This is to be expected because the Rician distribution becomes Gaussian at high signal-to-noise ratios. (By using the approximation

$$I_n(x) = (\exp x)/\sqrt{(2\pi x)}$$

for large x in eqn. (3.12.4) E_1 can be shown to be A.)

Let us now evaluate the output of a level detector operating on the rectified noise only, no signal being present. It is the expected value of rectified noise E_0, say.

The pdf of rectified noise is the Rayleigh distribution given in eqn. (3.7.9):

$$P_0[R] = \frac{R}{N} \exp(-R^2/2N)$$

R being the rectified noise.

The expected value E_0 is

$$E_0 = \int_0^\infty R \cdot \frac{R}{N} \exp(-R^2/2N)dR$$

$$= - \int_0^\infty Rd[\exp(-R^2/2N)]$$

Integration by parts leads to

$$E_0 = - [R \exp(-R^2/2N)]_0^\infty - (-\int_0^\infty \exp(-R^2/2N)dr)$$

$$= \frac{1}{2} \frac{\sqrt{\pi}}{\sqrt{(1/2N)}}$$

by standard integral

$$\therefore \quad E_0 = \sqrt{N} \sqrt{(\pi/2)}$$

Substituting for N with S and A leads to:

$$E_0 = A\sqrt{(\pi/2S)} \qquad (3.12.5)$$

This function is also plotted in Fig. 3-12.3.

The optimum threshold as shown in Fig. 3-12.3 can be seen to approximate the average of E_1 plus E_0. In fact it is found by using the relationship

$$x = -\frac{0.92\,E_1 + F_0}{2A} \tag{3.12.6}$$

a very close approximation can be made to the minimum error rate possible for a given signal-to-noise ratio. The table in Fig. 3-12.4 compares the error rates given by using eqn. (3.12.6) with the ideal-threshold eqn. (3.12.1).

Eqn. (3.12.6) involves a linear combination of E_1 and E_0 and can be easily implemented in hardware. It was therefore used in the test system.

3.13   ERROR DETECTOR AND COUNTER

The error detector

The transmitted data used in the test system was the output of a pseudo-random generator. The error detector had an identical generator whose sequence could be synchronised to that of the transmitter. Bit-to-bit comparisons could therefore be made with the incoming bit stream to produce a pulse for each error detected.

The error counter

Three simultaneous error counts were made:

(1) Datum error count - It was felt that it would be useful to establish the error rate caused by factors other than intersymbol interference. Reference to section 3.1.3 shows that, if the bit synchroniser is locked onto the pulse from the first multipath to arrive, the set-up pulse and the first bit in the new block of data must be free of intersymbol interference. Hence, the start of the leading edge of the set-up pulse and its centre were sampled in the case of ASK, in the knowledge that they should have been "1" and "0" respectively. The first bit of the new data block was also sampled, its value depending on the pseudo-random sequence. Hence, for ASK, three bits per frame could be sampled for this datum error count. For PSK though a "0" is a negative pulse and so only two bits, the set-up pulse and the first bit in each data block, could be sampled per frame.

(2) Errors after equalisation

(3) <u>Errors before equalisation</u> - For this measurement a separate
decision-maker was used.

The uses of these three counts may be summarised thus:

A comparison of the error rates from measurements (1) and
(3) gives an indication of the intersymbol interference present.

A comparison of the error rates from (1) and (2) indicates
how closely the performance of the equaliser was to the optimum.

Comparing (2) and (3) shows whether the counteraction of
multipath by the equaliser compensates for the amount by which it
amplifies the noise present in the signal from the channel.

APPENDIX 3.1

THE FOURIER TRANSFORM OF THE GENERAL RAISED-COSINE FUNCTION

The one-sided spectrum $H(f)$ of the raised-cosine function is:

$$H(f) = 1 , \qquad\qquad 0 < f < f_c - f_x$$

$$H(f) = \frac{1}{2} \left[1 + \cos \pi(f + f_x - f_c)/2f_x \right] , \qquad f_c - f_x < f < f_c + f_x$$

$$H(f) = 0 , \qquad\qquad f > f_c + f_x$$

Its fourier transform $h'(t)$ is given by:

$$h'(t) = \int_{-\infty}^{\infty} H(f) \exp(j2\pi ft) df$$

$$= \int_{-\infty}^{\infty} H(f) \cos 2\pi ft \, df + j \int_{-\infty}^{\infty} H(f) \sin 2\pi ft \, df$$

Since $H(f)$ is an even function we have:

$$h'(t) = 2 \int_{0}^{\infty} H(f) \cos 2\pi ft \, df$$

$$= 2 \int_{0}^{f_c - f_x} \cos 2\pi ft + 2 \int_{f_c - f_x}^{f_c + f_x} \frac{1}{2} \left[1 + \cos \pi(f + f_x - f_c)/2f_x \right] \cos 2\pi \, ft \, df$$

$$= 2 \left[\frac{\sin 2\pi ft}{2\pi t} \right]_{0}^{f_c - f_x} + \left[\frac{\sin 2\pi ft}{2\pi t} \right]_{f_c - f_x}^{f_c + f_x} + \int_{f_c - f_x}^{f_c + f_x} \cos \pi(f + f_x - f_c)/2f_x \cos 2\pi ft \, df$$

$$= \frac{\sin 2\pi f_c t \, \cos 2\pi f_x t}{\pi t} + \int_{f_c - f_x}^{f_c + f_x} \cos \pi(f + f_x - f_c)/2f_x \cos 2\pi ft \, df$$

Evaluation of the second integral leads to:

$$\int_{f_c-f_x}^{f_c+f_x} \cos (f+f_x-f_c)/2f_x \cos 2\pi ft \, df$$

$$= \frac{4\pi t}{(\pi^2/4f_x^2)-4\pi^2 t^2} \sin 2\pi f_c t \cos 2\pi f_x t$$

Hence h'(t) is given by:

$$h'(t) = \frac{\sin 2\pi f_c t \cos 2\pi f_x t}{\pi t} + \frac{4\pi t}{(\pi^2/4f_x^2) - 4\pi^2 t^2} \sin 2\pi f_c t \cos 2\pi f_x t$$

$$= \frac{\pi^2/4f_x^2}{\pi t(\pi^2/4f_x^2 - 4\pi^2 t^2)} \sin 2\pi f_c t \cos 2\pi f_x t$$

$$\therefore \quad h'(t) = 2f_c \frac{\sin 2\pi f_c t}{2\pi f_c t} \frac{\cos 2\pi f_x t}{1-(4f_x t)^2}$$

Normalising h'(t) to h(t) such that

$$h(t) = \frac{h'(t)}{h'(0)}$$

leads to

$$h(t) = \frac{\sin 2\pi f_c t}{2\pi f_c t} \frac{\cos 2\pi f_x t}{1-(4f_x t)^2}$$

HF LOG-PERIODIC ANTENNA
SERIES 747F

HORIZONTAL LOG-PERIODIC OFFERS VARIABLE TAKE-OFF ANGLES FROM 4 TO 30 MHz

- Performance-proven single-tower construction

- Withstands harsh environments--winds to 120 mph

- Occupies minimum space

- Gain of 10 to 13 dBi

- VSWR 2.1:1 over entire frequency range

- Eight versions available

GENERAL

747F series is a horizontally polarized, antenna system that operates over the frequency range 2 to 30 MHz.* Series 747F antennas covering the 4 to 30 MHz frequency range are also available. It is used in point-to-point circuits and for shore-to-ship and ground-to-air communications.

This antenna, like others in Granger Associates' 747 series, offers carefully controlled radiation patterns. Optimum take-off angles are built into the 747F; they vary in a predetermined manner with frequency. At 4 MHz, for example, the take-off angle is approximately 50°. At 30 MHz, it decreases to 27°. This permits one-hop transmissions over distances from 400 miles (at 4 MHz) to approximately 900 miles (at 30 MHz).

Maximum VSWR of Model 747F is 2.1:1. This antenna is available in eight versions (high and low power combinations and a receive-only model) to fit most communications needs.
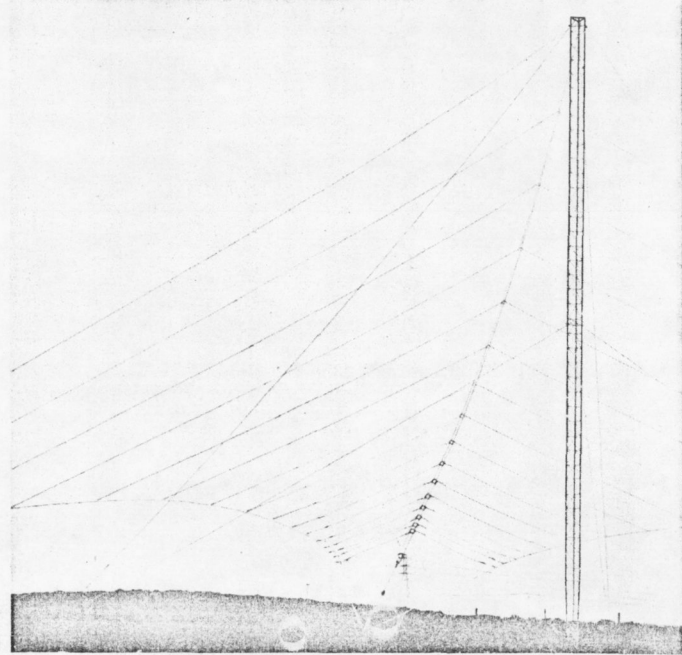
ENVIRONMENTAL

747F series is supported by a single steel tower, 75 feet high. The tower is supplied in knockdown form, and is extremely easy to assemble in the field. Like the antenna itself, the tower is fabricated to give long-term service under extremely harsh environmental conditions. Insulators are high-purity alumina, impervious to weather and unaffected by arcing. The 747F antenna will withstand 120 mph winds (no ice), or 100 mph winds with 1/2 inch of radial icing. It will also endure the highly corrosive tropical atmosphere with no degradation in performance.

LAND REQUIREMENTS

G/A 747F series occupies an area 192 feet by 315 feet, an unusually small area. Its radiation patterns are not appreciably affected by local ground conditions. No ground screen is needed. Even when erected over poor ground, the antenna will perform largely to specification. Where available real estate is scarce, or where extensive site preparation would be too expensive, the 747F series can offer significant cost savings as well as high-quality performance.

*A separate Delta antenna supported from the same tower provides omnidirectional coverage, high angle radiation in the frequency range from 2 to 4 MHz.

Granger Associates

ELEVATION AND AZIMUTH PLANE PATTERNS
Azimuth Patterns at or Near Beam Maximum.

APPENDIX 3.2
(Continued)
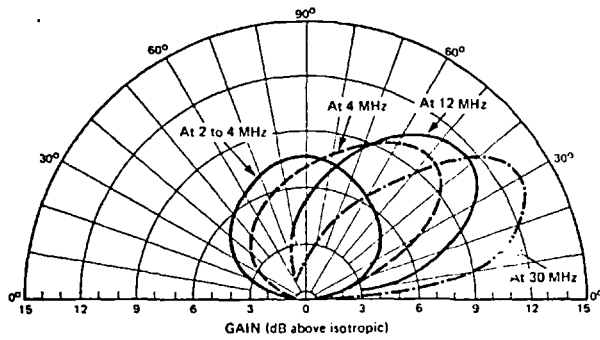
Elevation Plane Radiation Pattern

GAIN (dB above isotropic)
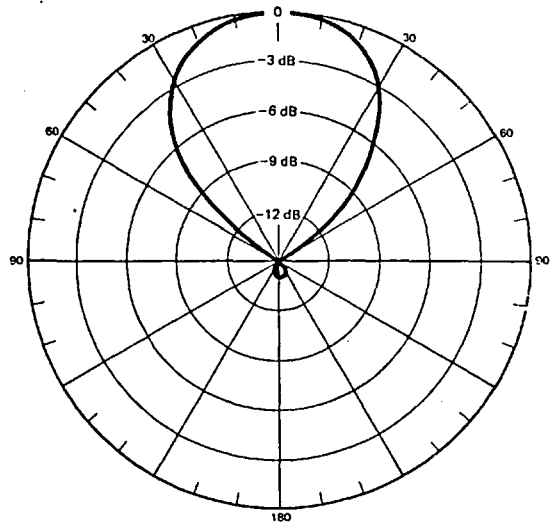
Azimuth plane radiation pattern from 4 MHz to 30 MHz (plotted in dB below maximum gain). From 2 to 4 MHz, the pattern in this plane is essentially omnidirectional.

SPECIFICATIONS: G/A SERIES 747F

MODEL NUMBER	747F-14K	747F-15K	747F-16K	747F-1K	747F-2K	747F-3K	747F-4K	747F-9K
Frequency (MHz)	2-30 *	2-30 *	2-30 *	4-32	4-32	4-32	4-32	4-32
Receive, Transmit	Rx only	Rx Tx	Rx Tx	Rx Tx	Rx only	Rx Tx	Rx Tx	Rx Tx
Impedance (ohms)	50	50	50	200	50	50	50	50
Power (KW)	Receive	30 Peak, 2.5 average	30 Peak, 10 average	40 Peak, 20 average	Not Applicable	30 Peak, 10 average	40 Peak, 20 average	30 Peak, 2.5 average
Connectors	Type N Female	Type LC Female	1-5/8" EIA	Balanced, Openwire	Type N Female	1-5/8" EIA Female	1-5/8" EIA Female	Type LC Female

*A separate Delta antenna supported from the same tower provides omnidirectional coverage, high angle radiation in the frequency range from 2 to 4 MHz.

ELECTRICAL

Polarization. Horizontal

Take-off angle. 50° at 4 MHz, decreasing to 27° at 30 MHz

Pattern beamwidth. Azimuthal plane, 60° nominal
Elevation plane, 65° at 4 MHz, 32° at 30 MHz

Front-to-back ratio. 14 dB

Side-lobe level. −14 dB

Directive gain. 10 dB relative to isotropic at 4 MHz, increasing to 13 dB relative to isotropic at 30 MHz (over perfect ground)

VSWR 2.1:1 from 4 to 30 MHz

MECHANICAL

Environmental. 120 mph wind, no ice (192 km/hr)
100 mph wind, 1/2 inch radial ice (161 km/hr 12.7mm)

Tower height. 75 feet (22.8 m)

Land area required. 192 feet (58.6 m) by 315 feet (96.1 m)

Shipping weight (approx.) 1000 lbs (454 kg)

Shipping volume (approx.) 60 cu. ft. (1.68 m^3)

Accessories.

Tower erection fixture, Part No. 0915313-02
Tower lighting kit, Part No. 0910334-03

Fig 3-0.1   Transmitter system and radio link
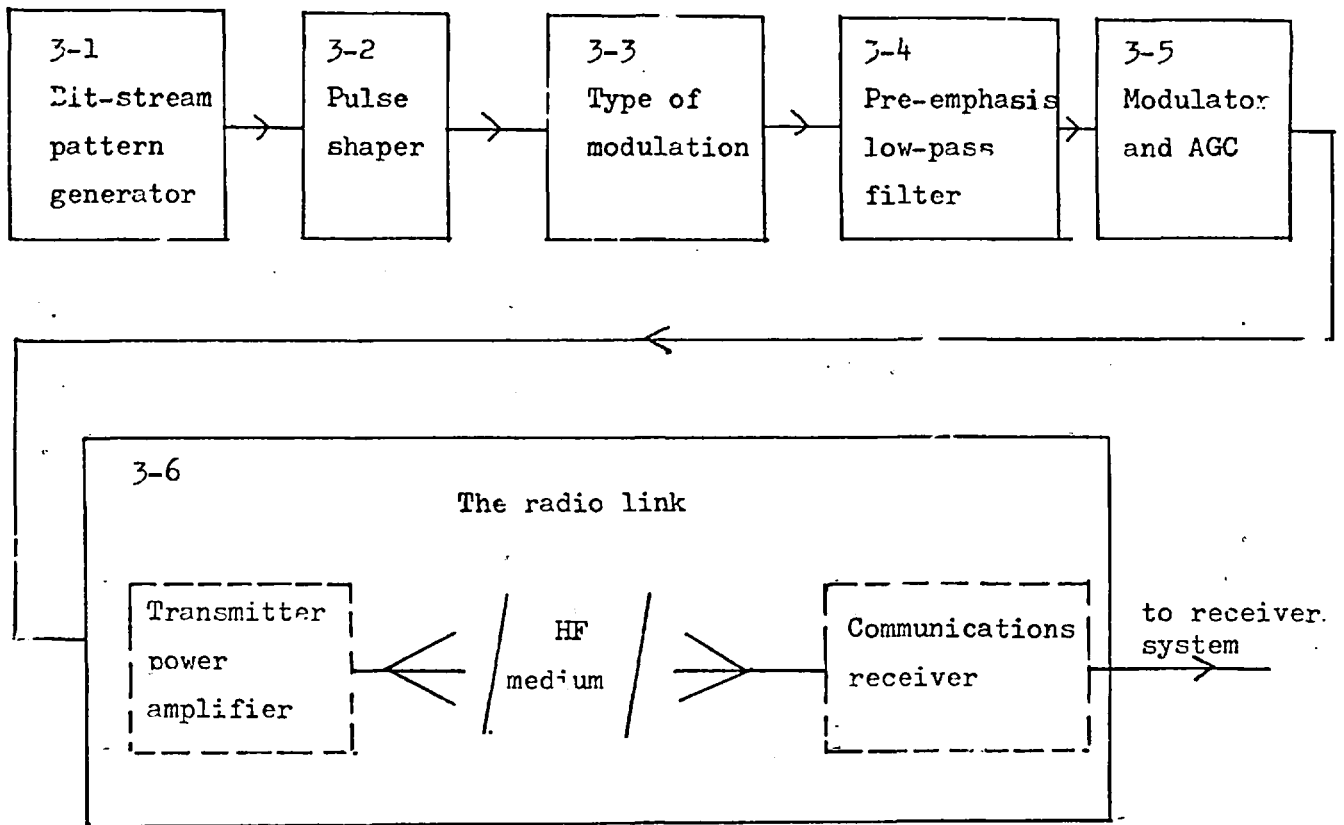
Fig 3-0.2 <u>Receiver system</u>
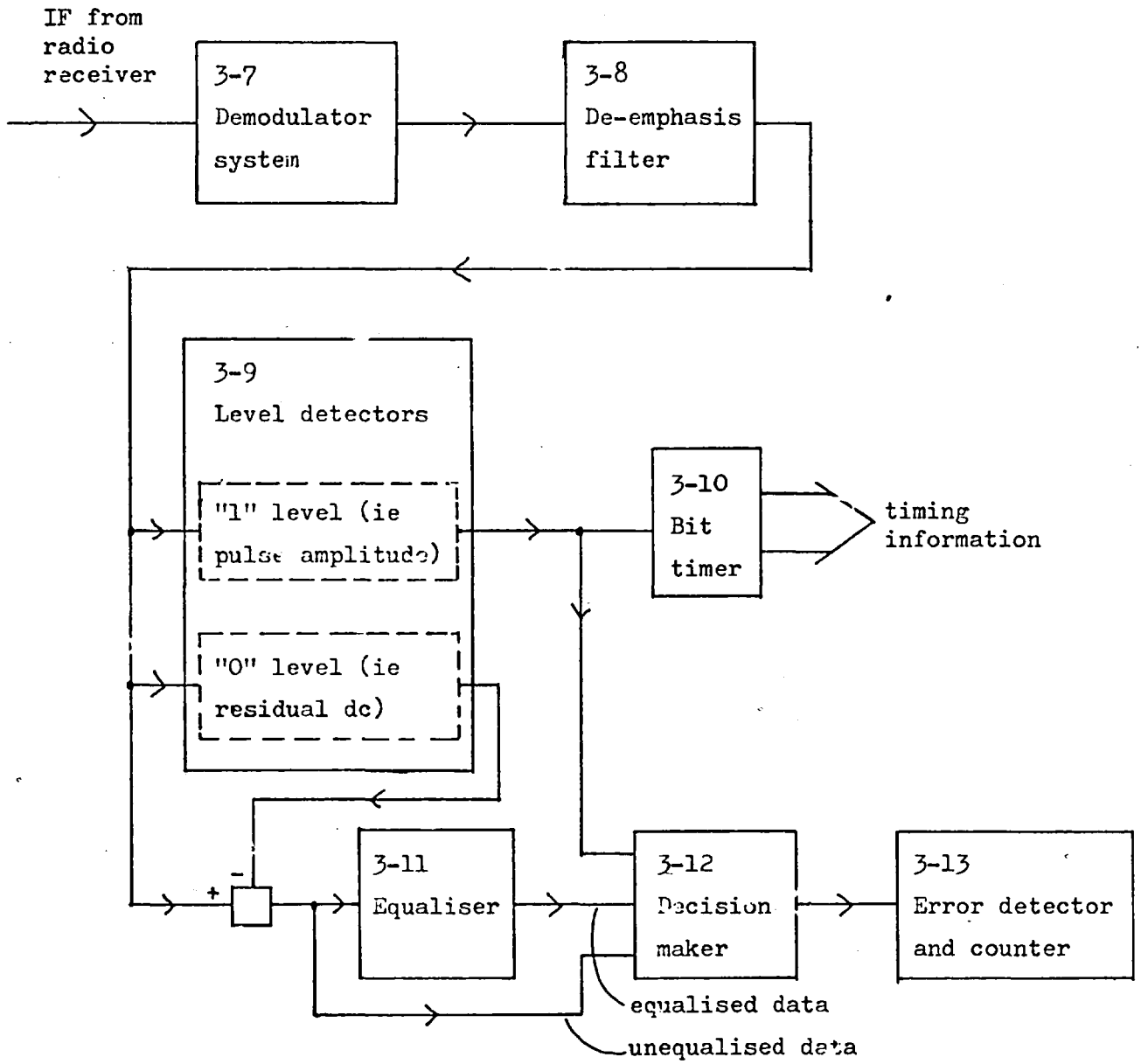
Fig 3-0.3   Geometry of the E- and F_2-layer single-hop propagation
             modes for the case of transmission over 1000km
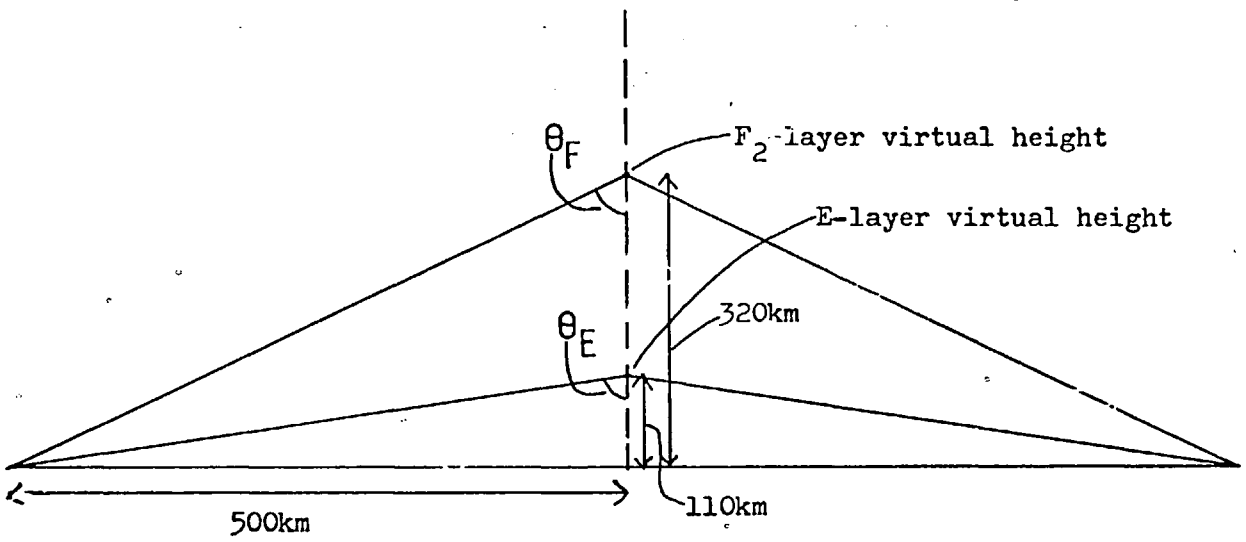
Fig 3-1.1   Data scrambler and de-scrambler

Scrambler(at transmitter)
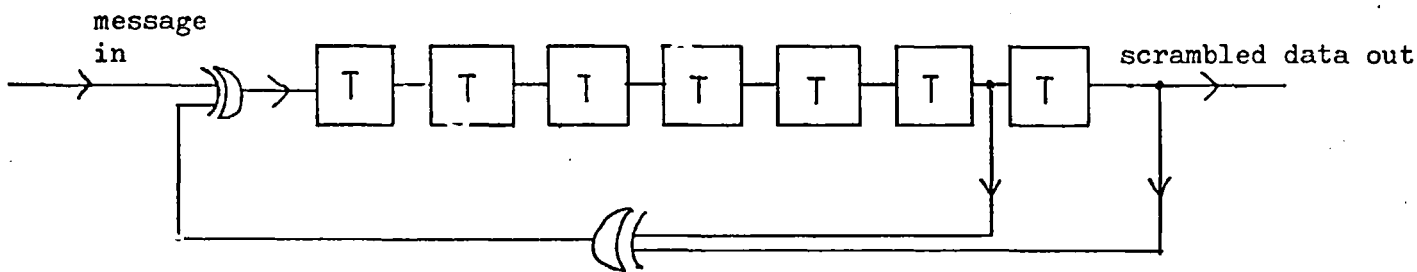
De-scrambler(at receiver)
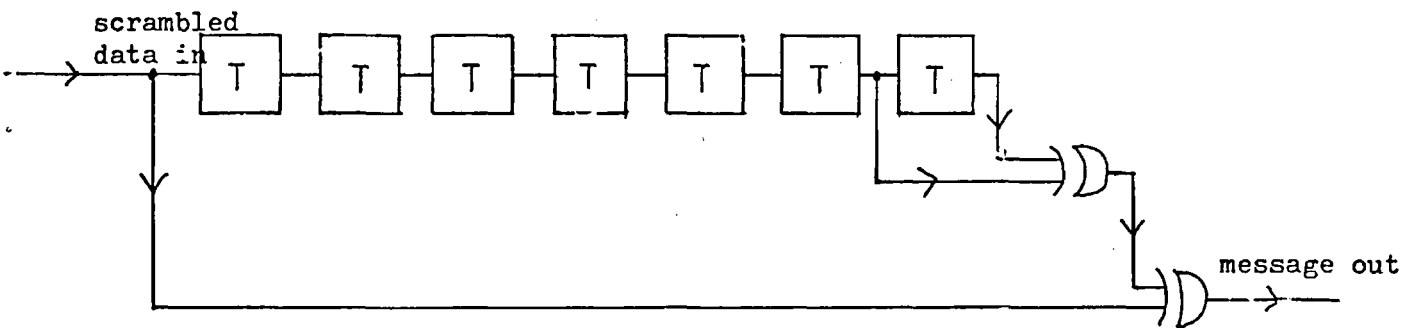
—[T]— denotes shift-register stage
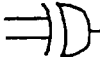
—⟩⟩— denotes EXCLUSIVE-OR gate

-209-

Fig 3-1.2   Set-up pulse with worst-case multipath
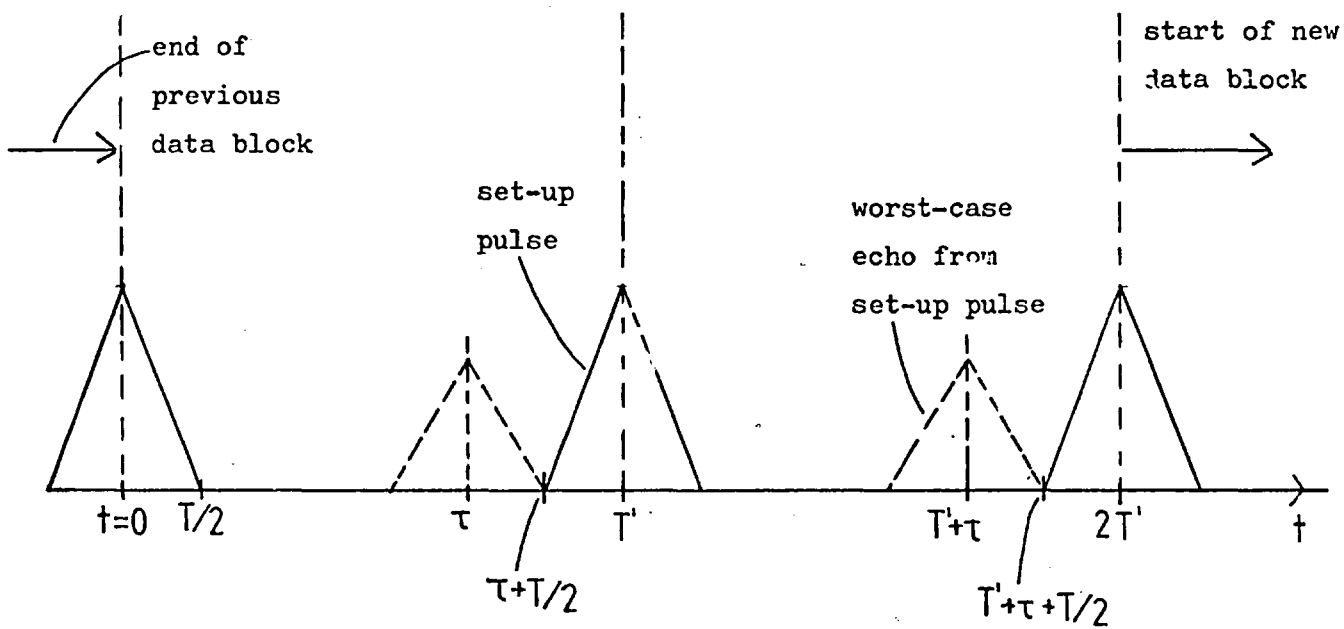
Fig 3-2.1   Raised-cosine spectrum
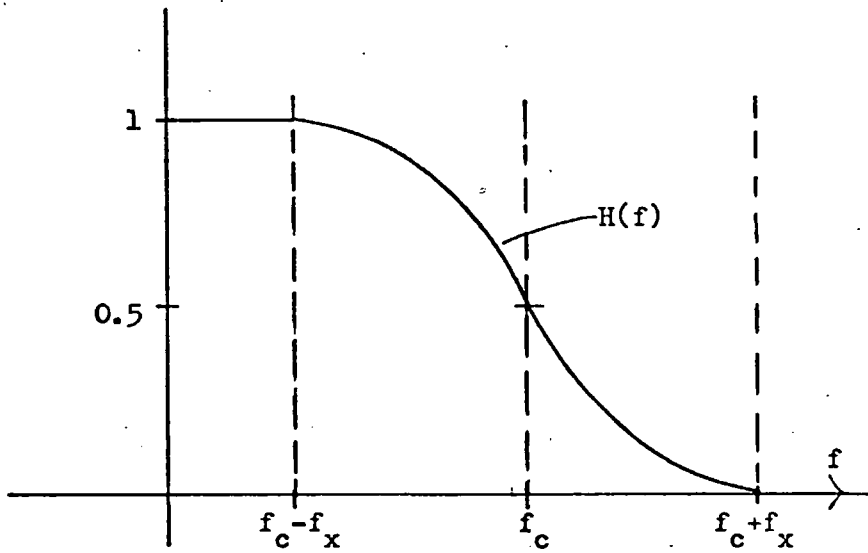
Fig 3-2.2   Pulse shape having raised-cosine spectrum
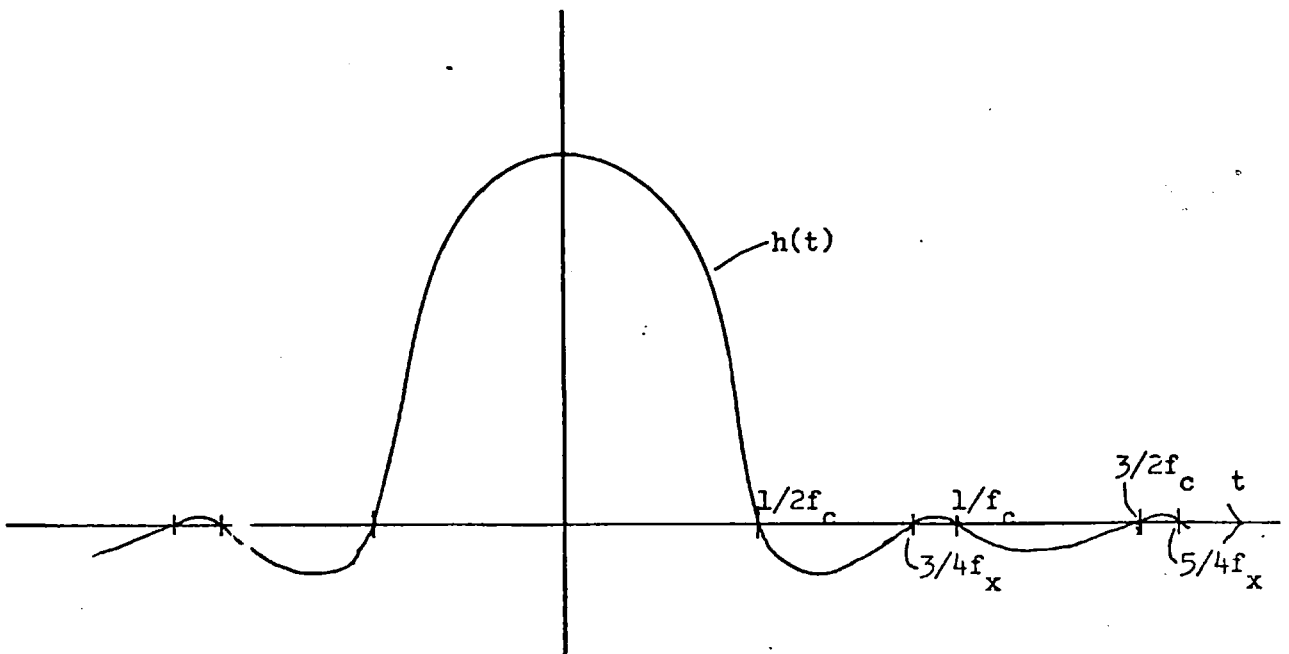
Fig 3-2.3   Sampling-pulse train
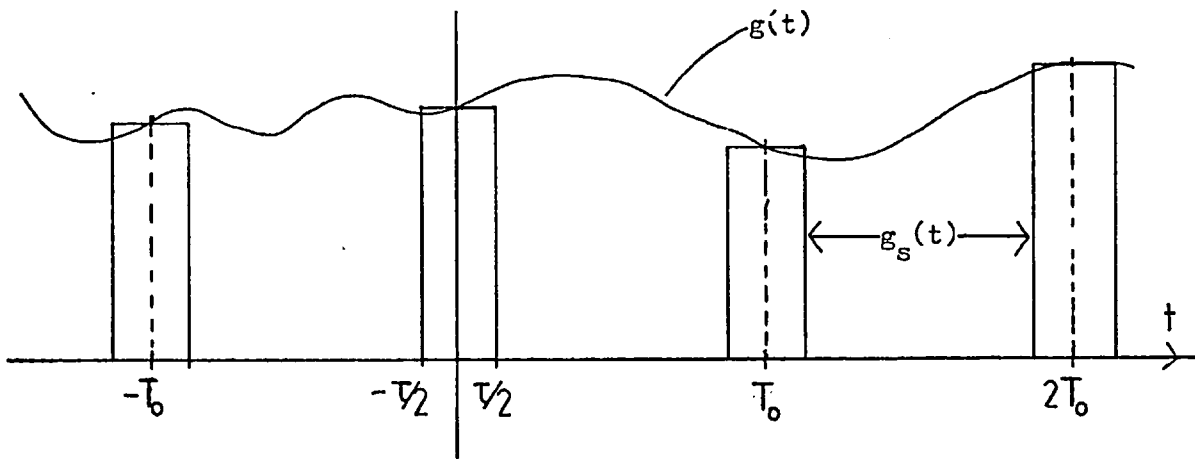
Fig 3-2.4   Spectrum of sampling-pulse train
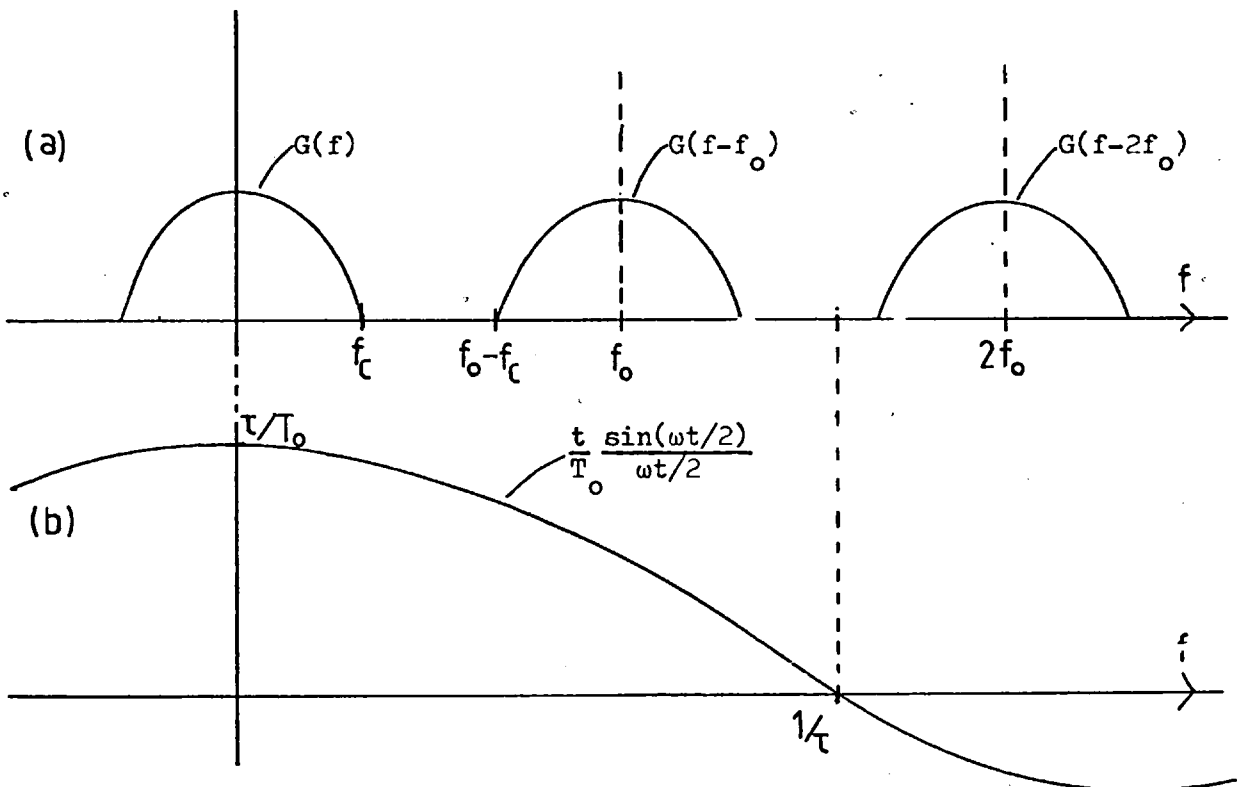
Fig 3-2.5   100% roll-off raised-cosine spectrum

$h(f)=(1+\cos\pi f/2f_c)/2$
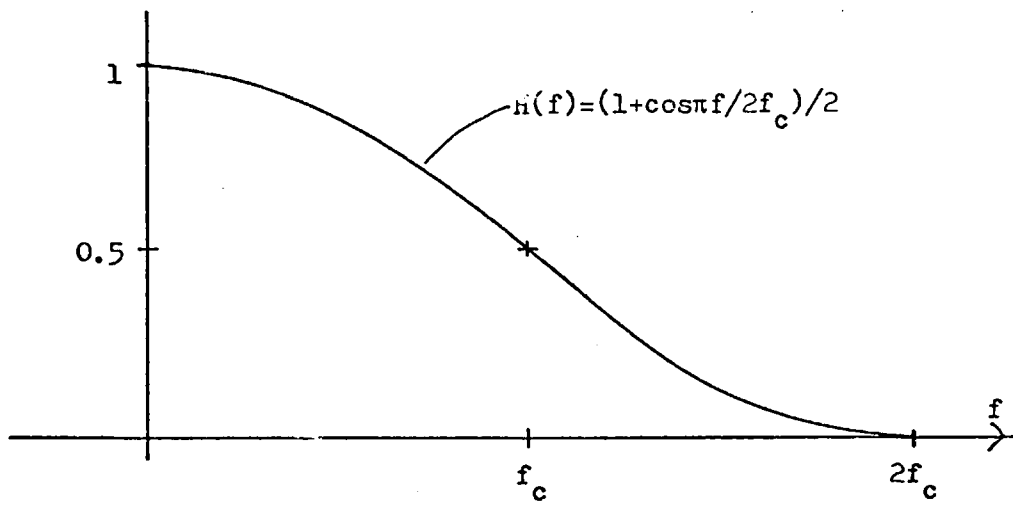
Fig 3-2.6   Shape of pulse having 100% roll-off spectrum
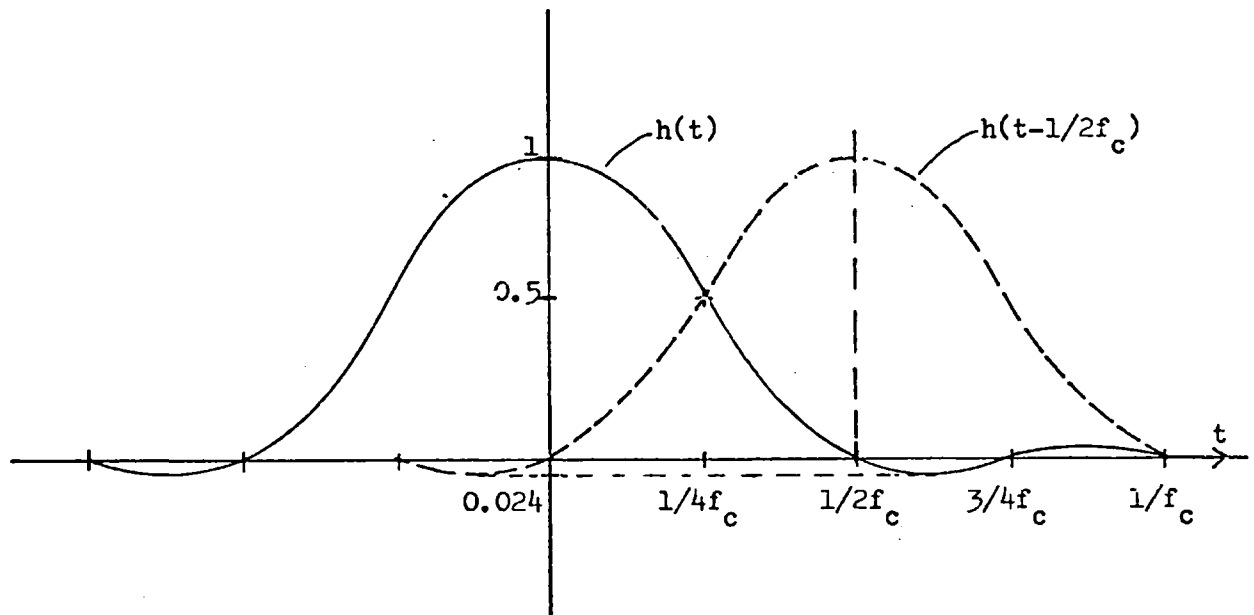
Fig 3-2.7   Sampled pulse having 100% roll-off raised-cosine spectrum
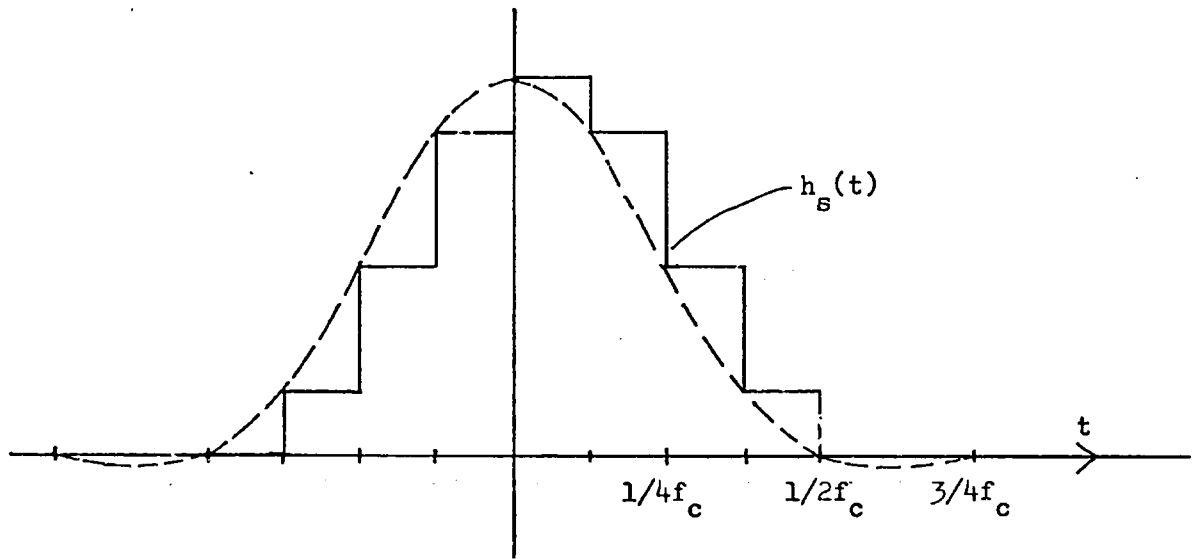
Fig 3-2.8   Spectrum of 100% roll-off sampled pulse
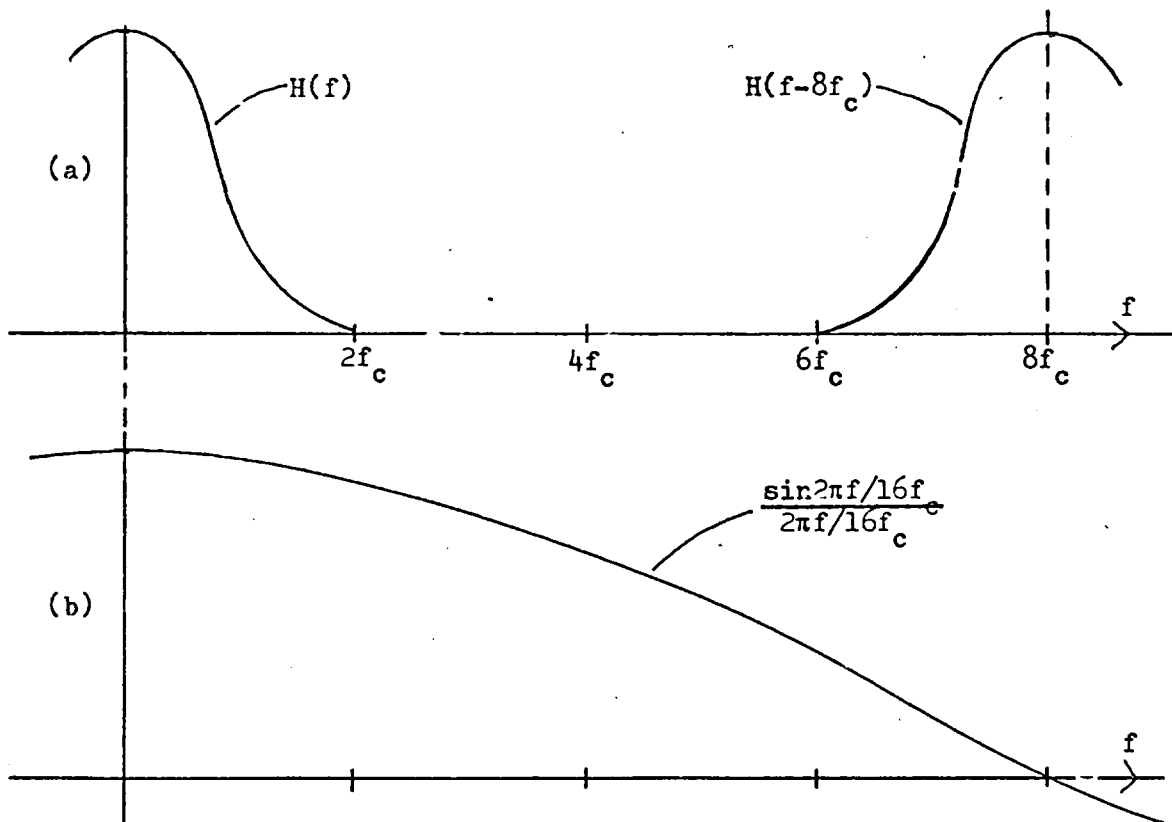
Fig 3-2.9   <u>33% roll-off raised-cosine spectrum</u>
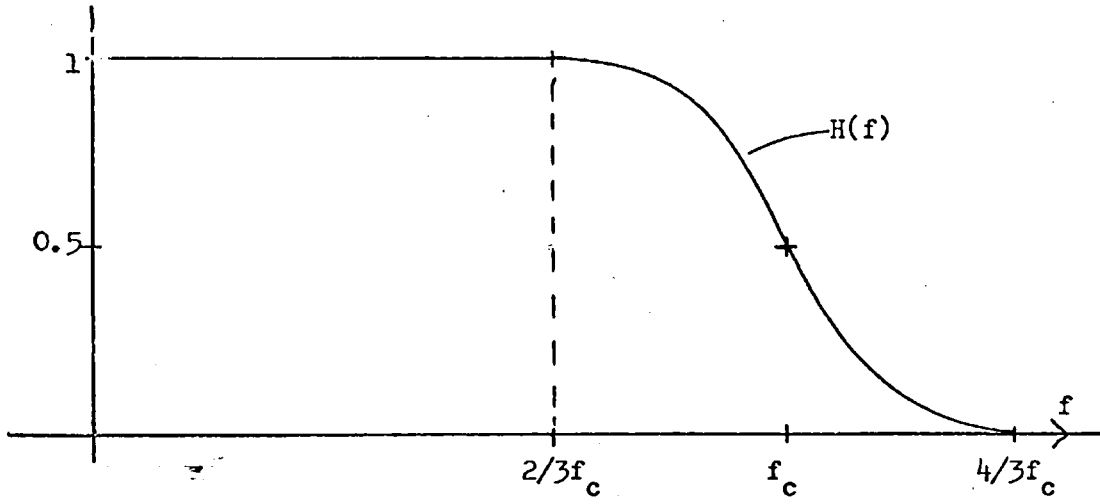
Fig 3-2.10   <u>Shape of pulse having 33% roll-off spectrum</u>
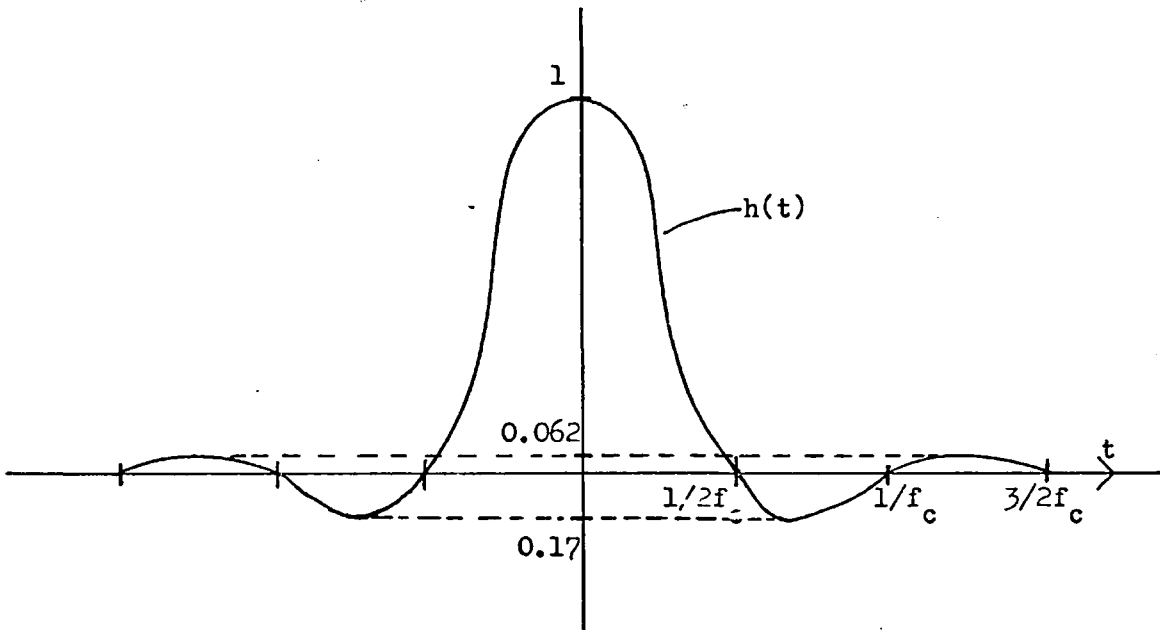
Fig 3-2.11   Sampled pulse having 33% roll-off spectrum
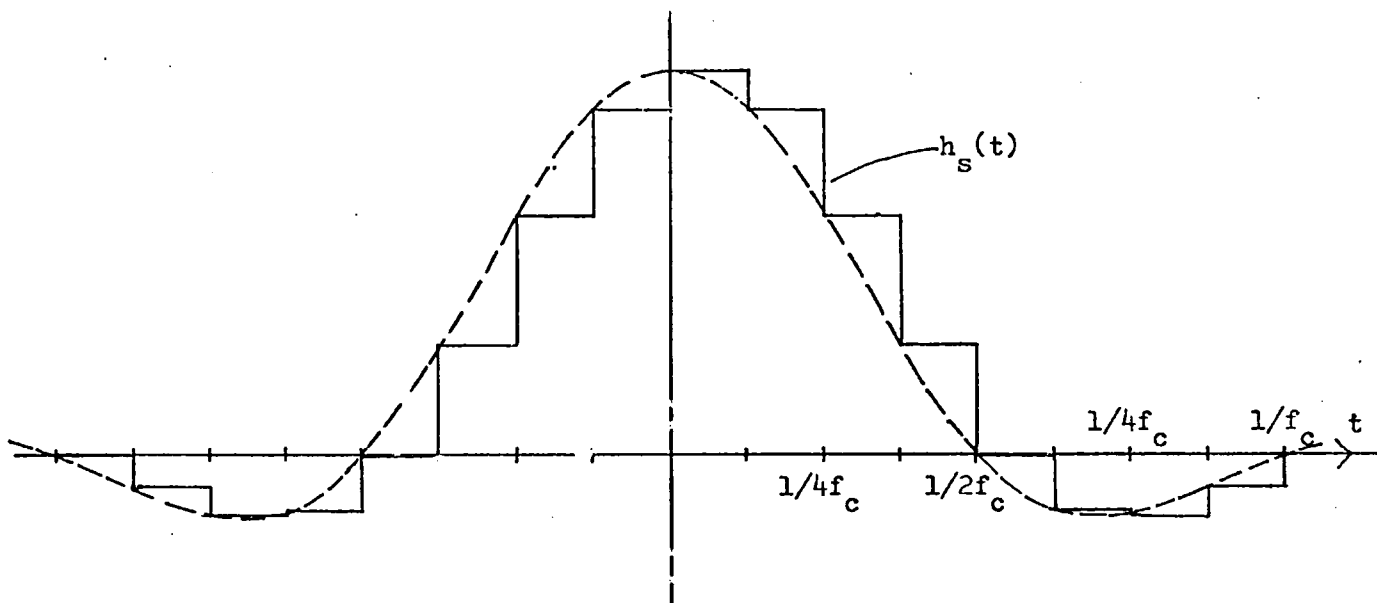
Fig 3-2.12   Spectrum of 33% roll-off sampled pulse
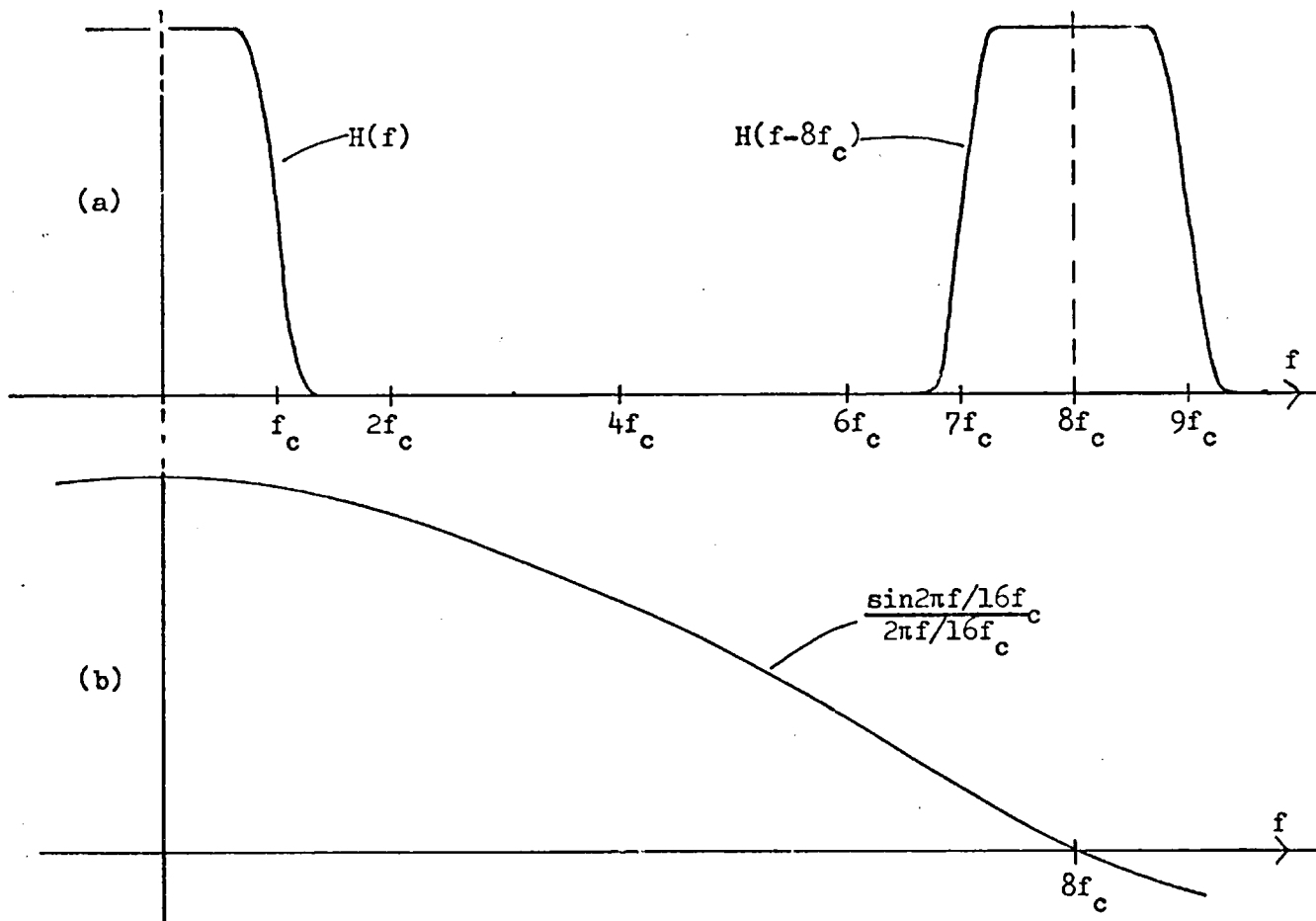
Fig 3-2.13 Synthesis of sampled pulse having 33% roll-off spectrum

Sampled pulse generated
-33% roll-off
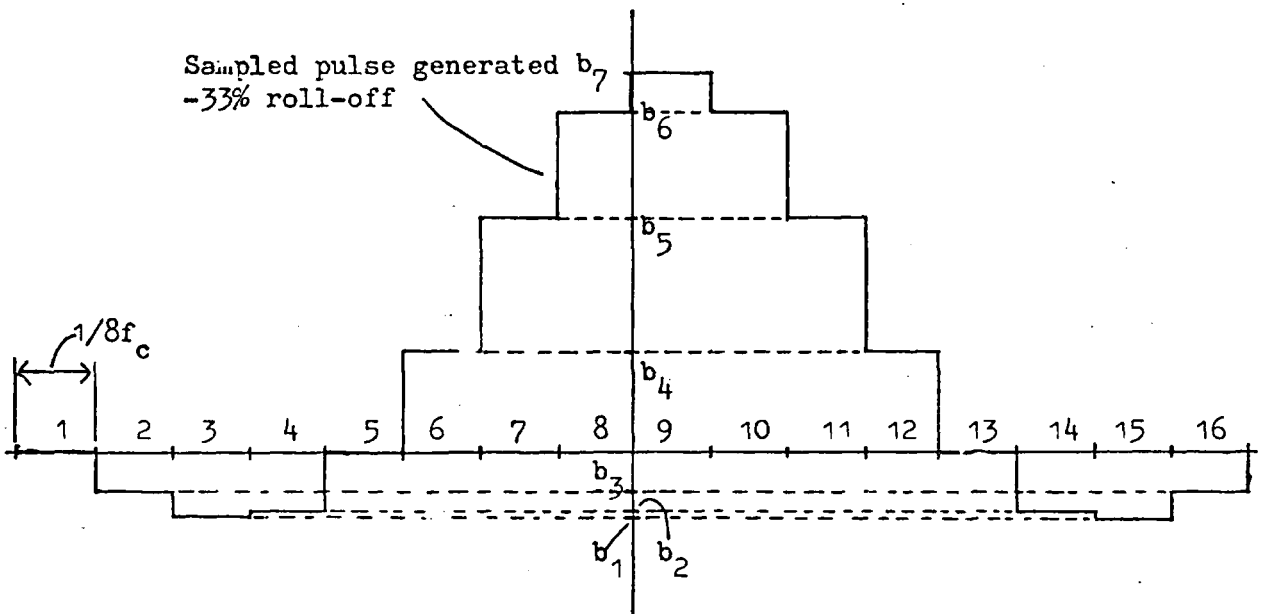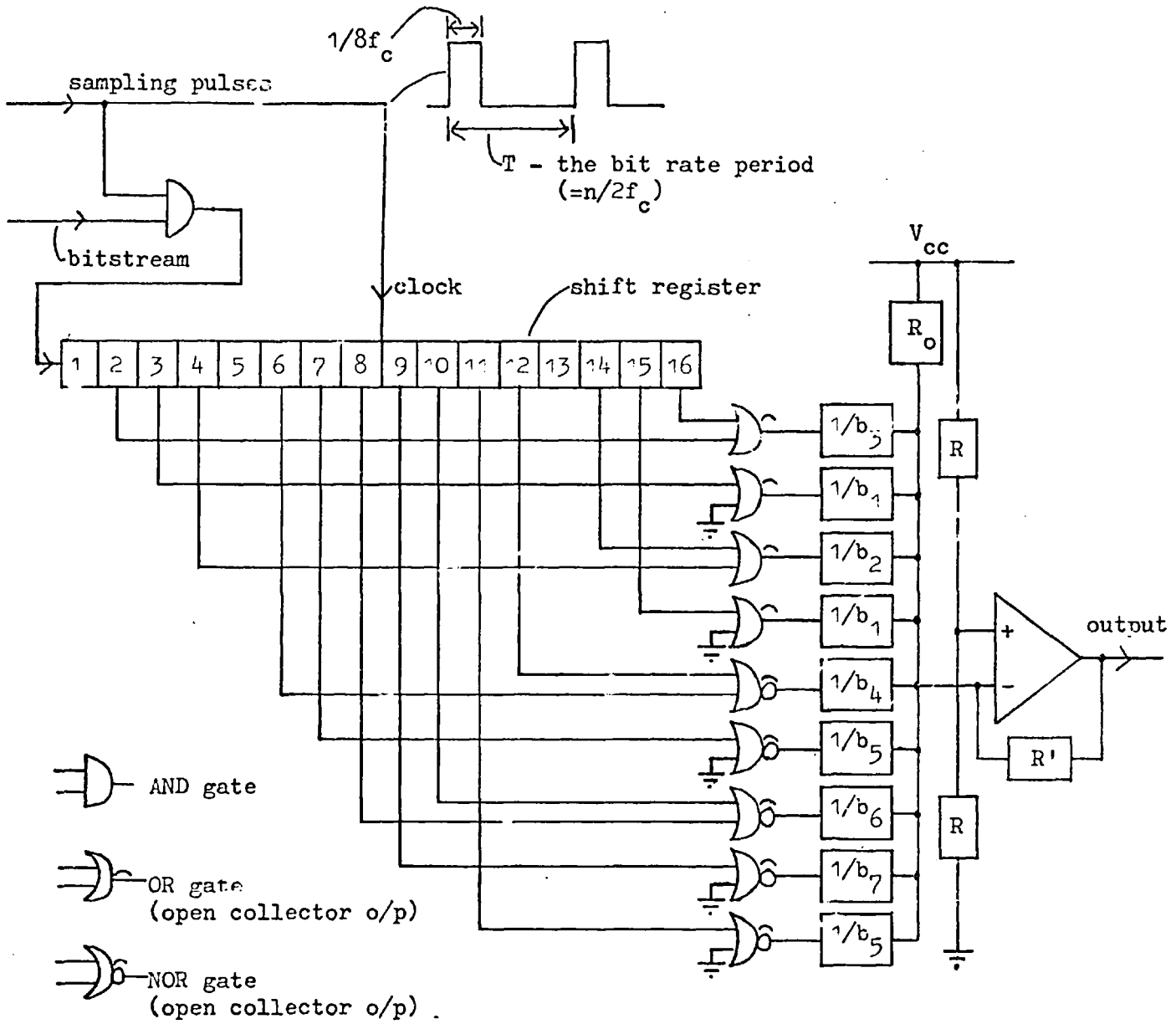
Fig 3-2.14   Table of bit-rate values

(a) Pulse stream having 100% roll-off raised-cosine spectrum

Bit rate (kb/s)	Bit-rate reduction factor 1/N
2.4	x1
1.2	x1/2
0.6	x1/4
0.4	x1/6

(b) Pulse stream having 3½% roll-off raised-cosine spectrum

Bit rate (kb/s)	Bit-rate reduction factor 1/N
3.6	x1
1.8	x1/2
0.9	x1/4
0.6	x1/6

Fig 3-3.1   General modulating signal f(t)
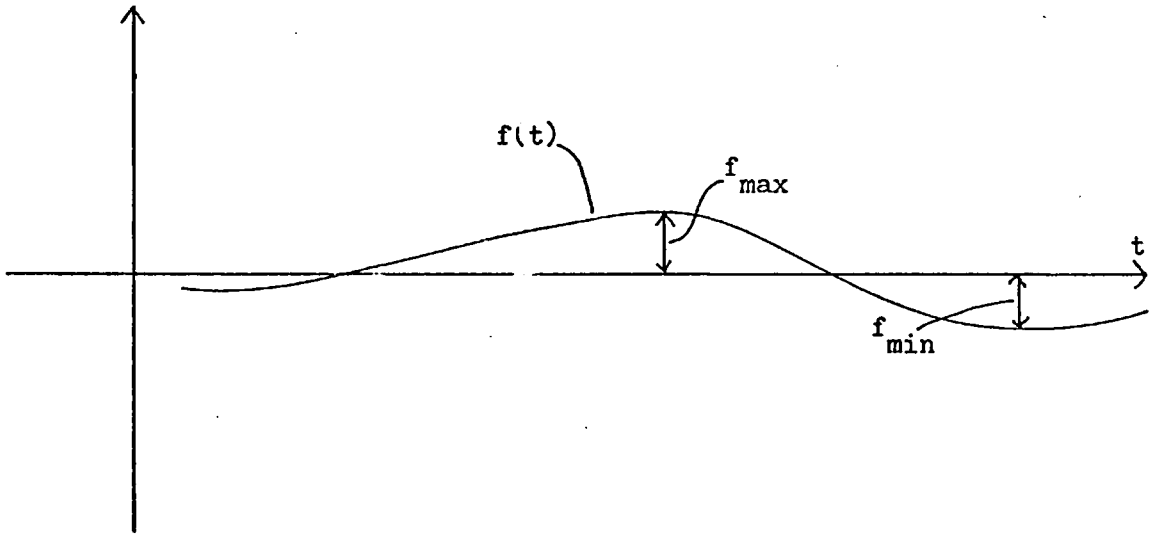
Fig 3-3.2   f(t) amplitude modulated - carrier added
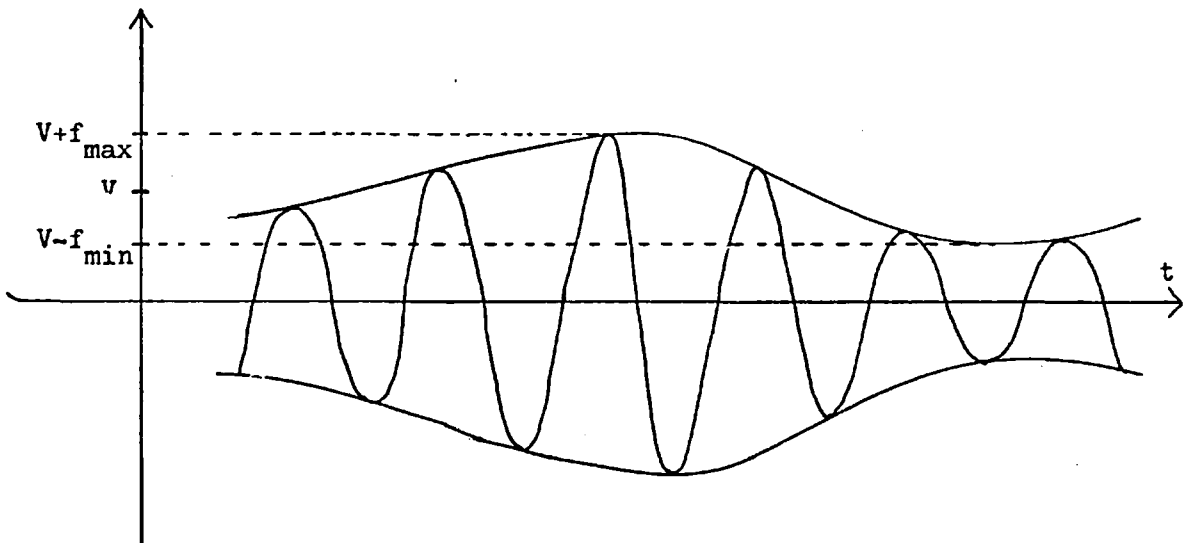
Fig 3-3.3   Pulse-stream diagrams showing the modulation modes

x1 bit-rate reduction factor

x1/2 bit-rate reduction factor

x1/4 bit-rate reduction factor
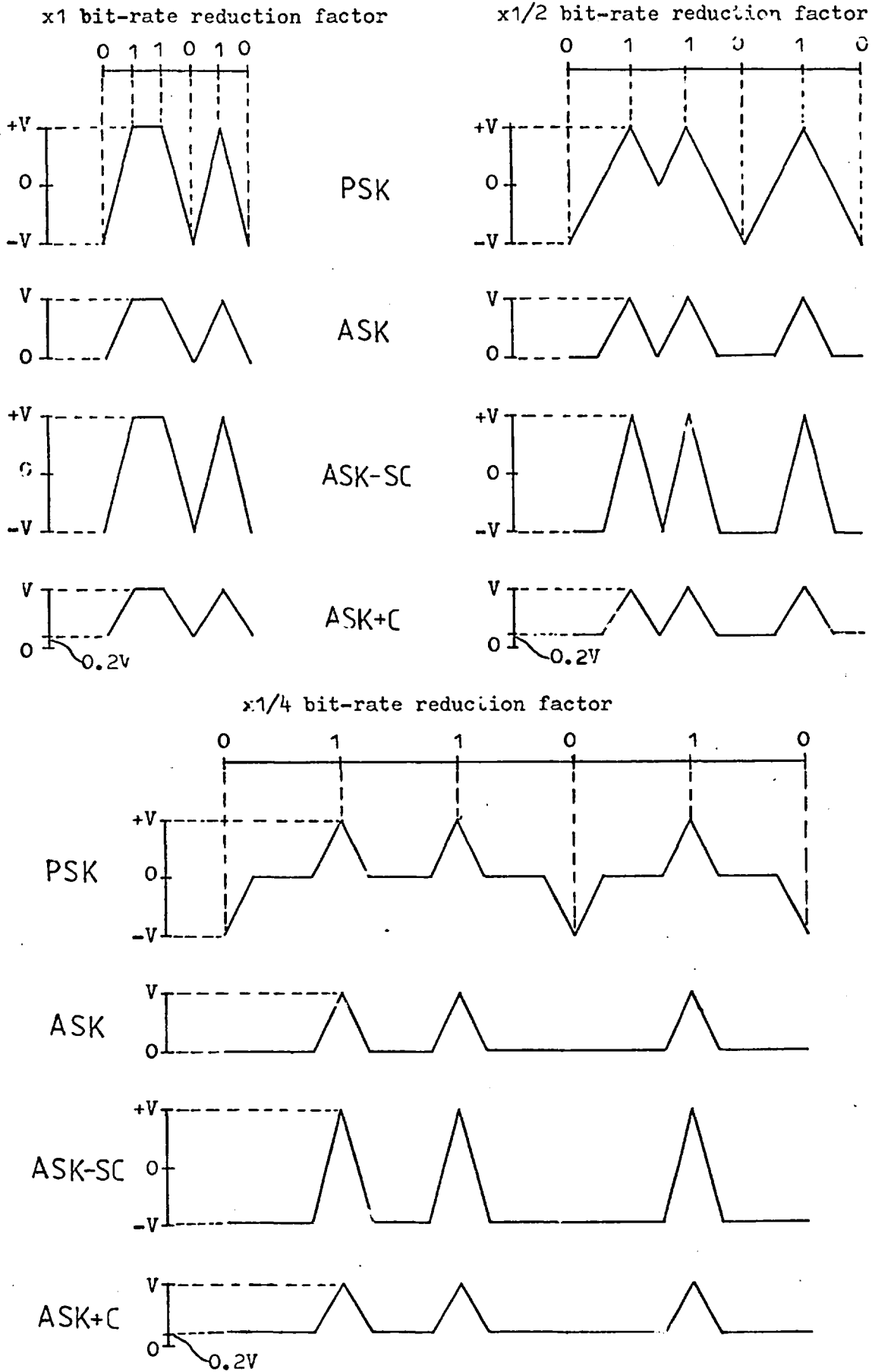
Fig 3-3.4   Pulse-stream diagrams illustrating the set-up pulse
(bit-rate reduction factor x1/2)
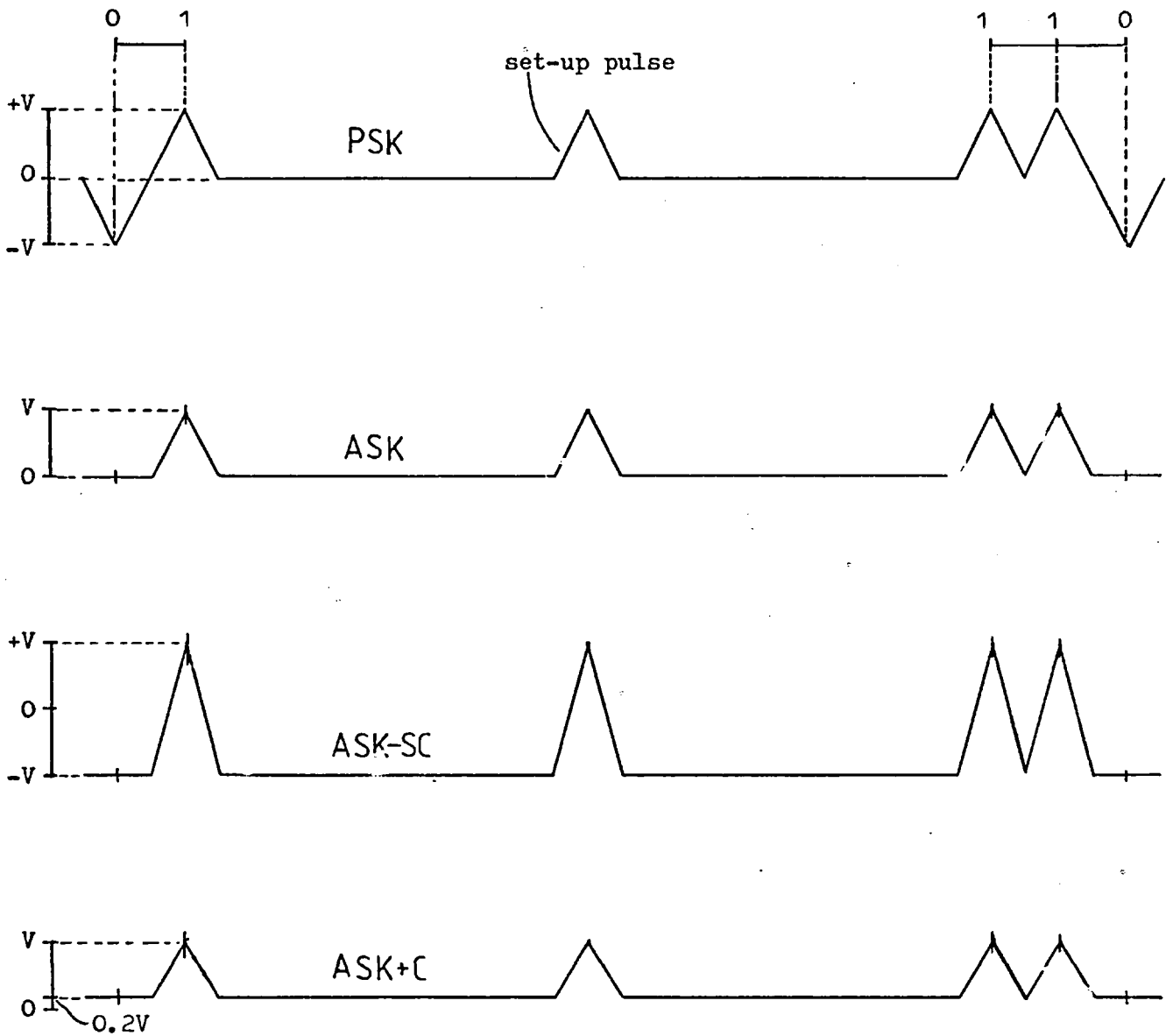
Fig 3-3.5   Receiver-operation considerations (1) to (4)

Mode	(1) Eye opening level	(3) Residual dc content	(4) Phase of mod. carrier
PSK	2V	No	2-phase
ASK	V	No	2-phase
ASK-SC	2V	Yes	2-phase
ASK+C	0.8V	Yes	1-phase

(2) Variation with bit rate of carrier-envelope level

Average voltage of carrier envelope

Bit-rate reduction factor
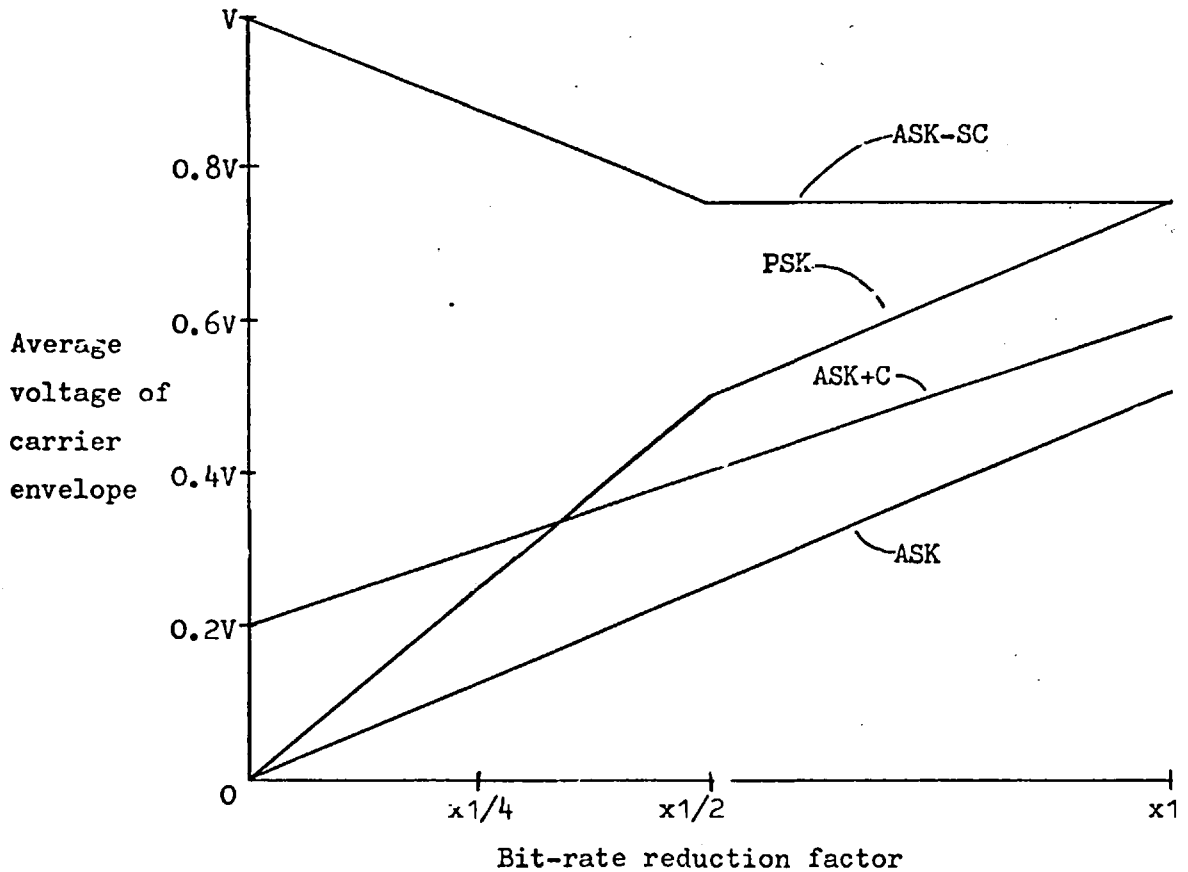
V is the average voltage of the unmodulated carrier envelope.
This graph assumes an equal density of "1"s and "0"s.

Fig 3-3.6   Phasor diagram for FM signal in two-path propagation
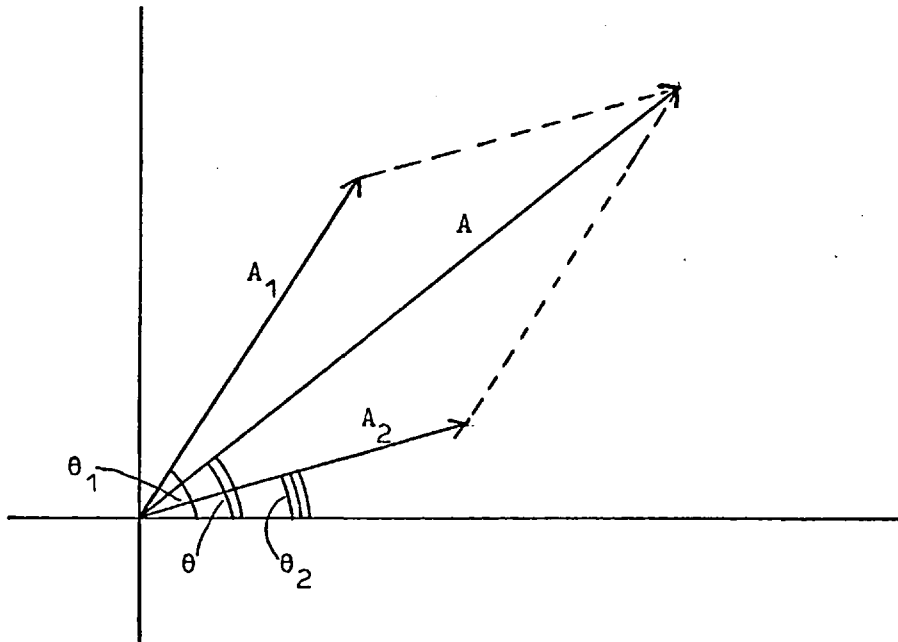
Fig 3-3.7   FM receiver system
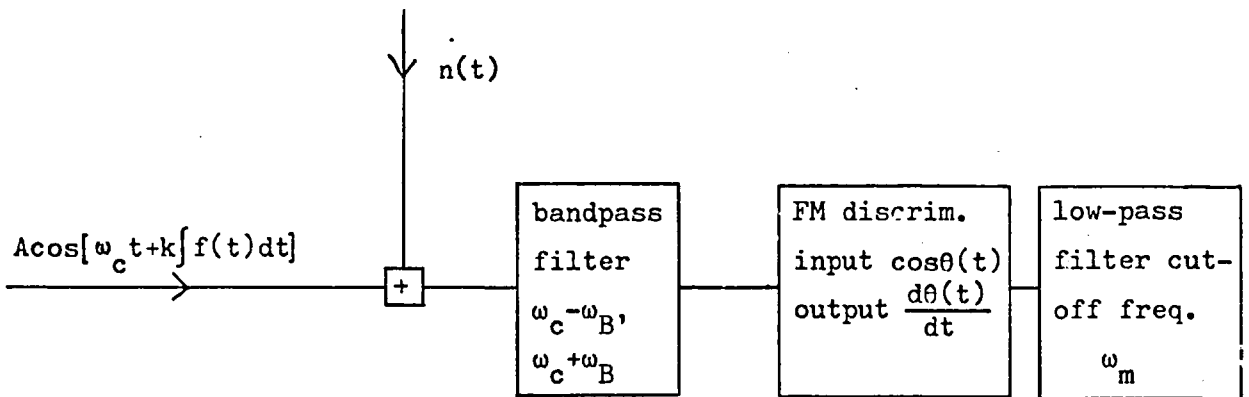
Fig 3-3.8   Power density spectrum at FM discriminator output
             when input is bandpass gaussian noise

Fig 3-3.9   <u>AM receiver</u>

$n(t)$

$f(t)\cos\omega_c t + n_c(t)\cos\omega_c t + n_s(t)\sin\omega_c t$

$f(t)\cos\omega_c t$

$+$

bandpass
filter -
$\omega_c-\omega_m$,
$\omega_c+\omega_m$

coherent
detector

$f(t)+n_c(t)$
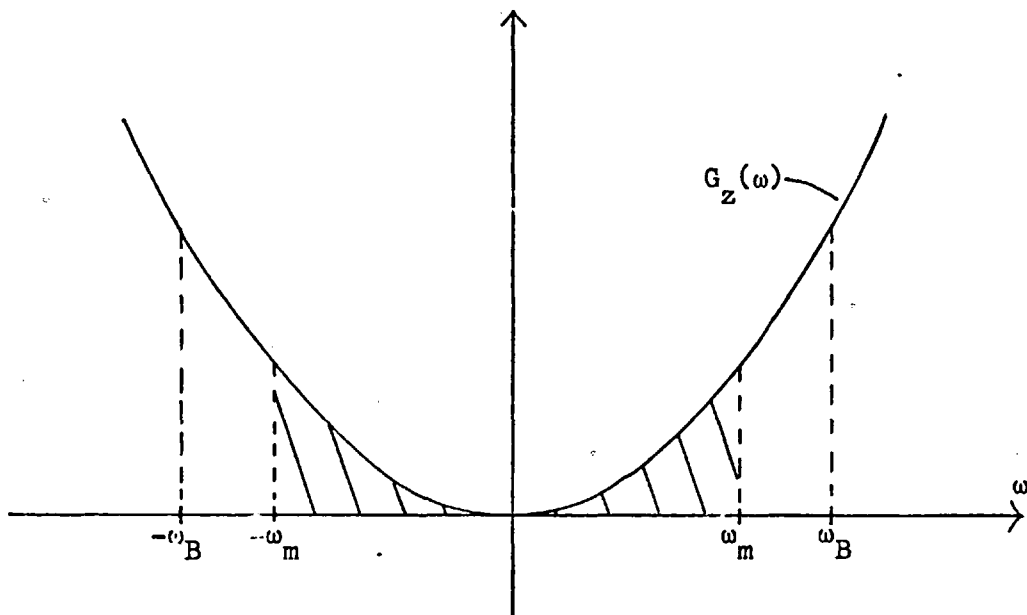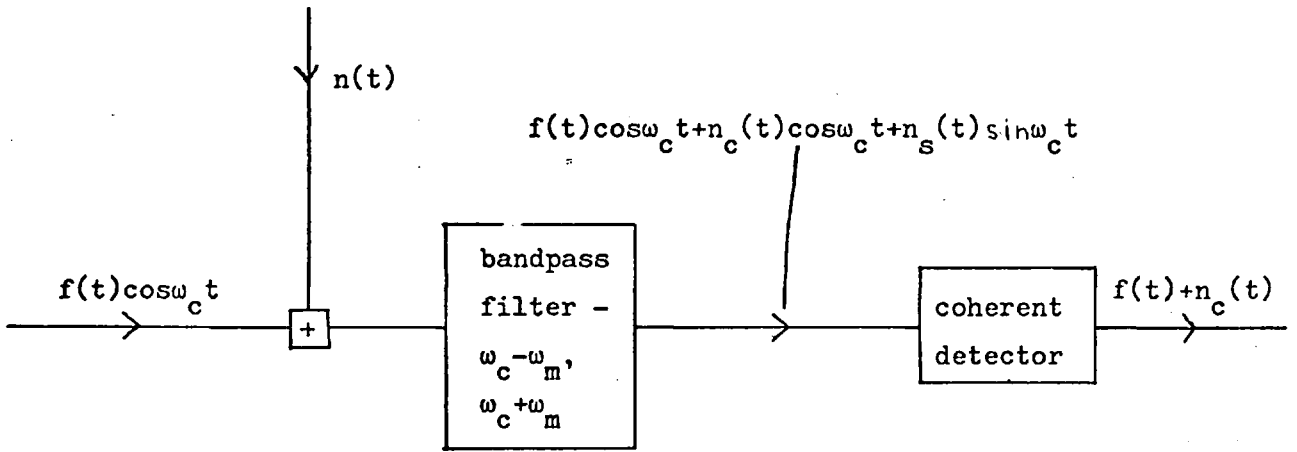
Fig 3-3.10   <u>FM signal spectrum using a square pulse as the modulating signal</u>

$-\omega_i$

$\omega=\omega_c$

ω_i

$4\omega_m$

ω
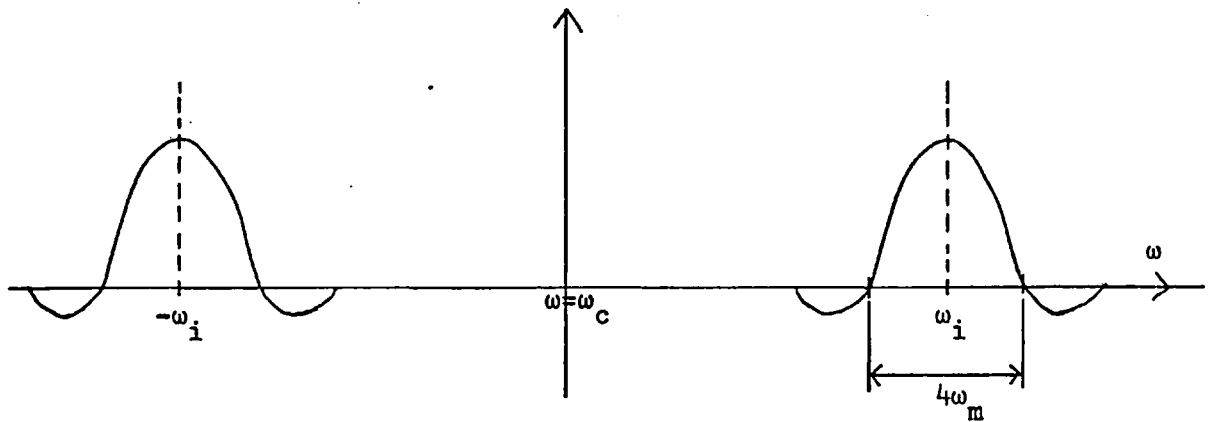
Fig 3-7.1 <u>Variation of ASK and PSK error probabilities with</u>
<u>signal-to-noise ratio</u>
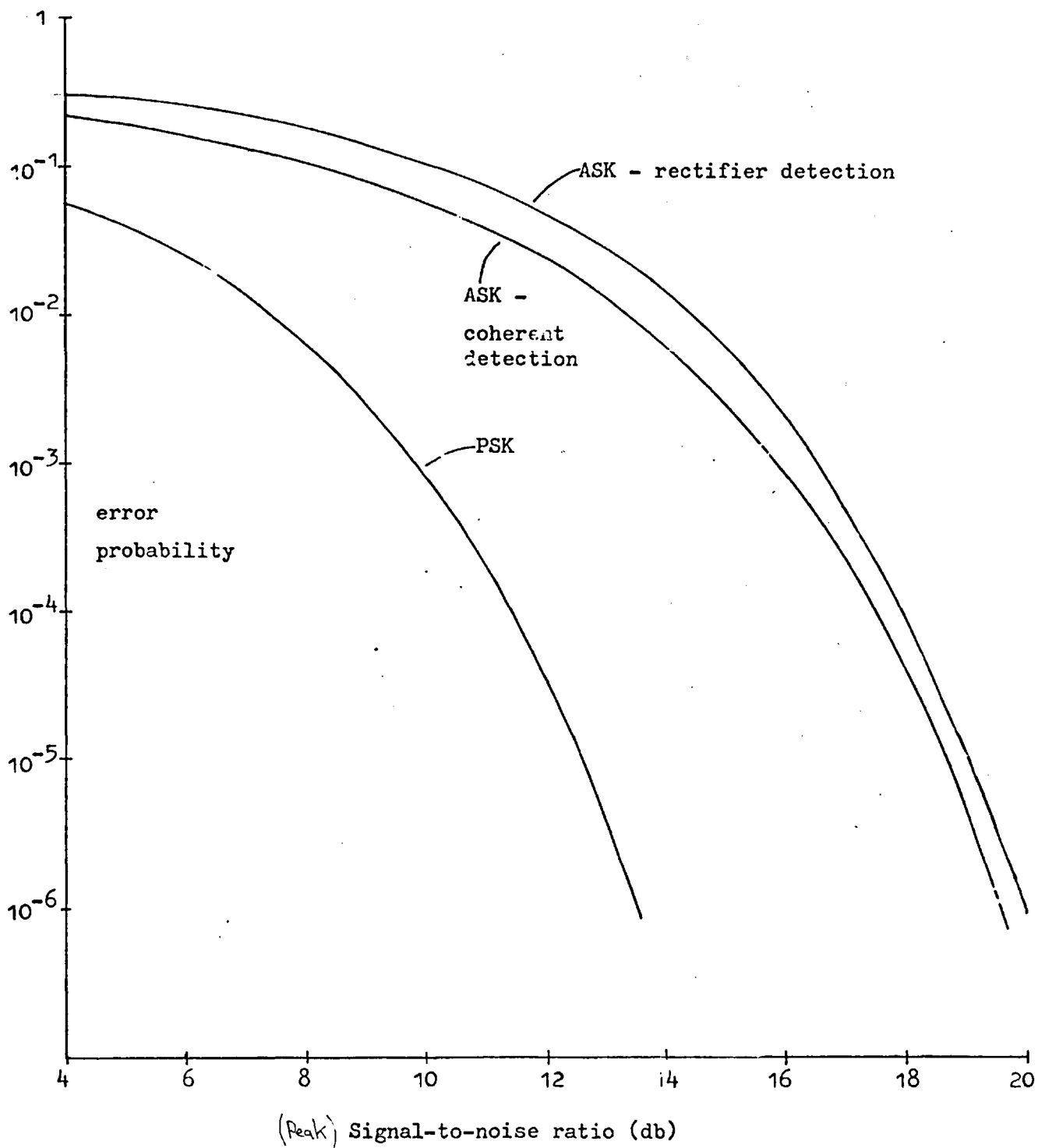
Fig 3-7.2   The phase-locked loop

Carrier

$Acos(\omega_c t+\theta_i(t))$ → [Hard limiter] → X_{HL} → [Monostable] → X_M

[Phase comparator] → I_{PC} → [Loop filter] → V_{LF}

X_{VCO}, phase = $\omega_c t+\theta_o(t)$

[VCO]

Waveforms

Carrier

X_{HL}

τ

X_M
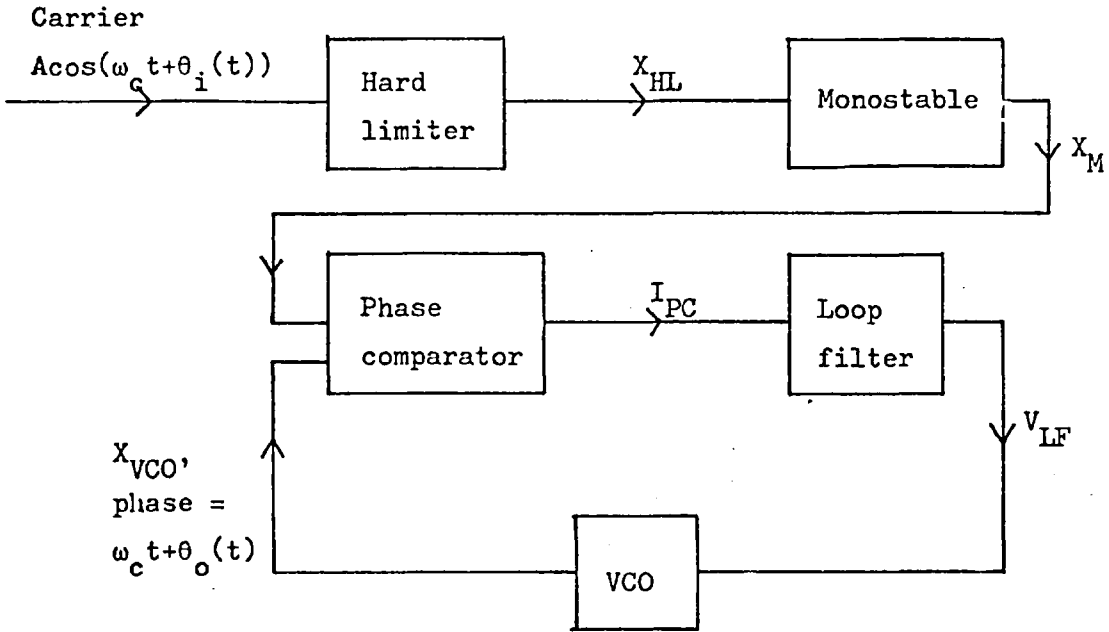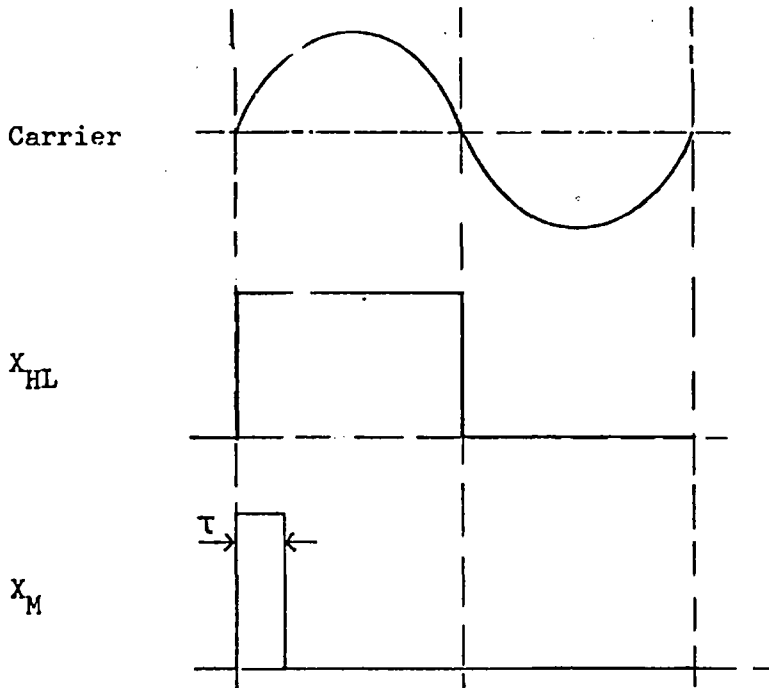
Fig 3-7.3   <u>PLL phase comparator waveforms</u>

(a) PLL out of lock(1)
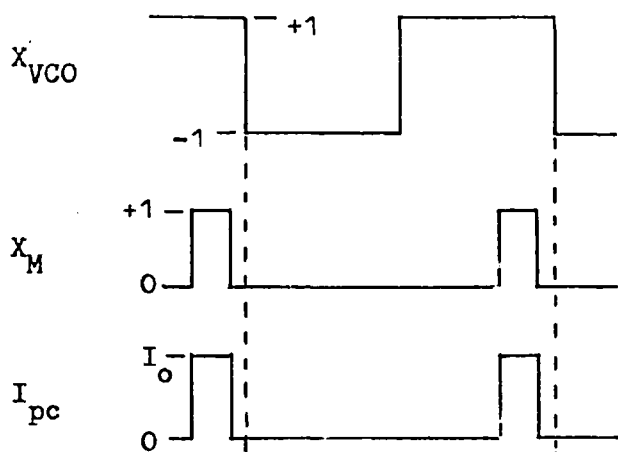
X_{VCO} is shifted to the left by phase comparator feedback to VCO

$$I_{pc} = I_o \times X_{VCO} \times X_M$$

(b) PLL out of lock(2)
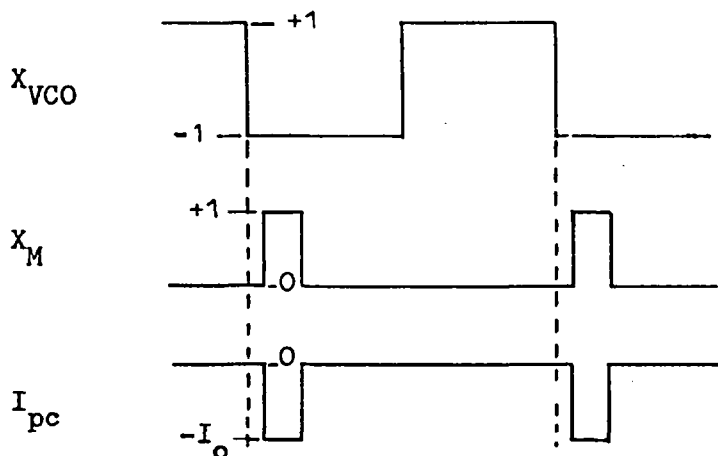
X_{VCO} is shifted to the right by phase comparator feedback

(c) PLL in lock
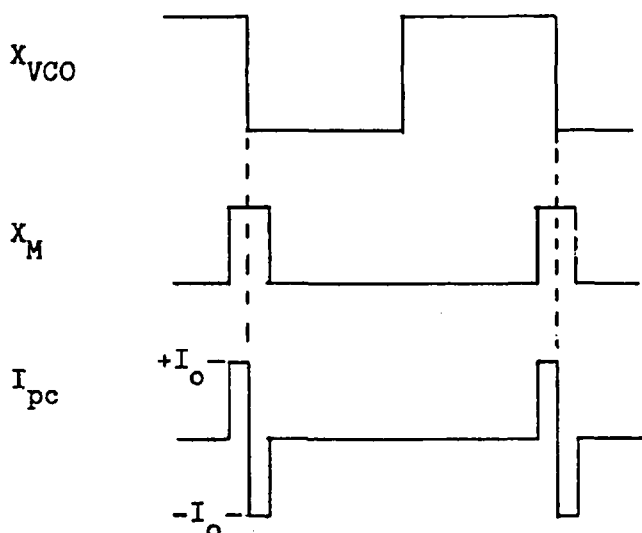
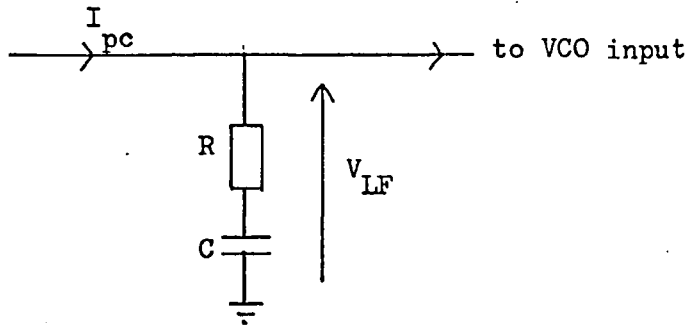Fig 3-7.4   PLL loop filter
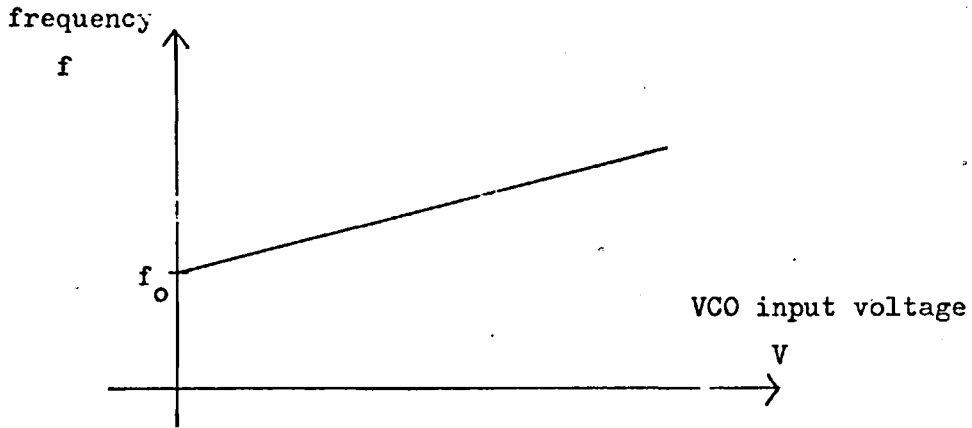

Fig 3-7.5   VCO voltage-frequency relationship

Fig 3-7.6   VCO input and output waveforms
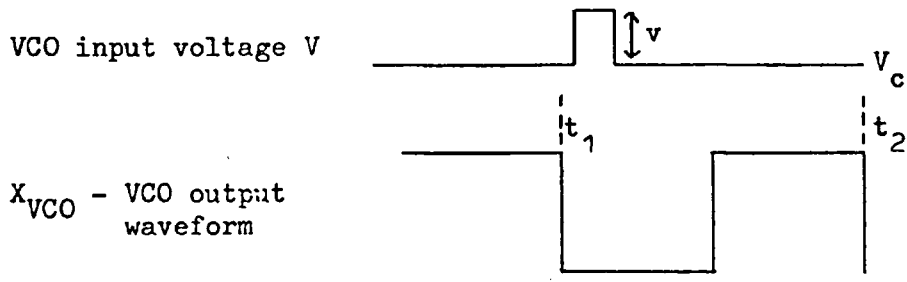
Fig 3-7.7   Variation of PLL phase increment $\Delta\theta$ with input phase θ_i
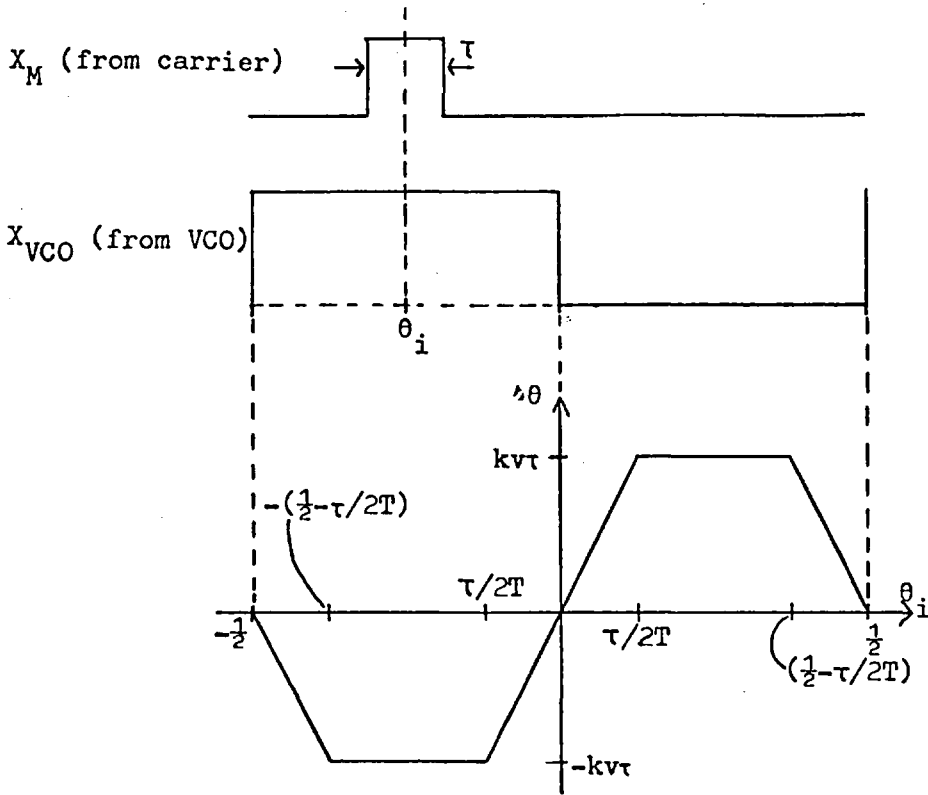

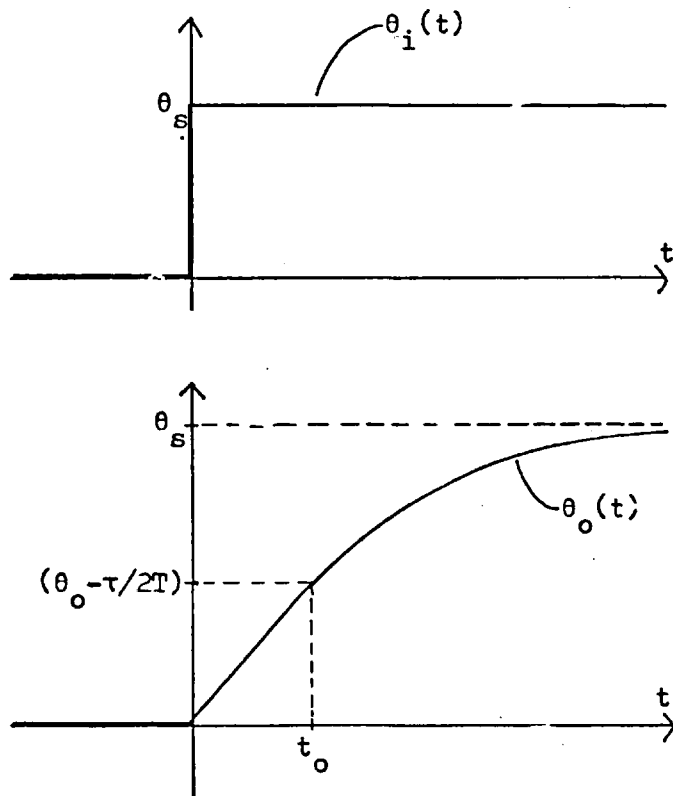Fig 3-7.8   Step response of PLL

Fig 3-7.9   Delta-modulator equivalent circuit of the PLL
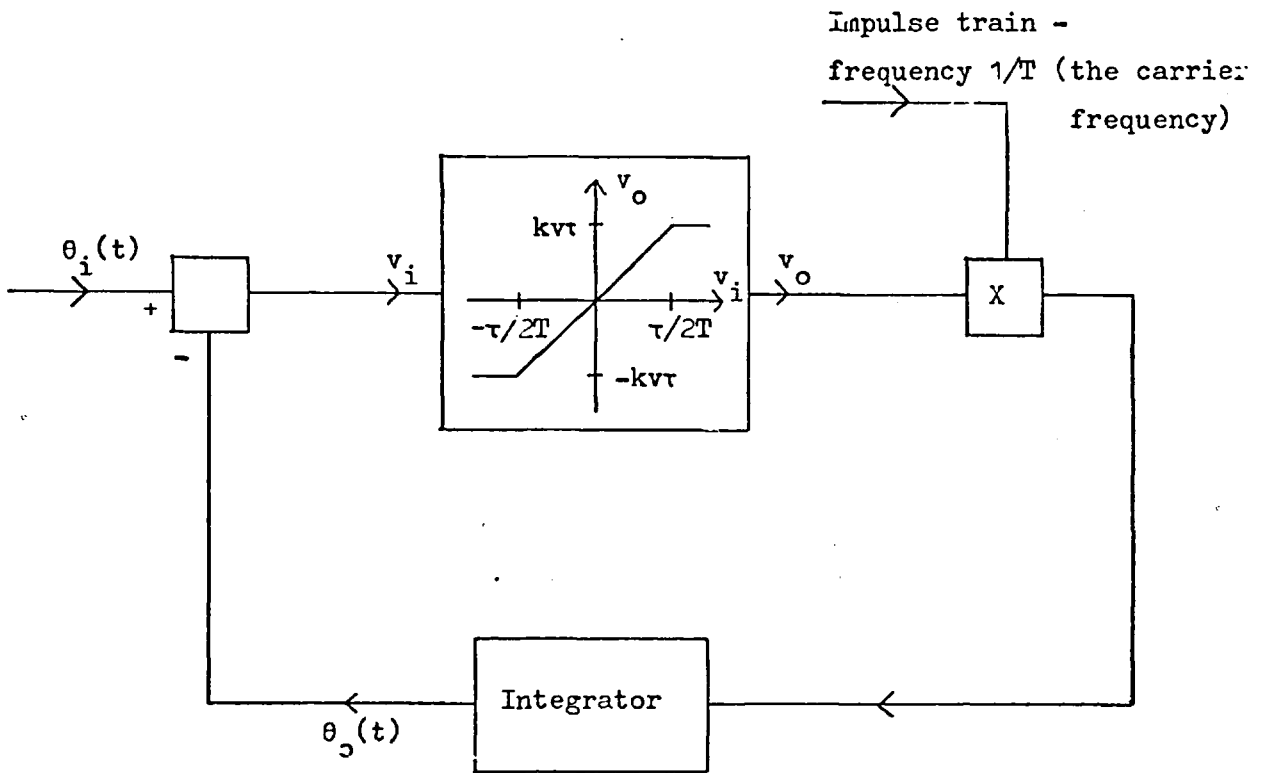
Fig 3-7.10   Test circuit for determining the performance of the PLL
in the presence of noise

Fig 3-7.11   <u>PLL performance in the presence of noise</u>
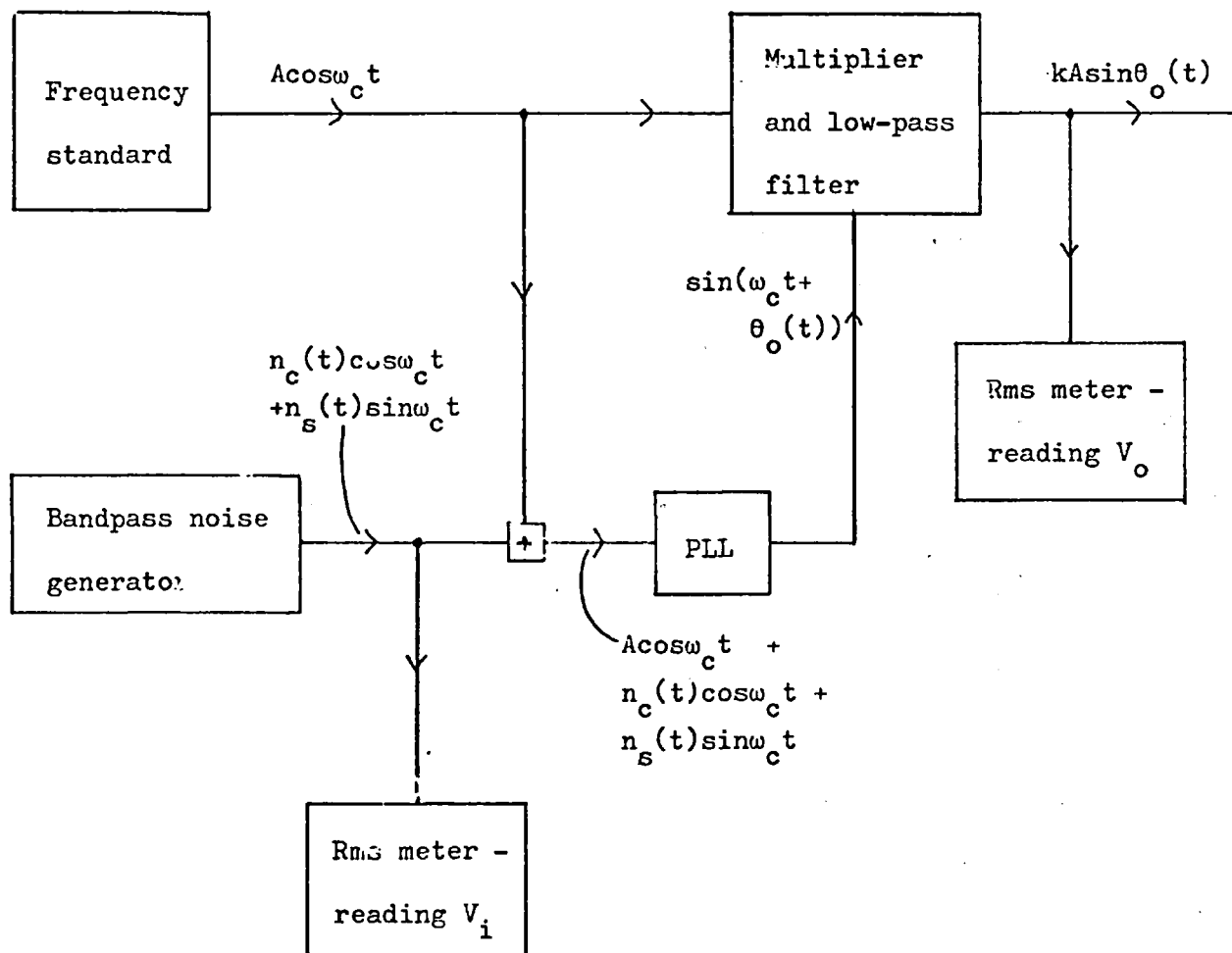

Fig 3-7.11   <u>PLL performance in the presence of noise</u>
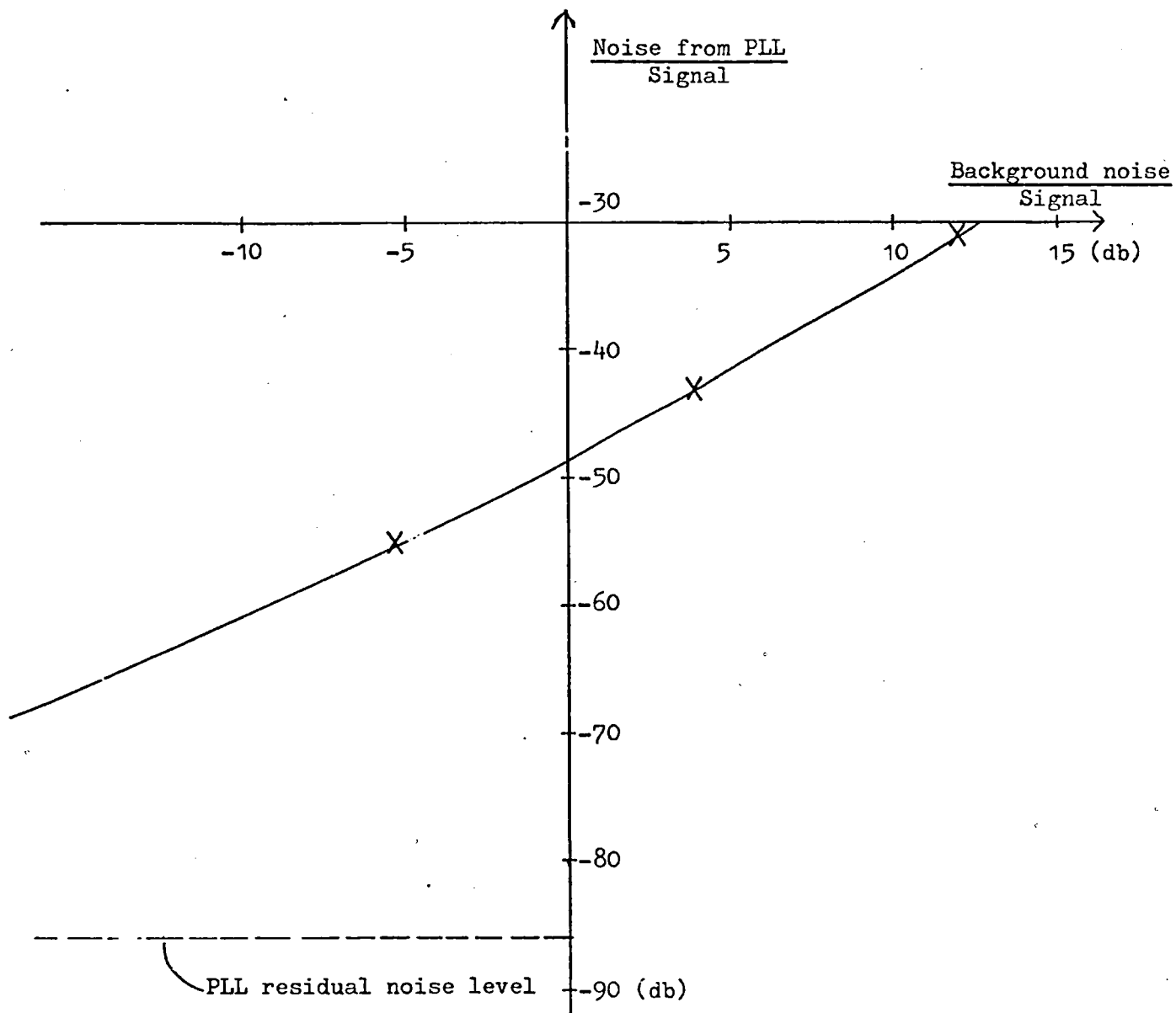
Fig 3-7.12   Waveforms of PLL when stable in 0 and π modes
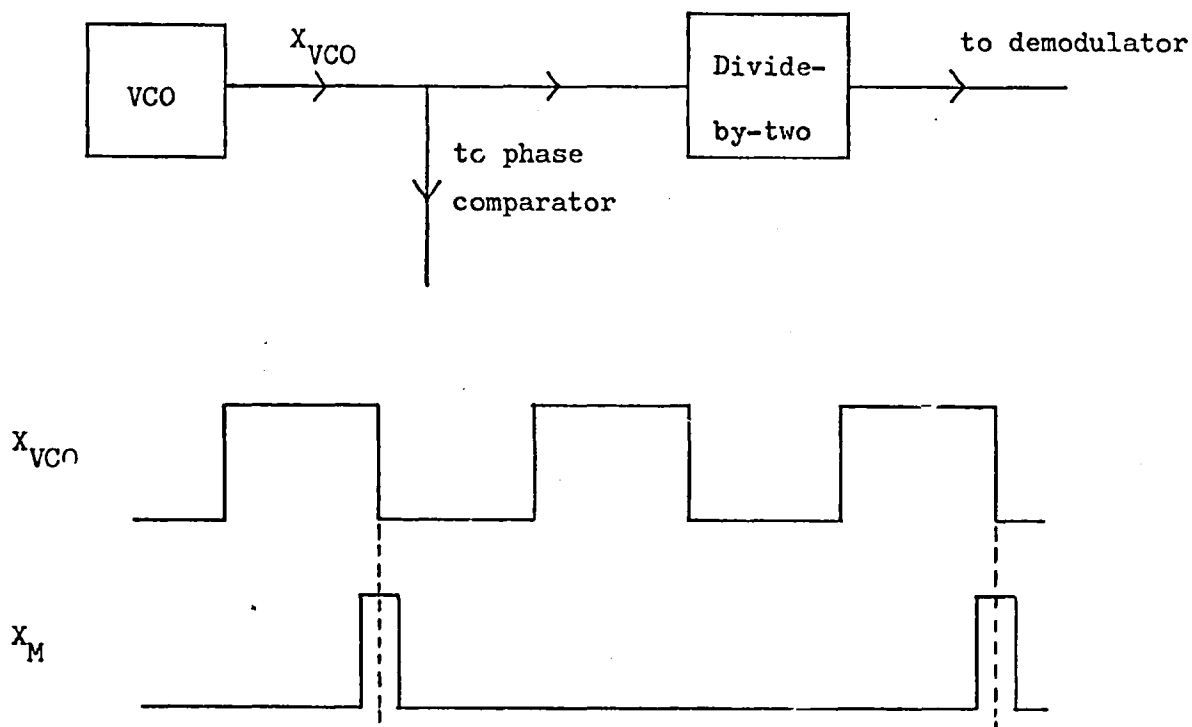
Fig 3-7.13   Pdfs of rectified carrier— plus— bandpass noise
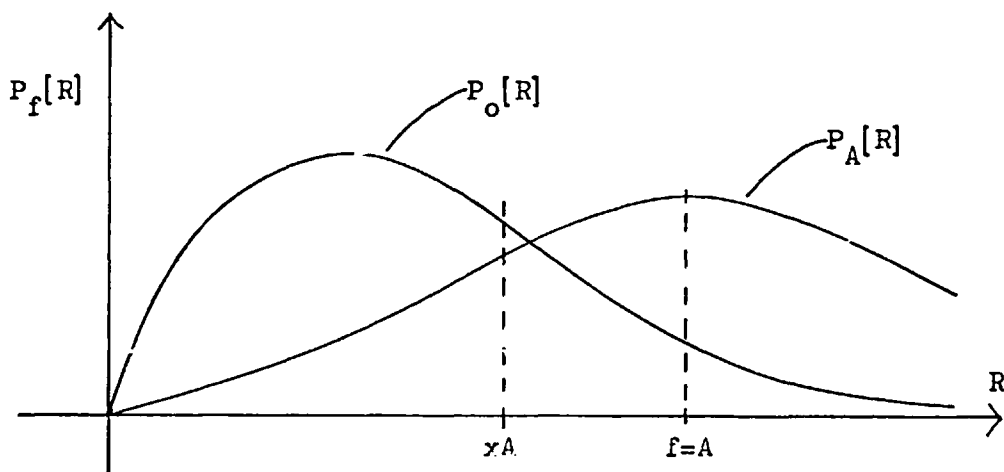
Fig 3-8.1 <u>Matched-filter correlator detector</u>

s(t), the transmitted pulse

r(t), the correlating waveform

r(t).s(t)

f(t), the correlator output
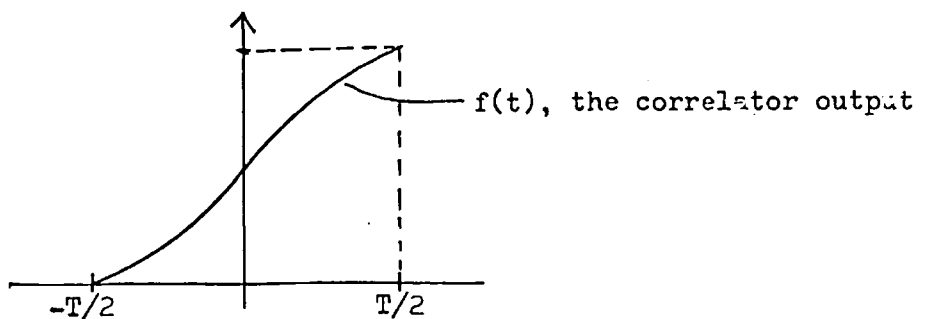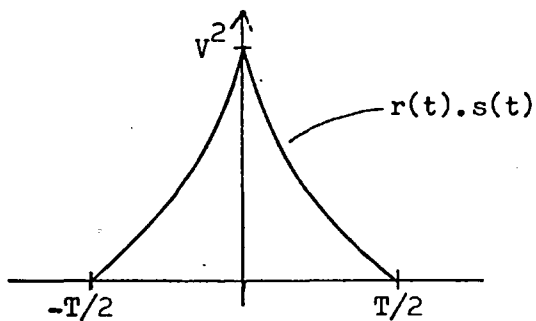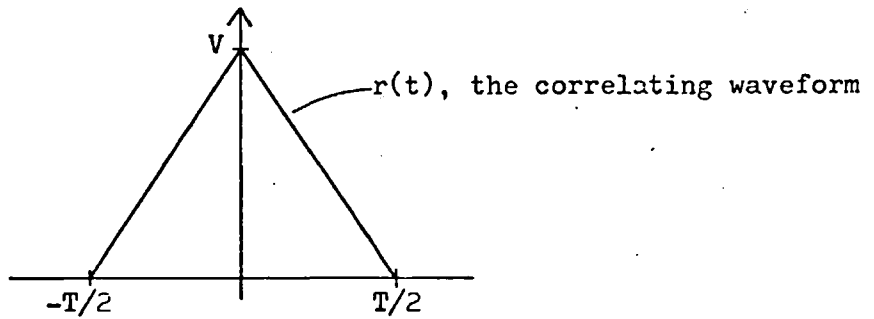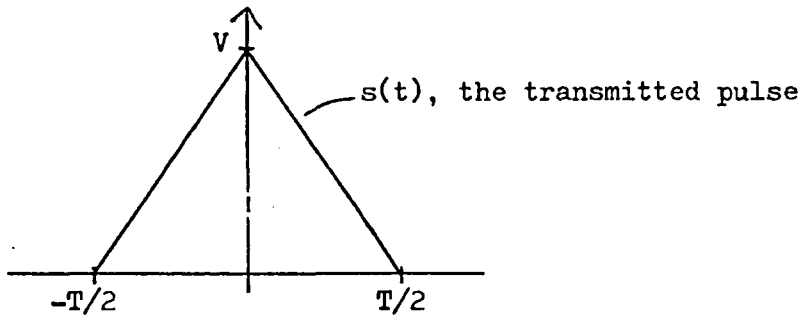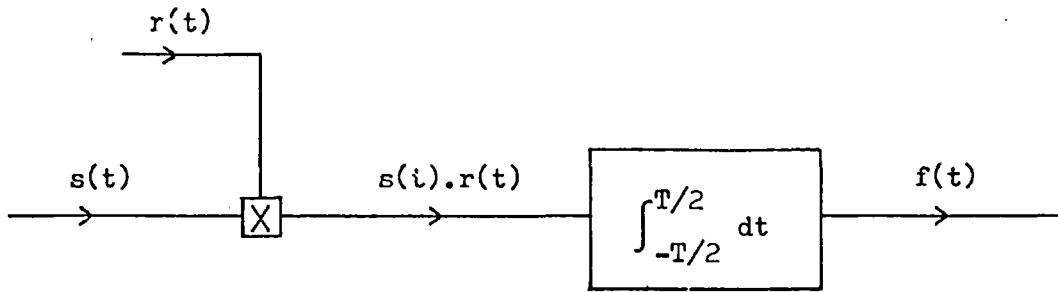
Fig 3-8.2   Multipath example for matched filter evaluation
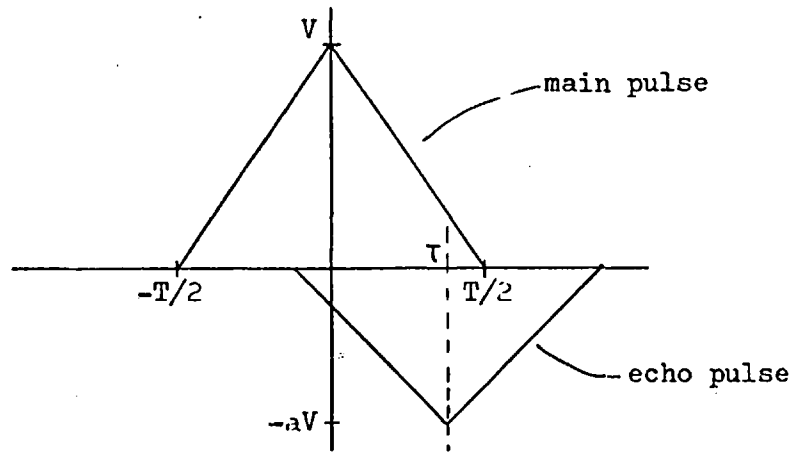
Fig 3-8.3 Effective signal-to-noise ratio at output of matched filter
         for two-multipath example
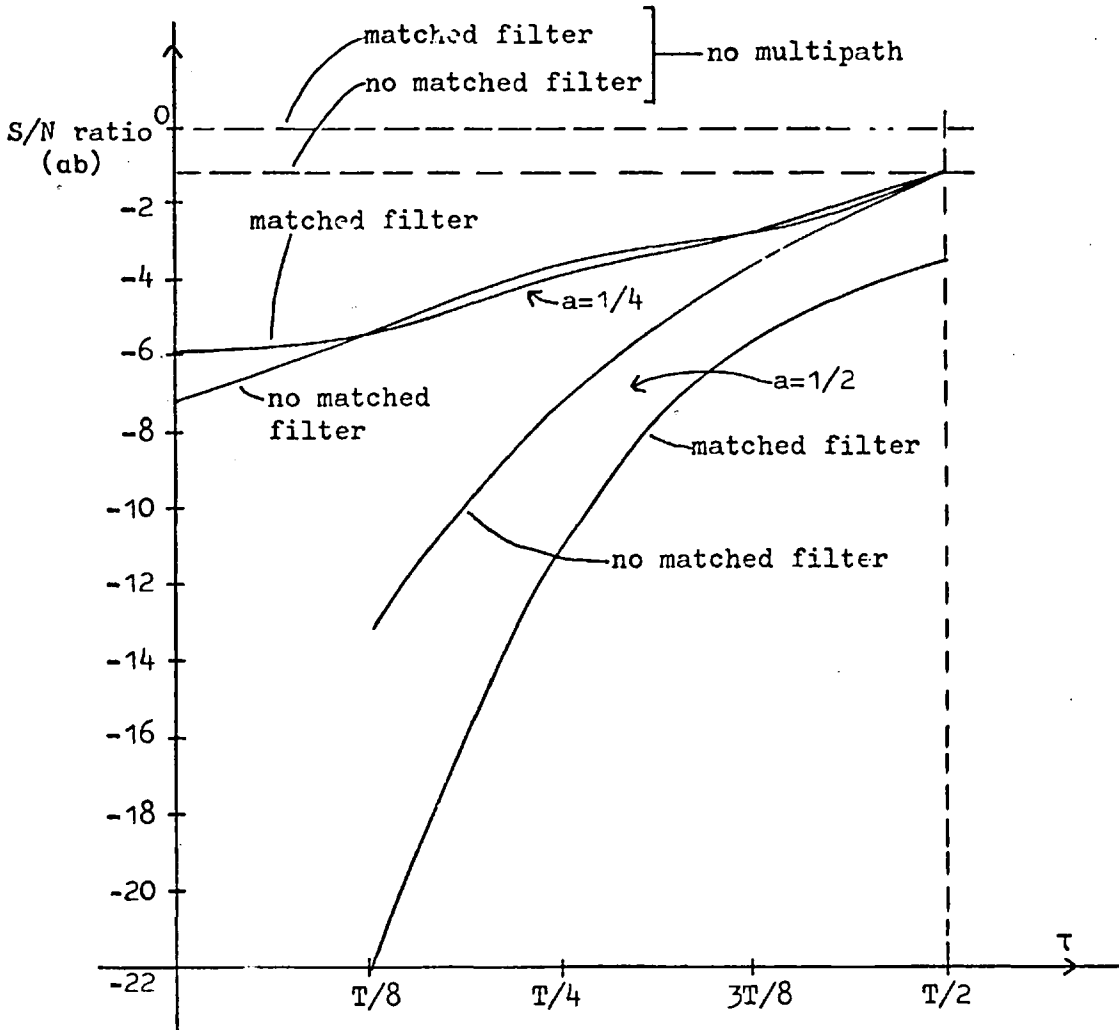
Fig 3-8.4   <u>Positioning of pre-emphasis, de-emphasis filters</u>

noise

signal → | Pre-emphasis filter, $H_1(f)$ | → + [ ] (Channel) → | De-emphasis filter, $H_2(f)$ | →
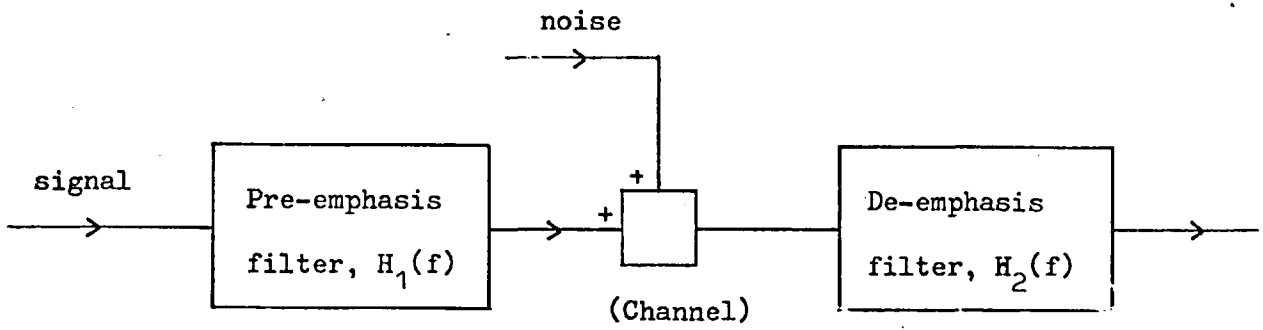
Fig 3-8.5   <u>Required pre-emphasis filter response</u>
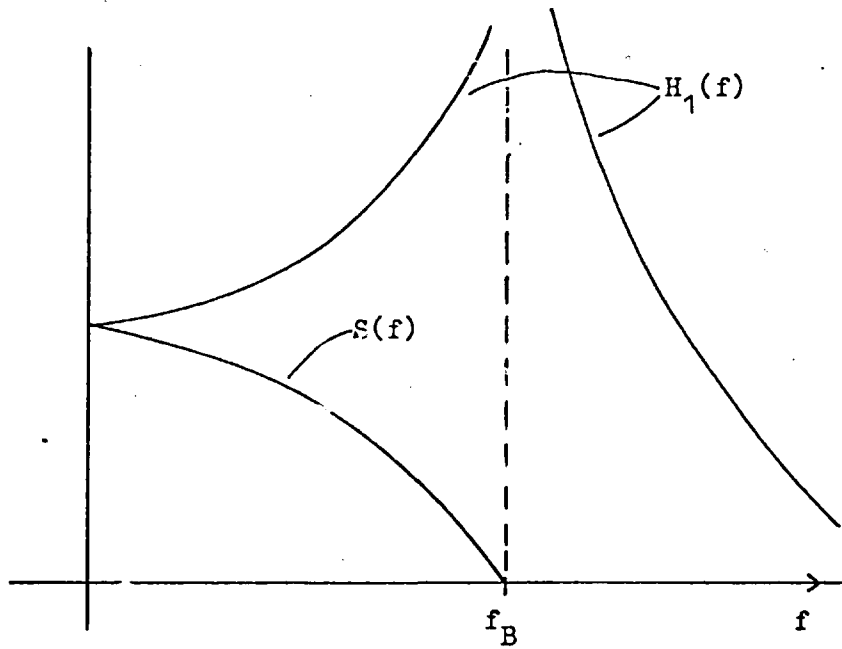
Fig 3-8.6 <u>Pre-emphasis, de-emphasis filter responses together</u>
          <u>with the pulse spectrum shapes</u>
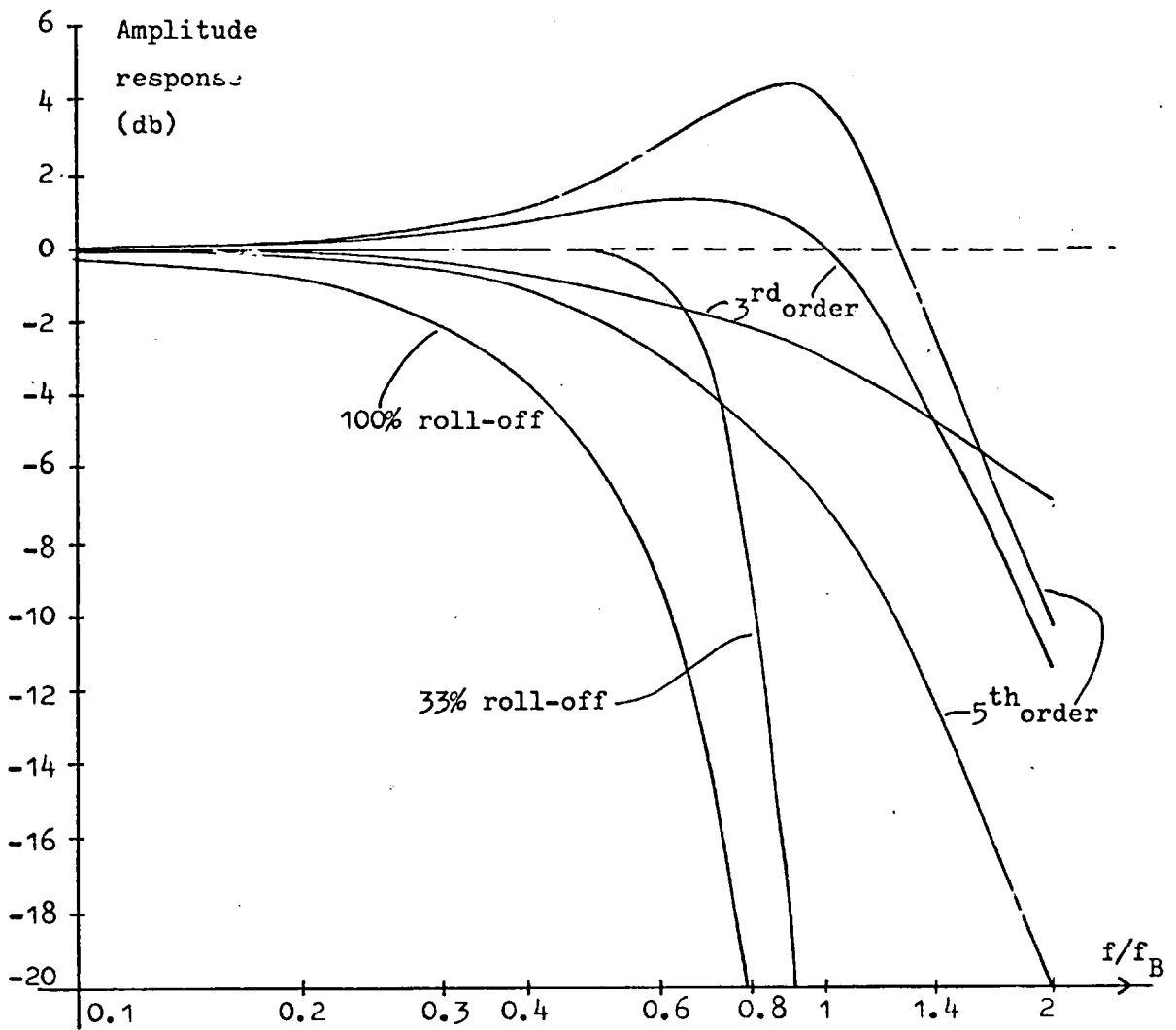
Fig 3-8.7 <u>Experimental system used to assess the performance of the</u>
<u>postulated pre-emphasis, de-emphasis filter shapes</u>

Fig 3-8.8   Results of pre-emphasis, de-emphasis filters - 33% roll-off

Fig 3-8.8   Results of pre-emphasis, de-emphasis filters - 33% roll-off
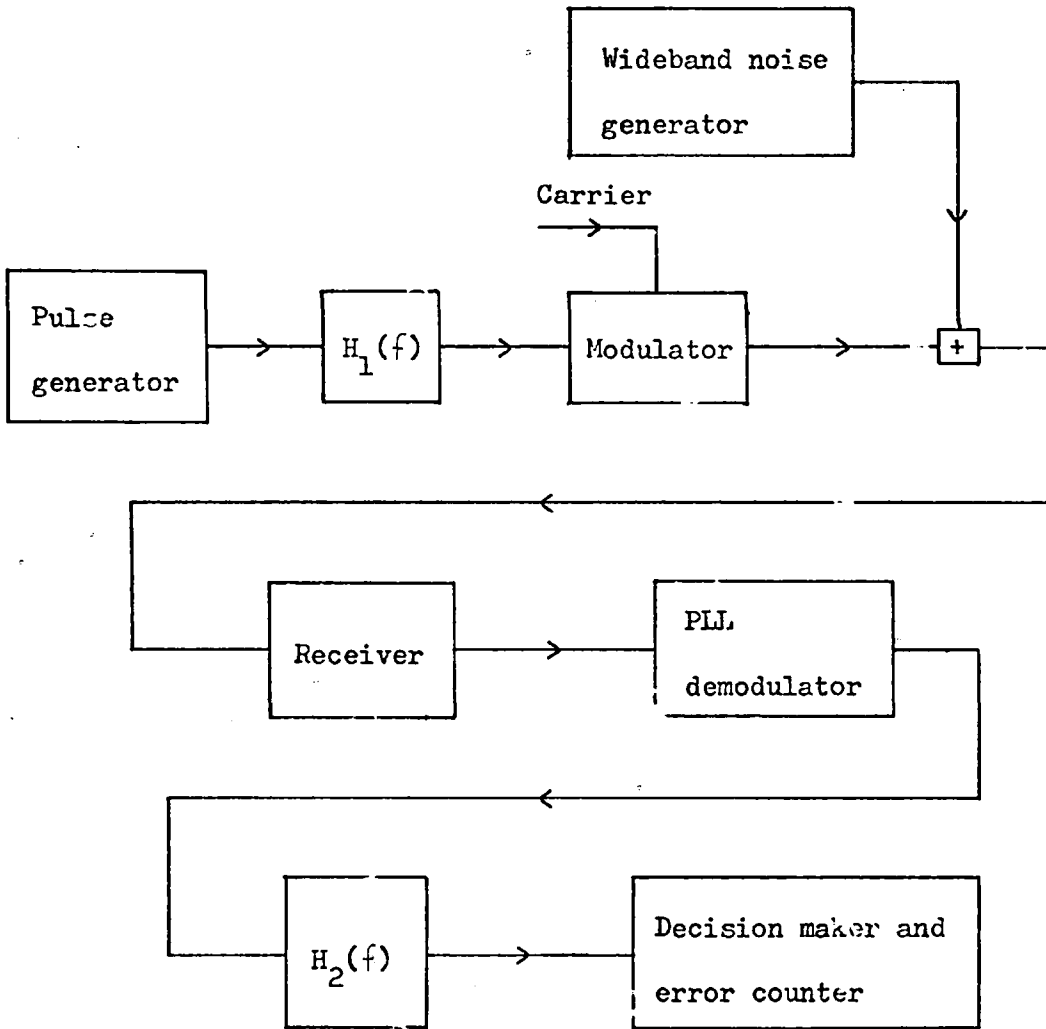
l - low noise

m - medium noise
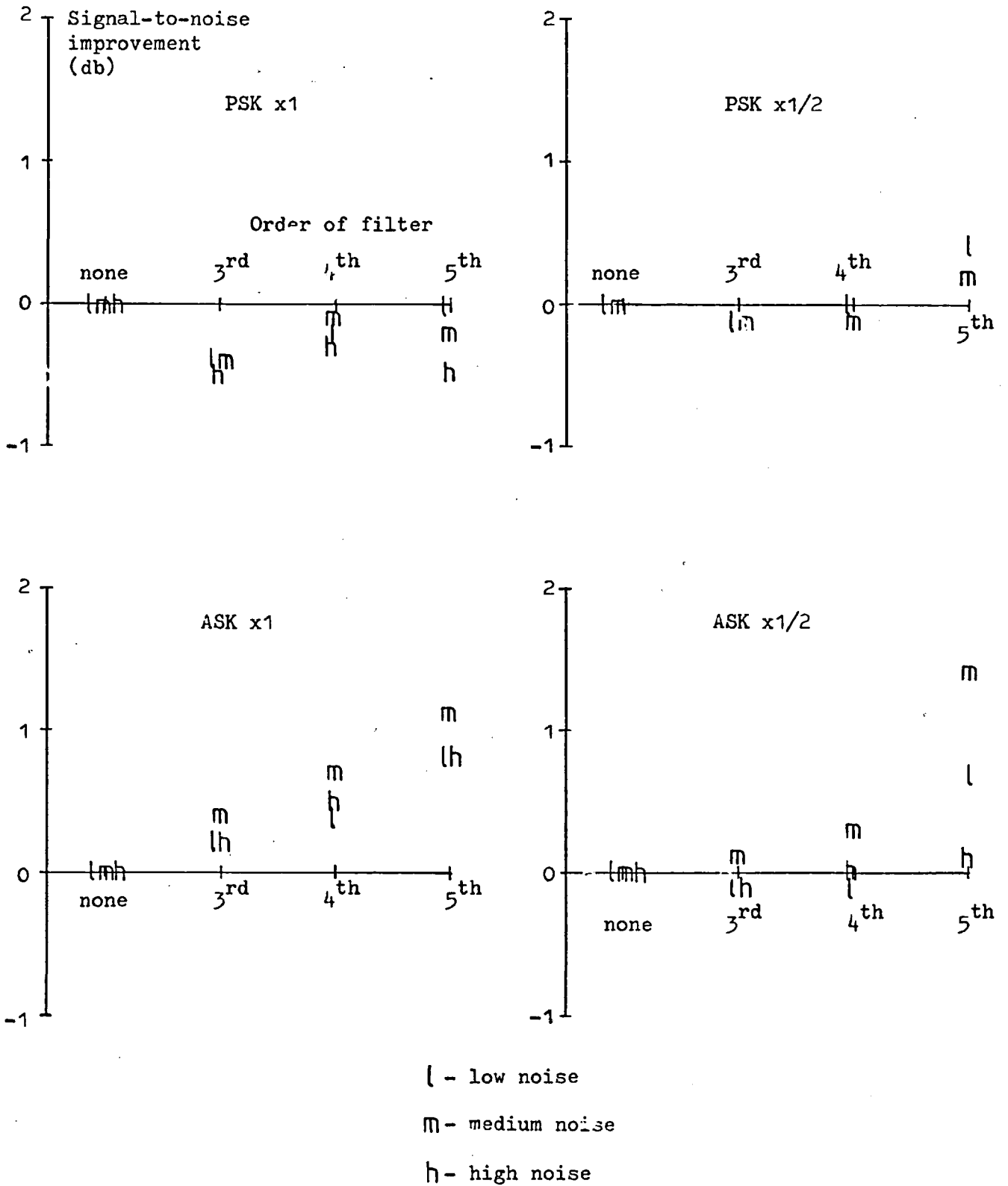
h - high noise

Fig 3-8.9 <u>Results of pre-emphasis, de-emphasis filters - 100% roll-off</u>
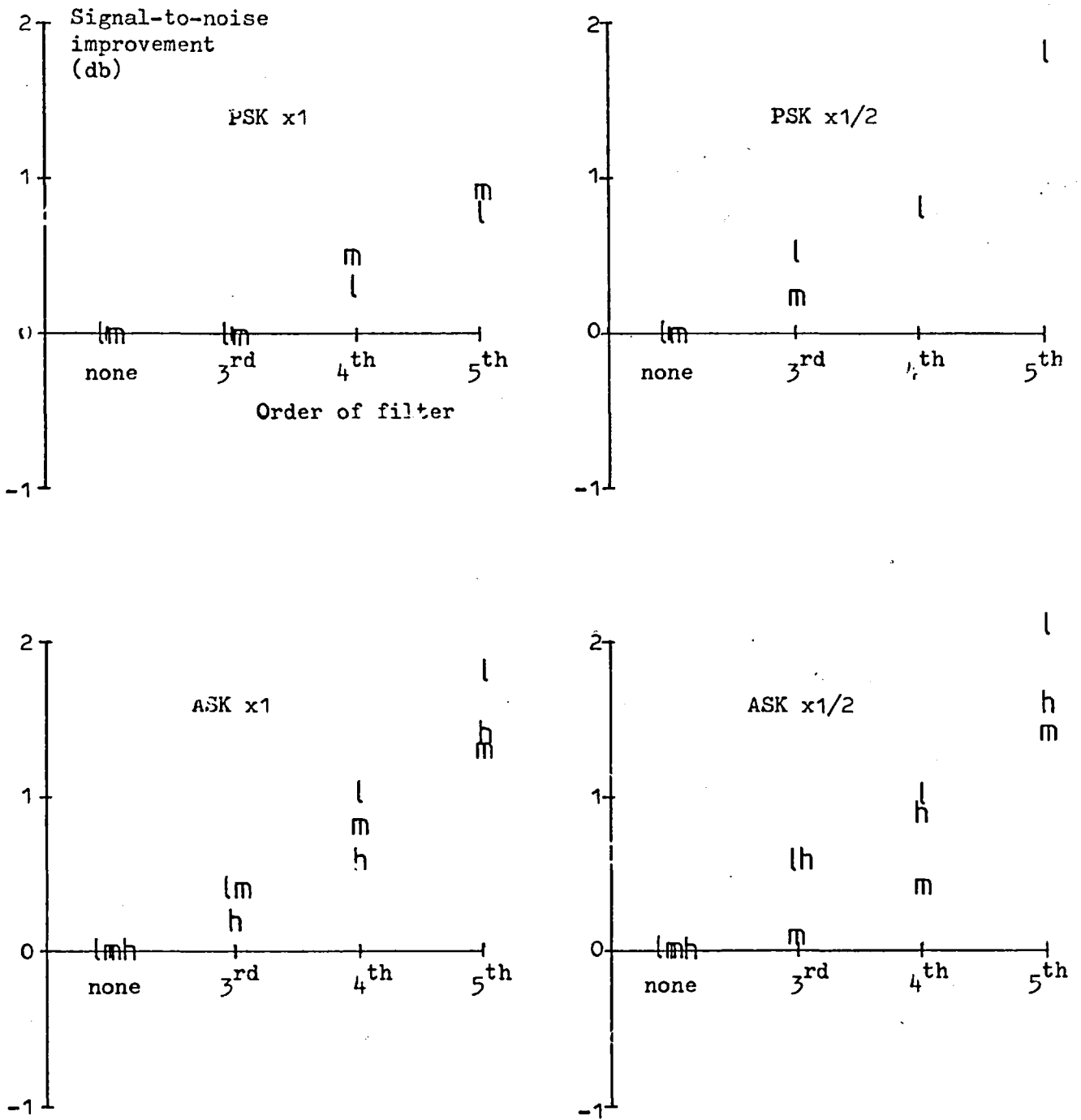
l - low noise

m - medium noise

h - high noise

Fig 3-9.1   The level detector

signal
being
detected → [Sample and hold] → [Low-pass filter - cut-off freq. f_c] → amplitude of the level

trigger - frequency f_s
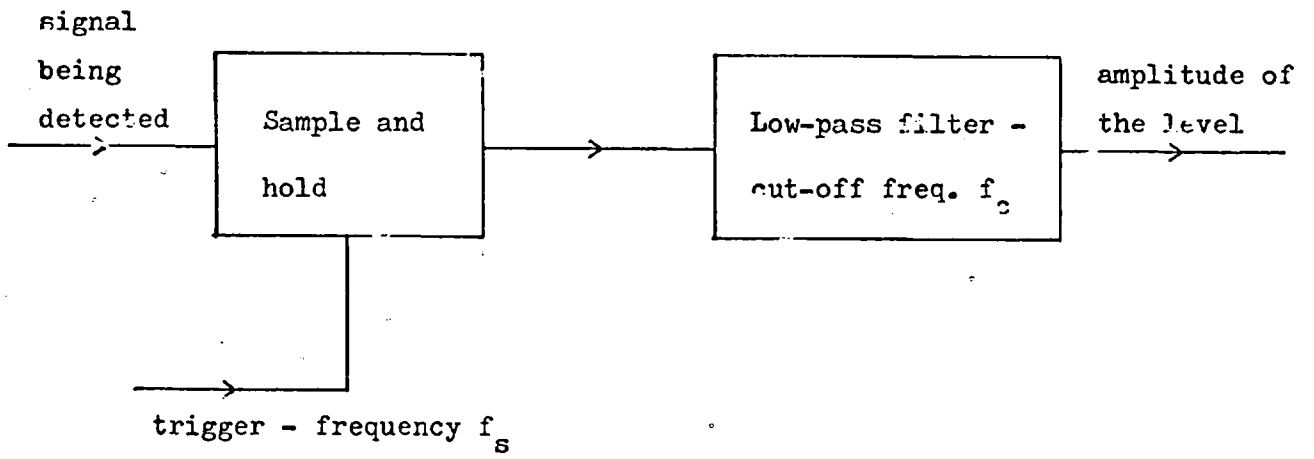
Fig 3-10.1   Positioning of bit-synchroniser window with respect
               to the set-up pulse
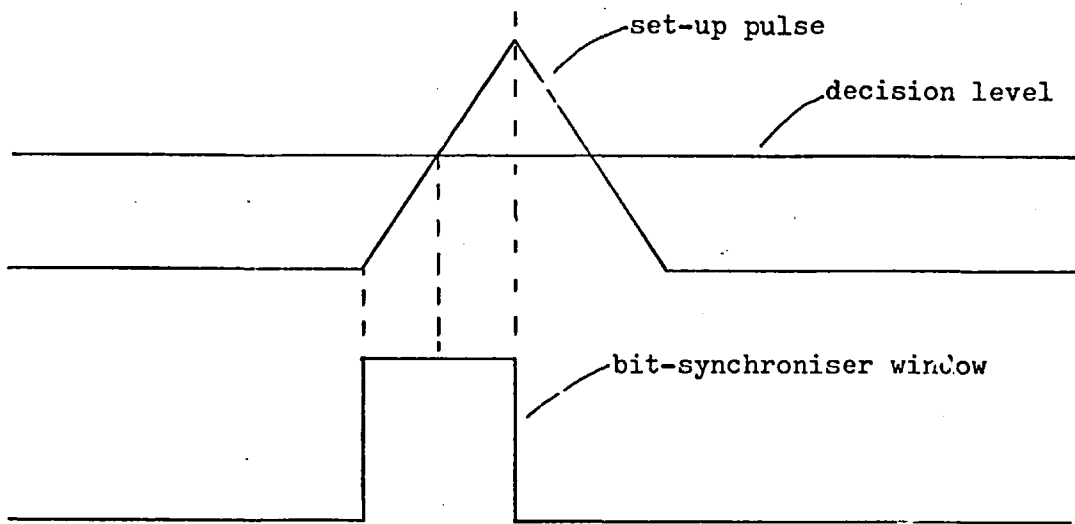
Fig 3-10.2   Realisation of the bit synchroniser
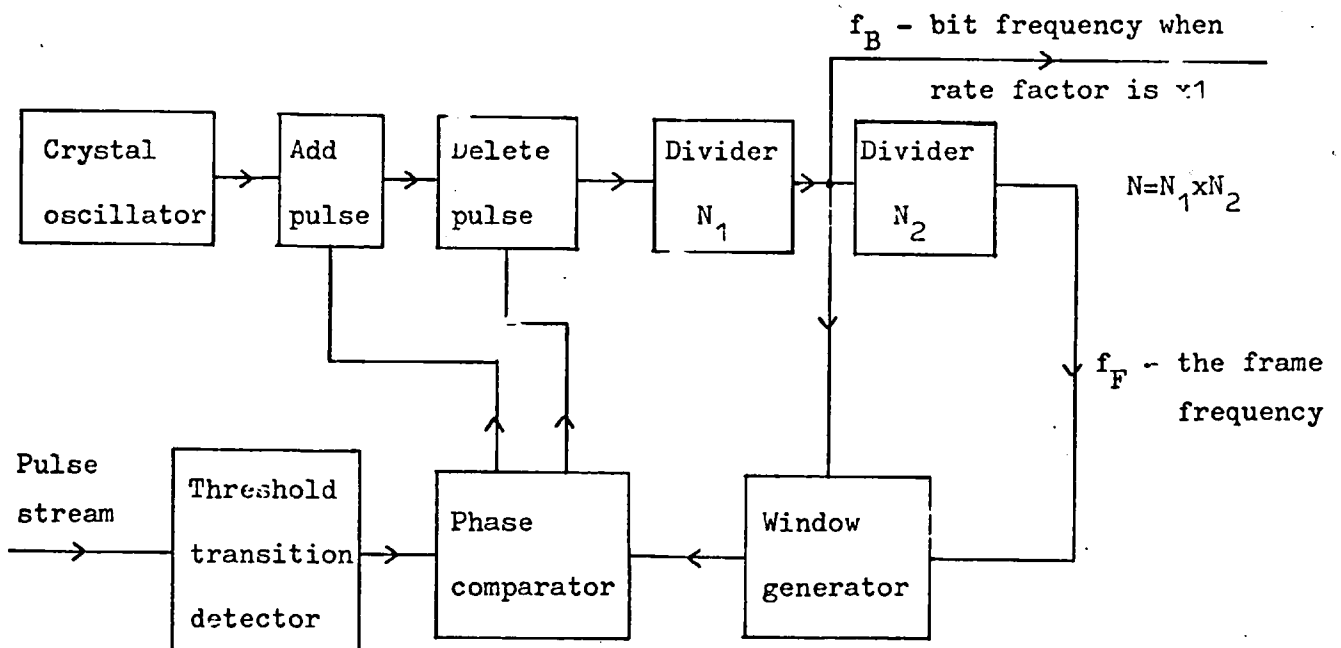
Fig 3-11.1   <u>Analogue impulse response of multipath channel</u>

Transmitter   Receiver
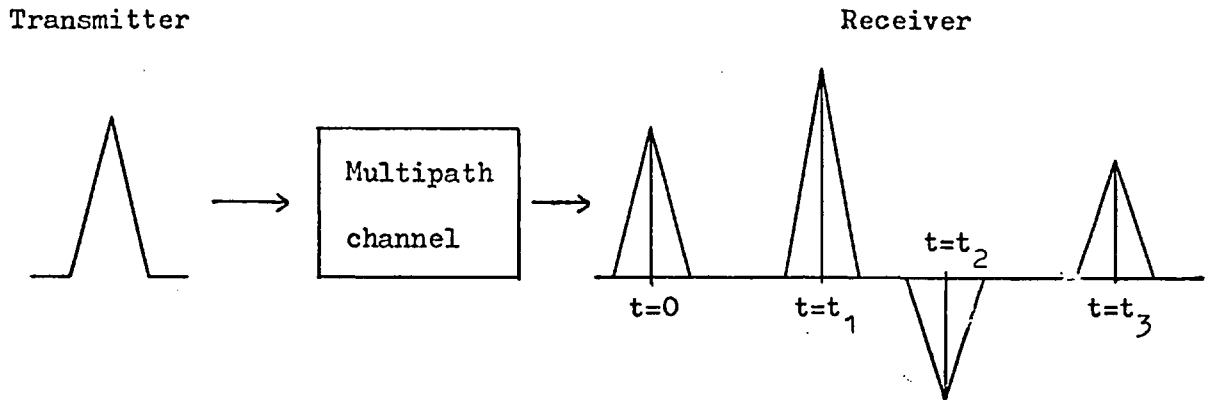
Fig 3-11.2   <u>Digital impulse response of multipath channel</u>
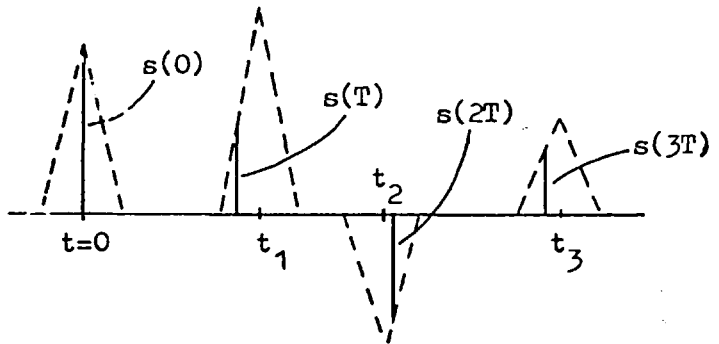
Fig 3-11.3   <u>Digital-filter model of multipath channel</u>
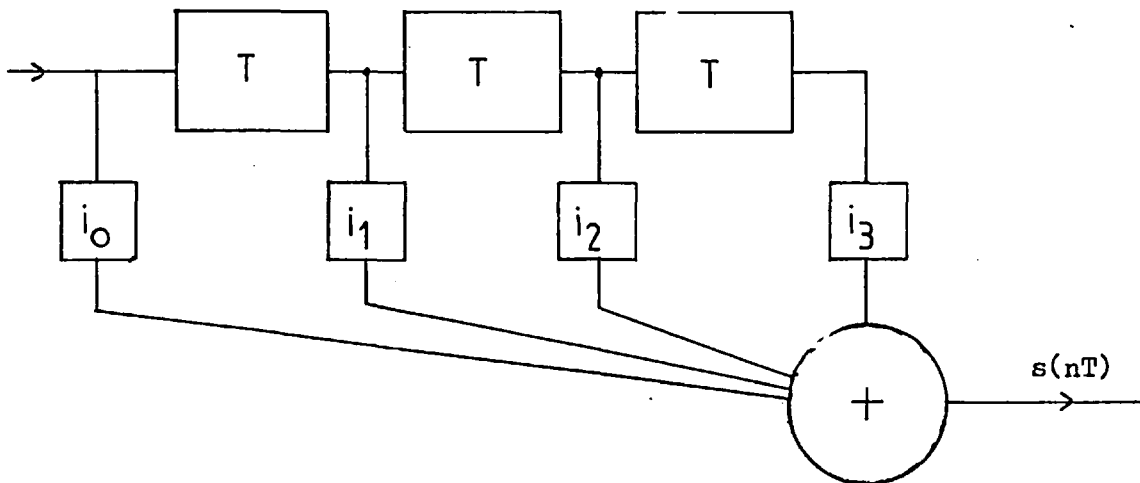
Fig 3-11.4 <u>z-transform equivalent of transmitted data, channel,</u>
<u>equaliser, and received data</u>
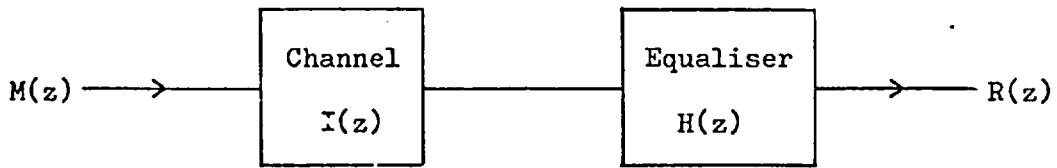
Fig 3-11.5   <u>Recursive digital filter</u>
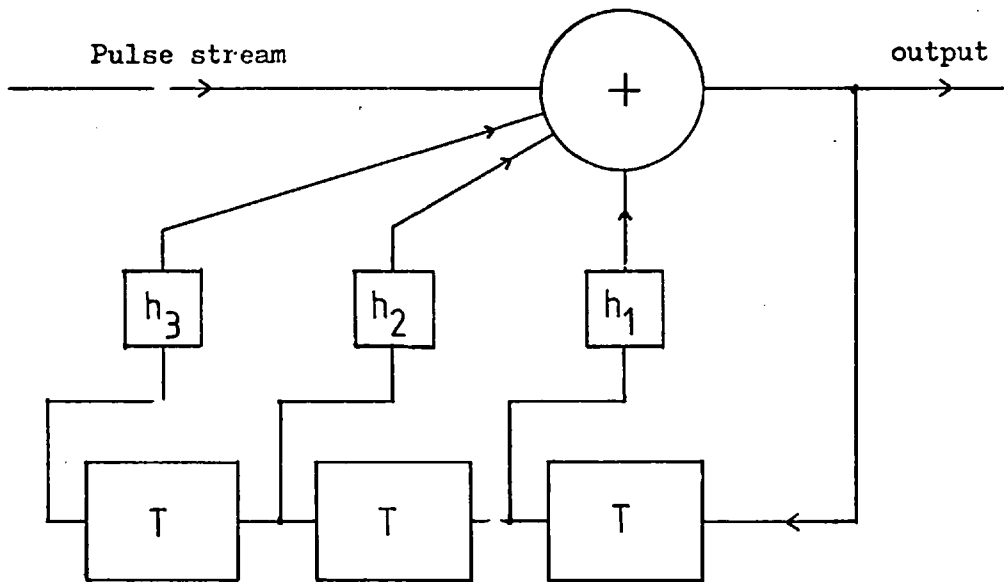
Fig 3-11.6 Performance of 5-tap transversal filter over 2-path channel

(a) MSE minimisation
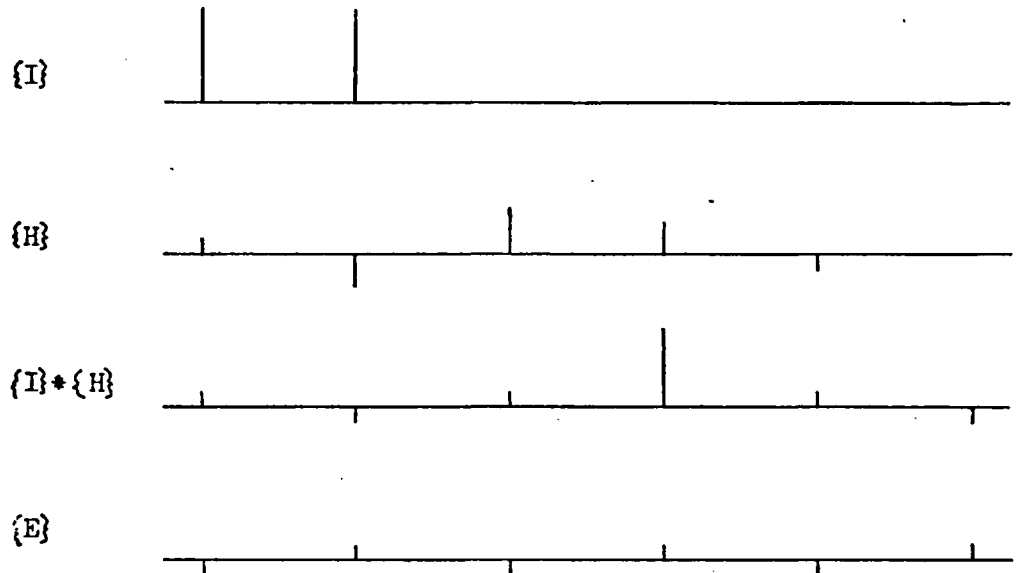
(b) Zero-forcing
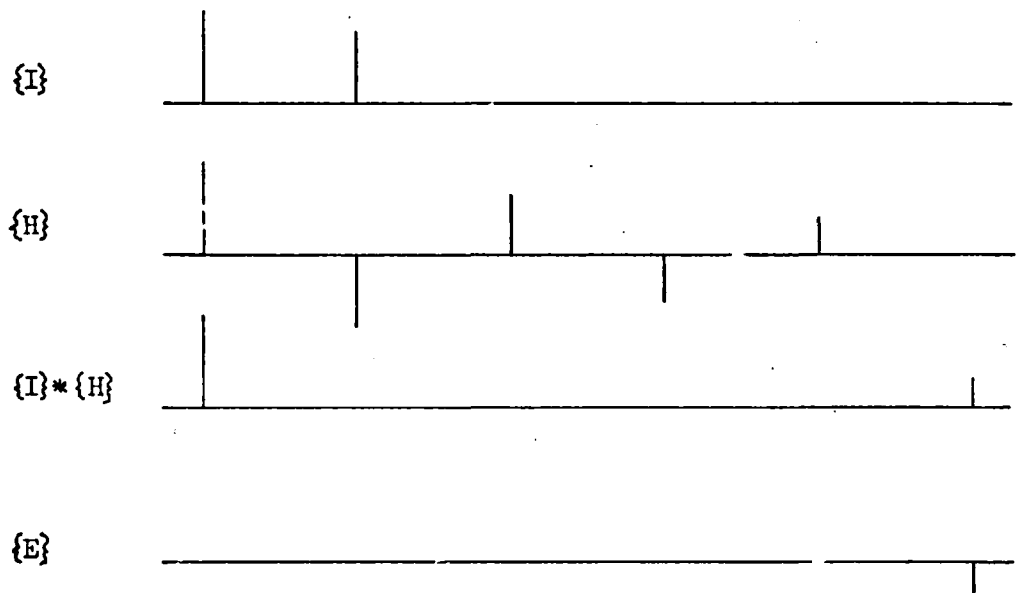
Fig 3-11.7   Decision-feedback equaliser structure
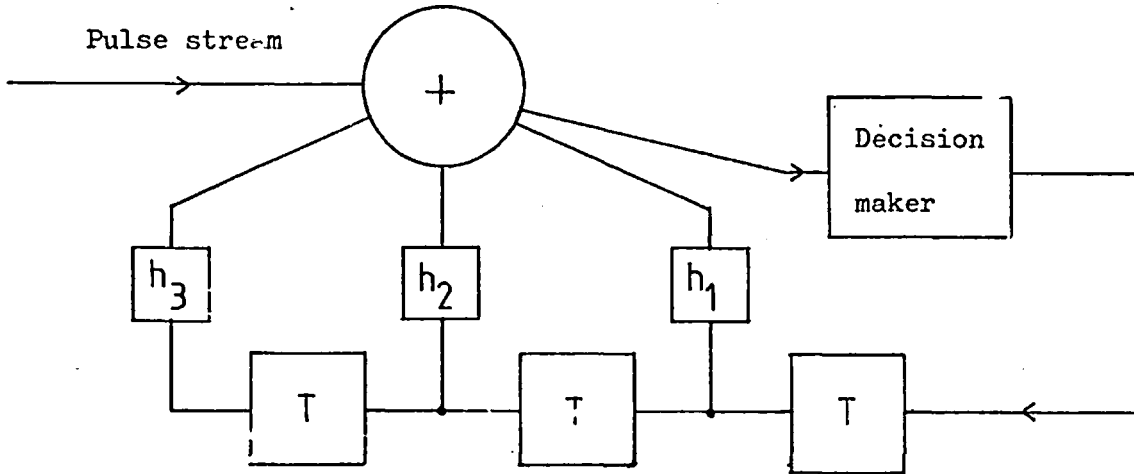
Fig 3-11.8   Evaluation of the channel impulse response by correlation
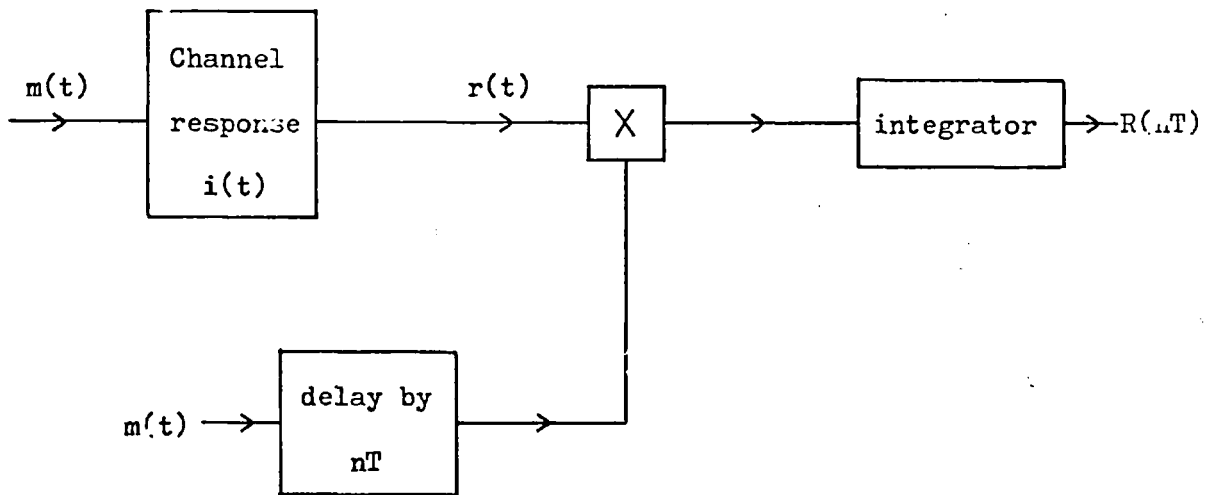
Fig 3-11.9   Estimation of the channel impulse response by correlation
              using the transmitted data
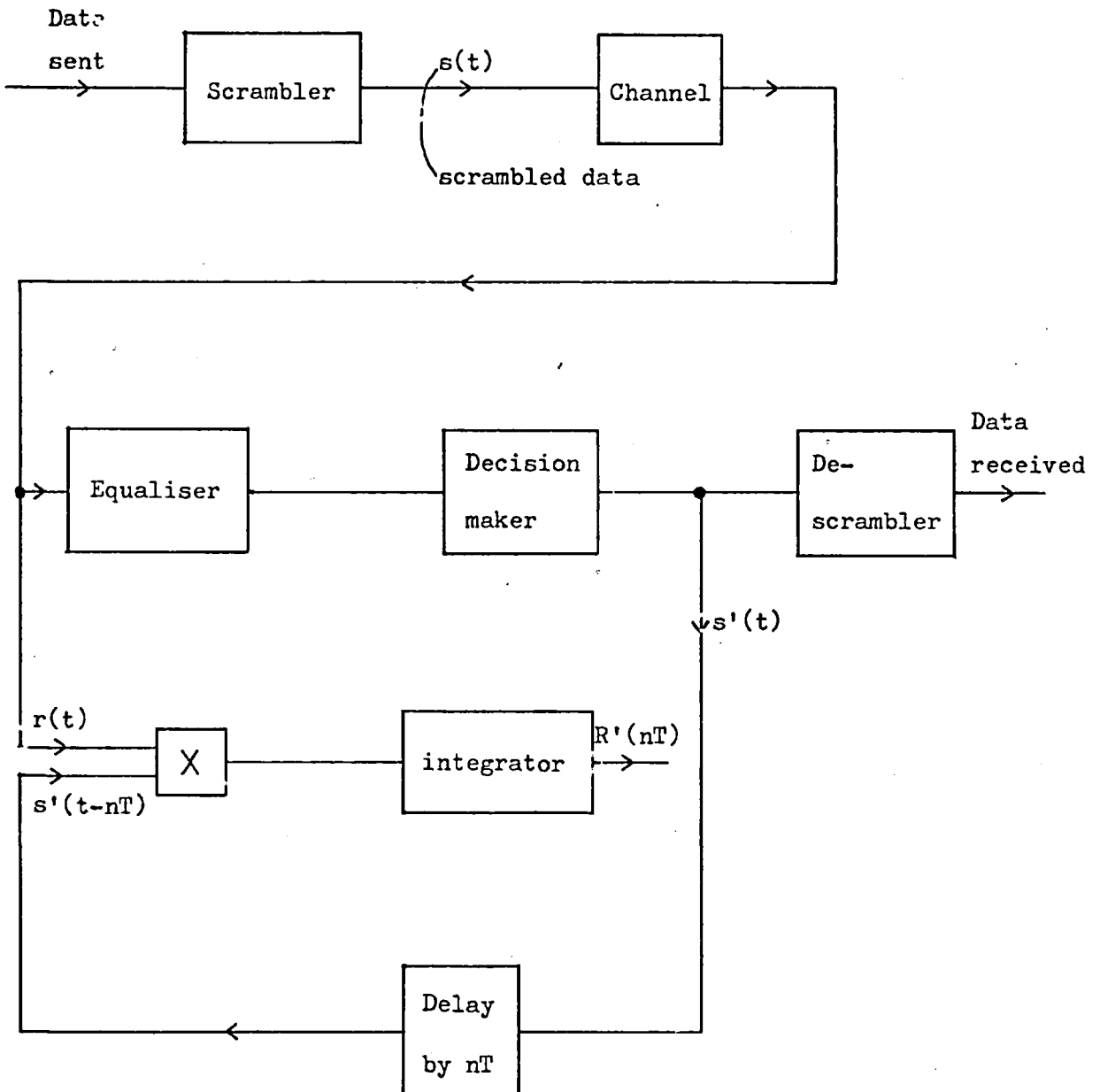
Fig 3-11.10   Display of channel impulse response in data-interrupt period
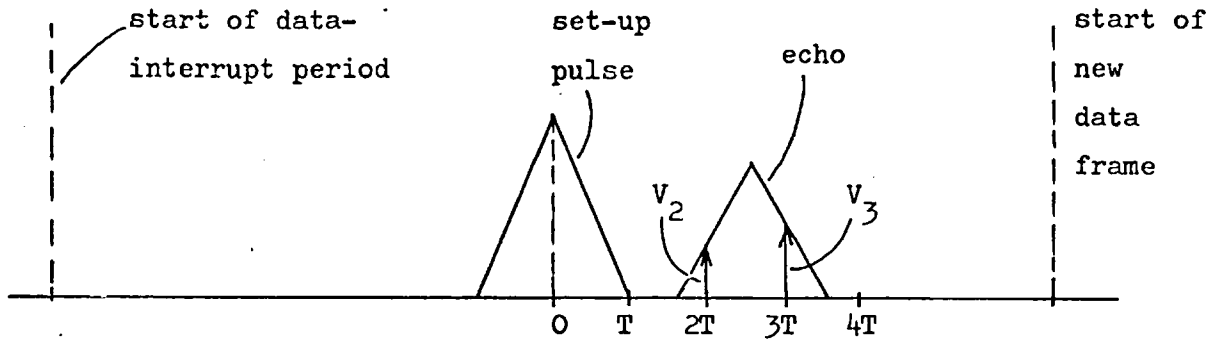
Fig 3-11.11   The test-system equaliser structure
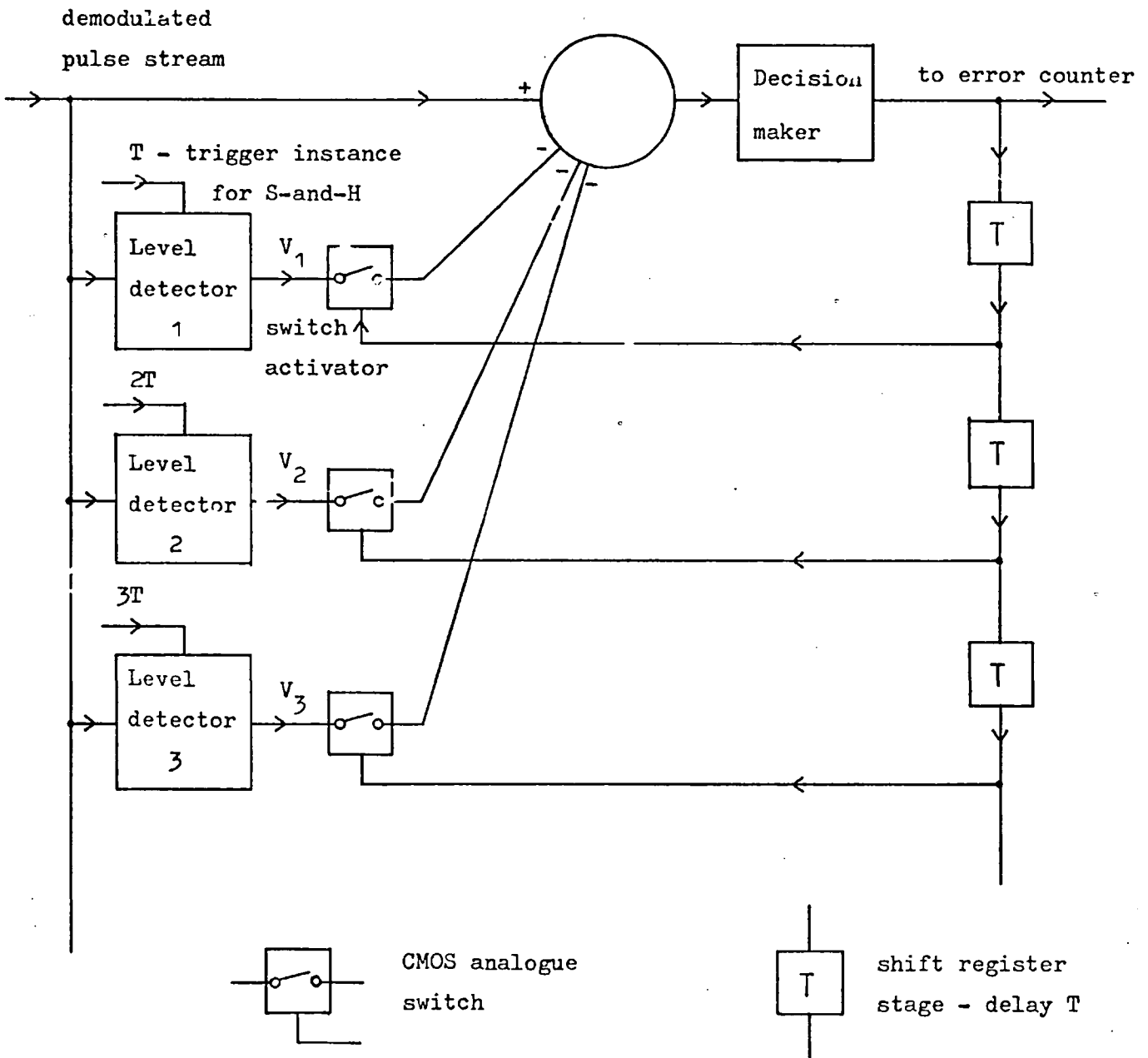
Fig 3-12.1   The decision maker
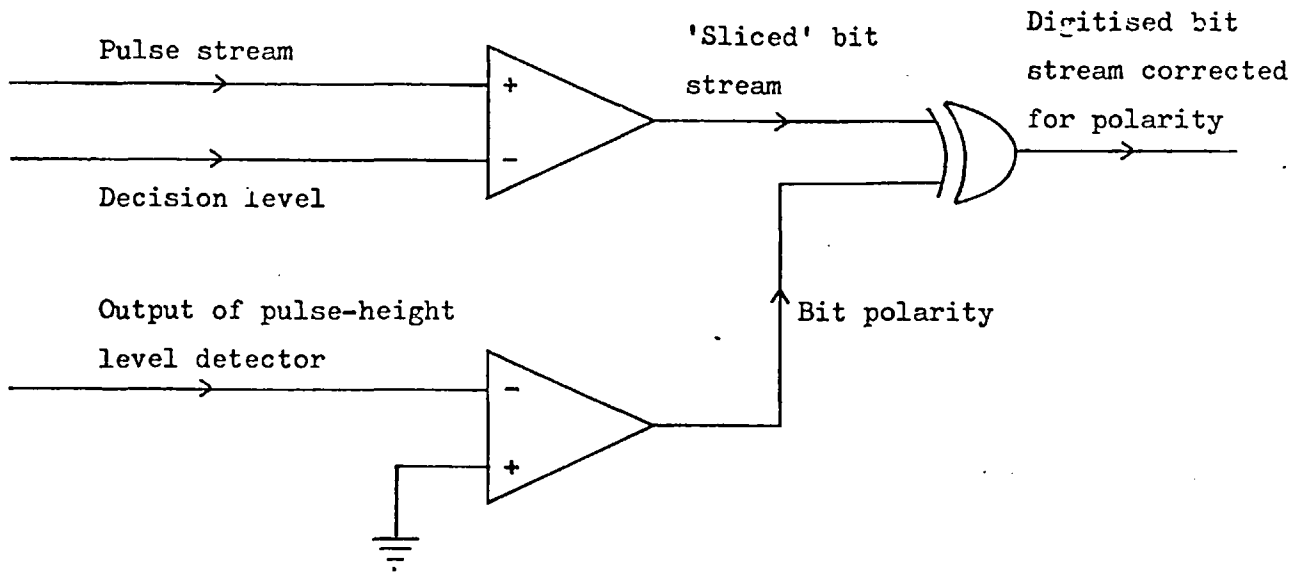
Fig 3-12.2 Probability density of binary signal in gaussian noise
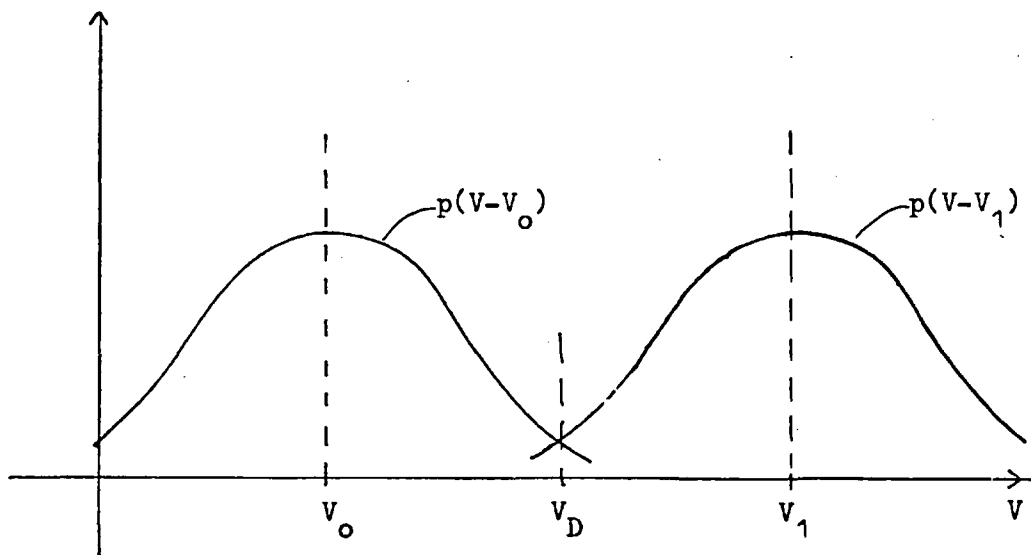
Fig 3-12.3   Variation in decision threshold with S/N ratio for rectifier-
demodulated ASK, together with the variation in the expected
values of rectified noise and rectified signal-plus-noise

Fraction of

pulse amplitude
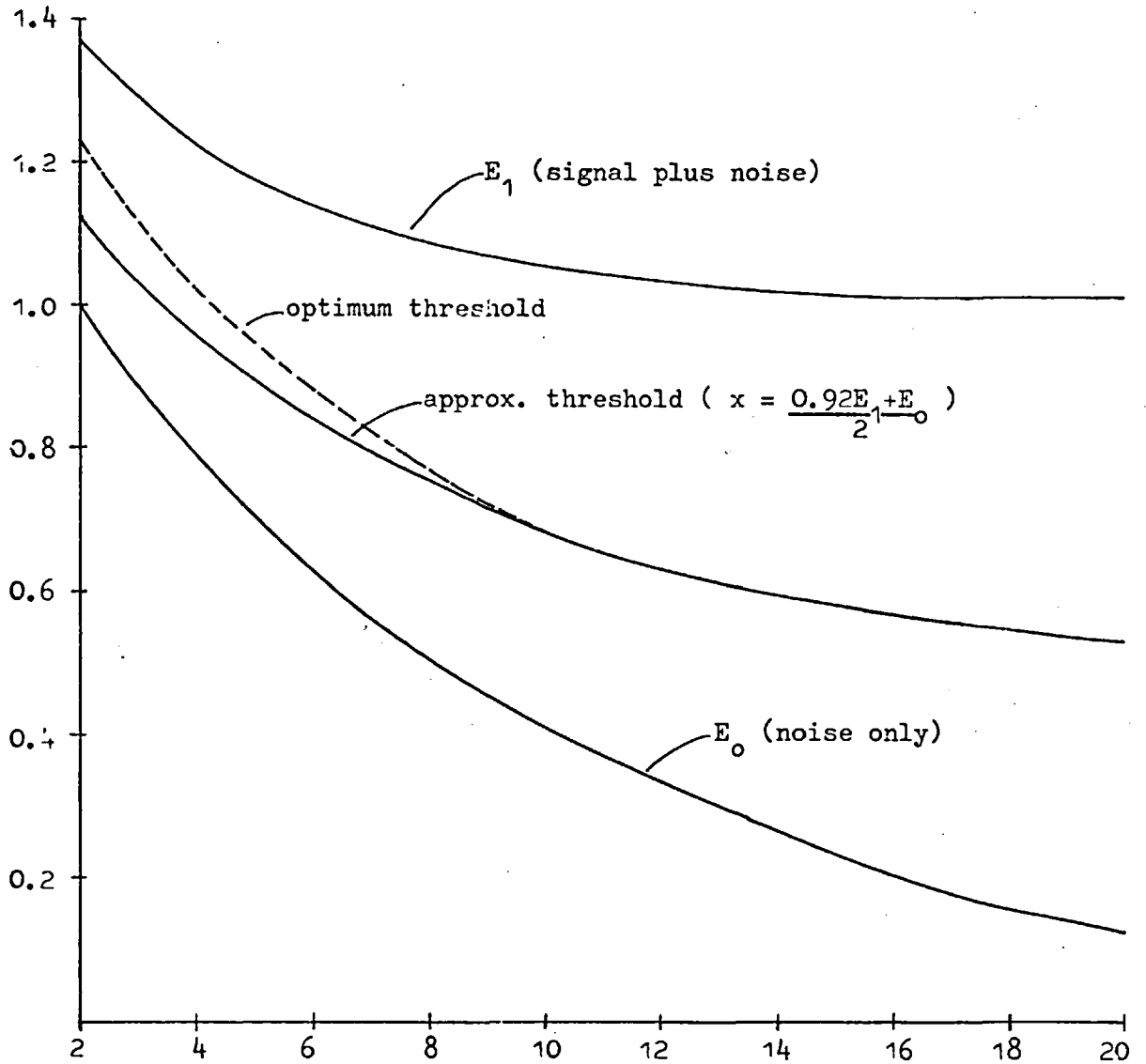
Fig 3-12.4   <u>Error performance of rectifier-demodulated ASK</u>
<u>comparing optimum and approximate threshold</u>

S/N ratio in dbs $(=10\log_{10}S)$	Error prob. using optimum threshold	Error prob. using approx. threshold
2	3.8×10^{-1}	3.8×10^{-1}
4	3.2×10^{-1}	3.3×10^{-1}
6	2.6×10^{-1}	2.6×10^{-1}
8	1.8×10^{-1}	1.8×10^{-1}
10	1.07×10^{-1}	1.07×10^{-1}
12	4.7×10^{-2}	4.8×10^{-2}
14	1.37×10^{-2}	1.37×10^{-2}
16	1.99×10^{-3}	1.96×10^{-3}
18	9.6×10^{-5}	9.6×10^{-5}
20	9.2×10^{-7}	9.9×10^{-7}

CHAPTER FOUR

PRESENTATION AND DISCUSSION OF THE RESULTS

INTRODUCTION

We start this chapter in section 4.1 by describing the experimental procedure adopted. We see that error measurements were performed on each permutation of the variables (modulation and demodulation modes, bit rates) in three-minute runs, the runs of the various permutations being tested sequentially to help average out long-term variations in the channel conditions. We see that three types of error measurements on the received data were performed, and that periodic estimates of the channel conditions were made.

We preface the presentation of the results by describing, in section 4.2, observations made during the trials. These were of two forms; of the channel conditions and of the test-system performance.

In section 4.3 we describe the results of the channel-condition estimates. The channel multipath structure is given, followed by the signal-to-noise conditions.

We start to analyse the error measurements in section 4.4 by examining how the error probability for the various modulation modes varied with bit rate. These are overall results for the 95% of the trial's duration when the signal-to-noise was high enough to permit reception. We start this section by deducing the channel conditions from the results. We then investigate the performance of such aspects of the test system as the dc extractor, the type of demodulation, and the type of pulse spectrum, to see how these aspects affected the results. We examine the difference in the error rate at the highest

and lowest bit rates, and finally we discuss the performance of each of the modulation modes.

In section 4.5 we present the error-probability distributions of PSK (the best-performance mode) and of ASK-rec.det.

In section 4.6 we examine the frequency and the amount by which the equaliser improved the error performance for PSK and ASK.

In section 4.7 we see how the error probability varied with time of day.

Finally, in section 4.8, we summarise the main results.

We start then by seeing the experimental procedure.

4.1 THE EXPERIMENTAL PROCEDURE

We start this section in 4.1.1 by summarising the variables in the test system, viz. the bit rates and the modulation and demodulation modes. We continue by describing how the permutations of the variables were tested in a series of three-minute runs, the runs of the different permutations being interleaved in a cyclic sequence. Overall readings for each permutation could then be obtained by aggregating the runs for that particular mode. We find out how the available radio channels were utilised, and what the duration of the field trials was.

In section 4.1.2 the measurements taken in the trials are described. We see they were of two types:

(1) error counts performed on the test system variables,

(2) periodic estimations of the channel conditions, these being of the multipath structure and the signal-to-noise ratio.

4.1.1 Method of Testing the System Variables

There were two sets of variables in the test system:

(1) the bit rates,

(2) the modulation and demodulation modes.

(1) Bit rates

We have seen in Chapter Three that two types of data-pulse spectra were used in the investigations. For each type of pulse four bit-rate variations were made, as summarised in the table in Fig. 4.1.1 (reproduced from Chapter Three). Thus in all there were eight bit rates.

Each of these rates was modulated and demodulated in each of the modes described below.

Modulation and demodulation modes

There were four different modulation modes as described in section 3.3, Chapter Three. They were:

PSK

and three different forms of ASK:

ASK
ASK-SC
ASK+C

All of these modes were demodulated by a phase-locked loop. Additionally the ASK mode was demodulated by rectifier. Thus there were five modulation and demodulation modes. They are summarised in the table in Fig. 4-1.2.

The total number of permutations of the test-system variables amounted therefore to forty.

In the trials error measurements were performed on each permutation sequentially in a series of three-minute test runs. The complete set of permutations were tested therefore every two hours. Interleaving the tests in this manner with a series of short runs had two advantages. Firstly, the effects of long-term changes, of the order of an hour or more, in the channel conditions would have tended to have been spread out amongst all the permutations. Secondly, the fact that short-term measurements were taken allows histogram results to be presented.

The duration of the field trials was from March 18 to April 18, 1977. Approximately fifty of the three-minute runs were made for each permutation of the test-system variables. Testing was performed in the hours from sunrise to sunset.

As far as the radio-transmission system was concerned channels for it were allocated in the range from 3.5 MHz to 8.5 MHz. During the March/April period of the tests the MUF varied from 3.5 MHz at sunrise and sunset to 9 MHz at mid-day. Each channel was separated by about 1 MHz which meant that the carrier frequency could be kept close to the MUF, and reasonably optimum propagation conditions could be maintained.

We now described the measurements taken.

4.1.2 The Measurements Taken

We have seen that the system parameters were tested in a series of three-minute runs. At the end of each run the three error measurements described in section 3.13, Chapter Three, were taken. These were:

  - the datum error count

  - the equalised-data error count

  - the unequalised-data error count

Further to these, measurements were made of the channel conditions. These were of the form of estimates of the average multipath structure and of the fluctuations in the signal-to-noise ratio. They were performed during each rectifier-detected mode run (which occurred every fifteen minutes). An example of the type of multipath-distribution recording is shown in Fig. 4-1.3. (The width of the data pulses was always found to be sufficiently narrow to "resolve" the multipaths.) The main signal multipath is that shown at time zero, i.e. the first multipath to be received. The remaining multipaths are displayed as post-echoes. The amplitude of the strongest path was normalised to unity; hence the amplitudes of

the other paths are with respect to this path*. Any paths having

an average strength during the three-minute run of less than

0.2 were not recorded. Since the estimates were made during

rectifier- detection they were of the magnitudes of the multipath

strengths. The results are presented in section 4.3.1.

The levels of the fluctuations in the signal-to-noise ratio

were also estimated during the rectifier-detected mode, ASK-rec. det.,

the ratio being taken as that of the pulse height to the noise variance*.

These results are presented in section 4.3.2.

Before we see the results of the measurements described above

we present some general observations that were made on the channel

conditions and on the performance of those parts of the receiver

system that could have been affected by the channel conditions.

---

*Both of these conditions could be readily observed on an oscilloscope
during the intervals on either side of the set-up pulse

4.2   OBSERVATIONS ON THE FIELD TRIALS

In section 4.2.1 we comment on the appearance of the channel conditions. In section 4.2.2 we comment on the performance of the receiver system.

4.2.1 Observations on the Channel Conditions

The channel was observed to be in a predominantly-multipath state. Usually from two to four paths gave significant signals, and the dispersion time was between one and two milli-seconds. The signal from the shortest-delay multipath was in general not the strongest path but always had at least 50% of the magnitude of the strongest path. The amplitude of the signal from each multipath was estimated to vary in the short-term (i.e. a few seconds) by 6 to 15 dB about a specular component. Longer-term changes in the relative strengths of the multipaths were also seen to occur.

The signal-to-noise ratio was estimated to be 10 to 15 dB overall. Again short- and long-term variations on these figures were observed to occur, the level of these variations being approximately -10 to +20 dB. These variations appeared to be due both to flat fading of the signal and to changes in the noise level. It was apparent that the noise was composed mostly of interference from other users. The very poor signal-to-noise levels were usually due to either a particularly strong interfering signal or to impulse noise apparently associated with local electrical storms. The signal-to-noise ratio was observed to be lowest at sunrise and within three hours of sunset.

4.2.2 Observations on the System Performance

In this section some consideration is given over to those parts of the receiver system whose performance was dependent on the channel conditions.

Phase-locked loop

The PLL proved to be stable under the channel conditions experienced, except at low bit rates (0.5 Kb/s) both for the ASK mode and, to a lesser extent, for the PSK mode. Cycle skipping was generally so infrequent that the resulting error bursts had a minimal effect on the overall error-rate results. The performance of the PLL was in fact assisted by the prevalence of multipath propagation because of the resultant increase in the average power of the receiver signal.

The demodulation system was seen to be defective, however, when the multipath structure was such that the first path (the signal) was weak in comparison to the "echo" paths. The PLL would lock onto the resultant of the phases of the carriers from the "echo" paths. Whenever the relative time delays of the multipaths were such that the carrier phase of the first path was at $\pi/2$ to this resultant an error burst ensued since the signal demodulated from this path would be zero. Thus, even at overall large signal-to-noise ratios, greater than 20 dB, the average error rate could be high because of this phenomenon. It is discussed in more detail in section 4.4.4.

Bit synchroniser

The requirement here was for the bit synchroniser to have sufficient gain to track time shifts in the signal-path delay caused by movement in the ionosphere. This was seen to be so.

Equaliser and level-detector bandwidths

The cut-off frequency of the low-pass filters in the equaliser
and threshold level detectors was made 5 Hz as a result of a series
of tests over a 100 Km HF link. In the Wick trials it was apparent
that this figure was too high and that the cut-off frequency should
have been made 3 to 4 Hz. This would have produced an improvement
in error performance due to the reduced effect of the amplification
of background noise by the equaliser.

We now proceed in this chapter by giving the results of the
measurements taken in the Wick trials.

4.3   RESULTS OF THE ESTIMATES MADE ON THE CHANNEL CONDITIONS

4.3.i   The Channel Multipath Structure

The method of recording the multipath structure was explained in section 4.1.2.   To reiterate, the position in time and the average strength of each multipath was estimated during each three-minute rectifier-detected run.   The strengths were normalised to the strongest multipath, and paths having a strength factor of less than 0.2 of the strongest path were ignored.

In Fig. 4-3.1 we see a histogram shown of the number of paths that were present in the channel. We note that in only 0.25% of the recordings was there single-path propagation; this illustrates the prevalence of multipath conditions.   The average number of paths works out to be three.

The distribution (in time) of the multipath amplitudes is shown in Fig. 4-3.2.   Here the origin is the first multipath to be received, the signal path.   The amplitudes given in each of the 0.25 mS slots after the origin are the result of grouping and then averaging the readings of the echoes made in each of the multipath recordings.

The distribution of the positions in time of the significant* multipath echoes are shown in Fig. 4-3.3.   The "average" multipath state of the channel may be expressed as in Fig. 4-3.4, by interpreting Figs 4-3.2 and 4-3.3.   The implication from Fig. 4-3.4 is that the first multipath (i.e. that at the origin) was the E-layer single-hop reflection and the second path was the F-layer single-hop reflection. This is because their delay difference was 0.5 mS (which it should be by considering the path geometry) and also because they were the two

---

*i.e. those having amplitudes greater than 0.2 of the strongest path

stronyest modes. The subsequent paths shown in Fig. 4-3.4 would have been multiple hops.

The distribution of the position of the strongest multipath is shown in Fig. 4-3.5. Again, the implication is that the first two paths received were the E-layer and F-layer reflections because the strongest path was usually either the first one or else a path delayed from the first one by 0.5 mS to 0.75 mS.

In the design of the hardware system the assumption was made that the first multipath to be received would always have substantial strength so that it could be treated as the signal path. The results shown of the multipath-structure measurements give experimental justification to this assumption. We see in Fig. 4-3.4 that the first path was on average 70% as strong as the largest-amplitude path whilst Fig. 4-3.5 shows that the first path was itself the strongest 35% of the time.

4.3.2 The Signal-to-Noise Ratio Measurements

These are presented in Fig. 4-3.6 in the form of a histogram. We see that the signal-to-noise ratio, defined as the ratio of the pulse height to the noise variance in the ASK(rectifier-detected) mode, was usually in the range from 5 dB to 20 dB with the average being 10-15 dB. When the signal-to-noise was less than 0 dB no reception was possible because the signal disappeared into the noise; the error rate here was therefore 0.5. This occurred 5% of the time. The results of the error measurements that are presented in the following sections are for the remaining 95% of the experiments.

We start by seeing how the error probability for each mode varied with bit rate.

4.4   THE VARIATION OF ERROR PROBABILITY WITH BIT RATE

The results for each of the five modulation and demodulation modes are presented in Figs. 4-4.1 to 4-4.5. They are overall results for all the runs for 95% of the experiments where reception was possible (i.e. when the S/N ratio was greater than 0 dB). The variations in these average results can be interpreted from the error-probability distributions given in section 4.5.

In Figs. 4-4.1 to 4-4.5 the probability variations are given for each of the three error measurements made, viz. datum errors, equalised errors, and unequalised errors.   It is apparent that for each of the modes the datum error rate was the lowest. This is to be expected since the datum bits did not experience the effects of, in the case of equalised data, the amplification of the background noise by the equaliser nor, in the case of unequalised data, the effects of intersymbol interference caused by the (almost-continuously) multipath conditions. It is also apparent that the equaliser gave an improvement in error performance for the PSK and ASK modes (Figs. 4-4.1 and 4-4.2) and to a lesser extent for the ASK-SC mode. These are general conclusions. More detailed analysis of these results is given in the following sub-sections. In sections 4.4.1 and 4.4.2 we deduce the prevailing channel conditions from the results and see how these deductions compare to the measurements presented in section 4.3. In section 4.4.3 we observe that the residual-dc extractor had a significant effect on the operation of the equaliser. We investigate various aspects of the performance of the PLL in sections 4.4.4 to 4.4.6 (viz. in comparison to rectifier-detection, the effect of the bit rate, and its reaction to whether it was stable in the one- or the two-phase state). In section 4.4.7 we see whether

the type of pulse spectrum used caused any difference in performance, whilst in section 4.4.8 we note how the error probability differed between the highest and lowest bit rates. Finally, in section 4.4.9, we compare the performance of the various modulation modes.

We start by inferring the channel conditions.

4.4.1 The Multipath Conditions

We observe that for the five modes as a whole the unequalised-data error rate was greater than 0.2 at around 3 Kb/s whilst the datum error rate figure was roughly 0.05 to 0.07. This indicates the continuous presence of multipath propagation in the trials since this discrepancy in error rate could only have been caused by intersymbol interference.

We then find that by again considering the five graphs as a whole the datum and data error rates are almost converged when the data bit rate is approximately 0.4 Kb/s. This can only mean that no intersymbol interference was present in the data stream at this rate. The bit-rate period must have been just such that the multipath echoes fell between pulses. Now at 0.4 Kb/s overall rate the bit-rate period is 2 mS, allowing for the 25% amount of redundancy due to the frame method of data transmission. The pulse width was 0.6 mS (100% spectrum roll-off) so the dispersion time must have been about 1.5 mS, in agreement with the results in section 4.3.1.

4.4.2 Level of Background Noise

An equivalent level of Gaussian noise can be inferred from the ASK-rec.det. results. The datum error rate is used since these bits were free from intersymbol interference. The average error probability for rectifier-detected ASK is then, from Fig. 4-4.5, given by:

$$P_e = 0.05$$

The equivalent average S/N ratio for Gaussian noise is found from section 3.12 to be 12 dB. This figure again is in agreement with the results of the signal-to-noise estimates given in section 4.3.2.

We now turn our attention towards the performance of aspects of the system.

4.4.3 Effect of the Residual-dc Extractor on the Performance of the Equaliser

We see in the results for PSK (Fig. 4-4.1) and ASK (Fig.4-4.2) that the use of the equaliser led to an overall improvement in the error performance. Defining the ratio of unequalised-data error count to equalised-data error count as the improvement factor, the value at a bit rate of 2.0 Kb/s to 2.5 Kb/s is 1.7 for PSK and 1.45 for ASK.

With the ASK-SC and ASK+C modes the error-rate improvement from the equaliser was small or negligible.

The difference that the two pairs of modes had in common was that PSK and ASK had no residual dc component whilst ASK-SC and ASK+C did. These latter two modes required the residual-dc extractor to be in operation, and hence it must have been this that affected the performance of the equaliser. The reason for this is as follows.

As well as the dc level the output of the residual-dc level detector contained low-frequency noise components synthesised from the background noise. The output was subtracted from the pulse stream to extract the dc (see Fig.3-0.2), and so low-frequency noise components were effectively added to the background noise. The level detectors in the equaliser had the same bandwidth as these components; hence to the equaliser this noise appeared as a slowly-varying level superimposed on to the signal. Thus its operation in measuring the

channel impulse response was impaired leading to incorrect equalisation of the pulse stream.

We now examine three aspects of the PLL.

4.4.4 Comparison of the PLL and the Rectifier Methods of Detection

For the purpose of this comparison the datum error results are used since these bits were free from the effects of the equaliser-operation and of intersymbol interference. Errors here were caused mainly by background noise.

The datum error rate P_{AR} for the rectifier-detected mode, ASK-rec.det., can be seen in Fig. 4-4.5 to be on average:

$$P_{AR} \simeq 0.05$$

The datum error rate P_{AP} when ASK was detected by the PLL is seen in Fig. 4-4.2 to be at best* (at 2.4 Kb/s):

$$P_{AP} = 0.07$$

For PSK, which had twice the eye opening of ASK at the receiver, the best datum error rate P_{pp} is given by (from Fig. 4-4.1):

$$P_{pp} = 0.045$$

Hence, contrary to its theoretical advantage over rectifier-detection as regards noise performance, PLL carrier extraction here resulted in a worsening of the error rate. The cause of this was noted in section 4.2.2 to be the effect of multipath on the action of

---

*The reason for this variation with data bit rate is discussed later

PLL. It is now analysed more rigorously by considering a three-multipath channel as an example.

Let the modulated signal arriving at the receiver be expressed as

$$f(t+t_1)A_1\cos w(t+t_1) + f(t+t_2)A_2\cos w(t+t_2) + f(t+t_3)A_3\cos w(t+t_3)$$

where A_1, A_2, and A_3 are constants proportional to the transmission coefficients of the three multipaths, t_1, t_2, and t_3 are the propagation delays of the multipaths with $t_1 < t_2 < t_3$, w is the carrier frequency, and f(t) is the transmitted pulse stream.

The receiver system treated the signal from the shortest multipath, i.e. $f(t+t_1)A_1\cos w(t+t_1)$, as the data. Thus let us aerive the component of the demodulator output that contains this signal

$$\text{Let} \quad \theta_1 = wt_1 - 2\pi n,$$

and

$$\theta_2 = wt_2 - 2\pi n_2$$

and

$$\theta_3 = wt_3 - 2\pi n_3$$

with n_1, n_2, and n_3 being integers of value such that θ_1, θ_2, and θ_3 lie between 0 and 2π.

Let us assume that the magnitude A_1 of the first path is substantially less than A_2 and A_3. Then the carrier component from this path will be "swamped" by intersymbol interference from paths A_2 and A_3. In this case the PLL will lock onto the phase

resultant of the carriers from the A_2 and A_3 paths. The phasor
diagram of this postulated multipath structure is shown in Fig.4-4.6.
Here A_2 and A_3 are made approximately equal. Thus the PLL-output
phase θ_{PLL} is given by:

$$\theta_{PLL} = \frac{\theta_2 + \theta_3}{2}$$

The output of the demodulator for the signal path A is the
reflection of this phasor on the PLL output phasor. Hence it is
proportional to

$$f(t + t_1) \cos(\theta_1 - \theta_{PLL})$$

$$= f(t + t_1) \cos[\theta_1 - (\theta_2 + \theta_3)/2]$$

Now since θ_1, θ_2, and θ_3 are functions of the path time
delays (t_1, t_2, and t_3) any changes caused by the ionospheric layers
shifting and altering the path lengths will make the "signal" phasor
A_1 rotate around the phasor diagram origin. When

$$\theta_1 - (\theta_2 + \theta_3)/2 \simeq \pi/2 \text{ or } 3\pi/2$$

the resolved component of the A_1 phasor on the PLL "In-phase" axis
(the demodulation output for the signal) will be close to zero. In
this region the signal will be small relative to the noise (given the
noise conditions described in sections 4.2.1, 4.3.2 and 4.4.2) and so
the error rate will be high.

In summary, the problem was that the first multipath was not
necessarily in phase with the "average" phase of the composite
carrier from all the multipaths. Hence, despite the fact that the

PLL was observed to be stable in locking to the "average" phase,

the signal-to-noise ratio of the demodulated signal was low at

times even in low noise conditions. A solution to this problem

of coherent demodulation is outlined in Chapter Five, in the

suggestions for further work.

4.4.5 Effect of the Bit Rate on the Phase-Locked Loop

It was noted in Chapter Three that the value of the average modulated-

signal to noise at the receiver could have an effect on the performance of

the phase-locked loop. The variation with bit rate of the average modulated

signal level was drawn for each mode in Fig. 3-3.5. In the results we

can see this phenomenon manifesting itself. Considering again the

datum error results we see that at the low bit rates (about 0.5 Kb/s)

the error rate increased for ASK (Fig. 4-4.2) and PSK (Fig.4-4.1).

These two modes had no residual carrier component and so their average

signal level was proportional to the bit rate. The deterioration in

the error performance is especially marked for ASK since this had half

the average level of PSK. The ASK+C mode (Fig. 4-4.4),having some

residual carrier, suffered only a slight degradation in its datum

error performance at low rates, whilst the ASK-SC mode (Fig. 4-4.3),

whose average signal level increased with decreasing bit rate (see

Fig. 3-3.5), displayed no such tendency at low rates.

4.4.6 Comparison of the Bi-phase and Uni-phase Stable Modes of the PLL

This comparison can be made by examining the performance of the

ASK and ASK+C modes. ASK had to be detected by a PLL which was stable

in both the π and the 2π phase positions because the zero crossings

in the base-band pulse stream, from the pulse side lobes, caused π

phase shifts in the modulated carrier. In the ASK+C mode sufficient carrier was added to remove these phase shifts enabling the PLL to be made stable in just the single-phase position.

The theoretical advantages of ASK+C are:

(1) the single-phase stable PLL is less likely to cycle skip than the bi-phase type which can flip between the two phase states,

(2) the additional carrier helps the PLL to lock at low bit rates.

Its disadvantages are:

(1) the eye-opening is reduced because of the added carrier,

(2) the added carrier requires a residual-dc extractor to be in operation because the demodulated signal has a dc component.

The results for ASK and ASK+C in Fig. 4-4.2 and Fig. 4-4.4 show that ASK performed in a superior fashion above 1 Kb/s for the best data performance. that of equalised data. As was discussed in section 4.4.3 the effect of the dc extractor was mainly responsible for this. Below 1 Kb/s the PLL tended to fail for ASK, and the extra signal power of ASK+C resulted in a somewhat better error rate. Hence, in conclusion, cycle skipping in a two-phase stable PLL is not a problem except at low bit rates.

We now turn to the error performance of the two types of pulse spectrum used.

4.4.7 Error Performance of the Two Types of Pulse Spectrum

We saw in Chapter Three that two types of spectrum were employed for the data pulses:

(1) 33% roll-off raised cosine

(2) 100% roll-off raised cosine

The 33% roll-off pulse had a main-lobe width of 0.4 mS opposed to a figure of 0.6 mS for the 100% roll-off case. Therefore the 33% roll-off pulse was more likely to resolve the multipath and so not suffer from distorted pulses. The 100% roll-off pulse had insignificant sidelobes and so produced a larger eye-opening at the receiver than the 33% roll-off pulse whose sidelobes were substantial (see Fig. 3-2.10).

In order to get a comparison of the performance of the two types of spectrum we use the ASK-rec.det. datum error results. The rectifier-detection mode is chosen because the PLL demodulation problem described in section 4.4.4, whilst the results for datum bits are chosen because they did not experience the effects of intersymbol interference from multipath or the effects of the equaliser. The results in Fig. 4-4.5 show virtually no difference between the two types of spectrum. The 1.2 Kb/s and 2.4 Kb/s 100% roll-off bit rates show marginally higher error rates indicating that pulse distortion may have been a slightly more significant factor than eye-opening degradation but, on the whole, the conclusion must be that the 33% roll-off spectrum is preferable simply because it allows a higher basic bit rate.

Having used the overall results presented in Fig. 4-4.1 to 4-4.5 to highlight particular aspects of data transmission we now discuss the results in these two final sections more generally.

4.4.8   The Difference in the Error Rate at the Highest and Lowest Bit Rates

As was discussed in Chapter Three the bit rate was lowered by keeping the pulse-width narrow and leaving blank gaps between bits. The pulse bandwidth and hence the receiver-filter bandwidth remained constant meaning that the noise content of the received signal was not reduced with decreasing bit rate as would have been the case where the bit rate is lowered by pulse widening. As was explained in section 3.1.1, though, it was felt that the method of widening the pulse would have led to severe pulse distortion if multipath was present, a situation which would have reduced the effective signal level and which could not have been rectified by a conventional equaliser*. Pulse distortion was also felt to be liable to present problems in obtaining bit synchronisation.

Since the receiver bandwidth was held constant it might be thought that the error rate would not have reduced with bit rate because the signal-to-noise at the receiver remained the same. In fact, as can be seen in Figs. 4-4.1 to 4-4.5, some reduction took place in the data error rate. With regard to the equalised data, the error rate reduced because, as the bit rate was lowered, less equaliser stages were necessary to be in circuit. Thus the amplification of the background noise caused by the equaliser level

---

*The likely performance of the pulse-widening method in the channel conditions experienced in the trials is discussed in the conclusions in Chapter Five.

detectors was less marked. With regard to the unequalised data, the multipath echoes encompassed fewer bits as the bit rate was lowered, hence reducing the amount of intersymbol interference.

We now examine the amount by which the error rate changed as the bit rate was reduced. At 3.6 Kb/s the best error-performance figure was for PSK, after equalisation,of:

$$P_e = 0.14$$

At 0.4 Kb/s the best figure was for ASK-SC, of:

$$P_e = 0.057$$

Hence for a nine-fold decrease in the bit rate the error rate was reduced by a factor of only two or three. This raises the question of whether the error rate at the lower bit rates would be improved more by transmitting data in code. For this system the fundamental rate used would be 3.6 Kb/s, and the level of code redundancy would then be made appropriate to achieving the desired data rate. For example, let us use a majority-decision code having redundancy factors of 3,5 and 7. The bit rates would then be respectively 1.2 Kb/s, 720 b/s and 510 b/s. Let us use the error probability at 3.6 Kb/s for PSK found in the trials, of:

$$P_e = 0.14$$

Then using the expression derived in Appendix 4.1 the resultant error rates for the codes can be calculated. They are plotted in Fig. 4-4.7 together with the equalised-data error rate for PSK. We can see the improvement afforded to the performance at the lower bit rates by utilising even a simple form of coding.

In this final sub-section we discuss the performance of each

mode in these overall results.

4.4.9 Comparison of the Modulation Modes for Overall Equalised-Data Error Performance

We start this section by presenting an order-of-merit table in

Fig. 4-4.8 for the equalised-data error performance of each mode.

The equalised data results are chosen because they were never worse

than the results for unequalised data and usually better. The table

is given for high (3.6 Kb/s), intermediate (1.2 Kb/s), and low

(0.4 Kb/s) bit rates, with the best mode shown at the top of each list.

This table provides a point of reference as we now examine the

performance of each mode.

PSK

We see in Figs. 4-4.1 to 4-4.5 that PSK gave the best performance

of all the modes except at low bit rates where the PLL became slightly

unstable because of the lack of average signal power. At the high

and intermediate bit rates its lack of a residual-dc component together

with its advantage of having double the eye-opening vis-a-vis ASK

enabled PSK to out-perform the other modes.

ASK

The results show ASK to have been the best of the amplitude-

shift-keying type of modes at all except low bit rates. As with

PSK its lack of a residual dc component gave it the advantage at the

higher rates. It had the minimum average signal power of all the

modes though (see Fig. 3-3.5, Chapter Three) which caused the worsening

of its performance at the low bit rates as the PLL lost its "grip".

ASK-SC

As was seen in Chapter Three (see Figs. 3-3.3 and 3-3.4) ASK-SC had the advantage of having twice the eye-opening of the other ASK modes. However its residual-dc content required the residual-dc extractor to be employed. As was noted in section 4.4.3 this affected the operation of the equaliser so severely that ASK-SC had inferior performance to ASK at the higher bit rates. We can see in Fig. 3-3.5, Chapter Three, though that ASK-SC had the highest average signal power at low rates of all the modes. Hence the deterioration in error performance caused by the PLL which was experienced by the other "coherently"-detected modes did not occur with ASK-SC, and it gave the best results at low rates.

ASK+C

For this mode sufficient carrier was added so that the π phase shifts present in the ASK modulated carrier did not occur. Hence the PLL needed to be stable in only the one phase mode which assisted its performance. As was discussed in section 4.4.6 the necessary connection of the dc extractor worsened the operation of the equaliser and, since ASK+C had the smallest eye-opening of all the modes as well, its error performance was inferior to all the other equalised modes, as can be seen in Figs. 4-4.1 to 4-4.4.

ASK-rec.det.

Rectifier detection of the ASK mode was included in the test system as an additional parameter mainly to operate as a check on the operation of the PLL. Its usefulness in this respect was seen in section 4.4.4. As far as its data error performance was concerned we see in Fig. 4-4.5 that it was inferior to all the PLL-detected methods. This was because rectifier detection is non-linear and

precludes the use of a conventional equaliser. The prevalence of multipath conditions therefore caused the error rate to be the highest except at low bit rates where the gap between bits was such that intersymbol interference was considerably reduced. Then the asynchronous nature of rectifier detection can be seen to have been of some advantage as compared to the PLL-detected ASK+C and ASK modes.

The results presented in Figs. 4-4.1 to 4-4.5 are averages over the time when reception was possible which was 95% of the total duration of the experiments. In order to get a feel of the variation in the error performance results we now give the error-probability distributions for PSK and ASK-rec.det. by making use of the fact that the error measurements were made over three-minute intervals.

4.5   THE ERROR-PROBABILITY DISTRIBUTIONS FOR PSK AND ASK-rec.det.

The distributions are given for PSK, because it was seen in section 4.4 to give the best overall performance, and for ASK-rec.det. because:

(1)   its datum error performance gives an indication of the variation in the signal-to-noise ratio,

(2)   the comparison of the data and datum error distributions indicates the variation that was present in the level of intersymbol interference caused by multipath.

The probability-density and the cumulative-probability distributions for equalised and unequalised PSK are shown in Figs 4-5.1 and 4-5.2, and for ASK-rec.det. (data and datum bits) they are shown in Figs. 4-5.3, 4-5.4, and 4-5.5.

For the purpose of presenting the probability-density distributions the error rate is divided into the following "slots": 5×10^{-1} to 2×10^{-1}, 2×10^{-1} to 10^{-1}, 10^{-1} to 3×10^{-2}, 3×10^{-2} to 10^{-2}, etc. up to 10^{-4} to 3×10^{-5}, a further slot being provided for error readings of less than 3×10^{-5}. The error rate for each of the three-minute period-of-operation runs was calculated and the number having an error rate associated with a given slot was counted and expressed as a fraction of the total number of runs for the mode and bit rate in question.

The four bit rates for which the distributions are given are:

(1)   3.6 Kb/s   33% spectrum roll-off, x 1 bit-rate reduction factor

(2)   2.1 Kb/s   from combining the 2.4 Kb/s (100% roll-off, x 1 rate) and the 1.8 Kb/s (33% roll-off, x 1/2 rate) results

(3) 1.05 Kb/s   from combining the 1.2 Kb/s (100% roll-off,

x 1/2 rate) and the 0.9 Kb/s (33% roll-off,

x 1/4 rate) results

(4) 0.6 Kb/s   from combining the 0.6 Kb/s results from

the 100% spectrum roff-off, x 1/4 rate,

results and the 33% spectrum roll-off, x 1/6

rate, results.

(As we have seen in section 4-4.7 there was no discernible

difference in the results for the two types of pulse spectrum, and

the combining of them for the rates (2), (3) and (4) results in better

averaging.)

The cumulative-distribution figure for the error probability

P_e is taken as being the fraction of the total number of runs that

the error rate is less than or equal to P_e.

The results are now interpreted.

PSK

The histograms for all four bit rates are shown in Fig.4-5.1

for PSK. The cumulative distribution is shown in Fig. 4-5.2. We

see in the histogram results that the presence of the equaliser

tended to reduce the percentage of very high error rates. The

overall improvement afforded by the equaliser is especially marked

at the 2.1 Kb/s bit rate. This can be seen more particularly in the

cumulative-distribution results for the 2.1 Kb/s rate where the

equalised and unequalised-data curves are separated by the largest

amount at this rate.

ASK-rec.det.

In the case of ASK-rec.det. the histograms and cumulative distributions for data are given in Figs. 4-5.3 and 4-5.4 respectively. The histograms show that there were a high percentage of runs with error rates greater than 0.2 for the 3.6 Kb/s, 2.1 Kb/s, and 1.03 Kb/s bit rates. This indicates the catastrophic effect of the intersymbol interference caused by the multipath conditions. For the 0.6 Kb/s rate the bit rate period was usually greater than the channel dispersion time (shown in section 4.3.1) and hence the histogram more closely resembles that for the datum bits which is shown in Fig. 4-5.5. The cumulative distribution for datum bits is also shown in Fig. 4-5.6.

We have seen that the equaliser improved the data error performance for PSK and ASK on average. We now see the frequency and the amount with which this improvement was sustained by presenting a histogram of the equaliser performance.

4.6   EQUALISER PERFORMANCE ILLUSTRATED BY HISTOGRAM

The term "the equaliser error-reduction factor" B, defined
as

$$B = \frac{errors\ \underline{before}\ equalisation}{errors\ \underline{after}\ equalisation}$$

was calculated for each three-minute test run of the PSK and ASK modes.
The results are then presented in the form of histograms for each of
the four bit rates described in the previous section. The equaliser
histograms for PSK are shown in Fig. 4-6.1, and for ASK they are
shown in Fig. 4-6.2.

The results contained in Figs. 4-6.1 and 4-6.2 can be
interpreted as follows. The equaliser can be seen to have performed
best at 2.1 Kb/s. At 3.6 Kb/s there were 75% more equaliser taps
switched in and the additional noise from these taps overrided the
extra intersymbol interference cancelled as compared to the 2.1 Kb/s
rate.   Below 2.1 Kb/s the concentration of multipath was relatively
less because the bit-rate period tended towards the same length as
the channel dispersion time.   Hence the effectiveness of the
equaliser decreased at the lower rates.

Figs. 4-6.3 and 4-6.4 display for PSK and ASK respectively
the fraction as plotted against bit rate of the total number of runs
for which the equaliser improved the error performance.   At the bit
rate of peak-performance improvement, viz. 2.1 Kb/s, the large
fractions of 80% to 90% indicate the almost continuous presence of
multipath during the trials.

We now see how the error probability varied with the time
of day.

4.7 THE VARIATION IN ERROR PROBABILITY WITH TIME OF DAY

The results for PSK for the bit rates defined in section 4.5 are shown in Fig. 4-7.1. Times of day are in British Summer Time. Sunrise occurred at 06.30 and sunset at 19.50 during the trials. These are again average results for all the readings taken*.

Apart from an increase at 07.00 and a smaller increase at 16.00 the error rate is shown to have been fairly independent of the time of day. The results imply that these increases were due to a lowering of the signal-to-noise ratio rather than a change in the level of intersymbol interference due to multipath because the datum error rates increased in step with the data error-rate curves. This was confirmed by observation as mentioned in section 4-2.1.

We now conclude this chapter by giving a summary of the main results.

---

*No finer detail is possible because of the scatter of the measurements

4.8 SUMMARY OF THE MAIN RESULTS

Channel conditions (section 4.3)

- The channel was in a multipath state 99.5% of the time.

- The average multipath structure (shown in Fig.4-3.4) was of three paths spread over 1.5 mS.

- The first significant multipath was always of significant strength.

- The signal-to-noise ratio was too low to make reception possible for 5% of the duration of the trials.

- For the remaining 95% of the experiments the signal-to-noise ratio varied from 0-6dB to greater than 25dB, being on average 10-15dB (see Fig.4-3.6).

Results of error-rate measurements (sections 4.4 to 4.7) (these are for the time when reception was possible which was 95% of the total duration of the experiments).

Relative performance of the modes

- PSK was found to be the best-performance mode except at bit rates of less than 0.5 Kb/s.

- ASK-SC performed best when the bit rate was less than 0.5 Kb/s.

- The rectifier-detected mode, ASK-rec.det., performed worse than the PLL-detected modes because the intersymbol interference resulting from the almost continuous presence of multipath meant that the data stream (unequalised) had a high error rate.

- The comparison to rectifier detection showed that the PLL output phase was not necessarily coherent to the phase of the signal multipath. This caused the datum error rate to be slightly higher for the PLL-detected modes than for the rectifier-detected mode.

- Performance of the ASK+C and ASK-SC modes suffered because of the necessary presence in circuit of the dc extractor. This affected the operation of the equaliser.

Effectiveness of the equaliser

- The equaliser was found to significantly improve the performance of the two modes without residual-dc components, PSK and ASK. At the bit-rate region of peak improvement, 2.0 to 2.5 Kb/s, the error rate was improved by a factor of 1.7 for PSK and 1.45 for ASK.

- Calculations of the frequency with which the equaliser improved the error performance (section 4.6) showed that, for PSK and ASK, the improvement took place on 70% to 85% of the test runs made at bit rates between 1.4 Kb/s and 3.6 Kb/s.

Variation of the error probability with bit rate

For the best performance modes the error rate varied
from:

at 3.6 Kb/s, PSK with equalisation, P_e = 0.14

at 1.2 Kb/s, PSK with equalisation, P_e = 0.07

at 0.4 Kb/s, ASK-SC with equalisation, P_e = 0.06

It was shown that the error probability at the lower
rates would improve more by using coding. The results
for employing a simple majority-decision code are (with
a fundamental rate of 3.6 Kb/s using PSK with equalisation,
where the error probability was 0.14)

at 1.2 Kb/s P_e = 0.053

at 0.5 Kb/s P_e = 0.0094

Range of variation of the error probability

PSK bit rate	equalised-data error probability P_e	fraction of total transmission time the error probability was better than P_e
3.6 Kb/s	1.4×10^{-1}	50%
3.6 Kb/s	4.8×10^{-2}	25%
3.6 Kb/s	3.5×10^{-3}	10%
2.1 Kb/s	6.7×10^{-2}	50%
2.1 Kb/s	2.5×10^{-2}	25%
2.1 Kb/s	1.8×10^{-3}	10%
1.05 Kb/s	4.8×10^{-2}	50%
1.05 Kb/s	1.4×10^{-2}	25%
1.05 Kb/s	1.3×10^{-3}	10%
0.6 Kb/s	4.6×10^{-2}	50%
0.6 Kb/s	1.3×10^{-2}	25%
0.6 Kb/s	2.8×10^{-4}	10%

Variation of the error probability with time of day (section 4.7)

No variation was found except for slight increases in the error rate at 07.00 and 16.00. (Sunrise was at 06.50 and sunset at 19.50 during the trials. Measurements were taken between 06.30 and 16.30.)

APPENDIX 4.1

PROBABILITY OF ERROR IN A MAJORITY-DECISION CODE

The probability of there being an error in a block of n bits, $P(m,n)$, is given by the product of the number of ways that m bits can be located in a block of n bits, the probability of m errors, and the probability of (n-m) errors not occurring, i.e.

$$P(m,n) = {}^{n}C_{m} P_{e}^{m} (1-P_{e})^{n-m}$$

where P_{e} is the basic error probability.

In a majority decision code there will be an error provided

$$m \geqslant \frac{n-1}{2} \qquad \text{(n is odd)}$$

Hence the error probability of the coded bit is given by

$P(n,m)$ occurring or $P(n-1,n)$ occurring or $P(n-2,n)$ occurring or .... $P(m,n)$ occurring, i.e.

$$P(n,n) + P(n-1,n)[1-P(n,n)] + P(n-2,n)[1-P(n,n)-P(n-1,n)[1-P(n-2,n)]]$$

$$\text{....... } P(m,n) \ [1-P(n,n)-P(n-1,n)[1-P(n,n)]-P(n-2,n)[1-P(n,n)-P(n-1,n)]...]$$

Fig 4-1.1   Summary of bit-rate values

(a) Pulse stream having 100% roll-off raised-cosine spectrum

Bit rate (kb/s)	Bit-rate reduction factor 1/N
2.4	x1
1.2	x1/2
0.6	x1/4
0.4	x1/6

(b) Pulse stream having 33% roll-off raised-cosine spectrum

Bit rate (kb/s)	bit-rate reduction factor 1/N
3.6	x1
1.8	x1/2
0.9	x1/4
0.6	x1/6

Fig 4-1.2   Summary of modulation and demodulation modes

(1)   PSK   bipolar ASK

(2)   ASK   on - off keying

(3)   ASK-SC   ASK with suppressed carrier

(4)   ASK+C   ASK with sufficient carrier added to remove
                                phase shifts caused by pulse sidelobes

(5)   ASK-rec. det.   ASK detected by rectifier

         Modes (1) to (4) coherently-detected by pll

Fig 4-1.3   Example of multipath-structure recording
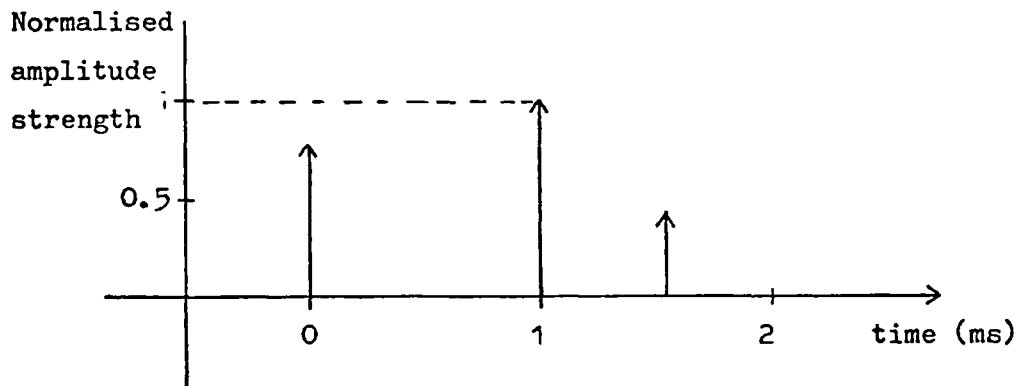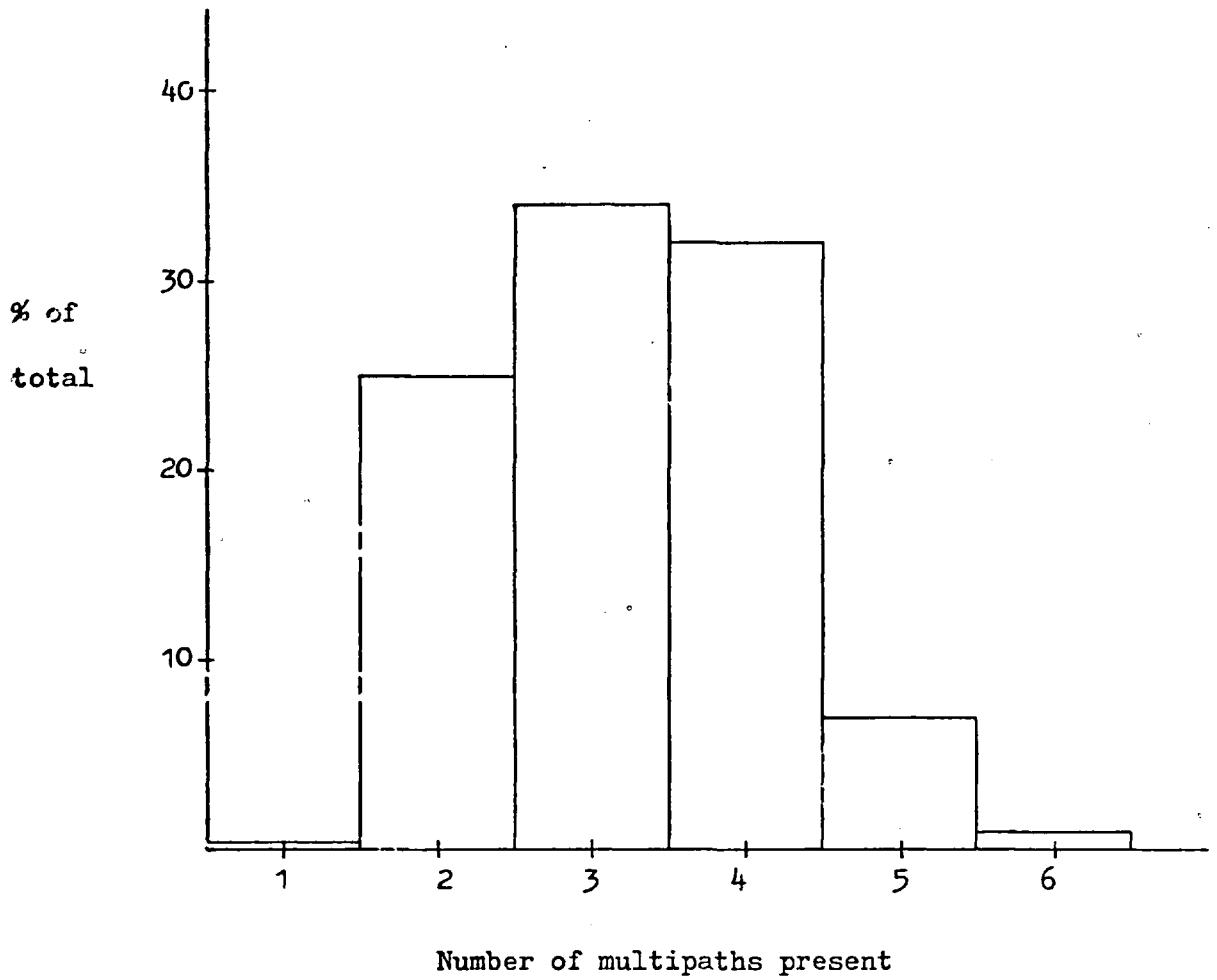
Fig 4-3.1   <u>Histogram of number of multipaths</u>

Number of multipaths present
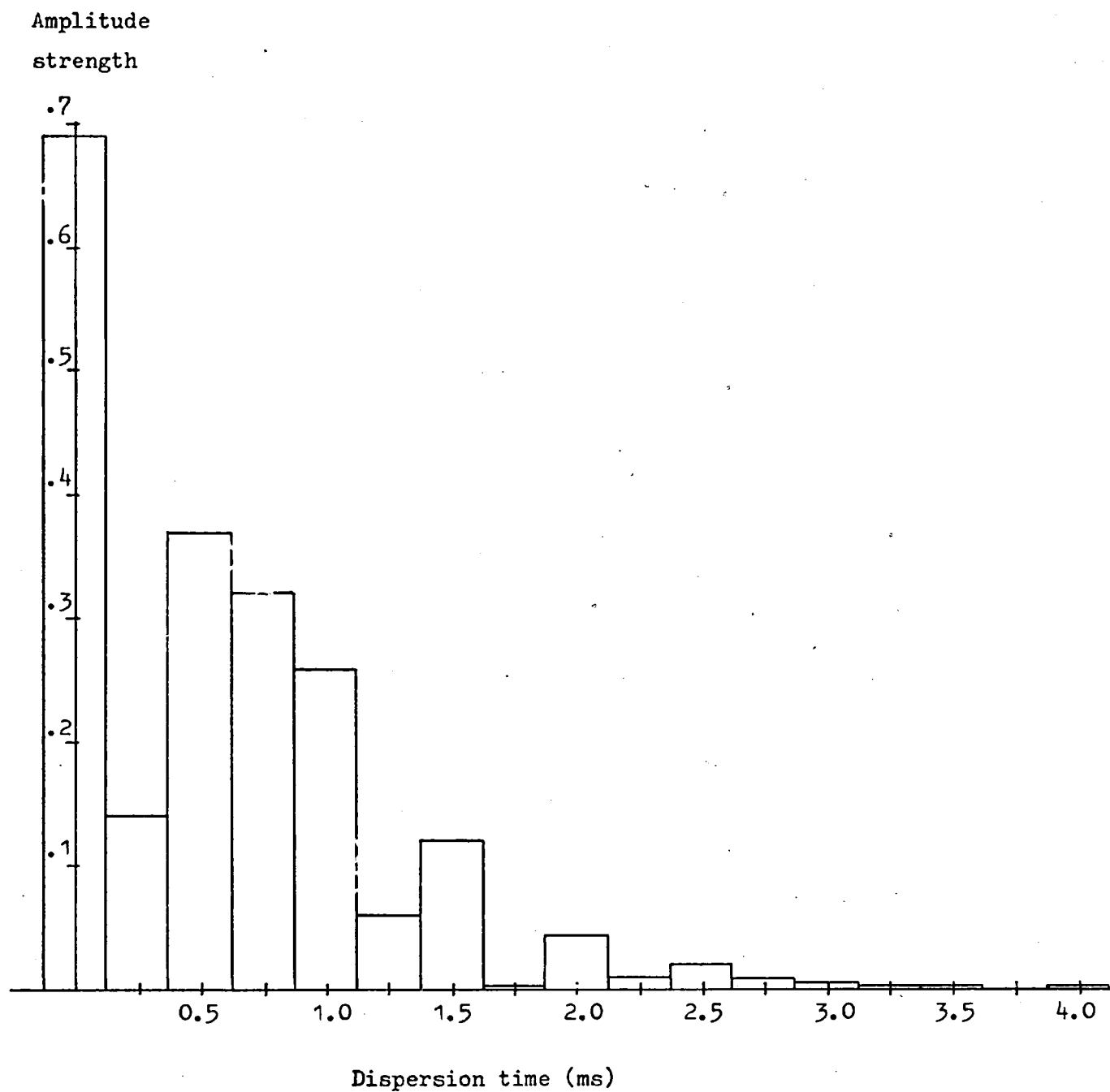
Fig 4-3.2   Distribution of the multipath amplitudes

Dispersion time (ms)

Fig 4-3.3   Position distribution of the significant multipaths

Dispersion time after signal multipath (ms)
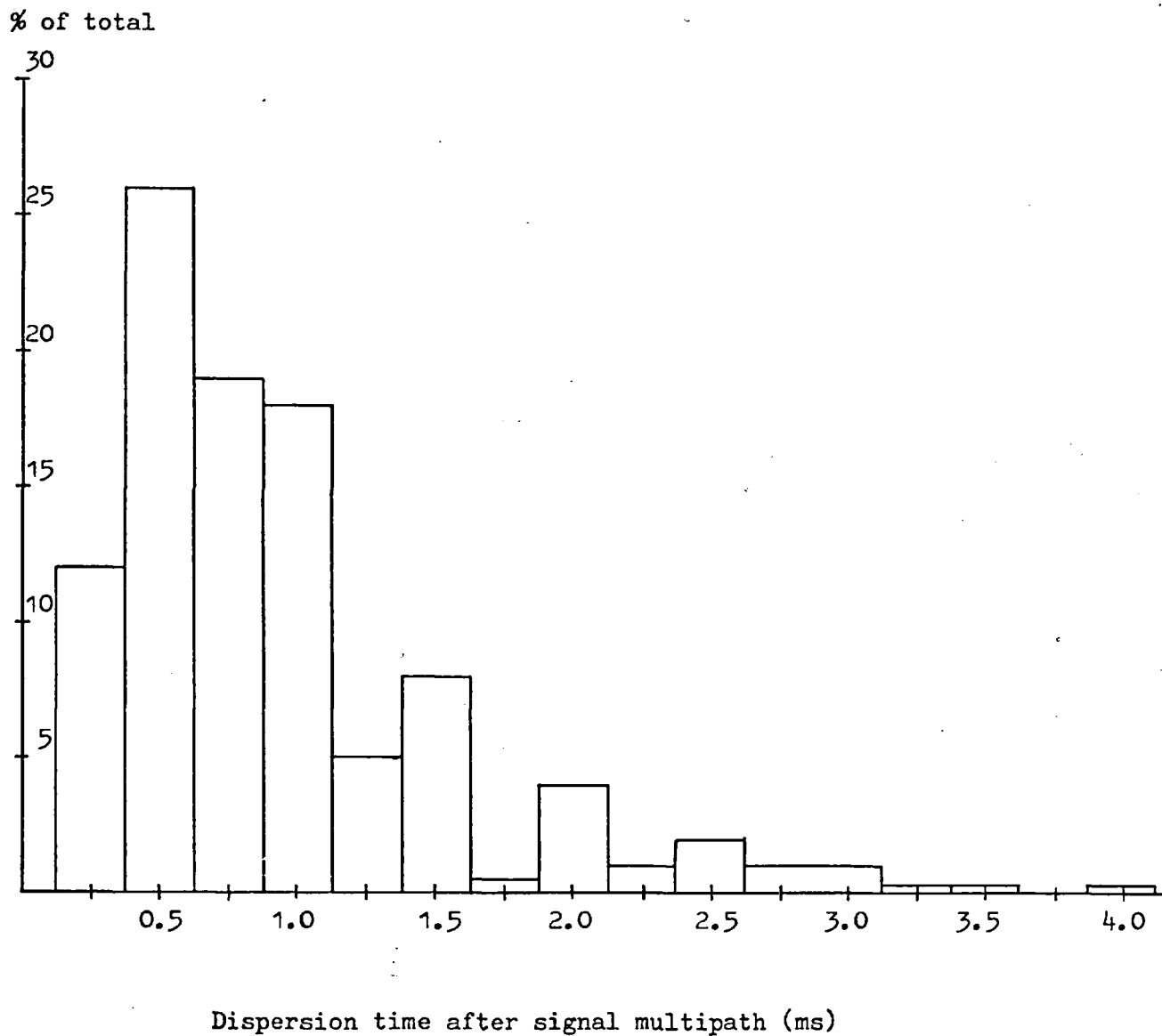
Fig 4-3.4   <u>Average state of multipath structure</u>

signal multipath

Multipath strength

Dispersion time (ms)
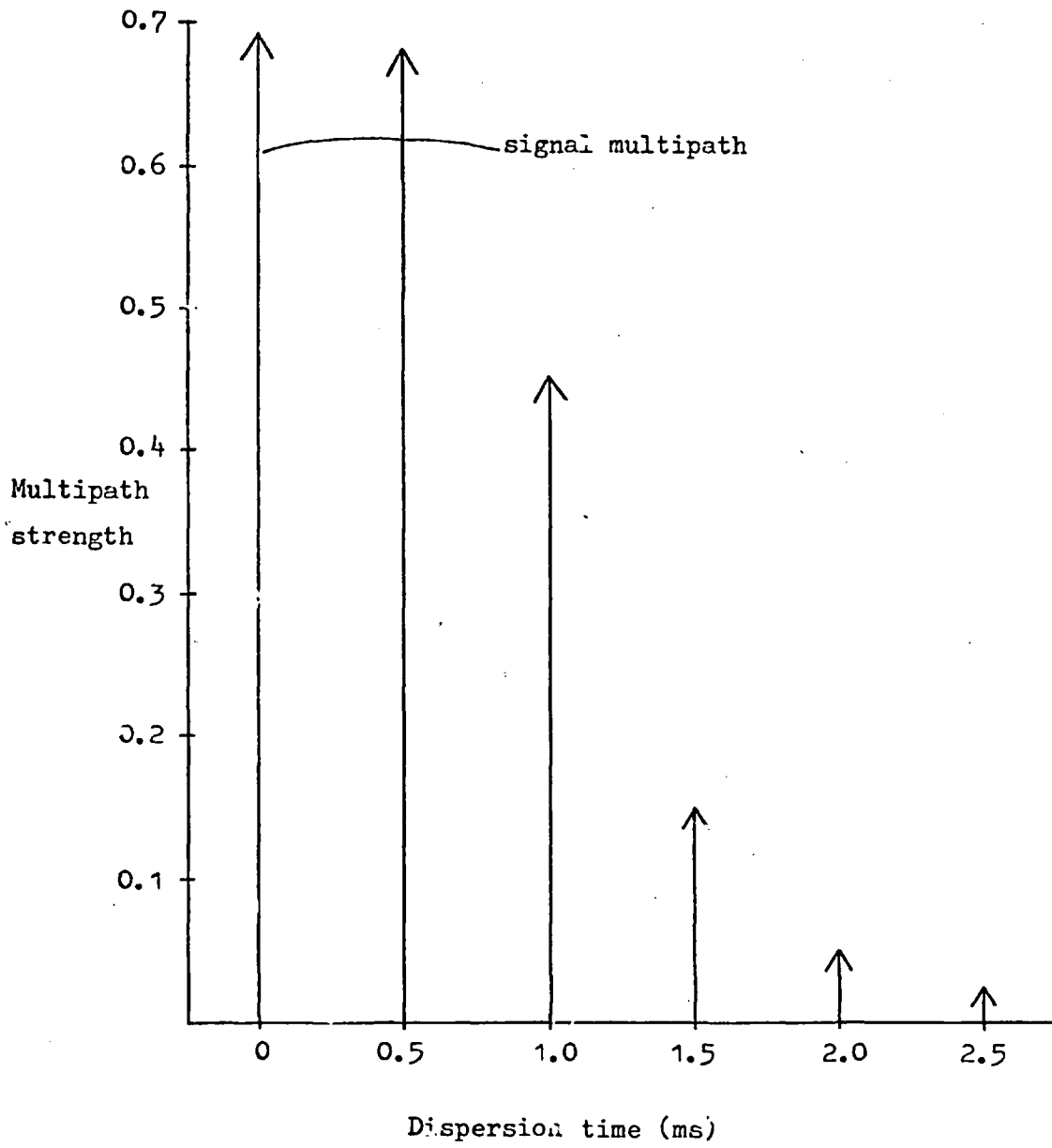
Fig 4-3.5   Position distribution of the strongest multipath

Dispersion time (ms)
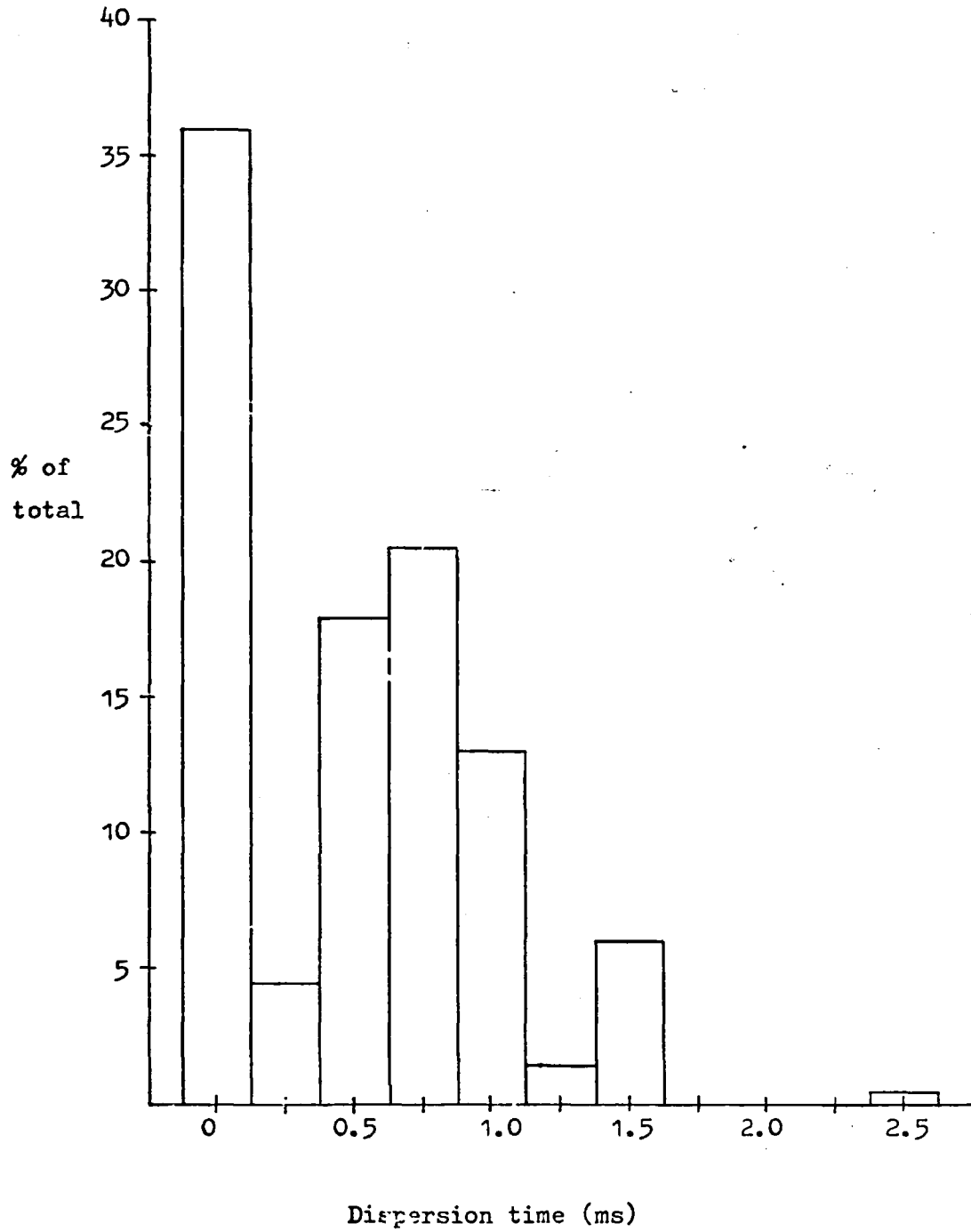
Fig 4-3.6   Distribution of signal-to-noise measurements

% of total

measurements

Signal-to-noise ratio (db)

no reception
was possible
in this region

Fig 4-4.1   PSK - variation in error probaoility with bit rate

U - unequalised data
e - equalised data
d - datum bits

Bit rate (kb/s)

— 33% roll-off
.— 100% roll-off

Fig 4-4.2   ASK - variation in error probability with bit rate
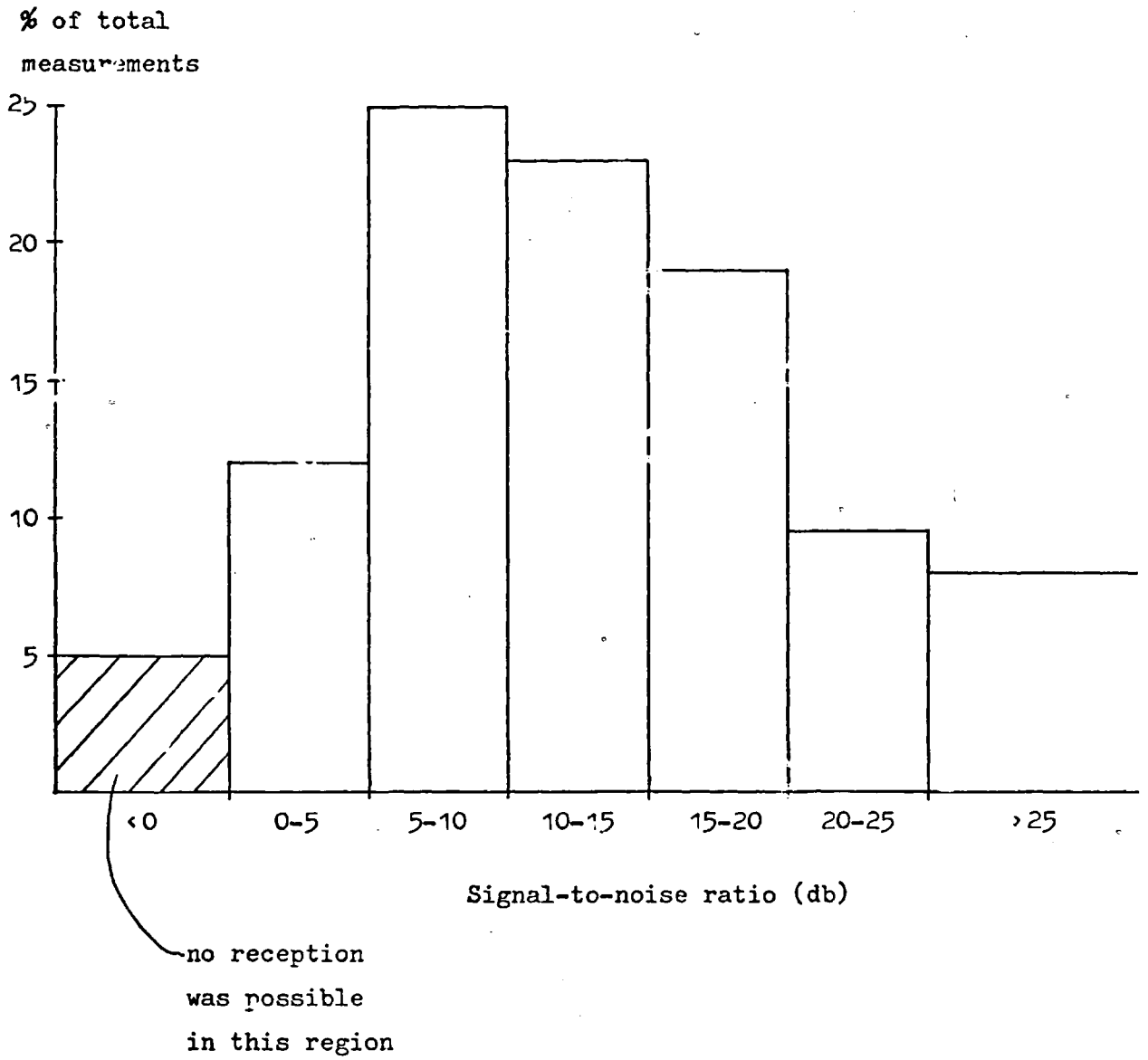

Bit rate (kb/s)

— 33% roll-off
— 100% roll-off
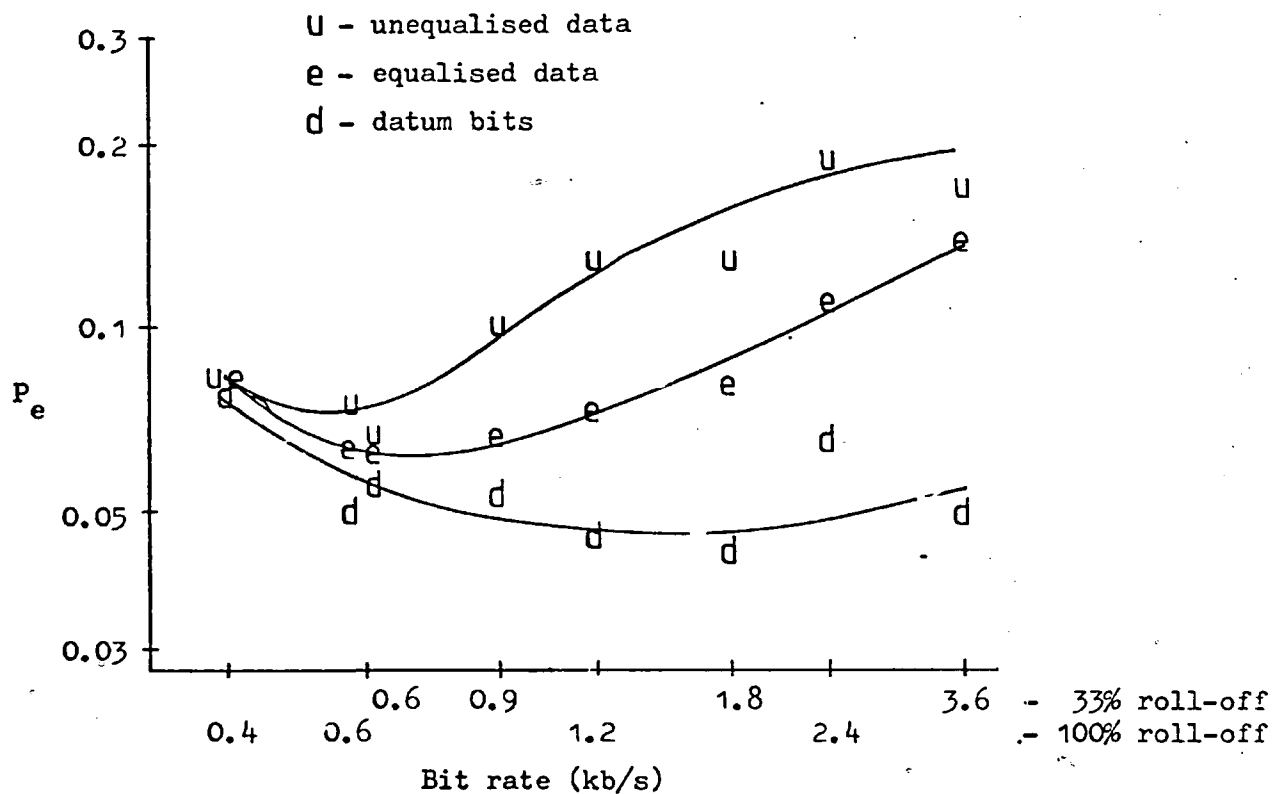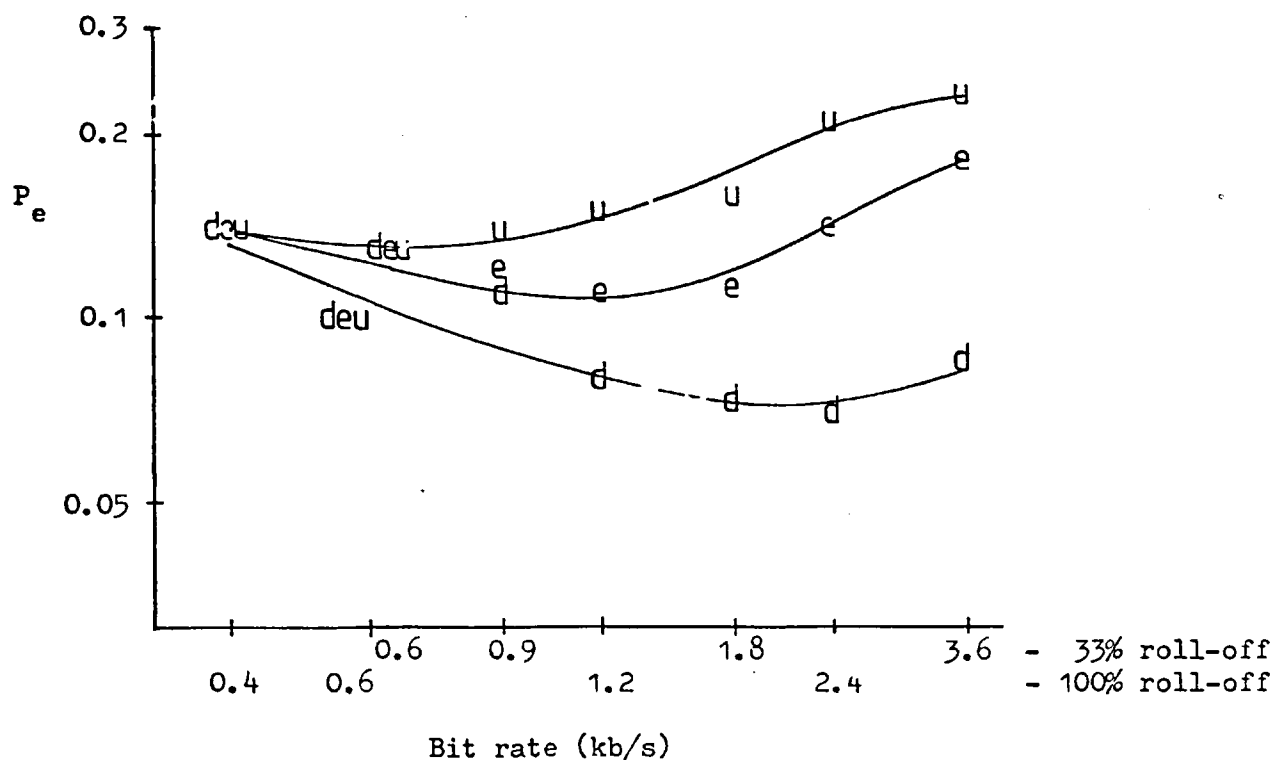
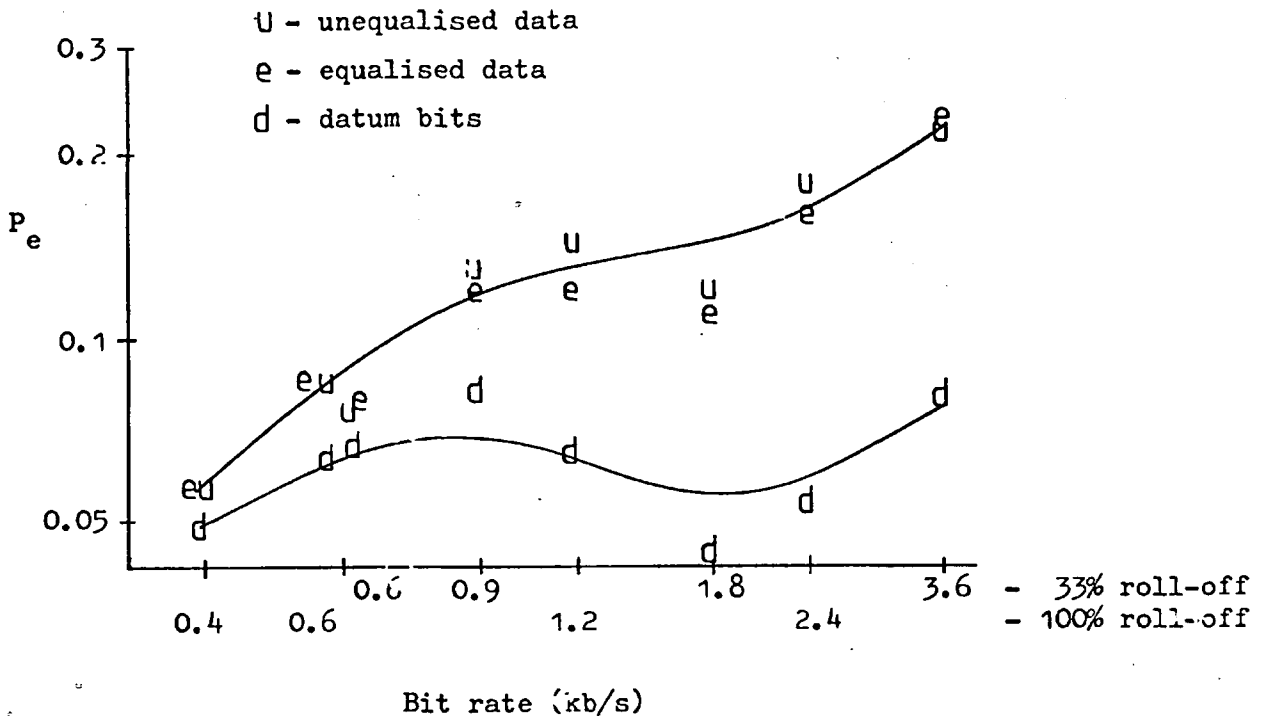Fig 4-4.3   ASK-SC - variation of error probability with bit rate

Fig 4-4.4   ASK+C - variation of error probability with bit rate

Fig 4-4.5 <u>ASK-rec. det. - variation in error probability with bit rate</u>
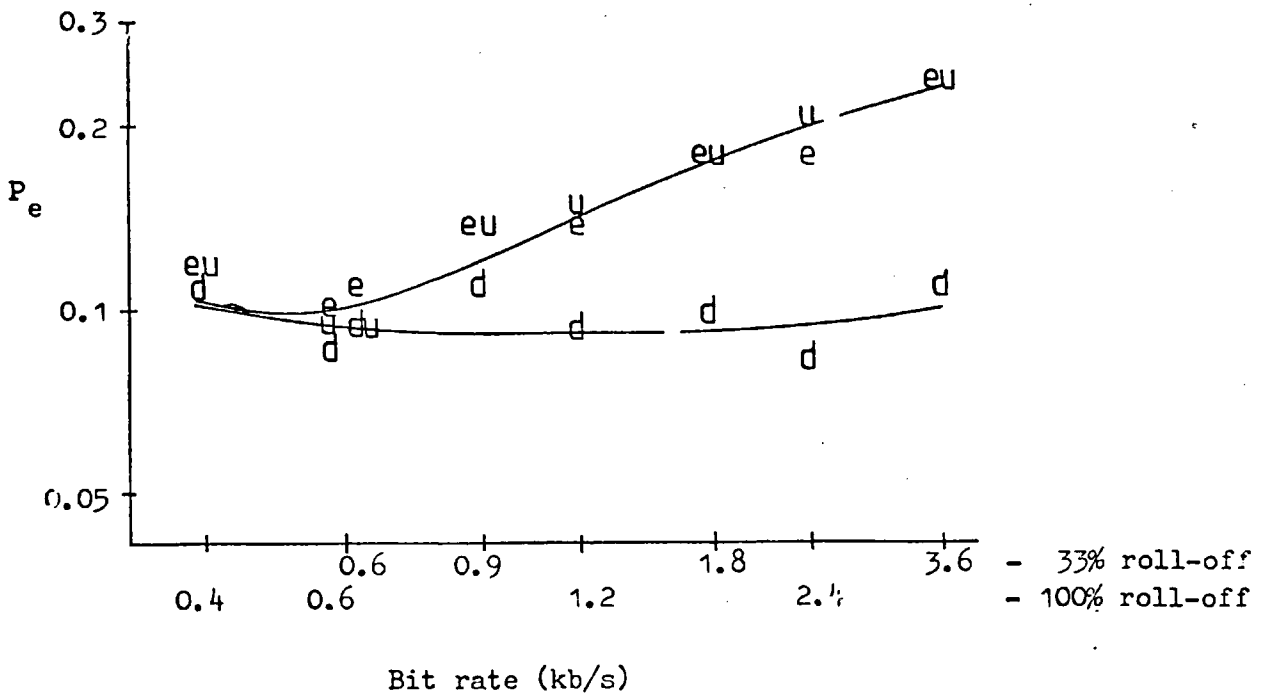

Bit rate (kb/s)
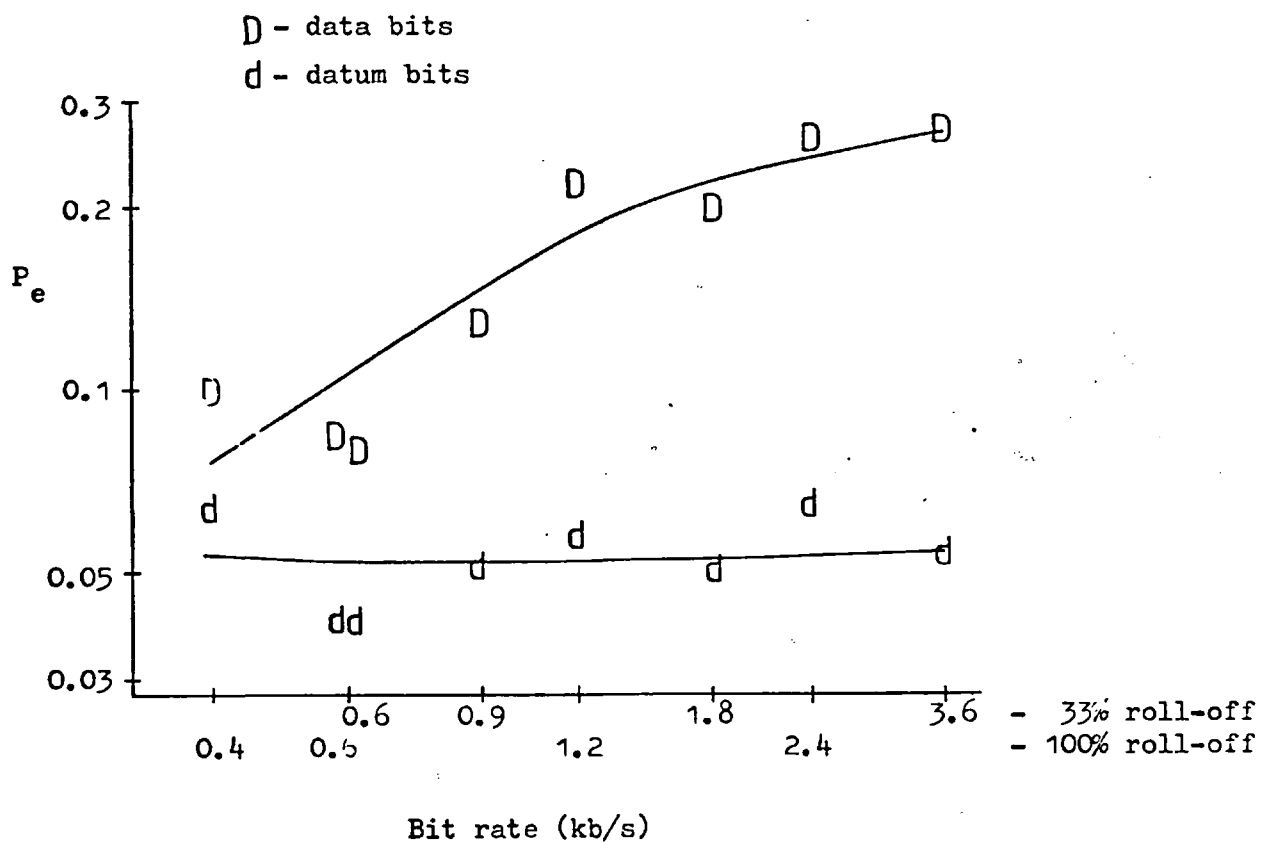
Fig 4-4.6   Phasor diagram of carriers from three-path multipath structure - PLL output also displayed
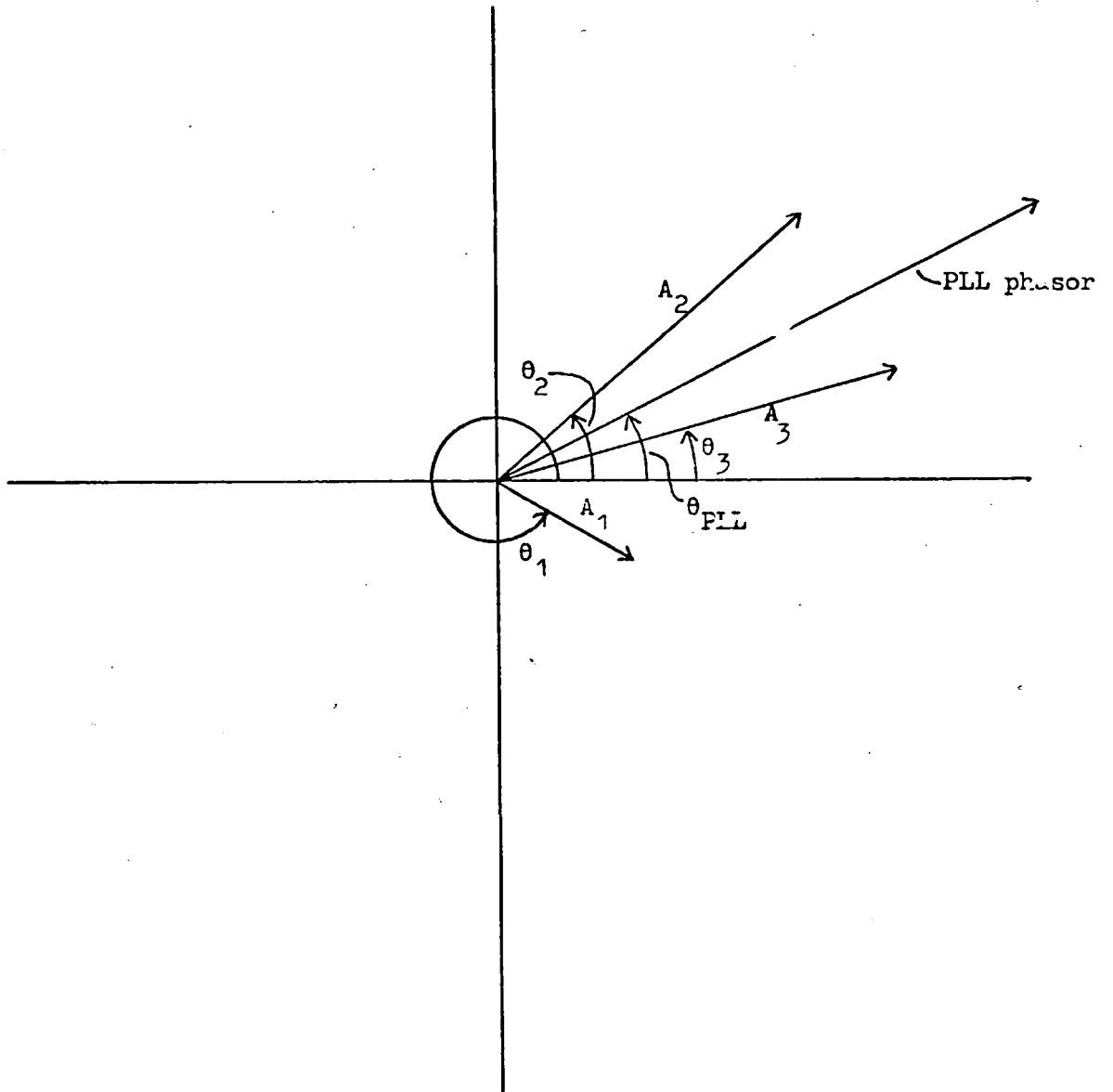
Fig 4-4.7 Reducing the data rate by coding compared with
reducing the raw bit rate

Data rate (kb/s)
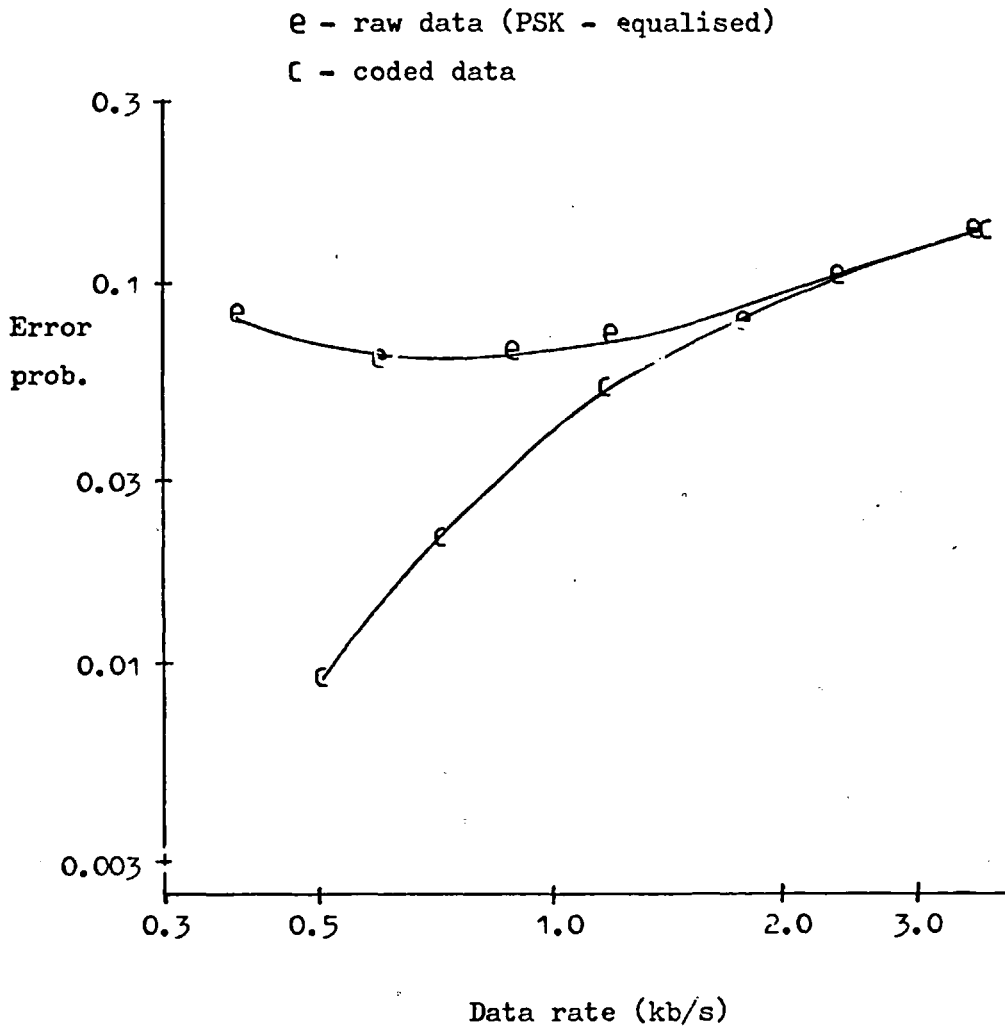
Fig 4-4.8   Order-of-merit table of modulation modes for error performance of equalised data (best first)

Bit rate

3.6kb/s	1.2kb/s	0.4kb/s
PSK	PSK	ASK-SC
ASK	ASK	PSK = ASK-rec. det.
ASK-SC	ASK-SC	
ASK+C	ASK+C	ASK+C
ASK-rec. det.	ASK-rec. det.	ASK

Fig 4-5.1   Error-probability distribution for PSK
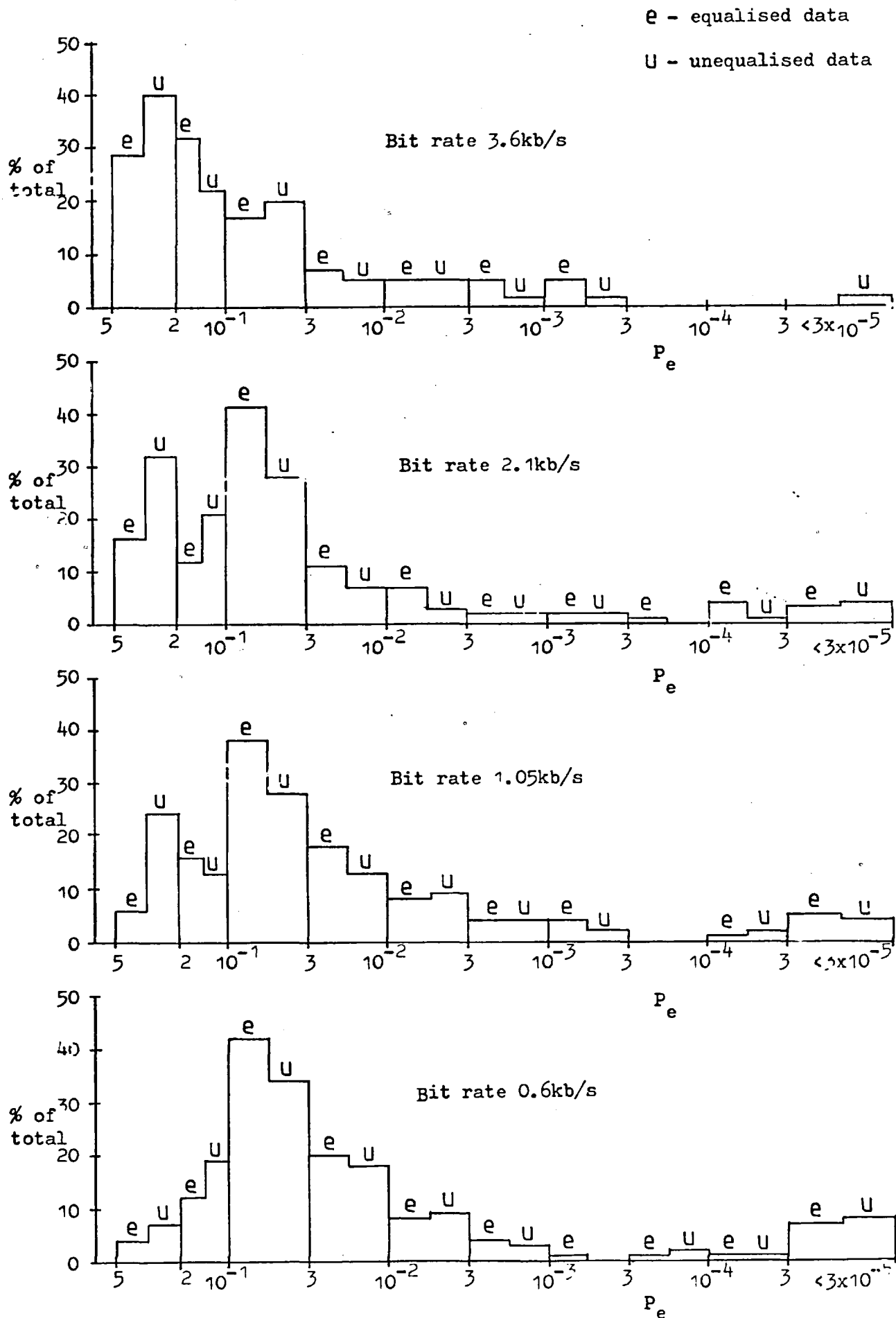
Fig 4-5.2   Cumulative distribution for PSK error rate

Fig 4-5.3

Bit rate 3.6kb/s

Bit rate 2.1kb/s

Bit rate 1.05kb/s

Bit rate 0.6kb/s
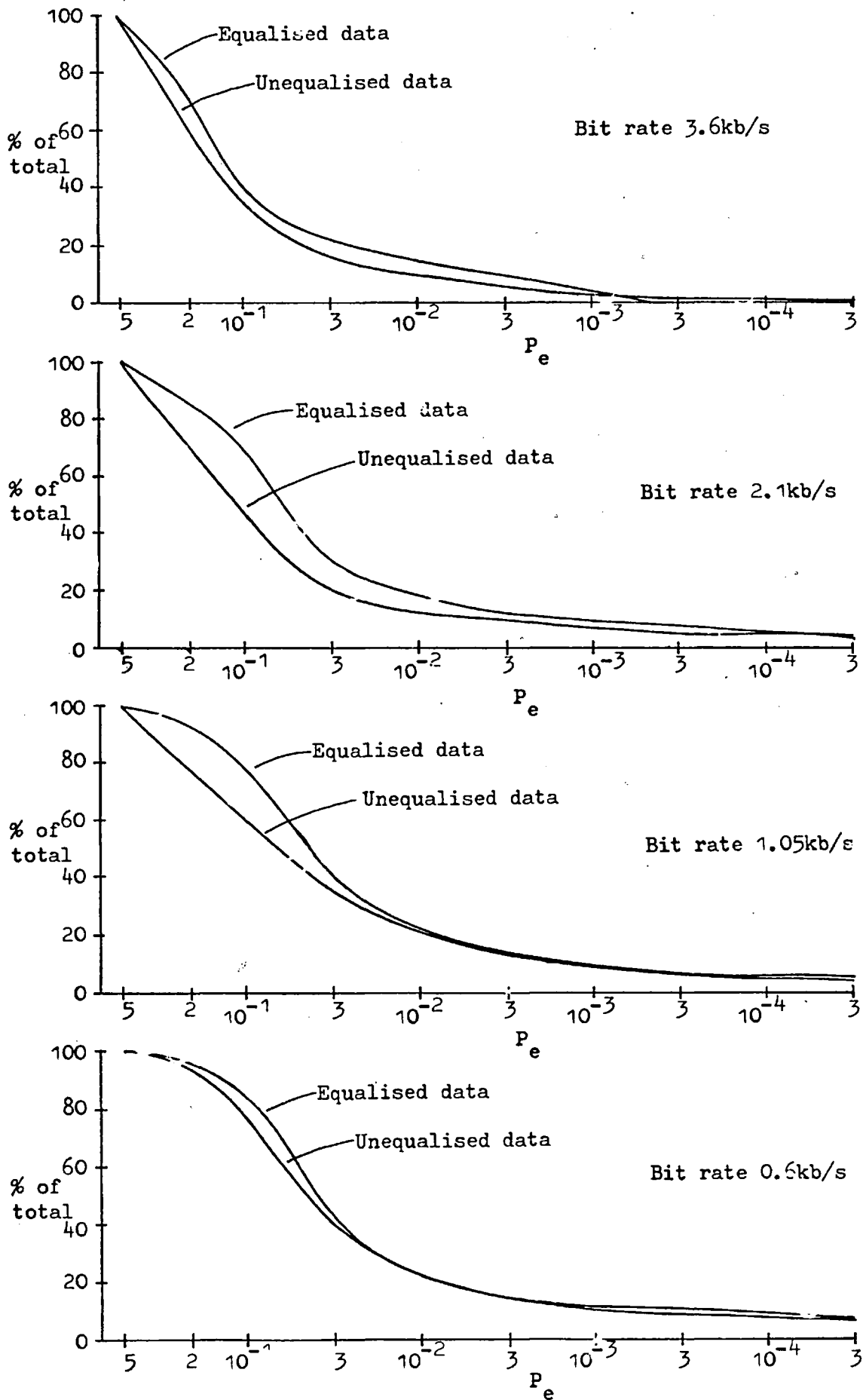
Fig 4-5.4 <u>Cumulative distribution for ASK-rec. det. error rate</u>

Bit rate 3.6kb/s

Bit rate 2.1kb/s

Bit rate 1.05kb/s

Bit rate 0.6kb/s
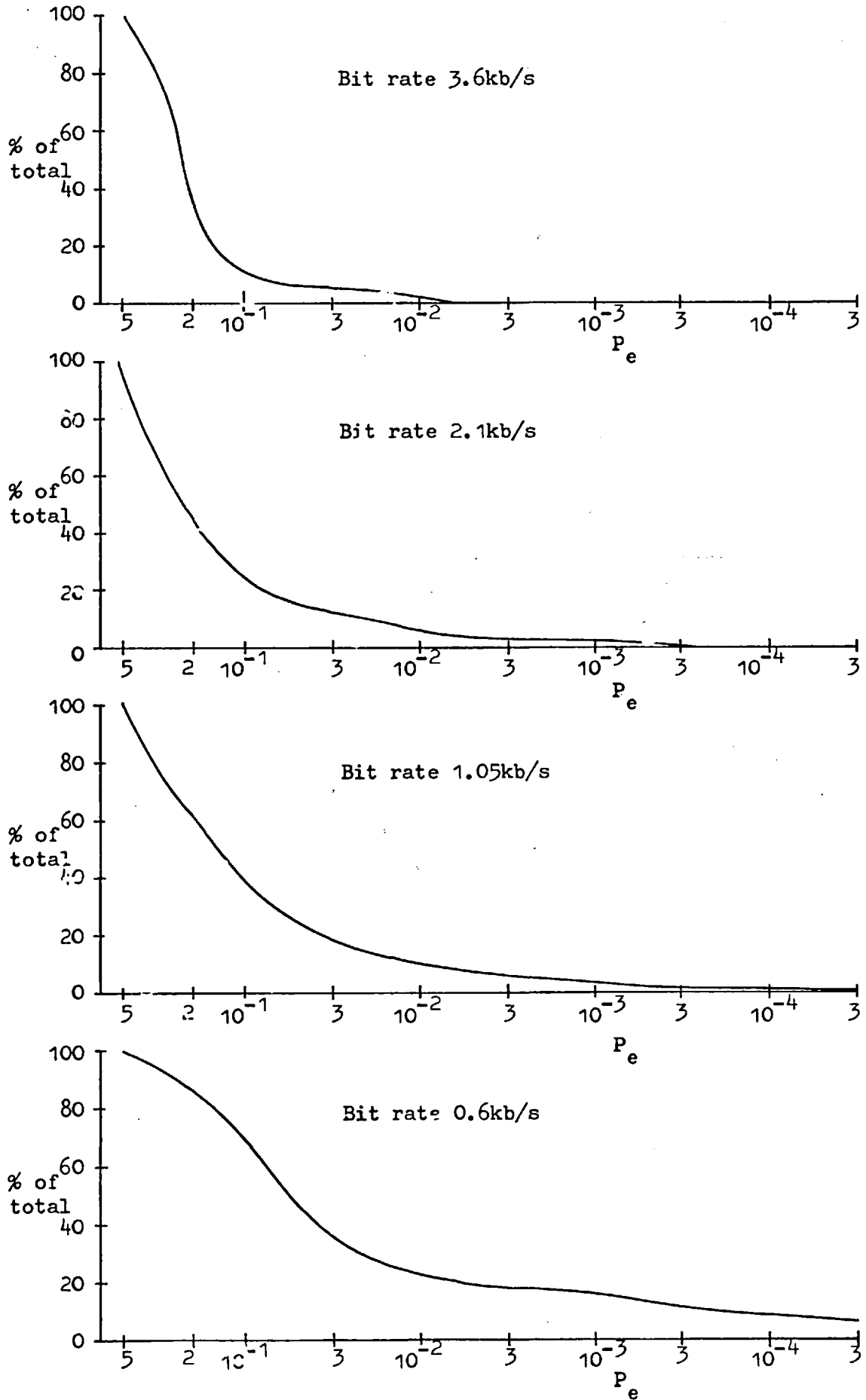
Fig 4-5.5   Error-probability distributions for ASK-rec. det. datum bits

Density distribution

Cumulative distribution
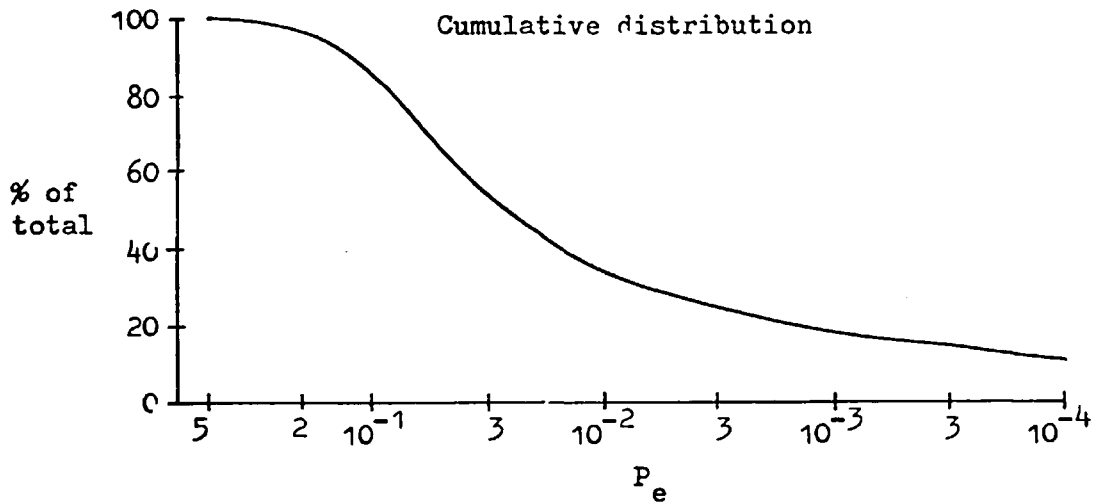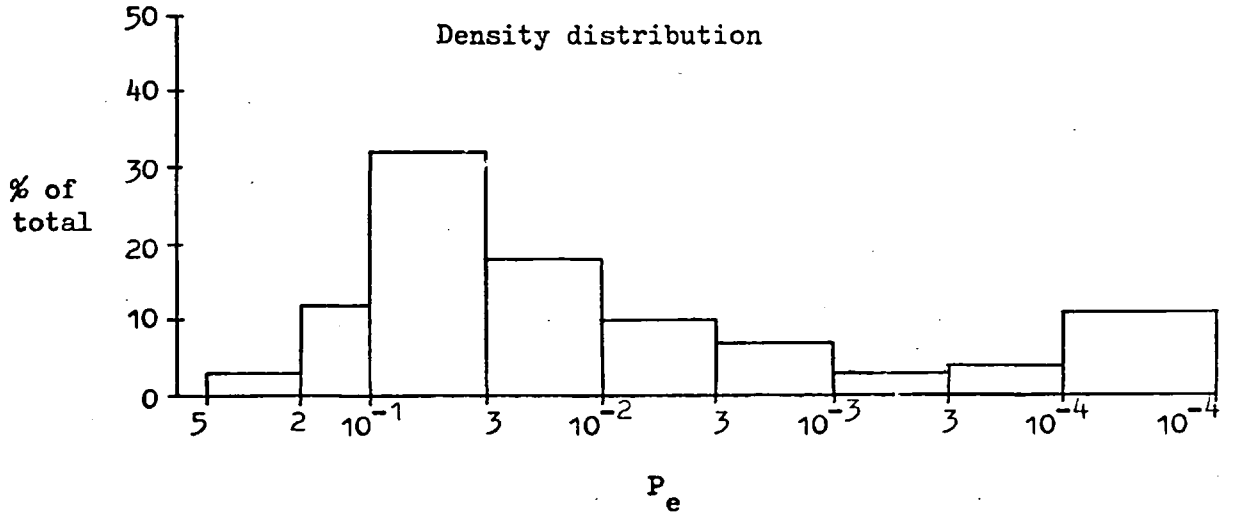
Fig 4-6.1   Histogram for PSK of equaliser error reduction factor B

Equaliser degrades
error performance

Equaliser improves
error performance

$B = \dfrac{\text{Errors before eq.}}{\text{Errors after eq.}}$
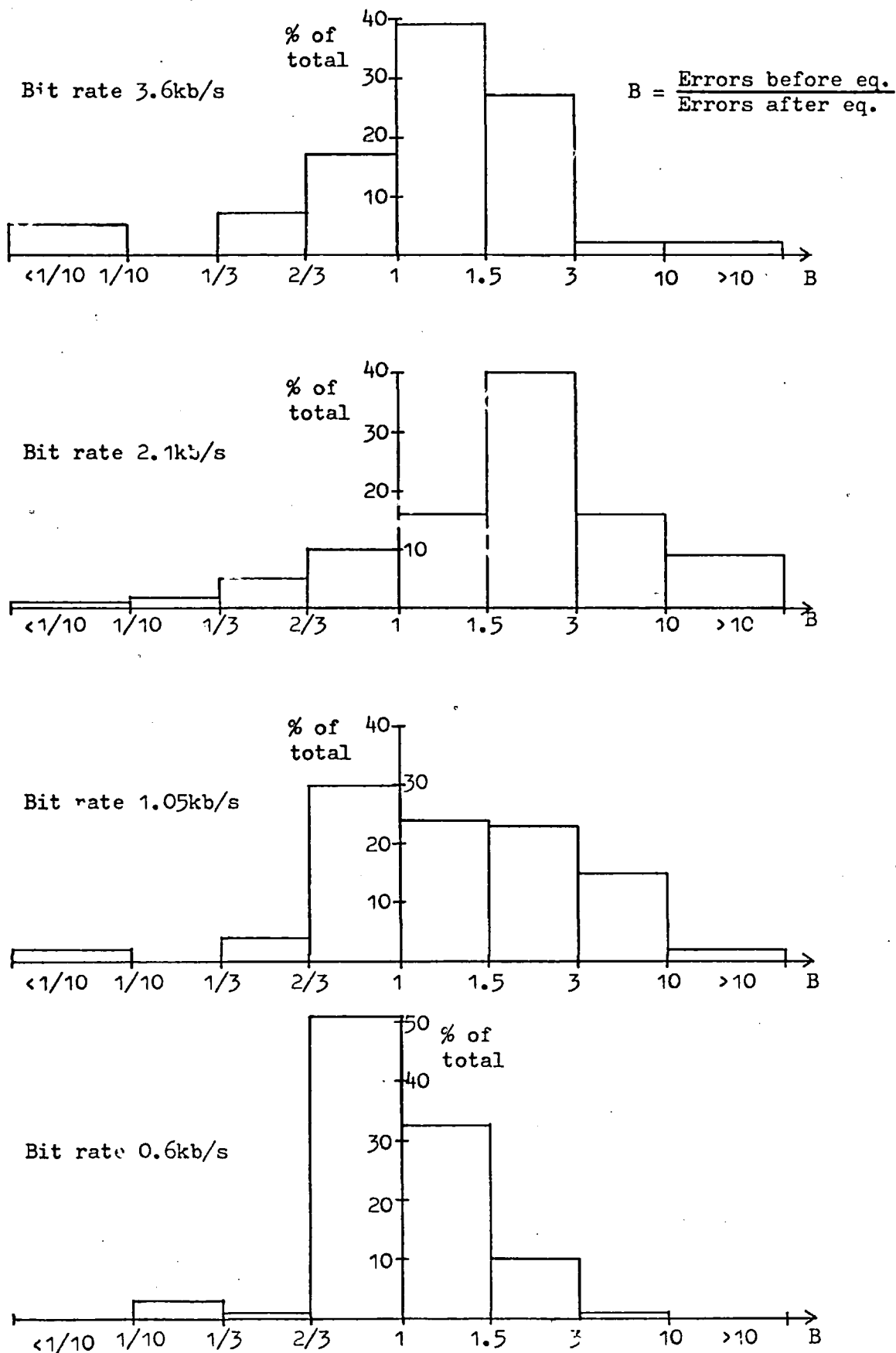
Fig 4-6.2   Histogram for ASK of equaliser error reduction factor B

Equaliser degrades · error performance

Equaliser improves error performance

Bit rate 3.6kb/s

$B = \dfrac{\text{Errors before eq.}}{\text{Errors after eq.}}$

Bit rate 2.1kb/s

Bit rate 1.05kb/s

Bit rate 0.6kb/s
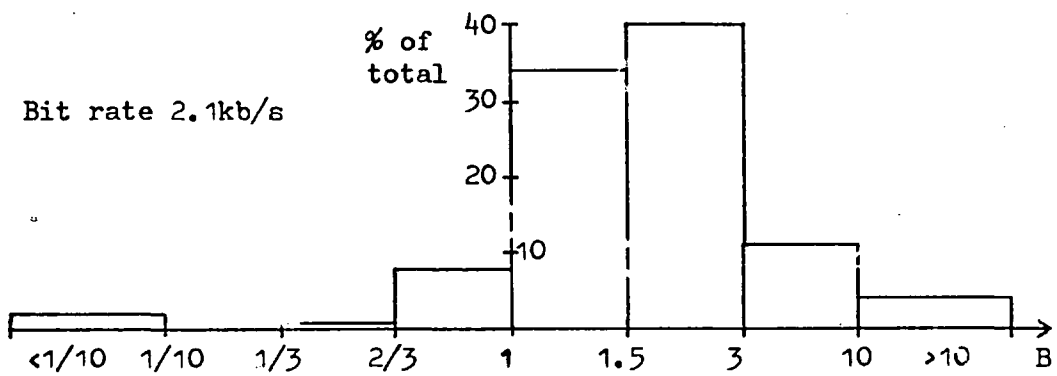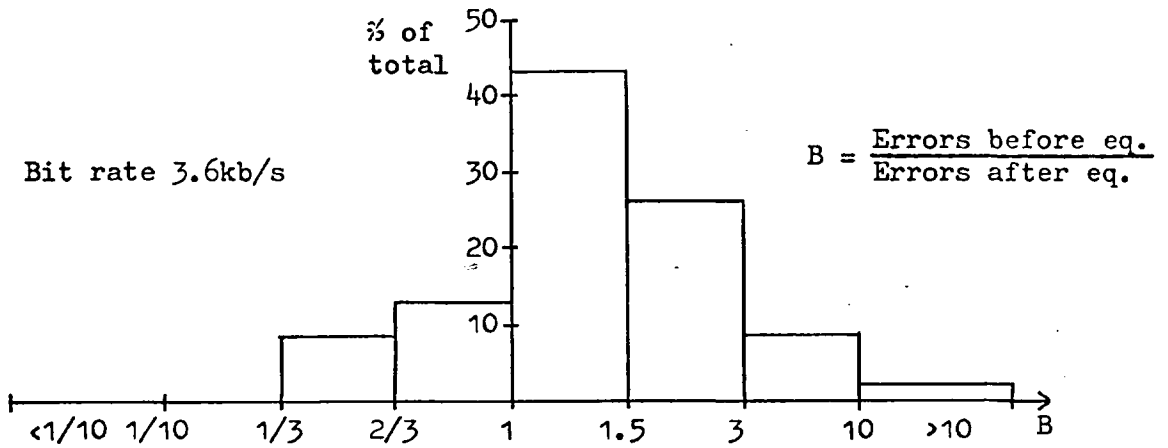
Fig 4-6.3 Fraction of time that equaliser improved error performance for PSK

% of runs for
which equaliser
improved error
performance
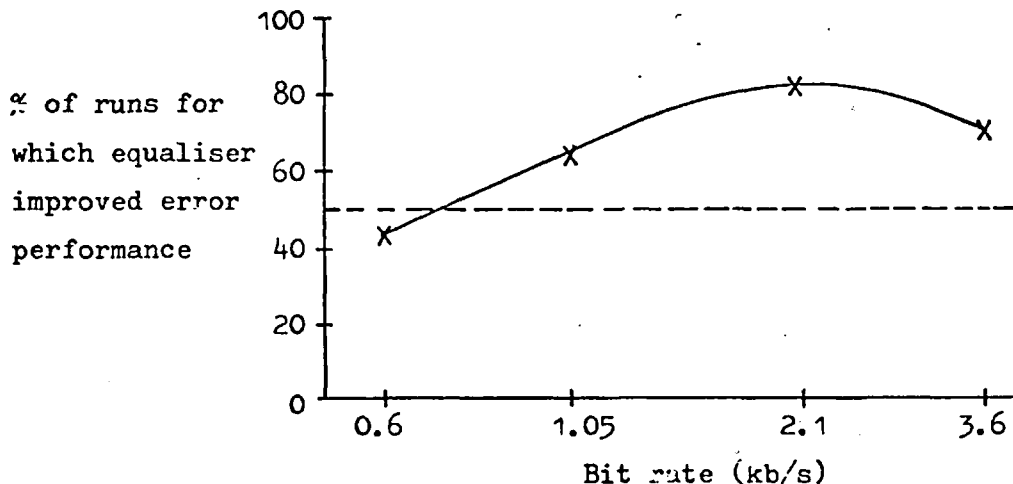
Fig 4-6.4 Fraction of time that equaliser improved error performance for ASK

% of runs for
which equaliser
improved error
performance
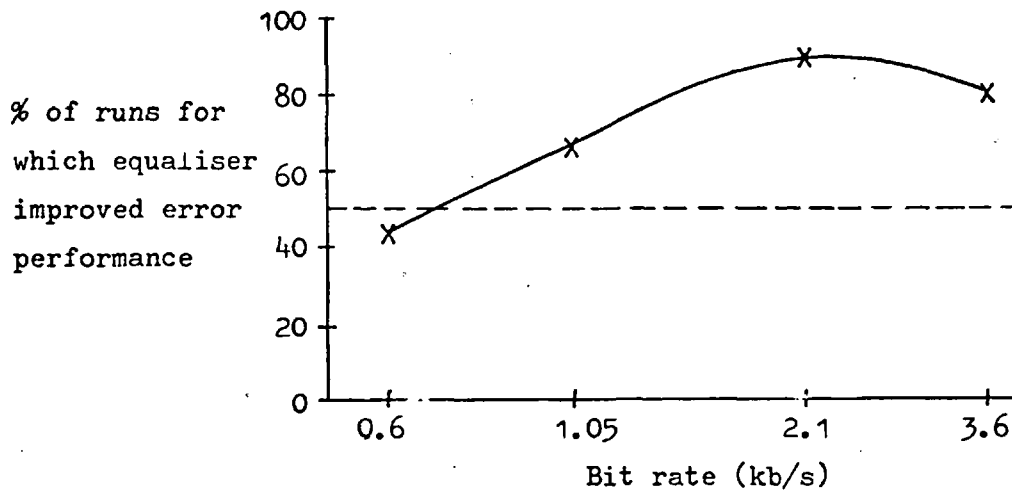
Fig 4-7.1   Variation in error probability with time of day for PSK

- unequalised data
- equalised data
- datum bits
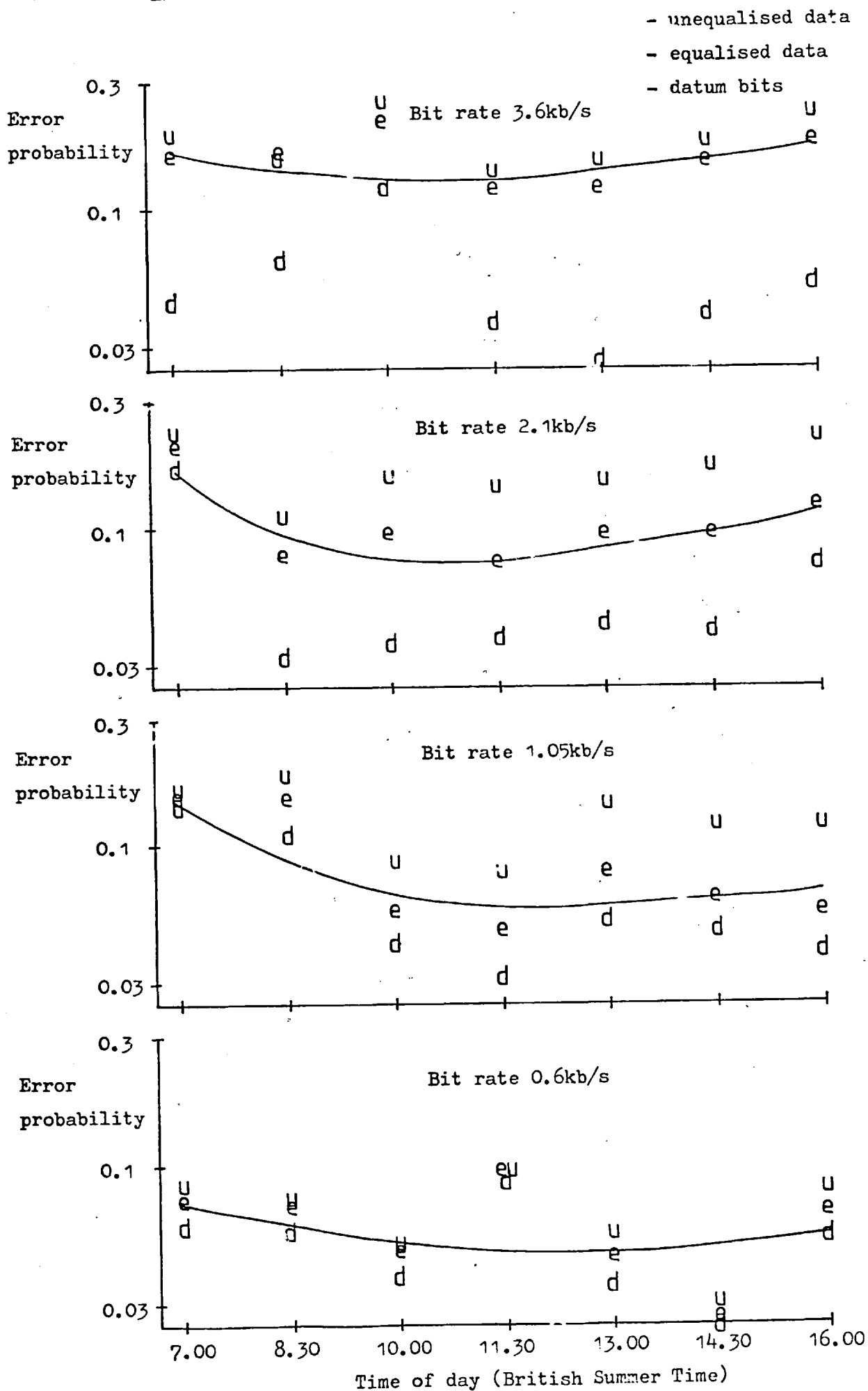
CHAPTER FIVE

CONCLUSIONS

We saw in Chapter One that this research arose out of an appreciation of the history of radio communications. We considered the growth in digital processing that has taken place since the Second World War, and saw the requirement for digital communication channels. We argued that the HF radio channel had a particular role tc play in providing the medium for long-distance digital communications, and observed that, because of channel congestion, the available channel bandwidth was limited.

Having taken the decision to concentrate this work on means of investigating efficient digital communication over HF radio we then, in Chapter Two, considered the properties of HF propagation in order to understand the problems that the digital user faces. We then reviewed the previous work in order to see what lines other researchers had followed. We found that the method of serial binary data transmission had received very little attention, and that there was a need to evaluate the performance of the fundamental systems involved in such a method of transmission.
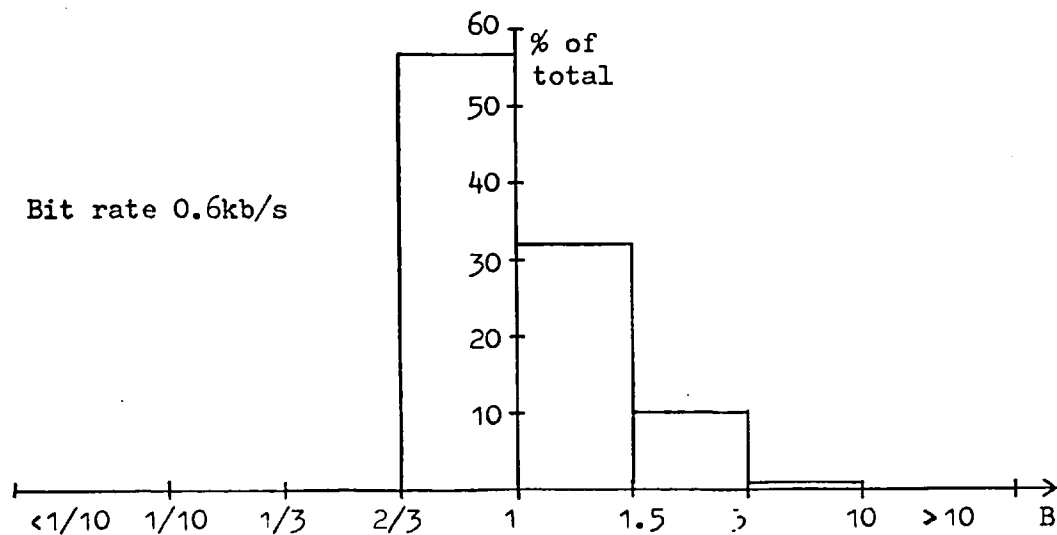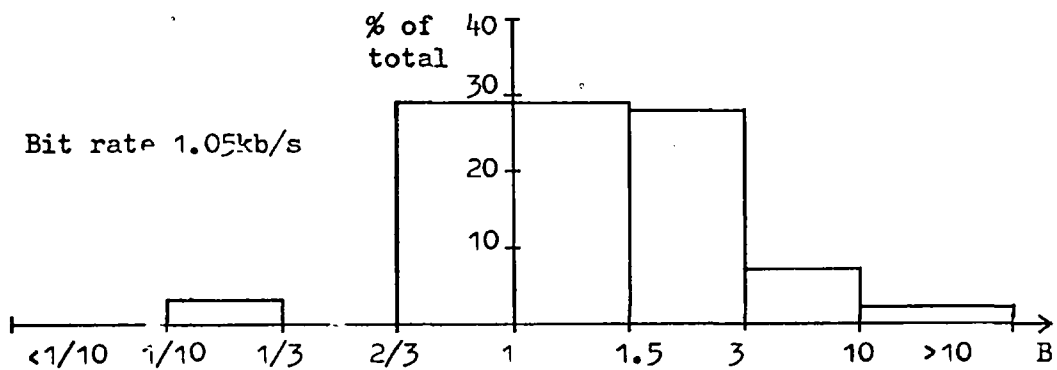
We now proceed to present the conclusions from this work in its evaluation of the techniques involved in serial data transmission over the HF radio channel. In section 5.1 we start by recounting the likely channel propagation conditions that were anticipated in Chapters Two and Three, and we then described briefly the conditions that were encountered in the actual trials. In section 5.2 we then re-iterate the main features of the test system that was developed to

to investigate the techniques of serial data transmission, and we end this section by recounting their performance in the field trials.

We summarise the overall conclusions in section 5.3. This is done by re-stating the precise objectives of this research that were given in section 2.3, Chapter Two, and then answering the four points that were raised one by one. Finally this chapter is concluded in section 5.4 by the presentation of some ideas for further research.

We begin with the channel conditions.

5.1   THE CHANNEL CONDITIONS

We found in Chapter Two that radio propagation at HF frequencies occurs by a process equivalent to reflection from the ionosphere. The "reflection" of the radio signal is at peaks of electron density within the relationship of electron density to height above the earth. These peaks are equivalent to reflecting layers since they are circumferential to the earth, and we saw that two layers in particular are responsible for propagation; the E-layer, altitude approximately 100 Km, and the F_2-layer, whose altitude is approximately 300 Km. We found in Chapter Three that for the radio link to be used in the trials, which had its transmitter and receiver separated by 1000 Km, propagation was possible from both the E-layer and the F_2-layer. Since further reflecting modes can exist by the ducting of the signal between the layers and the earth's surface it was expected that multiple-mode, or multipath, propagation would be present in the experiments. (The bearing of this phenomenon on the receiver operation is reviewed in the next section.)

The measurements made in the trials of the channel conditions, which are displayed in Chapter Four, Figs. 4-3.1 to 4-3.5, showed that the channel was predominantly (99.5% of the time) in a multipath state. There were on average three strong multipaths (the first three to arrive at the receiver) dispersed over a time of 1.5 mS. In addition there were sometimes two or three weaker ones spread over a further 1 mS interval after the third strong path (Fig. 4-3.4).

In order to facilitate the operation of the receiver the
assumption was made about the multipath structure that the strength
of the first multipath to be received would always be significant
so that it could be treated as the signal path. The signals from
the remaining multipaths would then appear in the channel impulse
response entirely as post-echoes. The results justified this assumption.
The first multipath was found to be the strongest 35% of the time
whilst on average it had a strength factor of 0.7 of the strongest
path. Short-term fluctuations in the strength of this (and the other
multipaths) was estimated by observation to be 6dB to 15dB, implying
that there was usually a strong specular component.

The background noise was observed to consist mainly of
interference from other users. The signal-to-noise ratio was
estimated periodically and the results are shown in Fig. 4-3.6,
Chapter Four. It was found that the average signal-to-noise ratio
(defined as the ratio of the main pulse amplitude to the noise
variance) was some 10dB to 15dB. Variations on this average due to
fluctuations in the individual multipath strength, in the flat fading
component, and in the level of background noise, were from -10dB to
+20dB. Reception was not possible for 5% of the duration of the
experiments; this was when the signal-to-noise ratio was too low
(less than 0dB), usually because of either a particularly strong
interfering signal or else high-level ionospheric noise apparently
associated with local electrical storms.

Having summed up the state of channel conditions that were
prevalent in the experiments we now turn to the techniques of serial
data transmission used by discussing the performance of the test
system.

5.2　THE TEST SYSTEM

The test system was designed to answer the queries about serial data transmission raised at the end of Chapter Two. To reiterate, these were:

(1) What is the best form of modulation - PSK, ASK, or FSK

(2) What is the best form of demodulation - Coherent or incoherent

(3) How does the error rate vary with bit rate

(4) How useful is adaptive equalisation in combating the effects of multipath

We proceed in this section (in 5.2.1) by reviewing the main features of the test system. Then, in section 5.2.2 conclusions are given with respect to the performance of the various aspects of the test system.

5.2.1　The Design of the Test System

We first review the design of the test system with regard to the queries raised above about serial transmission.

The type of modulation

Four types of binary modulation were employed:

PSK　-　this was realised using bi-polar ASk

ASK　-　on-off keying of the carrier using pulses having raised-cosine spectra

ASK-SC - a residual-dc component was subtracted

from the ASK pulse stream such that at

baseband the "1" and "0" levels were equally

spaced on either side of zero

ASK+C - sufficient carrier was added to the ASK pulse

stream to prevent the pulse sidelobes from

causing zero crossings in the baseband waveform.

FSK was not used because a theoretical analysis in Chapter

Three showed it to have an excessive bandwidth requirement for serial

transmission, and also because its demodulation process was non-linear

and would not therefore have left it amenable to equalisation.

The type of demodulation

All four modes mentioned above were demodulated coherently

by phase-locked loop. In addition the ASK mode was demodulated

incoherently by full-wave rectification.

The variation of the bit rate

The method used to lower the bit rate from the maxima

determined by the two types of pulse spectra and by the channel bandwidth

was to keep the pulse width constant (and therefore narrow) and to

leave blanks after each bit. It was felt that the method of lowering

the rate by pulse widening would have the pulses vulnerable to the

effects of multipath causing distortion. It was felt that this would have

outweighed the advantage of being able to lower the noise level by

narrowing the filter bandwidth.

The range of variation of the bit rate in the test system by the method of pulse blanking was from 0.4Kb/s to 3.6Kb/s.

Incorporation of an equaliser

It was found during a series of preliminary trials that symbol smear from multipath caused high error rates at bit rates where the period between bits was less than the multipath dispersion time. This effect restricted the transmission rate to 250 b/s. In order to combat this an equaliser was incorporated into the test system. It was of the decision-feedback type and, because the signal from the first multipath to arrive was taken as the main signal, it operated on post-echoes only.

We have reviewed above the main features of the test system with regard to the evaluation of the objectives of this work. One further important aspect of the test system that should be stressed was the adoption of a frame method of data transmission.

The frame method of data transmission

One of the main features of the hardware system in operation was for it to have the data stream periodically interrupted in order that an isolated pulse (termed the set-up pulse) could be transmitted. The gaps between the end of the previous data block and the set-up pulse, and between the set-up pulse and the start of the next block, were made equal to the sum of the maximum expected channel dispersion time plus one pulse's width.

This frame format reduced the effective channel capacity by 25%. However, together with the fact that the first multipath to arrive was to be treated as the signal path, it meant the set-up pulse was received free from intersymbol interference and could be put to the following uses:

(1) Bit synchronisation could be obtained by a threshold transition detector type of phase-locked loop triggered from the leading edge of the set-up pulse. This part of the pulse was not distorted by multipath since the pulse width was always sufficiently narrow to resolve the multipaths.

(2) Level detection for decision-making purposes could be established by measuring the pulse height by means of a sample-and-hold and low-pass filter arrangement as described in Chapter Three. Likewise, any residual-dc in the demodulated waveform (present in the ASK-SC and ASK+C modes) could be removed by sampling just before the set-up pulse.

(3) The reception of a known pulse meant that a polarity reference could be established at the receiver for PSK, obviating the need to differentially encode the data with the consequent doubling of the error rate caused by the decoding process.

(4) The channel impulse response could be determined for the equaliser by a series of the level detectors (of the type used in (2)) operating in the gap between the set-up pulse and the start of the next data block.

(5) The reception of two bits per frame* free of inter-symbol interference (the set-up pulse and the first bit in the next data block) was used to establish

---

*A third bit was available in the case of ASK, a "0". The data stream was decision-sampled just before the set-up pulse for this bit.

the error rate termed the datum error rate, so termed because errors here were caused by effects other than intersymbol interference.

The conclusions with regard to the performance of the test system are now presented.

5.2.2　The Performance of the Test System

We first of all deal with the phase-locked loop and the equaliser. We then discuss the modulation modes and how the error rate varied with bit rate. Finally, we compare these results to some of those obtained in other work.

The phase-locked loop

The phase-locked loop was found to operate satisfactorily in so far as locking to the average phase of the carrier from the sum of the multipaths. Some instability was present at low bit rates for the ASK modes (and to a lesser extent the PSK mode). This was because of the low average-carrier-to-noise ratio. For ASK at the "x 1/6" bit-rate reduction factor the ratio of the period of no carrier to the period with carrier was on average two to one. However longer periods without any carrier would occur because of the pseudo-random nature of the test bit stream.

The overall disadvantage of the coherent-demodulator system adopted was when the signal multipath had a strength weaker than the remaining echoes. Then the carrier from the signal multipath would not necessarily be in phase with the PLL output which would lock onto the average carrier phase from all the paths. Hence the demodulated signal-to-noise ratio was unnecessarily low when the signal-carrier phase and the PLL output phase differed by angles approaching $\pi/2$, resulting in sub-optimum performance. This situation was not the

fault of the PLL as such; as stressed it was perfectly stable in locking to the average phase of the received signal. The situation would have been the same if the technique of extracting a carrier pilot tone had been used. It was more of a demodulation fault, and a method of overcoming it is given in section 5.4 where ideas for further work are presented. It is estimated that the error probability figures obtained in the experiments could be improved upon by a factor of from two to five if such a method could be evolved.

The equaliser

The assumption made in the design of the receiver that the first multipath to arrive could always be used as the signal path was found in the experiments to have been justified. The equaliser was therefore able to afford improvement to the error rate for ASK and ASK modes. The improvement was most marked at around 2.1 Kb/s, where the error rate was better by a factor of 1.7 for PSK and 1.45 for ASK At faster bit rates the signal-to-noise ratio was not sufficiently high for the counter-action by the equaliser of intersymbol interference to be greater than the added noise it introduced because of the extra taps switched in at these rates. Some ideas for lowering the amount by which it amplifies the noise are given in section 5.4.

As far as the modes with a residual-dc component are concerned, ASK-SC and ASK+C, the equaliser gave no significant improvement in error performance. This was because of the effect of the dc extractor on the operation of the equaliser.

The modulation modes

The results showed that PSK gave the best overall performance. Its advantage of having both twice the eye-opening of ASK and also no residual-dc component was decisive as was the fact that the data format meant that differential encoding and decoding was not necessary.

ASK proved to the best of the remaining modes at medium to high data rates whilst the greater signal power of ASK-SC meant it had the best performance overall at low rates. The ASK+C mode was included to see whether cycle skipping by the PLL had a significant effect on the error rate. This was done by comparing ASK+C, for whom the PLL had to be stable in only one phase, with ASK, for whom two-phase stability was required. The results show that ASK+C gave inferior performance to ASK and so the conclusion is that cycle skipping was not a problem.

In comparison to the ASK (PLL detected) results the rectifier detection of ASK gave poor performance because of the predominantly multipath nature of the channel. Rectifier detection prevents the implementation of a conventional equaliser because it is a non-linear process. However the datum error results were better for ASK-rec.det. showing the advantage of its asynchronous nature as compared to the method of coherent demodulation employed.

Variation of error rate with bit rate

The results showed that,using the method of pulse blanking, the error rate only reduced by a factor of two or three when the bit-rate was lowered from 3.6 Kb/s to 0.4 Kb/s. This is for PSK, the best mode, for data after equalisation. The error rate could have been expected to reduce with bit rate for equalised data because less equaliser stages were switched in. The results were affected

by the fact, as mentioned above, that the coherent demodulator's output was not necessarily in phase with the signal's carrier phase, therefore raising the errror rate at all bit rates. It was shown in the results that the utilisation of coding would be a superior way to get an improved error performance at the lower bit rates with the results obtained.

The alternative method of reducing the bit rate by widening the pulse has the advantage that the noise level is reduced because the receiver bandwidth can be lowered. However it cannot be expected to perform well under the type of multipath conditions experienced. Let us take the multipath structure as shown in Fig. 4 3.4, Chapter Four and consider only the first two paths. Let us use, for example, a 0.5 Kb/s bit rate. Here the width of the pulse's main lobe using a 100% spectrum roll-off factor and allowing for the frame system of data transmission is 2.4 mS. Then it is apparent since the first two paths in Fig. 4-3.4 are separated by only 0.5 mS that pulses will overlap and so pulse distortion will be severe. These effects are likely to more than negate the four-fold drop in the noise level*.

Comparison of these error measurements with those from other research

We refer back to section 2.2.3, Chapter Two, where the results of previous work are given.

We see that Kirsch et al.using their Kathryn modem[29] obtained error rate results varying from 10^{-2} to 10^{-6}. However their recorded signal-to-noise ratios were in the region from 20dB to 50dB, i.e. some

---

*Some results involving this method of reducing the bit rate are given in Appendix 5.1. They were obtained using the improved demodulation system described in section 5.4.

20dB higher than those recorded for this work. The transmitted power can be assumed to have been significantly higher, and so no useful comparison is possible.

Chase's results[31] were of the order of 10^{-3} for the 2400b/s rate his Codem system transmitted. This on the face of it is considerably better than the results obtained in this work, since his transmitted power of 4 Kw was only some 6dB greater.

In general though the validity of such comparisons is questionable for reasons that, for example, the radio link parameters such as the gain of the aerials and their directivity may not have been specified and that atmospheric conditions may vary between the times of the particular trials.

We now sum up the main conclusions by answering those questions raised in the statement of the objectives in Chapter Two.

5.3   SUMMARY OF THE MAIN CONCLUSIONS

(1)   What is the best form of modulation for serial binary oata transmission

The conclusion of this work is that PSK is the best form of modulation.

(2)   Is coherent or incoherent demodulation better?

It was found that coherent demodulation gave the better performance because the demodulated pulse stream could be equalised. It was noted, however, that the asynchronous nature of incoherent detection meant that its actual demodulation process was not affected by the presence of multipath as was that of the PLL carrier extractor.

(3)   How does the error rate vary with bit rate?

The error rate was found to reduce by only two or three times when the bit rate was lowered from 3.6 Kb/s to 0.4 Kb/s. It was seen that coding would prove a superior method.

(4)   How useful is adaptive equalisation?

The use of equalisation was found to improve the error performance significantly. At 2.1 Kb/s the error rate for PSK was improved by a factor of 1.7.

However the techniques of serial data transmission described in this work should be regarded as being in their first stage of development. Whilst the conclusions expressed above are felt to be valid the error performances obtained can be improved upon by further work, and so ideas for this are now presented.

-324-

5.4   SOME IDEAS FOR FURTHER WORK

As mentioned these ideas involve development of the serial
transmission system described in Chapter Three. It is felt that
the most should be obtained from this design philosophy before
any attention is concentrated on other methods of digital-data
transmission over HF.

Therefore this section starts in 5.4.1 with a method being
outlined of overcoming the coherent demodulation problem detailed in
Chapters Four and Five. It is termed the "dual equaliser" since two
separate equalisers are required. In section 5.4.2 ways of improving
the performance of the basic equaliser are proposed. Finally, in
section 5.4.3, it suggested that alternative methods of varying the
bit rate should be tried. Reference is given to some results that
have been obtained using a channel simulator where some of these ideas
were implemented in the test system. They are detailed in Appendix 5.1.

Firstly we describe the dual equaliser.

5.4.1   The Dual Equaliser

This system is designed to overcome the problem   experienced in
multipath conditions with the coherent demodulator which was analysed
in section 4.4.4 of Chapter Four. To recap, it was seen that the phase
of the carrier from the signal multipath was not necessarily the same
as that of the resultant from the sum of all the multipaths, which was
the phase that the PLL locked onto. Whenever the phases differed by
a value approaching $\pi/2$ the resolved component of the signal with
respect to the PLL output became small leading to a high error rate.

The dual equaliser idea developed thus. The PLL output referred to above is the "in-phase" output. Whenever this is at $\pi/2$ to the signal carrier the resolved component of the signal with respect to the PLL "quadrature" output will be at a maximum. Hence a method of removing the impairment to the demodulator system is to operate an additional equaliser and threshold "slicer" on the component of the signal multipath that has been demodulated by the quadrature output of the PLL. Using the terms defined in section 4.4.4, this component will be given by:

$$f(t+t_1) \, A_1 \, \sin \left[\theta_1-(\theta_2+\theta_3)/2\right] \quad .$$

A comparator is then used to feed the input of the decision-maker to the output of whichever of the two equaliser and slicer systems is operating on the greatest resolved component.

A block diagram of the proposed dual-equaliser system is shown in Fig. 5-4.1. We see that the moduli of the pulse-amplitude level-detector outputs are compared for the purpose of deciding the input for the decision-maker. Fig. 5-4.2 shows the variation of the resolved component of the signal (viz. $A_1 f[t+t_1]$) for a complete cycle of $\left[\theta_1-(\theta_2+\theta_3)/2\right]$ from 0 to 2π. The variation for the single-equaliser system used in the trials is drawn for comparison. We see that for the dual equaliser the sliced component sampled by the decision-maker is never less than 0.707 of the optimum*.

---

*This dual system has been tested in comparison to the single-equaliser using a channel simulator, and some results are given in Appendix 5.1

The ability of the level detectors to track time variations in the amplitudes that they are measuring means that, in principle, the PLL could be replaced by a local oscillator. If this local oscillator and the receiver IF frequency could be adjusted (and remain stable) to within the bandwidth of the level detectors and the equaliser channel response estimators (i.e. 5 Hz as used for the test system in the trials) the whole system could follow the amplitude variations caused by the beating effect of the two frequencies. The comparator would ensure that the decision-maker input was always from the greatest resolved component. Such a dual equaliser system is somewhat similar to the pass-band type of equaliser described in references [51,52].

5.4.2 Improvements for the Equaliser

These must resolve around reducing the noise added by the equaliser from low-frequency noise components present in the impulse-response level detectors.

The first thing to be done must be to lower the bandwidth of the low-pass filters. It was noted in the observations (section 4.2.2 of Chapter Four) that the figure of 5 Hz was too high and that 3.5 Hz would be sufficient.

A second method would involve some form of tap thresholding. At the 3.6 Kb/s bit rate twenty equaliser taps are switched in. Since there were typically only two echo multipaths in the experiments only four of these taps would have been in effective operation. The remaining sixteen would simply have been amplifying the noise. A simple method of tap thresholding would be to compare the output of each equaliser level detector to a (to be determined) fraction of the pulse-amplitude level detector.

5.4.3   Improved Error Performance at the Lower Bit Rates

As already stated, reducing the bit rate by leaving blanks after each bit did not significantly improve the error rate. Whilst it is felt that the improvement would have been more significant if the dual equaliser type of demodulator was in operation* it would be instructive to obtain results using the following two methods in comparison:

(1) To widen the pulse keeping the bit- rate reduction factor constant at "x1". As mentioned previously this method has the advantage that the noise level can be reduced with bit rate because the bandwidth of the receiver filters can be lowered. However it leaves the pulses more vulnerable to distortion because the multipaths will no longer tend to be resolved as the bit rate is lowered*.

(2) To reduce the bit rate by coding. For example, a simple majority decision code was proposed in section 4.4.8 of Chapter Four and seen to be effective theoretically. It would be necessary for the implementation of such a coding scheme to establish the correlation between the separation of errors. Obviously if errors tended to occur in blocks then simply repeating the same bit n times in succession (for an n-bit code) would not result in the theoretical improvement in the error rate. It would be necessary to stagger the repetitions according to the correlation response.

---

*See Appendix 5.1 for some results obtained (with channel simulation) in comparison with pulse blanking using the dual equaliser.

APPENDIX 5.1

FURTHER RESULTS USING A CHANNEL SIMULATOR

The dual equaliser and the pulse-widening method of bit-rate variation have been tested in comparison to the single equaliser (as used in the trials) and to the pulse blanking method, using a channel simulator. The philosophy of the channel simulator was to re-create the conditions experienced in the trials. As such, a three-path multipath structure was realised, the strengths of each path being modulated by independent low-frequency "Rician" waveforms. The extent of the strength variations was some 10dB, each path having the same average amplitude. The two "echo" paths were slowly and independently time-swept in the range 0.35 mS to 3.5 mS after the "main" path. The level of the additive (Gaussian) noise was adjusted to give approximately the same error rates for the single equaliser as those obtained in the Wick trials detailed in Chapter Four. Because the noise was introduced before the simulator, and so was the result of "summing" the three multipaths, the variance of the noise changed continuously as was the case with the background noise in the trials.

Some results are given below. They are for the single equaliser with pulse-blanking bit-rate reduction; for the dual equaliser with pulse blanking; and for the dual equaliser with pulse widening. PSK was the modulation mode used.

APPENDIX 5.1 Continued

	Single equaliser	Dual equaliser	
Bit rate	Pulse blanking	Pulse blanking	Pulse widening
3.6 Kb/s	2.24×10^{-1}	1.8×10^{-1}	(1.8×10^{-1})
1.2 Kb/s	3.5×10^{-2}	2.5×10^{-3}	4.0×10^{-2}
0.45 Kb/s	1.6×10^{-2}	2.0×10^{-4}	2.5×10^{-2}

These results indicate the considerable improvement afforded by the dual equaliser, and also show that pulse blanking is indeed a superior method of decreasing the bit rate, as was anticipated in section 3.1.1 of Chapter Three.

Fig 5-4.1   The dual equaliser
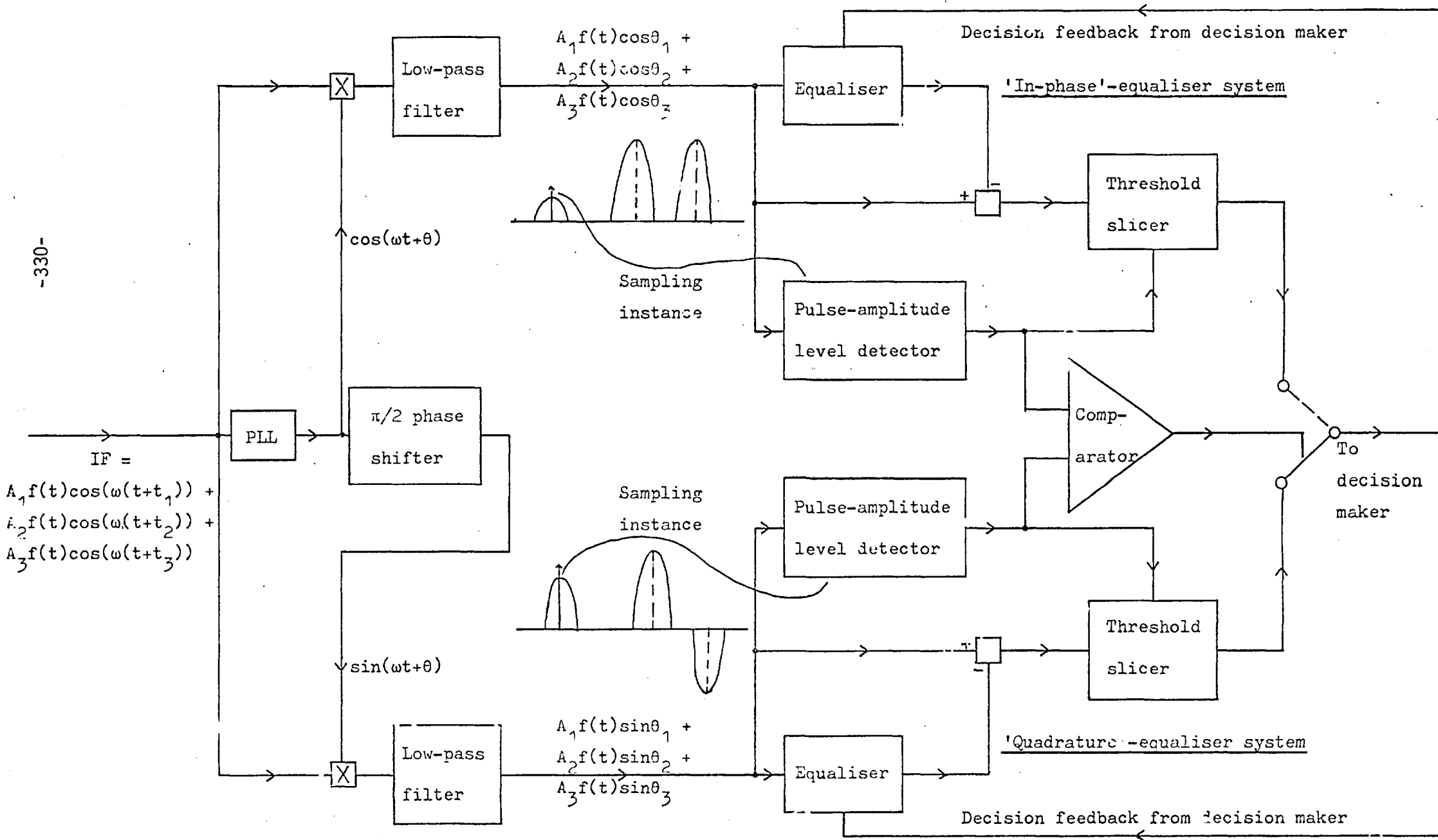
Fig 5-4.2   Variation in resolved signal component with relative phase
angle for dual- and single-equaliser systems

Dual-equaliser variation

Modulus of
resolved
component of
$A_1 f(t+t_1)$

| In-p. syst. | Quadrature system | In-phase system | Quadrature system | In-p. syst. |

Relative phase angle $(= \theta_1 - \frac{\theta_2+\theta_3}{2})$
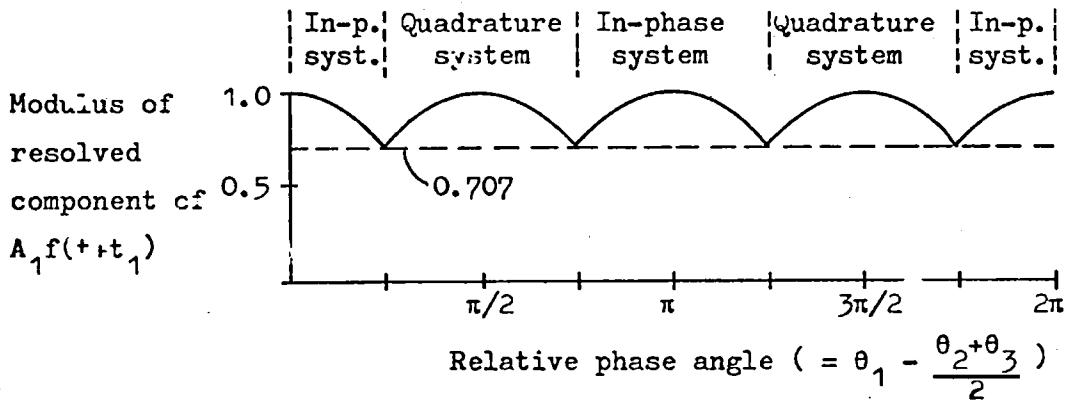
Single-equaliser variation

Modulus of
resolved
component of
$A_1 f(t+t_1)$

Relative phase angle $(= \theta_1 - \frac{\theta_2+\theta_3}{2})$
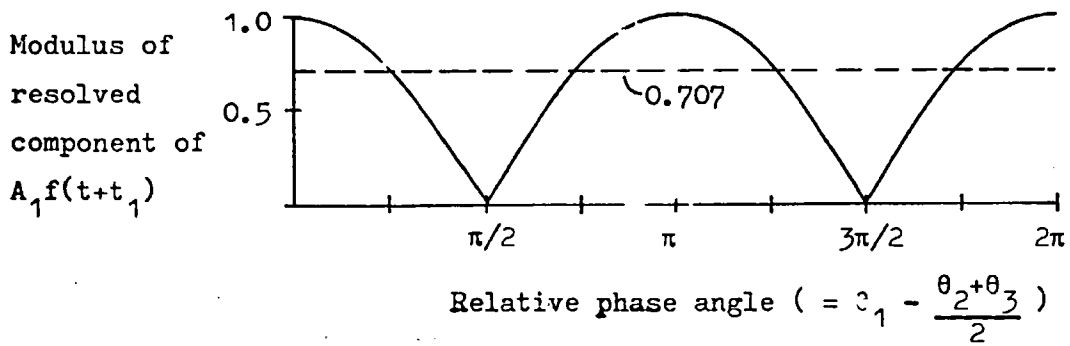
References

1.   A.H. Reeves, "The Single Sideband System Applied to Short
      Wavelengths", Elec. Comm., Vol. 10, No..1, pp.3-19, July 1931

2.   E.V. Appleton and R. Naismith, "Some Measurements of Upper-
      Atmospheric Ionisation", Proc. Royal Soc., Vol. 137, Series A,
      pp. 36-54, 1932

3.   S.K. Mitra, "The Upper Atmosphere", (The Royal Asiatic Society
      of Bengal)

4.   E.V. Appleton and R. Naismith, "Weekly Measurements of Upper-
      Atmospheric Ionisation", Proc. Phys. Soc., Vol. 45, Part 3,
      pp. 389-398, May 1933

5.   E.V. Appleton and R. Naismith, "Some Further Measurements of
      Upper-Atmospheric Ionisation", Proc. Royal Soc., Vol. 150,
      Series A, pp. 685-708, 1935

6.   C.E. Shannon, "A Mathematical Theory of Communication", BSTJ,
      Vol. 27, pp. 379-423,July 1948, and pp. 623-656, October 1948

7.   C.E. Shannon, "Communication in the Presence of Noise", Proc.
      IRE, Vol. 37, pp. 10-21, January 1949

8.   H.C.A. Van Duuren, "Typendruktelegrafie over Radioverkindingen
      (Tor)", Tijdschr. Nederlands Radiogenoot, Vol. 16, No. 2,
      pp. 53-67, March 1951

9.   R.E. Lacy, M. Acker, J.L. Glaser, "Performance of Space and
      Frequency Diversity Receiving Systems", IRE Nat. Conv. Record,
      Part II, pp. 148-152, 1953

10. R. Naismith, H.C. Bevan, P.A. Smith, "A New Method of Ionospheric Forecasting", Proc. IEE, Vol. 1C9B, No. 44, pp. 125-130, March 1962

11. J.A. Betts, "High Frequency Communication", (book) The English Universities Press Ltd., 1968

12. A. Picquenard, "Radio Wave Propagation", (book), The Macmillan Press Ltd., 1974

13. K. Davies, "Ionospheric Radio Propagation", United States Dep. of Commerce, National Bureau of Standards Monograph 80, 1965

14. B. Goldberg, "300KHz-30MHz MF/HF", IEEE Trans. on Comm. Tech., Vol. COM-14, No. 6, pp.767-784, December 1966

15. R.W.E. McNicol, "The Fading of Radio Waves at Medium and High Frequencies", Proc. IEE, Vol. 96, pt. IIi, pp. 517-524, Nov. 1949

16. S.O. Rice, "Mathematical Analysis of Random Noise", Bell System Tech. J., Vol. 23, pp. 282-333, July 1944, and Vol. 24, pp. 96-157, January 1945

17. H. Nyquist, "Certain Topics in Telegraph Transmission Theory", Trans. AIEE, Vol. 47, pp. 617-644, April 1928

18. W.R. Bennett and J.R. Davey, "Data Transmission", (book) McGraw-Hill Book Company, 1965

19. R.R. Mosier and R.G. Clabaugh, "Kineplex, A Bandwidth-Efficient Binary Transmission System", Trans. AIEE. (Part I: Communications and Electronics), Vol. 76, pp. 723-728, January 1958

20. "Kathryn HF Radio Teletype and Data System", Final Rept., Phase I, General Atronics Corp., Rept. 938-230-14(AD269032L), September 25, 1961

21. M.S. Zimmerman and A.L. Kirsch, "The AN/GSC-10 (KATHRYN) Variable Rate Data Modem for HF Radio", IEEE Trans. Comm. Tech., Vol. COM-15, pp.197-204, April 1967

22. P.A. Bello, "Selective Fading Limitations of the Kathryn Modem and Some System Design Considerations", IEEE Trans. Comm. Tech., Vol. COM-13, pp. 320-333, September 1965

23. G.C. Porter, "Error Distribution and Diversity Performance of a Frequency-Differential PSK HF Modem", IEEE Trans. Comm. Tech., Vol. COM-16, pp.567-575, August 1968

24. M.J. Di Toro, "Adapticom", Rec. 1st IEEE Ann. Comm. Conv. p.763, June 1965

25. M.M. Goutmann, "Intersymbol Interference as a Natural Code", IEEE Trans. Comm. Tech. (Corresp.), Vol. COM-20, pp. 1033-1037, October 1972

26. M.M. Goutmann and M.M. Gutman, "High Rate Serial Transmission Experiment Over the HF Channel", National Telecommunications Conf. 1974, pp. 750-754

27. P.G. Farrell and G. Andjargholi, "A Spread Spectrum Digital Transmission System for Reliable Communication in the High Frequency Band", IEE Colloquium on HF Communication Systems, 1976

28.   A.R. Schmidt, "A Frequency Stepping Scheme for Overcoming the Disastrous Effects of Multipath Distortion on High Frequency FSK Communication Circuits", IRE Trans. on Comm. Sys., pp. 44-47, March 1959

29.   A.L. Kirsch, P.R. Gray, and D.W. Hanna, Jr., "Field-Test Results of the AN/GSC-10 (KATHRYN) Digital Data Terminal", IEEE Trans. Comm. Tech., Vol. COM-17, pp. 118-128, April 1969

30.   P. McManamon, R.V. Janc, and S. Tsai, "Comparison of Diversity, Non-diversity, and Coding on a Parallel FSK HF Channel", IEEE Trans. Comm. Tech., Vol. COM-17, pp. 355-367, June 1969

31.   D. Chase, "A Combined Coding and Modulation Approach for Communication over Dispersive Channels", IEEE Trans. Comm., Vol. COM-21, pp. 159-174, March 1973

32.   M. Darnell, "Medium-Speed Digital Data Transmission over HF Channels", IEEE-IEE-IRE Conf. on Signal Processing, Loughborough, 1977

33.   M . Schwartz, W.R. Bennett, and S. Stein, "Communication Systems and Techniques", (book) McGraw-Hill Book Company, 1966

34.   K. Bullington, "Radio Propagation Fundamentals", Bell Syst. Tech. J., Vol. 36, pp. 593-626, May 1957

35.   P.F. Panter, "Modulation, Noise, and Spectral Analysis", (book) McGraw-Hill Book Company, 1965

36.   M. Schwartz, "Information, Transmission, Modulation   and Noise", (book) McGraw-Hill Book Company, 1970

37. H. Taub and D.L. Schilling, "Principles of Communication Systems", (book) McGraw-Hill Kogakusha Ltd., 1971

38. B.P. Lathi, "Communication Systems", (book) John Wiley and Sons, Inc., 1968

39. J.I. Marcum, "Tables of the 'Q' Function", Rand Corporation Memo RM 399, January 1950

40. J.P. Costas, "Synchronous Communications", Proc. IRE, vol. 44, pp. 1713-1748, December 1956

41. K.H. Mueller and M. Müller, "Timing Recovery in Digital Synchronous Data Receivers", IEEE Trans. Comm., Vol. COM-24, No. 5, pp. 516-531, May 1976

42. R.W. Lucky, "A Survey of the Communication Theory Literature: 1968-1973", IEEE Trans. Inf. Theory, Vol. IT-19, No. 5, pp.725-736, November 1973

43. J.L. Lawson and G.L. Uhlenbeck, "Threshold Signals", (book) McGraw-Hill Book Company, 1950

44. D. Hirsch and W.J. Wolf, "A Simple Adaptive Equaliser for Efficient Data Transmission", IEEE Trans. Comm. Tech., Vol. COM-18, pp.5-12, February 1970

45. C.W. Niessen and D.K. Willim, "Adaptive Equaliser for Pulse Transmission", IEEE Trans. Comm. Tech., Vol. COM-18, pp.377-395, August 1970

46. A. Lender, "Decision-Directed Digital Adaptive Equalisation Technique for High-Speed Data Transmission", IEEE Trans. Comm. Tech., Vol. COM-18, pp.625-632, October 1970

47. P. Monsen, "Feedback Equalisation for Fading Dispersive Channels", IEEE Trans. Inf. Theory, Vol. IT-17, pp.56-65, January 1971

48. R.R. Bowen, D.A. George, and J.R. Storey, "An Adaptive Decision-Feedback Equaliser", IEEE 1970 Annual Comm. Conf.

49. D.A. George, D.C. Coll, A.R. Kaye, and R.R. Bowen, "Channel Equalisation for Data Transmission", Journal of the Engineering Institute of Canada, May 1970

50. P. Monsen, "Digital Transmission Performance on Fading Dispersive Channels", IEEE Trans. Comm. Tech., Vol. COM-19, pp.33-39, January 1973

51. R.D. Gitlin, E.Y. Yo, and J.E. Mazo, "Passband Equalisation of Differentially Phase-Modulated Data Signals", Bell Sys. Tech. J., Vol. 52, No. 2, pp. 219-230, February 1973

52. D.D. Falconer, "Jointly Adaptive Equalisation and Carrier Recovery in Two-Dimensional Digital Communication Systems", Bell Sys. Tech. J., Vol. 55, pp. 317-334, March 1976

53. R.M.F. Goodman, and P.G. Farrell, "Data Transmission with Variable-Redundancy Error Control over a High-Frequency Channel", IEE Proc., Vol. 122, pp. 113-118, February 1975